



UNIVERSIDADE FEDERAL DE LAVRAS

**ESTUDO DA DIVERGÊNCIA GENÉTICA EM  
ACESSOS DE ARROZ ATRAVÉS DE  
MARCADORES MORFOLÓGICOS E  
MOLECULARES (RAPD)**

**ANTÔNIO TADEU DA SILVA**

1999

ANTÔNIO TADEU DA SILVA

ESTUDO DA DIVERGÊNCIA GENÉTICA EM ACESSOS DE ARROZ  
ATRAVÉS DE MARCADORES MORFOLÓGICOS E MOLECULARES  
(RAPD)

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do programa de Pós-graduação em Agronomia, área de concentração em Fitotecnia, para a obtenção do título de Doutor.

Orientador  
Prof. Dr. Moacir Pasqual

LAVRAS  
MINAS GERAIS - BRASIL  
1999

Ficha Catalográfica preparada pela Secção de Classificação e Catalogação  
da Biblioteca Central da UFLA.

Silva, Antônio Tadeu.

Estudo da divergência genética em acessos de arroz através de  
marcadores morfológicos e moleculares (RAPD)/ Antônio Tadeu da Silva.

– Lavras: UFLA, 1999.

185p. : il.

Orientador: Moacir Pasqual.

Tese de Doutorado - UFLA.

Bibliografia.

1. Arroz de Sequeiro - 2. Divergência genética 3. Correlação 4.  
Preservação 5. Banco de germoplasma 6. Marcadores moleculares  
(RAPD) e morfológicos 7 Cereal.

I. Universidade Federal de Lavras. II. Título.

CDD-633.1823

**ANTÔNIO TADEU DA SILVA**

**ESTUDO DA DIVERGÊNCIA GENÉTICA EM ACESSOS  
DE ARROZ ATRAVÉS DE MARCADORES  
MORFOLÓGICOS E MOLECULARES (RAPD)**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Agronomia, área de concentração em Fitotecnia, para a obtenção do título de Doutor.

Aprovada em 28 de Outubro de 1999

**Dr. Augusto Ramalho de Moraes**

**DEX/UFLA**

**Dr. Jaime Roberto Fonseca**

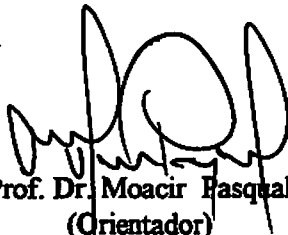
**Embrapa Arroz e Feijão**

**Dr. Rui Américo Mendes**

**Embrapa Recursos  
Genéticos e Biotecnologia**

**Dr. Wagner Pereira Reis**

**DAG/UFLA**



**Prof. Dr. Moacir Pasqual  
(Orientador)  
DAG/UFLA**

**Lavras  
MINAS GERAIS - BRASIL**

ESSES, QUE APESAR DE ANÔNIMOS  
AINDA FAZEM E ACEITAM CRÍTICAS

OFEREÇO ...

ÀQUELES QUE  
CONHECEM E VIVEM MINHA TRAJETÓRIA

DEDICO...

AO CRÉADOR.

SINCEROS AGRADECIMENTOS,  
PELA PASSAGEM NESTE PLANO E  
TER CONSEGUIDO O QUE MAIS SONHAVA  
*"CONSEGUIR, ENTENDER, O QUE É O SABER."*

NÃO QUERO CITAR NENHUM NOME,

POIS FORAM TANTOS QUE CONTRIBUÍRAM PARA A REALIZAÇÃO DISSO.

NO ENTANTO, AQUELES VOLUNTÁRIOS PERMANECERÃO

COMO MEUS AMIGOS

#####

## RECONHECIMENTOS

À Universidade Federal de Lavras, em todo seu contexto,  
especialmente o Departamento de Agricultura/ setores de  
Sementes e Cultura de Tecidos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível  
Superior (CAPES)

À Embrapa por sua unidades/Arroz e Feijão, bem como  
Recursos Genéticos e Biotecnologia.

O autor reconhece que na ausência destas Instituições,  
nada disto seria possível, e mais, não existiria treinamento  
ou instrução alguma.

## SUMÁRIO

	Página
LISTA DE SIGLAS, ABREVIATURAS E SÍMBOLOS	v
RESUMO	vi
ABSTRACT	vii
1 - CAPÍTULO 1	01
1.1 - INTRODUÇÃO GERAL	01
1.2 - OBJETIVOS GERAIS	03
1.3 - REFERENCIAL TEÓRICO	03
1.3.1 - Importância Social do Arroz	03
1.3.2 - Origem e Evolução das Espécies de <i>Oryzae</i>	04
1.3.2.1 - Classificação Botânica e Citogenética das Espécies Cultivadas de Arroz	05
1.3.2.2 - Características Morfo-Agronômicas da Planta de Arroz	08
1.3.3 - Histórico do Cultivo de Arroz no Brasil	12
1.3.3.1 - Sistemas de Cultivo das Lavouras de Arroz	13
1.3.4 - Interação com o Ambiente	15
1.3.5. - Base Genética e Ganho das Culturas	16
1.3.5.1 - Variabilidade Genética ou Diversidade do Arroz	20
1.3.5.2 - Fontes de Novos Reservatórios de Genes	23
1.3.5.3 - Conservação e Utilização do Germoplasma	25
1.3.5.3.1- Importância das Coletas na Conservação de Germoplasma	26
1.3.6 - Caracterização dos Acessos através de Marcadores Morfo-Agronômicos	27
1.3.6.1 - Caracterização através de Marcadores Genéticos	29
1.3.6.2 - Da Técnica de PCR aos Marcadores RAPD	34
1.3.6.3 - Tipos de Marcadores a Nível de DNA	39
1.3.7 - Banco Ativo de Germoplasma (BAG)	42
1.3.7.1 - Descarte de Acessos-duplicatas ou Duplicados	43
1.3.7.2 - Biometria Aplicada no Descarte de Acessos-duplicatas	44
1.3.7.3 - Emprego da Divergência Genética Correlações e Heterose	46
1.3.8 - Distância Genética Baseada na Distancia Estatística	49
1.3.8.1 - Distância Generalizada de Mahalanobis	50
1.3.8.2 - Distância Euclidiana	53
1.3.9 - Técnicas Multivariadas no Estudo da Divergência Genética	55
1.3.10 - Divergência Genética por Variáveis Canônicas	56
1.3.11 - Importância Relativa dos Caracteres	59
1.3.12 - Métodos de Agrupamentos dos acessos	61
1.3.12.1 - Método do Vizinho mais Próximo (VMP)	64
1.3.12.2 - Método de Agrupamento de Tocher	65

1.3.12.3 - Outros métodos de Agrupamento	67
1.4 - REFERENCIAS BIBLIOGRÁFICAS	69
2 - CAPITULO 2	77
EMPREGO DE MARCADORES MORFOLÓGICOS NO ESTUDO DA DIVERGÊNCIA GENÉTICA EM ACESSOS DE ARROZ DO GRUPO TRÊS MESES	77
RESUMO	77
ABSTRACT	78
2.1 - INTRODUÇÃO	79
2.2 - MATERIAL E MÉTODOS	82
2.2.1 - Localização, Instalação e Condução do Expirimento	82
2.2.2 - Análise Estatística dos Dados	86
2.3 - RESULTADOS E DISCUSSÃO	90
2.3.1 - Análise Preliminar dos Caracteres Qualitativos	90
2.3.1.1 - Análise Univariada dos Caracteres Qualitativos	92
2.3.1.2 - Avaliação da Divergência Genética entre os Acessos por Técnicas Multivariadas	95
2.3.1.3 - Emprego das Variáveis Canônicas (vc's)	97
2.3.1.4 - Análise de agrupamentos	101
2.3.1.5 - Importância entre as Variáveis Originais	107
2.4 - CONCLUSÕES	110
2.5 - REFERÊNCIAS BIBLIOGRÁFICAS	111
3 - CAPÍTULO 3	116
USO DE MARCADORES MOLECULARES (RAPD) NO ESTUDO DA DIVERGÊNCIA GENÉTICA EM ACESSOS DE ARROZ	116
RESUMO	116
ABSTRACT	117
3.1 - INTRODUÇÃO	118
3.2 - MATERIAL E MÉTODOS	122
3.2.1. Preparo do Material Vegetal para Extração de DNA Total	122
3.2.2 - Teste em Eletroforese de Gel de agarose e Densidade do DNA Extraído de Plantas Jovens	124
3.2.3. - Processo de Amplificação dos Fragmentos de DNA e Eletroforese em Gel	124
3.2.4 - Tabulação e Análise Estatísticas dos Marcadores RAPD	126
3.2.5 - Estimativas das Distâncias Genéticas pelo Complemento Aritmético de Jaccard (Cii')	127
3.2.5.1 - Análise de Agrupamentos pelo Método das Ligações Médias entre Pares não Ponderados (UPGMA)	128
3.3 - RESULTADOS E DISCUSSÃO	129
3.3.1 - Classificação e Identificação dos Marcaores de RAPD	132
3.3.2 - Avaliação de Divergência Genética entre os Acessos de Arroz	136



3.3.3 - Caracterização dos Acessos e Agrupamentos	138
3.3.3.1 - Estabelecimento do Grau de Parentesco	141
3.4 - CONCLUSÕES	144
3.5 - REFERÊNCIAS BIBLIOGRÁFICAS	145
4 - CAPITULO 4	149
COMPARAÇÃO DE DIVERGÊNCIA GENÉTICA EM ARROZ AVALIADA ATRAVÉS DE MARCADORES MORFOLÓGICOS E MOLECULARES (RAPD)	149
RESUMO	149
ABSTRACT	150
4.1 - INTRODUÇÃO	151
4.2 - MATERIAL E MÉTODOS	154
4.2.1 - Obtenção dos Acessos para Análise	154
4.2.2.- Aspectos Gerais sobre a Coleta de Dados Morfo- Agronômicos Qualitativos	155
4.2.3 - Análise Estatística dos Dados	156
4.2.4 - Obtenção e Análise de DNA por RAPD	157
4.2.5 - Estimativas da Divergência Genética por Meio de Marcadores Quantitativos	157
4.3 - RESULTADOS E DISCUSSÃO	158
4.3.1 - Análise Multivariada dos Dados Binários	159
4.3.2 - Distância Genéticas por Dados Morfo-Agronômicos	159
4.3.3 - Agrupamento Pelo Método "UPGMA"	160
4.3.4 - Caracterização Qualitativa por Marcadores Morfo- Agronômicos	162
4.3.5 - Comparação dos Grupos Formados entre os Métodos	165
4.3.6 - Discriminação entre Grupos de Germoplasmas	167
4.3.7 - Comparação da Divergência Genética entre os Marcadores Morfo-Agronômicos e Marcadores RAPD	170
4.4 - CONCLUSÕES	179
4.5 - REFERÊNCIAS BIBLIOGRÁFICAS	180

## LISTA DE ABREVIações E SÍMBOLOS

ANAVA	Análise de variância univariada
ANOVA	Análise de variância multivariada
AP-PCR	Reação de fragmentos de DNA em cadeias
BAG	Banco ativo de germoplasma
BSA	Albumina bovina solúvel
C + G	Citocinina mais guanina
CIAT	Centro Internacional de Agricultura Tropical
CNA	Centro Nacional do Arroz
CNPAF	Centro Nacional de Pesquisa de Arroz e Feijão
CTAB	Tampão para extração de DNA
cv' e cv's	Cultivar e cultivares
CV - CVg CVe	Coefficiente de variação e coef. Variação genético e ambiente
D <sup>2</sup> i'	Distância generalizada de Mahalanobis
DAG	Departamento de agricultura UFLA
DCA	Departamento de Ciências dos Alimentos - UFLA
dCTT	Dinucleotídio trifosfato de timina
DNA	Ácido desoxirribonucléico
dPTA	Dinucleotídio trifosfato de adenina
dPTC	Dinucleotídio trifosfato de citosina
dPTG	Dinucleotídio trifosfato de guanina
EEA	Estação Experimental do Arroz
hab.	Habitante
I	Relação entre coeficiente de variação genético e ambiental
ICA	Instituto Colombiano de Agropecuária
IRA	Idem IRRI
IRGA	Instituto Riograndense do Arroz
IRRI	Instituto Internacional de Pesquisa do Arroz - Filipinas
pb	pares de bases
PCR	idem AP-PCR
QTL	Marcador molecular de locus quantitativos
RAPD	reação de fragmentos polimórficos de DNA em cadeia
Taq-polimerase	enzima extraída da bactéria
TBE	tampão de purificação de DNA
UPGMA	método de ligação não ponderada pela média aritmética entre pares
vc' e vc's	variável canônica e variáveis canônicas

## RESUMO

SILVA, Antônio Tadeu da. **Estudo da Divergência Genética em Arroz Através de Marcadores Morfológicos e Moleculares (RAPD)**. Lavras: UFLA, 1999. 185p. ( Tese - Doutorado em Fitotecnia).<sup>12</sup>

Alguns acessos de arroz no BAG não passam de simples duplicatas, sendo catalogados como novos acessos apenas pela data e local de coletas. Em razão de reduzir custos , número e espaço para flexibilizar os materiais mais promissores foi conduzido este estudo com 36 acessos de arroz de nomes regionais 'três meses' (CV's Obsoletas) coletados pelo Território Brasileiro, para avaliar a divergência genética por meio de marcadores morfológicos qualitativos e quantitativos, bem como moleculares (RAPD). Os marcadores morfológicos foram coletados em parcela útil a campo de multiplicação de sementes dos acessos de arroz do CNPAF, Goiânia, GO. Para extração do DNA dos mesmos acessos, o plantio foi em casa de vegetação, em vasos de 500 g de substrato. A extração, reação de amplificação e corrida em gel seguiu o protocolo de Doyle e Doyle (1990). Após as análises dos dados procedeu-se a uma comparação dos grupos através da aglomeração com base na distância de Mahalanobis e Índice de Jaccard, as quais mostraram existir ampla distância genética entre os acessos, embora sejam estes, um conjunto de cultivares que já recebeu grande pressão de seleção pelos agricultores. Não foi observada grande associação entre os acessos e a procedência geográfica, nem com os nomes comuns. Qualquer dos métodos mostrou-se eficiente na detecção da divergência genética neste conjunto de acessos de arroz denominados 'três meses'.

**Palavras Chaves:** Divergência genética, marcador, acessos, arroz, *Oryza sativa* e germoplasma.

---

1 Comitê de Orientação: Moacir Pasqual - UFLA. (Orientador), Augusto Ramalho de Moraes- UFLA

## ABSTRACT

**SILVA, Antônio Tadeu da. Study of the genetic divergence in rice accessions across morphological and molecular (RAPD) markers. Lavras: UFLA, 1999. 185p. (Thesis - Doctorate in Agronomy / Plant Science).<sup>\*3</sup>**

Some rice accessions in BAG were simple duplicate-accessions, their being labeled as new accessions only by the date and location of harvest. Owing to reducing costs, number and space to become flexible the most promising materials, this study was conducted on 36 rice accessions the name 'three month' ( CV's Obsoletas) collected throughout the Brazilian territory, to evaluate the genetic divergence by means of qualitative and quantitative morphological markers as well as molecular. ( RAPD). The morphological markers were collected in useful plot in a seed multiplication field of the rice accessions of CNPAF/Embrapa, Goiania, GO. The DNA-extraction in even across the plants were grow in greenhouse in pot 500 g of substrate. The DNA extraction of were tissue liofilized, amplification reaction and gel run was otimization, conform Doyle e Doyle (1990) protocols. After the analyses of the data, a comparison of the cluster formed through the clustering on the basis of the Mahalanobis distance and Jaccard Index which showed to exist a wide genetic distance among the accession , although these are a set of cultivars which has already given a great selection pressure by farmers, No great association was observed among th e accessions and the geographical origin nor with the common names . Any of the methods proved effective for the detection of the genetic divergence in these rice accessions 'three month'

**Index terms:** Genetic divergence, accessions, markers, rice, *Oryza sativa* and germoplasm.

---

1 Guidance Committee: Moacir Pasqual - UFLA (Major Professosr) , Augusto Ramalho de Moraes - UFLA.

# CAPITULO 1

## 1.1 - INTRODUÇÃO

Dentre as culturas alimentares o arroz (*Oryza sativa* L.) é o alimento básico para mais de 40% da humanidade, com produção mundial de 525,5 milhões de t; é a maior produção dentre os cereais, provendo de 12 a 35% de calorias e de 9 a 20% de proteínas totais consumidas, respectivamente, tanto para os asiáticos, como para os latino-americanos. O cultivo ainda ocupa posição de destaque entre os cereais cultivados, sendo superado somente pelo trigo, em área cultivada, e pelo milho em produção de grãos (FAO, 1998).

A China e a Índia ocupam posições de destaque na produção mundial de arroz, enquanto o Brasil é modesto com apenas 2 - 2,5% do total produzido, embora o País tenha uma produção elevada, pelo fato de produzir 10,951 milhões de t do produto em casca no ano agrícola de 1998, numa área de 3,931 milhões de ha (Agroanalysis, 1998). O mercado mundial é pouco dinâmico, pois anualmente, apenas 12 milhões de t são movimentados, devido ao elevado consumo por parte dos países que o produzem. A Tailândia é um modesto produtor ao lado dos países asiáticos, no entanto é o principal exportador no mundo, de tal modo que as cotações de preços, advêm da bolsa de valores que opera em maior proporção apoiada nas cotações do arroz Tailandês.

Apesar da introdução do arroz no Brasil ter sido realizada pelos portugueses no início do Séc. XVIII (1745) no litoral do Pará e no Maranhão, somente em 1792 tornou-se cultura comercial; a partir de 1908 passou a ser cultivado com bases científicas.

Iniciaram-se no Rio Grande do Sul os plantios em sistemas irrigados, adotando-se cultivares americanas e italianas até que, por volta de 1918, foi fundada a Estação Experimental do Arroz (EEA) onde se realizavam os trabalhos de seleção nas introduções (Pedroso, 1982). Os espanhóis também

colaboraram nessa tarefa, uma vez que trouxeram o arroz para a América Andina, fazendo com que chegasse até ao Brasil pelo lado Oeste. A cultura encontra-se hoje difundida pelo Território Nacional, em sistemas de irrigado e terras altas, sendo cultivada por inúmeros produtores e estabelecimentos agrícolas; somente no MA, GO, MT e MS mais de um milhão de famílias o plantam em área de 912, 4 mil ha (Abbud, 1991; Zhu et al, 1998).

Lopes (1984) comenta que o arroz, como qualquer planta, tem seu crescimento e desenvolvimento influenciados tanto por fatores genéticos como por fatores ambientais, que são responsáveis pela determinação dos caracteres vegetativos e reprodutivos. Dentre os fatores ambientais, destacam-se como barreiras externas mais importantes a competição imposta pelos nutrientes do solo, respostas às condições climáticas e também por plantas daninhas, fazendo com que apareçam as variedades dentro da espécie.

Com a preocupação em conservar, utilizar e fazer a manutenção dos recursos genéticos de arroz foram criados os Bancos Ativos de Germoplasma (BAG's), com a missão de impedir a erosão genética nesse germoplasma, com o resgate por meio de coletas e a incorporação dos acessos da espécie. Atualmente, na avaliação de genótipos podem ser usadas técnicas estatísticas e computacionais, que permitem avaliar várias combinações de modelos, podendo suas hipóteses serem testadas, comprovando-as ou contestando-as, gerando com isso um caminho novo para a pesquisa, que pode avaliar com bastante segurança, as relações de similaridade ou dissimilaridade existentes entre os acessos de qualquer espécie.

Na caracterização, descarte das duplicatas e seleção de acessos ou genótipos é importante ter conhecimentos sobre a natureza da variabilidade genética ou grau de parentesco existente na população, a associação entre os diferentes caracteres quantitativos e a extensão da influência do ambiente sobre esses caracteres (Tingey e Del Tufo, 1992).

## **1.2 - OBJETIVOS GERAIS**

- 1- Avaliar o grau de similaridade dos acessos de arroz coletados em diversos Estados do País, tidos como CV's Obsoletas e denominadas três meses, possuindo homonímia.**
- 2- Adotar os marcadores morfológicos quantitativos para estabelecer a divergência genética entre os pares de acessos, com base na distância generalizada de Mahalanobis.**
- 3- Utilizar de marcadores qualitativos e moleculares para estudo da divergência genética entre os acessos de arroz, baseados no índice de Jaccard e complemento aritmético.**
- 4- Comparar os grupos de acessos encontrados e relacioná-los com a divergência geográfica e parentesco.**
- 5- Avaliar a contribuição de alguns caracteres morfo-agronômicos na quantificação da variabilidade genética entre acessos.**

## **1.3 - REFERENCIAL TEÓRICO**

### **1.3.1 - Importância Social do Arroz**

No Brasil o arroz é parte essencial da dieta básica da população, contribuindo com cerca de 24,2% do total de calorias e 17,9% de proteínas, sendo consumido por todas as camadas sociais. As regiões Sul e Sudeste respondem por 75% do consumo de arroz, preferem-no produzido em sistema irrigado, por ser um produto de melhor qualidade, fazendo com que sobre nos estoques o arroz colhido em terras altas (Agroanalysis, 1998; Ferreira, 1995).

O cultivo se faz em todo Território Nacional; em sistema de terras altas, terras altas favorecido (várzea) e irrigado, numa área de 3,843 milhões ha, com produtividade média em torno de 2,656 t/ha (Anuário...1998). Todavia, constata-se que a produção brasileira, às vezes é insuficiente para suprir a demanda, provocando a importação de 1,18 milhões t do produto,

principalmente do Uruguai e Argentina para formar o estoque regulador, no entanto, tem-se notado uma estabilidade no consumo, conseqüência de redução no consumo "per capita", que passou de 97 kg/hab/ano de arroz em casca no ano 1965 para cerca de 45 kg/hab/ano. Hoje o arroz compõe uma proporção de 6,8 a 12% da cesta básica (FAO; 1998; Rocha, 1987).

### 1.3.2 - Origem e Evolução das Espécies de *Oryzae*

Admite-se que as espécies do gênero *Oryzae* tenham surgido de dois Centros de origem e domesticação, sendo o principal e mais extenso o Sudoeste Asiático, indo de Norte de Bangladesh até a Tailândia, onde a Índia é preconizada como o primeiro País a cultivá-lo, devido à abundância de espécies selvagens encontradas. Provavelmente há mais de 10 mil anos, aparece a África como centro secundário (Vaughan, 1989; de Datta, 1981).

A divisão atual do gênero *Oryzae* segue a mais recente feita por Vaughan (1989). Os detalhes da caracterização que é realizada através das espiguetas são sempre discutido, com algumas controvérsias, tendo algumas espécies primitivas, em restrita distribuição geográfica, geralmente isoladas. Ao passo que outras, mais evoluídas, se agrupam em complexos de espécies que parecem ter sofrido especialização mais recente, apresentando evidências de relação evolutiva estreita.

A julgar pela morfologia, citogenética, cruzabilidade e genética molecular, aceitam-se duas espécies recentemente descritas, *Oryza indandamanica* e *O. rhyzomatis*, que ainda não foram agrupadas oficialmente ao gênero (Oliveira, 1992). Todavia, Vaughan (1989) comenta que os híbridos entre variedades domesticadas e raças selvagens, ou como as raças invasoras de *O. nivara* e *rufipogon* (denominado grupo de Arroz Vermelho, devido a cariópse vermelha), adaptam-se às condições de cultivo, infestam todas as lavouras rizícolas, causando prejuízo às comerciais.

Deve-se salientar que o arroz cultivado no Estado do MA, de cascas



douradas, também denominado de 'arroz-vermelho' não é o mesmo arroz que infestam as lavouras cultivadas no Sul e Sudeste.

Do ponto de vista biológico, *O. nivara* e *O. rufipogon* e todos seus híbridos pertencem ao conjunto gênico primário de *O. sativa*, o que traz discussão sobre a presença de arroz no Brasil antes do desenvolvimento, ou seja, arroz "nativo". No continente Americano, onde a cultura de *O. sativa* não é vizinha das populações de grupos selvagens do arroz, o Vermelho é originário apenas de raças mutantes da forma cultivada.

### 1.3.2.1 - Classificação Botânica e Citogenética das Espécies Cultivadas de Arroz

O gênero *Oryza* pertence à Família Poaceae (gramineae), Subfamília Oryzoideae, possui espécies nas regiões tropicais e temperadas (Vaughan, 1989). As espécies que possuem seus genomas diplóides ( $2n=2x=24$ ) são as que mais intercruzam e geram o arroz comercial, enquanto que nas tetraplóides ( $2n = 4x=48$ ) está a maioria dos cruzamentos que geram os híbridos estéreis. Os cruzamentos interespecíficos entre as espécies de *O. sativa* L. e *O. glaberrima* Steud, mostram pequenas diferenças, porém, com muitos materiais que são estéreis (Camargo 1982; Chang, 1976).

O arroz cultivado é uma gramínea (poaceae) anual, de reprodução sexuada, com número de cromossomos  $2n = 24$  (nas diplóides), grupo das autógamas, normalmente a taxa de fecundação cruzada é menor que 5,0% (de Datta, 1989).

O complexo *sativa*: segundo Vaughan (1989), esse é o grupo taxonomicamente confuso do gênero, composto por oito espécies diplóides, com genoma AA (Tabela 1), incluindo as duas espécies cultivadas (*O. sativa*, *O. glaberrima*). Todas as espécies do complexo se intercruzam, apresentando diferentes graus de isolamento reprodutivo (raramente se completam). Os cromossomos se pareiam razoavelmente na meiose do híbrido, mas a homologia

não é completa, sendo o grupo genômico AA subdivido nos grupos AA, AgAg, AIAI, AgpAgp, AmAm.

TABELA 1 - Espécies mais estudadas no complexo dentro do gênero *Oryzae* Lineau, com número 2n e formulas genômicas (Vaughan, 1989).

Complexo	Espécie	2n	formula genômica
<i>Officinalis</i>	<i>O. officinalis</i> Wall.	24	CC
	<i>O. minuta</i> Presl.	48	BBCC
	<i>O. eichingeri</i> A. Peter	24	CC
	<i>O. punctata</i> Kotschy	24/48	BB/BC
	<i>O. latifolia</i> Desv.	48	CCDD
	<i>O. alta</i> Swallen	48	CCDD
	<i>O. Glandiglumis</i> (Doell) Prod.	48	CCDD
	<i>O. australiensis</i> Domin.	24	EE
<i>Sativa</i>	<i>O. sativa</i> L.	24	AA
	<i>O. nivara</i>	24	AA
	<i>O. rufipogon</i> Griff.	24	AA
	<i>O. glaberrima</i> Steud.	24	A <sup>9</sup> A <sup>9</sup>
	<i>O. barthii</i> A. Chev.	24	A <sup>9</sup> A <sup>9</sup>
	<i>O. longistaminata</i> A.Chev. et Roehr	24	A <sup>1</sup> A <sup>1</sup>
	<i>O. glumaepatula</i>	24	A <sup>9p</sup> A <sup>9p</sup>
	<i>O. meridionalis</i>	24	A <sup>m</sup> A <sup>m</sup>

*Oryza sativa* foi domesticada na Ásia e espalhou-se pelo mundo, principalmente pelos trópicos, tomando-se o principal cereal do terceiro mundo; *O. glaberrima* Steud foi domesticada na África, no Oeste da região abaixo do Saara, e sua área de cultivo restringe-se a essa região; que vem sendo substituída.

As espécies selvagens dos grupos AA (*O. nivara*, *O. rufipogon*, *O. longistaminata*, *O. glumaepatula*, *O. meridionalis*) são as mais aparentadas com *O. sativa*, e consideradas por alguns autores como um só grupo, do complexo *O. perennis* Moench. Esse grupo é composto por raças geográficas continentais, respectivamente asiática, africana, sul-americana e australiana. Embora usem formalmente os nomes dados por Vaughan (1989), biologicamente é uma espécie pantropical em vias de especiação, devido ao isolamento geográfico e as barreiras reprodutivas entre si; porém, os

taxonomistas consideram o nome *O. perennis* desaconselhável, para qualquer das espécies ou grupo.

Algumas espécies tidas como correlatas se inter cruzam e produzem sementes férteis, cujos híbridos geram mais variantes genéticos dentro da população de *O. sativa*, considerados fontes de novos genes. Segundo Camargo (1982), tais espécies são: *rufipogon*, *glaberrima*, *longistaminata*, *gandighumis*.

O arroz possui muitos parentes silvestres, a exemplo das quatro espécies que habitam a América Latina: *O. glumaepatula* (*O. perennis*), *O. alta*, *O. latifolia* e *O. grandighumis*, as quais não sofreram nenhuma domesticação, mas oferecem vários aspectos interessantes para estudos, além de constituírem um reservatório gênico a ser explorado no melhoramento do arroz. A *O. glumaepatula* (única diplóide) apresenta cruzamento natural com *O. sativa*. Estas quatro espécies são tidas como excelentes forrageiras, especialmente para ambientes alagadiços (Oliveira, 1992).

O gênero *Oryza* é composto de 28 espécies no geral, mas 22 são coerentes e as outras são tidas como variabilidade intraespecífica. No entanto, apenas duas espécies diplóides ( $2n = 2x = 24$ ) são cultivadas: *O. sativa* L. e *O. glaberrima* Steud., tal que, a segunda que é originária e cultivada na África está sendo substituída gradualmente pela *O. sativa*; esta é subdividida em duas subespécies, ou raças geográficas que são cultivadas: Indica e Japônica; e uma que não é cultivada, Javânica ou Bulú, tem aplicação para pesquisas (Yoshida, 1981).

As CV's tradicionais provenientes do tipo Indica são as mais cultivadas nos trópicos, com as principais características: porte alto, vigorosas, muitos perfilhos, folhosas, de coloração mais clara e folhas largas, leve pubescência ou glabras, colmo fraco com tendência ao acamamento, grãos mais longos, estreitos, translúcidos, que na cocção formam "grãos soltos". Na maioria das vezes não apresenta aristas e degrana facilmente, sensibilidade variável ao

fotoperíodo.

As do tipo Japônica são cultivadas nas zonas temperadas e apresentam porte menor, produtivas, maior resposta a N, folhas estreitas, verde-escuras e pilosas, grãos mais arredondados (cariópse arredondada), que se aglutinam fácil na cocção. Oliveira (1992) salienta que o formato de grãos que confere qualidade às cultivares brasileiras e americanas é do tipo 'agulhinha' (grão longo-fino) o que é o oposto do padrão de grãos adotados nas orientais, sendo geralmente provenientes da Japônicas.

A subespécie denominada Javânica ou Bulu (ocorre em zonas intermediárias, em relação as outras duas) apresenta: folhas largas e erectas, coloração verde-clara, cariópse longa e espessa, baixo perfilhamento, porte alto, lema e pálea com pêlos longos, resiste ao acamamento e pouco sensível ao fotoperíodo (Chang e Bandenas, 1965). Sabe-se que as variedades brasileiras foram confundidas com as do grupo Índica, por alguns autores, mas estudos recentes evidenciam que são do grupo Japônica (Liang et al., 1994).

### 1.3.2.2 - Características Morfo-Agronômicas da Planta de Arroz

O arroz é uma planta hidrófila, pertencente à família das poaceae (gramíneas), única cultura que sobrevive em alagados, tem hábito terrestre ou aquático, habitando florestas úmidas e regiões alagadas dos trópicos e subtropicais, que se reproduz tanto por via sexuada e seminal, quanto vegetativa, pelos caules decumbentes. Consegue bem veicular suas sementes pela água, com a propriedade da dormência, isto é, aguarda condições adequadas para germinação (Fornasieri Filho e Fornasieri, 1993).

Sistema radicular: o arroz possui sistema radicular do tipo fasciculado, com 60% ocupando a superfície do solo, possui raízes seminais ou nodais frouxas e as adventícias. O número aumenta no sistema radicular do início de crescimento até o florescimento, em seguida diminui até a maturação. As

cultivares precoces têm menos sistema radicular que as tardias.

As raízes da planta de arroz são providas de cavidades aeríferas no parênquima cortical que vão da coroa à extremidade, fazendo com que o oxigênio absorvido pela parte aérea e pela coroa das raízes na forma gasosa ou dissolvido, seja transportado às extremidades mais finas das raízes, através dessas cavidades existentes. O arroz é, assim, anatomicamente adaptado ao crescimento em solos submersos.

Parte aérea da planta: o caule é constituído por um colmo redondo, principal mais numero variável de secundários e terciários. Cada colmo possui regiões ocas (entrenós) e maciças (nós). Os entrenós inferiores são curtos, e possuem gemas de onde germinam os perfilhos primários e destes os secundários e assim por diante, ao passo que os superiores são longos.

O comprimento do caule oscila entre 40 cm (anãs) a mais de 700 cm nos sistemas flutuantes, com estrutura frouxa a cespitosa e até rizomatosos. O ultimo entrenó se liga a base da panícula. Os entrenós variam quanto a sua estrutura, tendo nos mais baixos maior diâmetro, enquanto que nos superiores tem maior espessura.

Folhas: sésseis, de limbo foliar linear, plano, longo a estreito, lanceolado, achatado, de margens lisas ou pubescentes (escábricas). Lígulas e aurículas membranosas a escariosas exsertas, se distribuem alternadas ao longo do caule, com bainhas das folhas normalmente involutas. Aquela, imediatamente abaixo da panícula, denomina-se folha 'bandeira', que difere das demais pelo ângulo de inserção no colmo. Normalmente as folhas possuem: limbo foliar, lígulas membranosas e escariosas exsertas, aurículas e bainhas involutas no colmo.

Órgão Floral: parte reprodutiva - a inflorescência do arroz é localizada na extremidade do colmo (terminal) em forma de panícula, e é subtendida pela folha-bandeira. A panícula se subdivide nas seguintes partes: ráquis, ráquila e

espiguetas. Ráquis é o eixo principal, é ereto na emergência da panícula e habitualmente recurvado ou pendente, na maturação. Ráquila desarticulada abaixo do flósculo inferior. Dos ramos laterais da panícula finalizam em espiguetas, dispostos espiraladamente. O comprimento da panícula, assim como o número de ramificações primárias (ráquila) diferem entre as variedades do gênero.

A flor do arroz encontram-se envolvida pelas glumelas, que compreende um perianto reduzido (lodículas) que é de estrutura transparente, além de órgão masculino e feminino. O feminino ou gineceu consiste de um ovário sésil (um óvulo), glabo com estilete curto terminando em dois estigmas plumosos, com pelos ramificados, com variação na coloração de branco a violeta.

Órgão masculino ou androceu compreende as anteras, em numero de seis, as quais contém os grãos de pólen e o filete, que as ligam à base da flor. No grão maduro as duas brácteas superiores denominam-se glumelas (lema e pálea) e as duas inferiores glumas.

Glumelas férteis (lema e pálea) crustáceo-pergaminosas, pergaminoso-papiráceas, coriáceo-cartáceas, compresso-carinadas, quase do mesmo comprimento; a inferior, é mais longa (lema), assimétrica, quinquenervada, mútica ou com arista. Esta é delgada e curta chegando a grossa e longa, sempre terminal; cobre uma parte das margens da pálea. Na parte superior a pálea, é mais estreita, às vezes mais longa, acuminada ou aristada, aguda ou obtusa, trinervada com nestas laterais, próximas as margens.

Glumas são duas, muito pequenas, reduzidas à anulares ou bilobadas na extremidade do pedicelo, livres uma da outra. Possuem panículas laxas ou contratas, ramos simples ou divididos, em forma de rácemo.

Lodículas: são duas e compreende o perianto reduzido, que são glabras, inteiras ou bilobadas. Estames são 6 ou menos, com filete curto e capilar.

Anteras estreitas e lineares.

Espigüeta é um conjunto de dois pares de brácteas que envolvem a flor de arroz, após a formação do grão, essa estrutura formará a casca. Pela esta estrutura morfológica, taxonomicamente mais valiosa, pode se definir o grau de evolução, especialmente, devido aos genes que foram desativados no curso do tempo, mas ainda se expressam nas anormais, fazendo surgir rudimentos de flóculos nas bases dos lemas estéreis; sobre esta teoria existem várias hipóteses (Oliveira, 1992).

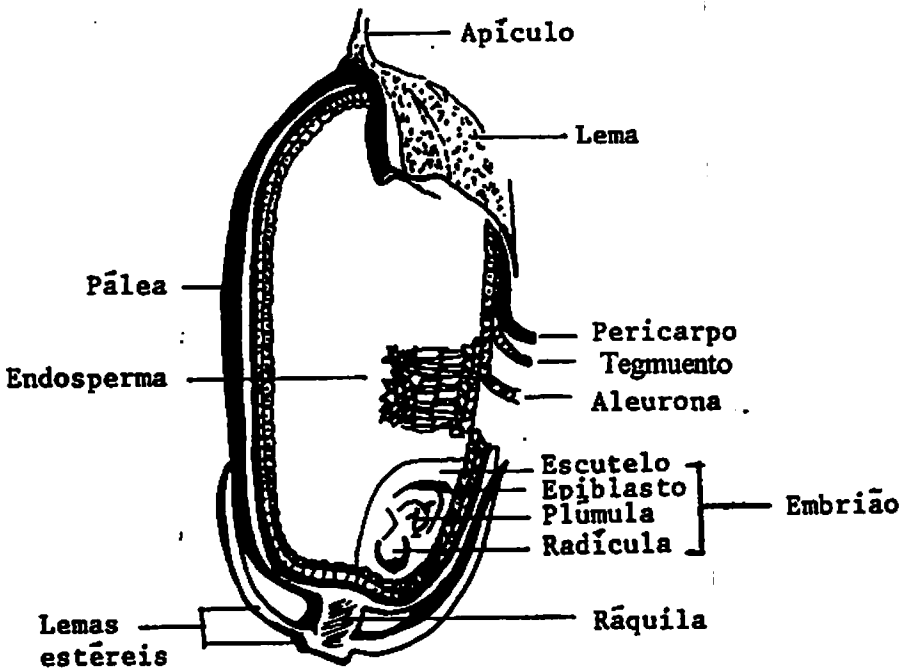


FIGURA 1 - Estrutura de uma espigüeta da planta de arroz na fase da maturação.

São curtas pediceladas, obliquas ou horizontalmente, que normalmente são isolados, em cada nó das ramificações. Também, possui articulação decídua e depressão fortemente lateral, estreito-oblongas ou lanceolado-elíptico-oblongas, com um flósculo terminal hermafrodita composto de uma pálea e um

lema, sobre dois flósculos basais estéreis compostos de apenas um lema estéril cada. Lemas estéreis com 1/8 a 1/2 do comprimento das espiguetas, que são lineares a linear-lanceolados, subculados ou cetáceos, eretos sem nervura (umi a qüinqüenervados), escariosos, subcoriáceos (finamente membranosos).

Fruto: é uma cariópse (Figura 1) lateralmente compressa, envolta pelas glumelas e ráquila (formando a espigüeta), que tem importância na classificação e identificação das formas cultivadas, e que serve para diferenciar as subespécies de *O. sativa* L. Existe uma classificação botânica que coincide o comprimento e a largura da espigüeta na identificação das formas cultivadas.

Ainda com relação do comprimento da espigüeta se classificam o grão em cinco proles distintas (Camargo, 1982; CIAT, 1975): perbrevis (< 7,0 mm), curta (7 - 8 mm), média (8 - 9 mm), longa (9-10 mm), perlonga (> 10 mm).

Escutelo oblíquo, ocupando quase sexta parte da cariópse, pode ser maior com bordas voltadas para frente. O Eixo do embrião é fusiforme cilíndrico.

### 1.3.3 - Histórico do Cultivo de Arroz no Brasil

O arroz é um dos alimentos mais antigos, pois há constatação de seu uso entre os chineses desde 3000 a.C. Atualmente, admite-se que sua difusão tenha acontecido no Sudoeste Asiático, da Índia para a China. Antes de espalhar-se no Território Chinês, também chegou para Indonésia e Ceilão, passando da China para a Coreia e para o Japão; a introdução na Ásia Ocidental e na Costa Mediterrânea foi mais recente, o que aconteceu pela expansão do Império Persa, que o cultivou da Mesopotâmia até ao Eufrates, Turquestão e Síria.

Os Árabes já no séc. IV da era cristã introduziram-no ao Norte da África e Egito. No séc. VIII foi introduzido na Península Ibérica, onde-se denominou arroz, tanto a planta como o produto, que deriva da palavra árabe 'roz'. Na América Central e parte da Sul foi introduzido pelos espanhóis e portugueses; em África Ocidental nas viagens áureas. No séc. XVIII, os



holandeses e portugueses levaram-no à América do Norte e Central, conseqüentemente ao Brasil, bem como à Austrália e Ilhas do Pacífico (Oliveira, 1992).

No Brasil, pelos estudos de Fornasieri Filho e Fornasieri (1993) e Oliveira (1992), desde o início do séc. XVI existem referências que podem ser confundidas entre o arroz nativo daqui [*Oriza grandiglumis* (Doell) Prod e *O. subculata* (Ness)] e usado pelos silvícolas, com *O. sativa* L. introduzido. Todavia, isso demonstra que foi o arroz-vermelho trazido pelos ilhéus, provenientes dos Açores, o qual teve a cultura desenvolvida até 1766, a partir de quando foi substituído pelo arroz-branco vindo da Carolina do Norte, EUA.

No séc. XVIII e início do XIX, o arroz alcançou grande destaque no Pará, Maranhão, Bahia, Rio de Janeiro e São Paulo, mantendo-se índice na pauta de exportações; porém, no séc. XX as coisas se inverteram. O Brasil passou de exportador a importador e em escala crescente; isso pelo fato do aumento de população e pela economia gerada na cultura cafeeira, mas a imposição de tarifas pretensionistas fez a importação cair rapidamente, elevando-se os preços para o consumidor interno. Durante a I Guerra a escassez de alimentos com preços altos culminou na expansão de área cultivada no Brasil, iniciando, aí o que se tem hoje como País produtor de arroz (Fornasieri Filho e Fornasieri, 1993; Abbud, 1991).

#### 1.3.3.1 - Sistemas de Cultivo das Lavouras de Arroz

Para melhor entendimento da situação da cultura do arroz devem-se considerar as peculiaridades dos diversos sistemas de cultivo (Fonseca, Rangel e Prabhu, 1981). No Brasil existem os três sistemas de cultivo, em função do fornecimento de água durante o ciclo da cultura: em sistema de terras altas e irrigado, ganhando, recentemente, interesse os sistemas de cultivos de arroz irrigado por aspersão ou pivô e em várzeas úmidas (terras altas favorecido).

A produção física anual de arroz em casca no Brasil, nos últimos 10

anos está com média em torno de 10, 2 milhões de t, produzidas em 4,41 milhões de ha, com predominância do sistema de terras altas, ocupando 73,7% do total da área rizícola, embora com menor produção (30%). Entretanto, nota-se que a maior produção vem do arroz irrigado (60%), que ocupa apenas 9,56 % da área rizícola, tem alta produtividade (5,35 t/ha), segundo Anuário... (1998).

A lavoura irrigada é a principal forma de cultivo no RS e SC e nas regiões de clima temperado, chega a ocupar quase 1,22 milhão de ha, com produtividade alcançando de 5,09 mil kg/ha. A cultura arrozeira movimentada em torno 174 milhões de dólares anuais (Agroanalysis, 1998).

Das pesquisas realizadas na Embrapa Arroz e Feijão (CNPAF) resultou um tipo ideal de planta que pode ser cultivada em sistema de cultivo em terras altas favorecido e mostra que as plantas devem ser intermediárias entre a tradicional de terras altas e a moderna (irrigado por inundação); ainda ser produtiva (3-5 t/ha), ciclo precoce a médio, resistências a brusone e mancha parda, viabilidade da soca, certo grau de dormência das sementes e grãos tipo de mercado (agulhinha). O denominado de terras altas favorecido já ocupa 16,80% da área do País e produz 10% do total (Anuário...1998).

Cultivares adotadas para cada sistema de cultivo diferem em suas características morfológicas e agrônômicas (Oliveira, 1992, Pedroso, 1982), tais como: cor das folhas, porte da planta e perfilhamento, ciclos de florescimento e maturação definida (precoce, média, tardia), cor das glumelas (casca do grão), dimensões e tipos de grãos, tamanho da panícula, peso de 100 grãos, resistências à seca e acamamento, brusone, helmintosporiose e cercosporiose

Ainda Abbud (1991) comenta que os materiais geneticamente melhorados para serem recomendados precisam possuir alguns quesitos para que possam se tornar uma cultivar: ter grãos do tipo agulhinha, gene do nanismo, tolerância à seca, ou mecanismo de escape.

Algumas vezes na literatura constata-se que as características de

resistência para seca e brusone (*Pyricularia oryzae* CAV) estão associados, visto que quase sempre as CV's adaptadas para o irrigado, quando plantadas no sequeiro se tornam mais susceptíveis, e de modo geral as CV's brasileiras de terras altas são mais tolerantes à brusone, isso se sob inundações. Outros atributos são: precocidade, suportar certas adversidades com relação aos sais do solo (Abbud, 1991).

#### 1.3.4 - Interação com o Ambiente

O arroz, como qualquer planta tem seu crescimento e desenvolvimento influenciados tanto por fatores genéticos como pelos ambientais, que são os responsáveis pela determinação dos caracteres vegetativos e reprodutivos da planta. O comportamento diferencial de genótipos em ambientes geográficos distintos, até mesmo situações diferentes, pelas técnicas culturais, tem mostrado que a variabilidade genética pode ser também utilizada para seleção de genótipos adequados a cada ecossistema e/ou emprego de técnica que modifique o ambiente (Stuber, 1992; Yoshida, 1981).

A interação genótipo X ambiente parte do princípio de que o ambiente é constituído de todos os fatores que afetam o desenvolvimento das plantas que não são de origem genética.

As causas da interação têm sido atribuídas a fatores fisiológicos e bioquímicos próprios de cada genótipo cultivado; como os genótipos se desenvolvem em sistemas dinâmicos onde ocorrem constantes mudanças, desde a semeadura até a maturação, há geralmente um comportamento diferenciado dos mesmos em termos de respostas às variações ambientais. Para avaliar a interação genótipo X ambiente é necessário avaliar em mais de um ambiente (Singh e Unrea, 1995). A agricultura tem evoluído para uma situação de maior controle das condições de ambiente, através de uso de insumos e técnicas modernas, para minizar os efeitos ambientais.

A relação entre variáveis são de origem genética e ambiental. Quando a

correlação é genética, denomina-se pleiotrópica, e se ambiental torna-se causa de alteração em dois ou mais caracteres consecutivos e se diz interação ambiental. A manifestação fenotípica é o resultado da ação da CV' sob interferência ambiental; entretanto, quando se considera uma série de ambientes, detecta-se um efeito adicional, resultante da interação deles. Entretanto, se os genótipos não são avaliados em ambientes representativos da Macrorregião, para onde se dirige a seleção, pode acontecer que as estimativas da variância genética não representam o seu valor verdadeiro.

Assim, como outras formas de interação, aquela entre genótipo e ambiente pode resultar da escala com a qual se mede e representa o carácter, desde que as linhagens não reagem da mesma maneira à mudança de uma variável do ambiente onde o indivíduo se encontra (Yoshida, 1981).

### **1.3.5 - Base Genética e Ganho das Culturas**

O melhoramento genético do arroz para sistema irrigado se iniciou no séc. XX em Ásia e foi baseado na seleção entre materiais locais, nos campos de agricultores, onde foram selecionadas plantas de maior produtividade e qualidade de grãos. O método consistia em eliminação de material genético fora do padrão adotado, ou seja, as misturas varietais. Em prosseguimento, era realizada seleção massal em tais materiais, posteriormente realizado a hibridação para combinar os fenótipos desejáveis em uma mesma cultivar. Para isso foram feitas introduções de materiais de países e regiões vizinhas, que foram cruzados com as variedades locais (de Datta, 1981).

Os primeiros trabalhos científicos foram realizados em uma única estação experimental, com avaliação das famílias em experimentos sem repetição. Há informações de que pesquisadores da Indonésia foram os primeiros a realizar seleção de materiais híbridos em estações de pesquisa, em

Boga e Jaca, provavelmente, o que proporcionou uma evolução de cultivares adaptadas para toda região de Jaca com seus diferentes tipos de solo e clima.

Porém, em Índia, os trabalhos de melhoramento tiveram início em 1911 em "Dast Bengala" (Bangladesh) e visaram a obter cultivares com diferentes ciclos, alta produtividade, boa qualidade de grãos, tolerâncias a inundações, salinidade e seca, além de resistências ao acamamento e brusone (Jennings, Coffman e Kauffman, 1985).

No Ocidente, o melhoramento do arroz iniciou-se por volta de 1909, nos EUA, que propuseram desenvolver cultivares com maturação precoce, grãos longos, porte moderado e adaptadas à aplicação de altas doses de fertilizantes. Nessa época grande parte da área plantada nos EUA era com as CV's Bluebelle, Della e Belle Patna.

A época que marca a estratégia da hibridação entre as subespécies Indica X Japônica, com seus objetivos claramente definidos e planejamento sistemático, representa uma transição do melhoramento antigo para o moderno; isso porque, como já foi salientado, as CV's do tipo Japônica apresentam grãos com características de comércio favorável, principalmente para o cultivo nos trópicos. Todavia, deve ser observado que a introdução direta das Japônicas nos trópicos não foi favorecida por várias razões.

O avanço no melhoramento do arroz aconteceu mais recentemente, pelo emprego das cultivares semi-anãs na China (Dee-geo-woo-gen, I-geo-tze e Tainchung Native 1). Todas estas portam o mesmo alelo recessivo do gene que reduz o tamanho dos internódios, sem que afete as panículas e as espiguetas. O alelo da Dee-geo-woo-gen para nanismo foi introduzido em inúmeras CV's e linhas melhoradas da raça indica, mais recente nas Japônicas (Jennings, Coffman e Kauffman, 1985).

Dos cruzamentos realizados, o destaque foi a 'Tainchung Native' (TN1) lançada em 1956, resultante da hibridação entre 'Dee-geo-woo-gen' X 'Tsai-

yuan-chung' de alta resposta a N, permitindo a produtividade de até 8,1 t/ha, que era quantidade bem acima da época. Com isso, a TN1 ficou sendo considerada a primeira CV' Indica de alta produtividade e seu desenvolvimento foi considerado um dos eventos mais significativos na História do melhoramento do arroz e de vegetais em geral. Com isso, também mostrou ser possível aumentar o potencial de rendimento do arroz indica através do melhoramento realizado apenas dentro do grupo, definindo um novo caminho para o melhoramento do arroz tropical.

O IRRI, fundado em 1960, pelas Fundações Ford e Rockefeller juntamente com o governo Filipino, iniciou em 1962 um programa de melhoramento, com ênfase a respostas crescentes a níveis de N, resistência a doenças e insetos, insensibilidade ao fotoperíodo, dormência de grãos, alto rendimento no beneficiamento, qualidade de grãos, lenta senescência de folhas e tolerância a baixas temperaturas.

Das hibridações realizadas, foi selecionada a 'IR-8' que fora resultante entre 'Dee-geo-woo-gen' e 'Peta' (originária da Indonésia, e de alto perfilhamento), por isso a 'IR-8' herdou esse carácter, ainda de folhas eretas, insensibilidade ao fotoperíodo, semi-anã e colmos mais rígidos, muito produtiva pela alta demanda de N, pois não sofre acamamento das plantas. Foi a primeira CV' de arroz indica altamente produtiva adaptada a climas tropicais, tomando-se conhecida como o arroz anão e milagroso, deixando inúmeras cultivares como grupo 'IR-5' com as demais 20, 22, 24 (Yoshida, 1981). O que as tornou CV's Obsoletas foram os problemas de doenças e pragas, devido ao cultivo contínuo, o que levou a pesquisa a se preocupar com esses fatores.

Em conseqüência, o IRRI lançou a linhagem 'IR-36', que chegou a ser cultivada em mais de 10 milhões de ha pelo mundo, pelas características de resistência às principais doenças e pragas do arroz, além de tolerar condições adversas de solo, e possuindo boa qualidade de grãos. Por causa disso, muitas

cultivares espalharam-se pelo mundo, tendo como ancestrais em programas de melhoramento genético do arroz as 'IR-8' e a 'TN1'.

Na América do Sul, especialmente em Colômbia, nos trabalhos desenvolvidos pelo CIAT e pelo ICA desenvolveram-se CV's superiores, com potencial produtivo, melhorando a arquitetura da planta, sementes com vigor, estatura semi-anã e folhas erectas, precocidade (90-120 dias), resistência à brusone, "ferrugem da bainha", boa qualidade de grãos. Dai surgiram as CV's 'Cica-4', provenientes de 'IR-930-31-1-1b', 'IR-22' e 'Cica-9', esta última sendo cultivada no sistema irrigado na Colômbia, e no de terras altas na Costa Rica e Panamá (Heiser, 1997; Abbud, 1991; de Datta, 1989).

A maioria dos estoques de germoplasma disponíveis no mundo é proveniente de poucas coletas, mormente realizadas em 1961, pelo Prof. Hiko-Ichi Oka, do Instituto Nacional de Genética, no Japão, de modo que a variabilidade não é muito grande, e como consequência da estreita base genética das mesmas. Hoje em dia, paradoxalmente, existem cada vez mais estudos sobre genética molecular desse material, enquanto que a caracterização de raças geográficas e outros tópicos simples não mereceram muitos esforços da pesquisa.

Na América Latina e Caribe as variedades de arroz que foram liberadas no período de 1977 - 89 têm na sua constituição genealógica, apenas 14 cv's distintos que foram provenientes de sete países, e receberam destes 69% de sua constituição genética (Cuevas-Perez et al., 1992).

No Brasil, o arroz de terras altas tem base genética formada por apenas seis ancestrais, e a partir da década de 80 tiveram que ampliar a base mediante a incorporação de germoplasma de África, Ásia e EUA (Heiser, 1997). Sempre estão dentro de grupo com as características de interesse agrônomo, originados de combinação de poucas variedades.

Silva (1994) comenta que a base genética do arroz cultivado no Brasil é

muita restrita pelo fato do melhoramento genético ter usado por muito tempo, quase somente genótipos considerados nacionais, pelo estudo de genealogia dessas CV's mostra existir um grau de parentesco acentuado entre as mesmas. A utilização de variedades melhoradas, mais adaptadas às nossas condições de cultivo, é uma alternativa que pode aumentar a produtividade da cultura do arroz, bem como a qualidade do produto.

Aceita-se como principal causa da queda do ganho genético, o estreitamento da base genética das cultivares utilizadas (Sristava e Nema, 1993); outra causa poderia ser o fato de que no Brasil utiliza-se constantemente o método genealógico para a condução das populações segregantes; no entanto, originalmente proposto depende em grande parte da eficiência da seleção visual nas primeiras gerações segregantes, a qual, no entanto, é de baixa eficiência; ainda assim, o método genealógico não permite a recombinação dos materiais segregantes que foram selecionados, para aproveitar os alelos favoráveis que se situam em indivíduos diferentes (Teixera, 1996; Camargo, 1982).

O uso freqüente do mesmo alelo de semi-nanismo no melhoramento genético poderá reduzir a diversidade genética do arroz por todo mundo, pois as CV's modernas contêm o mesmo alelo para nanismo e citoplasma semelhante. Devido ao uso intenso de CV's semi-anãs como fêmeas nos cruzamentos, existe a preocupação no sentido de incorporar germoplasma, materiais de origens diversas com novas fontes de ananismo, para ampliar a base genética. Sabe-se que por análise de RFLP, conseguiu-se identificar um alelo novo para semi-nanismo.

#### **1.3.5.1 - Variabilidade Genética ou Diversidade do Arroz**

Entende-se como sendo a capacidade de uma espécie, população ou mesmo progênie expressar diferentes fenótipos, que são fundamentado parcial ou totalmente na composição genética dos indivíduos; ainda pode a soma total da ação dos diferentes alelos em um certo ambiente.



A importância da diversidade genética para o melhoramento da espécie reside no fato de que cruzamentos que envolvem progenitores geneticamente diferentes são os mais convenientes para produzir alto efeito heterótico e também maior variabilidade genética em gerações segregantes (Rocha, 1987; Rao et al., 1981).

A variabilidade, que é imprescindível ao melhoramento, pode ser observada tanto entre quanto dentro das espécies, mesmo nas predominantemente autógamas. Por isso, as estimativas periódicas das variâncias genéticas e fenotípicas das populações sob seleção se prestam para orientação no melhoramento e na condução da seleção, avaliando o potencial genético da população e orientando a utilização de acessos de um BAG.

Na seleção de genótipos é importante ter conhecimentos sobre a natureza da divergência genética existente na população, a associação entre os diferentes caracteres quantitativos e a extensão da influência do ambiente sobre esses caracteres que serão selecionados.

A diversidade genética dentro das espécies cultivadas é essencial para sustentar os níveis de produtividade. Muitas vezes, quando se obtém um avanço tecnológico dentro de uma cultivar para produtividade com qualidade, isso sempre advém de poucos genes, ou apenas de um, gerando uma uniformidade genética desta CV obtida, objeto da pesquisa (Liang et al., 1994).

Um exemplo clássico da vulnerabilidade genética em plantas uniformes cultivadas aconteceu com a CV 'Vitória' de aveia que detinha resistência à doença da ferrugem, nos EUA. Pela sua característica de resistência à doença passou a ser usada como progenitor comum e fonte a resistência da raça virulenta. A variedade Vitoria possuía o gene *Pc-2* para o tipo de rusticidade na família 'vitoria', que tolerava a doença, até que novamente foi infestada pelo outro fator da ação do fungo, que está completamente ligado com o gene *Hv* para susceptibilidade à ferrugem (*Helminthosporium victoriae*). Como

conseqüência, todas as variedades derivadas da 'Vitória' foram infestadas, sofrendo os prejuízos da epidemia (Chang, 1976).

Outro exemplo aconteceu com as CV's de milho oriundas do Texas, que detinham a macho-esterilidade citoplasmática, quando a raça virulenta T do fungo (*Helminthosporium maydis*) atacou o "Corn Belt" na década de 70, causando queda na produção de milho (Chang, 1976).

No caso do arroz, o uso freqüente do mesmo alelo de semi-nanismo no melhoramento pode reduzir a diversidade genética dos materiais levando à vulnerabilidade genética, pois grande número de CV's possui o mesmo citoplasma, o que incrementou a aplicação de fertilizantes no sistema de irrigado favorecendo a multiplicação dos organismos patogênicos e pragas que se disseminam pelas lavouras devido à prática da intensidade cultural e o monocultivo.

A uniformidade para caracteres agrônômicos desejáveis nas variedades tem permanecido às expensas de uma redução na variabilidade genética e estão conduzindo ao que se conhece como vulnerabilidade à adversidade ambiental, ao ataque de pragas e a doenças (Beer et al., 1993). Há exemplos das perdas de fatores, devido à vulnerabilidade genética em todas espécies cultivadas que já foram estudadas; exemplos nos cereais como: aveia (Souza e Sorvells, 1989); arroz (Berrio e Cuevas-Peres, 1989).

De maneira geral, a variabilidade genética em muitas cultivares, parece ser a exploração agrônômica de uma base genética reduzida e a presença de pequenos grupos de ancestrais nos processos de melhoramento, a tentativa de obter uma variedade uniforme.

Pedroso (1982) menciona que o melhoramento genético de plantas tem sido praticado com sucesso desde os primórdios da civilização. Avanços podem ser alcançados a partir do momento em que exista variabilidade genética, e que o efeito ambiental não mascare por completo a expressão, enquanto que a

seleção e recombinação de genótipos superiores possam ser realizadas com o fim de se estabelecer a próxima geração. Sristava e Nema (1993) comentam que os materiais genéticos para receberem uma manutenção mais intensa deve ter uma variabilidade genética considerável.

Os programas de melhoramento, tanto os convencionais como os modernos, que utilizam técnicas de biotecnologia, dependem da disponibilidade da variabilidade, que são acessadas através de coletas de germoplasma de espécies silvestres ou não (CV's tradicionais) com a incorporação ao processo de seleção. Coletas de germoplasma e sua conservação em coleções garantem uma ampla base genética para a seleção e o melhoramento.

Para as espécies autógamas, nos casos em que a população desenvolvida, melhorada ou não, seja a fonte de coleta dos acessos, o estudo da divergência genética é fundamental para conhecer o comportamento "per se"; uma vez que em populações de base genética ampla, por apresentarem variabilidade aditiva de maior magnitude, respondem mais à seleção efetiva do que as populações de base restrita ou já previamente melhorada (Morais, 1992).

#### 1.3.5.2 - Fontes de Novos Reservatórios de Genes

Os novos reservatórios "pools" de genes podem ser indígenas ou introduzidos de uma fonte estrangeira, em outro país. Para o arroz existem as seguintes categorias, caracterizando-se em várias fontes de genes (Chang, 1976).

1 - CV's modernas (elite): altamente produtivas, provém de programas de melhoramento por hibridação e possuem o gene do ananismo, traz fontes de resistência às doenças e tolerância ao estresse do ambiente.

2 - Varietades Comerciais Principais: geralmente são bem adaptadas para um local, com maior produtividade por área. Possuem especialidade como: qualidade de grãos, tipo comercial, características intermediárias entre "elite" e

as de tipos minoritárias.

3 - Variedades Minoritárias: estas variedades não foram melhoradas, são inúmeras dentro de uma região, provavelmente contêm duplicatas, e entre regiões encontrarão suficiente diversidade que, às vezes, mostram diferenças somente nas características da planta e grãos. Isso que leva os produtores a continuarem o seu cultivo em pequenas áreas para apenas suprir suas necessidades. Não aparecem em plantios comerciais.

4 - Tipos Especiais: significante tolerância a uma endemia, frio, sal e solo adverso sempre são pobres em qualidade de grãos.

5 - CV's do Tipo Obsoletas: já foram variedade ou CV' especial a algum tempo ou lugar; porém, seu cultivo comercial está em desuso.

6 - Estoque para Melhoramento (germoplasma "elite"): são híbridos de CV's melhoradas com um grande número de caracteres desejados, mas ainda falta ser lançadas e nomeadas, ou seja, completar a pesquisa. Incluem-se neste grupo os F1, F2 ..., retrocruzamento, cruzamento múltiplo e suas progênies.

7 - Mutantes de genótipos: são provenientes de mudanças genéticas em um material conhecido, após uma mudança induzida por tratamento com agentes mutagênicos. Ocasionalmente um mutante espontâneo aparece naturalmente e pode surgir num programa de melhoramento.

8 - Tipos Primitivos: CV's que possuem caracteres primitivos, tais como: panículas laxas, pigmentos de antocianina, dormência de sementes, alta degrana, raiz com crescimento diverso, ou perfilhos no meio do internódio, alto nível de resistência a endemias, forte habilidade em competição, tolerância a inundação.

9 - Raças Invasoras: são raças que crescem como companheiras das cultivadas e em áreas adjacentes, as quais encontram condições de crescimento. Estas contêm várias integridades dentre as cultivares e as formas selvagens.

10 - Espécies Selvagens: dentro do gênero *Oryza* são dezoito, adaptadas

a muitos habitats, as quais são válidas, como a maioria são diplóides ( $2n = 24$ ) e poucas são tetraplóides ( $2n = 48$ ). Devido à ação antrópica em seus habitats, muitos ecótipos foram extintos (Vaughan, 1989).

### 1.3.5.3 - Conservação e Utilização de Germoplasma

A conservação genética almeja preservar o máximo das variações que estão disponíveis, dentro de uma dada população. Nestes termos, implica em um mínimo de seleção ou remoção de acessos, exceto para duplicatas evidentes e plantas atípicas autênticas; ainda, a multiplicação e uso de amostras de germoplasma podem realizar-se com os seguintes objetivos: incremento dos estoques de sementes, aplicação de descritores morfo-agronômicos, registro sistemático e eliminação de duplicatas por meio de comparação peculiar morfo-agronômica e evolução preliminar (Chang, 1976; Ram e Panwar, 1970).

O conjunto de acessos em um BAG deve ser caracterizado e catalogado para que possa ser utilizado pela pesquisa; às vezes o entrave consiste em definir quais os descritores que devem ser adotados para uma melhor caracterização.

Muitos critérios foram adotados, porém o que acontece normalmente é que as instituições adotam um número excessivo destes, podendo acontecer o inverso também. Deste modo é importante estabelecer critérios científicos sobre quais devem ser adotados para a otimização dos recursos, conseguindo uma caracterização mais direta e que atenda aos anseios dos usuários.

A finalidade e requerimento da sistemática dos descritores provêm de bases para: 1- caracterização de CV' ou melhoramento de linhas dentro de um interesse regional; 2- diferenciação entre acessos com nomes idênticos ou similares; 3- identificação de acessos tendo as características herdadas; 4- classificação comercial de variedades com critérios idôneos; 5- descobrir a interrelação entre grupo de caracteres geográficos; 6 - estimação da variabilidade existente comparando-se com uma coleção varietal; 7- avaliação

de genótipos dihaplóides.

Várias técnicas auxiliam na caracterização e discriminação biológica, que podem ser processadas através de descritores agrônômicos, morfológicos e bioquímicos, por técnica de micropropagação de propágulos; coeficientes de regressão para dados morfológicos quantitativos; sistemas de isoenzimas e citológicos; marcadores de DNA e a cultura de tecidos; para plantas de polinização cruzada (alógamas) utilizam-se embriões e ápices caulinares e até pedaços de folha (Rafaski e Willians, 1993; Kongkiatngam et al., 1995; Tingey e Del Tufo, 1992).

#### **1.3.5.3.1 - Importância das Coletas na Conservação de Germoplasma**

Na coleta de germoplasma consideram-se como prioridade aqueles recursos genéticos que apresentam maior urgência de preservação; quando se trata de plantas cultivadas, buscam-se esses recursos nas variedades locais a fim de minimizar as perdas de material devido à erosão genética causada pela substituição por variedades melhoradas e híbridos. No caso de espécies silvestres, a preservação é para evitar a ameaça de extinção (Coradini, 1980).

A importância das coletas está na utilização dos recursos genéticos em intercâmbio e no melhoramento, pelo fato de ampliar a base genética da cultura, possibilitando com isso, superar alguma barreira de ordem biológica ou ambiental. Há discussão sobre as estratégias das coletas, principalmente no que se refere às coletas das amostras em locais onde se observa variação no solo, vegetação, clima. Ainda pode haver outros fatores que possam induzir uma variação genética no material considerado (Ghaderi, Adams e Nassib, 1984).

A descrição sistemática da coleção requer plantio, mensuração, tratos culturais e registro de uma coleção varietal sob condições uniformes. Somente as diferenças em traços peculiares, que são registradas, representam verdadeiras características varietais que se expressa sob tais condições ambientais.

As peculiaridades para serem registradas devem mostrar possuir valor na taxonomia, utilização agrônômica ou investigação acadêmica. Essas peculiaridades têm mostrado ser de alta herdabilidade; somente assim podem as descobertas ser razoavelmente verificadas. A caracterização biológica, propriamente dita, é baseada em atributos qualitativos de alta herdabilidade (Valls, 1989).

Para o arroz existe uma relação confiável de caracteres morfológicos e agrônômicos, tais como: arquitetura da planta, parte aérea, folhas e os caracteres dos grãos que permitem o estudo e a expectativa de ampliação da base genética, a qual é constantemente citada como restrita ou estreita (Silva, 194).

### **1.3.6 - Caracterização dos Acessos através de Marcadores Morfo-Agrônômicos**

A caracterização fenotípica é realizada por meio de comparações de características morfológicas dos indivíduos, como cor de flor, cor de semente, cor e tipo de pubescência, dentre outras. Embora de fácil monitoramento, depende, na obtenção dos dados, de um processo dispendioso e demorado, uma vez que envolve as fases do desenvolvimento das plantas, além de serem na sua maioria afetados pelas condições ambientais, o que pode mesclar a informação (Sing et al., 1996; Kaw, 1995).

A caracterização e discriminação dessas populações, a níveis genéticos ou fenotípicos é importante para conhecer a variabilidade e estrutura genética existente no germoplasma disponível, também quantificar aqueles caracteres que sejam ligados ao processo produtivo (Sing et al., 1996). Ainda fornecer informações para a pesquisa, através do uso de descritores a fim de individualizar fenotipicamente cada acesso, bem como permitir maior eficiência no manejo do BAG (Beer et al., 1997).

Os primeiros marcadores utilizados em plantas foram os marcadores morfológicos ou fenotípicos (Keim et al., 1990a), que contribuíram para o

melhoramento vegetal convencional que é sustentado na seleção fenotípica de indivíduos segregantes oriundos de cruzamentos sexuais. A seleção é usualmente baseada em fenótipos e nos caracteres mensuráveis (Rafalski, Tingey e Williams, 1991).

A eficácia dessa seleção é mascarada pela influência do meio e restringe-se a regiões codificadoras do DNA; também pelo complexo multigênico envolvido na determinação das características fenotípicas; por isso, a seleção com base no fenótipo tem sido superada pela avaliação direta do genótipo, usando marcadores moleculares, por estes detectam e exploram a reação natural do polimorfismo na seqüência de nucleotídeos no DNA (Waugh, Baird e Powell, 1992; Souza e Soevelis, 1989).

Outra dificuldade discutida, nos estudos de genética clássica, pelo uso exclusivo de descritores morfológicos no auxílio à caracterização de germoplasma, refere-se a seu número disponível ser limitado; podendo ainda ser interpretados erroneamente, uma vez que são expressos em determinado período do desenvolvimento da planta, estando sujeitos a variações ambientais. O pequeno número de marcadores distintos numa linhagem reduz a probabilidade de se encontrar associações significativas entre descritores e caracteres de importância econômica, através do estudo de populações segregantes (Souza, 190; Hu e Quiros, 1991).

Além disso, a disponibilidade de marcadores morfológicos é essencialmente restrita às poucas espécies de plantas utilizadas como sistemas modelo para estudos de genética, tais como milho, tomate e ervilha, nas quais a intensidade de estudos e a disponibilidade de informações genéticas são maiores. Um enorme espectro de espécies vegetais carecia (e ainda carece) de informações a esse nível.

Para avaliar a divergência genética utilizando os marcadores morfológicos, deve-se anotar o maior número de caracteres dos materiais



envolvidos. O problema aparece quando se tenta utilizar toda essa informação simultaneamente, para se obterem as estimativas da divergência. Isto é possível com o emprego de estatística multivariada (Cruz, 1990; Ferreira, 1995).

Quando o número de acessos é grande, com poucas informações seguras sobre o comportamento dos descritores principais, adota-se um grande número de caracteres, sem estudo algum nem critérios relativos a contribuição destes para a variabilidade genética, fazendo gerar duplicação da informação, trazendo uma interpretação confusa e difícil.

Quanto aos marcadores citológicos, se extraídos da planta viva, podem alterar seriamente o fenótipo da planta, produzindo indivíduos que não podem ser utilizados, tem o monitoramento dificultado entre os cruzamentos elaborados. Devido a isso, só ocasionalmente que os marcadores citológicos e morfológicos ligados a genes de importância econômica são identificados, limitando seu emprego no melhoramento (Resende, 1991)..

Outros marcadores fenotípicos são as isoenzimas, as quais pelo fato de resultarem da interação da constituição genética com o meio, podem ser classificadas como tal; embora constituam as formas diferentes de uma mesma enzima, são controladas por um ou vários genes, podendo ser cadeias peptídicas muito parecidas ou diferentes, mas que têm a mesma função (Brummer, Kochert e Bouton, 1991).

As isoenzimas são consideradas superiores em relação aos caracteres morfológicos, por serem derivadas diretamente da expressão dos genes e pelo meio, estão mais próximo da expressão fenotípica final, ainda assim, esses marcadores são encontrados em quantidade limitada e apresentam pouco polimorfismo, podendo sofrer variações com o estágio de desenvolvimento da planta e apresentar diferenças entre os tecidos utilizados.

#### **1.3.6.1 - Caracterização através de Marcadores Genéticos**

Existem diversos tipos de marcadores que são produtos gênicos e são

intensamente usados em estudos de genética molecular; destacando-se os morfológicos, fisiológicos e bioquímicos. Por outro lado, existem os marcadores que se baseiam diretamente do DNA, seja em relação ao cromossomo como um todo (análise citogenética) ou em relação à molécula de DNA isoladamente (Gottlib, 1992.).

No primeiro caso, se apresentam em um número praticamente ilimitado e com alto grau de polimorfismo; no segundo, esses marcadores não são influenciados por condições ambientais, pois o polimorfismo é verificado a nível de DNA, que abrange tanto seqüências codificadoras como não codificadoras e os sítios de controle da expressão gênica; ao contrário dos demais marcadores, em que é analisado o produto gênico. Finalmente, esses marcadores não apresentam qualquer efeito fenotípico pleiotrópico de uma expressão na planta (Brummer, Kochert e Bouton, 1991).

Ainda mais, as regiões de éxons estão sob controle da seleção natural com mais intensidade do que os íntrons e, por isso, estes últimos são mais variáveis, podendo acumular grande diversidade; assim sendo, o número de marcadores moleculares possíveis de se encontrar é muito maior do que os marcadores fenotípicos, ampliando as possibilidades de estudo e manipulação do material genético da espécie, a qual se estuda (Stuber 1992; Keim et al., 1990a; ). Esses marcadores têm permitido um rápido avanço na genética molecular, detectando polimorfismo na seqüência de DNA, viabilizando estudos da herança genética, comportamento do indivíduo no ambiente e auxiliando em outras áreas do conhecimento (Beer et al., 1993).

Keim et al. (1990a) comentam ainda a vantagem de poder usar os marcadores genéticos em qualquer tecido e a qualquer estágio de desenvolvimento da planta: devido ao fato de não depender da expressão do gene, desde que suficiente quantidade de DNA de qualidade possa ser obtida, podendo desta forma, identificar diretamente genes antes da expressão

fenotípica, tomando por base esses pré-requisitos, vários tipos de marcadores genéticos foram encontrados.

A identificação de genótipos em estádios iniciais de desenvolvimento da planta abre a possibilidade de acelerar o processo de seleção e recombinação dos indivíduos desejados; como consequência, reduz-se o tempo necessário para completar uma geração de melhoramento, o que aumenta consideravelmente a eficiência do programa (Jack e Mayes, 1993). Conseguem-se marcar muitos locos com a preparação de DNA, o qual poderá ser estocado por longo período de tempo.

A revolução neste quadro iniciou-se com o desenvolvimento de marcadores isoenzimáticos, e, posteriormente, com a utilização de enzimas de restrição foi possível a análise de polimorfismo de comprimento de fragmentos de restrição de DNA (RFLP), segundo Saiki, Scharf e Faloona et al. (1985).

Com isso as vantagens desses marcadores são os seguintes: número elevado de marcadores, alto grau de polimorfismo, ausência de interação com o ambiente, não apresentam o efeito pleiotrópico; ainda mais, algumas alternativas promissoras dos marcadores moleculares sobre marcadores morfológicos são listadas a seguir:

(1) Muito esforço e planejamento são necessários na construção de mapa genético a partir de dados fenológicos, porque o número reduzido destes na linhagem restringe a cobertura do genoma. O polimorfismo limitado de marcadores morfológicos geralmente leva o pesquisador a recorrer a um grande número de cruzamentos para efeito de estudo de ligação gênica.

Por outro lado, um grande número de locos de marcadores moleculares pode ter seus alelos estudados em populações segregantes de cruzamentos específicos, facilitando o desenvolvimento de mapas genéticos. O nível de polimorfismo desses marcadores é geralmente alto para cada loco estudado, enquanto marcadores morfológicos possuem baixo nível de polimorfismo

(Kongkiatngam et al., 1995; Zabeau, 1993).

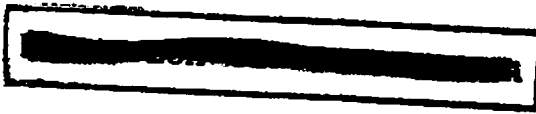
(2) Marcadores moleculares são, em geral, neutros em relação a efeitos fenotípicos, com mínimo ou nulo efeito epistático ou pleiotrópico. Marcadores morfológicos muitas vezes coíbem ou restringem o desenvolvimento normal da planta (albinos, mutantes clorofíticos), e seu controle genético pode afetar, ou ser afetado, por genes controlando outros caracteres, dificultando a caracterização dos genótipos.

(3) Marcadores moleculares, às vezes são co-dominantes, contendo maior quantidade de informação genética por loco, embora, por sua vez, são em sua maioria dominantes ou recessivos (Moser e Lee, 1994; Stuber, 1992).

Contrários a esses aspectos, marcadores morfológicos apresentam ainda, a desvantagem de ser na maioria identificados somente, a nível de planta inteira ou adulta. Enquanto que, marcadores bioquímicos ou moleculares como isoenzimas ou fragmentos de DNA, podem ser utilizados para caracterizar o genótipo de um indivíduo a partir de amostras de células ou de tecidos.

Pelos relatos recentes de Grodzicker, Adams e Hill (1992), que mais recentemente, estudaram o assunto, o ajuste e desenvolvimento do processo de amplificação em cadeia utilizando uma DNA-polimerase (PCR), levou à descrição de outras classes de marcadores moleculares, aliadas às técnicas de clonagem e de sequenciamento de DNA, possibilitando acúmulo de informações sobre a estrutura de genomas de eucariotos. O número de marcadores genéticos disponíveis foi ampliado em pelo menos uma ordem de magnitude, e a aplicabilidade da técnica passou a incluir potencialmente todas as espécies de plantas.

Atualmente, com o advento de técnicas de biologia molecular surgiram os marcadores genéticos, tomando a vez dos outros, destacando-se, principalmente, o RFLP e o RAPD, que são devidos à variação natural que ocorre na seqüência de bases do DNA. Os marcadores RAPD apresentam várias



vantagens com relação aos citados anteriormente; conseguem-se amplificar regiões do genoma que não são acessíveis à análise de RFLP, devido à presença de seqüências repetitivas no DNA. O uso desses marcadores tem aumentado consideravelmente o potencial para a detecção de diversidade entre indivíduos, pois são medidas variações existentes a nível de DNA (Kongkiatngam et al., 1995; Hu e Quiros, 1991).

Algumas das vantagens do RAPD sobre os demais são: simplicidade da técnica, usa elementos fluorescentes ao invés de radioativos, gasta quantidades inferiores de DNA, baixo custo por análise, alto nível de polimorfismo (Weeden, Muehlbauer e Ladizinsky, 1992).

A maneira de observar a variabilidade intra-específica do genoma de plantas é pelo polimorfismo nas seqüências de DNA, que podem ser seqüências altamente repetitivas, genes de cópia única ou pequenas famílias de genes. Dentre as várias aplicações das técnicas de RAPD citam-se: construção de mapas detalhados (Tinker, Fortin e Mather, 1993), acelerar programas de retrocruzamentos, selecionar indivíduos mais semelhantes ao progenitor recorrente (Mazurier, Godshach e Timothy, 1994), e ainda permite avaliar a variabilidade genética entre os organismos.

Esses marcadores, de maneira geral, estão sendo usados para caracterização e discriminação genotípica e identificação de marcadores ligados às características desejáveis em vegetais. O uso desses marcadores pode se tornar uma ferramenta complementar na identificação e caracterização de acessos, associando os caracteres de interesse a um programa de seleção, e ainda traz a faculdade de marcar características de efeito poligênico, por não sofrer a interferência do ambiente.

Esse aspecto é particularmente importante no melhoramento de espécies perenes de frutíferas e florestais, que requerem vários anos ou mesmo décadas para que uma geração de plantas melhoradas seja completada, ao contrário de

culturas anuais onde é possível obter mais de uma geração por ano.

### 1.3.6.2 - Da Técnica de PCR aos Marcadores RAPD

A introdução da técnica de reação em cadeia da polimerase (PCR) revolucionou os métodos tradicionais de obtenção de marcadores moleculares em nível de DNA. A técnica PCR alcançou uso extenso nas mais diversas áreas da biologia, por basear-se nas propriedades de replicação do DNA, o que lhe permite maior facilidade, rapidez e sensibilidade na detecção dos fragmentos que são os marcadores (Grodzicker, Adams e Hill, 1992; Williams et al., 1990).

As técnicas moleculares que são usadas para marcadores genéticos baseiam-se em uma propriedade intrínseca do DNA, a complementação de bases entre as duas fitas da molécula. Todavia, as que empregam eletroforese em géis, para separação de fragmentos de DNA, podem ser classificadas em dois grupos distintos: no primeiro, os fragmentos são gerados pela ação de enzimas de restrição, separados em gel e hibridizados à sondas apropriadas; no segundo grupo, os fragmentos de DNA são sintetizados "in vitro" por DNA polimerase termoestável, a partir de oligonucleotídeos iniciadores (primers) que se ligam a sítios específicos em moléculas-alvo.

O estudo dessa técnica foi desenvolvida por dois grupos de pesquisas, paralelas e independentes. O grupo de Williams et al. (1990) denominou-a de análise com RAPD (amplificação ao acaso de fragmentos polimórficos do DNA), enquanto que o grupo de Welsh e McClelland (1990), Califórnia, denominou-a por AP-PCR (reação em cadeia iniciada pela polimerase). Ambos chegaram à técnica revolucionária para gerar marcadores moleculares, com o uso de PCR, que fora descrita por Saiki, Scharf e Faloona (1985).

Nesse estudo, com a PCR que consiste em reações de amplificação simultânea das duas fitas de DNA, numa seqüência específica, utilizando-se dois "primers", que são complementares às extremidades do segmento a ser amplificado, através dos quais hibridizam ao DNA-molde, aleatoriamente.

Esses "primers" geralmente são seqüências de oligonucleotídios, com dez nucleotídios, de seqüência aleatória e uma freqüência de C+G superior a 50% (Williams et al., 1990). Esses iniciadores (primers) são peças fundamentais por possuírem dez nucleotídios e de 50 a 80% de C +G; essas características devem-se a testes feitos por Williams et al. (1990), que conseguiram sintetizar iniciadores de conteúdo de C+G variando de 0 - 100%. No entanto, foi amplificado somente quando essa percentagem estava acima de 40%. Quanto ao tamanho, os autores testaram iniciadores de 6 - 10 bases, sendo que o número mínimo necessário destas foi de nove.

A reação de RAPD é uma metodologia que tem como principio a técnica de PCR que foi desenvolvida por Saiki, Scharf e Faloona (1985). A técnica de RAPD baseia-se na amplificação de segmentos de DNA fragmentados pela enzima de restrição, utilizando iniciadores únicos de seqüência de nucleotídios conhecida e arbitrária, para amplificar segmentos específicos do DNA, que sejam diferentes do PCR (polymerase chain reaction).

Os iniciadores são orientados de modo que a síntese ocorra na região compreendida entre os mesmos. Os ciclos de amplificação de DNA compreendem a desnaturação do DNA-molde, o pareamento dos primers aos sítios que lhes são homólogos e a síntese ou a polimerização dos fragmentos de DNA, são repetidos varias vezes e permitem a amplificação exponencial do fragmento de DNA.

A enzima termoestável (Taq-polimerase) permanece ativa após incubação a 95 °C, a qual prepara a reação, que resiste a altas temperaturas sem ser inativada. Com isso os ciclos de amplificação podem variar, mas basicamente seguem a seguinte seqüência: um min a 94 °C, para que ocorra desnaturação das fitas de DNA e 2 min a 72 °C para extensão das fitas de DNA, que é o método de Williams et al. (1990). A seqüência de um minuto a 94°C, um minuto a 36 °C e dois minutos a 72 °C por 45 ciclos é a mais comumente utilizada, embora haja bastante variação dessas condições descritas na literatura

para atingir a otimização em cada espécie.

Em se tratando de RAPD, não requer o conhecimento do DNA a ser amplificado por um único tipo de iniciador (primer) de seqüência arbitrária o que é diferente do PCR, que se liga ao acaso no DNA-genômico em dois sítios, em fitas opostas do DNA-molde, se estes sítios estiverem separados por uma distância amplificável (geralmente entre 200 e 2.000 pares de bases) ilustrado na Figura 2. Com isso, acontece a amplificação da referida seqüências de nucleotídeos.

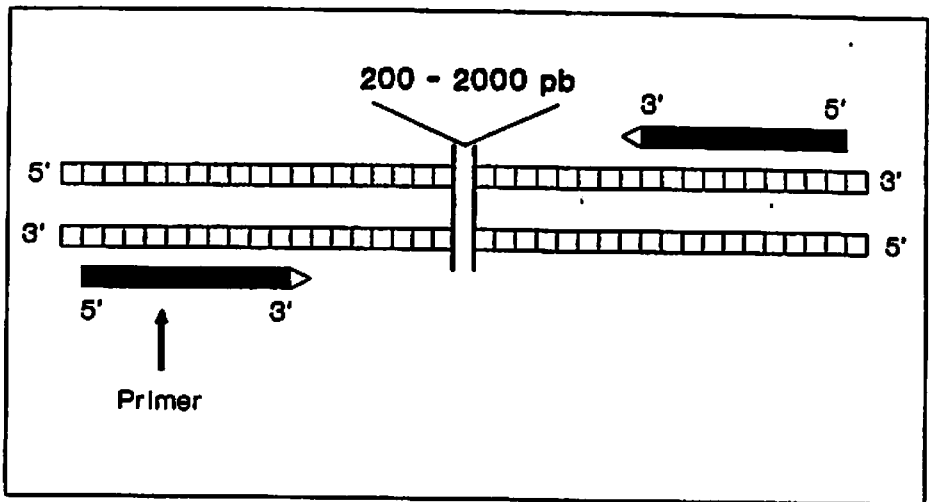


Figura 2 - Esquema de ligação dos “primers” à fita de DNA-molde e de amplificação das bandas polimórficas, de acordo com a técnica de Williams et al. (1990).

Os produtos da amplificação são geralmente separados em gel de agarose e visualizados pela coloração com brometo de etídio. Alternativamente, as bandas podem ser detectadas por coloração com prata em géis de poliacrilamida (Beer et al., 1993); esse método é mais sensível, permitindo uma melhor resolução das bandas. Entretanto, é um método ainda mais complexo, mais demorado e de custo mais elevado.



Para um dado iniciador pode-se identificar a presença de uma determinada banda polimórfica (produto de amplificação) em um indivíduo e a ausência em outro. Isso se deve a vários fatores, dentre estes citam-se:

a) a inexistência no indivíduo, daquele sítio de inserção para o iniciador ou a modificação deste por alguma mutação, não sendo mais reconhecido pelo iniciador;

b) a ocorrência de uma inserção entre os sítios, distanciando-os, leva à não amplificação; ou mesmo que isto ocorra, o fragmento amplificado terá tamanho diferente aparecendo a banda em outra posição;

c) a existência de um sítio alternativo, semelhante, porém não igual ao sítio real, mas que é reconhecido pelo iniciador levando à amplificação de fragmentos inteiramente diferentes.

Uma das aplicações direta dessa técnica consiste em correlacionar os polimorfismos das seqüências nucleotídicas com caracteres qualitativos e quantitativos, uma vez que os produtos de amplificação de cada iniciador poderão estar segregando juntamente alguma característica de interesse. Se isso for verificado, então o referido iniciador será um marcador para essa característica (Waugh, Baird e Powell, 1992; Williams et al., 1990).

Dentre os estudos que utilizaram o método de RAPD pode citar-se a identificação de cultivares de brócolos e couve-flor (Tinker, Fortin e Mather, 1993), herança e compatibilidade dos marcadores RAPD (Weeden, Muehlbauer e Ladizinsky, 1992); RAPD sem necessidade de purificar DNA (Martin et al., 1995; Mazurier, Godshach e Timothy, 1994).

A natureza dos fragmentos que são amplificados é altamente dependente da seqüência de nucleotídios do iniciador e da seqüência de DNA do genoma analisado, pois se um iniciador difere em apenas um nucleotídio, dá origem a diferentes bandas amplificadas. Quando há diferenças no genoma em um ou em ambos os sítios de ligação dos iniciadores resulta no desaparecimento

de uma banda amplificada. Assim, a amplificação de DNA com iniciador de seqüências aleatórias é um método altamente sensível para descobrir polimorfismo distribuído aleatoriamente no genoma (Martín et al., 1995).

Um iniciador usualmente amplifica várias bandas, cada uma originada de uma localização diferente do genoma (Rafalski, Tingey e Williams, 1991). Além do polimorfismo gerado pela modificação de sítios de ligação dos iniciadores, mediante mutações de ponto, o polimorfismo pode ser gerado também por deleções que incluam um ou ambos os sítios de ligação e inserções entre os sítios de ligação. Com isso podem distanciá-los de tal forma, que impossibilitam a amplificação ou mesmo, pequenas inserções e deleções, as quais apesar de não impedirem a amplificação, teriam o tamanho modificado do fragmento amplificado (Williams et al., 1990).

A freqüência de polimorfismo em marcadores RAPD mostrou-se ser de 0,3 banda polimórfica por iniciador, em *Arabidopsis thaliana*; 0,5 banda/iniciador em soja; uma banda/iniciador em milho; 1,68 banda Cao e Oard, (1997) em arroz e 2,5 bandas em *Neurospora crassa* (Tingey e Del Tufo, 1992).

Segundo Williams et al. (1990), os marcadores dominantes, tipo RAPD, podem ser usados para distinguir indivíduos heterozigotos de indivíduos homozigotos, desde que se usem marcadores RAPD fortemente ligados em 'trans', usados em pares. A confiança de que os indivíduos heterozigotos podem ser identificados depende da proximidade dos marcadores utilizados. Portanto, o uso de marcadores dominantes em pares para detectar heterozigotos requer duas vezes mais marcadores do que o necessário, utilizando-se marcadores co-dominantes.

O RAPD detecta a presença de apenas um alelo em determinado locus, que é o alelo representado pela presença de uma banda amplificada. A ausência de amplificação representa todos os outros alelos naquele locus; portanto, são

predominantemente dominantes e não podem ser utilizados individualmente para distinguir locos heterozigotos (Rafalski, Tingey e Williams, 1991).

Com uma frequência relativamente baixa, marcadores RAPD co-dominantes podem aparecer. Em um estudo com *Neurospora crassa*, de 88 marcadores RAPD identificados, 84 eram dominantes e quatro co-dominantes (Williams et al., 1990). Stuber (1992) cita que, dos marcadores RAPD mapeados, 98% eram dominantes e 2% co-dominantes. O polimorfismo é usualmente notado como presença ou ausência de produtos amplificados a partir de um único locus.

Os marcadores RAPD são dominantes e duas cópias do mesmo alelo (indivíduo heterozigoto) não são distinguidas quantitativamente daqueles com somente uma cópia do alelo. Em termos médios, cada iniciador pode detectar diretamente a amplificação de muitos loci discretos no genoma, fazendo da análise um eficiente caminho para marcar seqüências polimórficas de nucleotídeos entre indivíduos (Tingey e Del Tufo, 1992; Williams et al., 1990).

#### 1.3.6.3 - Tipos de Marcadores a Nível de DNA

O primeiro marcador molecular estudado, a nível de DNA de genomas complexos foi a partir de clivagem enzimática do DNA, a qual se tornou bem difundida como marcador genético, após o desenvolvimento da técnica de transferência de fragmentos de DNA separados em géis, para membrana de nitrocelulose, nas quais são realizadas hibridações com sondas apropriadas (Brummer, Kochert e Bouton, 1991). O autor ainda conseguiu determinar cerca de 200 fragmentos de DNA, gerados pela ação da enzima de restrição e polimorfismo, que poderiam ser usados como base para mapa de ligação (Paterson, Tanksley e Sorrells, 1991).

A técnica de RFLP, a qual tem algumas desvantagens em relação ao RAPD, pois envolve hibridação, "southern blot", o que a torna indesejável para aplicação em melhoramento, não permite a automação de algumas etapas. Uma

hibridização com DNA “blot” em geral requer muito tempo e trabalho intensivo (Tingey e Del Tufo, 1992), apresentando custos operacionais elevados, demora na obtenção de resultados, além de detectar poucos polimorfismos em espécies com estreita base genética, as quais são fatores limitantes para sua utilização em larga escala (Hu e Quiros, 1991).

Outra limitação, no caso de RFLP é a discriminação de sondas de DNA, que identifique a variabilidade genética a ser avaliada. Os marcadores RFLP têm sido usados para estudos de taxonomia, localização de seqüências específicas em genoma de tomate (Vallejos, Sahiyama e Chase, 1992), mapeamento genético de alface (Welsh e McClelland, 1990), análise genética da dureza do grão e diversidade genética em soja (Keim et al., 1990a), diversidade genética e mapeamento em arroz (Mackill e Lei, 1997; Dilday, 1990); mapeamento do genoma utilizando linhas isogênicas (Muehlbauer et al., 1991). A disseminação do seu uso contribuiu para a descoberta e estudo de diversas classes de seqüências repetitivas de DNA, chamadas mini e microssatélites, outra rica fonte de polimorfismo genético.

Hoje, um número virtualmente ilimitado de marcadores moleculares altamente polimórficos podem ser obtidos em qualquer organismo vivo, através de diversas técnicas. Os genomas de eucariotos são caracterizados pela ocorrência de grandes quantidades de seqüências repetidas de varias classes, uma delas consiste de seqüências de DNA curtas (2 a 4 pb), repetidas em tandem, denominadas de microssatélites ou seqüências únicas repetidas.

Os microssatélites são caracterizados por alto grau de polimorfismo. Muitos resultam de diferentes números de repetições de uma seqüência básica, podendo destacar sua característica de multialelismo que amplia as possibilidade de estudos genéticos em plantas (Morgante e Olivieri, 1993).

Os microssatélites podem ser alocados em grupos de ligação genética junto com RFLP, com RAPD e com marcadores fenotípicos, por serem altamente

polimórficos, podendo ser complementares aos marcadores de RFLP e RAPD, principalmente em espécies onde estes são de difícil detecção. O alto grau de polimorfismo foi confirmado por Mazurier, Godshach e Timothy (1994) e ainda mostraram que são mais variáveis que os RFLP.

A análise de AFLP (Amplified Fragment Length Polymorfism) é a tecnologia mais recente para a obtenção de grande número de marcadores moleculares distribuídos pelos genomas de procariotos e eucariotos; está sendo utilizada para construção de "fingerprinting", mapeamento genético localizado (BSA) e mapas genéticos, principalmente em espécies cultivadas que possuem baixa taxa de polimorfismo (Zabeau, 1993).

Os marcadores moleculares de locos controladores de caracteres quantitativos (denominados QTL), constituem a primeira estratégia para medida direta (DNA) da diversidade; também são usados para detectar e estimar o efeito de poligenes ligados a genes únicos e, mais recentemente, estão sendo usados para análises de correlação entre os marcadores moleculares e características quantitativas (Tinker, Fortin e Mather, 1993).

Particularmente relevante é o uso de marcadores moleculares com a finalidade de obter informações sobre genes que determinam características agrônômicas importantes, facilitando, sobremaneira, os esforços em melhoramento de plantas (Paterson, Tanksley e Sorrells 1991). É possível identificar, por exemplo, quantos genes influenciam dada característica; onde esses genes estão localizados nos cromossomas; o quanto cada gene afeta a característica; se genes individuais afetam mais de uma característica; se os genes funcionam similarmente em diferentes ambientes; se o efeito de um gene influencia o efeito de outros; e se um indivíduo com duas cópias de um gene é diferente de um indivíduo com apenas uma cópia.

Outros tipos de marcadores moleculares, mas que podem sofrer interação do ambiente no qual se encontram, são as isoenzimas (Stuber, 1992;

Vallejos, Sahiyama e Chase, 1992) e as proteínas de reserva de sementes, as quais podem ser detectadas mediante análises por eletroforese (Kongkiatngam et al., 1995) e por cromatografia líquida de alta resolução (Brummer, Kochert e Bouton, 1991). Tais marcadores apresentam a vantagem sobre os marcadores morfológicos, pelo fato de serem produtos diretos da expressão de genes específicos. A principal desvantagem deles é a quantidade limitada de polimorfismo que são capazes de evidenciar entre genótipos relacionados (Hu e Quiros, 1991).

### 1.3.7 - Banco Ativo de Germoplasma (BAG)

Entende-se por germoplasma o conjunto de genes que representa todos os alelos existentes em uma referida espécie ou o somatório total dos materiais de uma espécie (Berding e Roach, 1987), constituindo-se em material indispensável ao progresso das pesquisas e principalmente ao futuro da sobrevivência do homem. Segundo Giacometti e Fonseca (1980), o germoplasma pode ser utilizado na forma de sementes, estacas, gemas, bulbos, rizomas, tubérculos, meristemas, embrião, calos, plântulas, pólen, células, ou ainda estrutura mais simples, sendo o local de conservação, os denominados de BAG (banco ativo de germoplasma).

A preocupação no BAG é relevante para não deixar extinguir os recursos biológicos contidos num germoplasma que após uma coleta entram para o BAG como acesso daquela espécie. Kumari e Rangasamy (1997) comentam que praticamente todo arroz cultivado no mundo foi proveniente da Ásia, centro de diversidade da espécie que é detentora da grande variabilidade genética de arroz cultivado. Basicamente, o germoplasma de *O. sativa* tem proporcionado o conjunto gênico primário para os trabalhos de melhoramento. Situa-se no IRRRI o maior BAG da espécie, com amostras coletadas por todos os continentes.

No Brasil, a coleção ativa de arroz é mantida no BAG da Eembrapa

Arroz e feijão, Goiania,GO, que mantem acessos com ampla variabilidade genética, visando à preservação e multiplicação a curto prazos, em quantidade suficiente para atenderas solicitações. A ampliação da coleção é feita através da introdução de germoplasma, solicitação e doação espontânea, fazendo-se necessário sua multiplicação, pois as amostras são normalmente recebidas em pequena quantidade (Giacometti e Fonseca, 1980).

O plantio inicial das amostras obtidas em expedições de coletadas se completa com as seguintes tarefas: aumento do estoque de sementes, descrição morfo-agronômica pela chave sistemática, eliminação de acessos-duplicados óbvios pela descrição morfo-agronômicas e de local de coleta, avaliação preliminar para seleção conveniente dos acessos que possam receber intensa avaliação para ser conservados, formando o sistema de banco de dados que poderá ser lido e avaliado em outros trabalhos de pesquisa ou não, permitindo formar os catálogos para suporte as pesquisas em todas as áreas do conhecimento.

#### **1.3.7.1 - Descarte de Acessos-Duplicatas ou Duplicados**

Duplicatas são acessos extras de alguma variedade ou parental muito idênticos, que são variantes de uma mesma variedade. Estes acessos surgem quando uma variedade tenha se expandido muito, por ser cultivada em muitos locais; com isso pode ser encontrada com nomes mudados pelas localizações, resultando em:

1- Variedades morfologicamente idênticas com históricos diferentes e mesmo nome comum, provenientes de áreas de microclimas ou mesmo de outro País.

2 - Variedades com nomes diferentes e locais distintos ou diferentes nomes que deixam descendentes morfo-agronômicos essencialmente idênticos.

3 - Eco-raças tendo o mesmo nome mas que podem ser diferenciadas pelas suas respostas fisiológicas às condições adversas.

### 1.3.7.2 - **Biometria Aplicada no Descarte de Acessos-Duplicatas**

O uso da Biometria para estudo de biologia vem sendo ampliado para várias espécies e tem sido de grande utilidade para avaliação de acessos em bancos de germoplasma. Tem sido utilizada para determinar a divergência no descarte de duplicatas, identificação de progenitores para cruzamentos, estabelecimento da relação entre diversidade genética e geográfica, avaliação da variabilidade total disponível em grupos geneticamente relacionados (Kaw, 1995; Singh et al., 1993; Afifi e Clarck, 1984).

O termo taxonomia numérica é definido como o agrupamento, por métodos numéricos, de unidades taxonômicas de organização (OTU's), com base nos estudos dos caracteres avaliados. Um caracter, neste contexto, significa qualquer propriedade que possa variar entre unidades taxonômicas e os valores que lhe são atribuídos chamam-se estados daquele caracter (Cannings e Hoppensteadt, 1982).

Desde o início do século, pesquisadores vêm aplicando métodos numéricos à taxonomia. Neste processo, um dos primeiros métodos estatísticos de interesse para sistemática foi o "Coeficiente de Semelhança Racial". Esse coeficiente consiste em um tipo de coeficiente de similaridade, tendo sido desenvolvido, posteriormente, por Mahalanobis, na forma de "Distância Generalizada". A partir desses trabalhos, vários outros foram desenvolvidos, resultando em grande incremento nessa área, tanto pelo desenvolvimento, quanto na aplicação desses novos métodos.

Entre as várias medidas de dissimilaridade, empregadas em estudos de divergência genética, as distâncias Euclidianas e de Mahalanobis ( $D^2_{ii}$ ) têm sido as mais amplamente utilizadas (Afifi e Clarck, 1984; Peter e Rai, 1976).

Os métodos de natureza quantitativa de avaliação da divergência, ou da heterose manifestada nos híbridos, usam as análises dialélicas, que necessitam



de um número  $p$  de progenitores e de todas as suas combinações possíveis  $[p(p-1)/2]$ , ou parte delas tomando impraticável, em alguns casos a utilização quando  $p$  é de valor elevado, principalmente em plantas autóгамas.

Inúmeras medidas de dissimilaridade satisfazem as propriedades matemáticas conhecidas como métricas. Assim, considerando a dissimilaridade  $d_{ij}$  entre as OTU's  $i$  e  $j$ , possuem propriedades, segundo Cannings e Hoppensteadt (1982), Sneath e Sokal, (1973), tais como:

a) simetria:  $d_{ij} = d_{ji}$  maior ou igual 0, ou seja, a dissimilaridade entre  $i$  e  $j$  é igual independentemente da direção em que é medida, devendo ser positiva quando as duas OTU's não são idênticas ou ter valor zero quando são idênticas;

b) distinção entre OTU's não idênticas: se  $d_{ij}$  diferente 0, então  $i$  diferente  $j$ , ou seja,  $i$  e  $j$  não são a mesma OTU;

c) não distinção entre OTU's idênticas:  $d_{ij} = 0$ , ou seja, quando  $i$  e  $j$  são idênticas a dissimilaridade  $d_{ij}$  é zero;

d) desigualdade triangular: dadas 3 OTU's  $i, j$  e  $k$ , as dissimilaridades entre elas devem satisfazer a "desigualdade", assim  $d_{ik} \leq d_{ij} + d_{jk}$ .

Com a finalidade de separar os organismos em grupos, são empregados métodos estatísticos que permitem o cálculo das medidas de dissimilaridade ou distâncias entre OTU's e métodos que permitem o agrupamento das diferentes OTU's em grupos homogêneos. Tais técnicas foram amplamente desenvolvidas nas três últimas décadas e neste trabalho serão discutidas as mais empregadas em estudos taxonômicos; a outra alternativa é o uso das análises multivariadas, após ter sido identificados os marcadores, por técnicas de agrupamento.

Na caracterização de acessos e seleção de progenitores para cruzamentos, às vezes toma-se a diversidade geográfica como indicador de divergência genética, critério que recebe muitas críticas, pelo fato de não quantificar a divergência real entre as populações. Em muitos acessos, a diversidade geográfica não retrata a divergência genética. Há estudos de

relacionamento entre a divergência dos indivíduos, estimada com base em  $D^2_{ii}$  com locais das origens ou coletas dos acessos. Esses estudos são importantes para entendimento da evolução da espécie, além de dar informações sobre os recursos naturais e facilitar o intercâmbio de germoplasma (Cruz, 1990). As informações a esse respeito, em sua maioria relataram a inexistência de qualquer relação coerente (Singh et al, 1993; Muehlbauer et al, (1991), enquanto que outros trabalhos mostraram que a divergência obedeceu ao padrão de distribuição geográfica das variedades, com isso demonstrando respostas não generalizadas (Singh e Unrea, 1995; Rao et al. 1981).

### **1.3.7.3 - Emprego da Divergência Genética, Correlação e Heterose**

Os primeiros trabalhos em taxonômia numérica não consideravam os níveis taxonômicos específicos ou intraespecíficos e, conseqüentemente, uma das restrições seria que as OTU's fossem indivíduos ou exemplares e não amostras de populações. Em segundo lugar, o cálculo de distâncias generalizadas para um grande número de caracteres e/ou amostras requeria recursos computacionais não disponíveis na época.

Além das restrições anteriores, a distância generalizada empregava caracteres contínuos e não caracteres discretos, usualmente utilizados em taxonômia numérica. Atualmente, estas três objeções são irrelevantes e a distância generalizada tem-se mostrado bastante útil para taxonomistas numéricos (Van Beuningen e Busch, 1997; Moser e Lee, 1994).

A grande maioria dos trabalhos sobre divergência genética procura avaliar o grau de divergência existente numa população ou material em estudo e as suas relações de parentesco, diversidade de origem geográfica e heterose. Para tanto, são empregadas diversas fórmulas para compreender a oscilação que ocorre em populações quando esta sofre imigração, mutação e seleção, o que faz mudar o equilíbrio das mesmas (Kaw, 1995; Stuber, 1992).

A divergência genética entre dois grupos ou mais indivíduos é medida pela diferença nas frequências alélicas. Essa medida é importante em vários aspectos, mas sobretudo num BAG e no melhoramento de plantas visando identificar combinações mais promissoras, isto é, aquelas que, quando cruzadas, expressam maior heterose (Ghaderi, Adams e Nassib, 1984).

Na hibridação, ao utilizar progenitores que possuem entre si caracteres favoráveis e necessários para obter as recombinações especificamente desejadas, o melhorista de plantas cria populações nas quais a seleção terá maior probabilidade de êxito. Os resultados são mais rápidos e seguros quando a população-base submetida à seleção tem alta média para o carácter a ser selecionado e ampla variabilidade. Esta amplitude genética em uma população segregante é função da divergência genética entre os progenitores envolvidos nos cruzamentos (Ghaderi, Adams e Nassib, 1984).

A divergência genética de uma população pode ser estudada através de: a) genealogia - b) diversidade geográfica - c) análise dialélica - d) técnicas multivariadas (marcadores morfológicos) e marcadores moleculares de DNA e, fisiológicos com emprego de proteínas e isoenzimas (Autrique et al., 1996). Desses, os menos onerosos são os baseados em genealogia e os que empregam marcadores morfológicos.

Considerando que a avaliação por meio da análise genealógica frequentemente se torna impossibilitada por falta de registro de todas as populações de interesse, a diversidade geográfica, ao contrário, tem sido muito utilizada na inferência da diversidade genética (Singh e Unrea, 1995; Zhu et al, 1998).

Deste modo, a análise da divergência pode trazer importantes contribuições à caracterização e descarte de acessos, principalmente para investigar a natureza e suas relações entre a divergência genética, parentesco, diversidade de origem geográfica, capacidade de combinação e heterose entre

cultivares e/ou linhagens de plantas (Ghaderi, Adams e Nassib, 1984).

Uma medida de divergência genética entre introduções e/ou cultivares que possa ser obtida antes de serem realizados os cruzamentos, permitiria ao melhorista esforços no sentido de que aquelas combinações apresentassem maiores chances de sucesso. Essa medida pode ser determinada por procedimentos de análise multivariada: em que diversas características podem ser dimensionadas simultaneamente nos acessos, medidas essas que são tomadas como estimativas da divergência genética (Singh e Chaudhary, 1979).

A respeito da correlação ambiental, que mede o grau de associação entre caracteres devido a fatores ambientais, a mesma ocorre sempre que dois ou mais caracteres forem influenciados pelos mesmos fatores ambientais. Será positiva quando o efeito ambiental atuar na mesma direção em ambos os caracteres e negativa em caso contrário. Sabe-se que estas duas correlações, muitas às vezes, diferem em magnitude e até um sinal.

Quando a diferença for detectada pelo sinal será vista como causas genéticas, pois as causas ambientais afetam os caracteres por rotas fisiológicas diferentes (Bussab, Miazaki e Andrade, 1991). Deve-se sempre, preocupar-se com a correlação simultânea para os caracteres; sabendo que o sinal e a magnitude da correlação determinam a direção e a intensidade da resposta, ocasião que ocorrem respostas correlacionadas.

A heterose é considerada como a expressão fenotípica dos efeitos genéticos da hibridação entre indivíduos com genealogias diferentes. Com isso, o interessante é a manifestação variável, que pode ser expressa quando os parentais de um híbrido têm diferentes alelos num loco e há algum nível de dominância entre alelos de cada loco. Por isso a hibridação intencional passou a ser um método científico de aperfeiçoar e melhorar plantas (Heiser, 1997). Este autor ainda acrescenta que desde os primeiros trabalhos de hibridação em arroz, diversos caracteres têm sido utilizados para estudar a manifestação da

heterose nos híbridos específicos.

Uma estimativa de a variância genética alta indica a existência de variação nas comparações entre genótipos, mas quando a variância é baixa, uma estimativa que tem sido bastante utilizada para a inferência é o coeficiente de herdabilidade ( $h^2$ ). Esta estimativa permite antever a possibilidade de sucesso com a seleção, por refletir a proporção da variação fenotípica que pode ser herdada, e pode medir o valor fenotípico como indicador do valor reprodutivo (Ram e Panwar, 1970).

A divergência genética e relações de caracteres têm sido consideradas para estudos de heterose. Isto porque, dependendo da espécie, pode existir uma relação entre a expressão da heterose e a divergência genética dos parentais. Diante disso maior divergência genética, até um certo limite, pode expressar maior heterose (Souza, 1990; Muehlbauer et al., (1991). Por isso, diversos trabalhos têm sido conduzidos no sentido de avaliar o grau de associação existente entre diferentes métodos preditivos de estimação da diversidade genética, entre os quais estes: Autrique et al. (1996), em trigo; em soja; Beer et al. (1993), em aveia; Alfenas et al. (1981), em fungos; Cao e Oard (1997) e Dilday (1990), também Zhu et al. 1998) no germoplasma de arroz

### **1.3.8 - Distância Genética Baseada na Distância Estatística**

As diferentes formas genéticas entre as populações podem ser medidas com sucesso pela estatística multivariada, isto é da distância genética que é utilizada em análises de agrupamentos (análise multivariada). Por esse método procura-se detectar padrões de distribuição da diversidade.

Desse modo, os métodos aglomerativos diferem dos demais, em razão de dependerem fundamentalmente de medidas de dissimilaridade, estimadas previamente, como: euclidiana simples, euclidiana média e quadrática, ou a  $D^2_{ii}$ , além de outras. No método de agrupamento dos componentes principais e da variáveis canônicas, o objetivo é avaliar a similaridade dos indivíduos por

intermédio de uma dispersão gráfica, nos quais se consideram, em geral, dois eixos cartesianos (Cruz e Regazzi, 1994).

Os métodos preditivos, são aqueles que tomam por base as diferenças em morfologia, fisiologia e bioquímica mostradas pelos progenitores na determinação da divergência, geralmente quantificada pela estatística multivariada da distância ( $D^2_{ii'}$ ). Um outro exemplo é quando se faz a inferência da divergência genética com base na distribuição geográfica; é um método preditivo também, adotando técnicas de análise quantitativa.

Várias são as técnicas para medidas de similaridade ou dissimilaridade. Morais (1992) e Singh et al. (1993) sugerem, dentre outras, as distâncias Euclidianas e  $D^2_{ii'}$ , que são destacadas, e os coeficientes das correlações como medidas de dissimilaridade. Sua escolha acontece subjetivamente, mas considerando vários fatores, como: natureza das variáveis avaliadas ou a escala de mensuração, também a precisão das estimativas e facilidade de computar os dados.

Embora as distâncias sejam uma medida de dissimilaridade, sempre são citadas como medida de semelhança (Cruz, 1990; Afifi e Clark, 1984).

#### 1.3.8.1 - Distância Generalizada de Mahalanobis ( $D^2_{ii'}$ )

Pelos relatos de Cannings e Hoppensteadt (1982), a medida  $D^2_{ii'}$  é mais apropriada quando existe correlação entre os caracteres em estudo. A técnica tem sido largamente usada na avaliação da divergência genética, por intermédio da estatística multivariada e é comum a transformação das variáveis originais em um conjunto de variáveis padronizadas e não-correlacionadas. A estimativa da distância ( $D^2_{ii'}$ ) entre dois indivíduos  $i$  e  $i'$ , é definida por:

$$D^2_{ii'} = \sum_{i=1}^n \sum_{i'=1}^n W_{ii'} d_i d_{i'}$$

em que:

$n$  : número de caracteres;

$W_{ii}$ : elemento da  $i$ -ésima linha e  $i$ ' coluna da inversa da matriz de variâncias e co-variâncias residuais entre os genitores;

$d_i$  : diferenças entre as médias do  $i$ -ésimo caractere nas duas populações ou indivíduos considerados;

$$\text{Matricialmente pode se escrever: } D_{ii}^2 = \delta' E^{-1} \delta$$

em que:

$\delta'$  = vetor linha  $[d_1, d_2, \dots, d_i]$ , sendo  $j, d_j = (X_i - X_j)$ , para cada  $i$ ;

$i$ : numero de caracteres estudados;

$E$  = matriz de variâncias e co-variâncias residuais entre variáveis originais.

$$\delta = \text{vetor-coluna } \begin{vmatrix} d_1 \\ d_2 \\ \vdots \\ d_i \end{vmatrix} \quad \text{sendo } d_i = (X_i - X_j), \text{ para cada } i;$$

A alternativa para o uso da distância de Mahalanobis é por intermédio da condensação Pivotal e por apresentar algumas outras propriedades úteis, como fornecimento de combinação linear, para a transformação do conjunto de variáveis correlacionadas  $X_j$  em variáveis independentes  $Z_j$ .

O processo de condensação pivotal consiste em justapor, à direita da matriz que se está operando, a matriz identidade; posteriormente transformam-se, por operação nas linhas, os elementos de cada coluna, de tal forma que esta tenha o elemento 1 na diagonal e 0 abaixo dela. A seqüência de elementos nas linhas da matriz justaposta a direita, após cada condensação, corresponde aos coeficientes de transformação linear das variáveis não-correlacionadas e o elemento da diagonal que foi transformado na unidade correspondente, à variância daquela variável não-correlacionada. Com isso tem-se:

$$D_{ii}^2 = \delta' \delta = \sum_{i=1}^n (Z_{ii}' - Z_{ii})^2$$

em que:

$Z_{ii}'$ : média da  $i$ -ésima variável, obtida pelo processo de condensação pivotal. As variáveis  $Z$ 's têm variâncias residuais iguais a 1,0 e são não correlacionadas.

Segundo revisões de Singh e Chaudhary (1979), uma vez realizada a transformação, o processo de estimação equívale ao descrito para componentes principais, além do mais, os autores comentam que essa distância é mais apropriada quando há correlação residual entre os caracteres em estudo.

O cálculo da distância generalizada pressupõe a normalidade multivariada, ou seja, uma distribuição multinormal  $J$  dimensional (em que  $J$  é o número de caracteres) e a homogeneidade da matriz de covariância das unidades amostrais (Rao et al., 1981), o que restringe seu uso. Quando estas suposições não são violadas, a distância generalizada toma-se uma ferramenta útil para taxonomistas numéricos.

Arunachalan (1981), em estudos sobre divergência genética avaliou muitos caracteres com graus significantes de correlação, desaconselhando o uso da distância Euclidiana. Nestes casos, para avaliar a distância genética deve-se adotar a  $D^2$ , que considera as correlações dos caracteres por meio de covariâncias residuais entre os caracteres e, para o cálculo de  $D^{2'}$ , pressupõe-se a distribuição normal multidimensional e homogeneidade da matriz de covariância das unidade amostrais.

Nas situações em que as variáveis a serem utilizadas numa análise de similaridade são obtidas de delineamentos experimentais apropriados, envolvendo casualizações e repetições dentro das unidades amostrais e, principalmente, quando os caracteres são correlacionados, a distância  $D^{2'}$  têm sido a mais recomendada.

Quando as correlações entre caracteres são baixas (tendendo a zero),  $D^{2'}$  será semelhante ao quadrado da distância Euclidiana, considerando as variáveis padronizadas (Cannings e Hoppensteadt, 1982). Isso também é verdadeiro quando todas as correlações são iguais (Riboldi, 1986).

Ferreira (1995) comenta que quando são identificadas correlações residuais significativas entre os caracteres, deve-se preferir a distância  $D^{2'}$ .



Porém, é preciso observar a distribuição normal multidimensional para os cálculos.

Outros autores citam a  $D^2_{ii}$  como a opção mais útil, além de fornecer uma analogia entre a análise de agrupamento e outras análises multivariadas. Desta forma, o primeiro eixo transformado terá a direção da maior variabilidade entre as médias de diferentes taxas. A segunda deverá ser ortogonal a primeira e ter inclinação na direção da segunda maior variabilidade. O mesmo ocorre com a terceira e subsequentes (Cannings e Hoppensteadt, 1982), sendo que a análise da coordenada principal da matriz  $D^2_{ii}$  é idêntica a obtida por análise de variáveis canônicas.

#### 1.3.8.2. - Distância Euclidiana

A medida de distância mais amplamente utilizada em taxonomia numérica é a Distância Euclidiana Simples porque considera o universo Euclidiano; o que é vantajoso em taxonomia numérica (Cannings e Hoppensteadt, 1982), por ser mais familiar e amplamente difundida, uma vez que é derivada do Teorema de Pitágoras (Sneath e Sokal, 1973).

O cálculo da mesma obedece às propriedades métricas e também ao teorema de Pitágoras. Assim, a distância entre a  $OTU_i$  e a  $OTU_j$ , em relação a dois caracteres, a e b, será dada pela expressão:  $d_{ij} = [1/n \sum_1 (X_{ia} - X_{ja})^2]^{1/2}$ ; padronizados. Neste caso, aceita-se que i caracteres são representados por eixos ortogonais (ou seja, formando ângulos retos um em relação ao outro). Entretanto, na prática, devido à correlação entre caracteres, isto não se justifica e a distância euclidiana não medirá a distância real entre as OTU's.

Deve-se ressaltar que tal medida apresenta o inconveniente de ser influenciada não somente pelo número e pela escala, mas pelas correlações não-genéticas entre as variáveis em estudo, quando a mesma é estimulada a partir de variáveis originais. Em função disso, os estados de carácter devem ser padronizados antes de calculá-la (Cannings e Hoppensteadt, 1982; Singh e

Chaudhary, 1979). A padronização é obtida por:  $Z_{ik} = \frac{X_{ik}}{\sigma(X_k)}$ ;

em que:

$\sigma(X_k)$  é o desvio padrão dos dados do carácter k;

$X_{ik}$ : média padronizada do acesso i em relação ao carácter j.

Por outro lado, o fato de a pressuposição da independência entre os caracteres é o maior inconveniente quando da utilização da distância Euclidiana, segundo Riboldi (1986) e Lopes (1984), a distância Euclidiana, apesar das restrições inconvenientes citadas, ainda é muito usada, devido ao fato de suas propriedades matemáticas ser bem conhecidas e estar inseridas na maioria dos softwares de análise de agrupamento.

Arumachalan (1981) comenta que a distância euclidiana deve ser utilizada quando as variáveis disponíveis são independentes, como se verifica no caso em que as unidades para cálculo são escores de componentes principais ou de variáveis canônicas. Entretanto, a estimação da distância euclidiana a partir das variáveis originais é inadequada, porque não leva em consideração a interdependência que normalmente existe entre estas.

Assim sendo, pode ser obtida por meio das observações individuais dos progenitores, sem a necessidade de experimentos com delineamentos estatísticos. No entanto, quando se dispõe de ensaios planejados, pode-se ignorar o delineamento e calcular a distância através das médias das repetições, mas Cruz e Regazzi (1994) desaconselham esse procedimento por ser menos preciso que aquele que se baseia na distância  $D^2_{ii}$ .

As medidas de similaridade entre unidades taxonômicas (referidas como unidades taxonômicas operacionais, OTU's,) dependerão do tipo do carácter estudado e da forma que esta informação está sendo codificada. Assim, as medidas de similaridade só consideram o fato se os caracteres são binários (utilizando para este caso o coeficiente de Jaccard, por exemplo), se os caracteres são qualitativos com mais de dois estados e, finalmente, se os

caracteres são quantitativos.

### **1.3.9 - Técnicas Multivariadas no Estudo da Divergência Genética**

Através de análise de variância multivariada é possível testar a hipótese de que o conjunto de acessos contenha alguma divergência genética considerando a variabilidade global existente nos múltiplos caracteres; segundo Rao et al. (1981) a técnica permite avaliar preliminarmente a existência de variabilidade genética entre os recursos genéticos, disponíveis..

Esta é semelhante a uma análise univariada, considerando, simultaneamente, todas as variáveis. Morais (1992) fazem referências destas técnicas, as quais podem ser aplicadas a qualquer tipo de delineamento experimental. A mesma possibilita mostrar a nível multivariado se os materiais avaliados são diferentes; em seguida, procura-se interpretar as possíveis causas para os padrões identificados, fornecendo subsídios para as posteriores análises de agrupamento. Por constituir-se numa análise unificadora, permitindo integrar as múltiplas informações de um conjunto de caracteres, extraídas das unidades experimentais, a técnica tem sido amplamente utilizada para quantificar a divergência genética, com maior oportunidade de escolha de parentais divergentes em programas de melhoramento e na caracterização de acessos nos BAG's.

Vale salientar que essa análise estatística é o ramo direcionado para estudo de amostras e distribuição multidimensional, ou seja, métodos apropriados nos estudos em que várias características são consideradas simultaneamente. O critério de Wilks tem sido utilizado predominante entre os testes existentes, para verificar se houve diferenças entre OTU's.

No entanto, apesar destas técnicas terem eficiência comprovada e proporcionarem enriquecimento das informações extraídas de dados experimentais, é necessária, para sua aplicação, a disponibilidade de recursos de

informática. Atualmente, o método tem-se tornado de fácil disponibilidade aos pesquisadores; anteriormente era limitado seu repasse.

Segundo Cruz e Regazzi, (1994), as técnicas mais utilizadas para estudo da divergência genética são: a) componentes principais (para informações sem repetições); b) variáveis canônicas (quando se dispõe de experimentos com delineamento); c) os métodos aglomerativos e de agrupamentos, cuja aplicação depende da aplicação de uma medida de dissimilaridade, previamente estudada; permite avaliar com isso a similaridade dos indivíduos por meio de uma dispersão gráfica em plano cartesiano bidimensional).

### 1.3.10 - Divergência Genética por Variáveis Canônicas

A análise multivariada com base em variáveis canônicas (vc's) trata de um processo alternativo para a avaliação do grau de similaridade genética entre acessos, que considera, tanto a matriz de co-variância residual quanto a de co-variância entre as médias fenotípicas dos caracteres avaliados.

A análise consiste em transformar o conjunto original de  $p$  caracteres por outro de  $n$  variáveis canônicas, de maneira tal que as primeiras variáveis possam concentrar a maioria das informações incluídas no conjunto original. A principal diferença para os componentes principais é considerar as possíveis diferenças na dispersão sobre as médias. Desta forma, tem-se a vantagem de manter o princípio da análise de agrupamento, adotando-se a  $D^2_{ii}$ , qual seja a de considerar as correlações residuais existentes entre a média dos acessos.

Cada vc' é uma combinação linear das variáveis originais; além disso, são independentes entre si e estimadas com o propósito de reter, em ordem de estimação, o máximo da informação em termos de variação total, contida nos dados iniciais. A técnica possui princípios semelhantes aos do processo de agrupamento baseados na distância  $D^2_{ii}$ . Hussaini, Goodman e Timothy (1977) referem-se à análise por vc's como uma seqüência de transformações ortogonais que visam a maximizar a razão da variação entre e dentre grupos de

acessos, e que a classificação dos grupos homogêneos por meio da análise possibilite a maximização da herdabilidade.

Nessa análise são requeridos diversos caracteres, que são medidos nos diferentes cultivares, visando a reproduzir a variabilidade total. De modo geral, parte dessa variação pode ser avaliada em menor número de vc's, mantendo quase a totalidade da informação existente quando os vários caracteres iniciais foram obtidos, à semelhança dos componentes principais (Cannings e Hoppensteadt, 1982).

A técnica pode reduzir os K caracteres a um novo conjunto, que são combinações lineares das variáveis originais e se tomam não correlacionadas entre si; as seguintes propriedades devem ser atendidas:

a) Se  $Y_{ii}'$  é uma variável canônica, então:  $Y_{ii}' = a_1X_{i1} + a_2X_{i2} + \dots + a_{mp}X_{in}$ ;

b) Se  $Y_{ii}''$  é outra variável canônica, então:  $Y_{ii}'' = b_1X_{i1} + b_2X_{i2} + \dots +$

$$b_{mp}X_{in};$$

$$e \sum_{i=1}^n \sum_{j=1}^n a_i a_j' \delta_{ij}' = \sum_{i=1}^n \sum_{j=1}^n b_i b_j' \delta_{ij}' = 1$$

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j' \delta_{ij}' = 0$$

em que:

$\delta_{ij}'$  é a covariância residual entre os caracteres

c) Entre todas as variáveis canônicas,  $Y_{i1}$  apresenta a maior variância,  $Y_{i2}$  a segunda maior e, assim, sucessivamente.

As variâncias de cada variável canônica e os coeficientes de ponderação dos caracteres podem ser estimados pela solução do sistema linear homogêneo:

$$(T - \lambda_i E) \delta_i = 0$$

..e a  $i$ -ésima variância, pela raiz característica de mesma ordem ( $\lambda_i$ ) obtida pela solução de:  $\det |(T - \lambda_i E)| = 0$

em que:

T : matriz de variâncias e covariâncias fenotípicas, provenientes da avaliação entre genótipos;

$\lambda_i$  : raízes características (autovalores) da matriz  $E^{-1}$

$E$  : matriz de variâncias e co-variâncias residuais;

$\delta_i$  : autovetor associado a cada estimativa das raízes características, cujos elementos são os coeficientes de ponderação;

As funções de transformação para os parâmetros descritos ao estabelecidos por um processo denominado de condensação pivotal, de modo que as novas variáveis apresentam co-variâncias residuais nulas e variância iguais a 1,0. Desta forma, sua importância é considerada em função do percentual da variância total que ela explica. Assim, a primeira  $vc'$  é a mais importante por reter maior quantidade da variação existente nos dados originais. A segunda  $vc'$ , a segunda maior porção, e assim sucessivamente (Hussaini, Goodman e Timothy, 1977).

Espera-se que a configuração do grupo para identificação dos acessos semelhantes ou não, que é realizada por intermédio de gráficos de dispersão Bi ou Tridimensionais, possa ser adequadamente representada pelos primeiros dois ou três vetores canônicos. Cruz e Regazzi (1994), entretanto, afirmam que este valor deve ser acima de 80% da variação total, para avaliar a dissimilaridade por meio de uma dispersão gráfica, considerando o estudo da divergência genética por meio das distâncias geométricas entre indivíduos.

Assim, a divergência que é influenciada a princípio por um conjunto de  $n$ -dimensionais ( $n$  = número de variáveis), passa a ser avaliada por um complexo bidimensional de fácil interpretação geométrica. Quando mais de 70% da variação entre os cultivares podem ser atribuídas às primeiras variáveis canônicas, os caracteres originais podem ser substituídos por estas variáveis (Ferreira, 1995).

De acordo com Ram e Panwar (1970) e Peter e Rai (1976), as variáveis que apresentam os mais altos coeficientes nas duas primeiras variáveis canônicas têm mais impacto sobre a diversidade observada. Sumarizando, a utilização desta técnica; Cruz (1990) salientou que as variáveis canônicas têm dois objetivos: a) resumir o número de caracteres em um grupo de variáveis

independentes entre si, de menor dimensão e que possuem uma fácil explicação biológica; b) avaliar a importância de cada caracter para verificar a viabilidade de seu descarte, através de seu coeficiente de correlação com cada  $vc'$ .

Davis (1985) avaliou a divergência genética entre 50 híbridos de feijão a fim de identificar genótipos a serem inter cruzados para estabelecer uma população de base genética ampla. Adotou apenas as duas primeiras  $vc'$ s, que explicavam 73% da variação total. Com isso, poderia eliminar 24 híbridos para o inter cruzamento, mantendo-se a mesma amplitude de variação genética do grupo inicial.

No caso do arroz, vários estudos e trabalhos de genética foram e estão sendo realizados com respostas coerentes, mostrando ser o arroz uma eficiente espécie para reproduzir os produtos amplificados e promissora para os estudos de genética molecular ( Moser e Lee, 1994).

### **1.3.11 - Importância Relativa dos Caracteres**

O emprego das  $vc'$ s ainda permite avaliar a importância relativa dos caracteres para o descarte daquelas variáveis que pouco contribuíram para os estudos, tratando-se de um processo alternativo para a avaliação do grau de similaridade genética entre indivíduos.

Em razão de se dispor da estimativa da matriz de co-variância residual estuda-se a importância relativas dos caracteres por meios das  $vc'$ s. Utilizando a sugestão de Singh (1981), a qual o produto do elemento do autovetor pelo desvio padrão residual é a medida relativa da importância do caracter. A identificação dos caracteres de menor importância pode ser realizada a partir das últimas variáveis canônicas.

Baseando-se no critério de identificar a importância relativa das  $vc'$ s decresce da primeira para a última a variável com maior coeficiente de ponderação no componente de menor autovetor; podem ser descartados, por apresentar relevância numa  $vc'$  de participação pequena, aqueles que implicam

em mínima variação total. O interesse em avaliar a importância relativa dos caracteres reside na possibilidade de descartar alguns que contribuam pouco para a caracterização, reduzindo, desta forma, mão-de-obra, tempo e custo na experimentação.

Esse fato foi comprovado entre nove caracteres estudados em milho e relatado por Cruz (1990), que identificou, pela ordem, os quatro que menos contribuíram para a divergência dos genótipos. Também Mardia et al. (1979), recomendam para descarte de caracteres esse tipo de critério na análise de componentes principais. Analogamente, estudos de Hussaini, Goodman e Timothy, (1977), em acessos de milheto permitiram o descarte de onze caracteres, entre vinte nove considerados, inicialmente, pela análise de Componentes Principais.

Os caracteres dispensáveis em estudos de divergência genética são: aqueles relativamente invariantes entre os indivíduos estudados, fortemente afetados pelo ambiente apresentando instabilidade com mudanças ambientais ou são redundantes, por estarem correlacionados a outros caracteres (Afifi e Clarck, 1984).

Todavia, as variáveis selecionadas e descartadas devem apresentar correlações significativas entre si, ou seja, as descartadas devem ser as redundantes; pois, em determinados casos, muitos caracteres são tomados sem a existência de estudo prévio da contribuição para a variabilidade (Morais, 1992).

Por outro lado, as variáveis selecionadas que mais contribuem para a divergência dos acessos podem ser identificadas pelas vc's e devem ter baixas correlações entre si, de forma tal que dada variável preservada na análise será responsável por um tipo de informação biológica exclusiva e a ação conjunta das mesmas será complementar para a descrição geral dos indivíduos estudados (Singh e Unrea, 1995). Os acessos são dispostos em gráficos bidimensionais, utilizando-se como eixos representativos as primeiras vc's respectivas.



A análise univariada é necessária para descobrir se há diferenças entre as OTU's e deve ser realizada para todos os caracteres morfológicos quantitativos avaliados, caso seja conduzida em delineamentos experimentais, enfatizando discriminar os descritores de maior potencial sob vários contextos, principalmente o genético, que significa maior chance de êxito na caracterização de indivíduos e de progenitores divergentes.

Jonhson e Wichem (1982) comentam que outros procedimentos, como: análise por componentes principais, coeficiente de parentesco (Malecót), existem para estimativa da divergência genética. Ainda auxiliam nesses estudos função discriminante é utilizada para verificar a adequação da partição obtida nos grupos (Gottlib, 1992.); medidas de dissimilaridade aplicadas à quimiosistemática, que segundo Johnson e Wichem (1987) e coeficiente de semelhança racial (CSR) desenvolvido para medir a semelhança entre amostras, que considera a variabilidade intraespecífica, de forma que médias e variâncias de caracteres contínuos (comprimento, largura, etc). A análise de covariância pode ser utilizada no agrupamento, devido à simplicidade de cálculo e à sua utilidade nos testes de significância.

### 1.3.12 - Métodos de Agrupamento dos Acessos

Realizada a análise de aglomeração, que é a obtenção de uma medida de distância, na qual é estimada para cada par de OTU's. A etapa seguinte é adoção de uma técnica de agrupamento para estabelecer os grupos (Bussab, Miazaki e Andrade, 1991). Como existem vários algoritmos, cabe ao pesquisador decidir qual o mais adequado ao seu propósito, uma vez que cada técnica pode levar a uma solução diferente. Segundo (Cruz, 1990), a utilização da técnica complementa as informações da dispersão gráfica e permite, desta forma, o estabelecimento de grupos menos subjetivos do que os verificados por exames visuais.

A divergência seria avaliada por uma estrutura classificatória ou de

reconhecimento da existência de grupos; dentro do subgrupo das “elites”. Com isso, permite encontrar outros subgrupos, mais homogêneos, a partir de algum critério preestabelecido de similaridade ou dissimilaridade. A estratificação além de auxiliar o processamento dos dados, é de grande praticidade na interpretação dos acessos e possibilitando considerável simplificação de dados na computação, sintetizando as informações disponíveis (Cruz 1990; Riboldi, 1986).

Neste tipo de análise, algumas questões podem ser levantadas dos indivíduos em grupos. Por exemplo: Qual a medida de dissimilaridade a ser utilizada? Qual o número de grupos desejados? Uma vez que tem-se um grande número de acessos, e ainda um grande número de variáveis que são avaliadas por algum critério de classificação.

Entre as muitas medidas de aglomeração que existem na literatura, distâncias e coeficientes predominam em estudos de sistemática. O coeficiente de genética, estudado por Nei, baseia-se na identidade de genes entre populações e mede as diferenças alélicas acumuladas por loco, a partir de um ancestral comum. Por isso é conhecida como uma medida biológica, enquanto que o segundo é o coeficiente de similaridade (S) de Rogers, que mede a distância geométrica entre as frequências alélicas que possui propriedades métricas (Peter e Rai, 1976).

Estes autores, fizeram o uso da técnica em coleções de germoplasma e verificaram que quando não se dispõe de informações sobre as introduções dos acessos, a análise torna-se muito importante na classificação de acordo com seus respectivos grupos (“pools”) gênicos.

Alternativamente, a técnica tem a finalidade de agrupar indivíduos com base em medidas tomadas de um conjunto de unidades experimentais de acordo com seus padrões de similaridade mútuas. Não pressupõe a existência de grupos e os diversos processos que envolvem essa ampla metodologia, a fim de

transformar um conjunto de dados amostrais em unidades heterogêneas.

Segundo isto, algum critério de similaridade ou dissimilaridade pelos caracteres considerados de maior importância, deve ser adotado após proceder a divisão em vários subgrupos, que caracterizam a homogeneidade interna e pela heterogeneidade externa ou entre grupos (Ferreira, 1995).

As técnicas de agrupamento mais comuns são os hierárquicos, que são subdivididos em: aglomerativos (os quais procedem uma série de fusões sucessivas de  $n$  OTU's em grupos) ou divisivos (quando formam grupos independentes entre si). Inicia-se a formação de  $n$  grupos, cada qual contendo um único indivíduo. Com isso, combina-se a formação de  $n-1$  grupos. Também existe os de otimização que separam um grupo de  $n$  OTU's sucessivamente em menores grupos (Cruz e Regazzi, 1994; Oliveira, 1992; Sneath e Sokal, 1973).

Os métodos de otimização, nos quais os grupos são formados pela otimização do critério que se adota manter a distância média intragrupo sempre inferior a qualquer distância intergrupos. Enquanto que, os grupos remanescentes são combinados para originar  $n-2$  grupos e, assim sucessivamente, até formar um grupo contendo os  $n$  indivíduos, os quais são subdivididos em aglomerativos.

Nesse tipo de agrupamento, usando algoritmo de menor distância, denomina método do vizinho mais próximo (single linkage method) e se a distância for máxima denomina-se vizinho mais distante (complete linkage method). O método da ligação média (average linkage), que podem ser ponderados ou não; de centróide também ponderado ou não (Sneath e Sokal, 1973), 'de variância mínima' e ainda possuem o método de Ward e os métodos divisivos, como os apresentados por Mardia et al., (1979). Porém, Sneath e Sokal (1973) afirmam que os resultados vão depender da confiabilidade dos dados.

Essas técnicas sofrem diferentes classificações. Segundo Riboldi

(1976), cinco grande grupos são os mais comentados: a - técnicas de hierarquização; b - técnicas de partição ou otimização; c - técnicas de densidade; d - técnicas de agrupamento com interseção; e - outras que não se enquadram nestas.

Bussab, Miazaki e Andrade (1991) comentam que a estrutura básica da aplicação de técnicas de análise de agrupamentos pode ser decomposta nas seguintes etapas: 1) definição de objetivos, critérios e escolha dos caracteres; 2) obtenção e tratamentos do dados; 3) escolha dos critérios de similaridade ou dissimilaridade; 4) adoção e execução de um algoritmo de análise de agrupamento dos indivíduos, com o mais alto grau de similaridade; 5) a presença dos resultados, com facilidade de visualizar em dispersões gráficas; 6) avaliação e interpretação dos resultados.

Deve-se salientar que essas etapas não são independentes. Às vezes, torna-se necessário voltar a etapas anteriores para correção e aprimoramento das etapas posteriores. Assim, numa primeira etapa apesar de existirem várias medidas de similaridade (Mardia et al., 1979) tem sido de uso mais rotineiro a distância euclidiana média ou a distância  $D^2_{ii}$ . A segunda, têm tido maior preferência em estudos de divergência genética, destinados à identificação de progenitores para cruzamentos.

#### **1.3.12.1 - Método do Vizinho Mais Próximo (VMP)**

Os métodos de agrupamento mais utilizados nestes estudos e de melhoramento são os hierárquicos, que se baseiam no princípio de que com  $n$  acessos. Empregam tanto matriz de similaridade quanto a de dissimilaridade entre as OTU's. Inicialmente, através da matriz de dissimilaridade genética, faz a identificação de duas OTU's mais similares (menor  $D_2$ ), as quais constituirão o primeiro grupo e mais semelhantes.

A seguir, calcula-se a distância desse grupo em relação aos demais indivíduos. Nas etapas sucessivas calcula-se a distância entre esse grupo e as

demais, ou a distância em relação a outros grupos. Num estágio mais avançado, cada estágio opera com uma matriz de similaridade ou distância de dimensão reduzida, até que finalmente todos os indivíduos estejam agrupados.

A distância entre um acesso  $k$  e um grupo formado pelos acessos  $i$  e  $i'$ , que é dada por:  $d(ii')_k = \min [dik; di'k]$ ,

em que:

$\min (dik; di'k)$  é a menor distância entre os grupos de acessos  $ik$  e  $i'k$ ;

A distancia entre dois grupos formados, respectivamente, pelos indivíduos  $(i$  e  $i')$  e  $(j$  e  $k)$  é fornecido pelo menor elemento do conjunto formado pelas distâncias entre os pares de acessos  $(i$  e  $j)$ ,  $(i$  e  $k)$ ,  $(i'$  e  $j)$  e  $(i'$  e  $k)$ .

Nos métodos hierárquicos aglomerativos, que são mais usados em taxonomia numérica, as OTU's são agrupadas por processo que repete em vários níveis até que seja estabelecido um dendograma ou diagrama de árvore, sem a preocupação com o número ótimo de grupos que formaram. O interesse é na árvore toda e suas ramificações, analisando-se os pontos de mudanças de níveis visualmente. Ao efetuar cortes nesses pontos do dendograma, estabelecem-se os grupos e os números de indivíduos de cada grupo.

### 1.3.12.2 - Método de Agrupamento de Tocher

Nos métodos de otimização, os grupos são estabelecidos pela adequação de algum critério de agrupamento que adote manter a distância média intragrupo (par de acessos) sempre inferior a quaisquer distâncias intergrupos. Diferem-se dos métodos hierárquicos, pelo fato de os grupos formados serem mutuamente exclusivos, ou sob contexto da teoria de conjuntos, em razão de subdividir o grupo original em subgrupos não vazios, cuja intercessão é nula e a união reconstitui o conjunto total.

No método de Tocher adota o critério de que a média dos valores de  $Dij'$  intergrupos deve ser menor que os valores de  $Dij'$  intergrupos. O método



requer inicialmente a obtenção da matriz de distancias generalizadas de Mahalanobis, sobre a qual é identificado o par de acesso mais próximo, ou seja, o que apresenta o menor valor de  $D_{ij}$ . Estes acessos darão origem ao primeiro grupo. A partir daí é avaliada a possibilidade de inclusão de novos acessos adotando-se o critério anteriormente citado (Jonhson e Wichern; Rao, 1952).

Inicialmente, para a formação do primeiro grupo, identifica-se o par de génotipos que apresenta o menor valor de distância ( $D^2_{ii'}$ ) na matriz de dissimilaridade entre os acessos. A partir desta etapa, avalia-se a possibilidade de inclusão de outros acessos no primeiro grupo, respeitando-se o critério citado anteriormente. Com a entrada de um novo indivíduo num grupo, aumenta o valor médio da distância intragrupo, e o valor máximo de  $D^2_{ii'}$  denominado  $\Theta$ ;

$\Theta$  corresponde ao maior elemento do conjunto de menores distâncias, em relação aos acessos. A inclusão, ou não, deste indivíduo no grupo será permitida se o acréscimo no valor da distância média intragrupo não ultrapassar um valor máximo permitido. Este valor que pode ser arbitrariamente estabelecido, ou pode corresponder ao valor máximo de  $D^2_{ii'}$ , obtido no conjunto de menores distâncias envolvendo cada par de indivíduos.

Assim as diferenças entre as técnicas de otimização são desenvolvidos em função dos métodos usados para uma partição inicial de critério de agrupamento utilizado para otimização. A inclusão de um acesso em um grupo é possível quando:

$$\frac{D^2(ii')}{\frac{k}{n}} < \Theta,$$

em que:

$D^2(ii')k$  é a distância entre o grupo  $ii'$  e o acesso  $k$ , obtida pela:

$D^2(ii')k = D^2_{ik} + D^2_{i'k}$ ;

$D^2_{ik}$  é a distância entre os acessos  $i$  e  $k$ ;

$D^2_{i'k}$  é a distância entre os acessos  $i'$  e  $k$ ;

$n$  é o número de acessos que constitui o grupo original.

### 1.3.12.3 - Outros Métodos de Agrupamento

Outros métodos de agrupamento foram desenvolvidos a fim de que sejam evitados os extremos propostos pelo VMP ou CLM. Desses, o mais comum é o de ligação não ponderada pela média aritmética entre os pares, que recebe uma sigla em inglês 'UPGMA'. O mesmo determina a similaridade ou dissimilaridade média de uma OTU em relação a um grupo, formulando peso igual para cada OTU daquele grupo, independentemente de sua subdivisão estrutural (Sneath e Sokal, 1973).

Do ponto vista biológico, a técnica do agrupamento considera duas situações: a primeira se relaciona com a estimação de uma medida de similaridade (ou dissimilaridade) entre indivíduos a serem considerados; a segunda, com a adoção de uma técnica de agrupamento para a formação dos grupos. Segundo Oliveira (1992), são usadas na elaboração de dendogramas no estudo de divergência genética em plantas ou em qualquer OTU (unidade taxonômica de organização).

Desta forma procede, até que seja estabelecido um dendograma de árvore. Neste caso, não há preocupação com o número ótimo de grupos, uma vez que o interesse maior está no dendograma e nas ramificações que são obtidas. Cruz e Regazzi (1994) comentam que as limitações podem estar na visualização dos pontos de mudança de nível, que define os acessos em cada grupo.

O dendograma que representa os métodos hierárquicos de indivíduos ou unidades amostrais, denomina-se diagrama de árvore. O mesmo mostra uma síntese dos resultados, tomando a informação mais fácil de ser manipulada e armazenada, com importância para comparação, classificação e discussão dos fenômenos biológicos (Camin e Sokal, 1965). O autor, que pelo dendograma obtido, além de mostrar alta variabilidade entre os materiais mostrou ser um método eficiente de seleção, e ainda um indicador de fontes desejáveis para os

trabalhos de melhoramento.

Uma questão que surge é o número ideal de grupos que, em alguns métodos pode ser adotado para equacionar os pares. Entretanto, é mais coerente utilizar maior número de grupos e por algum critério de otimização selecionar os mais convenientes. Ou adotar a análise da função discriminante para verificação da adequação da partição relutante. Outra alternativa mais simples e eficiente é a escolha pelo número de grupos na análise gráfica.



## REFERÊNCIAS BIBLIOGRÁFICAS

- ABBUD, N.S. Melhoramento genético do arroz de sequeiro *Oryza sativa L* no Estado do Paraná de 1975 a 1989. Piracicaba: ESALQ, 1991.141p. (Tese de Doutorado em Genética e Melhoramento de Plantas).
- AFIFI, A.A; CLARCK, V. Computer-aided multivariate analysis. California, Lifetime Learning Publications, 1984. 458p.
- AGROANALISIS, Produtos vegetais: arroz. FGV: Rio de Janeiro, n.12, v.6, p.2-5, 1998.
- ALFENAS, A.C.; PETERS, I.; BRUNE, W.; et al. Eletroforese de proteínas de fungos e essências florestais. Viçosa, UFV: 1989. 134p.
- Anuário estatístico do Brasil. Rio de Janeiro: FIBGE, v.53, 1998.
- ARUNACHALAN, V. Genetic distance in plant breeding. Indian Journal Genetics and Plant Breeding. New Delhi, v.2, n.41, p.226-236, Jun, 1981
- AUTRIQUE, E.; NACHIT, M.M.; MONNEVEUX, P.; et al. Genetic diversity in durum wheat based on RFLP, morphophysiological traits, and coefficient of parentage. Crop Science, Madison, v.3, n.36, p.735-742, May -Jun, 1996.
- BEER, S.C.; GOLFREDA, J.; PHILLIPS, T.D.; et al. Assessment of genetic variation in *Avena sterilis* using morphological traits, isozymes and RFLPs. Crop Science, Madison, v. 34, n.5, p.123-129, Set- Out, 1993.
- BERDING, N.; ROACH, B.T. Germoplasm collection, maintenance and use. In: HEINZ, D. (ed.) Sugarcane. Improvement Through Breeding, New York, Elsevier, 1987, cap.4, p.143-210.
- BERRIO, L.E.; CUEVAS-PERES, F.E. Cultivar differences in milling yields under delayed harvesting of rice. Crop Science. Madison, v. 29, n.6, p.1510-1512, Nov -dez, 1989.
- BRUMMER, E.C.; KOCHERT, G.; BOUTON, J.H. RFLP variation in diploid and tetraploid alfalfa. Theoretical and Applied Genetics, Berlin, v.83, n.1, p.89-96, Jan. 1991.
- BUSSAB, W.O.; MIAZAKI, E.S.; ANDRADE, D.F. Introdução a análise de agrupamentos. In: SIMPOSIO NACIONAL DE PROBABILIDADE E

- ESTAÍSTICA**, n.5, 1990. São Paulo. Anais... Piracicaba: Ceres, 1991. p.105.
- CAMARGO, O.B.A.** Estudo agrobotânico de culturas de arroz (*Oryza sativa* L), em diferentes épocas de semeadura. Piracicaba: ESALQ, 1982. 105p. - (Dissertação de Mestrado em Fitotecnia).
- CANNINGS, C.; HOPPENSTEADT, F.** An introduction to mathematical taxonomy. Cambridge: Cambridge University Pres, 1982. 152p.
- CENTRO INTERNATIONAL DE AGRICULTURA TROPICAL. CIAT,** Sistema de evaluación estandar para arroz. Programa de Pruebas Internacinalis de Arroz para America Latina. Cali, 1975. 62p.
- CHANG, T. T.** Manual on genetic conservation of rice germoplasm for evaluation and utilization. Manila, Philippines: The Internacional Rice Research Institute, 1976, 77p.
- CHANG, T.T.; BANDENAS, E.A.** The morfology and varietal characteristics of the rice plant. Los Baños: IRRI, 1965. 40p. (Technical Bulletin, 4)
- CORADINI, L.** CENARGEN: Sua ação na coleta de germoplasma. In: SIMPÓSIO DE RECURSOS GENÉTICOS VEGETAIS. Sazão I - Base atual de germoplasma. n.1, 1979. Resumos ...Brasília: CENARGEN/EMBRAPA, 1980. p.1-6.
- CRUZ, C. D.; REGAZZI, A. J.** Métodos biométricos aplicados ao melhoramento genético. Viçosa: UFV. 1994. Cap.6, p.287-324.
- CRUZ, C.D.** Aplicação de algumas técnicas multivariadas no melhoramento de plantas. Piracicaba: ESALQ, 1990. 188p. (Tese de Doutorado em Genética e Melhoramento de Plantas).
- CUEVAS-PERES, F.E.; GUIMARAES, E.P.; BERRIO, L.E.; et al.** Genétic base of irrigated rice in Latin America and the Caribbean, 1971 a 1989. *Crop Science*, Madison, v.32, n.4, p.1054 - 1059, Jul - Ag, 1992.
- DAVIS, J.R.C.** Cluster analysis as an aid to selection for diversity in dry beans. *Annual Report Bean Improvemenn*. Colombia, v.18, p.21-23, Out. 1985.
- de DATA, S. K.** Principles and practices of rice production. New York: John Wiley, 1981. 618p.

- DILDAY, R.H.** Contribution of ancestral lines in the development of new cultivars of rice. *Crop Science*. Madison, v.30, n.2, p.123-128, May - Apr, 1990.
- FAO, Organización de las Naciones Unidas para la Agricultura y la Alimentación, Roma, 1998, v. 51 (Boletim n. 142).**
- FERREIRA, R. F.** Análises biométricas da tolerância do arroz (*Oryza sativa* L.) à toxidez de alumínio. Viçosa: UFV, Imp. Univ., 1995. 123p. (Tese - Doutorado em Genética e Melhoramento de Plantas).
- 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. (Circular Técnica, 14).
- FORNASIERI FILHO, D.; FORNASIERI, E.** Manual da cultura do arroz. Jaboticabal: UNESP/FUNEP, 1993. 221p.
- 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, v.24, n.1, p.37-42, Jan - Feb, 1984.
- GIACOMETTI, D. C.; FONSECA, J.N.L.** Introdução, intercambio e quarentena de pós- entrada de germoplasmas. In: SIMPOSIO DE RECURSOS VEGETAIS, Sessão 1 - Bancos ativos de germoplasma. n.1, 1979. Resumos ...Brasília: CENARGEN/EMBRAPA, 1980. p15-18.
- GRODZICKER, J.R.C.; ADAMS, M.V.; HILL, R.R.** Multivariate analysis in black gram (*Vigna mungo* L.) Hepper). *Indian Journal of Genetics*. New Delhi, v.44, n. 1, p.243-47, 1992.
- HEISER, C. B.** Sementes para a civilização: a história da alimentação humana. Trad. de Sylvio Uliana. São Paulo: EDUSP, 1997. 253p.
- HU, J.; QUIROS, C.F.** Identification of broccoli and cauliflower cultivars with RAPD markers. *Plant Cell Report*. New York, v.10, n.1, p.505-511, Oct, 1991.
- HUSSAINI, GOODMAN E TIMOTHY, 1977, S. H.; GOODMAN, M. M.; TIMOTHY, D. H.** Multivariate analysis and the geographical distribution of the world collection of finger millet. *Crop Science*, Madison, v.17, n.6 p.257-263, Nov - Dez, 1977.

- INTERNATIONAL RICE RESEARCH INSTITUTE , IRRI, Annual Report, Manila: IRRI, 1975. 34p.**
- JACK, P.L.; MAYES, S. Use of molecular markers, for oil palm breeding. II - Use of DNA markers (RFLPs). *Oléagineux*, v.48, n.1, p.1-8, 1993.**
- JARADAT, A.A. Phenotypic divergence for morphological and yield-related traits among landrace genotypes of durum wheat from Jordan. *Euphytica*. Wageningen, v.3, n.52, p.155-164, Oct, 1991.**
- JENNINGS, P. R.; COFFMAN, W. R.; KAUFFMAN, H. E. El mejoramiento del arroz. In : TASCÓN, E. J.; GARCÍA, E. D. (eds ). *Arroz: investigación e producción*. Cali: CIAT, 1985, p. 205-232.**
- JOHNSON, R. A.; WICHERN, D.W. Applied multivariate statistical analysis. Englewood Cliffs: Prentice-Hall, 1982. 593p.**
- KAW, R.N. Analysis of divergence in some cold-tolerant rices. *Indian Journal of Genetics and Plant Breeding*. New Delhi, v.55, n.1, p.84-89, Jan, 1995.**
- KEIM, P.; DIERS, B.W.; OLSON, T.C.; et al. RFLP mapping in soybean: Association between markers loci and variation in quantitative traits. *Genetics*. New York, v.126, n. 1, p.735-42, Jan, 1990a.**
- KONGKIATNGAM, P.; WATERWAY, M.J.; FORTIN, M.G.; et al. Genetic variation within between two cultivars of red clover (*Trifolium pratense L.*): comparisons of morphological, isozyme and RAPD markers. *Euphytica*. Netherlands, v.84, n. 1, p.237-246, Jul - Ag, 1995.**
- KUMARI, R.U.; RANGASAMY, P. Studies on genetic diversity in international early rice genotypes. *Annals of Agricultural Research*. V.18, n.1, p.29-33. Jan, 1997.**
- LIANG, C.Z.; GU, M.H.; PAN, X.B; et al. RFLP tagging of a new semidwarfing gene rice. *Theoretical and Applied Genetics*, Berlin, v.88, n.6-7, p.898-900, Oct, 1994.**
- LOPES, A. M. Análise genética dos componentes de produção num dialeto entre seis cultivares de arroz (*Oryza sativa L.*) em dois regime hidricos. Viçosa, UFV, Imp. Univ., 1984. 135p. (Tese - Doutorado em Genética e Melhoramento de Plantas).**

- MACKILL, D.J.; LEI, X. Genetic variation for traits related to temperature adaptation of rice cultivars. *Crop Science*. Madison, v.37, n.4, p.1340-1346, Jul-Ag, 1997.
- MARDIA, K.V.; KENT, J.T.; BIBBY, J.M. *Multivariate analysis*. London: Academic Press, 1979. 512p.
- MARTIN, J.M.; TALBERT, L.E.; LANNINH, S.P.; et al. Hybrid performance in wheat as related to parental diversity. *Crop Science*, Madison, v.1, n.35, p.104-108, Jan - Feb, 1995.
- MAZURIER, J.R.C.; GODSHACH, B.; TIMOTHY, D.H. Factor and principal component analysis an alternatives to index selection. *Theoretical and Applied Genetics*. Berlin, v.76, n. 2, p.352-60, Mar, 1994.
- MOSEER, H.; LEE, M. RFLP variation and genealogical distance, multivariate distance, heterosis, and genetic variance in oats. *Theoretical and Applied Genetics*. Berlin, v.87, n.8, p.947-956, Dez, 1994.
- MORAIS, O.P. de. *Análise multivariada da divergencia genética dos progenitores, índices de seleção e seleção combinada numa população de arroz oriunda de intercrossamento, usando macho-esterilidade*. Viçosa: UFV, 1992. 251p. (Tese de Doutorado em Genética e Melhoramento de Plantas).
- MORGANTE, M.; OLIVIERI, A. M. PCR-amplified microsatellites as markers in plant genetics. *The Plant Journal*, n3, p.175- 182, 1993.
- MUEHLBAUER, G.J.; STASWICK, P.E.; SPECHT, J.E.; et al. RFLP mapping using nearisogenic lines in the soybean (*Glycine max* (L) Merr.). *Theoretical and Applied Genetics*, Berlin, v.81, n.1, p.189-198, Jul, 1991.
- OLIVEIRA, G. C. X. *Padrões de variação fenotípica e ecológica de Oryzae (Poaceae) selvagem da Amazonia: Piracicaba: ESALQ, 1992. 303p. (Dissertação de Mestrado em Genética e Melhoramento de Plantas)*.
- PATERSON, A.H.; TANKSLEY, S.D.; SORRELLS, M.E. DNA markers in plant improvement. *Advance Agronimics*. New york, v.46, p.39-90, 1991.
- PEDROSO, B.A. *Arroz Irrigado: obtenção e manejo de cultivares*, Porto Alegre: Ed. Sagra, 1982. 175p.

- PETER, K. V.; RAI, B. Genetic divergence in tomato. **Indian Journal of Genetics and Plant Breedings**. New Delhi, v.36, n.3, p.379-83, 1976.
- RAFALSKI, J.A.; TINGEY, S.V.; WILLIAMS, J.G.K. RAPD markers - a new technology for genetic mapping and plant breeding. **Agrow Biotechnology News Information**, v.3, n.4, p.645-648, 1991.
- RAM, J.; PANWAR, D.V.S. Intrespecific divergence in rice. **The Indian Journal of Genetics e Plant Breeding**, New Delhi, v.30, p.1-10, 1970.
- RAO, A. V. A.; PRASAD, A. S. R.; SAI KRISHNA, T.; et al. Genetic divergence among some brown hopper resistant rice varieties. **Indian Journal of Genetic and Plant Breeding**, New Delhi, v.2, n.2, p.179-185, 1981.
- RIBOLDI, J. Análise de agrupamento "cluster analysis" e suas aplicações. Piracicaba: USP/ESALQ, 1986. 33p (monografia) 1986.
- ROCHA, S. B. Produção de semente híbrida de arroz. **Lavoura Arrozeira**, Porto Alegre, v. 40, n.736, p.3-6, nov./dez. 1987.
- SAIKI, R.K.; SCHARF, S.; FALOONA, F.; et al. Enzymatic amplification of beta-gliobin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. **Science**. Londres, v.230, p.1350-1354, 1985.
- SILVA, E. F. Caracterização de linhagens mutantes de arroz de sequeiro obtidas com a utilização de azida sódica (NaN<sub>3</sub>). Piracicaba: ESALQ, 1994. 57p. (Dissertação de Mestrado em Genética e Melhoramento de Plantas).
- SINGH, A.K.; SINGH, S.B.; SINGH, S.M. Genetic divergence in scented and fine genotypes of rice (*Oryza sativa* L.). **Annals of Agricultural Research**. v.17, n.2, p.163-166, 1996.
- SINGH, D.; UNREA, A.C.; Inter and intraracial hidridization and selection for seed yield in early generations of common bean, *Phaseolus vulgaris* L. **Euphytica**, Wageningen, v.81, n.21, p.131-137, Oct, 1995.
- SINGH, M.; CECCARELLI, S.; HAMBLIN, I. Estimation of heritability from varietal trials data. **Theoretical and Applied Genetics**, Berlin, v.86, n.2, p.437-441, Jul, 1993.

- SINGH, R. K.; CHAUDHARY, B. D. **Biometrical methods in quantitative Genetic analysis**, New Delhi, Kayami Publishers, 1979. 304p.
- SNEATH, P.; SOKAL, R.R. **Nemerial taxonomy:the principles and practice of numerical classification**. San Francisco: W.H. Freeman, 1973. 513p.
- SOUZA, N. R. **Divergência Genética e correlação em cultivares de arroz (*Oryza sativa* L.) em diferentes níveis de competição com plantas daninhas**. Viçosa: UFV, 1990. 83p. (Dissertação de Mestrado em Genetica e Melhoramento de Plantas).
- SOUZA, E.; SORVELIS, M.E. **Inheritance and frequency of a nullallele for diaphorase activity in North American oat cultivars**. **Journal Hereditary**. Bektimore, v.80, n.2, p.501-503, Jul - Ag, 1989.
- STUBER, C.W. **Biochemical and molecular markers in plant breeding**. In:DUDLEY, J.W.; HALLAUER, A.R. e RYDER, M (eds) **Plant breeding reviews**. N. York:John wiley. 1992. 132p.
- TINGEY, S.V.; DEL TUFO, J.P. **Genetic analysis with Random Amplified Polymorphic DNA markers**. **Plant Physiology**, v.101, p.349-352, 1992.
- TINKER, N.A.; FORTIN, M.G.; MATHER, D.E. **Randon amplified polymorphic DNA and pedigree relationships in spring barley**. **Theoretical Applied Genetics**. Berlin, v.85, n.3, p.976-984, Jul, 1993.
- VALLEJOS, C.E.; SAHIYAMA, N.S.; CHASE, C.D. **A molecular marker-based linkage map of *Phaseolus vulgaris* L**. **Genetics**. New York, v.131, n. 5, p.733-40, May, 1992.
- VALLS, J.F.M. **Caracterização morfológica, reprodutiva e bioquímica de germoplasma vegetal**. In: ENCONTRO SOBRE RECURSOS GENÉTICOS, I, Jaboticabal, 1988. **Anais...** Jaboticabal: UNESP/FCAVJ, 1989, p.106-128.
- VAN BEUNINGEN, L.T.; BUSCH, R.H. **Genetic diversity among North American Spring wheat cultivars. I. Analysis of the coefficient the parentage matrix**. **Crop Science**. Madison, v.37, n.2, p.479-483, Mar -Apr, 1993.
- VAUGHAN, D.A. **The genus *Oryza* L.: current status of taxonomy**. Manila: IRRI, 1989. 21p. (Research Paper Series, 138).
- WAUGH, R.; BAIRD, E.; POWELL, W. **The use of RAPD markers for the**

detection of gene introgression in potato. **Plant Cell Report**. New York, n.11, n.5, p.466-69. Jan, 1992.

**WEEDEN, N.F.; MUEHLBAUER, FJ.; LADIZINSKY, G.** Extensive conservation of linkage relationships between pea and lentil genetic maps. **Journal of heredity**. Baltimore, v.83, n.1, p.123-129, Jan, 1992.

**WELSH, J.; McCLELLAND, M.** Fingerprinting genomes using PCR with arbitrary primers. **Nuclear Acids Researchs**, v.18, n.24, p.213-218, 1990.

**WILLIAMS, J. G.K.; KUBELIK, A.R.; LIVAK, K.J.; RAFALSKI, J.A.; et al.** DNA polymorphism amplified by arbitrary primers are useful as genetic markers. **Nucleic Acids Research**. v.18, p.6531-35, 1990.

**YOSHIDA, S.** **Fundamentals of rice crop science**. Los Bãnos: IRRI, 1981. 269p.

**ZABEAU, M.** Selective restriction fragment amplification: a general method for DNA fingerprinting. **European Patent Application**, v.53, p.48-58. 1993.

**ZHU, J.; GALE, M.D.; QUARRIE, S.; JACKSON, M.T.; et al.** AFLP markers for the study of rice biodiversity. **Theoretical and Applied Genetics**. Berlin, v.85, n.3, p.976-984, Jul, 1993.



## CAPITULO 2 -

### **Emprego de Marcadores Morfológicos no Estudo da Divergência Genética em Acessos de Arroz Denominados Três Meses**

#### **RESUMO**

A divergência genética foi estimada utilizando-se 36 acessos de arroz de nome regional 'três meses', coletados em áreas onde é tradicional o cultivo de arroz e não é prática comum a aquisição de sementes certificadas por parte dos agricultores. O objetivo desse estudo foi agrupar os acessos similares, após analisados por meio da relação de 23 caracteres morfo-agronômicos (11 qualitativos e 12 quantitativos). Foram empregadas, como técnicas estatísticas para cálculo da distância genética, a análise multivariada, a distância generalizada de Mahalanobis, variáveis canônicas, adotando-se duas técnicas de agrupamentos de Tocher e Vizinho mais próximo como medida de dissimilaridade para construir os dendrogramas de agrupamento. Foi verificada a existência de divergência genética entre os acessos. Ao nível de 50% de divergência formaram-se seis grupos distintos, com muitos acessos no primeiro grupo. Por isso preferiu-se o agrupamento pelos subgrupos dentro de alguns grupos. Por este procedimento pode agrupar-se aqueles mais similares e que não mostraram alta associação com as procedências. Os caracteres que mais contribuíram para a divergência genética foram altura da planta, pesos de panículas e 100 grãos, número de dias até florescimento e ciclo de maturação das sementes. As correlações entre produção de grãos e os demais caracteres mostraram coerência dos dados.

**Palavras Chaves:** Variabilidade genética, análise multivariada, agrupamento, poaceae e *Oryza sativa*, cereal, Poaceae

## **Assessment of Morphological Markers in The Study of Genetic Divergence in Rize Accessions**

**ABSTRACT** - The genetic divergence was evaluated by utilizing 36 rice accessions with common name of the 'three months', in material that was collected in areas where rice growing is traditional and the purchase of certified seeds isn't common practice on the part of the farmers. The objective of this study was to group together the similar accessions after analysed by means of the ratio of 23 morpho-agronomic characters ( 11 qualitative and 12 quantitative). As statistical techniques for the calculation of the genetic distance , were used multivariate analysis, Mahalanobis generalised distance and canonical variable, by adopting two clustering techniques: of Tocher groupment and of the the more adjacent neighbour as a dissimilarity measure to build the clustering dendograms. The existence of genetic divergence among the accessions was verified . At the level of 50% of divergence, six distinct groups were formed with many accession in the first group. Therefore, the clustering by the subclusters within some groups was preferred . Through this procedure , one can cluster together those ones more similar and that did not show a high association with the origins . The characters that most contributed to the genetic divergence were plant size, weight of panicles and of 100 grains , number of days to flowering and to maturation of seeds. The correlation between grain yield and the other characters showed data consistency.

**Index Terms:** Genetic variability, multivariate analysis, cluster, poaceae, cereal and *Oriza sativa*

## 2.1 - INTRODUÇÃO

O arroz é o principal alimento no mundo, em termos de quantidade produzida e consumida. No Brasil, além de constituir o alimento básico, caracteriza-se por ser cultivado por um enorme grupo de agricultores, estando presente na maioria das propriedades de pequeno e médio porte, contribuindo para a dispersão das cultivares existentes (Abbud, 1991). Com a introdução de cultivares modernas descendentes daquelas que possuem gene para o nanismo, os produtores vão abandonando aquelas que estão em cultivo, tornando, assim, recursos genéticos que uma vez coletado são conservados em BAG's de arroz, necessitando de avaliação e caracterização para aferir suas propriedades intrínsecas e mantê-lo como um acesso da espécie, que irá sofrer manutenção correta para conservar esse recurso em coleções de arroz.

O conhecimento da diversidade genética pode auxiliar na caracterização e indicação de progenitores e desenvolvimento de novos genótipos, uma vez que se a diversidade for insuficiente, os genitores utilizados em cruzamentos reduziriam a variabilidade genética dos caracteres; e em razão disso, o progresso das pesquisas com as características selecionadas poderia ser limitado (Singh, Singh e Singh, 1996; Lopes, 1984; Yoshida, 1981).

Grande parte das características de interesse econômico é complexa, poligênica e a bioquímica de difícil definição. São os chamados caracteres quantitativos, extremamente sujeitos à variação ambiental, geralmente estudadas por métodos biométricos que avaliam a magnitude das variâncias genéticas e ambientais e suas interações, o que torna necessário um grande número de marcadores para os estudos, embora tenha, nos longos anos, sustentado o melhoramento clássico, mas demandando tempo e custo (Newbury e Ford-Lloyd, 1993; Waugh, Baird e Powell, 1992).

A pesquisa hoje conta com metodologias capazes de estimar a variância e seus componentes, os parâmetros genéticos e ambientais, podendo com isso estudar corretamente delineamentos especiais, aplicação da genética molecular na caracterização e no mapeamento das espécies.

A utilização de marcadores morfológicos para estudo da divergência genética em arroz é uma prática bastante empregada nos Centros Internacionais e tem mostrado informações coerentes sobre o grau de parentesco de acessos conservados entre indivíduos de uma população e as correlações entre caracteres que são de grande valia para um programa de melhoramento de plantas. No entanto, no Brasil não há tanto enfoque ainda, o que às vezes dificulta as comparações entre certos grupos de acessos ou mesmo cultivares nativas (Cuevas-Peres et al., 1992).

Esses marcadores morfológicos (descritores), os quais, após eleitos, devem ser os que, em menor número, permitam a melhor discriminação) por (Valls, 1989). Os marcadores morfológicos ideais refletem marcadores genéticos com segregação mensurável, ainda que os fenotípicos de fácil visualização estão relacionados com a reprodução ou produtividade sejam úteis. Adicionalmente estudou-se a importância relativa das características na predição da divergência genética (Kaw, 1995).

Nesse contexto há um grupo de espécies com mais privilégio, pois já existem para elas manual internacional de caracterização e listas de descritores adaptados à variabilidade presente nas diferentes regiões, sabendo que, a conservação é melhor verificada por um processo mínimo de ciclos de cultivo com as sementes rejuvenescidas sob um ambiente semelhante e mais estático para possibilitar que a população se manifeste nesse habitat diferente do nativo. Por outro lado, a evolução e utilização necessita da identificação dos recursos

genéticos disponíveis nos organismos para evidenciar o proveito de um genótipo envolvendo seleção e purificação (Santos, 1996; Hillis, 1987; Campbell e Kerne, 1982).

A divergência genética entre os acessos pode ser determinada pela análise das variáveis canônicas, que adota como critério identificar a importância dos descritores a partir dos maiores coeficientes de correlação dos descritores originais com as últimas variáveis canônicas, aquelas de menor importância, ou seja, aquelas que implicam em mínima variação total (Van Beuningen e Busch, 1993).

Isso foi comprovado entre nove caracteres em milho, cujos quatro caracteres que menos contribuíram para a divergência dos genótipos podem ser descartados. Também Smith e Smith (1989); Mardia et al. (1979) recomendam para descarte de caracteres este tipo de procedimento, na análise de componentes principais. Enquanto que, a análise de agrupamentos é empregada para classificar o grau de estabilidade genotípica, com base na similaridade do comportamento fenotípico, que é delimitada pela distância estatística.

Diversos trabalhos que visam à determinação da divergência genética de cultivares e/ou genótipos de arroz, trigo, cevada, soja e outras culturas têm sido relatados, na forma de estudos, envolvendo esses mesmos aspectos, destacando-se: análise de variância multivariada e análise com base em variáveis canônicas (Martin et al., 1995; Jaradat, 1991, Smith e Smith, 1989).

A análise de agrupamento também tem tido aplicação em Bancos de Germoplasma. Nesse contexto, o uso da análise de agrupamento em coleções de germoplasma foi realizado por Smith e Smith (1989); Peter e Rai (1976), os quais verificaram que, quando não se dispõe de informações sobre as introduções, a análise torna-se muito importante na classificação dos acessos dentro de seus

respectivos reservatório (“pools”) gênicos, uma vez que a divergência seria avaliada dentro do subgrupo das elites.

Com isso é possível identificar outros subgrupos, além de auxiliar na interpretação das informações sobre os acessos e permitir uma considerável simplificação dos dados na computação, sintetizando as informações disponíveis (Kim e Ward, 1997; Autrique et al., 1996, Gupta et al., 1996)

No entanto, quando existe um grande número de acessos, inúmeras variáveis; sendo avaliadas, as análises de aglomerados (agrupamento, variáveis canônicas ou componentes principais) para identificação dos indivíduos ou dispersão ficarão prejudicadas. Adota-se então um processo das técnicas de hierarquização para delimitação dos grupos (método do vizinho mais próximo) e ainda a técnica de otimização proposta por Tocher, para caracterização dos acessos, de um grupo original que é dividido em subgrupos pelos caracteres considerados de maior importância, para posterior divisão (Kim e Ward, 1997; Gupta et al., 1996; Murphy e Phillips, 1994).

Em vista disso, o objetivo desse trabalho foi avaliar a divergência genética por meio de caracteres morfológicos quantitativos e qualitativos, através de técnicas de análise multivariada, em 36 acessos de arroz considerados como CV's Obsoletas e que estão sendo conservados (*ex-situ*) no BAG/Embrapa Arroz e Feijão por mais de uma década, as quais foram coletadas em vários Estados e possuem nomes regionais semelhantes (homônimas).

## **2.2 - MATERIAL E MÉTODOS**

### **2.2.1 - Localização, Instalação e Condução do Experimento**

O ensaio de multiplicação de sementes e caracterização de acessos de arroz foi conduzido no campo experimental da Fazenda Experimental

“Capivara”, pertencente a Embrapa Arroz e Feijão, localizado no Município de Santo Antônio de Goiás, GO, tendo as coordenadas geográficas: 16° 28' Lat. Sul, Long. 49° 17' W. Grw, altitude de 823,77 m na Estação Climatológica. O clima, pela classificação de Köppen é Aw, tropical de savana, megatérmico. O solo predominante na área é o LVE, textura argilosa fase Cerradão subperenifólio, relevo plano. O preparo do mesmo seguiu o convencional, constituindo-se de uma aração seguida de gradagem 60 dias antes do plantio e de uma gradagem de nivelamento para facilitar as operações, às vésperas da instalação dos experimentos.

O plantio sob o regime de cultivo de verão, na safra 94/95, foi efetuado com suplemento de água por aspersão convencional, pH corrigido, adubado com 600 kg/ha da seguinte formulação 4-30-10, mais 40 kg/ha de FTE, BR-12, após análise do mesmo. A semeadura aconteceu em único dia (29/11/1994), com solo úmido, no espaçamento de 40 cm entre linhas na densidade de 60 sementes por m. Duas adulações de cobertura (20 e 50 dias ) foram realizadas com sulfato de amônio, 50 kg de N/ha. O controle de plantas daninhas foi com o uso de um herbicida seletivo pré-emergente dois dias pós-plantio, e três capinas durante o ciclo das plantas. O controle de doenças foi preventivo com o uso de produtos químicos (fungicidas e cupinicida) apenas nas sementes. Colheita manual panicula por panicula quando 85-95% dos grãos estavam maduros.

O delineamento experimental utilizado foi em blocos casualizados (DBC), com três repetições. Os tratamentos constaram de 36 acessos de arroz tidas como CV's Obsoletas e praticamente homônimas, as quais foram coletadas em vários Estados do Brasil (Tabela 2). As parcelas experimentais foram compostas por 3 fileiras de 3,0 m de comprimento por 0,40 m de largura,

**TABELA 2 - Relação dos Acessos de arroz denominados de três meses, para estudo do grau da divergência genética. CNPAF, Goiânia -GO, 1998.**

N. ordem	N. Protocolo	Nome popular	Procedência
A01	CA760016	90 dias	SP
A02	CA760020	Branco três meses	SP
A03	CA760021	Três meses branco	SP
A04	CA760100	Três meses branco	SP
A05	CA760353	Três meses branco	RR
A06	CA810072	Arroz 90 dias	GO
A07	CA820080	Três meses branco	AC
A08	CA820088	Três meses branco	AC
A09	CA820090	Arroz três meses	AC
A10	CA830044	Arroz três meses	*
A11	CA850049	Branco três meses	MS
A12	CA860075	Três meses	GO
A13	CA860082	Três meses branco	GO
A14	CA860096	Três meses	GO
A15	CA860098	Arroz três meses	GO
A16	CA860099	Arroz três meses	GO
A17	CA860105	Arr. três meses antigo	*
A18	CA870002	Arroz três meses	*
A19	CA870022	Três meses branco	GO
A20	CA870046	Três meses branco	*
A21	CA870051	Três meses	GO
A22	CA870052	Três meses	GO
A23	CA870055	Três meses antigo	GO
A24	CA870066	Três meses precoce	GO
A25	CA870089	Três meses branco	GO
A26	CA870093	Arroz três meses	MG
A27	CA870099	Três meses branco	MG
A28	CA870111	Três meses Ipiaçú	MG
A29	CA870114	Branco três meses	MG
A30	CA870145	Precoce três meses	GO
A31	CA870148	Arroz três meses	*
A32	CA880056	Três meses de abril	MT
A33	CA880063	Arroz três meses	MT
A34	CA880065	Três meses branco	MT
A35	CA880070	Três meses branco	MT
A36	CNA-6613	Três meses branco	*

\* Acessos sem informação das procedências geográficas (coletas).



sendo a área útil da parcela de composta de 1 linha central, totalizando 2,0 m<sup>2</sup> de área, eliminando 0, 5 m em ambos lados.

As avaliações seguiram o que é preconizada pelo CIAT para ensaios de arroz, com avaliação de diferentes materiais coletados (CIAT, 1975). Os marcadores morfológicos qualitativos (Tabela 6) e quantitativos foram avaliados em dez plantas da parcela.

Foram avaliados também os seguintes marcadores quantitativos, de acordo com a recomendação de (Embrapa 1977).

- 1- Número de dias para a emergência (DEm); período em dias gasto pelas sementes emergirem do solo, após o plantio em solo úmido, o que está relacionado com o grau de dormência em arroz.
- 2- Número de dias até o florescimento (DAF), que foi tomado da data da germinação das plântulas até a data em que mais de 50% das espiguetas floresceram (início da antese). Este estágio envolve as fases vegetativa e reprodutiva, sendo a primeira fase bastante variável e a segunda com duração de cerca de 35 dias.
- 3- Ciclo total da planta para maturidade (CTP), compreende o período entre a germinação e a maturação completa, mais de 85 - 90% das espiguetas estavam maduras. Com este dado determina-se o ciclo do genótipo. De acordo com a classificação proposta pela Embrapa(1977), os materiais são divididos nos seguintes grupos:
  - Precoce Até 105 dias e se .Semi-precoce de 106 a 120 dias
  - Média de 121 a 135 dias e se: .Semi-tardia de 136 a 150 dias, e como, Tardia acima de 150 dias.

- 4- Altura da planta na maturidade (APM), A medida da altura da planta foi tomada do nível do solo até a extremidade da panícula do colmo mais alto, na época da colheita e expressa em centímetros.
- 5- Comprimento da “folha-bandeira” (CFB), foi medida do limbo, da lígula ao ápice da folha bandeira, à época da colheita, em centímetros, ou seja, distanciado no inferior à extremidade da raquis principal, em centímetros. Este caráter é importante por estar relacionado à produtividade.
- 6- Largura da “folha-bandeira” (LFB), refere-se a medida, em centímetros, na região de maior largura do limbo da folha bandeira, à época da colheita.
- 7- Comprimento da panícula (CPa), medida em cm do último nó até o ápice da última espiguetta.
- 8- Peso da panícula (PPa), cada foi pesada em gramas, com todas as partes e os grãos chocho.
- 9- Peso de cem grãos (P100), peso médio de cinco amostras de 100 grãos com casca, corrigido para 13% de umidade, em balança de precisão de centígrama.
- 10- Produção de grãos por parcela (Prod), determinada em gramas, pela pesagem do total de grãos colhidos na parcela útil, após limpeza e secagem uniforme, corrigida para 13%.
- 11- Comprimento do grão Sem casca (CGS), após descascar 10 grãos de arroz procedeu-se a medida com paquímetro, de uma extremidade a outra extremidade do grão, em mm
- 12- Largura do grão Sem casca (LGS), com uso do paquímetro procedeu a leitura no mesmo de anterior da largura em mm

### 2.2.2 - Análise Estatística dos Dados

Os dados obtidos foram submetidos a quatro tipos de tratamentos estatísticos, mostrados detalhadamente em seguida..

### a- Análise Univariada

Foram realizada análise de variância univariada nos marcadores quantitativos com a finalidade de identificar diferenças entre acessos para cada um dos caracteres avaliados. As análise seguiram o seguinte modelo linear.

$$Y_{ij} = \mu + b_j + A_i + \varepsilon_{(ij)}$$

em que:

$Y_{ij}$ : valor fenotípico observado, no  $i$ -ésimo acesso da  $j$ -ésima repetição;

$\mu$  = média geral do caracter;

$b_j$  = efeito do  $j$ -ésimo bloco, com  $j = 1,2,3$ ; fator ao acaso;

$A_i$  = efeito fixo atribuído ao  $i$ -ésimo acesso, com  $i = 1,2, \dots, 36$ ;

$\varepsilon_{ij}$  = erro experimental associado à observação  $Y_{ij}$ , considerados

independente e normalmente distribuídos com média zero e variância constante.

O esquema está da análise de variância (Tabela 3) mostra as fontes de variações, grau de liberdade, quadrados médios e esperança matemática de quadrados médios, apresentando na Tabela 3.

TABELA 3 - Esquema da análise de variância univariada para os caracteres quantitativos avaliados em 36 acessos de arroz com nomes regionais 'três meses'. CNPAF, Goiânia, GO, 1998.

Fontes de variação	GL	QM	E(QM)
Bloco	$j - 1$	QMB	
Acessos	$j(i-1)$	QMA	$\delta^2 + 3\delta^2 t$
Resíduo	$(j-1)(i-1)$	QMR	$\delta^2$
Total	$I(j-1)$	-	

sendo que:

QMB: quadrado médio de blocos;

QMA: quadrado médio de acessos;

QMR: quadrado médio do resíduo;

$\mu$  : média do caracter em questão;

$\delta$  = desvio padrão da média;

O coeficiente de variação experimental (CVe) foi determinado pela seguinte equação:  $CVe(\%) = \sqrt{\delta} / \mu \cdot 100$ . Enquanto que, o CVg (coeficiente de

variação genética) entre os 36 acessos avaliados foi estimado através da expressão:

Ainda foi avaliado o índice (I) que calcula o máximo valor da relação entre coeficiente de variação genética e experimental, ou seja:  $I = CVg/ CVe$ .

Variabilidade genética de um caractere, em função da variabilidade ambiental, que se torna importante na avaliação de indivíduos, pois é calculado pela menor variação residual. O CVg é um valioso indicador da grandeza relativa das possíveis mudanças que podem ser conseguidas com cada caractere, por meio de uma seleção.

#### **b - Análise Multivariada**

Neste procedimento de análise, as variáveis originais ( $X_1, X_2, X_3, \dots, X_i$ ) foram transformadas em um conjunto de variáveis independentes ( $Y_1, Y_2, Y_3 \dots Y_j$ ), em que Y é uma função linear de X (Van Beuningen e Busch, 1993). As variáveis canônicas foram também usadas para a disposição dos acessos em gráfico bidimensional, utilizando-se a dispersão dos valores dos acessos, em relação aos eixos representativos das primeiras variáveis.

Esta análise foi realizada por meio do software 'GENES' (Cruz, 1998), que possibilitou estimar as distâncias generalizadas de Mahalanobis, compreendendo todas as combinações entre os 36 acessos de arroz. As técnicas de análise multivariada foram usadas, visando a estudar a divergência genética entre os 36 acessos, os quais foram avaliados por marcadores morfológicos quantitativos. Utilizou-se do critério de Wilks, para verificar se ocorreram diferenças entre os vetores médios dos acessos. Este é calculado por:

$$\lambda = \frac{\text{Det. (E)}}{\text{Det. (E+H)}}$$

sendo que:

Det (E) = determinante da matriz E;

Det. (E+H) = determinante da matriz E+H;

E.: matriz soma de quadrados e produtos residuais;

H: matriz soma de quadrados e produtos entre cultivares

Para o agrupamento dos acessos, utilizou-se como medida de dissimilaridade, a distância de Mahalanobis ( $D^2_{ij}$ ), e o método de agrupamento do Vizinho mais Próximo (VMP), também de Tocher, que já foram comentado antes. Adotando o software GENES para as análises.

#### **c) Projeção no Plano ou Dendrograma**

Com auxílio do programa STATISTIC, permitiu-se agrupar os acessos e formar o dendrograma de árvore. Adotando o método do vizinho mais próximo. A similaridade entre indivíduos foi então julgada pela proximidade ou pela distância que eles apresentam nesse hiperespaço (projeção em dendrograma bidimensional). Em seguida a esses procedimentos que permitiram fazer a separação em grupos morfológicamente similares, com padrão de afinidades intergrupais, e ainda pode-se fazer a alocação de indivíduos em grupos preexistentes (Campbell e Kerne, 1981).

#### **d) Importância Relativa dos Caracteres**

A identificação do caractere que mais contribui para a divergência dos acessos pode ser feita pelas variáveis canônicas, que é dado pela razão entre a variância por ela explicada e o total de variância disponível. Adotou-se o software Genes para análises, baseando na distância generalizada de Mahalanobis, identificando aqueles que são dispensáveis para o estudo da divergência, pois ficam invariáveis ou redundantes entre os acessos avaliados, por estarem correlacionados com outros (Cruz e Reggazi, 1994).

## **2.3 - RESULTADOS E DISCUSSÃO**

### **2.3.1 - Análise Preliminar dos Caracteres Qualitativos**

Preliminarmente os acessos foram submetidos à avaliação pelos caracteres qualitativos para verificar a existência de variabilidade desses, por entre as três repetições. Essa análise teve por finalidade reunir os 36 acessos de arroz em tantos grupos quanto fossem necessários, de tal modo a existir homogeneidade dentro e heterogeneidade entre o conjunto de acessos (dados não mostrado).

As características que apresentam distribuição descontínua, portanto, não foram submetidas à análise de variância, sendo analisados pela frequência relativa apresentada pelas plantas nas parcelas (Tabela 4). Esses caracteres qualitativos que mostraram semelhança entre os acessos são estáveis para a espécie, e como os acessos são CV's obsoletas de arroz, carregam certo grau homogeneidade entre si ou seja, caracteres que são ligados ao crescimento e à produção, pode mostrar certa uniformidade devido os ciclos de cultivo. Todavia, Singh et al. (1993) mencionam que esses caracteres qualitativos devem ser bem conhecidos pelo pesquisador e em muitos casos alguns desses caracteres qualitativos estão ligados à fatores genéticos de interesse.

Verificou-se que os acessos mostraram estabilidade para os seguintes caracteres: CCo, FLi, CAu e CGM, uma vez que em 100% dos casos mostraram a mesma característica, isso é não sofreram nenhuma variação entre as três repetições.

Todavia, o comportamento da maioria dos caracteres qualitativos avaliados apresentaram diferenças acentuadas (Tabela 4), e acontecendo tanto entre como dentro das parcelas. Esses caracteres foram AFB que apresentaram as três formas (erecta, intermediária e horizontal) proporcionalmente.

**TABELA 4 - Descrição das variáveis qualitativas adotadas e a frequência de ocorrência para avaliar a estabilidade entre 36 acessos de arroz de codinome 'três meses'. CNPAF, Goiânia, GO, 1996.**

Símbolo	Classe	Freq. Observada nas Parcelas	Frequência Relativa
CCo	incolore verde	108	100,00 %
FLi	fendida	108	100,00 %
CAu	incolore	108	100,00 %
AFB	ereta	50	30,86 %
	intermediária.	58	35,80 %
	horizontal	54	33,33 %
FPa	aberta	2	1,68 %
	intermediária	103	85,83 %
	compacta	15	12,50 %
EPa	justa	83	42,78 %
	média	103	55,67 %
	longa	3	1,55 %
TAr	mútica	10	8,06 %
	micro aristada	106	92,06 %
CAp	preta	10	8,06 %
	pardo	36	29,06 %
	palha	78	62,90 %
PuF	glabra	101	93,52 %
	pubescente	9	8,33 %
FGI	rudimentar	0	0,00 %
	estreita	96	91,43 %
CGM	vermelha	9	8,57 %
	palha	108	100,00 %

*Cor do colar (CCo); forma da lígula (FLi); cor da aurícula (CAu); ângulo folha andeira (AFB); exerceção da panicula (EPa); tipo de panicula (TPa); tipo arista (TAr); cor do apículo na floração (CAp); pubescência da folha (PuF); forma das glumas (FGI); cor da glumela na maturação (CGM).*

Enquanto que EPa apresentaram com maior frequência de justa e média (42,78% e 55,67% respectivamente); para CAp a cor palha apareceu para 62,90% dos casos, sobressaindo sobre a parda, com percentual de 29,06%. As

de coloração preta apareceram com apenas 8,06%. Também mostrou alta frequência (91,43%) a FGI de formato estreito.

Quanto a desuniformidade dos caracteres qualitativos (Singh, Singh e Singh, 1996) mencionam que são influenciados pelo ambiente, mas expressam a síntese genética da planta. Em se tratando de marcadores morfológicos, Martin et al. (1995) e Santos (1996) ressaltam que esses carregam grande influência do ambiente, assim como uma avaliação em outros ambientes poderia contornar essas dúvidas de que são instáveis geneticamente, ou sofrem o efeito do ambiente; mas pelo fato de serem acessos e são constantemente multiplicados, nota-se grande estabilidade dos caracteres.

Sabe-se que os caracteres que prestam para expressar as classes dos acessos, uma vez que, em sua grande maioria é de herança simples; sofre pouca ou nenhuma ação do ambiente; os indivíduos segregam numa razão coerente pela dominância, mas, segundo Fonseca (1993) por que são caracteres qualitativos de baixo valor para estudos da diversidade genética. Devido a essa grande oscilação, esses caracteres, por si só, não são confiáveis para avaliar os acessos de um BAG.

Com análises destes caracteres, os valores mostram que os acessos são diferentes, levando a inferir que existe alguma variabilidade entre os acessos. Porém, Souza (1990) comenta que essa desuniformidade dos caracteres, que acontece no arroz, é um resultado até certo ponto, esperado; considerando que os acessos provem de ambientes diferentes, e os caracteres oscilam devido a interferência da evolução e seleção que sofreram.

### **2.3.1.1 - Análise Univariada dos Caracteres Quantitativos**

Os resultados obtidos na análise de variância para os marcadores morfológicos quantitativos, bem como média, CVe e a razão dos coeficientes



(CVg e CVe = D), expressos na Tabela 5, revelou existirem diferenças significativas entre os acessos.

TABELA 5 - Resumos da ANOVA de doze caracteres avaliados em trinta e seis acessos de arroz do grupo três meses. Embrapa Arroz e Feijão, Goiânia, GO, 1998.<sup>1</sup>

Caracteres	Quadrados Médios			Parâmetros Genéticos		Cve (%)
	bloco	acessos	erro	média	I = CVg/CVe	
DEm	0,250	0,9595 NS	0,58334	7,47	0,4636	10,22
DAF	11,148	79,9656**	35,6791	76,148	0,8425	7,84
CTPI	177,750	142,3690**	11,8071	113,69	1,9199	3,02
API	99,966	91,5283**	15,6194	109,84	1,2728	3,60
CFB	3,977	19,1231*	5,44796	32,38	0,9147	7,21
LFB	0,631	0,1124 NS	0,08306	1,79	0,3433	16,08
CPa	1,291	4,07695*	1,5031	23,26	0,7555	5,27
PPa	0,1756	1,7723*	0,4015	3,2695	1,0667	6,77
P100	0,4761	0,1814 NS	0,04597	7,43	0,9911	2,89
CGS	0,5511	0,0169 NS	0,0152	1,86	0,1975	6,62
LGS	0,03211	0,0922 NS	0,03949	2,78	0,6675	7,13
Prod	41936,33	36419,323**	4071,838	302,94	1,8273	21,06

\* e \*\* significativo pelo teste de F, respectivamente, aos níveis de 5% e 1% de probabilidade.

*Dias para emergência (DEm); dias até o florescimento (DAF); ciclo total da planta até a maturidade (CTPI); altura da planta (API); comprimento e largura da folha-bandeira (CFB; LFB); comprimento e peso da panícula (CPa; PPa); peso de 100 grãos (P100); comprimento e largura do grão sem casca (CGS; LGS); produção de grãos (Prod).*

Verificou-se que alguns caracteres quantitativos (DEm, LFB, P100, CGS e LGS) não apresentaram diferenças entre os acessos, significando que são uniformes e isso acontece pelo fato destes apresentarem importância para a produção, e como estes acessos foram CV's um dia, já receberam seleção continuada pelos rizicultores, trazendo uma margem de estabilidade para esses acessos.

Mardia et al. (1979) comentam que quando de baixa magnitude do CVE, confirma a precisão experimental do ensaio, aqueles caracteres de mais alto valor indicam maior interferência pelo ambiente.

Quanto àqueles caracteres: DAF, API, CPa, PPA e Prod que diferiram e apresentaram CVE de valor baixo, exceto LFB e Prod, confirmando a performance desses caracteres; e que será comentado mais a frente. Por outro lado, segundo Singh, Ceccarelli e Hamblin (1993) esses caracteres comportam como marcadores verdadeiros e prestam para confirmar a dissimilaridade entre os acessos avaliados, ou mesmo podem mostrar a interferência do ambiente na espécie.

As diferenças encontradas neste estudo, denotadas pelas observações com análise de variância os marcadores morfológicos quantitativos (Tabela 4), confirmando o previsto daquela análise entre os acessos (Tabela 3) com os marcadores qualitativos, que apresentaram diferentes frequências entre e dentro das parcelas, descartando a hipótese da nulidade, ou seja ausência de variabilidade entre os acessos de arroz com nomes regionais similares. Mackill e Lei (1997) relata que em arroz essa variabilidade é baixa, devido a pequena base genética da espécie que é cultivada. Conforme os resultados de Souza (1990), que mesmo o arroz ser cultivado há anos e ser uma autógama, mesmo assim mostra alguma diversidade entre as CV's, por incorporar as reações do ambiente.

Esta variação entre os acessos, quanto a essas características, permite inferir dados sobre a evolução da espécie, a qual, quanto maior foi o tempo de domesticação ou mesmo a área de dispersão pelo globo, são incorporados alterações ou estabilização, que dependem das reações com o ambiente, e o grau de interesse econômico; ainda, ressalta-se que são caracteres de herança simples e sofrem baixíssima interferência do ambiente (Jonhson e Wichen, 1982).

Carbonera (1990) comentou que, para aqueles caracteres importantes para cultura do arroz, existe ampla variação das variáveis que caracterizam a espécie. Essas variações são devidas à função do ambiente em que as plantas são cultivadas, mas indicam também a condição genética para que mostre a variabilidade presente. E que o método usado no cálculo da estimativa, que pode ou não encontrar essa variabilidade genética existente entre os acessos, porque alguns caracteres pouco refletem a divergência, devido ser controlados por poucos genes.

### **2.3.1.2 - Avaliação da Divergência Genética entre os Acessos por Técnicas Multivariadas**

Através da análise multivariada o valor da estatística de Wilks ( $\Lambda$ ) = 0,7249, altamente significativo ( $P < 0,01$ ), adotando o teste ora referido, com  $n_1 = 187$ ;  $n_2 = 254$ ;  $F(\text{obs}) = 1,41$ ; indicando a existência da variação global entre os acessos avaliados. A análise de variância univariada não foi efetiva na detecção da variabilidade, para todos os caracteres avaliados, mas estes mesmos caracteres, no total de doze, foram analisados conjuntamente pela análise multivariada, mostrando diferenças altamente significativas ao nível de 1% de probabilidade para o efeito agregado de todos os caracteres.

A esse respeito, Demétrio (1985) faz uma observação pertinente: em muitos experimentos, são tomados dados de um conjunto de variáveis individualmente, sem considerar as correlações existentes entre estas. Nas análises individuais são observadas o o comportamento das variedades para cada uma das variáveis, enquanto que, na análise multivariada há uma avaliação dos tratamentos no conjunto de todas as variáveis, assim, é considerado um nível conjunto de significância, pois ao se adotarem as análises conjuntas dos caracteres, uma única resposta é formulada, a qual geralmente contém mais

informações sobre o efeito total de tratamentos do que se fosse uma série de respostas consideradas separadas.

As distâncias Generalizadas de Mahalanobis para todas combinações possíveis entre pares de acessos mostram um intervalo amplo em unidades de distância (u.d.) com valores de 18,93 a 243312; embora sejam valores adimensionais e difíceis de interpretar, melhor é representá-los em termos percentuais, o que equivale de 0,008 a 100%. Por este valor de amplitude nas distâncias fica comprovado que os acessos são mesmo divergentes, o que foi previsto pela análise univariada e multivariada, concordando com os relatos de Montalvan del Aguila (1990), sempre haver divergência entre cultivares.

Estes valores são próprios para os acessos nas condições em que foram conduzidos e avaliados, pois fica difícil compará-los com os de outros autores. Cabe relatar que a formação de grupos em um estudo no mesmo local não deve ser comparada com outro estudo do mesmo gênero em iguais condições (Dudley, 1994; Van Beuningen e Busch, 1993). O valor  $\alpha$  de divergência, que é a máxima distância entre as mínimas, serve para analisar os agrupamentos a respeito deste percentual entre os pares de acessos avaliados, verifica-se ser um valor arbitrário de magnitude, pois em termos de acessos deveria estar correlacionado com alguma característica de interesse no indivíduo.

Pelos resultados em distância estatística pôde-se observar que o acesso A7 (CA820080) foi o de maior distância entre os demais. Porém, adotando-se  $\alpha = 50\%$  de divergência total, inúmeros acessos mantêm distância dele (ex. A27; A19; A33; A23; A12; A5; A24; A25), o que será discutido posteriormente, com base no dendrograma. Muito próximo do A7, com menos de 2% de divergência encontram-se: os acessos A10; A34, A36, os quais como parentes próximos, têm as menores distâncias; pelos relatos de Tingey et al. (1991) isso indica a

presença de vários alelos em comum e esses alelos estão presentes na maioria dos demais.

Ainda podem-se verificar os acessos que estão numa distância inferior a 0,1%, ou seja, são similares os pares: A1 e A29; A13 e A15; A14 e A26; A21 e A22; A30 e A32, essa proximidade indica que os acessos são provenientes de mesmo progenitor ou cruzamentos próximos (Abdelmoor et al., 1995).

No que se refere à distância genética e procedências geográficas, geralmente, materiais de um mesmo Estado ou região têm menor distância que quando contrastados com procedências de outras regiões. A coleta do A7 foi no Estado do Acre, por isso expressou distância conforme com sua procedência original, no entanto, o A5 que foi coletado em Roraima, não mostrou associação com a procedência, com valores  $\geq 1,0\%$  de distância entre os pares A8, A14, A21, A22, A6 e A26, que são de coletas procedentes em GO, AC e MG.

No que tange à relação de distância, Montalvan-Aguila (1990) comenta que a distância parece crescer em termos gerais, correlacionada com genealogia, procedência geográfica, assim como materiais oriundos de um mesmo local têm distância menor do que quando confrontados com materiais de locais diferentes e nível de melhoramento. Neste estudo é mais provável que o nível de melhoramento seja o que mais pesa para todo o conjunto, uma vez que os acessos constituíram CV's Obsoletas. Isso se deve ao fato da interação ambiental interferir muito nas culturas domésticas principalmente.

### **2.3.1.3 - Emprego das Variáveis Canônicas (cv's)**

Em prosseguimento ao estudo da divergência genética entre os acessos foram calculadas as variáveis canônicas usando de todos 12 caracteres avaliados. Por esta análise, apura-se a divergência genética, que é evidenciada

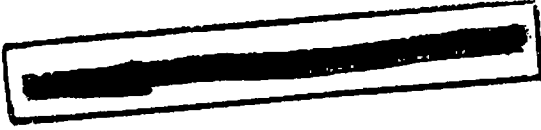
pela dispersão dos escores em gráficos, e pelas contribuições dadas pelas respectivas primeiras duas v.c's.

A primeira vc' explica apenas 39,60% da variação total contida no conjunto dos caracteres analisados; a segunda explica 21,56% da variação total, indo até a decima primeira que contribuiu para variação contida nos acessos (Tabela 6). Seguindo a orientação de Morais (1992); Sneath e Sokal (1973), deve-se observar o percentual acumulado das duas primeiras v.c's. Nessa análise, o percentual das duas primeiras atingiu a 61,16% de variação total. A terceira chegou-se a 70,74%, com a quarta a 79,07%, o que dificulta explicar e visualizar num plano multidimensional.

**TABELA 6 - Variâncias, Variâncias percentuais e acumuladas estimadas para as variáveis canônicas obtidas de 12 caracteres avaliados, para estimar a divergência genética, entre 36 acessos de arroz 'três meses'. CNPAF, Goiânia, GO, 1998.**

V. C.	Variâncias	% de Variância	% Variância acumulada
01	7.0818	39.6096	39.6096
02	3.8543	21.55783	61.1675
03	1.7127	9.57985	70.7473
04	1.4874	8.31921	79.0665
05	0.9462	5.29214	84.3587
06	0.7968	4.45661	88.8152
07	0.6459	3.61312	92.4284
08	0.4645	2.59828	95.0267
09	0.2830	1.58309	96.6097
10	0.2530	1.41516	98.0249
11	0.2007	1.12301	99.1479
12	0.1524	0.85206	100.00

Foram necessárias o número de quatro v.c's para atingir os 80%, o que é considerado o mínimo da variação total e aceitável. Segundo Cruz e Regazzi (1994) é fácil visualizar e mesmo aceitável quando, com apenas duas variáveis, atinge-se o percentual acumulado de pelo menos 75%, sendo desta forma, usadas



para a identificação de caracteres de menor importância. Daí os resultados não serem muito expressivos, necessitando de explorar mais métodos que auxiliem na explicação da divergência genética entre os acessos.

Jaradat (1991) trabalhando com trigo, também usou mais de duas v. c's, ou seja, dos 11 caracteres estudados utilizou-se de 8 para encontrar 92% da variação total. No entanto, é comum encontrar o uso de mais de duas variáveis canônicas e com isto utilizar um número mais significativo de caracteres. Porém, Campbell e Kerne (1982) encontraram 87,5% da variação total entre os caracteres nas duas primeiras raízes canônicas, mostrando com isso que a técnica apresentou resultados similares, em termos de padrão da divergência genética.

Utilizando a técnica das v.c's, Kaw (1995) estudou dezenove caracteres e verificou que as quatro primeiras v.c's explicaram 73,3% da variação global, entre as cultivares de arroz avaliadas. Baseando-se nos coeficientes de correlação entre os caracteres originais e usando-se as quatro v.c's selecionadas, foram descartadas, após identificação, três caracteres originais (comp. internódio, largura da folha e número de grãos por fileira) após verificar que não estavam correlacionadas significativamente com as v.c's de maior importância, indicando que não havia nenhuma contribuição direta para a divergência canônica.

Em estudo com 19 caracteres em cinco populações de arroz, Singh, Singh e Singh (1996) notaram que as três primeiras variáveis canônicas acumularam mais de 95% da variação total. Ao traçar os coeficientes de correlação entre caracteres e as três variáveis canônicas de maior importância, constatou-se que apenas quatro estavam em correlação significativa: pesos de nós e albúmens, diâmetro equatorial e percentagem de albúmen no fruto e água, mostrando que somente esses contribuíram para aquele valor.

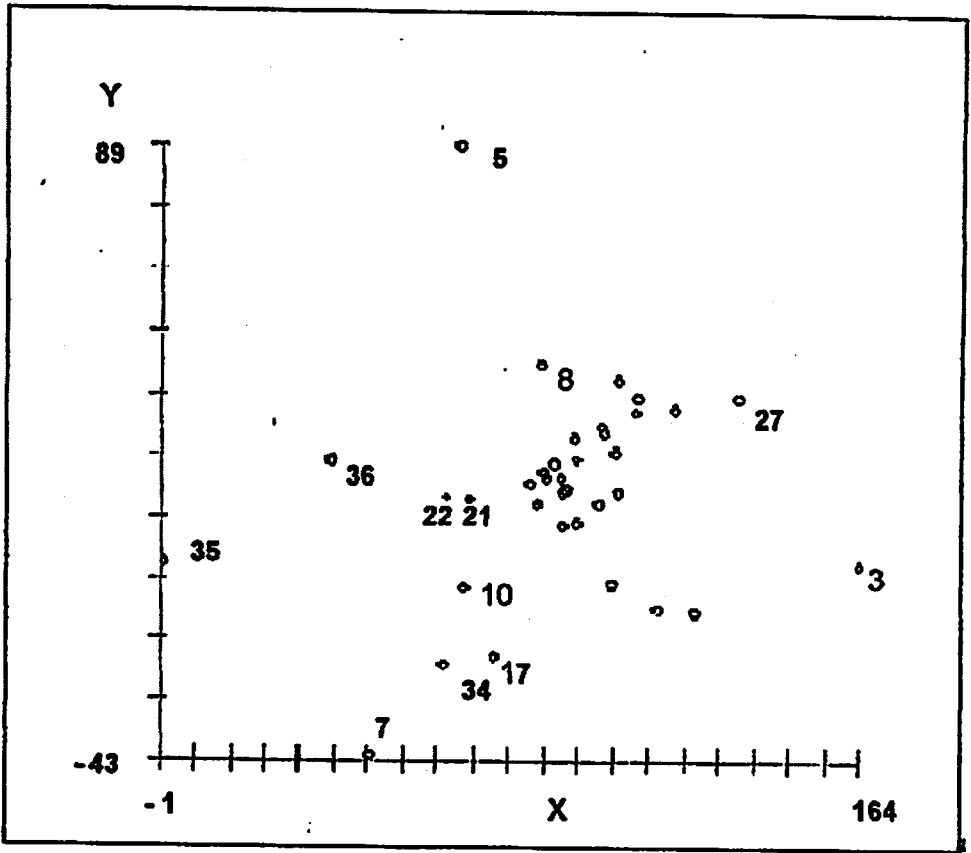


FIGURA 3 - Dispersão gráfica dos 36 acessos de arroz 'três meses', com base nos primeiro e segundo escores, das variáveis canônicas obtidos com uso de 12 caracteres avaliados. CNPAF, Goiânia, GO, 1998.

Os escores obtidos para o arroz a partir de dois autovalores foram utilizados para a representação gráfica no plano bidimensional. Em princípio, é oportuno salientar que a dispersão gráfica dos escores possibilitou a identificar aqueles acessos mais distintos, ou mesmo formar grupos isolados com alguns acessos (Figura 3), tendo como referência a dispersão que mostrou os mais divergentes: A3; A5; A35; A36; A7.



Também são mostrados os grupos pequenos de A21; A22, e ainda A10, A17; A34, os restantes agrupados; nota-se que o local de coleta e os nomes coincidentes não interferiram na divergência genética. Daher (1993), ao comparar a divergência genética em *Paspalum* pelas v.c's e  $D_{ij}^2$ , encontrou diferença na composição dos grupos, sem estarem associados à procedência geográfica. No entanto, Julquifar et al. (1992), ao avaliarem 100 CV's de arroz por meio da  $D_{ij}^2$  e v.c's encontraram grupos formados por 1 até 36 CV's, com os vetores de maior importância para a produção, peso de 1000 grãos, dias para a maturação e altura das plantas.

Resultados semelhantes com a dispersão dessas v.c's aconteceram com os acessos concordantes, uma vez que análise preliminar com os qualitativos mostrou que havia diferenças no conjunto os acessos. Então confirma que os acessos: A5; A7; A21; A22; A34; A35; A36, de fato são divergentes, pois permaneceram em grupos diferentes.

#### **2.3.1.4 - Análise de Agrupamento**

##### **a) Método do vizinho mais próximo (VMP)**

Para fins de classificação e caracterização, as relações de divergência entre os 36 acessos de arroz são melhor expressadas em agrupamentos. Pelos dados da distância generalizada de Mahalanobis foi possível agrupar os acessos com características semelhantes, utilizando o método do VMP. Os grupos formados por este método mostrando a divergência genética encontram-se na Figura 4. A análise do dendrograma agrupou os 36 acessos ao nível de apenas 38% de divergência genética apenas, formando 3 grupos distintos, muito

desequilibrados, com um muito grande. Seria mais coerente formar maior número de grupos, devido à oscilação dos marcadores morfológicos.

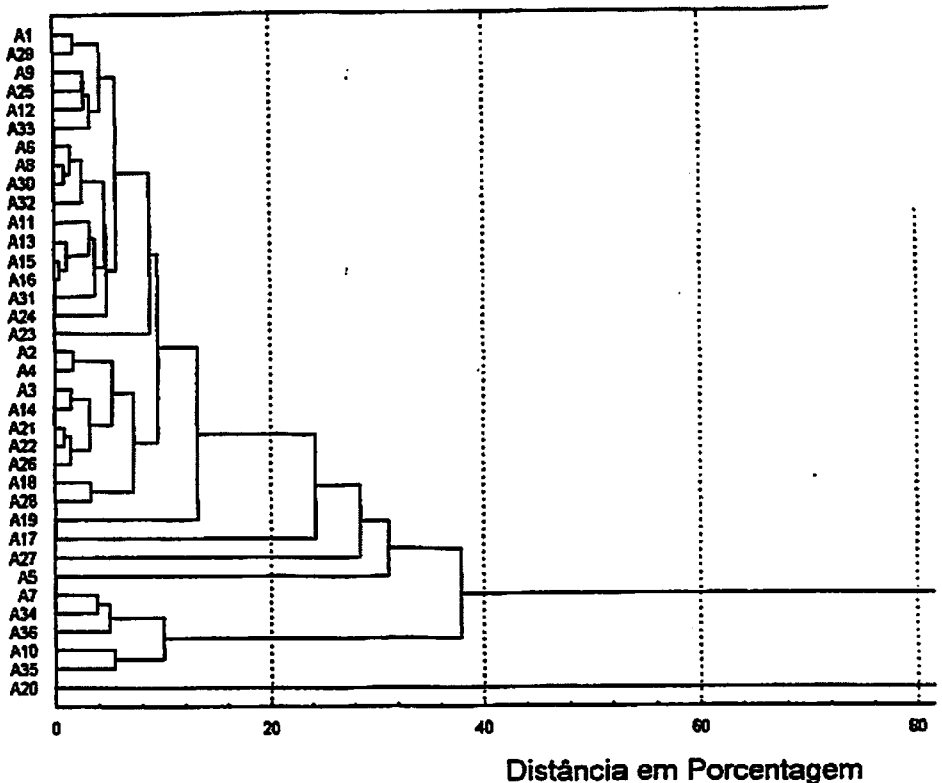


FIGURA 4 - Dendrograma resultante da Análise de Agrupamento dos 36 acessos de arroz 'três meses', baseado nas distâncias de Mahalanobis e no método de vizinho mais próximo. CNPAF, Goiânia, GO, 1998.

Para discussão adotou-se  $\alpha = 20\%$ ; com isso os acessos foram agrupados em 6 grupos (A - F) que foram subdivididos. Os grupos A e F foram subdivididos em subgrupos (11 e 2 respectivamente) com uma divergência de apenas 5% (Tabela 7). Desse grande número de acessos no grupo A era esperado, uma vez que os acessos foram coletados dentro de mesmas áreas ou próximas

destas, as quais fazem cultivo corriqueiro de arroz, tendo sido variedades cultivadas anteriormente (hoje são obsoletas). Segundo Dilday (1990) e Chang e Bandenas (1965), essa é uma razão da grande similaridade entre grupos de acessos.

**TABELA 7 - Agrupamento dos 36 acessos de arroz do 'três meses', baseado nas distâncias de Mahalanobis e no método de Vizinho mais Próximo. CNPAF, Goiânia, GO, 1998.**

Grupo s	$\alpha$ (%)	Subgrupos	ACESSOS	$\alpha$ (%)	nome comum	Origem
A	25%	onze		-		
		I-A	A1 e A29	5%		
		II-A	A9, A12, A25, A33	4,5%		
		III-A	A6, A8, A30	3%	Três meses branco	AC, GO
		IV-A	A13, A15, A16	1,5%	Arroz três meses	GO
		V-A	A11, A24, A31, A32	7%		
		VI-A	A23	7,5%		
		VII-A	A2, A4	4,5%	Três meses antigo	SP
		VIII-A	A3, A14	2,0%		GO, SP
		IX-A	A21, A22, A26	2,5%	Três meses	GO, MG
		X-A	A18, A28	5%		
XI-A	A19	13%				
B	25%		A17			
C	38%	-	A27			
D	32%	-	A5			
E	38%	dois				
		I-E	A7, A34, A36	5%	Três meses branco	-
		II-E	A10, A35	6%		-
F	100%	-	A20			

Deve-se ressaltar que a nível de cultivares a informação contida nos dendrogramas tem mais sentido que as distâncias analisadas entre os pares isolados. A identificação dos grupos e utilização permitirá uma melhor conservação e a classificação dos acessos para melhor agilização no BAG (Kim e Ward, 1997). Essa informação também pode ser usada para dar sustentação na hibridização entre os grupos. É comum a informação na literatura de que dentro de um grupo a divergência genética é pequena, e quando ocorre a nível de

subgrupo é menor ainda, isso certamente é devido aos ascentrais que são comuns (Arumachalan, 1981).

Verifica-se no dendrograma que os acessos dentro dos subgrupos são praticamente similares com divergência genética muito pequena, ou insignificante, no subgrupo III-A: A6, A8; A30 ( $\alpha = 3,0\%$ ); no subgrupo IV-A: A13, A15; A16 ( $\alpha = 1,5\%$ ); no subgrupo VIII-A: A21, A22 e A26 ( $\alpha = 2,5\%$ ), o que foi confirmado pela distância  $D^2_{ij}$ . Associando esses subgrupos e a região de coletas, verifica-se que ocorreu o efeito da procedência geográfica, pois os mais próximos sempre apresentam procedência de Goiás.

Noutro extremo ficou o acesso A20 que foi o mais divergente de todos com base nos seus atributos morfológicos para estimativa da posição perante o grupo. Em seguida observa-se o subgrupo I-E com os acessos A7, A34, A36 ( $\alpha = 5,0\%$ ) e II-E composto por A10 e A35. ( $\alpha = 6,0\%$ ). Para mais indagações, os grupos foram subdivididos em subgrupos, os quais mostraram coincidência entre os métodos de agrupamento foram A1 e A29; A21 e A22. Tomando como base os nomes regionais, raramente se repetem para os agrupamentos. A diversidade geográfica não mostrou contribuir para a divergência genética a nível de grupo, mostrando que o arroz tem uma variabilidade considerada por si mesmo.

### **B) Método de Tocher**

O conjunto de acessos também foram agrupados pelo método de Tocher (Tabela 8), considerando-se o conjunto dos caracteres antes mencionados, conforme Rao et al. (1981). O agrupamento realizado por este método, mostra os grupos de divergência um pouco diferente daquela formada pelo método agrupamento do VMP (Tabela 7).

Nesse caso formaram-se 7 grupos distintos; um grande grupo (A) com 29 acessos, embora com muitos acessos. Os quais têm suficiente grau de semelhança

que permite essa aglomeração, uma vez que são homônimos e ainda mais, como CV's, devem ter sofrido uma forte pressão de seleção, por longo tempo que foram cultivados e outros 6 grupos para os 7 acessos restantes (grupos B a F).

**TABELA 8 - Composição de grupos de 36 acessos de arroz 'três meses' calculado pelo método de Tocher. CNPAF, Goiânia, GO, 1998.**

Grupos	Subgrupos	$\alpha$	ACESSOS	Nome comum
A	I-A	< 3,0%	A21, A22, A23, A25, A30, A31, A16, A10	
	II-A	< 5,0%	A1, A29, A33	
	III-A	< 6,0%	A2, A4, A6, A11, A13, A19, A24, A32	
	IV-A	< 7,0%	A12, A14, A27, A28	
	V-A	< 10%	A18, A26	Arroz três meses
	VI-A	< 12%	A8, A17, A20	Três meses branco
B		< 25%	A7, A34, A9	Três meses branco
C		< 34,1%	A36	
D		< 38%	A3	
E		< 40%	A5	
D		32%	A7	
F		40%	A35	

Esse método é mais rigoroso, por isolar um grande grupo de acessos próximos, ficando os outros com poucos acessos ou apenas um, como nos outros estudos. A explicação para o fato segundo é dada em termos de seleção e adaptação ao ambiente natural, uma vez que ambas ocasionam diferenças na composição genética, as quais são pareadas e passadas para os descendentes.

Também usando Tocher, Maurya e Singh (1977) discriminaram 9 grupos em 43 variedades de arroz; Rao et al. (1981) encontraram 32 grupos em 120 variedades de arroz e Nadaf e Singh (1986) conseguiram discriminar 5 grupos em 120 linhagens de elite. Segundo estes últimos autores, quanto mais cultivado for o material, maior será o número de grupos discriminados.

Há limitações de informações sobre os acessos em questão, mas existem alguns aspectos interessantes que podem ser verificados nos agrupamentos: A7 e A34, A35, A36, quando avaliados pelos dois métodos de agrupamentos pois ficaram no mesmo grupo, divergindo do restante. Quando associados pela procedência, A34 e A35 se mostraram similares e foram coletados no Mato Grosso. Nessa análise, o A20 manteve-se isolado dos restantes. Por último, A10 e A35 estão em subgrupos coincidentes. Permaneceram isolados nos grupos os acessos: A3, A5, A35, A36.

No entanto, Viana (1990) utilizou-se da análise de agrupamento para estudar a divergência entre 20 clones de cana-de-açúcar através da distância Mahalanobis, calculada a partir de dados morfométricos. Foi aplicado o método de otimização de Tocher para agrupar os materiais. O método não apresentou eficiência prática, possivelmente devido a interferência de fatores ambientais na expressão fenotípica dos caracteres avaliados. O autor também não encontrou associação entre divergência e diversidade de origem geográfica.

Nesse estudo com o uso de Tocher possibilitou a formação de seis subgrupos distintos com distribuição variável dos acessos dentro de cada um. No outro método (VMP) formaram-se 13 subgrupos distintos de acessos; este fato mostra não haver coerência nas comparações entre metodologias. Com uso de Tocher conseguiu-se menor número de subgrupos, mostrando ser mais rigoroso; esta não coincidência da classificação dos grupos de acessos utilizando diferentes metodologias é discutida por Cruz e Regazzi (1994) como uma dificuldade não contornável.

Entretanto, dentro de maior rigor do método de Tocher, alguns acessos (A5, A7, A34, , A35, A36) mostraram ser divergentes pelos dois métodos. A relação dos grupos formados pelos métodos não pode ser demonstrada

estatisticamente com estes dados, porém, o que se observa é que o vizinho mais próximo para estes acessos mostrou uma distribuição mais precisa e com grupos mais coerentes, apesar de que alguns acessos permaneceram invariáveis, o que serviu para consolidar sua classificação, também Montalvan del Aguila (1990) trabalhando com arroz encontrou menor número de grupos para Tocher.

Cruz e Regazzi (1994) afirmam que, pelas técnicas que calculam distância genética, somente ocorre coincidência quando não há correlação entre os caracteres avaliados; como neste trabalho obteve-se alto índice de correlação entre algumas das características, esperava-se que não houvesse tantas coincidências.

#### 2.3.1.5 - Importância entre as Variáveis Originais

No estudo da divergência genética adota-se o número de  $v. c$ 's necessário para explicar o mínimo de 80% da variação entre os acessos estudados. Os escores das variáveis canônicas selecionadas são calculados e dispostos em planos cartesianos para a análise preliminar dos grupos.

Os caracteres que apresentam diferenças significativas são os que mais contribuem, e aqueles que não diferiram sempre aparecem com menor peso para cálculo dos autovalores, podendo ser descartados quando há muita informação (valores não publicados). As maiores contribuições para a divergência genética em arroz, tomadas através das maiores estimativas da correlação pela  $v. c$ 's foram ciclo para maturação, altura, comprimento da folha bandeira, pesos de 100 grãos e paniculas, e os que menos contribuíram foram comprimento e largura dos grãos (CGS e LGS). Ainda há relatos desses caracteres sofrerem alta interferência do ambiente onde está inserido ou mesmo de ordem genética (Souza, 1990; Rao et al., 1981).

Os valores encontrados para os caracteres mostram algum interesse para o melhoramento, como a qualidade de grãos e o ciclo de produção. Também Julquifar, (1981) avaliaram a natureza da divergência genética em 18 cultivares de duas raças geográficas de arroz, "Japônica" e "Índica", representando um amplo espectro de variação, concluíram que altura de planta e comprimento da panícula foram os caracteres que mais contribuíram para divergência entre as cultivares.

Por outro lado, Maurya e Sing (1977), em estudos semelhantes, verificaram que dias para maturação, altura de planta e perfilhamento foram os caracteres que mais contribuíram para a divergência genética entre 43 variedades de arroz.

Estudando a divergência genética entre 35 variedades de arroz de porte baixo para um conjunto de 12 caracteres adaptativos e relacionados com a produção, Singh, Singh e Singh (1996) verificaram que altura de planta, área da segunda folha, comprimento do primeiro internódio e peso de grãos foram as características que mais influenciaram a divergência genética entre as cultivares.

Os coeficientes de correlação que envolvem estas quatro primeiras v.c's como caracteres mais importantes, mostraram-se significativos, pelo menos com um dos caracteres avaliados, indicando que todos são importantes na descrição dos acessos. Cammings e Hoppensteadt (1982) sugerem que nos casos em que muitos caracteres são correlacionados, que se deve tentar encontrar aqueles não correlacionados pelo método de componentes principais ou v.c's, reduzindo, assim, o número de caracteres sem perda significativa da informação.

Correlações entre produção de grãos e outros 10 caracteres, em 24 CV's e linhagens de arroz foram estudadas por Arunachalan (1981), o qual obteve correlações positivas de produção de grãos com número de perfilhos,



comprimento de panícula, número de espiguetas/panícula e peso de 100 grãos, não ocorrendo associações antagônicas entre os caracteres. O autor concluiu que se poderia melhorar linearmente a produtividade pela seleção individual ou simultânea desses caracteres e todos cuidados adicionais se devem ser tomados com relação à obtenção dos descritores, principalmente para os caracteres quantitativos, os quais devem ser obtidos em experimentos com repetição, o que aconteceu nesse trabalho.

Analisando a relação I, Tabela 5 que dá idéia da variabilidade genética do caracter, em função da expressão fenotípica do mesmo, mostrou maior valor para o ciclo de maturação 1,97, e para produção de 1, 63. Pelo CVg, que fornece a variação total de um caractere, é de interesse devido mostrar a porção herdável (variação de natureza genética) os caracteres que mostraram menores valores foram: comprimentos da panícula e de grãos, indicando baixa interferência do ambiente para esses caracteres. Os coeficientes de variação genética (VGg) que mostraram a maior segregação gênica, estão dentro dos limites esperados para mostrar a estabilidade dos caracteres nas CV's de arroz, porém os seguintes caracteres: CTPI, API, PPa, Prod foram os que apresentaram valores maior que a unidade.

Pelos CVe's que foram de maiores valores DEm (10,22%), LFB (16,08%) e Prod (21,06) mostram que esses caracteres mesm com estes valores foram eficientes para diferenciar os acessos. Pelos estudos de divergência genética em híbridos de arroz, pelos mesmos procedimentos, Camargo (1982) conseguiu descartar 4 caracteres entre os 13 avaliados.

Porém, Resende (1991) empregou distância de Mahalanobis, variáveis canônicas e agrupamento pelo "vizinho mais próximo", em caracteres morfológicos e enzimáticos em 11 espécies de *Laelia parviflorae* e conseguiu

selecionar 6 caracteres morfológicos mais importantes entre os 13 avaliados. O grau e a natureza da associação entre os caracteres mais importantes indicaram que a seleção baseada em número e peso de grãos, conseqüentemente, peso de panícula seria efetiva.

#### 2.4. CONCLUSÕES

- a - O uso dos marcadores morfológicos, adotando-se a distância Mahalanobis, variáveis canônicas e agrupamentos (Tocher e vizinho mais próximo) nas análises, foram todos eficientes na determinação da divergência genética, entre acessos de arroz, mostrando coincidência para os acessos divergentes.
- b - Na formação dos grupos, o método de Tocher mostrou-se de fato ser mais rigoroso. Assim sendo, os acessos A7, A5, A35, A36 foram os mais divergentes e apresentam coincidências dos métodos, o que acontece também com os acessos similares A21, A22 e com A1, A19.
- c- Os caracteres ciclo de maturação, altura da planta, comprimento da folha-bandeira, pesos da panícula e de 100 grãos foram os mais importantes para a estimativa da divergência genética dos acessos.
- d - Comparando os acessos com o local das coletas e os nomes semelhantes, não se pode formar associação clara e confiável.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ABBUD, N.S. Melhoramento genético do arroz de sequeiro *Oryza sativa* L no Estado do Paraná de 1975 a 1989. Piracicaba: ESALQ, 1991.141p. (Tese de Doutorado em Genética e Melhoramento de Plantas).
- ABDELNOOR, R.V.; BARROS, E.G.; MOREIRA, M.A. Determination of genetic diversity within brazilian soybean germplasm using RAPD technique and comparative analysis with pedigree data. *Revista Brasileira de Genética*. Ribeirão Preto, v.18, n.2, p.265-273, Jun, 1995.
- ARUNACHALAN, V. Genetic distance in plant breeding. *Indian Journal Genetics and Plant Breeding*, New Delhi, v.2, n.41, p.226-36, 1981
- AUTRIQUE, E.; NACHIT, M.M.; MONNEVEUX, P.; et al. Genetic diversity in durum wheat based on RFLP, morphophysiological traits, and coefficient of parentage. *Crop Science*. Madison, v.36, n.3, p.735-742, May-Jun, 1996.
- CAMARGO, O.B.A. Estudo agrobotânico de culturas de arroz (*Oryza sativa* L), em diferentes épocas de semeadura. Piracicaba: ESALQ, 1982. 105p. - (Dissertação de Mestrado em Fitotecnia).
- CAMPBELL, L. G.; KERNE, J. J. Cultivar x environmental interaction in sugarbeet yula trials. *Crop Science*. Madison, v.22, n.2, p. 932-935, 1982.
- CANNINGS, C.; HOPPENSTEADT, F. *An introduction to mathematical taxonomy*. Cambridge University Pres, 1982. 152p.
- CARBONERA, R. Heterose e divergência genética em genótipos de arroz de sequeiro (*Oryza sativa* L). Piracicaba: CENA, 1990. 104p. (Dissertação de Mestrado em Energia Nuclear na Agricultura).
- CENTRO INTERNACIONAL DE AGRICULTURA TROPICAL. CIAT, Sistema de evaluación estandar para arroz. Programa de Pruebas Internacionales de Arroz para America Latina. Cali, s.d. 62p. 1975.
- CHANG, T.T.; BANDENAS, E.A. The morfology and varietal characteristics of the rice plant. Los Baños, IRRI, 1965. 40p. (Technical Bulletin, 4)

- CRUZ, C. D.; REGAZZI, A. J.** Métodos biométricos aplicados ao melhoramento genético. Viçosa: UFV. 1994. Cap.6, p.287-324.
- CUEVAS-PERES, F.E.; GUIMARAES, E.P.; BERRIO, L.E.; et al.** Genétic base of irrigated rice in Latin America and the Caribbean, 1971 a 1989. *Crop Science*, Madison, v.32, n.4, p.1054 - 1059, 1992.
- DAHER, R.F.** Diversidade genética e enzimática em capim-elefante *Pennisetum purpureum* Scchum. Viçosa: UFV, 1993. 110p. (Dissertação de mestrado em Agroquímica).
- DEMÉTRIO, C.G.B.** Análise multidimensional para dados de cana-de-açúcar. Piracicaba: ESALQ, 1985. 144p. (Tese de Doutorado em Genética e Melhoramento de Plantas).
- DUDLEY, J.W.** Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. *Crop Science*, Madison. V.31, n.4, p.718-723, Jul-Ag, 1994.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA.** Centro Nacional de Pesquisa de Arroz e Feijão (Goiânia, GO). Manual de métodos de pesquisa em arroz. Goiania, CNPAF, 1977. 106p.
- 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. (CIRCULAR TÉCNICA, 14).
- GUPTA, A.K.; MITAL, R.K; ZIAUDDIN, A.; et al.** Genetic divergence analysis in sprint wheat (*Triticum aestivum* L). *Indian Journal of Genetics and Plant breeding*, Nova Dehli, v.4 n.56, p.556-562, 1996.
- HILLIS, D.M.** Molecular versus morphological approaches to systematics. *Annual Review of Ecology and Systematics*, Palo Alto, v.18, p.23-42, 1987.
- JARADAT, A.A.** Phenotypic divergence for morphological and yeld-related traits among landrace genotypes of durum wheat from jordan. *Euphytica*. Wageningen, v.52, n.3, p.155-164, Oct, 1991.
- JOHNSON, R. A.; WICHERN, D.W.** Applied multivariate statistical analysis. Englewood Cliffs: Prentice-Hall, 1982. 593p.

- JULQUIFAR, F.R. **Numerical taxonomy: the principles and practice of numerical classification.** San Francisco: W.H. Freeman, 1992. 513p.
- JULQUIFAR, F.R. Genetic divergence in Rice. **The Indian Journal of Genetics and Plant Braecding.** New Delhi, v.37, n.3, p.395- 402, 1981.
- KAW, R.N. Analysis of divergence in some cold-tolerant rices. **The Indian Journal of Genetics and Plant Braecding.** New Delhi, v.55, n.1, p.84-99, Jan, 1995
- KIM, H.S.; WARD, R.W. Genetic diversity in Eastern USA soft winter wheat (*Triticum aestivum* L) basead on RFLPs and coefficient of parentage. **Theoretical and Applied Genetics.** Berlin, v.94, n.4, p.472-479, 1997.
- LOPES, A. M. **Análise genética dos componentes de produção num dialelo entre seis cultivares de arroz (*Oryza sativa* L) em dois regime hidricos.** Viçosa, UFV, Imp. Univ., 1984. 135p. (Tese - Doutorado em Genética e Melhoramento de Plantas).
- MACKILL, D.J.; ;LEI, X, Genetic variation for traits related to temperature adaptation of rice cultivars. **Crop Science.** Madson, v.37, n.4, jul-Ag, 1997.
- MARDIA, K.V.; KENT, J.T.; BIBBY, J.M. **Multivariate analysis.** London, Academic Press, 1979. 512p.
- MARTIN, J.M.; TALBERT, L.E.; LANNINH, S.P.; et al. Hybrid performance in wheat as related to parental diversity. **Crop Science.** Madison, v.1, n.35, p.104-108, 1995.
- MAURYA, D.M.; SINGH, D.P. Genetic divergence in Rice. **The Indian Journal of Genetics and Plant Braecding,** v.37, n.3, p.395- 402, 1977.
- MONTALVAN DEL AGUILA, R. **Determinação de divergencia genética em germoplasma de arroz (*Oryza sativa* L) através de eletroforética de proteínas de grão.** Piracicaba: ESALQ, 1990. 80p. (Dissertação de Mestrado em Genética e Melhoramento de Plantas).
- MORAIS, O.P. de. **Análise multivariada da divergencia genética dos progenitores, índices de seleção e seleção combinada numa população de arroz oriunda de intercruzamento, usando macho-esterilidade.** Viçosa:

- UFV, 1992. 251p. (Tese de Doutorado em Genética e Melhoramento de Plantas).
- MURPHY, J.P.; PHILLIPS, T.D. A comparison of RAPD and isozyme analyses for determining the genetic relationships among *Avena sterilis* L. accessions. *Theoretical and Applied Genetics*. Berlin, v.87, n.5, p.689-696, Oct, 1994.
- NADAF, S.K.; SING, I.S. Genetic distance and multiple disease resistance studies in rice (*Oryza sativa* L). *International Rice Res. Newsletter*, v.11,n.4, p.16-17, 1986.
- NEWBURY, H.J.; FORD-LLOYD, B.V. The use of RAPD for assessing variation in plants. *Plant Growth Regulator*. v. 12, p.43-51, 1993.
- PETER, K. V.; RAI, B. Genetic divergence in tomato. *Indian Journal of Genetics and Plant Breedings*, New Delhi, v.36, n.3, p.379-383, 1976.
- RAO, A. V. A.; PRASAD, A. S. R.; SAI KRISHNA, T.; et al. Genetic divergence among some brown planthopper resistant rice varieties. *Indian Journal of Genetic and Plant Breeding*, v.2, n.2, p.179-85, 1981.
- RESENDE, R.M.S. Aplicação de técnicas de análise multivariada e eletroforese de Isoenzimas em estudos de relações fenéticas no gênero *Laelia* seção *Parviflorae*. Piracicaba: ESALQ, 1991. 130p. (Tese de Doutorado em Genética e Melhoramento de Plantas).
- SANTOS, P.G. Estimativas de parâmetros genéticos e fenotípicos em populações segregantes de arroz irrigado por inundação contínua. Lavras: UFLa, 1996, 75p. (Dissertação de Mestrado em Genética e Melhoramento de Plantas. ).
- SINGH, M.; CECCARELLI, S.; HAMBLIN, I. Estimation of heritability from varietal trials data. *Theoretical and Applied Genetics*, Berlin, v.86, n.2, p. 437-441, Jul, 1993.
- SINGH, A.K.; SINGH, S.B.; SINGH, S.M. Genetic divergence in scented and fine genotypes of rice (*Oryza sativa* L.). *Annals of Agricultural Research*. v.17, n.2, p.163-166, 1996.

- 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, v.2, n.34, p.141-150, 1989.
- SNEATH, P.; SOKAL, R.R. *Nemerial taxonomy.:the principles and practice of numerical classification*. San Francisco: W.H. Freeman, 1973. 513p.
- SOUZA, N. R. *Divergência Genética e correlação em cultivares de arroz (Oryza sativa L.) em diferentes níveis de competição com plantas daninhas*. Viçosa: UFV, 1990. 83p. (Dissertação de Mestrado em Genetica e Melhoramento de Plantas).
- TINGEY, S.V. .RAFALSKI, J.A.; WILLIAMS, G.K. Soybean genome analysis: DNA polymorphism are identified by oligonucleotide primers of arbitrary sequence. In: HERMANN, R.G. e LARKINS, B.A. (eds). *Plant Molecular biology 2*. New York, Plenum Press, 1991. P. 263-268.
- VALLS, J.F.M. Caracterização morfológica, reprodutiva e bioquímica de germoplasma vegetal. In: ENCONTRO SOBRE RECURSOS GENÉTICOS, I, Jaboticabal, 1988. *Anais... aboticabal, UNESP/FCAVJ*, 1989, p.106-128.
- VAN BEUNINGEN, L.T.; BUSCH, R.H. Genetic diversity among North American Spring wheat cultivars. I. Analysis of the coefficient the parentage matrix. *Crop Science*. Madison, v.37, n.2, p.479-483, Mar - Apr, 1993.
- VIANA, J. M. S. *Divergência genética, estabilidade e adaptabilidade de clones de cana-de-açúcar (Sacharum spp.)*. Viçosa: UFV, 1990. 108 p. (Dissertação de Mestrado em Agroquímica).
- WAUGH, R.; BAIRD, E.; POWELL, W. The use of RAPD markers for the detection of gene introgression in potato. *Plant Cell Report*. New York, v.17, n.11, p.466-469, Jun, 1992.
- YOSHIDA, S. *Fundamentals Of rice crop science*. Los Bãnos, IRRI, 1981. 269p.

## CAPÍTULO 3 -

### Uso de Marcadores Moleculares (RAPD) no Estudo da Divergência Genética em Acessos de Arroz

#### RESUMO

O DNA total de plantas jovens de trinta seis acessos de arroz foi extraído, em seguida purificado e amplificado com uso de 13 decanucleotídeos iniciadores “primers” de seqüência aleatória. Dentre estes, apenas oito iniciadores produziram bandas polimórficas e monomórficas intensas. Foram tidos como positivos os seguintes “primers” (OP: A02, A10, A20, B01, B02, C20, L18, N05 ), obtendo-se 467 produtos amplificados, para o conjunto de acessos de arroz avaliados, ou seja, de 1 a 4 por “primer” grau médio de polimorfismo produzido. Apenas o acesso (A25) não gerou produtos amplificados. A análise da divergência genética foi estimada com auxílio do complemento aritmético de Jaccard ( $C_{ij} = 1 - J_{ij}$ ) e adotando o método de agrupamento UPGMA. As distâncias genéticas entre os 36 acessos analisados dois a dois variaram de 0 - 76%. Com base nessas distâncias construiu-se dendrograma que possibilitou a separação dos acessos em seis grupos em nível de 50% de distância relativa. Devido à grande aglomeração dos acessos dentro dos grupos, foi possível subdividi-los em subgrupos para que se identificassem os acessos mais próximos, o que pode ser transformado em um único acesso, evitando assim a duplicidade de amostras na coleção de germoplasma do arroz. Essas informações obtidas são de utilidade para orientação na conservação das coleções de germoplasma existentes.

**Palavras Chaves:** RAPD, descritores morfo-agronômicos, germoplasma, análise multivariada, variabilidade genética, cereal e *Oriza sativa*



## Use of Molecular Markers (RAPD) in the Genetic Divergence Study in Rice Accessions

**ABSTRACT** - Total DNA of young plants of each one of the 36 rice accessions was extracted, next it was purified and amplified by using 13 primers decanucleotide of random sequence. Among these, only eight primers which generated intense polymorphic and monomorphic bands were taken into consideration. The following primers ( Op: AO2, A10, A20, BO1, BO2, C20, L18, NO5) were regarded as positive and optimum, obtaining so 467 amplified products for the set of rice accessions evaluated, namely, 1 to 4 per average degree primer of polymorphisms produced. Only the access (A25) did not generate any amplified product. The analysis of genetic divergence was estimated with the aid of Jaccard's arithmetic complement ( $C_{ij} = 1 - J_{ij}$ ) and by adopting the UPGMA clustering method. The genetic distances among the 36 accessions analysed two by two ranged from 0-76%. On the basis of those distances, a dendrogram which enabled the separation of the accessions in six groups at a 50% level of relative distance was built. Due to the great clustering of the accessions inside the groups, it was possible to subdivide them into subgroups for the closest accessions to be identified, what may be transformed into a single accession, avoiding this way the duplicity of samples in the rice germoplasm collection. This information obtained are of use for guidance in the conservation of existing germoplasm collection.

**Index Terms:** RAPD, markers, morphoagronomics, germoplasma, genetic variability, cereal and *Oriza sativa*.

### 3.1 - INTRODUÇÃO

O cultivo do arroz no Brasil é amplo e em várias diversidade de ecossistemas, de inundado até terras altas; desde cultura de subsistência, com total ausência de tecnologia e insumos modernos, até um estágio, com processo de produção altamente mecanizado, desde o preparo do solo até a colheita, com alto nível de produtividade.

A utilização de cultivares mais modernas que surgem dos programas de melhoramento e também devido à aplicação tecnológica, traz como consequência o desuso de muitas cultivares. Pelo lançamento de novas cultivares, surge a preocupação de conservar-se o germoplasma, visando diminuir a erosão genética. Por isso, deve-se manter em BAG a maioria possível de acessos existentes (Landry et al., 1992; Cuevas-Peres et al., 1992).

O sucesso na conservação de germoplasma depende da disponibilidade de técnicas confiáveis para a caracterização dos indivíduos de uma população, permitindo o descarte de acessos-duplicatas no BAG. Além disso, possibilita comprovar a existência da diversidade genética no material disponível, para estimar a variabilidade e chances de sucesso, em qualquer trabalho de melhoramento (Lanza et al., 1997).

Há possibilidade de aplicação das técnicas moleculares tanto na conservação e manutenção de acessos de uma espécie, bem como na caracterização da variabilidade genética. Isso contribui para a síntese de híbridos ao concentrar somente as possíveis combinações promissoras (Kim e Ward, 1997; Rohrer et al., 1994). Coleções de germoplasma onde a quantidade de marcadores convencionais é reduzida e o custo de conservação é elevado, cada vez mais se utiliza dos marcadores moleculares, identificando por esse método indivíduos que assegurem a base genética tão ampla quanto possível, e

possibilitando também o descarte das duplicatas.

A melhor maneira para estabelecer relações de ligação gênica é analisar a segregação de uma população para um grande número de marcadores (Kim e Ward, 1997). Técnicas de genética molecular, como RFLP (polimorfismo no comprimento do fragmento de DNA) e RADP (polimorfismo de DNA amplificado ao acaso) são capazes de fornecer este grande número de marcadores (Williams et al., 1990).

O desenvolvimento das técnicas de marcadores moleculares em estudos de genética e na prática de melhoramento de plantas se tornou uma ferramenta valiosa na caracterização de indivíduos ou população ("fingerprints"), possibilitando com isso a seleção indireta precoce para o rastreamento de caracteres de interesse ao melhoramento genético. Possibilita a detecção da variação mutacional ao nível de DNA, nas regiões codificadas e não codificadoras do genoma (Kim e Ward, 1997).

Os princípios básicos envolvidos na obtenção e detecção de cada classe de marcador molecular são descritos em vários exemplos ou fornecidos como ilustração. Uma análise comparativa das características das principais classes de marcadores moleculares utilizadas na caracterização de acessos e melhoramento genético de plantas é apresentada por vários autores e detalhada na discussão específica de cada classe de marcador molecular (Weissenbach et al., 1992).

Para várias espécies de plantas, tem-se usado principalmente de isoenzimas, proteínas e os marcadores de DNA ou genético para identificar genes responsáveis por características qualitativas e quantitativas (Landry et al., 1992).

Como as isoenzimas e proteínas são limitadas em número de marcadores, além de apresentarem na maioria das vezes, expressão específica que é afetada facilmente pelo ambiente, pesquisadores propuseram a utilização dos marcadores

de DNA. Essa técnica tem mostrado grande avanço em potencial, para detecção da diversidade entre seres vivos, uma vez que o DNA é a expressão verdadeira do indivíduo sem os efeitos da interação ambiental.

Atualmente, técnicas de genética molecular estão sendo empregadas juntamente com técnicas de automação no estudo do melhoramento convencional, a fim de obter resultados de forma mais rápida, contribuindo para a caracterização dos indivíduos, visando a conservar os germoplasma, bem como atuar ao lado da domesticação das espécies.

Os marcadores moleculares tem recebido muita atenção e são quatro as técnicas mais citadas e que estão em uso no mapeamento e na identificação de indivíduos em plantas: RFLP (Bernatzisk e Tanksley, 1986); RAPD (Williams et al., 1990); AFLP (Amplificação de Comprimentos de Fragmentos Polimórficos) comentados por Zabeau, (1993); mini e microssatélites (Weissenbach et al., 1992); sabe-se que essas técnicas baseiam-se na análise do ácido desoxiribonucléico que, normalmente, é extraído de amostras de folhas ou brotações, sementes, embriões e outras partes das plantas (Lanza et al., 1997; Brummer et al., 1991).

As principais vantagens do uso do método baseado em marcadores RAPD são de não precisar conhecer as genealogias dos parentais e a obtenção de estimativas de diversidade isentas de efeitos ambientais; esta última é comum aos coeficientes de parentesco (Lindhout et al., 1994). Sabe-se que, esses marcadores apresentam herança mendeliana simples, assim sendo, o fenótipo de presença e ausência de bases, expresso por um acesso deve coincidir com aquele presente em pelo menos um de seus progenitores:

a) pode-se usar grande número de iniciadores em uma grande variedade de espécies; não há necessidade de trabalhos iniciais quanto ao isolamento e

marcação de sondas, a digestão de DNA, b) cada marcador é equivalente a um sítio alvo da seqüência de DNA, o que simplifica a informação, c) pode ser automatizado, o que leva a obtenção mais rápida de mapas genéticos; d) utiliza pequena quantidade de DNA; gera um grande número de polimorfismos entre genótipos bem próximos; e) menos oneroso que a técnica de RFLP.

Os marcadores RAPD, quando usados para avaliação da diversidade genética, podem fornecer subsídios úteis à preservação de germoplasma, bem como auxiliar na seleção de populações básicas a serem utilizadas no melhoramento vegetal (Lindhout et al., 1994), isso, porque os marcadores geram grande quantidade de caracteres adicionais que, combinados com características fenotípicas, fornecem um quadro mais completo para o agrupamento de genótipos e o planejamento dos cruzamentos.

Comparados ao RFLP, os marcadores RAPD apresentam algumas vantagens. Dado o número reduzido de passos, as análises com RAPD são menos trabalhosas, possibilitam a obtenção de resultados em tempo bastante inferior ao obtido com RFLP. Além disso, as análises com RAPD exigem quantidade de DNA muito menor, possibilitando o uso de técnicas de extração de DNA de melhor qualidade e rapidez. Outra vantagem se refere ao fato dos "primers" utilizados em uma espécie poderem ser usados em qualquer outra, seja de origem animal, vegetal ou, microorganismos, ao passo que as sondas de RFLP geralmente são específicas (Newbury e Ford-Lloyd, 1993).

Tanksley (1993) comenta que com o emprego dos marcadores moleculares pode-se acelerar a caracterização biológica e a interação de dados de origem diversa tende a ampliar o sucesso nessa etapa, e que a escolha das metodologias precisas, deve-se levar em conta as limitações físicas, temporais, humanas e financeiras.

O objetivo do presente trabalho foi avaliar a diversidade genética entre acessos nacionais de arroz, coletados em vários Estados e com nomes regionais 'três meses', por meio da utilização de marcadores moleculares RAPD, e assim fornecer subsídios para facilitar a incorporação de acessos ao BAG de arroz.

### **3.2 - MATERIAL E MÉTODOS**

#### **3.2.1 - Preparo do Material Vegetal para Extração de DNA Total**

Os acessos de arroz utilizados neste trabalho foram os mesmos utilizados no capítulo 2, Tabela 2, ressaltando que as sementes foram conseguidas no BAG da Embrapa Arroz e Feijão, as quais foram plantadas em vasos de 500 g de substrato preparado com uma mistura de terra + esterco + areia (3:2:1) e conduzidas em casa de vegetação do DAG/UFLA. Decorridos 30 dias da germinação, aquelas plantas vigorosas (3- 4 plantas/acesso) foram colhidas e colocadas em tubos de vidro com tampa de baixa pressão, e de imediato colocados a -80 °C por 24 h. Para inativação das reações de enzimas e catalisadores. Posteriormente, foram colocadas para liofilização a -40 °C por 48 h, e mantidas em dessecadores com sílica gel, até o momento da maceração.

A extração do DNA total foi realizada empregando-se 1-2g do material vegetal jovem e liofilizado, realizada com base no protocolo descrito por Doyle e Doyle (1990), com pequenas modificações, visando a obter uma maior quantidade de DNA de melhor qualidade. O material vegetal liofilizado foi macerado em gral de porcelana contendo nitrogênio líquido + 1 mg de PVP 1% (polivinilpirrolidone) e areia fina esterilizada, obtendo-se um pó fino.

Adicionaram a esse pó 5 ml de tampão de extração CTAB 2% (brometo de hexadeciltrimetil amônio); Tris-HCl 20 mM; (pH 8,0), 20 mM de EDTA; 1,4 M NaCl; 1% de PVP; 0,5% de sarkosyl; acrescido de 20 µl de 2-β-

mercaptoetanol, previamente aquecido a 65 °C, em banho-Maria, pH 8,5 (para fugir do pH ótimo de enzimas que reagem na presença de oxigênio). Posteriormente o macerado foi transferido para tubos de polivinil (50 ml), misturando-se gentilmente e com auxílio de uma pipeta foram adicionados mais 5 ml do tampão de extração CTAB (2%). O material foi incubado a 65 °C, por 30-40 min, agitando levemente por duas vezes para homogeneização; vertendo-se os tubos algumas vezes e estes foram colocados na água até atingir temperatura ambiente.

A seguir procedeu-se a extração, quando as proteínas foram removidas pela adição com igual volume de clorofórmio/álcool-isoamílico (24:1), agitando levemente os tubos por 5 min e posteriormente centrifugados a 10.000 rpm durante 5 min. Recolheu-se o sobrenadante, e foi feita uma segunda desproteínação, com o mesmo volume de clorofórmio/álcool-isoamílico. Os ácidos nucleicos foram precipitados da fase aquosa, com adição de 2/3 do volume de uma solução gelada da mistura álcool 95% + acetato de amônio 7,5 M (6:1), a qual foi mantida a -20 °C até que o DNA se precipitasse (2 -12 h). Após precipitação do DNA, que foi coletado com auxílio de pipeta com ponta adaptada, foi feito tratamento com proteinase K (10 mg/ml) a 37 °C, por 30 min. Em seguida o DNA foi transferido para um tubo tipo "eppendorf" de 1,5 ml, centrifugado a 5000 rpm, durante 10 min, seco ao ar e ressuspensão em 300 µl com tampão TE (10 mM Tris-HCl, 0,1 mM EDTA), pH 8,0. Novamente, no conteúdo dos tubos foram efetuadas três sucessivas precipitações em álcool absoluto, centrifugado e seco, para purificação, finalmente efetuada a ressuspensão do DNA, para volume de 300 µl com TE, a pH 8,0, para formar a solução estoque.

Com o DNA extraído de 36 acessos de arroz formou-se um conjunto de

DNA (10  $\eta$ g) que foi quantificado e depois usado como molde para as reações de amplificação na seleção de iniciadores de DNA (primers). Após a resolução dos produtos da amplificação, estes foram separados por eletroforese em géis de agarose 1,0%. Corados com brometo de etídio (0,5  $\mu$ g/ ml) e depois exposto à luz ultra-violeta.

### **3.2.2 - Teste em Eletroforese de Gel de Agarose e Densidade do DNA Extraído de Plantas Jovens**

A quantificação das amostras de DNA-estoques foi estimada pela relação A260/A280  $\eta$ m, feitos em espectrofotômetro VU/VIS, para leituras de absorvância a 260 e 280  $\eta$ m em que cada unidade de absorvância corresponde à concentração de 50  $\mu$ l/ml de DNA de fita dupla (Sambrook et al., 1989); enquanto que a integridade foi avaliada por eletroforese em gel de agarose 0,8% em TE a pH 8,0 juntamente com o DNA padrão de concentração e pureza confiáveis, como padrão do laboratório da Embrapa Recursos Genéticos e Biotecnologia. Para visualização das bandas, o gel foi corado com brometo de etídio (0,5  $\mu$ g/ml) e escaneado para tela de vídeo de microcomputador e impresso em papel próprio para scanner.

### **3.2.3 - Processo de Amplificação dos Fragmentos de DNA e Eletroforese em Gel**

Após testar a quantidade dos componentes da mistura, variaram-se as concentrações até que se chegasse a um protocolo que funcionasse; para se processarem as reações de amplificação, que foram feitas em um volume total a fim de processar as 36 reações (acrescidos 10% do volume das reações). A reação de amplificação foi baseada no método descrito por Williams et al. (1990) usando primers de 10 bases de sequência arbitrária, procedendo algumas modificações, onde as reações foram otimizadas para obtenção de produtos



amplificados claros. Cada reação continha 3 µl de DNA de cada acesso de concentração 10 ng/µl e mais 10 µl de uma mistura (2,8 µl de água destilada; 2,0 µl de tampão Pharmacia, 2,0 µl de cada um dos deoxinucleotídeos (dATP, dTTP, dGTP e dCTP), 3,0 µl do olinucleotídeo iniciador “primer” e uma unidade da enzima Taq-polimerase Pharmacia, obtendo um volume final de 13 µl de reação. A amplificação foi efetuada no termociclador Modelo “Perkin-Elmer (Gene Amp PCR System 2400)”, programado para 1 ciclo a 94 °C por 5 min e mais 45 ciclos. Cada ciclo consistiu temperatura de desnaturação 94 °C por 15 segundos, alongamento da cadeia a 36 °C por 30 segundos e pareamento (renaturação) a 72 °C por 1 min. Após os 45 ciclos foram acrescentados mais dois ciclos finais para manutenção, a 72 °C por 5 min, finalmente, a temperatura baixava a 4 °C, mantendo-se nessa temperatura até que as amostras fossem retiradas. Os produtos de amplificação foram separados em géis 1,0% de agarose e corados com brometo de etídio 0,5 µg/ml, visualizados em transluminador de luz UV.

Para teste e seleção de primers foram utilizados 25 iniciadores de 10 nucleotídios, entre estes, fez-se o aferimento de uma listagem específica, gentilmente nos cedida por Embrapa Recursos Genéticos e Biotecnologia e indicadas para reação de amplificação do DNA de plantas de arroz, foi aproveitado daqueles disponíveis no laboratório de sementes do DAG/UFLA.

Os primers utilizados foram selecionados dentre aqueles citados em literatura, para o arroz, que é uma espécie de ótima habilidade para sofrer amplificação do DNA com marcadores moleculares (Dolyle e Doyle, 1990), e consistiram dos 25 primers seguintes: A01, A02, A04, A10, A15, A17, A20, B01, B02, B08, C14, C17, C20, D04, D12, D16, K03, K13, L18, M04, M07, N05, N07, N10, N15, todas da marca “Operon Technologies”.

Em seguida à amplificação, os fragmentos obtidos foram separados por eletroforese em gel de agarose 1,5% (preparado em microondas por 3 cocções nos tempos de 4'2'1' seqüencialmente, sob baixa potência). Após a temperatura do gel atingir a 55 °C, despejava-se o seu conteúdo na cuba de acrílico com os pentes, para gelatinização e formar as fossas. Posteriormente, o gel foi imerso em tampão TBE (90 mM de Tris-borato, 1 mM de EDTA pH 8,5) 10X TBE. A corrida foi em gel de agarose (1,5%) em tampão TBE 10X (Tris-Borato 0,09 M e EDTA 0,002 M), por 2 h a 80 volts e 8 Ampères, tempo suficiente para atingir o extremo do gel.

Para estimar os tamanhos dos fragmentos foram utilizados os padrões de tamanho do DNA do bacteriófago "lambda", clivado com as enzimas de restrição Hin III, Bam HI e Eco RI de tamanho de fragmento em pares de bases (pb). Corados e visualizados como anteriormente (Landry et al., 1992).

### 3.2.4 - Tabulação e Análise Estatística dos Marcadores RAPD

As bandas polimórficas mais proeminentes foram tabuladas, codificando-se 1 para presença e 0 para ausência de determinada banda. As variáveis da expressão foram obtidas conforme demonstrado no esquema, expresso na Tabela 9. Essa matriz foi usada para cálculo da similaridade ou dissimilaridade genética em dois acessos comparados.

TABELA 9 - As variáveis da expressão binária foram tomadas conforme expressos na tabela para cada um dos acessos avaliados. UFLA, Lavras - MG, 1999.

Acesso i	acesso j	
	1	0
	a (1,1)	b (1,0)
	c (0,1)	d (0,0)

Os critérios para adotar um índice baseiam-se na inclusão ou não das co-

ocorrências negativas no coeficiente, estando essa inclusão muito relacionada com os tipos de características com que se está trabalhando (Alfenas, 1989). Existem vários índices para interpretar dados do tipo presença ou ausência (dados binários). O índice de Jaccard é um que exclui as co-ocorrências negativas, pelo fato delas não necessariamente representarem semelhança entre os genótipos, que são calculados pela seguinte expressão:

$$J_{ii}' = 100 \cdot \frac{a}{a + b + c} \quad (i)$$

em que:

- a : número de casos que o caracter está presente nas duas populações simultaneamente (1,1);
- b : número de casos em que o caracter está presente somente na população i (1,0);
- c : número de casos em que o caracter está presente somente na população i' (0,1);
- d) os casos em que o caracter está ausente nas duas populações, o qual é descartado.

### 3.2.5 - Estimativas das Distâncias Genéticas pelo Complemento Aritmético de Jaccard (Cii')

Dependendo da sintaxe adotada, a dissimilaridade pode ser estimada pela transformação adotando o complemento aritmético do índice de Jaccard, pela expressão:  $C_{ii}' = (100 - J_{ii}')$ , cujo a soma seja positiva e igual a 100%, a distância genética pode assim ser calculada e comparada em dendogramas pelos pares.

em que:

$J_{ii}'$ : é o coeficiente de Jaccard, e é determinado anteriormente;

$C_{ii}'$ : é a medida de dissimilaridade, através da matriz binária, seguindo o modelo [0 1 0 1 ... 1].

0 é definido como ausência; 1 a presença da expressão fenotípica em cada um dos indivíduos.

### 3.2.5.1 - Análise de Agrupamento pelo Método das Ligações Médias entre Pares não Ponderados (UPGMA)

Adotou-se o Software NTSYS - pc (Rohlf, 1992 ) para os cálculos. Como método de análise multivariada usou-se o índice de similaridade de Jaccard (Jii'), o qual segundo Alfenas et al. (1989), é adotado para fenótipos não interpretáveis, em níveis de locos e alelos. Isso porque o índice é próprio para dados do tipo presença ou ausência de banda polimórfica e agrupados através do método hierárquico não ponderado de agrupamento aos pares pela média aritmética (UPGMA), ou seja, caso de marcadores moleculares, onde não existe uma média legítima para iniciar o agrupamento.

#### a) Técnica de Agrupamento pelo Método 'UPGMA'

Este método requer como similaridade ou parecnça entre os dois grupos aquela dada pelos dois membros mais parecidos (Johnson e Wichem, 1982). Assim, entre todos coeficientes de parecnça entre os elementos de um grupo e de outro, escolhe se o mais semelhante como o coeficiente entre os dois grupos, ou se já, na matriz de distância genética  $D = (dik)$  é identificado o par mais próximo (U e V) de acessos, esses são formar o grupo UV. As distâncias entre UV e qualquer outro grupo W são determinadas por:

$$d(UV, W) = (\sum_i \sum_k dik) / N_{(UV)} N_W$$

em que:

$dik$ : é a distância entre o acesso i no grupo UV e o acesso k no grupo W;

$N_{(UV)}$  e  $N_W$ : são os números de acessos nos grupos UV, respectivamente

Após calcular e identificar a distância entre os pares de acessos mais próximos, formando o primeiro grupo, adota-se o procedimento UPGMA para inclusão de novos acessos no grupo, comparando-se o acréscimo no valor médio da distância dentro do grupo e o nível máximo permitido ( $\alpha$ ) para que a distância média intragrupo seja inferior a qualquer distância intergrupos.

## **b) Elaboração de Dendrograma**

A dissimilaridade entre os acessos foi avaliada com base na distância genética pelo índice de Jaccard (dados qualitativos) e dendrogramas baseados em agrupamento UPGMA para os dados qualitativos.

Esse é um método aglomerativo hierárquico; parte dos  $n$  acessos separados e os agrupam em conjuntos sucessivos até que, na última etapa estejam agrupados em um só conjunto, cujo gráfico é o dendrograma. A técnica adota a distância entre dois grupos como a média das distâncias entre todos os pares de acessos, sendo que os pares são formados por um acesso de cada grupo. Os acessos foram agrupados pela similaridade que é resultante da divergência genética, dada pelo inverso da expressão:

$$DG = (1,0 - S). 100$$

em que:

$S$  é a similaridade entre os pares, com a análise gráfica para facilitar a interpretação dos resultados.

## **3.3 - RESULTADOS E DISCUSSÃO**

A fim de estudar a variabilidade entre os acessos, utilizou-se a técnica de RAPD, sendo que um dos requisitos básicos refere-se à qualidade do DNA, o qual será molde nas reações. O método de extração com CTAB mostrou-se eficiente para os tecidos de arroz, obtendo-se grande quantidade de DNA e de qualidade. Esta foi avaliada ao se submeter as amostras à eletroforese em gel de agarose 0,8%, em TBE 1X pH 8 (0,9 M Tris Borato, 0,002 M EDTA) a 50 volts. O tempo de corrida no gel foi em torno de 40 min, e visualizados em luz UV.

As amostras recém-extraídas foram comparadas com outras amostras de DNA de arroz mantidos como padrão de qualidade e integridade no laboratório da Embrapa Recursos Genéticos e Biotecnologia (Figura 5). Pelos testes qualitativos

e comparativos notou-se pouca degradação e alta concentração no estoque formado.

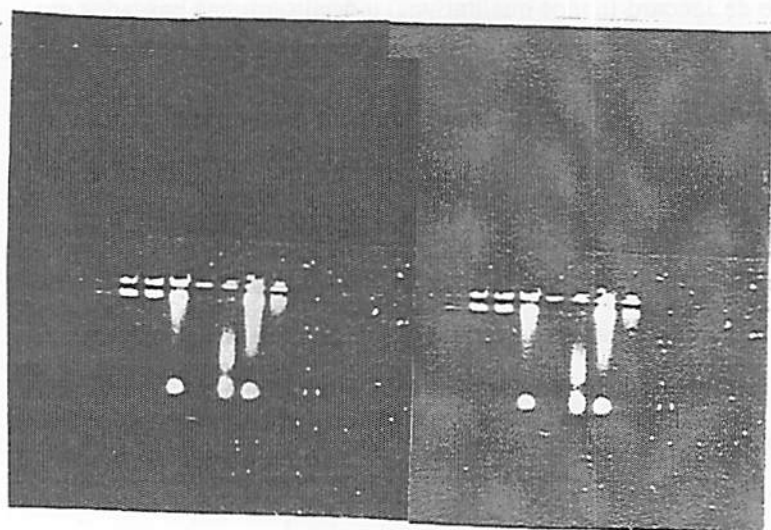


FIGURA 5 - Determinação de integridade de amostras de DNA total por eletroforese em gel de agarose (0,8%), utilizando-se de 3  $\mu\text{g}$  de DNA de arroz e comparadas com amostras de concentrações conhecidas (lado esquerdo). Embrapa Recursos Genéticos e Biotecnologia, Brasília - DF, 1999.

Embora o DNA de algumas amostras tenha apresentado manchas de degradação, as reações de amplificação, em que se utilizaram essas amostras, não foram afetadas, visto que o padrão geral de bandas foi indistinguível daqueles em que o DNA estava íntegro. O que possibilitou a reação foi acertar a concentração que estava muito alta ( $> 800 \mu\text{l}$ ). Para preparo da solução-trabalho procederam-se às leituras e diluições da solução estoque para um volume final de 300  $\mu\text{l}$  com TE, e uma concentração de 1,0  $\eta\text{g}/\mu\text{l}$ .

Pelos resultados expressos na Tabela 10, verifica-se que a concentração de DNA estimada em leituras no espectrofotômetro, absorvância a 260 e 280 $\eta\text{m}$

foi variável entre 51,5 µg a 2,5 mg, e 3,60 µg a 8,28 mg, respectivamente.

TABELA 10 - Leituras no espectrofotômetro a 260 nm e 280 nm, razão 260/280 e concentração de DNA de cada amostra de cada acesso estudado. UFLA, Lavras - MG, 1999.

ACESSO	Absorbância a 260 nm.	Absorbância a 280 nm.	razão 260/280	concentração (µg/µl)
A01	0.6786	0.3464	2.0822	0.9462
A02	0.5152	0.2441	2.1608	0.7382
A04	0.2748	0.1387	2.0808	0.3880
A05	0.0602	0.0305	2.0553	0.0862
A06	1.0060	0.4766	2.1513	1.4490
A07	0.7138	0.3601	2.1198	0.9856
A08	2.0160	1.0669	2.0723	2.7193
A09	0.4591	0.2320	2.0454	0.6616
A10	0.2502	0.1316	2.0982	0.3348
A11	0.7564	0.4040	1.9758	1.0747
A12	0.1154	0.0596	2.0668	0.1602
A13	0.1064	0.0572	2.0662	0.1414
A14	0.5609	0.2658	2.1460	0.8102
A15	0.3188	0.1529	2.1198	0.4620
A16	3.1054	1.6353	1.9579	4.5394
A17	0.5508	0.2699	2.1322	0.7770
A18	0.9038	0.4413	2.1126	1.2939
A19	0.6521	0.3265	2.1140	0.9100
A20	0.3483	0.1823	2.0936	0.4696
A21	0.5649	0.2873	2.0448	0.8090
A22	1.5065	0.7275	2.1061	2.1875
A.23	0.8847	0.4847	2.0936	1.1884
A.24	2.0902	1.0653	2.0287	3.0174
A25	0.2940	0.1478	2.0530	0.4215
A26	0.0769	0.0448	2.1075	0.0901
A27	1.5667	0.7896	2.0581	2.2460
A28	1.3528	0.6710	2.1094	1.9108
A30	0.6675	0.3236	2.1115	0.9628
A31	0.1631	0.0884	1.9212	0.2368
A32	1.8059	0.8967	2.0752	2.8005
A33	0.4286	0.2159	2.0392	0.6219
A34	0.0341	0.0217	1.7336	0.0462
A35	0.0857	0.0481	1.9936	0.1132
A.36	0.0103	0.0072	1.4284	0.0176

A natureza das amostras quando a razão seja entre 1,6 a 2,0, indica que o DNA apresenta boa pureza, ou seja livre de contaminações de proteínas (a razão < que 1,6) e de clorofórmio (razão > 2,0), segundo Rohrer et al. (1994) e Sambrook et al. (1989). Nesse trabalho, a razão variou em todos os intervalos,

naqueles onde a razão estava fora da faixa tolerável procedeu-se mais uma etapa de purificação no DNA estoque, com TBE e outra centrifugação para melhorar a qualidade, ficando a razão próxima aos valores 1,6 a 2,0, indicando que a pureza estava dentro da faixa tolerável.

O arroz cultivado é preconizado como uma espécie de alto grau de similaridade, o que foi confirmado por Cao e Lei (1997) ao caracterizar CV's de arroz americanas, empregando marcadores de RAPD.

### **3.3.1 - Classificação e Identificação dos Marcadores de RAPD**

Do DNA extraído dos 36 acessos de arroz formou-se um conjunto de DNA (1,0 ng de cada) que foi usado como molde para a amplificação com os 25 primers (oligonucleotídeos iniciadores) testados com o conjunto de DNA; alguns mostraram bandas fracas, enquanto alguns não reagiram, mas 13 primers amplificaram o DNA e foram usados para amplificação de cada amostra, considerando apenas aqueles com capacidade de formar reação de amplificação e que geraram pelo menos uma banda polimórfica em cada acesso analisado. Waugh et al. (1992) comentam que, teoricamente, o número de segmentos amplificados do 'primer' e do genoma alvo é baseado na probabilidade que uma dada ausência de DNA (complementar a do primer) possa ocorrer no genoma, em cadeias e orientações opostas dentro de uma distância amplificável pelo PCR.

Como o teste foi realizado com conjunto de DNA de plantas de arroz, alguns iniciadores relacionados formaram polimorfismo de respostas confiáveis (Figura 6). Inúmeros autores citam essa ocorrência, por exemplo: Dudley (1994), Tinker et al. (1993) em cevada; Bentolita et al. (1992), em milho; Singh e Unrea (1995), em feijão; Vasconcelos (1996), Abdelnoor, Barros e Moreira (1995), Tingey, Rafalsky e Williams (1991) em soja.

Com base nos resultados das amplificações de DNA observadas nos



amplificações de DNA observadas nos acessos de arroz com os primers relacionados (Tabela 11), distinguiram-se bem os acessos para procedimentos das análises; entre os primers testados, alguns foram eficientes na amplificação do DNA, nos quais aconteceram bandas intensas sem as obscuras. Totalizando 460 amplificações, com uma média de 1,65 bandas/acesso/primer. Com isso descreveram-se suas respectivas seqüências de bases.

**TABELA 11** - Relação de oligonucleotídios iniciadores (primers) utilizados neste trabalho e suas respectivas seqüências de nucleosídios associados e o número de produtos amplificados (NPF) através do "pool" de DNA extraído de plantas jovens de acessos de arroz, elegendo como padrão para reação o primers A20. UFLA, Lavras - MG, 1999.

Primer OP	Seqüência de bases (5'-3')	NPF	Primers OP	seqüência de bases (5'-3')	NPF
01 - A01	CAGGCCCTTC	4	09 - C20	ACTTCGCCAC	30
02 - A02	GTCCGAGCTC	57	10 - D04	TCTGGTGAGG	14
03 - A04	AATCGGGCTG	13	11 - D12	CACCGTATCC	8
04 - A10	GTGATCGCAG*	60	12 - D16	CACGCGTAAG	7
05 - A20	GTTGCGATCC	54	13 - K13	GGTTGTACCC	4
06 - B01	GTTTCGCTCC	57	14 - L18	ACCACCCACC	35
07. B02	TGATCCCTGG	42	15 - N05	ACTGAACGCC	67
08. B08	GTCCACACGG	8			

**Observação:** Um mesmo "primer" gerou até 4 bandas polimórficas

Os demais A01, A04, B08, D04, D12, D16, K13, não geraram produtos amplificados claramente, por isso não foram considerados. De forma geral, o número de bandas amplificadas foi de 1 a 4 para cada primer utilizado, sendo que o número de bandas polimórficas dominou. Segundo Lanza et al. (1997) podem-se com marcadores RAPD discriminar bandas monomórficas e uma polimórfica de 750 pb de nucleotídios.

A Figura 6 mostra géis de agarose como produtos amplificados provenientes de 35 acessos de arroz, mais o conjunto de DNA, utilizando o primer OP-A20, podendo-se comparar as bandas monomórficas e polimórficas.

É oportuno mencionar que o acesso A25 não formou reação em nenhum dos iniciadores utilizados. Williams et al. (1990) comentam que mudança única em um nucleotídeo no DNA-genômico pode impedir a amplificação toda, mal emparelhamento entre o segmento de DNA e o oligonucleotídeo, embora isso não implique em que todas as amplificações sejam resultado de um perfeito pareamento. Outras fontes de polimorfismo podem incluir deleção ou inserção no sítio de ligação do “primer” que pode torná-lo além da distância amplificável ou mudar o tamanho do fragmento sem impedir sua amplificação.

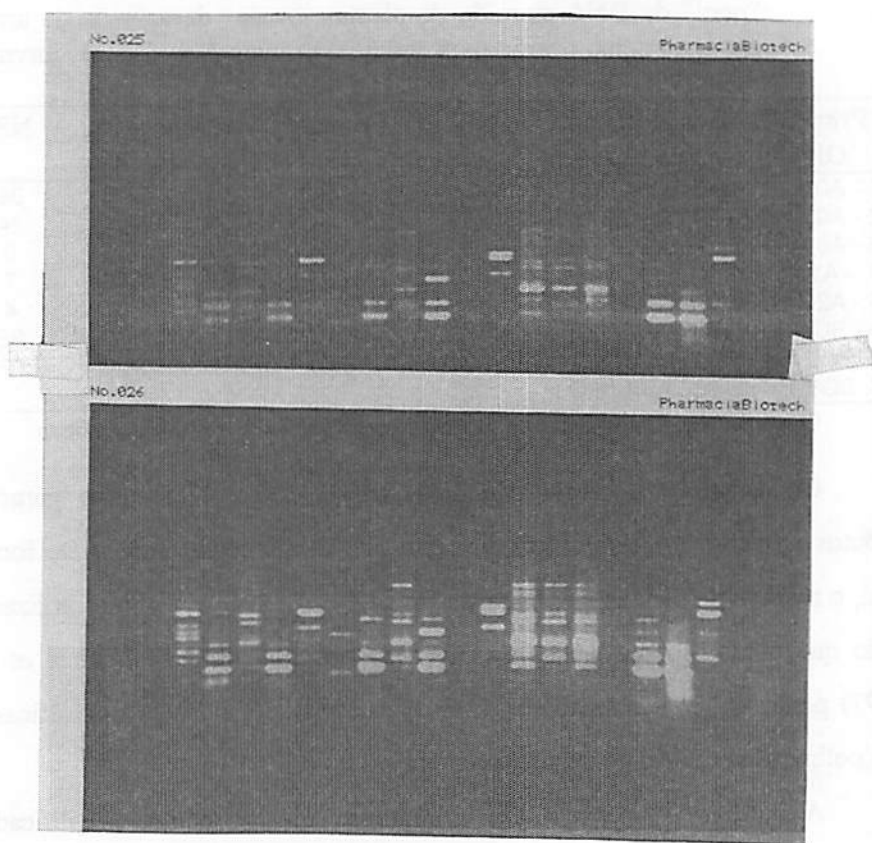
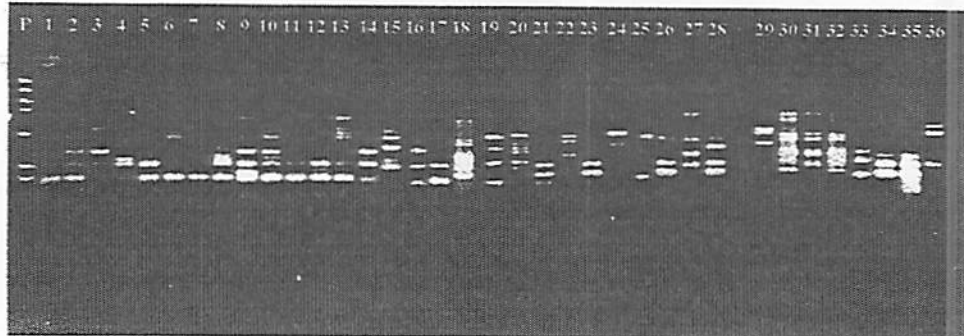


FIGURA 7 - Exemplo de “fringerprinter” de plantas jovens de arroz amplificadas com o “primer” OP-N05. UFLA, Lavras - MG, 1999.

Do total de 36 acessos de arroz avaliados apenas o (A25) não foi amplificado com nenhum dos 13 'primers' positivos e utilizados. Segundo Lindhout et al. (1994), várias hipóteses podem ser levantadas para explicar o fato de que, em alguns locus analisados, o alelo que está presente em um indivíduo não está presente noutro. As explicações mais simples são: segregação quando em híbridos, mutações no genoma e ainda epistásia. Nesse trabalho ocorreu para maioria dos acessos, resultados distintos ao comparar um com outro, não tendo como afirmar para qual hipótese os dados mais tenderam.

Weeden, Muehlbauer e Ladizinsky (1992) comentam que ao comparar bandas fortes e bem definidas existe uma correlação com outros marcadores, tanto como RFLP. As fracas podem produzir erros de 2-7%, dependendo da acuidade desejada dos resultados; devendo-se usar repetições nas análises e interpretar somente bandas intensas. A aplicação dos marcadores RAPD para seleção do germoplasma exige um maior cuidado, pois a quantificação da diversidade pode apresentar valores altos devido a variações não genéticas, as quais estão ligadas a erros inerentes à manipulação da própria técnica.

Neste estudo de avaliação da diversidade genética para os testes de amplificação de DNA do arroz foram realizadas reações com o padrão de amplificação de DNA com seus respectivos pesos moleculares e todos acessos em proporções iguais, com o primer OP-N05, após a resolução dos produtos da amplificação por eletroforese em géis de agarose 1,5%, os quais foram fotografados para ilustração (Figura 7). O RAPD é uma ferramenta confiável para medir a divergência genética, porém a correlação dos grupos com caracteres importantes agronomicamente e heterose ainda é difícil de acontecer ou interpretar. (Lanza et al., 1997).



fotocopia padrão

FIGURA 8 - Padrão em eletroforese dos fragmentos de DNA de acessos de arroz, amplificados com o primer OP-N05. Os números dispostos acima da foto correspondem aos acessos utilizados de acordo com dados da Tabela 2 (pág. 84), números à direita correspondem ao tamanho dos fragmentos de DNA amplificados em pares de base. O padrão está assinalado por p. UFLA, Lavras - MG, 1999.

A análise dos perfis de fragmentos de DNA, de 35 acessos do BAG de arroz, que amplificaram com os primers relacionados como positivos e migraram nos géis de agarose. Os 08 iniciadores citados anteriormente como positivos, permitiram formar uma matriz binária de dados. As bandas avaliadas com o padrão apontaram tamanho variando entre 500 a 3000 pb, as mais intensas a (Figura 8). Esses números estão na faixa de tamanho para o DNA do arroz, que é uma espécie considerada ótima para amplificação por RAPD.

### 3.3.2 - Avaliação da Divergência Genética entre os Acessos de Arroz

Para a estimativa da diversidade genética foram considerados dados daqueles primers positivos, os quais amplificaram pelo menos uma banda polimórfica entre os acessos analisados; transformados em dados binários gerados a partir de marcadores RAPD para cálculo das distâncias estatísticas (dados não apresentados) porém, expressos em dendrograma (Figura 9).

Pelas distâncias detectaram-se maior grau de similaridade; a qual foi transformada em porcentagens do total entre todos os pares possíveis de acessos

de arroz. Com essa análise obteve a estimativa da distância genética entre os acessos a partir de dados binários. Os valores variaram de 0,15 a 0,79, com maior valor para o acesso A05, sendo que um grupo de acessos (A13, A18 e A34) ficou com distância próxima de zero, ou seja, os similares extremos.

A magnitude das distâncias transformada para porcentagem foi ampla, com a mínima de 0,06% (entre os pares A13 e A18; A13 e A34; A18 e A34) e máxima de 72% (entre os pares A5 e A13), estes valores são coerentes e exclusivos para esse conjunto de acessos, de acordo com Rao et al. (1981).

As análises preliminares com marcadores qualitativos já havia mostrado discrepância entre o conjunto de acessos. Os valores para as similaridades genéticas foram obtidos com base nas diferenças encontradas entre dois acessos em relação ao total de bandas analisadas. Pelos relatos de Tingey, Rafalski e Williams (1991), com esse procedimento obtém-se a matriz de similaridade, que fornece as unidades de distâncias relativas, e mostram o grau de parentesco dentro de um grupo, quando um grupo realmente difere do outro.

Isso é importante quando se deseja caracterizar acessos em populações de cultivares, assim pela distância genética podem-se indicar quais acessos são duplicatas. Também Tinker, Fortin e Mather (1993), trabalhando com 33 primers aleatórios, encontraram 19 que evidenciaram polimorfismo em cevada com coeficiente de parentesco em 92%. Concluíram os autores que a predição de similaridade baseada nos marcadores moleculares forneceu mais informações do que mesmo o método de pedigree com marcadores morfológicos. Hu e Quiros (1991) observaram ainda que, as características morfológicas são de interpretação às vezes ambígua, tendo o uso limitado também por envolver, em muitos casos o crescimento das plantas, o que é dispendioso, trabalhoso e susceptível de variações ambientais

### 3.3.3 - Caracterização dos Acessos e Agrupamentos

Procedeu-se a análise de agrupamento pelo método UPGMA com base nos dados de distâncias entre os acessos calculadas pelo complemento aritmético do índice de Jaccard, a qual possibilitou-se a formação de um dendograma (Figura 9). No qual, ao se fazer o corte à distância do intervalo de  $\alpha = 50\%$ , separou o conjunto em 3 grupos ao nível de 70% de similaridade, ficando o grupo I intenso com inúmeros acessos, dificultando relacionar os grupos caracterizados.

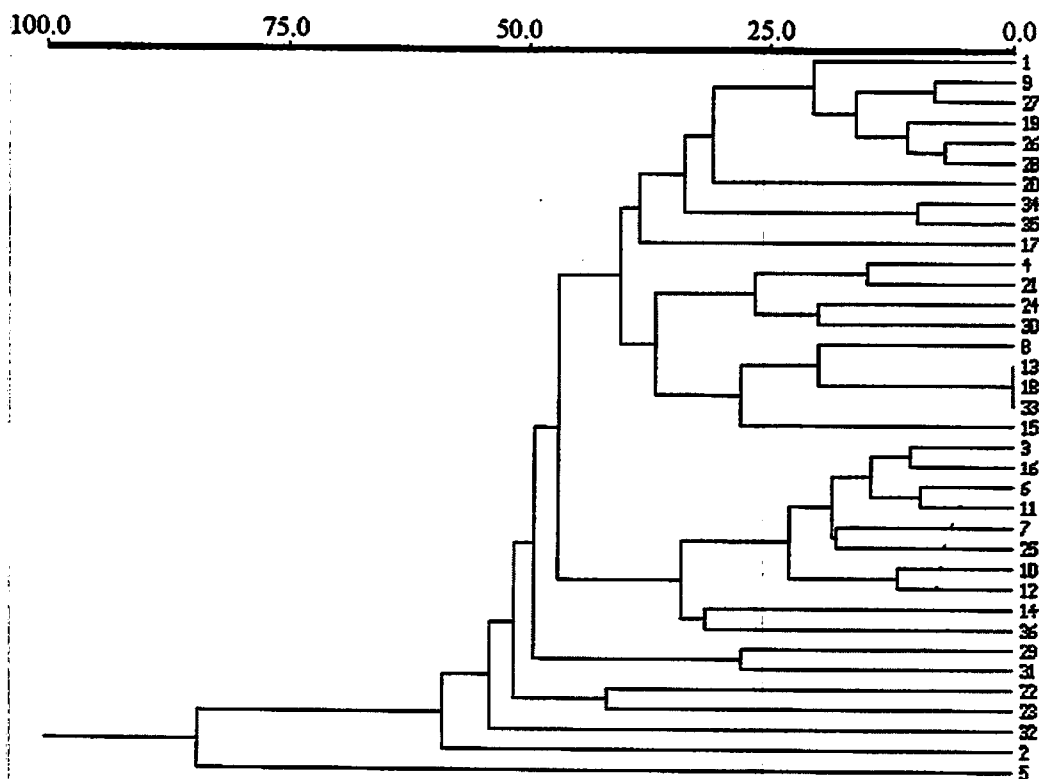
Deve-se ressaltar os relatos de Singh e Unrea (1995) que as informações dispostas em dendrograma são mais fáceis de visualização e tem mais sentido para análise de grupos de genótipos pela síntese das informações, do que avaliar cada grupo isoladamente pelas suas características quantitativas.

Efetuando-se o corte para obter a maior divergência entre grupos, acontece no valor de 50% na escala, formando 6 grupos distintos (A até F), compostos por 3 grupos isolados, 2 grupos com 2 acessos cada, e um grupo intenso, com o restante que se subdivide em 2 subgrupos ao nível  $\alpha$  de 40% de divergência genética; isto não está incoerente, mas por se tratar de acessos (grupo de cultivares) é melhor analisá-los a uma divergência menor, que neste caso será de 25%, para formar grupos e valores menores para os subgrupos.

Passando a adotar  $\alpha$ , de valor menor que 50%, pôde-se agrupar e discutir os resultados, por ter formado subgrupos dentro do grupo I. Nesse comportamento que indicou uma variabilidade não muito alta no conjunto, pode-se interpretar que, com os marcadores moleculares (RAPD), conseguiu-se eficiência e na detecção da variabilidade genética nesse conjunto de indivíduos, que é parente ao nível de cultivar.

Portanto, a hipótese de similaridade, entre todo o conjunto avaliado, que

possui além de homonímia, comportamento bastante próximo, também foi possibilitada com análise de marcadores morfológicos quantitativos (capítulo 2). Por isso foram considerados similares por essa metodologia apenas aqueles que apresentaram menor de 1,0% de distância, o que aconteceu para os pares: A13 e A18; A13 e A34; A18 e A34, para todas as possíveis combinações entre pares de acessos avaliados.



**FIGURA 9 - Dendrograma construído a partir do cálculo da distância euclidiana média entre os acessos de arroz baseados em marcadores moleculares (RAPD), como distâncias simples, dados padronizados e método de aglomeração "UPGMA". UFLA, Lavras - MG. 1999.**

A baixa diversidade do germoplasma de arroz foi comprovada por Tanksley (1993), pelo uso de marcadores moleculares (RFLP), por intermédio da análise de 53 genótipos de arroz. Verificou-se que grande parte dos locos encontrados eram monomórficos entre os genótipos, ou seja, apresentavam o mesmo alelo. Dos locos restantes, considerados polimórficos, aproximadamente 25% apresentaram um tipo alélico presente em mais de 90% dos genótipos estudados. Por outro lado Pinto (1996), avaliando dendogramas de clones de milho comparou 'primers' individuais e não encontrou um agrupamento coincidente para os pares de clones, e comenta que os resultados diferentes são sempre esperados, pois o polimorfismo detectado em um genótipo por um 'primer' não está em outro, ou por outro 'primer'.

Apenas o A5 apresentou a maior divergência genética pelos marcadores moleculares perante o grupo de acessos, uma vez que em todos os pares foi ele que distanciou mais, a exceção dos pares com A9 e A17. As distâncias formadas nos pares com a presença do A5 são superiores a 75%, confirmando que o restante dos acessos possui pouca divergência genética, o que é uma informação coerente pois são todos CV's Obsoletas; certamente sofreram grande pressão de seleção pelos produtores que os cultivaram (Dudley, 1994; Fonseca, Rangel e Prabhu, 1981).

Analisando-se o dendograma (Figura 9) ao nível de 25% de  $\alpha$  (percentual de divergência genética) é possível formar 16 grupos de A até P, e 11 subgrupos com divergência variando de 0% a 23%, valor que é significativo para discriminar acessos, conforme relatos de Tanksley (1993); Singh, Ceccarelli e Hamblin (1993), obtiveram resultados interessantes procedendo estudos com



genótipos de arroz. Ainda ocorreu uma exceção com acesso 'A25' que teve comportamento isolado perante as reações, devido não ter conseguido uma única reação para formar produto de amplificação, com o uso destes primers.

### 3.3.3.1 - Estabelecimento do Grau de Parentesco entre os Acessos

Os grupos foram compostos com base nas divergências genéticas entre os acessos avaliados com esse tipo de agrupamento. Nota-se que há pouca divergência entre os acessos, o que está coerente, pois os acessos são CV's Obsoletas e certamente sofreram grande pressão de seleção pelos produtores que as cultivaram, segundo relatos de Abbud (1991). Uma vez estabelecidos esses grupos de similaridade, pode-se relacioná-los com os caracteres de maior coincidência ou importância.

Baseando-se nessas informações, na prática a conservação e multiplicação de sementes dos acessos podem ser conduzidas em conjunto. Com isso, o número de acessos e espaço ficariam menores, evitando custos, de acordo com grupos e subgrupos de acessos que ficariam assim distribuídos como na Tabela 12.

Comparando-se com os caracteres mais importantes para cada grupo e subgrupo formados, o que se observa na coluna da direita, alguns acessos mostram coincidência, o que pode ser relacionado com as associações importantes da planta. Rao et al. (1981) avaliaram a divergência genética entre 170 variedades de arroz, com relação a nove caracteres agrônômicos. As duas primeiras variáveis (peso de panícula e dias para florescimento) envolveram cerca de 87% da variação total, o que poderia ser correlacionado com os marcadores moleculares, auxiliando na seleção de plantas com base nesses caracteres.

A seleção indireta pode ser usada, com base na resposta correlacionada,

permitindo um progresso mais rápido quando comparada a seleção direta da característica desejada. Porém, surgem dificuldades quando duas características apresentam correlação positiva significativa e uma delas é desejável ou quando as duas características são desejáveis, mas apresentam correlações negativas significativas (Lanza et al., 1997; Singh et al., 1981).

**TABELA 12 - Relação dos grupos e subgrupos formados com base pelos marcadores moleculares (RAPD), adotando o complemento aritmético de Jaccard como divergência genética e agrupamento pelo método UPGMA, para 35 acessos de arroz denominado 'três meses'. UFLA, Lavras - MG, 1999.**

Grupos	$\alpha$ (%)	Subgrupos	ACESSOS	$\alpha$ (%)	Caracteres comuns
A	33	três			
		I-A	A1	23%	Comprimento e largura de grãos Ciclos para florescimento e total
		II-A	A9, A28,	5%	
III-A	A19, A27, A29	7%			
B	33	-	A20		
C	36	-	A35, A36		
D	40	-	A17		
E	42	dois			
		I-E	A4, A21	14%	Peso de 100 grãos e panículas
II-E	A24, A31	23%			
F	26	dois			
		I-F	A8	20%	Caracteres de grãos e ciclo total
II-F	A13, A18, A34	0%			
G	26	-	A15		
H	34	quatro			
		I-H	A3, A16	10%	Dias para emergência
		II-H	A6, A11	8%	Idem
		III-H	A7, A26	20%	idem
IV-H	A10, A12	12%	idem		
I	30	-	A14		
J	20	-	A30		
K	26	-	A32		
L	40	-	A22		
M	40	-	A23		
N	55	-	A33		
O	60,8	-	A2		
P	79	-	A5		

Devido ao comportamento de cultivares ser alterado em função do ambiente, os estudos de divergência que se baseiam em um único ambiente têm sido criticados, com isso, a caracterização e divergência por marcadores moleculares é recomendada. A interação genótipo x ambiente pode prejudicar a consistência do agrupamento nos diferentes locais em que são testados (Thormann et al., 1993; Lanza et al., 1997).

Os resultados obtidos por esse procedimento (RAPD) confirmam as inferências anteriores realizadas pelas análises de marcadores moleculares e os quantitativos. Realmente, os acessos divergentes são os que permanecem isolados em um grupo, e como idênticos ou muito similares, os acessos que permanecem dentro de subgrupo e com distância mínima. Observados com os seguintes acessos: CNA860082, 870002 e 880065 (A13, A18 e A34 respectivamente) os quais embora tenham procedências distintas mas com a mesma constituição genética e até mesmo nome regional.

Nos grupos A, E, H formaram-se subgrupos que, ao analisar os caracteres quantitativos, nota-se que são similares dentro de subgrupos nos caracteres que aparecem à direita da Tabela 12. Isso confirma o potencial do RAPD para o estudo da divergência genética entre indivíduos de plantas, que no caso em estudo sabe-se que são CV's Obsoletas; esse potencial foi constatado por Hu e Quiros (1991), os quais com apenas 4 'primers', puderam distinguir 14 CV's de brócolos e 12 de couve-flor dentre centenas de genótipos.

Quanto à procedência geográfica, não aparece em qualquer grupo, apenas os pares A24 e A31 que estão no subgrupo II-E, mostram ter a mesma origem, confirmando assim que o nível de parentesco e a seleção realizada pelos rizicultores foram os que mais contribuíram para a alta similaridade encontrada. Nos estudos de Tingey e Del Tufo (1992) com marcadores RAPD, salientam que

estes apresentam ainda maior capacidade de detectar polimorfismo que os marcadores de RFLP, especialmente para espécies com estreita base genética, como acontece nas autógamias, principalmente. Apresentando como única alternativa para avaliação da variabilidade genética, mapeamento e determinação de sistema de cruzamento em diversas plantas perenes tropicais. Um dos meios para se obter maior efetividade na seleção, principalmente, quando o caráter desejado ainda não está selecionado, pela dificuldade na sua identificação, medição ou baixa herdabilidade, e empregando um caráter correlacionado, com alta herdabilidade e facilmente medido.

### 3.4 - CONCLUSÕES

- a - Os marcadores moleculares (RAPD) produziram baixa quantidade de bandas polimórficas. No entanto, mostram-se eficientes para discriminação, o que permite encontrar a divergência genética entre o conjunto de acessos de arroz.
- b - A obtenção do DNA foi fácil em quantidade e qualidade, no entanto dos 23 primers usados apenas 10 são bandas intensas e outros produzindo bandas leves. Os primers A10, A20 e N05 foram tidos como primers para o padrão.
- c - Pelo agrupamento UPGMA, com base nos dados gerados por RAPD, permitiu-se formar três grupos, evidenciando-se a similaridade para a maioria dos acessos e justificando alta homonímia (A - O) com distância variando de 25% a 75%.
- d - Ao abaixar o nível de divergência para menos de 25% dentro dos grupos, é possível formar subgrupos similares dentro dos grupos A, E, H.
- é - O uso de RAPD confirmou alto grau de similaridade dos acessos de arroz.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ABBUD, N.S. Melhoramento genético do arroz de sequeiro *Oryza sativa L* no Estado do Paraná de 1975 a 1989. Piracicaba: ESALQ, 1991.141p. (Tese de Doutorado em Genética e Melhoramento de Plantas).
- ABDELNOOR, R.V.; BARROS, E.G.; MOREIRA, M.A. Determination of genetic diversity within brazilian soybean germplasm using RAPD technique and comparative analysis with pedigree data. *Revista Brasileira de Genética*, v.18, n.2, p.265-273, Jun, 1995.
- ALFENAS, A.C.; PETERS, I.; BRUNE, W.; et al. Eletroforese de proteínas de fungos e essências florestais. Viçosa, UFV, 1989. 134p.
- BENTOLITA, S.; HARDY, T.; GUITTON, C.; et al. Comparative genetic analyses of F2 plants and anther culture derived plants of maize. *Genome*, v.35, p.575-582, 1992.
- BERNATZIKY, R.; TANKSLEY, S.D. Toward a saturated linkage map tomato based on isoenzymes and random cDNA sequences. *Genetics*, v.112, p.887-98, 1986.
- BRUMMER, E.C.; KOCHERT, G.; BOSTON, J.H. RFLP variation in diploid and tetraploid alfalfa. *Theoretical and Applied Genetics*. Berlin, v.83, n.6 p.89-96, 1991.
- CAO, D.; OARD, J.H. Pedigree and RAPD analysis of comercial US rice cultivars. *Crop Science, Madison*. v.37, n.5, p.1630- 1635, Sep-Oct, 1997.
- CUEVAS-PERES, F.E.; GUIMARAES, E.P.; BERRIO, L.E.; et al. Genétic base of irrigated rice in Latin America and the Caribbean, 1971 a 1989. *Crop Science, Madison*, v.32, n.4, p.1054 - 1059, 1992.
- DOYLE, J.J.; DOYLE, J.L. Isolation of plant DNA from fresh tissue. *Focus*, v.12, p.13-15. 1990.
- DUDLEY, J.W. Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. *Crop Science, Madison*. v.31, p.718-

723, 1994.

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. (CIRCULAR TÉCNICA, 14).

HU, J.; QUIROS, C.F. Identification of broccoli and cauliflower cultivars with RAPD markers. **Plant Cell Report**, v.10, p.505-511, 1991.

JOHNSON, R. A.; WICHERN, D.W. **Applied multivariate statistical analysis.** Englewood Cliffs: Prentice-Hall, 1982. 593p.

KIM, H.S.; WARD, R.W. Genetic diversity in Eastern USA soft winter wheat (*Triticum aestivum* L) based on RFLPs and coefficient of parentage. **Theoretical and Applied Genetics**. Berlin, v.4, n.94, p.472-479, 1997.

LANDRY, B.S.; HUBERT, N.; CRETE, R.; et al. A genetic map for *Brassica oleracea* based on RFLP markers detected with expressed DNA sequences and mapping of resistance genes to race 2 of *Plasmodiophora brassicae* (Woronin). **Genome**. Berlin, v.35, p.409-420, 1992.

LANZA, L.L.B.; SOUZA Jr., C.L. de; OTTOBON, L.M.M.; et al. Genetic distance of inbred lines and prediction of maize single-cross performance using RAPD markers. **Theoretical and Applied Genetics**. Berlin, v.94, p.1023-1030, 1997.

LINDHOUT, P.; HRUSDEN, S.V.; PET, G. et al. Perspectives of molecular marker assisted breeding for earliness in tomato. **Euphytica**. Wageningen, v.79, n.14, p.279-286, 1994.

NEWBURY, H.J.; FORD-LLOYD, B.V. The use of RAPD for assessing variation in plants. **Plant Growth Regulator**. v. 12, p.43-51, 1993.

PINTO, L.R. **Avaliação das relações de similaridade genética entre linhagens de milho com e baixa capacidade de combinação através de marcadores moleculares.** Jaboticabal, FCAV, UNESP, 1996. 76p. (Dissertação de Mestrado em Agronomia).

RAO, A. V. A.; PRASAD, A. S. R.; SAI KRISHNA, T.; et al. Genetic divergence among some brown planthopper resistant rice varieties. **Indian**

- Journal of Genetic and Plant Breeding**. New Delhi, v.2, n.2, p.179-85, 1981.
- ROHRER, G.A.; ALEXANDER, L.J.; KEELE, J.W.; et al.** A microsatellite linkage map the porcine genome. **Genetics**. Berlin, n.136, p.231-245, 1994.
- SAMBROOK, J.; FRITSCH, E.F.; MANIATIS, T.** Molecular cloning: laboratory manual. 2nd edition CSHL, Cold Spring Harbor, NY. 1989. 123p.
- SINGH, D.; UNREA, A.C.;** Inter and intraracial hidridization and selection for seed yield in early generations of common bean, *Phaseolus vulgaris* L. **Euphytica**, Wageningen, v.81, n.21, p.131-137, 1995.
- SINGH, M.; CECCARELLI, S.; HAMBLIN, I.** Estimation of heritability from varietal trials data. **Theoretical and Applied Genetics**, Berlin, v.86, p. 437-441, 1993.
- SINGH, Y. P.; KUMAR, A.; CHAUHAN, B.P.S.** Genetic divergence in pearl millet. **The Indian Journal of Genetics & Plant Breeding**, New Delhi, v.41, p.186-90, 1981.
- TANKSLEY, S.D.** Mapping polygenes. **Annual Review of the Genetics**, v. 27, p.205-233, 1993.
- THORMANN, R.; STEPHAN, B.R.** Interpretation of isozyma patterns of amlate dehydrogenase in scots pine using two different methods. **Silvae Genetica**, Frankfurt, v.42, n.1, p.5-8, 1993.
- TINGEY, S.V.; DEL TUFO, J.P.** Genetic analysis with Random Amplified Polymorphic DNA markers. **Plant Physiology**, v.101, p.349-352, 1992.
- TINGEY, S.V.; RAFALSKI, J.A. e WILLIAMS, G.K.** Soybean genome analysis: DNA polymorphism are identified by olygonucleotide primers of arbitrary sequence. In: **HERMANN, R.G. e LARKINS, B.A. (eds).** **Plant Molecular biology 2**. New York, Plenum Press, 1991. P. 263-268.
- TINKER, N.A.; FORTIN, M.G.; MATHER, D.E.** Randon amplified polymorphic DNA and pedigree relationships in spring barley. **Theoretical Applied Genetics**. Berlin, v.85, p.976-984, 1993.

- VASCONCELOS, M.J.V.; BARROS, E.G.; MOREIRA, M.A.; et al., Genetic diversity of the bean *Phaseolus vulgaris* L. Determined by DNA-based molecular markers. *Revista Brasileira de Genética*. v.19, p.447-451, 1996.
- WAUGH, R.; BAIRD, E.; POWELL, W. The use of RAPD markers for the detection of gene introgression in potato. *Plant Cell Report*, v.11, p.466-469. 1992.
- WEEDEN, N.F.; MUEHLBAUER, F.J.; LADIZINSKY, G. Extensive conservation of linkage relationships between pea and lentil genetic maps. *Journal of heredity*, v.83, p.123-129, 1992.
- WEISSENBACH, J.; GYPAY, G.; DIB, C.; et al. A second-generation linkage map of the human genome. *Nature*. Londres, n.359, p.794-801. 1992.
- WILLIAMS, J.G.K.; KUBELIK, A.R.; LIVAK, K.J.; et al. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acid Research*. v.18, p.6531-535, 1990.
- ZABEAU, M. Selective restriction fragment amplification: a general method for DNA fingerprinting. *European Patent Application*, v.53, p.48-58. 1993.



## CAPITULO 4 -

### **Comparação de Divergência Genética em Arroz Avaliada Através de Marcadores Morfológicos e Moleculares (RAPD)**

#### **RESUMO**

O Arroz é uma espécie de boa performance para pesquisas de laboratório (cultura de células, protoplasto e anteras; técnicas moleculares), como também para a caracterização por meio de marcadores moleculares. O objetivo do presente trabalho foi comparar a caracterização de 36 acessos de arroz (cultivares obsoletas e homônimas) por meio de marcadores morfo-agronômicos qualitativos, com aquelas usando de marcadores morfológicos quantitativos e moleculares (RAPD). A distância genética para RAPD foi realizada com auxílio do índice de Jaccard e agrupamento pelo método UPGMA, enquanto que para os caracteres quantitativos adotou-se da distância generalizada de Mahalanobis e agrupamento pelo método do vizinho mais próximo, com a projeção em dendrogramas. As três metodologias adotadas detectaram a divergência genética nos acessos, sendo que os mais divergentes apresentaram o mesmo comportamento, isolando dos demais. O mesmo aconteceu com os similares (< 5% de divergência genética). Verificou-se que o agrupamento utilizando de caracteres morfo-agronômicos ficou menos conglomerado, obteve-se o maior número de grupos se comparado aos outros métodos, mostrando que os acessos de arroz denominados 'três meses' possuem estabilidade de caracteres e há acessos homônimos, porém divergentes.

**Palavra Chaves:** Agrupamento, acessos, divergência genética, marcadores, RAPD, cereal e *Oryza sativa*.

## **Comparison RAPD and Quantitative, Quantitative Morphological Analyses for Determination of the Genetic Divergence on the Rice.**

**ABSTRACT-** The rice *O. Sativa L.* is a good material for research on cells, protoplast and anthers cultures, molecular techniques and of behavior performance characterizations through molecular markers as well. The objective of this research was compare qualitative morfo-agronomics markers and morphological quantitative and molecular markers (RAPD) to characterization of the 36 rice accessions obsoletes and homonymous cultivars. The genetic distance for qualitative markers and RAPD was done by Jacard's index process and groupment formation by UPGMA, while quantitative characteres were done by Mahalanobis'generalized distance technique and group formation by nearest more adjacent neighbour which dendograms projection. The three methodologies tested in this research detected genetic divergences and the most divergent accessions had the same performance and were detached from the others that had less than 5% of genetic divergence, which also had similar performance among them. Groups established through morpho-agronomic techniques clustered less and had more groups than other methods, indicating that 'three month' rices have characters stability and and that there is homonymous accession, but nevertheless divergents.

**Index Terms:** Cluster, accessions, genetic divergence, markers, RAPD, cereals and *Oryza sativa*

## 4.1 - INTRODUÇÃO

Variedades cultivadas ou cultivares são aquelas variações que aparecem dentro das espécies domesticadas e que possuem um grau de parentesco entre si, o qual pode ser quantificado em termos de distância genética. Numa coleção de acessos é importante conhecer ou estimar a divergência genética e correlações existentes, pois essas informações são de grande valia para um programa de conservação de germoplasma, bem como para fornecer os materiais para outras pesquisas (Nienhus et al., 1995). A coleção ativa de germoplasma de arroz é mantida no BAG da Embrapa Arroz e feijão, Goiânia, GO, que mantém milhares de acessos com ampla variabilidade genética, visando a preservação a curto e médio prazos, em quantidade suficiente para atender as solicitações.

A variabilidade que as espécies trazem em si é a causa do sucesso da adaptação em ambientes naturais com gradiente para os fatores edafoclimáticos. Essa variabilidade que ostenta os programas de melhoramento, e é proveniente através de coletas, introdução e conservação do germoplasma da espécie, garantindo com isso a base genética para seleção e obter ganhos de produtividade (Liang et al., 1994).

Caracterizar materiais considerados como "recursos fitogenéticos" em um banco ativo de germoplasma (BAG) desempenha papel fundamental, pois é de onde se obtém características qualitativas e quantitativas que permitem eliminar acessos-duplicados e manter aqueles acessos divergentes, que podem ser usados em seleção e fomentar um programa de melhoramento (Autrique et al., 1996). No caso do arroz, a caracterização é fundamental, pois é uma espécie autógama, que pode ter sido cultivado em condições distintas por um tempo considerado; ainda não é de praxe entre os rizicultores comprar e substituir as sementes de ano para ano, além das possíveis trocas de sementes entre produtores nas diferentes regiões, o que pode gerar uma forte seleção pelos anos de cultivo (Heun, Murphy e Phillips, 1994; Skrock, Tyvang e Nienhuis, 1992).

O estudo da divergência genética torna-se importante na caracterização e conservação de acessos, nas hibridações por fornecerem parâmetros para identificar progenitores que quando adotados nas pesquisas possibilitam maior êxito; Kumari e Rangasamy (1997) e Moraes et al. (1992) comentam que, sem dúvida, a técnica de estimação da divergência pode identificar os genes que habilitam para esses fatores.

Assim Nienhus et al. (1995) defendem a organização sistemática de coleções de germoplasma com base nos estudos de parentesco, como uma forma de aumentar a eficiência na utilização dos recursos genéticos disponíveis, pois a caracterização do germoplasma, tanto a nível fenotípico quanto genotípico é de importância para conhecer a estrutura gênica e aquelas características que estão ligadas ao mecanismo da produção.

A caracterização propriamente dita requer que os acessos a serem descritos sejam plantados com cuidados culturais completos, conservando a constituição genética da coleção varietal, sob condições uniformes, diferindo apenas dentro da constituição genética representada pelas verdadeiras características varietais, expressadas sob tais condições em um ambiente. Os relatórios em catálogos de germoplasma devem informar uma pertinente descrição das condições climáticas, tipo de solo e propriedades químicas, produtos culturais, data e local de plantio, para que os interessados possam fazer comparações nesse sentido.

Muitas das complicações da análise fenotípica podem ser mitigadas através da identificação direta de genótipos. Um sistema de diagnóstico baseado em marcadores que segregam conjuntamente com os genes de interesse pode ser utilizado para esse fim. Da mesma forma, a utilização de informação molecular baseada na análise de diversidade genética e de relacionamentos filogenéticos com germoplasma não domesticado, tem o potencial de facilitar o monitoramento e ampliação da base genética de populações em melhoramento.

As características morfológicas são de interpretação às vezes ambígua, tendo o uso limitado também por envolver, em muitos casos o crescimento das plantas, o que é dispendioso, trabalhoso e susceptível de variações ambientais (Hu e Quiros, 1991). Inúmeros são os métodos que avaliam a caracterização do germoplasma de um BAG, mas todos baseiam-se nos caracteres que expressam a constituição genética do indivíduo, esses marcadores genéticos podem ser: morfológicos, bioquímicos e moleculares (Fukuoka et al., 1992).

A comparação de métodos vem sendo adotada por muitos pesquisadores, visando utilizar aqueles mais seguros (componentes principais, variáveis canônicas e distâncias estatísticas). É importante para avaliar com maior segurança, as relações de similaridade ou dissimilaridade existentes entre os acessos dentro de qualquer espécie. Entre as medidas de similaridades, tem-se as distâncias multivariadas, que têm como finalidade agrupar acessos semelhantes (Liang et al., 1994).

O uso de características morfológicas, proteínas e isoenzimas para identificação e quantificação da diversidade genética apresentam limitações, principalmente no que se refere ao número insuficiente de marcadores disponíveis. Contudo, apesar dessas análises determinarem a variabilidade genética com precisão, ainda encontram pouca aplicação prática para o melhoramento, uma vez que o interesse principal está nos caracteres quantitativos poligênicos, que dependem das interações gênicas.

A estimativa da distância genética pode atualmente, ser calculada a partir de marcadores moleculares, com a vantagem de obter um grande número de informações de dados, permitindo com isso maior eficiência no cálculo da variabilidade genética dos indivíduos, principalmente quando a população em estudo já sofreu uma seleção direcionada para certos caracteres do fenótipo, identificando genótipos possuidores de uma frequência elevada de alelos favoráveis para os caracteres de interesse.

O emprego de marcadores moleculares tem também permitido rápido avanço na genética molecular, detectando polimorfismo na seqüência do DNA, proporcionando estudos de herança genética, mapeamento do genoma, identificação de cultivares e auxílio à construção de mapas saturados (Landry et al., 1992; Vasconcelos et al., 1996; Graner et al., 1991), estudos de genes ligados à resistência das plantas; como exemplo: estudo da diversidade genética (Livini et al., 1992; Tingey, Rafalski e Willians, 1991; Fukuoka et al., 1992), identificação varietal (Fukuoka et al., 1992; Hu e Quiros, 1991), mapeamento de características agronômicas (Kumari e Rangasamy, 1997) e auxílio a programas de melhoramento por retrocruzamento (Vasconcelos et al., 1996).

Na prática, nem sempre é fácil avaliar a divergência, pois exige envolvimento de vários caracteres ao mesmo tempo. Para tanto, tem sido adotados vários métodos bioquímicos, como o uso de isoenzimas, que é dificultado pelo seu pequeno número disponível, não permitindo medidas precisas da divergência e os fragmentos de DNA, que é uma técnica simples, aliada à rápida obtenção de resultados, tomou-se a mais promissora, embora seja de maior custo (Laila, 1996).

Em vista do exposto, o objetivo desse trabalho foi obter informações a respeito da divergência genética entre acessos de arroz, no sentido de agrupar os similares. Adotando-se a caracterização através de marcadores morfo-qualitativos e comparando os grupos formados com aqueles grupos oriundos de marcadores quantitativos e moleculares RAPD.

## **4.2 - MATERIAL E MÉTODOS**

### **4.2.1 - Obtenção dos Acessos para Análises**

Na realização desse trabalho foram usados 36 acessos de arroz, coletados entre produtores em diversos Estados do Brasil (MG, SP, RR, AC, MT, GO).