

Universidade de Brasília
Instituto de Ciências Biológicas
Departamento de Biologia Celular

TESE DE MESTRADO

Descoberta e validação de marcadores SNPs por sequenciamento de alta performance do genoma estrutural e por genotipagem por sequenciamento (GBS) de arroz de sequeiro (*Oryza sativa* spp. *japonica*)

Pedro Italo Tanno Silva

Orientador: Dr. Márcio Elias Ferreira

Brasília, 30 de abril de 2012

Dissertação de autoria de Pedro Ítalo Tanno Silva, intitulada “**Descoberta e validação de marcadores SNPs por sequenciamento de alta performance do genoma estrutural e por genotipagem por sequenciamento (GBS) de arroz de sequeiro (*Oryza sativa* spp. *japonica*)**”, apresentada como requisito parcial para obtenção do grau de Mestre em Biologia Celular pela Universidade de Brasília, em 30 de abril de 2012, defendida e aprovada pela banca examinadora abaixo assinada:

Prof. Dr. Márcio Elias Ferreira
Orientador
Embrapa Recursos Genéticos e Biotecnologia

Prof. Dr. Dario Grattapaglia
Embrapa Recursos Genéticos e Biotecnologia

Dr. Paulo Hideo Nakano Rangel
Embrapa Arroz e Feijão

Brasília, 30 de abril de 2012

AGRADECIMENTOS

A realização dessa dissertação somente foi possível graças ao apoio recebido de diversas instituições e pessoas, mas gostaria de agradecer:

Ao Dr. Márcio Elias Ferreira, da Embrapa Recursos Genéticos e Biotecnologia, pela orientação, apoio incondicional, incentivo e paciência em todas as etapas do processo e pelos ensinamentos;

Ao Dr. Paulo Hideo Nakano Rangel, da Embrapa Arroz e Feijão, membro da banca, pelos treinamentos concedidos e pela disponibilidade de ajudar sempre que lhe foi solicitado;

Ao Dr. Dario Grattapaglia, da Embrapa Recursos Genéticos e Biotecnologia, membro da banca, pela orientação, pelos ensinamentos e por todo o suporte dado;

A todos os amigos do Laboratório de Genética Vegetal da Embrapa Recursos Genéticos e Biotecnologia, pelo excelente clima de trabalho, pela ajuda nos momentos de dificuldade, pelos momentos de descontração.

Ao Departamento de Biologia Celular da UnB, pela concessão dos treinamentos e ensinamentos;

À Embrapa Recursos Genéticos e biotecnologia, pela infraestrutura de trabalho e apoio recebido;

A CAPES pelo apoio financeiro;

A todos que participaram de maneira direta ou indireta para a realização dessa dissertação;

E finalmente à minha família, responsável pela minha formação e educação, a quem dedico todas as minhas realizações.

Resumo

Entre as recentes tecnologias que podem ser aplicadas na detecção e desenvolvimento de marcadores moleculares, destaque tem sido dado à tecnologia NGS (*Next Generation Sequencing*), que possibilita a detecção de polimorfismo de DNA através da comparação de milhões segmentos de leitura (*reads*) do genomas estruturais de diferentes cultivares de uma espécie. O polimorfismo de DNA mais abundante, revelado por esta estratégia, é o polimorfismo de base de DNA (substituição ou transição de base), conhecido como SNP (*Single Nucleotide Polymorphism*). Neste trabalho foram utilizados resultados de sequenciamento do genoma estrutural de seis variedades de arroz *japonica* tropical para detectar, selecionar e desenhar painéis multiplex para genotipagem automatizada e simultânea de centenas de marcadores SNP em acessos de arroz de sequeiro. Foram desenvolvidos dois painéis multiplex (SNP1 e SNP2) com 768 SNPs distribuídos pelo genoma de arroz. Os painéis foram avaliados quanto à eficiência de clusterização de intensidade do sinal de fluorescência para a definição dos genótipos, à acurácia de genotipagem em testes de prova e contraprova, à capacidade de discriminação de acessos do Banco de Germoplasma e, finalmente, ao potencial emprego na construção de mapas genéticos de cruzamentos de interesse do programa de melhoramento genético. Os testes realizados comprovaram a eficiência da genotipagem e a elevada concordância de genótipos em testes de prova e contraprova. Observou-se que erros de genotipagem ocorrem a uma taxa muito baixa (0,33%). Cerca de 70% dos SNPs possuem o alelo mais comum com frequência entre 0,50 e 0,75, e são poucos os SNPs (<3%) com frequência do alelo mais comum acima de 0,95. O poder combinado dos painéis na confirmação de identidade genética ou na discriminação de genótipos é extremamente elevado. A segregação e a herança dos alelos dos SNPs que compõem os dois painéis foi testada em duas populações de linhagens puras recombinantes, possibilitando a construção de dois mapas genéticos (cruzamentos Chorinho x Puteca e Chorinho x Amaro). O processo de genotipagem utilizando painéis SNP em genotipador automático é simples, rápido e eficiente. Os painéis SNP1 e SNP2, portanto, possuem ampla aplicação no melhoramento varietal e são recomendados para análise genética de acessos *japonica* tropical de arroz do Brasil. Foi testada ainda uma nova metodologia de genotipagem por sequenciamento (*Genotyping by Sequencing* – GBS) que emprega sequenciamento em escala do genoma estrutural após a redução de complexidade do genoma pelo uso de enzimas de restrição. O emprego de GBS possibilita amostrar o polimorfismo de SNPs em regiões amplamente distribuídas ao longo do genoma utilizando, por exemplo, enzimas sensíveis a metilação, evitando dessa forma o sequenciamento massal de regiões repetitivas e complexas, que dificultam o processo de análise. Os experimentos de GBS foram realizados com acessos de arroz de sequeiro com alta diversidade genética, que compõem a Coleção Nuclear Temática de Tolerância a Seca de arroz da Embrapa. Os resultados obtidos neste trabalho comprovam que a metodologia de genotipagem por sequenciamento tem amplo potencial na detecção e seleção de um alto número de

marcadores SNPs. Contudo, verificou-se nas condições testadas um aparente excesso de genótipos heterozigotos nas amostras avaliadas. Além disso, em testes de prova e contraprova, o erro de genotipagem estimado foi muito elevado em algumas comparações, independentemente da cobertura genômica mínima utilizada na detecção dos SNPs ou do valor de qualidade mínima do SNP detectado. Assim, verificou-se que o ensaio GBS proporciona um alto número de marcadores SNPs capazes de detectar polimorfismo de DNA em acessos de arroz, mas a acurácia da genotipagem foi considerada baixa nas comparações de prova e contraprova realizadas. Portanto, é necessário um aperfeiçoamento do ensaio GBS para a seleção de marcadores SNP fidedignos, passíveis de uso em escala na genotipagem de arroz.

Abstract

Among the recent technologies that can be used to detect and develop molecular markers, Next Generation Sequencing (NGS) has received special attention. This technology allows the detection of DNA polymorphism by the comparison of millions of DNA fragments (reads) obtained from different individuals of a species. The most abundant DNA polymorphism revealed by this strategy is known as SNP (Single Nucleotide Polymorphism). In this work, the genome sequencing results of six different varieties of tropical japonica rice have been analyzed to detect, select and design hundreds of SNP markers to construct multiplex panels for automated simultaneous genotyping of rice germplasm accessions. Two multiplex panels (SNP1 and SNP2), composed of 768 SNPs evenly distributed across the genome, were developed and validated for rice genotyping. The panels were initially evaluated for clustering efficiency of fluorescence signal intensity for genotype calling and for genotype accuracy in test-retest reliability assays. Then, they were examined for the ability to discriminate accessions of the rice germplasm collection and for the potential use in genetic mapping. The results confirmed that the panels have a high genotyping efficiency and accuracy. Data was obtained in more than 96% of the SNPs evaluated and a very low genotyping error rate was observed (0.33%). Most SNPs (70%) presented a frequency between 50-75% for the most common allele, with only a few SNPs having frequency above 95%. The combined power of SNP panels to confirm genetic identity or to discriminate germplasm accessions was very high. Allelic segregation and inheritance of SNP markers was tested in different recombinant inbred line populations, allowing for the construction of two genetic maps, derived from the crossings between Chorinho x Puteca and Chorinho x Amaro. Plant genotyping based on SNP panels in automated genotypers is a simple, fast and efficient method. Both SNP panels have a broad application in breeding and are recommended for use in genetic analyses of tropical japonica rice germplasm. A new genotyping by sequencing (GBS) methodology was also tested in the present study. This methodology is based on high scale genome sequencing, after total DNA is submitted to complexity reduction using enzymatic restriction digestion. GBS is used to detect polymorphism in regions distributed across the whole genome. GBS assays were conducted using tropical japonica rice varieties with high genetic diversity, which compose Embrapa's Drought Tolerance Core Collection. The results indicated that the GBS methodology is useful for the detection and selection of a high number of SNP markers. However, a very high amount of heterozygous genotypes was detected in the analyzed samples. Also, genotype accuracy estimates indicated that the genotyping error rate of the GBS assay was too high, independently of the minimum genomic coverage used for SNP detection or the minimum quality value of the detected SNP. Therefore, the GBS assay can provide a high number of SNP markers capable of detecting DNA polymorphism in tropical japonica rice, but the accuracy was considered low in the test and re-test comparisons performed in the present study. Improvements in the GBS assay are necessary in order to obtain accurate genotypes in large scale that can be useful to japonica rice breeding.

Sumário

1. Introdução.....	1
1.1. O gênero <i>Oryza</i> e o arroz cultivado.....	1
1.2. A importância econômica e social do arroz.....	2
1.3. Sistemas de produção de arroz no Brasil.....	3
1.3.1. Arroz irrigado.....	3
1.3.2. Arroz de sequeiro (arroz de terras altas).....	3
2. O genoma de arroz.....	4
3. Tecnologias de sequenciamento de nova geração (<i>NGS – Next Generation Sequencing</i>).....	6
3.1. Princípios de funcionamento do sistema Roche 454.....	6
3.2. Princípios de funcionamento do sistema SOLiD.....	7
3.3. Princípios de funcionamento do sistema Illumina GAII e HiSeq.....	7
4. Descoberta de marcadores SNP.....	8
4.1. SNPs como marcadores moleculares.....	8
4.2. Descoberta de SNPs.....	10
4.2.1. Descoberta de SNPs a partir de sequências ESTs em bancos de dados	10
4.2.2. Descoberta de SNPs a partir de re-sequenciamento e montagem do genoma estrutural através de tecnologia NGS.....	11
5. Métodos de genotipagem de marcadores SNP.....	12
5.1. Hibridização alelo-específica e tecnologias afins.....	12
5.2. Extensão de primer e tecnologias afins.....	13
5.3. Ligaçãõ de bases nitrogenadas e tecnologias afins.....	14
6. Descoberta e detecção simultânea de SNPs através de Genotipagem por Sequenciamento (<i>GBS – Genotyping by Sequencing</i>).....	16
7. Objetivos.....	19
8. Objetivos específicos.....	19
9. Fluxograma experimental.....	20
CAPÍTULO 1.....	22
Desenvolvimento de painéis multiplex de marcadores SNP para genotipagem em escala de acessos de arroz <i>japonica</i> tropical (<i>Oryza sativa</i> L.).....	22
1. Resumo.....	22

2. Introdução.....	23
3. Objetivos	25
4. Material e Métodos.....	25
4.1. Identificação e seleção de SNPs, a partir de dados de sequenciamento genômico de alta performance, para compor painéis multiplex de genotipagem ..	25
4.2. Montagem dos genomas re-sequenciados e detecção de SNPs.....	26
4.3. Montagem de painéis multiplex de SNPs	28
4.4. Validação de painéis multiplex de SNPs	30
5. Resultados e Discussão	34
5.1. Detecção de marcadores SNPs	34
5.2. Validação de painéis multiplex de SNPs	41
6. Bibliografia	53
CAPÍTULO 2	62
Avaliação da metodologia de Genotipagem por Sequenciamento (<i>Genotyping by Sequencing – GBS</i>) e seu potencial de uso em programa de melhoramento genético de arroz de sequeiro (<i>Oryza sativa</i> L. spp. <i>japonica</i>)	62
1. Resumo	62
2. Introdução.....	63
3. Objetivos	65
4. Material e Métodos.....	65
4.1. Metodologia do ensaio GBS em amostras de arroz <i>japonica</i> tropical	65
4.2. Sequenciamento de alta performance no sequenciador GAII/Illumina utilizando protocolo GBS.....	65
4.3. Detecção, validação e avaliação de performance de marcadores SNPs em ensaio GBS	66
5. Resultados e Discussão	67
6. Bibliografia	85
ANEXOS.....	87
6.1. Anexo 1 – Parâmetros comparativos de avaliação genética entre os marcadores SNP que compõem os dois painéis 384-plex com base na genotipagem de uma amostra de 65 acessos da Coleção Nuclear Temática de Tolerância à Seca de arroz.....	88
Anexo 2 – Parâmetros qualitativos, localização e genótipos observados em 8 acessos de arroz de um total de 768 locos SNP que compõem os dois painéis 384-plex de arroz.....	113

Lista de figuras

Introdução geral

Figura 1: Esquema ilustrativo de hibridização alelo-específica empregada na detecção de alelos em sítios SNP – Sonda Alelo Específica (fonte: www.appliedbiosystems.com).....13

Figura 2. Esquema ilustrativo de hibridização alelo-específica empregada na detecção de alelos em sítios SNP – Primer Alelo Específico (fonte: www.appliedbiosystems.com).....13

Figura 3. Esquema ilustrativo da técnica de ligação de oligonucleotídeos empregada na detecção de alelos em sítios SNP – Método VeraCode (fonte: www.illumina.com).....15

Capítulo 1 - Desenvolvimento de painéis multiplex de marcadores SNP para genotipagem em escala de acessos de arroz *japonica* tropical (*Oryza sativa* L.)

Figura 1. Métodos de detecção de SNPs em comparações de genomas re-sequenciados de seis variedades de arroz: (a) método “pseudo-híbrido”; (b) método “sequência consenso”.....29

Figura 2: Distribuição de marcadores SNPs comuns em combinações de comparações de sequências de variedades que tiveram o genoma re-sequenciado, identificados pelo método “sequência consenso”. As comparações de sequências de variedades incluem três cruzamentos de arroz de sequeiro: (I) IAC 165 x Primavera; (II) Catetão x Azucena; (III) Ligeiro x Moroberekan.....35

Figura 3: Distribuição de marcadores SNPs comuns em combinações de comparações de sequências de variedades que tiveram o genoma re-sequenciado, identificados pelo método “pseudo-híbrido”. As comparações de sequências de variedades incluem três cruzamentos de arroz de sequeiro: (I) IAC 165 x Primavera; (II) Catetão x Azucena; (III) Ligeiro x Moroberekan.....36

Figura 4: Exemplo de sítio SNP comum entre os genitores dos três cruzamentos analisados: IAC 165 (IAC) x Primavera (PRI); Ligeiro (LIG) x Moroberekan (MOR); Catetão (CAT) x Azucena (AZU). O trecho é uma representação de uma região no cromossomo 1, posição 5.187.704. Sequência no topo corresponde a pseudo-molécula de referência, no caso a cultivar Nipponbare.....37

Figura 5: Representação esquemática da distribuição dos SNPs detectados ao longo do genoma de arroz nos seis parentais usados na seleção (IAC 165 x Primavera; Catetão x Azucena; Ligeiro x Moroberekan). Na parte superior da figura observa-se a distribuição de frequências de SNPs detectados no cromossomo 1, enquanto a parte inferior observa-se a distribuição de frequência de SNPs no cromossomo 12 (detecção pelo método “sequência consenso). Cada barra corresponde a um intervalo de 100kb.....38

Figura 6 – Distribuição dos marcadores SNP que compõem os dois painéis 384-plex (painéis SNP1 e SNP2) ao longo dos 12 cromossomos de arroz. Cada círculo representa um cromossomo, com origem e término no mesmo ponto. A posição de cada marcador SNP no cromossomo é representada por traços internos (vermelho). Os traços externos (em preto) indicam a posição (pb) em relação à origem do cromossomo. A região destacada em cinza dentro do anel representa a região centromérica.....39

Figura 7 - Distribuição de 768 marcadores SNP que compõem os painéis SNP1 e SNP2 nos 12 cromossomos de arroz.....40

Figura 8: Clusterização de intensidade de sinal de marcadores SNPs utilizando o software GenomeStudio. Quadros superiores(a): Marcadores SNP que apresentam alta qualidade de sinal e clusterização adequada, possibilitando acurácia na determinação dos alelos no loco analisado. Quadros inferiores(b): Marcadores SNP com baixa qualidade de sinal e clusterização inconsistente, gerando dúvida na determinação dos alelos no loco analisado.....42

Figura 9. Representação das relações de vínculo genético entre 65 acessos da Coleção Nuclear Temática de Tolerância à Seca e oito variedades utilizadas no desenvolvimento de populações de linhagens puras recombinantes de arroz. O dendrograma foi construído com base nas estimativas de similaridade genética par-a-par inferidas pela proporção de alelos em comum (*band coefficient*), submetidas à análise de agrupamento pelo método de UPGMA.....48

Figura 10 – Mapa genético construído com base na população RIL derivada do cruzamento entre as variedades Chorinho e Amaro. Os cromossomos, numerados de 1 a 12, são representados por barras, com a identificação dos marcadores SNP (painéis 384-plex) e microssatélites à direita e a distância em cM à esquerda. Marcadores SNP iniciam com “s” seguida do número do cromossomo ao qual pertencem, e da posição em pb no mapa físico. Foram mapeados 1440 cM do genoma de arroz.....50

Figura 11 – Mapa genético construído com base na população RIL derivada do cruzamento entre as variedades Chorinho e Puteca. Os cromossomos, numerados de 1 a 12, são representados por barras, com a identificação dos marcadores SNP (painéis 384-plex) e microssatélites à direita e a distância em cM à esquerda. Marcadores SNP iniciam com “s” seguida do número do cromossomo ao qual pertencem, e da posição em pb no mapa físico. Foram mapeados 1035,5 cM do genoma de arroz.....51

Capítulo 2 - Avaliação da metodologia de Genotipagem por Sequenciamento (Genotyping by Sequencing – GBS) e seu potencial de uso em programa de melhoramento genético de arroz de sequeiro (*Oryza sativa* L. spp. *japonica*)

Figura 1. Decréscimo do número de SNPs detectados pela metodologia GBS em função da cobertura genômica em sítios SNP para uma amostra de 24 acessos de arroz de sequeiro da Coleção Nuclear Temática de Tolerância à Seca. O número de SNPs detectados em cada cobertura genômica (4x, 7x, 11x e 15x) é similar nas duas repetições. (a) REP1 – distribuição de número de SNPs observada na repetição 1 e (b) REP2 – distribuição de número de SNPs observada na repetição 2, do ensaio GBS.....74

Figura 2. Distribuição de SNPs por classes de valor de qualidade em duas repetições (esquerda, REP1; direita, REP2) em função do aumento da cobertura genômica (4x, 7x, 11x e 15x) em cada sítio de DNA.....79

Lista de Tabelas

Capítulo 1 - Desenvolvimento de painéis multiplex de marcadores SNP para genotipagem em escala de acessos de arroz *japonica* tropical (*Oryza sativa* L.)

Tabela 1. Amostra de 65 acessos da Coleção Nuclear Temática de Tolerância à Seca, utilizados na avaliação dos 768 SNPs que compõem os painéis multiplex SNP1 e SNP2.....32

Tabela 2: Rendimento de SNPs detectados por dois métodos de detecção de polimorfismo de base única em comparações pareadas de sequências genômicas de variedades que tiveram o genoma re-sequenciado.....34

Tabela 3 – Avaliação de acurácia de genotipagem com os dois painéis 384-plex (SNP1 e SNP2) em 24 amostras que tiveram o DNA extraído (prova) e re-extraído (contraprova) para comparação de genótipos.....44

Tabela 4 – Parâmetros comparativos de avaliação genética dos marcadores SNP que compõem os dois painéis 384-plex com base na genotipagem de uma amostra de 65 acessos da Coleção Nuclear Temática de Tolerância à Seca de arroz.....46

Capítulo 2 - Avaliação da metodologia de Genotipagem por Sequenciamento (Genotyping by Sequencing – GBS) e seu potencial de uso em programa de melhoramento genético de arroz de sequeiro (*Oryza sativa* L. spp. *japonica*)

Tabela 1. Número total de sequências (*reads*) obtido para cada um dos 24 acessos de arroz de sequeiro em duas repetições (REP1 e REP2) de ensaio GBS.....68

Tabela 2. Número de SNPs observados em 24 acessos de arroz sequeiro submetidos a ensaio GBS. Os SNPs foram detectados com variações crescentes de cobertura mínima (4x, 7x, 11x e 15x) em cada sítio de DNA. Onde: AA – genótipo homozigoto para o primeiro alelo detectado no loco SNP; BB – genótipo homozigoto para o segundo alelo detectado no loco SNP; AB – genótipo heterozigoto no loco SNP.....69

Tabela 3. Estimativa do percentual de SNPs em comum em duas repetições do ensaio GBS realizadas com 24 acessos de arroz de sequeiro.....75

Tabela 4. Distribuição de valores de qualidade de SNPs detectados pela tecnologia GBS em função do aumento da cobertura genômica (média de duas repetições).....	78
Tabela 5 – Estimativa de homozigose de acessos de arroz de sequeiro com base na amostragem de marcadores SNP em ensaio GBS.....	80
Tabela 6 – Estimativa de homozigose de acessos de arroz de sequeiro com base na genotipagem com dois painéis 384-plex (ensaio GoldenGate, Illumina).	81
Tabela 7 – Estimativa de erro de genotipagem em ensaio GBS utilizando extração e re-extração de DNA (prova e contraprova) em três variedades de arroz de sequeiro. Foram utilizados nesta análise apenas os SNPs com valor de qualidade acima de 1.000.....	84

Descoberta e validação de marcadores SNPs por sequenciamento de alta performance do genoma estrutural e por genotipagem por sequenciamento (GBS) de arroz de sequeiro (*Oryza sativa* spp. *japonica*)

1. Introdução

1.1. O gênero *Oryza* e o arroz cultivado

O gênero *Oryza*, pertencente à família Gramineae (sub-família Oryzoideae), possui 24 espécies que podem ser classificadas em quatro grandes complexos: *O. sativa*, *O. officinalis*, *O. ridleyi* e *O. granulata*. Esta divisão em complexos agrupa as diferentes espécies com base em características comuns que incluem: descritores botânicos e morfológicos, avaliações citogenéticas, ciclo de vida, tipo e tamanho do genoma e facilidade de transferência de genes para o arroz cultivado (Jena, 2010; Vaughan, 2005).

Os genomas das espécies de arroz (ex. AA, BB, CC, EE, FF, AABB e CCDD) são classificados de acordo com a possibilidade de intercruzamento e avaliação de pareamento cromossômico em análise citogenética (Vaughan, 2005). O genoma AA (complexo *O. sativa*), por exemplo, inclui espécies importantes sob o ponto de vista da domesticação do arroz, e inclui as duas espécies cultivadas (*O. sativa* e *O. glaberrima*). Todas as espécies do complexo *O. sativa* são diplóides ($2n=24$) e intercruzáveis através de polinização controlada. Além de *O. sativa* e *O. glaberrima*, este complexo inclui seis espécies silvestres, entre elas *O. glumaepatula*, que é uma espécie nativa do Brasil; *O. rufipogon*, considerada o parente silvestre mais próximo de *O. sativa*, encontrada na Ásia; e *O. barthii*, o parente silvestre mais próximo *O. glaberrima* (Jena, 2010), encontrada na África. O arroz domesticado de origem africana (*O. glaberrima*) é cultivado na região ocidental do continente africano, enquanto o arroz domesticado de origem asiática (*O. sativa* L.) é cultivado em todo o mundo, compondo a base alimentar das populações humanas há milhares de anos.

Os chineses construíram uma civilização tendo o arroz como referência na dieta. Há muito reconhecem dois grupos de variedades, chamadas de *Hsien* e *Keng*, desde a dinastia Han (Século II A.C.). Esses grupos correspondem à classificação moderna das subespécies *japonica* e *indica* (Kato, 1928). A domesticação do arroz asiático ocorreu em duas grandes vertentes. As variedades do grupo *indica* apresentaram uma dispersão partindo da Índia para África oriental e depois ocidental, com uma frente também no sudeste asiático e na direção norte, rumo à China (Khush, 1997). Já variedades do grupo *japonica*, que provavelmente foram domesticadas na China, foram disseminadas para o norte, formando um novo grupo, hoje conhecido como *japonica* temperado. As duas subespécies apresentam, em geral, características morfológicas distintas, grande divergência genética e um isolamento parcial por reprodução sexuada, devido a processos diferenciados de domesticação e seleção (Han & Xue, 2003).

No Brasil, acredita-se que os portugueses tenham introduzido variedades *japonica* tropical e *indica* de terras baixas ainda no século XVI (Pereira, 2002). O ponto de entrada do arroz no Brasil foi provavelmente o Estado do Maranhão. Posteriormente, o arroz foi disseminado em direção ao Sul. Hoje, a produção brasileira de arroz divide-se, majoritariamente, no emprego de variedades do grupo *indica* na Região Sul (Rio Grande do Sul e Santa Catarina), com base no modo de produção de arroz irrigado, e de variedades do grupo *japonica*, especialmente no Centro-Oeste (Mato Grosso, Goiás), Norte (Maranhão) e Sudeste (Minas Gerais), através do modo de produção de sequeiro (Pereira, 2002).

1.2. A importância econômica e social do arroz

O arroz é um dos alimentos básicos mais importantes para a grande parte da população mundial, especialmente nas regiões sul e leste da Ásia, Oriente Médio e América Latina. Segundo dados da FAO (FAO, 2012), o arroz é o segundo cereal mais produzido no planeta, atingindo 672 milhões de toneladas em 2010, atrás apenas da produção de milho e ultrapassando a produção de trigo. Contudo, deve ser ressaltado que praticamente todo o arroz produzido no mundo é usado na alimentação humana, enquanto milho e trigo têm uso na alimentação animal, humana e em outras áreas, como bioenergia. Portanto, o arroz é o alimento mais consumido diretamente pela população humana.

O país com maior produção de arroz é a China (197 milhões de toneladas em 30 milhões de hectares plantados anualmente), seguido da Índia (120 milhões de toneladas em 36 milhões de hectares plantados a cada ano). Em 2009, a média mundial de produtividade foi de 4.328 toneladas por hectare. O Brasil ocupa a nona posição em produção total, com cerca de 11 milhões de toneladas em 2,57 milhões de hectares plantados, fato que insere o Brasil como único país não asiático na lista dos 10 maiores produtores de arroz.

A produtividade média nacional para a safra 2011/2012 foi de 4.805 kg/ha, 13,9% maior do que a alcançada no ano anterior (4.218 kg/ha) (Conab, 2012). Deve ser mencionado que este valor é uma média geral das produtividades alcançadas em diferentes sistemas de produção (irrigado, várzea, sequeiro) em diferentes regiões do país. O valor médio de produtividade para arroz de sequeiro é bem menor (cerca de 2.000 Kg/ha), enquanto que de arroz irrigado está acima de 6.500g/ha (Conab, 2012). No modo de produção irrigado prevalecem as variedades com *background* genético do grupo *indica*, principalmente nos estados do sul do Brasil. A alta produtividade de arroz irrigado no Sul do Brasil se deve ao clima favorável, ao uso de variedades melhoradas, à aplicação de tecnologia, e à antecipação da época de plantio para diminuir os efeitos danosos de frentes frias na época de florescimento. A produtividade média do Rio Grande do Sul, por exemplo, é de 7.600 kg/ha, comparável ou até superior à média dos maiores produtores mundiais. A região Sul é responsável por aproximadamente 75% da produção nacional de arroz, e apenas o Rio Grande

do Sul contribui com aproximadamente 60% da produção nacional (IBGE, 2011). Na região Norte e Central do Brasil, o cultivo da subespécie *japonica*, também conhecido como arroz de sequeiro, ou arroz de terras altas, ocupa a maior parte da área plantada com arroz.

1.3. Sistemas de produção de arroz no Brasil

1.3.1. Arroz irrigado

O arroz é cultivado em aproximadamente 150 milhões de hectares em todo o mundo, sendo a irrigação por alagamento o principal sistema de produção adotado. O arroz irrigado corresponde a mais de 90% da área total cultivada, e contribui com a maior parte da produção mundial desse cereal.

No Brasil, ao contrário dos países asiáticos, o arroz irrigado ocupa hoje cerca de 50% da área plantada com a espécie e, conforme mencionado, este modo de produção responde por aproximadamente 75% da produção nacional. O município de Uruguaiana (RS) é o maior produtor nacional de arroz irrigado, seguido pelo município de Itaqui (RS). Ambos fazem parte do bioma Pampa, com características peculiares de clima, solo e recursos hídricos.

O cultivo de arroz irrigado é realizado preferencialmente em solos com relevo plano a suave-ondulado, típico de baixadas, geralmente mal drenados, em razão da facilidade de manejo da irrigação por inundação para a cultura. O custo de produção de arroz irrigado é superior ao de arroz de sequeiro. Vários fatores contribuem para este encarecimento, incluindo o preço da terra, a sistematização dos tabuleiros, o bombeamento da água, a formação e manutenção dos canais, etc. Preocupações crescentes com a conservação do meio-ambiente e impacto deste tipo de produção na conservação de mananciais estão na pauta de discussão do futuro da produção de arroz irrigado no Brasil. Há estudos que apontam uma redução significativa do plantio no Brasil na próxima década (Brasil, 2011), o que aumenta a pressão pelo incremento de produtividade da cultura nos próximos anos.

1.3.2. Arroz de sequeiro (arroz de terras altas)

Lavouras de arroz de terras altas ou arroz de sequeiro representam mais de 50% da área total destinada a produção de arroz no Brasil. Porém, devido à sua baixa produtividade, o arroz de sequeiro contribui com apenas 25% da produção brasileira (Conab, 2011).

Essa baixa produtividade é atribuída, entre outros fatores, ao pequeno investimento feito neste modo de produção na correção de fertilidade, em insumos de produção e em técnicas modernas de produção. Boa parte da produção de arroz de sequeiro ainda advém de cultivo de subsistência. Há também o impacto do emprego, ainda costumeiro, do arroz de sequeiro na

abertura de novas áreas agrícolas, com conseqüente diminuição de produtividade, visto que essa prática geralmente não é acompanhada do uso de insumos. O arroz de sequeiro é plantado principalmente na região Central e Norte do Brasil, predominantemente em latossolos vermelho-escuro e latossolos vermelho-amarelo, conhecidos pela grande deficiência em nutrientes, pela alta concentração de óxidos de ferro e alumínio e pela baixa capacidade de troca catiônica e de retenção de água. Nestas condições, é importante o investimento em correção (ex. calagem, gessagem) e em insumos para aumento de produtividade.

A ocorrência de curtos períodos de estiagem durante o período chuvoso, conhecidos como “veranicos”, também afeta negativamente a produtividade do arroz de sequeiro. Um déficit de água durante a floração ou enchimento de grãos pode causar uma perda de até 40% na produção. O emprego de variedades adaptadas ao efeito de veranico (ex. variedades com sistema radicular mais profundo e mais desenvolvido) pode minimizar a quebra de produção.

Estes fatores contribuem para uma menor produtividade de arroz de sequeiro em relação a arroz irrigado. Ao mesmo tempo, demonstram que com maior tecnificação da produção e emprego de variedades melhoradas, o arroz de sequeiro pode assumir patamares médios de produção muito acima dos hoje obtidos. A duplicação da produtividade média de arroz de sequeiro teria forte impacto na produção nacional, levando a um superávit na produção que seria providencial para a exportação de grãos. Mas vale lembrar que, por razões econômicas e acordos financeiros entre Brasil e outros países sul-americanos, importamos cerca de 1 milhão de toneladas de arroz por ano para suprir o mercado interno.

O lançamento de cultivares melhoradas com melhor resposta ao uso de insumos tem proporcionado um aumento significativo na produção e qualidade dos grãos de arroz de sequeiro. Hoje, em algumas áreas especializadas no plantio de arroz de sequeiro, é possível atingir produtividades acima de 4.500 kg/ha, com uma qualidade dos grãos similar à observada em sistema irrigado. Isto demonstra o potencial de aumento de produtividade da cultura do arroz de sequeiro no Brasil. Como a área plantada com arroz de sequeiro no país é significativa (50% do total plantado com arroz), o aumento de produtividade de cultivares adaptados ao cultivo de sequeiro representará um novo paradigma para a produção nacional de cereais.

2. O genoma de arroz

Os cereais divergiram de um ancestral comum há aproximadamente 60 milhões de anos. Porém, apesar desse longo período de evolução independente, observa-se uma alta conservação do número e do repertório de genes nos genomas de diferentes espécies de gramíneas (Devos and Gale, 1996). Comparações de mapas físicos e genéticos de gramíneas demonstram essa conservação de conteúdo e orientação de sequências gênicas (Chen, 1997; Gale, 1998). Deve ser observado, contudo, que apesar dessa similaridade e sintonia entre genomas, eles diferem consideravelmente quanto ao tamanho. Por exemplo, as espécies de sorgo, milho, cevada e trigo possuem um tamanho estimado de genoma de aproximadamente

700, 3.000, 5.000 e 16.000 Megabases (Mbp), enquanto arroz possui um genoma muito menor, de ~430 Mbp, dividido em 12 cromossomos ($2n = 24$), que correspondem a aproximadamente 1500 cM observados em um mapa genético baseado em recombinação gênica (McCouch et al., 1987).

O arroz é considerado como espécie modelo para estudos em diversas áreas. O fato de possuir um genoma relativamente pequeno e ter uma importância tão grande como alimento contribui para justificar o grande interesse na pesquisa genética de arroz. O genoma do arroz tem sido usado como referência para estudos de genômica e biologia molecular em gramíneas tendo em vista: (1) a importância econômica e social da cultura; (2) a disponibilidade do banco de dados de sequências do seu genoma estrutural completo (Goff et al., 2002; Yu et al., 2002), (3) a sintenia genômica com outras espécies de cereais, possibilitando estudos de genômica comparativa entre importantes espécies agrícolas, (4) o grande número de acessos cultivados e silvestres conservados em Bancos de Germoplasma, assim como de estoques genéticos adequados a estudos biológicos (painéis de mutantes, populações de mapeamento, linhagens de introgressão e substituição, linhagens quase isogênicas, coleções nucleares, populações de melhoramento, etc.); (5) a disponibilidade de milhares de sequências ESTs (*Expressed Sequence Tags*) e *Full Length cDNA* em bancos de domínio público, (6) o fácil acesso a diversos bancos de dados genômicos de análise e comparação entre cereais (i.e. www.gramene.org), e (7) a existência de uma grande comunidade internacional de pesquisa em arroz (Ferreira, 2003).

As sequências geradas pelos programas de sequenciamento do genoma permitiram prever que o número de sequências gênicas (regiões codificadoras) no genoma de arroz varia de 40.000 a 50.000 genes funcionais, em um total de 15.000 famílias gênicas (Goff et al., 2002; Yu et al., 2002). Este número é duas vezes maior do que o estimado para *Arabidopsis thaliana*. Estima-se também que cerca de 50% do genoma do arroz é composto por sequências repetitivas (Kurata et al., 1994; Goff et al., 2002), que incluem microssatélites (motivos de repetição menores que 10pb), minissatélites (motivos de repetição menores que 40pb), satélites (motivos de repetição de centenas de pares de base) e transposons e retrotransposons de diversas famílias - LINE (Noma, 1999), SINE (Mochizuki, 1992), RIRE (Kumekawa, 2001), TOS (Hirochika, 2001), MITE (Mao, 2000), e diversos outros tipos de sequências degeneradas (Goff et al., 2002).

A disponibilidade de sequências genômicas em bancos de dados e a existência de um genoma referência para a espécie, baseado no sequenciamento e montagem do genoma da cultivar Nipponbare (IRGSP, 2005), tem sido fundamental para a realização de experimentos de genômica e genética molecular de arroz, como os desenvolvidos neste trabalho e descritos a seguir.

3. Tecnologias de sequenciamento de nova geração (NGS – Next Generation Sequencing)

Dois experimentos da década de 1970 sobre sequenciamento de DNA estabeleceram um marco na ciência que ditaria o rumo a ser seguido nos próximos anos nesta área do conhecimento. Tais experimentos possibilitaram o desenvolvimento de métodos de obtenção da sequência de fragmentos de DNA em laboratório (Maxam and Gilbert, 1977; Sanger and Coulson, 1975). Um dos métodos, chamado de terminação da cadeia ou sequenciamento dideoxi, passaria a ser conhecido como método Sanger de sequenciamento. Na década de 1980, com o desenvolvimento dos primeiros protótipos de sequenciadores semi-automáticos de DNA, este método se tornaria o método mais comum de sequenciamento de DNA, utilizado em rotina até os dias de hoje. O concomitante desenvolvimento de algoritmos de montagem de genomas a partir de fragmentos sequenciados ao acaso tornaria o método Sanger a principal ferramenta do projeto de sequenciamento do genoma humano e de outros organismos nos anos seguintes (Lander et al., 2001; Venter et al., 2001).

O conceito químico do método Sanger desde a primeira publicação do método até hoje permanece basicamente o mesmo, com o bônus de ter sido adaptado a sistemas automatizados de sequenciamento. Equipamentos atuais, como o sequenciador de DNA ABI 3730, permitem o sequenciamento de fragmentos de DNA em arranjos de 96 capilares em um tempo de corrida relativamente curto.

Nos últimos anos, porém, o sequenciamento Sanger vem sendo substituído por metodologias de “sequenciamento de nova geração”, ou NGS (*Next Generation Sequencing*). Estas metodologias proporcionam um aumento exponencial na quantidade de bases sequenciadas e, conseqüentemente, alteram significativamente a relação custo/benefício de obtenção de sequências de DNA, oferecendo um preço final por base sequenciada mais atraente e em menor tempo. As plataformas de sequenciamento de nova geração comercialmente disponíveis incluem o sistema Roche 454 (Roche Applied Science), GenomeAnalyser Iix (Illumina, Inc.), HiSeq (Illumina, Inc.) e SOLiD (Applied Biosystems).

3.1. Princípios de funcionamento do sistema Roche 454

A primeira plataforma de sequenciamento de nova geração disponibilizada no mercado foi a Roche 454 (Margulies et al., 2005) utilizando um método de amplificação do DNA conhecido como PCR em emulsão (Tawfik, 1998). Na PCR em emulsão, o DNA total é fragmentado ao acaso e ligado por adaptadores a *beads* (contas esféricas) contendo estreptavidina. Cada conta esférica é capturada em microgotas, sendo que cada microgota age como um reator de amplificação individual, produzindo ~100 cópias clonais de um único molde de DNA por conta esférica (Tawfik & Griffiths, 1998). Cada conta esférica contendo um molde

de DNA é transferida para um poço da placa de sequenciamento (*picotiter plate*) e analisada usando reação de pirosequenciamento. Dessa forma, é possível que centenas de milhares de reações de pirosequenciamento sejam conduzidas em paralelo, aumentando significativamente o número de sequências geradas em curto período de tempo. O pirosequenciamento é uma técnica de sequenciamento por síntese que mede a quantidade que pirofosfato inorgânico (*PPi*) liberado. Com o template de DNA imobilizado, as soluções de dNTPs são adicionadas uma de cada vez. A cada vez que um nucleotídeo é incorporado, há emissão de luz, que é detectada. A sequência de DNA é então determinada por um pirograma, que corresponde à ordem em que os diferentes nucleotídeos foram incorporados na sequência. Uma vez que a quimiluminescência é proporcional à quantidade de fosfato liberado na reação, a quantidade de bases incorporadas também será proporcional.

A plataforma Roche 454 mais avançada em comercialização pela Roche Applied Science é capaz de gerar 700 Mb de sequências de até 1.000 pares de base em uma corrida de 23 horas.

3.2. Princípios de funcionamento do sistema SOLiD

A plataforma *Sequencing by Oligo Ligation and Detection* (SOLiD) começou a ser comercializada em 2007. Nela, os fragmentos de DNA são ligados a sequências adaptadoras e, em seguida, à superfície de contas esféricas (um fragmento em cada conta esférica). Neste sistema, apenas uma microgota de água em emulsão em óleo contém todos os reagentes de amplificação. Após a desnaturação do DNA, as contas esféricas são imobilizadas em uma superfície e os iniciadores se anelam ao adaptador. Uma mistura de octâmeros também é hibridizada ao DNA e são adicionados os reagentes da ligação. No octâmero são caracterizadas as bases duas-a-duas, de acordo com a fluorescência na extremidade da sequência ligada. Depois da detecção a sequência é clivada logo após a última base sequenciada e o processo de repete. Com esse método, é possível se chegar a fragmentos de 40 pares de base de extensão. Apesar do baixo comprimento, o método possibilita o sequenciamento em paralelo mais de 50 milhões de agrupamentos (*clusters*) por corrida.

3.3. Princípios de funcionamento do sistema Illumina GAII e HiSeq

A plataforma de sequenciamento da empresa Solexa começou a ser comercializada em 2006 e, logo em seguida, foi adquirida pela empresa Illumina Inc., que investiu no desenvolvimento de analisadores automáticos de genoma. O princípio de funcionamento das duas plataformas comercializadas pela Illumina (GA e HiSeq) baseia-se na química de sequenciamento-por-síntese, com nucleotídeos terminadores reversíveis. Neste método, cada

base é marcada com um corante fluorescente diferente e uma DNA polimerase é capaz de incorporá-los à medida em que o sequenciamento prossegue. Inicialmente, o DNA é fragmentado e ligado a adaptadores em ambas as extremidades do fragmento. Após uma desnaturação, os fragmentos são imobilizados em uma superfície densamente coberta por oligos complementares aos adaptadores. Cada fragmento fita simples imobilizado cria uma “ponte”, ou seja, a outra extremidade do fragmento hibridiza com os oligos da superfície sólida, sendo clonados e formando um agrupamento (*cluster*). São adicionados então os reagentes de amplificação, e os adaptadores servem como iniciadores. Novamente se faz necessário uma reação de amplificação para que haja um aumento na intensidade do sinal luminoso emitido. A cada incorporação são detectados a posição na superfície e o nucleotídeo incorporado através da fluorescência emitida. Os reagentes são retirados e adicionados novamente, repetindo o ciclo de síntese. A quantidade de ciclos que se deseja obter é ajustável, variando de 40 até 150. O sequenciamento pode ser feito utilizando sequências de extremidade pareadas ou através de uma extremidade simples. Com essa tecnologia é possível sequenciar centenas de milhões de *clusters* simultaneamente (sequenciamento em paralelo). O analisador de genomas GAllx permite sequenciar cerca de 600 milhões de sequências de 76 pares de base por corrida em cada célula de fluxo (*flow cell*), enquanto o analisador HiSeq atinge, em média e com maior acurácia, cerca de 3 bilhões de sequências de 76 pb por corrida.

4. Descoberta de marcadores SNP

4.1. SNPs como marcadores moleculares

O sequenciamento de nova geração permite revelar polimorfismo de DNA em uma escala inédita através da comparação de sequências dos genomas sequenciados. Um marcador SNP (*Single Nucleotide Polymorphism*) pode ser definido como um sítio do DNA onde foi observada a substituição de uma única base entre amostras de indivíduos de uma mesma população (Risch and Merikangas, 1996). Um SNP é entendido ainda como uma posição no DNA onde diferentes sequências alternativas (alelos) co-existem em indivíduos em uma população, e o alelo menos frequente possui frequência maior que 1% (Brookes, 1999). Ao considerar a frequência alélica na definição de SNPs, distingue-se esta classe de marcador de meras mutações ao acaso que ocorrem no genoma ou mesmo de erros de sequenciamento que, pela baixa frequência, dificultariam o uso da tecnologia em análise genética. Por definição, marcadores SNP não incluem pequenas inserções e deleções (*indels*), embora alguns autores ainda mantenham esse dois tipos de polimorfismo (*SNP e indels*) na mesma categoria.

Em teoria, é possível a existência de quatro alelos diferentes para cada nucleotídeo em um sítio SNP, uma vez que existem quatro bases nitrogenadas que compõem o DNA (A, C, T,

G). Porém, na prática, o que se observa é a presença em maior frequência de apenas duas possíveis variações, fato que pode ser explicado pela ocorrência desigual de substituições de base do tipo transição ($A \leftrightarrow G, T \leftrightarrow C$) e transversão ($A \leftrightarrow C, A \leftrightarrow T, G \leftrightarrow C, G \leftrightarrow T$). Assim, apesar do número possível de transversões ser duas vezes maior do que o de transições, o que se observa na prática é que a ocorrência de transições é cerca de duas vezes maior do que de transversões (Vignal, 2002), fato que pode ser explicado pela alta taxa de deaminação espontânea da 5-metil-citosina em timina em dinucleotídeos CpG (Coulondre, 1978). Portanto, marcadores SNP normalmente possuem uma natureza bialélica e a tecnologia desenvolvida para detectar este polimorfismo é também baseada no bialelismo. Os equipamentos e a química hoje utilizadas para a detecção de SNPs detecta apenas dois alelos em cada loco.

Em princípio, a análise de polimorfismo bialélico definido pela variação de um sítio de DNA não é novidade, visto que marcadores RAPD também se baseiam no mesmo tipo de polimorfismo bialélico (presença e ausência de banda de DNA). Contudo, o comportamento no caso de marcadores RAPD é de dominância, isto é, não é possível distinguir os heterozigotos (presença/ausência de banda) de uma classe homozigota (presença/presença de bandas). SNPs têm comportamento co-dominante, isto é, no genótipo heterozigoto os dois alelos são identificados. O interesse e a procura por SNPs nos diferentes organismos aumentou muito recentemente. O sequenciamento completo do genoma de espécies modelo e o aumento do número de sequências disponíveis em bancos de dados públicos vem estimulando o interesse da comunidade científica por este tipo de marcador.

A natureza bialélica dos SNPs, ou seja, a detecção de apenas dois possíveis alelos em cada loco SNP de uma planta diplóide os torna menos informativos do ponto de vista genético quando comparados a marcadores microssatélites. Isto pode ser compensado pela abundância de SNPs, encontrados em larga escala ao longo do genoma (Kruglyak, 1997). O SNP é o polimorfismo de DNA mais abundante no genoma. A ocorrência e distribuição de SNPs no genoma de plantas têm sido amplamente estudadas. Em *Arabidopsis*, por exemplo, observa-se 1 SNP a cada 3.3 kilobases (Drenkard et al., 2000); em soja, 1 SNP a cada 200 pb (Graef & Diers, 2004); em milho são ainda mais frequentes, podendo chegar a 1 SNP a cada 70 pb (Bhatramakki, 2000). Em arroz, observa-se em média 1 SNP a cada 232 pb entre variedades escolhidas ao acaso (Nasu, 2002). A estimativa do número de SNPs observados em uma espécie depende, naturalmente, das relações de vínculo genético entre as amostras de acessos utilizadas na análise. Se as amostras apresentam grande diversidade genética, a tendência é se observar maior número de SNPs a cada kpb analisado.

A grande abundância de SNPs no genoma vem estimulando o seu emprego em pesquisa básica e aplicada de diferentes espécies. SNPs têm sido utilizados com sucesso, por exemplo, na construção de mapas genéticos de alta densidade, caracterizados por uma ocorrência no genoma muito além da capacidade dos marcadores tradicionais (Hyten et al., 2010). A elevada densidade de SNPs nos mapas genéticos em teoria aumenta a possibilidade de que SNPs próximos a genes de interesse agrônômico sejam amostrados e, com isso,

associações entre genótipo e fenótipo sejam mais prontamente encontradas, o que facilitaria a seleção assistida para características complexas.

Deve ser frisado que a frequência de SNPs é heterogênea ao longo do genoma, diferindo entre região codificadora/região não-codificadora, por exemplo, ou entre sequências gênicas ou mesmo dentro de genes (éxon/intron). Em geral, os SNPs são menos abundantes em regiões gênicas (codificadoras) do que em outras regiões (Fladung & Buschbom, 2009).

4.2. Descoberta de SNPs

Enquanto milhares de marcadores SNP são amplamente usados em rotina em análises genômicas nas áreas animal e humana, o seu uso em plantas e programas de melhoramento só ganhou força nos últimos anos, quando foram publicados trabalhos em grandes culturas (Baird et al., 2008; Collard et al., 2005; Eckert et al., 2008; Hyten et al., 2010; Huang et al., 2009a; Hyten et al., 2010; Price et al., 2002; Sanchez et al., 2002; Yu et al., 2011). Em geral, estes estudos avaliaram milhares de SNPs em centenas de amostras de plantas, tornando possível analisar o genoma alvo com uma alta densidade de marcadores, com o objetivo de tentar buscar associação dos marcadores com uma característica quantitativa de interesse. Este é o princípio básico dos estudos de genética associativa (*Genome Wide Association Studies* – GWAS).

Existem diversas metodologias de detecção e desenvolvimento de SNPs. As primeiras abordagens foram baseadas no alinhamento de sequências geradas pelo método Sanger, através de comparações do mesmo segmento genômico entre diversos indivíduos. Esse processo, contudo, é considerado pouco eficiente, devido à limitada capacidade de geração e análise de sequências. O recente desenvolvimento de novas tecnologias de sequenciamento mudou esse cenário, baixando custos, acelerando a velocidade na geração de dados de sequências e facilitando o processo de comparação de genomas. Atualmente, os principais métodos de descoberta de SNPs podem ser separados em: detecção de SNPs a partir de sequências de EST em bancos de dados e detecção de SNPs a partir do re-sequenciamento e montagem do genoma estrutural via sequenciamento de nova geração (NGS).

4.2.1. Descoberta de SNPs a partir de sequências ESTs em bancos de dados

Várias espécies de plantas possuem um alto número de sequências ESTs (*Expressed Sequence Tags*), disponíveis em bancos de sequência, como o GeneBank, NCBI, EUA. O número de ESTs com sequência publicada no *NCBI EST database* pode variar de centenas para espécies menos estudadas até milhões de sequências para espécies modelo ou de maior importância econômica (Benson et al., 2010). Apesar da potencial vantagem da existência de

sequências EST para um organismo de interesse, a descoberta de SNPs a partir de banco de dados de ESTs deve levar em consideração três aspectos: (1) a qualidade das sequências no banco de dados deve ser alta. Isto porque o valor de *Phred score* mínimo para ESTs depositados em Bancos de Dados é 20, o que significa que pode ocorrer um erro na sequência a cada 100pb. Isto pode favorecer a detecção espúria de SNPs através da comparação de sequências de qualidade inferior. Neste caso, seriam necessários vários ESTs do mesmo gene do mesmo indivíduo para aumentar a confiabilidade de detecção; (2) a dificuldade de separar sequências ortólogas (derivadas do mesmo gene) de sequências parálogas (derivadas de genes relacionados ou da mesma família gênica). Nesta situação, a ausência de informação pode levar à falsa detecção de SNPs; (3) o número de indivíduos usados na obtenção de ESTs, visto que é comum a obtenção de diferentes bibliotecas de ESTs de um mesmo genótipo ou de genótipos muito aparentados (quase-isogênicos), o que pode induzir falsos positivos. Como resultado dessas limitações, o número de SNPs identificados e validados usando essa metodologia de comparação de sequências ESTs depositadas em bancos de dados tem sido relativamente baixo, com taxas de erro entre 15 e 50% (Benson et al., 2010).

4.2.2. Descoberta de SNPs a partir de re-sequenciamento e montagem do genoma estrutural através de tecnologia NGS

O desenvolvimento de novas tecnologias de sequenciamento em alta escala mudou o cenário da geração de sequência, com custos que baixam a cada ano, tendo aumentado exponencialmente a velocidade de geração de dados de sequências nos últimos anos. As chamadas tecnologias de sequenciamento de segunda geração ou NGS (*Next Generation Sequencing*), incluindo Roche 454 (Margulies et al., 2005); Solexa/Illumina (Bennett, 2004) e ABI Solid (Valouev et al., 2008), são capazes de produzir uma quantidade enorme de dados de sequência, da ordem de bilhões de bases, em um curto intervalo de tempo, em sistemas altamente automatizados, e a um baixo custo por base sequenciada (Ganal, 2009). Dessa forma, essas tecnologias aumentaram significativamente a velocidade de descoberta de SNPs e, conseqüentemente, estimularam a sua aplicação em diversas áreas da ciência, inclusive no melhoramento genético de plantas, como em cevada (Wicker et al., 2008), eucalipto (Novaes et al., 2008), pinus (Cronn et al., 2008) e milho (Emrich et al., 2007). A tecnologia NGS oferece oportunidades para a análise do genoma estrutural de plantas, com ou sem um genoma de referência. Ela pode ser aplicada para o estudo de uma variedade bastante ampla de questões biológicas, desde a análise da estrutura do genoma ou de variações de níveis de expressão de transcritomas, até a descoberta de microssatélites e de SNPs, ou mesmo análise de interações DNA-proteína.

5. Métodos de genotipagem de marcadores SNP

Nos últimos anos, diversas metodologias de genotipagem de SNPs foram desenvolvidas baseadas em diferentes técnicas de discriminação alélica e diferentes plataformas de detecção. Devido à rápida e contínua evolução dos métodos, torna-se difícil identificar a melhor metodologia a ser utilizada, visto que a melhor adequação depende de diversas variáveis, como espécie estudada, nível de informação genômica, custo, natureza do ensaio, etc. Não existe um método ideal de genotipagem de SNPs, porém espera-se que o método utilizado satisfaça alguns quesitos como: (1) alta capacidade multiplex, possibilitando a genotipagem de vários sítios SNP ao mesmo tempo; (2) alta acurácia na detecção de polimorfismo, com identidade de genótipos entre réplicas próxima de 100%; (3) capacidade de genotipagem simultânea de diversas amostras; (4) rapidez na obtenção de dados; (5) baixo custo operacional e por genótipo obtido.

Os ensaios de genotipagem de SNPs, quanto ao mecanismo molecular de detecção, incluem as seguintes categorias: hibridização alelo-específica (*Allele Specific Oligonucleotide Hybridization - ASH*), extensão de primer (*Primer Extension*) e ligação de oligonucleotídeos (*Oligonucleotide Ligation*).

5.1. Hibridização alelo-específica e tecnologias afins

Nesse sistema duas sondas alelo-específicas são desenhadas, de forma que a base polimórfica fique no centro do oligonucleotídeo. Durante o anelamento das sondas ao DNA-alvo, apenas as sondas com hibridização completa são estáveis, enquanto os híbrido não-perfeitos não são estáveis, possibilitando a detecção do polimorfismo. Dentre os sistemas que utilizam esse mecanismo, um dos mais utilizados e com maior capacidade de multiplexagem é o sistema TaqMan (Applied Biosystems, 2002) (Figura 1). Nesse sistema, a discriminação alélica depende de duas sondas alelo-específicas, marcadas com um corante fluorescente sonda-específico e com um sequestrador (*quencher*). Durante a amplificação das sequências flanqueadoras do SNP, as sondas perfeitamente complementares ao DNA alvo são clivadas pela atividade 5'-exonuclease da *Taq* polimerase, tendo então emissão de fluorescência. Os genótipos são atribuídos através da análise da intensidade da fluorescência dos dois corantes.

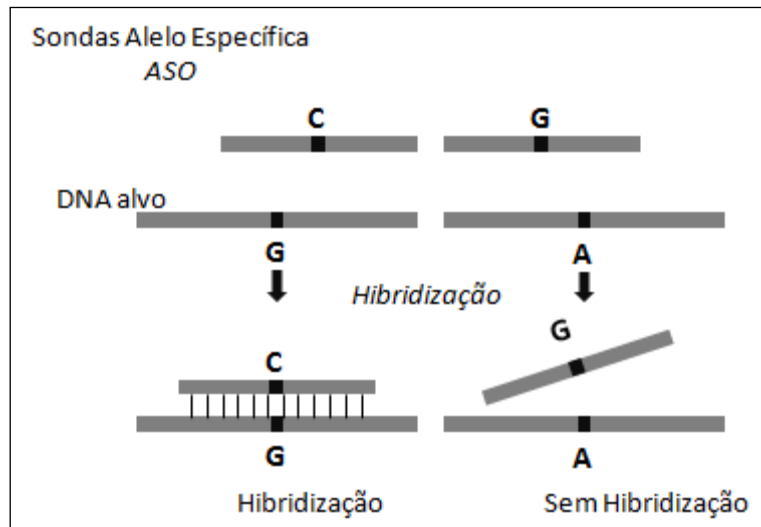


Figura 1: Esquema ilustrativo de hibridização alelo-específica empregada na detecção de alelos em sítios SNP – Sonda Alelo Específica (fonte: www.appliedbiosystems.com).

5.2. Extensão de primer e tecnologias afins

Também conhecida como mini-sequenciamento, a extensão de primer baseia-se na extensão de um iniciador anelado adjacente ao SNP alvo, sendo a incorporação do nucleotídeo polimórfico detectada através da emissão de fluorescência. Alguns sistemas como o SNaPshot (Applied Biosystems) (Figura 2) envolvem uma marcação fluorescente diferente para cada nucleotídeo, permitindo a identificação do nucleotídeo incorporado.

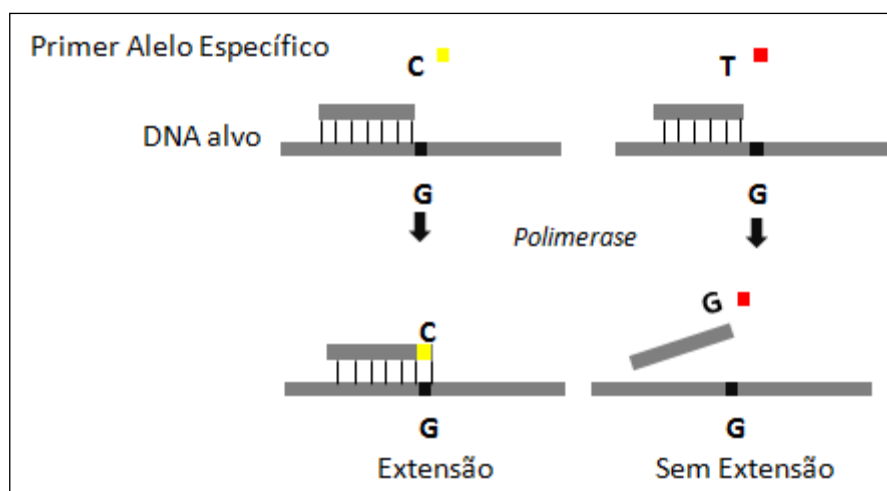


Figura 2. Esquema ilustrativo de hibridização alelo-específica empregada na detecção de alelos em sítios SNP – Primer Alelo Específico (fonte: www.appliedbiosystems.com).

5.3. Ligação de bases nitrogenadas e tecnologias afins

Esta metodologia baseia-se na capacidade de uma enzima DNA ligase “juntar” dois oligonucleotídeos hibridizados em sítios próximos um do outro. Requer que três sondas sejam desenhadas: duas alelo-específicas e uma comum aos dois alelos. A sonda comum se anela imediatamente depois (*downstream*) do sítio SNP. Cada uma das duas sondas alelo-específicas possui na posição final 3' uma base complementar a cada um dos alelos analisados. Sendo assim, apenas as sondas aneladas perfeitamente ao sítio do SNP será ligada à sonda comum. Nesse caso, a discriminação alélica é feita com uma marcação diferencial com corante para cada alelo. Dos diversos sistemas comercialmente disponíveis com essa metodologia, dois se destacam: SNPlex (Tobler et al., 2005) e GoldenGate (Fan et al., 2003) (Figura 3). A diferença entre os dois sistemas está centrada basicamente nos corantes utilizados e no método de detecção da fluorescência.

Grande atenção tem sido dada ao sistema GoldenGate para a detecção de SNPs. A tecnologia GoldenGate da Illumina (*GGGT – GoldenGate Genotyping Technology*) utiliza o mecanismo molecular de ligação dos oligonucleotídeos para discriminar os alelos. Existem dois sistemas de genotipagem comercializados pela Illumina: (1) *GoldenGate assay*, composto por arranjos de pequena a média densidade, podendo variar de 48 a 3.072 SNPs por ensaio, com a opção de emprego metodologias BeadArray ou VeraCode; (2) o *Infinium array*, baseado em microarranjos de DNA, permite ensaios de alta densidade, podendo chegar até 1 milhão de marcadores por arranjo.

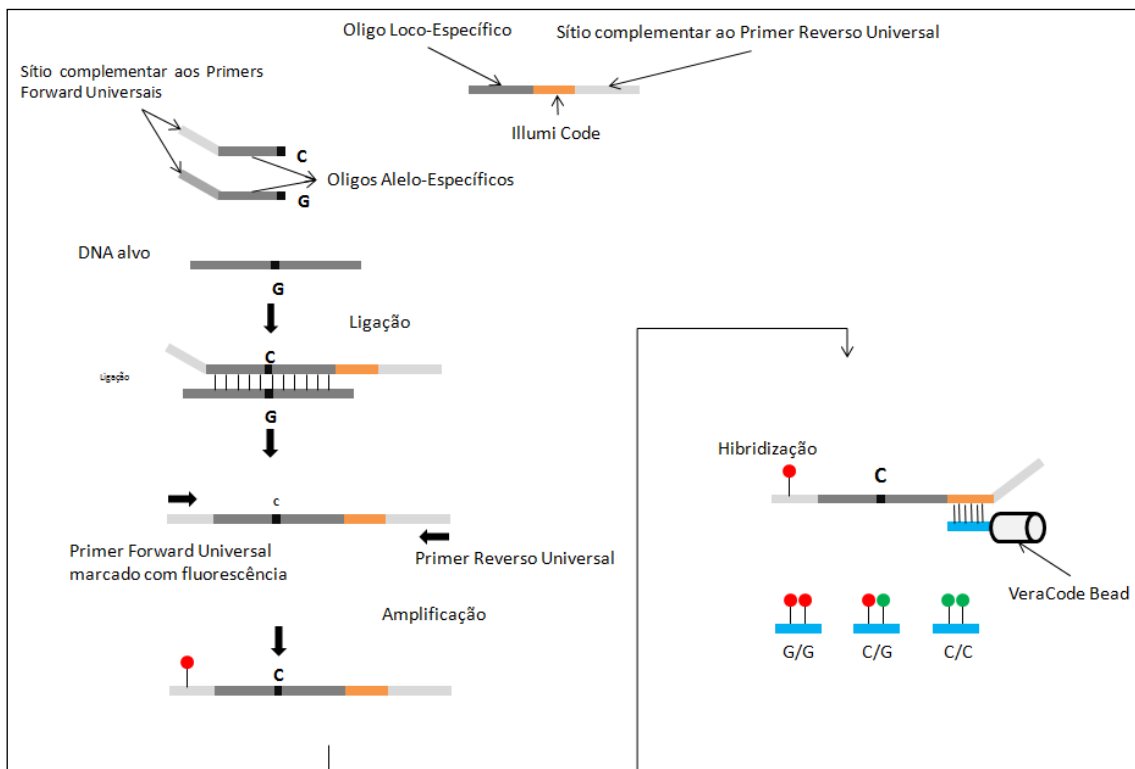


Figura 3. Esquema ilustrativo da técnica de ligaçao de oligonucleotídeos empregada na detecçao de alelos em sítios SNP – Método VeraCode (fonte: www.illumina.com).

No presente trabalho, os ensaios de detecçao de SNP empregaram a metodologia VeraCode. Neste caso, como já mencionado anteriormente, o mecanismo de discriminaçao por ligaçao requer dois primers (oligonucleotídeos) alelo-específicos (*Allele Specific Oligonucleotides* – ASO) e um oligonucleotídeo comum (*Locus specific oligonucleotide* – LSO). Os dois ASOs além de serem complementares à região anterior (*upstream*) ao SNP analisado, contém uma sequência complementar a um iniciador universal, assim como o LSO, que também contém uma sequência complementar a um iniciador universal, para discriminar entre as duas possíveis variantes alélicas para cada SNP. Os nucleotídeos de cada alelo-específico que são complementares à base que constitui o SNP estão localizados no terminal 3' dos ASOs. O outro oligo, que é específico para cada loco, mas comum a qualquer um dos dois alelos, anela após a posicão do SNP e possui um código de sequência (IllumiCode) único que identifica e diferencia cada SNP. Após o anelamento, o oligonucleotídeo alelo específico com a base 3' complementar a base correspondente ao seu alelo específico é estendido e ligado ao oligonucleotídeo loco específico, formando um DNA molde, que será posteriormente amplificado pela reação da polimerase em cadeia. A reação de ligaçao é realizada em múltiplos locos simultaneamente. Os moldes então gerados são amplificados via PCR utilizando um conjunto de três iniciadores universais, comuns a todos os SNPs. Dos três iniciadores utilizados, os iniciadores *forward* são marcados no terminal 5' com um fluorocromo diferente para cada primer (Cy3 ou Cy5), sendo que cada um anela com uma das duas sequências alternativas dos ASO. O terceiro primer anela com uma sequência comum no LSO. Os

produtos de PCR são, em seguida, anelados a oligonucleotídeos complementares com as seqüências IllumiCode no LSO utilizados para reconhecer cada sítio de SNP e estes imobilizados em contas esféricas, que contém um código de barras específico para cada SNP. As intensidades das fluorescências Cy3 e Cy5 são utilizadas para determinar o genótipo presente no sítio de SNP. Intensidades proporcionais de Cy3/Cy5 correspondem a genótipos heterozigotos enquanto que a predominância de apenas uma das fluorescências corresponde a cada um dos genótipos homozigotos.

Em plantas, painéis de SNPs para diferentes aplicações vêm sendo desenvolvidos ao longo dos últimos anos, com destaque para a construção de mapas genéticos com alta densidade de SNPs em diversas espécies, como pinus (Eckert et al., 2008), milho (Yan et al., 2010), soja (Hyten et al., 2010) e tomate (Shirasawa et al., 2010).

Recentemente, alguns painéis de marcadores SNP foram desenvolvidos e testados para análise de polimorfismo de DNA em arroz. Cinco painéis de marcadores SNPs com diferentes densidades (variação de 96 a 384 SNPs por painel) foram desenvolvidos para serem usados para detectar diversidade entre diferentes grupos ou populações de arroz (*indica x japonica*; *indica x aus*; *japonica* tropical dos EUA; *indica x O. rufipogon*; *japonica x O. rufipogon*) (Thompson et al., 2011). Estes painéis tiveram como base cerca de 1% dos SNPs detectados pelo projeto Pelergen/OryzaSNP (McNally et al., 2009; Zhao et al., 2010), que culminou na detecção de ~160.000 SNPs através do sequenciamento parcial de 20 acessos arroz pertencentes a diferentes grupos e tipos da espécie. No Japão, outro painel foi desenvolvido para a detecção de polimorfismo entre variedades de arroz *japonica* temperada (Yamamoto et al., 2010; Nagasaki et al., 2010). É importante observar que os painéis mencionados têm sido eficientes na detecção de polimorfismo nos grupos/populações para os quais foram desenhados. A expectativa de emprego eficiente destes painéis em outras populações é baixa devido à significativa subestruturação populacional encontrada em arroz (Glaszmann, 1987; Garris et al., 2005). Testes específicos para verificar o grau de eficiência de painéis em outros grupos/populações ainda não foram realizados.

6. Descoberta e detecção simultânea de SNPs através de Genotipagem por Sequenciamento (GBS – Genotyping by Sequencing)

As recentes reduções nos custos de sequenciamento têm permitido um rápido desenvolvimento e avanço de tecnologias de genotipagem em larga escala, o que possibilita a obtenção de genótipos em milhares de locos em um único experimento. Metodologias eficientes em custo e velocidade de genotipagem de um grande número de amostras em milhares de locos eram inexistentes até recentemente. As metodologias que hoje integram conjuntos de marcadores SNP para construção de painéis ou arranjos que permitem a detecção de genótipos em centenas (RT-PCR, Sequenom MassARRAY, Illumina GoldenGate)

ou até dezenas de milhares de marcadores SNPs (Affymetrix, Agilent) apresentam custos cada vez menores. Porém, o custo unitário de cada amostra genotipada ainda não é acessível à maioria dos programas de melhoramento genético.

Algumas metodologias recentemente testadas, baseadas em sequenciamento de nova geração (NGS), combinam a descoberta de polimorfismo e genotipagem em um único passo (Baird et al., 2008; Huang et al., 2009; Elshire et al., 2011). Isto significa que, ao invés de definir *a priori* quais são os SNPs que serão genotipados ao longo do genoma, essas novas metodologias primeiro detectam e genotipam os SNPs, para uma seleção *a posteriori* dos SNPs que serão usados na análise com base em um conjunto de critérios. As metodologias citadas (Baird et al., 2008; Huang et al., 2009; Elshire et al., 2011) geram uma alta quantidade de marcadores, amplamente distribuídos ao longo do genoma, além de poderem ser aplicadas em *pools* ou grupos de amostras. Outra vantagem dessas metodologias consiste na independência de genoma de referência, fato que contribui enormemente para a sua aplicação em organismos não-modelo, que carecem de informação genômica.

O método proposto por Huang et al. (2009) é baseado no re-sequenciamento completo do genoma e é aparentemente mais simples do que os outros, não envolvendo etapas de purificação, seleção de tamanho e digestão. Porém, como qualquer método baseado em construção de uma biblioteca de sequenciamento com protocolo padrão Illumina (fragmentação e posterior ligação do DNA a sequências adaptadoras), é um método relativamente caro, trabalhoso e que exige uma quantidade de DNA genômico inicial relativamente alta, o que em alguns casos pode ser um problema. Além disso, a metodologia apresenta um poder de multiplexagem ou *pooling* das amostras considerado baixo, quando comparado com os outros dois métodos.

Os métodos de RAD-seq ou *Restriction site Associated DNA* (Baird et al., 2008) e GBS (*Genotype by Sequencing*) (Elshire et al., 2011) envolvem uma etapa inicial de redução de complexidade do genoma alvo. Para isto o DNA deve ser cortado com uma ou mais enzimas de restrição para fragmentar o genoma da espécie alvo. Os fragmentos resultantes da digestão enzimática são, em seguida, separados por eletroforese e uma faixa de tamanho de fragmentos é escolhida para possibilitar a redução da complexidade. Nestas metodologias, a eficiência de detecção de SNPs é altamente dependente da etapa de redução de complexidade. A análise pode ser realizada com amostras multiplexadas ou com amostras únicas. O polimorfismo detectado nos fragmentos sequenciados (SNPs) pode ser usado como marcador genético, considerando o marcador como dominante (presença/ausência do sítio de restrição) ou como co-dominante (SNP dentro do fragmento sequenciado).

A escolha de um método ou outro deve levar em conta os tipos de marcadores gerados e a disponibilidade ou não de genoma de referência. Por exemplo, para estudos de populações silvestres, estruturação, filogenia, e em que não existe genoma de referência, um marcador passível de genotipagem e presente na maioria dos indivíduos analisados é o mais indicado. RAD-seq parece ser uma metodologia apropriada para estes tipos de estudos. Para ensaios de seleção assistida por marcadores, mapeamento de QTLs, onde os genótipos parentais são

conhecidos, existe um genoma de referência, o uso de GBS parece ser mais recomendado. Porém, trata-se apenas de uma indicação com base nos artigos originais, visto que as duas técnicas parecem ter aplicação ampla e diferentes variáveis devem ser levadas em consideração na escolha de uma ou outra, incluindo tempo e custo de genotipagem por amostra.

Marcadores do tipo RAD foram concebidos como marcadores para serem implementados em microarranjos, sendo a metodologia adaptada para sequenciamento de alta escala. Os marcadores consistem em sequências flanqueadoras de um sítio de restrição da enzima utilizada. Os produtos da digestão são então fragmentados ao acaso, ligados a adaptadores e amplificados por PCR. Esses fragmentos amplificados então são sequenciados, geralmente em sequenciadores Illumina (HiSeq ou GAIIx). Para espécies com carência de genoma de referência é possível sequenciar fragmentos maiores (até 150pb) e pareados (2x 150pb) para uma montagem *de novo* mais eficiente.

A metodologia de GBS (Elshire et al., 2011) envolve uma etapa de redução de complexidade do genoma através da digestão por enzima de restrição com sítio frequente no genoma alvo. O experimento original utilizou a enzima ApeKI para milho e sorgo, uma enzima com corte frequente nos dois genomas, sensível à metilação, termoestável e com sítio de restrição pouco frequente em regiões repetitivas do genoma. O ensaio GBS faz uso de adaptadores com “barcode” e adaptadores comuns, que são ligados às extremidades de corte dos fragmentos. O “barcode” é uma sequência de DNA específica inserida no adaptador para identificar cada genótipo em ensaios multiplex. Desta forma, a capacidade de multiplexagem do método pode chegar a grupos de amostras de até 384 genótipos (indivíduos) diferentes em cada reação de sequenciamento. Porém, quanto maior o número de genótipos, menor a quantidade de sequências obtidas de cada genótipo, o que pode dificultar a análise posteriormente.

O adaptador contendo *barcode* e o adaptador comum são misturados e usados nas reações de ligação. Porém, nem todos os fragmentos ligados aos adaptadores serão sequenciados, uma vez que não tendo os dois adaptadores complementares (adaptador com *barcode* e o adaptador comum), o fragmento não pode ser amplificado em ponte (*bridge PCR*) de maneira eficiente, não formando então grupos (*clusters*) de sequenciamento. Fragmentos longos (> 1kpb) também não são amplificados pelo impedimento de tamanho na formação das pontes e não são sequenciados. A metodologia GBS permite a detecção de centenas de milhares de SNPs no conjunto de genótipos analisado. A seleção dos SNPs úteis é feita seguindo vários critérios, o que leva a uma redução significativa do número de SNPs considerados para análise. Mesmo assim, ao final do procedimento obtém-se uma quantidade significativa de marcadores SNP (milhares), com grau de polimorfismo elevado.

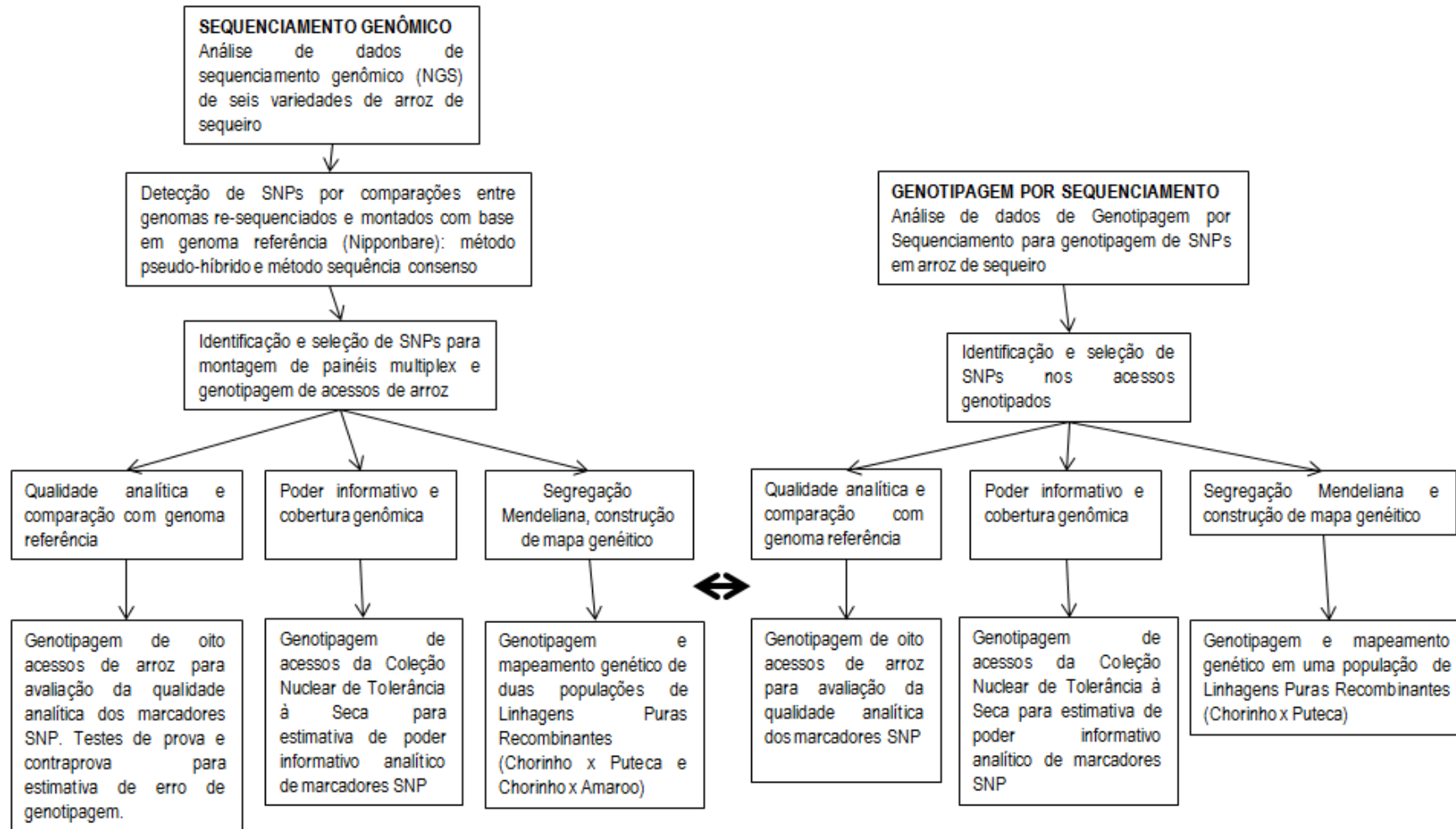
7. Objetivos

- Selecionar e desenvolver painéis de marcadores SNP para análise genética de acessos de arroz empregando tecnologia de sequenciamento de nova geração (*NGS – Next Generation Sequencing*).
- Testar tecnologia de Genotipagem por Sequenciamento (*GBS – Genotyping by Sequencing*) para a detecção e genotipagem de SNPs em acessos de arroz.

8. Objetivos específicos

- Utilizar dados de sequenciamento de DNA baseado em tecnologia de sequenciamento de nova geração (*NGS – Next Generation Sequencing*) para obter a sequência do genoma estrutural de seis cultivares de arroz;
- Montar o genoma estrutural de seis cultivares de arroz de sequeiro com base no genoma referência da cultivar Nipponbare;
- Detectar e selecionar SNPs pela comparação dos genomas re-sequenciados de seis cultivares de arroz;
- Desenvolver painéis de marcadores SNPs para genotipagem em larga escala de acessos de arroz.
- Analisar a qualidade analítica, a acurácia, o conteúdo informativo e o potencial de emprego dos painéis de marcadores SNPs selecionados em análise genética de arroz;
- Detectar e selecionar SNPs no genoma de arroz por *GBS (Genotyping by Sequencing ou Genotipagem por Sequenciamento)* e avaliar o potencial de emprego dos marcadores SNPs selecionados em análise genética de arroz.
- Comparar as tecnologias envolvidas na genotipagem de marcadores SNP (painéis multiplex e *GBS*) para emprego em rotinas de programas de melhoramento e de conservação de germoplasma de arroz

9. Fluxograma experimental



CAPÍTULO 1

**Desenvolvimento de painéis multiplex de marcadores SNP para genotipagem em
escala de acessos de arroz *japonica* tropical (*Oryza sativa* L.)**

CAPÍTULO 1

Desenvolvimento de painéis multiplex de marcadores SNP para genotipagem em escala de acessos de arroz *japonica* tropical (*Oryza sativa* L.)

1. Resumo

Entre as recentes tecnologias que podem ser aplicadas na detecção e desenvolvimento de marcadores moleculares, destaque tem sido dado à tecnologia NGS (*Next Generation Sequencing*). A tecnologia NGS possibilita a detecção de polimorfismo de DNA através da comparação de sequências dos genomas estruturais de diferentes cultivares de uma espécie. O polimorfismo de DNA mais abundante, revelado por esta estratégia, é o polimorfismo de base de DNA (substituição ou transição de base), conhecido como SNP (*Single Nucleotide Polymorphism*). Neste trabalho foram utilizados resultados de sequenciamento do genoma de seis variedades de arroz *japonica* tropical para detectar, selecionar e desenhar painéis multiplex para genotipagem automatizada e simultânea de centenas de marcadores SNP em acessos de arroz de sequeiro. Foram desenvolvidos dois painéis multiplex (SNP1 e SNP2), com 384 marcadores SNP cada, totalizando 768 SNPs distribuídos pelo genoma de arroz. Os painéis foram avaliados quanto à eficiência de clusterização de intensidade do sinal de fluorescência para a definição dos genótipos, à acurácia de genotipagem em testes de prova e contraprova, à capacidade de discriminação de acessos do Banco de Germoplasma e ao potencial emprego na construção de mapas genéticos de cruzamentos de interesse do programa de melhoramento genético. Os testes realizados comprovaram a eficiência da chamada de genótipos e a elevada concordância de genótipos em testes de prova e contraprova. Observou-se que erros de genotipagem ocorrem a uma taxa muito baixa (0,33%). Cerca de 70% dos SNPs possuem o alelo mais comum com frequência entre 0,50 e 0,75, e são poucos os SNPs (<3%) com frequência do alelo mais comum acima de 0,95. O poder combinado dos painéis na confirmação de identidade genética ou na discriminação de genótipos é extremamente elevado. A segregação e a herança dos alelos dos SNPs que compõem os dois painéis foi testada em duas populações de linhagens puras recombinantes, possibilitando a construção de dois mapas genéticos. O processo de genotipagem utilizando painéis SNP em genotipador automático é simples, rápido e eficiente. Os painéis SNP1 e SNP2, portanto, possuem ampla aplicação e são recomendados para análise genética de acessos *japonica* tropical de arroz do Brasil.

2. Introdução

O arroz (*Oryza sativa* L.) é a principal cultura agrícola destinada à alimentação direta da população humana. A cultura tem especial importância na dieta dos países asiáticos, com alta densidade populacional, onde vem sendo cultivada por milênios. O maior produtor e consumidor de arroz entre os países ocidentais é o Brasil. O arroz é alimento consumido diariamente pela população brasileira. Porém, ao contrário dos países asiáticos, onde predomina o plantio de arroz irrigado para a produção de grãos, cerca de 50% da área destinada à produção de arroz no Brasil é plantada com arroz de sequeiro.

O arroz de sequeiro é cultivado na região de Cerrados durante a estação chuvosa e a irrigação das lavouras é realizada pela água oriunda de precipitação. A produtividade média de arroz de sequeiro é cerca de 3x menor que a de arroz irrigado. O incremento da produtividade média de arroz de sequeiro teria forte impacto na produção nacional, proporcionando um superávit que seria providencial para o consumo interno e para a exportação de grãos.

O melhoramento genético e a adoção crescente de tecnologia agrícola têm contribuído para um aumento significativo da produção e da qualidade dos grãos de arroz de sequeiro. Hoje, em algumas áreas especializadas no plantio de arroz de sequeiro é possível atingir produtividades acima de 4.500 kg/ha, com uma qualidade dos grãos similar à observada em sistema de arroz irrigado. Isto demonstra o potencial de aumento de produtividade da cultura do arroz de sequeiro no Brasil. Como a área plantada neste modo de produção é significativa, o aumento de produtividade de cultivares adaptados ao cultivo de sequeiro representará um novo paradigma para a produção brasileira.

A ocorrência de curtos períodos de estiagem durante o período chuvoso, conhecidos como “veranicos”, afeta negativamente a produtividade do arroz de sequeiro. O déficit de água durante a floração ou enchimento de grãos pode causar decréscimo de até 40% da produção. O uso de cultivares adaptadas ao efeito de veranico (ex. cultivares com sistema radicular mais profundo e mais desenvolvido) pode minimizar a quebra de produção. Os programas de melhoramento genético de arroz de sequeiro podem usufruir de avanços recentes da genética molecular e da genômica para a seleção de genótipos superiores para desenvolvimento radicular. O uso de marcadores moleculares abundantemente distribuídos por todo o genoma pode ser importante recurso na identificação de alelos associados a fenótipos de interesse, como desenvolvimento radicular, em programas de seleção assistida para esta característica. Este trabalho visa explorar a detecção e emprego de marcadores moleculares em escala na análise do germoplasma de arroz de sequeiro do Brasil.

Entre as várias técnicas detecção e desenvolvimento de marcadores moleculares, recente destaque tem sido dado à tecnologia NGS (*Next Generation Sequencing*). Esta tecnologia proporciona um aumento exponencial na quantidade de bases sequenciadas em um único ensaio de sequenciamento e, conseqüentemente, altera significativamente a relação

custo:benefício, oferecendo um preço final por base sequenciada mais atraente. A tecnologia NGS possibilita a detecção de polimorfismo de DNA através da comparação de sequências dos genomas estruturais de diferentes cultivares de uma espécie. O polimorfismo de DNA mais abundante, revelado por esta estratégia, é o polimorfismo de base de DNA (substituição ou transição de base), conhecido como SNP (*Single Nucleotide Polymorphism*). SNP pode ser definido como um sítio do DNA onde foi observada a substituição de uma única base entre amostras de indivíduos de uma mesma população (Risch and Merikangas, 1996). Um SNP é entendido ainda como uma posição no DNA onde diferentes sequências alternativas (alelos) coexistem em indivíduos de uma população, e o alelo menos frequente possui frequência maior que 1% (Brookes, 1999).

Os esforços iniciais de sequenciamento ao acaso do genoma de arroz culminaram na montagem da primeira versão do genoma da espécie, baseada na cultivar *japonica* Nipponbare (Goff et al., 2002) e da variedade *indica* 9311 (Yu et al., 2002). Em seguida, o sequenciamento de alta acurácia da cultivar Nipponbare, ancorado por mapas genéticos e sequenciamento de pontas de clones BACS, foi concluído pelo Consórcio Internacional de Sequenciamento do Genoma de Arroz (IRGSP, 2005). Desde então, a cultivar Nipponbare vem sendo usada como referência para o genoma do arroz, e a comparação dos genomas de Nipponbare e 9311 (*japonica x indica*) vem sendo usado para a detecção e seleção de SNPs (Shen et al., 2004; Fetus et al., 2004). SNPs do genoma de arroz têm sido identificados ainda através da análise do banco de dados do projeto de alinhamento de mapas de *Oryza* (OMAP – *Oryza Map Alignment Project*, Ammiraju et al., 2006). Estes SNPs são resultantes do re-sequenciamento de pontas de BAC utilizados na montagem do mapa físico do genoma de doze espécies silvestres de arroz usando Nipponbare como referência. Outro projeto usado como referência para a detecção e seleção de SNPs tem sido o projeto OryzaSNP (www.oryzasnp.org), que gerou recentemente um banco de dados com ~160.000 SNPs derivados do re-sequenciamento parcial de 20 variedades de arroz usando tecnologia de sequenciamento da Perlegen BioSciences (McNally et al., 2009). O projeto, porém, conseguiu cobrir apenas 100 Mpb do genoma de arroz e os segmentos com cobertura suficiente para a detecção de SNPs na comparação entre os 20 acessos re-sequenciados foi limitada a ~11% deste total (McCouch et al., 2010).

Esforços adicionais são necessários para a detecção de SNPs em *pools* gênicos de interesse específico de programas de melhoramento genético, como é o caso de arroz de sequeiro do Brasil. A subestruturação das populações de arroz, analisada em diferentes níveis, é significativa (Glaszmann, 1987; Garris et al., 2005), caracterizando grupos e tipos dentro da espécie com diferenças significativas de frequências alélicas. Estes resultados indicam que um conjunto limitado de marcadores pré-selecionados tende a ser altamente polimórfico apenas para determinados grupo/tipo da espécie, o que significa a necessidade, ao optar por painéis de SNPs selecionados *a priori*, de um esforço especial de identificação de marcadores que

revelem polimorfismo no germoplasma de interesse. Este parece ser o caso das variedades *japonica* tropical do Brasil, utilizadas pelo programa de melhoramento da Embrapa.

Os resultados de re-sequenciamento do genoma de arroz têm sido usados para detectar, selecionar e desenhar painéis multiplex de genotipagem automatizada e simultânea de dezenas, centenas ou milhares de marcadores SNP para uso em genética e melhoramento da espécie. Uma opção tecnológica bastante usada para este fim é a tecnologia GoldenGate, em genotipador BeadXpress (Illumina), embora testes com a tecnologia Affymetrix e Kasp (KBioscience) estejam em andamento (McCouch et al., 2010). Recentemente, alguns painéis de marcadores SNP foram testados em análise de polimorfismo de DNA através da tecnologia GoldenGate, Illumina. Cinco painéis de marcadores SNPs com diferentes densidades (variação de 96 a 384 SNPs por painel) para serem usados para detectar diversidade entre diferentes grupos ou populações de arroz (*indica x japonica*; *indica x aus*; *japonica* tropical dos EUA; *indica x O. rufipogon*; *japonica x O. rufipogon*) (Thompson et al., 2011). Este painéis foram montados com base em SNPs dectados pelo projeto Pelergen/OryzaSNP (Zhao et al., 2010). Outro painel foi desenvolvido para a detecção de polimorfismo entre variedades *japonica* temperada do Japão (Yamamoto et al., 2010; Nagasaki et al., 2010). É importante frisar que estes painéis têm sido eficientes na detecção de polimorfismo nos grupos/populações para os quais foram desenhados. A expectativa de emprego eficiente destes painéis em outras populações é baixa devido à possibilidade de subestruturação populacional entre acessos arroz, mencionada anteriormente. Testes específicos para verificar o grau de eficiência de painéis em outros grupos/populações ainda não foram realizados.

Neste trabalho, dados de re-sequenciamento do genoma de seis variedades de arroz de sequeiro (*japonica* tropical) foram obtidos através de tecnologia NGS e analisados para a detecção, seleção e validação de marcadores SNPs na análise de polimorfismo de DNA de variedades de arroz de sequeiro do Brasil.

3. Objetivos

- Analisar dados de re-sequenciamento do genoma de seis variedades de arroz de sequeiro obtidos através de tecnologia NGS;
- Detectar e selecionar SNPs através da montagem e alinhamento do genoma das variedades re-sequenciadas.
- Validar painéis multiplex de marcadores SNPs para genotipagem automática e análise de polimorfismo de DNA de variedades de arroz de sequeiro do Brasil (*japonica* tropical).

4. Material e Métodos

4.1. Identificação e seleção de SNPs, a partir de dados de sequenciamento genômico de alta performance, para compor painéis multiplex de genotipagem

Inicialmente, oito variedades de arroz de sequeiro (Chorinho, Puteca, Primavera, IAC 165, Catetão, Azucena, Moroberekan e Ligeiro), pertencentes ao grupo *japonica* tropical, foram selecionadas para experimentos de re-sequenciamento e montagem do genoma estrutural de arroz usando tecnologia NGS (*Next Generation Sequencing*) (Ferreira et al., 2010). Com exceção de Primavera, estas variedades possuem variabilidade para tolerância à seca, apresentam sistema radicular robusto e bem desenvolvido, e resistência a diferentes raças de brusone do arroz, causada pelo patógeno *Magnaporthe oryzae*. A cultivar Primavera é a cultivar padrão de qualidade de grãos de arroz de sequeiro no Brasil. As oito variedades foram usadas no desenvolvimento de populações de linhagens puras recombinantes (RILs) em cruzamentos duplos, além de população de seleção recorrente e RIL-octaparental (8-way) derivadas de cruzamentos pareados. Estas populações estão sendo usadas no mapeamento QTLs e identificação de regiões genômicas associadas a tolerância à seca em arroz de sequeiro do Brasil.

No presente trabalho, a seleção dos SNPs para a construção de painéis multiplex foi baseada no alinhamento do genoma re-sequenciado das variedades Primavera, IAC 165, Catetão, Azucena, Moroberekan e Ligeiro. Detalhes da obtenção dos dados de sequenciamento, de montagem e análise dos genomas das cultivares estão disponíveis para consulta (Ferreira et al., 2010).

4.2. Montagem dos genomas re-sequenciados e detecção de SNPs

Foram desenvolvidos e avaliados dois métodos de descoberta de marcadores SNP utilizando os dados re-sequenciamento de seis variedades de arroz de sequeiro (*Oryza sativa* spp. *japonica*), obtidos com sequenciador Illumina Genome Analyser Iix. Todas as operações (montagens, detecção de SNPs, anotações) foram feitas em plataforma Windows, utilizando o software *CLC Genomics Workbench* v. 4.6 (CLC bio, Aarhus, Denmark). Os dois métodos de descoberta de SNPs testados são baseadas em alinhamentos de sequências do genoma dos seis genitores (Primavera, IAC 165, Catetão, Azucena, Moroberekan e Ligeiro). Os alinhamentos foram feitos par-a-par, de acordo com cruzamentos realizados entre estas variedades para geração de populações segregantes, ou seja, Primavera x IAC 165, Catetão x Azucena e Moroberekan x Ligeiro.

1- *Método Pseudo-Híbrido*: o procedimento baseou-se, inicialmente, na detecção de polimorfismo em sítios SNP através do mapeamento no genoma referência de arroz (cv. Nipponbare – v. MSU 6.1) das sequências de duas variedades (A e B) separadamente. Em seguida, foi realizado o mapeamento no genoma referência das sequências do “pseudo-híbrido” AB (agrupamento das sequências de A e de B como se fossem do mesmo genoma).

Após o mapeamento de A, B e de AB, foi feita a detecção de SNPs homocigotos e heterocigotos nas comparações A x Nipponbare, B x Nipponbare e AB x Nipponbare. A lista dos SNPs homocigotos detectados na comparação AB x Nipponbare identificou os SNPs com mesmo alelo para as variedades A e B, mas polimórficos entre estas variedades e o genoma referência. Ao subtrair desta lista os SNPs homocigotos identificados da comparação A x Nipponbare e B x Nipponbare, o resultado obtido foi a lista de SNPs homocigotos para cada variedade e polimórficos entre A e B (Figura 1). Parâmetros para a seleção dos SNPs incluíram uma cobertura mínima da região do sítio SNP de 10x e ausência de outro SNP no intervalo de 100 pares de base acima e abaixo do sítio SNP. Parâmetros de mapeamento no genoma referência incluíram: Fração da Sequência Mapeada (FSM)= 50%; Identidade Mínima da Sequência testada na referência (IMS) = 80%). Para a detecção de polimorfismo usou-se como parâmetro as variações de base com frequência >5% e cobertura mínima na região analisada de 10x. Foi utilizada uma frequência baixa do alelo alternativo (5%) para se retirar a tabela de *locos* em heterocigose na etapa de subtração, causada por heterocigose real ou por artefato de sequenciamento.

2 - *Método Sequência Consenso*: o procedimento baseou-se, inicialmente, no mapeamento das sequências da variedade A tendo como base o genoma referência de arroz (cv. Nipponbare – v. MSU 6.1), seguido da obtenção da sequência consenso do genoma da variedade A. O próximo passo foi a comparação das sequências da variedade A com a sequência consenso de A, visando a identificação de SNPs heterocigotos no genoma A, que foram removidos da análise. O mapeamento das sequências da variedade B, tendo como base o genoma referência de arroz, foi feito em seguida para a obtenção da sequência consenso do genoma da variedade B (“consenso B”). A comparação das sequências da variedade B com a sequência consenso de B permitiu a identificação de SNPs heterocigotos no genoma B, que foram removidos da análise. Comparou-se, então, a sequência consenso da variedade B (sem heterocigotos) com a sequência consenso de A (sem heterocigotos), visando a identificação de SNPs que revelam polimorfismo entre A e B, e que são homocigotos (Figura 1). Os parâmetros para a seleção do SNP incluíram uma cobertura mínima da região do SNP de 10x e ausência de outro SNP no intervalo de 100 pares de base acima e abaixo do sítio SNP considerado. Parâmetros de mapeamento no genoma referência incluíram: Fração da Sequência Mapeada (FSM)= 50%; Identidade Mínima da Sequência testada na referência (IMS) = 80%). Para a detecção de polimorfismo usou-se como parâmetro as variações de base com frequência >5% e cobertura mínima na região analisada de 10x. Foi utilizada uma frequência baixa do alelo alternativo (5%) para se retirar a tabela de *locos* em heterocigose na etapa de subtração, causada por heterocigose real ou por artefato de sequenciamento.

4.3. Montagem de painéis multiplex de SNPs

Foram selecionados 768 marcadores SNP pela comparação das sequências montadas do genoma das variedades Primavera, IAC 165, Catetão, Azucena, Moroberekan e Ligeiro. Os critérios utilizados na seleção dos 768 marcadores SNPs para compor dois painéis multiplex de 384 SNPs (painel SNP1 e SNP2) foram os seguintes:

- (1) SNPs obrigatoriamente homozigotos em cada comparação entre variedade A e B;
- (2) cobertura mínima no sítio do SNP igual ou maior a 10x;
- (3) o marcador SNP revela polimorfismo entre os genitores em pelo menos dois de três cruzamentos (Primavera x IAC 165; Catetão x Azucena; Moroberekan x Ligeiro);
- (4) SNP localizado em região gênica (éxon) do genoma de arroz, sempre que possível;
- (5) *SNP score* elevado (>6.0), definido por critérios do ensaio GoldenGate (www.illumina.com - *Illumina Assay Design Tool*). O *SNP score* do ensaio GoldenGate varia de 0 a 1,00 e reflete a habilidade de desenvolver um ensaio com sucesso com aquele marcador SNP, de acordo com o algoritmo desenvolvido pelo fornecedor. Para integrar um ensaio GoldenGate, o marcador SNP deve: (a) apresentar primers com *Tm* dentro da faixa de temperatura ótima do ensaio; (b) não ser duplicado no genoma e não estar posicionado em região repetitiva; (c) possuir dado em região de pelo menos 100 pb acima e abaixo do sítio SNP para desenho dos primers; (d) estar localizado a mais de 60 pb de outro SNP do mesmo ensaio; (e) ser bi-alélico (o ensaio não é apropriado para SNPs tri ou tetra-alélicos); (f) *indels* não são aceitáveis e bases indefinidas na região do sítio não são admitidas pelo algoritmo.

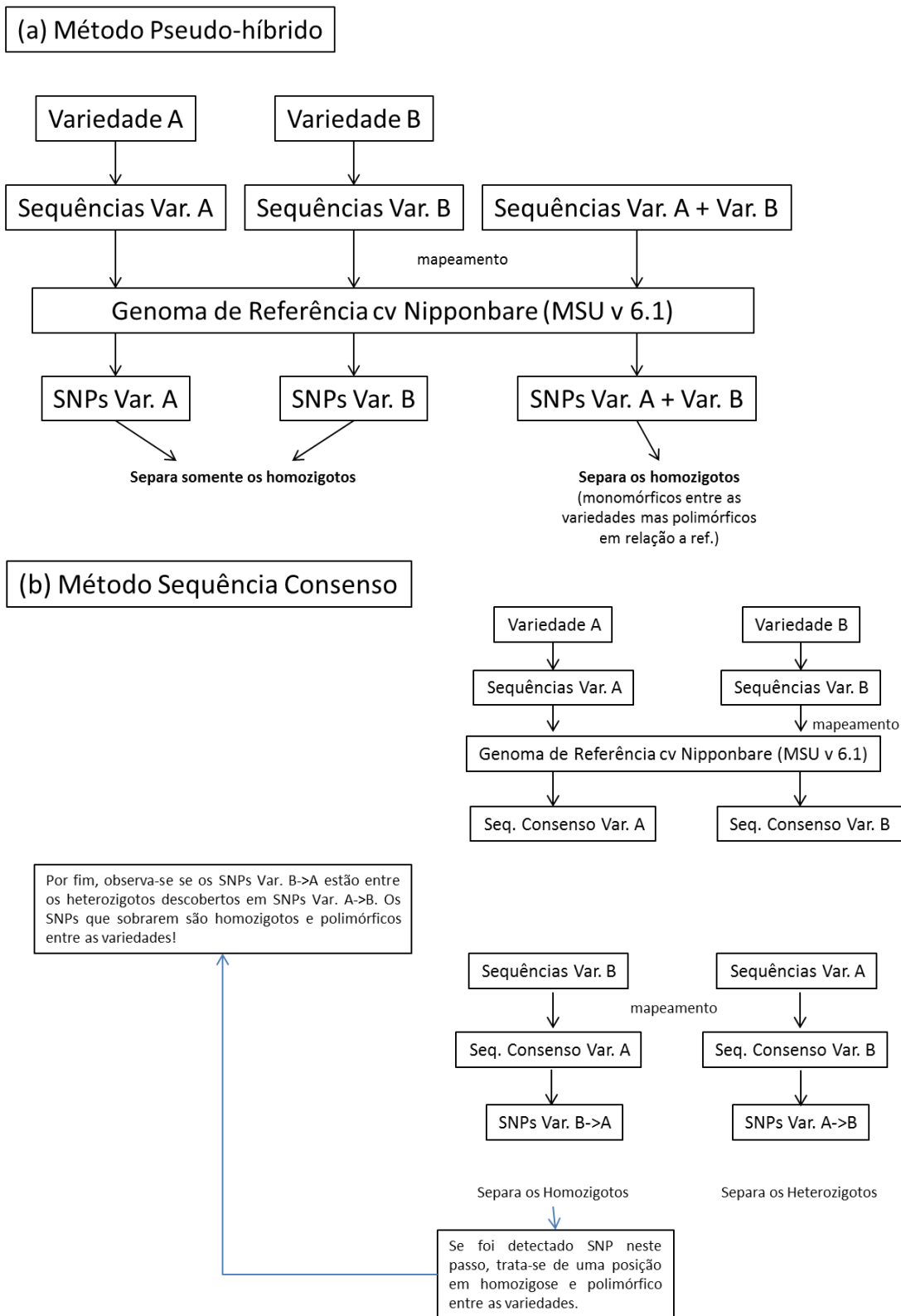


Figura 1. Métodos de detecção de SNPs em comparações de genomas re-sequenciados de seis variedades de arroz: (a) método “pseudo-híbrido”; (b) método “sequência consenso”.

4.4. Validação de painéis multiplex de SNPs

O DNA genômico foi extraído a partir de plântulas recém-germinadas em vaso com vermiculita por um método rápido com CTAB (Ferreira & Grattapaglia, 1998). A qualidade do DNA foi avaliada por eletroforese em gel de agarose 1%. A quantificação foi feita em fluorômetro NanoDrop 3300 (Thermo Scientific), utilizando PicoGreen como corante. O DNA foi diluído em tampão TE para uma concentração final de 50 ng/ μ L.

Os grupos de ensaios com oligonucleotídeos para detecção de SNPs (*SNP Oligonucleotide Pools Assays - OPAs*) foram sintetizados pelo fornecedor (Illumina) de acordo com os critérios de seleção descritos anteriormente. A validação dos marcadores SNP foi realizada em ensaio GoldenGate, utilizando um genotipador BeadXpress (Illumina) nas instalações Laboratório de Genética Vegetal – Embrapa Recursos e Biotecnologia. O protocolo utilizado seguiu as recomendações do fornecedor (Fan et al., 2006). Os dados foram analisados utilizando o software GenomeStudio V2009.2. A qualidade da genotipagem foi avaliada utilizando como principais parâmetros os valores de *Call Rate* e *GenTrain Score* (Illumina).

O teste e validação dos painéis multiplex com amostras de arroz *japonica* tropical foram realizados com os seguintes acessos:

(a) Os oito parentais que tiveram os genomas re-sequenciados (Chorinho, Puteca, Primavera, IAC 165, Catetão, Azucena, Moroberekan e Ligeiro) foram utilizados para avaliar parâmetros de qualidade analítica dos painéis multiplex, com base no padrão de clusterização dos SNPs em classes genotípicas, realizada como o software GenomeStudio V2009.2. O nível de polimorfismo detectado entre os parentais usados na seleção dos SNPs (Primavera, IAC 165, Catetão, Azucena, Moroberekan e Ligeiro) foi comparado com o nível de polimorfismo entre parentais não usados na seleção (Chorinho e Puteca).

(b) Vinte e quatro acessos do Banco de Germoplasma de Arroz foram usados para estimar o nível de erro de genotipagem dos dois painéis 384-plex. Para isso, uma única planta foi amostrada em cada um dos vinte e quatro acessos. O DNA de cada planta foi extraído, quantificado e diluído em duas extrações independentes (prova e contraprova). As genotipagens de prova e contraprova foram realizadas com os dois 384-plex, totalizando 768 marcadores SNPs genotipados na prova e na contraprova de cada planta. Os dados foram utilizados para avaliar o número de marcadores SNPs com genótipo idêntico em cada uma das 24 comparações.

(c) Uma amostra da Coleção Nuclear Temática de Tolerância à Seca (Pessoa Filho et al., 2010), composta por 65 acessos de *japonica* tropical (Tabela 1), foi utilizada para avaliar o

nível de polimorfismo entre acessos de arroz de sequeiro e estimar parâmetros genéticos dos marcadores SNP, visto que esta população apresenta alta diversidade genética. Os parâmetros genéticos foram estimados para cada marcador com base na genotipagem das 65 amostras utilizando o software Powermarker 3.25 (Liu and Muse, 2005), e incluíram, entre outros, a frequência alélica, frequência genotípica, diversidade gênica e PIC (conteúdo de informação polimórfica). Para avaliar a eficiência dos marcadores SNPs utilizados na análise, foi realizada uma estimativa dos valores individuais médios da Probabilidade de Exclusão (PE) em cada loco marcador, bem como da Probabilidade de Exclusão Combinada (PE) de conjuntos de marcadores. Valores de similaridade genética em comparações par-a-par entre os acessos da coleção nuclear foram estimados com base na razão entre a soma das proporções de alelos em comum entre dois acessos (P_s) nos locos microssatélites e duas vezes o número de marcadores testados, utilizando o coeficiente de banda (*band coefficient*; Lynch, M. 1990) implementado no programa NTSYSpc versão 2.10z (Rohlf, 2002). A matriz diagonal de similaridades genéticas foi submetida à análise de agrupamento pelo procedimento de SAHN (*Sequential Agglomerative Hierarchical Nested Cluster Analysis*), usando o método UPGMA (*Unweighted Pair-Group Method, Arithmetic Average*) e um dendrograma de similaridades genéticas foi construído com o programa NTSYSpc versão 2.10z (Rohlf, 2002).

(d) Duas populações de linhagens puras recombinantes (*Recombinant Inbred Lines*), derivadas de parentais diferentes dos usados na seleção dos SNPs foram usadas para avaliar a herança dos marcadores SNPs que compõem os dois painéis, através da construção de dois mapas de ligação. A população RIL Chorinho x Puteca, composta por 187 linhagens *japonica* tropical que segregam para tolerância à seca e outras características de importância econômica (Pessoa Filho, 2009; Schmidt, 2009), e População RIL Chorinho x Amaroo (Ohse, 2008; Lacerda, 2008; Schmidt, 2009), composta por 192 linhagens *japonica* tropical que segregam para tolerância ao frio e resistência à brusone. Os painéis multiplex foram testados nas duas populações para confirmar a segregação nos locos SNP e empregá-los na construção de mapas genéticos nos dois cruzamentos. Os locos SNP polimórficos entre os parentais de cada população RIL foram avaliados quanto ao seu modo de segregação, a fim de se detectar desvios da proporção esperada de 1:1 em uma população F7 de linhagens puras recombinantes. Os testes de distorção de segregação foram realizados utilizando-se o programa GQMOL (Cruz; Schuster, 2005). O teste de FDR, ou razão de falsas descobertas, foi utilizado como critério de ajuste para correção de erro do tipo I (Benjamini; Hochberg, 1995), e o nível de *significância* para os testes de distorção de segregação foi de 1%, após correção de Bonferroni.

Dois mapas de ligação foram construídos utilizando-se o programa MAPMAKER/EXP 3.0 (Lander et al., 1987). Os marcadores SNP selecionados para cada cruzamento foram avaliados em conjunto com 158 marcadores microssatélite previamente genotipados na população Chorinho x Amaroo (Schmidt, 2009), e com 83 marcadores microssatélite na população Chorinho x Puteca (Pessoa Filho, 2008). Os respectivos mapas genéticos construídos com marcadores microssatélite foram utilizados como frameworks na construção dos novos mapas.

Linhagens e locos com mais de 20% de dados faltantes foram excluídas da análise. Um valor de LOD mínimo de 5,0 e uma fração de recombinação (θ) máxima de 0,35 foram utilizados como parâmetros iniciais no estabelecimento dos grupos de ligação pelo comando “group”. A ordem mais provável dos marcadores dentro de cada grupo de ligação foi estimada através dos comandos “three point”, “compare”, “try” e “ripple”. As ordens finais dos locos em cada grupo foram checadas com a ordem dos locos no mapa físico do genoma de arroz, através do banco de dados Gramene (Jaiswal et al., 2006). No caso de incongruências entre mapa físico e mapas genéticos, foi considerada a ordem estimada a partir dos dados experimentais no processo de mapeamento. Os grupos de ligação detectados foram nomeados de acordo com os cromossomos do genoma do arroz. Frequências de recombinação entre marcadores foram convertidas em distâncias de mapa, em centiMorgan (cM), utilizando-se a função de mapeamento de Kosambi (Kosambi, 1943). A representação gráfica dos mapas de ligação foi desenhada utilizando-se o programa MapChart 2.2 (Voorrips, 2002).

Tabela 1. Amostra de 65 acessos da Coleção Nuclear Temática de Tolerância à Seca, utilizados na avaliação dos 768 SNPs que compõem os painéis multiplex SNP1 e SNP2.

ítem	Registro no Banco de Germoplasma	Nome Comum
1	CA780044-1	Cateto Seda
3	CA780059	Maruim
4	CA780158	Cajueiro Liso
5	CA780033	Enche Tulha
6	CA780170	Formosa
7	CA780202-1	Arroz Maranhao
9	CA780287	Paulista Dourado
10	CA780299	Meruim Ligeiro
13	CA780217	Puteca
14	CA780329	Comum Creolo
15	CA780336	Pratao Goiano
16	CA780261	Prata Branco
17	CA790032	Palha Murcha
19	CA790282	Chatao Vermelho
20	CA790301	Ligeiro
21	CA780295	Santo Americo
22	CA790328	Vermelho
23	CA780324	Cacho Duplo
24	CA790346	Nenen
25	CA800068	Amarelao/Douradao
26	CA800081	Brejeiro/Nenenzinho
28	CA800091	Quebra Cacho
29	CA800127	Bico De Rola
30	CA790337	Muruim Branco

31	CA800020	Douradao/Amarelao
32	CA800034	Iguape Sem Aresta
34	CA810050	Samabaia Amarelo
36	CA820069	Paulista
37	CA820103	Arroz Carolino
38	CA830133	Arroz Catetao
40	CA840049	Arroz Do Sul
41	CA860089	Arroz Branco Bico Preto
42	CA870050	Branco 4 Meses
46	CA840148-1	Fundo Roxo
49	CA870078	Meses Branco/3 Meses Amarelo
52	CA870139	Noventinha
53	CA870152	Carolina
57	CA880053	Bico Ganga Cana Roxa
58	CNA0005659	Arroz Ligeiro
59	CA880093	Arroz Piriquito
61	CA880081	Catetao
62	CA980023	Arroz Canela De Ferro
63	CNA0000027	Agulha Esav
64	CNA0000937	Catalao
65	CNA0001420	Carreon
66	CNA0004487	Makouta
67	CNA0005667	Beira Campo Dourado
69	CNA0004762	Pana
70	CA800143	Agulhinha Tardio
72	CA790359	Toró Vermelho
73	CA800015	Douradao
75	CA870068-1	Arroz Comprido
78	CA970012	Agulhinha Vermelho
79	CA980005	Arroz Toro Graudo
81	CA980029-1	Arroz Pindare
83	CNA0002878	Rexoro
55	CA870177	Arroz Roxo ou Caqui
86	CNA0003056	Tapuripa
87	CNA0004510	Piedad
88	CNA0008429	Agulhinha Ipameri
89	CA960008	Caiana Grande
90	CA790238-1	Jatoba
92	CA780357	Aracatuba
93	CA790176-1	Buriti Vermelho
95	CNA0004576	Zenith

5. Resultados e Discussão

5.1. Detecção de marcadores SNPs

Foram testadas duas metodologias de descoberta de SNPs utilizando o software CLC Bio Genomics Workbench, Dinamarca: (1) método “pseudo-híbrido” e (2) método “sequência consenso”. Os resultados indicam que os dois métodos são equivalentes na detecção de SNPs para composição de painéis multiplex. Os dois métodos permitiram a identificação de aproximadamente 80.000 SNP em cada comparação pareada (Primavera x IAC 165; Catetão x Azucena; Moroberakan x Ligeiro). Estes 80.000 SNPs atendem aos critérios mínimos estabelecidos para seleção (ex. cobertura mínima de 10x no sítio de SNP ou localização do SNP em região gênica). Cerca de 25% dos SNPs são comuns entre os três pares de variedades. Outros ~60.000 SNPs são exclusivos de cada par, independentemente da estratégia de descoberta utilizada (Tabela 2).

Tabela 2: Rendimento de SNPs detectados por dois métodos de detecção de polimorfismo de base única em comparações pareadas de sequências genômicas de variedades que tiveram o genoma re-sequenciado

Método de Detecção de SNP	Genitores	Número de SNPs Detectados	Eficiência Relativa
Sequência Consenso	IAC-165 x Primavera	94.611	100,0%
Sequência Consenso	Catetão x Azucena	62.654	66,2%
Sequência Consenso	Moroberakan x Ligeiro	84.193	89,0%
	Média	80.486	85,1 %
Pseudo-híbrido	IAC-165 x Primavera	82.762	87,2%
Pseudo-híbrido	Catetão x Azucena	58.179	61,3%
Pseudo-híbrido	Moroberakan x Ligeiro	94.928	100,0%
	Média	78.623	82,3%

A eficiência relativa média dos dois métodos mostrou-se similar (~80%) nos três cruzamentos avaliados (Tabela 2). O método “pseudo-híbrido” possibilitou a detecção do maior número de SNPs entre todas as combinações cruzamento/método testadas, atingindo 94.928 SNPs no cruzamento 3 (Moroberakan x Ligeiro), seguido pelo método “sequência consenso” no cruzamento 1 (Primavera x IAC 165), com 94.611 SNPs detectados. Os menores rendimento foram observados pelos métodos “pseudo-híbrido” e “sequência consenso” no cruzamento Catetão x Azucena, que apresentaram 58.178 e 62.653 SNPs, respectivamente.

O mapeamento dos SNPs detectados pelo método “sequência consenso” no genoma referência indicou que cerca de 18% estão localizados em éxons de regiões codificadoras (média dos três cruzamentos). Valor similar (20%) foi detectado pelo método “pseudo-híbrido”. É interessante observar que este percentual é quase uma ordem de magnitude superior ao

esperado, visto que a região codificadora do genoma de arroz perfaz apenas 2,5% do genoma total (Yu et al., 2002; Goff et al., 2002). A causa deste excesso de SNPs em éxons, observado nos dois métodos de detecção utilizados, não é conhecida.

Tanto o método “sequência consenso” quanto o método “pseudo-híbrido” possibilitaram a detecção de 4.000 SNPs comuns aos três cruzamentos entre as seis variedades consideradas na análise que tiveram o genoma re-sequenciado (Figuras 2 e 3). Os marcadores em comum naturalmente são passíveis de uso direto em programas de mapeamento genético das três populações de linhagens puras recombinantes derivadas destes seis genitores (Primavera x IAC 165; Catetão x Azucena; Moroberakan x Ligeiro). Espera-se, nas três populações, que o marcador SNP revele o polimorfismo esperado, conforme ilustrado na Figura 4. Observa-se, neste exemplo, que o marcador é polimórfico em três populações ao mesmo tempo: Primavera x IAC 165; Catetão x Azucena; Moroberakan x Ligeiro. Também é possível que este subconjunto de SNPs represente marcadores com maior transferibilidade para outros cruzamentos da subespécie *japonica*.

Os critérios e os métodos de detecção utilizados permitiram a seleção de um SNP a cada 3.8 Kb no genoma de arroz entre as seis variedades re-sequenciadas. Este valor é bastante conservador, muito acima dos valores de detecção descritos na literatura. Isto já era esperado devido ao rigor dos critérios de seleção empregados, e ao fato das variedades utilizadas em outros trabalhos serem geneticamente mais distantes, incluindo variedades de diferentes subespécies (*indica* x *japonica*) (Shen et al., 2004; Fetus et al., 2004), o que implica em um grau de polimorfismo maior.

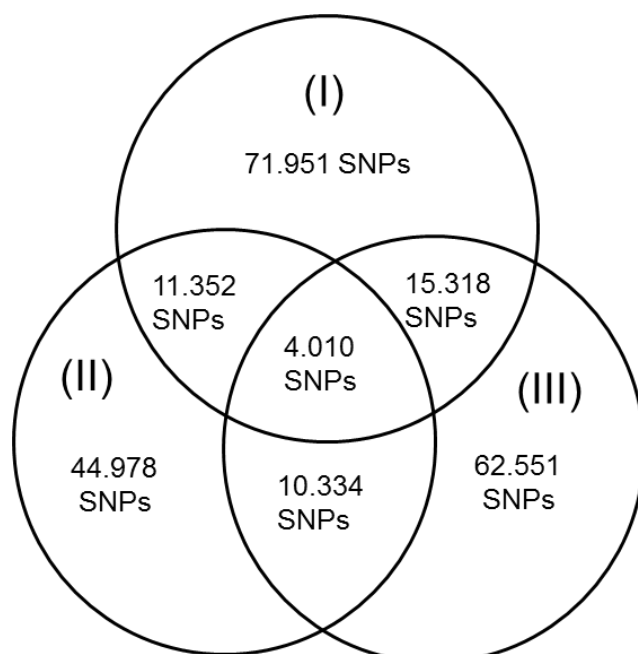


Figura 2: Distribuição de marcadores SNPs comuns em combinações de comparações de sequências de variedades que tiveram o genoma re-sequenciado, identificados pelo método “sequência consenso”. As comparações de sequências de variedades incluem três cruzamentos de arroz de sequeiro: (I) IAC 165 x Primavera; (II) Catetão x Azucena; (III) Ligeiro x Moroberekan.

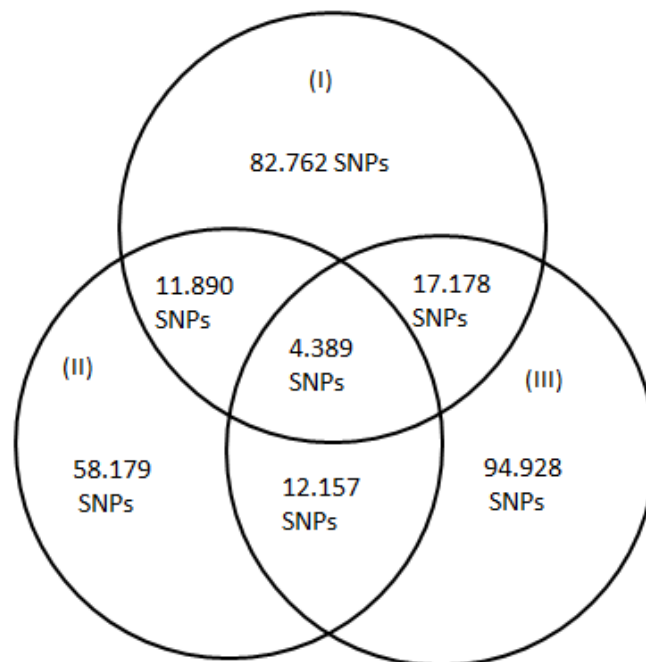


Figura 3: Distribuição de marcadores SNPs comuns em combinações de comparações de sequências de variedades que tiveram o genoma re-sequenciado, identificados pelo método “pseudo-híbrido”. As comparações de sequências de variedades incluem três cruzamentos de arroz de sequeiro: (I) IAC 165 x Primavera; (II) Catetão x Azucena; (III) Ligeiro x Moroberekan.

Observa-se que a distribuição dos sítios SNPs detectados nas duas metodologias, embora cubra todo o genoma, não é homogênea (Figura 5). Nas três comparações (IAC 165 x Primavera; Catetão x Azucena; Ligeiro x Moroberekan), observa-se a existência de regiões mais ricas em SNPs do que outras ao longo de cada cromossomo. As regiões ricas em SNP encontram-se quase sempre conservadas no genoma das seis variedades que tiveram o genoma sequenciado. Em outras palavras, quando detecta-se uma região rica em SNPs no genoma de uma das variedades, observa-se que a mesma região também é rica em SNPs nas demais variedade usadas para seleção destes marcadores.

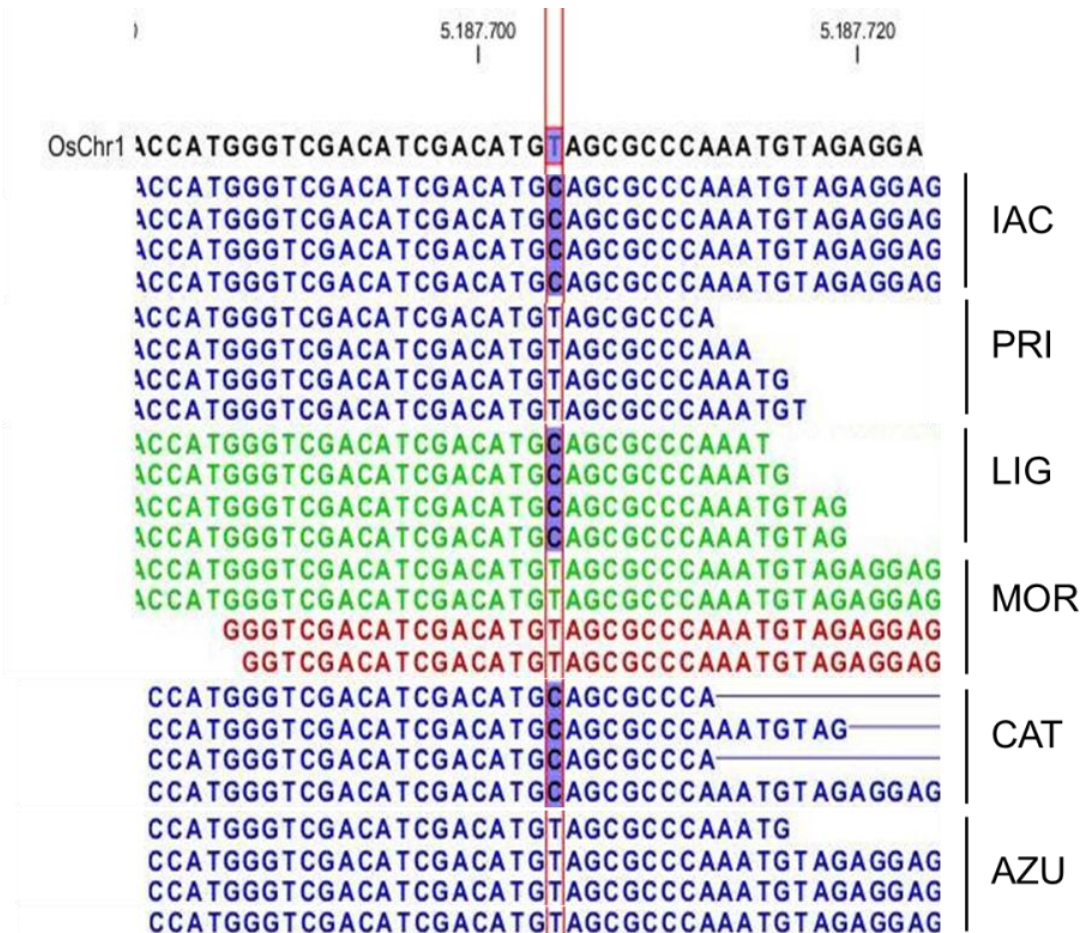


Figura 4: Exemplo de sítio SNP comum entre os genitores dos três cruzamentos analisados: IAC 165 (IAC) x Primavera (PRI); Ligeiro (LIG) x Moroberekan (MOR); Catetão (CAT) x Azucena (AZU). O trecho é uma representação de uma região no cromossomo 1, posição 5.187.704. Sequência no topo corresponde a pseudo-molécula de referência, no caso a cultivar Nipponbare.

A abundância de SNPs detectados pelo método “pseudo-híbrido” e “sequência consenso” permite maior liberdade na escolha dos marcadores moleculares de acordo com a necessidade do programa de melhoramento ou da análise genética. Uma bateria de 768 SNPs foi selecionada para validação e uso em análise genética de arroz. Procurando atender a um conjunto de critérios que maximizasse o potencial sucesso do emprego da bateria de SNPs em genotipagem de acessos de arroz, dois painéis de 384 SNPs foram montados (painéis SNP1 e SNP2).

Os dois painéis possuem construção semelhante, isto é, atendem aos mesmos critérios de seleção de SNPs por cromossomo, sem repetir marcadores entre si. Os dois painéis são complementares, visto que o emprego dos dois painéis ao mesmo tempo diminui proporcionalmente a distância entre os marcadores selecionados dentro de cada cromossomo. Não há repetição de SNPs entre um painel e outro. O painel SNP1 inclui 83 marcadores SNPs localizados em éxons de regiões gênicas distribuídas nos 12 cromossomos de arroz, juntamente com outros 301 distribuídos ao acaso em regiões não codificadoras do genoma (Figura 6; relação de SNPs em ANEXOS); (b) o painel SNP2 inclui 83 marcadores SNPs

localizados em éxons de regiões gênicas distribuídas nos 12 cromossomos de arroz, juntamente com outros 301 distribuídos ao acaso em regiões não codificadoras do genoma (Figura 6; relação de SNPs em ANEXOS).

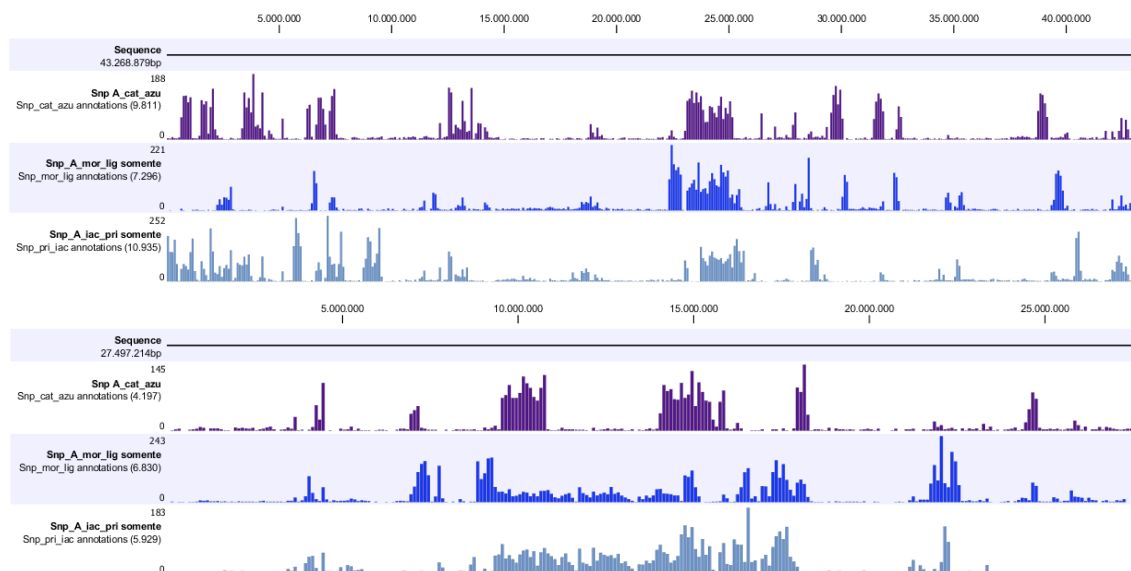


Figura 5: Representação esquemática da distribuição dos SNPs detectados ao longo do genoma de arroz nos seis parentais usados na seleção (IAC 165 x Primavera; Catetão x Azucena; Ligeiro x Moroberekan). Na parte superior da figura observa-se a distribuição de frequências de SNPs detectados no cromossomo 1, enquanto a parte inferior observa-se a distribuição de frequência de SNPs no cromossomo 12 (detecção pelo método “sequência consenso”). Cada barra corresponde a um intervalo de 100kb.

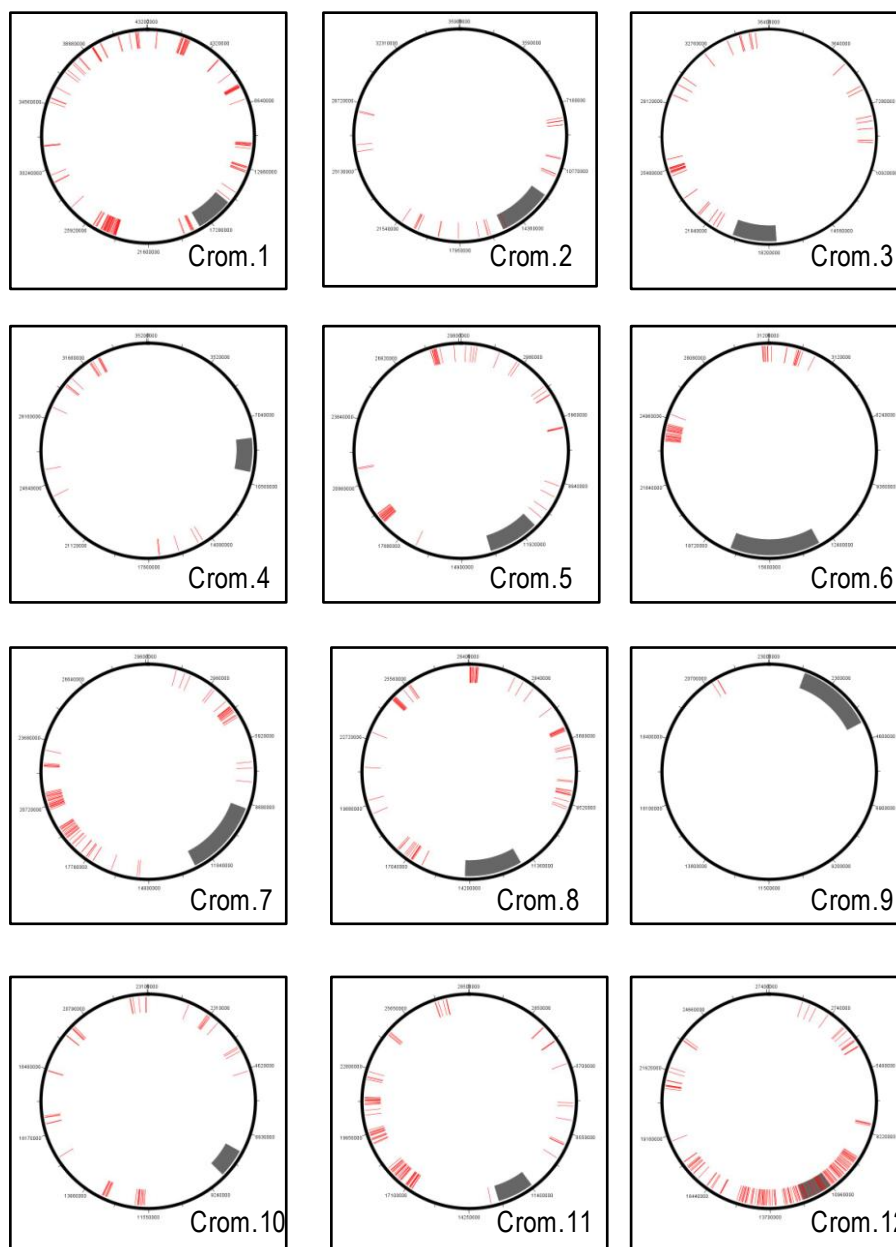


Figura 6 – Distribuição dos marcadores SNP que compõem os dois painéis 384-plex (painéis SNP1 e SNP2) ao longo dos 12 cromossomos de arroz. Cada círculo representa um cromossomo, com origem e término no mesmo ponto. A posição de cada marcador SNP no cromossomo é representada por traços internos (vermelho). Os traços externos (em preto) indicam a posição (pb) em relação à origem do cromossomo. A região destacada em cinza dentro do anel representa a região centromérica.

Os marcadores SNP selecionados apresentam uma distribuição considerada razoável, mas não ideal, ao longo dos cromossomos de arroz (Figura 5), com exceção do cromossomo 9, que está sub-representado. Os marcadores estão em geral espaçados, cobrindo a extensão dos cromossomos. Há grande variação, contudo, no número de marcadores por cromossomos (Figura 7). Inicialmente, é importante frisar que o número de marcadores em cada cromossomo não deve ser necessariamente igual, visto que os cromossomos possuem tamanhos distintos.

Apenas os cromossomos 1, 2 e 3, por exemplo, respondem por cerca de 30% do tamanho do genoma de arroz (Keller and Feuillet, 2004). Nestes três cromossomos o número de marcadores deveria ser necessariamente maior do que nos outros para se manter uma densidade proporcional de marcadores ao longo do genoma. O cromossomo 1, por ser significativamente maior que os demais cromossomos, deveria ter o maior número de marcadores.

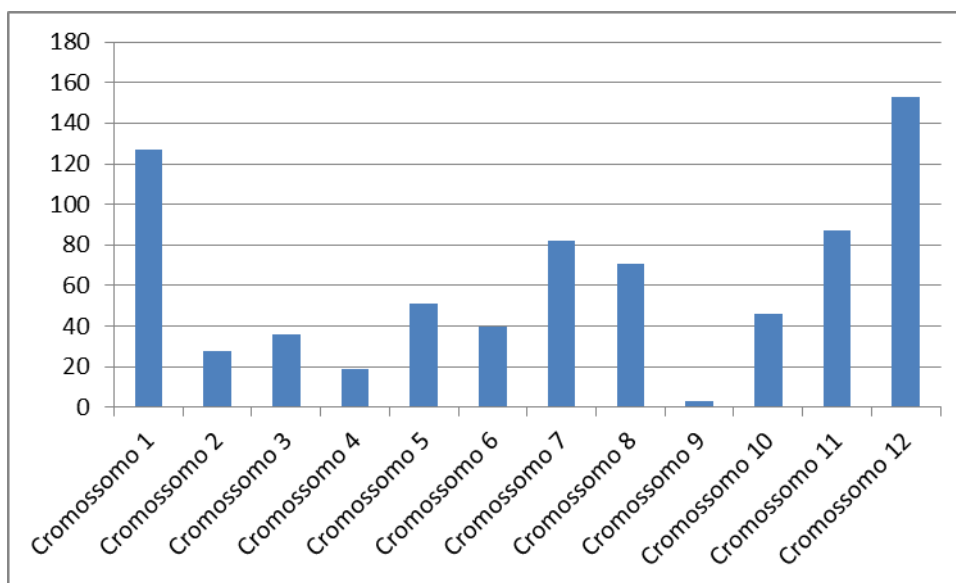


Figura 7 - Distribuição de 768 marcadores SNP que compõem os painéis SNP1 e SNP2 nos 12 cromossomos de arroz.

Contudo, um dos critérios de seleção de SNPs na comparação entre os genomas das seis variedades foi a posição do SNP em éxons de regiões gênicas. Este critério pode afetar significativamente o número de marcadores SNP por cromossomo, visto que o número de genes preditos é muito variável entre os cromossomos de arroz, e a sua localização é também heterogênea. Por exemplo, o cromossomo 10 é o menor cromossomo de arroz, porém possui ~4x mais genes preditos do que o cromossomo 9, que é o segundo menor cromossomo (IRGSP, 2005).

Nos dois painéis 384-plex, o número de marcadores SNP variou de 19 (cromossomo 4) a 156 (cromossomo 12) (Figura 7), com uma média de 62 marcadores SNP por cromossomo. Houve uma sub-representação significativa de SNPs no cromossomo 9, que conta com apenas 3 marcadores nos painéis. Ao final do processo de detecção, desenvolvimento e submissão aos critérios de síntese de primers (*SNP score* mínimo do ensaio GoldenGate), apenas estes 3 marcadores foram considerados para análise no cromossomo 9. Este é o cromossomo com o

menor número de genes no genoma, e com maior representação de sequências repetitivas (Goff et al., 2002; Yu et al., 2002).

5.2. Validação de painéis multiplex de SNPs

Os dois painéis multiplex de marcadores SNP (SNP1 e SNP2) foram testados para avaliação de qualidade analítica, detecção de polimorfismo e emprego em análise genética de arroz.

Inicialmente, as oito variedades empregadas como parentais nas populações de linhagens puras recombinantes (Chorinho, Puteca, Primavera, IAC 165, Catetão, Azucena, Moroberekan e Ligeiro) foram usadas para avaliar a qualidade analítica dos dois conjuntos de marcadores SNP desenvolvidos neste trabalho. A análise dos dados de genotipagem em genotipador BeadXpress permitiu observar que do total de 768 SNPs genotipados, 743 apresentaram uma alta qualidade na chamada dos alelos, ou seja, uma clusterização de intensidade de sinal que não gera dúvida quanto ao genótipo atribuído (Figura 8a). Apenas 25 SNPs foram removidos da análise por apresentarem um padrão de clusterização inadequado, onde não houve como atribuir um genótipo de maneira segura usando o software GenomeStudio no genotipador BeadXpress (Figura 8b). Entre os 25 marcadores excluídos da análise encontram-se alguns que podem ter falhado devido a problema na hibridização alelo-específica ou à química do ensaio. Portanto, após a eliminação dos marcadores que apresentaram algum tipo de problema ou não permitiram uma genotipagem acurada, observou-se que o ensaio teve uma eficiência de 96,74%, caracterizada por alta qualidade de sinal na definição dos genótipos.

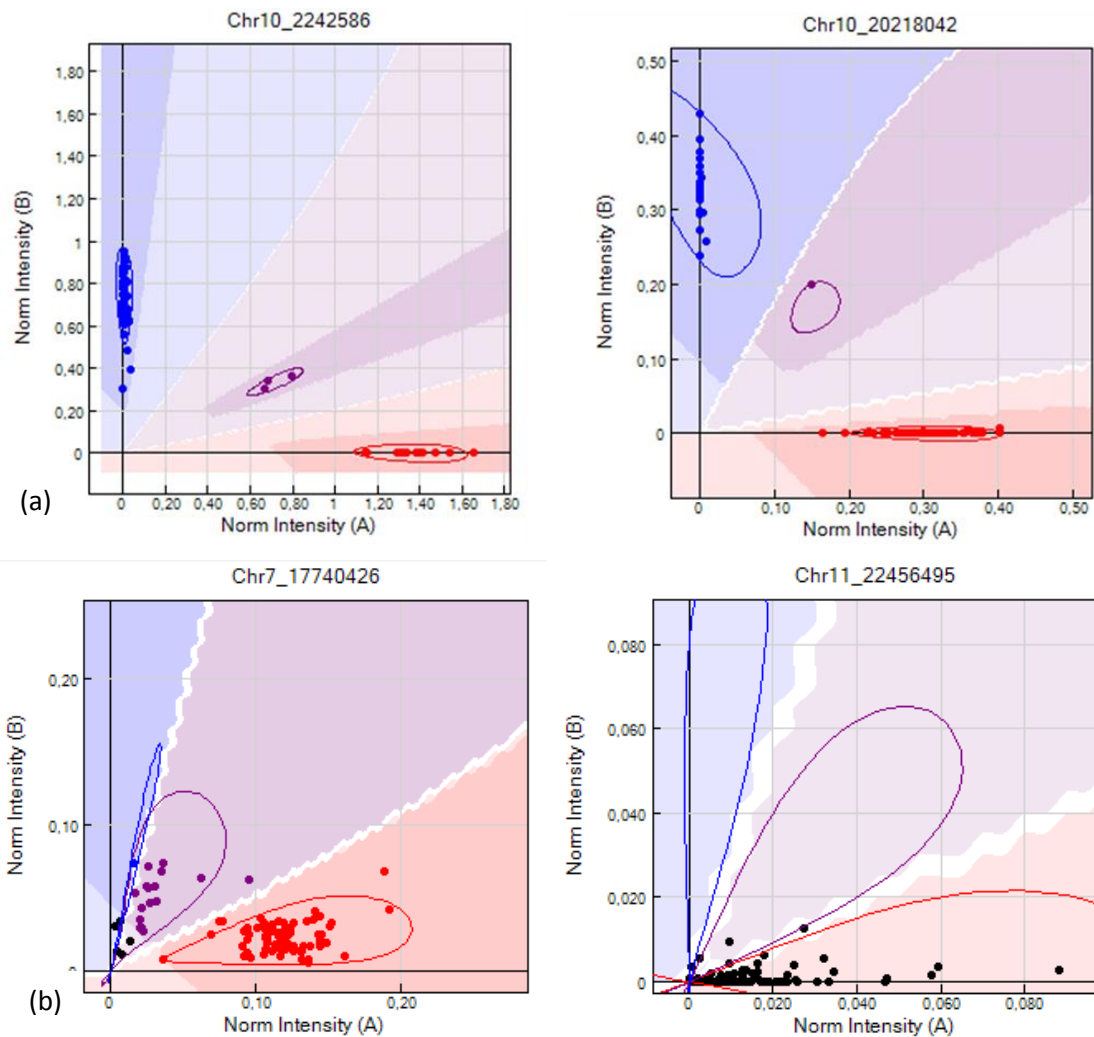


Figura 8: Clusterização de intensidade de sinal de marcadores SNPs utilizando o software GenomeStudio. Quadros superiores(a): Marcadores SNP que apresentam alta qualidade de sinal e clusterização adequada, possibilitando acurácia na determinação dos alelos no loco analisado. Quadros inferiores(b): Marcadores SNP com baixa qualidade de sinal e clusterização inconsistente, gerando dúvida na determinação dos alelos no loco analisado.

A qualidade dos SNPs testados foi ainda avaliada pelo valor GenTrain de cada loco, usando o software GenomeStudio. O GenTrain é uma medida de confiabilidade do marcador baseada nas distribuições dentro e entre classes genotípicas, variando de 0 a 1.0, sendo 1.0 o valor que representa o melhor valor possível para o marcador no conjunto de amostras avaliadas. Na validação utilizando apenas os oito parentais (Chorinho, Puteca, IAC 165, Primavera, Catetão, Azucena, Moroberekan e Ligeiro), a média dos valores de GenTrain de todos os marcadores amostrados (sem nenhuma seleção) foi de 0,75, com um valor mínimo encontrado de zero e atingindo o máximo de 0,98. Na média, portanto, os SNPs apresentam um valor de GenTrain elevado. Os valores de GenTrain na amostra de 743 marcadores SNP que foram selecionados pelo critério de alta qualidade do sinal e acurácia do genótipo foram

analisados em separado. Observou-se que nesta amostra a média dos valores de GenTrain foi de 0,76, com um valor mínimo encontrado de 0,3 e atingindo o máximo de 0,98.

A avaliação do polimorfismo de DNA revelado pelos dois painéis 384-plex ao analisar os parentais das populações RILs (IAC 165 x Primavera; Catetão x Azucena; Ligeiro x Moroberekon; Chorinho x Puteca) indicou grande diversidade de polimorfismo em algumas populações, permitindo inferir que os painéis terão alta eficiência de detecção de polimorfismo em alguns cruzamentos. O cruzamento Primavera x IAC 165, por exemplo, apresentou o maior número de SNPs polimórficos entre os 743 SNPs considerados na análise, atingindo 99,6% de polimorfismo. Portanto, o emprego dos dois painéis multiplex na genotipagem da população RIL Primavera x IAC 165 possibilitará uma alta eficiência de detecção de alelos segregantes, com ampla cobertura de marcadores por todo o genoma, o que será muito positiva para a construção de mapa genético para esta população. Vale lembrar que este é um cruzamento de alto interesse para o programa de melhoramento genético, visto que há um contraste significativo de desenvolvimento radicular e tolerância à seca entre as variedades Primavera e IAC 165. O polimorfismo revelado pelos dois painéis nos cruzamentos Catetão x Azucena e Ligeiro x Moroberekan foi intermediário e similar entre os cruzamentos, atingindo 54,51% e 64,47%, respectivamente. Por fim, os dois painéis foram menos eficientes na detecção de polimorfismo entre as variedades parentais Chorinho e Puteca, que não haviam sido incluídas na análise genômica para detecção e seleção dos SNPs. No cruzamento Chorinho e Puteca, o nível de polimorfismo de marcadores SNPs usados nos dois painéis atingiu apenas 25,3%. Este resultado não é surpreendente, visto que estudos anteriores demonstraram que Chorinho e Puteca são duas variedades tradicionais com similaridade genética maior do que a observada entre os parentais dos outros três cruzamentos. Isto pode ser observado também na análise de similaridade genética realizada com base no polimorfismo de SNPs dos dois painéis, conforme descrito abaixo.

Vinte e quatro acessos do Banco de Germoplasma tiveram o DNA extraído em regime de prova e contraprova para avaliar a acurácia da genotipagem realizada com os dois painéis 384-plex. Como a quantidade total de marcadores SNPs avaliados é significativa, o teste representa um critério simples de avaliação do erro de genotipagem com a bateria de SNPs. Os dados deste teste indicaram que em 9 das 24 comparações de prova e contraprova, o percentual de locos com genótipo idêntico atingiu 100% dos 690 marcadores testados (Tabela 3). Em outras 6 comparações de prova e contraprova, apenas um erro de genotipagem foi observado em 690 locos testados. Em outras 6 comparações, foram observadas três a cinco discrepâncias de genótipos em 690 locos testados. O menor percentual de genótipos idênticos foi observado na comparação 4x43, estimada em 98,7%. Os dados, combinados, contabilizam 54 erros em 16.506 genotipagens realizadas, ou um erro de aproximadamente 0,33%. Portanto, a acurácia dos dois painéis combinados é elevada e adequada à genotipagem de acessos de arroz.

Tabela 3 – Avaliação de acurácia de genotipagem com os dois painéis 384-plex (SNP1 e SNP2) em 24 amostras que tiveram o DNA extraído (prova) e re-extraído (contraprova) para comparação de genótipos.

Prova e Contraprova	Locos idênticos	Locos com dados faltantes	Locos com genótipos distintos	Total de locos comparados	% locos com genótipos idênticos
(1) 6x77	689	12	1	690	99,855%
(2) 30x71	685	0	5	690	99,275%
(3) 17x38	690	0	0	690	100,000%
(4) 37x68	690	0	0	690	100,000%
(5) 19x33	689	2	1	690	99,855%
(6) 49x50	690	2	0	690	100,000%
(7) 44x52	683	97	7	690	98,986%
(8) 4x43	681	0	9	690	98,696%
(9) 31x80	689	11	1	690	99,855%
(10) 34x35	690	0	0	690	100,000%
(11) 34x73	687	0	3	690	99,565%
(12) 35x73	687	0	3	690	99,565%
(13) 10x12	689	1	1	690	99,855%
(14) 10x60	689	1	1	690	99,855%
(15) 12x60	690	2	0	690	100,000%
(16) 11x16	689	4	1	690	99,855%
(17) 11x18	685	34	5	690	99,275%
(18) 16x18	684	34	6	690	99,130%
(19) 26x27	690	6	0	690	100,000%
(20) 26x85	690	6	0	690	100,000%
(21) 27x85	690	6	0	690	100,000%
(22) 55x56	690	0	0	690	100,000%
(23) 55x51	685	1	5	690	99,275%
(24) 56x51	685	1	5	690	99,275%

Para testar o poder informativo dos marcadores SNP selecionados em análise genética de arroz, uma amostra de 65 acessos da Coleção Nuclear Temática de Tolerância à Seca, composta por variedades tradicionais de arroz do Brasil (Tabela 1), apresentando alta diversidade alélica e variabilidade para tolerância à seca (Pessoa-Filho et al., 2010) foi genotipada com marcadores SNP. A análise de clusterização e de valores de GenTrain inicialmente revelou um percentual de marcadores com alta qualidade analítica similar ao percentual observado na análise das oito variedades parentais (~95%). Foram removidos da análise os SNPs com mais de 20% de dados faltantes na amostra da Coleção Nuclear de Tolerância à Seca. Assim, de um total de 768 marcadores SNPs nos dois painéis, 703 marcadores (91%) apresentaram uma alta qualidade e confiabilidade na detecção de alelos

neste grupo de acessos. A amostra de 65 acessos da coleção foi então utilizada para estimar parâmetros genéticos para os marcadores SNPs (Tabela 4; veja tabela completa em ANEXOS). A frequência média do alelo mais comum em 703 SNPs avaliados foi 0,69, com valor mínimo de 0,50 e máximo de 0,98. Observou-se que 69,39% dos marcadores SNPs testados possuem alelos mais comuns com frequência variando entre 0,5 e 0,75. Apenas 3,9% dos marcadores SNP analisados apresentam frequência do alelo mais comum acima de 0,95.

Aproximadamente 62% dos marcadores SNP detectaram pelo menos um genótipo heterozigoto entre os 65 acessos testados. A Heterozigosidade média observada, no entanto, foi bastante baixa ($H_{obs} = 0,03$), com valor mínimo de zero e máximo de 0,80. A Diversidade Gênica variou de 0,03 a 0,05, com valor médio igual a 0,39. O Conteúdo de Informação Polimórfica (PIC) apresentou valor médio de 0,31, variando de 0,03 a 0,38 na amostra analisada. Os valores de PIC estão dentro do esperado para a espécie (Zhao et al., 2010; Thompson et al., 2011), significativamente inferiores aos valores de PIC estimados para marcadores microssatélites usando uma amostra similar de acessos *japonica* tropical (Pessoa Filho, 2008).

Tabela 4 – Parâmetros comparativos de avaliação genética dos marcadores SNP que compõem os dois painéis 384-plex com base na genotipagem de uma amostra de 65 acessos da Coleção Nuclear Temática de Tolerância à Seca de arroz. **

Marcador SNP	Frequência Alelo mais Comum	No. Genótipos	Dados Observados	No. Alelos	Diversidade Gênica	Heterozigosidade observada	PIC
Chr1_11272869	0,62	3	65	2	0,47	0,03	0,36
Chr1_11312519	0,70	2	64	2	0,42	0,61	0,33
Chr1_11339617	0,61	3	64	2	0,48	0,03	0,36
Chr1_11374629	0,67	3	65	2	0,44	0,05	0,34
Chr1_11426073	0,72	3	65	2	0,41	0,02	0,32
Chr1_11465807	0,62	3	65	2	0,47	0,03	0,36
Chr1_11469994	0,72	3	65	2	0,41	0,02	0,32
Chr1_11653926	0,79	3	65	2	0,33	0,05	0,27
Chr1_12867748	0,53	3	59	2	0,50	0,02	0,37
Chr1_12897299	0,60	3	65	2	0,48	0,06	0,36
Chr1_12940593	0,52	3	65	2	0,50	0,02	0,37
Chr1_12976776	0,60	3	65	2	0,48	0,06	0,36
Chr1_13135677	0,53	3	65	2	0,50	0,05	0,37
Chr1_13148124	0,52	3	65	2	0,50	0,08	0,37
(...)							
Chr12_9979324	0,71	2	65	2	0,41	0,58	0,33
Chr12_9982878	0,58	2	65	2	0,49	0,00	0,37
Média	0,69	2,62	64,27	2,00	0,39	0,03	0,31
Mínimo	0,50	2,00	52,00	2,00	0,03	0,00	0,03
Máximo	0,98	3,00	65,00	2,00	0,50	0,80	0,38

**São apresentados apenas parte dos resultados. A tabela completa está apresentada em ANEXOS.

Os dados de genotipagem com SNPs da amostra de 65 acessos da Coleção Nuclear foram analisados em conjunto com os dados de polimorfismo observado nos mesmos locos nas oito variedades parentais (Chorinho, Puteca, Primavera, IAC 165, Moroberekan, Ligeiro, Catetão e Azucena). Um total de 687 marcadores SNPs foi empregado nesta análise, após a eliminação de marcadores com mais de 20% de dados faltantes. Foram realizadas estimativas de similaridade genética em comparações pareadas entre os 73 acessos utilizando o coeficiente *Band* (Lynch, 1990). Os resultados confirmam a ampla diversidade desta amostra de acessos *japonica* tropical (Figura 9). As oito variedades parentais (Chorinho, Puteca, Primavera, IAC 165, Moroberekan, Ligeiro, Catetão e Azucena) estão posicionadas em diversos subgrupos ao longo da distribuição. As variedades Chorinho e Puteca apresentam a maior similaridade genética entre as variedades utilizadas em cruzamentos para desenvolvimento de populações RILs, corroborando análises anteriores desenvolvidas com marcadores microssatélites (Pessoa Filho, 2008). As variedades Primavera e Azucena, embora distantes geneticamente, formam um grupo com algumas outras variedades tradicionais brasileiras utilizadas no estudo (ex. Agulhinha Ipameri e Araçatuba).

Foi observada identidade genética entre os acessos IAC 165 e a variedade CNS 49. A variedade CNS 49, que compõe a Coleção Nuclear Temática de Tolerância à Seca (Tabela 1), está identificada no Banco de Germoplasma de Arroz com o número CA870078 e traz como nome original de coleta a denominação “3meses branco/3meses amarelo”. Foi ainda observada a identidade genética entre Catetão e CNS 61, mas isto já era esperado, visto que CNS 61 está identificada no Banco de Germoplasma de Arroz com o número CA8780081 e traz como nome original de coleta a denominação “Catetão”. As plantas de Catetão utilizadas no desenvolvimento das populações RILs foram selecionadas dentro de amostras de plantas de CNS 61. O mesmo se aplica à alta similaridade genética observada entre Puteca e CNS 13, e entre Ligeiro e CNS 20. Neste caso, plantas de Puteca e Ligeiro foram selecionadas nos acessos originais CNS 13 e CNS 20, respectivamente, para derivar as populações RILs. Estes dados corroboram a eficiência dos painéis SNP1 e SNP2 na estimativa de relações de vínculo genético entre acessos, comprovando a identidade genética entre acessos que são idênticos, ou separando-os eficazmente de acordo com o polimorfismo de DNA observado.

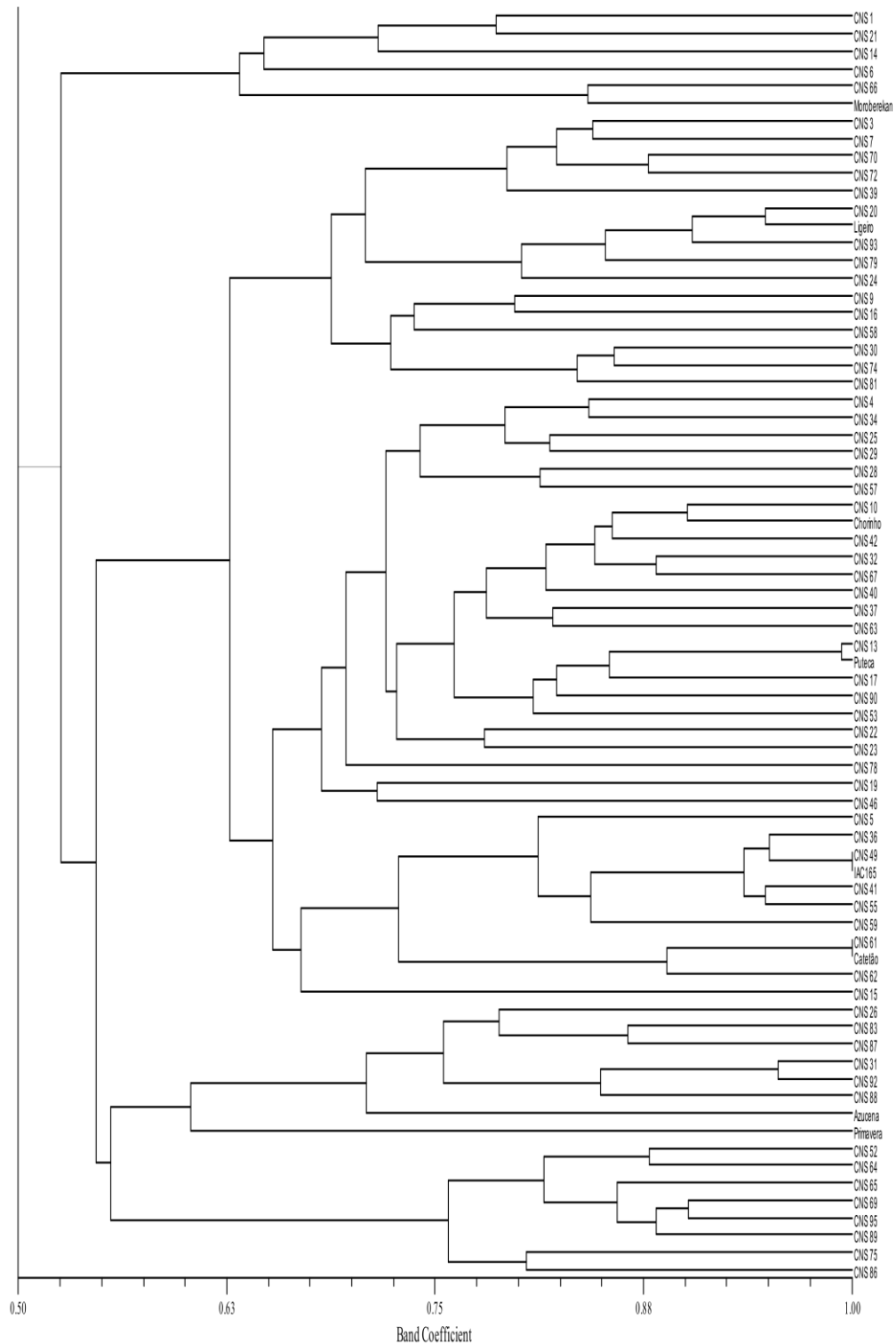


Figura 9. Representação das relações de vínculo genético entre 65 acessos da Coleção Nuclear Temática de Tolerância à Seca e oito variedades utilizadas no desenvolvimento de populações de linhagens puras recombinantes de arroz. O dendrograma foi construído com base nas estimativas de similaridade genética par-a-par inferidas pela proporção de alelos em comum (*band coefficient*), submetidas à análise de agrupamento pelo método de UPGMA.

Os dois painéis 384-plex também foram testados em duas populações de linhagens puras recombinantes (*Recombinant Inbred Lines*), derivadas de parentais diferentes dos usados na seleção original dos marcadores SNP, com o objetivo de avaliar a herança dos marcadores que compõem os dois painéis, através da construção de dois mapas de ligação. O primeiro mapa genético foi construído para a população RIL Chorinho x Amaro (Ohse, 2008; Lacerda, 2008; Schmidt, 2009), composta por 192 linhagens *japonica* tropical que segregam para tolerância ao frio e resistência à brusone (Figura 10). Os marcadores SNPs foram integrados a dados de segregação de marcadores microssatélites e um mapa genético com 1440 cM foi construído. O mapa contém 242 marcadores moleculares distribuídos nos 12 cromossomos de arroz.

O segundo mapa genético foi construído para a população Chorinho x Puteca (Figura 11). Conforme já mencionado anteriormente, apenas 188 (25,3%) marcadores SNP dos dois painéis 384-plex foram polimórficos neste cruzamento. Os marcadores SNPs foram integrados a dados de segregação de marcadores microssatélites e um mapa genético com 1035,5 cM foi construído. O mapa contém 182 marcadores distribuídos nos 12 cromossomos de arroz.

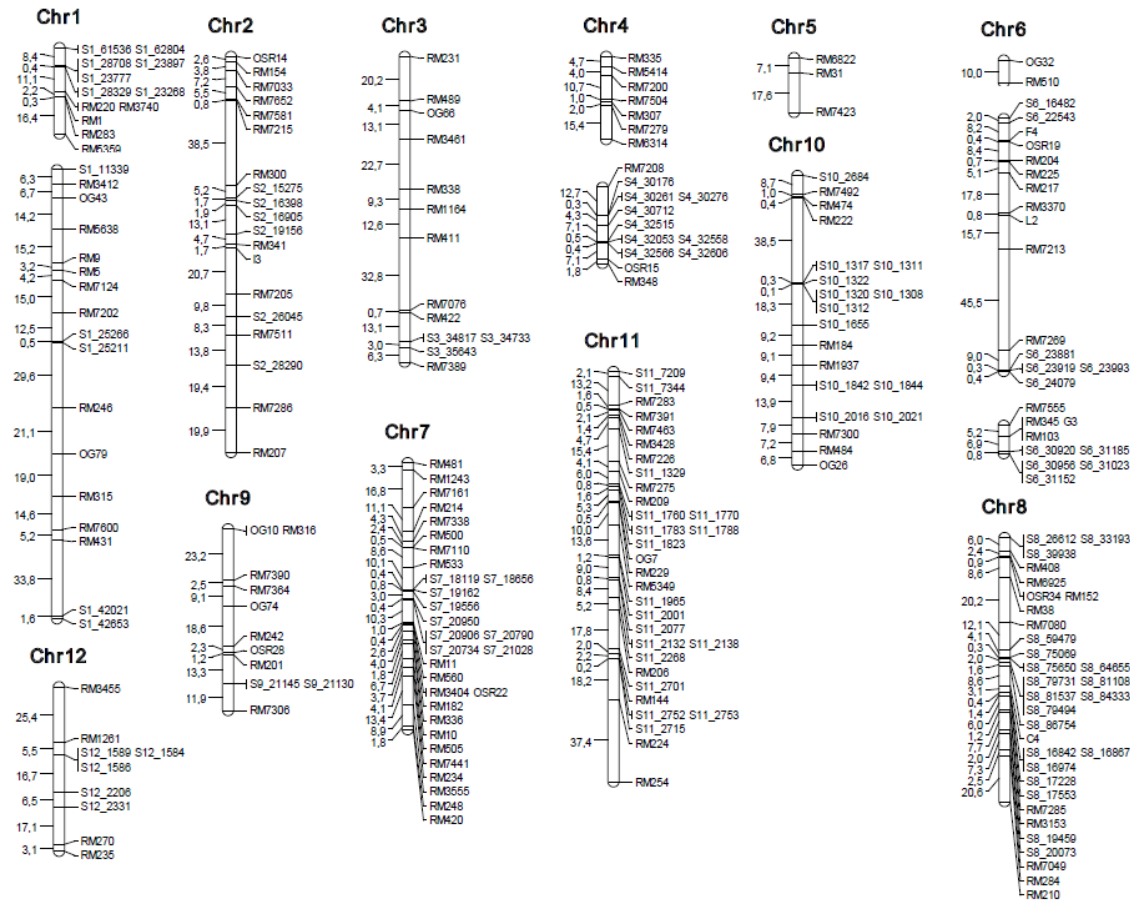


Figura 10 – Mapa genético construído com base na população RIL derivada do cruzamento entre as variedades Chorinho e Amaro. Os cromossomos, numerados de 1 a 12, são representados por barras, com a identificação dos marcadores SNP (painéis 384-plex) e microssatélites à direita e a distância em cM à esquerda. Marcadores SNP iniciam com “s” seguida do número do cromossomo ao qual pertencem, e da posição em pb no mapa físico. Foram mapeados 1440 cM do genoma de arroz.

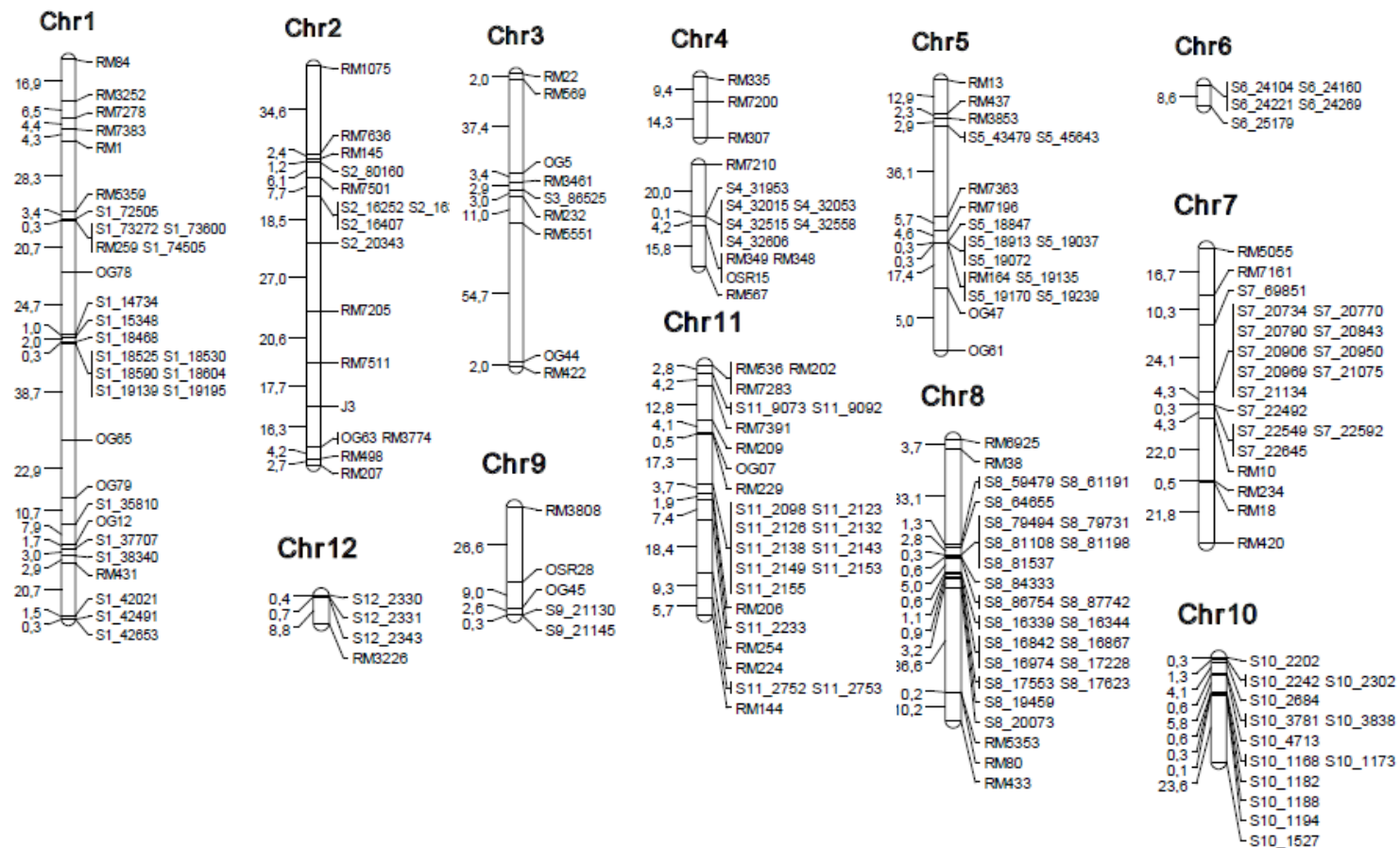


Figura 11 – Mapa genético construído com base na população RIL derivada do cruzamento entre as variedades Chorinho e Puteca. Os cromossomos, numerados de 1 a 12, são representados por barras, com a identificação dos marcadores SNP (painéis 384-plex) e microsatélites à direita e a distância em cM à esquerda. Marcadores SNP iniciam com “s” seguida do número do cromossomo ao qual pertencem, e da posição em pb no mapa físico. Foram mapeados 1035,5 cM do genoma de arroz.

Os SNPs descritos neste trabalho foram selecionados utilizando um conjunto de critérios que favoreceram a detecção de polimorfismo de DNA em acessos de arroz *japonica* tropical. As variedades que serviram de referência para a detecção dos SNPS (IAC 165, Primavera, Catetão, Azucena, Moroberekan e Ligeiro) são variedades *japonica* tropical e foram anteriormente utilizadas no desenvolvimento de populações segregantes para uso no programa de melhoramento genético. Os testes realizados comprovaram a eficiência da chamada de genótipos, a acurácia da genotipagem, o poder dos painéis na confirmação de identidade genética ou na discriminação de genótipos, assim como a segregação e herança dos alelos no mapeamento genético dos locos SNPs. O processo de genotipagem utilizando painéis SNP em genotipador BeadXpress é simples, rápido e eficiente. A tecnologia, portanto, tem muito a oferecer como sistema de genotipagem em rotina de programas de conservação de germoplasma e de melhoramento genético.

Deve ser mencionado, por fim, que o desenvolvimento de painéis multiplex de marcadores SNP para detecção de polimorfismo de DNA no genoma envolve uma série de etapas até o eventual emprego dos marcadores em genotipagem de rotina e em escala. Este processo caracteriza uma estratégia de obtenção *a priori* de marcadores SNP, que determina um grupo fixo de SNPs para emprego em genotipagem. No próximo Capítulo, outra metodologia de genotipagem de SNPs (GBS – *Genotyping by Sequencing*) é avaliada. Nessa metodologia não há necessidade de seleção *a priori* de SNPs, visto que o processo de detecção e genotipagem ocorrem em uma única etapa, possibilitando a seleção *a posteriori* dos marcadores informativos.

6. Bibliografia

- Antonius, K., Karhu, S., Kaldmäe, H., Laci, G., Rugenius, R., Baniulis, D., Sasnauskas, A., et al. (2011). Development of the Northern European Ribes core collection based on a microsatellite (SSR) marker diversity analysis. *Plant Genetic Resources*, 1-4. Retrieved from http://journals.cambridge.org/abstract_S1479262111000980
- Applied Biosystems. (2002). Taqman MGB Probes.
- Baird, N. a, Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. (J. C. Fay, Ed.) *PloS one*, 3(10), e3376. Public Library of Science. doi:10.1371/journal.pone.0003376
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. (J. C. Fay, Ed.) *PloS one*, 3(10), e3376. Public Library of Science. doi:10.1371/journal.pone.0003376
- Belaj, A., Dominguez-García, M. del C., Atienza, S. G., Martín Urdíroz, N., Rosa, R., Satovic, Z., Martín, A., et al. (2011). Developing a core collection of olive (*Olea europaea* L.) based on molecular markers (DArTs, SSRs, SNPs) and agronomic traits. *Tree Genetics & Genomes*, 1-14. Springer Berlin / Heidelberg. doi:10.1007/s11295-011-0447-6
- Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate - A practical approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological*, v.57, n.1, p.289-300, 1995.
- Bennett, S. (2004). Solexa Ltd. *Pharmacogenomics*, 5(4), 433-8. doi:10.1517/14622416.5.4.433
- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2010). GenBank. *Nucleic acids research*, 38(Database issue), D46-51. doi:10.1093/nar/gkp1024
- Bhatramakki, D. (2000). Conserved single nucleotide polymorphism (SNP) haplotypes in maize. *Proceedings of the Plant and Animal Genome Conference*.
- Boualaphanha, C., Daygon, V. D., Calingacion, M. N., Sanitchon, J., Jothityangkoon, D., Mumm, R., Hall, R. D., et al. (2011). Use of New Generation Single Nucleotide Polymorphism Genotyping for Rapid Development of Near-Isogenic Lines in Rice. *Crop Science*, 51(5), 2067. The Crop Science Society of America, Inc. doi:10.2135/cropsci2011.02.0103
- BRASIL 2011 Projeções do Agronegócio Brasileiro 2010/2011 a 2020/2021. Assessoria de Gestão Estratégica. Ministério da Agricultura, Pecuária e Abastecimento. 59 p. Brasília, DF
- Brookes, A. J. (1999). The essence of SNPs. *Gene*, 234(2), 177-86. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10395891>
- Brown, A. H. . (1995). The core collection at the crossroads. Core collections of plant genetic resources (pp. 1-19). IPGRI, Roma (Italia).

- Chen, M. (1997). Microcolinearity in sh2-homologous regions of the maize, rice, and sorghum genomes. *Proceedings of the National Academy of Sciences*, 94(7), 3431-3435. doi:10.1073/pnas.94.7.3431
- Clark, R. M., Schweikert, G., Toomajian, C., Ossowski, S., Zeller, G., Shinn, P., Warthmann, N., et al. (2007). Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science (New York, N.Y.)*, 317(5836), 338-42. doi:10.1126/science.1138632
- Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*, 142(1-2), 169-196. Springer Netherlands. doi:10.1007/s10681-005-1681-5
- Conab. (2011). Acompanhamento da Safra Brasileira - Grãos.
- Coulondre, C., Miller, J. H., Farabaugh, P. J., & Gilbert, W. (1978). Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature*, 274(5673), 775-780. doi:10.1038/274775a0
- Cronn, R., Liston, A., Parks, M., Gernandt, D. S., Shen, R., & Mockler, T. (2008). Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic acids research*, 36(19), e122. doi:10.1093/nar/gkn502
- Cruz, C. D.; Schuster, I. GQMOL - Genética Quantitativa e Molecular. 2005.
- Deulvot, C., Charrel, H., Marty, A., Jacquin, F., Donnadiou, C., Lejeune-Hénaut, I., Burstin, J., et al. (2010). Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea. *BMC genomics*, 11(1), 468. doi:10.1186/1471-2164-11-468
- Drenkard, E., Richter, B. G., Rozen, S., Stutius, L. M., Angell, N. a, Mindrinos, M., Cho, R. J., et al. (2000). A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in *Arabidopsis*. *Plant physiology*, 124(4), 1483-92. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1539302&tool=pmcentrez&rendertype=abstract>
- Eckert, A. J., Pande, B., Ersoz, E. S., Wright, M. H., Rashbrook, V. K., Nicolet, C. M., & Neale, D. B. (2008). High-throughput genotyping and mapping of single nucleotide polymorphisms in loblolly pine (*Pinus taeda* L.). *Tree Genetics & Genomes*, 5(1), 225-234. Springer Berlin / Heidelberg. doi:10.1007/s11295-008-0183-8
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. (L. Orban, Ed.) *PloS one*, 6(5), e19379. Public Library of Science. doi:10.1371/journal.pone.0019379
- Emrich, S. J., Barbazuk, W. B., Li, L., & Schnable, P. S. (2007). Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome research*, 17(1), 69-73. doi:10.1101/gr.5145806
- Fan, J.-B., Gunderson, K. L., Bibikova, M., Yeakley, J. M., Chen, J., Wickham Garcia, E., Lebruska, L. L., et al. (2006). Illumina universal bead arrays. *Methods in enzymology*, 410, 57-73. doi:10.1016/S0076-6879(06)10003-8

- Fan, J.-B., Oliphant, A., Shen, R., Kermani, B. G., GARCIA, F., Gunderson, K. L., Hansen, M., et al. (2003). Highly Parallel SNP Genotyping. *Cold Spring Harbor Symposia on Quantitative Biology*, 68(1), 69-78. doi:10.1101/sqb.2003.68.69
- FAO. (2010). *The Second Report on The State of the World's Plant Genetic Resources for Food and Agriculture* (p. 372).
- FAO. (2012). *FAO statistical yearbook 2010*. Retrieved from <http://www.fao.org/economic/ess/ess-publications/ess-yearbook/ess-yearbook2010/en/>
- Feltus, F. A., Wan, J., Schulze, S. R., Estill, J. C., Jiang, N., & Paterson, A. H. (2004). An SNP resource for rice genetics and breeding based on subspecies *indica* and *japonica* genome alignments. *Genome research*, 14(9), 1812-9. Cold Spring Harbor Laboratory Press. doi:10.1101/gr.2479404
- Ferreira, M., & Grattapaglia, D. (1998). *Introdução ao uso de marcadores moleculares em análise genética* (Documentos., p. 220p). EMBRAPA-CENARGEN.
- Fladung, M., & Buschbom, J. (2009). Identification of single nucleotide polymorphisms in different *Populus* species. *Trees*, 23(6), 1199-1212. Springer Berlin / Heidelberg. doi:10.1007/s00468-009-0359-3
- Frankel, O. H. (1984). *Genetic perspectives on germplasm conservation. Genetic manipulation: impact on man and society* (pp. 161-170). Cambridge University Press. Retrieved from <http://books.google.com/books?hl=pt-BR&lr=&id=8oWTPPpsUMQC&pgis=1>
- Gale, M. D. (1998). Comparative genetics in the grasses. *Proceedings of the National Academy of Sciences*, 95(5), 1971-1974. doi:10.1073/pnas.95.5.1971
- Ganal, M. W., Altmann, T., & Röder, M. S. (2009). SNP identification in crop plants. *Current opinion in plant biology*, 12(2), 211-7. doi:10.1016/j.pbi.2008.12.009
- Gepts, P. (1995). Genetic markers and core collections. *Core collections of plant genetic resources* (pp. 127-146). IPGRI, Roma (Itália).
- Goff, S. a, Ricke, D., Lan, T.-H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science (New York, N.Y.)*, 296(5565), 92-100. doi:10.1126/science.1068275
- Graef, D. W. G. L., & Diers, A. M. P. B. W. (2004). Identification of putative QTL that underlie yield in interspecific soybean backcross populations. *Theoretical and Applied Genetics (TAG)*, 458-467. doi:10.1007/s00122-003-1449-z
- Grauke, L. J., Thompson, T. E., & Marquard, R. D. (1995). Evaluation of pecan [*Carya illinoensis* (Wangenh.) K. Koch] germplasm collections and designation of a core subset. *Hortscience*. Retrieved from <http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-0028830037&partnerID=40&rel=R8.2.0>
- Grenier, C., Bramel-Cox, P. J., & Hamon, P. (2001). Core Collection of Sorghum: I. Stratification Based on Eco-Geographical Data. *Atomic Energy*, 41(February), 379-380. Retrieved from <http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-0035128540&partnerID=40&rel=R8.2.0>

- Han, B., & Xue, Y. (2003). Genome-wide intraspecific DNA-sequence variations in rice. *Current Opinion in Plant Biology*, 6(2), 134-138. doi:10.1016/S1369-5266(03)00004-9
- Hirochika, H. (2001). Contribution of the Tos17 retrotransposon to rice functional genomics. *Current opinion in plant biology*, 4(2), 118-22. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11228433>
- Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., Guan, J., Fan, D., Weng, Q., Huang, T., Dong, G., Sang, T., & Han, B. (2009a). High-throughput genotyping by whole-genome resequencing. *Genome research*, 19(6), 1068-76. doi:10.1101/gr.089516.108
- Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., Guan, J., Fan, D., Weng, Q., Huang, T., Dong, G., Sang, T., & Han, B. (2009b). High-throughput genotyping by whole-genome resequencing. *Genome Research*, 1068-1076. doi:10.1101/gr.089516.108.
- Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature genetics*, 42(11), 961-7. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved. doi:10.1038/ng.695
- Hyten, D. L., Choi, I.-Y., Song, Q., Specht, J. E., Carter, T. E., Shoemaker, R. C., Hwang, E.-Y., et al. (2010). A High Density Integrated Genetic Linkage Map of Soybean and the Development of a 1536 Universal Soy Linkage Panel for Quantitative Trait Locus Mapping. *Crop science*, 50(3), 960. Crop Science Society of America. doi:10.2135/cropsci2009.06.0360
- Hyten, D. L., Choi, I.-Y., Song, Q., Specht, J. E., Carter, T. E., Shoemaker, R. C., Hwang, E.-Y., et al. (2010). A High Density Integrated Genetic Linkage Map of Soybean and the Development of a 1536 Universal Soy Linkage Panel for Quantitative Trait Locus Mapping. *Crop Science*, 50(3), 960. doi:10.2135/cropsci2009.06.0360
- IBGE. (2011). LEVANTAMENTO SISTEMÁTICO DA PRODUÇÃO AGRÍCOLA.
- Jaiswal, P.; NI, J.; Yap, I.; Ware, D.; Spooner, W.; Youens-Clark, K.; Ren, L.; Liang, C.; Zhao, W.; Ratnapu, K.; Faga, B.; Canaran, P.; Fogelman, M.; Hebbard, C.; Avraham, S.; Schmidt, S.; Casstevens, T. M.; Bucker, E. S.; Stein, L.; McCouch, S. Gramene: a bird's eye view of cereal genomes. *Nucleic Acids Research*, v.34, n.Database issue, p.D717-23-D717-23, 2006.
- Jena, K. K. (2010). The species of the genus *Oryza* and transfer of useful genes from wild species into cultivated rice, *O. sativa*. *Breeding Science*, 60, 518-523. Retrieved from http://www.jstage.jst.go.jp/article/jsbbs/60/5/60_518/_article
- Kato, S., Kosaka, H., & Hara, S. (1928). On the affinity of rice varieties as shown by fertility of hybrid plants. *Bull Sci Fac Agric Kyushu Univ*, 3, 132–147. Retrieved from <http://scholar.google.com.br/scholar?q= On the affinity of rice varieties as show by fertility of hybrid plants#0>
- Khush, G. S. (1997). Origin, dispersal, cultivation and variation of rice. *Plant Molecular Biology*, 35(1), 25-34. Springer Netherlands. doi:10.1023/A:1005810616885
- Kosambi, D. D. The estimation of map distances from recombination values. *Annals of Human Genetics*, v.12, n.1, p.172-175, 1943.

- Kruglyak, L. (1997). The use of a genetic map of biallelic markers in linkage studies. *Nature genetics*, 17(1), 21-4. doi:10.1038/ng0997-21
- Kumekawa, N., Ohmido, N., Fukui, K., Ohtsubo, E., & Ohtsubo, H. (2001). A new gypsy-type retrotransposon, RIRE7: preferential insertion into the tandem repeat sequence TrsD in pericentromeric heterochromatin regions of rice chromosomes. *Molecular genetics and genomics* MGG, 265(3), 480-488. Retrieved from <http://www.mendeley.com/research/new-gypsytype-retrotransposon-rire7-preferential-insertion-tandem-repeat-sequence-trsd-pericentromeric-heterochromatin-regions-rice-chromosomes/>
- Kurata, N., Nagamura, Y., Yamamoto, K., Harushima, Y., Sue, N., Wu, J., Antonio, B. A., et al. (1994). A 300 kilobase interval genetic map of rice including 883 expressed sequences. *Nature genetics*, 8(4), 365-72. doi:10.1038/ng1294-365
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., et al. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 860-921. Macmillian Magazines Ltd. doi:10.1038/35057062
- Lander, E. S.; Green, P.; Abrahanson, J.; Barlow, A.; Daly, M. J.; Lincoln, S. E.; Newburg, L. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*, v.1, n.2, p.174-81, 1987.
- Li, Xiaobai, Yan, W., Agrama, H., Hu, B., Jia, L., Jia, M., Jackson, A., et al. (2010). Genotypic and phenotypic characterization of genetic differentiation and diversity in the USDA rice mini-core collection. *Genetica*, 138(11-12), 1221-30. Springer Netherlands. doi:10.1007/s10709-010-9521-5
- Lin, C. H., Yeakley, J. M., McDaniel, T. K., & Shen, R. (2009). Medium- to high-throughput SNP genotyping using VeraCode microbeads. *Methods in molecular biology* (Clifton, N.J.), 496, 129-42. doi:10.1007/978-1-59745-553-4_10
- Mao, L. (2000). Rice Transposable Elements: A Survey of 73,000 Sequence-Tagged-Connectors. *Genome Research*, 10(7), 982-990. doi:10.1101/gr.10.7.982
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057), 376-80. doi:10.1038/nature03959
- Maxam, A. M. (1977). A New Method for Sequencing DNA. *Proceedings of the National Academy of Sciences*, 74(2), 560-564. doi:10.1073/pnas.74.2.560
- McNally, K. L., Childs, K. L., Bohnert, R., Davidson, R. M., Zhao, K., Ulat, V. J., Zeller, G., et al. (2009). Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proceedings of the National Academy of Sciences of the United States of America*, 106(30), 12273-8. doi:10.1073/pnas.0900992106
- Mochizuki, K., Umeda, M., Ohtsubo, H., & Ohtsubo, E. (1992). Characterization of a plant SINE, p-SINE1, in rice genomes. *Idengaku zasshi*, 67(2), 155-66. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1326298>

- Mohammadi, S. A., & Prasanna, B. M. (2003). Analysis of Genetic Diversity in Crop Plants—Salient Statistical Tools and Considerations. *Crop Science*, 43(4), 1235. Crop Science Society of America. doi:10.2135/cropsci2003.1235
- Moragues, M., Comadran, J., Waugh, R., Milne, I., Flavell, A. J., & Russell, J. R. (2010). Effects of ascertainment bias and marker number on estimations of barley diversity from high-throughput SNP genotype data. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 120(8), 1525-34. doi:10.1007/s00122-010-1273-1
- Nasu, S. (2002). Search for and Analysis of Single Nucleotide Polymorphisms (SNPs) in Rice (*Oryza sativa*, *Oryza rufipogon*) and Establishment of SNP Markers. *DNA Research*, 9(5), 163-171. Oxford University Press. doi:10.1093/dnares/9.5.163
- Noma, K., Ohtsubo, E., & Ohtsubo, H. (1999). Non-LTR retrotransposons (LINEs) as ubiquitous components of plant genomes. *Molecular & general genetics : MGG*, 261(1), 71-9. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10071212>
- Novaes, E., Drost, D. R., Farmerie, W. G., Pappas, G. J., Grattapaglia, D., Sederoff, R. R., & Kirst, M. (2008). High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC genomics*, 9, 312. doi:10.1186/1471-2164-9-312
- Pereira, J. (2002). *Cultura do arroz no Brasil - subsídios para a sua história*. Embrapa Meio Norte.
- Pessoa-Filho, M. A. C. D. P. Identificação de regiões genômicas associadas ao controle de tolerância à seca em variedades tradicionais de arroz de sequeiro (*Oryza sativa* spp. *japonica*). f. 250. Doutorado (Tese de Doutorado). Instituto de Biologia - Departamento de Biologia Celular, Universidade de Brasília, Brasília, 2008.
- Pessoa-Filho, M., Rangel, P. H. N., & Ferreira, M. E. (2010). Extracting samples of high diversity from thematic collections of large gene banks using a genetic-distance based approach. *BMC plant biology*, 10(1), 127. BioMed Central Ltd. doi:10.1186/1471-2229-10-127
- Price, A. H., Townend, J., Jones, M. P., Audebert, A., & Courtois, B. (2002). Mapping QTLs associated with drought avoidance in upland rice grown in the Philippines and West Africa. *Plant molecular biology*, 48(5-6), 683-95. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11999843>
- R Tobler, A., Short, S., R Andersen, M., M Paner, T., C Briggs, J., M Lambert, S., P Wu, P., et al. (2005). The SNPlex genotyping system: a flexible and scalable platform for SNP genotyping. *Journal of biomolecular techniques JBT*, 16(4), 398-406. The Association of Biomolecular Resource Facilities. Retrieved from <http://www.mendeley.com/research/the-snp-plex-genotyping-system-a-flexible-and-scalable-platform-for-snp-genotyping-1/>
- Risch N. & Merikangas K. 1996 The future of genetic studies of complex human diseases. *Science* 273:1516-1517
- Sanchez, A. C., Subudhi, P. K., Rosenow, D. T., & Nguyen, H. T. (2002). Mapping QTLs associated with drought resistance in sorghum (*Sorghum bicolor* L. Moench). *Plant molecular biology*, 48(5-6), 713-26. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11999845>

- Sanger, F., & Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology*, 94(3), 441-8. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1100841>
- Schmidt, A. B. Desenvolvimento de Painéis Multiplex de Marcadores Microsatélites e Mapeamento de QTLs de tolerância à seca e ao frio em Linhagens Puras Recombinantes de arroz (*Oryza sativa* L.). f. 378. Doutorado em Ciências - Recursos Genéticos Vegetais (Tese de Doutorado). Centro de Ciências Agrárias, Programa de Pós-Graduação em Recursos Genéticos Vegetais, Universidade Federal de Santa Catarina, Florianópolis, 2009.
- Schoen, D. J. (1993). Conservation of Allelic Richness in Wild Crop Relatives is Aided by Assessment of Genetic Markers. *Proceedings of the National Academy of Sciences*, 90(22), 10623-10627. doi:10.1073/pnas.90.22.10623
- Shen, Y.-J., Jiang, H., Jin, J.-P., Zhang, Z.-B., Xi, B., He, Y.-Y., Wang, G., et al. (2004). Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. *Plant physiology*, 135(3), 1198-205. doi:10.1104/pp.103.038463
- Shirasawa, K., Isobe, S., Hirakawa, H., Asamizu, E., Fukuoka, H., Just, D., Rothan, C., et al. (2010). SNP discovery and linkage map construction in cultivated tomato. *DNA research : an international journal for rapid publication of reports on genes and genomes*, 17(6), 381-91. doi:10.1093/dnares/dsq024
- Tanksley, S. D., & McCouch, S. R. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. *Science (New York, N.Y.)*, 277(5329), 1063-6. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9262467>
- Tawfik, D. S., & Griffiths, A. D. (1998). Man-made cell-like compartments for molecular evolution. *Nature biotechnology*, 16(7), 652-6. Nature Publishing Company. doi:10.1038/nbt0798-652
- Tung, C.-W., Zhao, K., Wright, M. H., Ali, M. L., Jung, J., Kimball, J., Tyagi, W., et al. (2010). Development of a Research Platform for Dissecting Phenotype–Genotype Associations in Rice (*Oryza* spp.). *Rice*, 3(4), 205-217. Springer New York. doi:10.1007/s12284-010-9056-5
- Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., et al. (2008). A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome research*, 18(7), 1051-63. doi:10.1101/gr.076463.108
- Vaughan, D. A., Kadowaki, K., Kaga, A., & Tomooka, N. (2005). On the phylogeny and biogeography of the genus *Oryza*. *Breeding Science*, v. 55(2) p. Retrieved from <http://agris.fao.org/agris-search/search/display.do?f=2006/JP/JP0516.xml;JP2006001202>
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., et al. (2001). The sequence of the human genome. *Science (New York, N.Y.)*, 291(5507), 1304-51. doi:10.1126/science.1058040

- Vignal, A., Milan, D., SanCristobal, M., & Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics, selection, evolution* : GSE, 34(3), 275-305. doi:10.1051/gse:2002009
- Voorrips, R. E. MapChart: software for the graphical presentation of linkage maps and QTLs. *The Journal of heredity*, v.93, n.1, p.77-78, 2002.
- Wang, M. L., Sukumaran, S., Barkley, N. A., Chen, Z., Chen, C. Y., Guo, B., Pittman, R. N., et al. (2011). Population structure and marker-trait association analysis of the US peanut (*Arachis hypogaea* L.) mini-core collection. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 123(8), 1307-1317. Springer Berlin / Heidelberg. doi:10.1007/s00122-011-1668-7
- Wicker, T., Narechania, A., Sabot, F., Stein, J., Vu, G. T. H., Graner, A., Ware, D., et al. (2008). Low-pass shotgun sequencing of the barley genome facilitates rapid identification of genes, conserved non-coding sequences and novel repeats. *BMC genomics*, 9(1), 518. BioMed Central Ltd. doi:10.1186/1471-2164-9-518
- Xu, Y., McCouch, S. R., & Zhang, Q. (2005). How can we use genomics to improve cereals with rice as a reference genome? *Plant molecular biology*, 59(1), 7-26. Springer Netherlands. doi:10.1007/s11103-004-4681-2
- Yan, J., Yang, X., Shah, T., Sánchez-Villeda, H., Li, J., Warburton, M., Zhou, Y., et al. (2010). High-throughput SNP genotyping with the GoldenGate assay in maize. *Molecular Breeding*, 25(3), 441-451. Springer Netherlands. doi:10.1007/s11032-009-9343-2
- Yu, H., Xie, W., Wang, J., Xing, Y., Xu, C., Li, X., Xiao, J., et al. (2011). Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. *PloS one*, 6(3), e17595. doi:10.1371/journal.pone.0017595
- Yu, Jun, Hu, S., Wang, J., Wong, G. K.-S., Li, S., Liu, B., Deng, Y., et al. (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science (New York, N.Y.)*, 296(5565), 79-92. doi:10.1126/science.1068037
- Zhang, H., Zhang, D., Wang, M., Sun, J., Qi, Y., Li, J., Wei, X., et al. (2011). A core collection and mini core collection of *Oryza sativa* L. in China. *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, 122(1), 49-61. Springer Berlin / Heidelberg. doi:10.1007/s00122-010-1421-7
- Zhao, K., Tung, C.-W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., Norton, G. J., et al. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nature communications*, 2, 467. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved. doi:10.1038/ncomms1467

CAPÍTULO 2

Avaliação da metodologia de Genotipagem por Sequenciamento (*Genotyping by Sequencing – GBS*) e seu potencial de uso em programa de melhoramento genético de arroz de sequeiro (*Oryza sativa* L. spp. *japonica*)

CAPÍTULO 2

Avaliação da metodologia de Genotipagem por Sequenciamento (*Genotyping by Sequencing – GBS*) e seu potencial de uso em programa de melhoramento genético de arroz de sequeiro (*Oryza sativa L. spp. japonica*)

1. Resumo

Com o recente avanço nas tecnologias de sequenciamento de alto desempenho, os custos de sequenciamento de DNA caíram vertiginosamente quando comparados à década passada. A aplicação de sequenciamento para genotipagem em larga escala de diferentes indivíduos de uma mesma espécie torna-se agora mais atraente. O emprego de metodologias de sequenciamento de nova geração (NGS – *Next Generation Sequencing*) visa acelerar o uso de informação genômica no melhoramento genético, facilitando o mapeamento genético de características de importância econômica e adaptando a análise de polimorfismo de DNA às rotinas do programa. O primeiro passo para inserir esse tipo de informação em programas de melhoramento genético é testar e adaptar as metodologias recém-desenvolvidas para aplicação no organismo de escolha. As novas metodologias de genotipagem por sequenciamento baseiam-se, em sua grande maioria, na redução de complexidade do genoma através do uso de enzimas de restrição. É possível amostrar regiões amplamente distribuídas ao longo do genoma utilizando, por exemplo, enzimas sensíveis a metilação, evitando, dessa forma, o sequenciamento massal de regiões repetitivas e complexas, que dificultam o processo de análise. Essa é uma estratégia considerada reproduzível, específica e rápida. Neste trabalho foi testada em arroz de sequeiro a metodologia proposta por Elshire (2011), denominada Genotipagem por Sequenciamento (*GBS – Genotyping by Sequencing*). Os desenhos dos adaptadores (com código de barra ou comum), digestões, ligações, construção dos *pools* de amostra, amplificação por PCR e reações de purificação de fragmentos foram feitos seguindo o protocolo sugerido pelos proponentes da metodologia. Para o teste, foram selecionados acessos de arroz de sequeiro com alta diversidade genética, que compõem a Coleção Nuclear Temática de Tolerância a Seca de arroz. Os resultados obtidos neste trabalho comprovam que a metodologia de genotipagem por sequenciamento tem amplo potencial na detecção e seleção de um alto número de marcadores SNPs. Contudo, verificou-se um aparente excesso de genótipos heterozigotos nas amostras avaliadas. Além disso, em testes de prova e contraprova, o erro de genotipagem estimado foi elevado em algumas comparações, independentemente da cobertura genômica mínima utilizada na detecção dos SNPs ou do valor de qualidade mínima do SNP detectado. O ensaio GBS proporciona, portanto, um alto número de marcadores SNPs capazes de detectar polimorfismo de DNA em

acessos de arroz, mas a acurácia da genotipagem foi considerada baixa nas comparações de prova e contraprova realizadas. É necessário um aprofundamento em todas as etapas do ensaio GBS para a seleção de marcadores SNP fidedignos, passíveis de uso em escala na genotipagem de arroz.

2. Introdução

O emprego em escala de modelos preditivos de fenótipo com base no polimorfismo de DNA em programas de melhoramento de plantas depende da descoberta e genotipagem de um alto número de marcadores moleculares de forma rápida e acessível. O marcador de escolha, no momento, para atingir este objetivo é o polimorfismo de base única – SNP, por serem abundantes, amplamente distribuídos ao longo do genoma, com genotipagem passível de automação. Para organismos modelo existem diferentes projetos de descoberta e validação em escala de SNPs, como em *Arabidopsis thaliana* (<http://walnut.usc.edu/2010>), arroz (<http://irfgc.irri.org>) e milho (<http://www.panzea.org/>).

As plataformas de sequenciamento em alta escala têm permitido um rápido desenvolvimento e avanço de tecnologias de genotipagem de genomas inteiros. Até recentemente existia uma lacuna de metodologias que fossem eficientes tanto do ponto de vista de custo quanto de tempo, para experimentos que demandassem uma alta densidade de marcadores para a genotipagem de um grande número de amostras. As metodologias disponíveis atualmente são baseadas na seleção de um conjunto de marcadores SNP para construção de painéis ou arranjos que permitem a detecção de genótipos em centenas (RT-PCR, Sequenom MassARRAY, Illumina GoldenGate) ou até dezenas de milhares de marcadores SNP (Affymetrix, Agilent), porém a um custo ainda elevado para uso em rotina em programas de melhoramento genético.

Recentemente, novas metodologias de genotipagem de SNP baseadas em sequenciamento de nova geração (NGS) foram propostas, combinando a descoberta de polimorfismo e genotipagem em um único passo (Baird et al., 2008; Huang et al., 2009; Elshire et al., 2011). Isto significa que, ao invés de definir *a priori* quais são os SNPs que serão genotipados ao longo do genoma (veja Capítulo 1), essas novas metodologias primeiro detectam e genotipam os SNPs, para uma seleção *a posteriori* dos SNPs que serão usados na análise com base em um conjunto de critérios. As metodologias citadas (Baird et al., 2008; Huang et al., 2009; Elshire et al., 2011) geram uma alta quantidade de marcadores, amplamente distribuídos ao longo do genoma, além de poderem ser aplicados em *pools* ou grupos de amostras. Outra vantagem dessas metodologias consiste na independência de genoma de referência, fato que contribui enormemente para o melhoramento genético de organismos não-modelo, que carecem de informação genômica.

Para a identificação de marcadores com qualidade é desejável que se tenha um sequenciamento de alta cobertura em diferentes amostras. Para se alcançar esse aumento de cobertura sem aumentar significativamente o custo com sequenciamento, o número de regiões do genoma a serem sequenciadas pode ser minimizado através da redução de complexidade do genoma. Essa redução de complexidade pode ser feita pelo uso de enzimas de restrição na construção da biblioteca de sequenciamento. A redução de complexidade de genoma pela construção de bibliotecas de representação reduzida (RRL – *Reduced Representation Libraries*) é tema recorrente na descoberta de SNPs em alta escala em diversos organismos (Barbazuk et al., 2005; Wiedmann et al., 2008).

A redução de complexidade possibilita que um conjunto menor de fragmentos do genoma seja sequenciado, o que permite não apenas aumentar a cobertura de sequenciamento nessas regiões, mas também sequenciar simultaneamente múltiplas amostras de DNA de diferentes acessos, através da estratégia de sequenciamento multiplexado (Craig et al., 2008; Cronn et al., 2008). Com isso é possível processar milhares de locos em dezenas ou até centenas de amostras simultaneamente (Baird et al., 2008; Huang et al., 2009; Elshire et al., 2011)

A metodologia de GBS (Elshire et al., 2011) envolve uma etapa de redução de complexidade do genoma através da digestão por enzima de restrição com sítio frequente no genoma alvo. A metodologia GBS permite a detecção de centenas de milhares de SNPs no conjunto de genótipos analisado. A seleção dos SNPs úteis é feita seguindo vários critérios, o que leva a uma redução significativa do número de SNPs considerados para análise. Mesmo assim, ao final do procedimento obtém-se uma quantidade significativa de marcadores SNP (milhares), com grau de polimorfismo elevado.

Apesar de ser um método eficiente do ponto de vista da geração de um alto número de dados em um curto intervalo de tempo e de baixo custo e alto benefício por amostra genotipada, pouco se sabe sobre a qualidade e a confiabilidade dos genótipos gerados. Apesar de alguns estudos apresentarem resultados animadores quanto aos genótipos gerados (Baird et al., 2008; Elshire et al., 2011; Huang et al., 2009; Myles et al., 2010), os autores não são claros quanto aos parâmetros de qualidade utilizados na seleção dos marcadores. Por exemplo, são precárias as informações sobre o valor de cobertura mínimo a ser considerado por *loco* por amostra para se poder obter uma genotipagem confiável dada a taxa de erro de sequenciamento, ou ainda o valor de qualidade mínimo do marcador SNP selecionado na atribuição do genótipo. A mensuração do erro de genotipagem, utilizando réplicas de DNA, também não é clara nesses estudos. Esse tipo de informação é vital para o uso em rotina da metodologia em programas de melhoramento genético, pois dizem respeito à reprodutibilidade e à confiabilidade dos dados, o que pode afetar diretamente a tomada de decisão nos programas.

Neste trabalho, dados de genotipagem de marcadores SNP em 24 acessos de arroz de sequeiro (*japonica* tropical) foram obtidos através de metodologia de GBS. Os dados obtidos foram analisados e avaliados quanto ao número, polimorfismo e acurácia

(reprodutibilidade) dos marcadores SNPs para emprego em estudos genéticos de arroz de sequeiro.

3. Objetivos

- Analisar dados de genotipagem de vinte e quatro variedades de arroz de sequeiro obtidos através de metodologia GBS (“*Genotyping by Sequencing*”);
- Avaliar a eficiência de detecção de polimorfismo do método GBS em arroz de sequeiro;
- Avaliar a acurácia dos genótipos detectados em locos de marcadores SNPs;
- Comparar a eficiência na utilização dos marcadores gerados por GBS com os marcadores validados em plataforma Illumina BeadXpress (Capítulo 1).

4. Material e Métodos

4.1. Metodologia do ensaio GBS em amostras de arroz *japonica* tropical

O DNA genômico foi extraído a partir de plântulas recém germinadas em vaso com vermiculita, por um método rápido com CTAB, conforme descrito em (Ferreira & Grattapaglia, 1998). A qualidade do DNA foi avaliada por eletroforese em gel de agarose 1%. A quantificação foi feita em fluorômetro NanoDrop 3300 (Thermo Scientific), utilizando PicoGreen como corante. O DNA foi diluído em tampão TE para uma concentração final de 50 ng/μL.

Para o teste e validação da metodologia de GBS em arroz de sequeiro, foram usados vinte e quatro acessos do Banco de Germoplasma de Arroz. Estes acessos foram selecionados de uma coleção de 700 acessos de variedades tradicionais brasileiras de arroz de sequeiro (*Oryza sativa* spp. *japonica*) com base na genotipagem com marcadores microssatélites (Pessoa Filho, 2008). O emprego de três algoritmos para definição de coleções nucleares (Corex, MStrat, PowerCore) possibilitou a seleção de um grupo de linhagens representativas da diversidade genética de acessos de arroz *japonica* tropical (Pessoa Filho et al., 2010). Os 24 acessos selecionados neste trabalho são comuns nos três métodos de obtenção de coleção nuclear e, portanto, representam um grupo com alta diversidade genética de arroz de sequeiro.

4.2. Sequenciamento de alta performance no sequenciador GAII/Illumina utilizando protocolo GBS

Todos os passos utilizado neste trabalho para a construção da biblioteca de sequenciamento de GBS tomou como base o protocolo desenvolvido por Elshire (2011). Os desenhos dos adaptadores (com código de barra ou comum), digestões, ligações, construção

do *pools* de amostra, amplificação por PCR e reações de purificação de fragmentos foram feitos seguindo o protocolo descrito por Elshire (2011).

O protocolo, em síntese, consiste de uma redução da complexidade do genoma através de uma digestão com enzima de restrição, seguida da ligação dos fragmentos aos adaptadores, montagem do *pool* de amostras, PCR com iniciadores complementares aos adaptadores, seleção por tamanho dos fragmentos em gel de agarose e extração e purificação dos fragmentos.

A redução de complexidade do genoma foi feita com reação enzimática utilizando a enzima *ApeKI*, seguida da ligação dos fragmentos de DNA aos adaptadores comuns e com código de barras e, posteriormente, da montagem do *pool* com todas as amostras analisadas. Com os fragmentos ligados e em *pool*, foi feita uma PCR (*Polymerase Chain Reaction*) com uma DNA polimerase de alta fidelidade, visando a amplificação seletiva do DNA genômico. Após a PCR, foi feita uma etapa de seleção dos fragmentos por eletroforese horizontal em gel de agarose (Agarose Metaphor 2% em tampão SB) para a seleção de fragmentos de DNA com tamanho específico (redução genômica). Neste trabalho foram selecionados fragmentos com tamanho entre 200 e 400 pb, tendo em vista a limitação técnica do sequenciador onde fragmentos maiores podem diminuir a eficiência e a qualidade do sequenciamento. Após a extração dos fragmentos 200-400 pb do gel e posterior purificação, a biblioteca ficou pronta para o sequenciamento NGS.

O sequenciamento da biblioteca construída foi feito em equipamento GA IIx (Illumina), pela plataforma Genômica-DF, seguindo o protocolo padrão de sequenciamento fornecido pelo fabricante para o kit *TruSeq v3 Technology®* para sequências de terminação simples (*single end*) de 76 ciclos de amplificação. Foi feito o sequenciamento em duplicata (REP 1 e REP 2) da biblioteca para permitir uma comparação dos genótipos entre as diferentes repetições.

4.3. Detecção, validação e avaliação de performance de marcadores SNPs em ensaio GBS

Como o sequenciamento foi conduzido em equipamento Genome Analyser IIx, o sequenciador não gera arquivos de sequência que podem ser imediatamente utilizados, mas sim arquivos de “probabilidade e posição”. Os arquivos de probabilidade consistem em arquivos que contém a informação de probabilidade de uma base na sequência de DNA ser uma das quatro bases nitrogenadas possíveis, ou seja, A, C, G ou T, tendo em vista a detecção da incorporação ou não de base a cada ciclo de sequenciamento. Os arquivos de posição fornecem os dados de coordenada dentro de cada canaleta de sequenciamento que permitem a localização da base incorporada. As informações dos arquivos de probabilidade e o de posição permitem identificar qual base foi incorporada em um determinado “cluster” a cada ciclo de sequenciamento. Portanto, os arquivos gerados pelo equipamento (extensão *bcf*) foram transformados para “*qseq*” – formato legível em qualquer plataforma de análise. Essa

transformação foi feita utilizando o software BCL Converter v. 1.7, com configurações padrão para o equipamento utilizado e sem nenhum filtro de qualidade. Toda a análise subsequente foi feita com os arquivos de sequência *qseq* não-filtrados.

De posse dos arquivos com os genótipos, iniciou-se uma etapa de filtragem dos dados brutos de acordo com os dados de qualidade da chamada das bases em cada loco SNP. Todo o processamento de filtragem foi feito se utilizando um script em *perl*. Foram retiradas, inicialmente: (a) as sequências que continham “N”, ou seja, sequências para as quais não foi atribuir nenhuma base de maneira confiável em pelo menos uma posição da sequência de 76 pb; (b) as sequências sem a construção correta esperada – código de barra da amostra seguido de sítio de restrição da enzima. Em seguida, as sequências contendo bases com valor de qualidade iguais a “B” foram truncadas a partir da primeira base com essa qualidade. Finalmente, todas as sequências foram truncadas a 64 pares de base, eliminando extremidades com menor qualidade. Para se atribuir uma sequência a determinada amostra, foi realizado um passo de demultiplexagem, isto é, identificação das origem amostral da sequência com base no código atribuído a cada uma, utilizando o software *Novobarcodes* (Novocraft Inc.), com as configurações padrão sugeridas pelo *software*.

As sequências originárias de GBS, após a filtragem de qualidade e alocação ao acesso de origem, foram mapeadas em relação ao genoma de referência de arroz cv. Nipponbare (MSU v. 6.1) utilizando o software *Novoalign* (Novocraft). Neste passo, utilizou-se os parâmetros-padrão de *mismatches* e *gaps* recomendados pelo *software*. Uma vez mapeado no genoma referência, os arquivos de mapeamento foram processados usando o *software* GATK (McKenna et al., 2010). Este programa possibilita uma re-calibração das bases baseada nos valores de qualidade obtidos no sequenciamento Illumina GAI, o realinhamento local de inserções e deleções, a descoberta e genotipagem de SNPs, possibilitando a comparação entre todas as amostras testadas simultaneamente utilizando os parâmetros de filtragem e recalibração padrão (DePristo et al., 2011). Depois de descobertos e genotipados, todos os passos seguintes de filtragem e seleção dos marcadores foram feitos utilizando o *software* *Vcftools* (Danecek et al., 2011).

5. Resultados e Discussão

Foram selecionados 24 acessos de arroz para análise com marcadores SNPs através da metodologia GBS. Estes acessos foram selecionados de uma coleção de 700 acessos de variedades tradicionais brasileiras de arroz de sequeiro (*Oryza sativa* spp. *japonica*) com base na genotipagem com marcadores microssatélites (Pessoa Filho, 2008). Os 24 acessos selecionados neste trabalho representam um grupo com alta diversidade genética de arroz de sequeiro. Espera-se que o polimorfismo de DNA nos locos SNP desta amostra seja elevado. Dois acessos (CNS 10 e CNS 49) tiveram o DNA extraído em duplicata (prova e contra-prova). Um terceiro acesso (CNS 51) teve o DNA extraído em triplicata, também para testes de prova e

contraprova. O objetivo destas extrações independentes de DNA de um mesmo genótipo foi avaliar o potencial erro de genotipagem (acurácia) entre amostras extraídas de um mesmo acesso.

O rendimento em sequência de DNA (*reads*) do ensaio GBS indicou que dois acessos (CNS6 e CNS17) apresentaram rendimento de sequenciamento limitado, com um número final de sequências (*reads*) muito abaixo do esperado (Tabela 1). É possível que isto seja resultante de deficiências no preparo inicial da amostra de DNA ou de algum passo no processo de ligação de adaptadores na construção das bibliotecas. Portanto, foram considerados para análise de dados gerados por GBS os 22 diferentes acessos de arroz de sequeiro. E os acessos CNS 10, CNS 49 e CNS 51 tiveram repetições de contraprova dentro de cada repetição do ensaio.

O experimento foi desenhado com duas repetições. A repetição 1 ou REP1 resultou na obtenção de 34.061.047 fragmentos (*reads*) de DNA, totalizando 2.588.639.572 pb sequenciados (Tabela 1). O segundo ensaio (repetição 2 ou REP2) resultou na obtenção de 33.981.093 fragmentos (*reads*) de DNA, totalizando 2.582.563.068 pb sequenciados.

Tabela 1. Número total de sequências (*reads*) obtido para cada um dos 24 acessos de arroz de sequeiro em duas repetições (REP1 e REP2) de ensaio GBS.

Acesso	REP1	REP2
CNS6	6.153	5.697
CNS7	831.916	835.524
CNS9	528.880	522.416
CNS10	1.459.753	1.475.834
CNS10A	406.968	414.509
CNS14	1.351.184	1.323.416
CNS17	529.877	540.375
CNS29	468.765	448.545
CNS37	473.075	476.140
CNS41	853.015	860.248
CNS44	1.079.359	1.092.403
CNS46	667.908	648.516
CNS49	658.143	639.708
CNS49A	2.524.295	2.515.548
CNS51	5.660.877	5.656.426
CNS53	1.686.628	1.717.227
CNS51A	2.026.559	1.984.970
CNS51B	1.143.647	1.114.006
CNS58	1.833.621	1.878.794
CNS62	1.951.629	1.911.254
CNS63	1.606.203	1.579.365

CNS69	1.588.894	1.580.938
CNS77	2.122.043	2.171.700
CNS87	942.238	939.586
CNS88	612.738	622.116
CNS90	1.046.679	1.025.832
TOTAL	34.061.047	33.981.093

Para a detecção de SNPs, procurou-se definir inicialmente a cobertura genômica mínima em um sítio de DNA para identificar um SNP verdadeiro, que não fosse resultante de erros de sequenciamento, procurando minimizar a possibilidade de identificação de SNPs espúrios. Para isto, foi analisado o efeito da cobertura mínima de cada sítio na quantidade de SNPs detectados, variando a cobertura de 4x a 15x (coberturas de 4x, 7x, 11x, 15) nos milhares de sítios SNP. O aumento de cobertura em um sítio SNP torna a detecção mais conservadora, diminuindo a possibilidade de identificação de falsos SNPs. O objetivo desta análise foi identificar uma faixa de cobertura adequada e segura para a detecção de SNPs em arroz de sequeiro. Os dados iniciais para avaliação do efeito da cobertura são computados na Tabela 2.

Tabela 2. Número de SNPs observados em 24 acessos de arroz sequeiro submetidos a ensaio GBS. Os SNPs foram detectados com variações crescentes de cobertura mínima (4x, 7x, 11x e 15x) em cada sítio de DNA. Onde: AA – genótipo homocigoto para o primeiro alelo detectado no loco SNP; BB – genótipo homocigoto para o segundo alelo detectado no loco SNP; AB – genótipo heterocigoto no loco SNP.

Acesso	Repetição	Cobertura	Falhas	AA	AB	BB	TotalSNPs	Tx.Homozig.
CNS10	Rep1	04x	25.495	44.650	6.158	11.961	62.769	90,19%
	Rep2	04x	24.554	45.098	6.666	11.947	63.711	89,54%
	Rep1	07x	14.093	17.849	2.803	4.406	25.058	88,81%
	Rep2	07x	13.696	18.196	3.082	4.265	25.543	87,93%
	Rep1	11x	5.742	5.273	1.158	1.241	7.672	84,91%
	Rep2	11x	5.518	5.555	1.219	1.244	8.018	84,80%
	Rep1	15x	2.415	1.608	504	411	2.523	80,02%
	Rep2	15x	2.287	1.703	542	1.703	3.948	86,27%
CNS10A	Rep1	04x	76.124	8.637	1.406	2.097	12.140	88,42%
	Rep2	04x	75.535	9.049	1.477	2.204	12.730	88,40%
	Rep1	07x	37.267	1.241	338	305	1.884	82,06%
	Rep2	07x	37.331	1.266	342	300	1.908	82,08%
	Rep1	11x	13.218	90	53	53	196	72,96%
	Rep2	11x	13.316	117	56	47	220	74,55%
	Rep1	15x	4.888	18	15	17	50	70,00%
	Rep2	15x	4.881	19	18	19	56	67,86%
CNS14	Rep1	04x	31.132	40.010	5.790	11.332	57.132	89,87%
	Rep2	04x	32.644	38.712	6.044	10.865	55.621	89,13%
	Rep1	07x	16.697	15.733	2.667	4.054	22.454	88,12%
	Rep2	07x	17.557	15.150	2.611	3.921	21.682	87,96%
	Rep1	11x	6.501	4.738	980	1.195	6.913	85,82%
	Rep2	11x	6.721	4.667	985	1.163	6.815	85,55%
	Rep1	15x	2.477	1.552	454	455	2.461	81,55%

	Rep2	15x	2.603	1.490	440	1.490	3.420	87,13%
CNS29	Rep1	04x	73.697	9.999	2.075	2.493	14.567	85,76%
	Rep2	04x	74.501	9.523	1.761	2.480	13.764	87,21%
	Rep1	07x	36.679	1.555	501	416	2.472	79,73%
	Rep2	07x	37.069	1.369	412	389	2.170	81,01%
	Rep1	11x	13.167	114	67	66	247	72,87%
	Rep2	11x	13.302	129	57	48	234	75,64%
	Rep1	15x	4.885	18	13	22	53	75,47%
	Rep2	15x	4.876	24	19	24	67	71,64%
CNS37	Rep1	04x	75.922	10.567	1.093	682	12.342	91,14%
	Rep2	04x	75.458	10.936	1.236	635	12.807	90,35%
	Rep1	07x	37.492	1.381	166	112	1.659	89,99%
	Rep2	07x	37.484	1.425	194	136	1.755	88,95%
	Rep1	11x	13.273	92	19	30	141	86,52%
	Rep2	11x	13.367	103	25	41	169	85,21%
	Rep1	15x	4.906	9	6	17	32	81,25%
	Rep2	15x	4.906	9	6	9	24	75,00%
CNS41	Rep1	04x	52.614	24.143	4.317	7.190	35.650	87,89%
	Rep2	04x	52.710	23.912	4.542	7.101	35.555	87,23%
	Rep1	07x	28.452	7.188	1.552	1.959	10.699	85,49%
	Rep2	07x	28.558	7.151	1.600	1.930	10.681	85,02%
	Rep1	11x	10.974	1.534	475	431	2.440	80,53%
	Rep2	11x	10.999	1.565	515	457	2.537	79,70%
	Rep1	15x	4.261	367	198	112	677	70,75%
	Rep2	15x	4.256	348	210	348	906	76,82%
CNS44	Rep1	04x	39.488	34.143	4.911	9.722	48.776	89,93%
	Rep2	04x	39.144	34.316	5.262	9.543	49.121	89,29%
	Rep1	07x	22.136	11.835	2.108	3.072	17.015	87,61%
	Rep2	07x	21.969	11.977	2.202	3.091	17.270	87,25%
	Rep1	11x	8.976	2.933	735	770	4.438	83,44%
	Rep2	11x	8.903	3.066	772	795	4.633	83,34%
	Rep1	15x	3.566	780	332	260	1.372	75,80%
	Rep2	15x	3.538	852	318	852	2.022	84,27%
CNS46	Rep1	04x	63.923	9.343	9.478	5.520	24.341	61,06%
	Rep2	04x	64.387	9.352	9.298	5.228	23.878	61,06%
	Rep1	07x	32.753	2.201	2.991	1.206	6.398	53,25%
	Rep2	07x	33.168	2.108	2.768	1.195	6.071	54,41%
	Rep1	11x	12.166	388	664	196	1.248	46,79%
	Rep2	11x	12.408	377	558	193	1.128	50,53%
	Rep1	15x	4.641	85	153	59	297	48,48%
	Rep2	15x	4.665	105	114	105	324	64,81%
CNS49	Rep1	04x	63.486	9.549	10.245	4.984	24.778	58,65%
	Rep2	04x	63.931	9.528	10.083	4.723	24.334	58,56%
	Rep1	07x	32.590	2.294	2.960	1.307	6.561	54,88%
	Rep2	07x	33.077	2.180	2.817	1.165	6.162	54,28%
	Rep1	11x	12.125	412	670	207	1.289	48,02%
	Rep2	11x	12.362	401	581	192	1.174	50,51%
	Rep1	15x	4.633	89	154	62	305	49,51%
	Rep2	15x	4.663	104	119	104	327	63,61%
CNS49A	Rep1	04x	6.878	56.864	8.793	15.729	81.386	89,20%
	Rep2	04x	7.062	56.437	9.437	15.329	81.203	88,38%
	Rep1	07x	2.457	26.052	4.235	6.407	36.694	88,46%
	Rep2	07x	2.385	26.076	4.474	6.304	36.854	87,86%

	Rep1	11x	607	8.954	1.792	2.061	12.807	86,01%
	Rep2	11x	569	9.083	1.833	2.051	12.967	85,86%
	Rep1	15x	199	3.154	806	779	4.739	82,99%
	Rep2	15x	171	3.191	825	3.191	7.207	88,55%
CNS51	Rep1	04x	2.185	60.723	8.605	16.751	86.079	90,00%
	Rep2	04x	2.090	60.375	9.355	16.445	86.175	89,14%
	Rep1	07x	599	27.850	3.909	6.793	38.552	89,86%
	Rep2	07x	572	27.842	4.122	6.703	38.667	89,34%
	Rep1	11x	172	9.420	1.632	2.190	13.242	87,68%
	Rep2	11x	184	9.473	1.706	2.173	13.352	87,22%
	Rep1	15x	75	3.318	741	804	4.863	84,76%
	Rep2	15x	71	3.327	754	3.327	7.408	89,82%
CNS53	Rep1	04x	19.204	48.373	7.054	13.633	69.060	89,79%
	Rep2	04x	18.374	48.815	7.713	13.363	69.891	88,96%
	Rep1	07x	9.612	20.887	3.393	5.259	29.539	88,51%
	Rep2	07x	9.278	21.104	3.567	5.290	29.961	88,09%
	Rep1	11x	3.583	6.760	1.430	1.641	9.831	85,45%
	Rep2	11x	3.456	6.995	1.428	1.657	10.080	85,83%
	Rep1	15x	1.401	2.295	642	600	3.537	81,85%
	Rep2	15x	1.314	2.390	621	2.390	5.401	88,50%
CNS51A	Rep1	04x	12.890	53.596	7.162	14.616	75.374	90,50%
	Rep2	04x	13.258	53.007	7.673	14.327	75.007	89,77%
	Rep1	07x	5.058	24.625	3.429	6.039	34.093	89,94%
	Rep2	07x	5.340	24.306	3.644	5.949	33.899	89,25%
	Rep1	11x	1.478	8.504	1.429	2.003	11.936	88,03%
	Rep2	11x	1.655	8.404	1.503	1.974	11.881	87,35%
	Rep1	15x	525	2.991	682	740	4.413	84,55%
	Rep2	15x	518	3.013	689	3.013	6.715	89,74%
CNS51B	Rep1	04x	36.003	37.058	5.303	9.900	52.261	89,85%
	Rep2	04x	37.200	35.978	5.550	9.537	51.065	89,13%
	Rep1	07x	19.800	13.676	2.300	3.375	19.351	88,11%
	Rep2	07x	20.505	13.241	2.296	3.197	18.734	87,74%
	Rep1	11x	8.035	3.704	783	892	5.379	85,44%
	Rep2	11x	8.396	3.511	745	884	5.140	85,51%
	Rep1	15x	3.399	937	325	277	1.539	78,88%
	Rep2	15x	3.492	877	284	877	2.038	86,06%
CNS58	Rep1	04x	21.825	57.339	5.332	3.768	66.439	91,97%
	Rep2	04x	20.541	57.989	5.970	3.765	67.724	91,18%
	Rep1	07x	10.602	25.171	1.911	1.467	28.549	93,31%
	Rep2	07x	9.846	25.740	2.185	1.468	29.393	92,57%
	Rep1	11x	3.996	8.350	596	472	9.418	93,67%
	Rep2	11x	3.589	8.786	700	461	9.947	92,96%
	Rep1	15x	1.673	2.854	202	209	3.265	93,81%
	Rep2	15x	1.533	2.964	244	2.964	6.172	96,05%
CNS62	Rep1	04x	26.352	25.449	12.374	24.089	61.912	80,01%
	Rep2	04x	26.779	25.158	12.728	23.600	61.486	79,30%
	Rep1	07x	12.552	10.782	3.970	11.847	26.599	85,07%
	Rep2	07x	12.791	10.803	3.982	11.663	26.448	84,94%
	Rep1	11x	4.448	3.512	1.463	3.991	8.966	83,68%
	Rep2	11x	4.638	3.503	1.476	3.919	8.898	83,41%
	Rep1	15x	1.692	1.199	617	1.430	3.246	80,99%
	Rep2	15x	1.679	1.224	642	1.224	3.090	79,22%
CNS63	Rep1	04x	25.155	43.764	6.824	12.521	63.109	89,19%

	Rep2	04x	26.160	42.834	7.282	11.989	62.105	88,27%
	Rep1	07x	12.250	18.929	3.059	4.913	26.901	88,63%
	Rep2	07x	12.872	18.332	3.274	4.761	26.367	87,58%
	Rep1	11x	4.387	6.247	1.196	1.584	9.027	86,75%
	Rep2	11x	4.685	6.092	1.193	1.566	8.851	86,52%
	Rep1	15x	1.761	2.088	519	570	3.177	83,66%
	Rep2	15x	1.817	2.033	499	2.033	4.565	89,07%
CNS69	Rep1	04x	22.324	45.118	8.716	12.106	65.940	86,78%
	Rep2	04x	22.511	44.703	9.110	11.941	65.754	86,15%
	Rep1	07x	11.722	18.901	4.054	4.474	27.429	85,22%
	Rep2	07x	11.686	18.829	4.201	4.523	27.553	84,75%
	Rep1	11x	4.422	6.047	1.547	1.398	8.992	82,80%
	Rep2	11x	4.545	6.078	1.550	1.363	8.991	82,76%
	Rep1	15x	1.769	1.996	651	522	3.169	79,46%
	Rep2	15x	1.854	1.938	639	1.938	4.515	85,85%
CNS7	Rep1	04x	55.241	22.610	4.017	6.396	33.023	87,84%
	Rep2	04x	55.650	22.073	4.173	6.369	32.615	87,21%
	Rep1	07x	29.598	6.408	1.430	1.715	9.553	85,03%
	Rep2	07x	29.905	6.270	1.379	1.685	9.334	85,23%
	Rep1	11x	11.364	1.266	411	373	2.050	79,95%
	Rep2	11x	11.383	1.297	464	392	2.153	78,45%
	Rep1	15x	4.369	305	172	92	569	69,77%
	Rep2	15x	4.351	275	195	275	745	73,83%
CNS77	Rep1	04x	12.261	53.274	7.198	15.531	76.003	90,53%
	Rep2	04x	11.250	53.554	7.998	15.463	77.015	89,62%
	Rep1	07x	4.457	24.554	3.604	6.536	34.694	89,61%
	Rep2	07x	4.095	24.847	3.817	6.480	35.144	89,14%
	Rep1	11x	1.376	8.357	1.517	2.164	12.038	87,40%
	Rep2	11x	1.238	8.585	1.581	2.132	12.298	87,14%
	Rep1	15x	557	2.884	693	804	4.381	84,18%
	Rep2	15x	379	3.081	696	3.081	6.858	89,85%
CNS87	Rep1	04x	45.168	30.002	4.542	8.552	43.096	89,46%
	Rep2	04x	45.269	29.781	4.919	8.296	42.996	88,56%
	Rep1	07x	24.785	9.836	1.891	2.639	14.366	86,84%
	Rep2	07x	24.828	9.924	1.952	2.535	14.411	86,45%
	Rep1	11x	9.766	2.333	686	629	3.648	81,20%
	Rep2	11x	9.910	2.409	629	588	3.626	82,65%
	Rep1	15x	4.011	519	246	162	927	73,46%
	Rep2	15x	3.932	582	251	582	1.415	82,26%
CNS88	Rep1	04x	65.754	15.703	2.416	4.391	22.510	89,27%
	Rep2	04x	64.933	16.044	2.722	4.566	23.332	88,33%
	Rep1	07x	33.813	3.619	762	957	5.338	85,72%
	Rep2	07x	33.692	3.728	821	998	5.547	85,20%
	Rep1	11x	12.504	539	225	146	910	75,27%
	Rep2	11x	12.554	625	222	135	982	77,39%
	Rep1	15x	4.754	95	59	30	184	67,93%
	Rep2	15x	4.757	95	61	95	251	75,70%
CNS9	Rep1	04x	68.703	13.617	2.249	3.695	19.561	88,50%
	Rep2	04x	69.097	13.333	2.315	3.520	19.168	87,92%
	Rep1	07x	35.126	2.699	631	695	4.025	84,32%
	Rep2	07x	35.214	2.643	657	725	4.025	83,68%
	Rep1	11x	12.829	326	156	103	585	73,33%
	Rep2	11x	12.969	323	147	97	567	74,07%

	Rep1	15x	4.845	31	39	23	93	58,06%
	Rep2	15x	4.825	49	42	49	140	70,00%
CNS90	Rep1	04x	43.655	31.068	4.526	9.015	44.609	89,85%
	Rep2	04x	44.113	30.690	4.640	8.822	44.152	89,49%
	Rep1	07x	23.923	10.471	1.882	2.875	15.228	87,64%
	Rep2	07x	24.261	10.294	1.838	2.846	14.978	87,73%
	Rep1	11x	9.363	2.628	691	732	4.051	82,94%
	Rep2	11x	9.466	2.673	673	724	4.070	83,46%
	Rep1	15x	3.735	669	295	239	1.203	75,48%
	Rep2	15x	3.743	695	272	695	1.662	83,63%

Observou-se, inicialmente, grande congruência entre os resultados obtidos entre as repetições 1 e 2 (REP1 e REP2). O rendimento de sequências (*reads*) foi similar nas duas repetições (Tabela 1). A variação do número de SNPs detectados para cada acesso nas diferentes coberturas genômicas nas repetições 1 e 2 também foi muito similar (Tabela 2). Estimou-se o *p-value* para um Teste T entre os dois conjuntos de dados referentes às duas repetições (REP1 e REP2) para verificar se as repetições possuem a mesma distribuição. O *p-value* foi estimado em 0,473, portanto, não significativo. Isto indica que a quantidade de SNPs obtidos por GBS entre repetições experimentais foi similar. A plotagem dos dados para a REP1 (Figura 1a) e para a REP2 (Figura 1b) mostra claramente esta tendência. Independentemente da cobertura estabelecida para a chamada de um SNP, a quantidade de SNPs obtida entre diferentes repetições é mantida.

Como o número de SNPs obtidos em cada cobertura genômica não é diferente entre repetições, o próximo passo foi verificar se os SNPs obtidos em cada repetição são, de fato, os mesmos SNPs. Para verificar este ponto, foi realizada uma comparação de cada endereço dos milhares de SNPs detectados entre os conjuntos de dados em cada cobertura genômica (4x, 7x, 11x, 15x). Utilizou-se a posição do SNP no mapa físico de cada cromossomo em pares de bases em relação ao genoma referência da cultivar Nipponbare. Isto significou a comparação do conjunto de 145.768 SNPs identificados na REP1 (todas as coberturas) com o conjunto de 145.970 SNPs identificados na REP2 (Tabela 3). Na cobertura 4x observou-se, por exemplo, que dos 88.264 SNPs considerados para comparação, 64.252 são comuns entre a REP1 e REP2 (Tabela 3). Isto indica que nesta cobertura cerca de 72,80% dos SNPs são comuns entre as repetições. O mesmo foi realizado com os dados das outras coberturas (7x, 11x e 15x) onde o valor estimado de SNPs comuns entre a REP1 e a REP2 também foi de ~72% (Tabela 3). Os resultados indicam que ensaio GBS de 24 acessos de arroz de sequeiro, o número de SNPs detectado entre as repetições é similar e que, em média, cerca de 72% dos SNPs são comuns entre repetições.

Observou-se ainda, conforme o esperado, que à medida em que a cobertura aumenta o número de SNPs cai drasticamente (Figura 1). Por exemplo, para a variedade CNS10 (Tabela 1), observou-se que o número de SNPs detectados variou de forma decrescente, na média das duas repetições, de ~63.000 na cobertura 4x, para ~25.000 na cobertura 7x, 7.800

cobertura 11x e apenas 3.200 SNPs na cobertura 15x. Assim, ao aumentar-se a cobertura, procura-se diminuir a possibilidade de detectar SNPs espúrios mas, como consequência, diminui-se o número de SNPs selecionados para análise.

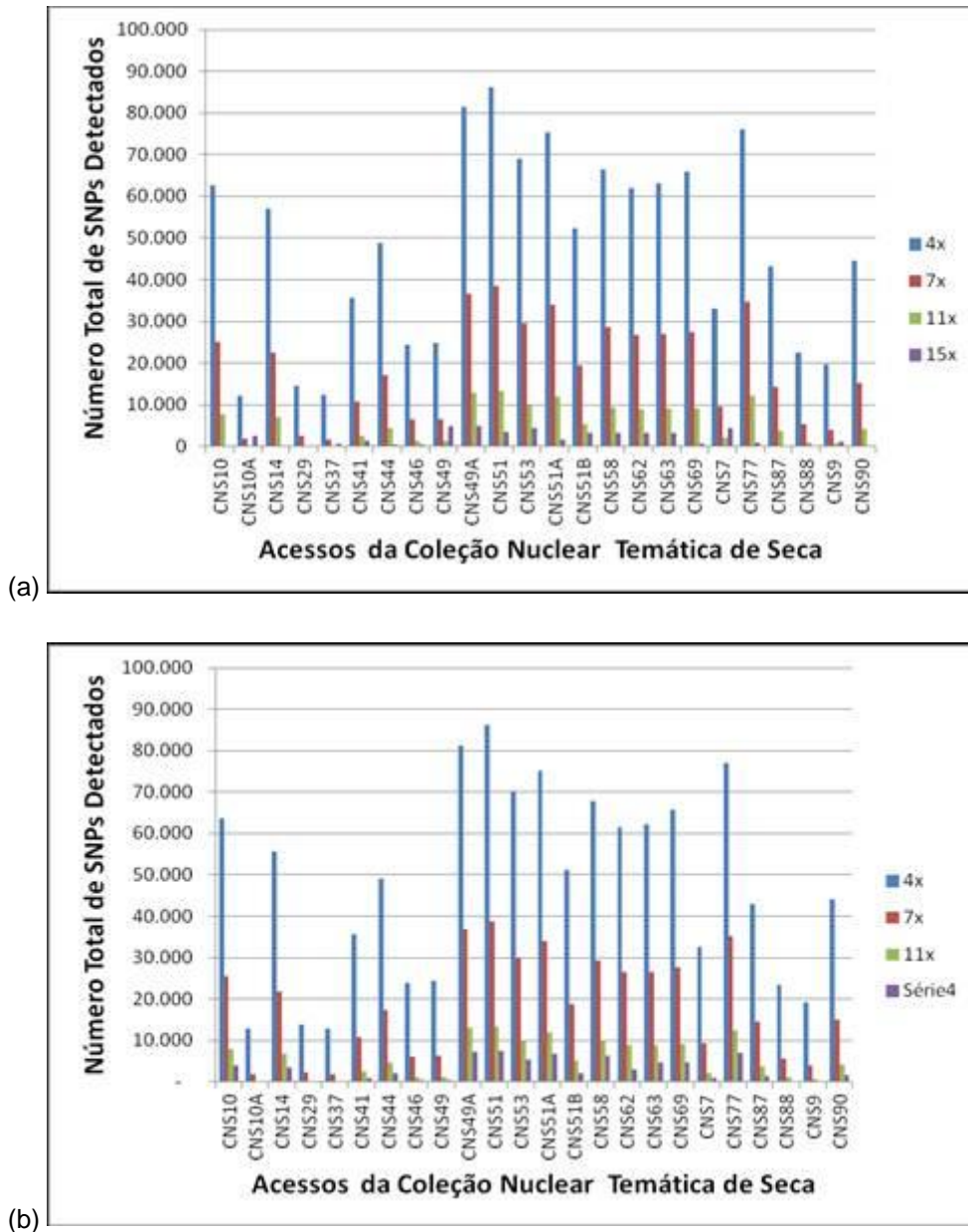


Figura 1. Decréscimo do número de SNPs detectados pela metodologia GBS em função da cobertura genômica em sítios SNP para uma amostra de 24 acessos de arroz de sequeiro da Coleção Nuclear Temática de Tolerância à Seca. O número de SNPs detectados em cada cobertura genômica (4x, 7x, 11x e 15x) é similar nas duas repetições. (a) REP1 – distribuição de número de SPSs observada na repetição 1 e (b) REP2 – distribuição de número de SNPs observada na repetição 2, do ensaio GBS.

Tabela 3. Estimativa do percentual de SNPs em comum em duas repetições do ensaio GBS realizadas com 24 acessos de arroz de sequeiro.

Cobertura	No. de SNPs (REP1)	No. de SNPs (REP2)	No. de comparações realizadas	Total de SNPs em comum	% de SNPs em comum entre REP1 e REP2
4x	88.264	88.265	88.265	64.252	72,79%
7X	39.152	39.230	39.239	28.483	72,59%
11X	13.414	13.536	13.536	9.859	72,84%
15X	4.938	4.939	4.939	3.580	72,48%

Se o número de SNPs detectados cai à medida em que a cobertura aumenta (Tabela 2, Figura 1), procurou-se verificar se esta queda seria compensada pela detecção de SNPs com maior qualidade de sequência em cada sítio amostrado. Em outras palavras, verificou-se se a perda significativa de SNPs com o aumento de cobertura resultava necessariamente em SNPs com maior qualidade de sequência e, portanto, menor possibilidade de seleção de falsos positivos. Haveria aí, desta forma, uma compensação: menor número de SNPs com maior cobertura, porém maior proporção de SNPs de maior qualidade analítica e, conseqüentemente, menor número de falsos positivos. As sequências oriundas de sequenciamento de DNA podem ser avaliadas quanto à qualidade de acordo com o sinal emitido por cada base, avaliado em pequenos conjuntos de bases nitrogenadas ao longo da fita de DNA. Portanto, para cada um dos SNPs detectados pode ser atribuído um valor de qualidade. Este valor é definido pela probabilidade de existência do polimorfismo no sítio considerado, de acordo com uma escala de qualidade (ex. escala *Phred*) das sequências naquele loco ($Phred = -10 \cdot \log(1-p)$). Assim, um valor de qualidade de SNP igual a 10 *Oryza sativa* que a probabilidade de erro de detecção do SNP seria 1 em 10, enquanto que um valor de qualidade de SNP igual a 100 indica como 1 em 100 a probabilidade de erro de detecção do SNP.

O SNPs detectados com a metodologia GBS foram classificados de acordo com o valor de qualidade estimado em cada sítio em classes variando de 0 a 100, 100 a 1.000, 1.000 a 5.000, 5.000 a 10.000 e acima de 10.000. Esta classificação foi feita para cada cobertura genômica (Tabela 4). Observou-se que o percentual de SNPs com valor de qualidade inferior a 100 diminui à medida em que a cobertura aumenta (Tabela 4). Por exemplo, na cobertura 4x foram identificados 29.730 SNPs com valor de qualidade inferior a 100, o que corresponde a 33,68% dos SNPs detectados nesta cobertura. Na cobertura 7x, este percentual cai para 25,74%, em 11x diminui para 18,24% e atinge 13,16% na cobertura 15x. A tendência inversa é observada para os SNPs com valor de qualidade acima de 10.000. Na cobertura 4x apenas 0,6% dos SNPs possuem este valor de qualidade, mas à medida que a cobertura aumenta, observa-se um aumento do percentual de SNPs nesta categoria, com percentual de 1,34% na cobertura 7x, 3,9% na cobertura 11x, atingindo 9,68% na cobertura 15x. Portanto, verifica-se um aumento de qualidade dos SNPs à medida em que a cobertura genômica aumenta. Observa-se, ainda, que cerca de 60% dos SNPs possuem valor de qualidade entre 100 e

5.000, independentemente da cobertura (Figura 2). Isto indica que a qualidade dos SNPs detectados em ensaio GBS é, aparentemente, elevada.

Assim, para a escolha da cobertura genômica a ser utilizada para a detecção de SNPs, observou-se que:

(a) não há diferença significativa entre o número de SNPs detectados em cada repetição, independentemente da cobertura genômica analisada;

(b) aproximadamente 70% dos SNPs têm o mesmo endereço no genoma entre as repetições realizadas;

(c) há um aumento de valor de qualidade do SNP com o aumento de cobertura, o que compensa a queda do número de SNPs verificada à medida em que a cobertura também aumenta. Levando-se em consideração estes dados, uma cobertura mínima de 7x parece ser adequada para a detecção de SNPs em ensaios GBS por possibilitar a obtenção de um número alto de SNPs (39.195, média das duas repetições), conservar ~70% dos SNPs entre repetições e favorecer a obtenção de SNPs com alto valor de qualidade em mais de 75% do conjunto de dados. A cobertura 7x parece ser um bom compromisso entre o grande número de SNPs detectados na cobertura 4x, mas possivelmente com maior probabilidade de seleção de SNPs espúrios, e do pequeno número de SNPs de coberturas superiores a 11x, mas com maior valor de qualidade de SNP.

Comparação de análise genética realizadas com SNPs detectados por ensaio GBS e genotipagem com painéis 384-plex de SNPs - Um outro ponto a ser avaliado para o potencial emprego da metodologia GBS em arroz de sequeiro é se os genótipos obtidos em cada marcador SNP são fidedignos, isto é, realmente representam o genótipo esperado naquele loco. Para isto, utilizou-se as amostras em duplicata e triplicata dos acessos CNS 10, CNS 49 e CNS 51 sob a hipótese de que a genotipagem de uma extração de DNA (prova) deve ser idêntica à genotipagem da re-extração de DNA (contraprova) da mesma planta. As variações observadas entre um genótipo multiloco e outro servem para estimar o erro de genotipagem da metodologia GBS, determinando a sua acurácia. Para avaliar a acurácia dos dados de genotipagem obtidos com a metodologia GBS foram realizadas ainda comparações entre os genótipos multiloco obtidos com esta metodologia e os dados de genotipagem com os dois painéis 384-plex (Capítulo 1).

Para isso, foram definidas, inicialmente, algumas etapas de análise e seleção de SNPs em ensaios GBS:

(1) procedeu-se a detecção de SNPs na cobertura mínima 7x. Neste ensaio, foram detectados 39.195 SNPs com esta cobertura (Tabela 4);

(2) selecionou-se, entre os 39.195 SNPs apenas aqueles com valor de qualidade acima de 100, restando, após este procedimento, 29.105 SNPs com alta qualidade;

(3) utilizou-se destes 29.105 SNPs apenas os SNPs comuns em duas repetições, o que significou manter na análise 21.608 SNPs;

(4) considerou-se para a análise genética apenas os SNPs com menos de 20% de dados faltantes e que eram polimórficos entre os 24 acessos (monomórficos foram retirados).

Portanto, qualquer marcador SNP com 5 ou mais dados faltantes entre os acessos foi retirado da análise. O resultado foi um conjunto de 1.572 marcadores SNPs selecionados em ensaio GBS após cumpridas as etapas (filtros) descritas acima.

É importante enfatizar que este último passo, que analisa os SNPs comuns entre o conjunto de acessos analisados, eliminando aqueles com excesso de dados faltantes, parece ser o que tem impacto mais significativo no conjunto final de SNPs considerado para análise. Isto porque os SNPs com dados faltantes observados em um acesso não coincidem, necessariamente, com os SNPs com dados faltantes de outro acesso. Portanto, ao comparar um acesso com os demais, apesar do número de SNPs ser alto em cada acesso, apenas parte dos SNPs são comuns entre a maioria dos acessos. À medida em que o número de acessos aumenta, diminui os SNPs em comum entre os acessos. O percentual de queda no número de SNPs em cada etapa de seleção serve para ilustrar este ponto. Por exemplo, a seleção dos SNPs com valor de qualidade acima de 100 reduziu em 25,74% o número inicial de SNPs detectados (de 39.195 SNPs para 29.105 SNPs). O emprego apenas dos SNPs comuns nas duas repetições reduziu em 25,75% o número de SNPs de maior qualidade (de 29.105 SNPs para 21.608 SNPs). Já o emprego de marcadores SNPs com no máximo 20% de dados faltantes no conjunto de acessos levou à redução de 92,72% no número de SNPs comuns nas duas repetições (de 21.608 SNPs para 1.572 SNPs). Ao final, dos 39.195 SNPs selecionados na cobertura 7x, apenas 1.572 foram selecionados para análise entre os 24 acessos de arroz de sequeiro, o que representa apenas 4% dos SNPs inicialmente selecionados. Este poderia ser considerado, portanto, o rendimento final esperado em um ensaio GBS.

Tabela 4. Distribuição de valores de qualidade de SNPs detectados pela tecnologia GBS em função do aumento da cobertura genômica (média de duas repetições).

Qualidade	Cobertura 4x		Cobertura 7x		Cobertura 11x		Cobertura 15x	
	No. SNPs	%	No. SNPs	%	No. SNPs	%	No. SNPs	%
<100	29.730	33,68%	10.090	25,74%	2.458	18,24%	650	13,16%
100<x<1000	35.151	39,82%	17.509	44,67%	6188	45,92%	2.088	42,29%
1000<x<5000	19.360	21,93%	7.621	19,44%	2535	18,82%	1.131	22,90%
5000<x<10000	3.496	3,96%	34.47	8,80%	1767	13,12%	591	11,97%
>10000	528	0,60%	527	1,34%	525	3,90%	478	9,68%
TOTAL	88.265	100,00%	39.195	100,00%	13.475	100,00%	4.938	100,00%

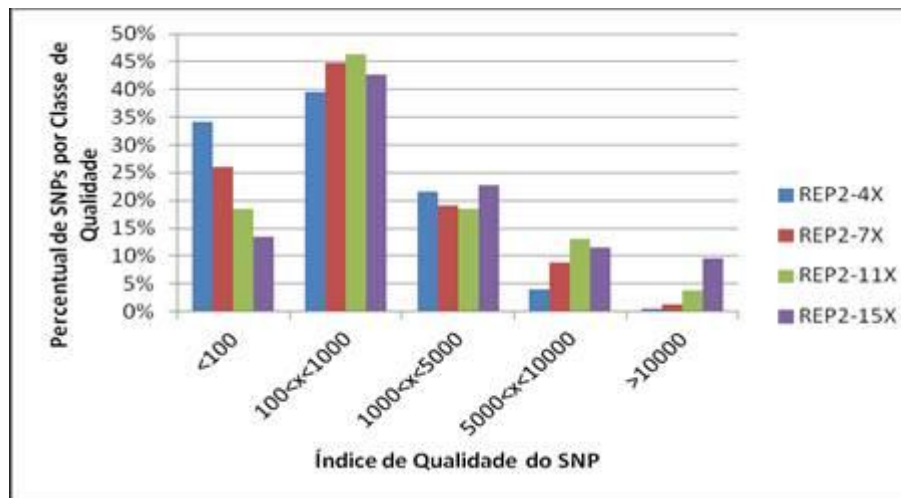
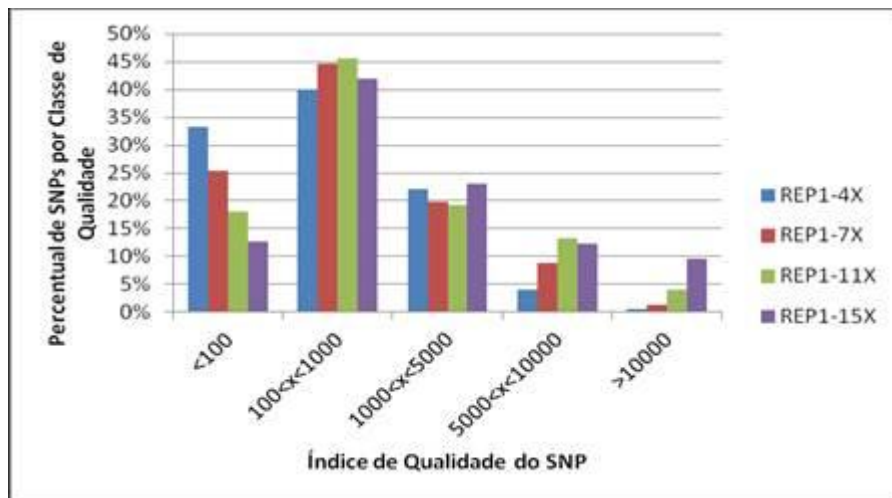


Figura 2. Distribuição de SNPs por classes de valor de qualidade em duas repetições (esquerda, REP1; direita, REP2) em função do aumento da cobertura genômica (4x, 7x, 11x e 15x) em cada sítio de DNA.

A Tabela 5 traz um resumo dos 1.572 SNPs selecionados para análise segundo os critérios descritos acima. Uma tabela similar foi preparada com os dados de genotipagem dos mesmos acessos com os dois painéis 384-plex descritos no Capítulo 1 (Tabela 6).

Tabela 5 – Estimativa de homozigose de acessos de arroz de sequeiro com base na amostragem de marcadores SNP em ensaio GBS.

Acessos	Falhas	Genótipo AA	Genótipo AB	Genótipo BB	Total SNPs	Homozigose
CNS-7	202	882	241	247	1.370	82,41%
CNS-9	676	572	159	165	896	82,25%
CNS-10	9	1.056	250	257	1.563	84,01%
CNS-10A	979	373	132	88	593	77,74%
CNS-14	25	1.019	249	279	1.547	83,90%
CNS-29	905	400	151	116	667	77,36%
CNS-37	1.111	377	47	37	461	89,80%
CNS-41	144	919	251	258	1.428	82,42%
CNS-44	52	996	246	278	1.520	83,82%
CNS-46	387	344	603	238	1.185	49,11%
CNS-49	374	352	615	231	1.198	48,66%
CNS-49A	-	1.030	287	255	1.572	81,74%
CNS-51	-	1.051	250	271	1.572	84,10%
CNS-53	1	1.046	256	269	1.571	83,70%
CNS-51A	1	1.055	244	272	1.571	84,47%
CNS-51B	49	1.025	228	270	1.523	85,03%
CNS-58	57	1.356	85	74	1.515	94,39%
CNS-62	36	464	284	788	1.536	81,51%
CNS-63	26	1.045	220	281	1.546	85,77%
CNS-69	7	1.012	293	260	1.565	81,28%
CNS-77	-	1.011	260	301	1.572	83,46%
CNS-87	92	978	237	265	1.480	83,99%
CNS-88	430	748	198	196	1.142	82,66%
CNS-90	73	971	249	279	1.499	83,39%
Mínimo	-	-	-	-	461	48,66%
Média	235	-	-	-	1.337	80,71%
Máximo	1.111	-	-	-	1.572	94,39%
Desvio-Padrão	343	-	-	-	343	10,35%

Tabela 6 – Estimativa de homozigose de acessos de arroz de sequeiro com base na genotipagem com dois painéis 384-plex (ensaio GoldenGate, Illumina).

Acessos	Falhas	Genótipo AA	Genótipo AB	Genótipo BB	Total SNPs	Homozigose
CNS_7	2	316	10	362	688	98,55%
CNS_9	3	323	10	354	687	98,54%
CNS_10	0	330	11	349	690	98,41%
CNS_10A	1	330	10	349	689	98,55%
CNS_14	3	337	13	337	687	98,11%
CNS_29	0	336	13	341	690	98,12%
CNS_37	0	322	9	359	690	98,70%
CNS_41	1	315	43	331	689	93,76%
CNS_44	49	273	12	356	641	98,13%
CNS_46	1	339	12	338	689	98,26%
CNS_49	0	342	8	340	690	98,84%
CNS_49A	2	341	8	339	688	98,84%
CNS_51	1	326	10	353	689	98,55%
CNS_53	1	333	10	346	689	98,55%
CNS_51A	0	329	9	352	690	98,70%
CNS_51B	0	329	9	352	690	98,70%
CNS_58	3	344	7	336	687	98,98%
CNS_62	1	280	92	317	689	86,65%
CNS_63	1	335	11	343	689	98,40%
CNS_69	57	254	7	372	633	98,89%
CNS_77	6	299	10	375	684	98,54%
CNS_87	4	306	12	368	686	98,25%
CNS_88	5	335	10	340	685	98,54%
CNS_90	1	327	14	348	689	97,97%
Mínimo	-	-	-	-	633	86,65%
Média	6	-	-	-	684	97,77%
Máximo	57	-	-	-	690	98,98%
Desvio-Padrão	15	-	-	-	15	2,63%

Não houve congruência entre a análise realizada através do ensaio GBS, após a seleção de 1.572 SNPs usando vários critérios, e análise utilizando um conjunto final de 690 SNPs dos dois painéis 384-plex descritos no Capítulo 1. A comparação entre os dados da Tabela 6 e da Tabela 7 revela que há um aparente excesso de genótipos heterozigotos nos marcadores SNP do ensaio GBS. Ao se estimar o nível de homozigose dos 22 acessos de arroz de sequeiro, a média utilizando SNPs do ensaio GBS foi 80,71%, com valor mínimo de 48,66% e máximo de 94,39%. No entanto, ao se estimar o nível de homozigose dos 22 acessos de arroz de sequeiro com os SNPs dos painéis multiplex, a média foi de 97,77%, com

valor mínimo de 86,65% e máximo de 98,98%. Estranha-se a potencial existência nível de homozigose tão baixo quanto os observados para alguns acessos usando o ensaio GBS, como CNS 46 (49,11%), CNS 49 (48,66%) ou CNS 29 (77,36%). O arroz é uma planta autógama e espera-se que acessos submetidos a autofecundações seguidas resultem em plantas com elevado nível médio de homozigose. Além disso, estes acessos foram submetidos a dois ciclos de autofecundação para multiplicação de sementes desde que as variedades foram selecionadas para compor a Coleção Nuclear de Tolerância à Seca. As estimativas de homozigose média realizadas com os dois painéis 384-plex parecem mais condizentes com o esperado na prática. O acesso com menor nível de homozigose foi CNS 62 (86,65%), seguido de CNS 41 (93,76%). O restante dos acessos tem homozigose média acima de 98%. Não há coincidência de ordem crescente de homozigose na comparação entre as Tabelas 6 e 7, o que indica que, se há um excesso de heterozigotos nos marcadores SNPs revelados por GBS, isto aparentemente ocorre ao acaso.

Deve ser notado ainda que no ensaio GBS, os acessos usados com duplicata possuem níveis de homozigose diferente, refletindo diferenças de genotipagem entre prova e contraprova. O acesso CNS 10 (prova) possui estimativa de homozigose média de 84,01%, enquanto CNS 10A (contraprova) possui 77,74% de homozigose. Diferenças significativas são também observadas entre CNS 49 e CNS 49A (48,66% e 81,74%), respectivamente. Contudo, as estimativas de homozigose para as amostras em triplicata (CNS 51, CNS 51A e CNS 51B) são similares, com valores de 84,10%, 84,47% e 85,03%.

Não há diferença aparente entre a homozigose média estimada para duplicatas e triplicatas utilizando os dados de genotipagem com os dois painéis 384-plex. Mais ainda, o genótipo multiloco destas amostras foi demonstrado idêntico, como seria esperado, com o emprego dos dois painéis 384-plex (Capítulo). Este parece ser um bom teste para ser realizado com as amostras em duplicata e triplicata: verificar se os genótipos multiloco de prova e contraprova são idênticos, sabendo que pequenas variações observadas poderiam ser usadas para estimar o erro de genotipagem ou acurácia. Para isto, os genótipos dos acessos em duplicata (CNS 10 e CNS 10A, CNS 49 e CNS 49A) e em triplicata (CNS 51, CNS 51A e CNS 51B) foram comparados em cada loco SNP nas coberturas 7x, 11x e 15x (Tabela 7). Para aumentar ainda mais a estringência desta análise, foram considerados na comparação apenas os SNPs com valor de qualidade acima de 1.000. Conforme pode ser observado, a taxa de genótipos diferentes em cada comparação é bastante variável (0% a 56,13%). Não houve efeito aparente na taxa de erro com o aumento da cobertura de 7x para 15x. O valor mínimo de qualidade de detecção de SNP também parece não ter influenciado muito, visto que a taxa de erro de genotipagem é, em média, bastante elevada. Os menores valores de erro foram observados entre as amostras em triplicata do acesso CNS 51, variando de 0% a 3,85%. Mesmo neste caso, o erro de genotipagem é considerado alto em duas das comparações (3,85%). Erros acima de 50% foram observados nas coberturas 7x, 11x e 15x.

É possível que os critérios usados para a seleção de SNPs via ensaio GBS não sejam ainda suficientemente estridentes ou adequados para a seleção de marcadores robustos, que possam ser usados em rotina na genotipagem de arroz de sequeiro. É evidente que parte dos genótipos obtidos é correta, idêntica nas análise de prova e contraprova, mas o conjunto de dados ainda apresenta contaminantes, SNPs espúrios que não devem ser considerados na análise. A inclusão de SNPs falso positivos na análise causa, naturalmente, distorções de toda a sorte na análise genética. O emprego de coberturas maiores, como as testadas no presente trabalho (11x e 15x) leva a um decréscimo substancial na quantidade de SNPs final, o que pode limitar o emprego da técnica. É necessária uma análise complementar, potencialmente usando acessos de arroz sabidamente homocigotos, como linhagens duplo-haplóide, para confirmar a existência de excesso de heterocigotos no ensaio GBS. É necessária ainda um aprofundamento em todas as etapas do ensaio GBS para a seleção de marcadores SNP fidedignos, passíveis de uso em escala na genotipagem de arroz.

Tabela 7 – Estimativa de erro de genotipagem em ensaio GBS utilizando extração e re-extração de DNA (prova e contraprova) em três variedades de arroz de sequeiro. Foram utilizados nesta análise apenas os SNPs com valor de qualidade acima de 1.000.

Comparação (prova e contraprova)	Cobertura	Total de locos SNP	Total de falhas	SNPs analisados	Genótipos distintos	Genótipos idênticos	Erro (%)
CNS 10 X CNS 10A	7x	856	512	344	81	263	23,55%
CNS 49 X CNS 49A	7x	856	195	661	371	290	56,13%
CNS 51 X CNS51 A	7x	856	0	856	134	722	15,65%
CNS 51 X CNS 51B	7x	856	24	832	164	668	19,71%
CNS 51A X CNS 51B	7x	856	24	832	118	714	14,18%
CNS 10 X CNS 10A	11x	120	79	41	12	29	29,27%
CNS 49 X CNS 49A	11x	120	23	97	48	49	49,48%
CNS 51 X CNS51A	11x	120	0	120	18	102	15,00%
CNS 51 X CNS 51B	11x	120	3	117	17	100	14,53%
CNS 51A X CNS 51B	11x	120	3	117	13	104	11,11%
CNS 10 X CNS 10A	15x	26	10	16	4	12	25,00%
CNS 49 X CNS 49A	15x	26	4	22	14	8	63,64%
CNS 51 X CNS51 A	15x	26	0	26	1	25	3,85%
CNS 51 X CNS 51B	15x	26	0	26	0	26	0,00%
CNS 51A X CNS 51B	15x	26	0	26	1	25	3,85%

6. Bibliografia

- Baird, N. a, Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. (J. C. Fay, Ed.) *PloS one*, 3(10), e3376. Public Library of Science. doi:10.1371/journal.pone.0003376
- Barbazuk, W. B., Bedell, J. A., & Rabinowicz, P. D. (2005). Reduced representation sequencing: a success in maize and a promise for other plant genomes. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 27(8), 839-48. doi:10.1002/bies.20262
- Craig, D. W., Pearson, J. V., Szelinger, S., Sekar, A., Redman, M., Corneveaux, J. J., Pawlowski, T. L., et al. (2008). Identification of genetic variants using bar-coded multiplexed sequencing. *Nature methods*, 5(10), 887-93. Nature Publishing Group. doi:10.1038/nmeth.1251
- Cronn, R., Liston, A., Parks, M., Gernandt, D. S., Shen, R., & Mockler, T. (2008). Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic acids research*, 36(19), e122. doi:10.1093/nar/gkn502
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., et al. (2011). The variant call format and VCFtools. *Bioinformatics (Oxford, England)*, 27(15), 2156-8. doi:10.1093/bioinformatics/btr330
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*, 43(5), 491-8. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved. doi:10.1038/ng.806
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. (L. Orban, Ed.) *PloS one*, 6(5), e19379. Public Library of Science. doi:10.1371/journal.pone.0019379
- Ferreira, M., & Grattapaglia, D. (1998). *Introdução ao uso de marcadores moleculares em análise genética* (Documentos., p. 220p). EMBRAPA-CENARGEN.
- Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., Guan, J., et al. (2009). High-throughput genotyping by whole-genome resequencing. *Genome Research*, 1068-1076. doi:10.1101/gr.089516.108.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., Garimella, K., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9), 1297-303. doi:10.1101/gr.107524.110
- Myles, S., Chia, J.-M., Hurwitz, B., Simon, C., Zhong, G. Y., Buckler, E., & Ware, D. (2010). Rapid genomic characterization of the genus vitis. *PloS one*, 5(1), e8219. doi:10.1371/journal.pone.0008219
- Pessoa-Filho, M., Rangel, P. H. N., & Ferreira, M. E. (2010). Extracting samples of high diversity from thematic collections of large gene banks using a genetic-distance based approach. *BMC plant biology*, 10(1), 127. BioMed Central Ltd. doi:10.1186/1471-2229-10-127

Wiedmann, R. T., Smith, T. P. L., & Nonneman, D. J. (2008). SNP discovery in swine by reduced representation and high throughput pyrosequencing. *BMC genetics*, 9(1), 81. doi:10.1186/1471-2156-9-81

ANEXOS

6.1. Anexo 1 – Parâmetros comparativos de avaliação genética entre os marcadores SNP que compõem os dois painéis 384-plex com base na genotipagem de uma amostra de 65 acessos da Coleção Nuclear Temática de Tolerância à Seca de arroz.

Marcador SNP	Frequência Alelo mais Comum	No. Genótipos	Dados Observados	No. Alelos	Diversidade Gênica	Heterozigosidade observada	PIC
Chr1_11272869	0,62	3	65	2	0,47	0,03	0,36
Chr1_11312519	0,70	2	64	2	0,42	0,61	0,33
Chr1_11339617	0,61	3	64	2	0,48	0,03	0,36
Chr1_11374629	0,67	3	65	2	0,44	0,05	0,34
Chr1_11426073	0,72	3	65	2	0,41	0,02	0,32
Chr1_11465807	0,62	3	65	2	0,47	0,03	0,36
Chr1_11469994	0,72	3	65	2	0,41	0,02	0,32
Chr1_11653926	0,79	3	65	2	0,33	0,05	0,27
Chr1_12867748	0,53	3	59	2	0,50	0,02	0,37
Chr1_12897299	0,60	3	65	2	0,48	0,06	0,36
Chr1_12940593	0,52	3	65	2	0,50	0,02	0,37
Chr1_12976776	0,60	3	65	2	0,48	0,06	0,36
Chr1_13135677	0,53	3	65	2	0,50	0,05	0,37
Chr1_13148124	0,52	3	65	2	0,50	0,08	0,37
Chr1_13217704	0,52	3	65	2	0,50	0,02	0,37
Chr1_13249852	0,60	3	65	2	0,48	0,06	0,36
Chr1_14734744	0,75	2	65	2	0,37	0,00	0,30
Chr1_15348002	0,52	2	60	2	0,50	0,00	0,37
Chr1_18468432	0,62	2	65	2	0,47	0,00	0,36
Chr1_18525132	0,54	3	63	2	0,50	0,03	0,37
Chr1_18530303	0,53	3	63	2	0,50	0,02	0,37
Chr1_18590506	0,61	3	65	2	0,48	0,02	0,36
Chr1_18604179	0,60	2	65	2	0,48	0,00	0,36

Chr1_18668242	0,61	3	65	2	0,48	0,05	0,36
Chr1_19139384	0,50	3	65	2	0,50	0,02	0,38
Chr1_19195812	0,51	3	64	2	0,50	0,08	0,37
Chr1_2279277	0,86	2	65	2	0,24	0,00	0,21
Chr1_2326868	0,84	3	65	2	0,27	0,02	0,23
Chr1_2377724	0,79	3	65	2	0,33	0,02	0,27
Chr1_23787556	0,74	2	65	2	0,39	0,00	0,31
Chr1_23809551	0,61	3	65	2	0,48	0,02	0,36
Chr1_23836460	0,69	2	65	2	0,43	0,00	0,34
Chr1_23857450	0,83	3	65	2	0,28	0,03	0,24
Chr1_23883508	0,59	3	65	2	0,48	0,02	0,37
Chr1_2388681	0,95	2	65	2	0,09	0,00	0,08
Chr1_2389723	0,88	2	65	2	0,22	0,00	0,19
Chr1_23912753	0,85	3	64	2	0,25	0,05	0,22
Chr1_23955243	0,76	3	65	2	0,36	0,02	0,30
Chr1_23965910	0,77	2	65	2	0,36	0,00	0,29
Chr1_24015434	0,67	3	65	2	0,44	0,08	0,34
Chr1_2403294	0,95	2	65	2	0,09	0,00	0,08
Chr1_24081577	0,78	2	59	2	0,34	0,00	0,28
Chr1_24138244	0,65	3	65	2	0,46	0,06	0,35
Chr1_24147898	0,69	3	64	2	0,43	0,03	0,34
Chr1_24227227	0,70	3	64	2	0,42	0,05	0,33
Chr1_24264642	0,71	2	65	2	0,41	0,00	0,33
Chr1_24313675	0,66	3	65	2	0,45	0,03	0,35
Chr1_24368856	0,72	3	65	2	0,41	0,02	0,32
Chr1_24381553	0,66	3	65	2	0,45	0,03	0,35
Chr1_24430990	0,82	3	65	2	0,29	0,05	0,25
Chr1_24511372	0,66	3	65	2	0,45	0,03	0,35

Chr1_24535665	0,72	3	65	2	0,41	0,02	0,32
Chr1_24598139	0,67	3	65	2	0,44	0,02	0,34
Chr1_24631118	0,62	3	65	2	0,47	0,03	0,36
Chr1_24653796	0,64	2	64	2	0,46	0,00	0,35
Chr1_24701256	0,62	3	65	2	0,47	0,05	0,36
Chr1_24727564	0,66	2	65	2	0,45	0,00	0,35
Chr1_24803304	0,66	2	65	2	0,45	0,00	0,35
Chr1_24826024	0,62	3	65	2	0,47	0,02	0,36
Chr1_25211220	0,85	3	65	2	0,25	0,05	0,22
Chr1_25266532	0,87	3	65	2	0,23	0,02	0,20
Chr1_25408222	0,92	2	65	2	0,14	0,00	0,13
Chr1_25457890	0,92	2	65	2	0,14	0,00	0,13
Chr1_2562473	0,95	2	65	2	0,09	0,00	0,08
Chr1_2572143	0,95	2	65	2	0,09	0,00	0,08
Chr1_2576718	0,95	2	65	2	0,09	0,00	0,08
Chr1_2616120	0,94	2	63	2	0,12	0,00	0,11
Chr1_2660700	0,94	2	65	2	0,12	0,00	0,11
Chr1_2705284	0,94	2	65	2	0,12	0,00	0,11
Chr1_2719765	0,88	3	65	2	0,20	0,02	0,18
Chr1_27288194	0,56	3	65	2	0,49	0,11	0,37
Chr1_2797145	0,90	3	65	2	0,18	0,02	0,16
Chr1_2832944	0,88	3	65	2	0,20	0,02	0,18
Chr1_2870838	0,88	3	65	2	0,20	0,02	0,18
Chr1_29275933	0,55	2	65	2	0,49	0,00	0,37
Chr1_29821621	0,59	3	65	2	0,48	0,02	0,37
Chr1_31779292	0,82	3	65	2	0,29	0,05	0,25
Chr1_31796289	0,74	2	65	2	0,39	0,00	0,31
Chr1_31841038	0,73	3	64	2	0,39	0,03	0,31

Chr1_34761138	0,55	2	65	2	0,49	0,00	0,37
Chr1_35006178	0,54	3	65	2	0,50	0,06	0,37
Chr1_35030275	0,63	2	59	2	0,47	0,00	0,36
Chr1_35810792	0,73	3	65	2	0,39	0,02	0,32
Chr1_36971673	0,97	2	65	2	0,06	0,00	0,06
Chr1_37183724	0,97	2	65	2	0,06	0,00	0,06
Chr1_37707861	0,60	2	65	2	0,48	0,00	0,36
Chr1_38285201	0,92	2	65	2	0,14	0,00	0,13
Chr1_38340675	0,57	2	65	2	0,49	0,00	0,37
Chr1_39376380	0,92	2	65	2	0,14	0,00	0,13
Chr1_39390907	0,92	2	65	2	0,14	0,00	0,13
Chr1_39422020	0,88	2	65	2	0,22	0,00	0,19
Chr1_39440721	0,91	2	64	2	0,17	0,00	0,16
Chr1_39464636	0,88	2	65	2	0,22	0,00	0,19
Chr1_39473259	0,97	2	65	2	0,06	0,00	0,06
Chr1_39988267	0,50	3	65	2	0,50	0,02	0,38
Chr1_40036706	0,63	3	65	2	0,47	0,03	0,36
Chr1_41244361	0,89	3	65	2	0,19	0,06	0,17
Chr1_41304470	0,56	3	65	2	0,49	0,05	0,37
Chr1_42021996	0,58	3	65	2	0,49	0,03	0,37
Chr1_42457631	0,59	3	65	2	0,48	0,17	0,37
Chr1_42491941	0,58	3	60	2	0,49	0,02	0,37
Chr1_42523627	0,58	3	60	2	0,49	0,05	0,37
Chr1_42604189	0,74	3	65	2	0,39	0,06	0,31
Chr1_42653198	0,54	3	65	2	0,50	0,06	0,37
Chr1_5130438	0,78	3	65	2	0,35	0,02	0,29
Chr1_5168074	0,90	2	60	2	0,18	0,00	0,16
Chr1_5194342	0,90	3	63	2	0,19	0,02	0,17

Chr1_615367	0,65	3	64	2	0,46	0,02	0,35
Chr1_628044	0,67	3	65	2	0,44	0,05	0,34
Chr1_6378875	0,63	2	65	2	0,47	0,00	0,36
Chr1_7250515	0,62	3	65	2	0,47	0,05	0,36
Chr1_7271290	0,56	3	65	2	0,49	0,02	0,37
Chr1_7327287	0,60	3	65	2	0,48	0,03	0,36
Chr1_7352717	0,52	3	65	2	0,50	0,02	0,37
Chr1_7360011	0,52	3	65	2	0,50	0,06	0,37
Chr1_7396465	0,55	2	58	2	0,49	0,00	0,37
Chr1_7425807	0,51	3	65	2	0,50	0,06	0,37
Chr1_7443772	0,51	3	64	2	0,50	0,02	0,37
Chr1_7450516	0,61	3	65	2	0,48	0,02	0,36
Chr1_7493335	0,52	3	65	2	0,50	0,02	0,37
Chr1_8310349	0,65	3	65	2	0,45	0,02	0,35
Chr10_11682240	0,70	3	65	2	0,42	0,02	0,33
Chr10_11737478	0,69	2	65	2	0,43	0,00	0,34
Chr10_11740834	0,69	3	65	2	0,43	0,03	0,34
Chr10_11789711	0,92	2	65	2	0,14	0,00	0,13
Chr10_11825111	0,59	3	65	2	0,48	0,05	0,37
Chr10_11886826	0,59	3	65	2	0,48	0,02	0,37
Chr10_11911606	0,94	3	65	2	0,12	0,03	0,11
Chr10_11920922	0,94	2	65	2	0,12	0,00	0,11
Chr10_11945290	0,59	3	65	2	0,48	0,05	0,37
Chr10_12019694	0,70	3	65	2	0,42	0,02	0,33
Chr10_13050766	0,83	3	65	2	0,28	0,03	0,24
Chr10_13085389	0,80	2	65	2	0,32	0,00	0,27
Chr10_13111426	0,89	3	65	2	0,19	0,06	0,17
Chr10_13116751	0,92	2	65	2	0,14	0,00	0,13

Chr10_13179272	0,91	3	65	2	0,17	0,03	0,15
Chr10_13206132	0,61	3	65	2	0,48	0,02	0,36
Chr10_13229937	0,92	3	65	2	0,14	0,03	0,13
Chr10_15272868	0,75	3	65	2	0,38	0,02	0,31
Chr10_16535481	0,56	3	62	2	0,49	0,03	0,37
Chr10_16556591	0,63	3	57	2	0,47	0,04	0,36
Chr10_16769865	0,58	3	65	2	0,49	0,02	0,37
Chr10_16782798	0,70	3	65	2	0,42	0,05	0,33
Chr10_16830602	0,56	3	65	2	0,49	0,05	0,37
Chr10_18425910	0,58	3	65	2	0,49	0,05	0,37
Chr10_18449742	0,57	3	65	2	0,49	0,03	0,37
Chr10_19842255	0,51	3	65	2	0,50	0,06	0,37
Chr10_19853566	0,52	3	65	2	0,50	0,02	0,37
Chr10_20153803	0,72	3	65	2	0,40	0,06	0,32
Chr10_20160250	0,72	3	65	2	0,41	0,02	0,32
Chr10_20179897	0,76	3	65	2	0,36	0,08	0,30
Chr10_20218042	0,75	3	65	2	0,38	0,02	0,31
Chr10_2242586	0,82	3	65	2	0,29	0,02	0,25
Chr10_22488323	0,58	2	65	2	0,49	0,00	0,37
Chr10_2302343	0,83	3	62	2	0,28	0,02	0,24
Chr10_23051382	0,52	3	65	2	0,50	0,02	0,37
Chr10_23083676	0,65	3	65	2	0,46	0,03	0,35
Chr10_2340630	0,50	3	64	2	0,50	0,59	0,38
Chr10_2684280	0,53	3	65	2	0,50	0,02	0,37
Chr10_3781853	0,57	3	65	2	0,49	0,03	0,37
Chr10_3838047	0,53	2	64	2	0,50	0,00	0,37
Chr10_3958580	0,61	3	65	2	0,48	0,29	0,36
Chr10_4713938	0,65	2	65	2	0,46	0,00	0,35

Chr11_13291181	0,75	3	65	2	0,38	0,02	0,31
Chr11_16881339	0,64	3	65	2	0,46	0,02	0,36
Chr11_16887564	0,86	3	57	2	0,24	0,04	0,21
Chr11_16954935	0,83	3	64	2	0,28	0,03	0,24
Chr11_17001212	0,80	2	64	2	0,32	0,00	0,27
Chr11_17009070	0,75	3	61	2	0,38	0,02	0,31
Chr11_17023333	0,98	2	52	2	0,04	0,00	0,04
Chr11_17108839	0,53	3	65	2	0,50	0,66	0,37
Chr11_17167169	0,67	3	65	2	0,44	0,02	0,34
Chr11_17200311	0,75	3	63	2	0,38	0,03	0,31
Chr11_17212089	0,76	3	65	2	0,36	0,02	0,30
Chr11_17464275	0,74	3	65	2	0,39	0,06	0,31
Chr11_17504733	0,75	2	65	2	0,37	0,00	0,30
Chr11_17516754	0,71	3	64	2	0,41	0,05	0,33
Chr11_17525120	0,66	3	64	2	0,45	0,09	0,35
Chr11_17597632	0,78	2	65	2	0,34	0,00	0,28
Chr11_17605614	0,66	2	65	2	0,45	0,00	0,35
Chr11_17647750	0,80	2	65	2	0,32	0,00	0,27
Chr11_17702355	0,65	3	65	2	0,45	0,02	0,35
Chr11_17789344	0,98	2	62	2	0,03	0,00	0,03
Chr11_17836032	0,65	3	65	2	0,45	0,02	0,35
Chr11_17882866	0,70	2	61	2	0,42	0,00	0,33
Chr11_17969837	0,94	3	63	2	0,10	0,02	0,10
Chr11_18009558	0,79	3	65	2	0,33	0,02	0,27
Chr11_18035544	0,89	2	65	2	0,19	0,00	0,17
Chr11_18085361	0,75	3	65	2	0,38	0,02	0,31
Chr11_18106947	0,88	2	60	2	0,21	0,00	0,18
Chr11_18111214	0,90	3	65	2	0,18	0,02	0,16

Chr11_18230919	0,56	3	65	2	0,49	0,02	0,37
Chr11_19462431	0,71	2	65	2	0,41	0,00	0,33
Chr11_19507931	0,58	3	65	2	0,49	0,02	0,37
Chr11_19513789	0,57	3	65	2	0,49	0,03	0,37
Chr11_19651123	0,55	3	65	2	0,50	0,02	0,37
Chr11_19671083	0,57	3	65	2	0,49	0,03	0,37
Chr11_19713489	0,62	3	65	2	0,47	0,02	0,36
Chr11_19714192	0,60	3	62	2	0,48	0,05	0,36
Chr11_19849611	0,59	3	63	2	0,48	0,03	0,37
Chr11_19875372	0,62	3	65	2	0,47	0,06	0,36
Chr11_19890334	0,55	3	65	2	0,49	0,03	0,37
Chr11_19892078	0,61	3	65	2	0,48	0,05	0,36
Chr11_19953558	0,62	3	65	2	0,47	0,02	0,36
Chr11_19981650	0,58	3	61	2	0,49	0,05	0,37
Chr11_20015123	0,56	3	63	2	0,49	0,02	0,37
Chr11_20092217	0,52	3	65	2	0,50	0,06	0,37
Chr11_20771608	0,50	3	65	2	0,50	0,02	0,38
Chr11_20982607	0,50	3	63	2	0,50	0,02	0,38
Chr11_21234402	0,52	3	65	2	0,50	0,02	0,37
Chr11_21322064	0,55	3	64	2	0,49	0,02	0,37
Chr11_21330402	0,54	3	64	2	0,50	0,02	0,37
Chr11_21386015	0,56	3	65	2	0,49	0,02	0,37
Chr11_21431933	0,63	2	65	2	0,47	0,00	0,36
Chr11_21496513	0,62	3	64	2	0,47	0,02	0,36
Chr11_21535425	0,57	3	64	2	0,49	0,02	0,37
Chr11_21554364	0,53	2	64	2	0,50	0,00	0,37
Chr11_22335713	0,50	2	58	2	0,50	0,00	0,38
Chr11_22684758	0,95	2	64	2	0,09	0,00	0,09

Chr11_24494344	0,86	2	65	2	0,24	0,00	0,21
Chr11_24546934	0,87	3	63	2	0,23	0,05	0,21
Chr11_24611954	0,79	2	62	2	0,33	0,00	0,28
Chr11_24653048	0,80	3	65	2	0,32	0,03	0,27
Chr11_27010793	0,80	2	65	2	0,32	0,00	0,27
Chr11_27025213	0,69	3	65	2	0,43	0,03	0,34
Chr11_27155526	0,66	3	63	2	0,45	0,05	0,35
Chr11_27365240	0,71	3	65	2	0,41	0,03	0,33
Chr11_27520584	0,55	3	65	2	0,50	0,05	0,37
Chr11_27534452	0,55	3	65	2	0,50	0,08	0,37
Chr11_27535474	0,55	3	65	2	0,50	0,05	0,37
Chr11_3568412	0,58	3	65	2	0,49	0,03	0,37
Chr11_3594150	0,72	2	60	2	0,41	0,00	0,32
Chr11_4332466	0,78	3	65	2	0,35	0,02	0,29
Chr11_4353161	0,85	3	65	2	0,25	0,02	0,22
Chr11_4373256	0,66	3	65	2	0,45	0,03	0,35
Chr11_5458459	0,68	3	65	2	0,44	0,06	0,34
Chr11_7209046	0,55	2	65	2	0,49	0,00	0,37
Chr11_7344165	0,64	3	62	2	0,46	0,02	0,36
Chr11_7977979	0,88	2	64	2	0,22	0,00	0,19
Chr11_9011218	0,67	3	63	2	0,44	0,05	0,34
Chr11_9073778	0,68	3	65	2	0,44	0,03	0,34
Chr11_9092055	0,69	3	65	2	0,43	0,03	0,34
Chr11_9731974	0,52	3	65	2	0,50	0,03	0,37
Chr12_10061622	0,84	2	65	2	0,27	0,32	0,23
Chr12_10122138	0,58	3	65	2	0,49	0,02	0,37
Chr12_10217975	0,71	2	65	2	0,41	0,00	0,33
Chr12_10255735	0,61	3	61	2	0,47	0,02	0,36

Chr12_10320565	0,67	3	59	2	0,44	0,02	0,34
Chr12_10341968	0,64	2	59	2	0,46	0,00	0,35
Chr12_10392447	0,61	3	65	2	0,48	0,08	0,36
Chr12_10395341	0,58	2	65	2	0,49	0,00	0,37
Chr12_10461268	0,60	2	65	2	0,48	0,00	0,36
Chr12_10469880	0,59	3	64	2	0,48	0,03	0,37
Chr12_10522396	0,70	2	64	2	0,42	0,00	0,33
Chr12_10640097	0,58	3	65	2	0,49	0,03	0,37
Chr12_10665499	0,58	2	65	2	0,49	0,00	0,37
Chr12_10761460	0,58	2	65	2	0,49	0,00	0,37
Chr12_10796464	0,58	3	65	2	0,49	0,02	0,37
Chr12_10831750	0,75	2	64	2	0,38	0,00	0,30
Chr12_10862056	0,62	3	65	2	0,47	0,02	0,36
Chr12_10901418	0,63	2	65	2	0,47	0,00	0,36
Chr12_10960570	0,73	3	65	2	0,39	0,02	0,32
Chr12_11050628	0,64	3	65	2	0,46	0,02	0,36
Chr12_11195005	0,60	3	65	2	0,48	0,03	0,36
Chr12_11250814	0,62	2	65	2	0,47	0,00	0,36
Chr12_11292981	0,62	3	65	2	0,47	0,05	0,36
Chr12_11346394	0,74	2	65	2	0,39	0,00	0,31
Chr12_11373615	0,75	3	65	2	0,38	0,02	0,31
Chr12_11580708	0,74	3	64	2	0,38	0,02	0,31
Chr12_11701306	0,66	3	65	2	0,45	0,62	0,35
Chr12_11786910	0,63	2	64	2	0,47	0,00	0,36
Chr12_11920597	0,73	3	65	2	0,39	0,02	0,32
Chr12_12085224	0,65	2	62	2	0,46	0,00	0,35
Chr12_12104433	0,66	3	63	2	0,45	0,02	0,35
Chr12_12131927	0,75	2	64	2	0,38	0,00	0,30

Chr12_12202599	0,70	3	65	2	0,42	0,02	0,33
Chr12_12277570	0,71	2	65	2	0,41	0,00	0,33
Chr12_12311981	0,70	3	65	2	0,42	0,02	0,33
Chr12_12397733	0,61	3	65	2	0,48	0,02	0,36
Chr12_12420891	0,70	3	65	2	0,42	0,02	0,33
Chr12_12457628	0,63	2	64	2	0,47	0,00	0,36
Chr12_12527055	0,65	3	65	2	0,46	0,18	0,35
Chr12_12695348	0,71	2	65	2	0,41	0,00	0,33
Chr12_12789673	0,70	3	65	2	0,42	0,02	0,33
Chr12_12815390	0,61	3	65	2	0,48	0,02	0,36
Chr12_12844381	0,62	2	65	2	0,47	0,00	0,36
Chr12_13010363	0,69	2	65	2	0,43	0,62	0,34
Chr12_13094813	0,71	2	65	2	0,41	0,00	0,33
Chr12_13434729	0,72	3	65	2	0,41	0,02	0,32
Chr12_13478115	0,72	2	65	2	0,40	0,00	0,32
Chr12_13494617	0,68	3	64	2	0,44	0,05	0,34
Chr12_13604466	0,75	2	65	2	0,37	0,00	0,30
Chr12_13695791	0,61	3	65	2	0,48	0,02	0,36
Chr12_13759597	0,71	2	65	2	0,41	0,00	0,33
Chr12_13845714	0,62	2	65	2	0,47	0,00	0,36
Chr12_13883943	0,70	3	65	2	0,42	0,02	0,33
Chr12_13946832	0,62	2	65	2	0,47	0,00	0,36
Chr12_13950048	0,61	3	65	2	0,48	0,02	0,36
Chr12_13987423	0,69	2	65	2	0,43	0,00	0,34
Chr12_14029565	0,68	3	65	2	0,44	0,03	0,34
Chr12_14046810	0,62	2	65	2	0,47	0,00	0,36
Chr12_14086939	0,70	3	65	2	0,42	0,05	0,33
Chr12_14160817	0,74	3	64	2	0,38	0,02	0,31

Chr12_14165906	0,77	3	64	2	0,36	0,03	0,29
Chr12_14220276	0,68	3	65	2	0,43	0,02	0,34
Chr12_14239375	0,66	3	64	2	0,45	0,02	0,35
Chr12_1433582	0,76	3	57	2	0,36	0,02	0,30
Chr12_14348122	0,81	2	58	2	0,31	0,00	0,26
Chr12_14468229	0,85	2	65	2	0,26	0,00	0,23
Chr12_14499933	0,78	2	65	2	0,34	0,00	0,28
Chr12_14693245	0,89	2	65	2	0,19	0,00	0,17
Chr12_14739304	0,60	2	65	2	0,48	0,80	0,36
Chr12_14748134	0,89	2	65	2	0,19	0,00	0,17
Chr12_14771715	0,94	2	65	2	0,12	0,12	0,11
Chr12_14772638	0,88	2	59	2	0,21	0,00	0,19
Chr12_14848121	0,89	2	65	2	0,19	0,00	0,17
Chr12_14889935	0,89	2	65	2	0,19	0,00	0,17
Chr12_14900289	0,88	3	65	2	0,20	0,02	0,18
Chr12_14951694	0,87	2	54	2	0,23	0,00	0,20
Chr12_15008734	0,89	2	64	2	0,19	0,00	0,18
Chr12_15039997	0,90	3	64	2	0,18	0,02	0,17
Chr12_15128536	0,61	3	65	2	0,48	0,02	0,36
Chr12_15844279	0,60	2	65	2	0,48	0,00	0,36
Chr12_15862305	0,67	3	65	2	0,44	0,02	0,34
Chr12_15893454	0,61	2	54	2	0,48	0,00	0,36
Chr12_16342124	0,94	2	65	2	0,12	0,00	0,11
Chr12_16497437	0,92	3	65	2	0,15	0,02	0,14
Chr12_16997356	0,87	3	62	2	0,22	0,03	0,20
Chr12_17274648	0,92	3	65	2	0,15	0,02	0,14
Chr12_17429998	0,76	3	65	2	0,36	0,02	0,30
Chr12_17435936	0,85	3	65	2	0,26	0,03	0,23

Chr12_1745586	0,81	3	65	2	0,31	0,02	0,26
Chr12_17494948	0,88	3	63	2	0,21	0,02	0,19
Chr12_17576482	0,77	3	65	2	0,36	0,03	0,29
Chr12_17611571	0,88	3	65	2	0,20	0,02	0,18
Chr12_17639213	0,85	3	65	2	0,25	0,02	0,22
Chr12_17692661	0,89	3	65	2	0,19	0,03	0,17
Chr12_17896641	0,90	3	65	2	0,18	0,02	0,16
Chr12_17914195	0,90	3	65	2	0,18	0,02	0,16
Chr12_18910429	0,95	2	65	2	0,09	0,00	0,08
Chr12_2041641	0,55	3	57	2	0,49	0,05	0,37
Chr12_21142588	0,70	3	64	2	0,42	0,03	0,33
Chr12_21260026	0,57	3	65	2	0,49	0,03	0,37
Chr12_21286804	0,73	2	65	2	0,39	0,54	0,32
Chr12_21297436	0,52	3	65	2	0,50	0,06	0,37
Chr12_21488294	0,59	3	65	2	0,48	0,08	0,37
Chr12_21500937	0,57	3	64	2	0,49	0,02	0,37
Chr12_21871704	0,69	2	65	2	0,43	0,62	0,34
Chr12_22061968	0,57	3	63	2	0,49	0,03	0,37
Chr12_23306465	0,55	3	65	2	0,50	0,05	0,37
Chr12_23319865	0,53	3	65	2	0,50	0,02	0,37
Chr12_23433733	0,54	3	60	2	0,50	0,02	0,37
Chr12_2620976	0,80	2	65	2	0,32	0,00	0,27
Chr12_3436634	0,67	3	65	2	0,44	0,05	0,34
Chr12_3500264	0,64	3	65	2	0,46	0,02	0,36
Chr12_3658931	0,97	2	65	2	0,06	0,00	0,06
Chr12_3970772	0,66	3	64	2	0,45	0,02	0,35
Chr12_4147113	0,72	2	65	2	0,40	0,00	0,32
Chr12_4194926	0,91	2	65	2	0,17	0,00	0,15

Chr12_4212494	0,72	2	65	2	0,40	0,00	0,32
Chr12_4215979	0,78	2	60	2	0,34	0,00	0,28
Chr12_4438655	0,79	3	65	2	0,33	0,02	0,27
Chr12_4477396	0,91	2	65	2	0,17	0,00	0,15
Chr12_7846049	0,90	3	57	2	0,17	0,02	0,16
Chr12_7879440	0,91	2	57	2	0,16	0,00	0,15
Chr12_7928753	0,82	3	65	2	0,29	0,02	0,25
Chr12_9340557	0,86	2	65	2	0,24	0,00	0,21
Chr12_9344536	0,65	3	65	2	0,45	0,02	0,35
Chr12_9394260	0,58	2	64	2	0,49	0,00	0,37
Chr12_9461465	0,59	2	64	2	0,48	0,00	0,37
Chr12_9498931	0,61	3	65	2	0,48	0,02	0,36
Chr12_9534086	0,55	2	65	2	0,49	0,00	0,37
Chr12_9538580	0,60	3	63	2	0,48	0,02	0,37
Chr12_9561938	0,66	2	64	2	0,45	0,00	0,35
Chr12_9638630	0,53	2	64	2	0,50	0,00	0,37
Chr12_9715551	0,63	3	60	2	0,46	0,03	0,36
Chr12_9760949	0,66	2	65	2	0,45	0,00	0,35
Chr12_9864641	0,63	2	64	2	0,47	0,00	0,36
Chr12_9870376	0,58	3	65	2	0,49	0,03	0,37
Chr12_9926782	0,83	2	65	2	0,28	0,00	0,24
Chr12_9979324	0,71	2	65	2	0,41	0,58	0,33
Chr12_9982878	0,58	2	65	2	0,49	0,00	0,37
Chr2_10273874	0,94	2	65	2	0,12	0,00	0,11
Chr2_10311118	0,97	2	65	2	0,06	0,06	0,06
Chr2_11188067	0,93	3	65	2	0,13	0,02	0,12
Chr2_11317462	0,80	2	65	2	0,32	0,00	0,27
Chr2_15275990	0,51	2	65	2	0,50	0,00	0,37

Chr2_16252795	0,85	2	65	2	0,26	0,00	0,23
Chr2_16398811	0,75	3	65	2	0,38	0,11	0,31
Chr2_16407237	0,83	2	65	2	0,28	0,00	0,24
Chr2_16905297	0,51	2	65	2	0,50	0,00	0,37
Chr2_18044260	0,75	3	65	2	0,37	0,03	0,30
Chr2_19156464	0,73	3	65	2	0,39	0,02	0,32
Chr2_19191264	0,72	3	65	2	0,41	0,05	0,32
Chr2_20343091	0,88	2	65	2	0,22	0,00	0,19
Chr2_20508336	0,79	3	65	2	0,33	0,02	0,27
Chr2_20527179	0,69	2	65	2	0,43	0,00	0,34
Chr2_20589489	0,71	3	65	2	0,41	0,03	0,33
Chr2_20601159	0,83	2	65	2	0,28	0,00	0,24
Chr2_21344613	0,75	2	65	2	0,37	0,00	0,30
Chr2_26045482	0,53	3	65	2	0,50	0,02	0,37
Chr2_26412422	0,68	3	65	2	0,44	0,03	0,34
Chr2_28236755	0,69	2	65	2	0,43	0,00	0,34
Chr2_28290282	0,61	3	65	2	0,48	0,08	0,36
Chr2_8016075	0,72	2	65	2	0,40	0,00	0,32
Chr2_8166854	0,86	2	65	2	0,24	0,00	0,21
Chr2_8192372	0,77	2	65	2	0,36	0,00	0,29
Chr2_8407535	0,86	2	65	2	0,24	0,00	0,21
Chr2_8410677	0,86	2	65	2	0,24	0,00	0,21
Chr3_21240205	0,85	2	65	2	0,26	0,00	0,23
Chr3_21456335	0,77	2	65	2	0,36	0,00	0,29
Chr3_21760722	0,75	3	65	2	0,38	0,02	0,31
Chr3_22408454	0,61	3	65	2	0,48	0,02	0,36
Chr3_22525694	0,56	3	60	2	0,49	0,08	0,37
Chr3_23666632	0,74	3	64	2	0,38	0,05	0,31

Chr3_25082494	0,89	2	65	2	0,19	0,00	0,17
Chr3_25252811	0,92	3	65	2	0,15	0,02	0,14
Chr3_25298901	0,95	2	65	2	0,09	0,00	0,08
Chr3_25337804	0,85	2	65	2	0,26	0,00	0,23
Chr3_25367581	0,94	2	65	2	0,12	0,12	0,11
Chr3_25460399	0,88	3	63	2	0,21	0,02	0,19
Chr3_25461760	0,85	2	65	2	0,26	0,00	0,23
Chr3_25469584	0,85	2	65	2	0,26	0,00	0,23
Chr3_25514669	0,85	2	65	2	0,26	0,00	0,23
Chr3_25553010	0,92	2	65	2	0,14	0,00	0,13
Chr3_26035785	0,86	2	65	2	0,24	0,28	0,21
Chr3_29673798	0,56	3	65	2	0,49	0,02	0,37
Chr3_30325078	0,54	2	57	2	0,50	0,00	0,37
Chr3_30795599	0,55	3	65	2	0,50	0,02	0,37
Chr3_32585678	0,85	3	65	2	0,26	0,03	0,23
Chr3_34008736	0,55	3	58	2	0,49	0,07	0,37
Chr3_34733242	0,63	3	65	2	0,47	0,03	0,36
Chr3_34744057	0,56	3	57	2	0,49	0,04	0,37
Chr3_34817453	0,53	2	58	2	0,50	0,00	0,37
Chr3_35285634	0,88	3	65	2	0,20	0,02	0,18
Chr3_35380640	0,92	2	65	2	0,14	0,00	0,13
Chr3_35643104	0,67	3	65	2	0,44	0,05	0,34
Chr3_4690701	0,55	2	56	2	0,49	0,00	0,37
Chr3_6184520	0,55	3	56	2	0,49	0,11	0,37
Chr3_6423861	0,63	2	65	2	0,47	0,00	0,36
Chr3_7953743	0,51	3	65	2	0,50	0,03	0,37
Chr3_8217923	0,58	2	65	2	0,49	0,00	0,37
Chr3_8652542	0,52	3	65	2	0,50	0,02	0,37

Chr3_9318774	0,75	2	65	2	0,37	0,00	0,30
Chr3_9450608	0,82	3	65	2	0,30	0,06	0,26
Chr4_14554232	0,72	2	64	2	0,40	0,00	0,32
Chr4_14811473	0,84	3	65	2	0,27	0,02	0,23
Chr4_15955626	0,80	3	63	2	0,32	0,02	0,27
Chr4_17006154	0,84	3	65	2	0,27	0,17	0,23
Chr4_17049568	0,60	2	65	2	0,48	0,00	0,36
Chr4_25431506	0,62	3	65	2	0,47	0,02	0,36
Chr4_28848232	0,54	3	65	2	0,50	0,03	0,37
Chr4_30176006	0,68	3	65	2	0,43	0,02	0,34
Chr4_30261133	0,72	3	65	2	0,41	0,02	0,32
Chr4_30276779	0,70	3	65	2	0,42	0,02	0,33
Chr4_30712250	0,61	3	65	2	0,48	0,08	0,36
Chr4_31953215	0,59	3	65	2	0,48	0,02	0,37
Chr4_32015102	0,55	3	65	2	0,50	0,05	0,37
Chr4_32053366	0,57	2	65	2	0,49	0,00	0,37
Chr4_32515050	0,61	3	63	2	0,48	0,02	0,36
Chr4_32558753	0,62	2	65	2	0,47	0,00	0,36
Chr4_32566460	0,60	3	65	2	0,48	0,03	0,36
Chr4_32606397	0,60	2	65	2	0,48	0,00	0,36
Chr5_10355428	0,65	3	65	2	0,45	0,02	0,35
Chr5_10774997	0,53	3	65	2	0,50	0,05	0,37
Chr5_1783164	0,52	3	65	2	0,50	0,09	0,37
Chr5_18847550	0,57	3	65	2	0,49	0,03	0,37
Chr5_18913012	0,53	3	63	2	0,50	0,02	0,37
Chr5_18946012	0,55	3	65	2	0,50	0,02	0,37
Chr5_19012922	0,55	3	63	2	0,50	0,02	0,37
Chr5_19037591	0,55	3	65	2	0,49	0,03	0,37

Chr5_19070959	0,55	3	65	2	0,50	0,02	0,37
Chr5_19072226	0,54	3	65	2	0,50	0,03	0,37
Chr5_19121218	0,53	3	65	2	0,50	0,02	0,37
Chr5_19135141	0,52	3	65	2	0,50	0,08	0,37
Chr5_19156139	0,53	3	65	2	0,50	0,02	0,37
Chr5_19170949	0,51	3	65	2	0,50	0,06	0,37
Chr5_19209308	0,52	3	64	2	0,50	0,02	0,37
Chr5_19239979	0,52	3	65	2	0,50	0,02	0,37
Chr5_19278004	0,70	3	65	2	0,42	0,02	0,33
Chr5_19349214	0,55	3	65	2	0,49	0,03	0,37
Chr5_212132	0,82	3	58	2	0,30	0,05	0,25
Chr5_21579071	0,53	3	65	2	0,50	0,05	0,37
Chr5_21635833	0,62	2	65	2	0,47	0,00	0,36
Chr5_2670803	0,56	3	65	2	0,49	0,02	0,37
Chr5_2814809	0,55	3	65	2	0,50	0,05	0,37
Chr5_28469441	0,65	3	65	2	0,46	0,03	0,35
Chr5_28477355	0,58	2	65	2	0,49	0,00	0,37
Chr5_28547843	0,57	2	65	2	0,49	0,00	0,37
Chr5_28576950	0,62	3	65	2	0,47	0,05	0,36
Chr5_28617101	0,57	2	65	2	0,49	0,00	0,37
Chr5_28647080	0,58	3	65	2	0,49	0,06	0,37
Chr5_28661593	0,57	2	65	2	0,49	0,00	0,37
Chr5_28700378	0,57	3	65	2	0,49	0,03	0,37
Chr5_28723244	0,57	2	65	2	0,49	0,00	0,37
Chr5_28759040	0,58	3	65	2	0,49	0,05	0,37
Chr5_28761699	0,58	2	64	2	0,49	0,00	0,37
Chr5_28920719	0,73	3	65	2	0,39	0,02	0,32
Chr5_29578251	0,55	2	65	2	0,49	0,00	0,37

Chr5_4347955	0,51	2	65	2	0,50	0,00	0,37
Chr5_4564362	0,55	3	65	2	0,49	0,03	0,37
Chr5_4795370	0,84	3	65	2	0,27	0,02	0,23
Chr5_4817119	0,86	2	65	2	0,24	0,00	0,21
Chr5_488532	0,66	3	65	2	0,45	0,03	0,35
Chr5_607720	0,80	3	65	2	0,32	0,03	0,27
Chr5_6390749	0,69	2	64	2	0,43	0,00	0,34
Chr5_6437282	0,55	2	65	2	0,49	0,00	0,37
Chr5_6448939	0,55	3	65	2	0,50	0,02	0,37
Chr5_739182	0,81	3	61	2	0,31	0,02	0,26
Chr5_9124745	0,51	3	64	2	0,50	0,02	0,37
Chr5_9651575	0,68	3	65	2	0,44	0,06	0,34
Chr6_113705	0,55	3	60	2	0,50	0,07	0,37
Chr6_1423131	0,58	2	65	2	0,49	0,00	0,37
Chr6_1431533	0,61	3	64	2	0,48	0,03	0,36
Chr6_1451810	0,58	2	65	2	0,49	0,00	0,37
Chr6_1516624	0,59	3	65	2	0,48	0,02	0,37
Chr6_1648271	0,55	3	65	2	0,49	0,03	0,37
Chr6_2254306	0,62	3	65	2	0,47	0,05	0,36
Chr6_23881784	0,52	3	65	2	0,50	0,03	0,37
Chr6_23919266	0,61	3	65	2	0,48	0,02	0,36
Chr6_23965052	0,63	2	65	2	0,47	0,00	0,36
Chr6_23993741	0,55	2	62	2	0,50	0,00	0,37
Chr6_24022203	0,54	3	63	2	0,50	0,03	0,37
Chr6_24079888	0,63	2	65	2	0,47	0,00	0,36
Chr6_24104114	0,54	3	65	2	0,50	0,03	0,37
Chr6_24136792	0,51	2	65	2	0,50	0,00	0,37
Chr6_24160787	0,56	3	65	2	0,49	0,02	0,37

Chr6_24221194	0,51	2	65	2	0,50	0,00	0,37
Chr6_24269866	0,50	3	65	2	0,50	0,02	0,38
Chr6_24349173	0,60	2	65	2	0,48	0,00	0,36
Chr6_24423272	0,54	2	65	2	0,50	0,00	0,37
Chr6_24461748	0,52	3	65	2	0,50	0,02	0,37
Chr6_24500380	0,56	2	64	2	0,49	0,00	0,37
Chr6_24541268	0,55	2	65	2	0,49	0,00	0,37
Chr6_24575902	0,63	2	65	2	0,47	0,00	0,36
Chr6_24628088	0,69	3	65	2	0,43	0,28	0,34
Chr6_24656852	0,63	2	65	2	0,47	0,00	0,36
Chr6_24700648	0,63	2	65	2	0,47	0,00	0,36
Chr6_25179399	0,55	2	65	2	0,49	0,00	0,37
Chr6_30920215	0,79	3	65	2	0,33	0,02	0,27
Chr6_30956408	0,82	2	65	2	0,30	0,00	0,26
Chr6_31002977	0,80	2	65	2	0,32	0,00	0,27
Chr6_31023148	0,80	2	65	2	0,32	0,00	0,27
Chr6_31152322	0,78	3	65	2	0,35	0,02	0,29
Chr6_31185457	0,78	3	65	2	0,35	0,05	0,29
Chr6_857879	0,60	2	65	2	0,48	0,00	0,36
Chr6_870656	0,62	3	65	2	0,47	0,08	0,36
Chr6_871216	0,68	2	65	2	0,44	0,00	0,34
Chr7_1293892	0,85	2	65	2	0,26	0,00	0,23
Chr7_15242701	0,98	2	62	2	0,03	0,00	0,03
Chr7_16555976	0,65	2	65	2	0,46	0,71	0,35
Chr7_1694721	0,67	3	65	2	0,44	0,02	0,34
Chr7_17462737	0,77	3	65	2	0,36	0,03	0,29
Chr7_17473541	0,77	3	65	2	0,36	0,03	0,29
Chr7_17813875	0,87	3	65	2	0,23	0,02	0,20

Chr7_17830135	0,65	3	65	2	0,45	0,45	0,35
Chr7_18108773	0,53	3	65	2	0,50	0,08	0,37
Chr7_18119507	0,60	3	61	2	0,48	0,02	0,37
Chr7_18508003	0,56	3	65	2	0,49	0,05	0,37
Chr7_18656825	0,64	3	65	2	0,46	0,05	0,36
Chr7_18698410	0,70	3	65	2	0,42	0,11	0,33
Chr7_18897918	0,55	3	65	2	0,49	0,03	0,37
Chr7_19028604	0,67	3	65	2	0,44	0,08	0,34
Chr7_19049590	0,57	3	65	2	0,49	0,03	0,37
Chr7_19162041	0,64	3	65	2	0,46	0,05	0,36
Chr7_19188931	0,64	3	60	2	0,46	0,08	0,35
Chr7_19231002	0,67	3	65	2	0,44	0,05	0,34
Chr7_19245970	0,65	3	65	2	0,45	0,08	0,35
Chr7_19305523	0,68	3	64	2	0,44	0,05	0,34
Chr7_19378781	0,68	3	65	2	0,44	0,09	0,34
Chr7_19416130	0,66	3	65	2	0,45	0,03	0,35
Chr7_19496801	0,58	3	65	2	0,49	0,06	0,37
Chr7_19529846	0,70	3	65	2	0,42	0,05	0,33
Chr7_19556942	0,68	3	65	2	0,44	0,06	0,34
Chr7_19609132	0,70	3	65	2	0,42	0,05	0,33
Chr7_19625790	0,69	3	65	2	0,43	0,06	0,34
Chr7_1974932	0,95	2	65	2	0,09	0,00	0,08
Chr7_20586169	0,80	3	65	2	0,32	0,06	0,27
Chr7_20607220	0,55	3	64	2	0,49	0,02	0,37
Chr7_20653073	0,51	3	65	2	0,50	0,12	0,37
Chr7_20677948	0,56	3	65	2	0,49	0,02	0,37
Chr7_20688426	0,52	3	65	2	0,50	0,03	0,37
Chr7_20734668	0,58	2	65	2	0,49	0,00	0,37

Chr7_20770081	0,50	3	65	2	0,50	0,08	0,38
Chr7_20790690	0,53	3	65	2	0,50	0,02	0,37
Chr7_20843856	0,50	3	65	2	0,50	0,05	0,38
Chr7_20906840	0,52	3	64	2	0,50	0,02	0,37
Chr7_20950472	0,59	3	65	2	0,48	0,05	0,37
Chr7_20969923	0,55	2	65	2	0,49	0,00	0,37
Chr7_21028789	0,58	3	65	2	0,49	0,03	0,37
Chr7_21075149	0,51	2	57	2	0,50	0,00	0,37
Chr7_21134321	0,52	2	65	2	0,50	0,00	0,37
Chr7_21149256	0,56	3	65	2	0,49	0,02	0,37
Chr7_21214128	0,77	3	65	2	0,36	0,03	0,29
Chr7_21220581	0,85	3	65	2	0,26	0,03	0,23
Chr7_21274715	0,78	3	64	2	0,34	0,03	0,28
Chr7_21314665	0,79	3	65	2	0,33	0,02	0,27
Chr7_22492679	0,76	3	65	2	0,36	0,02	0,30
Chr7_22549841	0,80	3	62	2	0,32	0,02	0,27
Chr7_22592770	0,76	3	65	2	0,36	0,02	0,30
Chr7_22645313	0,76	3	65	2	0,36	0,02	0,30
Chr7_23236831	0,70	3	65	2	0,42	0,02	0,33
Chr7_3126071	0,82	3	65	2	0,29	0,02	0,25
Chr7_3250498	0,75	2	65	2	0,37	0,00	0,30
Chr7_3963635	0,70	3	64	2	0,42	0,03	0,33
Chr7_4250127	0,97	2	65	2	0,06	0,00	0,06
Chr7_4267087	0,88	2	65	2	0,22	0,00	0,19
Chr7_4292360	0,95	2	59	2	0,10	0,00	0,09
Chr7_4333082	0,95	2	64	2	0,09	0,00	0,09
Chr7_4368717	0,95	2	65	2	0,09	0,00	0,08
Chr7_4389114	0,86	2	65	2	0,24	0,00	0,21

Chr7_4444412	0,95	2	65	2	0,09	0,00	0,08
Chr7_4446115	0,97	2	59	2	0,05	0,05	0,05
Chr7_4458778	0,95	2	65	2	0,09	0,00	0,08
Chr7_4509662	0,95	2	65	2	0,09	0,00	0,08
Chr7_4561734	0,97	2	65	2	0,06	0,00	0,06
Chr7_4568760	0,96	3	65	2	0,07	0,02	0,07
Chr7_4713093	0,94	2	64	2	0,12	0,00	0,11
Chr7_4756028	0,57	2	65	2	0,49	0,00	0,37
Chr7_6985139	0,63	3	58	2	0,47	0,02	0,36
Chr7_7267127	0,98	2	65	2	0,03	0,00	0,03
Chr7_7908930	0,91	3	58	2	0,17	0,02	0,16
Chr8_136903	0,62	3	65	2	0,47	0,02	0,36
Chr8_16339846	0,50	2	64	2	0,50	0,00	0,38
Chr8_16344216	0,51	3	65	2	0,50	0,06	0,37
Chr8_16842367	0,58	3	60	2	0,49	0,02	0,37
Chr8_16867954	0,74	3	60	2	0,38	0,35	0,31
Chr8_16882010	0,73	2	64	2	0,40	0,55	0,32
Chr8_16974973	0,60	3	60	2	0,48	0,07	0,36
Chr8_17156701	0,55	3	65	2	0,50	0,05	0,37
Chr8_17228130	0,55	3	65	2	0,50	0,02	0,37
Chr8_173704	0,51	2	65	2	0,50	0,00	0,37
Chr8_17533934	0,62	3	65	2	0,47	0,05	0,36
Chr8_17553509	0,52	3	65	2	0,50	0,02	0,37
Chr8_17623639	0,68	3	65	2	0,44	0,06	0,34
Chr8_19459381	0,62	3	65	2	0,47	0,02	0,36
Chr8_20073861	0,57	3	65	2	0,49	0,03	0,37
Chr8_20369	0,62	3	65	2	0,47	0,02	0,36
Chr8_205404	0,66	3	65	2	0,45	0,06	0,35

Chr8_21504738	0,52	3	65	2	0,50	0,02	0,37
Chr8_23071056	0,59	3	65	2	0,48	0,02	0,37
Chr8_24725921	0,88	2	65	2	0,22	0,00	0,19
Chr8_24756144	0,87	3	65	2	0,23	0,02	0,20
Chr8_2476886	0,85	2	65	2	0,26	0,00	0,23
Chr8_24793324	0,90	3	65	2	0,18	0,02	0,16
Chr8_24829288	0,88	2	65	2	0,22	0,00	0,19
Chr8_24854148	0,87	3	65	2	0,23	0,02	0,20
Chr8_24868153	0,91	2	65	2	0,17	0,00	0,15
Chr8_24908113	0,94	2	65	2	0,12	0,00	0,11
Chr8_25640586	0,62	2	60	2	0,47	0,00	0,36
Chr8_266125	0,58	3	65	2	0,49	0,02	0,37
Chr8_3001294	0,60	2	62	2	0,48	0,00	0,37
Chr8_311834	0,51	2	65	2	0,50	0,00	0,37
Chr8_331937	0,65	2	62	2	0,46	0,00	0,35
Chr8_344452	0,65	2	65	2	0,46	0,00	0,35
Chr8_367705	0,53	3	65	2	0,50	0,02	0,37
Chr8_399380	0,70	3	65	2	0,42	0,08	0,33
Chr8_404746	0,53	3	65	2	0,50	0,02	0,37
Chr8_4136020	0,51	2	65	2	0,50	0,00	0,37
Chr8_5134106	0,69	3	65	2	0,43	0,03	0,34
Chr8_5169779	0,74	2	62	2	0,38	0,00	0,31
Chr8_5186293	0,63	3	65	2	0,47	0,03	0,36
Chr8_5208685	0,75	3	65	2	0,38	0,02	0,31
Chr8_5227340	0,74	3	65	2	0,39	0,03	0,31
Chr8_5254167	0,75	3	65	2	0,38	0,02	0,31
Chr8_5284362	0,74	3	65	2	0,39	0,03	0,31
Chr8_5322211	0,76	3	65	2	0,36	0,02	0,30

Chr8_5947998	0,55	2	60	2	0,50	0,00	0,37
Chr8_6053118	0,65	3	65	2	0,45	0,02	0,35
Chr8_6119160	0,57	2	61	2	0,49	0,00	0,37
Chr8_6465504	0,70	2	60	2	0,42	0,00	0,33
Chr8_66412	0,64	3	65	2	0,46	0,05	0,36
Chr8_7506944	0,61	3	63	2	0,48	0,02	0,36
Chr8_7565092	0,72	2	65	2	0,40	0,00	0,32
Chr8_7949448	0,66	3	65	2	0,45	0,03	0,35
Chr8_7973169	0,66	2	65	2	0,45	0,00	0,35
Chr8_80456	0,63	2	65	2	0,47	0,00	0,36
Chr8_8110816	0,66	2	65	2	0,45	0,00	0,35
Chr8_8119838	0,66	3	65	2	0,45	0,03	0,35
Chr8_8153751	0,66	2	65	2	0,45	0,00	0,35
Chr8_8433342	0,65	3	65	2	0,45	0,02	0,35
Chr8_8675401	0,59	2	64	2	0,48	0,00	0,37
Chr8_8774226	0,61	3	65	2	0,48	0,02	0,36
Chr8_92154	0,64	3	65	2	0,46	0,05	0,36
Chr9_21130812	0,65	2	65	2	0,46	0,00	0,35
Chr9_21145279	0,62	3	65	2	0,47	0,03	0,36
Média	0,69	2,62	64,27	2,00	0,39	0,03	0,31
Mínimo	0,50	2,00	52,00	2,00	0,03	0,00	0,03
Máximo	0,98	3,00	65,00	2,00	0,50	0,80	0,38

Anexo 2 – Parâmetros qualitativos, localização e genótipos observados em 8 acessos de arroz de um total de 768 locos SNP que compõem os dois painéis 384-plex de arroz.

	GenTrain Score	Chromosome	Position	SNP	CNS 99	CNS 101	CNS 96	CNS 97	CNS 98	CNS 100	CNS 102	CNS 103
Chr7_20515692	0,9797	7	20515692	[T/C]	BB	BB	BB	AA	AA	AA	BB	--
Chr3_21760722	0,8197	3	21760722	[A/G]	BB	BB	BB	AA	BB	BB	AA	BB
Chr3_25082494	0,8477	3	25082494	[A/G]	AA	AA	AA	BB	AA	BB	AA	AA
Chr5_19349214	0,8252	5	19349214	[G/C]	BB	BB	AA	AB	AA	BB	AA	BB
Chr6_24221194	0,9692	6	24221194	[T/C]	AA	BB	AA	AA	BB	BB	AA	BB
Chr11_18106947	0,9288	11	18106947	[T/C]	BB	BB	BB	BB	AA	AA	BB	BB
Chr1_25457890	0,8769	1	25457890	[T/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr11_17789344	0,9758	11	17789344	[C/G]	BB	BB	BB	BB	BB	BB	BB	BB
Chr12_11050628	0,7798	12	11050628	[A/G]	BB	BB	AA	BB	BB	BB	AA	BB
Chr3_6423861	0,8795	3	6423861	[T/C]	AA	AA	BB	AA	BB	BB	AA	BB
Chr12_13010363	0,5083	12	13010363	[C/G]	AB	AB	BB	AB	AB	AB	BB	AB
Chr11_18035544	0,9739	11	18035544	[T/G]	BB	BB	BB	BB	AA	AA	BB	BB
Chr2_28236755	0,8626	2	28236755	[T/C]	BB	BB	BB	AA	AA	AA	BB	BB
Chr1_2389723	0,8013	1	2389723	[T/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr12_10901418	0,9544	12	10901418	[T/G]	AA	AA	BB	AA	AA	AA	BB	AA
Chr8_136903	0,7565	8	136903	[A/G]	BB	BB	AA	AA	AA	AA	BB	AA
Chr1_37707861	0,971	1	37707861	[A/G]	BB	AA	BB	AA	AA	AA	BB	BB
Chr12_14693245	0,9162	12	14693245	[T/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr10_13116751	0,8361	10	13116751	[A/G]	AA	AA	AA	BB	BB	BB	AA	AA
Chr1_2660700	0,715	1	2660700	[T/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr3_9318774	0,67	3	9318774	[T/C]	AA	BB	BB	BB	AA	BB	AA	BB
Chr12_15844279	0,8906	12	15844279	[T/C]	BB	BB	AA	BB	AA	BB	AA	AA
Chr8_8153751	0,952	8	8153751	[T/C]	AA	BB	AA	BB	BB	AA	BB	AA

Chr11_20771608	0,7797	11	20771608	[T/C]	BB	AA	AA	AA	AA	BB	AA	BB
Chr5_28617101	0,8306	5	28617101	[A/G]	BB	BB	BB	BB	BB	AA	BB	AA
Chr1_23940204	0,3292	1	23940204	[T/C]	BB	--	AB	AA	AB	AA	BB	BB
Chr11_19890334	0,6593	11	19890334	[T/C]	BB	BB	AB	BB	AA	AA	BB	BB
Chr11_27535474	0,8093	11	27535474	[A/G]	AA	BB	BB	AA	AA	AA	BB	BB
Chr1_31796289	0,9711	1	31796289	[T/G]	AA	AA	AA	AA	BB	BB	AA	AA
Chr12_13478115	0,8593	12	13478115	[T/C]	BB	BB	AA	BB	BB	BB	AA	BB
Chr8_5208685	0,8685	8	5208685	[A/G]	BB	BB	AA	AA	AA	AA	BB	BB
Chr11_19849611	0,6864	11	19849611	[A/G]	BB	BB	BB	BB	--	AA	BB	BB
Chr12_17274648	0,7789	12	17274648	[A/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr12_12085224	0,8821	12	12085224	[T/C]	AA	AA	BB	AA	AA	AA	BB	AA
Chr6_23993741	0,8819	6	23993741	[T/G]	AA	AA	AA	AA	BB	BB	AA	BB
Chr1_12867748	0,7577	1	12867748	[A/G]	BB	AA	AA	AA	BB	AA	BB	AA
Chr4_30176006	0,8494	4	30176006	[A/G]	BB	AA	BB	AA	BB	BB	AA	BB
Chr12_13604466	0,6351	12	13604466	[A/G]	BB	BB	AA	BB	BB	BB	AA	BB
Chr10_11920922	0,833	10	11920922	[T/A]	AA	AA	AA	BB	AA	BB	AA	AA
Chr12_13946832	0,7158	12	13946832	[T/A]	BB	BB	AA	BB	BB	BB	AA	BB
Chr8_17553509	0,8288	8	17553509	[T/C]	BB	AA	AA	AA	AA	BB	AA	BB
Chr7_18897918	0,6471	7	18897918	[T/G]	AA	AA	BB	AA	AA	BB	AA	AA
Chr12_4473647	0,7182	12	4473647	[A/G]	BB	BB	BB	AA	AA	AA	BB	BB
Chr2_8410677	0,7318	2	8410677	[T/C]	BB	BB	BB	AA	AA	AA	BB	BB
Chr6_24575902	0,8305	6	24575902	[T/G]	BB	BB	AA	BB	BB	AA	BB	AA
Chr12_17611571	0,7743	12	17611571	[A/T]	BB	BB	BB	AA	BB	AA	BB	BB
Chr1_42491941	0,8487	1	42491941	[A/C]	AA	--	AA	BB	AA	AA	BB	BB
Chr4_25431506	0,8211	4	25431506	[A/G]	AA	AA	BB	AA	BB	BB	AA	BB
Chr12_12397733	0,8447	12	12397733	[T/G]	BB	BB	AA	BB	BB	BB	AA	BB
Chr12_9982878	0,7825	12	9982878	[T/G]	BB	BB	AA	BB	AA	BB	AA	AA
Chr11_17023333	0,6545	11	17023333	[A/G]	AA	AA	AA	AA	--	--	AA	--

Chr1_7443772	0,854	1	7443772	[T/C]	AA	BB	BB	BB	BB	BB	AA	AA
Chr10_22488323	0,9691	10	22488323	[A/G]	BB	BB	AA	AA	AA	AA	BB	BB
Chr12_9340557	0,8117	12	9340557	[A/G]	BB	BB	BB	AA	AA	AA	BB	BB
Chr12_11346394	0,8998	12	11346394	[T/C]	BB	BB	AA	BB	BB	BB	AA	BB
Chr5_488532	0,8436	5	488532	[C/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr8_182028	0,3619	8	182028	[A/G]	BB	AB	AA	AA	AA	AA	BB	AA
Chr12_12277570	0,8994	12	12277570	[T/G]	AA	AA	BB	AA	AA	AA	BB	AA
Chr11_19507931	0,7912	11	19507931	[T/C]	BB	BB	BB	BB	AA	AA	BB	BB
Chr5_17072463	0,6799	5	17072463	[T/G]	BB	--	AA	BB	--	BB	AA	AA
Chr12_9926782	0,8333	12	9926782	[T/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr7_22645313	0,8956	7	22645313	[A/G]	AA	BB	AA	BB	BB	BB	AA	AA
Chr1_23965910	0,8203	1	23965910	[A/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr1_24727564	0,9655	1	24727564	[T/G]	AA	AA	AA	BB	BB	BB	AA	BB
Chr7_19416130	0,725	7	19416130	[C/G]	AA	AA	BB	AA	AA	BB	AA	AA
Chr1_24147898	0,5999	1	24147898	[G/C]	AA	AA	AA	AB	AB	BB	AA	AA
Chr8_7565092	0,867	8	7565092	[T/C]	BB	AA	BB	AA	AA	BB	AA	BB
Chr6_25179399	0,8049	6	25179399	[T/C]	AA	BB	BB	AA	BB	BB	AA	BB
Chr3_26035785	0,8136	3	26035785	[T/C]	BB	AB	BB	AB	BB	BB	AB	BB
Chr7_19231002	0,7314	7	19231002	[A/G]	BB	BB	AA	BB	BB	AA	BB	BB
Chr11_27010793	0,6376	11	27010793	[G/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr1_39422020	0,9773	1	39422020	[A/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr12_17914195	0,8549	12	17914195	[A/G]	AA	AA	AA	BB	AA	BB	AA	AA
Chr6_31152322	0,8261	6	31152322	[G/C]	BB	BB	BB	AA	AA	AA	BB	AA
Chr10_16769865	0,9802	10	16769865	[A/G]	BB	AA	AA	BB	AA	BB	AA	AA
Chr11_18230919	0,8305	11	18230919	[T/C]	BB	BB	BB	BB	AA	AA	BB	BB
Chr12_13987423	0,7468	12	13987423	[A/G]	BB	BB	AA	BB	BB	BB	AA	BB
Chr7_21149256	0,86	7	21149256	[T/G]	BB	AA	BB	AA	AA	AA	BB	AA
Chr2_16905297	0,8707	2	16905297	[T/G]	AA	AA	AA	BB	BB	AA	BB	BB

Chr7_19162041	0,8556	7	19162041	[T/C]	AA	AA	BB	AA	AA	BB	AA	AA
Chr11_22335713	0,9609	11	22335713	[T/C]	BB	AA	AA	BB	AA	BB	AA	BB
Chr12_14046810	0,7405	12	14046810	[A/G]	BB	BB	AA	BB	BB	BB	AA	BB
Chr8_80456	0,6519	8	80456	[A/G]	AA	AA	BB	BB	BB	BB	AA	BB
Chr1_35030275	0,97	1	35030275	[T/G]	AA	AA	BB	AA	BB	AA	BB	BB
Chr6_24136792	0,8396	6	24136792	[T/C]	AA	BB	AA	AA	BB	BB	AA	BB
Chr1_7493335	0,8209	1	7493335	[T/C]	AA	BB	BB	BB	BB	BB	AA	AA
Chr1_42604189	0,6391	1	42604189	[T/C]	BB	BB	BB	AB	BB	BB	AA	AA
Chr8_24782657	0,3282	8	24782657	[A/G]	BB	BB	AB	AB	AA	--	BB	BB
Chr8_331937	0,7083	8	331937	[T/G]	BB	BB	AA	AA	AA	AA	BB	AA
Chr7_4250127	0,9136	7	4250127	[T/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr1_29821621	0,6895	1	29821621	[G/C]	AA	BB	AA	AA	AA	AA	BB	AA
Chr3_34733242	0,7021	3	34733242	[A/G]	BB	BB	BB	AA	AA	AA	BB	BB
Chr7_1293892	0,9789	7	1293892	[A/G]	AA	AA	AA	BB	AA	BB	AA	AA
Chr7_22492679	0,6863	7	22492679	[A/C]	AA	BB	AA	BB	BB	BB	AA	AA
Chr12_9534086	0,8947	12	9534086	[A/G]	AA	AA	BB	AA	BB	AA	BB	BB
Chr12_12457628	0,8374	12	12457628	[A/G]	AA	AA	BB	AA	AA	AA	BB	AA
Chr11_21535425	0,8395	11	21535425	[T/C]	AA	BB	BB	BB	--	AA	BB	AA
Chr6_1423131	0,9603	6	1423131	[A/G]	AA	AA	BB	BB	BB	BB	AA	BB
Chr10_13206132	0,8199	10	13206132	[T/C]	BB	BB	BB	AA	AA	AA	BB	BB
Chr5_10355428	0,853	5	10355428	[T/C]	BB	BB	AA	BB	BB	BB	AA	AA
Chr12_15039997	0,6567	12	15039997	[A/G]	BB	BB	BB	AB	BB	AA	BB	BB
Chr12_9864641	0,6745	12	9864641	[A/G]	BB	BB	AA	BB	AA	BB	AA	AA
Chr11_17836032	0,7971	11	17836032	[T/C]	AA	AA	AA	AA	BB	BB	AA	BB
Chr6_24079888	0,8769	6	24079888	[A/G]	AA	AA	AA	AA	BB	BB	AA	BB
Chr11_17116084	0,8508	11	17116084	[T/C]	AA	AA	AA	AA	--	--	AA	--
Chr10_12019694	0,8126	10	12019694	[A/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr5_19012922	0,838	5	19012922	[A/G]	BB	BB	AA	BB	BB	BB	AA	BB

Chr1_11272869	0,8537	1	11272869	[T/C]	AA	AA	AA	AA	AA	AA	BB	AA
Chr5_4795370	0,788	5	4795370	[A/T]	AA	AA	AA	BB	AA	AA	BB	AA
Chr12_9638630	0,645	12	9638630	[A/G]	AA	AA	BB	AA	BB	AA	BB	BB
Chr11_4332466	0,7373	11	4332466	[T/C]	AA	AA	AA	AA	AA	BB	AA	AA
Chr1_12940593	0,8999	1	12940593	[A/G]	BB	AA	AA	AA	BB	AA	BB	AA
Chr10_11789711	0,8566	10	11789711	[T/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr4_17049568	0,7177	4	17049568	[A/G]	BB	BB	AA	BB	AA	AA	BB	BB
Chr5_2670803	0,8484	5	2670803	[T/C]	AA	AA	BB	BB	AA	AA	BB	AA
Chr11_17001212	0,8338	11	17001212	[T/C]	BB	BB	BB	BB	AA	AA	BB	AA
Chr1_29275933	0,8757	1	29275933	[A/T]	AA	BB	AA	AA	BB	AA	BB	AA
Chr3_8217923	0,7463	3	8217923	[A/G]	AA	BB	AA	BB	AA	AA	BB	AA
Chr3_35380640	0,9577	3	35380640	[T/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr10_19853566	0,8497	10	19853566	[G/C]	BB	BB	AA	BB	AA	AA	BB	AA
Chr3_25367581	0,5718	3	25367581	[T/A]	BB	BB	BB	AB	BB	AB	BB	BB
Chr12_17494948	0,6207	12	17494948	[A/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr1_39376380	0,9384	1	39376380	[T/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr4_15955626	0,9496	4	15955626	[T/C]	BB	BB	AA	BB	BB	AA	BB	BB
Chr1_19139384	0,9773	1	19139384	[A/G]	BB	AA	AA	BB	AA	BB	AA	AA
Chr5_19121218	0,7693	5	19121218	[A/G]	AA	AA	BB	AA	AA	AA	BB	AA
Chr4_30276779	0,6399	4	30276779	[T/C]	BB	AA	BB	AA	BB	BB	AA	BB
Chr8_16882010	0,7292	8	16882010	[T/C]	BB	AB	AB	AB	AB	BB	AB	BB
Chr12_10522396	0,9237	12	10522396	[A/G]	BB	BB	AA	BB	AA	BB	AA	AA
Chr12_10261394	0,4341	12	10261394	[T/C]	AB	AB	BB	AB	BB	AA	BB	BB
Chr7_4368717	0,8964	7	4368717	[A/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr3_25461760	0,8729	3	25461760	[T/A]	BB	BB	BB	AA	BB	AA	BB	BB
Chr1_7352717	0,8305	1	7352717	[T/C]	BB	AA	AA	AA	AA	AA	BB	BB
Chr12_14499933	0,9726	12	14499933	[A/G]	AA	AA	AA	BB	BB	BB	AA	AA
Chr12_7813541	0,6416	12	7813541	[A/G]	AA	--	--	BB	AA	BB	AA	AA

Chr12_9461465	0,7586	12	9461465	[T/C]	AA	AA	BB	AA	AA	AA	BB	BB
Chr7_19305523	0,9018	7	19305523	[A/G]	BB	BB	AA	BB	BB	AA	BB	BB
Chr8_24725921	0,8906	8	24725921	[A/G]	BB	BB	BB	BB	AA	AA	BB	BB
Chr12_22061968	0,9108	12	22061968	[A/G]	AA	AA	AA	BB	BB	BB	AA	BB
Chr12_10217975	0,8107	12	10217975	[A/G]	BB	BB	AA	BB	AA	BB	AA	AA
Chr1_23836460	0,9693	1	23836460	[T/G]	AA	AA	AA	BB	BB	BB	AA	AA
Chr1_6378875	0,6263	1	6378875	[T/C]	AA	AA	BB	BB	BB	BB	AA	BB
Chr7_20790690	0,8171	7	20790690	[A/G]	AA	--	AA	BB	BB	BB	AA	BB
Chr7_19049590	0,8901	7	19049590	[T/C]	AA	AA	BB	AA	AA	BB	AA	AA
Chr11_17969837	0,6255	11	17969837	[A/C]	BB	BB	BB	BB	AB	AA	BB	BB
Chr6_24349173	0,9739	6	24349173	[T/G]	BB	BB	BB	BB	AA	AA	BB	AA
Chr11_7977979	0,9256	11	7977979	[T/C]	AA	AA	AA	BB	BB	BB	AA	AA
Chr5_6437282	0,8761	5	6437282	[A/C]	AA	AA	BB	AA	BB	BB	AA	BB
Chr2_16398811	0,6442	2	16398811	[G/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr12_3970772	0,7841	12	3970772	[T/G]	AA	AA	BB	BB	AA	AA	BB	BB
Chr8_25640586	0,5982	8	25640586	[C/G]	AA	AA	AA	AA	BB	AA	BB	BB
Chr6_24656852	0,7884	6	24656852	[A/G]	AA	AA	BB	AA	AA	BB	AA	BB
Chr7_20734668	0,8469	7	20734668	[C/G]	AA	BB	AA	BB	BB	BB	AA	BB
Chr6_857879	0,8566	6	857879	[A/G]	AA	AA	BB	BB	BB	BB	AA	BB
Chr10_18449742	0,971	10	18449742	[A/G]	BB	AA	BB	AA	AA	BB	AA	BB
Chr1_23883508	0,8205	1	23883508	[T/C]	AA	AA	AA	BB	BB	BB	AA	AA
Chr5_4347955	0,9754	5	4347955	[G/C]	BB	AA	AA	AA	AA	BB	AA	BB
Chr8_5254167	0,6484	8	5254167	[T/C]	BB	BB	AA	AA	AA	AA	BB	BB
Chr1_24264642	0,5991	1	24264642	[A/G]	AA	AA	AA	BB	AA	BB	AA	AA
Chr10_3838047	0,8906	10	3838047	[A/G]	BB	AA	AA	BB	BB	BB	AA	BB
Chr11_16954935	0,6174	11	16954935	[A/G]	AA	AA	AA	AA	AB	BB	AA	BB
Chr5_29578251	0,8314	5	29578251	[T/C]	AA	AA	AA	AA	AA	BB	AA	AA
Chr12_10665499	0,9677	12	10665499	[A/G]	AA	AA	BB	AA	BB	AA	BB	BB

Chr10_20218042	0,8795	10	20218042	[A/G]	AA	AA	AA	BB	BB	BB	AA	BB
Chr1_24826024	0,8191	1	24826024	[T/C]	AA	AA	AA	BB	BB	BB	AA	BB
Chr12_21142588	0,7647	12	21142588	[T/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr11_17605614	0,9773	11	17605614	[T/G]	AA	AA	AA	AA	BB	BB	AA	BB
Chr8_8110816	0,9617	8	8110816	[T/A]	AA	BB	AA	BB	BB	AA	BB	AA
Chr10_16535481	0,9083	10	16535481	[C/G]	AA	BB	AA	AA	AA	AA	BB	AA
Chr12_14348122	0,8622	12	14348122	[T/C]	BB	BB	BB	AA	AA	AA	BB	BB
Chr5_21635833	0,6817	5	21635833	[A/G]	AA	BB	AA	AA	AA	AA	BB	BB
Chr6_23919266	0,8627	6	23919266	[T/C]	BB	BB	BB	BB	AA	AA	BB	AA
Chr12_12529802	0,666	12	12529802	[A/T]	BB	BB	AA	BB	BB	BB	AA	BB
Chr11_21322064	0,8183	11	21322064	[T/G]	BB	AA	AA	AA	BB	BB	AA	BB
Chr12_9760949	0,7625	12	9760949	[A/G]	BB	BB	AA	BB	BB	BB	AA	AA
Chr7_20677948	0,8155	7	20677948	[T/C]	AA	AA	AA	BB	BB	BB	AA	BB
Chr3_4690701	0,6635	3	4690701	[T/C]	BB	BB	AA	AA	BB	BB	AA	BB
Chr1_11339617	0,6194	1	11339617	[T/G]	BB	BB	BB	BB	BB	BB	AA	BB
Chr10_23051382	0,5032	10	23051382	[T/C]	AA	AA	BB	BB	BB	BB	AA	AA
Chr10_2684280	0,8429	10	2684280	[A/G]	BB	AA	AA	BB	BB	BB	AA	BB
Chr1_25266532	0,8655	1	25266532	[G/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr12_14748134	0,9005	12	14748134	[T/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr11_19953558	0,8469	11	19953558	[A/G]	BB	BB	BB	BB	BB	AA	BB	BB
Chr3_22525694	0,7817	3	22525694	[T/C]	BB	BB	BB	--	AB	--	AA	AA
Chr7_4713093	0,661	7	4713093	[A/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr12_3500264	0,8287	12	3500264	[A/G]	BB	BB	BB	AA	BB	BB	AA	BB
Chr12_11786910	0,91	12	11786910	[T/A]	BB	BB	AA	BB	BB	BB	AA	BB
Chr8_16842367	0,832	8	16842367	[T/C]	AA	--	BB	BB	BB	AA	BB	AA
Chr11_7209046	0,971	11	7209046	[A/G]	BB	BB	BB	AA	AA	AA	BB	BB
Chr3_25514669	0,9623	3	25514669	[A/G]	AA	AA	AA	BB	AA	BB	AA	AA
Chr7_3250498	0,8791	7	3250498	[T/C]	BB	BB	BB	BB	BB	BB	AA	BB

Chr8_2476886	0,9609	8	2476886	[T/G]	BB	BB	BB	AA	AA	AA	BB	BB
Chr12_14220276	0,7835	12	14220276	[A/G]	BB	BB	BB	AA	AA	AA	BB	BB
Chr3_34817453	0,9378	3	34817453	[A/G]	AA	AA	AA	BB	BB	BB	AA	AA
Chr7_21075149	0,9773	7	21075149	[T/C]	AA	BB	AA	BB	BB	BB	AA	BB
Chr7_19609132	0,9041	7	19609132	[T/C]	BB	BB	AA	BB	BB	AA	BB	BB
Chr5_19209308	0,8706	5	19209308	[A/G]	BB	BB	AA	BB	AA	BB	AA	BB
Chr8_19459381	0,8238	8	19459381	[A/G]	BB	AA	AA	AA	BB	BB	AA	BB
Chr7_15341312	0,3454	7	15341312	[A/G]	BB	BB	AB	AB	AB	AB	BB	BB
Chr8_21504738	0,6908	8	21504738	[T/C]	BB	BB	BB	BB	BB	AA	BB	AA
Chr11_9073778	0,8729	11	9073778	[C/G]	AA	BB	AA	AA	AA	AA	BB	AA
Chr2_8192372	0,8137	2	8192372	[T/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr11_21234402	0,8823	11	21234402	[A/C]	BB	AA	AA	AA	BB	BB	AA	BB
Chr1_11469994	0,7769	1	11469994	[T/C]	BB	BB	BB	BB	BB	BB	AA	BB
Chr11_24611954	0,8204	11	24611954	[T/C]	BB	AA	AA	AA	BB	--	AA	BB
Chr8_17228130	0,8445	8	17228130	[C/G]	AA	BB	BB	BB	BB	AA	BB	AA
Chr1_34761138	0,8289	1	34761138	[T/G]	AA	AA	BB	AA	AA	AA	BB	BB
Chr12_12131927	0,9135	12	12131927	[T/C]	BB	BB	AA	BB	BB	BB	AA	BB
Chr8_24829288	0,8417	8	24829288	[T/C]	BB	BB	BB	BB	AA	AA	BB	BB
Chr11_16881339	0,8872	11	16881339	[T/C]	AA	AA	AA	AA	BB	BB	AA	BB
Chr12_10395341	0,8207	12	10395341	[T/G]	AA	AA	BB	AA	BB	AA	BB	BB
Chr3_30325078	0,7077	3	30325078	[T/C]	AA	AA	AA	BB	BB	BB	AA	BB
Chr3_21240205	0,946	3	21240205	[T/C]	BB	BB	BB	AA	BB	BB	AA	BB
Chr7_4292360	0,8333	7	4292360	[T/A]	AA	AA	AA	BB	AA	BB	AA	AA
Chr2_19156464	0,8502	2	19156464	[T/C]	BB	BB	BB	BB	BB	AA	BB	AA
Chr11_17167169	0,6793	11	17167169	[A/G]	BB	BB	BB	BB	AA	AA	BB	AA
Chr12_14889935	0,8401	12	14889935	[A/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr1_42021996	0,8791	1	42021996	[T/C]	BB	AA	BB	AA	BB	BB	AA	AA
Chr12_13759597	0,9045	12	13759597	[T/C]	BB	BB	AA	BB	BB	BB	AA	BB

Chr5_19070959	0,9359	5	19070959	[T/C]	BB	BB	AA	BB	BB	BB	AA	BB
Chr7_21274715	0,937	7	21274715	[T/C]	AA	AA	AA	BB	BB	BB	AA	AA
Chr11_19713489	0,8377	11	19713489	[A/G]	BB	BB	BB	BB	BB	AA	BB	BB
Chr12_2620976	0,8622	12	2620976	[T/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr12_10461268	0,7445	12	10461268	[C/G]	AA	AA	BB	AA	BB	AA	BB	BB
Chr12_14951694	0,9545	12	14951694	[G/C]	AA	AA	--	BB	--	BB	AA	AA
Chr12_11250814	0,9276	12	11250814	[T/C]	AA	AA	BB	AA	AA	AA	BB	AA
Chr1_11426073	0,8683	1	11426073	[A/G]	AA	AA	AA	AA	AA	AA	BB	AA
Chr12_12695348	0,7335	12	12695348	[A/G]	AA	--	BB	AA	AA	AA	BB	AA
Chr1_24081577	0,7443	1	24081577	[T/C]	BB	BB	BB	AA	--	AA	BB	BB
Chr11_19651123	0,8388	11	19651123	[T/G]	AA	AA	AA	AA	BB	BB	AA	AA
Chr8_404746	0,521	8	404746	[A/G]	AA	AA	BB	BB	BB	BB	AA	BB
Chr7_7908930	0,9115	7	7908930	[T/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr2_15275990	0,9656	2	15275990	[T/C]	BB	BB	BB	AA	BB	BB	AA	BB
Chr5_9124745	0,8627	5	9124745	[A/G]	AA	AA	BB	AA	AA	AA	BB	BB
Chr12_9394260	0,9289	12	9394260	[T/C]	BB	BB	AA	BB	BB	BB	AA	AA
Chr6_31023148	0,6758	6	31023148	[A/G]	BB	BB	BB	AA	AA	AA	BB	AA
Chr11_17212089	0,8272	11	17212089	[A/C]	AA	AA	AA	AA	BB	BB	AA	BB
Chr12_23319865	0,8443	12	23319865	[A/C]	AA	BB	BB	AA	BB	AA	BB	AA
Chr8_6465504	0,8469	8	6465504	[T/C]	BB	AA	BB	AA	AA	BB	AA	BB
Chr1_13135677	0,6559	1	13135677	[A/C]	AA	BB	BB	BB	AA	BB	AA	BB
Chr7_20969923	0,9777	7	20969923	[G/C]	AA	BB	AA	BB	BB	BB	AA	BB
Chr1_2870838	0,6597	1	2870838	[T/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr8_8675401	0,5278	8	8675401	[A/G]	AA	BB	AA	BB	BB	AA	BB	AA
Chr1_25184568	0,3698	1	25184568	[T/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr5_28477355	0,8003	5	28477355	[T/G]	AA	AA	AA	AA	BB	BB	AA	BB
Chr1_2576718	0,8819	1	2576718	[T/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr6_2254306	0,6439	6	2254306	[A/C]	AA	AA	AA	BB	AB	BB	AA	AA

Chr8_4136020	0,9094	8	4136020	[A/G]	AA	AA	BB	BB	BB	BB	AA	AA
Chr2_11196198	0,8786	2	11196198	[T/G]	BB	BB	BB	--	AA	--	BB	BB
Chr11_17504733	0,8981	11	17504733	[T/C]	BB	BB	BB	BB	AA	AA	BB	AA
Chr4_31953215	0,8332	4	31953215	[T/C]	BB	AA	BB	AA	BB	BB	AA	BB
Chr2_20527179	0,9545	2	20527179	[T/C]	BB	BB	BB	BB	AA	AA	BB	BB
Chr11_27520584	0,8886	11	27520584	[T/A]	BB	--	AA	BB	BB	BB	AA	AA
Chr10_11737478	0,9529	10	11737478	[T/C]	AA	BB	BB	AA	AA	AA	BB	BB
Chr12_12844381	0,9211	12	12844381	[C/G]	BB	BB	AA	BB	BB	BB	AA	BB
Chr7_1974932	0,6477	7	1974932	[A/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr5_28723244	0,9234	5	28723244	[T/C]	BB	BB	BB	BB	BB	AA	BB	AA
Chr12_14772638	0,8312	12	14772638	[T/A]	AA	AA	AA	BB	AA	BB	AA	AA
Chr10_4713938	0,9344	10	4713938	[T/G]	AA	BB	BB	AA	AA	AA	BB	AA
Chr7_17830135	0,6707	7	17830135	[T/C]	AA	AA	BB	BB	AB	BB	AA	AA
Chr7_18119507	0,8649	7	18119507	[T/G]	BB	BB	BB	BB	AA	AA	BB	BB
Chr2_10311118	0,7181	2	10311118	[A/G]	AA	AA	AA	AB	AA	AB	AA	AA
Chr1_5168074	0,9541	1	5168074	[A/G]	BB	BB	BB	AA	AA	AA	BB	BB
Chr1_24368856	0,6917	1	24368856	[T/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr10_11886826	0,8601	10	11886826	[A/G]	AA	BB	BB	AA	AA	AA	BB	BB
Chr4_32053366	0,8391	4	32053366	[T/G]	AA	BB	AA	BB	AA	AA	BB	AA
Chr11_17525120	0,6879	11	17525120	[T/G]	AA	AB	AA	AA	BB	BB	AA	BB
Chr12_10761460	0,9387	12	10761460	[T/C]	AA	AA	BB	AA	BB	AA	BB	BB
Chr1_23787556	0,865	1	23787556	[T/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr8_20369	0,8171	8	20369	[A/G]	AA	AA	BB	BB	BB	BB	AA	BB
Chr6_1451810	0,8275	6	1451810	[T/C]	BB	BB	AA	AA	AA	AA	BB	AA
Chr11_21386015	0,8373	11	21386015	[A/C]	BB	AA	AA	AA	BB	BB	AA	BB
Chr7_4561734	0,9128	7	4561734	[T/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr4_14554232	0,9271	4	14554232	[A/G]	BB	BB	AA	BB	BB	AA	BB	BB
Chr8_17081149	0,4433	8	17081149	[T/C]	BB	AA	AA	AA	AA	BB	AA	BB

Chr1_13217704	0,8316	1	13217704	[A/G]	AA	BB	BB	BB	AA	BB	AA	BB
Chr1_38285201	0,9609	1	38285201	[G/C]	BB	BB	BB	AA	AA	AA	BB	BB
Chr12_13845714	0,9245	12	13845714	[T/C]	BB	BB	AA	BB	BB	BB	AA	BB
Chr5_28761699	0,7597	5	28761699	[T/A]	AA	AA	AA	AA	AA	BB	AA	BB
Chr6_30956408	0,9652	6	30956408	[A/G]	AA	AA	AA	BB	BB	BB	AA	BB
Chr8_266125	0,697	8	266125	[C/G]	BB	BB	AA	AA	AA	AA	BB	AA
Chr8_25749131	0,8578	8	25749131	[T/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr11_17702355	0,5835	11	17702355	[T/C]	AA	AA	AA	AA	BB	BB	AA	BB
Chr9_21130812	0,9407	9	21130812	[A/G]	BB	AA	BB	BB	BB	BB	--	AA
Chr5_6390749	0,7678	5	6390749	[T/C]	AA	AA	BB	AA	BB	BB	AA	BB
Chr3_25298901	0,7407	3	25298901	[T/C]	BB	--	BB	AA	BB	AA	BB	BB
Chr12_21260026	0,8101	12	21260026	[T/C]	AA	AA	AA	BB	AA	BB	AA	BB
Chr7_4760602	0,4181	7	4760602	[A/G]	BB	AA	AA	AA	AB	AA	BB	AA
Chr8_5169779	0,7494	8	5169779	[G/C]	AA	AA	BB	BB	BB	BB	AA	AA
Chr7_20607220	0,7216	7	20607220	[T/C]	AA	AA	AA	BB	BB	BB	AA	BB
Chr8_24922488	0,9094	8	24922488	[T/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr12_4194926	0,8797	12	4194926	[T/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr12_17429998	0,8108	12	17429998	[T/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr8_7973169	0,9136	8	7973169	[A/T]	AA	BB	AA	BB	BB	AA	BB	AA
Chr1_14734744	0,9692	1	14734744	[T/C]	BB	AA	BB	AA	AA	BB	AA	AA
Chr1_615367	0,8524	1	615367	[A/C]	AA	AA	AA	AA	BB	BB	AA	AA
Chr11_4373256	0,8482	11	4373256	[A/G]	AA	AA	BB	AA	AA	BB	AA	AA
Chr8_5322211	0,8945	8	5322211	[T/A]	BB	BB	AA	AA	AA	AA	BB	BB
Chr2_26045482	0,8368	2	26045482	[T/G]	AA	AA	AA	BB	AA	BB	AA	AA
Chr12_21500937	0,8333	12	21500937	[T/C]	BB	BB	BB	AA	BB	AA	BB	AA
Chr1_2562473	0,7312	1	2562473	[T/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr11_9731974	0,7243	11	9731974	[T/C]	BB	BB	AA	AA	AA	AA	BB	AA
Chr1_24598139	0,7842	1	24598139	[A/T]	AA	AA	AA	BB	BB	BB	AA	AA

Chr5_28661593	0,9779	5	28661593	[T/C]	BB	BB	BB	BB	BB	AA	BB	AA
Chr7_19529846	0,5407	7	19529846	[A/G]	BB	BB	AA	BB	BB	AA	BB	BB
Chr7_20906840	0,7402	7	20906840	[T/C]	AA	BB	AA	BB	BB	BB	AA	BB
Chr1_39988267	0,8275	1	39988267	[C/G]	AA	AA	AA	AA	BB	AA	BB	AA
Chr12_4215979	0,8578	12	4215979	[A/T]	AA	AA	AA	--	BB	--	AA	AA
Chr5_19156139	0,8669	5	19156139	[T/C]	BB	BB	AA	BB	BB	BB	AA	BB
Chr8_367705	0,6252	8	367705	[T/C]	AA	AA	BB	BB	BB	BB	AA	BB
Chr7_22549841	0,8913	7	22549841	[A/C]	AA	BB	AA	BB	BB	BB	AA	AA
Chr10_2302343	0,888	10	2302343	[T/G]	AA	BB	AA	BB	AA	BB	AA	AA
Chr6_24423272	0,8275	6	24423272	[T/G]	AA	AA	BB	AA	BB	BB	AA	BB
Chr1_41244361	0,6089	1	41244361	[A/G]	BB	BB	BB	AB	BB	AB	BB	BB
Chr12_7879440	0,9378	12	7879440	[C/G]	AA	AA	AA	BB	AA	BB	AA	AA
Chr12_9561938	0,9142	12	9561938	[A/G]	AA	AA	BB	AA	AA	AA	BB	BB
Chr7_17462737	0,6504	7	17462737	[A/G]	AA	--	BB	BB	AA	BB	AA	AA
Chr1_24535665	0,8543	1	24535665	[A/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr6_24500380	0,9545	6	24500380	[A/T]	AA	AA	BB	AA	AA	BB	AA	BB
Chr11_20015123	0,8192	11	20015123	[A/G]	AA	AA	AA	AA	--	BB	AA	AA
Chr1_2377724	0,8665	1	2377724	[A/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr11_27155526	0,8707	11	27155526	[A/G]	BB	BB	BB	AA	AA	AA	BB	BB
Chr12_15893454	0,8856	12	15893454	[T/A]	BB	BB	AA	BB	AA	BB	AA	AA
Chr8_24868153	0,6892	8	24868153	[T/G]	BB	BB	BB	BB	AA	AA	BB	BB
Chr12_10831750	0,7596	12	10831750	[T/C]	BB	BB	AA	BB	BB	BB	AA	BB
Chr12_17692661	0,6106	12	17692661	[C/G]	AA	AA	AA	BB	AA	BB	AA	AA
Chr1_36971673	0,9777	1	36971673	[A/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr5_18946012	0,5995	5	18946012	[T/C]	AA	AA	BB	AA	AA	AA	BB	BB
Chr1_39464636	0,8417	1	39464636	[A/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr1_2832546	0,6754	1	2832546	[T/C]	AA	AA	AA	AB	AA	BB	AA	AA
Chr5_28547843	0,7995	5	28547843	[T/C]	AA	AA	AA	AA	BB	BB	AA	BB

Chr11_3568412	0,8566	11	3568412	[T/C]	AA	AA	BB	AA	BB	BB	AA	AA
Chr8_6053118	0,798	8	6053118	[A/G]	BB	BB	BB	BB	BB	AA	BB	AA
Chr4_32606397	0,8906	4	32606397	[T/C]	AA	BB	AA	BB	BB	AA	BB	AA
Chr5_19278004	0,7465	5	19278004	[G/C]	AA	AA	BB	AA	BB	AA	BB	BB
Chr1_2719765	0,7696	1	2719765	[A/G]	AA	AA	AA	BB	AA	BB	AA	AA
Chr5_739182	0,9648	5	739182	[A/G]	AA	AA	AA	BB	AA	BB	AA	AA
Chr3_32585678	0,8748	3	32585678	[T/C]	BB	AA	BB	BB	BB	BB	AA	BB
Chr10_2242586	0,8664	10	2242586	[G/C]	BB	AA	BB	AA	BB	AA	BB	BB
Chr1_24653796	0,8303	1	24653796	[T/C]	AA	AA	AA	BB	BB	BB	AA	BB
Chr12_10341968	0,633	12	10341968	[T/C]	AA	AA	BB	AA	BB	AA	BB	BB
Chr2_20343091	0,849	2	20343091	[G/C]	AA	BB	BB	BB	BB	AA	BB	BB
Chr7_4458778	0,9521	7	4458778	[A/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr1_7271290	0,845	1	7271290	[T/C]	AA	BB	BB	BB	BB	BB	AA	AA
Chr10_13085389	0,9701	10	13085389	[A/C]	AA	AA	AA	BB	BB	BB	AA	AA
Chr12_11580708	0,8094	12	11580708	[A/G]	BB	BB	AA	BB	BB	BB	AA	BB
Chr7_4444412	0,9739	7	4444412	[A/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr1_7396465	0,8578	1	7396465	[T/C]	AA	BB	BB	BB	BB	BB	AA	AA
Chr10_16830602	0,6585	10	16830602	[A/G]	AA	BB	BB	AA	AB	AA	BB	BB
Chr12_1745586	0,6906	12	1745586	[A/T]	BB	BB	BB	AA	BB	AA	BB	BB
Chr12_16342124	0,8469	12	16342124	[T/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr12_21297436	0,7162	12	21297436	[A/G]	AA	AA	AA	BB	AA	BB	AA	BB
Chr4_32558753	0,9162	4	32558753	[T/C]	AA	BB	AA	BB	AA	AA	BB	AA
Chr12_14160817	0,8146	12	14160817	[T/C]	BB	BB	BB	AA	AA	AA	BB	BB
Chr7_18656825	0,9588	7	18656825	[T/G]	AA	AA	BB	AA	AA	BB	AA	AA
Chr7_21214128	0,6845	7	21214128	[A/C]	BB	BB	BB	AA	AA	AA	BB	BB
Chr8_16339846	0,5542	8	16339846	[A/G]	AA	BB	BB	BB	BB	AA	BB	AA
Chr11_24494344	0,9329	11	24494344	[A/G]	AA	BB	BB	BB	AA	AA	BB	AA
Chr1_18468432	0,6033	1	18468432	[T/G]	AA	BB	AA	BB	AA	AA	BB	BB

Chr2_20601159	0,9545	2	20601159	[T/C]	AA	AA	AA	AA	AA	BB	AA	AA
Chr10_20160250	0,9126	10	20160250	[A/G]	AA	AA	AA	BB	BB	BB	AA	BB
Chr1_18604179	0,6475	1	18604179	[T/C]	BB	AA	BB	AA	BB	BB	AA	AA
Chr12_13094813	0,855	12	13094813	[T/C]	AA	AA	BB	AA	AA	AA	BB	AA
Chr7_6985139	0,7179	7	6985139	[T/C]	BB	AA	AA	BB	BB	BB	AA	AA
Chr1_2279277	0,9407	1	2279277	[A/G]	AA	AA	AA	BB	AA	BB	AA	AA
Chr1_18530303	0,8539	1	18530303	[A/C]	BB	AA	BB	AA	BB	BB	AA	AA
Chr2_8016075	0,8673	2	8016075	[A/C]	AA	BB	BB	AA	AA	BB	AA	BB
Chr7_17740426	0,429	7	17740426	[A/G]	AA	AA	--	AA	AB	AB	AA	AA
Chr6_871216	0,8361	6	871216	[T/C]	AA	AA	BB	BB	BB	BB	AA	BB
Chr1_31779292	0,7191	1	31779292	[A/G]	AA	AA	AA	AA	BB	BB	AA	AA
Chr12_10255735	0,7329	12	10255735	[T/C]	AA	AA	BB	AA	BB	AA	BB	BB
Chr12_14086939	0,7369	12	14086939	[T/C]	AA	AA	AA	BB	AB	BB	AA	AA
Chr11_19462431	0,8267	11	19462431	[T/C]	BB	BB	BB	BB	BB	AA	BB	BB
Chr12_12420891	0,7975	12	12420891	[A/G]	AA	AA	BB	AA	AA	AA	BB	AA
Chr11_20982607	0,8326	11	20982607	[T/C]	AA	BB	BB	BB	AA	AA	BB	AA
Chr12_23306465	0,6945	12	23306465	[A/G]	AA	BB	BB	AA	AB	AA	BB	AA
Chr12_9715551	0,8312	12	9715551	[A/T]	AA	AA	BB	AA	AA	AA	BB	BB
Chr8_20073861	0,8109	8	20073861	[A/G]	AA	BB	BB	BB	AA	AA	BB	AA
Chr8_25709537	0,9005	8	25709537	[G/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr11_17779079	0,4376	11	17779079	[A/G]	BB	BB	BB	AB	AA	AA	AB	AA
Chr11_20092217	0,7483	11	20092217	[A/T]	BB	BB	BB	BB	AA	AA	BB	BB
Chr7_20688426	0,769	7	20688426	[A/C]	BB	BB	BB	AA	AA	AA	BB	AA
Chr1_7360011	0,7224	1	7360011	[T/C]	BB	AA	AA	AA	AB	AA	BB	BB
Chr10_18425910	0,7734	10	18425910	[A/T]	AA	BB	AA	BB	BB	AA	BB	AA
Chr10_22806614	0,7736	10	22806614	[A/G]	AA	AA	BB	BB	BB	BB	AA	AA
Chr12_21488294	0,5809	12	21488294	[T/C]	BB	BB	BB	AA	BB	AA	BB	AA
Chr8_24756144	0,7737	8	24756144	[T/C]	AA	AA	AA	AA	BB	BB	AA	AA

Chr11_16887564	0,6439	11	16887564	[A/C]	BB	BB	BB	BB	AA	AA	BB	AA
Chr7_22592770	0,7701	7	22592770	[C/G]	BB	AA	BB	AA	AA	AA	BB	BB
Chr12_12104433	0,8175	12	12104433	[A/G]	AA	AA	BB	AA	AA	AA	BB	AA
Chr3_34008736	0,7443	3	34008736	[A/G]	AA	AA	AA	BB	BB	BB	AA	BB
Chr7_21155658	0,3925	7	21155658	[A/C]	AB	BB	AB	BB	BB	BB	AA	BB
Chr12_21286804	0,7271	12	21286804	[G/C]	AA	AA	AA	AB	AB	AB	AA	AB
Chr7_18508003	0,6434	7	18508003	[A/G]	AA	AA	BB	AA	AA	BB	AA	AA
Chr1_24701256	0,8305	1	24701256	[T/C]	AA	AA	AA	BB	AB	BB	AA	BB
Chr8_399380	0,5762	8	399380	[G/C]	AA	AA	BB	BB	AB	BB	AA	BB
Chr4_30712250	0,6006	4	30712250	[A/G]	AA	BB	AA	BB	AB	AA	BB	AA
Chr11_22417338	0,301	11	22417338	[T/C]	AB	BB	AB	AA	AB	AA	BB	AA
Chr2_20508336	0,6978	2	20508336	[A/G]	BB	BB	BB	BB	AB	AA	BB	BB
Chr10_16556591	0,8196	10	16556591	[T/C]	BB	AA	BB	BB	BB	BB	AA	BB
Chr11_19671083	0,7314	11	19671083	[A/C]	AA	AA	AA	AA	BB	BB	AA	AA
Chr4_23936450	0,6402	4	23936450	[A/G]	BB	BB	BB	BB	BB	BB	BB	BB
Chr10_16782798	0,8155	10	16782798	[T/C]	AA	BB	BB	AA	BB	AA	BB	BB
Chr1_24430990	0,7268	1	24430990	[T/C]	BB	BB	BB	AA	AA	AA	BB	BB
Chr5_28647080	0,8793	5	28647080	[T/C]	BB	BB	BB	BB	BB	AA	BB	AA
Chr5_4564362	0,6133	5	4564362	[A/G]	BB	AA	AA	AA	AA	BB	AA	BB
Chr5_607720	0,8333	5	607720	[A/G]	AA	AA	AA	BB	AA	BB	AA	AA
Chr10_2290710	0,6785	10	2290710	[A/G]	BB	AA	BB	AA	BB	AA	BB	BB
Chr1_5194342	0,8057	1	5194342	[T/C]	BB	BB	BB	AA	AB	AA	BB	BB
Chr8_24793324	0,6553	8	24793324	[T/C]	AA	AA	AA	AA	BB	BB	AA	AA
Chr12_13027259	0,3745	12	13027259	[A/G]	AA	AB	BB	AB	AB	AA	BB	AA
Chr3_6184520	0,6636	3	6184520	[A/G]	AA	BB	AA	BB	BB	BB	AA	AA
Chr10_19842255	0,6348	10	19842255	[A/G]	AA	AA	BB	AA	AB	BB	AA	BB
Chr12_10796464	0,8467	12	10796464	[A/G]	AA	AA	BB	AA	AA	AA	BB	BB
Chr1_39473259	0,9632	1	39473259	[T/C]	AA	AA	AA	BB	AA	BB	AA	AA

Chr12_10862056	0,7498	12	10862056	[A/G]	AA	AA	BB	AA	AA	AA	BB	AA
Chr8_7506944	0,8143	8	7506944	[A/G]	BB	AA	AA	AA	AA	BB	AA	BB
Chr8_92154	0,817	8	92154	[T/C]	BB	BB	AA	AA	AA	AA	BB	AA
Chr1_8310349	0,8046	1	8310349	[T/C]	BB	BB	BB	BB	AB	BB	AA	AA
Chr10_13111426	0,6281	10	13111426	[A/G]	BB	BB	BB	AA	AA	AA	BB	BB
Chr6_24461748	0,8438	6	24461748	[T/G]	BB	BB	AA	BB	BB	AA	BB	AA
Chr3_35285634	0,7124	3	35285634	[T/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr12_16497437	0,6539	12	16497437	[T/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr12_10960570	0,7793	12	10960570	[A/C]	BB	BB	AA	BB	BB	BB	AA	BB
Chr1_31841038	0,7558	1	31841038	[T/C]	AA	AA	AA	AA	BB	BB	AA	AA
Chr7_18108773	0,719	7	18108773	[T/C]	BB	BB	BB	BB	AA	AA	BB	BB
Chr11_17516754	0,8022	11	17516754	[T/C]	AA	AA	AA	AA	BB	BB	AA	BB
Chr4_32015102	0,7086	4	32015102	[A/G]	BB	AA	BB	AA	BB	BB	AA	BB
Chr1_11374629	0,879	1	11374629	[A/G]	AA	AA	AA	AA	AA	AA	BB	AA
Chr5_28545448	0,339	5	28545448	[A/G]	AB	AB	AB	AB	AA	AA	BB	AA
Chr12_14771715	0,8311	12	14771715	[T/C]	AA	AA	AA	AB	AB	AB	AA	AA
Chr1_35006178	0,682	1	35006178	[A/G]	BB	BB	AA	BB	AB	BB	AA	AA
Chr12_10640097	0,7861	12	10640097	[T/A]	AA	AA	BB	AA	BB	AA	BB	BB
Chr7_23236831	0,7853	7	23236831	[A/C]	AA	BB	AA	BB	BB	BB	AA	BB
Chr2_11188067	0,8011	2	11188067	[A/G]	AA	AA	AA	BB	AB	BB	AA	AA
Chr5_18847550	0,7172	5	18847550	[C/G]	BB	BB	AA	BB	BB	BB	AA	AA
Chr1_24883583	0,4712	1	24883583	[T/C]	BB	BB	BB	AA	AA	AA	BB	AA
Chr12_12789673	0,5933	12	12789673	[A/G]	AA	AA	BB	AA	AA	AA	BB	AA
Chr1_18668242	0,6706	1	18668242	[T/G]	AA	AA	AA	BB	AA	AA	BB	AA
Chr8_3001294	0,6511	8	3001294	[A/G]	AA	AA	BB	BB	BB	BB	AA	AA
Chr12_21871704	0,6578	12	21871704	[T/G]	AB	AB	AB	BB	AB	BB	AB	BB
Chr12_9979324	0,708	12	9979324	[T/A]	AB	AB	AA	AB	AA	AB	AA	AA
Chr7_21314665	0,8623	7	21314665	[A/G]	BB	BB	BB	AA	AA	AA	BB	BB

Chr11_17647750	0,7334	11	17647750	[T/C]	BB	BB	BB	BB	AA	AA	BB	AA
Chr8_8774226	0,8274	8	8774226	[A/G]	BB	AA	BB	AA	AA	BB	AA	BB
Chr7_18698410	0,4816	7	18698410	[T/C]	AA	AA	BB	AA	AA	BB	AA	AA
Chr8_5947998	0,8969	8	5947998	[A/C]	AA	BB	BB	BB	AA	AA	BB	AA
Chr6_24628088	0,5925	6	24628088	[A/G]	BB	BB	AA	BB	BB	AA	BB	AB
Chr10_11911606	0,8265	10	11911606	[T/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr11_9011218	0,5655	11	9011218	[A/C]	BB	AA	BB	BB	AB	--	AA	BB
Chr3_8652542	0,8343	3	8652542	[A/G]	AA	BB	BB	BB	AA	AA	BB	BB
Chr11_27365240	0,8348	11	27365240	[T/G]	AA	BB	BB	AA	BB	AA	BB	BB
Chr2_18044260	0,6558	2	18044260	[T/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr12_11373615	0,8289	12	11373615	[T/C]	BB	BB	AA	BB	BB	BB	AA	BB
Chr8_16837566	0,4079	8	16837566	[A/G]	BB	AA	AB	AB	AB	BB	AA	BB
Chr10_3958580	0,6502	10	3958580	[T/C]	AB	BB	BB	AA	AA	AB	BB	AA
Chr5_2814809	0,8539	5	2814809	[A/C]	BB	BB	AA	AA	BB	BB	AA	BB
Chr12_11701306	0,7351	12	11701306	[A/G]	AB	AB	BB	AB	AB	AB	BB	AB
Chr12_13950048	0,6447	12	13950048	[A/G]	AA	AA	BB	AA	AA	AA	BB	AA
Chr1_13148124	0,6032	1	13148124	[C/G]	BB	AA	AA	AA	BB	AA	BB	AA
Chr7_21028789	0,5133	7	21028789	[T/C]	BB	AA	BB	AA	AA	AA	BB	AA
Chr6_23965052	0,6148	6	23965052	[A/G]	BB	--	BB	BB	AA	AA	BB	AA
Chr11_21330402	0,8059	11	21330402	[T/G]	AA	BB	BB	BB	AA	AA	BB	AA
Chr3_29673798	0,8413	3	29673798	[A/C]	AA	AA	BB	BB	BB	BB	AA	AA
Chr1_24015434	0,8422	1	24015434	[T/G]	AA	AA	AA	BB	BB	BB	AA	AA
Chr1_40036706	0,6201	1	40036706	[T/C]	BB	BB	BB	BB	AB	BB	AA	BB
Chr7_19245970	0,7889	7	19245970	[T/C]	BB	BB	AA	BB	BB	AA	BB	BB
Chr6_24269866	0,6216	6	24269866	[A/G]	BB	AA	BB	BB	AA	AA	BB	AA
Chr7_20586169	0,7891	7	20586169	[A/G]	BB	BB	BB	AA	AA	AA	BB	BB
Chr1_24227227	0,8044	1	24227227	[T/A]	AA	AA	AA	BB	BB	BB	AA	AA
Chr8_5134106	0,8294	8	5134106	[T/C]	AA	AA	BB	BB	BB	BB	AA	AA

Chr5_19239979	0,8911	5	19239979	[A/T]	AA	AA	BB	AA	BB	AA	BB	AA
Chr4_28848232	0,6971	4	28848232	[T/C]	AA	BB	AA	BB	AB	AA	BB	AA
Chr6_24022203	0,8475	6	24022203	[A/C]	BB	BB	BB	BB	AA	AA	BB	AA
Chr12_15128536	0,8117	12	15128536	[T/C]	AA	AA	BB	AA	BB	AA	BB	BB
Chr12_18910429	0,8981	12	18910429	[A/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr7_3126071	0,7426	7	3126071	[C/G]	BB	BB	BB	AA	BB	BB	AA	BB
Chr6_24395844	0,7347	6	24395844	[T/C]	BB	BB	--	BB	AA	AA	BB	AA
Chr11_4353161	0,8208	11	4353161	[T/C]	BB	BB	BB	BB	BB	AA	BB	BB
Chr8_205404	0,6925	8	205404	[A/G]	BB	BB	AA	AA	AB	AA	BB	AA
Chr7_16555976	0,601	7	16555976	[A/G]	AB	AB	AB	BB	AB	BB	AB	AB
Chr8_24908113	0,7961	8	24908113	[T/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr1_18525132	0,7707	1	18525132	[T/C]	AA	BB	AA	BB	AA	AA	BB	BB
Chr1_2797145	0,8809	1	2797145	[A/G]	AA	AA	AA	BB	AA	BB	AA	AA
Chr1_41304470	0,8458	1	41304470	[T/G]	BB	BB	AA	BB	BB	BB	AA	BB
Chr7_4852994	0,3748	7	4852994	[T/C]	AA	BB	BB	BB	AB	BB	AA	BB
Chr3_7953743	0,7414	3	7953743	[T/C]	AA	BB	AA	BB	AA	AA	BB	AA
Chr7_19028604	0,8037	7	19028604	[C/G]	BB	BB	AA	BB	BB	AA	BB	BB
Chr12_14739304	0,4416	12	14739304	[T/G]	AB	AB	AB	BB	AB	BB	AB	AB
Chr7_19556942	0,7251	7	19556942	[A/G]	BB	BB	AA	BB	BB	AA	BB	BB
Chr11_19875372	0,6728	11	19875372	[A/G]	AA	AA	AA	AA	AA	BB	AA	AA
Chr8_17156701	0,6814	8	17156701	[T/G]	AA	BB	BB	BB	BB	AA	BB	AA
Chr2_20589489	0,6162	2	20589489	[T/G]	AA	AA	AA	AA	AA	BB	AA	AA
Chr10_11740834	0,6641	10	11740834	[A/G]	AA	BB	BB	AA	AA	AA	BB	BB
Chr12_17639213	0,8268	12	17639213	[G/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr5_19037591	0,6673	5	19037591	[G/C]	AA	AA	BB	AA	AA	AA	BB	AA
Chr1_7250515	0,7949	1	7250515	[T/G]	BB	AA	AA	AA	AA	AA	BB	BB
Chr1_7327287	0,8508	1	7327287	[A/T]	BB	AA	AA	AA	AA	AA	BB	BB
Chr2_8166854	0,6416	2	8166854	[A/G]	BB	BB	BB	AA	BB	AA	BB	BB

Chr1_5130438	0,7066	1	5130438	[T/G]	BB	BB	BB	AA	AA	AA	BB	BB
Chr10_11945290	0,8099	10	11945290	[T/G]	BB	AA	AA	BB	BB	BB	AA	AA
Chr12_14540102	0,7338	12	14540102	[T/C]	AA	AA	AA	BB	AB	BB	AA	AA
Chr7_4267087	0,9136	7	4267087	[A/G]	AA	AA	AA	BB	AA	BB	AA	AA
Chr11_7344165	0,8252	11	7344165	[A/T]	AA	AA	AA	BB	BB	BB	AA	AA
Chr10_13050766	0,7999	10	13050766	[A/G]	BB	BB	BB	AA	AA	AA	BB	BB
Chr2_28290282	0,6399	2	28290282	[A/T]	AA	AA	BB	AA	AB	BB	AA	BB
Chr12_12579443	0,4493	12	12579443	[T/C]	AA	AB	BB	AB	AB	AA	BB	AA
Chr12_12815390	0,8138	12	12815390	[A/G]	AA	AA	BB	AA	AA	AA	BB	AA
Chr12_21250812	0,4274	12	21250812	[T/C]	AB	AB	BB	AA	BB	AA	BB	AA
Chr3_34744057	0,8272	3	34744057	[A/G]	AA	AA	AA	BB	BB	BB	AA	AA
Chr12_1433582	0,6675	12	1433582	[A/G]	AA	AA	AA	BB	BB	BB	AA	AA
Chr3_23666632	0,8099	3	23666632	[T/C]	BB	BB	BB	BB	BB	BB	AA	BB
Chr12_11195005	0,7946	12	11195005	[A/G]	AA	AA	BB	AA	AB	AA	BB	AA
Chr6_31185457	0,5073	6	31185457	[T/C]	BB	BB	BB	AA	AB	AA	BB	AA
Chr10_11825111	0,7972	10	11825111	[A/C]	AA	BB	BB	AA	AA	AA	BB	BB
Chr11_18111214	0,7303	11	18111214	[A/G]	AA	AA	AA	AA	AB	BB	AA	AA
Chr3_35643104	0,8513	3	35643104	[T/C]	BB	BB	AA	AA	AA	AA	BB	BB
Chr2_8407535	0,5593	2	8407535	[T/C]	BB	BB	BB	AA	AA	AA	BB	BB
Chr6_31080223	0,421	6	31080223	[A/G]	AB	BB	AB	AA	AA	AA	AB	AA
Chr8_2064973	0,4864	8	2064973	[A/G]	AA	AA	AB	AB	AA	BB	AA	BB
Chr7_1694721	0,7131	7	1694721	[A/C]	BB	BB	BB	BB	BB	BB	AA	BB
Chr12_13842863	0,3467	12	13842863	[A/G]	AA	AA	BB	AA	AA	AA	AB	AA
Chr12_9870376	0,8749	12	9870376	[T/A]	AA	AA	BB	AA	BB	AA	BB	BB
Chr2_10273874	0,9339	2	10273874	[T/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr11_17882866	0,9009	11	17882866	[A/G]	AA	AA	AA	AA	BB	BB	AA	BB
Chr7_20843856	0,8894	7	20843856	[A/G]	AA	BB	AA	BB	BB	BB	AA	BB
Chr7_21134321	0,7352	7	21134321	[A/G]	AA	BB	AA	BB	BB	BB	AA	BB

Chr8_16867954	0,5451	8	16867954	[A/G]	AB	BB	BB	BB	BB	AA	BB	AA
Chr1_42523627	0,7823	1	42523627	[A/G]	AA	--	AA	BB	AA	AA	BB	BB
Chr8_173704	0,6236	8	173704	[A/G]	AA	AA	BB	BB	BB	BB	AA	BB
Chr2_16407237	0,7333	2	16407237	[A/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr11_24546934	0,6183	11	24546934	[T/C]	AA	BB	BB	BB	AA	AA	BB	AA
Chr1_39390907	0,9596	1	39390907	[A/G]	AA	AA	AA	BB	AA	BB	AA	AA
Chr11_21268786	0,3968	11	21268786	[A/G]	BB	AA	AA	AA	AB	BB	AA	BB
Chr12_9611515	0,7384	12	9611515	[A/G]	BB	BB	AA	BB	AB	BB	AA	AA
Chr5_19135141	0,6357	5	19135141	[A/G]	AA	AA	BB	AA	AA	AA	BB	AA
Chr12_14468229	0,9575	12	14468229	[A/G]	AA	AA	AA	BB	BB	BB	AA	AA
Chr11_19892078	0,8444	11	19892078	[G/C]	AA	AA	AA	AA	AA	BB	AA	AA
Chr8_5284362	0,7356	8	5284362	[T/C]	BB	BB	AA	AA	AA	AA	BB	BB
Chr11_13291181	0,8179	11	13291181	[A/G]	AA	AA	AA	BB	BB	BB	AA	BB
Chr1_24803304	0,9018	1	24803304	[T/C]	BB	BB	BB	AA	AA	AA	BB	AA
Chr1_29302358	0,8855	1	29302358	[T/G]	AA	AA	AA	BB	AA	BB	AA	AA
Chr3_21456335	0,8261	3	21456335	[G/C]	AA	AA	AA	BB	AA	AA	BB	BB
Chr1_35810792	0,75	1	35810792	[A/G]	BB	AA	BB	AA	BB	BB	AA	BB
Chr2_19191264	0,671	2	19191264	[G/C]	BB	BB	BB	BB	BB	AA	BB	AA
Chr11_17140364	0,8381	11	17140364	[T/C]	AA	AA	AA	AA	BB	BB	AA	BB
Chr7_15242701	0,5727	7	15242701	[T/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr1_2832944	0,8069	1	2832944	[C/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr6_1516624	0,7598	6	1516624	[A/C]	AA	AA	BB	BB	AA	BB	AA	BB
Chr3_25553010	0,8566	3	25553010	[A/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr11_27534452	0,8671	11	27534452	[A/G]	AA	BB	BB	AA	AA	AA	BB	BB
Chr12_14900289	0,5201	12	14900289	[A/G]	AA	AA	AA	BB	AB	BB	AA	AA
Chr5_9651575	0,6116	5	9651575	[A/T]	AA	AA	BB	AA	AA	AA	BB	BB
Chr6_113705	0,7755	6	113705	[A/G]	AA	AA	BB	BB	BB	--	AA	BB
Chr12_17576482	0,7368	12	17576482	[A/G]	AA	AA	AA	BB	AA	BB	AA	AA

Chr8_7949448	0,5815	8	7949448	[T/C]	AB	AA	BB	AA	AA	BB	AA	BB
Chr5_28700378	0,6096	5	28700378	[A/G]	AA	AA	AA	AA	AA	BB	AA	BB
Chr10_2340630	0,6365	10	2340630	[A/G]	AB	BB	AB	BB	AB	BB	AB	AA
Chr5_28469441	0,579	5	28469441	[T/A]	BB	BB	BB	BB	AA	AA	BB	AA
Chr12_15862305	0,6924	12	15862305	[T/C]	BB	BB	AA	BB	AA	BB	AA	AA
Chr12_10392447	0,5838	12	10392447	[A/T]	AA	AA	BB	AA	AB	AA	BB	BB
Chr7_4389114	0,9093	7	4389114	[T/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr12_12202599	0,5759	12	12202599	[T/G]	AA	AB	BB	AA	AA	AA	BB	AA
Chr7_4568760	0,5384	7	4568760	[G/C]	AA	AA	AA	BB	AB	BB	AA	AA
Chr7_4756028	0,7027	7	4756028	[T/C]	AA	BB	BB	BB	AA	BB	AA	BB
Chr1_25408222	0,938	1	25408222	[T/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr8_25373283	0,9692	8	25373283	[T/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr7_4509662	0,9545	7	4509662	[A/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr3_25469584	0,8615	3	25469584	[T/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr7_19188931	0,6851	7	19188931	[A/G]	BB	BB	AA	BB	BB	AA	BB	BB
Chr12_14029565	0,7671	12	14029565	[T/C]	BB	BB	AA	BB	BB	BB	AA	BB
Chr6_1431533	0,6879	6	1431533	[T/C]	BB	BB	AA	AA	AA	AA	BB	AA
Chr1_24631118	0,845	1	24631118	[T/C]	AA	AA	AA	BB	BB	BB	AA	BB
Chr11_21554364	0,8729	11	21554364	[T/C]	BB	AA	AA	AA	AA	BB	AA	BB
Chr3_25337804	0,8672	3	25337804	[A/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr12_16997356	0,6415	12	16997356	[T/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr1_2388681	0,849	1	2388681	[T/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr11_19714192	0,8371	11	19714192	[T/C]	BB	BB	BB	BB	--	AA	BB	BB
Chr1_11465807	0,9143	1	11465807	[A/G]	BB	BB	BB	BB	BB	BB	AA	BB
Chr5_19072226	0,7988	5	19072226	[A/G]	BB	BB	AA	BB	BB	BB	AA	BB
Chr8_24854148	0,6926	8	24854148	[A/G]	BB	BB	BB	BB	AA	AA	BB	BB
Chr3_25252811	0,5424	3	25252811	[A/G]	AA	AA	AA	BB	AA	BB	AA	AA
Chr7_19496801	0,7557	7	19496801	[T/A]	BB	BB	AA	BB	BB	AA	BB	BB

Chr6_31002977	0,8119	6	31002977	[T/C]	BB	BB	BB	AA	AA	AA	BB	AA
Chr12_9344536	0,62	12	9344536	[A/G]	BB	BB	AA	BB	BB	BB	AA	AA
Chr11_9092055	0,8647	11	9092055	[A/G]	AA	BB	AA	AA	AA	AA	BB	AA
Chr11_3594150	0,6316	11	3594150	[A/G]	BB	BB	BB	BB	AA	AA	BB	BB
Chr11_17464275	0,7062	11	17464275	[T/C]	BB	BB	BB	BB	AB	AA	BB	AA
Chr7_20770081	0,7593	7	20770081	[A/C]	AA	BB	AA	BB	AB	BB	AA	BB
Chr6_24104114	0,8005	6	24104114	[A/G]	AA	BB	AA	AA	BB	BB	AA	BB
Chr1_7450516	0,74	1	7450516	[C/G]	BB	AA	AA	AA	AA	AA	BB	BB
Chr12_12527055	0,5649	12	12527055	[T/C]	BB	BB	AA	BB	BB	AB	AA	BB
Chr10_3781853	0,7132	10	3781853	[T/C]	AA	AB	BB	AA	AA	AA	BB	AA
Chr12_10122138	0,6192	12	10122138	[T/C]	BB	BB	AA	BB	AA	BB	AA	AA
Chr5_6408414	0,5489	5	6408414	[A/G]	BB	BB	AB	BB	AB	AA	BB	AB
Chr2_26412422	0,5913	2	26412422	[G/C]	AA	AA	AA	BB	AB	BB	AA	AA
Chr5_28920719	0,7885	5	28920719	[T/C]	BB	BB	BB	BB	AB	AA	BB	AA
Chr1_37848648	0,5877	1	37848648	[T/A]	AA	AB	--	--	--	BB	AA	AA
Chr11_17009070	0,8107	11	17009070	[A/G]	BB	BB	BB	BB	AA	AA	BB	AA
Chr1_24313675	0,707	1	24313675	[T/C]	BB	BB	BB	AA	AA	AA	BB	BB
Chr1_23857450	0,8222	1	23857450	[G/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr1_24138244	0,7482	1	24138244	[T/C]	BB	BB	BB	AA	AA	--	BB	BB
Chr11_18009558	0,7836	11	18009558	[T/C]	AA	AA	AA	AA	BB	BB	AA	AA
Chr12_7846049	0,5416	12	7846049	[G/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr2_11317462	0,9218	2	11317462	[A/G]	BB	BB	BB	AA	AA	AA	BB	BB
Chr12_10469880	0,8308	12	10469880	[T/C]	BB	BB	AA	BB	AA	BB	AA	AA
Chr7_19378781	0,7316	7	19378781	[T/C]	BB	BB	AA	BB	AB	AA	BB	BB
Chr12_17435936	0,7017	12	17435936	[T/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr12_11431372	0,971	12	11431372	[A/G]	AA	AA	--	AA	AA	AA	--	AA
Chr12_17896641	0,7355	12	17896641	[A/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr1_2705284	0,9797	1	2705284	[A/C]	AA	AA	AA	BB	AA	BB	AA	AA

Chr8_66412	0,8449	8	66412	[A/G]	BB	BB	AA	AA	AA	AA	BB	AA
Chr12_10709287	0,5615	12	10709287	[T/C]	BB	BB	AA	BB	AA	BB	AA	AA
Chr7_3963635	0,7392	7	3963635	[A/C]	BB	BB	BB	AA	AA	AA	BB	BB
Chr8_23071056	0,8312	8	23071056	[A/G]	BB	BB	BB	BB	BB	BB	AA	AA
Chr2_21344613	0,7951	2	21344613	[T/G]	BB	BB	AA	BB	BB	AA	BB	BB
Chr12_23433733	0,8276	12	23433733	[A/G]	BB	AA	AA	BB	AA	BB	AA	BB
Chr11_5458459	0,7907	11	5458459	[A/G]	BB	BB	BB	AA	AB	AA	BB	AA
Chr11_18085361	0,8371	11	18085361	[T/C]	AA	AA	AA	AA	BB	BB	AA	AA
Chr1_18590506	0,7121	1	18590506	[A/G]	BB	AA	BB	AA	BB	BB	AA	AA
Chr7_7267127	0,8149	7	7267127	[T/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr4_17006154	0,7677	4	17006154	[A/C]	BB	BB	BB	BB	BB	AA	BB	BB
Chr12_7928753	0,5762	12	7928753	[T/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr9_20913240	0,4463	9	20913240	[A/G]	BB	BB	BB	AA	AB	AA	BB	BB
Chr10_20153803	0,5793	10	20153803	[A/G]	AA	AA	AA	BB	BB	BB	AA	BB
Chr7_20950472	0,7321	7	20950472	[A/G]	BB	AA	BB	AA	AA	AA	BB	AA
Chr7_17813875	0,8186	7	17813875	[T/C]	AA	AA	BB	BB	AA	BB	AA	AA
Chr11_21496513	0,7138	11	21496513	[T/G]	BB	AA	AA	AA	BB	BB	AA	BB
Chr6_24700648	0,9219	6	24700648	[C/G]	AA	AA	BB	AA	AA	BB	AA	BB
Chr1_25211220	0,6439	1	25211220	[A/G]	AA	AA	AA	BB	AA	BB	AA	AA
Chr12_15008734	0,8173	12	15008734	[A/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr3_22408454	0,8176	3	22408454	[A/G]	BB	BB	BB	BB	BB	BB	AA	AA
Chr12_3436634	0,5664	12	3436634	[A/G]	AA	AA	AA	AB	AB	BB	AA	AA
Chr10_2202318	0,7611	10	2202318	[A/G]	BB	AA	BB	AA	AB	AA	BB	BB
Chr8_311834	0,6841	8	311834	[T/C]	AA	AA	BB	BB	BB	BB	AA	BB
Chr1_13249852	0,7464	1	13249852	[T/C]	AA	BB	BB	BB	--	BB	AA	BB
Chr8_6119160	0,8464	8	6119160	[C/G]	AA	BB	BB	BB	BB	AA	BB	AA
Chr6_30920215	0,8055	6	30920215	[T/C]	AA	AA	AA	BB	BB	BB	AA	BB
Chr12_11292981	0,6555	12	11292981	[A/G]	AA	AA	BB	AA	AA	AA	BB	AA

Chr10_15272868	0,8511	10	15272868	[T/G]	BB	BB	AA	BB	BB	AA	BB	BB
Chr11_24653048	0,8145	11	24653048	[C/G]	AA	BB	BB	BB	AA	AA	BB	AA
Chr9_21145279	0,7969	9	21145279	[A/G]	AA	BB	AA	AA	AA	AA	BB	BB
Chr1_23809551	0,8007	1	23809551	[C/G]	AA	AA	AA	BB	BB	BB	AA	AA
Chr11_22684758	0,9664	11	22684758	[G/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr7_17473541	0,9009	7	17473541	[T/C]	BB	BB	AA	AA	BB	AA	BB	BB
Chr4_32566460	0,5883	4	32566460	[A/G]	AA	BB	AA	BB	AA	AA	BB	AA
Chr3_25460399	0,5678	3	25460399	[T/C]	BB	BB	BB	AB	BB	AA	BB	BB
Chr12_11920597	0,7211	12	11920597	[A/G]	BB	BB	AA	BB	BB	BB	AA	BB
Chr8_16974973	0,8901	8	16974973	[A/C]	BB	AA	AA	AA	AA	BB	AA	BB
Chr1_2572143	0,8111	1	2572143	[A/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr12_4147113	0,8529	12	4147113	[T/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr12_4212494	0,9136	12	4212494	[A/T]	AA	AA	AA	BB	BB	BB	AA	AA
Chr1_24511372	0,8885	1	24511372	[A/G]	BB	BB	BB	AA	AA	AA	BB	BB
Chr4_14811473	0,787	4	14811473	[T/C]	BB	BB	AA	BB	BB	AA	BB	BB
Chr11_19981650	0,7361	11	19981650	[T/G]	AA	AB	AA	AA	AA	BB	AA	AA
Chr5_212132	0,8068	5	212132	[A/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr5_1783164	0,8956	5	1783164	[A/G]	BB	BB	AA	AA	BB	BB	AA	BB
Chr12_14848121	0,8655	12	14848121	[A/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr5_21579071	0,7558	5	21579071	[T/A]	BB	AA	BB	BB	BB	BB	AA	AA
Chr1_42653198	0,7896	1	42653198	[A/G]	BB	AA	BB	AA	AA	BB	AA	AA
Chr8_16344216	0,7454	8	16344216	[T/C]	BB	AA	AA	AA	AB	BB	AA	BB
Chr12_13434729	0,6188	12	13434729	[T/C]	BB	BB	AA	BB	BB	BB	AA	BB
Chr8_5227340	0,8652	8	5227340	[T/C]	AA	AA	BB	BB	BB	BB	AA	AA
Chr7_19625790	0,8292	7	19625790	[T/G]	AA	AA	BB	AA	AA	BB	AA	AA
Chr3_30795599	0,7728	3	30795599	[A/G]	AA	AA	BB	BB	AA	BB	AA	AA
Chr3_9450608	0,5718	3	9450608	[A/C]	AA	AA	AA	BB	AA	AA	BB	BB
Chr1_12897299	0,7686	1	12897299	[T/C]	BB	AA	AA	AA	BB	AA	BB	AA

Chr1_42457631	0,5974	1	42457631	[A/G]	AA	BB	AA	BB	AB	AA	BB	BB
Chr8_17623639	0,7859	8	17623639	[A/C]	AA	BB	BB	BB	BB	AA	BB	AA
Chr6_1648271	0,588	6	1648271	[A/G]	BB	BB	AA	AA	BB	AA	BB	AA
Chr1_27288194	0,5563	1	27288194	[G/C]	AA	AA	BB	AA	AB	BB	AA	BB
Chr8_5186293	0,6881	8	5186293	[A/G]	AA	AA	BB	BB	AB	BB	AA	AA
Chr12_4438655	0,5788	12	4438655	[A/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr12_9498931	0,5702	12	9498931	[T/G]	BB	AB	AA	BB	BB	BB	AA	AA
Chr12_12311981	0,8104	12	12311981	[G/C]	BB	BB	AA	BB	AB	BB	AA	BB
Chr1_23955243	0,7135	1	23955243	[A/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr1_12976776	0,6254	1	12976776	[T/C]	BB	AA	AA	AA	BB	AA	BB	AA
Chr6_23881784	0,7179	6	23881784	[A/G]	AA	AA	AA	AA	BB	BB	AA	BB
Chr11_21431933	0,8384	11	21431933	[T/C]	BB	AA	AA	AA	BB	BB	AA	BB
Chr5_10774997	0,7452	5	10774997	[A/G]	BB	BB	AA	BB	AB	BB	AA	AA
Chr12_4477396	0,8203	12	4477396	[G/C]	BB	BB	BB	AA	AA	AA	BB	BB
Chr8_17533934	0,5863	8	17533934	[A/G]	AA	BB	BB	BB	AB	AA	BB	AA
Chr5_6448939	0,6002	5	6448939	[T/C]	BB	BB	AA	BB	AA	AA	BB	AA
Chr6_24160787	0,8003	6	24160787	[A/G]	BB	AA	BB	BB	AA	AA	BB	AA
Chr11_17807716	0,7272	11	17807716	[A/G]	BB	BB	BB	BB	AB	AA	BB	AB
Chr1_37183724	0,9545	1	37183724	[A/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr7_4333082	0,8729	7	4333082	[T/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr5_18913012	0,8681	5	18913012	[T/C]	AA	AA	BB	AA	AA	AA	BB	BB
Chr11_17597632	0,9201	11	17597632	[T/C]	BB	BB	BB	BB	AA	AA	BB	AA
Chr11_27025213	0,8671	11	27025213	[T/C]	AA	AA	AA	BB	BB	BB	AA	AA
Chr12_14165906	0,5907	12	14165906	[A/G]	BB	BB	BB	AA	AA	AA	BB	BB
Chr1_15348002	0,631	1	15348002	[T/C]	AA	BB	AA	BB	AA	AA	BB	BB
Chr7_20653073	0,6996	7	20653073	[T/C]	AA	AA	AA	AB	AB	BB	AA	BB
Chr11_19513789	0,7725	11	19513789	[T/C]	AA	AA	AA	AA	BB	BB	AA	AA
Chr12_9538580	0,8036	12	9538580	[A/G]	BB	BB	AA	BB	--	BB	AA	AA

Chr12_10320565	0,7538	12	10320565	[A/G]	BB	BB	AA	BB	AA	BB	AA	AA
Chr8_8433342	0,8065	8	8433342	[T/C]	BB	AA	BB	AA	BB	BB	AA	BB
Chr10_13229937	0,7424	10	13229937	[T/A]	AA	AA	AA	BB	BB	BB	AA	AA
Chr10_20179897	0,6004	10	20179897	[C/G]	AA	AA	AA	BB	AB	BB	AA	BB
Chr7_19107296	0,44	7	19107296	[A/G]	BB	BB	AA	AB	AB	AA	BB	BB
Chr1_2616120	0,8672	1	2616120	[T/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr12_13695791	0,8203	12	13695791	[T/C]	AA	AA	BB	AA	AA	AA	BB	AA
Chr12_13494617	0,7763	12	13494617	[T/G]	AA	AA	BB	AA	AB	AA	BB	AA
Chr11_17108839	0,5036	11	17108839	[T/C]	AB	AB	AB	AB	AA	AA	AB	AA
Chr5_19170949	0,6606	5	19170949	[A/G]	AA	AA	BB	AA	AB	AA	BB	AA
Chr1_24381553	0,8006	1	24381553	[A/G]	BB	BB	BB	AA	AA	AA	BB	BB
Chr12_9818661	0,4081	12	9818661	[T/G]	BB	BB	AA	AB	AA	BB	AA	AA
Chr4_30261133	0,8187	4	30261133	[T/C]	BB	AA	BB	AA	BB	BB	AA	BB
Chr11_16979704	0,7438	11	16979704	[A/G]	BB	BB	BB	BB	BB	BB	BB	BB
Chr6_870656	0,696	6	870656	[A/G]	AA	AA	BB	BB	AB	BB	AA	BB
Chr1_23912753	0,6883	1	23912753	[T/C]	AA	AA	AA	BB	AA	BB	AA	AA
Chr11_17200311	0,7649	11	17200311	[G/C]	AA	AA	AA	AA	BB	BB	AA	BB
Chr10_23083676	0,841	10	23083676	[A/G]	BB	BB	AA	AA	AA	AA	BB	BB
Chr10_11682240	0,78	10	11682240	[A/G]	BB	AA	AA	BB	BB	BB	AA	AA
Chr1_628044	0,6551	1	628044	[C/G]	AA	AA	AA	AA	BB	BB	AA	AA
Chr12_10061622	0,6347	12	10061622	[T/C]	AA	AA	AB	AA	AB	AA	AB	AB
Chr12_13883943	0,6807	12	13883943	[T/A]	AA	AA	BB	AA	AA	AA	BB	AA
Chr6_24541268	0,94	6	24541268	[T/C]	AA	AA	BB	AA	AA	BB	AA	BB
Chr1_11312519	0,7191	1	11312519	[A/G]	AB	AB	AB	AB	AB	AB	BB	AB
Chr7_21220581	0,8236	7	21220581	[T/G]	AA	AA	AA	BB	BB	BB	AA	AA
Chr1_2326868	0,815	1	2326868	[A/G]	BB	BB	BB	AA	BB	AA	BB	BB
Chr8_344452	0,6144	8	344452	[T/C]	AA	AA	BB	BB	BB	BB	AA	BB
Chr8_8119838	0,6032	8	8119838	[T/C]	BB	AA	BB	AA	AA	BB	AA	BB

Chr2_16252795	0,6855	2	16252795	[A/G]	BB	BB	BB	AA	AA	AA	BB	BB
Chr12_2041641	0,6229	12	2041641	[A/G]	BB	AA	BB	BB	BB	AA	BB	AA
Chr6_1358478	0,7354	6	1358478	[C/G]	AA	AA	BB	AB	AB	BB	AA	BB
Chr12_3658931	0,8622	12	3658931	[T/C]	BB	BB	BB	AA	AA	AA	BB	BB
Chr12_9433836	0,4252	12	9433836	[C/G]	BB	AB	AA	AB	AB	BB	AA	AA
Chr5_28576950	0,8015	5	28576950	[T/C]	BB	BB	BB	BB	BB	AA	BB	AA
Chr7_4446115	0,562	7	4446115	[A/G]	AA	AA	AA	AB	AA	AB	AA	AA
Chr1_2403294	0,9797	1	2403294	[T/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr4_32515050	0,8361	4	32515050	[T/C]	AA	BB	AA	BB	AA	AA	BB	AA
Chr1_11653926	0,7042	1	11653926	[C/G]	BB	BB	AA	AA	BB	AA	BB	AA
Chr1_19195812	0,6819	1	19195812	[A/C]	AA	BB	BB	AA	AB	AA	BB	BB
Chr5_4817119	0,8478	5	4817119	[G/C]	AA	AA	AA	BB	AA	AA	BB	AA
Chr1_38340675	0,7303	1	38340675	[A/G]	AA	BB	BB	BB	BB	BB	AA	AA
Chr1_39440721	0,7965	1	39440721	[T/C]	BB	BB	BB	AA	BB	AA	BB	BB
Chr12_14239375	0,7312	12	14239375	[T/C]	BB	BB	BB	AA	AA	AA	BB	BB
Chr5_28759040	0,7581	5	28759040	[T/A]	AA	AA	AA	AA	AA	BB	AA	BB
Chr1_7425807	0,7198	1	7425807	[G/C]	BB	AA	AA	AA	AA	AA	BB	BB
Chr10_13179272	0,5551	10	13179272	[A/G]	BB	BB	BB	AA	AA	AA	BB	BB
Chr10_22507870	0,3996	10	22507870	[A/C]	AB	AB	AA	AA	AA	AA	AB	BB
Chr7_22533576	0,4225	7	22533576	[T/C]	AA	BB	AA	BB	BB	BB	AB	AA