



UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA CELULAR

“DIVERSIDADE GENÉTICA E MAPEAMENTO DE QTLs DO SISTEMA GÊNICO DE MACHO-ESTERILIDADE TERMO-SENSÍVEL (TGMS) DO GENOMA DE ARROZ (*Oryza sativa* L.)”

ANTÔNIO ALBERTO NEVES DE ALCOCHETE

ORIENTADOR: MÁRCIO ELIAS FERREIRA, Ph.D.

BRASÍLIA - DF, 2005

UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA CELULAR

“DIVERSIDADE GENÉTICA E MAPEAMENTO DE QTLs DO SISTEMA GÊNICO DE MACHO-ESTERILIDADE TERMOSENSÍVEL (TGMS) DO GENOMA DE ARROZ (*Oryza sativa* L.)”

ANTÓNIO ALBERTO NEVES DE ALCOCHETE

ORIENTADOR: MÁRCIO ELIAS FERREIRA, Ph.D.

Tese apresentada ao Departamento de Biologia Celular do Instituto de Biologia da Universidade de Brasília como requisito parcial para a obtenção do grau de doutor em Ciências Biológicas, Área de Concentração: Biologia Molecular.

BRASÍLIA - DF, 2005

Trabalho de tese desenvolvido no Laboratório de Genética de Plantas da Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF, e no Campo Experimental da Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO, sob orientação do Doutor Márcio Elias Ferreira.

Tese apresentada e aprovada em 24/10/2005 por:

Doutor Márcio Elias Ferreira
Embrapa Recursos Genéticos e Biotecnologia

Doutora Gláucia Salles Corotpassi Buso
Embrapa Recursos Genéticos e Biotecnologia

Paulo Hideo Nakata Rangel
Embrapa Arroz e Feijão

Orlando Peixoto de Moraes
Embrapa Arroz e Feijão

Doutor Renato de Oliveira Resende
Universidade de Brasília

Suplente:
Doutor Dário Grattapaglia, Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF.

“Conseguir a auto-suficiência na produção de arroz e a manutenção da estabilidade do preço são importantes objetivos nos países de baixa-renda, onde o arroz como fundamental fornece a base para a segurança alimentar nacional e gera emprego e renda para os pobres. A Ásia produz e consome 90% do arroz mundial. A maioria dos países produtores de arroz nesta região tem conseguido satisfazer as suas necessidades em arroz no passado utilizando a tecnologia da Revolução Verde. Mas o futuro oferece um grande desafio. Até 2010, o mundo deve produzir 70% mais arroz do que produziu em 1995 para satisfazer a procura resultante do aumento populacional e do aumento em recursos financeiros. Este aumento de produção deve ser atingido em menos área, com menos mão de obra, com menos água, e menos pesticidas, e deve ser sustentável. Aumentar o potencial de produção das variedades de arroz é um objetivo importante para a realização desse desafio. Nos anos 70, cientistas chineses demonstraram que o uso de híbridos de arroz aumentou a produção de arroz na China em 15-20%. Arroz híbrido é atualmente utilizado intensivamente na China. Mas esses híbridos não eram inicialmente adaptados às condições tropicais nem estavam disponíveis gratuitamente para outros países. Assim, o IRRI iniciou uma investigação em 1979 para explorar o potencial de adaptação desta tecnologia nos trópicos. Rapidamente concluiu-se que os híbridos em arroz ofereciam uma opção importante para aumentar a produção das variedades nos trópicos. Isto encorajou vários países produtores de arroz a desenvolverem esta tecnologia quer independente ou em colaboração com o IRRI e/ou com a China. Presentemente, 17 programas nacionais estão envolvidos no desenvolvimento de híbridos de arroz: Bangladesh, Brasil, Colômbia, Egito, Índia, Indonésia, Coreia do Norte, Coreia do Sul, Japão, Malásia, Myanmar, Paquistão, Filipinas, Sri Lanka, Tailândia, Estados Unidos da América, e Vietnã. Os programas de pesquisa de híbridos de arroz nos sistemas nacionais de pesquisa agrícola (NARS) estão em diferentes estágios de desenvolvimento. Embora a China tenha a maior rede nacional em híbridos de arroz, os programas no Brasil, Índia, Japão, e Estados Unidos da América estão bem estabelecidos. Outros países poderão levar 1-5 anos para desenvolver tais programas. Algumas companhias privadas no Brasil, Índia, Japão, e E.U.A. têm desenvolvido uma forte pesquisa básica em híbridos de arroz enquanto algumas estão sendo desenvolvidas e outras estão envolvidas somente na produção de sementes”.

Dr. G.H.L. Rothschild

Commonwealth Scientific and Industrial Research Organization, Canberra, Australia

Terceiro Simpósio Internacional sobre Híbridos de Arroz. 1997.

DEDICATÓRIA

Aos meus pais,
Henriques e Graciete,
e irmãos
pelo apoio na minha formação acadêmica.

Aos meus filhos, Leo e Gerson,
e minha esposa, Francisca Pereira,
pela compreensão e apoio na longa ausência,
em especial ao Leo por lhe ter faltado nestes
primeiros anos de sua vida.

AGRADECIMENTOS

Aos governos de Angola e Brasil, pelo intercâmbio na formação em cursos de pós-graduação e apoio financeiro, tornando possível a criação de cientistas em várias áreas de desenvolvimento.

À Universidade Agostinho Neto, através da Faculdade de Ciências, e ao Centro Nacional de Recursos Fitogenéticos, pela pronta disponibilidade e auxílio na indicação da candidatura ao curso de pós-graduação.

À Universidade de Brasília, em especial ao Departamento de Biologia Celular, pela oportunidade de realizar este curso.

A EMBRAPA Recursos Genéticos e Biotecnologia, em especial ao Laboratório de Genética de Plantas, pela disponibilidade de condições técnicas e materiais para a realização do projeto.

A EMBRAPA Arroz e Feijão pela disponibilidade do material biológico utilizado neste trabalho e na realização dos experimentos de campo.

Ao Doutor Márcio Elias Ferreira pela orientação, sugestões e apoio durante a realização deste trabalho, e, sobretudo pela oportunidade de realizar um trabalho gratificante e enriquecedor à minha formação profissional.

Ao Doutor Afonso Candeia Valois, pioneiro no intercâmbio das relações técnico-científicas entre Brasil-Angola, tornando possível o estabelecimento de uma parceria na área de recursos fitogenéticos.

Aos Doutores Orlando da Mata e Augusta Silva, Decano da Faculdade de Ciências e Chefe de Departamento de biologia da FC, pelo apoio na candidatura e durante a realização deste programa.

A Doutora Elizabeth Matos, Chefe do Centro Nacional de Recursos Fitogenéticos, por ter aberto as portas na área da genética e da conservação e gestão dos recursos fitogenéticos.

Ao Doutor Paulo H.N. Rangel pela ajuda no desenvolvimento das populações segregantes e na avaliação fenotípica da característica TGMS.

Ao Doutor Péricles C.F. Neves pela ajuda com a avaliação fenotípica da característica TGMS.

Aos Doutores Gláucia S. Buso, Orlando Peixoto de Moraes, Renato de Oliveira Rezende e Paulo H. N. Rangel, pela disponibilidade como membros da banca examinadora desta tese, pelas correções, sugestões, e discussões levantadas para melhorar a finalização deste trabalho.

Aos Doutores Gláucia Buso e Ana Ciampi, e Márcio Morethzon, do Laboratório de Genética Vegetal do Cenargen-Embrapa, pela ajuda prestada durante a realização deste trabalho.

Aos professores do Departamento de Biologia Celular da Universidade de Brasília pelos ensinamentos e receptividade.

Aos colegas do Projeto Orygens e do Laboratório de Genética de Plantas do CENARGEN pela ajuda

pronta, em especial ao Leocir Welter e Marco Pessoa Filho, pelo auxílio nos experimentos e sugestões.

ÍNDICE

1. INTRODUÇÃO.....	7	
1.1. O Arroz.....	7	
1.2. Importância Econômica do Arroz.....	7	
1.3. O Arroz no Brasil.....	9	
1.4. O genoma	11	
1.5. Germoplasma de Arroz.....	13	
1.6. Análise de Diversidade Genética e Estruturação Genética	15	
1.7. Coleções Nucleares		22
1.8. Melhoramento Genético de Arroz e Produção de Híbridos.....	23	
1.9. O Sistema TGMS em Arroz.....	26	
1.10. Efeito de Temperatura na Esterilidade do Pólen.....	30	
1.11. Linhagens TGMS em Arroz.....	36	
1.12. Marcadores Moleculares.....	38	
1.13. Reações Múltiplas (Multiplex) de Microssatélites	41	
1.14. Mapeamento Genético	44	
1.15. “Bulk Segregant Analysis” (BSA)		45

1.16. Detecção de QTLs.....	47
1.17. QTL Associados ao Controle de TGMS em Arroz.....	48
2. OBJETIVOS.....	50
3. FLUXOGRAMA.....	51
4. RESULTADOS.....	52
4.1. CAPÍTULO 1 – Estrutura Genética de Amostras de um Banco de Germoplasma de Arroz	52
4.2. CAPÍTULO 2 - Mapeamento de QTLs para a Macho-Ersterilidade Gênica Termo-Sensível (TGMS) em Arroz <i>indica</i>	87
5. DISCUSSÃO GERAL.....	105
6. CONSIDERAÇÕES FINAIS.....	117
7. BIBLIOGRAFIA	

ANEXOS

I - Dados de Genotipagem.

II – Matriz de Distância Genética

II – Dendograma

IV – Estruturação

V – Fertilidade do Pólen

VI – Dados de Mapeamento

VII – Distorção dos Marcadores Microssatélites

VIII – Genetic Structure of Rice Samples from a Bank of Germplasm

IX - Mapping of QTLs for Thermosensitive Genic Male Sterility (TGMS) in Indica Rice

LISTA DE TABELAS

Tabela 1 - Maiores produtores mundiais de arroz (SAFRA 2002/03 - Base casca, em milhões de toneladas).	7
Tabela 2 – Área e Produção Brasileira de Arroz	9
Tabela 3 - Produtividade de Arroz no Brasil (Kg/ha.)	27
Tabela 4 - diferenças entre linhagens CMS e TGMS na produção de híbridos	37
Tabela 5 - Descrição de linhagens macho-estéreis termo-sensíveis utilizadas em programas de melhoramento genético	55
Tabela 6 - Acessos de arroz do Banco de Germoplasma de Arroz (Embrapa, Brasil) utilizadas na análise.	63
Tabela 7 - Marcadores microssatélites marcados com fluorescência utilizados nos três painéis múltiplos	69

Tabela 8 - Estatística descritiva de 16 locos marcadores microssatélites genotipados em painéis múltiplos para uma coleção de 298 acessos de arroz	
Tabela 9 - Banco de dados de frequência alélica para acessos de arroz utilizando 16 marcadores microssatélites	72
	-
	73
Tabela 10 - Estimativa dos parâmetros F de Wright para a coleção de germoplasma. Os agrupamentos de acessos foram definidos de acordo com análise Neighbour-Joining dos dados de distância genética baseadas no polimorfismo de microssatélites; Intervalos de confiança de 95% foram estimados baseados na análise de bootstrap (1000x).	80

LISTA DE FIGURAS

Figura 1 – Electroferograma mostrando o genótipo do acesso controle ARR1 obtido com 10 marcadores microssatélites analisados com três painéis múltiplos. O eixo horizontal representa as estimativas do tamanho em pares de bases e o eixo vertical indica a intensidade de fluorescência medida pelo seqüenciador de DNA ABI Prism	9
Figura 2 – Amplitude alélica dos 16 marcadores microssatélites dos três painéis múltiplos (cinza escuro e branco – pentaplex e hexaplex, Beló, 2001; cinza claro; – pentaplex; Pessoa Filho, 2004).	11
Figura 3 – Histograma de barras da frequência alélica dos 16 marcadores microssatélites	11
Figura 4 – Análise de estrutura genética de 298 acessos de arroz <i>O. sativa</i> L. baseada no polimorfismo de DNA de 16 locos microssatélites. Estimativas de distância genética, baseadas no coeficiente de alelos partilhados e análise de agrupamento, foram calculadas utilizando o método Neighbour-Joining (Saitou & Nei, 1987). A análise de estruturação genética foi baseada no método de agrupamento Bayesiano (barra colorida) e efetuada de acordo com Pritchard et al. (2000a).	28
Figura 5 – Gráfico de plotagem da riqueza alélica versus tamanho da coleção nuclear composta de acessos de arroz mais diversos	38
Figura 6 - Análise fatorial clássica dos 298 acessos de arroz <i>O. sativa</i> L. baseadas no polimorfismo de DNA em 16 locos polimórficos. As estimativas de distância genética foram baseadas no coeficiente de “shared allele distance” e análise de agrupamento foram calculadas utilizando o método Neighbour-Joining (Saitou & Nei, 1987).	56

Figura 7 – Análise de estrutura genética dos 48 acessos de arroz *O. sativa* L. baseada no polimorfismo de DNA de 16 locos microssatélites, extraídos como coleção nuclear. As estimativas de distância genética foram baseadas no coeficiente “shared allele distance” e análise de agrupamento e calculadas utilizando o método Neighbour-Joining (Saitou & Nei, 1987). A análise de estruturação genética foi baseada no método de agrupamento Bayesiano (barra colorida) e de acordo com Pritchard et al. (2000a). 64

Figura 8 - Distribuição da fertilidade do pólen a população segregante F2 do cruzamento entre as linhagens parentais IR68945-4-33-4-14 (TGMS) x IR50 (fértil) 70

Figura 9 - Genotipagem de marcadores microssatélites de uma população F2 derivada do cruzamento entre as linhagens parentais *O. sativa* IR68945-4-33-4-14-48 (P1) x *O. sativa* IR50 (P2). (a) Segregação dos alelos do marcador OG11 em eletroforese de gel de poliacrilamida 4% corada com nitrato de prata; (b) Segregação dos alelos do marcador OS9 em eletroforese de gel de agarose (3,5%) corado com brometo de etídio; (c) Electroferogramas do marcador RM7012 mostrando os genótipos da linhagem IR50 (fértil), IR68945-4-33-4-14-48 (TGMS) e sua F1. Os eixos verticais indicam intensidade de fluorescência medida por um Seqüenciador de DNA ABI Prism 377. O eixo horizontal representa a estimativas dos tamanhos dos produtos em pares de base. M = 1 Kb Ladder 73

Figura 10 – Mapeamento de QTL TGMS QTL baseado na análise de mapeamento por Intervalo Simples dos cromossomos de arroz: (a) QTL TGMS localizado no cromossomo 1 (LOD = 3,90); (b) QTL TGMS localizado no cromossomo 7 (LOD = 3,19); (c) QTL TGMS localizado no cromossomo 12 (LOD = 3,92). 81

Resumo:

O arroz (*O. sativa* L.) é a principal cultura alimentar e um dos alimentos diários de mais de metade da população mundial. Este cereal fornece cerca de 20% das calorias consumidas mundialmente (Khush, 2001). Com o rápido crescimento populacional, há a necessidade de um incremento dos atuais níveis de produtividade do arroz para satisfazer os cerca de 8,9 bilhões de pessoas no planeta estimadas para 2010 (Tanksley & McCouch, 1997). O aumento da produtividade do arroz é dificultado por fatores como a estreita base genética dos programas de melhoramento. Este incremento pode ser atingido utilizando variedades híbridas desenvolvidas em programas do melhoramento genético da cultura.

O cruzamento de linhagens puras pertencentes a grupos heteróticos distintos para a obtenção de variedade apresentando vigor híbrido, isto é, uma produtividade superior à produtividade das linhagens genitoras, tem sido uma estratégia de grande impacto na agricultura desde o início do século XX. O fenômeno de vigor híbrido ou heterose tem sido eficazmente explorado em espécies alógamas (ex., milho) e, classicamente, esta estratégia tem sido associada ao melhoramento genético de espécies de polinização cruzada. Nos últimos anos, no entanto, o emprego de heterose tem tido grande apelo comercial também em algumas espécies autógamas, como o arroz. O desenvolvimento de variedades híbridas de arroz representa uma revolução no incremento de produtividade da cultura. A combinação de linhagens elite de programas de melhoramento de arroz híbrido pode elevar a produtividade de linhagens puras de arroz irrigado hoje estagnada em patamares médios de 5.000 kg/ha no Brasil, para médias significativamente superiores.

A análise de relações genéticas em plantas agrícolas é um componente importante dos programas de melhoramento, visto que ela fornece informação acerca da diversidade genética, e é uma plataforma para a amostragem estratificada das populações de melhoramento. O estudo adequado dos níveis e padrões de diversidade genética pode ser de grande valor no melhoramento de plantas para diferentes aplicações, tais como análise de variabilidade genética em cultivares, identificação de combinações parentais diversas para a criação de progênies segregantes com o máximo de diversidade genética para posterior seleção, e introgressão de genes de germoplasma diverso em linhagens elites e populações de melhoramento. Em nível de coleção de germoplasma, a análise de diversidade genética pode facilitar a real classificação dos

acessos, bem como a identificação de subgrupos de coleções nucleares com utilidade potencial de acordo com objetivos específicos de melhoramento. O desenvolvimento de uma coleção nuclear pode promover a avaliação e uso dos recursos genéticos, e uma gestão mais eficiente dos bancos de germoplasma, bem como o estabelecimento de novas estratégias de coleta.

Por ser uma espécie autógama, a produção de arroz híbrido depende essencialmente da existência de um mecanismo de esterilização de pólen sem que haja dano ao sistema reprodutor feminino. A descoberta do sistema de macho-esterilidade genética citoplasmática (CMS - "cytoplasmic-genetic male sterility") permitiu o desenvolvimento de uma estratégia de produção de híbridos de arroz pelo sistema de três linhagens, ou seja, linhagem macho-estéril, linhagem mantenedora e linhagem restauradora. Mais recentemente, a identificação do sistema gênico de macho-esterilidade termo-sensível (TGMS - "thermosensitive genic male sterility") vem permitindo a obtenção de linhagens macho-estéreis sob certos regimes de temperatura, abrindo novos horizontes para a exploração de arroz híbrido. O sistema TGMS tem simplificado o procedimento da produção de sementes híbridas em arroz porque as linhagens TGMS podem ser usadas como macho-estéreis e como mantenedoras, e têm grande valor devido ao amplo espectro de restauradores e seu grande potencial na produção de sementes híbridas (Lang et al., 1999; Jia et al., 2001). O uso do sistema TGMS é potencialmente simples, econômico, eficiente e elimina as limitações associadas com o sistema CMS. A alteração da fertilidade é condicionada pelas oscilações de temperatura e pela fase de desenvolvimento morfológico da planta (IRRI 1997) devido à ação de vários genes dispersos pelo genoma.

O objetivo deste estudo foi estudar a diversidade genética de acessos de arroz do banco de germoplasma da Embrapa e localizar regiões do genoma responsáveis pelo controle genético de macho-esterilidade termo-sensível. A análise de diversidade genética baseou-se em uma coleção de 298 acessos de arroz, *O. sativa* L., utilizando painéis múltiplos de 16 microssatélites marcadores com fluorescências e genotipagem automática em seqüenciador automático de DNA. Os dados de polimorfismo de DNA foram submetidos à análise de distância genética entre os acessos em comparações par-a-par e, em seguida, avaliados por métodos de agrupamento por distância genética contrastados com estimativas de estruturação populacional. Uma coleção nuclear foi desenvolvida, baseada na maximização da distância genética entre os acessos genotipados e avaliada quanto à riqueza genética em relação à diversidade genética do grupo total de acessos.

Os resultados da análise de amostras do banco de germoplasma mostram ampla diversidade genética entre os acessos analisados. Um banco de dados de frequências alélicas foi desenvolvido com base nos dados de genotipagem utilizando locos microssatélites. As relações genéticas entre acessos possibilitaram observar o agrupamento de acessos com variedades controle pertencentes às subespécies *indica* ou *japonica*. Foi possível ainda seleção de duas linhagens de arroz (IRTGMS-68945 e IR50) com suficiente diferenciação genética para a geração de uma população segregante F2 visando o estudo de mapeamento de QTLs para o controle genético de TGMS.

Com base na análise de distância genética pelo método de “shared allele distance” (Bowcock et al., 1994) e de agrupamento baseado em “Neighbour-Joining” (Saitou & Nei, 1987) os acessos foram distribuídos em subgrupos. Os agrupamentos observados serviram de modelo para detectar estruturação genética na população. A análise de estruturação empregou as estatísticas F de Wright (Wright, 1978; Weir and Cockerham, 1984) e estimativas de índice de fixação com base na Cadeia de Markov e Monte Carlo (MCMC) (Pritchard et al., 2000). Os resultados indicaram uma correlação não significativa entre o modelo de agrupamento baseado em distância genética e diferenciação da coleção com base na estratégia MCMC. As estimativas das estatísticas F de Wright revelaram um alto valor de coeficiente de endogamia (F_{IS}) e de índice de fixação geral (F_{IT}), porém um nível moderado de índice de fixação (F_{ST}). Uma coleção nuclear definida a partir de dados de distância genética e de riqueza alélica indicou que uma subamostra da coleção composta por apenas 48 acessos (16,66%) mantinha 81,31% da riqueza alélica detectada. Níveis moderados de diferenciação genética foram observados na coleção nuclear.

Uma população F2 derivada do cruzamento entre linhagens *indica* férteis e TGMS foi usada para construir um mapa genético de arroz baseado em marcadores microssatélites. O fenótipo TGMS analisado apresentou uma variação contínua na população segregante. Um baixo nível de distorção da segregação foi detectado na população segregante (14,65%) e a causa da distorção demonstrada como sendo a seleção zigótica. Não houve evidência de relação de causa-efeito entre seleção zigótica e o controle de TGMS neste cruzamento. Um mapa de ligação compreendendo 1.213,3 cM foi construído baseado nos dados de segregação da população F2. Noventa e cinco marcadores microssatélites de um total de 116 marcadores testados foram

reunidos em 11 grupos de ligação com uma média de 12,77 cM entre os locos marcadores adjacentes. Os dados fenotípicos e genotípicos permitiram a identificação de três novos QTLs para a macho-esterilidade gênica termossensível em arroz *indica*. Dois dos QTLs foram mapeados em cromossomos que não tinham ainda sido associados ao controle genético da característica TGMS (cromossomos 1 e 12). O terceiro QTL foi mapeado no cromossomo 7, onde um loco TGMS (*tms2*) foi recentemente mapeado. Testes alélicos deverão ser realizados para verificar se as regiões mapeadas são as mesmas.

ABSTRACT

Rice (*O. sativa* L.) is a staple food for more than half of the world population, providing about 20% of the calories found in the human diet (Khush, 2001). The rapid growth of the human population demands a substantial increase on rice yield to feed the estimated 8,9 billions of people in the planet by 2010 (Tanksley & McCouch, 1997). The increase on rice yield is constrained by factors such as the narrow genetic basis of breeding programs. This increase can be reached using hybrid varieties developed by the breeding programs of rice.

The cross between inbred lines from distinct heterotics groups to obtain a new variety expressing hybrid vigor, i.e., showing a higher yield than observed in the parental lines is one of the most important methods in use by the seed industry since the beginning of the 20th century. The phenomenon of hybrid vigor or heterosis has been efficiently exploited in allogamous species (ex: maize) and, classically, this strategy has been associated with genetic improvement of outcrossing species. In the last years, however, the use of heterosis has had a great commercial appeal for autogamous species such as rice. The development of rice hybrid varieties represents a revolution on yield improvement of rice. The combination of elite lines to develop new hybrid varieties can potentially increase the yield of irrigated rice, stabilized on average levels of 5,000 Kg/ha in Brazil, to significantly higher averages.

Accurate assessment of levels and patterns of genetic diversity can be invaluable in crop breeding for diverse applications including the analysis of genetic variability in cultivars, the identification of diverse parental combinations to create segregating progenies with maximum genetic variability for further selection and introgression of desirable genes from diverse germplasm into the available genetic base. An understanding of genetic relationships among inbred lines or

pure lines can be particularly useful in planning crosses, in assigning lines to specific heterotic groups, in defining parental lines to be used in the development of new recombining populations or to enhance the use of germplasm in breeding programs. The study of genetic diversity of germplasm collections, coupled with genetic differentiation estimates, can facilitate the reliable classification of the accessions, the establishment of their pairwise and group genetic relationships, the selection of representative samples which capture the genetic diversity of the collection and the detection of patterns of differentiation in the whole collection as well as in samples drawn from it. DNA polymorphism can be very useful to define a group of accessions (core collection) that could genetically represent the usually large germplasm collection from which it is derived. A core collection represents a sample of the whole collection composed of a limited number of accessions, which together should represent the maximum possible level of the genetic diversity existing in the whole collection.

As autogamous species, the rice hybrid production depends essentially on pollen sterility mechanism without affecting the female sexual system. The discovery of cytoplasmic-genetic male sterility (CMS) allowed the development of rice hybrid strategy by means of tree-lines system, i.e., male-sterile line, maintainer line and restorer line. The identification of a pollen sterility system dependent on temperature, known as the Thermosensitive Genic Male Sterility (TGMS), has contributed to obtain male-sterile lines under specific ranges of temperature, opening new perspectives for hybrid rice exploitation. The TGMS system has the potential to simplify the procedures required by the hybrid seed production in rice since lines TGMS can be used as male sterile and maintainers. The use of TGMS system is potentially simple, cost effective, and efficient and eliminates most limitations associated with the CMS system. Pollen sterility is conditioned by variations in temperature and by the plant morphologic stage due to genetic effects of several genes dispersed in the rice genome.

The objective of this study was to evaluate the genetic diversity of rice accessions included in EMBRAPA's Germplasm Gene Bank and to locate genomic regions associated with the genetic control of thermo sensitive genic male sterility. Analysis of genetic diversity was based on a collection of 298 accessions of rice, *O. sativa* L., using multiplex panels of 16 microsatellite markers labeled with fluorochromes and genotyped on DNA sequencer. The DNA polymorphism data was submitted to analysis of genetic distance in pairwise comparisons and, then, evaluated by clustering methods of genetic distance and contrasted with estimates of population structure. A core collection

was developed based on levels of allele richness measured in groups of accessions defined according to maximum genetic distance.

The results show high levels of genetic diversity between the evaluated accessions. A database of allelic frequencies was constructed based on genotypic data using microsatellite markers. Control accessions of the *indica* or *japonica* subspecies were used to indicate the background composition of the remaining accessions upon cluster analysis. Two rice lines (IRTGMS68945 and IR50) with enough genetic differentiation to develop an F2 segregating population to study the genetic control of the TGMS trait were selected based on this approach.

A Neighbour-Joining based dendrogram was used as model to define clusters and infer possible genetic structuring of the collection. The analysis of the genetic relationships of these accessions suggested no significant correlation between clustering based on distance data and subpopulation differentiation based on MCMC approach. The estimates of Wright's F-statistics revealed a high value of inbreeding coefficient (FIS) and a relatively high overall fixation index (FIT) but only moderate levels genetic differentiation (FST) of subpopulations defined according to the genetic distance clustering model. A core collection based on estimated pairwise genetic distances and allele richness was successfully obtained, characterized by a subset of rice accessions representing only 16.66% of the whole collection while still maintaining 81.31% of the allele richness detected. Moderate levels of genetic differentiation were observed also in the core collection. It is possible that the use of germplasm conserved in rice germplasm collections could be enhanced if molecular characterization and population genetics approaches could be more intensively applied to better define the sample of accessions which would be more appropriately suited to different genetic and breeding purposes.

An F2 population derived from the cross between fertile and TGMS *indica* lines was used to construct a microsatellite-based genetic map of rice. The TGMS phenotype showed a continuous variation in the segregant population. A low level of segregation distortion was detected in the F2 (14,65%) and the cause of distortion was found to be zygotic selection. There was no evidence suggesting a cause-effect relationship between zygotic selection and the control of TGMS in this cross. A linkage map comprising 1,213.3 cM was constructed based on the segregation data of the F2 population. Ninety-five out of 116 microsatellite markers tested were assembled into 11 linkage groups with an average of 12.77 cM between adjacent marker loci. The phenotypic and genotypic

data allowed for the identification of three new QTLs for thermosensitive genic male sterility in indica rice. Two of the QTLs mapped on chromosomes that so far have not been associated with the genetic control of the TGMS trait (chromosomes 1 and 12). The third QTL mapped on chromosome 7, where a TGMS locus (*tms2*) has recently been mapped. Allelic tests will have to be developed in order to clarify if the two regions are the same or not.

1. INTRODUÇÃO

1.1. O Arroz

O gênero *Oryza* é uma angiosperma monocotiledônea da família Gramineae (Poaceae), subfamília Oryzoideae (Bambusoideae), e tribo Oryzae. Este grupo possui somente duas espécies de arroz cultivado, *Oryza sativa* L. ($2n=24$; AA) e *O. glaberrima* Steud. ($2n=24$; AA), e aproximadamente 20 espécies silvestres distribuídas nos trópicos da África, Ásia, Américas e Austrália (Chang, 1976; Vaughan, 1994).

A região de origem de *O. sativa* L. parece incluir Nordeste da Índia, Bangladesh, Myanmar, Tailândia, Camboja, Vietnã e Laos, países que apresentam grande diversidade genética de ecótipos primitivos e parentes silvestres da espécie. Já *O. glaberrima* Steud é originária do Oeste Africano. Os ancestrais destas duas espécies cultivadas são respectivamente *O. nivara* e *O. rufipogon*, e *O. barthii* e *O. breviligulata* (Morishima, 1994; Vaughan, 1994; OECD, 1999).

A domesticação de *O. sativa* L. é considerada como tendo ocorrido entre 15.000 a 10.000 A.C., e esta espécie tem sido cultivada na China e na Índia por mais de 8.000 anos. Existem duas subespécies de *O. sativa* L., que são *indica* e *japonica*, como resultado às variações ambientais e seleção entre diferentes áreas (Khush, 1977).

1.2. Importância Econômica do Arroz

O arroz (*O. sativa* L.) é a principal cultura alimentar e um dos alimentos diários de mais de metade da população mundial. Este cereal fornece cerca de 20% das calorias consumidas mundialmente (Khush, 2001). Com o rápido crescimento populacional há a necessidade de um incremento dos atuais níveis de produtividade do arroz para satisfazer os cerca de 8,9 bilhões de

peças no planeta estimadas para 2010 (Tanksley & McCouch, 1997). O aumento da produtividade do arroz é dificultado por fatores como a estreita base genética dos programas de melhoramento. O arroz, como grão, é exclusivamente consumido pelo ser humano. Cerca de 90% da produção mundial de arroz é consumido na Ásia, onde esta cultura constitui dieta base e a principal fonte de carboidratos, proteínas, lipídeos e minerais. A matéria verde é também uma importante ração animal em muitos países (Khush, 1977; Rouhi, 1997; OECD, 1999).

Além do consumo pelo homem e da utilização como ração animal, o grão e a planta do arroz podem ser transformados numa grande variedade de produtos. A casca do arroz é utilizada, na Tailândia, como combustível para os geradores vaporizados. O farelo do arroz é extremamente nutritivo e uma fonte natural de vitamina E; 40% do farelo são usados na extração de óleo e 60% é utilizado na alimentação animal. O óleo extraído do farelo do arroz é um óleo comestível de alta qualidade que pode diminuir o colesterol no sangue. O arroz quebrado, cerca de 70%, é utilizado na produção de ingredientes ou na produção de “noodles” instantâneos ou aperitivos. A farinha de arroz, produzida a partir do arroz quebrado, é utilizada na produção de pasta de arroz, aperitivos e cereais. O amido, principal nutriente do arroz, é utilizado como um componente que torna os molhos espessos, nas sobremesas, e xaropes. A palha é utilizada, em países asiáticos, como meio de crescimento de cogumelos, na alimentação animal, na produção de papel, em processos de produção de alimentos bem como na confecção de sandálias e de chapéus. O arroz é utilizado na produção de várias bebidas alcoólicas, incluindo o saké ou saki (vinho de arroz) e cerveja. O papel de arroz, produzido a partir de celulose extraída da planta do arroz é utilizado no fabrico de cigarros ou de balas, e a cola de arroz, produzida por dissolução de arroz em água fervendo, são outros produtos do arroz. (www. fao.org). O arroz integral é mais nutritivo do que o arroz branco (polido), mas não tão popular quanto o arroz branco que é mais saboroso e fácil de cozinhar.

A maior produção de arroz está centralizada na Ásia, incluindo o subcontinente indiano, que é também a principal região consumidora. Por exemplo, China e Índia permanecem no topo dos produtores mundiais (Tabela 1), mas são também os países com as maiores populações mundiais.

Tabela 1 – Maiores produtores mundiais de arroz (SAFRA 2002/03 - Base casca, em milhões de toneladas).

PAÍS	PRODUÇÃO	PARTICIPAÇÃO %
CHINA	177,5	30,50
INDIA	120,0	20,62
INDONÉSIA	51,6	8,86
BANGLADESH	38,5	6,61
VIETNÃ	34,1	5,86
TAILÂNDIA	25,8	4,43
MYANMAR	21,9	3,76
FILIPINAS	13,2	2,27
AFRICA SUBSAARIANA	12,0	2,06
JAPÃO	11,1	1,91
BRASIL	10,7	1,84
SUBTOTAL	516,4	88,7
OUTROS	65,6	11,28
TOTAL	582,0	100,0

Fonte: FAO, 2003.

A produção mundial de arroz, em 2003, foi de 582,0 milhões de toneladas (Mt), tendo a China com o principal produtor, com cerca de 180 Mt, enquanto o Brasil é o décimo - primeiro produtor, com cerca de 10 Mt de arroz (FAO, 2003).

1.3. O Arroz no Brasil

Foram, provavelmente, os portugueses quem introduziram esse cereal na África Ocidental, e os espanhóis, os responsáveis pela sua disseminação nas Américas. Entretanto, o arroz silvestre já ocorria no Brasil antes da chegada dos portugueses e é, ainda hoje, encontrado no Pantanal

Matogrossense, na Amazônia e nas margens dos igarapés (Buso, 1998a). O arroz era o “milho d’água” (*abati-uaupé*) que os tupis já colhiam nos alagados próximos ao litoral. O arroz silvestre brasileiro pertence às espécies *O. glumaepatula* (2n=25), *O. grandiglumis* (2n=48), *O. latifolia* (2n=48) e *O. alta* (2n=48), (Black, 1950; Rangel et al., 1996; Marin, 1998; Buso, 1998a; Pott & Pott, 2000; Pereira, 2002).

A prática da orizicultura no Brasil, de forma organizada e racional, só aconteceu em meados do século XVIII (Pereira, 2002). O país é o décimo - primeiro maior produtor de arroz no mundo. Na América Latina e no Caribe, o Brasil é o maior produtor de arroz, embora seja também o maior importador (FAO, 2003). O país ainda importa anualmente cerca de 10% do arroz que consome.

O arroz é cultivado em praticamente todos os estados do Brasil, destacando-se inicialmente o cultivo de variedades *japonica* e, mais recentemente, variedades *indica*. Existem dois principais ecossistemas de cultura do arroz no Brasil, várzeas e terras altas (sequeiro). O ecossistema de várzeas representou cerca de 40% da área total sob a cultura, de 1,5 milhões de ha., contribuindo com 60% da produção de 10 milhões de toneladas, na safra 98/99. Nesse ecossistema predomina o sistema de cultivo com irrigação controlada, que ocupa cerca de 1 milhão de hectares na região subtropical (Rio Grande do Sul e Santa Catarina), onde a cultura é manejada sob alto nível tecnológico e apresenta rendimento médio ao redor de 5,5 t/ha. O ecossistema de terras altas desempenhou um papel de grande relevância na produção de arroz sob o sistema de cultivo de sequeiro, nas décadas de 60 a 80, em que a cultura chegou a ocupar 4,5 milhões de hectares. Devido à sua rusticidade e adaptação a solos ácidos, foi uma alternativa altamente satisfatória para o desbravamento dos cerrados. Mas desde a década de 80 a área sob a cultura no sistema de sequeiro vem decrescendo, devido a fatores ecológicos e econômicos, atingindo menos de 2 milhões de hectares na safra 98/99. Em 2000, o consumo de arroz *per capita* no Brasil foi de 58 kg de arroz integral.

O levantamento da safra 2002/03, realizado em abril, pela Companhia Nacional de Abastecimento do Ministério da Agricultura, Pecuária e Abastecimento, indicou uma redução da área cultivada com arroz de 1,5%, passando de 3.219,6 mil hectares para 3.172,1 mil hectares. A produção foi de 10,6 milhões de toneladas, 4,5% inferior à safra 2001/02, devido a problemas climáticos na Região Sul, especialmente no RS (<http://www.irga.rs.gov.br>).

A Região Sul destaca-se com 36,8% da área cultivada, 58,6% da produção brasileira de arroz e produtividade 59% superior à média nacional, considerando a estimativa da atual safra (Tabelas 2 e 3). A Região Nordeste detém a segunda maior área cultivada, porém a mais baixa produtividade entre as regiões. Em termos de produção, a região Centro Oeste detém a segunda melhor produção. As Regiões Norte e Nordeste detêm praticamente a mesma participação na produção nacional, respectivamente com 11,2 e 11,5%. Com uma percentagem de apenas 4,3% da área e 3,1% da produção brasileira encontra-se a Região Sudeste que, por outro lado, é a maior consumidora de arroz (www.irga.rs.gov.br).

Tabela 2 – Área e Produção Brasileira de Arroz

REGIÃO	ÁREA - 1.000 ha.			PRODUÇÃO- 1.000 t		
	2001/2002	2002/2003	Var. %	2001/2002	2002/2003	Var. %
Centro Oeste	605,2	593,1	-2,0	1.650,9	1.652,9	0,2
Sul	1.203,8	1.169,1	-2,9	6576,6	6.221,7	-5,4
Nordeste	735,3	733,3	-0,3	966,4	1.223,2	26,6
Norte	529,5	541,1	2,2	1.090,0	1.193,0	9,4
Sudeste	145,8	135,5	-7,1	343,0	325,3	-5,2
TOTAL	3.219,6	3.172,1	-1,5	10.626,1	10.616,1	-0,1

Fonte: CONAB, IRGA, IBGE (2002-2003).

Tabela 3 – Produtividade de Arroz no Brasil (Kg/ha.)

REGIÃO	2001/2002	2002/2003	Var. %
Centro Oeste	5.463	5.322	-2,6
Sul	2.727	2.787	2,2
Nordeste	1.314	1.668	26,9
Norte	2.059	2.205	7,1
Sudeste	2.353	2.401	2,0
Média	3.300	3.347	1,4

Fonte: CONAB, IRGA, IBGE.

1.4. Genoma

Atualmente, o estudo da genômica do arroz conta com vários recursos tais como a seqüência completa do genoma estrutural, mapas genéticos densos, mapas físicos completos,

mapas comparativos extensos ancorados em outras gramíneas de importância econômica, mais de 4.000 marcadores de DNA de uso integrado entre diferentes laboratórios, várias bibliotecas de cromossomos artificiais de bactéria (BAC), bibliotecas de cromossomos artificiais derivados de P-1 (PAC) e bibliotecas de cromossomos artificiais de leveduras (YAC), populações de mutantes insercionais de TDNA e de transposons, linhagens mutantes derivadas de mutação química e por irradiação, bancos extensos de germoplasma depositados em diferentes países. Esses dados revelam características importantes da estrutura genética e física do genoma do arroz e a evolução dos cromossomos dos cereais (Kurata et al., 2002).

O genoma de arroz (*Oryza sativa*; genoma AA) é composto de 12 cromossomos ($2n = 24$) e tem um tamanho de 430 Mb (mega pares de bases = 1.000.000 pares de bases) correspondendo a cerca de 1.500 cM (centiMorgans, unidade de medida da frequência de recombinação na meiose genética) (Kurata et al., 1994b; Ferreira & Grattapaglia, 1998; Harushima et al., 1998). Assim, em arroz, cada cM corresponde à média de 290 kb. Esta relação varia entre as espécies e tamanhos de genomas, como por exemplo, o milho (3.000 Mb); cevada (3.500 Mb); trigo (uma espécie hexaplóide com genomas A, B e D) (~16.000 Mb) (Keller & Feuillet, 2000).

O genoma da subespécie *japonica* de arroz foi seqüenciado e agrupado pelo sequenciamento utilizando a estratégia “shotgun” (Goff et al., 2002; RGISC, 2005). As seqüências agrupadas cobriram 93% das 420 megabases do genoma. A predição de genes nas seqüências agrupadas sugere que o genoma contém 32.000 a 50.000 genes. A sintonia e homologia gênica entre arroz e os genomas de outros cereais são grandes. Noventa e oito por cento de proteínas homólogas de milho, trigo, e cevada foram também encontradas em arroz. Por outro lado, a sintonia com *Arabidopsis* é limitada, embora seja possível identificar e atribuir a funcionalidade de genes candidatos de arroz ortólogos aos genes de *Arabidopsis* (Goff et al., 2002; RGISC, 2005).

Por outro lado, uma seqüência “rascunho” do genoma de arroz mais cultivado na China, *Oryza sativa* L. ssp. *indica*, foi produzido também por sequenciamento “shotgun” (Yu et al., 2002). O genoma foi estimado em 466 Mb, com um número de genes entre 46,022 a 55,615. A cobertura funcional em seqüências agrupadas foi de 92,0%. Cerca de 40% do genoma é aparentemente constituído por repetições oligoméricas de 20-nucleotídeos, e a maioria dos transposons localizada nas regiões intergênicas. Embora 80,6% dos genes preditos de *Arabidopsis thaliana* tenham homologia em arroz, esta elevada proporção de homologia representa apenas 49,4% dos genes de

arroz (Yu et al., 2002).

O arroz tem inúmeras características que a tornam uma espécie modelo entre as gramíneas. O genoma de pequeno tamanho, compatível com as tecnologias e esforços de sequenciamento, tornou possível determinar a seqüência completa de cada um dos cromossomos nucleares de arroz. Ferramentas e recursos, tais como programas de predição de genes, ESTs e cDNAs, e identificação de recursos moleculares para mapeamento, têm sido desenvolvidos e melhorados para maximizar a interpretação das seqüências genômicas (Yuan et al., 2001). Por ser uma das culturas alimentares mais importantes no mundo e pelo tamanho pequeno do seu genoma, o arroz serve como modelo de pesquisa genômica. A sua relação de sintenia com os genomas de outras gramíneas torna a informação genômica do arroz de grande utilidade no melhoramento de outros cereais como o milho, trigo, cevada, sorgo, milheto, etc (Barry, 2001).

1.5. Germoplasma de Arroz

A variabilidade genética em arroz conservada em bancos de germoplasma (*O. sativa* L.) é apenas uma pequena fração da variabilidade encontrada na natureza (Tanksley & McCouch, 1997). Ela consiste na diversidade de parentes silvestres do arroz, populações locais (landraces) ou cultivares primitivas, cultivares absoletas, linhagens experimentais, linhagens produzidas por mutações e outros produtos dos programas de melhoramento, cultivares modernas (Nass, 2001), e as novas fontes de diversidade, como as plantas transgênicas (Bellon et al., 1996).

A diversidade genética do gênero *Oryza* (particularmente espécies do genoma AA) tem sido estudada utilizando descritores morfológicos, análise citogenética, hibridização interespecífica, e marcadores bioquímicos e moleculares. Existem duas espécies cultivadas e vinte e uma espécies silvestres no gênero *Oryza* agrupadas nos seguintes “pool” gênicos: o “pool” gênico primário é constituído pelas espécies silvestres *O. rufipogon* Griff., *O. nivara* Sharma et Shastry, *O. glumaepatula* Steud., *O. meridionalis* Ng., *O. breviligulata* Chev. Et Rhoer., *O. longistaminata* Chev. et Rhoer., e pelas espécies cultivadas *O. sativa* L. e *O. glaberrima*. Estas espécies possuem o genoma AA e a transferência gênica entre elas pode ser feita através da hibridização convencional e seleção; o “pool” gênico secundário compreende as espécies que possuem os genomas BB, CC, BBCC, CCDD, EE e FF, havendo homologia limitada com o genoma AA e, por isso, a transferência

gênica é bastante dificultada; o “pool” gênico terciário engloba espécies que raramente produzem híbridos viáveis com o arroz (Khush, 1977; Buso, 1998a; Buso et al., 1998b; Brondani, 2000; Buso et al., 2001).

Os melhoristas de plantas reconhecem a importância das espécies silvestres como fonte de genes de resistência a pragas, doenças, e aos estresses ambientais. A introdução destes genes em variedades economicamente importantes evita perdas de produtividade. Mas, para atingir este objetivo, é necessária a existência de um amplo acervo de germoplasma de arroz disponível aos melhoristas (Ko et al., 1994). A riqueza dos recursos genéticos de arroz é a base para o progresso no melhoramento genético da cultura face ao racionamento de recursos biológicos e físicos, sendo que o germoplasma silvestre das espécies utilizadas na agricultura constitui um dos maiores tesouros biológicos do homem (Chang, 1984).

O arroz possui uma das maiores coleções de germoplasma no mundo. Aproximadamente 84.000 acessos estão conservados no Centro Internacional de Germoplasma de Arroz (Internacional Rice Germplasm Centre - IRGC) nas Filipinas (Jackson & Huggan, 1993) e o total conservado em diferentes países pode atingir 230.000 acessos. A disponibilidade internacional de germoplasma de arroz tem constituído a base para a maioria dos resultados obtidos no melhoramento genético do arroz.

O germoplasma, que representa a diversidade genética das plantas cultivadas, tem sido ativamente conservado nos bancos genéticos existentes em várias partes do mundo. A conservação e utilização dos recursos genéticos de plantas são fatores essenciais para a manutenção e desenvolvimento da produção agrícola. A necessidade de se conservar essa diversidade tem sua base na demanda agrícola, e, portanto, no melhoramento vegetal. Para uma melhor utilização do germoplasma conservado é necessário que se caracterize a diversidade nele existente, quer morfológicamente, via descrição de caracteres morfológicos, quer, em nível do genoma, utilizando técnicas moleculares.

A genética molecular tem tido um papel importante na conservação e caracterização da diversidade genética, com o fim de melhorar a coleta, manutenção e utilização de germoplasma. Esta função é possível pelos estudos de estimativas de parâmetros genéticos de populações e da análise da relação genética entre as amostras, acessos ou espécies. O emprego de marcadores

moleculares em genética de populações representa um grande avanço pela possibilidade de se identificar genótipos dos indivíduos analisados, permitindo a estimativa de parâmetros genéticos para o estudo de populações.

Com o avanço tecnológico no campo da biologia molecular, nos últimos 25 anos, a avaliação e utilização da diversidade genética sofreram uma grande mudança. A análise molecular permite uma visão do polimorfismo genético que muitas vezes difere do polimorfismo derivado da análise fenotípica. A tecnologia molecular tem permitido a identificação de similaridade no DNA de gêneros separados há milhares de anos por evolução e cujo fenótipo mascara sua herança genética comum (Jauhar, 1996; Phillips & Vasil, 1994; Sobral, 1996). O poder da genética molecular traz consigo a necessidade de se repensar na classificação dos acessos de germoplasma existente em bancos de germoplasma.

1.6. Análise de Diversidade Genética e Estruturação Genética

Diversidade genética - A análise das relações genéticas nas espécies de plantas alimentares é um componente importante dos programas de melhoramento porque fornece informação sobre a diversidade genética e, conseqüentemente, uma base de amostragem de variedades, linhagens, cultivares, acessos ou populações envolvidas nesse processo (Mohammadi & Prasanna, 2003). A avaliação correta dos níveis e padrões da diversidade genética pode ser de grande valor para o melhoramento através de aplicações diversas tais como análise da diversidade genética em cultivares (Smith, 1984; Cox et al., 1986), identificação de combinações parentais para a criação de progênies segregantes com o máximo de variabilidade genética para posterior seleção (Barret & Kidwell, 1998), e introgressão de genes de germoplasma diverso em background genético conhecido (Thompson et al., 1998). Por outro lado, o entendimento das relações genéticas entre linhagens tem um papel importante no planejamento de cruzamentos, na classificação de grupos heteróticos e na identificação de plantas para a proteção de variedades (Hallauer & Miranda, 1988).

A análise de diversidade em coleções de germoplasma pode facilitar a classificação, mais correta, dos acessos e a identificação de sub-amostras de acessos com possível utilidade em objetivos específicos do melhoramento (Mohammadi & Prasanna, 2003). Particular atenção tem sido dada à cultura do arroz na relação entre as espécies cultivadas e silvestres, entre as

subespécies, variedades, cultivares, linhagens do arroz cultivado, e entre os diferentes acessos que constituem o grande acervo de germoplasma de arroz mundial (Ni et al., 2002; Pessoa Filho, 2004; Zenga et al., 2004).

O estudo da diversidade genética é o processo pelo qual a variação entre indivíduos ou grupos de indivíduos é analisada por um método específico ou por uma combinação de métodos. Os dados envolvem medições numéricas e combinações de diferentes tipos de variáveis, sendo que os mais importantes são os dados de pedigrees (Bernardo, 1993; Messmer et al., 1993; van Hintun & Haalman, 1994), dados de passaporte e dados morfológicos (Smith & Smith, 1992; Bar-Hen et al., 1995) dados bioquímicos de isoenzimas (Hamrick & Godt, 1997) e de proteínas reserva (Smith et al., 1987) e, mais recentemente dados de marcadores baseados em DNA que permitem uma diferenciação mais correta dos genótipos. E, porque cada tipo de dados fornece diferente tipo de informação, a escolha do(s) método(s) analítico(s) depende do(s) objetivo(s) do experimento, do nível de resolução desejado, dos recursos e infra-estrutura tecnológica disponível, e das dificuldades operacionais e de tempo (Karp et al., 1997a; 1997b).

A distância genética é a diferença entre duas entidades que podem ser descritas pela variação alélica (Nei, 1973). Esta definição foi, posteriormente, mais elaborada por Nei (1987) e refinada por Beaumont et al. (1998) como “qualquer medida quantitativa da diferença genética, em nível de seqüência ou freqüência alélica, que é calculada entre indivíduos, populações ou espécies”. A distância genética, entre genótipos, populações ou indivíduos, pode ser calculada por várias medidas estatísticas dependendo do tipo de dados. A distância Euclidiana ou “linha direta” é comumente utilizada para dados morfológicos. Grower (1971) descreveu um coeficiente (Distância de Grower) para diferentes tipos de caracteres, dicotômicos, qualitativos e quantitativos. Várias medidas de distância genética foram propostas para análise de dados produzidos por marcadores moleculares para a análise de diversidade genética. Para marcadores SSRs e RFLPs a distância genética pode ser calculada através da distância de Rogers (RD) (Rogers, 1972), Coeficiente de Nei & Li (1979), Coeficiente de Jaccard (1908), Coeficiente de “Simple Match” (Sokal & Michener, 1958) ou Distância Baseada em Alelos Partilhados (DAS – Distance of Shared Alleles; Jin & Chakraborty, 1993). O Coeficiente de Jaccard leva em consideração “coincidências” entre banda (alelo) presentes e ignora as ausentes em ambos os indivíduos de uma comparação par-a-par. O Coeficiente de Nei e Li mede a proporção de bandas (alelos) partilhadas como resultado da herança a partir de ancestral, e representa a proporção de bandas (alelos) presente em cada

indivíduo; a Distância Euclidiana, por sua vez, leva em consideração as igualdades e desigualdades das bandas, e atribui igual peso às mesmas (Mohammadi & Prasanna, 2003). Já a Distância Baseada em Alelos Partilhados leva em conta os alelos compartilhados por dois indivíduos e é calculada como $D_{sa} = 1 - P_{sa}$, com $P_{sa} = \frac{\sum_s S}{S_r}$, onde o número de alelos partilhados é somado sobre todos os locos r (Jin & Chakraborty, 1993).

A escolha de medida de distância genética, na base do tipo de variável e da escala da medição, é um componente importante na análise da diversidade genética entre genótipos. Os coeficientes de Nei & Li e de Jaccard, diferem na atribuição do peso dos marcadores polimórficos dominantes e codominantes. Embora ambas as medidas levem a níveis idênticos de distância genética entre linhagens puras, a distância genética pode diferir quando se analisa locos heterozigotos em híbridos (Link et al., 1995) ou no caso de populações quando heterozigotos ocorrem normalmente. Para marcadores co-dominantes (RFLP e SSR), o Coeficiente de Nei & Li é uma função linear do coeficiente de co-ancestralidade (Melchinger, 1993). Para marcadores dominantes, esta propriedade é aplicada ao Coeficiente de Jaccard (Link et al., 1995). A distância de Igualdade Simples tem propriedades métricas euclidianas que possibilitam o seu uso em estratégias de agrupamento hierárquico (Ward, 1963) e AMOVA (Excoffier et al., 1992), que podem ser utilizadas para calcular os componentes de variância entre e dentro de grupos.

Com o aumento do tamanho das amostras do material de melhoramento e acessos de germoplasma utilizados em programas de melhoramento, os algoritmos estatísticos multivariados utilizados para classificar e ordenar a variabilidade genética estão assumindo grande importância. O seu uso é estratégico na classificação de germoplasma, na ordenação da variabilidade de um grande número de acessos, ou na análise das relações genéticas entre material de melhorado. Técnicas analíticas multivariadas, que analisam simultaneamente medições múltiplas em cada indivíduo, são amplamente utilizados independentemente do tipo de dados (morfológico, bioquímico ou molecular). Entre esses algoritmos, a Análise de Agrupamento, Análise do Componente Principal (PCA), Análise da Coordenada Principal (PcoA), e Escala Multidimensional (MDS) são os mais usados (Melchinger, 1993; Johns et al., 1997; Tompson et al., 1998; Brown-Guedira et al., 2000).

Estruturação Genética - O estudo da estrutura genética de populações naturais é fundamental ao entendimento da sua ecologia e as forças evolutivas que as afetam. O

conhecimento da estrutura genética de populações de espécies silvestres que representam recursos genéticos para plantas melhoradas é essencial para a conservação *ex-situ* e *in-situ*. O objetivo de genética de população é entender as forças que formam padrões de variabilidade genética dentro e entre espécies. A informação sobre diversidade e estrutura de população é essencial na ajuda aos melhoristas na seleção de pais para cruzamentos, fornecendo uma base mais racional para expandir o acervo de genes no material melhorado e para identificar materiais que abrigam alelos de valor para o melhoramento. É também de interesse entender as etapas de especiação e como as populações divergem e diferem no tempo. Os dados biológicos e genéticos das populações permitem entender melhor a estrutura desta diversidade e as forças agindo sobre ela, a fim de melhorar sua conservação e seu uso para bem estar humano (Garris et al., 2005; Semon et al., 2004).

Em arroz, a análise da estrutura genética de populações possibilita a caracterização da divergência intra-específica de acessos de germoplasma e permite estimar os relacionamentos evolutivos entre grupos dentro da espécie estudada. A compreensão da estruturação da diversidade genética intra-específica é de importância para examinar a base genética de domesticação, adaptação, desenvolvimento da planta, e desempenho agrícola (Garris et al., 2005).

A análise de agrupamento refere-se ao grupo de técnicas multivariadas cujo propósito primário é agrupar indivíduos ou objetos baseado nas características que eles possuem, de modo que indivíduos com descrições similares sejam matematicamente agrupados (Hair et al., 1995). Os grupos resultantes devem exibir grande homogeneidade interna e grande heterogeneidade externa. No caso de classificações bem sucedidas, os indivíduos dentro do grupo deverão estar próximos e grupos diferentes deverão estar afastados (Hair et al., 1995). Dois métodos de agrupamento são amplamente conhecidos: (1) Métodos baseados em distâncias, nos quais uma matriz de distâncias entre pares é utilizada para análise por um algoritmo específico (Johnson & Whichern, 1992), levando a uma representação gráfica (conhecida como árvore ou dendrogramas), e (2) método baseado em modelos, em que as observações de cada grupo são assumidas como amostragens aleatórias do modelo paramétrico, e inferências sobre os parâmetros de cada grupo e seu membro, são realizadas simultaneamente usando métodos estatísticos.

O método de agrupamento baseado em distâncias pode ser dividido em grupo hierarquizado e não-hierarquizado. Os métodos hierarquizados são os mais utilizados na análise

de diversidade genética de plantas agrícolas. Os métodos de agrupamento *Neighbor-Joining* (NJ) (Saitou & Nei, 1987) e *Sequential, Agglomerative, Hierarchical, e Nested* (SAHN) (Sneath & Sokal, 1973) procedem de uma série de “uniões” sucessivas ou por uma série de divisões de um grupo de indivíduos. Este último, também conhecido como método hierárquico aglomerativo, tem como referência o método UPGMA (do inglês *Unweighted Paired Group Method Arithmetic Averages*) (Sneath & Sokal, 1973; Panchen, 1992) que é um dos algoritmos mais adotados, seguido do Método de Variância de Ward (Ward, 1963). Os métodos não hierarquizados não envolvem dendrogramas ou árvores, e são também conhecidos como “K-means clustering” (Everitt, 1980). São métodos raramente utilizados na análise genética intra-específica em plantas agrícolas por falta de informação sobre o número ótimo de grupos necessário para a correta inserção dos indivíduos. Dendrogramas baseados em UPGMA têm sido amplamente referenciados pela sua consistência no agrupamento de materiais biológicos com relações computadas de diferentes tipos de dados (Ajmone-Marsan et al., 1992; 1998; Mohna et al., 1992; Murmm et al., 1994). Entretanto, algumas considerações em relação à congruência entre resultados de diferentes métodos, e o poder do teste e suas debilidades têm sido verificadas (Mohammadi & Prasanna, 2003).

Segundo Franco et al. (1997) os seguintes pontos devem ser cuidadosamente considerados na aplicação de medidas de diversidade genética e métodos multivariados para a análise de diversidade genética: (1) “sensato” e efetivo uso de diferentes tipos de variáveis contínua, discreta, ordinal, multiestado, binomial, etc; (2) aplicação de grupos de dados múltiplos (morfológico, bioquímico e molecular); (3) escolha correta da medida de distância genética e algoritmos de agrupamento; e (4) determinação do número ótimo de grupos.

Muitos são os softwares disponíveis para a análise da diversidade genética (Labate, 2000), embora NTSYS-pc (Rohlf, 1998) e PHYLIP (Felsenstein, 2004) proporcionem meios eficazes para tipos de dados diversos em plantas e animais. Estes programas oferecem análises na base de modelos evolucionários relevantes e a facilidade de uso para o usuário. Vários dos programas recentemente desenvolvidos oferecem opções de análise do nível de polimorfismo, frequências alélica e genotípica, homozigidade e heterozigidade, heterogeneidade, padrões de agrupamento, e reamostragem numérica (*bootstrap* e *jackknife*). Outros possibilitam o estudo da estruturação genética em populações envolvendo parâmetros estatísticos como o F de Wright (Wright, 1978) através das metodologias de (1) G_{st} de Nei, (2) θ de Weir & Cockerham, e (3) os Bayesianos.

O uso do parâmetro estatístico G_{st} é recomendado para distribuição da variação dentro e entre populações quando se está interessado em características de populações que incluem nas amostras efeitos fixos de amostragem. Por outro lado, a utilização do estatístico θ relaciona-se à amostragem aleatória das amostras, representando características de um grande grupo de populações (Corander et al., 2003). O método Bayesiano é usado para estimar a sub-estruturação populacional utilizando marcadores moleculares multilocus e informação geográfica fornecido pela amostragem. A distribuição posterior para a estimativa do subestruturação utilizando marcadores moleculares multilocus e as informações geográficas são fornecidas pelo modelo de amostragem. A distribuição posterior da subestruturação e frequência alélica das respectivas populações são disponíveis numa forma analítica quando o número de populações é pequeno, enquanto que uma aproximação baseada na simulação da Cadeia de Markov Monte Carlo pode ser obtida para número grande de populações. A maior vantagem desse método é que o número de populações é tratado como um parâmetro desconhecido. O que é tradicionalmente considerado como duas populações geneticamente distintas, quer recentemente formadas ou ligadas pelo fluxo gênico, é aqui tratada como uma única população com certa probabilidade baseada nos dados de marcadores moleculares (Holsinger 1999, Edwards & Beerli 2000; Kitada *et al.* 2000; Pritchard *et al.* 2000a; Lawson & Belkhir 2001).

A técnica de Cadeia de Markov Monte Carlo foi desenvolvida pela estatística física com diferente ênfase na computação para estudos com algoritmos aleatórios, estatística espacial e estatística aplicada, como é o caso do contexto Bayesiano. Entretanto, o trabalho teórico do desenvolvimento permitiu o desenvolvimento de inferências estatísticas para a maioria dos problemas de (bio)estatística, como por exemplo, análise de dados, análise de imagens, padrão espacial aleatório, modelos estatísticos financeiros, doenças epidêmicas contagiosas e na genética (Kendall et al., 2004). A Cadeia de Markov e Monte Carlo (MCMC) descreve uma classe de métodos que se baseiam em um tipo especial de simulação do processo estocástico, conhecido como cadeia de Markov, para estudar propriedades de uma distribuição complexa de probabilidade que não podem ser facilmente estudadas utilizando os métodos analíticos. A Cadeia de Markov gera uma série de variáveis aleatórias tais como a probabilidade de distribuição de estados esperada completamente determinados pelo estado da corrente em qualquer ponto da cadeia. Sob certas condições, a Cadeia de Markov terá uma distribuição estacionária, significando que se a cadeia está interagindo por um período suficiente de tempo, o estado visitado tenderá a uma probabilidade específica de distribuição que não dependerá do número de interações ou do estado

inicial da variável. A forma mais simples do MCMC é a integração de Monte Carlo (MC). A idéia básica da integração MC é que as propriedades das variáveis aleatórias podem ser estudadas simulando muitos exemplos de uma variável e analisando os resultados. Cada réplica da simulação é independente e o procedimento é equivalente a tomadas de várias amostras da Cadeia de Markov que é estacionária em pontos que são suficientemente separados e sem correlação. A integração MC tem sido amplamente aplicada em estatística genética e a sua vantagem é que as estimativas não são imparciais e o erro padrão das estimativas pode ser corretamente calculado porque as variáveis aleatoriamente simuladas são independentes e identicamente distribuídas. A desvantagem é que, para variáveis multidimensionais complexas, tem de haver um grande estado de espaço, elevado número de replicações para se obter uma correta estimativa do parâmetro (Beaumont & Rannala, 2004).

Vários são os programas computacionais que envolvem o algoritmo MCMC para inúmeras análises genéticas incluindo análise de estruturação genética em população: MDIV (<http://www.binf.ku.dk/users/rasmus/webpage/programs.html>; Nielsen & Wakeley, 2001); MicSat (<http://www.maths.abdn.ac.uk/~ijw/downloads/micsat/micsat.htm> ; Wilson & Balding, 1998); Migrate (<http://evolution.genetics.washington.edu/lamarc/migrate.html>; (Beerli & Felsenstein 1999; 2001); MrBayes (<http://morphbank.ebc.uu.se/mrbayes/>; Ronquist & Huelsenbeck, 2003); FSTMET (<http://www.rdg.ac.uk/statistics/genetics/software.html>; Balding & Nichols, 1997); HARDY (<http://www.stat.washington.edu/thompson/Genepi/Hardy.shtml>; Guo & Thompson, 1992); HWMET (<http://www.rdg.ac.uk/statistics/genetics/software.html>; Ayres & Balding, 1998); LDMET (<http://www.rdg.ac.uk/statistics/genetics/software.html>; Ayres & Balding, 2001); LEA (<http://www.cnrs-gif.fr/pge/bioinfo/lea/index.php?lang=en>; Langella et al., 2001); WinBug (<http://www.mic-bsu.cam.ac.uk/bugs/winbug/contents.shtml>); BAPS (Corander et al., 2003) e STRUCTURE 2.1 (Pritchard et al., 2000a; Pritchard & Wen, 2004; Falush et al., 2003).

O programa STRUCTURE versão 2.1 é um software utilizado na análise de estruturação de populações com base em dados de genotipagem de multilocus. Suas aplicações incluem inferências sobre a presença de populações distintas, designação de indivíduos às populações, estudo de zonas de híbridos, identificação de migrantes e indivíduos não puros, e cálculos de frequências alélicas de população em situações onde muitos indivíduos são migrantes ou híbridos. A análise tem envolvido na sua maioria dados de genotipagem de marcadores moleculares microsatélites, RFLPs e SNPs (Pritchard et al., 2000a). O poder deste método de associação

estruturada baseia-se na análise efetiva da estrutura da população, agrupamento correto e inclusão dos indivíduos em grupos apropriados, utilizando um número pequeno de marcadores moleculares não ligados, e a identificação de migrantes ou indivíduos com mistura genética. Este programa foi inicialmente utilizado com sucesso em estudos envolvendo populações humanas (Zöllner et al., 2004; Rosenberg et al., 2002), testes de associação (Falush et al., 2003; Wall & Pritchard, 2003; Pritchard & Rosenberg, 1999), mas tem sido também aplicado em plantas agrícolas tais como milho (Pritchard, 2001) e arroz (Semon et al., 2005).

1.7. Coleções nucleares

A diversidade genética em culturas alimentares pode ser analisada em diferentes níveis intra-específicos tais como linhagens ou clones, variedades tradicionais, híbridos e populações, e em nível de espécie. Em cada um dos casos, essa diversidade varia de acordo com a natureza do material genético. As estratégias de amostragem ao nível de população, ao contrário de linhas puras, são complicadas por fatores como ligação gênica, tamanho efetivo, modo de reprodução, migração, deriva e existência de subestruturação. A maioria das medidas de diversidade genética depende do número de indivíduos amostrados na população, do número de locos amostrados, da composição genotípica e alélica da população, do sistema de polinização, e do tamanho efetivo da população (Nei & Chesser, 1983; Namkoong, 1988).

Frankel et al, (1995) sugeriu dois conceitos de variação genética que são aplicados a nível populacional: (1) riqueza da população ou amostra, que corresponde ao número total de genótipos ou alelos presente na população ou amostra, e (2) frequência dos diferentes tipos de alelos na população ou amostra. A riqueza alélica é calculada tendo em conta o número médio de alelos por locos e a percentagem de locos polimórficos. Este parâmetro é sensível à presença ou ausência de alelos raros (~5%) na população ou amostra, bem como ao elevado grau de erro de amostragem associados aos alelos raros (Nei, 1987; Namkoong, 1988; Sjogren & Wyone, 1994). A percentagem de alelos polimórficos na população é um parâmetro crucial da variação genética por estar sujeita a um grande erro de amostragem e pode ser confiável quando um grande número de locos é amostrado (Brown & Weir, 1983).

A amostragem de indivíduos para a análise de diversidade genética realça a discriminação dos indivíduos marginais e é proporcional ao tamanho da amostra (Marshall & Brown, 1975; Brown, 1989b; Frankel et al., 1995). Em estudos de análise populacional torna-se necessário balancear o tamanho da amostra, contra a amostragem de várias populações, e o número de locos. Parece não haver recomendações simples sobre o tamanho da amostra, número de amostra, e número de locos ideais. Entretanto, Marshall e Brown (1975) recomendaram amostras com número igual ou superior a 59 gametas ou mais (amostra aleatória de 50 indivíduos diplóides) como sendo suficientemente grandes para ter uma probabilidade de 95% de detectar todos os alelos com uma frequência igual ou superior a 5%. Outras sugestões foram apresentadas por Crossa et al., (1993), Warburton et al. (2002) e Baverstock & Moritz (1996).

No caso de grande número de acessos de germoplasma conservados *ex-situ*, a amostragem pode envolver a seleção de acessos a partir de uma coleção inteira ou de plantas entre acessos que sejam representativas. A formação de coleções nucleares pode constituir uma etapa importante na solução desse objetivo. O conceito de coleção nuclear (do inglês, *core collection*) foi proposto por Frankel (1984) e, mais tarde Frankel & Brown (1984) e Brown (1989a, 1989b, 1994). Uma coleção nuclear é definida como “um grupo limitado de acessos derivado de uma coleção de germoplasma, escolhido para representar o espectro genético da coleção inteira”. A coleção nuclear deve incluir tanta diversidade genética quanto possível. O conceito, baseado na teoria de alelos neutros (Kimura & Crow, 1964), uma coleção nuclear deve representar cerca de 10% dos acessos de toda coleção e incluir (com uma probabilidade de 95%) aproximadamente 70% da riqueza alélica (alelos comuns e raros) disponíveis no acervo gênico inicial (Ewens, 1972; Brown, 1989a).

1.8. Melhoramento Genético de Arroz e Produção de Híbridos

Os principais objetivos da maioria dos programas de melhoramento de arroz são o desenvolvimento de cultivares com alto potencial produtivo, estabilidade na produção, resistência a doenças e qualidade de grãos. Outras características específicas têm sido também objeto de estudo de alguns programas, como o desenvolvimento de variedades adaptadas às condições ambientais de uma determinada área, como a tolerância à salinidade, ao frio, à seca e à toxidade por ferro. Estes objetivos poderão ser atingidos com o desenvolvimento e aplicação da tecnologia de produção de híbridos em arroz.

O híbrido do arroz é uma planta oriunda do cruzamento entre dois genitores geneticamente diferentes. Na produção de híbridos, o fenômeno da heterose é muito importante porque o híbrido resultante mostra superioridade agrônômica sobre os seus genitores. O termo heterose foi introduzido por Shull, em 1914, (Shull, 1948) para explicar o aumento de vigor, tamanho, produção de frutos, tempo de crescimento, resistência a doenças e pragas, ou a fatores climáticos, manifestada pelo cruzamento dois organismos diferentes. Duas hipóteses principais foram propostas para explicar a base genética da heterose: a dominância (Davenport, 1908) e a sobre-dominância (East 1908, 1936).

A hipótese de dominância sustenta que a heterose é devido à acumulação de genes dominantes no híbrido derivado de dois genitores, enquanto a hipótese de sobredominância sugere que o heterozigoto é mais vigoroso e produtivo do que qualquer um dos homozigotos. No caso de caracteres poligênicos, a heterose é resultado da dominância (Xiao et al 1995), parcial ou completa, sobredominância (Stuber et al 1992; Yu et al 1997; Li et al 2000), ou epistasia (Zhang et al., 2001), ou ainda, ela pode resultar da combinação de interações alélicas dentro e entre locos gênicos (Comstock & Robinson 1952). A heterose pode ter também como causa o efeito positivo específico do citoplasma materno sobre a componente nuclear do parente paterno (Virmani et al., 2003). Entretanto, a heterose é expressa na primeira geração, varia de acordo com o nível da diversidade parental e ou presença de blocos gênicos heteróticos nas linhagens parentais e, pode ser positiva ou negativa. A heterose pode ser classificada em três categorias, dependendo da referência usada para a comparação da performance de um híbrido: 1) a heterose da média dos genitores - que é o aumento ou diminuição na performance de um híbrido em comparação com valor médio dos genitores; 2) heterobeltiose - aumento ou diminuição da performance de um híbrido com o melhor genitor do cruzamento; e 3) heterose padrão - medida em relação à variedade padrão da região. Do ponto de vista prático, este último tipo de heterose é interessante porque ajuda o desenvolvimento de híbridos melhores que as variedades mais produtivas (IRRI, 1997; Virmani et al., 2003).

Comercialmente, a heterose tem sido explorada na produção de híbridos de arroz resultantes de cruzamentos intra-subespecífico, tais como *indica* × *indica* e *japonica* × *japonica*, e

muito raramente interespecíficos. Entretanto, recentemente foi desenvolvido um híbrido interespecífico (superhíbrido) que explora a heterose dos cruzamentos entre *indica* e *japonica*. Bons híbridos têm um potencial de produtividade de 15-20% maior do que a melhor variedade autofecundada crescendo sob condições similares (IRRI, 1997; Bong et al., 2000). O interesse na produção de híbridos surge devido à estagnação da produtividade das variedades de arroz, a redução de terras destinadas à produção, à maior produtividade em relação às variedades autofecundadas sob condições de campo, e a uma potencial capacidade de produzir melhor sob condições adversas de seca e salinidade (IRRI, 1997). Nos últimos 25 anos a tecnologia de híbridos em arroz tem contribuído significativamente para a segurança alimentar, proteção ambiental, e redução de desemprego em países como a China. Desde 1990, esta tecnologia tem sido desenvolvida e introduzida também na Índia, Vietnã, Filipinas, Bangladesh, e Estados Unidos e, atualmente vários outros países estão desenvolvendo a mesma tecnologia.

Sendo o arroz uma planta autopolinizada, o desenvolvimento de híbridos requer o uso de um sistema de esterilidade masculina que produza pólen inviável sendo, por isso, as espiguetas incapazes de produzir sementes por autopolinização. A esterilidade masculina pode ser definida como uma condição em que o grão de pólen é inviável ou não pode germinar e fertilizar normalmente, para a produção de sementes. Os sistemas de esterilidade-masculina conhecidos em arroz são os sistemas Citoplásmico Genético de Macho-esterilidade (CMS – Cytoplasmic-Genetic Male Sterility), Macho-Esterilidade Gênica Termo-sensível (TGMS – Thermosensitive Genic Male Sterility) e Macho-esterilidade Induzida Quimicamente.

O sistema macho estéril mais utilizado em programas de desenvolvimento de híbridos de arroz é o sistema de três linhagens que utiliza linhagens macho-esteréis citoplasmático (IRRI, 1997; Bong et al., 2000). Este tipo de esterilidade é causado por uma interação de fatores genéticos localizados no DNA citoplasmático e no DNA nuclear (IRRI, 1997; Koh et al., 1999). A produção de híbridos CMS envolve uma linhagem CMS (A), uma linhagem mantenedora (B) e uma linhagem restauradora (R), justificando a denominação de sistema de três linhagens. São poucos os genes CMS conhecidos em arroz (Koh et al., 1999), os quais têm sido utilizados em programas de melhoramento para a produção de sementes híbridas comerciais. O uso de CMS em termos práticos é complexo, com dificuldades na manutenção de linhagem macho-estéril pura, de linhagens restauradoras que possibilitem combinação heterótica adequada e de produção de

sementes em grande escala. A produção de sementes híbridas é realizada em duas etapas, sendo que na primeira se multiplica a linhagem CMS e, na segunda, as sementes são produzidas. Na etapa de multiplicação da linhagem CMS, esta linhagem é cruzada manualmente ou no campo por isolamento no espaço e no tempo, com uma linhagem mantenedora de forma que ocorra sincronização na floração de ambas as linhagens. Na etapa seguinte, produção de sementes, a linhagem CMS é plantada juntamente com uma linhagem restauradora, numa proporção fêmea:macho específica. O plantio deve permitir a sincronização das duas linhagens para permitir a polinização (Virmani et al., 2003).

A utilização do sistema CMS na produção de híbridos tem sido realizada com sucesso em arroz. Híbridos produzidos com este sistema têm produtividade maior do que as linhagens puras convencionais (Yuan, 1994a, b) e têm sido extensivamente usados em programas de melhoramento de vários países para produzir híbridos em vários *backgrounds* genéticos (Virmani, 1994). Embora efetivo, o sistema CMS é pouco econômico e trabalhoso, requerendo a seleção de linhagens mantenedoras e restauradoras apropriadas para o desenvolvimento de híbridos férteis (IRRI, 1997; Bong et al., 2000). Por outro lado, cerca de 90% das variedades híbridas na China fazem uso da mutação *WA* (*wild abortive*) de CMS (Yao et al., 1997). A dependência de um único tipo de fonte de macho-esterilidade citoplasmática, como o sistema *WA*, representa alta vulnerabilidade genética à ocorrência súbita de pestes e doenças. Basta lembrar do impacto negativo do uso em escala do citoplasma T em milho, no início dos anos 70, que possibilitou um ataque epidêmico de grandes proporções para a produção de milho nos Estados Unidos. Assim, a diversificação de fontes de macho-esterilidade deve ser um componente importante de um programa de melhoramento de híbridos de arroz (IRRI, 1997). Neste sentido, o Sistema Gênico de Macho-Esterilidade Termo-sensível (TGMS) apresenta-se como uma interessante opção ao uso indiscriminado do sistema CMS.

1.9. O Sistema TGMS em Arroz

O Sistema Gênico de Macho - esterilidade Termo-sensível (TGMS) de arroz é o sistema em que a fertilidade do pólen nas plantas é influenciada pelas flutuações de temperatura que ocorrem no ambiente de cultivo da espécie. Este tipo de esterilidade foi primeiro observado em pimenta por Martin & Crawford, em 1951, mas tem sido explorado comercialmente só em arroz devido ao trabalho pioneiro dos chineses (Virmani et al., 2003). Existem dois tipos de categorias

para o sistema TGMS: (i) TGMS (*temperature-sensitive genic male sterility*), propriamente dito e, (ii) rTGMS ou TGMS reverso (*reverse temperature-sensitive genic male sterility*).

A descoberta do sistema TGMS, atribuída ao professor Shi Ming Shang da Universidade de Hubei, na China (Yuan, 1987), e a aplicação do sistema na produção de híbridos representa uma potencial revolução na metodologia de produção de sementes híbridas de arroz. O sistema TGMS tem simplificado o procedimento da produção de sementes híbridas em arroz porque as linhagens TGMS podem ser usadas como macho-estéreis e como mantenedoras, isto é, não dependem de uma linhagem mantenedora específica como no sistema CMS, e têm grande valor devido ao amplo espectro de restauradores e seu grande potencial na produção de sementes híbridas (Lang et al., 1999; Jia et al., 2001). O uso do sistema TGMS é potencialmente simples, econômico, eficiente e elimina as limitações associadas com o sistema CMS. O arroz TGMS é usado na produção de híbridos do arroz na China e em alguns países asiáticos do Sul (Dong et al., 2000).

O sistema TGMS tem mostrado ser vantajoso em relação ao CMS por não necessitar de linhagens mantenedoras para a multiplicação genética, utilizar qualquer linhagem fértil como genitor masculino, não mostrar efeitos negativos da esterilidade induzida pelo citoplasma, poder ser incorporado em qualquer "background" genético, não necessitar de genes restauradores no genitor masculino, e não aumentar a vulnerabilidade genética à incidência de epidemias, como é plausível com o uso intensivo do sistema CMS (IRRI, 1997; Bong et al., 2000) (Tabela 4). Uma boa linhagem TGMS deve apresentar 100% de plantas macho-estéreis numa população de mais de 1000 plantas, 99,5% de pólen estéril em cada planta analisada, ter regimes de alteração esterilidade/fertilidade claramente definidos, ter a duração da fase macho-estéril por mais de 10 dias consecutivos, apresentar um bom potencial de produção de sementes durante a fase fértil para a multiplicação da linhagem, e indução da macho-esterilidade a uma temperatura o mais baixa possível (Maruyama et al., 1991; IRRI, 1997; Bong et al., 2000).

Há evidências de correlação entre efeitos de temperatura e fotoperíodo na expressão da macho-esterilidade termo-sensível, ou seja, não se observa a existência de tipos estritamente sensíveis à temperatura ou ao fotoperiodismo visto que esses fatores interagem em graus

diferentes baseados no *background* genético no qual o gene foi transferido (Reddy et al., 2000; Ku et al., 2001). Esta é uma questão ainda passível de debate visto que, quando em condições de campo, as plantas estão sujeitas simultaneamente à ação da temperatura e do fotoperíodo qualquer que seja a sua fase de desenvolvimento (Ku et al., 2001).

Tabela 4 – Diferenças entre linhagens CMS e TGMS na produção de híbridos

LINHAGENS CMS	LINHAGENS TGMS
Macho-esterilidade causada por uma interação de fatores genéticos no citoplasma e no núcleo	Macho-esterilidade causada por genes nucleares e influenciada pelas flutuações de temperatura e estado de desenvolvimento da panícula na planta
Processo tradicional na produção de híbridos	Novo processo na produção de híbridos
Envolve uma linhagem CMS (A), uma linhagem mantenedora (B) e uma linhagem restauradora (R), - sistema de três - linhagens.	Envolve uma linhagem TGMS e uma linhagem normal – sistema de duas-linhagens
Reduzido número de genes	Conhecidos vários genes e QTLs dispersos pelo genoma
Dificuldade de manutenção da linhagem macho-estéril e identificação de restaurador;	Não necessita de linhagem B e de restauradores específicos. Não apresenta efeitos negativos da esterilidade induzida pelo citoplasma
Pouco econômico e trabalhoso requerendo a escolha de linhas mantenedoras e restauradoras apropriadas	Uso é potencialmente simples, econômico, eficiente; tem simplificado o procedimento da produção de sementes híbridas.
Dependência excessiva de um único tipo de fonte de citoesterilidade, como por exemplo, o sistema WA, causando potencial vulnerabilidade genética a ataque de patógenos.	Pode ser incorporado em qualquer “background” genético, limitando a vulnerabilidade genética.

No sistema TGMS, a alteração esterilidade/fertilidade é condicionada por diferentes regimes de temperatura (IRRI, 1997; Lang et al., 1999) sendo que as linhagens TGMS permanecem estéreis em temperaturas altas (>30°C) e restauram a fertilidade em temperaturas mais baixas (<25°C; IRRI, 1997). Entretanto, a temperatura ambiental em relação a estádios específicos de desenvolvimento da planta tem sido referida como críticas para a indução e reversão da esterilidade. Bong et al. (2000) destacaram que na maioria das linhagens TGMS a temperatura crítica ocorria logo após o início da panícula e que a temperatura que condicionava a esterilidade seria superior a 25°C, enquanto que a fertilidade era restaurada à temperatura mais baixas. Reddy et al. (2001) demonstraram que as linhagens TGMS SA2 e 1024 eram completamente estéreis acima de 28°C e altamente férteis entre 20°C e 26°C. Estas duas linhagens sofriam também a influência do fotoperíodo. Uma outra linhagem, denominada IR68945 apresenta macho-esterilidade induzida por temperaturas de 30°C a 35°C, sem efeito de fotoperíodo (Reddy et al., 2001). A esterilidade de linhagens TGMS e PGMS (Photoperiod Genic Male Sterility) foi induzida a 32°C/26°C (dia/noite) com 14 horas de luz natural, enquanto que a fertilidade ocorria a 26°C/20°C (dia/noite) com 10 horas de luz natural. A observação microscópica das anteras estéreis revelou a falha completa do desenvolvimento do pólen (Ku et al., 2001). A alteração da fertilidade é condicionada também pela fase de desenvolvimento morfológico da planta que corresponde a uma fase termo-sensível crítica que varia entre 15 a 25 dias antes ou 5 a 15 dias depois do surgimento da panícula (IRRI 1997).

No desenvolvimento de linhagens TGMS, linhagens TGMS e doadoras de pólen são tipicamente autofecundadas e as respectivas progênes resultantes são intercruzadas. O híbrido resultante é um TGMS. Por envolver duas plantas, estes híbridos são também denominados sistemas de híbridos de duas linhagens (IRRI, 1997).

Mecanismo de Esterilidade Masculina Gênica Termo-sensível em Arroz - O arroz é cultivado principalmente em regiões tropicais e subtropicais, mas tanto a temperatura alta (Matsui & Omassa, 2002) quanto à temperatura baixa (Imin et al., 2004) ou o estresse hídrico (Sheoran & Saini, 1996) durante o florescimento podem induzir a esterilidade e limitar a produção de grãos. A antecipação de temperaturas elevadas pode induzir a esterilidade da flor e aumentar a instabilidade da produção de arroz (Horie et al. 1996). A principal causa da esterilidade induzida pela temperatura elevada na floração é a indeiscência das anteras (Satake & Yoshida, 1978; Mackill et al., 1982; Matsui et al., 2000a, b; 2001). Temperaturas elevadas na floração inibem a intumescência dos grãos de pólen (Matsui et al., 2000b), que é a força motora da deiscência das

anteras (Matsui et al., 1999a, b). Assim, as anteras não abririam durante a floração e o pólen não é liberado para polinizar o estigma, ocorrendo a macho-esterilidade.

Temperaturas baixas também induzem a esterilidade masculina durante o desenvolvimento do grão de pólen, no estágio de “*emborrachamento*” (Nishiyama 1995). Neste estágio, o frio causa limitações da produção em até 40% (Angus & Lewin 1991; Nishiyama 1984). A causa dessa perda é a esterilidade do pólen resultante do efeito pós-meiótico da temperatura baixa, no estágio de tétrades, ou estágio de micrósporos jovens (Satake & Hayase 1970). Estudos citológicos e histológicos revelaram anormalidades nas anteras (Terao, 1940, Sakai, 1943 citados por Imin et al., 2004). As anteras tornam-se pequenas, as células do tapetum que envolvem e são responsáveis pela transferência de nutrientes para os grãos de pólen sofrem hipertrofia e morte. Como resultado, não ocorre o desenvolvimento normal do pólen, e esses grãos contendo pouco ou nenhum amido, tornam-se funcionalmente estéreis (Nishiyama, 1984). Em nível bioquímico verifica-se uma diminuição da respiração, sacarose acumula-se nas anteras, os níveis protéicos caem e a composição de aminoácidos se altera, incluindo uma acentuada diminuição de prolina e aumento de asparagina (Ito, 1970). Alguns compostos químicos foram sugeridos como tendo efeito na macho-esterilidade, como a ação da β -1,3-glucanase na liberação das tétrades (Stieglitz, 1973; Tsuchiya et al., 1995), a ação da oxidase no desenvolvimento do pólen (Kitashiba 1999) e a diminuição da atividade da invertase nas anteras sob estresse hídrico durante a meiose, sugerindo que um distúrbio no metabolismo de carboidratos pode estar envolvido na macho-esterilidade. Embora, no passado se tenha sugerido a causa da esterilidade masculina induzida pelo frio como sendo a quebra do metabolismo de açúcares, Imin et al. (2004) sugerem outros compostos (síntese e formação da estrutura secundária e terciária de proteínas, biogêneses de lipídios e degradação de proteínas).

Por sua vez, o estresse hídrico inibe a deiscência da antera, germinação do pólen e reduz a sua viabilidade (O’Toole & Namuco 1983; Ekanayake et al. 1990). A perda do grão em plantas sob estresse está associada à falha no desenvolvimento da panícula a partir da folha de bandeira (O’Toole & Namuco 1983; Cruz & O’Toole 1984). O déficit de água durante os estágios precoces da meiose nas células-mãe do pólen (PMC) também reduz a produtividade de grãos (Sato 1954; Namuco & O’Toole 1986). O estresse no estágio de meiose causa um comportamento anormal dos cromossomos e interrupção precoce do desenvolvimento da flor.

1.10. Efeito da Temperatura na Esterilidade do Pólen

Dentre os sistemas de macho-esterilidade mais conhecidos, o sistema de macho-esterilidade termo-sensível é visto com grande entusiasmo para a produção de híbridos devido as suas vantagens sobre a esterilidade citoplasmática. No entanto, o aproveitamento deste sistema no melhoramento de plantas só será pleno e eficaz quando se conhecerem efetivamente os fenômenos bioquímicos e moleculares que estão na sua base. A esterilidade masculina e a redução da produtividade de sementes têm sido reportadas em algumas plantas agrícolas, tais como arroz (Matsui & Kagata, 2003), milho, soja, algodão, girassol, trigo, arábida, brassica (Morrison & Stewart, 2002) e petúnia (Napoli et al., 1999), trigo (Saini & Aspinall, 1981), milho (Schoper et al., 1987a; 1987b; Mitchell & Petolino, 1988), algodão (Kittock et al., 1988), feijão frade (Ahmed & Hall, 1993), e ervilha (Guilioni et al., 1997). Em paralelo a esses estudos, várias publicações foram feitas sobre o desenvolvimento dos órgãos masculinos nas plantas, o efeito da temperatura no desenvolvimento das anteras, e a presença de algumas proteínas como resposta às oscilações de temperatura.

Vários genes TGMS recessivos e alguns poucos dominantes (Wang et al., 1995; Subudhi et al., 1997; Lang et al., 1999; Li et al., 1999; Bong, et al., 2000; Jia, et al, 2000) foram, até o momento, identificados. A fase do desenvolvimento das anteras/pólen em que esses genes são expressos precisa ser esclarecida. O estudo do desenvolvimento da antera é importante porque não só ajuda a compreender a regulação gênica na reprodução sexual das plantas com flores, mas também devido ao seu potencial na agricultura como macho-estéril. Esse desenvolvimento pode ser dividido em duas fases: (1) estabelecimento da morfologia da antera e geração das tétrades dos micrósporos, e (2) diferenciação do pólen e antera, degeneração do tecido, deiscência e liberação do pólen. Esses processos são acompanhados por diversos genes, mas somente 10-20% são antera-específicos. Jeon et al. (1999) identificaram um clone de cDNA específico da antera da planta de arroz, RA8, e indicaram que o transcrito desse gene está presente especificamente nas anteras e o nível do transcrito aumenta com a maturação da planta. O clone de RA8 codifica uma proteína com 264 aminoácidos com uma região N-terminal hidrofóbica cuja seqüência codante não mostra homologia significativa com qualquer outra seqüência conhecida.

Matsui et al. (2001) argumentam que, por ser o arroz uma planta autógama, não há autopolinização se não houver a deiscência das anteras. Temperaturas elevadas por ocasião da

floração inibem a intumescência dos grãos de pólen (Matsui et al., 2000b), enquanto que temperaturas baixas, no estágio de "rebentos" impedem o desenvolvimento do pólen (Shimazaki et al., 1964). Visto que a força que determina a deiscência da antera é o aumento de volume do pólen na altura da floração (Matsui et al., 1999a, 1999b), cogita-se que o estresse da alta temperatura reduz a percentagem de anteras abertas na floração (Shimazaki et al., 1964; Matsui et al., 2000b). Assim, temperaturas muito altas e/ou muito baixas resultam numa polinização reduzida e perda de produtividade (Horie et al., 1992, 1996).

Ku et al., (2001) testaram várias linhagens TGMS quanto ao grau de fertilidade de pólen sob diferentes condições de temperatura, e verificaram que em duas linhagens ocorria a inativação do pólen apenas por efeito térmico, enquanto em outras linhagens foi constatado um efeito fotoperiodismo sugerindo que a esterilidade masculina é provavelmente controlada por mais de um gene. Anteras dissecadas sob microscopia de transmissão mostraram anteras estéreis com lóculos completamente vazios, mas os restantes tecidos normais. No estágio de células-mãe do pólen as anteras macho-estéreis não mostraram qualquer diferença estrutural em relação às anteras controle, sugerindo que a falha de desenvolvimento de pólen ocorre depois da meiose. Kalaiyarasi & Vaidyanathan (2003) ao estudarem o mecanismo de esterilidade em linhagens TGMS e CMS, verificaram que o estágio crítico para a fertilidade do pólen era o estágio de "primórdios do pistilo e estame" e que genes nucleares envolvidos têm ação pré- e pós-meiótica, causando a esterilidade do pólen e do esporófito.

Ku et al. (2001) revelaram que o tapetum, tecido que provê nutrientes para a formação do pólen, apresenta importante papel no desenvolvimento do mesmo. O tapetum cujas células sofrem uma morte celular programada (DCD) está completamente reduzido na altura da maturação completa do pólen. Observações citológicas revelaram que as anteras de arroz TGMS, nos estágios precoces de desenvolvimento do pólen, iniciam a morte prematura das células (PCD) que continua até a completa degradação das células do tapetum.

Ogihara et al. (1997) reportaram o fenômeno de pistiloidia (morfogênese de outros órgãos em pistilos) no trigo aloplásmico com citoplasma de *Aegilops crassa*, o que induz a esterilidade masculina. Análise de hibridização com DNA e RNA mitocondrial mostrou que as linhagens aloplásmicas diferem do trigo selvagem e de *Aegilops*. Esta diferença deve-se às substituições nucleares que induzem as alterações estruturais no genoma mitocondrial nas regiões cox III e

ORF25. Os padrões de transcrição da ORF25 estão associados com a reversão do fenômeno de PCMS, mostrando a relação entre o genoma nuclear e o condrioma. A disfunção mitocondrial parece ser causada por uma falha de um controle cooperativo da expressão dos genes mitocondriais durante a morfogênese da flor, especialmente o processo que determina a identidade do órgão.

Em brássicas (*B. napus* e *B. rapa*), Morrison e Stewart (2002) constataram que o período de uma a duas semanas depois do aparecimento da primeira flor é crítico para a produção de sementes. Esse dado foi comprovado por Richards e Thurling (1978) e McGregor (1981), que reportaram que uma demora na plantação, resultando em temperaturas altas durante a floração, causou baixa de produtividade em brássicas. Young et al. (2004) verificaram que plantas de brássica quando expostas ao estresse por aumento de temperatura (HTS) na floração, reduziam consideravelmente o desenvolvimento de sementes e frutos. Na cessação do HTS, as plantas compensavam a falta de fruto e sementes com o aumento do número de inflorescências laterais. Durante o HTS, a viabilidade do pólen e a germinação foram reduzidas. O crescimento *in vitro* do tubo polínico a 15°C parece anormal, embora *in vivo* o crescimento do tubo aparente normalidade. Os perfis de expressão de um subgrupo de proteínas "heat shock" (HSP101, HSP70, HSP17.6) mostrou que o micro- e megagametófito eram termos-sensíveis. HTS durante a floração resulta na redução da produção de sementes em plantas mono e dicotiledôneas, por exemplo, *Brassica napus* (L.) (Nuttal et al., 1992; Morrison, 1993; Angadi et al., 2000), *Linum usitatissimum* (L.) (Gusta et al., 1997), *Lycopersicon esculentum* (Mill.) (Peet et al., 1998; Sato et al., 2002), *Phaseolus vulgaris* (L.) (Shonnard & Gepts, 1994), *Triticum aestivum* (L.) (Saini et al., 1983), and *Zea mays* (L.) (Herrero & Johnson, 1980; Carlson, 1990). O número de espécies afetadas pelo HTS durante os estágios reprodutivos sugere que algum mecanismo comum possa estar envolvido. Uma resposta universal ao HTS é a produção de proteínas HSPs (HSPs, Nagao et al., 1986). HSPs pequenas (Malik et al., 1999; Park & Bong, 2002), HSP90 (Ludwig Muller et al., 2000) e HSP101 (Hong et al., 2000, 2001; Queitsch et al., 2000) são necessárias para melhorar a termo-tolerância em plantas. Em *Glycine max* (L.) Merrill (Kimpel & Key, 1985) e *Malus domestica* Borkh (Ferguson et al., 1998), HSPs foram sintetizadas em folhas sob condições de campo em resposta a HTS. Entretanto, não se sabe se HTS induz a expressão de HSP em gametófitos de *B. napus*.

Em ervilha (*Pisum sativum* L.), Escobar Galvis et al. (2001) verificaram que a resposta a estresse de temperatura envolvia a interação de uma enzima mitocondrial, Nucleosídeo Difosfato

Quinase (mtNDPK), e uma proteína de 86 kilodaltons. A isoforma mitocondrial da enzima mtNDPK é muito abundante nos tecidos reprodutivos e jovens da planta. A expressão dessa enzima não é por si só afetada por estresse algum, mas associa-se a uma proteína de 86-KD, sintetizada sob efeito da temperatura alta. Estudos sobre oligomerização revelaram que mtNDPK apresenta uma flexibilização oligomérica. Esta flexibilidade possibilita à enzima interagir com a proteína 86-KD, atuando como seu modulador. As enzimas NDPK são enzimas ubiquitinas que transferem grupos fosfatos de nucleosídeos trifosfatos para nucleosídeos trifosfatos (Agarwal & Parks, 1971). Em plantas, resposta do fitocromo B (Choi et al., 1999), sinalização da luz UV-B (Zimmermann et al., 1999), e respostas hormonais (Nato et al., 1997; Novikova et al., 1999) são processos nos quais as isoformas de NDPK estão envolvidas. Não obstante, a resposta a temperaturas altas nas plantas tem envolvido enzimas NDPK (Moisyadi et al., 1994) e BiP (binding protein) (Hurkman et al., 1998). Os estudos sobre estresse de temperaturas altas em mitocôndrias de plantas têm focado na caracterização de proteínas *heat shock* (HSPs) (Neumann et al., 1993; Lund et al., 1998). Somente uma pequena proteína HSP mitocondrial, que proporciona termo-tolerância ao complexo respiratório I (Downs & Heckathorn, 1998), foi até o momento relatada.

Napoli et al., (1999) verificaram em um mutante de petunia, caracterizado por possuir anteras brancas, não acumulava flavonóides no pólen, e induzia a esterilidade masculina. A complementação transgênica com cDNA de chalcone sintase A (ChsA) sugere que a lesão genética responsável pelo fenótipo mutante está no gene Chs, o gene responsável pela primeira enzima da via biossintética de flavonol. O estudo mostrou que apenas dois genes Chs (A e J) são expressos nos tecidos reprodutivos e são diferencialmente regulados na corola e antera. Os flavonóides são compostos específicos das plantas que se acumulam virtualmente em todos os tecidos vegetais (Koes et al., 1994). Esses compostos são classificados de acordo com estado de oxidação do anel pirano central. Os flavonóides camferol e quercetina (Vogt & Taylor, 1995) são essenciais para a germinação do pólen e crescimento do tubo polínico em petunia (*Petunia hybrida*) e milho (Mo et al., 1992; Taylor & Jorgensen, 1992; Pollak et al., 1993). As plantas mutantes com a falta de chalcone sintase (CHS), que cataliza a primeira etapa da síntese de flavonóides, não acumulam flavonóides e é macho-estéril. O pólen mutante pode ser induzido a germinar pela suplementação com camferol, um flavonóide aglicone. A complementação bioquímica ocorre naturalmente quando o pólen mutante é cruzado com o tipo normal. Os autores reportaram que a polinização depende da maturidade do estigma, indicando que a acumulação do flavonol aglicone pode ser regulada pelo desenvolvimento.

A via biossintética de flavonóide em petunia, como em muitas outras espécies, foi bem caracterizada, e foram identificados muitos genes estruturais e reguladores (Dooner et al., 1991; Martin et al., 1991; Forkmann et al., 1994). Em petunia foram identificados 8 a 10 cópias de genes Chs expressos, mas apenas dois deles, ChsA e ChsJ são expressos nas anteras e corola (Koes et al., 1989). Também foram identificados sítios reguladores, que controlam na transcrição a expressão temporal e tecido-específico dos genes flavonóides estruturais. O gene *An4* controla produção de flavonóide nas anteras regulando ChsJ mas não ChsA (Quattrocchio et al., 1993). Em planta homocigota recessiva para o alelo *an4* observa-se a expressão de ChsA e ChsJ nas corolas, mas os tecidos da antera acumulam mRNA de ChsA exclusivamente (Quattrocchio et al., 1993). A falta da proteína CHS nas anteras de petunia e de milho, resulta em pólen branco que é destituído de todo o flavonol e não pode germinar ou produzir um tubo de pólen funcional em autopolinizações (Mo et al., 1992; Pollak et al., 1993). O elemento bioativo da petunia selvagem, identificado nos estigmas é, novamente, o camferol. Embora o mecanismo de indução do flavonol na germinação do pólen seja completamente desconhecido, a complementação bioquímica com CMF ("resgate de pólen") estabelece que a resposta dos flavonóis exógenos é sensível e específica aos aglicones de flavonol (Vogt & Taylor, 1995). O sucesso da polinização depende da maturidade do estigma, indicando que a acumulação de aglicone de flavonol pode ser regulada durante o desenvolvimento.

Em *Arabidopsis*, um componente essencial de sinalização para ao desenvolvimento e fertilidade do pólen é o jasmonato (JA). Os mutantes que bloqueiam a via sintética de JA resultam em macho-estéreis. Mandaokar et al. (2003) estudaram os processos de maturação das anteras e pólen por microarranjos e expressão diferencial dos genes selvagem e mutante macho-estéril. A análise por "Northern blot" confirmou a baixa expressão de *opr1* comparado com o gene selvagem. Quatro genes eram induzidos nas anteras *opr1* dentro de 0.5-1 h de tratamento com JA, enquanto que os genes restantes eram regulados depois de 1-8 h. Os quatro genes de indução precoce, que codificam a lipoxigenase, um provável fator de transcrição *bHLH*, uma proteína especificadora do epitóio (ESP – Epithiospecifier protein) que especifica a reação dos produtos e não dos substratos no sistema mirosinase-glucosinolato (sistema defensivo das plantas) e uma proteína desconhecida, podem estar envolvidos na sinalização de JA e na iniciação do processo de desenvolvimento. Os quatro genes de indução tardia, que codificam uma proteína semelhante à extensina, um transportador de peptídeos e duas proteínas desconhecidas, podem representar componentes

necessários na fase final da maturação da antera e do pólen.

Guan et al. (2001) estudaram a localização da atividade da ATPase em anteras férteis e estéreis de arroz, analisando a precipitação de Pb^{2+} . Os autores verificaram que na fase de desenvolvimento do pólen denominada de vacúolo grande, o tapetum das anteras férteis mostra grande atividade da ATPase no seu núcleo. No pólen fértil, a ATPase foi localizada na face externa da exina e no núcleo, nas fases de grande vacúolo e de dois-núcleos do pólen. Nas anteras estéreis, o tapetum estava completamente degradado na fase de grande vacúolo. Nos pólenes estéreis, a ATPase estava localizada na plasmembrana e na intina. Este fenômeno prolongava-se até a fase de bi-nucleado. A maioria do pólen estéril não mostrou enditina bem desenvolvida. Estes resultados sugerem que a degradação do *tapetum* não fornece suficiente nutriente para o desenvolvimento do pólen, e a elevada atividade da ATPase, na membrana do plasma e na intina, pode resultar na escassez de ATP nas anteras estéreis.

Goetz et al. (2001) reportaram o papel crítico dos carboidratos no desenvolvimento do pólen e da antera. Os carboidratos, fornecidos pelo tapetum, são utilizados para suprir o crescimento e os sinais que influenciam o desenvolvimento *in vivo* e *in vitro*. Várias linhagens macho-estéreis são caracterizadas por apresentarem um metabolismo de carboidratos alterado. Os açúcares assimiláveis são produzidos fotossinteticamente e, via espaço apoplásmico, transportado até as anteras. Existem invertases extracelulares, cujas isoenzimas já foram identificados em tomate, batata e tabaco (Goetz et al., 2001), que são expressas espacial e temporalmente nas anteras e ajudam a clivagem de açúcares ajudando a sua regulação. Plantas transgênicas contendo uma seqüência antisense da invertase são bloqueadas na fase precoce do desenvolvimento do pólen, causando a esterilidade masculina. Portanto, existe uma relação entre a clivagem extracelular de sacarose e o desenvolvimento da antera e do pólen.

Dixon et al. (1995) estudaram o metabolismo de composto fenilpropanol como reposta ao estresse ambiental e constataram que um grande número de fenilpropanóis induzidos pelo estresse era derivado do esqueleto flavonóide C15, sintetizado através da chalcone sintase. Os fenilpropanóis são derivados do ácido cinâmico, que é formado a partir de fenilalanina pela ação da fenilalanina amônia-liase (PAL), a ramificação enzimática entre o metabolismo primário (via Shiquimato) (Hermann, 1995) e secundário (fenilpropanol) (Harbone, 1988; Hahlbrock & Scheel, 1989; Lewis & Yamamoto, 1990; Dixon et al., 1992).

Em resumo, podemos dizer que a esterilidade do pólen pode ser resultado da resposta da planta aos efeitos ambientais (temperatura e umidade) ou a efeitos genéticos. Como resposta a flutuações de temperatura alta e baixa umidade, as anteras não liberam os grãos de pólen impossibilitando o autocruzamento das plantas. Estes fatores ambientais provocam respostas das plantas, através de proteínas termo-tolerantes, ação "heat shock", etc., através da expressão de genes nucleares que podem ter algum efeito nas mitocondriais responsáveis pelo fornecimento de energia à planta e identidade dos órgãos maduros. A severidade desses fatores pode tornar a planta estéril. Por outro lado, mutações gênicas podem levar ao bloqueio de vias biossintéticas e tornar as plantas deficientes em certos metabolitos vitais (camferol, jasmonate, etc.) para o desenvolvimento dos órgãos reprodutivos masculinos da plantas, tornando-as macho-estéreis. Muitos dos genes com efeito na esterilidade têm ação pré-meiótica, pois resultam em anteras sem pólen, enquanto outros têm ação pós-meiótica, ou seja, o pólen é formado, mas não se desenvolve. Como se pode constatar, o fenômeno da esterilidade masculina é um complexo e desafiante sistema de regulação gênica.

1.11. Linhagens TGMS de Arroz

A descoberta do sistema TGMS tem simplificado a produção de sementes híbridas de arroz porque, quando se compara com o sistema CMS, elas podem ser utilizadas ao mesmo tempo como linhagens macho-estéreis e linhagens mantenedoras. As linhagens TGMS têm feito uma grande contribuição e dado benefícios econômicos à produção de arroz híbrido do sistema de duas-linhagens. Na China, a área de cultivo de variedades híbridas 2-linhagens aumentou de 60,000 hectares em 1994 para 330,000 hectares em 1998 e para 2.000.000 hectares in 2001. As variedades híbridas TGMS e PGMS (por exemplo, as variedades "Pei ai ST" e "Pei ai 49") tiveram uma vantagem de produtividade de 5 a 10% maior do que as variedades híbridas de 3-linhagens, e 20 a 30% maior do que as variedades modernas de arroz (ICARDA, 2001; webmaster@agroviet.gov.vn).

A característica TGMS é controlada, na maioria dos casos, por genes TGMS monogênicos recessivos e não alélicos, tornando difícil a sua aplicação no melhoramento de arroz (Jia et al., 2001). No entanto, existem também linhagens TGMS (exemplo, "Pingxiand" e "8987") com o fenótipo controlado por genes TGMS dominantes (Tabela 5). Várias linhagens TGMS têm sido

obtidas tanto por seleção de mutações espontâneas como por mutações induzidas por radiação, nos países asiáticos (China, Japão, Tailândia, Filipinas e Vietnã) onde o desenvolvimento de variedades híbridas em arroz é grande (Dong et al., 2001; Li et al., 1999a; 1999b) (Tabela 5).

Um dos problemas relacionados com os locos de macho-esterilidade termo-sensível é a sua influência na distorção de segregação. Reddy et al., (2000) revelaram que as linhagens TGMS ID24, SA2 e IR68945 apresentaram a maioria da sua progênie expressando segregação transgressiva em relação à esterilidade ou fertilidade, causando instabilidade da fertilidade e baixa reversibilidade, que pode, segundo os autores do estudo, ser devida ao grande número de loco QTL simples e a interação epistática entre esses locos. He et al. (1999) reportaram que a reversão da esterilidade para a fertilidade era influenciada pela duração do dia, embora a instabilidade fosse causada pela oscilação da temperatura. Embora existam linhagens TGMS espontâneas, algumas linhagens têm sido desenvolvidas também por mutação induzida (IRRI, 1992; Subudhi et al., 1998; Lang et al., 1999).

Tabela 5 - Descrição de linhagens macho-estéreis termo-sensíveis utilizadas em programas de melhoramento genético

Linha TGMS	Indução TGMS	Derivação	Temperatura Estéril/Fértil	Gene TGMS	Origem	Referência
------------	--------------	-----------	-------------------------------	--------------	--------	------------

J207S	Espontânea	(IR64xZhacang)xJing101	<11C/>11C	Rtms1	China	Jia et al., 2001
AnnongS-1	-	-	-	-	-	Jia et al., 2001
TGMS-VN1	-	-	-	tms4(t)	Vietnan	Dong et al., 2000
N422S	Mutagênese	cultivar Chiem Bau	-	-	-	Li et al., 1999
Peiai 64S	-	-	-	-	-	Li et al., 1999
IR12164TGMS	Raios gama	-	-	-	-	Subudhi et al., 1998
IR12164	-	IR12164	-	tms1	Filipinas	Subudhi et al., 1998
ID24	-	-	-	tms4	Índia	Reddy et al., 2000
SA2	-	-	28C/20-26C	-	Índia	Reddy et al., 2000
IR68945	-	-	30-35C	tms2	Filipinas	Reddy et al., 2000
NorinPL 12*	-	NorinPL12	-	tms1	Filipinas	Lopez et al., 2000,
54605	-	-	-	-	Filipinas	Reddy et al., 2000
Pingxiang	dominante	-	-	-	-	-
8987	dominante	-	-	-	-	-
IRTGMS-68945-433-4-14-48	-	-	27°C	QTLs	IRRI	Reddy et al., 2000

* mutante *japonica* TGMS temperado, Norin PL 12, ancestral de IRTGMS-68945-433-4-14-48 (Lopez & Virmani, 2000)

1.12. Marcadores Moleculares

O desenvolvimento de marcadores moleculares veio dar um impulso maior à pesquisa genética, propiciando a detecção de polimorfismo de DNA com comportamento Mendeliano, passível de ser utilizado em diferentes áreas da genética e do melhoramento de plantas (Ferreira & Grattapaglia, 1998). O surgimento da técnica de PCR (Polimerase Chain Reaction) tornou possível aliar o poder de informação gerada por marcadores moleculares e a rapidez do processo de genotipagem, baseada em ciclos contínuos de desnaturação e amplificação das fitas de DNA mediados pela enzima DNA polimerase (Mullis & Faloona, 1987). Entre os marcadores moleculares baseados na reação de polimerase em cadeia destacam-se os marcadores microssatélites ou SSR (Simple Sequence Repeats, Weber & May, 1989; Litt & Luty, 1989).

A técnica de microssatélites baseia-se na detecção de variação existente entre locos de seqüências repetitivas (di-, tri-, tetra- penta- e hexa-nucleotídeos). Esta variação é dada pela diferença do número de repetições em tandem da seqüência repetitiva microssatélite. As seqüências que flanqueiam estes locos são mais conservadas, permitindo que primers específicos a elas sejam sintetizados, e utilizados para amplificar especificamente essa região via PCR. Os marcadores microssatélites são altamente informativos, permitindo detectar polimorfismo molecular com grande eficácia (Ferreira & Grattapaglia, 1998; McCouch et al., 1997). Os SSR apresentam as seguintes características: são marcadores codominantes; estão amplamente distribuídos pelo

genoma de eucariotos, são multialélicos, são amplificados via PCR, o que facilita a sua obtenção mesmo com pouca quantidade de DNA, os primers podem ser facilmente compartilhados entre laboratórios, primers marcados com fluorescência possibilitam sistemas múltiplos de detecção, e conseqüentemente, avaliar rapidamente um grande número de locos por indivíduo (Ferreira & Grattapaglia, 1996; Akagi et al., 1996; McCouch et al., 1997).

O processo de isolamento de locos microsatélites úteis é laborioso e, em geral, de alto custo (Holton, 2002). O método padrão para o isolamento de SSRs envolve: 1) a criação de uma pequena biblioteca de insertos genômicos; 2) a análise da biblioteca para seqüências microsatélites por hibridização; 3) a localização da posição da seqüência microsatélite dentro do clone positivos utilizando PCR ancorado; 4) o sequenciamento de clones onde a posição da seqüência microsatélite é adequada ao desenvolvimento de primers para a sua amplificação por PCR; 5) desenho de primers; 6) a seleção dos marcadores SSR analisados via PCR; e 7) identificação dos marcadores polimórficos (Rafalski et al., 1996). A eficiência de isolamento de microsatélites pode ser aumentada através de um processo de enriquecimento das bibliotecas (Ritschel et al., 2004). Para espécies que apresentam um grande número de dados de seqüências gênicas devido ao sequenciamento genômico ou de ESTs, é possível identificar microsatélites através do exame de Bancos de Seqüências de DNA. A vantagem de utilizar microsatélites presentes em ESTs é a possibilidade de mapeamento de genes com funções conhecidas.

A detecção dos marcadores microsatélites, a determinação do tamanho dos seus fragmentos e do tamanho do polimorfismo pode ser feita por eletroforese de gel de agarose, poliacrilamida ou por sistema de capilares em seqüenciadores de DNA. A coloração com brometo de etídio é comumente utilizada para detectar produtos de PCR em géis não desnaturantes, i.e., géis de agarose. A resolução em agarose é, no entanto, limitada em comparação com poliacrilamida. Vale comentar que géis de agarose MetaPhor 2% e UltraPura 1% possibilitam maior distinção de diferenças alélicas (Becker & Heun, 1995). A resolução de nucleotídeos simples requer o uso de géis de poliacrilamida desnaturante ou por capilaridade utilizando coloração de prata, marcação radioativa ou fluorescência. A marcação com fluorescências, além de apresentar vantagens sobre a marcação radioativa e a coloração com prata, permite, combinada com um sistema automático eletroforético, a mais alta resolução na análise dos tamanhos dos alelos SSR.

A aplicação de marcadores microsatélites em estudos de diversidade alélica tem

possibilitado a identificação de acessos de arroz em bancos de germoplasma e em programas de melhoramento. Olufowote et al. (1997) discriminaram setenta e uma linhagens de arroz geneticamente relacionadas com apenas seis marcadores microssatélites. O custo inicial para o desenvolvimento de marcadores microssatélites é plenamente compensado pela facilidade de transferência destes marcadores para outros grupos de pesquisa, através da divulgação das seqüências de primers. Devido ao contínuo desenvolvimento de bibliotecas genômicas específicas para o desenvolvimento de "primers" SSR e ao sequenciamento do genoma de arroz (Goff et al., 2002; Yu et al., 2002; RGISC, 2004), cerca de 50.000 seqüências microssatélites de arroz foram identificadas, das quais 24% são repetições dinucleotídeas, 59% são trinucleotídeas e 17% tetranucleotídeas. Os motivos mais freqüentes são AC/CT, CGG/CCG e ATCG/CGAT. A principal limitação da técnica de microssatélites é o custo inicial de desenvolvimento dos primers específicos às regiões flangeadoras das seqüências repetitivas.

Os marcadores moleculares podem ser úteis no estabelecimento da identidade genética e no diagnóstico da heterogeneidade (Ferreira & Grattapaglia, 1998). Algumas técnicas estão disponíveis para a detecção de diferenças genéticas dentro e entre cultivares. Microssatélites ou "Simple Sequence Repeats" (SSR) são, presentemente, os marcadores de escolha para a maioria de espécies, possibilitando a construção de mapas genéticos altamente densos (Senior & Heun, 1993; Taramino & Tingey, 1996). O atual nível de cobertura genômica em arroz dada pelos marcadores microssatélites possibilita, por exemplo, o fingerprinting de DNA e a proteção de variedades (Olufowote et al., 1997; Rongwen et al., 1995; Smith & Helentjaris, 1996), estudos relativos à conservação e utilização de germoplasma (Bretting & Widrechner, 1995; Brown & Kresovich, 1996; Hahn & Grifo, 1996; Olufowoote et al., 1997; Powell et al., 1996; Xiao et al., 1996a; 1996b; Yang et al., 1994; Brondani et al., 2002; Belo, 2001; Pessoa Filho, 2004), análise de genes e locos QTLs (Blair & McCouch, 1997; Koh et al., 1996; Xiao, 1996a); análises de pedigree e melhoramento assistido por marcadores (Ayres et al., 1997; Wooster et al., 1994; Xiao et al., 1996a; Belo, 2001; Brondani et al., 2002).

Marcadores microssatélites têm mostrado grande potencialidade como metodologia de discriminação de genótipos de plantas cultivadas (Yang et al., 1994; Russel et al., 1997). Tais marcadores podem ser utilizados para a determinação da diversidade genética, do sistema de cruzamento, estudos de genética de populações, teste de paternidade e fluxo gênico. O uso de microssatélites na identificação de variedades se deve a sua capacidade de detectar, repetida,

certeira e eficientemente, grandes números de alelos (Smith & Helentjaris, 1996). Os marcadores microssatélites têm sido, também, utilizados com sucesso na determinação de acessos e cultivares puros, na manutenção apropriada (detecção de duplicatas, mistura de sementes, polinização cruzada não esperada e deriva genética), para determinar a relação entre indivíduos ou grupos de acessos (Belo, 2001) e para clarificar a estrutura genética, ou partição da variação entre indivíduos, acessos, populações e espécies, e na determinação da presença de genes ou complexos de genes num acesso particular (Olofowote et al., 1997; Yang et al., 1994; Wu & Tanksley, 1993; Ayres et al., 1997; Blair & McCouch, 1997; Koh et al., 1996; Xiao et al., 1996a; 1996b; McCouch et al. 1997). Um dos problemas da análise de diversidade genética em plantas agrícolas utilizando marcadores moleculares, particularmente microssatélites, é a falha na amplificação de certos alelos. Segundo Robinson & Harris (1999) nunca se sabe se a falha se deve aos alelos nulos e, a não ser que se tenha a certeza, tal falha deve ser considerada como um dado genotípico faltando durante a análise (Warburton & Crossa, 2000) para minimizar erros de interpretação.

1.13. Reações Múltiplas de Microssatélites (Multiplex)

PCR múltiplo ou multiplex é uma variante da PCR na qual dois ou mais locos são simultaneamente amplificados na mesma reação ou, ainda, amplificados em reações distintas, mas genotipados simultaneamente. O desenvolvimento de um sistema multiplex requer: 1) seleção de marcadores de acordo com a variação de tamanho de alelos em cada loco, evitando sobreposição de faixas de alelos em marcadores marcados com o mesmo fluorocromo; 2) PCR de cada marcador sob as mesmas condições de amplificação; 3) PCR múltiplo numa mistura equimolar dos primers; 4) ajuste do tempo e temperatura de extensão; 5) ajuste do tempo e temperatura de anelamento; 6) ajuste da quantidade de marcadores em multiplex; e 7) ajuste da concentração do tampão. Esta técnica tem permitido a ampliação simultânea de vários locos na mesma reação e tornou-se um meio rápido e conveniente para a pesquisa laboratorial (Henegariu et al., 1997). A técnica de PCR múltiplo tem sido também aplicada com sucesso em análises de deleção (Henegariu et al., 1994), mutações (Shuber et al., 1995) e polimorfismos ou em ensaios quantitativos (Mansfield et al., 1994) e PCR-Transcrição Reversa (Crisan, 1994).

A capacidade de montar uma PCR com vários pares de primers em uma única reação é necessária para reduzir o tempo, trabalho e custo da genotipagem (Olivo, et al., 2000). A

otimização das condições da PCR múltipla pode ser entediante e demorada visto que os primers interagem e afetam a quantidade e qualidade dos produtos, produzem bandas inespecíficas, e produzem dímeros de primers (Jordan et al., 1999). Em compensação, a mistura de vários marcadores pode melhorar o rendimento final permitindo a visualização em cada linha de um gel de muitos fragmentos polimórficos de DNA, aumentando a informação produzida pela eletroforese (Armour et al., 2002).

A montagem de painéis de reações múltiplas impõe que microssatélites marcados com a mesma fluorescência possuam variação de tamanho de alelos diferentes e aqueles com faixa de tamanho em sobreposição sejam marcadores com fluorescência diferente (Ziegle et al., 1992). A mistura dos microssatélites pode ser feita após amplificação ou alelos em locos microssatélites podem ser amplificados juntos numa mesma reação de PCR (Mitchell et al., 1997). No caso de reações múltiplas universais, é essencial ter em conta a variação do tamanho de alelo de cada loco com base na análise de acessos de germoplasma, bem como o tipo de marcação fluorescente a ser usada por cada conjunto de marcadores analisados conjuntamente (Blair et al., 2002; Belo, 2001; Pessoa Filho, 2004).

Em arroz, Blair et al. (2002) reportou o uso de 27 marcadores SSR marcados com fluorescência organizados em 4 painéis para análise da diversidade em espécies de *Oryza*. A combinação dos pares de primers foi feita de forma tal a acomodar a análise de cultivares geneticamente diferentes, permitindo um método robusto e flexível para a detecção de variabilidade intra e inter-específica (Coburn et al., 2002). A análise múltipla, no entanto, foi realizada com amplificações PCR de cada marcador individualmente, seguida de misturas específicas antes da eletroforese. Por outro lado, os painéis múltiplos mais interessantes são aqueles que reúnem um bom número de marcadores em uma única reação PCR (Belo, 2001; Pessoa Filho, 2004). Dois painéis múltiplos baseados em reação única de PCR, combinando 5 e 6 locos para análise genética de germoplasma, de arroz foram desenvolvidos por Beló (2001) e um terceiro, mais recentemente, por Pessoa Filho (2004). Para identificação dos locos aos quais pertenciam os alelos evitou-se a sobreposição dos mesmos, através da amplitude alélica que determinou a diferença entre o menor e o maior alelo de cada loco.

O sistema automatizado de determinar o tamanho dos alelos SSR marcados com fluorescência é um sistema em que um dos primers PCR para um loco SSR é marcado com um

marcador fluorescente colorido. Os produtos da PCR carregando diferentes marcadores fluorescentes são separados em gel de poliacrilamida. O uso de microssatélites marcados com fluorescência para a genotipagem automática em seqüenciadores de DNA oferece muitas vantagens em relação à análise usando técnicas de detecção com prata. Uma das vantagens é a possibilidade de analisar vários locos ao mesmo tempo em uma única coluna de gel. Um segundo benefício é o significativo aumento na acurácia na mensuração do tamanho do alelo, obtido pelo uso de um padrão interno em cada coluna de gel e analisado com o uso de algoritmo automatizado (“allele calling”). De um modo geral, a automatização aumenta a velocidade e a acurácia da coleta de dados e o processamento da informação. A alta sensibilidade da detecção também reduz o volume necessário (e, sobretudo, o custo) da reação de PCR e permite a detecção dos locos que apresentam amplificação mais difícil (Mitchell et al., 1997; Coburn et al., 2002). Também elimina o uso de radioatividade na visualização dos alelos.

O uso da análise de fragmentos de restrição automática baseada na fluorescência foi primeiro reportado por Carrano et al., 1989. Este método foi adaptado e melhorado para a análise com microssatélites (Edwards et al., 1991; Ziegel et al., 1992). Os métodos automáticos de genotipagem com SSR estão gradualmente substituindo os sistemas manuais de genotipagem no melhoramento de plantas e na pesquisa genética. Esses métodos facilitam a aplicação eficiente de marcadores microssatélites para o mapeamento “high-throughput” (Rhodes et al., 1998; Ponce et al., 1999), análise de *pedigree* (Lexer et al., 1999), *fingerprinting* de acessos (Carrano et al., 1989) e avaliação da diversidade genética (Diwan & Cregan, 1997; Macaulay et al., 2001). A tecnologia pode melhorar a eficiência da gestão de uma coleção de germoplasma, facilitar a manutenção de “stocks” de sementes puras para os melhoristas, e fornecer a base da proteção da propriedade intelectual (Mitchell et al., 1997).

Marcadores SSR têm uma grande probabilidade de funcionar como uma ferramenta para a identificação de cultivares, combinando a alto conteúdo informativo com a precisão de definição automática de tamanho de alelos. Esta combinação oferece um sistema rápido e padrão para a identificação de genótipos como alternativa à identificação convencional de alelos microssatélites usando outros tipos de marcação (Diwan & Cregan, 1997). Este método de genotipagem tem se mostrado eficiente e econômico (Ziegle et al., 1992; Manfield et al., 1994; Blair et al., 2002).

Em plantas, a técnica de detecção fluorescente de microssatélites tem sido aplicada em

arabidopsis (Bell & Ecker, 1994; Ponce et al., 1999), em soja (Akkaya et al., 1995; Diwan & Cregan, 1997; Morgante & Olivieri, 1993, Cregan et al., 1994, Rongwen et al., 1995, Maughan et al., 1995), em milho (Senior et al., 1996; Smith et al., 1997), *Brassica* sp (Kresovich et al., 1995), em tomate (Bredemeijer et al., 1998), em uva (Thomas & Scott, 1993) e em arroz (McCouch et al., 1997; Akagi et al., 1997; Brondani, 2001, 2002; Beló, 2001; Pessoa Filho, 2004), mostrando que a genotipagem de homozigotos e heterozigotos é rápida e que os SSR são herdados de forma Mendeliana.

1.14. Mapeamento Genético

O desenvolvimento e aprimoramento das técnicas de marcadores moleculares baseados em PCR tem permitido a obtenção de mapas genéticos saturados, mais rápida e economicamente mais viável, evidenciando associação entre marcadores moleculares e QTLs de importância econômica. O sucesso do mapeamento genético envolve cinco etapas importantes: 1) escolha cuidadosa dos parentais, em geral contrastantes para a característica de interesse; 2) obtenção da população segregante, maximizando o potencial de evidenciar desequilíbrio de ligação entre um loco marcador e região genômica associada ao controle da característica de interesse; 3) a genotipagem dos indivíduos que compõem a população segregante; 4) a avaliação fenotípica na população segregante (direta, quando possível e representativa, ou indireta, através da análise da progênie de cada indivíduo); 5) análise estatística da co-segregação marcador/fenótipo.

Vários tipos de populações, tais como populações RILs (linhagens puras recombinantes) (Nakata et al., 2002; Zhuang et al., 2002), populações F2 (Welter, 2003), populações de retrocruzamentos (Brondani et al., 2002), populações duplo-haplóides (Ferreira et al., 1994), são utilizadas na análise e mapeamento de QTLs. A informação genética é maximizada com a utilização de populações segregantes do tipo F2, desenvolvida por autofecundação (ou por intercruzamento, no caso de espécies de polinização cruzada) entre indivíduos F1. Os indivíduos F2 apresentam o dobro de informação genética esperada num indivíduo de uma população de retrocruzamento (RC1).

Outras populações segregantes utilizadas em mapeamentos genéticos são populações F1, populações de retrocruzamento (RC1 ou RC2), desenvolvidos pelo cruzamento da F1 com um dos genitores utilizados no cruzamento inicial, populações de linhagens puras recombinantes (RIL),

desenvolvidas por autofecundações sucessivas de uma planta a partir de indivíduos de uma população F₂. Essas linhagens podem ser utilizadas para o mapeamento porque é essencialmente uma população F₂ imortalizada, com uma ilimitada possibilidade de mapeamento e que também podem ser analisadas para caracteres morfológicos ou quantitativos. Linhagens RIL têm um grande potencial para a análise de caracteres quantitativos porque as réplicas podem ser analisadas usando material genético idêntico. Desta forma, os dados gerados podem ser usados para determinar se marcadores moleculares estão ligados à característica. As sucessivas gerações de recombinação diminuem, no entanto, o desequilíbrio de ligação em relação ao encontrado nas gerações F₂ e retrocruzamento (Tanksley et al., 1991). Populações F₂ ou retrocruzamento são efêmeras, sendo que a fonte de tecido para isolar o DNA se esgotará com o tempo e o mapeamento terá de ser iniciado numa outra população. Populações duplo-haplóides (DH) são populações utilizadas no mapeamento genético porque reúnem as vantagens das populações RILs e das populações de retrocruzamento. Os genótipos dos indivíduos duplo-haplóides são completamente homocigotos. Quando existe dificuldade, quer devido ao longo período de geração quer por depressão, na obtenção de linhagens endogâmicas, pode-se recorrer à análise de progênies resultantes de cruzamentos heterocigotos (Moretzsohn et al., 2001).

Cada tipo de população utilizada no mapeamento apresenta uma proporção de segregação alélica específica nos locos analisados. O conhecimento dessas proporções é importante para determinar se a população está expressando uma segregação de acordo com o esperado (proporções Mendelianas) em cada loco. Se o loco for um marcador co-dominante segregando nos três tipos de populações (retrocruzamento, RIL e F₂), a proporção de segregação esperada será, respectivamente, 1:1, 1:1 e 1:2:1. Se o marcador for dominante, as proporções esperadas serão, respectivamente, 1:1, 1:1 e 3:1.

1.15. “Bulked Segregant Analysis”(BSA)

O método Bulked Segregant Analysis (BSA), ou análise de grupos segregantes, é um procedimento rápido para identificar marcadores ligados em locos de interesse em regiões específicas do genoma (Michelmore et al., 1991). Este método é uma alternativa ao mapeamento molecular saturado que é uma atividade cara, trabalhosa e que demanda muito tempo. O método de BSA surgiu como uma modificação feita por Michelmore e colaboradores à técnica anteriormente proposta (Arnheim et al., 1985), que consistia na análise de misturas de DNA de

indivíduos homocigotos para um determinado loco através da identificação de RFLP em desequilíbrio de ligação de loco associados com doenças em seres humanos (Ferreira & Grattapaglia, 1998).

A Análise de Bulks Segregantes envolve a comparação de duas amostras misturadas de DNA uma população segregante de um determinado cruzamento. Em cada grupo ou mistura de DNA, os indivíduos são idênticos para a característica ou gene de interesse, mas são arbitrários para todos os outros genes. Os dois grupos contrastantes são analisados para identificar marcadores que os distinguem. Os marcadores que são polimórficos entre os grupos estarão potencialmente ligados aos locos que controlam a característica usada para construir os grupos. Esta evidência de ligação gênica entre o marcador polimórfico e o loco alvo é confirmada através de uma análise de segregação de todos os indivíduos da população segregante. Estima-se, desta forma, o valor da frequência de recombinação entre o marcador e o loco genético que controla o caráter (Michelmore et al., 1991; Ferreira & Grattapaglia, 1996).

O método BSA utiliza uma seleção genômica prévia para reduzir o esforço necessário à identificação de ligação gênica, e focaliza a análise diretamente na região do genoma associado à característica de interesse. Os marcadores identificados estão em geral numa janela genética de 25 cM de cada lado do alvo (Ferreira & Grattapaglia, 1998).

Este método tem duas aplicações imediatas no desenvolvimento de mapas genéticos: (1) o desenvolvimento de mapas genéticos detalhados através da análise da segregação de marcadores moleculares selecionados em populações; e (2) a localização rápida de genes que não segregam em populações inicialmente usadas para gerar o mapa genético (Michelmore et al., 1991).

O método BSA tem sido aplicado com sucesso na identificação de genes associados à resistência ao míldio (*Bremia lactucae*) em alface, a *Pseudomonas syringae* em tomate, a *Puccinia graminis* em aveia, e a *Uromyces phaseoli* em feijão (Ferreira & Grattapaglia, 1998). Entretanto a técnica de BSA pode não ser muito efetiva na detecção de QTLs de pequeno efeito, pelo que a análise requer a utilização de um maior número de indivíduos nas populações segregantes para possibilitar uma classificação fenotípica mais eficiente e um maior número de

marcadores moleculares que permitam uma cobertura maior do genoma (Brunelli et al., 2002; Linch & Walsh, 1998).

1.16. Detecção de QTLs

QTLs são locos gênicos que controlam características de herança complexa, ou características quantitativas, geralmente identificados por estratégias de mapeamento genético utilizando marcadores moleculares. As características quantitativas são geralmente afetadas por vários QTLs distribuídos por todo o genoma, com forte interação com o ambiente. O número de QTLs para uma determinada característica é variável. A capacidade de detecção de QTLs é limitada pelo tamanho da população, estringência estatística para detecção dos locos, número de marcadores utilizados no mapeamento e herdabilidade da característica em estudo.

A interação epistática entre QTLs tem sido verificada e a sua natureza poderá ser esclarecida com a utilização de marcadores moleculares e o desenvolvimento de novas populações tais como linhagens isogênicas ou linhagens de substituição. O uso de marcadores moleculares e seus mapas genéticos têm permitido o mapeamento de genes que expressam características monogênicas e poligênicas (Patterson, 1991; Tanksley et al., 1991). Patterson et al., (1988) descreveram a dissecação genética de vários QTLs de tomate em fatores Mendelianos simples, lançando luz à utilização de marcadores de DNA em estudos de várias culturas (de Vicente & Tanksley, 1991; Li et al., 1998; He et al., 1999; Zhuang et al., 2001; Nagata et al., 2002; Zhuang et al., 2002; Wang et al., 2002).

Vários são os testes estatísticos comumente utilizados nos diferentes estudos de mapeamento visando a detecção de QTLs. A avaliação da associação entre marcadores moleculares e QTL pode ser uma análise individualizada, ou seja, centrada na avaliação de um marcador de cada vez (ex. "single marker analysis" ou análise de marcador individualizado por regressão linear), ou uma análise simultânea, onde vários marcadores são utilizados em conjunto na análise (ex. mapeamento de intervalo simples, mapeamento de intervalo composto).

A análise de marcador individualizado por regressão linear (Haley & Knott, 1992) não requer o conhecimento prévio da posição dos marcadores no genoma. Entretanto, segundo Tanksley (1991), existe grande dificuldade em detectar um QTL utilizando este método quando a distância entre o marcador e o QTL é grande. Há ainda a dificuldade de diferenciar QTLs de pequenos efeitos próximos de marcas simples e QTLs de grandes efeitos mais distantes (Lynch & Walsh, 1998). Tanksley (1991) aponta uma solução para essas limitações que é a utilização de um maior número de marcadores cobrindo o genoma numa janela genômica média de aproximadamente 15 cM.

O princípio do mapeamento por intervalo simples é testar a possível presença de QTL em várias posições dentro de um intervalo flanqueado por dois marcadores moleculares (Lander & Bolstein, 1989). Isto pode ser realizado por estimativas de máxima verossimilhança ou por regressão linear. Já o mapeamento por intervalo composto (Zeng, 1994) é uma modificação da análise de marcador individualizado onde as variâncias de outros QTLs no genoma são levadas em consideração na análise ao incluir coeficientes de regressão parcial de outros marcadores no modelo.

1.17. QTLs Associados ao Controle de TGMS em Arroz

Em arroz, a análise de QTLs com marcadores moleculares, baseados em mapas genéticos (Cause et al., 1994; Kurata et al., 1994b; McCouch et al., 1998; Saito et al., 1991; Brondani, 2000) tem sido utilizada na detecção de regiões genômicas associadas com várias características complexas. Este tipo de análise tem sido aplicado às características tais como percentagem de estigma extruso (Li et al., 2001), atraso na senescência da folha (Jiang et al., 2003), altura da planta (Ishimaru et al., 2003), absorção de Na(+) e K(+), controle de tolerância à seca pelas raízes (Li et al., 2001), qualidade do grão (Septiningsih et al., 2001a; Li et al., 2001), produtividade e seus componentes (Septiningsih et al., 2001b; Brondani et al., 2002), entre outras.

O desenvolvimento de pesquisa sobre genes ou QTLs que controlam a fertilidade/esterilidade em arroz levou à identificação e localização de vários QTLs dispersos por vários cromossomos. Dados recentes indicam a existência de cerca de trinta e oito QTLs para características diretamente relacionadas com a fertilidade/esterilidade *sensu lato* em arroz. Estes

QTLs incluem: três regiões associadas ao controle de desenvolvimento embrionário (*esa-1*, *esa-2* e *esa-3*), localizados nos cromossomos 6 e 12 (Liu et al., 2001); dois QTLs para a esterilidade da F2 (*F2GST*), localizados nos cromossomos 1 e 8 (Wang et al., 2002); quatro para a restauração da fertilidade (*qRf-1*, *qRf-5*, *qRf-10-1*, *qRf-10-2*), nos cromossomos 1, 5 e 10; oito para a restauração da fertilidade masculina (*qRf-1*, *qRf-7*, *qRf-10* e *qRf-11*) nos cromossomos 1, 7, 10 e 11 (Zhuang et al., 2001; Xie et al., 2002); dois para a saliência da panícula (*qPEN-1* e *qPEN-4*), nos cromossomos 1 e 4 (Hittalmani et al., 2003), quatro para a esterilidade masculina sensível ao fotoperíodo (*pms1* e *pms1*), nos cromossomos 7 e 12 (Mei et al., 1999), quatro para a fertilidade do pólen (*PF*), nos cromossomos 1, 7 e 10 (Cai & Morishima, 2002) e doze para a fertilidade da espiguetta (*qSF-1*, *SPKFRT-1*, *SPKFRT-4b*, *SPKFRT-6*, *qSF-10*, *qSF-10*, *ste10.2*, *SPKST-1*, *SPKST-2*, *SPKST-5* e *SPKST-12*), nos cromossomos 1, 4, 5, 6, 10 e 12 (Moncada et al., 2001; Lafitte et al., 2002; Nagata et al., 2002; Zhuang et al., 2002).

A complexidade da característica TGMS em arroz impõe um conhecimento profundo da interação dos locos ou genes envolvidos na sua expressão. A detecção desses fatores genéticos é grandemente limitada quando são utilizados métodos tradicionais de melhoramento de plantas. Entretanto, a disponibilidade de estratégias genômicas e técnicas moleculares tem permitido clarificar os mecanismos reguladores de muitas características agronomicamente importantes e influenciadas pelo ambiente. A associação entre a expressão dessas características e os fatores genéticos, envolvidos no seu controle, tem sido possível através do uso de marcadores moleculares. A análise de diversidade genética com marcadores moleculares e a caracterização fenotípica extensiva permitem a correção de distorções e uma melhor correlação entre fenótipo e genótipo.

2.

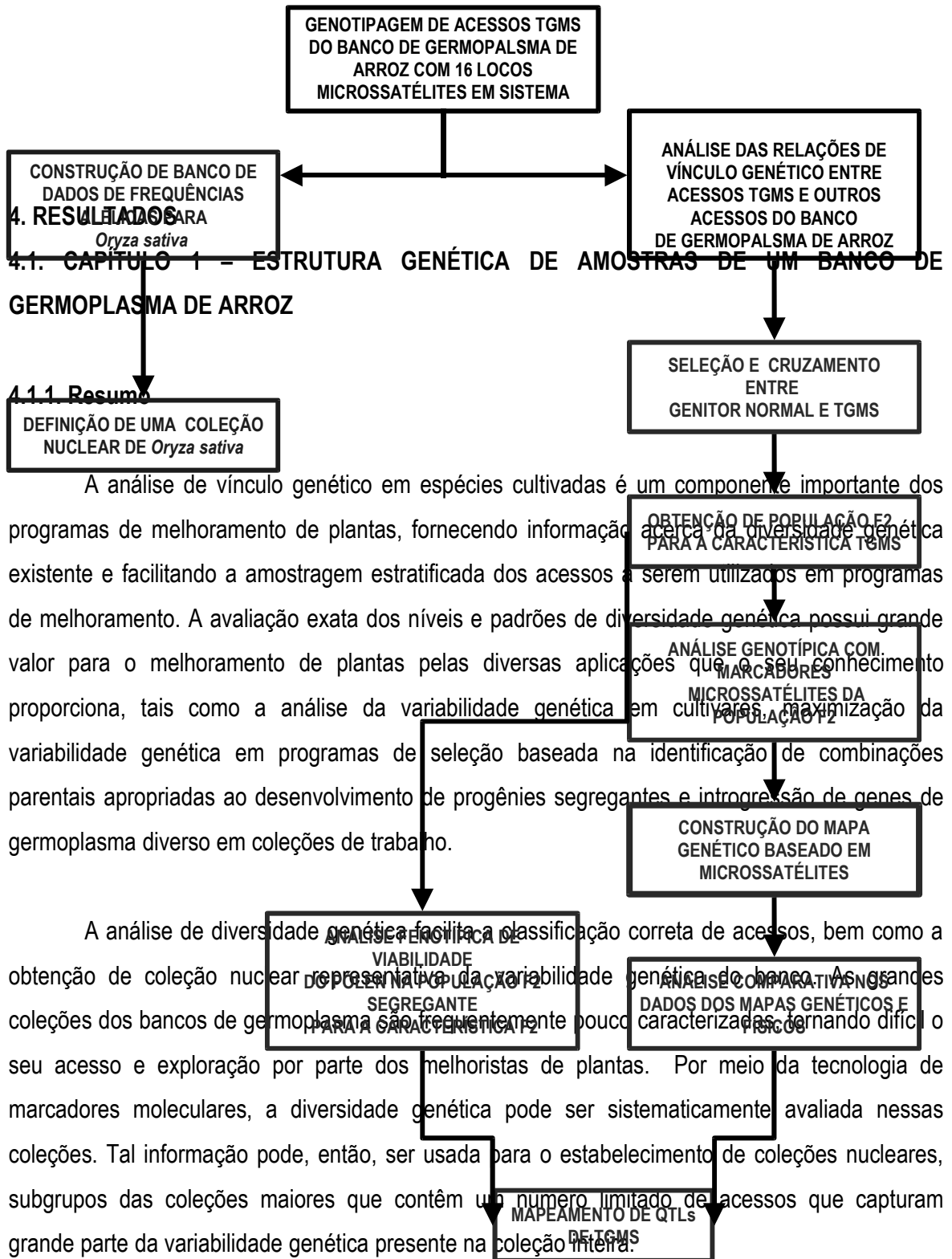
OBJETIVOS

Os objetivos gerais deste estudo são:

1. Avaliação das relações de vínculo genético entre acessos TGMS do Banco de Germoplasma e acessos de arroz das subespécies *indica* e *japonica*;
2. Análise da frequência alélica em locos microssatélites para a subespécie *indica*, e contraste com dados da subespécie *japonica*;
3. Seleção de linhagens normal e TGMS e construção de um mapa genético intraespecífico de arroz baseado em marcadores microssatélites;
4. Mapeamento e análise de QTLs para macho-esterilidade termo-sensível em arroz;

Os objetivos específicos são:

1. Genotipagem automática de uma coleção de germoplasma que inclui acessos TGMS utilizando painéis múltiplos de marcadores microssatélites;
2. Análise de distância genética entre acessos do Banco de Germoplasma de Arroz;
3. Definição de coleção nuclear com base em dados de distância genética para a subespécie *indica*;
4. Desenvolvimento de populações segregantes para a característica TGMS através de um cruzamento intra-específico de linhagens *Oryza sativa* ssp *indica* (normal) x *O. sativa* ssp *indica* (TGMS), selecionadas com base em dados de distância genética;
5. Desenvolvimento mapa genético utilizando marcadores moleculares microssatélites;
6. Mapeamento de QTLs que controlam macho-esterilidade gênica termo-sensível (TGMS) em arroz.



A análise de vínculo genético em espécies cultivadas é um componente importante dos programas de melhoramento de plantas, fornecendo informações acerca da diversidade genética existente e facilitando a amostragem estratificada dos acessos a serem utilizados em programas de melhoramento. A avaliação exata dos níveis e padrões de diversidade genética possui grande valor para o melhoramento de plantas pelas diversas aplicações que o seu conhecimento proporciona, tais como a análise da variabilidade genética em cultivares, a maximização da variabilidade genética em programas de seleção baseada na identificação de combinações parentais apropriadas ao desenvolvimento de progênes segregantes e introgressão de genes de germoplasma diverso em coleções de trabalho.

A análise de diversidade genética facilita a classificação correta de acessos, bem como a obtenção de coleção nuclear representativa da variabilidade genética do banco. As grandes coleções dos bancos de germoplasma são frequentemente pouco caracterizadas, tornando difícil o seu acesso e exploração por parte dos melhoristas de plantas. Por meio da tecnologia de marcadores moleculares, a diversidade genética pode ser sistematicamente avaliada nessas coleções. Tal informação pode, então, ser usada para o estabelecimento de coleções nucleares, subgrupos das coleções maiores que contêm um número limitado de acessos que capturam grande parte da variabilidade genética presente na coleção inteira.

Neste estudo, três painéis de marcadores microsatélites (painéis multiplex) constituídos por marcados marcados com fluorocromos foram utilizados para a genotipagem semi-automática

em seqüenciador de DNA de 298 acessos de arroz, em parte coletados em diferentes áreas do Brasil e conservados no banco de germoplasma. Dezesesseis locos marcadores distribuídos por todo o genoma de arroz foram genotipados utilizando os painéis multiplex e os dados utilizados para o cálculo de distâncias genéticas entre os acessos. Um dendrograma baseado no método de Neighbour-Joining foi utilizado como modelo para definição de agrupamentos e para a inferência de possível estruturação genética da coleção de germoplasma. A análise das relações genéticas desses acessos sugere uma correlação não significativa entre os agrupamentos baseados nos dados de distância genética e de diferenciação das subpopulações baseadas no método Cadeia de Markov e Monte Carlo (MCMC). As estimativas da estatística F de Wright revelaram um alto valor do coeficiente de endogamia (Fis) e um índice de fixação geral (Fis) relativamente alto, mas somente níveis moderados de diferenciação genética (Fst) das subpopulações definidas de acordo com o modelo de agrupamento de distância genética. Uma coleção nuclear baseado nas estimativas de distância genética par-a-par e na riqueza alélica da coleção foi obtida com sucesso, caracterizada por possuir um subgrupo de acessos representando somente 16,66% de toda a coleção enquanto mantém 81,31% da riqueza alélica detectada. Níveis moderados de diferenciação genética foram também observados na coleção nuclear. A caracterização molecular em escala possibilita a análise rápida e eficiente de coleções inteiras de germoplasma, proporcionando a definição de coleções nucleares e a possibilidade de incremento do uso do germoplasma conservado nas coleções de germoplasma de arroz nos programas de melhoramento.

Palavras-chave: microssatélites; diversidade genética; coleção nuclear; genotipagem múltipla.

4.1.2. Introdução

Avaliação correta dos níveis e padrões de diversidade genética pode ser inestimável em melhoramento de plantas devido a aplicações diversas, incluindo (i) análise de variabilidade genética em cultivares (Smith, 1984; Cox et al., 1986), (ii) maximização da variabilidade genética em programas de seleção baseada na identificação de combinações parentais apropriadas ao desenvolvimento de progênies segregantes (Barrett & Kidwell, 1998), e (iii) e introgressão de genes de germoplasma diverso em coleções de trabalho (Brondani et al., 2002). O entendimento das relações genéticas entre acessos pode ser particularmente útil no planejamento de cruzamentos, na inclusão de linhagens em grupos heteróticos específicos, na definição de

linhagens parentais a serem utilizadas no desenvolvimento de novas populações recombinantes ou para permitir a utilização do germoplasma em programas de melhoramento (Ferreira & Grattapaglia, 1998).

O estudo da estrutura genética de populações de plantas é fundamental para a compreensão da sua ecologia e forças evolucionárias que as afetam (Hartl, 1987). Porquanto a estruturação genética reflete o histórico dos alelos trocado entre populações, ela tem grande consequência na composição genética dos indivíduos. Estimativas corretas de diferenciação genética são cruciais para a biologia da conservação, onde é muitas vezes necessário compreender se as populações estão geneticamente isoladas. No caso dos recursos genéticos, o conhecimento da estrutura genética dos acessos depositados nos bancos de germoplasma é essencial para a seleção daqueles que serão úteis para os diferentes propósitos genéticos e melhorísticos, potencializando assim o seu uso.

A análise da estrutura genética das populações ou grupos de acessos tem sido baseada em princípios subjacentes à estatística F de Wright (Wright, 1978; Weir & Cockerham, 1984). Os parâmetros da estatística F hierárquica são utilizados para medir a extensão de endogamia dentro das subpopulações (F_{is}), a extensão da diferenciação genética entre as subpopulações (F_{st}) e a redução média em heterozigosidade de um indivíduo em relação a toda população (F_{it}). Enquanto que esta metodologia tem sido muito utilizada na análise de populações naturais, sua aplicação na análise de bancos de germoplasma tem sido limitada.

O estudo da diversidade genética das coleções de germoplasma, em paralelo com as estimativas de diferenciação genética, pode facilitar a classificação correta dos acessos, o estabelecimento de relações genéticas par-a-par e de grupo, a seleção de amostras representativas que capturem a diversidade genética da coleção e a detecção de padrões de diferenciação em toda a coleção bem como nas amostras retiradas. O polimorfismo de DNA pode ser muito útil na definição de um grupo de acessos (coleção nuclear) que pode geneticamente representar a grande coleção de germoplasma do qual ela é originada. Uma coleção nuclear representa um subgrupo derivado de uma coleção inteira que é composta de um pequeno número de acessos que juntos devem representar o máximo possível da diversidade genética existente em toda a coleção (Frankel & Brown, 1984).

A análise genética molecular, baseada no polimorfismo de DNA detectado em locos marcadores tem sido útil na análise de diversidade genética das coleções de germoplasma. Os microssatélites ou repetições de seqüências simples (Weber & May, 1989; Litt & Luty, 1989; Tautz, 1989) são marcadores moleculares codominantes muito úteis, bem conhecidos pelo seu usual alto conteúdo de informação (Ferreira & Grattapaglia, 1998). Mais de 6.000 marcadores microssatélites foram desenvolvidos para a pesquisa em arroz, proporcionando uma grande oportunidade de aplicação em diferentes estudos genéticos e nos esforços de melhoramento. Marcadores microssatélites têm sido utilizados para exploração do genoma do arroz em temas tais como mapeamento genético de características economicamente importantes (Yamamoto et al., 1998; Brondani et al., 2002; Zhang et al., 2001; Sasaki et al., 2002), estudos de populações (Barbier 1989b; Gao et al., 2000a, b; Gao et al., 2001), análise de germoplasma (Glaszmann 1987; Ni et al., 2002; Parsons et al. 1999; Garris et al., 2003; Gao et al., 2000a, b; Gao et al., 2001) e avaliação do nível e estrutura de diversidade genética em cultivares de interesse e em populações naturais de *O. sativa* e outros parentes silvestres (Olufowote et al., 1997; Yang et al., 1994; Beló, 2001; Gao et al., 2000a, b; Gao et al., 2002; Ni et al., 2002; Garris et al., 2003, 2005). A PCR combinada de múltiplos marcadores numa única reação baseada em painéis de marcadores marcados com fluorescência aumenta grandemente a capacidade de genotipagem semi-automática de um grande número de acessos, tais como aqueles conservados em coleções de germoplasma. Esta estratégia tem sido confirmada como rápida e altamente informativa na caracterização de recursos genéticos de arroz (Beló, 2001; Pessoa Filho, 2004).

Neste estudo, três painéis combinando dezesseis marcadores microssatélites marcados com fluorescência foram usados para a genotipagem semi-automática de uma coleção de germoplasma e para a análise de sua estrutura genética. Estimativas par-a-par de distância genética e parâmetros de diversidade genética foram usados para o estabelecimento de relações genéticas entre os acessos. Os modelos de estruturação de populações baseados nas relações genéticas foram então testadas com estimativas do da estatística F. Uma coleção nuclear foi extraída a partir da coleção inteira baseada nos dados de distância genética e nas estimativas de riqueza alélica. A diferenciação genética foi então avaliada na coleção nuclear.

4.1.3. Material e Métodos

4.1.3.1. Material Biológico e Extração de DNA

Uma amostra de 298 acessos de arroz (Tabela 1), na sua maioria composta de variedades tradicionais coletadas em território brasileiro, foi avaliada neste estudo. Esta amostra incluiu também nove linhagens que apresentam a característica de macho-esterilidade gênica termo-sensível (TGMS) (TGMS70977_12, TGMS70977_26, TGMS68940, TGMS70987, TGMS68935, TGMS71018, TGMS70989, TGMS68944 e TGMS68945) e várias variedades de arroz comumente utilizadas em programas de melhoramento. Os acessos de *indica* IRGA 417 (ARR1) e IRGA422 (ARR3) foram utilizados em todos os géis de análise de tamanho dos fragmentos para controlar a variação de amplitude alélica entre cada corrida de eletroforese.

Cerca de 150 mg de folhas de pelo menos 10 indivíduos de cada acesso foram utilizadas para a extração de DNA baseada no protocolo CTAB (Ferreira & Grattapaglia, 1998). As folhas foram secas em tubos de eppendorf de 1,5 ml utilizando nitrogênio líquido e maceradas dentro dos tubos, seguida da adição de 700 uL de tampão de extração pré-aquecido (2% acetyldimethyl tiethylammonium bromide – CTAB, 1.4 M NaCl, 20 mM EDTA, 100 mM Tris-HCL pH=8.0, 1% polyvinylpyrrolidone, 0.2% 2-mercaptoethanol). A solução foi incubada por 40 min a 60°C com agitação ocasional. Então, 600 uL de clorofórmio: álcool isoamílico (24:1) foi adicionado em cada amostra, seguida de centrifugação a 12.000 rpm durante 10 min. A fase aquosa foi transferida para um novo tubo, e o DNA foi precipitado com 0.6 volumes de isopropanol frio. Uma centrifugação a 7.500 rpm durante 5 min possibilitou a obtenção do “pellet” de DNA e o descarte do sobrenadante. O “pellet” foi lavado num processo de duas etapas utilizando 1 mL etanol frio a 70% frio e etanol a 100% e secado numa centrífuga à vácuo (Centrivap Concentrator - LABCONGO) durante 10 min. O “pellet” foi então eluído com 5 uL de tampão TE (10mM Tris-HCl, ph=8.0, 1 mM EDTA) contendo RNase (10 ug/mL) e colocado a 37°C durante 120 min para digestão com RNA. A concentração de DNA foi estimada em eletroforese de gel de agarose 1% utilizando 10 ng de λ DNA (GibcoBRL) como padrão. O DNA foi diluído em tampão TE para a concentração final de 3ng/uL.

Tabela 6 – Acessos de arroz do Banco de Germoplasma de Arroz (Embrapa, Brasil) utilizadas na análise.

#	#GEL	# BANCO GERMOPLASMA	NOME COMUN	ANO DE COLETA	FONTE
1	AC1480	CA880070	3 MESES BRANCO	-	CNPAF
2	571_AL	CA960010	90 DIAS	2000	AL
3	298_RO	CA830032	A.BOLINHA/CATETINHO	1984	RO
4	A1729B	-	A1729B	-	CNPAF

5	A1744B	-	A1744B	-	CNPAF
6	A1802C	-	A1802C	-	-
7	A1809B	-	A1809B	-	-
8	A1860B	-	A1860B	-	-
9	AC1802b	-	AC1802b	-	-
10	ACC5	CA780002	AGULHA (1)	-	CNPAF
11	ACC310	CA790064	AGULHA (2)	-	CNPAF
12	ACC516	CA790300	AGULHA (2A)	-	CNPAF
13	AC1012	CA830126	AGULHA (3)	-	CNPAF
14	538_MT	CA880075	AGULHA DA TERRA	1989	MT
15	107_CE	CA780283	AGULHA DOURADO	1983	CE
16	ACC222	CA780342	AGULHINHA	-	CNPAF
17	ACC543	CA790326	AGULHINHA	-	CNPAF
18	AC878	CA820073	AGULHINHA (6)	-	CNPAF
19	AC235	CA780366	AGULHINHA (8)	-	CNPAF
20	482_MG	CA870120	AGULHINHA BRANCO	2000	MG
21	AC1257	CA860048	AGULHINHA-DO-BREJO	-	CNPAF
22	AC1928	BAG 94	AKITAKOMACHI	-	EPAGRI
23	AC1193	CA850050	AMARELAO	-	CNPAF
24	AC205	CA780313	AMARELAO PRECOCE	-	CNPAF
25	573_AL	CA960017	AMARELINHO	2000	AL
26	576_AL	CA960029	ANÃO	2000	AL
27	AC202	CA780313	ANAO (5)	-	CNPAF
28	ACC105	CA780134	ANAO (8)	-	CNPAF
29	AC189	CA780285	ANAO-DO-FIM	-	CNPAF
30	577_AL	CA960030	ANAOZINHO	2000	AL
31	-	-	ARR1	-	CNPAF
32	-	-	ARR3	-	CNPAF
33	AC1420	CA870158	ARROZ 4 MESES	-	CNPAF
34	AC1426	CA870168	ARROZ 51	-	CNPAF
35	AC1439	CA870191	ARROZ 9 ANOS	-	CNPAF
36	AC913	CA830002	ARROZ AMARELO	-	CNPAF
37	294_SC	CA830003	ARROZ AMARELO E BRANCO	1984	SC
38	549_MG	CA890001	ARROZ ANÃO	2000	MG
39	277_ES	CA810039	ARROZ BARRIGA BRANCA	1997	ES
40	491_GO	CA870153	ARROZ DA TERRA	1999	GO
41	51_MA	CA780157	ARROZ DE DEUS OU ARROZ JUDIANO	1983	MA
42	377_MS	CA850023	ARROZ DE MAIO	2000	MS
43	CC802A	CA810060	ARROZ MINEIRO	-	CNPAF
44	433_GO	CA870007	ARROZ PELUDO	1996	GO
45	689_GO	CNA0005564	ARROZ PIAUI	2000	GO
46	AC1809	GEN 1239	ARROZ PRETO	-	CNPAF
47	AC1218	CA860003	ARROZ PRETO (4)	-	CNPAF
48	493_MG	CA870162	ARROZ TRÊS MESES	1999	GO
49	AC1971	PB08	ARROZ VERMELHO	-	CPAMN
50	AC1964	PB01	ARROZ VERMELHO (1)	-	CPAMN
51	AC1965	PB02	ARROZ VERMELHO (2)	-	CPAMN
52	AC1969	PB06	ARROZ VERMELHO (3)	-	CPAMN
53	AC263	CA790003	ARROZ VERMELHO (5)	-	CNPAF

54	AC1967	PB04	ARROZ VERMELHO (7)	-	CPAMN
55	AC1974	PB11	ARROZ VERMELHO (8)	-	CPAMN
56	AC1975	PI01	ARROZ VERMELHO (9)	-	CPAMN
57	AC1968	PB05	ARROZVERMELHO (6)	-	CPAMN
58	AC1976	PE01	ARROZVERMELHO4	-	CPAMN
59	ACC338	CA790098	BACABA	-	CNPAF
60	257_MG	CA800128	BARRIGA BRANCA	1994	MG
61	A1722a	-	BASMATI 370	-	CNPAF
62	BG902	-	BG90-2	-	CNPAF
63	AC1409	CA870142	BICO GANGA	-	CNPAF
64	AC1811	RG033	BILLCLINTON	-	IRGA
65	AC1813	RG037	BLUE BONNET 50	-	IRGA
66	AC1812	RG036	BLUEBELLE	-	IRGA
67	AC1813	RG037	BLUEBONNET 502	-	IRGA
68	46_MA	CA780148	BRANCO PRECOCE	1983	MA
69	AC1345	CA870040	BRANCO-DE-BREJO	-	CNPAF
70	ACC250	CA780391	BRANQUINHO	-	CNPAF
71	ACC17	CA780017	BRANQUINHO 90 DIAS	-	CNPAF
72	662_CLB	CNA0004482	BRASILEIRO	1983	CLB
73	BRIRGA	-	BR-IRGA409	-	CNPAF
74	AC1814	RG047	BRS-BOJURU	-	IRGA
75	ACC403	CA790168	BURITI	-	CNPAF
76	62_RS	CA780180	CACHINBO DO SECO	1983	RS
77	AC1816	RG053	CACHINHO	-	IRGA
78	59_RS	CA780171	CACHINHO	1983	RS
79	625_BZL	CNA0000982	CALORO	2000	BZL
80	AC1817	RG055	CALORO 7985	-	IRGA
81	615_RS	CNA0000777	CALOURO	2000	RS
82	574_AL	CA960020	CANARINHO	2000	AL
83	AC1436	CA870186	CANAROXIA	-	CNPAF
84	318_RO	CA830125	CANELA CURTA / PINRABNAO	1984	RO
85	ACAC73	CA780091	CANELA-DE-ACO	-	CNPAF
86	ACC78	CA780099	CANELA-DE-FERRO	-	CNPAF
87	ACC577	CA790376	CANELA-DE-FERRO (6)	-	CNPAF
88	670_CLB	CNA0004573	CANILLA	1983	CLB
89	673_CLB	CNA0004600	CANUTO	1983	CLB
90	AC1818	RG057	CAPI-93	-	IRGA
91	AC48	CA780063	CAQUI	-	CNPAF
92	554_PB	CA910002	CAQUI	1996	PB
93	AC1819	RG059	CARNAROLI	-	IRGA
94	AC208	CA780319	CAROLINA (1)	-	CNPAF
95	AC462	CA790237	CAROLINA (2)	-	CNPAF
96	569_AL	CA960006	CAROLINA BRANCO	2000	AL
97	695_GO	CNA0006548	CATALAO 101	1994	GO
98	44_MA	CA780139	CATARINA	1983	MA
99	ACC449	CA790222	CATETO	-	CNPAF
100	575_AL	CA960027	CHATINHO	2000	AL
101	668_CLB	CNA0004546	CHATO RAYADO	1983	CLB

102	AC1168	CA850015	CHIFRE-DE-VEADO	-	CNPAF
103	666_CLB	CNA0004503	CHILENO	1983	CLB
104	ACC87	CA780109	CHILILICA	-	CNPAF
105	AC1820	RG061	CHINSEIASAHI	-	IRGA
106	AC1806	GEN 1236	CHIQUNHO	-	CNPAF
107	CICA 8	-	CICA8	-	CNPAF
108	AC1822	RG068	CINIA 900	-	IRGA
109	AC1823	RG069	CINIA 937	-	IRGA
110	AC1734	-	CIWINI	-	CNPAF
111	AC719	CA800146	COME-CRU	-	CNPAF
112	ACC295	CA790044	COMECRU (13)	-	CNPAF
113	AC55	CA780071	COMUM	-	CNPAF
114	AC67	CA780084	CUTIAO BRANCO	-	CNPAF
115	ACC307	CA790061	CUTIAO VERMELHO	-	CNPAF
116	AC79	CA780101	DEABRIL (1)	-	CNPAF
117	AC627	CA800028	DEABRIL (2)	-	CNPAF
118	AC1824	RG225	DELLA	-	IRGA
119	AC1424	CA870166	DESCONHECIDO	-	CNPAF
120	AC1040	CA840032	DESCONHECIDO (4)	-	CNPAF
121	AC1890	RG640	DESCONHECIDO12 (OKAMINORI 2)	-	IRGA
122	AC1825	RG226	DIAMANTE	-	IRGA
123	AC1729	-	Diamante	-	CNPAF
124	AC1826	RG227	DIAMANTE 1078	-	IRGA
125	AC1827	RG229	DOBLE CAROLINA	-	IRGA
126	270_ES	CA810016	DOIDÃO	1998	ES
127	669_CLB	CNA0004560	DONATO	1983	CLB
128	AC59	CA780075	DOURADINHO	-	CNPAF
129	AC1828	RG230	DREW	-	IRGA
130	AC1829	RG231	DULAR 22169-76	-	IRGA
131	AC1830	RG232	DULAR CNA1010	-	IRGA
132	AC1536	CA950029	EC1	-	CNPAF
133	AC1831	RG233	EEA201	-	IRGA
134	AC1832	RG234	EEA301	-	IRGA
135	AC1833	RG235	EEA304	-	IRGA
136	AC1834	RG236	EEA401LC	-	IRGA
137	AC1835	RG237	EEA404	-	IRGA
138	AC1836	RG238	EEA405	-	IRGA
139	AC1837	RG239	EEA406	-	IRGA
140	AC1838	RG240	EEA406 (MBL)	-	IRGA
141	AC1839	RG243	ELONI	-	IRGA
142	A1580a	CA980002	EMGOPINHA	-	CNPAF
143	21_AL	CA780055	ESAV-36	1994	AL
144	AC1555	CA960024	ESAVE	-	CNPAF
145	AC1840	RG249	EUROSE	-	IRGA
146	F1	-	F1	-	CNPAF
147	AC1841	RG250	FANNY21501-24	-	IRGA
148	AC1842	RG251	FANNY22193-	-	IRGA
149	AC1512	CA940007	FARROUPILHA	-	CNPAF
150	AC1843	RG253	FARROUPILHA10	-	IRGA

151	AC1379	CA870097	FERRAO PRETO	-	CNPAF
152	AC1370	CA870086	FERRAO PRETO (3M)	-	CNPAF
153	AC1330	CA870021	FERRUJAO	-	CNPAF
154	AC1844	RG270	FRANCES	-	IRGA
155	ACC130	CA780176	FRONTEIRA	-	CNPAF
156	AC1845	RG271	FUKU60	-	IRGA
157	AC1846	RG272	FUKUBOZU	-	IRGA
158	99_CE	CA780266	GAVIAO	1983	CE
159	AC1318	CA860138	GUAIRA	-	CNPAF
160	AC1116	CA840131	GUAIRA (4)	-	CNPAF
161	AC1343	CA870042	GUAIRA BRANCO	-	CNPAF
162	AC21	CA780022	GUAPINHA	-	CNPAF
163	AC1847	RG280	GULFMONT	-	IRGA
164	AC1847	RG280	GULFMONT (2)	-	IRGA
165	ACC41	-	HIBRIDO	-	-
166	268_ES	CA810003	HÍBRIDO	1997	ES
167	AC616	CA800013	HIBRIDO (6)	-	CNPAF
168	AC1850	RG287	HINOHKARI	-	IRGA
169	AC665	CA800071	HONDURAS	-	CNPAF
170	AC1738	-	HUAN-SEM-GO	-	CNPAF
171	ACC355	CA790117	IAC	-	CNPAF
172	AC1769	-	IAC 202	-	CNPAF
173	AC1768	-	IAC1246	-	CNPAF
174	AC1768	-	IAC165	-	CNPAF
175	AC1851	RG291	IACUBA18	-	IRGA
176	AC1852	RG292	IACUBA19	-	IRGA
177	AC1853	RG293	IACUBA20	-	IRGA
178	AC1855	RG295	ICTAQUIRIGUA	-	IRGA
179	AC31	CA780036	IGUAPE	-	CNPAF
180	AC1849	RG296	ILLABONG	-	IRGA
181	AC1859	RG300	INIA P10	-	IRGA
182	AC1860	RG301	INIA P11	-	IRGA
183	AC1856	RG297	INIACUARÓ	-	IRGA
184	AC1857	RG298	INIATACUARI	-	IRGA
185	AC1723	-	IR36	-	CNPAF
186	IR50	-	IR50	-	CNPAF
187	163_RS	CA780409	ITAQUI	1994	RS
188	AC1862	RG588	JACKSON	-	IRGA
189	274_ES	CA810030	JAGUARÃO OU TAQUARÃO	2000	ES
190	AC196	CA780296	JAPONES	-	CNPAF
191	ACC126	CA780169	JAPONES CLARO	-	CNPAF
192	AC700	CA800122	JAPONES-DA-VARZEA	-	CNPAF
193	578_AL	CA960033	JAQUARI	2000	AL
194	685_GO	CNA0004799	JARAGUA	1983	GO
195	AC1865	RG592	JEFFERSON	-	IRGA
196	AC1805	GEN 1235	JOÃO PACU	-	CNPAF
197	AC1866	RG593	JODON	-	IRGA
198	AC1867	RG594	JUMA 62	-	IRGA

199	AC1867	RG594	JUMA624	-	IRGA
200	AC1858	RG595	K39-96-1-1-1-2	-	IRGA
201	AC1868	RG596	KAMEJI	-	IRGA
202	AC1869	RG597	KANTO 51	-	IRGA
203	AC1872	RG601	KULINAR(GM)	-	IRGA
204	ACC285	CA790031	L.CORADIM (1484)	-	CNPAF
205	ACC350	CA790110	L.CORADIN (1560)	-	CNPAF
206	ACC553	CA790332	LAGEADO (1)	-	CNPAF
207	ACC850	CA820041	LAGEADO (2)	-	CNPAF
208	AC1038	CA840029	LAGEADO (3)	-	CNPAF
209	AC1135	CA840158	LAGEADO (4)	-	CNPAF
210	AC251	CA780392	LAGEADO (51)	-	CNPAF
211	AC366	CA790128	LAGEADO (62)	-	CNPAF
212	334_PI	CA840042	LAGEADO LIGÍTIMO	1996	PI
213	AC1876	RG609	LAGRUE	-	IRGA
214	AC1878	RG612	LEBONNET	-	IRGA
215	AC1484	CA880078	LEVANTA TESTA	-	CNPAF
216	345_PI	CA840069	LIGEIRINHO DESCONHECIDO	1985	PI
217	674_CLB	CNA0004601	LIGERITO	1983	CLB
218	324_PI	CA840022	MACABA MIUDO	1985	PI
219	ACC614	CA800011	MACARICO	-	CNPAF
220	ACC86	CA780108	MAIO	-	CNPAF
221	ACC210	CA780321	MANGOTE	-	CNPAF
222	AC63	CA780079	MARANHAO	-	CNPAF
223	234_MG	CA800041	MARELINHO/PAGA DÍVIDA	1994	MG
224	MET11B	-	METICA	-	CNPAF
225	AC1884	RG625	MIARA	-	IRGA
226	AC1743	-	MINAMIHATAMOCHI	-	CNPAF
227	579_AL	CA960036	MINEIRO	2000	AL
228	683_FLP	CNA0004763	MIRITI	1983	FLP
229	AC1453	CA880025	MISTURA	-	CNPAF
230	AC1929	BAG 98	MOCHIGOME	-	EPAGRI
231	ACC253	CA780396	MOGI	-	CNPAF
232	AC1885	RG626	MOLO	-	IRGA
233	AC1009	CA830122	MONTANHA	-	CNPAF
234	110_CE	CA780288	MOURIN AGULHA	1983	CE
235	55_MA	CA780166	MULATINHO	1983	MA
236	235_MG	CA800049	MUNDICERA	2000	MG
237	255_MG	CA800214	NANICO	1994	MG
238	AC1886	RG633	NIPPONBARE	-	IRGA
239	AC1742	-	NOURINMOCHI	-	CNPAF
240	AC1888	RG638	NP125	-	IRGA
241	AC1889	RG639	OBANASAWA (6)	-	IRGA
242	AC197	CA780297	OITENTAO	-	CNPAF
243	AC1891	RG641	OKAMINORI (3)	-	IRGA
244	AC1892	RG642	OOTORI	-	IRGA
245	AC1752	-	Oryzica1	-	CNPAF
246	AC1893	RG650	OU188	-	IRGA
247	AC628	CA800029	PAGA DIVIDA (L.HIPOTECA)	-	CNPAF

248	AC427	CA790196	PALHA MURCHA	-	CNPAF
249	AC626	CA800027	PANCHOLINA	-	CNPAF
250	AC1506	CA930003	PARAZINHO PALHA	-	CNPAF
251	AC879	CA820074	PAULISTA	-	CNPAF
252	47_MA	CA780143	PIRRACA	1983	MA
253	AC1898	RG669	PUNTAL(ÉFFEM)	-	IRGA
254	AC1902	RG678	QUILLA166002	-	IRGA
255	680_CLB	CNA0004637	RABO DE YEGUA	1983	CLB
256	AC372	CA790134	RABO-DE-BURRO	-	CNPAF
257	ACC83	CA780105	RECHORO	-	CNPAF
258	ACC129	CA780175	REETZ	-	CNPAF
259	AC1137	CA840160	REXORO	-	CNPAF
260	AC1905	RG682	RIKUTONORIN (11)	-	IRGA
261	AC1906	RG683	RIKUTOSHINRIKI	-	IRGA
262	AC1923	BAG 54	ROXO	-	EPAGRI
263	661_FLP	CNA0004319	SALUMPIKIT	1983	FLP
264	ACC99	CA780125	SANTA CATARINA	-	CNPAF
265	228_MG	CA800001	SANTA CATARINA	2000	MG
266	AC1907	RG691	SAN-YANG-AI (1)	-	IRGA
267	158_BA	CA780397	SAQUAREMA	1984	BA
268	AC1034	CA840018	SAQUAREMA (6)	-	CNPAF
269	AC1810	GEN 1240	SEUANTONIO	-	CNPAF
270	AC1910	RG700	SINALOAA80	-	IRGA
271	AC1912	RG717	TA-POO-CHO-Z	-	IRGA
272	40_MA	CA780128	TAQUARAO	1983	MA
273	AC1913	RG718	TASHIMINORI	-	IRGA
274	AC1916	RG721	TEXMONT	-	IRGA
275	-	-	TGMS68935	-	CNPAF
276	-	-	TGMS68940	-	CNPAF
277	-	-	TGMS68944	-	CNPAF
278	TGMS68945	-	TGMS68945	-	CNPAF
279	-	-	TGMS70977_12	-	CNPAF
280	-	-	TGMS70977_26	-	CNPAF
281	-	-	TGMS70987	-	CNPAF
282	-	-	TGMS70989	-	CNPAF
283	-	-	TGMS71018	-	CNPAF
284	AC1808	GEN 1238	TIRIRICA (11)	-	CNPAF
285	ACC81	CA780103	TIRIRICA (13)	-	CNPAF
286	AC1744	-	TomoeMochi	-	CNPAF
287	AC54	CA780070	TORO	-	CNPAF
288	24_PB	CA780061	TRES POTES	1983	PB
289	43_MA	CA780136	VASSOURINHA	1983	MA
290	236_MG	CA800050	VENES BRANCO	1994	MG
291	39_MA	CA780127	VENEZ ROXO	1983	MA
292	AC1566	CA960041	VERMELHAÇÃO	-	CNPAF
293	AC1580	CA980001	VERMELHO	-	CNPAF
294	AC1918	RG730	XIANGJING834	-	IRGA
295	ACC365	CA790127	ZEBU	-	CNPAF

296	AC1919	RG734	ZENITH	-	IRGA
297	AC1861	RG736	ZHAOTONMAXAIGU	-	IRGA
298	AC1920	RG737	ZHEN-GUI-AI-1	-	IRGA

4.1.3.2. Genotipagem utilizando marcadores fluorescentes

Três painéis múltiplos compostos de dezesseis marcadores microssatélites marcados com fluorocromos, desenvolvidos por Bélo (2000) e Pessoa Filho (2004), foram utilizados neste estudo (Tabela 2). Amplificações simultâneas por PCR foram realizadas num volume final de 15 uL contendo 6 ng de DNA genômico, 0.4 mM de cada dNTP, 0.2% ug/uL BSA, 3 mM MgCl₂, e 2 U Taq DNA Polimerase (Phoneutria-Brazil). Para o Painel múltiplo A, composto de 5 marcadores, as concentrações dos “primers” foram 0.2 uM (OS19 e RM248) e 0,13 uM (RM252, RM224 e OG44); para o Painel múltiplo B, composto de 6 marcadores, as concentrações dos “primers” foram 0.13 uM (OG101, OG05 e OG81), 0.2 uM (OG106), 0.23 uM (OG61) e 0.1 uM (RM263); e para o Painel C, composto de 5 marcadores, as concentrações dos “primers” foram 0.13 uM (RM335, RM420, RM418, RM259 e RM475). As reações foram realizadas em um termociclador GeneAmp 9600 (Perkin-Elmer) utilizando o seguinte programa de amplificação: (i) um “hot start” de 94°C durante 5 min, (ii) 30 ciclos de amplificação distribuídos em 1 min a 94°C, 1 min a 52°C, 2 min a 72°C, e (iii) uma etapa de extensão final de 7 min a 72°C. Uma alíquota de 5 uL de produto amplificado foi misturada com 3 uL de tampão de carregamento (98% formamida, 10 mM EDTA-blue dextran) e 2uL de um padrão de tamanho conhecido (Brondani & Grattapaglia, 2001), seguido de desnaturação a 95°C durante 5 min. Um total de 1,5 uL da amostra foi carregado em géis de poliacrilamida “Long Ranger” a 4% em tampão TBE 1X TBE durante 2.5 horas com potência constante de 30W num seqüenciador automático de DNA ABI Prism 377 (Applied Biosystems).

Tabela 7 – Marcadores microssatélites marcados com fluorescência utilizados nos três painéis múltiplos.

PAINEL	LOCO	FLUOROCROMO	COR	VARIAÇÃO DE TAMANHO ESPERADA*	CROMOSSOMO	MOTIF
A	RM252	TET	Green	194-274	4q	(ag)n
A	RM224	TET	Green	124-162	11q	(aag)8(ag)13g (ag)
A	OG44	6-FAM	Blue	152-172	3q	(aag)8(ag)22(gt) (gc)6
A	RM248	6-FAM	Blue	72-108	7q	(ag)n
A	OS19	HEX	Yellow	174-198	6p	(ag)n
B	OG101	TET	Green	95-161	2p	(ag)n
B	OG106	TET	Green	178-254	9p	(ag)n
B	OG81	6-FAM	Blue	71-89	1	(ag)n
B	RM263	6-FAM	Blue	147-195	2q	(ag)n
B	OG05	HEX	Yellow	170-194	3p	(ag)n
B	OG61	HEX	Yellow	96-154	5q	(ag)n
C	RM335	6-FAM	Blue	104-156	4p	(ctt)25
C	RM420	6-FAM	Blue	183-201	7q	(aaat)7
C	RM418	6-FAM	Blue	253-311	7q	(att)21
C	RM259	NED	Yellow	155-176	1p	(ct)17
C	RM475	NED	Yellow	345-387	3q	(attg)30

*tamanho esperado do alelo, em pares de bases, de acordo com Pessoa Filho (2004).

4.1.3.3. Análise dos dados

A determinação do tamanho dos fragmentos microssatélites foi realizada utilizando o programa GeneScan versão 3.1.2. (Applied Biosystems). Os tamanhos dos alelos amplificados foram determinados utilizando o programa Genotyper versão 2.5.2 (Applied Biosystems) baseado nos tamanhos dos fragmentos padrão internos em cada pista do gel. A chamada dos alelos foi realizada por arredondamento dos valores atribuídos aos genótipos para valores próximos do número de pares de bases inteiro para resultar numa estimativa de par de bases para cada alelo (ANEXO II). Como a maioria dos locos utilizados neste estudo possui repetições dinucleotídicas, o processo de chamada resultou algumas vezes em valores intermediários para os alelos indicados. Uma correção foi efetuada de tal forma que todos os valores seguem o tamanho esperado para

locos com motivo dinucleotídeo. Para esse fim, o fragmento alélico mais freqüente de cada loco foi considerado como referência para os valores esperados dos outros alelos no loco marcador.

4.1.3.4. Análise de diversidade genética

As freqüências alélicas, o número total de alelos observados, a amplitude alélica por loco, o número de alelos diferentes por loco, a diversidade genética observada (GDo) e a diversidade genética esperada (GDe) foram calculadas utilizando o programa Cervus versão 2.0 (Marshall et al., 1998). As medidas de diversidade genética esperadas foram obtidas seguindo a estimativa *unbiased* de Nei, que é uma reflexão da diversidade alélica e suas freqüências, com uma correção para o número de indivíduos amostrados, dada pela fórmula $[GDe] = (1 - \sum p_i^2) / (2n - 1)$, onde p_i é a freqüência do *i*-ésimo alelo para cada loco e n o número de amostras analisadas (Nei, 1987). O conteúdo de informação polimórfica (PIC) foi também estimado como a probabilidade de um indivíduo ser informativo em relação à segregação dos seus alelos herdados (Botstein et al., 1980). A probabilidade de genótipos idênticos, definido como $PI = \sum p_i^4 + \sum (2p_i p_j)^2$ foi inicialmente estimada para cada loco selecionado, e depois, para todos os dezesseis locos (Sefc et al., 1999).

As distâncias genéticas entre os acessos de arroz baseadas no polimorfismo dos marcadores microssatélites foram estimados pelo método "Shared Allele Distance" em comparações par-a-par. As estimativas foram baseadas na soma da proporção de alelos comuns entre dois acessos examinados em cada loco dividida pelo dobro do número de locos testados (Bowcock et al., 1994; Goldstein et al., 1995). A matriz diagonal de distância genética foi calculada utilizando o parâmetro $[-\ln(Ps)]$ da Calculadora de Distância Genética (<http://www2.biology.ualberta.ca/jbrzusto/sharedst.php>). A matriz diagonal foi então submetida a análise de agrupamento utilizando o Método de Neighbour-Joining (Saitou & Nei, 1987) realizado pelo programa NTSYSpc versão 2.10 (Rohlf, 1998). Gráficos de plotagem bi- e tri-dimensional foram obtidos utilizando os procedimentos Dcenter e Eigen no NTSYSpc (Rohlf, 1998) para melhor visualizar a distribuição dos acessos.

4.1.3.5. Análise de estrutura genética da coleção de germoplasma

Como finalidade de analisar a estrutura genética da coleção de germoplasma, os principais agrupamentos de acessos definidos pela análise de distância genética foram avaliados como subpopulações. As estimativas foram calculadas utilizando a metodologia de *bootstrapping* do

programa GENETIX versão 4.05 (Belkhir, 2001). Os parâmetros estatísticos, que se seguem, foram calculados (Weir & Cockerham, 1984): o coeficiente de endogamia (F_{IS}) foi calculado por $F_{IS}=(H_s-H_i)/H_s$ e utilizado para estimar a redução mínima em heterozigosidade de um indivíduo devido ao cruzamento aleatório dentro da população. O índice de fixação foi calculado por $F_{ST}=(H_T-H_s)/H_T$ e utilizado para estimar a extensão da diferenciação genética entre as subpopulações (agrupamentos). O índice de fixação geral (F_{IT}) foi calculado por $F_{IT}=(H_T-H_i)/H_T$ e utilizado para estimar a redução média de heterozigosidade de um indivíduo em relação à população total (coleção de germoplasma). Os acessos da coleção foram subdivididos de acordo aos resultados da análise de Neighbour-Joining: (a) dois agrupamentos principais e (b) nove agrupamentos.

A estrutura genética da coleção de germoplasma foi também analisada pelo programa STRUCTURE versão 2.1 (Pritchard et al., 2000a), com base em um modelo *a priori* de estruturação genética e um modelo *a priori* não definido. O programa permite a determinação do coeficiente de ancestralidade (Q) e divergência genética (F_{ST}) através do método de Cadeia de Markov e Monte Carlo (MCMC). É um programa baseado em modelos de estruturação que permite a comparação entre estruturação populacional determinada *a priori* pelo usuário com base em dados previamente disponíveis (ex. análise de agrupamento) com modelos de estruturação que incluem um número não definido de subpopulações. A análise de estruturação foi estabelecida para modelos mistos de ancestralidade, frequência correlacionada dos alelos, e 20.000, 50.000 e 2000.00 repetições MCMC depois de igual período de “burning” (Pritchard et al., 2004).

4.1.3.6. Determinação da coleção nuclear a partir da coleção da germoplasma

Uma coleção nuclear representa uma amostra de uma coleção inteira composta de um número limitado de acessos que juntos devem representar o nível máximo possível de diversidade genética existente em toda a coleção (Frankel & Brown, 1984). Inicialmente, os dados genotípicos de 16 locos microssatélites analisados em 298 acessos de arroz foi utilizado para estimar a riqueza alélica da coleção. Posteriormente, um grupo de acessos mais diversos na coleção foi definido utilizando o programa Corex (M.E.Ferreira, não publicado). O programa possibilita a definição de subgrupos de acessos selecionados com base na maximização da distância genética par-a-par entre os acessos. O tamanho de cada grupo de acessos é definido pelo usuário. Neste estudo, grupos compostos de 28, 56, 84, 112, 140, 168, 196, 224, 252 e 280 acessos foram pré-definidos no sentido de encontrar, em cada grupo, os acessos que foram mais divergentes baseados no

coeficiente “shared allele distance” (Bowcock et al., 1994; Goldstein et al., 1995). Cada grupo foi reamostrado 1.000 vezes e os acessos sempre presentes em cada uma das repetições foram selecionados para constituir o grupo final, o qual foi avaliado para a riqueza alélica. Um gráfico de plotagem do tamanho dos grupos obtidos versus a riqueza alélica foi então utilizado para determinar o tamanho do grupo ou coleção nuclear que retém, pelo menos, 80% da riqueza alélica de toda a coleção. A coleção nuclear resultante foi então comparada com uma coleção nuclear obtida por amostragem aleatória (100 repetições) do mesmo número de acessos a partir da coleção inteira. A amostragem foi efetuada sem reposição do acesso selecionado. Os parâmetros estatísticos obtidos pela coleção nuclear baseada em distância genética e daquela obtida pela amostragem aleatória foram comparados. Os dados foram também submetidos à análise de agrupamento seguindo os procedimentos descritos anteriormente. A estrutura genética da coleção nuclear baseada na análise de distância genética foi analisada utilizando o programa Structure 2.1 (Pritchard et al., 2000a; Falush et al., 2003).

4.1.4. Resultados e Discussão

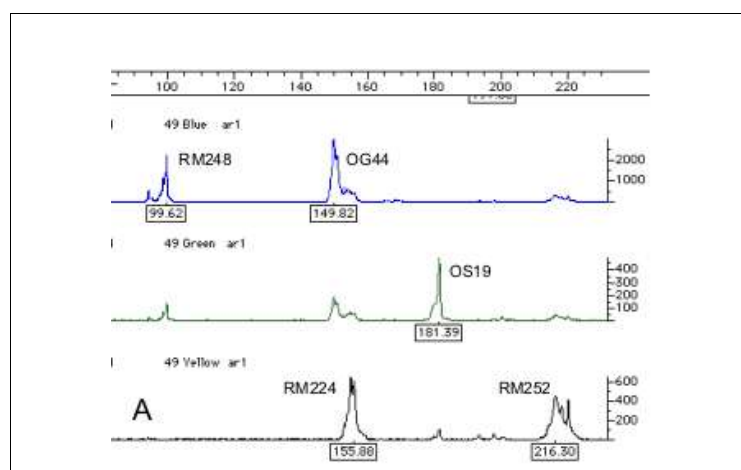
4.1.4.1. Genotipagem com multiplex de microssatélites

A genotipagem semi-automática de três painéis múltiplos de marcadores microssatélites fluorescentes (Figura 1) resultou em 305 alelos únicos (Tabela 2). A média do número de alelos por loco foi 19,06, variando de 7 para o marcador RM420 a 30 para o marcador RM250. O alelo mais freqüente (≥ 40 contagens) representou, em média, 54,47% de todos os alelos em cada loco marcador e variou de 22,27% no loco OG106 a 93,72% no loco RM420.

A maioria dos locos apresentou amplitude alélica esperada, de acordo com dados coletados em estudos anteriores (Beló, 2001; Pessoa Filho, 2004). As exceções foram os locos OG61 e OG81, com repetições dinucleotídicas, para os quais a amplitude alélica estendeu-se para 64 pb e 32 pb contra a esperada 58 pb e 18 pb, respectivamente. A amplitude alélica observada para os locos RM335 e RM418, marcadores baseados em repetições trinucleotídicas, foram também diferentes do esperado (66 pb ao invés de 73 pb esperado e 63 pb ao invés de 52 pb esperado, respectivamente). A amplitude observada dos alelos microssatélites por loco foi maior do que a relatada em estudos prévios utilizando outros tipos de marcadores tais como isoenzimas tais como isoenzimas (Glaszmann, 1987; Second, 1982), RFLPs (Wang & Tanksley, 1989; Zhang

et al., 1992) e microssatélites (Beló, 2001). Entretanto, seis marcadores de um total de 16 cujas amplitudes alélicas foram maiores do que as esperadas de acordo com Belo (2001) apresentaram amplitudes menores do que os números reportados por Pessoa Filho (2004). Isto pode ter sido causado por dois fatores: (a) uma mais correta genotipagem na determinação dos alelos utilizando genotipagem baseada em fluorescência em seqüenciador de DNA comparada à eletroforese de gel de poliacrilamida (PAGE) corada com nitrato de prata, como descrita por Beló (2001) e (b) um efeito do tamanho da amostra entre o estudo aqui descrito e o que foi desenvolvido por Pessoa Filho (2004).

Os microssatélites têm sido marcadores ideais para a caracterização de diversidade genética em arroz cultivado quer a nível inter-varietal (Yang et al. 1994; Xiao et al. 1996; Beló, 2001; Pessoa Filho, 2004) como intra-varietal (Olufowote et al. 1997). Acrescenta-se que, muitos marcadores microssatélites permitem a detecção de alelos numa amplitude muito próxima das espécies de arroz não domesticadas (Wu & Tanksley 1993; Panaud et al. 1996; Harrington 2000; Moncada et al. 2001; Ishii et al. 2001) e algumas vezes, em outras espécies de gramíneas mais distantes (Ishii & McCouch 2000; Chen et al. submetido). A extensão da diversidade alélica, detectada nos locos de marcadores microssatélites, depende do loco específico a ser avaliado, do número e diversidade dos genótipos amostrados e, em certo grau, da sensibilidade da técnica utilizada para a detecção de pequenas diferenças no peso molecular.



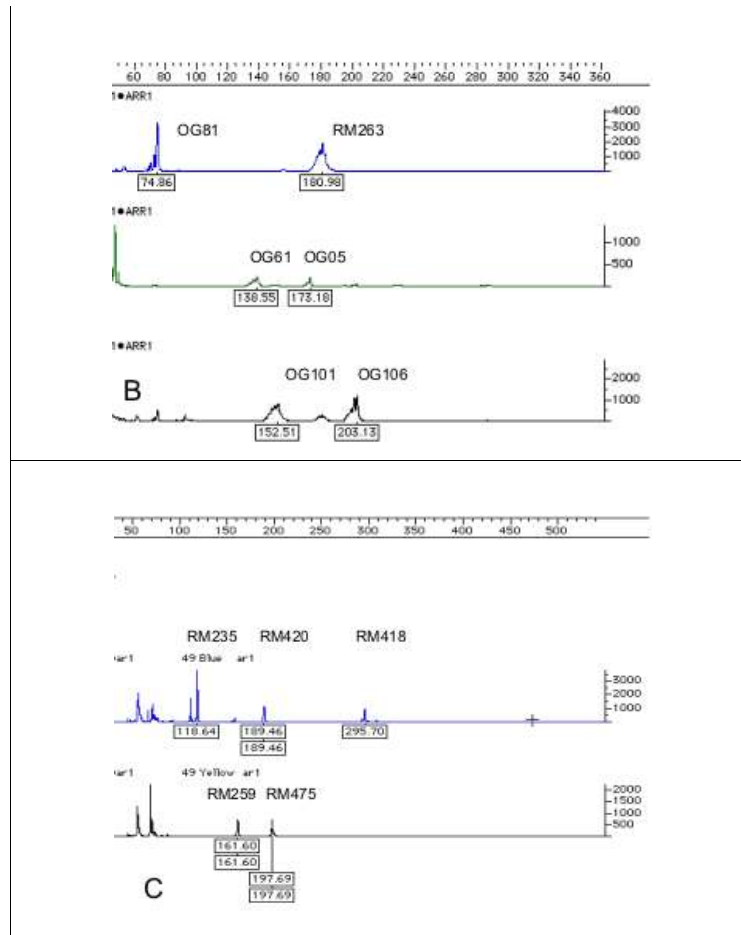


Figura 1 – Electroferograma mostrando o genótipo do acesso controle ARR1 obtido com 16 marcadores microssatélites analisados com três painéis múltiplos. O eixo horizontal representa as estimativas do tamanho em pares de bases e o eixo vertical indica a intensidade de fluorescência medida pelo seqüenciador de DNA ABI Prism.

4.1.4.2. Banco de dados de frequências alélicas e análise de diversidade genética

O polimorfismo de 16 locos marcadores genotipados com os três painéis múltiplos foi avaliado através das estimativas do número de alelos detectados em cada loco (k), diversidade genética observada (GDo), diversidade genética esperada (GDe) e valor de PIC (Tabela 3).

Tabela 8 – Estatísticas descritivas de 16 locos marcadores microssatélites genotipados em painéis múltiplos em uma coleção de 298 acessos de arroz.

Marcador	k	N	GDo	GDe	PIC
RM248	17	272	0,250	0,900	0,890
OG44	13	283	0,201	0,877	0,863
RM224	19	275	0,193	0,881	0,868
RM250	30	271	0,188	0,943	0,939
OS19	15	258	0,174	0,904	0,894
OG81	16	265	0,140	0,837	0,817
RM263	20	274	0,186	0,871	0,857
OG101	26	260	0,262	0,938	0,933
OG106	29	247	0,146	0,942	0,937
OG61	27	218	0,220	0,932	0,926
OG05	13	201	0,214	0,897	0,886
RM335	21	239	0,163	0,904	0,895
RM420	7	247	0,045	0,806	0,777
RM418	21	235	0,149	0,913	0,905
RM259	13	237	0,143	0,904	0,894
RM475	18	202	0,173	0,835	0,820
Média	19,06	-	0,1779	0,892	0,881
Total	305	-	-	-	-

K= número de alelos; N= número de acessos analisados; GDo=diversidade genética observada; GDe = diversidade genética esperada; PIC = Conteúdo de Informação do Polimorfismo

O número médio de alelos por loco foi 19, variando de 7 alelos para o marcador RM420 (motivo tetranucleotideo) a 30 alelos para o marcador RM250 (motivo dinucleotideo). Relatos recentes (Blair et al., 2002; Ni et al., 2002 e Pessoa Filho, 2004), embora amostrando locos marcadores diferentes, mostraram variações similares no número de alelos por locos, usualmente maiores para marcadores com repetições dinucleotídicas e menores para marcadores com repetições tetranucleotídicas. É importante enfatizar que o tamanho e diversidade das amostras de acessos de arroz usadas neste tipo de estudo têm de ser similar para fins de comparação.

Os valores de diversidade genética esperada (GDe) foram altos na maioria dos marcadores microssatélites, com um valor médio de 0,892, variando de 0,806 para RM420 a 0,943

para RM250. O valor médio de GDe (0,892) foi maior que o reportado por Pessoa Filho (2004) para uma coleção de germoplasma duas vezes maior do que a estudada neste estudo (0,786). Uma correlação significativa entre valores de GDe e o número de alelos por locos ($r = 0,750$, $P < 0,001$), e a amplitude alélica (Figura 2; $r = 0,550$, $P < 0,001$) foi observada, corroborando estudos prévios (Pessoa Filho, 2004).

Marcadores microssatélites baseados em seqüências repetitivas maiores (ex. tetranucleotídeos) tendem a ter um menor número de alelos por loco do que aqueles baseados em seqüências repetitivas menores (ex. dinucleotídeos), mas, no geral, apresentam uma variação de tamanho de alelo maior (Hall et al., 1996; Pessoa Filho, 2004). Visto que o número de marcadores com motivos tri- e tetranucleotídeos utilizados neste estudo foi limitado, não foi possível avaliar a relação entre o tamanho da seqüência repetitiva do microssatélite e a variação do tamanho de alelos do loco.

Os valores de PIC foram também altos para a maioria dos marcadores microssatélites, com um valor médio de 0,881, variando de 0,777 para RM420 a 0,939 para RM250. O valor médio de PIC (0,881) foi similar ao estimado por Blair et al. (2002) e Ni et al. (2002). Os 16 marcadores microssatélites foram usados para estimar a probabilidade de genótipos idênticos (Sefc et al., 1999), baseada o banco de dados de freqüências alélicas (Tabela 4a, b). A probabilidade de genótipos idênticos estimada com os três painéis multiplex foi $1,05 \times 10^{-23}$. Isto indica que os três painéis multiplex de microssatélites podem ser usados com acurácia para diferenciar indivíduos e acessos de arroz.

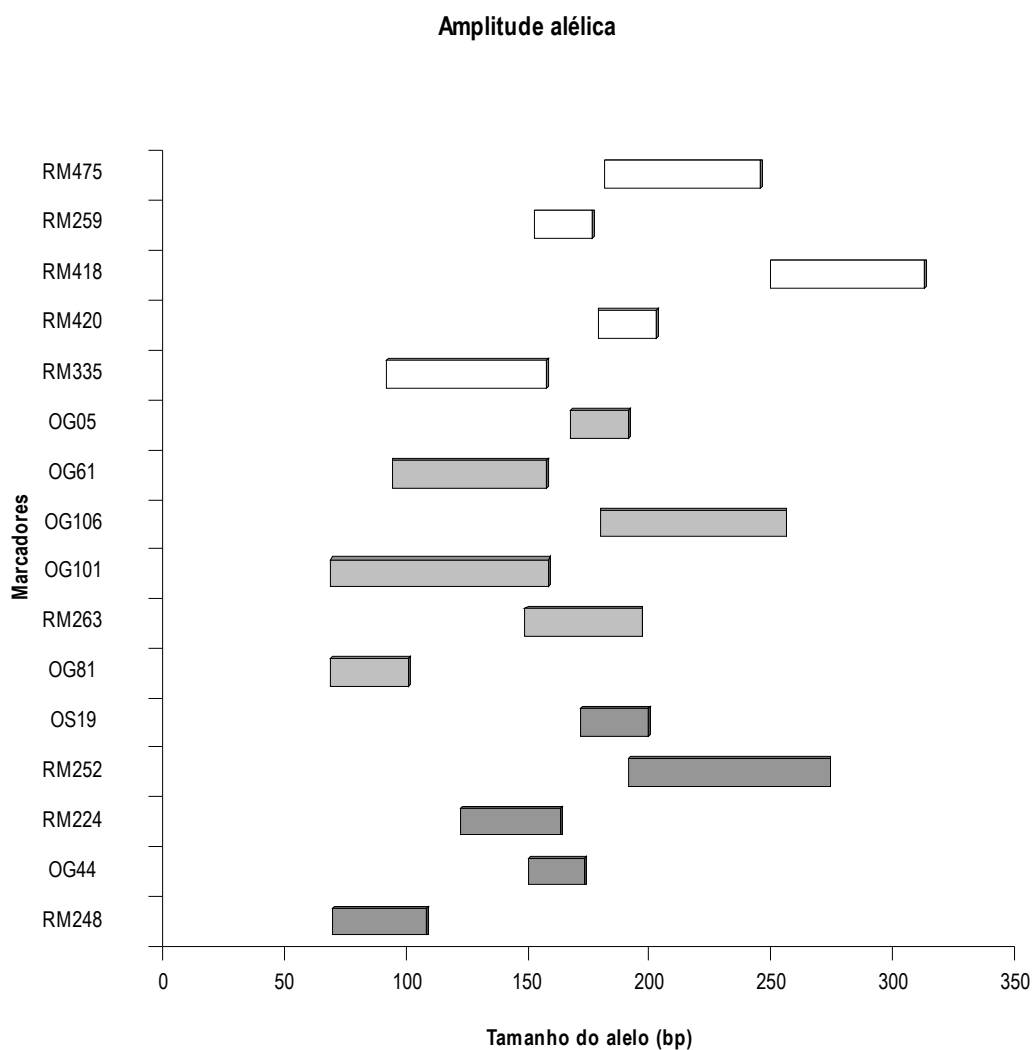


Figura 2 – Amplitude alélica dos 16 marcadores microssatélites dos três painéis múltiplos (cinza escuro e branco – pentaplex e hexaplex, Beló, 2001; cinza claro; – pentaplex; Pessoa Filho, 2004).

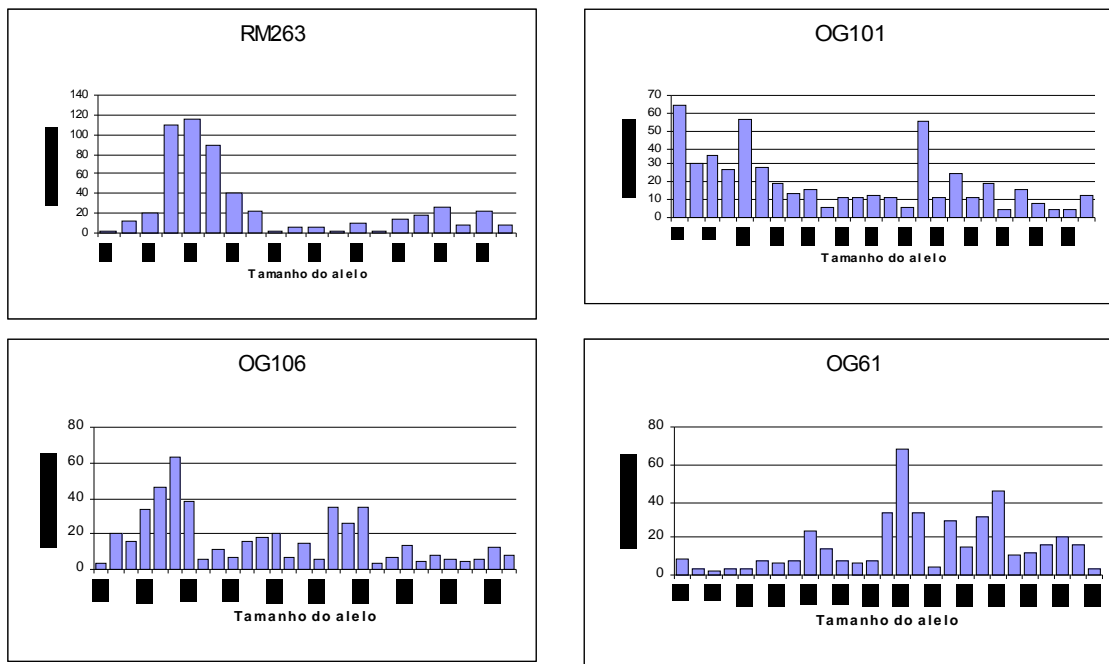
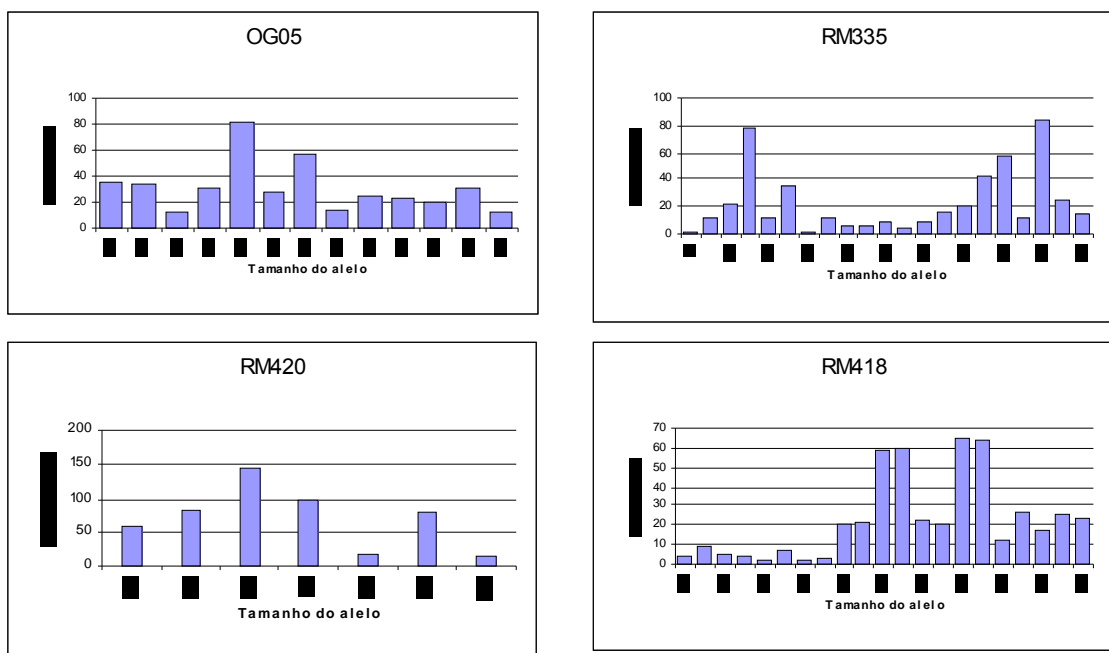


Figura 3 – Histograma de barras da frequência alélica dos 16 marcadores microsatélites utilizados para avaliar uma coleção de germoplasma de arroz.



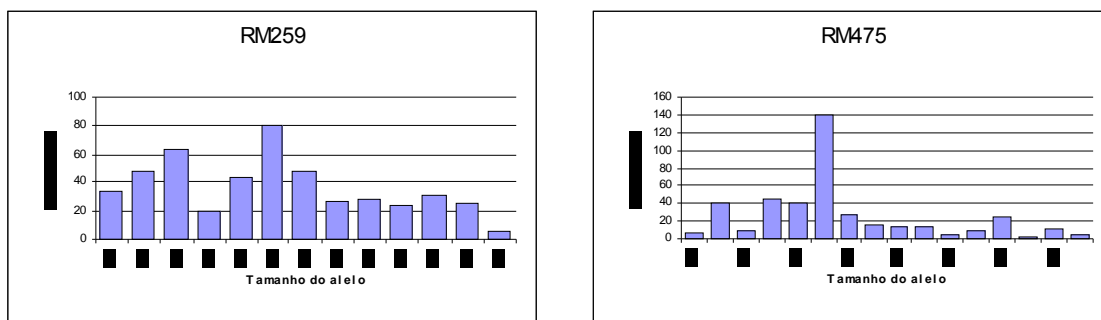


Figura 3 – Histograma de barras da frequência alélica dos 16 marcadores microssatélites utilizados para avaliar uma coleção de germoplasma de arroz. (cont.)

4.1.4.3. Análise de agrupamento dos acessos de germoplasma

A matriz diagonal de distância genética, baseada nas estimativas de “shared allele distance”, foi utilizada para a análise de agrupamento dos acessos de arroz baseada no método de Neighbour-Joining. O dendrograma possibilitou a identificação de grupos de acessos de arroz com vínculo genético mais próximo. O dendrograma de distância genética resultante (Figura 4) mostra os acessos distribuídos em dois grupos principais, um composto por 55 acessos (Grupo 1) e outro composto por 243 acessos (Grupo 2). O Grupo 1 inclui acessos que não são claramente classificados como variedades *indica* ou *japonica*. Nenhum dos acessos típicos de *japonica* incluídos neste estudo (ex. Tomoe Mochi, IAC 166) bem como os acessos típicos de *indica* (ex. IR-50, Cica 8, BG 90-2) estão próximos aos 55 acessos do Grupo 1. Em um estudo similar (Pessoa Filho, 2004) esses 55 acessos também formaram um grupo à parte dos acessos típicos de *indica* e

japonica.

O segundo grupo principal (Grupo 2), o maior deles, está composto de arroz *indica* e *japonica* subdividido em pequenos subgrupos. Em geral, os acessos conhecidos do *background* genético *japonica* agruparam com outros tipos de *japonica*. O mesmo se verifica para os tipos *indica*. Esta análise foi importante para ter a primeira indicação do *background* genético da maioria dos acessos tradicionais na coleção. Alguns exemplos chamam a atenção. Nove acessos de arroz vermelho foram genotipados pela primeira vez com os painéis múltiplos de marcadores microsatélites. Esses acessos foram coletados em pequenas vilas do Nordeste do Brasil (José Almeida Pereira, comunicação pessoal), onde tradicionalmente os agricultores dão mais valor ao arroz vermelho do que o arroz normal por razões culinárias.

Os acessos Arroz Vermelho 1, Arroz Vermelho 2, Arroz Vermelho 3 e Arroz Vermelho 9 formam um subgrupo com Ciwini e Basmati 370, conhecidos como tipo *indica* (Beló, 2001; Fatima et al., 2002). Basmati 370 é um arroz aromático, com qualidade fina e grão extra fino e que apresenta um peculiar alongamento depois de cozido. Entretanto, este subgrupo também inclui a variedade chamada Mochi Gome, que é um nome comum para variedades castanhas, glutinosas ("sticky") no Japão. Os outros acessos de Arroz Vermelho também agrupam com tipos *indica* em outros subgrupos. Por exemplo, Arroz Vermelho (7) e Arroz Vermelho (8) agrupam com Oryzica 1, Cica 8 e BG 90-2, enquanto Arroz Vermelho (4) agrupa com linhagens TGMS. Os acessos de Arroz Vermelho 5 e Arroz Vermelho 6 pertencem a um subgrupo diferente que não inclui qualquer tipo *indica* bem conhecido e é composto, na sua maioria, de acessos com nomes comuns bastante conhecidos. Pode ser observar, entretanto, que as variedades tradicionais denominadas Arroz Vermelho coletadas no nordeste do Brasil são muito diversas entre si e podem ser prontamente distinguidas baseadas na análise de genotipagem multilocos.

Todas as linhagens com característica de macho-esterilidade gênica termo-sensível (TGMS) agruparam-se no mesmo subgrupo (linhagens TGMS70977_12, TGMS70977_26, TGMS68940, TGMS70987, TGMS68935, TGMS71018, TGMS70989, TGMS68944 e TGMS68945), formando um grupo com as linhagens *indica* IR 36, IR50 e Metica 1. A linhagem TGMS68935 foi cruzada com IR50 (normal) para o desenvolvimento de populações segregantes que foram utilizadas para mapear locos associados ao controle de TGMS no genoma de arroz (ver Capítulo 2).

Foi interessante verificar que os acessos com o mesmo nome comum, algumas vezes agrupam no mesmo subgrupo (ex. Agulha (1), Agulha (2) e Agulha (3) ou Carolina (1) e Carolina (2)) enquanto que outras (ex. acessos com nome Caqui ou DeAbril) não. Isto é uma clara indicação que os genótipos com o mesmo nome comum geralmente são geneticamente diferentes, especialmente quando variedades tradicionais são consideradas. No universo da nomenclatura tradicional, alguns nomes mais óbvios ou similares são dados para variedades distintas cultivadas em diferentes localidades, como amarelão, amarelinho ou dourado, ou ainda três meses, 3 meses ou 90 dias.

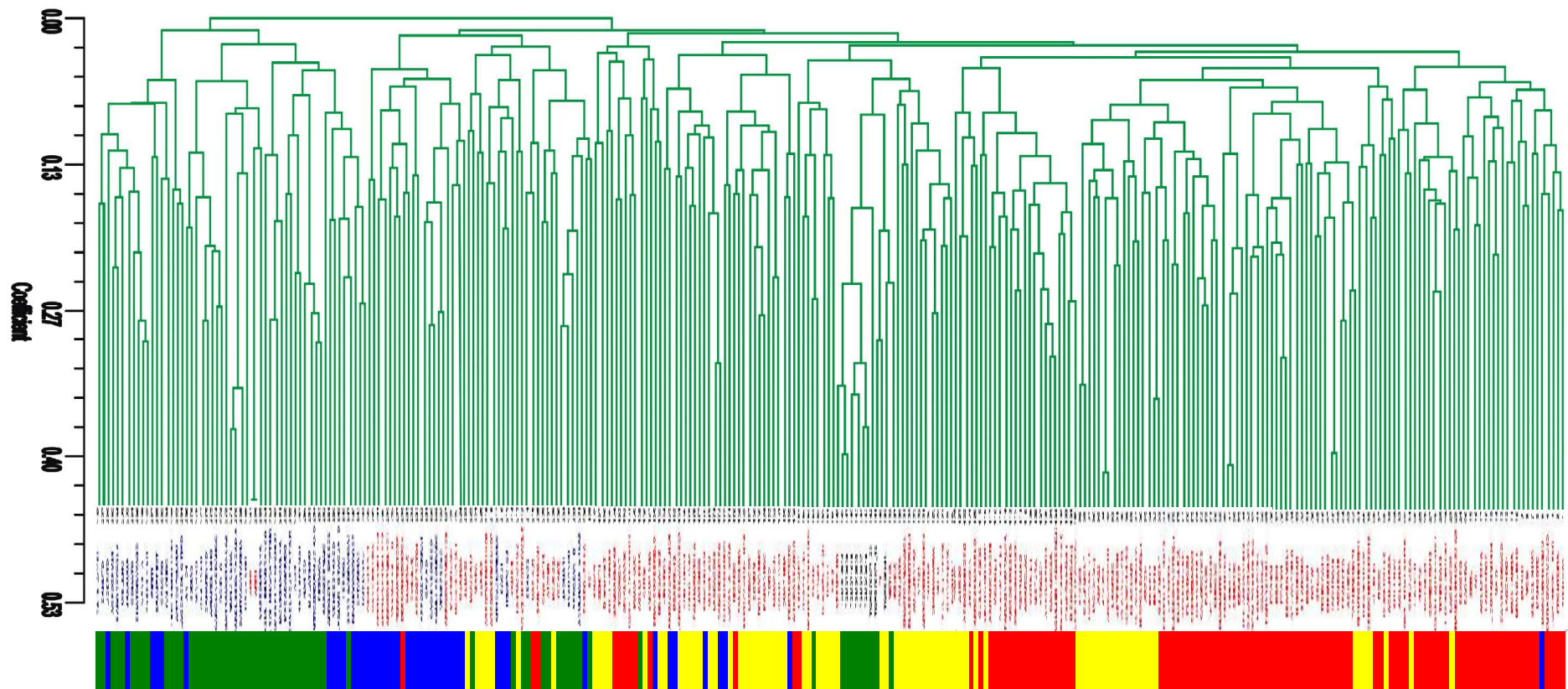
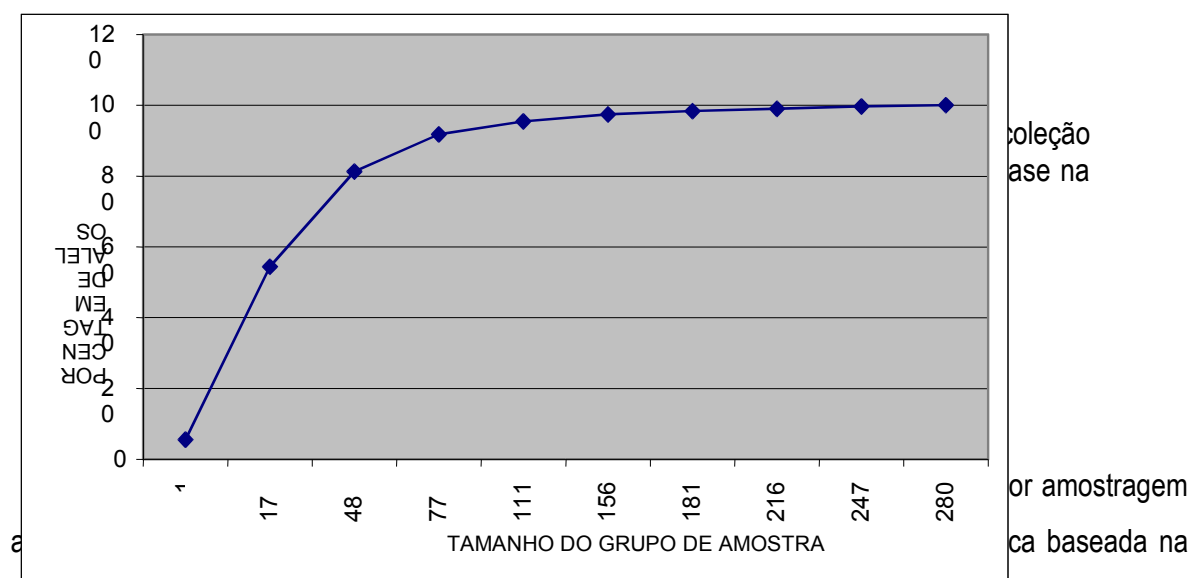


Figura 4 – Análise de estrutura genética de 298 acessos de arroz *O. sativa* L. baseada no polimorfismo de DNA de 16 locos microssatélites. Estimativas de distância genética baseadas no coeficiente de alelos partilhados (*allele sharing distance*) e análise de agrupamento foram calculadas utilizando o método Neighbour-Joining (Saitou & Nei, 1987). A análise de estruturação genética foi baseada no método de agrupamento Bayesiano (barra colorida) e efetuada de acordo com Pritchard et al. (2000a).

4.1.4.4. Coleção Nuclear

Distâncias genéticas par-a-par foram utilizadas para a extração, a partir da coleção de 298 acessos, daqueles acessos que foram consistentemente os mais distantes um do outro, independentemente do tamanho da subamostra extraída da coleção. Os acessos que foram mais divergentes baseados no coeficiente de “shared allele distance” foram selecionados para formar grupos de tamanho 28, 56, 84, 112, 140, 168, 196, 224, 252 e 280 acessos (Figura 5). Cada grupo foi reamostrado 1.000 vezes e os acessos presentes em todas as repetições foram selecionados para compor o grupo e examinados para a riqueza alélica. Um gráfico de plotagem do tamanho do grupo versus riqueza alélica (Figura 5) indicou que uma amostra de 48 acessos, baseada na distância genética par-a-par, constituindo 16,6% da coleção de germoplasma analisada, manteve a riqueza alélica acima 80,0% (81,31%), o que caracterizou este subgrupo de acessos como coleção nuclear por definição (Frankel & Brown, 1984).



amostragem aleatória foi calculada como $73,65 \pm 2,10$ %, significativamente menor ($t=11,49$; $p>0,001$) do que 81,31% da riqueza alélica obtida pela metodologia de distância genética.

Os resultados mostram que a seleção de subgrupos com elevada diversidade genética baseada na análise de distância genética é uma metodologia útil para determinar o tamanho e a composição das coleções nucleares baseadas num número mínimo de acessos que continuam

mantendo uma alta percentagem de alelos. Valores similares de Gdo, GDe e PIC foram estimados para a coleção de germoplasma (0,129, 0,911 e 0,892, respectivamente) e para a coleção nuclear (0,181, 0,883, e 0,859, respectivamente).

4.1.4.5. Estrutura genética da coleção

Com o objetivo de analisar a consistência dos agrupamentos obtidos com o método “Neighbour-Joining”, o nível de diferenciação genética entre os agrupamentos dos acessos obtidos com os dados de distância par-a-par foram estimados baseados nos parâmetros F de Wright (Tabela 10). A média de diferenciação genética ($F_{ST} = 0,08616$) entre grupos (Weir & Cockerham, 1984) foi considerada moderada (Wright, 1978), e estimada como 0,077 quando os dois grupos principais definidos pelo método “Neighbour-Joining” foi considerado e 0,095 quando os nove grupos foram considerados. O agrupamento dos acessos obtidos pela análise de distância, portanto, explica somente cerca de 8% da diferenciação genética.

Tabela 10 – Estimativa dos parâmetros F de Wright para a coleção de germoplasma. Os agrupamentos de acessos foram definidos de acordo com análise Neighbour-Joining dos dados de distância genética baseadas no polimorfismo de microssatélites; Intervalos de confiança de 95% foram estimados baseados na análise de bootstrap (1000x).

F de Wright	Dois grupos	Nove grupos
F_{IS}	0,786 (0,763 – 0,811)	0,773 (0,747 – 0,798)
F_{IT}	0,802 (0,782 – 0,827)	0,794 (0,772 – 0,820)
F_{ST}	0,077 (0,051 – 0,103)	0,095 (0,073 – 0,121)

As estimativas dos parâmetros F de Wright (Weir & Cockerham, 1984) revelaram um valor alto do coeficiente de endogamia quando duas ou nove subpopulações (grupos) são consideradas na coleção ($F_{IS}=0,786$ e $0,773$, respectivamente) (Tabela 10), assim como um índice total de fixação relativamente alto ($F_{IT} = 0,779$). Por outro lado, os valores de F_{ST} indicam diferenciação genética moderada quando duas ou nove subpopulações são considerados para compor a coleção ($F_{ST} = 0,077$ e $0,095$, respectivamente). Isto é uma indicação clara que a extensão da endogamia dentro de grupos é alta, mas a extensão de diferenciação genética entre grupos que compõem a coleção de germoplasma é somente moderada.

Análises posteriores foram realizadas utilizando o programa Structure 2.1 (Pritchard et al. 2000a; Falush et al. 2003) para inferir o nível de diferenciação genética entre subpopulações (grupos) definidos de acordo com a análise Neighbour-Joining (dois e nove grupos) e sem um modelo definido *a priori*. O melhor valor de K resultante (número de subpopulações) foi pela matriz sem modelo definido *a priori* e com modelo definido com base nos dois grandes grupos (Grupo 1 e Grupo 2) da análise do Neighbour-Joining. Para ambas as matrizes, a probabilidade 'best-fit' de 1,00 foi obtida para K=4, indicando uma diferenciação genética da coleção em quatro subpopulações (Figura 8).

A proporção de ancestralidade comum entre as quatro subpopulações geneticamente definidas detectada no total de 298 acessos mostra um padrão interessante. As estatísticas F para K=4 foram $F_{IS}= 0,637$; $F_{ST}= 0,035$ e $F_{IT}= 0,650$ (Weir & Cockerham, 1984). Novamente, os dados indicam que a extensão da endogamia dentro dos grupos é alta, mas a diferenciação genética entre os grupos é somente moderada. Os quatro grupos compreendem 63, 45, 106 e 81 acessos, respectivamente. O grupo 1 partilha 20,63% de ancestralidade com o grupo 2, 9,52% com o grupo 3 e menos de 5,0% com grupos ou mistura de grupos 2-3, 2-4, 3-4 e 4. O grupo 2 partilha 11,11% de ancestralidade com o grupo 1, 13,33% com o grupo 3 e 8,89% com o grupo 4. O grupo 3 partilha 10,38% de ancestralidade com o grupo 2, 8,49% com o grupo 4 e menos de 5,00% com os grupos ou misturas de grupos 1, 1-2, 2-4, 4-1 e 4-2. O grupo 4 partilha 11,11% de ancestralidade com o grupo 1, 12,00% com o grupo 2, 12,35% com o grupo 3, e menos de 5,00% com grupos ou mistura de grupos 1-2, 1-3 e 2-3. Uma análise fatorial clássica (Belkhir, 2001) possibilitou a separação espacial das quatro subpopulações retratadas pela metodologia de MCMC (Figura 6).

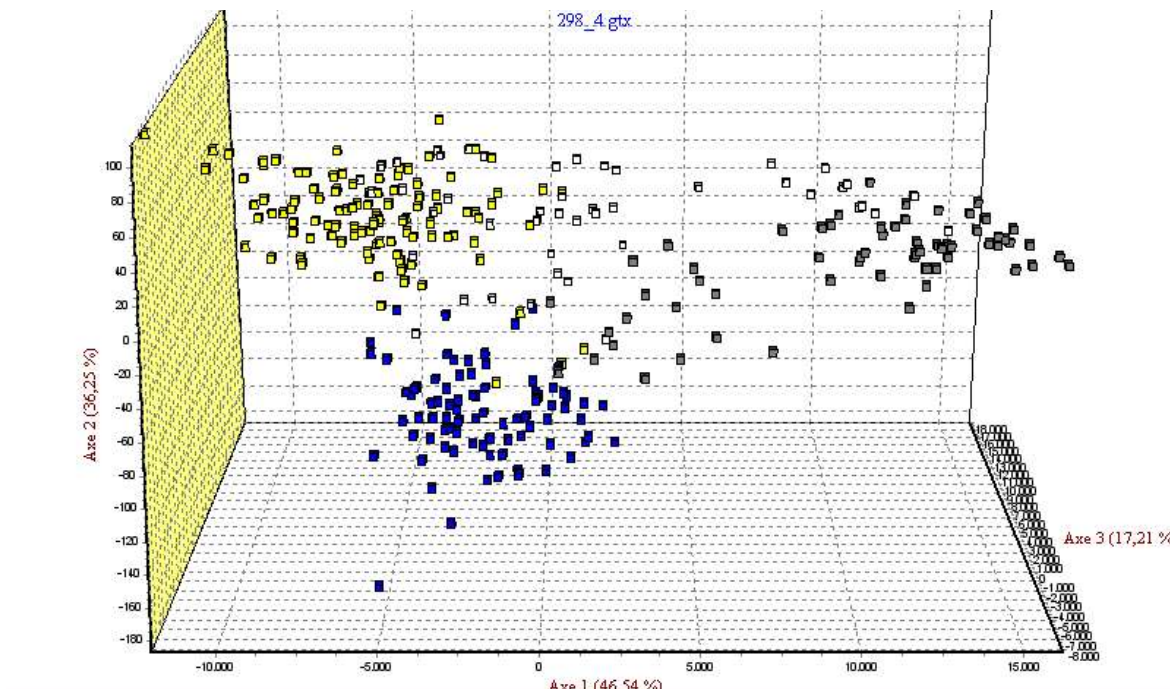


Figura 6 - Análise fatorial clássica dos 298 acessos de arroz *O. sativa* L. baseadas no polimorfismo de DNA em 16 locos polimorficos. As estimativas de distância genética foram baseadas no coeficiente de “shared allele distance” e análises de agrupamento foram calculadas utilizando o método Neighbour-Joining (Saitou & Nei, 1987).

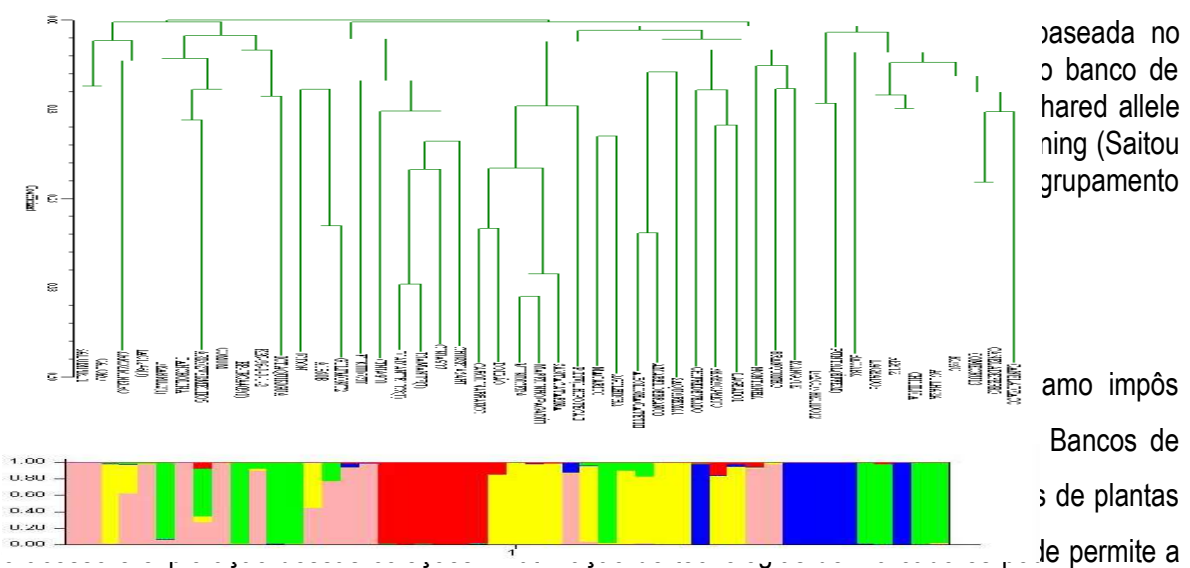
A estratificação genética sugerida pelo método Neighbour-Joining (Saitou & Nei, 1987) não correlaciona ($r= 0.75$, $P >0.10$) com as mesmas subpopulações (grupos) sugeridas pelo método MCMC (Figura 4) (Pritchard et al., 2000a; Falush et al., 2003). Esse resultado era esperado devido às diferenças de metodologias, uma vez que o método Neighbour-Joining agrupa os acessos baseado em distâncias genéticas (i.e., shared allele distances) e a metodologia MCMC agrupa de acordo com a probabilidade de dois alelos terem um ancestral comum. Cada método tem seus pontos fortes e debilidades, e não há uma estratégia única ou simples que atenda efetivamente a vários assuntos complexos relacionados à escolha de medida (s) de distância, métodos de agrupamento, determinação do número ótimo de grupos ou análise de indivíduos, e grupos de dados combinados por meio de várias ferramentas estatísticas. Entretanto, dados empíricos gerados recentemente através de diferentes estratégias têm proporcionado uma maior

compreensão dos assuntos citados, e meios efetivos razoáveis para analisar a diversidade genética nos diversos níveis (indivíduo, populações, ou espécie).

Diferentes níveis de estrutura genética têm sido documentados em populações naturais (Barbier 1989b; Gao et al., 2000a, b; Gao et al., 2001) e coleções de germoplasma de arroz (Glaszmann 1987; Ni et al., 2002; Parsons et al. 1999; Garris et al., 2003; Gao et al., 2000a, b; Gao et al., 2001). Em populações naturais de uma espécie autógama tal como o arroz, espera-se encontrar uma maior proporção de diversidade entre subpopulações do que dentro delas (Buso et al., 1998). As coleções de germoplasma, entretanto, são artificiais por natureza e várias forças, incluindo a deriva genética e a seleção artificial dirigida pelo homem, afetam a forma em que são estruturadas. Altos valores de F_{ST} obtidos em arroz silvestre, indicando forte diferenciação genética em populações naturais de *O. officinalis* ($F_{ST} = 0.882$; Gao et al., 2001), *O. rufipogon* f ($F_{ST} = 0.310$; Barbier 1989b; Gao et al. 2000a), *O. granulata* ($F_{ST} = 0.859$; Gao et al. 2000b) e *Oryza glumaepatula* (Buso et al., 1998). Valores altos do índice de fixação ($F_{ST} > 0.85$) têm sido reportados em estudos de variedades de arroz (ecotipos de *O. sativa*) oriundas de diferentes países e que apresentam diversidade para resistência à *Xanthomonas oryzae pv oryzae* (Garris et al., 2003). Da mesma forma, amostras de germoplasma compostas de acessos de arroz (*O. sativa*) proveniente de diversas partes do mundo mostraram valores de F_{ST} variando de 0,20 a 0,43 (Garris et al., 2005). Tem de ser realçado que a coleção aqui analisada não é representativa da diversidade genética do arroz e o nível de diferenciação observado é considerado moderado. Mas, é interessante notar a possibilidade de detecção de estruturação em populações com um grupo pequeno de marcadores microssatélites distribuídos em três painéis multiplex, corroborando a utilidade dos marcadores altamente informativos na análise de genética de populações.

De modo similar à coleção inteira, os 48 acessos que compõem a coleção nuclear foram submetidos à análise Neighbour-Joining baseadas em distância genética (Figura 7). Os principais grupos observados foram utilizados como modelo para a análise de estruturação genética utilizando a metodologia MCMC (Pritchard et al., 2000b; Falush et al, 2003) e estimativas da estatística F (Weir & Cockerham, 1984). A metodologia MCMC, desenvolvida sem um modelo de estruturação definido *a priori* indicou uma diferenciação em K=5 subgrupos (probabilidade “best-fit” =1,00). Um gráfico de plotagem das linhagens pertencentes a cada subgrupo contra o agrupamento obtido pelo método Neighbour-Joining (Figura 7) indicou que os resultados entre as duas metodologias não

foram congruentes ($r= 0.475$; $p= 0,45$). As estimativas dos parâmetros F de Wright revelam um alto valor do coeficiente de endogamia quando cinco subpopulações (grupos) são considerados na coleção nuclear ($F_{IS}= 0,70613$) bem como um alto índice de fixação geral ($F_{IT}= 0,74445$). Por outro lado, os valores de F_{ST} indicam uma diferenciação genética moderada quando cinco subpopulações são consideradas para constituir a coleção nuclear ($F_{ST} = 0,13040$). Os resultados indicam que a coleção nuclear de 48 acessos obtida a partir da coleção de germoplasma composta de 298 acessos apresentou grupos de acessos com diferenciação genética moderada.



avaliação sistemática da diversidade genética em tais coleções. Essa informação pode ser usada para o estabelecimento de coleções nucleares, subgrupos de grandes coleções que contêm um número limitado de acessos que capturam a maioria da variabilidade genética presente na coleção inteira. Esses subgrupos nucleares facilitam a gestão da coleção e contribuem para um aumento da utilização do germoplasma. É possível que o uso do germoplasma conservado de arroz nas coleções de germoplasma possa aumentar se a caracterização molecular e metodologias de genética de populações puderem ser mais intensivamente aplicadas para melhor definir as amostras de acessos que serão mais apropriadamente adaptados a diferentes estudos genéticos.

Com o recente desenvolvimento e uso dos métodos de agrupamento baseados em modelos na estatística Bayesiana (Pritchard, 2001; Thornsberry et al., 2001), a possibilidade de realizar estudos de associação em plantas agrícolas para a identificação de genes para características

complexas e agronomicamente importantes têm sido aumentada. Entretanto, continua havendo a necessidade de desenvolver programas estatísticos que facilitem a análise integrada dos diferentes grupos de dados para a geração de uma informação fidedigna acerca das relações genéticas, diversidade de germoplasma, e variação alélica favorável. Igualmente importante, e talvez mais desafiante, é a utilização concertada e planejada do germoplasma em programas de melhoramento de plantas com base em conhecimento obtido de estudos sobre diversidade genética (Mohammadi & Prasanna, 2003).

4.2. Capítulo 2 – Mapeamento de QTLs para a Macho-esterilidade Gênica Termo-sensível (TGMS) em Arroz *indica*

4.2.1. Resumo

O conhecimento do controle genético de macho-esterilidade termossensível (TGMS) em arroz é importante para o desenvolvimento de uma indústria de produção sustentável de sementes híbridas de arroz nos trópicos. Uma população F2 derivada do cruzamento entre linhagens *indica* fértil e TGMS foi usada para construir um mapa genético de arroz baseado em marcadores microssatélites. O fenótipo TGMS analisado apresentou uma variação contínua na população segregante. Um baixo nível de distorção da segregação foi detectado na população segregante (14,65%) e a causa da distorção demonstrada como sendo a seleção zigótica. Não houve evidência de relação de causa-efeito entre seleção zigótica e o controle de TGMS neste cruzamento. Um mapa de ligação compreendendo 1.213,3 cM foi construído baseado nos dados de segregação da população F2. Noventa e cinco marcadores microssatélites de um total de 116 marcadores testados foram reunidos em 11 grupos de ligação com uma média de 12,77 cM entre os locos marcadores adjacentes. Os dados fenotípicos e genotípicos permitiram a identificação de três novos QTLs para a macho-esterilidade gênica termossensível em arroz *indica*. Dois dos QTLs foram mapeados em cromossomos que não tinham ainda sido associados ao controle genético da característica TGMS (cromossomos 1 e 12). O terceiro QTL foi mapeado no cromossomo 7, onde um loco TGMS (*tms2*) foi recentemente mapeado. Testes alélicos deverão ser realizados para verificar se as regiões mapeadas são as mesmas.

Palavras-chave: microssatélite, arroz híbrido, mapa genético, *Oryza sativa* L.

4.2..3. Introdução

O desenvolvimento recente de variedades híbridas de arroz representa a mais importante contribuição ao aumento de produtividade em arroz desde a Revolução Verde. Relatos de que a produtividade dos híbridos de arroz é pelo menos 15-20% maiores do que a produtividade das

melhores variedades puras semi-anãs locais têm sido publicados em vários países (Virmani et al., 1994). O vigor híbrido ou heterose é observado em cruzamentos entre linhagens puras onde a F1 apresenta produtividade mais alta do que as linhagens parentais. Porém, em espécies autógamas, a autopolinização representa uma grande limitação à produção de sementes híbridas, a menos que um mecanismo, eficiente e econômico, de esterilidade do pólen seja utilizado em uma das linhagens parentais, enquanto o sistema reprodutivo feminino é mantido intacto.

A descoberta da Macho-Esterilidade Genética Citoplasmática (CMS) em arroz nos anos 60 tem contribuído para o desenvolvimento de uma estratégia de produção de híbridos de arroz utilizando o sistema de três-linhagens, isto é, macho-estéril (linhagem A), mantenedora (linhagem B) e restauradora (linhagem R). Este sistema, embora útil, usualmente (a) demanda um alto *input* de trabalho para a produção de híbrido, que ameaça a sustentabilidade da produção de sementes a preços acessíveis, (b) conta com uma pequena fonte de linhagens macho-estéreis, especialmente a macho-esterilidade induzida pelo citoplasma “wild abortive” (WA), que favorece a vulnerabilidade genética dos híbridos derivados, (c) requer grandes esforços dos programas de melhoramento, visto que as linhagens macho-estéreis têm de ser adequadamente mantidas pelas linhagens B e as linhagens restauradoras devem ser identificadas para serem eficientemente exploradas em combinações híbridas, e (d) uma vez que as cultivares de arroz *japonica* e arroz basmati aparentemente não carregam genes restauradores, a produção de híbridos envolvendo linhagens *japonica* é restrita a um longo processo de transferência de genes restauradores (Lang et al., 1999; Reddy et al., 2000; Dong et al., 2000; Lopez & Virmani, 2000; Virmani et al., 2003). Recentemente, a identificação de um sistema de esterilidade de pólen dependente da temperatura (Yuan, 1987; Sun et al. 1989; Zhang et al., 1991), conhecido como Macho-Esterilidade Gênica Termo-sensível (TGMS), tem contribuído para a obtenção de linhagens macho-estéreis sob variações específicas de temperatura, abrindo novas perspectivas para a exploração de híbridos de arroz.

No geral, as linhagens TGMS tornam-se estéreis quando a temperatura ambiental é maior do que 25-30°C, durante o estágio de iniciação da panícula e floração, a retornam à condição fértil quando a temperatura atinge valores inferiores a 25-30°C. O sistema TGMS é, portanto, a base para um sistema de duas linhagens para a produção de sementes híbridas em arroz, caracterizado

como simples, econômica e eficiente.

O conhecimento do controle genético da macho-esterilidade gênica termo-sensível (TGMS) em arroz é importante para o desenvolvimento sustentável de uma indústria de sementes de arroz híbrido nos trópicos. Como alternativa a algumas limitações do sistema CMS, o sistema TGMS não requer uma linhagem restauradora (linhagem B) na produção de sementes híbridas. As linhagens B do sistema CMS, por exemplo, devem ser rotineiramente verificadas ao longo de gerações para confirmar se elas são quase-isogênicas em relação às linhagens A, o que torna a rotina do sistema CMS mais trabalhoso (Beló, 2001). O sistema TGMS também não requer linhagens restauradoras (linhagem R), visto que a planta TGMS pode restaurar sua fertilidade se cultivada em temperatura apropriada. Acrescenta-se ainda o fato do sistema TGMS reduzir o risco de vulnerabilidade genética, visto que a expressão da macho-esterilidade não é dependente do citoplasma, podendo os híbridos ser desenvolvidos com bases genéticas diversas. Isto é de interesse especial depois de experiências que levaram a grandes perdas agrônômicas e econômicas devido ao uso intensivo de materiais com a mesma base genética, como o citoplasma-T utilizado na indústria de híbridos de milho, que foi atingido pela “southern corn leaf blight” há algumas décadas (Levings, 1993).

A identificação e uso de diferentes alelos TGMS em arroz proporcionará uma base genética mais ampla para o desenvolvimento de híbridos de arroz. Para a exploração do sistema de duas linhagens é necessário avaliar as linhagens que apresentam tal comportamento, estimar a robustez da expressão da característica numa faixa de temperaturas e também compreender o seu controle genético com vista à utilização do gene no desenvolvimento de cultivares.

A tecnologia genômica oferece um conjunto de técnicas e estratégias que podem potencialmente acelerar o desenvolvimento de variedades em programas de melhoramento. A tecnologia de marcadores moleculares tais como RAPD, RFLP e AFLP, associada à Análise de Misturas Segregantes (BSA – Bulk Segregant Analysis) (Michelmore et al., 1991), tem sido utilizada no mapeamento de vários locos TGMS. A metodologia de bulks ou misturas de DNA possibilitou a identificação e mapeamento de quatro locos TGMS conhecidos como *tms1* (cromossomo 8 - Wang et al., 1995), *tms3(t)* (cromossomo 6 - Subudhi et al., 1997), *tms4* (cromossomo 2 - Dong et al., 2000) e dois locos *tms4* (cromossomo 9 - Reddy et al., 2000). Os dois

locos denominados *tms4* não são certamente os mesmo visto que foram mapeados em cromossomos diferentes. Isto chama a atenção da necessidade de melhor discussão da nomenclatura dos locos TGMS em arroz, como já foi alertado anteriormente (Reddy et al., 2000). Um loco TGMS adicional denominado *tms2* (*cromossomo 7*- Maruyama et al., 1991) foi mapeado usando a metodologia de mapeamento de QTLs (Yamaguchi et al., 1997).

Entre os vários tipos de marcadores moleculares disponíveis, os microssatélites têm recebido recentemente maior atenção, especialmente para propósitos de melhoramento. Os marcadores microssatélites, também conhecidos como seqüências simples repetidas ou SSRs (Litt & Luty, 1989; Weber & May, 1989) são grupos de bases nucleotídeas curtas (usualmente 2 a 6) repetidas em tandem, distribuídas pelo genoma. Os marcadores microssatélites são em geral marcadores genéticos codominantes, multialélicos, e altamente polimórficos. A genotipagem dos alelos microssatélites requer pequenas quantidades de DNA para uma PCR direta e análise de gel de electroforese (Rafalski et al., 1996). Sua principal desvantagem é o elevado custo do investimento inicial necessário para o desenvolvimento dos marcadores. Entretanto, o número de marcadores microssatélites disponíveis para uma espécie modelo, tal como o arroz, é grande e os pesquisadores podem desta forma capitalizar nesta tecnologia na pesquisa genética de arroz. Centenas de marcadores microssatélites foram fisicamente mapeadas no genoma do arroz e podem potencialmente ser utilizado como marcadores “âncora” para a genética comparativa, mapeamento de caracteres de interesse econômico e isolamento genético.

O mapeamento de locos TGMS tem sido baseado na premissa de um controle genético simples da característica. Para este fim, as plantas na população segregante têm sido classificadas em duas classes, estéril ou fértil (Wang et al., 1995; Subudhi et al., 1997; Dong et al., 2000; Reddy et al., 2000). Deve ser mencionado que a esterilidade do pólen das populações segregantes derivadas de cruzamentos entre linhagens TGMS e férteis não seguem uma simples distribuição binária. Esforços devem ser empreendidos, portanto, no mapeamento de QTLs TGMS com base no percentual de pólen estéril medido na população segregante, ao invés da classificação das linhagens em apenas duas categorias (macho estéril ou fértil).

O objetivo deste trabalho foi a seleção e uso de marcadores microssatélites para mapear regiões do genoma associadas ao controle genético de TGMS em arroz.

4.2.4. Material e Métodos

4.2.4.1. Material Biológico e análise fenotípica da macho-esterilidade

Inicialmente, um grupo de nove linhagens TGMS tropicais (Lopez & Virmani, 2000) (IR68945-4-33-4-14-48, IR68935-16-6-27, IR68940-8-1-8-18-25-B, IR68944-6-23-5-4-15-B, IR1018-13-73-3-6, IR70977-16-5-4-1-12, IR90977-16-5-4-26, IR70989-3-19-3-14-3^A, IR70978-8-22-5-14-13), gentilmente cedidas pelo Instituto Internacional de Pesquisa em Arroz (IRRI), Filipinas, foram avaliadas para a fertilidade do pólen sob um espectro de temperaturas. Cem plantas de cada linhagem foram germinadas e cultivadas numa casa de vegetação até a fase de floração. Amostras de 20 plantas foram então transferidas para câmaras de crescimento Tecnal® Model 50 (Tecnal, Campinas, Brasil) ajustadas para as seguintes temperaturas: 21°C, 23°C, 25°C, 27°C e 29°C. Condições de fotoperíodo (12/12 hs luz/escuro) e umidade (70%) foram fixadas nas câmaras de crescimento. Após 14 dias sob essas condições, pelo menos 10 espiguetas por planta foram coletadas de cada linhagem, em cada regime de temperatura, para análise. A coleta foi realizada durante o período da manhã durante três dias consecutivos. As espiguetas foram fixadas em etanol a 70%, tratadas com solução de iodina iodeto de potássio (IKI) 1% e montadas em lâminas de vidro. Os grãos de pólen redondos e corados de marrom escuro foram registrados como férteis, enquanto os amarelos ou marrons claros e de forma irregular foram registrados como estéreis. Um total de 300 a 500 grãos de pólen foi examinado e submetido a cada temperatura específica e as percentagens de pólen estéril anotadas. Estes dados foram utilizados para a seleção de uma das nove linhagens TGMS para cruzar com uma linhagem fértil (IR50) para a obtenção de uma população F2 segregante para a característica TGMS. Esta população foi então usada para o mapeamento de regiões genômicas associadas ao controle genético da característica TGMS. O critério de seleção foi: (a) desempenho da linhagem em relação ao fenótipo TGMS numa determinada temperatura e (b) estimativa da distância genética entre a linhagem TGMS e a linhagem genitora normal (IR50), para maximizar a detecção de marcadores polimórficos destinados ao mapeamento. Estes dados estão presentes em um outro estudo (Alcochete, 2005 – Capítulo 1).

Uma população de noventa e três indivíduos do cruzamento entre duas linhagens *indica*

IR68945-4-33-4-14-48 (TGMS) x IR50 (fértil), foi usada neste estudo. A linhagem IR68945-4-33-4-14-48 torna-se estéril durante o estágio de floração quando a temperatura atinge 27°C e retorna à fertilidade quando a temperatura é inferior a 27°C. Trata-se de uma linhagem estéril estritamente sensível à temperatura (Reddy et al., 2000), derivada de um mutante TGMS de *japonica* de clima temperado denominado Norin PL12 (Lopez & Virmani, 2000). Ambas as linhagens parentais, e as plantas da F1 e da F2 foram germinadas e cultivadas numa casa de vegetação até o estágio de floração, quando elas foram transferidas para câmaras de crescimento Tecnal® Model 50 (Tecnal, Campinas, Brasil) ajustadas para 27°C. As estimativas da esterilidade do pólen (%) seguiram as condições e protocolos acima descritos.

64.2.4.2. Extração e quantificação de DNA

A extração de DNA foi realizada de acordo com os procedimentos descritivos por Ferreira e Grattapaglia (1998). A concentração de DNA foi quantificada utilizando um espectrofotômetro UV Ultrospec 2000 (Pharmacia Biotech) e/ou por visualização em gel de eletroforese de agarose a 0,8% corado com EtBr (10 mg mL⁻¹) e visualizado sob luz UV. Os géis foram fotodocumentados utilizando um equipamento Eagle Eye (Stratagene). A intensidade da banda foi comparada com padrões de Lambda DNA de 10, 50, 100 e 200 ng uL⁻¹ e a concentração final ajustada para 2,0 ng uL⁻¹.

4.2.4.3. Seleção de marcadores

Um total de seiscentos e cinquenta e nove marcadores microssatélites foram avaliados neste estudo, dos quais 462 (70,1%) eram baseados em pares de primers não marcados e analisados em géis de poliacrilamida ou em géis de agarose. Outros 197 marcadores (29,9%) foram marcados com fluorocromos e analisados com um seqüenciador de DNA ABI 377 (Applied Biosystems). Oitenta marcadores microssatélites foram selecionados de pesquisas recentes voltadas para o desenvolvimento de marcadores em arroz (Wu & Tanksley, 1993; Akagi et al., 1996; Panaud et al., 1996; Chen et al., 1997; Brondani, 2000; Temnykh et al., 2002). Os marcadores não marcados consistiram em 312 (67,53%) da série RM, 105 (22,73%) da série OG, 37 (8,01%) da série OS, e 8 (1,73%) da série STSG. Os primers fluorescentes incluíram 89 (45,18%) tetras, 20 (10,15%) tris, 38 (19,80%) dis, e 50 (25,38%) marcadores com motivos complexos.

4.2.4.4. PCR e análise de polimorfismo de DNA

Todos os pares de marcadores não marcados foram amplificados utilizando máquinas de 96 poços Thermal Cycler PT-100 (MJ research) em reações de 13 uL de volume final, contendo 9 ng DNA molde, 0,3 uM de cada primer, 0,1 uL de 50% DMSO, 0,25 mM de cada dNTP, 1,5 mM MgCl₂, tampão de reação 1X, e 1 unidade de Taq DNA polimerase. As condições de amplificação foram as seguintes: desnaturação durante 5 minutos a 94°C; 30 ciclos de 94°C durante 1 minuto, 56°C durante 1 minuto e 72°C durante 2 minutos; etapa de extensão final a 72°C durante 7 minutos (Ferreira & Grattapaglia, 1998; Brondani, 2000; Beló, 2001). As temperaturas de anelamento de alguns primers foram ajustadas para 52 ou 58°C. As reações de PCR para os marcadores fluorescentes foram realizadas separadamente com um volume final de 13 uL e amplificadas por 30 ciclos de 94°C durante 1 minuto, 55°C/58°C/60°C (de acordo com os primers) durante 1 minuto e 72°C durante 2 minutos.

4.2.4.5. Genotipagem da população

Noventa e três indivíduos da F2, as duas linhagens genitoras (IR50 e IR68945-4-33-14-48) e a geração F1 foram genotipados utilizando os marcadores informativos obtidos a partir do “screening” para polimorfismo dos marcadores de DNA. Os produtos de PCR dos marcadores não marcados foram separados e visualizados depois de eletroforese em gel de agarose 3,5% com 0,15 mg/mL de EtBr, ou em gel de acrilamida desnaturante 4% corados com nitrato de prata (AgNO₃). Os marcadores fluorescentes foram separados em gel de poliacrilamida desnaturante 4% utilizando um seqüenciador de DNA ABI Prism 377 (Applied Biosystems). Antes da separação, 3 uL do produto de PCR foram adicionados a igual volume de uma solução de formamida (4,5:0,5 uL de formamida: ROX 500) para prevenir que a alta concentração de sais influenciasse negativamente a corrida. Depois da eletroforese, o programa do seqüenciador ABI Prism 377 coletou os dados da detecção de fluorescência automática para a determinação do tamanho e quantificação dos fragmentos amplificados. O programa GeneScan Analysis v. 3.1.2 (Applied Biosystems) determinou o tamanho de cada produto marcado e comparou com a curva dos tamanhos padrão (ROX) com a curva dos fragmentos da amostra, enquanto que a avaliação dos alelos foi realizada utilizando o programa Genotyper v 2.5.2 (Applied Biosystems).

4.2.4.6. Análise estatística

4.2.4.6.1. Macho-Esterilidade Gênica Termo-Sensível – Os dados fenotípicos TGMS foram testados para a Normalidade, Curtose e Assimetria utilizando o Teste Lilliefors/Kolmogorov-Smirnov, com o programa GQMOL versão 2.1 (Cruz & Schuster, 2004).

4.2.4.6.2. Análise de segregação dos marcadores – A segregação dos marcadores foi avaliada através do programa GQMOL versão 2.1 (Cruz & Schuster, 2004) aplicando o teste estatístico qui-quadrado (χ^2) para uma proporção de segregação esperada de 1:2:1 para marcadores codominantes segregando numa população F2 com um nível crítico $p=0,001$ (2 g.l.). O Fator de Correção de Bonferroni foi aplicado aos dados de segregação (Cruz & Schuster, 2004). Os locos marcadores eventualmente mostrando distorção de segregação foram, posteriormente, analisados

para a natureza da distorção, se gamética ou zigótica. A natureza da distorção foi determinada pelo uso seqüencial de testes χ^2 (Lorieux et al., 1995). No primeiro teste, a hipótese nula assumindo igualdade das freqüências alélicas [Ho:p(A)=q(B)] foi testada por $\chi^2_{(GL=1)} = \left[\frac{(2np - n)^2 + (2nq - n)^2}{n} \right]$, onde n_{AA} é o número de homozigotos para alelos da variedade *O. sativa* spp. *indica* var. IR50, n_{AB} é o número de heterozigotos, n_{BB} é o número de homozigotos para alelos da variedade *O. sativa* spp. *indica* var. IR68945-4-33-4-14-48 (TGMS), $p(A) = n_{AA} + (n_{AB})/2$, $p(B) = n_{BB} + n_{AB}/2$ e n , o número total de indivíduos na população. Valores significantes de χ^2 são uma indicação inicial de seleção gamética, resultante, portanto, das diferenças nas freqüências alélicas ($p \neq q$) expressa através da vantagem seletiva de um genótipo sobre os outros. No segundo teste, uma avaliação do cruzamento aleatório sob Equilíbrio de Hardy-Weinberg (HWE) com uma distribuição esperada expressa pela equação $p^2+2pq+q^2$ foi testada por $\chi^2_{(GL=1)} = \left[\frac{(n_{AA} - np^2)^2}{np^2} + \frac{(n_{AB} - 2npq)^2}{2npq} + \frac{(n_{BB} - nq^2)^2}{nq^2} \right]$, onde p^2 é o número de homozigotos esperado para alelos da variedade IR50 (AA), $2pq$ é o número esperado de heterozigotos (AB), e q^2 é o número de homozigotos esperado para alelos da variedade IR68945-433-4-14-48 (BB). Valores significativos de χ^2 indicam que as freqüências genótípicas observadas diferem das esperadas. Ambos os testes são complementares e juntos eles indicam a natureza da distorção (Lorieux et al., 1995). Se os valores de χ^2 ou se o primeiro teste não for significativo, mas o segundo for, então o tipo de seleção é zigótica. Se os valores de χ^2 forem significativos para o primeiro teste e não significativos para o segundo, a seleção é gamética.

4.2.4.7. Mapeamento genético baseado em marcadores microsatélites

Um mapa genético de ligação baseado nos dados genotípicos da F2 para os marcadores microsatélites que foram polimórficos entre os parentais foi construído utilizando o programa MAPMAKER/EXP versão 3.0 (Lincoln et al., 1992a). Somente os marcadores genotipados em 77 ou mais indivíduos da F2 foram considerados para análise. Um valor de “LOD” de 3,0 e um máximo de frequência de recombinação de 0,30 foram utilizados para agrupar os marcadores microsatélites em potenciais grupos de ligação. Análise de três pontos e pontos múltiplos foi efetuada para encontrar a ordem mais provável dos locos dentro dos eventuais grupos de ligação observados. A ordem final dos locos em cada grupo de ligação foi testada pela remoção de um loco por vez e verificação das inconsistências. Os eventos de dupla recombinação foram examinados, e os resultados originais verificados novamente antes de se atribuir à ordem final dos grupos de ligação. As frequências de recombinação foram corrigidas baseadas na função de distância de mapa de Kosambi como executado pelo Mapmaker (Kosambi, 1944). A posição dos marcadores foi confirmada com os comandos “ripple” e “try” para permitir as comparações com os mapas de arroz publicados. O uso de marcadores mapeados por Brondani (2000) e aqueles publicados no Banco de Dados Gramene (Gramene, 2005) - Cornell SSR 2001 Map e IRMI 2003 Map – permitiram a rápida identificação dos grupos de ligação e a determinação das posições dos marcadores não ligados, de acordo com a nomenclatura dos cromossomos de arroz.

4.2.4.8. Detecção de locos TGMS pelo χ^2 de tabelas de contingência

A ligação entre locos marcadores e QTL TGMS foi testada pelo teste χ^2 numa tabela de

contingência 3x3 ($\chi^2 = \sum_1^k \frac{(o - e)^2}{e}$), com três classes de genótipos (AA; AB e BB) e três classes fenotípicas (0%-30%; 31-60% e >61% de fertilidade), nível crítico p = 0,001, 4 g.l., (Preacher, 2001). Esta análise foi realizada para comparar com os métodos descritos abaixo, baseados nos dados fenotípicos contínuos.

4.2.4.9. Análise de mapeamento de QTLs TGMS

O mapeamento de locos de características quantitativas baseado em dados fenotípicos contínuos foi realizado utilizando as análises de “single-point” e de “interval-mapping”. Um LOD mínimo de 3,0 foi assumido quando utilizando o programa MAPMAKER/QTL versão 1.1 (Lincoln et al., 1992b) para análise de mapeamento por intervalo. O programa QTL Cartographer (Basten et al., 2005) possibilita a análise de QTLs utilizando métodos de mapeamento “single-point” pelos testes de regressão linear e “F” utilizando planilhas mapmaker. A análise de “single-point” foi efetuada com regressão linear, onde a variável dependente foi o valor da característica (% de pólen estéril) e a independente foi o estado do alelo no loco. A proporção de variância fenotípica (σ^2_p) explicada pelo QTL foi calculada como o quadrado do coeficiente de correlação parcial (R^2). Os efeitos aditivos (d) e os efeitos de dominância (h) foram estimados como descrito por Lincoln et al. (1992b).

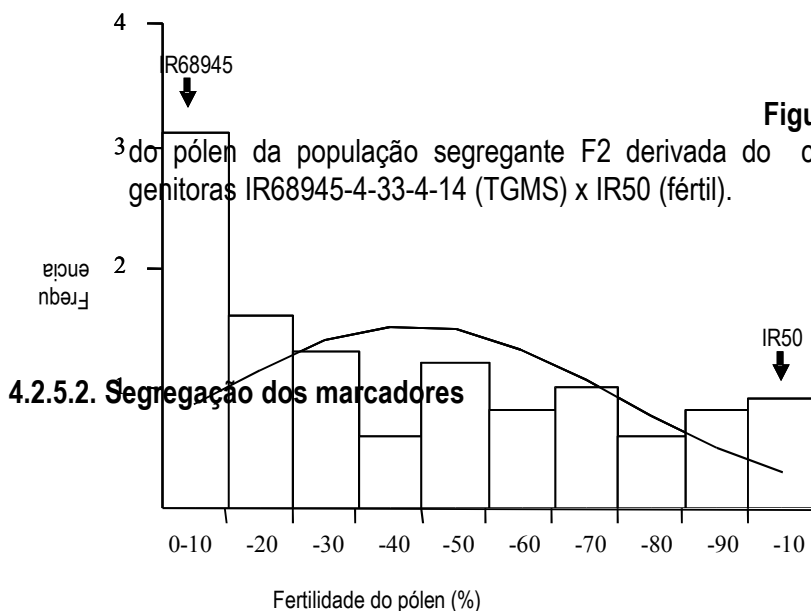
4.2.5. Resultados e Discussão

4.2.5.1. Distribuição e segregação dos dados de macho-esterilidade

A característica TGMS tem sido usualmente estudada sob premissa de que um ou poucos alelos podem estar envolvidos no seu controle. Na maioria dos estudos de mapeamento dos locos TGMS, a variação fenotípica tem sido tratada como dado binário (isto é, estéril x normal). Entretanto, apesar da observação de uma variação na fertilidade do pólen variando de 0 a 100% em plantas da F2 derivadas do cruzamento entre linhagens fértil e TGMS, cada planta tem sido simplesmente classificada em somente duas classes, isto é, estéril ou fértil. Esta simplificação pode ser vantajosa para detecção de genes principais, especialmente quando a estratégia de mistura de DNA tal como BSA (bulked segregant analysis) é utilizada (Michelmore et al., 1991), o que tem sido o caso de alguns estudos (Wang et al, 1995; Subudhi et al, 1997; Dong et al, 2000; Reddy et al., 2000). Nesses estudos, as plantas foram classificadas em apenas duas classes de acordo a percentagem de pólen corado pelo IKI. Por exemplo, plantas como menos de 5% de pólen corado foram consideradas estéreis, enquanto que plantas tendo mais de 50% de pólen corado foram classificadas como férteis no estudo desenvolvido por Dong et al. (2000). Uma escala ainda que ligeiramente diferente, considerou plantas com menos de 20% de pólen corado como estéreis,

enquanto plantas tendo mais de 20% de pólen corado como férteis (Reddy et al., 2000). Esta simplificação de classificação dos segregantes em apenas duas classes pode limitar a possibilidade de verificação da existência de mais de um loco TGMS na mesma população segregante.

A esterilidade do pólen parece ser uma característica complexa, controlada por vários locos que influenciam o número de vias metabólicas envolvidas no desenvolvimento do pólen. A distribuição contínua dos dados de esterilidade termo-sensível em populações segregantes F2 indica que a análise de QTLs (Lander & Botstein, 1989; Tanksley et al., 1989; Paterson et al., 1991) poderia ser apropriada ao mapeamento de regiões genômicas associadas com o controle TGMS. Os dados da macho-esterilidade gênica termo-sensível observados na população F2 estudada mostram uma distribuição com média e desvio padrão de 38,72 e 30,95% de pólen fértil, respectivamente (Figura 8; ANEXO V). Os dados não se ajustam a uma distribuição normal, mostrando uma assimetria de 0,47 ($p = 0,055$) (com ligeira inclinação para os valores mais baixos de fertilidade) e uma curtose de -0,96 ($p = 0,0012$) (relativamente achatada). O teste de Liilliefors para a distribuição normal indica que os dados não são normalmente distribuídos a um nível de significância de 5% ($D=0,13$; $p=0,079$). O coeficiente de variação (CV%) foi alto (79%). A mesma linhagem TGMS utilizada neste estudo (IR68945-4-33-4-14-48) foi também utilizada por Reddy et al. (2000) no desenvolvimento de uma população F2 para mapeamento de locos TGMS. Nesse estudo, a população F2 apresentou uma distribuição que não se ajusta a proporção 3:1 esperada para uma característica recessiva monogênica (Reddy et al., 2000). A população F2 segregando TGMS estudada mostrou uma distribuição fenotípica (Figura 8), portanto, que é contínua e não binária.



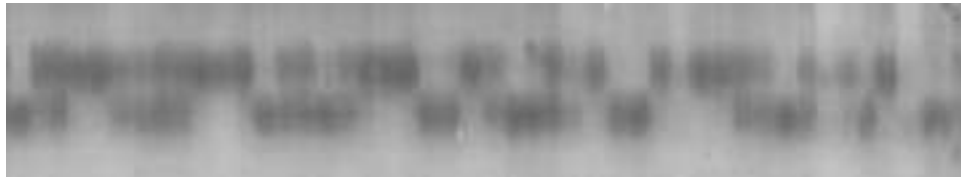
Embora um dos critérios usados para a seleção das linhagens genitoras tenha sido a maximização da distância genética entre a linhagem TGMS e a linhagem fértil, o polimorfismo de DNA entre as duas linhagens genitoras foi considerado baixo do ponto de vista do experimento, corroborando as estimativas de limitada distância genética entre elas (Alcochete, 2005 - Capítulo 1). Cento e dezesseis marcadores microssatélites foram selecionados para a análise de QTLs, mostrando o padrão claro de segregação entre das duas linhagens parentais. (Figura 8). A maioria dos marcadores mostrou uma proporção de segregação de acordo com as expectativas Mendelianas. Entretanto, dezessete marcadores (14,65%) mostraram segregação distorcida ($p < 0.001$), depois de aplicado o Fator de Correção de Bonferroni (Cruz & Schuster, 2004). Os marcadores distorcidos foram observados no cromossomo 2 (RM244, RM263B, OS1 e OG93), cromossomo 5 (RM7653 e RM7453), cromossomo 8 (RM7452, M38, OG85 e RM14B), cromossomo 9 (RM7576) e cromossomo 12 (RM7113, RM247 e RM28B). Um baixo nível de marcadores distorcidos foi esperado na população segregante visto que as duas linhagens parentais mostram uma alta similaridade genética (Alcochete, 2005 – Capítulo 1). A distorção de segregação observada (14,65%) é muito mais baixa do que a observada em cruzamentos envolvendo variedades *japonica* e *indica*, o qual pode variar de 17% a 70% (Xu et al., 1997). Entretanto, ela é mais alta do que a distorção (0,076%) recentemente reportada num cruzamento entre duas variedades com a característica de macho-esterilidade gênica fotoperíodo-sensível da subespécie *indica* (He et al., 1999).

Oito marcadores distorcidos (RM7452, RM7653, RM248, RM38, OG85, RM7453, RM244, RM214B) mostraram distorção a favor da linhagem genitora IR50; cinco marcadores (RM263b, RM7113, RM247, RM228a e RM7161) mostraram distorção em direção da linhagem genitora IR68945-4-33-4-14-48; três marcadores (OS11, OG93 e RM120) mostraram distorção em direção de heterozigotos; e o marcador RM7576 (0,86%) mostrou distorção em direção a ambos os homozigotos. Valores de χ^2 altamente significativos para todos os marcadores distorcidos foram observados ($p < 0.0001$) e a análise dos dados (Lorieux et al., 1995) indicou seleção zigótica como a causa da distorção da segregação. Harushima et al., (1996) e Lyttle (1991) enfatizam que o desvio das proporções de segregação Mendelianas esperadas tem sido observado em progênies de híbridos intra- e inter-específicos de arroz e sugeriram a causa desse desvio como sendo devida a fatores de distorção da segregação geneticamente ligados aos marcadores distorcidos, os quais

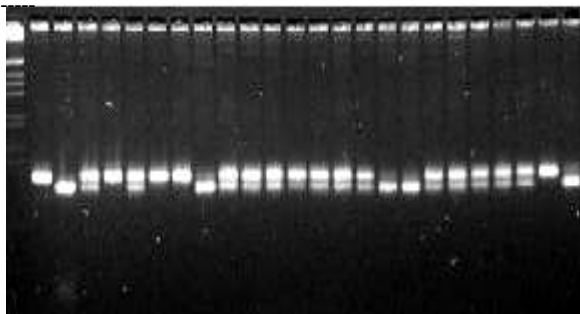
afetam a competição dos gametas ou o aborto do gameta ou do zigoto. Guiderdoni (1989) num estudo de progênies F2 obtidas de cinco híbridos entre arroz *indica* e *japonica* verificaram que a causa da distorção era a seleção gamética devido a esterilidade intervarietal durante a gametogênese. Não há evidências sugerindo uma relação de causa-efeito entre a seleção zigótica e o controle de TGMS no cruzamento entre IR68945-4-33-4-14-48 e IR50.

(a) 1Kb P1

P₁ P₂ F₁ ----- F₂



(b) M P₁ P₂ F₁ ----- F₂



(c)

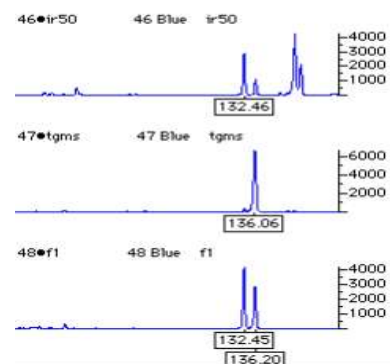


Figura 9 - Genotipagem de marcadores microssatélites de uma população F₂ derivada do cruzamento entre as linhagens genitoras *O. sativa* IR68945-4-33-4-14-48 (P₁) x *O. sativa* IR50 (P₂). (a)

Segregação dos alelos do marcador OG11 em eletroforese de gel de poliacrilamida 4% corada com nitrato de prata; (b) Segregação dos alelos do marcador OS9 em eletroforese de gel de agarose (3,5%) corado com brometo de etídio; (c) Electroferogramas do marcador RM7012 mostrando os genótipos da linhagem IR50 (fértil), IR68945-4-33-4-14-48 (TGMS) e sua F1. Os eixos verticais indicam intensidade de fluorescência medida por um seqüenciador de DNA ABI Prism 377 DNA. O eixo horizontal representa a estimativas dos tamanhos dos produtos em pares de base. M = 1 Kb Ladder.

4.2..5.3. Mapeamento de QTLs TGMS no genoma de arroz

Um mapa de ligação compreendendo 1.213,3 cM foi construído baseado nos dados de segregação da população F2. Noventa e cinco de um total de cento e dezesseis marcadores microssatélites testados foram reunidos em 11 grupos de ligação com uma média de 12,77 cM entre marcadores adjacentes e 8,6 cM marcadores por grupo de ligação. Para permitir a recombinação entre vários marcadores de posição conhecida no genoma de arroz, um valor de θ de 0,40 foi inicialmente usado. O mapa foi alinhado com informação de outros mapas genéticos do genoma de arroz (Wu & Tanksley, 1993; Akagi et al., 1996; Panaud et al., 1996; Chen et al., 1997; Brondani, 2000; Temnykh et al., 2002).

Associações significativas entre os genótipos marcadores e o fenótipo TGMS foram inicialmente testadas pelo teste de χ

² numa tabela de contingência. Para esse fim, indivíduos segregantes da F2 foram classificados em apenas três classes fenotípicas (0%-30%; 31-60% e >61% fertilidade), de forma similar ao que vem sendo feito noutros estudos. Usando esta metodologia, somente um marcador (RM7012), localizado no cromossomo 7, mostrou associação significativa com a característica TGMS ($\chi^2_{4df} = 6,567$, $p=0,0020$).

A análise de QTL inicialmente realizada pela análise individual de marcadores também discriminou o marcador RM7012 como associado a uma região controlando TGMS em arroz ($R^2 = 13.74$; $p = 0.0003$). A análise de mapeamento por intervalo detectou intervalos de marcadores associados de forma significativa com TGMS com valores de LOD acima de 3,00 (Figura 10). A região incluindo o marcador RM7012 no cromossomo 7 foi novamente detectada como associada ao controle de TGMS em arroz, mas duas outras regiões no cromossomo 1 (intervalo RM9-RM246) e cromossomo 12 (intervalo RM228A-RM247), que não haviam sido descritas ainda, foram também detectadas. A metodologia de mapeamento de QTLs baseada na distribuição contínua dos dados fenotípicos foi eficiente na detecção de outras regiões cromossômicas associadas ao controle da característica TGMS que não foram detectadas quando o fenótipo foi tratado como característica binária.

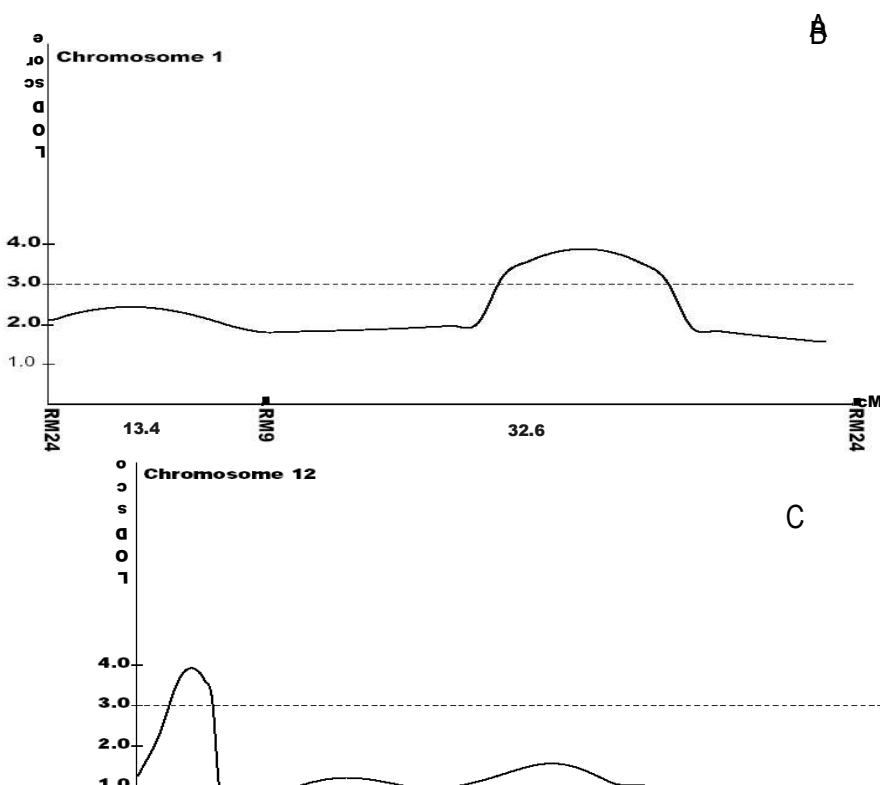


Figura 10 – Mapeamento de TGMS QTLs baseado na análise de mapeamento por Intervalo Simples dos cromossomos de arroz: (a) QTL TGMS localizado no cromossomo 1 (LOD = 3,90); (b) QTL TGMS localizado no cromossomo 7 (LOD = 3,19); (c) QTL TGMS localizado no cromossomo 12 (LOD = 3,92).

O principal TGMS QTL mapeado neste estudo, associado ao marcador RM7012, está localizado no cromossomo 7 (LOD = 3,90; $d = -20,73$; $h = 0,34$; $\sigma_{2p} = 19,20$), onde o loco *tms2* foi previamente mapeado (Yamaguchi et al., 1997; Lopez et al., 2003). A estimativa do efeito aditivo sugere que o alelo IR68945-4-33-4-14-48 contribui para a redução da fertilidade neste QTL, enquanto a dominância é negligível. O cromossomo 7 de arroz também inclui um loco QTL associado ao controle genético da macho-esterilidade gênica fotoperíodo-sensível (PGMS) e um QTL para a esterilidade do pólen (Doi et al., 1998). Análises suplementares são necessárias para a verificação das relações alélicas entre o QTL TGMS detectado no presente estudo e outras regiões previamente associadas a macho-esterilidade descritas no mesmo cromossomo (Yamaguchi et al., 1997; Doi et al., 1998; Lopez et al., 2003;).

No cromossomo 1 (Figura 10), os dados indicam um QTL TGMS na vizinhança dos intervalos incluindo os marcadores RM9-RM246 (LOD = 3,19; $d = -12,64$; $h = -51,35$; $\sigma_{2p} = 74,70$). A sobredominância é reconhecida numa localização cromossômica se a proporção do efeito de dominância em relação ao valor absoluto dos efeitos aditivos é maior do que a unidade. Os valores de d e h sugerem que o alelo IR68945-4-33-4-14-48 contribui para a redução da fertilidade do pólen neste QTL e mostra sobredominância. QTLs para esterilidade da espiguetta (Li et al., 1997; Nagata et al., 2002), esterilidade da geração F2 (Li et al., 1997; Wang et al., 2002) e fertilidade do pólen (Li et al., 1997) foram localizados neste mesmo cromossomo.

Os dados também indicam um QTL TGMS no intervalo RM228A-RM247 no cromossomo 12 (LOD = 3,92; $d = -10,99$; $h = 44,64$; $\sigma_{2p} = 69,40$). As estimativas do efeito aditivo para este QTL sugerem que alelo de IR68945-4-33-4-14-48 reduz a fertilidade do pólen. Entretanto, a sobredominância foi também detectada, um efeito do alelo proveniente da linhagem genitora IR50. Um QTL de macho-esterilidade gênica fotoperíodo-sensível (PGMS) foi identificado no cromossomo 12 (Mei et al., 1999) bem como um QTL para a fertilidade do pólen (He & Xu, 2000). Dois locos PGMS foram mapeados no cromossomo 7 e 12 (Mei et al., 1999). Experimentos posteriores deverão revelar a relação entre essas regiões e os QTLs TGMS detectados no presente estudo.

Os genes TGMS devem ser transferidos para linhagens com diferentes *background* genético para maximizar a chance de se obter boas combinações de híbridos. A introgressão de genes TGMS para linhagens de arroz geneticamente diversas pode ser facilitada pela informação de marcadores moleculares. O uso de marcadores moleculares no melhoramento assistido facilitará e acelerará o processo de introgressão de genes TGMS para linhagens com *background* genético mais amplo. Alguns esforços têm sido feitos recentemente nesta direção (Lopez et al., 2003). Os marcadores microssatélites flanqueando os QTLs TGMS serão úteis como ferramentas de seleção na obtenção de linhagens TGMS para a produção de híbridos de arroz. Os intervalos de marcadores descobertos neste estudo podem ser utilizados como um ponto de partida nestes esforços. O sistema de duas linhagens baseado na característica TGMS poderá ser útil na produção de arroz híbrido em países tropicais tal como o Brasil. Uma metodologia possível será o cultivo de linhagens TGMS e linhagens férteis em regiões onde a média da temperatura ambiental durante a fase de iniciação da panícula e floração seja maior do que 25-30°C. Há um grande número de regiões que se ajustam a este requisito em períodos específicos do ano no país. As linhagens TGMS deverão ser testadas para identificar as regiões geográficas e estações onde a expressão da esterilidade é completa e a expressão da fertilidade máxima.

5. DISCUSSÃO GERAL

5.1. Material Vegetal

O material vegetal utilizado neste estudo constituiu-se de cultivares de arroz *indica* conhecidos (METICA1, BR IRGA, BG90-2, CICA 8, IR50, IR36), linhagens TGMS (TGMS70977_12, TGMS70977_26, TGMS68940, TGMS70987, TGMS68935, TGMS71018, TGMS70989, TGMS68944 e TGMS68945), alguns acessos de arroz japonica (TomoeMochi, IAC 166) e uma vasta amostragem de variedades tradicionais de arroz coletadas no território brasileiro. Esse material biológico constitui parte dos cerca de 10,000 acessos de arroz do Banco de Germoplasma de Arroz da EMBRAPA Arroz e Feijão (Abadie et al., 2005). Parte desses acessos tem sido caracterizada morfológicamente para vários descritores morfológicos e para alguns índices agrônômicos. Este tipo de avaliação apresenta limitações concernentes à definição e validação de caracteres neutros, ao custo do estabelecimento de manutenção de um experimento de campo, tempo de avaliação e interação genótipo x ambiente. O recurso à caracterização molecular, envolvendo marcadores moleculares microssatélites, permite-nos estabelecer com maior clareza os vínculos genéticos entre os diferentes acessos, linhagens ou cultivares e, conseqüentemente, uma

melhor organização do germoplasma no Banco de Germoplasma. Permite ainda enriquecer o conhecimento para melhorar a estratégia de coleção futura de germoplasma, e possibilitar mais eficazmente a utilização desse material em programa de melhoramento.

Várias linhagens de arroz foram selecionadas para cruzamentos específicos com linhagens TGMS. O cruzamento envolvendo as linhagens IR50 e IRTGMS-68945-4-33-14-48 resultou em quantidades suficientes de sementes que possibilitaram a obtenção de uma população segregante F2, adequada a estudos de mapeamento genéticos. Esta população, constituída por 93 indivíduos, seus genitores e híbrido F1, foi genotipada utilizando os marcadores informativos obtidos do *screening* de marcadores microssatélites polimórficos no cruzamento, com vistas ao mapeamento de genes TGMS no genoma do arroz.

A análise de agrupamento baseada em estimativas de distância genética de alelos “partilhados” utilizando o método de Neighbour-Joining permitiu identificar grupos de acessos de arroz muito próximos geneticamente. O dendrograma resultante mostra os acessos distribuídos em dois grupos, um constituído por 55 acessos (Cluster 1) e outro composto por 243 acessos (Cluster 2). O Cluster 1 inclui acessos que não são claramente classificados como variedades indica ou japonica. Nenhum dos acessos tipicamente *japonica* incluídos neste estudo (ex. TomoeMochi, IAC 166) bem como acessos típicos de *indica* (ex. IR-50, Cica 8, BG 90-2) estão próximos aos 55 acessos do Cluster 1. Nota-se que num estudo similar (Pessoa Filho, 2004), esses 55 acessos formaram um cluster distinto dos acessos de indica e de japonica.

O segundo grupo principal (Cluster 2), o maior grupo, é composto de arroz *indica* e *japonica* subdivididos em cluster menores. No geral, os acessos de background genético conhecido como *japonica* agruparam com outros tipos de *japonica*. O mesmo foi observado em relação aos tipos indica. Este estudo foi importante para se ter uma primeira indicação do background genético da maioria dos acessos tradicionais na coleção de arroz. Alguns exemplos chamam a atenção, como é o caso de nove acessos de arroz vermelho que foram genotipados pela primeira vez com os painéis múltiplos de marcadores microssatélites. Esses acessos têm sido coletados em pequenas vilas do Nordeste do Brasil, onde tradicionalmente os agricultores dão maior valor ao arroz vermelho do que o arroz normal por razões culinárias. Os acessos Arroz Vermelho 1, Arroz Vermelho 2, Arroz

Vermelho 3 e Arroz Vermelho 9 formam um grupo com Ciwini e Basmati 370, conhecidos como tipo *indica* (Beló, 2001; Fatima et al., 2002). Basmati 370 é um arroz aromático, com boa qualidade e grãos extra-longos e alongamento depois de cozido. Entretanto, este cluster também inclui uma variedade chamada Mochi Gome, que é um nome comum para variedades de arroz marrom, glutinosas cultivadas no Japão. Os outros acessos de Arroz Vermelho também agrupam com tipo *indica* em outros grupos. Por exemplo, Arroz Vermelho (7) e Arroz Vermelho (8) agrupam com Oryzica 1, Cica 8 e BG 90-2, enquanto Arroz Vermelho (4) agrupa com as linhagens TGMS *indica*. Os acessos Arroz Vermelho 5 e Arroz Vermelho 6 pertencem a um grupo separado que não inclui nenhum tipo *indica* bem conhecido e é composto na sua maioria por acessos com nomes muito comuns. Pode-se notar que as variedades tradicionais chamadas de Arroz Vermelho são muito diversas e podem ser distinguidas com base na análise de genotipagem multilocos.

Todas as linhagens macho-estéreis gênicas termo-sensíveis (TGMS) estão agrupadas juntas (linhagens TGMS70977_12, TGMS70977_26, TGMS68940, TGMS70987, TGMS68935, TGMS71018, TGMS70989, TGMS68944 e TGMS68945), formando um grupo com as linhagens *indica* IR 36, IR50 e Metica 1. A linhagem TGMS68935 foi cruzada com IR50 (normal) com vista o desenvolvimento de populações segregantes utilizadas no mapeamento de locos gênicos do controle de TGMS no genoma de arroz.

É interessante notar que acessos com o mesmo nome comum, algumas vezes agrupam juntos (ex. Agulha (1), Agulha (2) e Agulha (3) ou Carolina (1) e Carolina (2)) enquanto que outras vezes não (ex. Caqui ou DeAbril). Isto é uma indicação clara de que os genótipos com o mesmo nome comum são geneticamente diferentes, especialmente quando variedades tradicionais são consideradas.

5.2. Genotipagem utilizando painéis múltiplos de marcadores

A genotipagem semi-automática utilizando painéis de microssatélites marcados com fluorescência e genotipados em seqüenciador de DNA oferece muitas vantagens sobre as técnicas tradicionais baseadas em PAGE/coloração com prata ou resolução em agarose corada com brometo de etídio. A maior vantagem é a maior quantidade de informação genotípica em cada gel

analisado pela a possibilidade de juntar o produto de várias PCRs ou o produto da PCR de vários primers simultâneos em um único gel. Outra vantagem desse sistema de genotipagem é a significativa melhoria da determinação do tamanho dos alelos devido ao uso de fragmentos internos de tamanho padrão, a disponibilidade de algoritmos de inferência do tamanho em pares de bases dos alelos amplificados e a automação da obtenção eletrônica dos dados obtidos no seqüenciador de DNA. Em geral, a genotipagem em seqüenciador de DNA aumenta a velocidade e a acurácia da coleta de dados e o seu processamento. A alta sensibilidade de detecção a laser dos produtos de PCR também reduz o volume necessário (e, portanto, o custo) da PCR e possibilita a detecção de locos de difícil amplificação. O uso de fluorescência baseada na análise automática de fragmentos de restrição foi primeiro reportado por Carrano et al. (1989). O método foi então adaptado e melhorado para a análise de microsatélites posteriormente (Edwards et al., 1991; Ziegler et al., 1992).

Os métodos automáticos de genotipagem com SSR estão substituindo gradualmente os sistemas manuais na pesquisa genética de plantas. Esses novos métodos facilitam a aplicação de marcadores microsatélites para o mapeamento (Rhodes et al., 1998; Ponce et al., 1999), análise de pedigree (Lexer et al., 1999), fingerprinting de acessos (Carrano et al., 1989), e avaliação de diversidade genética (Diwan & Cregan, 1997; Macaulay et al., 2001). A gestão mais eficiente de uma coleção de germoplasma é também facilitada (Mitchell et al., 1997; Belo, 2001; Pessoa Filho, 2004).

Em arroz, Blair et al. (2002) descreveu o uso de 27 marcadores SSR fluorescentes organizados em quatro painéis para o estudo de diversidade. Mais, recentemente Belo (2000) desenvolveu um pentaplex e um hexaplex e Pessoa Filho (2004) um segundo pentaplex para genotipagem de locos microsatélites de arroz, totalizando 16 locos distribuídos em diferentes cromossomos de arroz e passíveis de análise com apenas três reações de PCR. Como ferramenta para o “scan” de genomas, os painéis múltiplos baseados em marcadores microsatélites podem ser eficazmente utilizados, por exemplo, no mapeamento genômico, na avaliação da diversidade genética, e na estimativa das relações genéticas entre acessos do banco de germoplasma de *O. sativa*.

A principal limitação da genotipagem em grande escala utilizando microsatélites é a edição

manual da determinação alélica em cada loco. Embora a determinação automática dos alelos venha sendo aplicada eficazmente, a capacidade de realizá-la com acurácia só é assegurada pela inspeção manual dos electroferogramas para confirmação. A limitação é particularmente importante na análise de microsatélites baseados em repetições de dinucleotídeos, visto que neste caso o nível de stuttering observado na detecção alélica é maior do que em repetições de trinucleotídeos ou tetranucleotídeos. Pálsson et al. (1999) descreveram um método paramétrico (Decode-GT) para controlar a qualidade da chamada dos alelos do programa TrueAllele (TA). Este método possibilita a leitura dos dados e classifica os alelos em “bons”, “ruins” e “ambíguos”. O método descarta os alelos “ruins” e sugere uma revisão manual para os ambíguos, reduzindo o tempo de edição manual e o erro de chamada dos alelos para 1%. Esta estratégia pode ajudar a reduzir o tempo consumido na edição manual dos alelos, principalmente quando a genotipagem envolve várias centenas de acessos.

Os painéis múltiplos de microsatélites utilizados no Capítulo 1 já haviam sido testados em outros estudos (Beló, 2001; Pessoa Filho, 2004) e mostraram eficiência na separação de variedades, linhagens ou cultivares de arroz entre as subespécies *indica* e *japonica*, bem como seus híbridos, e na detecção de alto nível de polimorfismo. Resultados similares foram encontrados no presente estudo. A maioria dos locos mostrou amplitude alélica dentro do esperado com exceção de OG61 e OG81, marcadores baseados em repetições de dinucleotídeos, com amplitudes alélicas de 64 pb e 32 pb contra 58 pb e 18 pb esperados. Os locos RM335 e RM418, repetições de trinucleotídeos, também mostram variação dos tamanhos dos alelos diferente do esperado, 66 pb (73 pb esperados) e 63 pb (52 pb esperados), respectivamente. Seis marcadores apresentaram amplitude alélica maior do que o observado por Belo (2000) e menor do que o descrito por Pessoa Filho (2004). A causa deste resultado pode estar na diferença de tamanho amostral nos diferentes estudos e potencial detecção de alelos raros devido a diferenças de composição das amostras.

A quantidade de dados faltantes, na genotipagem, pode causar distorção durante a análise de comparação dos genótipos multilocos. Amostras com mais de quatro locos faltantes foram retiradas da análise. Assim, de um total de 364 amostras originalmente genotipadas, somente 298 acessos foram incluídos na análise final. Os dados faltantes têm sido descritos como alelos nulos, presentes nos acessos analisados, ou falhas de amplificação. De um modo geral as falhas

ocorreram nos locos microssatélites marcados com fluorescência amarela (HEX), principalmente os locos RM475 (31,9%), OG61 (29,9%), RM259 (13,3%) e um loco 6-FAM, RM418 (15,8%). Os marcadores não foram sintetizados ao mesmo tempo e o tempo pós-síntese pode ter efeito na eficiência da detecção da fluorescência.

5.3. Análise de Agrupamento e Estruturação das Amostras

A análise de agrupamento dos 298 acessos analisados pelo método Neighbour-Joining, mostrou a formação de dois grupos principais, com vários subgrupos neles contidos. O padrão de agrupamento foi utilizado como modelo para analisar a estrutura genética da coleção utilizando as estatísticas F de Wright (Wright, 1978; Weir & Cockerham, 1984) desenvolvidas no programa GENETIX 4.03 (Belkhir, 2004) e o método MCMC do programa STRUCTURE (Pritchard et al., 2000a) para determinar os níveis de diferenciação genética entre os subgrupos produzidos. Foram estimados os coeficientes de endogamia nos subgrupos (F_{IS}), coeficiente de endogamia total (F_{IT}) e o índice de fixação devido à endogamia (F_{ST}). Os valores médios de F_{IS} e F_{IT} revelaram níveis elevados de endogamia nos subgrupos e nos indivíduos (média $F_{IS} = 0,779$ e $F_{IT} = 0,798$). Entretanto, os valores médios do índice de fixação F_{ST} são relativamente moderados (0.086). Os valores baixos de F_{ST} parecem indicar um baixo nível de diferenciação entre os distintos grupos de acessos. Não foi detectada correlação significativa entre os agrupamentos utilizados como modelo, baseados em dados de distância genética, e os subgrupos que representam a subestruturação da coleção, determinados pela análise MCMC e as estatísticas F de Wright.

5.4. Diversidade Genética e Banco de Freqüências Alélicas

Um banco de dados de freqüências alélicas de arroz foi constituído, resultado da análise dos 16 locos microssatélites utilizados na genotipagem dos 288 acessos de arroz. Oito mil e trinta e seis alelos (média de 502,25) foram detectados pelos três painéis múltiplos (16 microssatélites), dos quais 4.386 (54,57%) representam os alelos mais freqüentes (com freqüência acima de 40%). Os alelos mais comuns representam 54,57% dos alelos genotipados, evidenciando uma alta porcentagem de alelos detectados. A média de alelos mais comuns detectados em cada loco foi de 274, 125, variando de 110 para o loco OG106 a 463 para o loco RM420. Estes números são

inferiores aos obtidos por Pessoa Filho (2004), que, no entanto, analisou um número maior de acessos (468). O desenvolvimento de bancos de frequências alélicas de locos microssatélites específicos para cada uma das subespécies do arroz (Beló 2001; Pessoa Filho, 2004) é uma ferramenta poderosa e com grande aplicação em estudos de diversidade genética, proteção de variedades, pureza genética e melhoramento genético.

5.5. Coleção Nuclear

Desde a sugestão de Frankel e Brown (1984) sobre a criação de coleções nucleares para melhorar a gestão de bancos e germoplasma e permitir o seu uso em programas de melhoramento, várias coleções nucleares para diversas culturas alimentares têm sido desenvolvidas. As estratégias são variadas e baseiam-se na diversidade de caracteres morfológicos, e têm evoluído de acordo com a evolução da biotecnologia, pelo que inclui dados isoenzimáticos, dados de marcadores RAPDs e mais recentemente dados de marcadores microssatélites. O Banco de germoplasma de Arroz da Embrapa possui uma enorme quantidade de acessos/cultivares e linhagens de arroz, tornando-se necessária o desenvolvimento de ferramentas tais como banco de frequências alélicas e coleções nucleares para a sua gestão eficaz. Este estudo constitui mais uma etapa na criação de coleções nucleares baseados na diversidade genética que satisfaçam os propósitos apresentados e representem toda coleção de germoplasma.

A coleção nuclear constituída neste trabalho satisfaz os requisitos básicos para a sua formação (70-80% riqueza alélica e 10-20% do tamanho da amostra inicial) e pode contribuir para o melhoramento da gestão do BGA. O desenvolvimento de programas de informática (ex. COREX – Core Extraction, M.E. Ferreira, não publicado) para a extração de subamostras, baseadas em parâmetros genéticos resultantes do uso de ferramentas genômicas (marcadores microssatélites) proporciona um mecanismo importante no desenvolvimento de coleções nucleares. As coleções nucleares passam a ser formadas, cada vez mais, com base na informação contida no genoma dos seus acessos. Uma coleção nuclear constituída pode facilitar uma melhor utilização dos recursos genéticos conservados em bancos de germoplasma.

5.6. O Sistema TGMS de Arroz

O entendimento do mecanismo de controle genético da macho-esterilidade gênica termo-sensível (TGMS) em arroz é importante para a produção sustentável de sementes de arroz híbrido nos trópicos. Como alternativa a algumas limitações do sistema CMS, o sistema TGMS não requer linhagens mantenedoras (B) na produção de sementes híbridas; não necessita de linhagens restauradoras (R), visto que plantas TGMS podem restaurar sua fertilidade quando sob temperatura apropriada; e reduz o risco de vulnerabilidade genética, pelo fato da macho-esterilidade não ser dependente do citoplasma, e os híbridos poderem ser desenvolvidos com *backgrounds* genéticos diversos. Esta última característica é de especial interesse, sobretudo depois de experiências que levaram a grandes perdas agrônômicas e econômicas devido ao uso intensivo de materiais com o mesmo *background* genético, tais como o citoplasma-T usado na indústria do milho híbrido, que foi afetado pela ferrugem da folha, há algumas décadas (Levings, 1993).

A identificação e uso de diferentes alelos TGMS em arroz possibilita uma ampla base genética para o desenvolvimento de arroz híbrido. Para podermos explorar este sistema de duas linhagens torna-se necessário avaliar as linhagens que apresentam tal comportamento, estimar a robustez da expressão da característica numa faixa de temperaturas e compreender o seu controle genético com o vista a utilização do gene no desenvolvimento de cultivares.

O uso de marcadores moleculares no melhoramento assistido irá certamente facilitar e acelerar o processo de introgressão de genes TGMS em background genéticos mais amplos (Lopez et al., 2003) e elucidar a base molecular da característica TGMS. A tecnologia dos marcadores moleculares tem sido efetiva na clonagem de genes baseada no mapeamento genético de característica monogênicas (Martin et al., 1993) bem como poligênicas (Frary et al., 2000). A clonagem posicional tem proporcionado alguns resultados importantes na identificação e isolamento de QTLs envolvidos no controle de caracteres complexos (Buckler & Thornsberry, 2002). A aplicação de associações genéticas em algumas culturas (Thornsberry et al., 2004; Rafalski, 2002; Morgante & Salamini, 2003) possibilita a associação de dados de marcadores e de sequenciamento de DNA com a variação fenotípica, sem a necessidade de populações segregantes, como exigido pela clonagem posicional.

5.7. Distribuição e Segregação dos dados de Macho-Esterilidade

A característica TGMS tem sido estudada sob a premissa de que apenas um ou alguns poucos locos estão envolvidos no seu controle. NA maioria dos estudos de mapeamento dos locos TGMS, a variação fenotípica tem sido tratada como binária (isto é, estéril x normal). Assim, não obstante o fato da fertilidade do pólen variar de 0 a 100% em plantas F2 derivadas de cruzamentos entre linhagens férteis e linhagens TGMS, cada planta é classificada em apenas uma de duas classes, isto é, estéril ou fértil. Esta simplificação pode ser desvantajosa para a detecção de genes de grande efeito, especialmente quando uma estratégia de mistura de DNA, tal como BSA (bulk segregant analysis) é utilizada (Michelmore et al., 1991), como tem sido visto em alguns estudos (Wang et al, 1995; Subudhi et al, 1997; Dong et al, 2000; Reddy et al., 2000). Nesses estudos, as plantas têm sido classificadas em apenas duas classes de acordo com a percentagem de pólen corado com IKI. Por exemplo, plantas com menos de 5% de pólen corado são consideradas estéreis, enquanto plantas com mais de 50% são classificadas como férteis (Dong et al., 2000). Uma escala ligeiramente diferente considera as plantas com mais de 20% de pólen corado como férteis foi também descrita (Reddy et al., 2000). Esta simplificação na classificação de segregantes em apenas duas classes pode limitar a possibilidade de detecção de mais de um loco TGMS numa população segregante.

A fertilidade do pólen parece ser uma característica complexa, controlada por vários locos que influenciam um número de vias metabólicas envolvidas no desenvolvimento do pólen. A distribuição contínua dos dados de esterilidade termo-sensível em populações segregantes F2 parece indicar o mapeamento de QTL (Lander & Botstein, 1989; Tanksley et al., 1989; Paterson et al., 1991) como sendo uma análise apropriada para o mapeamento de regiões genômicas do genoma de arroz associadas ao controle de TGMS. Os dados de macho-esterilidade termo-sensível observados na população F2 estudada mostra uma distribuição com uma média e desvio padrão de 38,72 e 30,95% de pólen fértil, respectivamente. Os dados não se adequam a uma distribuição normal, mostrando uma assimetria de 0,47 ($p=0,055$) (em direção aos valores mais baixos de fertilidade) e uma curtose negativa ($-0,96$; $p=0,0012$) (relativamente achatada). O teste de Lilliefors/Kolmogorov para a distribuição normal indica que os dados não estão normalmente distribuídos a um nível de significância de 5% ($D=0,13$; $p=0,079$). O coeficiente de variação (cv%) foi alto (79%). A linhagem TGMS utilizada neste estudo (IR68945-4-33-4-14-48) foi também utilizada em outro cruzamento com o objetivo de mapear locos TGMS e a população F2 derivada desse cruzamento apresentou uma distribuição da fertilidade do pólen que não se ajusta a proporção esperada 3:1 para uma característica monogênia recessiva (Reddy et al., 2000). A população segregante F2 TGMS utilizada neste estudo mostrou uma distribuição fenotípica (Figura 1) contínua, e que, portanto não é binária.

5.8. Segregação dos Marcadores

Embora um dos critérios utilizados para a seleção das linhagens genitoras tenha sido maximizar a distância genética entre a linhagem TGMS e a linhagem fértil, o polimorfismo de DNA entre as duas linhagens genitoras foi considerado baixo a partir dos dados do experimento, corroborando a relativa pequena distância entre elas (Alcochete, 2005). Cento e quarenta e seis (22.61%) marcadores microssatélites de um total de 659 testados mostraram polimorfismo entre as duas variedades genitoras. Desses marcadores polimórficos, noventa e seis (65.75%) eram não marcados e cinquenta (34.25%) eram marcados com fluorescência. Os marcadores não marcados dividiam-se em cinquenta e quatro (56.25%) do tipo RM, trinta e quatro (35,42%) OG, e oito (8.33%) OS. Os marcadores fluorescentes compreendiam vinte e dois (44,0%) tetra-, três (6,0%) tri-, seis (12,0%) di-, e dezenove (38.0%) do tipo complexo. Cento e dezesseis microssatélites foram

polimórficos e mostraram um padrão de segregação claro entre os dois genitores. A maioria dos marcadores mostrou uma proporção de segregação de acordo com as proporções Mendelianas esperadas. Entretanto, dezessete marcadores (14.65%) mostraram distorção da segregação ($p < 0.001$), depois de aplicado o Fator de Correção de Bonferroni (Cruz & Schuster, 2004). Os marcadores distorcidos foram observados nos cromossomos 2 (RM244, RM263B, OS1 e OG93), cromossomo 5 (RM7653 e RM7453), cromossomo 8 (RM7452, M38, OG85 e RM14B), cromossomo 9 (RM7576) e cromossomo 12 (RM7113, RM247 e RM28B). Devido à similaridade do genoma das linhagens genitoras, um baixo nível de distorção era esperado na população segregante. A distorção de segregação detectada (14,65%) é mais baixa que a observada no cruzamento envolvendo variedades *japonica* e *indica* (17% - 70%; Xu et al., 1997) Entretanto, ela é mais alta do que a distorção relatada (0.076%) em um cruzamento entre duas variedades de *indica* macho-estéreis fotoperíodo-sensíveis (He et al., 1999).

Oito marcadores distorcidos (RM7452, RM7653, RM248, RM38, OG85, RM7453, RM244, RM214B) favoreceram a linhagem genitora IR50; cinco (RM263b, RM7113, RM247, RM228a, RM7161) mostraram distorção na direção do genitor IRTGMS-68945-4-33-4-14-48; três (OS11, OG93, RM120) mostraram distorção na direção dos heterozigotos, e o marcador RM7576 (0.86%) mostrou seleção em favor de ambos os homozigotos. Foram obtidos valores altamente significativos de χ^2 para todos os marcadores distorcidos, sugerindo que a seleção zigótica foi a causa da distorção. Harushima et al., (1996) e Lyttle (1991) reportaram que o desvio da proporção de segregação Mendeliana esperada tem sido observado em progênies de cruzamentos intra- e inter-específicos e a causa desses desvios são os fatores de distorção da segregação que afetam a competição entre os gametas ou causam a falha do gameta ou zigoto. Guiderdoni et al., (2001) num estudo de progênies F2 e cultura de anteras de cinco híbridos intervarietais de arroz entre *indica* e *japonica* verificaram que os desvios se deviam à seleção gamética resultado da esterilidade durante a gametogênese. Não há evidência sugerindo a relação causa-efeito entre a seleção do zigoto e o controle de TGMS nos cruzamentos entre IR68945-4-33-4-14-48 e IR50.

5.9.. Mapeamento de QTLs para TGMS no Genoma de Arroz

Um mapa de ligação (1.261,90 cM) foi construído baseado nos dados de segregação da

população F2. Noventa e cinco marcadores microssatélites foram reunidos em 11 grupos de ligação com uma média de 12,7 cM entre locos marcadores adjacentes e 8,6 marcadores por mapa. Para permitir a recombinação entre vários marcadores de posição conhecida no genoma de arroz (www.gramene.org), utilizamos $\theta = 0.40$. O mapa foi alinhado com informação de outros mapas genéticos do genoma de arroz (Brondani, 2000; Wu & Tanksley, 1993; Panaud et al., 1996; Chen et al., 1997; Akagi et al., 1997; Temnykh et al., 2002).

Associação significativa entre os genótipos dos marcadores e a característica TGMS foi inicialmente testada pelo teste de χ^2 numa Tabela de Contingência. Para este fim, os indivíduos F2 foram classificados em apenas três classes fenotípica (0%-30%; 31-60% e >61% fertilidade), similar às utilizadas em estudos de mapeamento de locos TGMS. Utilizando este método, somente um marcador (RM7012), localizado no cromossomo 7, mostrou associação significativa com a característica TGMS ($\chi^2_{4df} = 16,567$, $p=0.0020$).

A análise de QTLs foi realizada inicialmente pelo método de análise simples e também discriminou o marcador RM7012 ($R^2=13,74$; $p=0.0003$). A análise por mapeamento de intervalo detectou intervalos de marcadores significativamente associados à TGMS com valores de LOD acima de 3,00. A região abrangendo o marcador RM7012 no cromossomo 7 foi novamente detectada como associada ao controle de TGMS em arroz, mas outras regiões genômicas no cromossomo 1 (intervalo RM9-RM246) e cromossomo 12 (RM228A-RM247), ainda não mapeadas, foram também detectadas. O mapeamento de QTLs baseado na distribuição continua dos dados fenotípicos foi eficiente na detecção de várias regiões cromossômicas associadas ao controle da característica TGMS.

O principal QTL TGMS mapeado neste estudo, associado ao marcador RM7012, parece estar numa posição cromossômica diferente do loco *tms2* (Yamaguchi et al., 1997; Lopez et al., 2003). O cromossomo 7 de arroz também contém um QTL associado ao controle genético da macho-esterilidade gênica fotoperíodo-sensível (PGMS) e um QTL para a fertilidade do pólen (Doi et al., 1998). Análises posteriores se fazem necessárias para verificação das relações alélicas entre os TGMS QTL detectados no presente trabalho e as outras regiões associadas a macho-esterilidade descrita (Yamaguchi et al., 1997; Lopez et al., 2003; Doi et al., 1998).

Os dois outros QTLs para TGMS foram descritos pela primeira vez nos cromossomos 1 e 12. No cromossomo 1, os dados indicam um QTL na proximidade do intervalo delimitado pelos marcadores RM9-RM246. QTLs para esterilidade da espiguetta (SPKST) (Li et al., 1997; Nagata et al., 2002), esterilidade da geração F2 (F2GST) (Li et al., 1997) e fertilidade do pólen (Li et al., 1997) foram localizados no mesmo cromossomo. Os dados também indicam um QTL para TGMS no intervalo RM228A-RM247 no cromossomo 12. Um QTL para PGMS foi identificado no cromossomo 12 (Mei et al., 1999) bem como um QTL para a fertilidade do pólen (He & Xu, 2000). Dois locos PGMS foram mapeados nos cromossomos 7 e 12 (Mei et al., 1999). Experimentos posteriores deverão revelar a relação entre essas regiões e o QTL para TGMS detectado no presente trabalho. Os dados identificaram três QTLs para a macho-esterilidade gênica termo-sensível no genoma de arroz. Dois desses QTLs foram pela primeira vez associados ao controle genético da característica TGMS (cromossomos 1 e 12). O terceiro QTL mapeia num cromossomo onde o loco *tms2* foi mapeado (Yamaguchi et al., 1997), mas não foram realizados testes alélicos para clarificar se essas duas regiões são as mesmas.

O sistema baseado em duas linhagens para a característica TGMS poderá ser utilizado na produção de arroz híbrido num país tropical como o Brasil. Uma metodologia poderá ser o crescimento de linhagens TGMS e a linhagem genitora fértil em regiões onde a média de temperatura ambiental durante a iniciação da panícula e estágio de antese é mais elevada do que 25-30°C. Há, no país, um grande número de regiões que se enquadram neste perfil em estações específicas do ano. As linhagens TGMS deverão ser testadas para identificar as regiões geográficas e estações em que a expressão da esterilidade é completa e a expressão da fertilidade máxima. Alelos TGMS deverão ser transferidos para linhagens com diferente background genético, com vista a maximizar a chance de se obter boas combinações de híbridos. A introgressão de alelos TGMS em backgrounds diversos pode ser facilitada pela informação obtida dos marcadores moleculares. Os marcadores microssatélites franqueadores de QTLs para TGMS serão úteis como ferramentas de seleção na obtenção de linhagens TGMS para a produção de híbridos de arroz.

6. CONSIDERAÇÕES FINAIS

- Os painéis múltiplos de marcadores microsatélites mostraram-se úteis para a genotipagem em escala de acessos/cultivares/linhagens de arroz conservados no Banco de Germoplasma. Um total de 298 acessos foi utilizado para análises genéticas e estatísticas, com um número mínimo de pelo menos 12 marcadores genotipados. A obtenção dos genótipos foi realizada eficientemente em seqüenciador automático de DNA.
- A genotipagem semi-automática utilizando painéis de microsatélites marcados com fluorescência e genotipados em seqüenciador de DNA proporciona maior quantidade de informação genotípica em cada gel analisado e aumento na acurácia de determinação alélica em cada loco.
- Foi constituído um banco de freqüências alélicas de locos microsatélites de arroz, passível de uso em análise genética da espécie.
- A genotipagem de locos de marcadores moleculares microsatélites permitiu estabelecer com maior clareza os vínculos genéticos entre os diferentes acessos, linhagens ou cultivares e, conseqüentemente, uma melhor organização do germoplasma no Banco de Germoplasma.
- O cruzamento envolvendo as linhagens IR50 e IRTGMS-68945-4-33-14-48, utilizado na obtenção de população segregante para mapeamento de locos TGMS de arroz foi baseado em estimativas de distância genética visando maximizar o polimorfismo detectado na população de mapa.
- A análise de agrupamento dos 298 acessos analisados pelo método Neighbour-Joining, mostrou a formação de dois grupos principais, com vários subgrupos neles contidos.
- O padrão de agrupamento foi utilizado como modelo para analisar a estrutura genética da coleção. Os valores de F_{IS} e F_{IT} revelaram níveis elevados de endogamia nos subgrupos e nos indivíduos. Entretanto, os valores do índice de fixação F_{ST} são relativamente moderados, indicando um baixo nível de diferenciação entre os distintos grupos de acessos.
- Não foi detectada correlação significativa entre os agrupamentos utilizados como modelo, baseados em dados de distância genética, e os subgrupos que representam a subestruturação da coleção, determinados pela análise MCMC e as estatísticas F de Wright.

- A coleção nuclear constituída neste trabalho satisfaz os requisitos básicos para a sua formação (70-80% riqueza alélica e 10-20% do tamanho da amostra inicial) e pode contribuir para o melhoramento da gestão do BGA. A coleção nuclear apresenta níveis moderados de estruturação genética.
- Um mapa genético parcial para regiões específicas do genoma permitiu a identificação de três regiões cromossômicas relacionadas com o controle genético de macho-esterilidade termo-sensível (TGMS).
- Dois desses QTLs foram mapeados em cromossomos ainda não associados ao controle genético da característica TGMS (cromossomos 1 e 12). O terceiro QTL mapeia num cromossomo onde o loco *tms2* foi anteriormente mapeado, mas não foram realizados testes alélicos para clarificar se essas duas regiões são as mesmas.
- Os dados gerados neste trabalho pretendem contribuir para a discussão sobre a caracterização e intensificação do uso de recursos genéticos depositados em bancos de germoplasma, em especial a métodos de amostragem de coleções para definição de acessos apropriados a uso específico em estudos genéticos e programas de melhoramento.

7. BIBLIOGRAFIA

1. ABADIE, T.;CORDEIRO,C.M.T.; FONSECA,J.R.; NEVES ALVES,R.B.; BURLE,M.L.; BRONDANI,C.; RANGEL,P.H.N.; CASTRO,E.M.; TORRES DA SILVA,H.; FREIRE,M.S.; ZIMMERMANN,F.J.P.; E MAGALHÃES, J.R. (2005).Construção de uma coleção nuclear de arroz para o Brasil Pesq. Agropec. Bras., Brasília, v.40, n.2, p.129-136,
2. AGARWAL, R. P. ANDPARKS, R. E. Jr.(1971). Purine Nucleoside Phosphorylase from Human Erythrocytes. V. CONTENT AND BEHAVIOR OF SULFHYDRYL GROUPS J. Biol. Chem. 1971 246: 3763-3768
3. AHMED, F. E. AND A. E. HALL. (1993). Heat injury during early floral bud development in cowpea. Crop Sci. 33: 764-767.
4. AJMONE-MARSAN, P., LIVINI, C.; MESSMER, M.M. MELCHINGER, A.E. AND MOTTO, M.. (1992). Cluster analysis of RFLP data from related maize inbred lines of the BSSS and LSC heterotic groups and comparison with pedigree data. Euphytica 60139–148.
5. AJMONE-MARSAN, P., P. CASTIGLIONI, F. FUSARI, M. KUIPER, AND M. MOTTO. (1998). Genetic diversity and its relationship to hybrid performance in maize as revealed by RFLP and AFLP markers. Theor. Appl. Genet. 96219–227
6. AKAGI, H.; YOKOZEKI, Y.; INAGAKI, A.; FUJIMURA, T. Microsatellite DNA markers for rice chromosomes. **Theoretical Applied Genetics**, v.93, p.1071-1077, 1996.
7. AKKAYA M.S., SHOEMAKER R.C., SPECHT J.E., BHAGWAT A.A., CREGAN P.B. (1995) Integration of simple sequence repeat (SSR) DNA markers into soybean linkage map. Crop Sci.;35:1439-1445
8. ANGADI, S.V., CUTFORTH, H.W., MILLER, P.R., MCCONKEY, B.G., ENTZ, M.H., BRANDT, S.A., VOLKMAR, K.M. (2000). Response of three *Brassica* species to high temperature stress during reproductive growth. *Canadian Journal of Plant Science* 80, 693–701.
9. ALCOCHETE, A.N.A. **Diversidade Genética e Mapeamento de QTLs do Sistema Gênico de Macho-Esterilidade Termo-Sensível (TGMS) do Genoma de Arroz (*Oryza sativa* L.)**. 2005. 167p. Dissertação (Doutorado) - Departamento de Biologia Celular. Universidade de Brasília, Brasília.
10. ANGUS, J. F. AND L.G., L. (1991) *Forecasting Ausdtralian rice yields. In 'Climatic Variation and Change: Implication for the Pacific Rim.'* (Eds Shu Geng and Casey Walsh Cady.) pp. 1-8 (University of California: Davis, CA, USA), .
11. ARNHEIM, N., STRANGE, C. AND ERLICH, H.(1985) Use of pooled DNA samples to detect linkage disequilibrium of polymorphicrestriction fragments and human disease: studies of the HLA class II loci. Proc. Natl. Acad. Sci., 82, 6970-6974
12. ARMOUR, JAL; BARTON, DE, COCKBURN, DJ AND TAYLOR, GR (2002)The detection of Large Deletions or Duplications in Genomics DNA. Human Mutation 20325-337
13. AYRES NM, MCLUNG, AM LARKIN,PD, BLIGH HFJ, JONES CA, AND PARK, WD (1997) Microsatellites and a single nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. Theor Appl Genet 94773-781
14. AYRES, K.L. & BALDING, D.J. (1998). Measuring departures from Hardy-Weinberg: a Markov chain Monte Carlo method for estimating the inbreeding coefficient. *Heredity* 80, 769-777

15. AYRES, K.L. & BALDING, D.J. (2001). Measuring gametic disequilibrium from multi-locus data. *Genetics* 157, 413 – 423.
16. BALDING, D.J. & NICHOLS, R.A. (1997). Significant genetic correlations among Caucasians at forensic DNA loci. *Heredity* 78 583-589
17. BARBIER P, (1989a). Genetic variation and ecotypic differentiation in the wild rice species *Oryza rufipogon*. I. Population differentiation in life-history traits and isozymic loci. *Jpn J Genet* 64:259–271.
18. BARBIER P, (1989b). Genetic variation and ecotypic differentiation in the wild rice species *Oryza rufipogon*. II. Influence of the mating system and life-history traits on the genetic structure of populations. *Jpn J Genet* 64:273–285.
19. BAR-HEN A, CHARCOSSET A, BOURGOIN M, GUIARD J (1995). Relationship between genetic markers and morphological traits in a maize inbred lines collection. *Euphytica* 84: 145-154.
20. BARRETT, B.A., AND K.K. KIDWELL. (1998.) AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. *Crop Sci.* 38:1261–1271.
21. BARRY, G.F. (2001). The Use of the Monsanto Draft Rice Genome Sequence in Research. *Plant Physiology*, March 2001, Vol. 125, pp. 1164–1165,
22. BAVERSTOCK, P.R., AND C. MORITZ. (1996). Project design. p. 17–27. *In* D.M. Hillis et al. (ed.) *Molecular systematics and PHYLIP* (J. Felsenstein, Sunderland, MA. University of Washington, Seattle, USA),
23. BASTEN, C. J.; WEIR, B. S.; ZENG, Z. B. **QTL Cartographer version 1.17. A Reference Manual and Tutorial for QTL Mapping**. North Carolina State University, Raleigh, USA. 2005. 190p.
24. BEAMONT, MA and RANNALA, B (2004). The Bayesian Revolution in Genetics. *Nature Reviews* v5, 251-261
25. BEAUMONT, M.A., IBRAHIM, K.M.; BOURSOT, P. AND BRUFORD, M.W.. (1998). Measuring genetic distance. p. 315–325. *In* A. Karp et al. (ed.) *Molecular tools for screening biodiversity*. Chapman and Hall, London.
26. BEERLI, P. AND FELSENSTEIN, J. (1999) Maximum likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152(2):763-773.
27. BEERLI, P. AND FELSENSTEIN, J. (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *PNAS* 98(8): 4563-4568.
28. BECKER, J & HEUN, M (1995) Barley microsatellites allele variation and mapping. *Plant Molecular Biology* 27:835-45
29. BELKHIR K, BORSA P, CHIKHI L, RAUFASTE N & BONHOMME F(2001). GENETIX 4.02, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).
30. BELL, C.J.; ECKER, J.R (1994). Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics* 19: pp 137-144
31. BELLON, M. R. , BRAR, D. S. , LU, B. R. AND PHAM, J. L. (1996) Rice Genetic Resources. *In* Fisher, KS *Caring for the biodiversity of tropical rice ecosystems*. Volume 2
32. BELÓ, A. **Avaliação de Recursos Genéticos para produção de híbridos de arroz (*Oryza sativa* L.)**. 2001. 104p. Dissertação (Mestrado) - Departamento de Ecologia. Universidade de Florianópolis. Brasil

33. BERNARDO, R. (1993). Estimation of coefficient of coancestry using molecular markers in maize. *Theor. Appl. Genet.* 851055–1062.
34. BLACK, G A (1950) Os capins aquáticos da Amazônia. Belém IAN, 53-94p (IAN.Boletim Técnico, 19)
35. BLAIR M. W., HEDETALE V., McCOUCH S. R. (2002) Fluorescent-labeled microsatellite panels useful for detecting allelic diversity in cultivated rice (*Oryza sativa* L.). *Theor Appl Genet* 105449-457
36. BONG BB. (2000). Genetic improvement of rice varieties for The Mekong Delta of Vietnam: Current status and future approaches. *Rice Research and Development in Vietnam for the 21st Century*. BB Bong (Ed.). Proc. of the Conf. On Rice Res. And Dev. In Vietnam for the 21st century – Aspects of Vietnam-India. Cantho, Sep. 18-19, 2000. CLRRI. Pp.123-134
37. BOTSTEIN D, WHITE RL, SKOLNICKM, DAVIS RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32: 314–331.
38. BOWCOCK, A.M., A. RUIZ-LINARES, J. TOMFOHRDE, E. MINCH, J.R. KIDD, AND L.L. CAVALLI-SFORZA. (1994). High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368: 455-457
39. BREDEMEIJER GMM, ARENS P, WOUTERS D, VISSER D, AND VOSMAN B, 1998. The use of semi-automated fluorescent analysis for tomato cultivar identification. *Theor Appl Genet* 97:584–590
40. BRETTING, P.K., WIDRLECHNER, M.P. (1995). Genetic markers and plant genetic resource management. *Plant Breed Rev* 13: 11–86
41. BRONDANI, C. **Desenvolvimento de Marcadores Microsatélites, Construção de Mapa Genético Interespecífico de *Oryza glumaepatula* x *O. sativa* e análise de QTLs para Caracteres de Importância Econômica.** 2000. 226 p. Dissertação (Doutorado) - Departamento Biologia Celular. Universidade de Brasília, Brasília.
42. BRONDANI, R.P.V. AND D. GRATTAPAGLIA.(2001). Cost-effective method to synthesize a fluorescent internal DNA standard for automated fragment sizing. *BioTechniques* 31:793-800.
43. BRONDANI C, RANGEL N, BRONDANI V, FERREIRA E.(2002) QTL mapping and introgression of yield-related traits from *Oryza glumaepatula* to cultivated rice (*Oryza sativa*) using microsatellite markers.*Theor Appl Genet.*; 104 (6-7)1192-1203. Epub 2002 Apr 06.
44. BROWN, A.H.D., AND B.S. WEIR. (1983). Measuring genetic variability in plant populations. p. 219–229. *In* S.D. Tanksley and T.J. Orton sure(s), clustering methods, determination of optimal (ed.). *Isozymes in plant genetics and breeding. Part A.* Elsevier,
45. BROWN, A. H. D. (1989a). The case for core collections. Pp. 136–156 in *The Use of Plant Genetic Resources*, A. H. D. Brown, O. H. Frankel, D. R. Marshall, and J. T. Williams, eds. New York: Cambridge University Press.
46. BROWN, A. H. D. (1989b). Core collections: A practical approach to genetic resources management. *Genome* 31 (2): 818–824. *In*: BROWN, A. H. D., O. H. FRANKEL, D. R. MARSHALL, AND J. T. WILLIAMS, (eds.). *The Use of Plant Genetic Resources*. New York: Cambridge University Press.
47. BROWN, J.K.M. 1994. Bootstrap hypothesis tests for evolutionary trees and other dendrograms. *Proc. Natl. Acad. Sci. (USA)* 9112293–12297.
48. BROWN-GUEDIRA, G.L., THOMPSON, J.A. NELSON, R.L. AND WARBURTON, M.L. (2000). Evaluation of genetic diversity of soybean introductions using RAPD and SSR markers. *Crop Sci.* 40815–823.
49. BROWN, S.M. & KRESOVICH, S. (1996). Molecular characterization for plant genetic resources conservation, pp. 85-93. *In* Paterson H (ed.), *Genome Mapping of Plants*. Academic Press, San Diego.

50. BRUNELLI, KR; SILVA, HO; CAMARGO, LEA (2002) Mapeamento de Genes de Resistência a Puccinia polysora em Milho. *Fitopatologia Brasileira*, vol27, n2134-140
51. BUCKLER, E.S.T.; THORNSBERRY, J.M. Plant molecular diversity and applications to genomics. **Current Opinion on Plant Biology**, v.5(2), p.107-11, 2002.
52. BUSO, G. S. C., RANGEL, P.H., FERREIRA, M. E. (1998a) Análise genética de espécies silvestres de arroz (*Oryza* spp.) nativas do Brasil estrutura de populações, diversidade genética e relações filogenéticas utilizando marcadores moleculares. Tese de Doutorado, UnB, Brasília, Brasil.
53. BUSO GSC, RANGEL PH, FERREIRA ME (1998b) Analysis of genetic variability of South American wild rice populations (*Oryza glumaepatula*) with isozymes and RAPD markers. *Molecular Ecology*, 7, 107–117.
54. BUSO, G.S.C.; RANGEL, P.H.N. AND FERREIRA, M.E. (2001) Analysis of random and specific sequences of nuclear and cytoplasmic DNA in diploid and tetraploid American wild rice species (*Oryza* spp.) *Genome/Génome* 44 (3): 476-494 (2001)
55. CAI W, MORISHIMA H.(2002)QTL clusters reflect character associations in wild and cultivated rice.Theor Appl Genet. 2002 Jun;104(8)1217-1228. Epub Apr 10. Related Articles, Links National
56. CARLSON, R.E. (1990). Heat stress, plant-available soil moisture, and corn yields in Iowa: a short- and long-term view. *Journal of Production Agriculture* 3, 293–297.
57. CARRANO, A.V., J. LAMERDIN, L.K. ASHWORTH, B. WATKINS, E. BRANSCOMB, T. SLEZAK, M. RAFF, P.J. DE JONG, D. KEITH, L. MCBRIDE, S. MEISTER, AND M. KRONICK. 1989. A high resolution, fluorescence-based, semi-automated method for DNA fingerprinting. *Genomics* 4:129–136.
58. CAUSSE M, FULTON TM, CHO YG, AHN SN, WU K, XIAO J, CHUNWONGSE J, YU Z, RONALD PC, HARRINGTON SE, SECOND GA, MCCOUCH, S.R., TANKSLEY, S.D. (1994). Saturated molecular map of the rice genome based on an interspecific backcross population. *Genetics* 138: 1251–1274
59. CHANG, T.T. 1976. Manual on genetic conservation of rice germplasm for evaluation and utilization. International Rice Research Institute, P.O. Box 933, Manila, Philippines.
60. CHANG, T. T. (1984b). The role and experience of an international crop-specific genetic resources center. Pp. 35–45 in *Conservation of Crop Germplasm—An International Perspective*, W.L. Brown, ed. Madison, Wis.: Crop Science Society of America.
61. CHEN, X.; TEMNYKH, S.; XU, Y.; CHO, Y.G; MCCOUCH, S.R. Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L). **Theoretical Applied Genetics**, v.95, p.553-567, 1997.
62. CHOI, G., YI,H., LEE,J., KWON,Y.K., SOH,M.S., SHIN,B., LUKA,Z., HAHN,T.R., AND SONG,P.S. (1999). Phytochrome signalling is mediated through nucleoside diphosphate kinase 2. *Nature* 401, 610-613.
63. COMSTOCK RE, ROBINSON HF. (1952). Estimation of average dominance of genes. In: *Heterosis*. Ames, Iowa (USA): Iowa State College Press. p 494-516.
64. COBURN J. R., TEMNYKH S. V., PAUL E. M., MCCOUCH S. R. (2002) Design and application of microsatellite marker panels for semiautomated genotyping of rice (*Oryza sativa* L.) *Crop Science* 42:2092-2099
65. CORANDER, J.; WALDMANN, P. AND SILLANPÄÄ, M. J. (2003) Bayesian Analysis of Genetic Differentiation Between Populations. *Genetics*, Vol. 163, 367-374.

66. COX, T.S., J.P. MURPHY, AND D.M. RODGERS. (1986). Changes in genetic diversity in the red winter wheat regions of the United States. *Proc. Natl. Acad. Sci. (USA)* 83:5583–5586.
67. CREGAN, P.B., BHAGWAT, A.A., AKKAYA, M.S., JIANG, R. (1994) Microsatellite fingerprinting and mapping of soybean. *Meth. Mol. Cell. Biol.*;5:49-61.
68. CRISAN, D.(1994). Molecular diagnostic testing for determination of myeloid lineage in acute leukemias. *Ann. Clin. Lab. Sci.* 24:355-363.
69. CROSSA, J., C.M. HERNANDEZ, P. BRETTING, S.A. EBERHART, AND S. TABA. (1993). Statistical genetic considerations for maintaining germplasm collections. *Theor. Appl. Genet.* 86:673–678
70. CRUZ RT, O'TOOLE JC (1984) Dryland rice response to an irrigation gradient at flowering stage. *Agron J* 76: 178–187
71. CRUZ, C.D.; SCHUSTER, I. GQMOL: genética quantitativa e molecular. Laboratório de Bioinformática. Universidade Federal de Viçosa. <<http://www.ufv.br/dbg/home3.html>> (2004).
72. DAVENPORT, C. B. (1908) Degeneration, albinism and inbreeding. *Science* 28, 454–455.
73. DE VICENTE, M. C. and S. D. TANKSLEY, (1991). Genome-wide reduction in recombination of backcross progeny derived from male versus female gametes in an interspecific backcross of tomato. *Theor. Appl. Genet.* 83:173-178.
74. DIXON RA, PAIVA NL (1992). Prospects of accessing DNA banks for the isolation of genes encoding biologically active proteins. In Adams RP, Adams JE (eds) *Conservation of Plant Genes DNA Banking and in vitro Biotechnology*, pp. 99–118. Academic Press, San Diego
75. DIXON, R.A., M.K. BHATTACHARYYA, AND N.L. PAIVA. (1995). Engineering disease resistance in plants: an overview. p. 249-270. In R.P. Singh and U.S. Singh (eds.) *Advanced Methods in Plant Pathology*. CRC Press, Boca Raton
76. DIWAN, N., AND CREGAN, P.B.. (1997). Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. *Theor. Appl. Genet.* 95:723–733.
77. DOI, K.; YOSHIMURA, A.; IWATA, N. RFLP mapping and QTL analysis of heading date and pollen sterility using backcross population between *Oryza sativa* L. and *Oryza glaberrima* Steud. **Breeding Science**, v.48, p.395-399, 1998.
78. DOONER, H.K., ROBBINS, T.P., AND JORGENSEN, R.A. (1991) Genetic and developmental control of anthocyanin biosynthesis. *Annu. Rev. Genet.* 25:173-199
79. DONG, N.V.; SUBUDHI, P.K.; LUONG, P.N.; QUANG, V.D.; QUY, T.D.; ZHEN, G. H.G.; WANG, B.; NGUYEN, H.T. Molecular mapping of a rice gene conditioning thermosensitive genic male sterility using AFLP, RFLP and SSR techniques. **Theoretical Applied Genetics**, v.100, p.727-734, 2000.
80. DONG, N.V.; SUBUDHI, P.K.; LUONG, P.N.; QUANG, V.D.; QUY, T.D.; ZHENG, H.G.; WANG, B.; NGUYEN, H.T. (2001) Molecular mapping of a rice gene conditioning Thermosensitive genic male sterility using AFLP, RFLP and SSR techniques. *Theor Appl Genet*, v103, p607–612,.
81. DOWNS, C. A., and S. A. HECKATHORN. (1998). The mitochondrial small heat-shock protein protects NADH:ubiquinone oxidoreductase of the electron transport chain during heat stress in plants. *Federation of European Biochemical Societies letters* 430: 246–250.

82. EAST, E.M. (1908) Inbreeding in corn. Rep. Conn. Agric. Exp. Stn 1907, 419–428.
83. EAST, EM (1936) Heterosis Genetics , 21: 375-396
84. EDWARDS, A.; CIVITELLO, A.; HAMMOND, H.A.; AND CASKEY, C.T. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am. J. Hum. Genet. 49:746–756.
85. EDWARDS, S. V., AND BEERLI, P. (2000). Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution 54:1839-1854.
86. EKANAYAKE IJ, STEPONKUS PL, DE DATTA SK (1990) Sensitivity of pollen to water deficits at anthesis in upland rice. Crop Sci 30: 310–315
87. ESCOBAR GALVIS, M.L. MARTTILA, S. HÅKANSSON, G. FORSBERG, J. AND KNORPP, C. (2001) Heat stress response in pea involves interaction of mitochondrial nucleoside diphosphate kinase (mtNDPK) with a novel 86 kDa protein, Plant Physiology; 126: 69-78.
88. EWENS, W. J. (1972). The sampling theory of selectively neutral alleles. Theor. Pop. Biol. 3: 87–112.
89. EVERITT, B. (1980). Cluster analysis. 2nd edition, Halstead Press, New York.
90. EXCOFFIER, L., P. SMOUSE, AND J. QUATTRO. 1992. Analysis of molecular application to human mitochondrial DNA restriction data. Genetics 131:479–491
91. FALUSH, D., M. STEPHENS and J. K. PRITCHARD, 2003 Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164:1567-1587.
92. FAO – Food and Agriculture Foundation. (2003) Database Statistics (<http://www.fao.org>)
93. FATIMA, T; JAN, A.; HUSNAIN, T. AND RIAZUDDIN, S. (2002). Initiation and Cryopreservation of Cell Suspension of Rice BAsmati Varieties. Pakistan Journal of Biological Sciences 5(6):679-682
94. FRANKEL, O. H., AND A. H. D. BROWN. (1984). Plant genetic resources today: A critical appraisal. Pp. 249–257 in Crop Genetic Resources: Conservation and Evaluation, J.H. W. Holden and J. T. Williams, eds. London: Allen and Unwin.
95. FELSENSTEIN, J. (2004) PHYLIP-Phylogeny Inference Package, Version 3.6. Department of Genome Sciences and Department of Biology University of Washington Box 57730 Seattle, WA 98195-7730 USA
96. FERGUSON IB, SNELGAR W, LAY YEE M, WATKINS CB, BOWEN JH. (1998). Expression of heat shock protein genes in apple fruit in the field. *Australian Journal of Plant Physiology* 25, 155–163
97. FERREIRA, M. E., P. H. WILLIAMS, and T. C. OSBORN, (1994) RFLP mapping of *Brassica napus* using doubled haploid lines. Theor. Appl. Genet 89:615-621.
98. FERREIRA, M.E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. 3.ed. Brasília: Embrapa -Cenargen, 1998. 220p.
99. FORKMANN, G.; DEDIO, J.; HENKEL, J.; MIN, B.; WASSENEGGER M. (1994) genetics, biosynthesis and molecular biology of flower colour of *dianthus caryophyllus* (carnation). ISHS Acta Horticulturæ 420: Ornamental Plant Improvement: Classical and Molecular
100. FRANCO, A. A., MUNDY, L. M., TRUCKSIS, M., WU, S., KAPER, J. B. & SEARS, C. L. (1997). Cloning and characterization of the *Bacteroides fragilis* metalloprotease toxin gene. Infect Immun 65, 1007±1013.
101. FRANKEL, O.H. 1984. Genetic perspectives of germplasm conservation. p. 161-170. In W.K. Arber et al. (ed.)

Genetic manipulation: impact on man and society. Cambridge Univ. Press. Cambridge, England.

102. FRANKEL, O.H., A.H.D. BROWN, AND J.J. BURDON. (1995). The conservation of plant biodiversity. Cambridge Univ Press, Cambridge, England.
103. FRARY, A.; NESBITT, T.C.; GRANDILLO, S.; KNAAP, E.; CONG, B.; LIU, J.; MELLER, J.; ELBER, R.; ALPERT, K.B.; TANKSLEY, S.D. fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. **Science**, v.289, p85-88, 2000.
104. GAO LZ, GE S, AND HONG DY, (2000a). Allozymic diversity and genetic structure of common wild rice *Oryza rufipogon* Griff. in China. *Theor Appl Genet* 101494–502.
105. GAO LZ, GE S, AND HONG DY, (2000b). Low levels of genetic diversity within populations and high differentiation among populations of a wild rice, *Oryza granulata* Nees et Arn. ex Watt. from China. *Int J Plant Sci* 161691–697.
106. GAO, L.Z.; GE, S. AND D.-Y. HONG, D.Y (2001) High Levels of Genetic Differentiation of *Oryza officinalis* Wall. ex Watt. From China. *The Journal of Heredity* 92(6) 511-516.
107. GARRIS, A., S. MCCOUCH, S. KRESOVICH, (2003) Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the xa5 locus of rice (*Oryza sativa* L.). *Genetics* 165: 759-769.
108. GARRIS, A. J.; TAI, T.H.; COBURN, J.; KRESOVICH, S. AND MCCOUCH, S. (2005) Genetic structure and diversity in *Oryza sativa* L. *Genetics: Published Articles Ahead of Print*, published on January 16, 2005 as 10.1534/genetics.104.035642
109. GLASZMANN, J.C. (1987) Isozymes and classification of Asian rice varieties. *Theor Appl Genet* 74 21–30 (1987).
110. GOFF S.A., RICKE D., LAN T.H., PRESTING G., WANG R., DUNN M., GLAZEBROOK J., SESSIONS A., OELLER P., VARMA H., HADLEY D., HUTCHISON D., MARTIN C., KATAGIRI F., LANGE B.M., MOUGHAMER T., XIA Y., BUDWORTH P., ZHONG J., MIGUEL T., PASZKOWSKI U., ZHANG S., COLBERT M., SUN W.L., CHEN L., COOPER B., PARK S., WOOD T.C., MAO L., QUAIL P., WING R., DEAN R., YU Y., ZHARKIKH A., SHEN R., SAHASRABUDHE S., THOMAS A., CANNINGS R., GUTIN A., PRUSS D., REID J., TAVTIGIAN S., MITCHELL J., ELDREDGE G., SCHOLL T., MILLER R.M., BHATNAGAR S., ADEY N., RUBANO T., TUSNEEM N., ROBINSON R., FELDHAUS J., MACALMA T., OLIPHANT A., BRIGGS S., (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica) *Science* 296 (5) 92-100
111. GOETZ, M. GODT, D. E., GUIVARCH, A., KAHMANN, U., CHRIQUI, D. AND ROITSCH, T. (2001) Induction of male sterility in plants by metabolic engineering of the carbohydrate supply *PNAS*, vol. 98, no. 11, 6522-6527
112. GOODMAN, M.N. (1972). Distance analysis in biology. *Syst. Zool.* 21 174–186.
113. GOLDSTEIN, D.B., A. RUIZ LINARES, L.L. CAVALLI-SFORZA, AND M.W. FELDMAN. (1995). An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139:463-471.
114. GRAMENE database. Disponível em: <<http://www.gramene.com>> Acesso em: [Abril 2005].
115. GROWER, J.C. (1971). A general coefficient of similarity and some of its generalized properties. *Biometrics* 27857–874.
116. GUAN HX, ZHU YG, LAN SY, XU ZX (2001) Ultrastructural localization of ATPase activity in fertile and sterile anther of rice (*Oryza sativa* L. cv. Marxie) *Shi Yan Sheng Wu Xue Bao*. 2001 Dec;34(4):323-7. Related Articles, Links
117. GUIDERDONI, E. Comparative gametic selection in sexual and anther culture derivatives from five distant F1 hybrids of rice. **Rice Genetic Newsletter**, v.6, p.91, 1989.
118. GUILIONI L, WERY J, TARDIEW F. (1997). Heat stress-induced abortion of buds and flowers in pea: is sensitivity linked to organ age or to relations between reproductive organs? *Annals of Botany* **80**, 159–168.
119. GUO, S-W. AND THOMPSON, E. A. (1992) Performing the exact test of Hardy-Weinberg equilibrium for multiple alleles. *Biometrics*, 48: 361-372.

120. GUSTA, L.V., O'CONNOR, B.J., BHATTY, R.S. (1997). Flax (*Linum usitatissimum* L.) responses to chilling and heat stress on flowering and seed yield. *Canadian Journal of Plant Science* 77, 97–99.
121. HAHLBROCK K, SCHEEL D (1989) Physiology and molecular biology of phenylpropanoid metabolism. *Ann.Rev. Plant Physiol.* 40:347-369.
122. HAIR, J. F. JR., ANDERSON, R. E. TATHAM, R. L. AND BLACK, W. C.(1995). *Multivariate Data Analysis: with Readings*. Prentice Hall, Englewood Cliffs, New Jersey, 4th edition, 1995.
123. HALL JM, LEDUC CA, WATSON AR, ROTER AH (1996) An approach to high-throughput genotyping. *Genome Res* 6:781–790
124. HALLAUER, A.R., AND J.B.MIRANDA. (1988). *Quantitative genetics in maize breeding*. 2nd edition, Iowa State University Press, Ames, IA.
125. HALLEY, C.S. and KNOTT, S.A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using least squares. *Genetics*, 136:1195-1207
126. HAMRICK, J.L., AND M.J.W.GODT. (1997). Allozyme diversity in cultivated crops. *Crop Sci.* 37:26–30
127. HAN, WJ, GRIFO FT (1996). Molecular markers in plant conservation genetics. In: Sobral BWS (ed) *The Impact of Plant Molecular Genetics*, pp. 113–136. Birkhauser, Boston
128. HARBONE, J.B. (1988). *The Flavonoids Advances in Research Since 1980*. Chapman & Hall, London
129. HARTL, D.L. 1987: *A primer of population genetics*. 2nd edition. Sinauer Associates, Massachusetts. 305 p.
130. HARRINGTON, S.(2000). "A survey of genetic diversity of eight AA genome species of *Oryza* using microsatellite markers." *MS thesis* Cornell University, Ithaca, NY.
131. HARUSHIMA ,Y.; KURATA, N.; YANO, M.; NAGAMURA, Y; SASAKI, T.; MINOBE, Y.; NAKAGAHRA, Y. Detection of segregation distortions in an indica-japonica rice cross using a high-resolution molecular map. **Theoretical Applied Genetics** , v.92, p.145-150, 1996.
132. HARUSHIMA Y, YANO M, SHOMURA A, SATO M, SHIMANO T, KUBOKI Y, YAMAMOTO T, LIN SY, ANTONIO BA, PARCO A, KAJIYA H, HUANG N, YAMAMOTO K, NAGAMURA Y, KURATA N, KHUSH GS, SASAKI T(1998) A high-density rice genetic linkage map with 2275 markers using a single F2 population. *Genetics* 148 479–494
133. HE, Y.Q.; YANG, J., XU, C.G.; ZHANG, Z.G.; ZHANG, Q. Genetic bases of instability of male sterility and fertility reversibility in photoperiod-sensitive genic male-sterile rice. **Theoretical Applied Genetiocs**, v.99, p.683-693,1999.
134. HE,Y.Q.; XU,C.G. Importance of epistasis as the genetic basis of sterility instability for photoperiod-sensitive genic male sterile rice. **Acta Botanica Sinica** , v.42, p.1062-1068, 2000.
135. HENEGARIU, O., HIRSCHMANN, P., KILIAN, K. (1994) Rapid screening of the Y chromosome in idiopathic sterile men, diagnostic for deletions in AZF, a genetic factor expressed during spermatogenesis. *Andrologia*, 26, 97–106.
136. HENEGARIU, O., N.A. HEEREMA, S.R. DLOUHY, G.H. VANCE, AND P.H. VOGT. (1997). Multiplex PCR Critical Parameters and step-by-step protocol. *Biotechniques* 23:504–511.
137. Hermann K.M. (1995) The shikimate pathway: early steps in the biosynthesis of aromatic compounds. *Plant Cell* 7: 907-919
138. HERRERO MP, JOHNSON RR. (1980). High temperature stress and pollen viability of maize. *Crop Science* 20, 796–800.

139. HITTALMANI S, HUANG N, COURTOIS B, VENUPRASAD R, SHASHIDHAR HE, ZHUANG JY, ZHENG KL, LIU GF, WANG GC, SIDHU JS, SRIVANTANEYAKUL S, SINGH VP, BAGALI PG, PRASANNA HC, MCLAREN G, KHUSH GS (2003). Identification of QTL for growth- and grain yield-related traits in rice across nine locations of Asia. *Theor Appl Genet.* 107(4) 679-90.
140. HOLSINGER, K. E. (1999). Analysis of genetic diversity in geographically structured populations: a Bayesian perspective. *Hereditas* 130:245-255.
141. HOLTON T A, CHRISTOPHER J, MCCLURE L, HARKER N AND HENRY R J (2002) Identification of polymorphic SSR markers from expressed gene sequences of barley and wheat. *Molecular Breeding* 9, 63-71.
142. HONG S, VIERLING E, HONG SW. (2000). Mutants of *Arabidopsis thaliana* defective in the acquisition of tolerance to high temperature stress. *Proceedings of the National Academy of Sciences, USA* 97, 4392–4397.
143. HONG S, VIERLING E, HONG SW. (2001). HSP101 is necessary for heat tolerance but dispensable for development and germination in the absence of stress. *The Plant Journal* 27, 25–35.
144. HORIE T, YAJIMA M, NAKAGAWA H. (1992). Yield forecasting. *Agricultural Systems* 40: 211–236
145. HORIE T, MATSUI T, NAKAGAWA H, OMASA K. (1996). Effect of elevated CO₂ and global climate change on rice yield in Japan. In: Omasa K, Kai K, Toda H, Uchijima Z, Yoshino M, eds. *Climate change and plants in East Asia*. Tokyo: Springer-Verlag, 39–56.
146. HURKMAN, W.J., M.F. DUPONT, S.B. ALTENBACH, A. COMBS, R. CHAN, C.K. TANAKA, M. REUVENI, AND J.E. BERNARDIN. (1998) BiP, HSP70, NDK and PDI in wheat endosperm. II. Effects of high temperature on protein and mRNA accumulation. *Physiol. Plant.* 103:80–90.
147. IBGE (2003) Anuário estatístico do Brasil.
148. ICARDA (2001) Hybrid rice in vietnam recent progress and issues. Ngoc Ha Street - Ba Dinh - Hanoi – Vietnam (webmaster@agroviet.gov.vn)
149. Imin, N., Weinman, J.J. and Rolfe, B.G. (2004). Investigation of the molecular mechanisms causing Cold induced sterility in rice. A report for the Rural Industries Research and Development Corporation. RIRDC Publication No 04/002. RIRDC Project No ANU-29^a. Rural Industries Research and Development Corporation. Australia
150. IRRI (1992) Drought resistance in crops with emphasis on rice. IRRI, Manila, Filipinas.
151. IRRI (ed.).1997.Hybrid Rice Breeding Manual. IRRI.Philippines
152. ISHII T, MCCOUCH SR (2000) Microsatellites and microsynteny in the chloroplast genomes of *Oryza* and eight other Gramineae species. *Theor Appl Genet* 100:1257–1266
153. ISHII T, XU Y, MCCOUCH SR (2001) Nuclear and chloroplast microsatellite variation in A-genome species of rice.
154. ISHIMARU K, ONO K, KASHIWAGI T. (2003) Identification of a new gene controlling plant height in rice using the candidate-gene strategy. *Planta.* 8 [Epub ahead of print].
155. Ito, N., H. Hayase, T. Satake, and I. Nishiyama, 1970: Male sterility caused by cooling treatment at the meiotic stage in rice plants. III. Male abnormalities at anthesis. *Proc. Crop Sci. Soc. Jpn.* 39, 60–64.
156. JACCARD, P. (1908). Nouvelles recherches sur la distribution florale. *Bull. Soc. Vaud. Sci. Nat.*, 44:223-270.
157. JACKSON MT, HUGGAN R. (1993). Sharing the diversity of rice to feed the world. *Diversity* 9: 22±25.

158. JAUHAR, PREM P. 1996. Genome Analysis: A Prologue. Page1-9 in Prem P. Jauhar, editor. Methods of Genome Analysis in Plants. CRC Press, Inc., Boca Raton, FL.
159. JEON, J.-S., CHUNG, Y.-Y., LEE, S., YI, G.-H., OH, B.-G., AND AN, G. (1999) Isolation and characterization of an anther-specific gene, *RA8*, from rice (*Oryza sativa* L.). *Plant Mol. Biol.* **39**:35-44
160. Jia JH, Li CY, Qu XP, Wang Q, Deng QY, Weng ML, Yuan LP, Wang B (2000) Constructing a genetic linkage map and mapping a new thermo-sensitive genic male-sterile gene (*tms5*) in rice. In the “**Abstract Book of the 1st National Plant Genome Meeting in China**”, 2000, Dalian, China. p.37,
161. Jia, J. H.; Zhang, D. S. ; Li , C. Y. ; Qu , X. P. ; Wang , S. W. ; Chamarek , V. ; Nguyen , H. T. ; Wang, B. (2001) Molecular mapping of the reverse thermo-sensitive genic male-sterile gene (*rtms1*) in rice. **Abstract**, v.103 (4), p.607-612
162. JIANG L, CAO YJ, WANG CM, ZHAI HQ, WAN JM, YOSHIMURA A (2003).[Detection and analysis of QTL for seed dormancy in rice (*Oryza sativa* L.) using RIL and CSSL population]Yi Chuan Xue Bao. ;30(5)453-8.
163. JIN, L. & CHAKRABORTY, R. (1993) *Estimation of Genetic Distance and Coefficient of Gene Diversity from Single-Probe Multilocus DNA Fingerprinting Data*. *Mol. Biol. Evol.*, vol. 11, no. 1, pages 120-127, 13 1993.
164. JOHNS, M.A., P.W. SKORCH, J. NIENHUIS, P. HINRICHSEN, G. BASCUR, AND C. MUNOZ-SCHICK. (1997). Gene pool classification of common bean landraces from Chile based on RAPD and morphological data. *Crop Sci.* 37:605–613.[
165. JOHNSON R. A. AND WHICHERN D. W., (1992). *Applied multivariate statistical analysis*, Prentice Hall Inc
166. JORDAN, H.; DARFLER, M.; SOLUS, J. (1999). A THERMOSTABLE DNA POLYMERASE FOR GENOTYPING OF DINUCLEOTIDE REPEAT POLYMORPHISMS WITHOUT NONTEMPLATED NUCLEOTIDE ADDITION . *Plant & Animal Genome VII Conference*.Town & Country Hotel, San Diego, CA, January 17-21, 1999.
167. KALAIYARASI, R. AND VAIDYANATHAN, P. (2003). Cytological screening of rice TGMS lines. *Plant Breeding* 122, 334—338
168. KARP, A., S. KRESOVICH, K.V. BHAT, W.G. AYAD, AND HODGKIN, T. (1997a).Molecular tools in plant genetic resources conservation A guide to the technologies. IPGRI, Rome
169. KARP A, EDWARDS KJ, BRUFORD M, FUNK S, VOSMAN B, MORGANTE M, SEBERG O, KREMER A, BOURSOT P, ARCTANDER P, TAUTZ D & HEWITT GM (1997b). Molecular technologies for biodiversity evaluation opportunities and challenges. *Nature Biotechnology*, 15625-628.
170. KELLER, B.; FEUILLET, C. Colinearity and gene density in grass genomes. *Trends in Plant Science*, London, v.5, n.6, p.246-251, 2000.
171. KENDALL, W.; LIANG, F. AND WANG, JIAN-SHENG (eds). *Markov Chain Monte Carlo: innovations and applications in statistics, physics, and bioinformatics* (1 - 28 Mar 2004), Singapore.
172. KHUSH, G.S. (1977). Origen, dispersal, cultivation and variation of rice. *Plant Mol. Biol.* 3525-34
173. KHUSH, G. (2001) Chair's introduction. In *Rice Biotechnology: Improving Yield, Stress Tolerance and Grain Quality* (Goode, J.A. and Chadwick, D., eds). Chichester: Wiley, pp. 11–12.

174. KIMPEL, J.A. AND J.L. KEY. (1985). Heat shock in plants. *Trends Biochem. Sci.* **10**: 353-357
175. KIMURA M, CROW JF (1964). The number of alleles that can be maintained in a finite population. *Genetics*, 49:725-738.
176. KITADA, S., HAYASHI, T. and KISHINO, H. (2000). Empirical Bayes procedure for estimating genetic distance between populations and effective population size. *Genetics* 156: 2063–2079
177. Kitashiba, H., Kitazawa, E., Kishitani, S. and Toriyama, K. (1999) Partial male sterility in transgenic tobacco carrying an antisense gene for alternative oxidase under the control of a tapetum-specific promoter. *Molecular Breeding*, **5**(3), 209-218.
178. KITTOCK, D.L., E.L. TURCOTTE, AND W.C. HOFMAN, (1988). Estimation of heat tolerance improvement in recent American pima cotton cultivars, *Journal Agronomy Crop Science*, 161, 305
179. KO, H.L.; COWAN, D.C.; HENRY, R.J.; GRAHAM, G.C.; BLAKENEY, A.B.; LEWIN, L.G. (1994). Random amplified polymorphic DNA analysis of Australian rice (*Oryza sativa* L.) varieties. *Euphytica*, Dordrecht, v.80, n.3, p.179-189, 1994.
180. KOES R. E., F. QUATTROCCHIO, J. N. M. MOL, (1994) The flavonoid biosynthetic pathway in plants: function and evolution *BioEssays* 16:123-132
181. KOH HJ, HEU MH, MCCOUCH SR (1996). Molecular mapping of the *ges* gene controlling the super-giant embryo character in rice (*Oryza sativa* L.). *Theor Appl Genet* 93: 257–261
182. KOH HJ, SON YH, HEU MH, LEE HS, MCCOUCH SR (1999) Molecular mapping of a new genic male-sterility gene causing chalky endosperm in rice (*Oryza sativa* L.). *Euphytica* 106:57–62
183. KOSAMBI, D.D. The estimation of map distances from recombination values. **Annals of Eugenetics**, v.12, p.172-175, 1944.
184. gene causing chalky endosperm in rice (*Oryza sativa* L.). *Euphytica* 106:57–62
185. KRESOVICH, S., SZEWC-MCFADDEN, A.K., BLIEK, S.M. (1995) Abundance and characterization of simple-sequence repeats (SSRs) isolated from a size-fractionated genomic library of *Brassica napus* L. (Rapeseed). *Theor. Appl. Genet.*;91:206-211.
186. KU, SJ; CHO, KH; CHOI,YJ; BAEK,WK; KIM, S; SUH, HS AND CHUNG, YY. (2001) Cytological observation of two environmental genic male-sterile lines of rice. *Mol cells* 31;12(3)403-6
187. KURATA N, NAGAMURA Y, YAMAMOTO K, HARUSHIMA Y, SUE N, WU J, ANTONIO BA, SHOMURA A, SHIMIZU T, LIN SY, INOUE T, FUKUDA A, SHIMANO T, KUBOKI Y, TOYAMA T, MIYAMOTO Y, KIRIHARA T, HAYASAKA K, MIYAO A, MONNA L, ZHONG HS, TAMURA Y, WANG ZX, MOMMA Y, UMEHARA Y, YANO M, SASAKI T, MINOBE Y. (1994a). A 300-kilobase interval genetic map of rice including 883 expressed sequences. *Nature Genet* 8: 365–375
188. KURATA, N. (1994). A 300-kilobase interval genetic map of rice including 880 expressed sequences. *Nature Genetics* **8**: 365–372
189. KURATA, N, NONOMURA, K-I and HARUSHIMA, Y (2002) Rice Genome Organization the Centromere and Genome Interactions. *Annals of Botany* 90 427±435, 2002
190. LABATE, J.A. (2000). Software for population genetic analysis of molecular marker data. *Crop Sci.* 40:1521–1528.

191. LAFITTE-H-R, COURTOIS-B, ARRAUDEAU-M (2002) Genetic improvement of rice in aerobic systems Progress from yield to genes. Source Field-Crops-Research, 75, pp.171-190
192. LANDER, E.S.; BOTSTEIN, D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics**, v.121, p.185-199, 1989.
193. LANG, NT; SUBUDHI, PK; VIRMANI, SS; BRAR, DS; KHUSH, GS; LI, Z; HUANG, N. (1999) Development of PCR-based markers for thermosensitive genetic male sterility gene tms3(t) in rice (*Oryza sativa* L.). *Hereditas* 131(2) 121
194. LANG, N.T.; SUBUDHI, P.K.; VIRMANI, S.S.; BRAR, D.S.; KHUSH, G.S.; LI, Z.; HUANG, N. (1999) Development of PCR-based markers for thermosensitive genetic male sterility gene tms3(t) in rice (*Oryza sativa* L.). **Hereditas**, v.131(2), p.121-127.
195. LANGELLA O., CHIKHI L., BEAUMONT M.A.. LEA (Likelihood-based estimation of admixture): a program to simultaneously estimate admixture and the time since admixture. *Molecular Ecology Notes*, 1(4): 357-358, 2001.
196. LAWSON, K. J. and K. BELKHIR, 2001 A Bayesian approach to the identification of panmictic populations and the assignment of individuals. *Genet. Res.* **78**:59-77.
197. LEVINGS, C. S. III .Thought on cytoplasmic male-sterility in cms-T maize. **Plant Cell**, v.5, p.1285-1290, 1993.
198. LEWIS, N.G., AND YAMAMOTO, E. (1990). Lignin: Occurrence, biogenesis and biodegradation. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **41**, 455–496.
199. LEXER, C., B. HEINZE, H. STEINKELLNER, S. KAMPFER, B. ZIEGENHAGEN, AND GLÖSSL, J. (1999). Microsatellite analysis of maternal half-sib families of of *Quercus robur*, pedunculate oak detection of seed contaminations and inference of the seed parents from the offspring. *Theor. Appl. Genet.* 99:185–191.
200. LI, H.B.; WANG,J.; LIU,A.M.; LIU,K.D.; ZHANG,Q; ZOU,J.S.; ZHANG,Q.F. Genetic basis of low-temperature-sensitive sterility in indica-japonica hybrids of rice as determined by RFLP analysis . **Theoretical Applied Genetics**, v.95, p1092-1097, 1997.
201. LI, S.G.; ZHOU, K.D.; ZHOU, L.H. Genetic analysis and molecular mapping of a dominant thermo-sensitive genic male-sterile gene in rice. **Chinese Science Bulletin**, v.44, p.955-958.1999.
202. LI, X. H. AND WANG, X. K. (1998). Differentiation and classification of parental lines and favorable genic interactions affecting F1 fertility in distant crosses of rice (*Oryza sativa*L.) *Theor Appl Genet* 96 526–538 (Springer-Verlag 1998
203. LI SG, ZHOU KD, ZHU LH (1999a) Genetic analysis and molecular mapping of a dominant thermo-sensitive genic male-sterile gene in rice. *Chinese Sci Bull* 44:955–958

204. LI, RH; XU, CG; LI, XH AND WANG, XK (1999b) Favorable genes and favorable genic interactions enhancing F1 fertility in indica/japonica hybrids. *Yi Chuan Xue Bao* (1999)26(3)228-38
205. LI ZK, LUO LJ, MEI HW, SHU QY, TABIEN R, ZHONG D B, YING CS, STANSEL JW, KHUSH GS, PATERSON AH. (2000). Genetic basis of inbreeding depression and heterosis in rice (*Oryza sativa* L.). In: Paper presented at the International Rice Research Conference, 31 March-3 April 2000. Los Baños (Philippines): International Rice Research Institute.
206. LI C, SUN CQ, MU P, CHEN L, WANG XK.(2001)[QTL analysis of anther length and ratio of stigma exertion, two key traits of classification for cultivated rice (*Oryza sativa* L.) and common wild rice (*O. rufipogon* Griff.)]*Yi Chuan Xue Bao*. Aug; 28(8)746-51.
207. LINCH, M; WALSH, B (1998) **Genetics and Analysis of Quantitative Traits**. Sunderland and Sinauer Associates, 980p
208. LINCHON, M; WALSH, B. *Genetics and Analysis of Quantitative Traits*. Sunderland Sinauer Associates, 1998, 980p
209. LINCOLN, S.; DALY, M.; LANDER, E.S. **Constructing genetic maps with MAPMAKER/EXP 3.0**. 3.ed. Cambridge: Whitehead Institute Technical Report, 1992a. 97p.
210. LINCOLN, S.; DALY, M.; LANDER, E.S. **Mapping genes controlling quantitative traits with MAPMAKER/QL 1.1**. Cambridge : Whitehead Institute Technical Report. 1992b. 73p.
211. LINK, W., C. DICKENS, M. SINGH, M. SCHWALL, AND A.E. MELCHINGER. (1995). Genetic diversity in European and Mediterranean faba bean germplasm revealed by RAPD markers. *Theor. Appl. Genet.* 90:27–32
212. LITT, M.; LUTY, J.A. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. **American Journal of Human Genetics**, v.44, p.397-401, 1989.
213. LOPEZ, M.T.; VIRMANI, S.S. Development of TGMS lines for developing two-line rice hybrids for the tropics. **Euphytica**, v.114, p.211-215, 2000.
214. LOPEZ, M.T.; TOOJINDA, T.; VANAVICHIT, A.; TRAGOONRUNG, S. Microsatellite Markers Flanking the tms2 Gene Facilitated Tropical TGMS Rice Line Development. **Crop Science**, v.43, p.2267-2271, 2003.
215. LORIEUX, M.; PERRIER, X.; GOFFIRET, B.; LANAUD, C.; GONZALEZ DE LEON, D. Maximum-likelihood models for mapping genetic markers showing segregation distortion. 2. F2 populations. **Theoretical Applied Genetics**, v.90, p.81-89, 1995.
216. LYTTLE, T.W. Segregation distorters. **Annual Review of Genetics**, v.25, p.511-557, 1991.

217. LUDWIG MULLER J, KRISHNA P, FORREITER C. (2000). A glucosinolate mutant of *Arabidopsis* is thermosensitive and defective in cytosolic HSP90 expression after heat stress. *Plant Physiology* 123, 949–958.
218. LUND, A. A., BLUM, P. H., BHATTRAMAKKI, D., ELTHON, T. E. (1998). Heat-Stress Response of Maize Mitochondria. *Plant Physiol.* 116: 1097-1110
219. LUO LG, ZHAI HQ, WAN JM. (2001) [Analysis of heading time genotype for a rice male sterile line Zhenshan 97A] *Yi Chuan Xue Bao.* Nov; 28(11) 1019-27
220. MACAULAY, M., L. RAMSAY, W. POWELL, AND R. WAUGH. (2001). A representative, highly informative genotyping set of barley SSRs. *Theor. Appl. Genet.* 102801–809. [ISI]
221. Mackill, D.J., Coffman, W.R. Rutger, J.N. (1982) Pollen shedding and combining ability for high temperature tolerance in rice. *Crop Science* 22: 730-733.
222. MALIK MK, SLOVIN JP, HWANG C, ZIMMERMAN JL, HWANG CH. (1999). Modified expression of a carrot small heat shock protein gene, Hsp 17.7, results in increased or decreased thermotolerance. *The Plant Journal* 20, 89–99.
223. MANDAOKAR, A.; KUMAR,V.D.; AMWAY, M. AND BROWSE,J. (2003) Microarray and differential display identify genes involved in jasmonate-dependent anther development. *Plant Molecular Biology* 52: 775–786, 2003.
224. MANSFIELD D. C., BROWN A. F., GREEN D. K., CAROTHERS A. D., MORRIS S. W., EVANS H-J (1994) Automation of genetic linkage analysis using fluorescent microsatellite markers. *Genomics* 24:225-233
225. MARSHALL, D.R., AND A.H.D. BROWN. (1975). Optimum sampling strategies in genetic conservation. p. 53–80. In O.H. Frankel and J.G. Hawkes (ed.) *Crop Genetic resources for today and tomorrow* Cambridge Univ. Press, Cambridge, England.
226. MARSHALL, TC, SLATE, J, KRUIK, LEB & PEMBERTON, JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* 7(5): 639-655.
227. MARTIN, C., PRESCOTT, A., MACKAY, S., BARTLETT, J., AND VRIJLANDT, E. (1991) Control of anthocyanin biosynthesis in flowers of *Antirrhinum majus*. *Plant J.* 1:37-49
228. MARTIN, G.B.; BROMMONSCHENKEL, S.H.; JUPAPARK, C.; FRARY, A.; GANAL, M.W.; SPIVEY,R.; W.U., T.; EARLE, E.D.; TANKSLEY, S.D. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. **Science**, v.262, p.1432-1436, 1993.
229. MARIN, R E A (1998) agricultura no delta do rio Amazonas colonos produtores de alimentos e Macaá no período colonial. In MARIN, REA. A escrita da história paraense. Belém UFPA, 53-91p.
230. MATSUI T, OMASA K, HORIE T. (1999b). Mechanism of anther dehiscence in rice (*Oryza sativa* L.). *Annals of Botany* 84: 501–506
231. MATSUI T, OMASA K, HORIE T. (2000a). Mechanism of septum opening in anthers of two-rowed barley (*Hordeum vulgare* L.). *Annals of Botany* 86: 47–51.
232. MATSUI T, OMASA K, HORIE T. (2001). The difference in sterility due to high temperatures during the

flowering period among japonica-rice varieties. *Plant Production Science* 4: 90–93

233. MATSUI, T and Omassa, K (2002) Rice (*Oryza sativa* L.) Cultivars tolerant to high temperature at flowering: anther characteristics. *Annals of Botany* 83: 683-687
234. MATSUI, T and KAGATA, H (2003) Characteristics of Floral Organs Related to Reliable Self-pollination in Rice (*Oryza sativa* L.). *Annals of Botany* 91: 473-477, 2003
235. MAUGHAN P.J., SAGHAI MAROOF M.A., BUSS G.R. (1995). Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean. *Genome* 1995; 38 715-723
236. MARUYAMA, K., ARAKI, H.; KATO, H. **Themosensitive genic male sterility induced by irradiation. Rice Genetics II**. International Rice Research Institute, Manila, Philippines. 1991. p.227-233.
237. MCCOUCH SR, CHEN X, PANAUD O, TAMNYKH S, XU Y, CHO YG, HUANG N, ISHII T, BLAIR M (1997) Microsatellite marker development, mapping and applications in rice genetics and breeding. *Plant Molecular Biol* 3589–99
238. MCCOUCH SR, KOCHERT G, YU ZH, WANG ZY, KHUSH GS, COFFMAN WR, TANKSLEY SD (1988) Molecular mapping of rice chromosomes. *Theor Appl Genet* 76815–829
239. MCGREGOR, D.I. (1981). Pattern of flower and pod development in rapeseed. *Can. J. Plant Sci.* 61:275-282
240. MCCLUNG, A.M., CHEN, M., BOCKELMAN, H.E., BRYANT, R.J., YAN, W., FJELLSTROM, R.G., (2004). Characterization Of A Core Collection Of Rice Germplasm And Elite Breeding Lines In The Us With Genetic Markers Associated With Cooking Quality. Meeting Proceedings. P. 127
241. MELCHINGER, A.E. 1993. Use of RFLP markers for analysis of genetic relationships among breeding materials and prediction of hybrid performance. p. 621–628. *In* D.R.Buxton et al. (ed.) *International crop science I*. ASSA, CSSA, and SSSA, Madison WI,
242. MESSMER, M.M., A.E. MELCHINGER, R.G. HERRMANN, AND J. BOPPENTHE (1993). Relationships among early European maize inbreds II. Comparison of pedigree and RFLP data. *Crop Sci.* 33944–950.
243. MEI,M.H.; DAI,X.K.; XU,C.G; ZHANG,Q. Mapping and genetic analysis of the genes for photoperiod-sensitive genic male sterility in rice using the original mutant Nongken 58S. **Crop Science** , v.39, p.1711-1715, 1999.
244. MICHELMORE, R. W., I. PARAN AND R. V. KESSELI, 1991. Identification of markers linked to disease resistance genes by bulk segregant analysis a rapid method to detect markers in specific genomic regions using segregating populations. *Proc. Natl. Acad. Sci. USA* 88 9828-9832.
245. MITCHELL, J. C., & PETOLINO, J. F., (1988) Selection for Increased Anther Culture Response in Maize, *Theor. Appl. Genet.* 76:157-159.
246. MITCHELL S. E., KRESOVICH S., JESTER C. A., HERNANDEZ C. J., SZEWC-MCFADDEN A. K. (1997) Application of multiplex PCR and fluorescence-based, semi-automated allele sizing technology for genotyping plant genetic resources. *Crop Science* 37617-624
247. MO, Y., NAGEL, C., AND TAYLOR, L.P. (1992) Biochemical complementation of chalcone synthase mutants defines a role for flavonols in functional pollen. *Proc. Natl. Acad. Sci. USA* 89:7213-7217
248. MOHAMMADI, S. A. AND PRASANNA, B. M. (2003). Analysis of Genetic Diversity in Crop Plants—Salient

- Statistical Tools and Considerations REVIEW & INTERPRETATION. *Crop Sci.* 43:1235–1248
249. MOHAMMADI, SA, AND BM PRASANNA. 2003. Analysis of genetic diversity in crop plants-salient statistical tools and considerations. *Crop Sci.* 43:1235–1248.
250. MOHNA, F.I., P. SHEN, S.C. JONG, AND K. ORIKONO. (1992). Molecular - evidence supports the separation of *Lentinula edodes* from *Lentinus* and related genera. *Can. J. Bot.* 70:2446–2452.
251. MOISYADI S, DHARMASIRI S, HARRINGTON HM, LUKAS TJ (1994) Characterization of a low molecular mass autophosphorylating protein in cultured sugarcane cells and its identification as a nucleoside diphosphate kinase. *Plant Physiol.* 104: 1401-1409.
252. MONCADA-P, MARTINEZ-C-P, BORRERO-J, CHATEL-M, GAUCH-H, GUIMARAES-E, TOHME-J, MCCOUCH-S-R (2001) Quantitative trait loci for yield and yield components in an *Oryza sativa* X *Oryza rufipogon* BC2F2 population evaluated in an upland environment . *Theoretical and applied genetics* , 102(1), pp.41-42
253. MORETZSOHN, MC; COELHO, PJA; AMARAL, ZPS; HERCOS, AP, TUPINAMBÁ, E.A. (2001). Desenvolvimento e usod e marcadores microssatélites na análise da variabilidade genética de ecótipos de coqueiro (*Cocos nucifera* L.). *Boletim de Pesquisa e Desenvolvimento da Embrapa*, 16, 25p.
254. MORGANTE M., and SALAMINI F. (2003) From plant genomics to breeding practice. *Current Opinion in Biotechnology* 14:1-6
255. MORGANTE, M.; SALAMINI, F. From plant genomics to breeding practice. **Current Opinion in Biotechnology** , v.14(2), p.214-9, 2003.
256. MORISHIMA, H. (1994). Observations at permanent study-sites of wild rice in the suburb of Bangkok. *Tropics* 3:227-233.
257. MORGANTE, M & OLIVIERI, A.M. (1993). PCR-amplified microsatellites as markers in plant genetics. *The Plant Journal*, 3 175-182.
258. MORRISON, M.J., D.W. STEWART AND P.B.E. MCVETTY (1992). Maximum area, expansion rate and duration of summer rape leaves. *Can. J. Plant Sci.* 72:117-126.
259. MORRISON MJ. (1993). Heat stress during reproduction in summer rape. *Canadian Journal of Botany* **71**, 303–308
260. MORRISON, M. J. AND STEWART, D. W.(2002) Heat Stress during Flowering in Summer Brassica *Crop Sci.*, May 1,; 42(3): 797 – 803
261. MULLIS KB AND FALOONA FA. (1987). Specific synthesis of DNA *in Vitro* via a polymerase-catalyzed chain reaction. *Methods in Enzymology* 155:335-350.
262. MUMM, R.H., J. HUBERT, AND J.W. DUDLEY. (1994). A classification of maize inbreds II. Validation of cluster analysis based on RFLPs. *Crop Sci.* 34:852–865.
263. NAGATA,K.; FUKUTA,Y.; SHIMIZU,H.; YAGI,T.; TERAQ,T. Quantitative Trait Loci for Sink Size and Ripening Traits in Rice (*Oryza sativa* L.). **Breeding Science**, v52, p.259-273, 2002.
264. NAMKOONG, G. (1988). Sampling for germplasm collections. Hortand Press, New York.
265. Namuco OS, O'Toole JC (1986) Reproductive stage water stress and sterility. I. Effect of stress during meiosis. *Crop Sci* 26: 317–321
266. NAGAO RT, KIMPEL JA, VIERLING E, KEY JL. (1986). The heat shock response: a comparative analysis. In:

- Mifflin BJ, ed. *Oxford surveys of plant molecular and cell biology*, Vol. 3. London: Oxford University Press, 384–438.
267. NAPOLI, CA; FAHY, D; WANG, H-Y; and P. TAYLOR; LT (1999) white anther: A Petunia Mutant That Abolishes Pollen Flavonol Accumulation, Induces Male Sterility, and Is Complemented by a Chalcone Synthase Transgene1. *Plant Physiol.* (1999) 120: 615-622
268. NASS, L.L. (2001). Utilização de Recursos Geneéticos no Melhoramento. In NAA, L.L.; Valois, A.C.C.; Soares de Melgo, I.; Valadares-Ingliis, M.;C. (eds). Recursos Geneéticos e Melkhoramento:Plantas.Fundação MT< Rondonópolis.29-55p.
269. NATO A, MIRSHAHI A, TICHTINSKY G, MIRSHAHI M, FAURE JEAN P, LAVERGNE D, DE BUYSER J, JEAN C, DUCREUX G, HENRY Y (1997) Immunological detection of potential signal-transduction proteins expressed during wheat somatic tissue culture. *Plant Physiol* **113**: 801-807
270. NEI, M., 1973 Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* **70**:3321-3323
271. NEI, M. & LI, W.H. Mathematical model for studying genetic variations in terms of restriction endonucleases. *Proceedings of the National Academy of Science, Washington*, **76**:5269-5273, 1979.
272. NEI, M. (1987). Molecular evolutionary genetics. Columbia University and heterosis data. *Maydica* 3753–60.
- 273.
274. NEI, M., CHESSER. R.K. (1983). Estimation of fixation indices and gene diversities. *Ann. hum. Genet.* 47 253-259.
275. NEI, M. AND R. K. CHESSER, (1983) Estimation of fixation indices and gene diversities. *Ann. Hum. Genet.* **47**:253-259
276. NI J, COLOWIT PM, MACKILL DJ. (2002). Evaluation of genetic diversity in rice subspecies using microsatellite markers. *Crop Sci.* 42:601-607
277. NIELSEN, R. AND J. W. WAKELEY. (2001). Distinguishing Migration from Isolation: an MCMC Approach. *Genetics* 158: 885-896.
278. Nishiyama, I. (1984) *Climatic influence on pollen formation and fertilization*. Japan scientific societies press Tokyo, Tokyo
279. NOVIKOVA GV, MOSHKOV IE, SMITH AR, KULAIEVA ON, HALL MA (1999) The effect of ethylene and cytokinin on guanosine 5'-triphosphate binding and protein phosphorylation in leaves of *Arabidopsis thaliana*. *Planta* 208: 239-246.
280. NUTTAL WF, MOULIN AP, TOWNLEY SMITH LJ. (1992). Yield response of canola to nitrogen, phosphorus, precipitation, and temperature. *Agronomy Journal* 84, 765–768.
281. OECD (1999) Consensus Document on the biology of *Oryza sativa* (Rice). Environmental Health and Safety Publications. Series on Harmonization of Regulatory Oversight in Biotechnology. Nº14. Environment Directorate. Organization for Economic Co-operation and Development. Paris
282. OGIHARA Y, FUTAMI K, TSUJI K, MURAI K. (1997) Alloplasmic wheats with *Aegilops crassa* cytoplasm which express photoperiod-sensitive homeotic transformations of anthers, show alterations in mitochondrial DNA structure and transcription. *Mol Gen Genet.* (1997);255(1):45-53. Related Articles, Links
283. OLIVO, C; BARRAGÁN, MC; CASTELLAANOS, C.; RODRÍGUEZ, MC; SILIÓ & TORO, MA (2000) Application of molecular markers (RAPD, AFLP and Microsatellites) to Iberian pig genotype characterization. Departamento de

- Mejora Genética y biotecnología, INIA, Ctra. Coruna. Spain
284. OLUFOWOTE JO, XU Y, CHEN X, PARK WD, BEACHELL HM, DILDAY RH, GOTO M & MCCOUCH SR (1997). Comparative evaluation of within-cultivar variation of rice (*Oryza sativa* L.) using microsatellite and RFLP markers. *Genome*, 40 370-378.
285. O'Toole JC, Namuco OS (1983) Role of panicle exertion in water stress induced sterility. *Crop Sci* 23: 1093–1097
- 286.
287. PANAUD O, CHEN X, MCCOUCH SR (1995) Frequency of microsatellite sequences in rice (*Oryza sativa* L.). *Genome* 38 1170–1176
288. PANAUD, O.; CHEN, X.; MCCOUCH, S.R. Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). **Molecular Gene Genetics**, v 252, p.597-607, 1996.
289. PANCHEN, A. L. (1992). *Classification, Evolution, and the Nature of Biology*. Cambridge University Press, New York, NY.
290. PARK SM, BONG HC. (2002). Class I small heat-shock protein gives thermotolerance in tobacco. *Journal of Plant Physiology* 159, 25–30.
291. PARSONS BJ, NEWBURY HJ, JACKSON MT, FORD-LLOYD BV (1999). The genetic structure and conservation of aus, aman and boro rices from Bangladesh. *Genetic Resources and Crop Evol.* 46: 587-598.
292. PATTERSON, A.H., E.S. LANDAU, J.D. HEW, S.E. LINCOLN, AND S.D. TANKSLEY. (1988). Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment polymorphisms. *Nature* 325:721–726.
293. PATERSON, A.H.; DAMON, S.; HEWITT, J.D.; ZAMIR, D.; RABIONOWITCH, H.D.; LINCOLN, S.E.; LANDER, E.S.; TANKSLEY, S.D. Mendelian factors underlying quantitative traits in tomato: comparison across species, generations and environments. **Genetics**, v127, p.181-197, 1991.
294. PEET MM, SATO S, GARDNER RG. (1998). Comparing heat stress effects on male-fertile and male-sterile tomatoes. *Plant, Cell and Environment* 21, 225–231.
295. PESSOA FILHO, M A C P (2004) Análise Genética de Variedades Tradicionais do Banco de Germoplasma de Arroz (*Oryza sativa* L.) com Marcadores Moleculares. Tese de Mestrado. Departamento de Biologia Molecular. Universidade de Brasília
296. PEREIRA, J.A. (2002) Cultura de Arroz no Brasil. Subsídios para a sua história. Embrapa
297. PHILLIPS, R.L.; VASIL, I.K. (Eds.;2001) DNA-Based Markers in Plants Series: Advances in Cellular and Molecular Biology of Plants, Vol. 6 2nd ed., , 532 p., ISBN: 0-7923-6865-7
298. PINTO, LR, VIEIRA, MLC, PEREIRA DE SOUSA, A E SOUSA JUNIOR, CL (2001) Isoenzimas e microsatélites em plantas aspectos técnicos e interpretação genética. In Biotecnologia, Ciência e Desenvolvimento. Nº20.
299. POLLAK PE, VOGT T, MO Y, TAYLOR LP (1993) Chalcone synthase and flavonol accumulation in stigmas and anthers of petunia hybrida. *Plant Physiol.* 102: 925-932.
300. PONCE, M.R.; ROBLES, P.; AND MICOL, J.L. (1999). High-throughput genetic mapping in *Arabidopsis thaliana*. *Mol. Gen. Genet.* 261408–415.

301. POTT, VJ E POTT, A (2000) Plantas aquáticas do Pantanal. Brasília Embrapa. 404p
302. POWELL W, MACHRAY GC, PROVAN J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1 215–222
303. PREACHER, K. J. Calculation for the chi-square test: An web-based interactive calculation tool for chi-square tests of goodness of fit and independence [Computer software].2001. <http://www.unc.edu/~preacher/chisq/chisq.htm>
304. PRITCHARD, JK AND ROSENBERG, NA (1999a). Use of unlinked genetic markers to detect population stratification in association studies. Am. J. of Hum. Gen. 65: 220-228.
305. PRITCHARD, J.K., M. STEPHENS, AND P. DONNELLY. (2000a). Inference of population structure using multilocus genotype data. Genetics 155:945–959.
306. PRITCHARD, J. K., STEPHENS, M., ROSENBERG, N. A., DONNELLY, P. (2000b) Association mapping in structured populations. American Journal of Human Genetics 67:170-181
307. Pritchard, J.K. (2001) Deconstructing maize population structure. Nature Genetics 28: 203-204
308. PRITCHARD, J.K. AND WEN, W. (2004). Documentation for structure software: version 2. Department of Human Genetics. University of Chicago 920 E 58th CLSC 507 Chicago IL 60637, USA (<http://pritch.bsd.uchicago.edu>)
309. PRITCHARD, J.K. (2001). Deconstructing maize population structure. Nature Genetics 28: 203-204.
310. QUATTROCCHIO, F., J.F. WING, H.T.C. LEPPEN, J. N.M. MOL, AND R.E. KOES. (1993). Regulatory genes controlling anthocyanin pigmentation are functionally conserved among plant species and have distinct sets of target genes. Plant Cell. 5:1497-1512.
311. QUEITSCH C, HONG S, VIERLING E, LINDQUIST S, HONG SW. (2000). Heat shock protein 101 plays a crucial role in thermotolerance in *Arabidopsis*. *The Plant Cell* 12, 479–492.
312. RAFALSKI, J.A., VOGEL, J.M., MORGANTE, M., POWELL, W., ANDRE, C.; TINGEY, S.V. Generating and using DNA markers in plants. In: BIRREN, B.; LAI, E., (Eds.). **Nonmammalian Genome Analysis. A Practical Guide**. Academic Press, San Diego, 1996, p.75-134.
313. RAFALSKI, A. Applications of single nucleotide polymorphisms in crop genetics. **Current Opinion on Plant Biology**, v5(2), p.94-100, 2002.
314. RANGEL, PHN (1998) Origem e evolução do arroz. Goiânia EMBRAPA-CNPAP. Palestra apresentada no 1º Curso Internacional de Melhoramento Genético de Arroz; Goiânia.
315. REDDY, O. U. K.; SIDDIQ, E. A. ; SARMA, N. P; ALI, J. ; HUSSAIN, A. J; NIMMAKAYALA, P. ; RAMASAMY, P. ; PAMMI, S.; REDDY, A. S. Genetic analysis of temperature-sensitive male sterility in rice. **Abstract**, v.100, n.5, p.794-801, 2000.
316. RICHARDS, R. A., AND THURLING, N. (1978). Variation between and within species of rapeseed (*Brassica campestris* and *Brassica napus*) in response to drought stress. II. Growth and development under natural drought stresses. Australian Journal of Agricultural Research 29, 479-490.
317. RGISC, (2005)The map-based sequence of the rice Genome. NATURE|Vol 436|11 August 2005
318. RITSCHEL, P.S.; LINS, T.; TRISTAN, R.L.; BUSO, G.C.S.; BUSO, J.A.; FERREIRA, M.E. Caracterização de marcadores de regiões hipervariáveis do genoma e comparação entre sistemas de marcadores baseados em PCR (RAPD e SSR) na análise genética de acessos de melão (*cucumis melo* L.) In: SIMPOSIO DE RECURSOS GENETICOS PARA A AMERICA LATINA E CARIBE-SIRGEALC, 3., 2001, Londrina. Anais... Londrina: IAPAR, 2001.p.392-394 Tipo: PL (581.15) (S612a)

319. ROBINSON, J. P. AND HARRIS, S. A. (1999) Amplified Fragment Length Polymorphisms and Microsatellites: A phylogenetic perspective Chapter 12 in Gillet, E.M. (ed.). . : Which DNA Marker for Which Purpose? Final Compendium of the Research Project Development, optimisation and validation of molecular tools for assessment of biodiversity in forest trees in the European Union DGXII Biotechnology FW IV Research Programme Molecular Tools for Biodiversity.
320. ROGERS, J. S. (1972). Measures of genetic similarity and genetic distance. *Studies in Genet.* VII. Univ. Texas Publ., 7213:145-153.
321. ROHLF, F.J. (1998). NTSYS-pc (Numerical Taxonomy and Multivariate performance. p. 621–628. *In* D.R. Buxton (ed.) *Proceedings of the Analysis System*). Version 1.70. Exeter, Setauket, NY. International Crop Science Congress, 1st, Ames, IA. July
322. RHODES, M., STRAW, R. ; FERNANDO, S. ; EVANS, A. ; LACEY, T.; DEARLOVE, A.; GREYSTRONG, J., WALKER, J.; WATSON, P.; WESTON, P. ; KELLY, M.; TAYLOR, D.; GIBSON, K.; MUNDY, C.; BOURGADE, F.; POIRIER, C.; SIMON, D. ; BRUNIALTI, A.L.B.; MONTAGUTELLI, X.; GUNET, J.; HAYNES, A. ;AND BROWN, S.D.M. (1998). A high-resolution microsatellite map of the genome. *Genome Res.* 8:531–542.
323. RONGWEN J, AKKAYA MS, BHAGWAT AA, LAVI U & CREGAN PB (1995). The use of microsatellite DNA markers for soybean genotype identification. *Theoretical and Applied Genetics*, 90 43-48.
324. RONQUIST, F. AND HUELSENBECK, J. P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *BIOINFORMATICS APPLICATIONS NOTE* Vol. 19 no. 12, pages 1572–1574
325. ROSENBERG, N.A.;_PRITCHARD, J.K.; WEBER, J.L.; CANN, H.M.; KIDD, K.K.; ZHIVOTOVSKY, L.A. AND FELDMAN, M.W. (2002).The genetic structure of human populations. *Science*, 298: 2381-2385.
326. ROUHI, A.M. 1997. Plants to the rescue. *Chem. Eng. News*. January 13. pp. 21-23.
327. RUSSEL, J; FULLER, J; YOUNG, G; THOMAS, B; TARAMINO, G; MACAULAY, M; WAUGH, R 7 POWELL, W (1997) Discriminating between barley genotypes using microsatellites markers. *Genome* 40:442-450
328. Saini, HS, Aspinall D (1982) Abnormal sporogenesis in wheat (*Triticum aestivum* L.) induced by short periods of high temperature. *Ann Bot* 49: 835–846
329. SAINI, HS, SEDGLEY M, ASPINALL D. (1983). effect of heat stress during floral development on pollen tube growth and ovary anatomy in wheat (*triticum aestivum* l.). *australian journal of plant physiology* 10, 137–144.
330. Sato K (1954) Relationship between rice crop and water. *Jpn J Breed* 4: 264–289
331. Satake T and Yoshida, S. (1978) High temperature induced sterility in Indica rice at flowering. *Jpn. J. Crop. Sci.* 47: 6-17;
332. SATO S, PEET MM, THOMAS JF. (2002). Determining critical pre- and post-anthesis periods and physiological processes in *Lycopersicon esculentum* Mill. exposed to moderately elevated temperatures. *Journal of Experimental Botany* 53, 1187–1195.
333. SATO T, UEDA T, FUKUTA Y, KUMAGAI T, YANO M. (2003) Mapping of quantitative trait loci associated with ultraviolet-B resistance in rice (*Oryza sativa* L.).*Theor Appl Genet.*107(6)1003-8.
334. SAITO, A., M. YANO, N. KISHIMOTO, M. NAKAGAHRA, A. YOSHIMURA, K. SAITO, S. KUHARA, Y. UKAI, M. KAWASE, T. NAGAMINE, S. YOSHIMURA, O. IDETA, R. OHSAWA, Y. HAYANO, N. LWATA AND M. SUGIURA, (1991). linkage map of restriction fragment length polymorphism loci in rice. *japan. j. breed.* 41: 665-670.
335. SAITOU, N. & NEI, M. (1987). The neighbor-joining method a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, 4:406-425.

336. SASAKI, S., M. OBARA, K. KASHIBA, T. SATO, M. YANO, T. EBITANI AND T. YAMAYA (2002). Linkage analysis and characterization for QTL on chromosome 2 that associated with cytosolic glutamine synthetase content and panicle weight in rice (*Oryza sativa* L.). *Plant and Cell Physiology* 43: S71-S71.
337. SCHOPER J.B., R.J. LAMBERT, B.L. VASILAS. (1987a). Pollen viability, pollen shedding, and combining ability for tassel heat tolerance in maize. *Crop Sci.* 27: 27-31.
338. SCHOPER, J.B., R.J. LAMBERT, B.L. VASILAS, M.E. WESTGATE. (1987b). Plant factors controlling seed set in maize. The influence of silk, pollen, and ear-leaf water status and tassel heat treatment at pollination. *Plant Physiol.* 83:121-125.
339. SECOND, G. (1982). Origin of the genic diversity of cultivated rice (*Oryza* spp.) Study of the polymorphism scored at 40 isozyme loci. *Jpn. J. Genet.* 57:25-57
340. SEFC, K., F. REGNER, E. TURETSCHKE, J. GLOESSL, AND H. STEINKELLNER. (1999). Identification of microsatellite sequences in *Vitis riparia* and their applicability for genotyping of different *Vitis* species. *Genome* 42: 376-373.
341. SEMON, M; NIELSEN, R; JONES, M. P. AND MCCOUCH, S. R. (2005) The Population Structure of African Cultivated Rice *Oryza glaberrima* (Steud.): Evidence for Elevated Levels of Linkage Disequilibrium Caused by Admixture with *O. sativa* and Ecological Adaptation. *Genetics*, 1639-1648
342. SENIOR, M.L. AND HEUN, M. (1993). Mapping maize microsatellites and polymerase chain reaction confirmation of the targeted repeats using a CT primer. *Genome* 36: 884-889
343. SENIOR, M.L., CHIN, E.C.L., LEE, M., STUBER, C.W., SMITH, J.S.C. (1996) Simple sequence repeat markers developed from maize sequences found in the Genbank database: map construction. *Crop Sci.* 1996;36:1676-1683.
344. SEPTININGSIH EM, TRIJATMIKO KR, MOELJOPAWIRO S, MCCOUCH SR (2003a) Identification of quantitative trait loci for grain quality in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. *Theor Appl Genet.* Sep 26.
345. SEPTININGSIH EM, PRASETIYONO J, LUBIS E, TAI TH, TJUBARYAT T, MOELJOPAWIRO S, MCCOUCH SR (2003b). Identification of quantitative trait loci for yield and yield components in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative *O. rufipogon*. *Theor Appl Genet.* Sep 26
346. SJOGREN, P., AND P.I. WYONE. (1994). Conservation genetics and detection of rare alleles in finite populations. *Conserv. Biol.* 8:267-270. 148 U.S.
347. SHEORAN, I.S. AND SAINI, H.S. (1996) Drought-induced male sterility in rice: changes in carbohydrate levels and enzyme activities associated with the inhibition of starch accumulation in pollen. *Sex Plant Reprod* (1996) 9:161-169
348. SHIMAZAKI Y, SATAKE T, ITO N, DOI Y, WATANABE K. (1964). Sterile spikelets in rice plants induced by low temperature during the booting stage. *Research Bulletin of the Hokkaido National Agricultural Experiment Station* (Japan) 83: 1-9
349. SHONNARD, G. C. AND P. GEPTS (1994). Genetics of heat tolerance during reproductive development in common bean. *Crop Science* 34(5): 1168-1175. {a} Dep. Agron. Range Sci., Univ. Calif., Davis, CA 95616-8515, USA
350. SHUBER, A. P., V. J. GRONDIN, AND K. W. KLINGER. (1995). A simplified procedure for developing multiplex PCRs. *Genome Res.* 5:488-493
351. SHULL, G. H., 1948 What is "heterosis"? *Genetics* 33:439-446

352. SMITH, J.S.C. (1984). Genetic variability within U.S. hybrid maize Multi-variate analysis of isozyme data. *Crop Sci.* 24:1041–1046.
353. SMITH, J.S.C., S. PASZKIEWICS, O.S. SMITH, AND J. SCHAEFFER. (1987). Electrophoretic, chromatographic and genetic techniques for identifying associations and measuring genetic diversity among corn hybrids.. p. 187–203. *In Proc. 42nd Annu. Corn Sorghum Res. Conf., Chisity.*
354. SMITH, O.S., AND J.S.C. SMITH. (1992). Measurement of genetic diversity among maize hybrids; A comparison of isozymic, RFLP, pedigree, and heterosis data. *Maydica* 37:53–60.
355. SMITH, S., HELENTJARIS, T. (1996). DNA fingerprinting and plant variety protection. In: Paterson AH (ed) *Genome Mapping in Plants*, pp. 95–110. R. G. Landes Company
356. SMITH, A.M., DENYER, K., AND MARTIN, C. (1997) The synthesis of the starch granule. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 48:67-87
357. SNEATH P. H. A. AND SOKAL R. R., (1973). *Numerical taxonomy: the principles and practice of numerical classification*, Freeman & Co.
358. SOBRAL, B.W.S. (1996). Characterization of Biodiversity and Technological Opportunities: High Throughput Screening Technologies. Proceedings of the PADCT/SBIO-funded Biodiversity Workshop, Campinas, Brazil, May 1996
359. SOKAL RR, MICHENER CD (1958) A statistical method for evaluating systematic relationships. *Univ Kansas Sci Bull* 38:1409-1438
360. Satake T, Hayase H (1970) Male sterility caused by cooling treatment at the young microspore stage in rice plants. V. Estimation of pollen development stage and the most stage to coolness. *Proc Crop Sci Soc Jpn* 39: 468–473
361. Stieglitz, H. and Stern, H. (1973) Functional analysis of a -1,3-glucanase gene (*Tag 1*) with anther-specific RNA and protein accumulation using antisense RNA inhibition. *Dev. Biol.*, 34, 169-192.
362. STUBER CW, LINCOLN SE, WOLFF DW, HELENTJARIS T, LANDER ES. (1992). Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers *Genetics* Nov 132(3): 823-839
363. SUBUDHI, P.K.; BØRKAKATI, R.K; VIRMANI, S.S.; HUANG. N. Molecular mapping of a thermo-sensitive genetic male sterility gene in rice using bulked segregant analysis. **Genome**, v.40, p.188-194, 1997.
364. SUBUDHI PK, VIRMANI SS, HUANG N (1998) A TGMS-linked nuclear DNA marker as originated from the mitochondrial genome in rice (*Oryza sativa* L.) *HEREDITY* 80: 285-292 Part 3 MAR 1998
365. SUN, Z.X.; MIN, S.K.; XIONG, Z.M. A temperature-sensitive male sterile line found in rice. **Rice Genetics Newsletter**, v.6, p116-117, 1989.
366. TANKSLEY SD, MCCOUCH S, CAUSSE M, FULTON T (1991) Molecular Mapping of the Rice Genome: Recent Advances. In: G.S.Khush and G.H. Toenniessen (eds.) *Rice Biotechnology*.
367. TANSKLEY, S.D.; YOUNG, N.D.; PATTERSON, A.H.; BONIERBALE, M.W. RFLP mapping in plant breeder new tools for an abdsience. **Biotechnology**, v.7, p.257-264, 1989.
368. TAUTZ D 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res* 17: 6463-6471.
369. TEMNYKH, S.; DECLERCK, G.; LUKASHOVA, A.; LIPOVICH, L.; CARTINHOOR, S.; MCCOUCH, S. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation,

- transposon associations, and genetic marker potential. **Genome Research**, v.11, p.1441-1452, 2002.
370. THORNSBERRY, J. M., GOODMAN, M. M., DOEBLEY, J., KRESOVICH, S., NIELSEN, D. AND BUCKLER, E. S. (2001) *Dwarf8* polymorphisms associate with variation with flowering time. *Nature Genetics*, 28, 286-289.
371. THORNSBERRY, J.M.; GOODMAN, M.M.; DOEBLEY, J.; KRESOVICH, S.; NIELSEN, D.; BUCKLER, E.S. IV: *Dwarf8* polymorphisms associate with variation in flowering time. **Nature Genetics**, v.28, p.286-289, 2004.
372. TANKSLEY SD AND SR MCCOUCH (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277:1063-1066
373. TARAMINO, G., AND S. TINGEY. (1996). Simple sequence repeats for germplasm analysis and mapping in maize. *Genome* 39:277-287
374. TAYLOR, L. P., AND JORGENSEN, R. (1992). Conditional male fertility in chalcone synthase-deficient petunia. *J. Heredity* 83, 11-17.
375. THOMAS M. R. N. S. SCOTT (1993) Microsatellite repeats in grapevine reveal DNA polymorphisms when analysed as sequence-tagged sites (STSs). *Theoretical and Applied Genetics* 86: 985-990
376. THOMAS MR, CAIN P & SCOTT NS (1994). DNA typing of grapevines. A universal methodology and database for describing cultivars and evaluating genetic relatedness. *Plant Molecular Biology*, 25 939-949.
377. THOMPSON, J.A., NELSON, R.L. AND VODKIN, L.O..(1998). Identification of diverse soybean germplasm using RAPD markers. *Crop Sci.* 38:1348–1355.
378. TSUCHIYA, T., TORIYAMA, K., YOSHIKAWA, M., EJIRI, S. AND HINATA, K. (1995) Tapetum-specific expression of the gene for an endo-beta-1,3-glucanase causes male sterility in transgenic tobacco. *Plant Cell Physiol*, 36,487-94.
379. VAN HINTUM, TH.J.L., AND D. HAALMAN. (1994). Pedigree analysis for composing a core collection of modern cultivars, with examples from barley (*Hordeum vulgare*). *Theor. Appl. Genet.* 88:70–74.
380. VAUGHAN DA (1994). *The Wild Relatives of Rice: A Genetic Resources Handbook*. InternationalRiceResearch Institute, P.O. Box 933, Manila, Philippines
381. VIRMANI, S.S.; KHUSH, G.S.; PINGALI, P.L. Hybrid Seed Production, Development and Security of Major Cereal Crops. In: PAROLA, R. S.; RAI, M) (Ed.). *Hybrid Research and Development Needs in Major Cereals in the Asia-Pacific Region*, 1994. p.61-86
382. VIRMANI, S.S.;SUN, Z.X.;MOU, T.M.;JAUHAR ALI, A.;MAO, C.X. **Two-Line Hybrid Rice Breeding Manual**. International Rice Research Institute. Los Banhos, Phillipines. 2003. 88 p.
383. VOGT T, TAYLOR LP (1995) Flavonol 3-O-glycosyltransferases associated with petunia pollen produce gametophyte-specific flavonol diglycosides. *Plant Physiol.* 108: 903-911.
384. WALL, J.D. AND PRITCHARD, J.K. (2003). Haplotype blocks and linkage disequilibrium in the human genome. *Nature Reviews Genetics*, 4: 587-597.
385. WANG, Z. Y., and S. D. TANKSLEY, 1989 Restriction Fragment Length Polymorphism in *Oryza sativa* L. *Genome* 32: 1113-1118.
386. WANG, B.; XU, W.W.; WANG, J.Z.; WU, W.; ZHENG, H.G.; YANG, Z.Y.; HUANG, N.; RAY, J.D.; NGUYEN, H.T. Tagging and mapping the thermo-sensitive genic male-sterile gene in rice with molecular markers. **Theoretical Applied Genetics**, v.91, p.1111-1114, 1995.
387. WANG, C.M.; YASUI,H.; YOSHIMURA,A.; WAN,J.M.; ZHAI,H.Q. Identification of quantitative trait loci controlling F2 sterility and heading date in rice. **Acta Genetica Sinica**, v.29, p.339-342, 2002.

388. WARBURTON, M., AND J. CROSSA. (2000). Data analysis in the CIMMYT Applied Biotechnology Center for Fingerprinting and Genetic Diversity Studies. CIMMYT, Mexico.
389. WARBURTON, M.L., X. XIANCHUN, J. CROSSA, J. FRANCO, A.E. MELCHINGER, LI- M. FRISCH, M. BOHN, AND D. HOISINGTON. (2002). Genetic characterization of CIMMYT inbred maize lines and open pollinated populations using large scale fingerprinting methods. *Crop Sci.* 42:1832–1840
390. WARD, J.H. Jr. (1963). Hierarchical grouping to optimize an objective function. *J. Am. Statist. Assoc.* 58:236–244.
391. WARD, J. H. (1963). Hierarchical grouping to optimize an objective function. *J. Am. Statist. Assoc.* 58, 236-244.
392. WEBER, J. L.; MAY, P. E. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. **American Journal of Human Genetics**, v.44, p.388-396, 1989.
393. WEIR, B. S. and C. C. COCKERHAM, 1984 Estimating *F*-statistics for the analysis of population structure. *Evolution* 38:1358-1370.
394. WELTER, L. J. (2003) Distorção de segregação mendeliana e mapeamento de loco de resistência ao TSWV utilizando marcadores microssatélites em um cruzamento interespecífico de pimenta (*Capsicum annum* L. x *C. chinense* L.). Tese de Mestrado. Universidade Federal de Santa Catarina. Florianópolis, 2003. 83p0
395. WILSON, J. AND BALDING, D. J. (1998). Genealogical inference from microsatellite data. *Genetics*, 150:499-510.
396. WOOSTER, R; CLETON-JANSEN, A.M. COLLINS, N., MANGICON, J., CORNELIS, R.S., COOPER DEVILEE, P. & STRAFTON, M.R. (1994) Instability of short tandem repeats (microsatellites) in human cancer. *Nature Genetics*, 6:152-156
397. WRIGHT, S (1978) ***Evolution and the Genetics of Populations, Vol. IV. Variability Within and Among Natural Populations.*** University of Chicago Press, Chicago
398. WU, K.S.; TANKSLEY, S.D. Abundance, polymorphism and genetic mapping of microsatellites in rice. **Molecular Gene Genetics**, v.241, p.225-235, 1993.
399. YAMAGUCHI, Y.; IKEDA, R.; HIRASAWA, H.; MINAMI, M.; UJIHARA, P. Linkage analysis of thermosensitive genic male sterility gene, *tms-2* in rice (*Oryza sativa* L.). **Breeding Science**, v.47, p.371-373, 1997.
400. YAMAMOTO, T., Y. KUBOKI, S. Y. LIN, T. SASAKI, and M. YANO, 1998 Fine mapping of quantitative trait loci, *Hd-1*, *Hd-2* and *Hd-3*, controlling heading date of rice as single Mendelian factors. *Theor. Appl. Genet.* 97:37-44.
401. YUAN, L.P. Strategy conception of hybrid rice breeding. **Hybrid Rice**, v.1, p.1-3, 1987.
402. YANG GP, SAGHAI MAROOF MA, XU CG, ZHANG Q, BIYASHEV RM (1994) Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. *Mol Gen Genet* 245:187–194
403. YAO, F.Y. , XU, C.G. , YU, S.B. , LI, J.X. , GAO, Y.J., LI , X.H. AND ZHANG, Q (1997)_ Mapping and genetic analysis of two fertility restorer loci in the wild-abortive cytoplasmic male sterility system of rice (*Oryza sativa* L.). *Euphytica* 98 183–187, 1997.
404. YOUNG, L. W.; WILEN, R.W.; AND BONHAM-SMITH,P.C. (2004). High temperature stress of *Brassica napus* during flowering reduces micro- and megagametophyte fertility, induces fruit abortion, and disrupts seed production *Journal of Experimental Botany*, Vol. 55, No. 396, pp. 485-495,
405. YU SB, ZI JX, XU CG, TAN YF, GAO YJ. (1997). Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. USA* 94:9226-9231.

406. 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., GENG J., HAN Y., LI L., LI W., HU G., LI J., LIU Z., QI Q., LI T., WANG X., LU H., WU T., ZHU M., NI P., HAN H., DONG W., REN X., FENG X., CUI P., LI X., WANG H., XU X., ZHAI W., XU Z., ZHANG J., HE S., XU J., ZHANG K., ZHENG X., DONG J., ZENG W., TAO L., YE J., TAN J., CHEN X., HE J., LIU D., TIAN W., TIAN C., XIA H., BAO Q., LI G., GAO H., CAO T., ZHAO W., LI P., CHEN W., ZHANG Y., HU J., LIU S., YANG J., ZHANG G., XIONG Y., LI Z., MAO L., ZHOU C., ZHU Z., CHEN R., HAO B., ZHENG W., CHEN S., GUO W., TAO M., ZHU L., YUAN L., YANG H. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*) *Science* 296 (5) 79-9
407. YUAN LP (1994a) Increasing yield potential in rice by exploitation of heterosis. In 'Hybrid rice technology: new developments and future prospects'. (Ed. SS Virmani) pp. 1-6. (International Rice Research Institute: Los Banos, Philippines)
408. YUAN LP, YANG ZY, YANG JB (1994b) Hybrid rice in China. In 'Hybrid rice technology: new developments and future prospects'. (Ed. SS Virmani) pp. 143-147. (International Rice Research Institute: Los Banos, Philippines)
409. YUAN Q, QUACKENBUSH J, SULTANA R, PERTEA M, SALZBERG SL & BUELL CR (2001). Rice bioinformatics. Analysis of rice sequence data and leveraging the data to other plant species. *Plant Physiol.*, 125: 1166-1174.
410. XIAO J, LI J, YUAN L, MCCOUCH SR, TANKSLEY SD: (1996a). Genetic diversity and its relationship to hybrid performance and heterosis in rice as revealed by PCR-based markers. *Theor Appl Genet* 92: 637–643
411. XIAO J, LI J, GRANDILLO S, AHN SN, MCCOUCH SR, TANKSLEY SD, YUAN L (1996b). A wild species contains genes that may significantly increase the yield of rice. *Nature* 384: 223–224
412. XIE JK, ZHUANG JY, FAN YY, TU GQ, XIA YW, ZHENG KL.(2002) Mapping of fertility-restoring genes with main effects and epistatic effects for CMS-DA in rice[Article in Chinese] *Yi Chuan Xue Bao.* Jul;29(7)616-21.
413. XU, Y.; ZHU, L.; XIAO, J.; HUANG, N.; MCCOUCH, S.R. Chromosomal regions associated with segregation distortion of molecular markers in F₂, backcross, doubled haploid, and recombinant inbred populations in rice (*Oryza sativa* L.). **Molecular Gene Genetics**, v.253, p.535-545, 1997
414. ZHANG, Z.G.; YUAN, S.C.; ZEN, H.L.; LI, Y.Z.; LI, Z.C., WEI C.L. **Preliminary observation of fertility changes in the new type temperature sensitive male sterile rice: TVA**. Hybrid Rice, Vol. 1. Hybrid Rice Research Center, Hunan, China, p.31-34,1991
415. ZHANG, Q.F., M.A.S. MAROOF, T.Y. LU, AND SHEN, B.Z. (1992). Genetic diversity and differentiation of *indica* and *japonica* rice detected by RFLP analysis. *Theor. Appl. Genet.* 83495–499.
416. ZHANG, W.P.; SHEN, X.Y.; WU, P.; HU, B. AND LIAO, C.Y. (2001a) QTLs and epistasis for seminal root length under a different water supply in rice (*Oryza sativa* L.). *Theor Appl Genet* (2001) 103:118–123
417. ZHANG, J., ZHENG H. G., AARTI A., PANTUWAN G., NGUYEN T.T., TRIPATHY J.N., SARIAL A.K., ROBIN S., BABU R.C., NGUYEN B. D., SARKARUNG S., BLUM A., NGUYEN H. T. (2001b) Locating genomic regions associated with components of drought resistance in rice comparative mapping within and across species. *Theoretical and Applied Genetics*, 10:319-29.
418. ZENG, Z.-B., (1994) Precision mapping of quantitative trait loci. *Genetics* 136:1457-1468
419. ZENGA,L.; TAEK-RYOUN KWON, T-R.; LIU, X.; WILSON, X.; GRIEVE, C.M.; GREGORIO, GB. (2004) Genetic diversity analyzed by microsatellite markers among rice (*Oryza sativa* L.) genotypes with different adaptations to saline soils. *Plant Science* 166 (2004) 1275–1285
420. ZIEGLE, J.S., Y. SU, K.P. CORCORAN, L. NIE, P.E. MAYRAND, L.B. HOFF, L.J. MCBIDE, M.N. KRONICK, AND S.R. DIEHL. (1992). Application of automated DNA sizing technology for genotyping microsatellite loci. *Genomics* 141026–1031.

421. ZHUANG JY, FAN YY, RAO ZM, WU JL, XIA YW, ZHENG KL. (2002) Analysis on additive effects and additive-by-additive epistatic effects of QTLs for yield traits in a recombinant inbred line population of rice. *Theor Appl Genet.*105(8)1137-1145
422. ZIMMERMANN S, BAUMANN A, JAEKEL K, MARBACH I, ENGELBERG D, FROHNMEYER H (1999) UV-responsive genes of *Arabidopsis* revealed by similarity to the Gcn4- mediated UV response in yeast. *J Biol Chem* **274**: 17017-17024
423. ZÖLLNER, S. ; WEN, X. ; HANCHARD, NA; HERBERT, MA ; OBER, C. AND PRITCHARD, J.K. (2004) Evidence for Extensive Transmission Distortion in the Human Genome.. *Am J. Hum Genet.*, 74:62-72 .