**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Diversidade genética em germoplasma de arroz filipino identificada
por marcadores moleculares e caracteres agromorfológicos**

Thiago Luiz da Mata

Dissertação apresentada para obtenção do título
de Mestre em Ciências. Área de concentração:
Genética e Melhoramento de Plantas

**Piracicaba
2010**

Thiago Luiz da Mata
Biólogo

**Diversidade genética em germoplasma de arroz filipino identificada por
marcadores moleculares e caracteres agromorfológicos**

Orientador:
Prof. Dr. **JOSÉ BALDIN PINHEIRO**

Dissertação apresentada para obtenção do título
de Mestre em Ciências. Área de concentração:
Genética e Melhoramento de Plantas

**Piracicaba
2010**

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP**

Mata, Thiago Luiz da

Diversidade genética em germoplasma de arroz filipino identificada por marcadores moleculares e caracteres agromorfológicos / Thiago Luiz da Mata. - - Piracicaba, 2010. 99 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2010. Bibliografia.

1. Análise multivariada 2. Arroz 3. Diversidade genética 4. Germoplasma vegetal Marcador molecular 6. Melhoramento genético vegetal I. Título

CDD 633.18
M425d

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

*À minha família,
que me deu toda a estrutura para que eu pudesse realizar este trabalho.*

*À Anna Carla,
por estar sempre ao meu lado, me apoiando em todos os momentos.*

*Aos Mestres Superiores,
por me concederem mais uma experiência de real aprendizado.*

AGRADECIMENTOS

À Universidade de São Paulo e ao Departamento de Genética da ESALQ;

À FAPESP pela bolsa de estudos;

Ao Professor José Baldin Pinheiro, por representar mais que o papel de orientador, mas de amigo, sempre disponível a qualquer necessidade;

À Professora Maria Imaculada Zucchi, pelo suporte na área molecular e, mais do que isso, pela tranquilidade que passa;

A todos os colegas do Laboratório de Diversidade Genética e Melhoramento que me ajudaram em todas as etapas, Milene Möller, em primeiro lugar, pela amizade, companheirismo, pela paciência, bom humor e por abrir todas as portas acadêmicas até então, Mariza Monteiro pela paciência e amizade, Miklos pelos risos e pela ajuda fundamental nas análises moleculares, Fátima pela ajuda fundamental no campo, nas análises, pela paciência, Glauber, grande companheiro de laboratório, campo e de perseverança, Aluana pela paciência, Carlos Batista, Marcelo Cavallari, Carlos Aguillar, grande companheiro de campo, juntamente a Camilla Montebelli, Jackeline, Talita Val e, no laboratório, Jéssica, Bruno Mulato entre tantos outros.

Aos funcionários do Departamento de Genética, em especial a Domingos de Sálvio Amaral e Márcio Araújo Silva;

A Eliana, João, Danielle, Carolina, Zilda e Luca;

SUMÁRIO

RESUMO.....	9
ABSTRACT	11
1 INTRODUÇÃO	13
2 DESENVOLVIMENTO.....	15
2.1 Revisão Bibliográfica	15
2.1.1 Origem e classificação botânica	15
2.1.2 Importância econômica	16
2.1.3 Base genética estreita	17
2.1.4 Descritores agromorfológicos	20
2.1.5 Marcadores moleculares	22
2.1.6 Análise estatística.....	25
2.2 Material e Métodos	27
2.2.1 Diversidade genética utilizando marcadores SSRs	27
2.2.1.1 Semeadura e coleta de tecido foliar	27
2.2.1.2 Extração de DNA.....	28
2.2.1.3 Análise com marcadores moleculares	29
2.2.1.3.1 Seleção de marcadores microssatélites	29
2.2.1.3.2 Reações de amplificação	29
2.2.1.3.3 Detecção de polimorfismo alélico	30
2.2.1.3.4 Análises estatístico-genéticas	30
2.2.2 Diversidade genética utilizando caracteres agromorfológicos	32
2.2.2.1 Instalação e condução dos experimentos	32
2.2.2.2 Caracterização e avaliação dos acessos	33
2.2.2.3 Análises estatístico-genéticas	35
2.3 Resultados e discussão.....	38
2.3.1 Análises de variância dos caracteres agromorfológicos	38
2.3.2 Matriz de distância e agrupamento pelo método de otimização de Tocher	40
2.3.3 Estimativas de distância intragrupos, intergrupos e de Mahalanobis	44
2.3.4 Variáveis canônicas.....	45
2.3.5 Análises moleculares.....	46
2.3.5.1 Análise de parâmetros genéticos e dendrograma	46
2.3.5.2 Análise através do programa Structure	58
2.3.6 Agrupamento geral pelo método de otimização de Tocher	62
2.3.6.1 Estimativas de distância intra e intergrupos	74
3 CONSIDERAÇÕES FINAIS	75
4 CONCLUSÕES	81
REFERÊNCIAS.....	83
ANEXOS	93

RESUMO

Diversidade genética em germoplasma de arroz filipino identificada por marcadores moleculares e caracteres agromorfológicos

A utilização intensiva de germoplasma melhorado de arroz reduziu a variabilidade genética necessária no processo de seleção causando a estagnação dos níveis de produtividade. Visando aumentar a base genética das variedades comerciais ocorre uma busca por genes nos BAG's. Para este fim, a avaliação de 144 acessos de arroz filipino provenientes do banco da ESALQ/USP foi realizada através de dezenove caracteres agromorfológicos e 15 marcadores microssatélite. As análises de agrupamento com base nos dados moleculares discriminaram um total de 3 grupos. Todos os 15 locos analisados na genotipagem dos acessos foram polimórficos, sendo encontrados 156 alelos. O número de alelos por loco variou de 3 a 20, com uma média de 10,4. Os acessos analisados possuem uma quantidade bastante significativa de alelos raros (70), sendo os genótipos 84 (Mum 1), 106 (Khao Phe Do), 68 (Khao Khane), 132 (Ku-79-1), 112 (E-Boot), 64 (Bikyat), 146 (BRS-Sertaneja), 121 (Puntas Claras), 85 (Khao Phe Py), 138 (Maraja), 133 (Unnamed), 105 (Os-6), 107 (Ba Ke Gonh) e 47 (Khao Mack Fay Khao) os que apresentaram alelos exclusivos. A heterozigosidade esperada teve seu maior valor no loco RM 257 (0,934) e seu menor valor no loco RM 277 (0,542). Os caracteres agromorfológicos foram submetidos a análises univariadas e multivariadas para a estimação da diversidade genética. Com base no teste F, diferenças significativas a 1% foram observadas em 12 das 14 variáveis quantitativas usadas no estudo. Aquelas que não apresentaram tais diferenças foram: comprimento de folha bandeira (CFB) e produtividade de grãos (PG). As duas primeiras variáveis canônicas explicaram cerca de 99,71% da variação total. Para a primeira variável as características mais importantes e de maior contribuição para a divergência foram: altura de planta na maturidade (APM) e comprimento de colmo (CC); para a segunda foram: comprimento de espiguetas (CE) e número de dias para o florescimento (NDF). A análise de agrupamento reunindo dados moleculares e agromorfológicos discriminou um total de 26 grupos, dos quais os 6 primeiros foram atribuídos como os principais por abranger 69% dos acessos. Desta forma, pode-se observar a presença de maior estruturação de grupos (26), em comparação aos outros agrupamentos envolvendo apenas dados agromorfológicos (10), referentes ao primeiro ano agrícola do estudo, e dados moleculares (3). Tal fato sugere o uso da maior quantidade de dados disponíveis, os quais proporcionarão resultados mais confiáveis e completos, em termos reais.

Palavras-chave: *Oryza sativa*; Germoplasma; Microssatélites; Melhoramento genético vegetal; Análise multivariada

ABSTRACT

Genetic diversity in philippine rice germplasm identified by molecular markers and agromorphological traits

The improved rice germplasm's intense use, with related genitors, reduced genetic variability in breeding programs. This intense use causes stagnation of productivity. Intending to enhance genetic bases in commercial cultivars, a search for genes occurs in germplasm. This study aimed to analyze 144 philippine rice accessions, from ESALQ-USP germplasm, across using nineteen agromorphological traits and fifteen microsatellite markers. The grouping analyses, based on molecular data, found an overall of three groups. All the fifteen markers analyzed in the genotyping of the accessions were polymorphic, and 156 alleles were found. The number of alleles per locus varied from three to 20, with an average of 10,4. The analyzed accessions possess a significant amount of rare alleles (70), being genotypes 84 (Mum 1), 106 (Khao Phe Do), 68 (Khao Khane), 132 (Ku-79-1), 112 (E-Boot), 64 (Bikyat), 146 (BRS-Sertaneja), 121 (Puntas Claras), 85 (Khao Phe Py), 138 (Maraja), 133 (Unnamed), 105 (Os-6), 107 (Ba Ke Gonh) and 47 (Khao Mack Fay Khao), the ones that had presented exclusive alleles. The expected heterozygosity had its highest value in RM 257 (0,934) and its lowest value in RM 277 (0,542). The agromorphological traits were submitted to univariate and multivariate analyses to estimate genetic diversity. Based on F test, 1% significant difference was observed in 12 of 14 quantitative traits used on the study. The traits that didn't present significant statistical differences were: flag leaf length (FLL) and grain productivity (GP). The first and the second canonical variables absorbed together 99,71% of the observed variation. The most important traits for the first canonical variable, which had most contributed for genetic difference, were: plant height in maturity (PHM) and culm length (CL). For the second canonical variable, grain length (GL) and number of days for flourishing (NDF), were the most important. Grouping analyzes, that brought together molecular data and agromorphological traits, found an overall of 26 groups. The main groups were the six first ones, responsible for 69% of the accessions. Using both, molecular data and agromorphological traits, it's noticed more structural patterns grouping (26) than in any other grouping connected with the agromorphological traits from the first harvest (10) and molecular data (3). The results suggest, as much as possible, the use of available data, which offers the most completed and trustable results.

Keywords: *Oryza sativa*; Germplasm; Microsatellite; Plant breeding; Multivariate analysis

1 INTRODUÇÃO

O arroz aparece no cenário mundial como uma das culturas base da alimentação, destacando-se pela área cultivada, cerca de 150 milhões de hectares, e sua produção de 590 milhões de toneladas (IBGE, 2009), principalmente de países em desenvolvimento da Ásia e Oceania. Na América Latina, o arroz também assume grande importância para a economia ocupando postos de destaque em produção e consumo. Neste contexto, o Brasil aparece como o principal produtor fora do continente asiático, estando entre os 10 maiores do mundo. Para a manutenção da grande produtividade mundial e o suprimento da enorme demanda, faz-se necessário o surgimento de variedades cada vez mais produtivas e adaptadas a diversos fatores bióticos e abióticos, cuja necessidade é constantemente suprida por inúmeros programas de melhoramento que desenvolvem novas variedades. Contudo, a utilização intensiva de germoplasma melhorado de arroz, com genitores aparentados em programas de melhoramento, reduziu a variabilidade genética necessária no processo de seleção. A conservação de acessos em bancos de germoplasma assume papel essencial, já que compõem fonte de genes para incremento da produtividade agrícola e adaptação, preservando a diversidade genética em espécies cultivadas, especialmente, a presente em populações remanescentes de ancestrais selvagens. Para que a diversidade genética disponível nos BAG's seja utilizada, é necessário que os acessos sejam caracterizados e documentados de forma que o melhorista identifique aqueles potencialmente úteis para seu programa de melhoramento. Para este fim, duas estratégias são amplamente utilizadas. A primeira é baseada no uso de características agromorfológicas na diferenciação e caracterização de acessos. Recentemente, com o progresso tecnológico das últimas décadas, novas ferramentas denominadas marcadores moleculares passaram a adquirir grande importância neste cenário, aumentando o número de informações disponíveis através da análise de diferenças ao nível do DNA. Desta forma, a caracterização do germoplasma disponível é essencial para a disponibilização da variabilidade existente, ou seja, no uso de acessos mantidos nos bancos de germoplasma e assim permitir o contínuo desenvolvimento do agronegócio nacional.

2 DESENVOLVIMENTO

2.1 Revisão Bibliográfica

2.1.1 Origem e classificação botânica

A espécie *Oryza sativa* é uma monocotiledônea pertencente à família Poaceae, possuindo características tais como caules ocos, flores reduzidas de cor verde e aquênios especializados ou cariopses como frutos (PINHEIRO, 1999). A origem da espécie, segundo Vavilov, é a região situada a sudeste do Himalaia, apesar de as regiões de Madras, na Índia, e Orissa, nas Filipinas, poderem ser também apontadas como centros primários e secundários da espécie (GALLI, 1978). Na Ásia, a domesticação deve ter ocorrido independentemente na Índia, Myanmar, Tailândia, Laos, Vietnã e China (KHUSH, 1997).

Dado ao processo evolutivo e de domesticação, a espécie se adaptou a diversas condições agroecológicas estando, a partir de 1928, subdividida em duas principais subespécies: *Indica* e *Japônica*. Em 1950, foi incluída mais uma nova subespécie *Javânica* as quais incluem as variedades “Bulu” e “Gundil” (LU; CHANG, 1980). O grupo *Javânica*, teria sido produto de seleção efetivada no grupo *Indica*, principalmente na indonésia (PEREIRA, 2002). O grupo *Indica*, segundo Dalrymple (1986), ocorre principalmente em regiões Tropicais, e caracteriza-se por possuir colmos longos, alta capacidade de perfilhamento, folhas longas e decumbentes e ciclo tardio. A este grupo pertencem as variedades brasileiras de arroz irrigado. O grupo *Japônica* ocorre principalmente em regiões temperadas possuindo colmos curtos e rígidos, mediana capacidade de perfilhamento, folhas escuras e eretas e ciclo precoce. Este grupo abrange as variedades de arroz de sequeiro (de terras altas) do Brasil, embora variedades recentes adaptadas a esta condição e de grande qualidade, são híbridos de *Indica* e *Japônica* (PINHEIRO, 1998).

Existem duas teorias principais que explicam a origem destas duas subespécies. A primeira sugere que a subespécie *Indica*, teria se originado primeiro de seus ancestrais não domesticados no Sul de Sudeste da Ásia, sendo o precursor da

subespécie *Japônica* que teria desenvolvido de um tipo adaptado a elevadas altitudes e latitudes. A segunda teoria considera que a diferenciação entre estas duas subespécies ocorreu entre os ancestrais silvestres antes que estes fossem domesticados (SECOND, 1982).

No Brasil a cultura do arroz foi trazida pelos portugueses no século XVI, na costa dos estados da Bahia e Maranhão (PEREIRA, 2002). Existem dois tipos de sistemas de plantio: terras altas e irrigado. O primeiro, não utiliza irrigação, sendo mais utilizado nas regiões do cerrado brasileiro. Já o sistema irrigado, também conhecido por terras baixas, o arroz é plantado em áreas naturalmente inundadas, várzeas com irrigação controlada ou em várzea úmida, e corresponde a 60% da produção nacional (PEREIRA, 2002).

A espécie *Oryza sativa* possui, no total, vinte e quatro cromossomos ($2n=24$) (CHANG, 1976). Trata-se um genoma relativamente pequeno com 466 Mb na subespécie *Indica*, estimando-se a presença de 46.022 a 55.615 genes (YU et al., 2002). Já a subespécie *Japônica*, possui 420 Mb estimando-se a presença de 32.000 a 50.000 genes (GOFF et al., 2002).

2.1.2 Importância econômica

O arroz, para cerca de 2,4 bilhões de pessoas, é alimento básico, principalmente, em países em desenvolvimento na Ásia e Oceania. Segundo estimativas, até 2050, haverá uma demanda pelo dobro desta população. Só no Brasil, o prognóstico para a safra de 2009/2010 é de que serão colhidos 12. 170,029 toneladas em uma área equivalente a 2. 881,754 hectares, um acréscimo de 0,8% em relação a área referente ao ano de 2008 (IBGE, 2009). A produção de arroz duplicou entre o período de 1966 e 1990, devido a utilização de germoplasma altamente produtivo (BRONDANI et al., 2004).

2.1.3 Base genética estreita

Desde o século XIX, os melhoristas têm aperfeiçoado, sistematicamente, populações locais. Deste processo culminaram as variedades avançadas, que culminaram nos cultivares modernos de alto rendimento. À medida que estes cultivares elite se popularizava, substituíam cada vez mais as populações locais que haviam coexistido e cruzado com parentes silvestres. Em relação a estas populações, as cultivares produzidas por programas de melhoramento genético são poucas e uniformes e, embora não possuíssem produtividade semelhante estas últimas, reuniam grande diversidade, como por exemplo, genes relacionados a fatores de estresse ambiental. Como resultado, ao longo do tempo, as culturas foram se tornando cada vez mais homogêneas. Como resultado de um fenômeno denominado erosão genética, a uniformidade genética pode acarretar sérios danos à agricultura mundial, aumentando a vulnerabilidade das culturas a doenças e insetos (HOYT, 1992).

A estagnação dos patamares de produtividade, alcançada pelas cultivares modernas de arroz, tem como principal causa a excessiva redução da base genética das culturas (TANKSLEY; MCCOUCH, 1997). No caso do Brasil, alguns estudos sugerem que apenas sete genitores contribuíram com 70% do conjunto gênico das variedades cultivadas (RANGEL *et al.*, 1996). Dentro desta realidade, a conservação de acessos em bancos de germoplasma assume papel essencial (GILBERT *et al.*, 1999). A conservação dos recursos genéticos das plantas cultivadas, assim como de seus parentes silvestres, ao longo do último século se tornou uma das questões de grande importância entre melhoristas e cientistas de áreas afins. Isto porque a exploração é válida tanto para programas de melhoramento genético, quanto para indústrias como a farmacêutica, de energia, programas de segurança alimentar, entre outros (CALLOW *et al.*, 1997).

Uma alternativa promissora, para ampliação da base genética, vem de programas de pré-melhoramento onde são desenvolvidas linhagens provenientes de cruzamentos entre genitores de base genética mais ampla, as quais, posteriormente, são cruzadas com germoplasma elite, sem romper os blocos gênicos que conferem fenótipos favoráveis (BRONDANI *et al.*, 2003).

A variabilidade genética inter e intra-específica representativa dos recursos genéticos vegetais são, em grande parte, manejada e organizada nos bancos de germoplasma, que se constituem em estrutura física, onde as coleções são conservadas na forma de células, tecidos, sementes ou plantas (NASS, 2001). Os bancos de germoplasma são úteis como fonte de genes para incorporar em materiais domesticados e/ou geneticamente melhorados características de interesse econômico e de importância na alimentação dos povos (DIOLA, 2005).

De maneira geral, o nível de uso dos bancos genéticos é muito baixo, girando em torno de 5% (NASS et al., 1993). Os melhoristas preferem trabalhar com material mais conhecido e com baixa variabilidade, a acessos contidos em bancos de germoplasma os quais podem proporcionar bons ganhos genéticos (NASS, 2001). A utilização destes acessos contidos em bancos só deverá aumentar quando esse material disponibilizar níveis expressivos de informações genéticas e utilitárias, relacionadas a adaptação ambiental, potencial genético, potencial industrial, conhecimento etnobiológico e potencial socioeconômico (VILELA-MORALES, 1999). Além dos bancos de germoplasma, existem os parentes silvestres e raças locais, que constituem as classes menos exploradas até o momento (HAWTIN et al., 1996). Duas alternativas são sugeridas pra o aumento do uso destes materiais tanto em programas de melhoramento quanto para pesquisadores de áreas afins: programas de pré-melhoramento (*pré-breeding*) e coleções nucleares (*core collections*).

Pré-melhoramento é o conjunto de atividades que visam a incorporação de características de materiais não melhorados em materiais elite. Este processo visa a identificação de características e ou genes de interesse em materiais não adaptados ou que não tenham passa do por nenhum processo de melhoramento, e sua posterior inclusão em materiais de elevado potencial produtivo. Portanto, a alternativa mais promissora para conectar os recursos genéticos aos programas de melhoramento é a intensificação das atividades relacionadas aos programas de pré-melhoramento (NASS; PATERNIANI, 2000). As contribuições mais importantes destes programas são: síntese de novas populações base, identificação de genes potencialmente úteis, identificação de novos padrões heteróticos, melhor conhecimento dos acessos *per se* e em

cruzamentos, auxílio no estabelecimento de coleções nucleares (NASS; NISHIGAWA, 2001).

A abordagem tradicional para a utilização de germoplasma exótico é fazer um *screening* dos acessos presentes em um banco de germoplasma à procura da característica desejada. Uma vez definida os genótipos com a característica desejada ele é cruzado com cultivares elites a fim de introduzir os genes do doador exótico nas cultivares. Essa abordagem funciona bem quando a característica de interesse é controlada por um ou poucos genes, como, por exemplo, resistências a doenças e insetos. Apesar de eficaz para certas características, essa estratégia analisa apenas uma pequena parte da variação genética presente no germoplasma exótico. A maioria das características de interesse agrônômico não é controlada por poucos genes, e sim por muitos deles, sendo chamadas de características quantitativas. Uma destas características, por exemplo, é a produtividade de grãos. A produtividade de acessos exóticos é geralmente muito menor do que das cultivares, graças ao grande sucesso no aumento da frequência de alelos favoráveis a essa característica em vários locos. Por isso, programas de melhoramento têm utilizado como estratégia predominante no desenvolvimento de novas cultivares, cruzamentos apenas entre variedades altamente produtivas, porém próximas geneticamente ou aparentadas (MULATO, 2009). Outra forma de abordagem é por meio do uso de técnicas de genética molecular, baseadas em tecnologia de marcadores moleculares, que possibilitam a detecção de diferenças ao nível do DNA, fornecendo medidas mais precisas da variabilidade genética, não somente entre acessos armazenados no banco, mas também dentro deles (BRONDANI; BRONDANI, 2004).

A utilização de bancos de germoplasma teve papel fundamental na expansão da agricultura brasileira ao longo das últimas três décadas. O arroz é citado como bom exemplo deste uso. O que antes era restrito a áreas irrigadas, com o uso de germoplasma adaptado a terras altas (sequeiro), passou a ser produzido no Centro Oeste brasileiro com excelente qualidade para o consumo. Com estas mudanças a área cultivada passou de 344.000 hectares, em 1999/2000, para 843.000 hectares em

2003/2004, um aumento de 132%, que gerou um benefício econômico direto, para a região, neste período, de R\$ 540 milhões (LOPES, 2005)

2.1.4 Descritores agromorfológicos

O arroz possui uma das maiores coleções de germoplasma dentre as espécies de interesse comercial, e estima-se que existam no mundo ao redor de 120.000 variedades diferentes de arroz (KHUSH, 1997). Entre as estratégias utilizadas na caracterização de um banco de germoplasma, destacam-se os descritores morfológicos e moleculares. Os marcadores mais antigos e amplamente difundidos são os que têm como base características morfológicas. Estes ainda continuam sendo aplicados, com eficiência, para certos tipos de germoplasma. Suas principais vantagens residem no fato de serem simples, rápidos e com baixo custo de análise (BRETTEING; WIDRLECHENER, 1995).

Os descritores morfológicos podem ser divididos em dois grupos de caracteres. O primeiro grupo, composto por caracteres ditos morfológicos, são muito úteis na identificação de variedades e tem por características a pouca influência ambiental e alta herdabilidade. São exemplos de caracteres morfológicos: pubescência das folhas, coloração da lígula e da aurícula, ângulo da folha-bandeira e dos perfilhos, cor do internódio e intensidade de antocianina nos nós do colmo. O segundo grupo é composto por caracteres ditos agronômicos, que por sofrer grande influência ambiental e apresentar baixa herdabilidade, sendo exemplos destes: cor da folha, altura da planta, comprimento e espessura do colmo, comprimento, tipo, exserção e degrane da panícula, data da floração e ciclo cultural, massa de 1000 grãos, rendimento de grãos inteiros no beneficiamento, acamamento e presença de arista (FONSECA et al., 2004)., As características fortemente influenciadas pelo ambiente em arroz são a presença de arista, comprimento e espessura do colmo, comprimento da panícula, peso de 1000 grãos, altura da planta, ciclo, tipo e exserção da panícula, degrane, rendimento de grãos inteiros e cor das folhas (FONSECA et al. , 2002; YAN et al., 1999).

Para o arroz, de acordo com o CGIAR (Consultative Groupe on International Agricultural Research) e o International Rice Research Institute (IRRI, 1980), são

citados 50 descritores da cultura. Dentre estes pode-se citar o tipo de endosperma; massa de 100 grãos; número de dias para o florescimento; número de perfilhos, acamamento, comprimento, ângulo e diâmetro do colmo; número de dias para a maturação; comprimento, cor e forma da lígula; comprimento, tipo, ramificação, exserção e degrane da panícula; comprimento, largura, pubescência, cor e ângulo das folhas; cor, pubescência, comprimento e largura da espiguetas; entre outros.

Com o intuito de registrar e proteger novas cultivares, o governo federal sancionou a Lei de Proteção de Cultivares (BRASIL, Lei nº. 9.456, de 25 de abril de 1997). Esta lei normaliza parâmetros botânicos e agrônômicos usados na caracterização e diferenciação dos cultivares. Dentre descritores morfológicos (qualitativos) e agrônômicos (quantitativos), é estabelecido um número mínimo necessário para a descrição de cada cultivar. Foram estabelecidos 27 descritores mínimos para a cultura do arroz. Com inclusões e alterações segundo Jennings et al. (1979), Fonseca e Bedendo (1984) e Freire et al. (1999), são estes: cor da folha, pubescência da folha, coloração da aurícula, coloração da lígula, ângulo da folha bandeira, altura da planta, comprimento do colmo, espessura do colmo, ângulo dos perfilhos, cor do internódio, presença e intensidade de antocianina nos nós do colmo, comprimento da panícula, tipo da panícula, exserção da panícula, degrane da panícula, arista, pubescência das glumelas, coloração do apículo na floração e na maturação, coloração das glumelas (casca), coloração das glumas estéreis, data da floração, ciclo cultural, massa de 1.000 grãos, comprimento do grão sem casca (cariopse), relação comprimento/largura (C/L) do grão sem casca, forma do grão (cariopse), cor do grão sem casca (cariopse), conteúdo de amilose, temperatura de gelatinização e rendimento de grãos inteiros no beneficiamento. Ainda que bastante numerosos e importantes, os descritores morfológicos apresentam limitações, surgindo então a necessidade de outras técnicas complementares.

Características agromorfológicas são fundamentais no melhoramento aplicado, e por isso, são extremamente usadas na caracterização de diversidade genética. Métodos de análise multivariada, como a análise por componentes principais, variáveis canônicas e métodos de agrupamento, podem ser aplicados na predição da divergência

genética usando estes caracteres. A análise por componentes principais permite inferir quais os caracteres responsáveis pela maior parte da divergência encontrada, possibilitando o descarte de caracteres que pouco contribuem na diferenciação dos genótipos avaliados, reduzindo significativamente o tempo e custo de avaliação de características agromorfológicas.

Quando o experimento apresenta resíduo, ou seja, possui repetições, as variáveis canônicas, semelhantes aos componentes principais, são mais robustas. Já os métodos de agrupamento são realizados em duas etapas. A primeira é composta pela estimação de medidas de dissimilaridade entre os genótipos estudados. A segunda etapa é composta pela utilização de um método para formação dos grupos. A distância generalizada de Mahalanobis e a distância Euclidiana são as duas medidas de dissimilaridade mais utilizadas, sendo que a apenas a primeira leva em conta as correlações entre os caracteres. Os métodos de agrupamento podem ser hierárquicos ou de otimização. Os métodos de otimização são melhores quando não se tem conhecimento prévio do material a ser examinado, sendo um dos mais comuns o método de otimização de Tocher (CRUZ, 1990).

A integração entre a análise fenotípica e a alta capacidade de genotipagem, através dos marcadores moleculares, possibilita muitos avanços para o desenvolvimento de cultivares superiores (FERREIRA, 2006).

2.1.5 Marcadores moleculares

Neste contexto surgem, no início da década de 70, os marcadores moleculares, atualmente classificados em dois grupos distintos, baseados em diferentes metodologias. O primeiro grupo é baseado na técnica de hibridização, e tem como representante o RFLP (*Restriction Fragment Length Polymorphism*). Com o desenvolvimento do princípio de reações de polimerização em cadeia (PCR - Polymerase Chain Reaction) (SAIKI et al., 1988), surgem os marcadores baseados na amplificação de DNA compondo o segundo grupo de marcadores, dentre os quais destacam-se: Microssatélites (ou SSR - “Simple Sequence Repeats”), AFLP (Amplified Fragment Length Polymorphism) e RAPD (“Random Amplified Polymorphic DNA”).

Os marcadores RAPD e AFLP possuem como características, serem aleatórios e dominantes. Já os microssatélites e RFLP além de apresentar posições no mapa genético do arroz bem conhecidas, são marcadores codominantes, ou seja, ambos os alelos de um indivíduo heterozigoto podem ser visualizados no gel. Comparando a informação genética oriunda de marcadores RFLP e microssatélites em arroz, Ni et al. (2002) encontraram maior eficiência por parte dos SSR's em detectar polimorfismo que, aliados a um maior número de alelos por loco, indicam maior eficácia na diferenciação de acessos em bancos de germoplasma, particularmente quando são aparentados (XU et al., 2004).

Nos últimos anos, tem sido observado um constante aumento na utilização de marcadores moleculares no mundo inteiro (FRALEIGH, 2006), principalmente, pelo fato de ser um método seguro, uma vez que não é afetado pelo ambiente e permite com maior segurança, descrever as diferenças entre os acessos, reunindo informações úteis ao melhoramento genético, além de servir como complementação à técnicas morfológicas, bioquímicas e citológicas (WEISING et al., 2005).

Informações sobre a diversidade genética do germoplasma podem auxiliar no enriquecimento da base genética, num programa de melhoramento, bem como na avaliação da redundância e deficiências das coleções de germoplasma, gerando dados sobre a eficiência do processo de coleta, manutenção, manejo e ampliação de um banco de germoplasma (PHILLIPS et al., 1993; NEWBURY; FORD-LLOYD, 1993). Assim, os marcadores microssatélites são considerados ideais à caracterização molecular de recursos genéticos, e têm sido utilizados com sucesso na avaliação da diversidade genética em bancos de germoplasma (VARSHNEY et al., 2005).

Os SSR (Simple Sequence Repeats) ou microssatélites são pequenas seqüências, repetidas lado a lado, de um a seis nucleotídeos amplamente distribuídos no genoma da maior parte dos eucariotos (FERREIRA; GRATTAPAGLIA, 1998). Os microssatélites são encontrados em regiões codantes e não codantes do genoma (ZANE et al., 2002). As regiões expressas do genoma apresentam uma alta freqüência de microssatélites, os quais estão predominantemente localizados nas regiões que não codificam proteínas como 3' e 5'UTR (MORGANTE et al., 2002).

Acredita-se que, durante a replicação, o mau-pareamento de motivos microssatélites ou deslizamento (*slippage*), seja o principal mecanismo por trás do surgimento e amplificação destas seqüências nos genomas. Durante a replicação, as fitas de DNA separam-se e ao se associarem novamente, o fazem de forma incorreta. Por meio da inserção ou deleção de uma unidade de repetição, são geradas cópias de trechos de DNA (alelos) com diferentes tamanhos ou números de repetições de um determinado motivo no próximo ciclo de replicação. (SCHLOTTERER; TAUTZ, 1992).

Ao contrário do que se pensava inicialmente a respeito das funções dos microssatélites (neutros e sem função no genoma), muitos estudos demonstram que os microssatélites estão envolvidos em diversos processos como na organização do cromossomo, na regulação de processos metabólicos e na regulação da atividade gênica (LI et al., 2002; LI et al., 2004).

Os marcadores SSR têm sido muito utilizados em estudos de diversidade genética em bancos de germoplasma, devido à robustez, confiabilidade, praticidade operacional e por serem marcadores mais informativos geneticamente. A utilização destes marcadores baseia-se na amplificação destas regiões contendo seqüências repetidas simples utilizando-se um par de *primers* específicos complementares a seqüências únicas que flanqueiam o microssatélite. O polimorfismo entre as bandas será decorrente dos números diferentes de elementos simples repetidos. Cada segmento diferente de DNA representa um alelo daquele loco específico (ALVES, 2002).

Os marcadores microssatélites apresentam diversas características desejáveis, pois, além de serem marcadores codominantes e possuírem ampla e uniforme distribuição no genoma, são multialélicos, amplificados via PCR e possuem *primers* totalmente compartilháveis que, se marcados com fluorescência, apresentam a vantagem de possibilitarem sistema multiplex, o que permite avaliar rapidamente um grande número de indivíduos para um grande número de locos em pouco tempo (SUGANUMA; CIAMPI, 2001).

O grande volume de trabalho necessário para o desenvolvimento de *primers* específicos para cada espécie representa o maior problema para a utilização deste marcador. Contudo, mais de 2.000 marcadores microssatélites já foram desenvolvidos

para a cultura do arroz (McCOUCH et al., 2002), como os publicados por Ni et al. (2002), Yu et al. (2003), Xu et al. (2004), Brondani et al. (2006a e 2006b), entre outros. Publicações recentes, neste sentido, podem ser verificadas em trabalhos como o apresentado por Lapitan et al. (2007), avaliando, com sucesso, 24 acessos de arroz filipino através de 151 locos SSR polimórficos, gerando alto número de dados de alta qualidade para avaliação de diversidade genética.

Ainda que quantidades expressivas de *primers* se encontrem disponíveis para a cultura do arroz, não muitos destes são necessários, segundo Ni et al. (2002), para a identificação de cultivares e na avaliação segura e eficiente da diversidade genética desta cultura. Estudando 38 variedades de arroz de particular interesse em programas de melhoramento e dois acessos de espécies selvagens, por meio da análise de 111 marcadores microssatélites, estes autores sugeriram que número bastante inferior, ou seja, 30 marcadores microssatélites são suficientes para uma análise segura, produzindo informações equivalentes.

2.1.6 Análise estatística

Existem duas maneiras básicas de avaliar a divergência genética: de natureza quantitativa ou preditiva (MIRANDA et al., 1988). Quando de natureza quantitativa, podem ser citadas as análises dialélicas, sendo necessária a avaliação dos genitores e de todas as combinações híbridas. Em culturas autógamas, como o arroz, o problema ainda é maior devido a dificuldades enfrentadas no processo de polinização manual gerando pouca probabilidade de êxito da semente híbrida (CRUZ, 1990). Considerando, basicamente, diferenças morfológicas apresentadas pelos genitores, os métodos preditivos têm recebido maior atenção, pois dispensam a obtenção prévia das combinações híbridas (CRUZ; REGAZZI, 1994). Vários métodos multivariados podem ser aplicados na predição da divergência genética, dentre os quais tem se destacado os métodos de agrupamento e as análises por componentes principais e variáveis canônicas. Esses métodos de análise têm sido utilizados por vários autores em arroz (SILVA et al., 1999, RANGEL et al., 1991; SINGH et al., 1996; MAURYA; SINGH, 1977).

As distâncias euclidiana média padronizada e generalizada de Mahalanobis entre os pares de genótipos, são amplamente utilizadas como medida de dissimilaridade nos métodos de agrupamento, sendo que a segunda oferece a vantagem de levar em consideração a existência de correlações entre os caracteres analisados por meio da matriz de variâncias e covariâncias residuais, porém, necessita de experimentos com repetições (CRUZ; CARNEIRO, 2003).

A distância generalizada de Mahalanobis (D_{ii}^2) é obtida pela transformação das variáveis originais, via condensação pivotal. Este processo 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 à direita, após cada condensação, corresponde aos coeficientes da transformação linear das variáveis não correlacionadas, e o elemento da diagonal que foi transformado na unidade correspondente à variância daquela variável não correlacionada. Posteriormente, as variáveis são padronizadas, por meio da razão entre os valores de cada característica avaliada e a raiz quadrada da respectiva variância da característica em questão (MORAIS, 1992; CRUZ; REGAZZI, 1994).

No método dos componentes principais, e também no de variáveis canônicas, o objetivo é avaliar a similaridade dos genitores por intermédio de uma dispersão gráfica, em que se consideram, em geral, dois eixos cartesianos, enquanto que a análise de agrupamento tem por finalidade reunir, por algum critério de classificação, os genótipos em vários grupos, de tal forma que exista homogeneidade dentro do grupo e heterogeneidade entre grupos (CRUZ; REGAZZI, 1994). A análise por variáveis canônicas mostra-se mais eficiente quando há resíduo no experimento, ou seja, quando há repetições. O delineamento em blocos aumentados permite esta abordagem. A importância relativa dos caracteres avaliados no material genotípico observado pode ser calculada por meio de coeficientes de ponderação (autovetores) estimados pela técnica de variáveis canônicas (MARDIA et al., 1979). O descarte de caracteres mediante essa técnica multivariada pode ser feito por meio dos maiores coeficientes a eles associados, partindo-se das últimas variáveis canônicas. Após a eliminação de

caracteres, a primeira variável canônica envolverá quase toda a variância estimada. Assim, os autovetores da primeira variável canônica associada aos caracteres não eliminados, ou seja, os maiores coeficientes relativos às primeiras variáveis canônicas podem indicar as variáveis de maior importância na avaliação dos genótipos.

Os estudos de divergência genética, com a utilização de análise multivariada, têm sido muito proveitosos para a constituição de grupos com alta similaridade, para a avaliação do nível evolutivo de espécies do gênero *Oryza* e para escolha de genitores divergentes nos programas de melhoramento (RANGEL et al., 1991).

2.2 Material e Métodos

O trabalho foi dividido em duas etapas principais, sendo a primeira definida pela identificação da diversidade genética em germoplasma de arroz filipino por meio marcadores microssatélites, e a segunda a partir de caracteres agromorfológicos utilizados com o mesmo objetivo. A segunda etapa pode ser subdividida em duas sub-etapas, pois a avaliação dos caracteres agromorfológicos foi realizada durante dois anos agrícolas consecutivos, 2007/2008 e 2008/2009.

2.2.1 Diversidade genética utilizando marcadores SSRs

2.2.1.1 Semeadura e coleta de tecido foliar

Foram analisados 144 acessos filipinos do banco de germoplasma de arroz do Departamento de Genética da ESALQ/USP e 6 genótipos testemunha (cultivares comerciais). Vinte dias após a germinação, amostras do tecido foliar foram coletadas para a extração de DNA genômico.

2.2.1.2 Extração de DNA

As extrações de DNA foram realizadas segundo protocolo descrito por (DOYLE; DOYLE, 1987), utilizando o método CTAB. A concentração de DNA das amostras foi estimada por eletroforese em gel de agarose 1%, corado com a tecnologia Syber Safe, por comparação visual com o DNA - padrão do fago lambda, de peso molecular conhecido. Logo após a quantificação, para cada amostra de DNA, foram adicionados 1µl de RNase. O DNA genômico foi então diluído a concentração de 3 ng/µl de acordo com os resultados apresentados por Borba (2005). Géis de agarose podem ser vistos nas figuras 1 e 2, nos quais DNAs-padrões estão presentes nas quatro primeiras e últimas canaletas, nas concentrações de 50, 100, 150 e 200ng/µl (da esquerda para a direita), permitindo a correta quantificação por comparação visual.

Figura 1 - Gel ilustrando a quantificação do DNA genômico de 130 acessos

Figura 2 - Gel ilustrando a quantificação do DNA genômico dos 20 acessos restantes

2.2.1.3 Análise com marcadores moleculares

2.2.1.3.1 Seleção dos marcadores microssatélites

As reações de amplificação foram realizadas para cada um dos 150 genótipos, utilizando *primers* específicos. Quinze pares de *primers* foram selecionados para uso neste trabalho, sendo 13 pares de SSRs genômicos (OG61, RM38, RM222, RM229, RM223, RM335, RM257, RM1, RM234, RM207, RM277, RM231 e RM252) e 2 pares de microssatélites de seqüências expressas (EST-SSRs), sendo estes RM190 e RM149. Com relação aos locos EST's, o primeiro (RM 190) é relacionado a síntese de amido, e o segundo (RM 149) é relacionado ao stress. A escolha individual dos 13 *primers* genômicos utilizados foi feita com base nos resultados apresentados por Borba (2005) a qual avaliou, com sucesso, 242 acessos da Coleção Nuclear Brasileira de Arroz. Desta maneira, foram selecionados *primers* que apresentaram melhor desempenho de informação, altos valores de diversidade genética de Nei (H_E), e ampla cobertura do genoma do arroz, contemplando todos os doze cromossomos da espécie.

2.2.1.3.2 Reações de amplificação

As reações de amplificação foram conduzidas para 14 dos 15 *primers* do estudo, segundo os melhores resultados obtidos por Borba (2005), com algumas modificações, em um volume final de 15 μ l, contendo 1,5 μ L de Tampão de reação (10x, acrescido de

1,5 mM de MgCl_2), 1,3 μL de dNTP (2,5 mM de cada), 0,2 μL de Taq Polimerase (5 unidades/ μL), 0,75 μL de *primer* SSR contendo as sequências *forward* e *reverse* nas concentrações de 0,5 μM e 10 ng de DNA. A avaliação do loco RM 1, o qual se deu por meio de seqüenciador automático, ocorreu em volume final de 15 μL , contendo 1,5 μL de Tampão de reação (10x, acrescido de 1,5 mM de MgCl_2), 1,3 μL de dNTP (2,5 mM de cada), 0,2 μL de Taq Polimerase (5 unidades/ μL), 0,3 μL de *primer* SSR *forward* e 0,6 μL de *primer* SSR *reverse* nas concentrações de 0,1 μM e 0,4 μM , respectivamente, 0,4 μM de FAM e 4,5 ng de DNA.

As reações foram conduzidas em termociclador com a seguinte programação: um pré-ciclo de 94°C por 5 minutos, em seguida 30 ciclos de 94°C por 1 minuto, 56°C por 1 minuto, 72°C por 1 minuto, e um passo final de 72°C por 7 minutos. A checagem da amplificação foi feita por eletroforese horizontal em gel de agarose 2% contendo tampão TBE 1X e os géis visualizados após coloração com Syber Safe sobre luz UV.

2.2.1.3.3 Detecção de polimorfismo alélico

Os produtos de amplificação foram separados sob condições desnaturantes em gel de poliacrilamida a 7%, uréia 7M, TBE1X, por 3-5 horas a uma wattagem constante de 70W. Os fragmentos foram detectados por coloração em nitrato de prata, segundo protocolo de Creste et al. (2001), e seus tamanhos estimados por meio de comparação com marcadores de pares de base. Fragmentos de diferentes tamanhos foram considerados como diferentes alelos. Apenas no caso do loco RM 1, este foi submetido a genotipagem utilizando seqüenciador automático ABI 3730 (Applied Biosystems) e o polimorfismo alélico detectado pelo software Peak Scanner software 1.0 (Applied Biosystems).

2.2.1.3.4 Análises estatístico-genéticas

A partir da genotipagem dos géis as freqüências alélicas, o número de alelos por loco (A), a heterozigotidade observada (H_o) e esperada (H_e) e a riqueza alélica, para

estimar diversidade genética, foram obtidas utilizando-se o programa FSTAT 2.9.3 (GOUDET, 1995).

A estruturação da variabilidade foi visualizada utilizando-se o programa NTSYS (ROHLF, 1989), por meio de dendrograma construído pela matriz de distâncias genéticas de Rogers-W (WRIGHT, 1978) e pelo critério de agrupamento UPGMA, (método de média aritmética não ponderada), utilizando, para isto, o programa TFGA (MILLER et al., 1997). A estabilidade dos agrupamentos foi testada através de 10.000 reamostragens *bootstrap*. O valor cofenético foi calculado através de um teste de Mantel com 1000 permutações no NTSYS com o objetivo de estimar quanto da matriz original foi explicada pelo dendrograma obtido.

O conteúdo de informação polimórfica (PIC – polymorphism information content), proposto por Anderson et al. (1993), foi calculado em função do número de alelos detectados, da sua distribuição e frequência na população estudada. Os valores de PIC por loco são determinados pela expressão:

$$PIC = 1 - \sum p_i^2$$

Nesta expressão, p_i é a frequência do alelo i na população. O cálculo é baseado no número de alelos detectados por determinado loco e a frequência relativa de cada alelo no conjunto total de acessos. Desta maneira o PIC está relacionado com o número de alelos, que, por sua vez, está diretamente associado à divergência genética e ao número de genótipos em estudo.

Foi também utilizado o programa Structure (PRITCHARD et al., 2000) para inferir o número de grupos (K) de indivíduos geneticamente relacionados através de métodos bayesianos, não necessitando de informações hierárquicas *a priori*, com o objetivo de visualizar a estrutura genética do banco germoplasma de arroz filipino. Devido à ausência de heterozigotos entre os genótipos analisados no estudo, a análise, através do programa Structure, foi conduzida segundo modelo haplóide. Os outros parâmetros utilizados neste estudo foram: 100.000 *burn-in*, 200.000 MCMC, com 10 arquivos de parâmetros contendo valores de $k=1$ à $k=10$ e cinco corridas para cada valor de K. O modelo utilizado foi o sem mistura (*no admixture*).

A determinação do número K mais provável em relação aos propostos foi realizada utilizando valores de ΔK , (EVANNO et al. 2005).

2.2.2. Diversidade genética utilizando caracteres agromorfológicos

2.2.2.1. Instalação e condução dos experimentos

Os experimentos para caracterização e avaliação dos acessos filipinos de arroz do banco de germoplasma foram conduzidos no campo experimental do Departamento de Genética da Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ/USP) localizado no município de Piracicaba, SP (Figura 3), sendo as semeaduras realizadas em 29 de outubro de 2007 e 2008, respectivamente. Os experimentos de campo foram avaliados durante dois anos agrícolas consecutivos.

Figura 3 - Experimento em campo referente ao ano agrícola 2007/2008

A densidade de semeadura utilizada foi de 60 sementes por metro linear, com espaçamento entre linhas de 40 cm. O delineamento experimental foi o de blocos aumentados com seis tratamentos comuns (BRS sertaneja, Curinga, IAC 202, Caiapó, IAC 201e IAC 1246), sendo estas cultivares comerciais. A parcela experimental foi composta por quatro fileiras de 2,0 m de comprimento, sendo a área útil da parcela composta pelas duas linhas centrais. Foram realizados os tratos culturais e fitossanitários recomendados para a cultura.

2.2.2.2. Caracterização e avaliação dos acessos

A caracterização e avaliação dos acessos foi feita de acordo com alguns descritores indicados pelo IBPGR-IRRI Rice Advisory Committee (1980), alterados por Fonseca et al. (1981) e por Fonseca e Bedendo (1984), com algumas inclusões e alterações julgadas oportunas.

Abaixo segue uma melhor descrição dos caracteres avaliados, de acordo com critérios quantitativos e qualitativos, respectivamente:

A) QUANTITATIVOS

- Número de dias para a emergência (NDE) – número de dias, desde a semeadura, até a emergência no solo.
- Número de dias para o florescimento (NDF) - número de dias da semeadura até o florescimento de 50% das panículas na parcela.
- Ciclo total da planta (CTP) – número de dias entre a semeadura e a maturação completa, ou seja, quando 85-90% das espiguetas estão maduras. Com esse dado classifica-se o genótipo em um dos grupos: precoce (até 105 dias), semi-precoce (de 106 a 120 dias), médio (de 121 a 135 dias), semi-tardio (de 136 a 150 dias) e tardio (acima de 150 dias).
- Massa de cem sementes (MCS) - determinada em balança de precisão. É a média de cinco amostras de 100 sementes com umidade de 13%.
- Produtividade de grãos (PG) - determinada pela pesagem dos grãos colhidos na parcela útil, após limpeza e secagem uniforme, com umidade corrigida para 13%.

- Comprimento do colmo (CC) - medida em cm do colmo principal do nível do solo ao nó ciliar da panícula, a partir do enchimento de grãos.
- Número de perfilhos (NP) - obtido em contagem realizada após a floração.
- Comprimento da folha bandeira (CFB) - medida do limbo, da lígula ao ápice da folha bandeira, à época da colheita, em centímetros.
- Largura da folha bandeira (LFB) - refere-se à medida, em centímetros, na região de maior largura do limbo da folha bandeira, à época da colheita.
- Altura da planta na maturidade (APM) - distância média, em centímetros, medida da superfície do solo até a extremidade da panícula do colmo mais alto, expressa em centímetros, a partir do enchimento dos grãos.
- Comprimento da panícula (CP) - distância, em centímetros, da base da panícula à ponta da última espiguetta, determinada na época da colheita.
- Comprimento da espiguetta (CE) - medido com paquímetro, dado em milímetros.
- Largura da espiguetta (LE) - medida com paquímetro, dada em milímetros.
- Número de ramificações por panícula (NRP) - obtido pela contagem das ramificações a partir do eixo da panícula.

B) QUALITATIVO

- Pubescência do limbo (PL) - observada entre o emborrachamento e a emissão da panícula, e classificada em: (1) Ausente, (2) Escassa, (3) Média, (4) Forte.
- Cor da aurícula (CA) - observada na penúltima folha, entre o emborrachamento e a antese, e classificada em: (1) Verde-claro, (2) Púrpura.
- Ângulo de inserção da folha bandeira (IFB) - refere-se ao ângulo formado pela folha bandeira e o colmo. É avaliado na época da floração, empregando-se a seguinte escala: (1) Ereto - menor que 30°, (2) Intermediário - entre 31 e 60°, (3) Horizontal - entre 61 e 90°, (4) Descendente - maior que 90°.
- Tipo de panícula (TP) - determinada na maturação e classificada em: (1) Compacta, (2) Intermediária, (3) Aberta.
- Comprimento da arista (CA) - esta determinação será feita mediante a escala: (1) Ausente, (2) Micro aristada, (3) Aristada.

A caracterização dos acessos no primeiro ano agrícola (2007 / 2008) foi realizada utilizando-se todos os 19 caracteres apresentados acima. No entanto, para o segundo ano agrícola (2008 /2009), foram utilizados, apenas, 8 caracteres quantitativos (NDF, CC, APM, CFB, LFB, CP, CTP e PG). O critério de escolha destes caracteres foi baseado em resultados obtidos nas análises do primeiro experimento. A partir da análise de variáveis canônicas estipularam-se quais foram as características que mais contribuíram na avaliação da diversidade genética, além da importância agronômica de alguns caracteres como a produtividade de grãos (PG), por exemplo.

2.2.2.3 Análises estatístico-genéticas

Com relação aos dados quantitativos e qualitativos, a divergência genética entre os acessos foi quantificada através de análises estatísticas univariadas (análise de variância) e multivariadas, utilizando-se a distância generalizada de Mahalanobis, para as variáveis quantitativas, complemento aritmético do coeficiente de Jaccard para as variáveis qualitativas, agrupamento dos acessos pelo método de otimização de Tocher e análise de variáveis canônicas.

O complemento aritmético do coeficiente de Jaccard é calculado de acordo com a seguinte fórmula: $1 - [(a / a + b + c)]$, onde “a” representa o número de concordâncias positivas do tipo 1-1, “b” o número de discordâncias do tipo 1-0 e “c” o número de discordâncias do tipo 0-1.

As estimativas das distâncias de Mahalanobis a partir de variáveis transformadas por condensação pivotal são obtidas por meio da seguinte expressão:

$$D_{ii'}^2 = \sum (z_{ij} - z_{i'j})^2$$

em que:

$D_{ii'}^2$: distância generalizada de Mahalanobis entre os genótipos i e i';

$z_{ij} - z_{i'j}$: diferença entre os progenitores i e i' em relação à j-ésima variável;

z_{ij} : média do i-ésimo genótipo em relação à j-ésima variável com variância igual a

1.

A transformação das variáveis originais em variáveis padronizadas é assim realizada:

$$Z = XV'$$

em que:

X: variável original;

Z: variável padronizada;

V': matriz de transformações.

A padronização das variáveis dá-se pela expressão:

$$z_j = Z_j / V(Z_j)^{1/2}$$

em que:

z_j : variável transformada padronizada;

Z_j : variável transformada;

$V(Z_j)$: variância da variável transformada (Z_j).

O procedimento fornece a matriz V, as variâncias das variáveis não correlacionadas (z_j) obtidas pela condensação pivotal e as estimativas das distâncias generalizadas.

A partir das distâncias generalizadas entre os pares de genótipos foi construída a matriz de dissimilaridade, para a definição dos agrupamentos conforme o algoritmo de otimização de Tocher. Sobre a matriz de dissimilaridade é identificado o par de genótipos mais similares, que formarão o grupo inicial. A partir daí é avaliada a possibilidade de inclusão de novos genótipos adotando-se o critério de que a média das medidas de dissimilaridade dentro de cada grupo deve ser menor que as distâncias médias entre quaisquer grupos. A entrada de um genótipo em um grupo sempre aumenta o valor médio da distância dentro do grupo. Assim, toma-se a decisão de incluir o genótipo em um grupo por meio da comparação entre o acréscimo no valor médio da distância dentro do grupo e um nível máximo permitido, que será o valor máximo da medida de dissimilaridade (α), encontrado no conjunto das menores distâncias envolvendo cada genótipo.

Uma vez formado o primeiro grupo, calculam-se as medidas de dissimilaridade entre esse grupo e os demais genótipos através da expressão:

$$d_{(ij)k} = d_{ik} + d_{jk}$$

em que:

$d_{(ij)k}$: medida de dissimilaridade entre o grupo ij e o genótipo k;

d_{ik} : medida de dissimilaridade entre os genótipos i e k;

d_{jk} : medida de dissimilaridade entre os genótipos j e k.

A inclusão do genótipo no grupo ij será feita verificando se a distância deste genótipo (k) em relação ao grupo, dividida pelo número de genótipos que já o constitui (n), é inferior ao máximo permitido (α), dessa forma:

- se $\alpha \geq d_{(ij)k} / n$, inclui-se o genótipo k no grupo ij.

- se $\alpha \leq d_{(ij)k} / n$, o genótipo k não é incluído no grupo ij.

A inclusão de novos genótipos no grupo, ou a formação de novo grupo é feita do mesmo modo.

A análise multivariada utilizando variáveis canônicas é semelhante, quanto ao aspecto funcional, à análise por componentes principais, e objetiva a simplificação dos dados amostrais, permitindo a identificação de grupos similares em estudos de divergência genética (CRUZ, 1990). A análise por variáveis canônicas é, conseqüentemente, um método de ordenação, possibilitando avaliar o grau de similaridade entre genótipos através de matrizes de variâncias e covariâncias residuais, bem como a partir de matrizes de variâncias e covariâncias entre médias fenotípicas dos caracteres avaliados de dados provenientes de experimentos com repetições. Comumente, as variáveis originais são transformadas em variáveis padronizadas e não-correlacionadas igualando-se a matriz de dispersão residual à matriz identidade. A partir de então, o procedimento para a obtenção das variáveis canônicas equivale ao utilizado para a obtenção dos componentes principais (CRUZ, 1990).

A análise por componentes principais é indicada a estudos com médias amostrais ou com dados de experimentos sem repetições. Já a análise por variáveis canônicas mostra-se mais eficiente quando há resíduo no experimento, ou seja, quando há repetições. A contribuição relativa dos caracteres para a divergência foi calculada utilizando o critério proposto por SINGH (1981). Desta forma, a importância relativa dos caracteres é estimada por meio da participação dos componentes de D^2 (distância generalizada de Mahalanobis), relativos a cada característica no total da dissimilaridade observada.

Inicialmente, foi realizada uma análise referente aos dados obtidos para o primeiro ano agrícola (2007 / 2008), resultando em um primeiro agrupamento dos acessos.

Com relação aos dados moleculares, os dados obtidos pela leitura dos quinze *primers* do estudo foram organizados e, a partir destes, foi originada uma matriz de distância adaptada ao uso de bancos de germoplasma, denominada matriz de Rogers-W.

Com o objetivo de se obter resultados mais robustos e mais próximos da realidade, ao final do processo, obtidas as matrizes independentes referentes aos dados quantitativos e qualitativos do primeiro ano juntamente as matrizes referentes aos dados moleculares (matriz de Rogers-W) e quantitativos do segundo ano (2008 / 2009), foi realizada a soma destas quatro matrizes com posterior agrupamento dos acessos pelo método de otimização de Tocher. Esta análise deu origem a um agrupamento envolvendo todos os dados obtidos pelo estudo, chamado de agrupamento total.

As análises foram realizadas no programa GENES (CRUZ, 2001).

2.3 Resultados e Discussão

2.3.1 Análises de variância dos caracteres agromorfológicos

Foi realizada uma análise de variância para cada uma das 14 variáveis quantitativas com o objetivo de verificar a existência de diferenças estatísticas significativas entre as médias dos genótipos. O delineamento utilizado foi o de blocos aumentados. Com base no teste F, diferenças significativas a 1% foram observadas em 12 das 14 variáveis quantitativas usadas no estudo. As variáveis que não apresentaram diferenças estatisticamente significativas foram: comprimento de folha bandeira (CFB) e produtividade de grãos (PG). Outros parâmetros genéticos foram estimados como o coeficiente de variação experimental (CV_e) e o coeficiente de variação genética (CV_g). O primeiro parâmetro (CV_e) mediu a magnitude da precisão experimental. Embora tenha apresentado valores bastante variáveis (Tabela 1), considera-se os valores

obtidos sejam satisfatórios uma vez que os genótipos avaliados são exóticos para as condições brasileiras. O segundo parâmetro, o coeficiente de variação genética ($CV_g\%$), razão entre o desvio padrão genético e a média da população, expressa em percentagem, constitui um indicador valioso da grandeza relativa das mudanças possíveis que podem ser conseguidas em cada característica, por meio da seleção (MORAIS, 1992), além de ser um indicativo da variabilidade genética encontrada entre os acessos para cada característica mensurada. Este parâmetro, assim como o primeiro, se mostrou bastante variável (Tabela 1) com destaque para o caractere CC (80,38), denotando alta diversidade genética.

Tabela 1 - Médias das testemunhas, dos genótipos filipinos, coeficientes de variação ambiental, genético, herdabilidade e Teste F dos caracteres avaliados

	NDE (dias)	NDF (dias)	NP (n°)	CC (cm)	APM (cm)	CFB (cm)	LFB (cm)	CP (cm)	NRP (n°)	CE (mm)	LE (mm)	CTP (dias)	MCS (g)	PG (kg/parcela)
M_t	7,93	105,33	5,46	66,43	87,28	26,22	1,58	20,91	2,2	9,2	2,69	111,53	2,51	0,63
M_g	7,76	120,06	4,3	92	105,9	24,21	1,51	20,07	1,82	8,68	3,38	134,35	2,83	0,41
CV_e	2,97	6,67	30,41	6,22	7,03	20,47	7,58	12,56	24,07	3,08	3,15	2,62	4,54	46,65
CV_g	7,48	8,58	32,79	80,38	9,6	5,33	14,53	5,89	35,56	10,25	20,11	12,34	19,03	5,63
h²	86,36	62,34	53,76	99,4	65,06	6,35	78,6	18,03	68,58	91,72	97,6	95,66	94,59	1,43
F	6,03	3,11	2,06	160,89	3,89	0,85*	4,08	1,20	2,96	12,72	46,84	28,56	20,24	0,85*

* Valores não significativos a 1%

Comparando as médias obtidas pelos genótipos comerciais (M_t) com as obtidas pelos genótipos filipinos (M_g), podemos observar que os filipinos apresentam número de dias para emergência semelhantes aos comerciais, florescimento mais tardio, menor número de perfilhos, colmos e plantas maiores, folhas bandeira mais delgadas, panículas menores e menos ramificadas, grãos mais curtos e largos, ciclos maiores, maior massa de cem sementes.

É certo que resultados como estes já eram, de certa forma, esperados por se tratar de uma comparação de médias entre genótipos não comerciais (filipinos) com genótipos de valor comercial que já passaram por intenso processo de seleção em programas de melhoramento. Embora com características menos desejáveis, é

importante notar que esses resultados indicam, de certa forma, a presença de variabilidade genética, o que possibilita a obtenção de ganhos por seleção, mesmo em características complexas como a produção de grãos.

2.3.2 Matriz de distância e agrupamento pelo método de otimização de Tocher

Com base em um conjunto de dados obtidos a partir de caracteres quantitativos e qualitativos referentes ao primeiro ano agrícola, realizou-se o agrupamento dos acessos pelo método de otimização de Tocher. Este método discriminou um total de 10 grupos (Tabela 2). Os 2 primeiros grupos foram atribuídos como os principais por abranger 82% dos acessos, sendo que somente o grupo 1 representou 60,6%; grupo 2, (21,3%); grupo 3, (8%); grupo 4, (3,3%); grupo 5, (2%); grupos 6 e 7, (1,3%) e os grupos 8, 9 e 10, (0,7%) com apenas um representante.

Tabela 2 - Agrupamento dos acessos pelo método de otimização de Tocher

Grupos	Acessos																												
1	107	141	7	57	102	78	85	12	77	113	79	47	35	144	21	123	124	133	81										
	101	126	24	67	142	40	135	138	38	3	32	105	84	68	93	6	4	100	33										
	118	59	18	66	41	63	48	56	36	112	116	9	76	150	119	16	61	25	148										
	60	31	8	53	127	37	23	134	50	74	71	110	62	73	88	29	108	114	146										
	92	15	75	22	20	19	104	14	117	91	128	106	80	30	44														
2	115	131	46	149	147	43	90	122	103	11	26	28	87	72	136	55	65	34	39										
	98	95	125	45	99	49	111	139	69	86	97	51	137																
3	5	129	58	52	83	70	140	64	82	143	94	109																	
4	2	120	54	132	130																								
5	17	27	1																										
6	42	89																											
7	13	121																											
8	145																												
9	10																												
10	96																												

Analisando os valores médios intragrupos (Tabela 3) obtidos para cada agrupamento, nota-se que no grupo 1 ocorre a predominância de acessos com ausência de pubescência no limbo (PL), ângulo de inserção de folha bandeira (IFB) compreendidos entre 31 e 60 graus, aristas ausentes e com panículas do tipo compacta. No grupo 1, assim como em todos os grupos, a cor da aurícula é verde, caracterizando o descritor cor de aurícula (CAU) como pouco discriminante dentro desta população e, portanto, merece menor atenção em estudos posteriores. Com relação às características qualitativas, o grupo 1, é o que mais se assemelha aos acessos comerciais.

Tabela 3 - Caracteres qualitativos referentes a cada agrupamento

Grupos	PL	CAU	IFB (graus)	TP	CA
1	Ausente	Verde	31-60	Compacta	Ausente
2	Escassa	Verde	31-60	Intermediária	Microaristada
3	Escassa	Verde	Menor 30	Aberta	Aristada
4	Escassa	Verde	31-60	Intermediária	Aristada
5	Média	Verde	Menor 30	Compacta	Microaristada
6	Escassa	Verde	Menor 30	Compacta	Microaristada
7	Ausente	Verde	61-90	Intermediária	Microaristada
8	Ausente	Verde	Menor 30	Aberta	Ausente
9	Escassa	Verde	61-90	Intermediária	Microaristada
10	Média	Verde	61-90	Aberta	Aristada

Tabela 4 - Caracteres quantitativos referentes a cada agrupamento

Grupos	NDE (dias)	NDF (dias)	NP	CC (cm)	APM (cm)	CFB (cm)	LFB (cm)	CP (cm)	NRP	CE (mm)	LE (mm)	CTP (dias)	MCS (g)	PG (Kg/ parcela)
1	8	119	4	83,88	103,78	24,32	1,49	20,12	2	8,77	3,31	132	2,83	0,43
2	8	119	4	86,06	106	23,3	1,52	20	2	8,88	3,30	131	2,87	0,42
3	8	118	4	89,32	110,30	24,56	1,64	21,22	2	8,94	3,16	133	2,85	0,48
4	8	114	5	73,69	93,82	25,15	1,766	19,74	2	7,54	3,28	140	2,39	0,34
5	7	120	6	81	101,1	24,75	1,45	20,1	2	8,57	2,95	127	2,36	0,47
6	9	132	3	89,49	105,08	20,96	1,59	19,30	2	8,05	3,59	151	2,88	0,47
7	8	115	7	83,5	102,075	31,035	1,68	18,575	3	4,68	6,41	126,5	2,9	0,42
8	8	109	5	116,10	141,3	28,62	1,59	25,20	2	9,77	3,53	107	3,45	0,48
9	7	109	5	116,1	141,3	28,62	1,59	25,2	2	9,77	3,53	107	3,45	0,48
10	8	109	3	97,05	127,85	18,35	1,25	20,5	1	11,51	2,97	121	3,62	0,43

O grupo 2 difere do grupo 1 por possuir escassa pubescência, microaristas e panículas intermediárias. Embora possua diferenças qualitativas, sob o ponto de vista quantitativo o grupo 2 é bastante semelhante ao primeiro grupo (Tabela 4). O terceiro grupo é o único a ter ângulos de folha bandeira inferiores a 30 graus juntamente com panículas abertas e aristadas. Destaca-se, sob o aspecto quantitativo, entre os mais produtivos (0,48 Kg), embora este número seja bastante inferior a média obtida pelos acessos comerciais (0,63 Kg), mas, estatisticamente, não houve diferença para esta característica (Tabela 1). O grupo 4 representou os acessos com os menores valores de CC e APM, com folhas bandeira mais largas e de ciclos mais longos (140 dias). O quinto grupo está entre os poucos a apresentar grau de pubescência médio. Apresenta o segundo maior número de perfilhos (6), superando, discretamente, a média obtida entre os acessos comerciais (5,46). O grupo 6 reúne acessos que guardam grandes semelhanças qualitativas ao grupo 5, diferindo apenas por apresentar menor grau de pubescência. Este grupo se destaca, entre todos os outros, por possuir emergência de plântula (NDE), número de dias para florescimento (NDF) e ciclo total (CTP) mais tardios, com 9, 132 e 151 dias, respectivamente. Isto se torna mais evidente ao compararmos esses números às médias referentes aos acessos comerciais utilizados

como testemunhas (Tabela 1), sendo estas, aproximadamente, 8, 105 e 111, respectivamente.

O grupo 7 tem como característica a presença em conjunto de microaristas, ângulos entre 61 e 90 graus, e ausência de pubescência. Destaca-se por possuir panículas com maior número de ramificações (3), maior valor médio em relação ao comprimento de folha bandeira, e os grãos são os mais curtos e largos, sendo os mais distantes do padrão comprido e fino apresentado pelos acessos comerciais. Outro destaque negativo do grupo é a baixa produtividade de grãos (0,42 Kg) em média. Embora possua aspectos negativos em relação a características como CE, LE e PG, o grupo 7, destaca-se positivamente por apresentar o maior número de perfilhos (7), superior a média apresentada pelos acessos comerciais (5,46). O grupo 8 assemelha-se, qualitativamente, ao primeiro grupo por possuir ausência de pubescência e arista, possuem ainda panícula aberta e ângulos de inserção (IFB) inferiores a 30 graus. Embora, qualitativamente, tão diferente do grupo 9, sob o ponto de vista quantitativo, os grupos 8 e 9, são praticamente idênticos. Tem por características apresentarem as mais altas plantas, com os maiores valores de CC e APM, longas folhas bandeira e os maiores comprimentos de panícula (25,2). Maior destaque é dado quando se refere ao CTP e CE. Estes grupos apresentam ciclos mais curtos (107 dias) que a média comercial (109 dias), além de grãos mais longos (9,77 mm) (Tabela 1). Por fim, o grupo 10 caracteriza-se por apresentar grau médio de pubescência, ângulos entre 61 e 90 graus, panículas abertas, aristada e juntamente com os grupos 8 e 9, possui o menor valor de (NDF).

Mesmo com plantas de estatura alta, conta com panículas curtas e com o menor número de ramificações entre todos os outros grupos (1). Suas folhas bandeira são longas e estreitas. Entretanto, devemos destacar seus grãos extremamente longos (11,51 mm) que superam expressivamente a média obtida pelos acessos comerciais (9,2 mm) sendo um indicativo potencial para possíveis retrocruzamentos.

2.3.3 Estimativas de distância intragrupos, intergrupos e de Mahalanobis

Com base na estimativa de distância de Mahalanobis, foi realizada uma análise individual, e não entre grupos, dentro da população em estudo. Esta análise baseou-se em dados referentes a 14 variáveis quantitativas, as quais contribuíram de maneira bastante variável na avaliação do grau de divergência dos acessos. A contribuição relativa de cada característica pode ser visualizada na Tabela 5, com destaque para o caracter comprimento do colmo (CC), que se mostrou de indiscutível importância para a análise contribuindo com mais de 98% da divergência. A amplitude das distâncias apresentou valor máximo de 10133250,03 entre os acessos 42 e 96, e o mínimo de 30.13 entre os acessos 59 e 88. Assim, sob o ponto de vista genético, conclui-se que os acessos 42 e 96 são os mais divergentes e os acessos 59 e 88 os mais similares.

Com relação às distâncias intra e intergrupos, com exceção dos grupos 8, 9 e 10 que possuem um único acesso, os valores de divergência intragrupos foram de aproximadamente 10158, 1282, 154, 18, 5, 1 e 2, para os grupos 1, 2, 3, 4, 5, 6 e 7 respectivamente. Esses valores apontam para uma provável presença de significativa diversidade principalmente dentro dos grupos 1, 2 e 3. Em termos de distância intergrupos, pode-se constatar maior divergência entre os grupos 1 e 2, 1 e 3, 1 e 4, 2 e 3 e 1 e 7, com os valores aproximados de 10099, 3952, 1655, 1314 e 1304 respectivamente. As menores divergências couberam aos grupos 9 e 10, 8 e 10, 8 e 9, 5 e 8 e 6 e 8, com os valores de 4, 4.6, 4.7, 9.9 e 11 respectivamente. Existem indícios, portanto, da presença de grande diversidade entre os diferentes grupos e mesmo dentro de cada grupo.

Tabela 5 – Contribuição relativa dos caracteres para a divergência, de acordo com Critério de Singh (1981)

Variável	Valor em %
CC	98,3408
CE	0,6791
LE	0,3033
CTP	0,1499
APM	0,1088
NDF	0,089
MCS	0,0694
NRP	0,0688
CP	0,0607
PG	0,0489
NDE	0,0308
LFB	0,0306
NP	0,0117
CFB	0,0083

2.3.4 Variáveis Canônicas

Com o objetivo de avaliar a influência de cada característica na análise de diversidade entre os acessos, a importância relativa das variáveis canônicas foi medida pela porcentagem de seus autovalores (variâncias) em relação ao total dos autovalores, ou seja, é a porcentagem da variância total que elas explicam (Tabela 6).

Como pode ser visto na tabela 6, as duas primeiras variáveis canônicas explicaram cerca de 99,71% da variação total, sendo desta 94,27% referente à primeira variável canônica e 5,43% referente a segunda. Para a primeira variável canônica as características mais importantes e que mais contribuíram foram altura de planta na maturidade (APM), comprimento de colmo (CC), concordando com os resultados encontrados pela distância de Mahalanobis (tabela 5) e comprimento de panícula (CP). As características mais importantes na segunda variável canônica foram comprimento de espiguetas (CE) e número de dias para o florescimento (NDF). Pela grande

importância relativa, portanto, as características APM, CC, CP, CE, e NDF merecem constar em estudos posteriores referentes à avaliação de acessos deste germoplasma.

Tabela 6 - Estimativa dos autovalores na análise de variáveis canônicas

Variável Canônica	Raiz (%)	% acumulada
1	94,2736	94,2735827
2	5,4385	99,7120805
3	0,1734	99,885477
4	0,04473	99,9302066
5	0,02935	99,9595568
6	0,01311	99,972664
7	0,00943	99,9820896
8	0,00549	99,9875818
9	0,00449	99,9920722
10	0,00301	99,9950827
11	0,00165	99,9967349
12	0,00152	99,9982506
13	0,00113	99,9993849
14	0,00062	100

2.3.5 Análises moleculares

2.3.5.1 Análise de parâmetros genéticos e dendrograma

Um conjunto de 15 marcadores (13 genômicos e 2 EST-SSR), altamente informativos e distribuídos por todos os cromossomos pertencentes ao genoma do arroz, foi selecionado para a caracterização dos 150 acessos de arroz filipino. Inicialmente, foi proposta a análise dos dados a partir de seis painéis multiplex a serem avaliadas em seqüenciador automático. No entanto, devido a problemas na qualidade de amplificação, contendo excesso de falhas e sobreposição dos locos, optou-se pela avaliação dos 15 locos individualmente, sendo que 14 deles por meio de géis de

poliacrilamida corados com nitrato de prata, e um deles (RM 1) avaliado em seqüenciador automático.

O índice de polimorfismo foi estimado para cada um dos 15 locos presentes no estudo (Tabela 7), o maior valor encontrado refere-se ao loco RM 257 (0,927), sendo assim o mais informativo, e o menor valor para o loco RM 277 (0,445), ambos são locos genômicos. A média encontrada foi de 0,763, valor esse superior ao encontrado por Lapitan et al. (2007), que encontrou valor médio de 0,68, avaliando 24 acessos de arroz filipino. Segundo a classificação de Botstein et al. (1980), este valor médio de PIC é considerado altamente informativo, classificando dessa forma valores superiores a 0,5.

Tabela 7 - Estimativas por loco dos índices de polimorfismo (PIC)

Locos	PIC
OG61	0,8794
RM38	0,8679
RM222	0,6937
RM229	0,6793
RM223	0,9103
RM335	0,9144
RM257	0,9272
RM1	0,6743
RM234	0,6596
RM207	0,8288
RM277	0,4459
RM231	0,7465
RM190	0,7574
RM252	0,6755
RM149	0,7977

Desta forma, tendo por base a classificação acima, todos os marcadores são altamente recomendados para avaliação da diversidade genética, visto seus altos índices de detecção do polimorfismo.

A numeração e a correspondência dos genótipos utilizados nas análises moleculares é a mesma apresentada para os caracteres agromorfológicos, variando de 1 a 150, sendo que os seis últimos números correspondem a variedades comerciais. A Figura 4 representa parte do gel referente ao *primer* RM 252, exemplificando a presença de polimorfismo entre os acessos.

Figura 4 - Gel de acrilamida do *primer* microsatélite genômico RM 252

Os acessos analisados apresentaram um total de 156 alelos distribuídos pelos 15 locos. O número de alelos por loco variou de três (RM 277) a 20 (RM 257 e RM 335) (Figura 5). A diversidade genética por loco foi máxima para o *primer* RM 257 (0.938) e mínima para RM 277 (0.544), o que já era esperado devido à variação alélica observada acima.

Figura 5 - Número de alelos por loco para cada um dos 15 *primers* utilizados no estudo

A média encontrada foi de 10,4 alelos por *primer*, sendo superior a encontrada por Coburn et al. (2002), Lapitan et al. (2007) e Ni et al. (2002), ao quais obtiveram 5, 5,8 e 6,8 alelos por loco em média, respectivamente, utilizando 159, 164 e 111 marcadores SSR em acessos de arroz. Esta média, também, foi bastante superior a encontrada por Silva et al. (2007), encontrando uma média de 5 alelos por loco analisando 7 populações de *Oryza glumaepatula* por meio de sete marcadores SSR.

Foram calculadas as freqüências alélicas dentro de cada loco as quais estão representadas pela figura 6.

Figura 6 - Freqüência alélica referentes a todos os *primers* utilizados no estudo (continua)

Figura 6 - Frequência alélica referentes a todos os *primers* utilizados no estudo (continuação)

Figura 6 - Frequência alélica referentes a todos os *primers* utilizados no estudo (continuação)

Figura 6 - Frequência alélica referentes a todos os *primers* utilizados no estudo (conclusão)

Dos 156 alelos totais encontrados, apenas dois alelos apresentaram frequência superior a 50%, o que mais uma vez comprova a grande divergência genética entre os acessos. Um alto índice de alelos raros foi encontrado (70 alelos) representando cerca de 45% do total.

Com relação aos alelos ditos privados ou exclusivos, observados em apenas um acesso no loco, foram encontrados apenas 14 (Tabela 8). Os *primers* que mais distinguiram este tipo de alelo foram os *primers* RM 229 e RM 1, apresentando 3 alelos cada, seguidos pelo RM 223 com 2 alelos e OG 61, RM 335, RM 257, RM 207, RM 190 e RM 252 com apenas 1 alelo exclusivo. O número total de alelos privados encontrados neste trabalho é bastante inferior ao apresentado por Borba (2005), que encontrou 89 destes alelos avaliando 242 acessos pertencentes à coleção nuclear brasileira de arroz (CNBA) utilizando 25 locos SSR.

Tabela 8 - Alelos exclusivos detectados entre os 150 acessos avaliados

Primers	Alelo	Acessos
OG 61	164	84
RM 229	136, 138, 164	106, 68, 132
RM 223	154, 192	112, 64
RM 335	138	146
RM 257	224	121
RM 1	94, 98, 124	85, 138, 133
RM 207	154	105
RM 190	130	107
RM 252	210	47

A presença de alelos privados pode ser representativa de um processo distinto de adaptação ambiental, independente de este alelo diferencial ser diretamente relacionado com uma característica fenotípica favorável para o melhoramento genético. Pois, apesar dos marcadores moleculares serem neutros, é possível que estejam ligados a um gene que apresenta maior adaptabilidade.

Além disso, a presença deste tipo de alelo sugere um possível potencial, desses acessos, para contribuir no aumento da base genética do arroz no Brasil. Neste sentido, os acessos possuidores de alelos privados podem ser prioritariamente incluídos nos blocos de cruzamento em dialelo, para avaliar sua capacidade geral e específica de combinação (Borba, 2005).

A heterozigosidade esperada, parâmetro também conhecido como diversidade genética de Nei (NEI, 1978), teve seu maior valor para o *primer* RM 257 (0,934) e seu menor valor (0,542) referente ao *primer* RM 277 (Tabela 9).

Tabela 9 - Valores de Heterozigosidade esperada para 15 locos SSR

Locos	H_E
OG61	0,892936
RM38	0,88295
RM222	0,739882
RM229	0,705214
RM223	0,919762
RM335	0,923252
RM257	0,934665
RM1	0,725691
RM234	0,698418
RM207	0,851023
RM277	0,54252
RM231	0,783295
RM190	0,791505
RM252	0,721759
RM149	0,823224

A média para os valores de H_E encontradas neste estudo foi de 0,795, sendo esta superior as encontradas por Xu et al. (2004), Coburn et al. (2002), Ni et al. (2002) e Silva et al. (2007), sendo estas 0,66, 0,67, 0,62, e 0,24, respectivamente, e pouco inferior as encontradas por Borba (2005) (0,81). Como já era esperado, devido ao modo de reprodução da espécie, não houve nenhum acesso heterozigoto sendo, portanto, H_o igual a zero.

Considerando todos os 15 marcadores SSR utilizados, foi construída uma matriz de distâncias de acordo com os parâmetros de Rogers-W a qual foi utilizada para a construção de um dendrograma pelo critério de agrupamento UPGMA.

Importante salientar que nenhuma informação a respeito da origem e ou localização dos genótipos esta disponibilizada.

A análise da figura 7 sugere a formação de três grupos principais (A, B e C). O primeiro grupo (A), em relação ao número de acessos, é o segundo maior contendo um total de 50 genótipos o que representa, aproximadamente, 33,3% do total. Este grupo se destaca por incluir todas as seis testemunhas usadas no estudo (145, 146, 147, 148, 149 e 150) e por conter o segundo maior número de acessos contendo alelos exclusivos, sendo estes os genótipos 85, 107, 121, 132, e 146. Estes acessos podem ser interessantes, pois, além de possuírem alelos exclusivos, são os mais semelhantes geneticamente às testemunhas, que possuem características notáveis como alto nível de produtividade, baixo acamamento, certo nível de tolerância a algumas doenças como o brusone de folha, brusone no eixo, tolerância a seca, além de boas qualidades culinárias, em relação a genótipos dos outros grupos e que contêm este tipo de alelo.

O segundo grupo (B) é o maior em relação ao número de acessos reunindo 70 genótipos, valor que corresponde a, aproximadamente, 46,6% do total. Este grupo, além de não conter nenhum genótipo testemunha, é o que menos reúne acessos portando alelos exclusivos.

O terceiro e último grupo (C) é o menor em número de componentes reunindo apenas 30 acessos, representando 20% do total. Assim como o grupo anterior não possui nenhum genótipo testemunha, mas se destaca por conter o maior número de acessos que apresentam alelos exclusivos (7), o que corresponde a 50% do total de genótipos com essa característica. São estes os acessos 138, 133, 112, 105, 68, 84 e 106.

De maneira geral, a maior semelhança genética foi encontrada entre os acessos 130 e 131, ambos pertencentes ao primeiro grupo (A), com distância genética de 0,28 segundo a matriz de Rogers modificada por Wright.

Alguns dos nós merecem algum destaque, pois reúnem testemunhas comerciais a outros genótipos não comerciais. Desta forma podemos destacar um nó contendo o genótipo 18, não comercial, junto a três testemunhas (147, 148 e 149). Estas três variedades comerciais possuem características muito interessantes. Podemos citar como exemplo a variedade 147 (Caiapó) que segundo a Embrapa, possui certa tolerância ao brusone de folha, tolerância moderada ao brusone no eixo, além de qualidade dos grãos e culinária.

Figura 7 - Dendrograma referente aos 15 locos, construído por distâncias de Rogers-W e pelo agrupamento UPGMA ($r = 0.69891$)

Já a testemunha 148 (Curinga) está relacionada a baixo acamamento, menor suscetibilidade ao brusone, tolerância ao percevejo do colmo e notável resistência a seca. Segundo o IAC, a variedade IAC 1246, aqui denominada como testemunha 149, além de grande produtividade, também está ligada a tolerância a seca. Outro nó interessante reúne o acesso 25 ao genótipo testemunha 150 (IAC 201) relacionada à boa qualidade e tipo de grão, com semelhança às variedades irrigadas, além de alto rendimento no beneficiamento, e excelente qualidade culinária. Também podemos destacar o nó que reúne os acessos 38 e 42 a testemunha 145 (IAC 202) a qual está relacionada à boa qualidade industrial e culinária, com boa arquitetura de planta.

Ao se analisar a matriz, pode-se verificar que estes acessos possuem as menores distâncias genéticas em relação aos acessos testemunha sendo de 0,67 para 18 x 149, 0,65 para 18 x 150 e 18 x 147, 0,62 para 18 x 148. O acesso 38 possui 0,67 em relação ao 145. A menor distância fica por conta do acesso 25, em relação à testemunha 150, apresentando distância de 0,53.

Ao fazer uma comparação superficial entre o dendrograma acima e o agrupamento obtido para os resultados de campo, pode-se ver algumas semelhanças e diferenças. Com relação às semelhanças, podemos destacar o acesso 18 que, como comentado acima possuía grande relação às testemunhas 150 e 148, IAC 201 e CURINGA respectivamente, o que também ocorreu segundo aquele agrupamento compondo o mesmo grupo. Este fato se repete para o acesso 25 o qual mais uma vez se agrupa a testemunha 150. Outro fato interessante é a formação de um nó envolvendo as testemunhas 147 e 149, CAIAPÓ e IAC 1246 respectivamente, que, com base no dendrograma, possuem distância genética de 0,55. Estes mesmos genótipos também se agrupam com base nos dados de campo, compondo um grupo que não envolve outros genótipos testemunha.

No entanto, diferenças ocorrem como o não agrupamento entre o acesso 18 e a testemunha 149, acontecendo o mesmo com os acessos 38 e 145, que, para caracteres agromorfológicos, este último compunha um grupo separado formado apenas pelo dado genótipo.

Apesar de se sugerir a formação de três grupos, de maneira geral, os acessos não possuem um padrão de estruturação clara, apresentando baixos valores de

bootstraps. Este fato sugere a existência de uma base genética diversificada, isto é, ocorre grande diversidade genética dentro do conjunto de acessos.

Através desta análise, também, não foi observada a existência de duplicatas dentro do grupo, o que reforça a necessidade de manutenção de todos os acessos pertencentes ao banco de germoplasma.

2.3.5.2 Análise através do programa Structure

Com base no mesmo conjunto de dados moleculares, os quais deram origem às análises anteriores, foi feito um novo agrupamento pelo programa computacional STRUCTURE 2.3.

Utilizando este programa, foram testados 10 arquivos de parâmetros contendo valores de $k=1$ à $k=10$. Foram utilizadas cinco corridas para cada valor de K , admitindo-se um modelo sem mistura, frequências alélicas independentes, 100.000 *burn-ins* e 200000 simulações de Monte Carlo de cadeias de Markov (MCMC). A seleção do número K , que se refere ao número de grupos mais provável para o dado conjunto de dados, foi realizada através dos valores de ΔK , de acordo com o método proposto por EVANNO et al. (2005). Os dados foram analisados segundo modelo haplóide.

Segundo este autor, o valor de K selecionado com o maior valor de ΔK foi de três (Figura 8), representada pelas cores vermelha, verde e azul. Esta informação corrobora de início, ao menos numericamente, a sugestão feita para a subdivisão do dendrograma (Figura 7).

Figura 8 - Valores de ΔK para cada valor de K, calculado nos acessos do banco de germoplasma da ESALQ, de acordo com o proposto por Evanno et al. (2005). O maior valor de ΔK corresponde ao K ótimo

Desta forma o agrupamento realizado por este programa teve como resultado os valores representados na figura abaixo:

Figura 9 - Teste de atribuição para os acessos de arroz filipino avaliados com (K=3). Acessos representados pelas barras verticais coloridas. A mesma cor em acessos diferentes indica que eles pertencem ao mesmo grupo. Cores diferentes no mesmo indivíduo indica a probabilidade deste pertencer a certo grupo K. Identificação dos acessos na Tabela 12

O eixo y, com valores que variam de 0 a 1, corresponde à probabilidade de determinado acesso se enquadrar ao respectivo grupo. Por exemplo, o acesso 140 possui cerca de 59,5% de probabilidade de pertencer ao grupo representado pela cor azul e cerca de 40,5% de pertencer ao grupo verde.

Inicialmente, pode-se sugerir uma boa correspondência entre este agrupamento e o observado para o dendrograma (Figura 7).

O grupo azul apresenta grande semelhança ao grupo A (Figura 7). Neste caso, reúne, no total, 58 acessos, oito a mais que o grupo A. Assim como este grupo, também reúne todas as testemunhas (145, 146, 147, 148, 149 e 150). Diferencia-se do grupo A pela presença de alguns genótipos (55, 72, 80, 76, 57, 43, 65, 8, 33, 52 e 54) alocados no grupo B. Isto pode ser explicado visto que, exceto os genótipos 55, 72, 80 e 76, esses acessos possuem certa probabilidade de pertencer ao grupo verde, semelhante ao grupo B.

O grupo verde guarda maior semelhança ao grupo B, também, devido a sua composição de acessos. Destaca-se como o maior dos três grupos, reunindo um total de 62 acessos. Desta forma, se mostra um pouco menor quando comparado ao grupo B, o qual reúne 71 acessos. Além disso, mantém a ausência de testemunhas e o menor número de alelos exclusivos. Os acessos 99, 2, 85 e 144 estão presentes no grupo verde, porém ausentes no grupo B. Já os acessos 8, 43, 17, 57, 33, 52, 54, 55, 72, 65, 76 e 80 constam no grupo B, porém, não fazem parte da composição do grupo verde. Embora estes genótipos não façam parte deste grupo, possuem certa probabilidade de pertencê-lo, com exceção aos acessos 17, 55 e 72.

O grupo vermelho guarda grande semelhança ao grupo C, reunindo todos os 29 acessos pertencentes a esse grupo, somado ao acesso 17, proveniente do grupo B. Sendo assim, também se destaca pela presença da maioria dos acessos portando alelos privados ou exclusivos.

Sendo assim, com base nos dados moleculares, organizados por meio de um dendrograma e do agrupamento realizado pelo programa STRUCTURE 2.3, podemos sugerir a existência de certo grau de estruturação, dando origem a três subdivisões. Embora os valores de *bootstraps*, referentes aos nós que compõem o dendrograma, tenham se apresentado de forma não significativa, em parte, a análise do Structure

suporta a sugestão de formação de três grupos principais, conferindo maior consistência a esta sugestão. Importante lembrar que a semelhança entre estas duas análises não se resume ao número de grupos, mas sim, principalmente, a sua composição de genótipos.

2.3.6 Agrupamento geral pelo método de otimização de Tocher

Com o objetivo de diminuir erros referentes à influência ambiental e aumentar a confiabilidade dos dados avaliados, foram utilizadas para esta análise o conjunto total de informações obtidas envolvendo dados agromorfológicos e moleculares.

Com base na soma das matrizes individuais, referentes aos dados quantitativos e qualitativos do primeiro ano agrícola, quantitativos do segundo ano agrícola de campo e moleculares, foi realizado o agrupamento com as distâncias desta matriz resultante segundo o método de otimização de Tocher.

Foram discriminados, por este método, um total de 26 grupos (Tabela 10). Os grupos 1, 2, 3, 5 e 6 podem ser considerados principais por abranger, aproximadamente, 69% do total de acessos, sendo que somente o grupo 3 representou cerca de 21%; grupo 1, (19%); grupo 2, (12%); grupo 6, (9%) e grupo 5 com 7%. Os grupos restantes reuniram um conjunto menor de acessos, sendo os grupos 7 e 8 com 6 genótipos, grupos 4 e 9 com 5 acessos e grupo 10 com 3 acessos. Um total de 6 grupos apresentaram 2 acessos cada e 10 grupos foram formados com apenas um representante.

Tabela 10 - Agrupamento dos acessos pelo método de Tocher envolvendo o conjunto total de dados

[illegible]

Partindo-se da análise do agrupamento apresentado acima pode-se verificar a importância de uma análise envolvendo aspectos agromorfológicos e moleculares. Desta forma, observa-se a presença de maior estruturação de grupos (26), em comparação aos outros agrupamentos envolvendo apenas dados agromorfológicos referentes ao primeiro ano agrícola do estudo (10) e dados moleculares (3). Estes dados sugerem o uso do maior número possível de dados disponíveis, os quais proporcionarão resultados mais confiáveis e completos chegando o mais próximo possível da realidade.

Destaca-se o grupo 1 por conter 5 (146, 147, 148, 149 e 150) das 6 testemunhas usadas no estudo. A única testemunha que não se enquadrou no grupo 1 foi o genótipo 145 o qual compõe o grupo 16 juntamente com o genótipo 82. Este resultado foi semelhante ao encontrado para o agrupamento agromorfológico do primeiro ano onde o 145 compunha um grupo isolado de apenas um acesso. A diferença entre este genótipo comercial e os outros cinco, pode ser observada, principalmente, com base nos dados agromorfológicos, já que compõe o mesmo grupo dos outros acessos comerciais sob o aspecto molecular. Já o acesso 82 compunha outro grupo, no qual não havia nenhuma testemunha.

Além de contar com acessos qualitativamente mais semelhantes aos genótipos comerciais, com ausência de pubescência no limbo, panículas compactas e não aristadas, sob o ponto de vista molecular, também se destacam pela similaridade genética abrangente dada pelos marcadores moleculares, com relação aos comerciais, uma vez que compõem, também, o mesmo grupo molecular.

De forma geral pode-se sugerir que os acessos pertencentes ao grupo 1 possuem semelhança agromorfológica e molecular às testemunhas. Essa semelhança é bastante interessante sugerindo um provável potencial destes genótipos, visto que os acessos comerciais possuem um conjunto de características apreciadas pelo mercado como já salientado neste trabalho.

O grupo 2, apesar de conter acessos semelhantes qualitativamente aos comerciais, se difere dos mesmos por não comporem o mesmo grupo molecular. No entanto, o grupo 2 possui importância devido à presença de acessos com maior número de alelos exclusivos. A única exceção no grupo é do acesso 27 que se diferencia

qualitativamente dos outros por possuir microaristas, ângulo de inserção de folha bandeira menor que 30 graus e pubescência média do limbo.

O terceiro e maior grupo, se caracteriza por apresentar 75% de seus componentes com semelhanças qualitativas aos comerciais, com exceção dos genótipos 97, 51, 111, 45, 139, 95 e 28, os quais possuem microaristas, panículas do tipo intermediária e pubescência escassa do limbo. Todos eles, no entanto, não se assemelham as testemunhas sob o ponto de vista molecular, uma vez que compõem um grupo caracterizado pela ausência de acessos comerciais além de menor frequência de alelos exclusivos. O único acesso pertencente a este grupo que compõe o grupo molecular das testemunhas é o de número 8. O agrupamento deste junto aos demais pode ser explicado pela porcentagem significativa de seu genoma semelhante ao grupo dos demais chegando a 35%.

Pode-se sugerir que o quarto grupo possui como critério principal de agrupamento sua semelhança genômica sob o ponto de vista molecular. Embora todos pertençam ao grupo cuja característica principal é a presença de alelos exclusivos, nenhum componente do grupo 4 possui tais alelos, além de pertencerem a grupos agromorfologicamente distintos.

O grupo número 5, embora seja composto, em sua maioria, por acessos qualitativamente semelhantes aos comerciais, porém com diferenças moleculares, assim como grande parte do grupo 3, estes se diferenciam por reunir acessos com altos valores de NDF (número de dias para o florescimento) e altíssimos valores de ciclo total de planta (CTP). Assim, foram agrupados pelo fato de reunirem acessos de ciclo extremamente longos, e bem acima da média.

O sexto grupo é bastante heterogêneo. Tem como característica reunir acessos de duas formas. A primeira é dada por acessos notavelmente produtivos, aliados com ciclos menores que a média dos acessos filipinos, característica esta de grande importância. Podemos destacar dentre estes os acessos 58, 64 e 94 que, além de apresentarem estas características, também possuem semelhança molecular a genótipos comerciais, visto que pertencem ao mesmo grupo sob este ponto de vista.

A outra é dada por acessos muito pouco produtivos, aliados a ciclos maiores. Desta forma podemos sugerir a união de extremos como critério importante para o agrupamento.

O grupo 7 é composto por 6 acessos todos pertencentes ao mesmo grupo molecular das testemunhas denotando proximidade a estas ultimas ,com relação ao genoma. Destaca-se por reunir acessos com, relativamente, boa produtividade, aliados a um baixo número de perfilhos e comprimento de folha bandeira. Além disso, possuem notável massa de cem sementes, significativamente acima da média.

Mesmo com a semelhança ao grupo anterior, em relação a presença de altos valores de MCS, o oitavo grupo se diferencia pela não semelhança molecular aos genótipos testemunha. Uma característica interessante a este grupo é o baixo valor em relação ao número de dias para o florescimento (NDF).

Aliados a um florescimento precoce, o grupo nove reúne acessos com boa largura de folha bandeira, panículas curtas, grãos mais largos, ciclos curtos e boa produtividade, acima da média em relação aos outros filipinos.

O décimo grupo reúne acessos qualitativamente semelhantes aos genótipos comerciais, semelhantes a estes também sob o ponto de vista molecular. Caracteriza-se por possuir maior período para emergência (NDE).

Com plantas menores e com um acesso (133) portando um alelo exclusivo, o grupo 11 se destaca por apresentar também ciclos longos, florescimento tardio, folhas bandeira delgadas, baixo valor de MCS e produtividade baixa.

O décimo segundo grupo apresenta acessos com baixo número de perfilhos, plantas significativamente altas e com largura de folha bandeira acima da média geral.

Já o grupo 13 apresenta acessos que já compunham o mesmo grupo agromorfológico do primeiro ano. São acessos de florescimento tardio, com baixo número de perfilhos, ciclos longos e produtividade baixa.

Os acessos do grupo 14 pertencem ao mesmo grupo agromorfológico e molecular, sugerindo semelhanças qualitativas e genômicas aos genótipos comerciais. Tem por característica possuírem longas folhas bandeira e com boa largura.

Acessos pertencentes ao grupo 15 possuem em comum o florescimento precoce, folha bandeira estreita, além de comporem o mesmo grupo agromorfológico, o qual

reúne 3 acessos comerciais, e mesmo grupo molecular caracterizado pela presença de maior número de alelos exclusivos.

Assim como discutido anteriormente, o grupo 16 reúne 2 acessos sendo um comercial (145) e outro filipino (82). O provável motivo de comporem o mesmo grupo deve se relacionar ao fato de serem altamente semelhantes sob o ponto de vista molecular, visto que são notavelmente distinto agromorfológicamente. Pelo fato de comporem um grupo a parte, sugere-se que estes acessos sejam bastante distintos, molecularmente, em relação aos outros.

A partir do grupo 17 até o último (26), todos são compostos por apenas um acesso.

O primeiro destes grupos (17), formado pelo acesso 5 (Figura 10), é pertencente ao grupo molecular onde estão todos os cultivares comerciais usados como testemunha. Trata-se de um acesso aristado, com médio número de perfilhos, colmo baixo, boa largura de folha bandeira, com panículas bem ramificadas.

Figura 10 - Aspecto dos grãos referentes ao acesso 5

O acesso 52 (Figura 11) compõe o grupo dezoito sendo pertencente ao grupo molecular das testemunhas, além de apresentar florescimento precoce, baixo

perfilamento com comprimento e largura de folha bandeira de larga extensão. Outra característica de interesse é o grande tamanho de suas panículas.

Figura 11 - Aspecto dos grãos referentes ao acesso 52

Como pode-se observar na figura abaixo (Figura 12), o acesso 9 que compõe o grupo 19 se caracteriza por possuir grãos curtos e largos. Além disso, possui ciclo longo, panícula curta, baixa produtividade de grãos, comprimento e largura de folha bandeira de menor extensão. Um aspecto positivo é a emergência precoce (NDE) sendo necessário 6 dias. Apenas outro acesso (27), dentre todos os 150, possui tal característica.

Figura 12 - Aspecto dos grãos referentes ao acesso 9

O vigésimo grupo é formado pelo acesso 106 (Figura 13), portador do alelo exclusivo 136, referente ao loco RM 229. Compõe o grupo C molecular caracterizado por reunir o maior número de acessos portadores de alelo exclusivo. Possui baixo perfilhamento, baixa massa de cem sementes e folha bandeira delgada. Além disso, apresenta ciclo longo e boa ramificação

Figura 13 - Aspecto dos grãos referentes ao acesso 106

Características de grande interesse podem ser vistas no acesso 57, componente único do grupo 21. Além de compartilhar semelhanças agromorfológicas e moleculares, junto aos genótipos comerciais, fazendo parte dos mesmos grupos que estes, se destacam, também, por outras características. Possui florescimento precoce, folhas bandeira longas e largas, panículas grandes e ramificadas, e o mais importante é que todos estes valores foram superiores às médias dos genótipos comerciais. Somando-se estas características com a boa produtividade faz com que esse acesso mereça atenção especial dos melhoristas, por ser um doador de alelos potencial aos programas de melhoramento. O único aspecto negativo se refere à forma do grão, sendo este ligeiramente largo e curto (Figura 14).

Figura 14 - Aspecto dos grãos referentes ao acesso 57

Ao contrário do grupo anterior, o acesso 137 que compõe o grupo 22 apresenta um conjunto de características pouco interessantes. Possui florescimento tardio, baixo padrão de perfilhamento, folhas bandeira curtas, panículas curtas e ramificadas, baixa massa de cem sementes, ciclo longo e baixa produtividade. Abaixo segue exemplo de panícula referente a este acesso (Figura 15).

Figura 15 - Aspecto de panícula referente ao acesso 137

O grupo 23 é composto pelo acesso 24, que por sua vez apresenta florescimento precoce, bom padrão de perfilhamento, folhas bandeira e panículas curtas, grãos largos (Figura 16) e boa produtividade.

Figura 16 - Aspecto dos grãos referentes ao acesso 24

O vigésimo quarto grupo é formado pelo acesso 121 (Figura 17), também portador do alelo exclusivo 224, referente ao loco RM 257. Trata-se de um acesso de

florescimento tardio, plantas baixas, ciclo longo, baixa produtividade e massa de cem sementes, com grãos notavelmente curtos. Um aspecto interessante ocorre em relação ao comprimento e largura de folha bandeira, ambos bastante superiores, inclusive, as médias encontradas para os genótipos comerciais.

Figura 17 - Aspecto dos grãos referentes ao acesso 121

O grupo 25, composto pelo acesso 124 se caracteriza por apresentar baixo padrão de perfilhamento, folhas bandeira curtas e largas, panículas pouco ramificadas, grãos largos (Figura 18) e com massa de cem sementes muito superior a média.

Figura 18 - Aspecto dos grãos referentes ao acesso 124

Finalmente, o grupo de número 26, composto unicamente pelo acesso 41, apresenta florescimento bastante tardio, plantas altas, panículas curtas e ramificadas, ciclo longo, grãos bem largos (Figura 19) e pouco produtivos. Por outro lado, possui bom padrão de perfilhamento, com folhas bandeira, apesar de delgadas, notavelmente longas, sendo superior a média encontrada pelos genótipos comerciais.

Figura 19 - Aspecto dos grãos referentes ao acesso 41

A alta estruturação encontrada por esta análise reflete pela alta diversidade encontrada, tanto durante as coletas de dados nos dois anos agrícolas estudados, quanto pelos padrões moleculares apresentados. Isto também pode ser observado, embora de maneira superficial, pela variação no aspecto dos grãos apresentados pelas imagens anteriores.

2.3.6.1 Estimativas de distância intra e intergrupos

Com relação às distâncias intragrupo, os maiores valores foram 6070, 5010, 1910, 982 e 652, para os grupos 3, 1, 2, 6 e 5 respectivamente. Esses resultados coincidem com os grupos que reúnem o maior número de genótipos sugerindo que, mesmo dentro destes grupos, ocorre significativa diversidade genética. Em termos de distância intergrupos, pode-se constatar maior divergência entre os grupos 1 x 3, 2 x 3, 1 x 2, 3 x 6, 1 x 6 e 3 x 5, com os valores aproximados de 13790, 8886, 7845, 6066, 5554 e 5237 respectivamente. Os menores valores foram encontrados entre os grupos 17 x 18, 17 x 19, apresentando o valor de 13, e 17 x 21, 18 x 21, 19 x 20, 21 x 23 e 25 x 26 com o valor 14.

3 CONSIDERAÇÕES FINAIS

Os resultados das análises de agrupamento confirmaram a importância de uma análise abrangente, utilizando para isso, dados agromorfológicos, moleculares e principalmente pela união de todas as informações disponíveis, visto que para cada uma delas, diferentes padrões de agrupamento foram verificadas. Este fato sugere que esses diferentes métodos avaliam de maneira distinta a presença da diversidade genética em acessos de arroz filipino e por esse motivo reafirmam sua importância para uma avaliação segura e robusta.

Dessa forma, os resultados obtidos nesse trabalho indicam que os caracteres agromorfológicos, incluindo dados quantitativos e qualitativos, tiveram maior poder de separação entre os acessos na análise de diversidade, em relação à análise molecular, pois geraram uma maior quantidade de grupos.

No entanto, assim como pode ser observado ao longo do estudo, o agrupamento que reuniu os dados agromorfológicos e moleculares teve como resultado uma maior estruturação dos acessos, resultado direto da somatória de todas as contribuições, proporcionando resultados mais próximos à realidade sendo de suma importância para qualquer processo de avaliação da diversidade genética. Outro motivo pelo qual pode se observar maior estruturação neste tipo de análise foi que, além dos dados agromorfológicos do primeiro ano e moleculares, foram inclusos dados agromorfológicos de um segundo ano agrícola, o que aumentou ainda mais o poder de separação dos acessos, não sendo, portanto, o resultado do agrupamento geral a soma dos agrupamentos individuais apresentados.

Com relação aos caracteres agromorfológicos, buscou-se aproveitar todo o seu potencial de avaliação por considerar tanto caracteres qualitativos quanto quantitativos. Os caracteres mais úteis na análise da diversidade genética, nos acessos de arroz avaliados, variáveis canônicas foram, altura de planta na maturidade (APM), comprimento de colmo (CC), comprimento de panícula (CP), comprimento de espiguetas (CE) e número de dias para o florescimento (NDF) os quais merecem destaque em estudos com este objetivo. Mesmo explicando pouco da variação, não se pode deixar de incluir a produtividade de grãos, devido a sua grande importância agrônoma. Já a característica cor de aurícula (CA), que não demonstrou variabilidade alguma, foi

considerada a menos discriminante, dentro do germoplasma, e, portanto, merece menor atenção em estudos posteriores. Alguns dos acessos que obtiveram destaque, ou seja, os quais apresentaram os maiores ou menores valores para cada caractere, no primeiro ano agrícola avaliado, estão relacionados abaixo (Tabela 11):

Tabela 11 - Maiores e menores médias dos acessos filipinos para os caracteres quantitativos avaliados

	NDE	NDF	NP	CC	APM	CFB	LFB	CP	NRP	CE	LE	CTP	MCS	PG
	(dias)	(dias)	(n°)	(cm)	(cm)	(cm)	(cm)	(cm)	(n°)	(mm)	(mm)	(dias)	(g)	(kg)
Maior	9	158	21	117	141,9	38,7	2,43	33,3	3	11,5	5,9	169	4,2	0,75
(acesso)	(47)*	-15	-14	-19	-19	-23	(73)*	-62	(41)*	-96	-41	-37	-62	-25
Menor	6	99	2	52,9	67	15,1	1	12,3	1	5,7	2,2	106	1,6	0,09
(acesso)	(9)*	(14)*	-117	-117	-117	-24	(34)*	-92	(18)*	-121	-46	-2	(46)*	-107

*valor encontrado para mais de um acesso.

Assim, por realizar uma análise da divergência genética considerando estes dois períodos, minimizou-se o efeito ambiental sobre as características quantitativas aumentando a precisão experimental, por meio da diminuição da interação genótipo x anos.

Foi constatada a presença de significativa diversidade dentro da maioria dos caracteres avaliados. Uma amostra desta variabilidade pode ser observadas nas figuras 20 e 21.

Figura 20 - Aspecto das panículas referentes aos acessos 54 (A), 150 (B), 149 (C), 34 (D), 86 (E) e 73 (F) representando parte da diversidade encontrada

Figura 21 - Aspecto dos grãos referentes aos acessos 146 (A), 54 (B), 5 (C), 123 (D), 86 (E) e 119 (F) representando parte da diversidade encontrada

A análise molecular também obteve resultados interessantes e avaliou a diversidade genética de maneira distinta aos caracteres agromorfológicos. Com base nos valores de PIC (*Polymorphism Information Content*), o marcador mais informativo utilizado, ou seja, aquele que melhor discriminou os dados moleculares foi o RM 257. Por outro lado, o *primer* menos informativo foi o RM 277, que mesmo sendo o menos importante, apresentou um índice de polimorfismo significativo. Tanto o marcador RM 257, quanto o RM 335, foram os que apresentaram maior número de alelos (20), sendo o que menos apresentou foi o RM 277, com apenas três alelos. Em relação a presença de alelos exclusivos, os marcadores que mais se destacaram foram RM 229 e RM 1, com 3 alelos em cada, sendo que os *primers* RM 38, RM 222, RM 234, RM 277, RM 231 e RM 149 não apresentaram este tipo de alelo.

Portanto, sugere-se que para uma boa análise de diversidade em arroz filipino, deve ser feita por meio do uso em conjunto tanto de marcadores microssatélites quanto de análises agromorfológicas, de preferência utilizando mais de um ano agrícola, diminuindo erros e para que se consiga acessar a diversidade genética total da espécie de maneira mais adequada e próxima da realidade.

A avaliação dos acessos deste estudo se mostrou bastante eficaz e condizente com seu objetivo que, além de melhor entender a relação entre os acessos do banco, contribuindo para uma melhor conservação deste patrimônio, ainda pode ser útil na incorporação de germoplasma exótico aos programas de melhoramento da espécie.

4 CONCLUSÕES

- Dentre todos os caracteres agromorfológicos analisados, altura de planta na maturidade (APM), comprimento de colmo (CC), comprimento de panícula (CP), comprimento de espigeta (CE) e número de dias para o florescimento (NDF) são os mais indicados para a análise da diversidade genética em acessos de arroz;
- Uma análise abrangente, envolvendo dados agromorfológicos, moleculares e, principalmente, a união destas informações, é muito importante, visto que, para cada uma delas, diferentes padrões de agrupamento foram verificados;
- Os acessos analisados, que possuem alelos exclusivos, são os genótipos 47, 64, 68, 84, 85, 105, 106, 107, 112, 121, 132, 133, 138 e 146;
- Com base na análise molecular, o marcador mais informativo utilizado foi o RM 257, sendo o *primer* RM 277 o menos informativo;

REFERÊNCIAS

- ALVES, R. M. **Caracterização genética de populações de cupuaçuzeiro, *Theobroma grandiflorum* (Willd. ex. Spreng.) Shum., por marcadores microssatélites e descritores botânico-agronômicos.** 2002, 159 p. Tese (Doutorado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura “Luiz de Queiroz” da Universidade de São Paulo, Piracicaba, 2002
- ANDERSON, J.A.; CHURCHILL, G.A.; AUTRIQUE, J.E.; TANKSLEY S.D.; SORRELLS M.E. Optimizing parental selection for genetic linkage maps. **Genome**, Toronto, v. 36, p. 181-186, 1993.
- BORBA, T.C.O.; **Análise por marcadores SSR dos grupos de genótipos melhorados e introduzidos da Coleção Nuclear Brasileira do Arroz.** 2005, 90p. Dissertação (Mestrado). Universidade Federal de Goiás, Goiânia, 2005.
- BOTSTEIN, D.; WHITE, R.L.; SKOLNIEK, M.; DAVIS, R.V. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **American Journal Human Genetic**, Boston, n. 32, p. 314-331, 1980.
- BRASIL. Decreto-lei nº 2.366, de 5 de novembro de 1997. Regulamenta a Lei n. 9.456, de 25 de abril de 1997, que institui a Proteção de Cultivares, dispõe sobre o Serviço Nacional de Proteção de Cultivares – SNPC, e dá outras providências. Diário Oficial [da] República Federativa do Brasil, Brasília, n. 216, p. 25342–25343, 7 nov. 1997. Seção 1.
- BRETTING, P.K.; WIDRLECHENER, M.P. Genetic markers and horticultural germplasm management. **HortScience**, Alexandria, v. 30, n. 7, p. 1349-1355, 1995.
- BRONDANI, C.; BORBA, T.C.O.; BRUNES, T.; RANGEL, P. H. N.; BRONDANI, R. P. V. Determination of genetic variability of traditional varieties of Brazilian rice using microsatellite markers. **Genetics and Molecular Biology**, Ribeirão Preto, v.29, n.4, 676-684, 2006a.

BRONDANI, C.; CALDEIRA, K. S.; BORBA, T. C. O.; RANGEL, P. N.; MORAIS, O. P.; CASTRO, E. M.; RANGEL, P. H. N.; MENDONÇA, J. A.; BRONDANI, R. P. V. Genetic variability analysis of upland rice genotypes from VCU experiment using SSR markers. **Crop Breeding and Applied Biotechnology**, Viçosa, v. 6, n. 1, p. 9-17, 2006b.

BRONDANI, C.; BRONDANI, R. P. V.; BORBA, T.C.O.; BRUNES, T.; RANGEL, P. H. N.; GUIMARÃES, E.P. **Utilização de marcadores microssatélites no melhoramento populacional do arroz**. Goiânia: EMBRAPA-CNPAP, 2004. 32 p. (EMBRAPA-CNPAP. DOCUMENTOS, 169).

BRONDANI, C.; BRONDANI, R.P.V. Germoplasma: base para a nova agricultura. **Ciência Hoje**. Rio de Janeiro, v. 35, n 207, p.70-73. 2004.

BRONDANI, C.; RANGEL, P.H.N.; BORBA, T.C.O.; BRONDANI, R.P.V. Transferability of microsatellite and sequence tagged site markers in *Oryza* species. **Hereditas**, Lund, v.138, p.187-192, 2003.

CALLOW, B.V.; FORD-LLOYD, B.; NEWBURY, H.J. Overview. In: CALLOW, B.V.; FORD-LLOYD, B.; NEWBURY, H.J. **Biotechnology and plant genetic resources: conservation and use**. Londres: CAB international, 1997. cap 01, p.05-10.

CHANG, T.T. **Manual of genetic conservation of rice germplasm for evaluation and conservation**. Manila: IRRI, 1976. 77 p.

COBURN, J. R.; TEMNYKH, S. V.; PAUL, E. M.; MCCOUCH S. R. Design and application of microsatellite marker panels for semiautomated genotyping of rice (*Oryza sativa* L.). **Crop Science**. Madison, v.42, p. 2092-2099. 2002.

CRESTE, S.; TULMANN NETO, A.; FIGUEIRA, A. Detection of Single Sequence Repeat Polymorphisms in Denaturing Polyacrylimide Sequencing Gels by Silver Staining. **Plant Molecular Biology Reporter**, New York, v. 19, n. 4, p. 299-306, 2001.

CRUZ, C.D. **Programa Genes – Versão Windows**. Viçosa: UFV, 2001. 642 p.

CRUZ, C.D.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa: UFV, 2003. 585 p.

CRUZ, C.D.; REGAZZI, A.J. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa: UFV, 1994. 642 p.

CRUZ, C.D. **Aplicação de algumas técnicas multivariadas no melhoramento de plantas**. 1990. 188p. Tese (Doutorado em Genética e Melhoramento de Plantas). Escola Superior de Agricultura "Luiz de Queiroz" - ESALQ, Universidade de São Paulo, Piracicaba, 1990.

DALRYMPLE, D. G. **Development and spread of high-yielding rice varieties in developing countries**. Washington: Agency for International Development, 1986. 117 p.

DIOLA, V. **Base de dados para os materiais genéticos vegetais disponíveis nas Estações Experimentais da Epagri**. 2005, 50 p. Tese (Doutorado) Universidade Federal de Santa Catarina, Florianópolis, 2005.

DOYLE, J.J.; DOYLE, J.L. Isolation of DNA from fresh tissue. **Focus**, Rockville, v.12, n.1, p. 13-15, 1987.

EVANNO, G.; REGNAUT, S.; GOUDET, J. Detecting the number of clusters of individuals using the software structure: a simulation study. **Molecular Ecology**, Vancouver, n. 14, p. 2611-2620, 2005.

FERREIRA, M.E. Molecular analysis of gene Banks for sustainable conservation and increased use of crop genetic resources. In: RUANE, J.; SONNINO, A. **The role of biotechnology in exploiting and protecting agricultural genetic resources**. Roma: FAO, 2006. p. 121-128.

FERREIRA, M. E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. 3ª ed. Brasília: Embrapa-CENARGEM, 1998. 220 p.

FONSECA, J.R.; CUTRIM, V.A.; RANGEL, P.H.N. **Descritores morfo agrônômicos e fenológicos de cultivares comerciais de arroz de várzea**. Brasília: Embrapa, 2002. 22 p. (Documento, 141).

FONSECA, J.R.; RANGEL, P.H.N.; PRABHU, A.S. **Características botânicas e agronômicas de cultivares de arroz (*Oryza sativa* L.)**. Goiânia: EMBRAPA-CNPAP, 1981. 33 p. (EMBRAPA-CNPAP. Circular Técnica,14).

FONSECA, J.R.; BEDENDO, I.P. **Características morfológicas, agronômicas e fenológicas de algumas cultivares de arroz**. Goiânia: EMBRAPA-CNPAP, 1984. 58p. (EMBRAPA-CNPAP. Boletim de Pesquisa, 3).

FONSECA, J. R.; DE CASTRO, E.M.; DE MORAIS, O.P. **Descritores morfo agronômicos e fenológicos de cultivares comerciais de arroz (*Oryza sativa* L.) de terras altas**. Santo Antônio de Goiás : Embrapa Arroz e Feijão, 2004. 28 p. – (Documentos / Embrapa Arroz e Feijão, Documentos, 162).

FRALEIGH, B. Global overview of crop genetic resources. In: RUANE, J.; SONNINO, A. **The role of biotechnology in exploiting and protecting agricultural genetic resources**. Roma: FAO, 2006, p. 21-32.

FREIRE, M. S; MORALES, E.A.V.; BATISTA, M.F. Diversidade genética. In: VIEIRA, N.R.A.; SANTOS, A.B.; SANT'ANA, E.P. **A cultura do arroz no Brasil**. Santo Antônio de Goiás: Embrapa Arroz e Feijão, 1999, p. 559-581.

GALLI, J. Origen, distribuição e domesticação do arroz. **Lavoura Arrouzeira**, Porto Alegre, v. 31, n. 307, p. 63-68, 1978.

GILBERT, J.E.; LEWIS, R.V.; WILKINSON, M.J.; CALIGARI, P.D.S. Developing an appropriate strategy to assess genetic variability in plant germplasm collections. **Theoretical and Applied Genetics**, Berlin, v. 98, p.:1125-1131, 1999.

GOFF, S.A. et al. A draft sequence of the rice genome (*Oryza sativa* L. Ssp. *Japonica*). **Science**, Washington, v. 296, p. 92-100, 2002.

GOUDET, J. **Fstat version 2.9.3.2**. Lausann, 2002. Disponível em: <<http://www2.unil.ch/popgen/softwares/fstat.htm>>. Acesso em: 25 nov. 2009.

HAWTIN, G.; IWANAGA, M.; HODGKIN, T. Genetic resources in breeding for adaptation. **Euphytica**, Dordrecht, v.92, p.255-266, 1996.

HOYT, E. Conservação dos parentes silvestres das plantas cultivadas. **International Board for Plant Genetic Resources**. Roma: IBPGR, 1992. 55 p.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. Disponível em: <<http://www.ibge.gov.br/home>>. Acesso em: 21 abr. 2009.

IBPGR-IRRI. **Descriptors for rice (*Oryza sativa* L.)**. Philippines, 1980. 21 p.

JENNINGS, P. R.; COFFMAN, W. R.; KAUFFMAN, H. E. **Rice improvement**. Los Baños: IRRI, 1979. 186 p.

KHUSH, G.S. Origin, dispersal, cultivation and variation of rice. **Plant Molecular Biology**, Dordrecht , v.35, p.25-34, 1997.

LAPITAN, V.C.; BRAR, D.S.; ABE, T.; REDONA, E.D. Assessment of Genetic Diversity of Philippine Rice Cultivars Carryng good Quality Traits using SSR Markers. **Breeding Science**, Tokyo, v. 57, p. 263-270, 2007.

LI, Y.C.; KOROL, A.B.; FAHIMA, T.; BEILES, A.; NEVO, E. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. **Molecular Ecology**, Vancouver, v. 11, n. 12, p. 2453-2465, 2002.

LI, Y.C.; KOROL, A.B.; FAHIMA, T.; NEVO, E. Microsatellites within genes: structure, function, and evolution. **Molecular Biology and Evolution**, Oxford, v. 21, n. 6, p. 991-1007, 2004.

LOPES, M.A. **Importância estratégica do intercâmbio de germoplasma para a embrapa e para a pesquisa agropecuária nacional**. Disponível em: <<http://www.cenargen.embrapa.br/cenargenda/pdf/intercambio.pdf>>. Acesso em: 12 nov 2008.

LU, J.J.; CHANG, T.T. Rice in its temporal and spatial perspectives. In: LUH, B.S. **Rice: production and utilization**. Westport, 1980. p. 1-74.

MARDIA, K.V.; KENT, J.T.; BIBBY, J.M. **Multivariate analysis**. New York: Academic, 1979. 521 p.

MAURYA, D.M.; SINGH, D.P. Genetic divergence in rice. **The Indian Journal of Genetics and Plant Breeding**, New Delhi, v. 37, n.3, p. 395-402, 1977.

McCOUCH, S.R.; TEYTELMAN, L.; XU, Y.; LOBOS, K.B.; CLARE, K.; WALTON, M.; FU, B.; MAGHIRANG, R.; LI, Z.; XING, Y.; ZHANG, Q.; KONO, I.; YANO, M.; FJELLSTROM, R.; DECLERK, G.; SCHNEIDER, D.; CARTINHO, S.; WARE, D.; STEIN, L. Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). **DNA Research**, Oxford, v. 9, p. 199-207, 2002.

MILLER, M. **Tools for population genetic analyses (TFPGA)**. Logan, 1997. Disponível em : <<http://www.marksgeneticsoftware.net/tfpga.htm>>. Acesso em: 25 nov. 2009.

MIRANDA, J.E.C.; CRUZ, C.D.; COSTA, C.P. Predição do comportamento dos híbridos de pimentão (*Capsicum annum* L.) pela divergência genética dos progenitores. **Revista Brasileira de Genética**, Ribeirão Preto, v. 11, p. 929-937, 1988.

MORAIS, O.P. **Análise multivariada da divergência genética dos progenitores, índices de seleção e seleção combinada numa população de arroz oriunda de intercruzamentos, usando macho-esterilidade**. Viçosa, 1992. 251 p. Tese (Doutorado em Genética e Melhoramento) - UFV, Viçosa, 1992.

MORGANTE, M.; HANAFEY, M.; POWELL, W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. **Nature Genetics**, New York, v. 30, n. 2, p. 194-200, 2002.

MULATO, B. M. **Diversidade genética em germoplasma de soja identificada por marcadores SSR, EST-SSR e caracteres agromorfológicos**. 2009, 88 p. Dissertação (Mestrado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura “Luiz de Queiroz” da Universidade de São Paulo, Piracicaba, 2009.

NASS, L.L. Utilização de recursos genéticos vegetais no melhoramento. In NASS, L.L.; MELO, I.S.; VALADARES-INGLIS, M.C. **Recursos genéticos e melhoramento de plantas**. Rondonópolis: Fundação MT, 2001, p.29-56.

NASS, L.L.; NISHIKAWA, M.A.N. Pre-melhoramento de germoplasma vegetal. In: GOES, M. **Tópicos em recursos genéticos**. Brasília: Embrapa, 2001. 1183 p.

NASS, L.L.; PATERNIANI, E. Pre-breeding: a link between genetic resources and maize breeding. **Scientia Agricola**, Piracicaba, v.57, p.581-587, 2000.

NASS, L.L.; PELLICANO, I.J.; VALOIS, A.C.C. Utilization of genetic resources for maize and soybean breeding in Brazil. **Brasilian Journal of Genetics**, Ribeirão Preto, v.16, p.983-988, 1993.

NEI, M. Estimation of average heterozygosity and genetic distance from a small number of individuals, **Genetics**, Pittsburgh, v. 89, n. 3, p. 583-590, 1978.

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

NI, J.; COLOWIT, P.M.; MACKILL, D.J. Evaluation of genetic diversity in rice subspecies using microsatellite markers. **Crop Science**, Madison, v. 42, p. 601-607, 2002.

PEREIRA, J.A. **Cultura do arroz no Brasil**: subsídios para a sua história. Teresina: Embrapa Meio-Norte, 2002, p. 17-26.

PHILLIPS, T. D.; MURPHY, J. P.; GOODMAN, M. M. Isozyme variation in germplasm accessions of the wild oat *Avena sterilis* L. **Theoretical and Applied Genetics**. Berlin, v. 86 p. 54-64, 1993.

PINHEIRO, B.S. Características morfofisiológicas da planta relacionadas à produtividade. In: VIEIRA, N.R.A.; SANTOS, A.B.; SANT'ANA, E.P. **A cultura do arroz no Brasil**. Santo Antônio de Goiás: Embrapa Arroz e Feijão, 1999. p. 116-147.

PINHEIRO, B. da S. **Morfologia e crescimento da planta de arroz**. In: CURSO INTERNACIONAL DE MELHORAMENTO GENÉTICO DE ARROZ GOIÂNIA, 1998, Goiânia. Goiânia: EMBRAPA-CNPAP, 1998. Palestra.

PRITCHARD, J. K.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, Pittsburgh, v. 155, n. 2, p. 945-959, 2000.

RANGEL, P.H.N.; GUIMARÃES, E.P.; NEVES, P.C.F. Base genética das cultivares de arroz (*Oryza sativa* L.) irrigado do Brasil. **Pesquisa Agropecuária Brasileira**, Brasília, v. 31, n.5, p. 349-347. 1996.

RANGEL, P.H.N.; CRUZ, C.D.; VENCovsky, R.; FERREIRA, R.P. Selection of local lowland rice cultivars based on multivariate genetic divergence. **Revista Brasileira de Genética**, Ribeirão Preto, v. 14, p. 437-453, 1991.

ROHLF, F.J. **NTSYS-Pc**: Numerical taxonomy and multivariate analysis system. New York: Exeter Publisher, 1989. 210 p.

SAIKI, R. K.; GELFAND, D. H.; STOFFEL, S.; SHARF, S. J.; HIGUCHI, R.; HORN, G. T.; MULLIS, K.; ERLICH, H. A. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. **Science**, Washington, v. 239 p. 487-491, 1988.

SCHLOTTERER, C.; TAUTZ, D. Slippage synthesis of simple sequence DNA, **Nucleic Acids Research**. London, v.20, n.2, p.211-215, 1992.

SECOND, G. Origin of the genetic diversity of cultivated rice (*Oryza spp.*): study of the polymorphism scored at 40 isozyme loci. **Japanese journal of genetics**. Tokyo, v. 57, p. 25-57, 1982.

SILVA, C.M.; KARASAWA; M.M.G.; VENCovsky, R.; VEASEY, E.A. Elevada diversidade genética em populações de *Oryza glumaepatula* Steud. (Poaceae) avaliada com microssatélites. **Biota Neotropica**. Campinas, v.7, p. 1-7, 2007.

SILVA, A.T. **Estudo da divergência genética em acessos de arroz através de marcadores morfológicos e moleculares (RAPD)**. 1999. 185 p. Tese. (Doutorado em Fitotecnia)- UFLA, Lavras, 1999.

SINGH, D. The relative importance of characters affecting genetic divergence. **The Indian Journal of Genetics and Plant Breeding**. New Delhi, v. 41, p.237-245, 1981.

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**. New Delhi, v. 17, n.2, p.163-166, 1996.

SUGANUMA, E.; CIAMPI, A. Y. **Análise genética populacional de jatobá (*Hymenaea ssp Leguminosaea*) utilizando microssatélites**. Disponível em: <http://www.redbio.org/portal/encuentros/enc_2001/posters/03/03pdf/03-007.pdf>. Acesso em: 13 nov. 2008.

TANKSLEY, S. D.; MCCOUCH, S. R. Seed banks and molecular maps: unlocking genetic potential from the wild. **Science**, Washington, v. 277, n. 5329, p. 1063-1066, 1997.

VARSHNEY, R.K.; GRANER, A.; SORRELLS, M.E. Genic microsatellite markers in plants: features and applications. **Trends in Biotechnology**, Cambridge, v.23, n.1, p. 48-55, 2005.

VILELA-MORALES, E.A. Principios geneticos da informacao para recursos geneticos vegetais. In: SIMPOSIO DE RECURSOS GENETICOS PARA AMERICA LATINA E CARIBE, 1999, Brasilia. **Anais...** Brasilia: Embrapa, 1999. CD-ROM.

WEISING, K.; NYBOM, H.; WOLFF, K.; KAHL, G. **DNA fingerprinting in plants – Principles, Methods, and Applications**. Germany: Taylor & Francis Group, 2005. 444 p.

WRIGHT, S. **Evolution and the genetics of populations: variability within and among natural populations**. Chicago: University of Chicago Press, 1978. 573 p.

XU, Y.; BEACHELL, H.; MCCOUCH, S.R. A marker-based approach to broadening the genetic base of rice in the USA. **Crop Science**, Madison, n. 44, p.1947-1959, 2004.

YU, J.; HU, S.; WANG, J.; WONG, G.K.; LI, S.; LIU, B.; DENG, Y.; DAI, L.; ZHOU, Y.; ZHANG, X.; CAO, M.; LIU, J.; SUN, J.; TANG, J.; CHEN, Y.; HUANG, X.; LIN, W.; YE, C.; TONG, W.; CONG, L.; et al. A draft sequence of the rice genome (*Oryza sativa* L. Ssp. Indica). **Science**, Washington, v. 296, p. 79-91, 2002.

YAN, J.; ZHU, J.; HE, C.; BENMOUSSA, M.; WU, P. Molecular marker-assisted dissection of genotype x environment interaction for plant type traits in rice (*Oryza sativa* L.). **Crop Science**, Madison, v. 39, p. 538-544, 1999.

YU, S.B.; XU, W.J.; VIJAYAKUMAR, C.H.M.; ALI, J.; FU, B.Y.; XU, J.L.; JIANG, Y.Z.; MARGHIRANG, R.; DOMINGO, J.; AQUINO, C.; VIRMANI, S.S.; LI, Z.K. Molecular diversity and multilocus organization of the parental lines used in the International Rice Molecular Breeding Program. **Theoretical and Applied Genetics**, Berlin, v. 108, p. 131-140, 2003.

ZANE, L.; BARGELLONI, L.; PATARNELLO, T. Strategies for microsatellite isolation: a review. **Molecular Ecology**, Vancouver, v. 11, n. 1, p. 1-16, 2002.

ANEXOS

Tabela 12 - Identificação dos acessos presentes no estudo

Número	Identificação	Nome
1	90F	Khao Kap Xang
2	54F	Ketji
3	266F	Lingkod
4	346F	Bata suduwee
5	301F	Betete
6	146F	Khao Pick
7	46F	Sawak
8	118F	Khao Dok Dou
9	292F	Poyeh
10	220F	Khao Xiou Khay
11	91F	Jao Khao
12	230F	Khao Met Nhay
13	213F	Khao Namma
14	487F	Hill Sel. X Rnbmt54v
15	123F	Mai Kai
16	129F	Khao Lay Nock
17	249F	Khao Sim
18	403F	Iac-47
19	414F	Ku-47
20	110F	Khao Nhouak Nhay
21	122F	Khao Dam
22	263F	Wiang
23	141F	Khao Eo Nhay
24	443F	Ku-94-2
25	49F	Putu
26	53F	Lokan
27	80F	Tjempo Tsino
28	166F	Pang Leu
29	107F	Khao kai

30	73F	Majang Djanggut
31	215F	Khao Toun
32	367F	Bandan-2
33	491F	Gendjah Kutu
34	120F	E-nawn
35	217F	Khao Xiou Khao
36	22F	Ku 55-1
37	167F	Khae
38	297F	Carcata
39	265F	Cabaysay
40	112F	Khao Khay
41	444F	Ku 86
42	318F	Liberian Coll. D3-167
43	115F	Ngah Chahng
44	173F	Pah Yan
45	142F	Khao Do
46	304F	Cokoba
47	111F	Khao Mack Fay Khao
48	216F	Khao Xien Py
49	136F	Gaen Fai
50	150F	Khao Hido
51	225F	Lai Dawk Doo
52	41F	Kyetek
53	85F	Khao Pio
54	33F	Djoro One
55	320F	Gbegbea
56	309F	Liberian Coll. D3-64
57	77F	Ligerito
58	287F	Tumendog
59	380F	Ku 101
60	21F	Ku 43-2
61	467F	Ku 70-1

62	119F	Khao None Nhay
63	64F	Makasar
64	276F	Bikyat
65	326F	Gbegbete
66	145F	Ja Ae
67	13F	3558
68	133F	Khao Khane
69	250F	Sew Glang Dong
70	377F	Menurun
71	459F	Ku 101
72	325F	Ngelegohun
73	144F	Khao Dohome
74	92F	Khao Chao Hom
75	134F	Khao Met Nhay
76	300F	Gette
77	369F	Melati
78	105F	Khao Kikhouei
79	221F	Khao Vanthang
80	31F	Ku 58
81	125F	Khao Kangkaynoi
82	188F	Khao Chao Met Nhay
83	114F	Khao Bangkhao
84	243F	Mum 1
85	219F	Khao Phe Py
86	86F	Khao Sumneua
87	187F	Khao Chao Hay
88	69F	Metan
89	316F	Bulahon
90	211F	Hahng
91	317F	Kpanpolohombon
92	476F	Selibon Siangan
93	284F	Kalinayan

94	61F	Lembese
95	106F	Leuang Noi
96	135F	Khaomu
97	222F	Khao Metto B
98	26F	Ku 91
99	59F	Menurun
100	288F	Simping
101	11F	Br-br-r (C6-1-25-12)
102	82F	Ba Djang Plon
103	78F	Iac-5544
104	247F	Graboon
105	488F	Os.6
106	138F	Khao Phe Do
107	83F	Ba Ke Gonh
108	190F	Khao Seng
109	299F	Finya
110	149F	Jan
111	113F	Khao Noi
112	244F	E-boot
113	184F	Ma Hing 267-2-18
114	305F	Fogbandi
115	76F	Iac-5100
116	421F	Wittie Sada (Pi-369815)
117	475F	Jeteh
118	345F	Bibilial
119	285F	Mangglutus
120	343F	Sole Yoe
121	84F	Puntas Claras
122	38F	Tangun
123	425F	Ku 10
124	34F	Ketan Tlasi
125	19F	Agbede

126	121F	Khao Nuan
127	183F	Khao Dok Pout
128	165F	Khao Nhn Py
129	163F	Khao Dock Keonoi
130	44F	Sibakas
131	404F	Iac-1131
132	452F	Ku 79-1
133	264F	Unnamed
134	310F	Liberian Coll. D3-101
135	130F	Khao Mone Khan
136	12F	Rogue (464799) 420
137	327F	Felue
138	63F	Maraja
139	101F	Daw Hen
140	60F	Majangan
141	241F	Daw Dan
142	35F	Ketan Pulosaren
143	344F	Bogutu
144	186F	Mawai
145	Testemunha	IAC 202
146	Testemunha	BRS SERT
147	Testemunha	CAIAPÓ
148	Testemunha	CURINGA
149	Testemunha	IAC 1246
150	Testemunha	IAC 201

* cultivares comerciais em negrito