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

**Diversidade, estruturação genética e mapeamento associativo em
germoplasma japonês de arroz utilizando marcadores DArT-seq**

Vanessa Rizzi

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

**Piracicaba
2017**

Vanessa Rizzi
Engenheira Agrônoma

Diversidade, estruturação genética e mapeamento associativo em germoplasma japonês de arroz utilizando marcadores DArT-seq
versão revisada de acordo com a resolução CoPGr 6018 de 2011

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

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

Piracicaba
2017

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Rizzi, Vanessa

Diversidade, estruturação genética e mapeamento associativo em germoplasma japonês de arroz utilizando marcadores DArT-seq / Vanessa Rizzi. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2017.

76 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. *Oryza sativa* L. ssp. *japonica* 2. Germoplasma 3. Diversidade genética 4. Estrutura populacional 5. Genotipagem por sequenciamento 6. GWAS I. Título

*À minha Mãe Rosania,
ao meu Pai Ibanez,
às minhas irmãs Tati, Dani e Alana,
ao Cesar, meu companheiro de vida.
Dedico*

AGRADECIMENTOS

À Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ/USP) e ao programa de Pós-graduação em Genética e Melhoramento de Plantas do Departamento de Genética pela oportunidade de realizar o curso de Doutorado;

Ao Prof. Dr. José Baldin Pinheiro pela orientação, convivência, conversas e por tudo que tive a oportunidade de aprender durante esses quatro anos;

Aos professores do Departamento de Genética pelo conhecimento e motivação para ensinar e aprender, vocês são pessoas e exemplos que vou levar para toda a vida;

À secretária do Departamento de Genética Candida Vanderleia de Oliveira que sempre me auxiliou com muita disposição e aos funcionários de campo pela colaboração durante a condução e colheita dos experimentos;

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de estudo;

À minha família maravilhosa, que apesar da distância torce para que eu consiga sempre ir mais longe e que acredita na educação como o meio de transformação das nossas vidas. Sem o sorriso de vocês em meu pensamento diário eu não teria chegado até aqui, obrigada por serem meu suporte, meu refúgio e meu impulso;

Ao meu companheiro e grande amor Cesar, que está sempre ao meu lado, apoiando seja com uma palavra, seja com um gesto ou um abraço amoroso. Agradeço todos os dias por você existir;

A toda família do Cesar, Olivia, Marco, Pedro, vó Isabel, vó Maria, vô Plínio, tias, tios, primos, madrinha, que me acolhem como uma filha torcendo e vibrando a cada conquista;

Aos colegas do Laboratório de Diversidade Genética e Melhoramento de Plantas pela convivência, cafés e conhecimentos compartilhados. Principalmente a Maisa, a Diane e ao João, meus grandes amigos que nunca mediram esforços para me ajudar.

Enfim, a todas as pessoas que de alguma forma participaram deste processo de aprendizagem e crescimento e que estarão sempre em meu coração. Muito obrigada!

*“Quando se tem um sonho
E vontade para fazer,
Não existe força ou pessoa,
Que não deixe acontecer.
Se chega onde se quer,
Basta ter amor e fé.”*

Autor desconhecido

SUMÁRIO

RESUMO	8
ABSTRACT	9
LISTA DE FIGURAS	10
LISTA DE FIGURAS	10
LISTA DE TABELAS	11
1. INTRODUÇÃO	13
REFERÊNCIAS.....	17
2. DIVERSIDADE E ESTRUTURA GENÉTICA EM ACESSOS JAPONESES DE ARROZ	21
RESUMO	21
ABSTRACT	21
2.1. INTRODUÇÃO.....	22
2.2. MATERIAL E MÉTODOS	24
2.2.1. Material vegetal.....	24
2.2.2. Extração de DNA.....	25
2.2.3. Genotipagem	25
2.2.4. Filtragem dos dados genotípicos	25
2.2.5. Prova de genotipagem.....	25
2.2.6. Diversidade genética da população	26
2.2.7. Estrutura genética da população	26
2.2.8. Análise de agrupamento.....	27
2.3. RESULTADOS.....	27
2.3.1. Marcadores	27
2.3.2. Diversidade genética.....	27
2.3.3. Estruturação genética.....	28
2.3.4. Agrupamento	30
2.3.5. Prova de genotipagem.....	31
2.4. DISCUSSÃO.....	32
2.4.1. Diversidade genética.....	32
2.4.2. Estrutura da população.....	33
2.4.3. Prova de genotipagem.....	35
2.5. CONCLUSÃO.....	35
REFERÊNCIAS.....	36
3. MAPEAMENTO ASSOCIATIVO DE CARACTERÍSTICAS AGRONÔMICAS EM ACESSOS JAPONESES DE ARROZ UTILIZANDO MARCADORES DART-SEQ (IN SILICO E SNP)	39
RESUMO	39
ABSTRACT	39

3.1. INTRODUÇÃO.....	40
3.2. MATERIAL E MÉTODOS.....	42
3.2.1. Material vegetal.....	42
3.2.2. Fenotipagem.....	42
3.2.3. Extração de DNA	42
3.2.4. Genotipagem DArTseq	42
3.2.5. Filtragem dos dados.....	43
3.2.6. Estrutura da população, matriz de parentesco e desequilíbrio de ligação	43
3.2.7. Mapeamento associativo	44
3.2.8. Anotação dos genes	44
3.3. RESULTADOS.....	45
3.3.1. Genotipagem	45
3.3.2. Estrutura da população, matriz de parentesco e desequilíbrio de ligação	46
3.3.3. Mapeamento associativo.....	48
3.3.4. Análise de associação utilizando marcadores SNPs.....	48
3.3.5. Análise de associação utilizando marcadores silico DArTs	51
3.3.6. Anotação dos genes relacionados	57
3.4. DISCUSSÃO	58
3.4.1. Matriz de parentesco e desequilíbrio de ligação	58
3.4.2. Mapeamento associativo	59
3.4.3. Anotação de genes	64
3.5. CONCLUSÃO.....	66
REFERÊNCIAS	66
APÊNDICES	71

RESUMO

Diversidade, estruturação genética e mapeamento associativo em germoplasma japonês de arroz utilizando marcadores DArT-seq

O conhecimento da diversidade genética e da estrutura populacional das variedades mantidas em bancos de germoplasma é de fundamental importância para sua efetiva utilização em programas de melhoramento. O mapeamento por associação, também conhecido como mapeamento por desequilíbrio de ligação, é um dos principais métodos para relacionar genes e alelos às características de interesse, através da co-segregação de marcadores genéticos polimórficos com os genes envolvidos na variação das características em estudo. O Banco de Germoplasma de Arroz do Departamento de Genética da ESALQ contém 192 acessos japoneses que foram estudados com o objetivo de entender sua diversidade, estruturação genética e determinar a associação genômica de caracteres agrônômicos relacionados a produção de grãos. A caracterização molecular foi conduzida através da tecnologia DArT-seq, que gerou dados de marcadores SNPs (single-nucleotide polymorphism) e silico DArTs. Em seguida, após a filtragem, 5.578 SNPs de alta qualidade foram utilizados para calcular as estimativas de diversidade no pacote *hierfstat* e a estrutura do painel de acessos através da análise discriminante de componentes principais (DAPC), que consiste em determinar existência de *cluster* em um grupo de genótipos em que não há informação *a priori* sobre existência de grupos. A diversidade genética nos acessos foi evidenciada pelo valor de heterozigosidade esperada (H_s) (0,0279) e a estruturação foi evidenciada pela formação de três subgrupos. O mapeamento associativo foi realizado com o uso do pacote GAPIT, sendo considerados seis caracteres: número de dias para florescimento (NDF), estatura de planta (EP), comprimento da panícula (CP), peso de parcela (PP), massa de mil grãos (MMG) e CICLO, bem como 24.266 marcadores silico DArTs e 1.965 marcadores SNPs. Foram detectadas um total de 113 associações significativas genótipo-fenótipo ($P < 0,001$) quando utilizado marcadores silico DArTs em todas as seis características analisadas e, um total de 21 associações significativas genótipo-fenótipo ($P < 0,001$) quando utilizado marcadores SNPs para apenas quatro das seis características analisadas: EP, CICLO, MMG e PP. Considerando-se os 113 silico DArTs associados significativamente na análise, 90 foram localizados em regiões intergênicas e 23 foram localizados dentro de genes. Enquanto que, dos 21 SNPs significativos, 11 foram localizados em regiões intergênicas e 10 foram localizados dentro de genes. A informação gerada neste estudo foi útil para testar associações ao longo do genoma do arroz. O modelo linear misto (MLM) empregado no mapeamento associativo acredita-se ter conseguido controlar eficientemente os falsos positivos no mapeamento utilizando os marcadores SNPs. As informações geradas neste estudo servem de base para avaliações mais aprofundadas, utilizando o conjunto de marcadores significativos como ponto de partida para determinação dos genes mais importantes para a produtividade em arroz.

Palavras-chave: *Oryza sativa* L.ssp. *japonica*; Germoplasma; Diversidade genética; Estrutura populacional; Genotipagem por sequenciamento; GWAS

ABSTRACT

Diversity, genetic structuring and association mapping in Japanese rice germoplasm using DArT-seq markers

The knowledge of the genetic diversity and population structure of varieties maintained in germplasm banks is crucial for their effective use in breeding programs. Association mapping, also known as linkage disequilibrium mapping, is one of the main methods for relating genes and alleles to the characteristics of interest, through the co-segregation of polymorphic genetic markers with the genes involved in the variation of the characteristics under study. The Rice Germplasm Bank of the Department of Genetics of ESALQ contains 192 Japanese accessions that were studied with the purpose of understanding its diversity, genetic structuring and determining the genomic association of agronomic traits related to grain production. The molecular characterization was conducted by DArTseq technology, which generated data of SNPs (single-nucleotide polymorphism) markers and silico DArTs. Then, after filtering, 5,578 high-quality SNPs were used to calculate the diversity estimates in hierfstat package and the accession panel structure through discriminant analysis of principal components (DAPC), which consists of determining the cluster existence in a group of genotypes where there is no a priori information about the existence of groups. The genetic diversity in the accessions was evidenced by the expected heterozygosity value (HS) (0.0279) and the population structure was evidenced by the formation of three clusters. The association mapping was performed using the GAPIT package, considering six characters: number of days for flowering (NDF), plant height (EP), panicle length (CP), plot weight (PP), mass of thousand grains (MMG) and CYCLE, as well as 24,266 silico DArTs markers and 1,965 SNPs markers. We detected a total of 113 significant associations genotype-phenotype ($P < 0.001$) when used silico DArTs markers in all six analyzed characteristics and a total of 21 significant associations genotype-phenotype ($P < 0.001$) when used SNPs markers for only four of the six analyzed characteristics: EP, CYCLE, MMG and PP. Considering the 113 silico DArTs significantly associated in the analysis, 90 were located in intergenic regions and 23 were localized within genes. While of the 21 significant SNPs, 11 were located in intergenic regions and 10 were located within genes. The information generated in this study was useful for testing associations throughout the rice genome. The mixed linear model (MLM) used in association mapping is believed to have been able to efficiently control false positives in the mapping using the SNPs markers. The information generated in this study serves as a basis for further evaluation using the set of significant markers as a starting point for determining the most important genes for rice yield.

Keywords: *Oryza sativa* L.ssp. *japonica*; Germplasm; Genetic diversity; Population structure; Genotyping by sequencing; GWAS

LISTA DE FIGURAS

- Figura 1. Análise discriminante dos componentes principais (DAPC) com 1 função discriminante nos 192 acessos do germoplasma de arroz, mais 10 controles e 8 PGs estruturados em 3 agrupamentos. Os agrupamentos foram nomeados como S_1 , S_2 e S_3 29
- Figura 2. *Scatterplot* da análise discriminante dos componentes principais (DAPC) com 2 funções discriminantes nos 192 acessos do germoplasma de arroz, mais 10 controles e 8 PGs estruturados em 3 agrupamentos. Os agrupamentos foram nomeados como S_1 , S_2 e S_3 . Inserido na parte superior direita o *Bayesian information criterion* (BIC) e na parte inferior o *Discriminant analysis eigenvalues* 29
- Figura 3. *Barplot* da análise discriminante dos componentes principais (DAPC) nos 192 acessos do germoplasma de arroz, mais 10 controles e 8 PGs estruturados em 3 agrupamentos e identificados em cada um dos subgrupos..... 30
- Figura 4. Dendrograma gerado pela análise de agrupamento de 202 genótipos obtidos através do método UPGMA com base na matriz de distâncias de (Rogers, 1972) modificada por (Wright, 1978)..... 31
- Figura 5. *Kinship* calculado pelo algoritmo *loiselle* utilizando marcadores SNPs e marcadores silico DArTs 47
- Figura 6. Decaimento do desequilíbrio de ligação conforme a distância genética entre os pares de marcadores calculado pelo pacote GAPIT através de marcadores SNPs e marcadores silico DArTs..... 47
- Figura 7. *Manhattan plot* e *Q-Q plot* da análise de associação para as características agronômicas EP, CICLO, MMG E PP utilizando 1.965 marcadores SNPs em 186 acessos de arroz da subespécie *japonica* 49
- Figura 8. *Manhattan plot* e *Q-Q plot* da análise de associação para as características agronômicas EP, CICLO, MMG utilizando 24.266 marcadores silico DArTs em 186 acessos de arroz da subespécie *japonica* (continua)..... 52
- Figura 8. *Manhattan plot* e *Q-Q plot* da análise de associação para as características agronômicas PP, NDF E CP utilizando 24.266 marcadores silico DArTs em 186 acessos de arroz da subespécie *japonica* (conclusão) 53

LISTA DE TABELAS

Tabela 1. Índices de diversidade genética	28
Tabela 2. Número de marcas SNPs e silico DArTs identificadas entre os 186 acessos de <i>Oryza sativa</i> L. spp. <i>japonica</i> em cada um dos cromossomos e total, imagem de cariótipo modificado de http://ensembl.gramene.org/oryza_sativa/location/genome	46
Tabela 3. Marcadores SNPs que apresentaram associação com as características de produtividade EP, CICLO, MMG e PP, identificados entre os 186 acessos de <i>Oryza sativa</i> spp. <i>japonica</i> . São apresentados: as marcas significativamente associadas (SNPs), o cromossomo, a posição do loco SNP no cromossomo (genoma de referência rice_v9 - phytosome assembly v9), o <i>p</i> -valor, a mínima frequência alélica (maf), a estimativa da variação fenotípica em % devido ao efeito do marcador (R^2), os alelos de cada loco SNP, a região gênica na qual o snp está inserido e o transcrito com base no <i>international rice genome sequencing project</i> (IRGSP).....	50
Tabela 4. Marcadores silico DArTs que apresentaram associação com as características de produtividade EP, CICLO, MMG, PP e CP, identificados entre os 186 acessos de <i>Oryza sativa</i> L. spp. <i>japonica</i> . São apresentados: as marcas significativamente associadas (SNPs), o cromossomo, a posição do loco SNP no cromossomo (genoma de referência rice_v9 - phytosome assembly v9), o <i>p</i> -valor, a mínima frequência alélica (MAF), a estimativa da variação fenotípica em % devido ao efeito do marcador (R^2), os alelos de cada loco SNP, a região gênica na qual o SNP está inserido e o transcrito com base no <i>international rice genome sequencing project</i> (IRGSP) (continua)	55
Tabela 5. Marcadores silico DArTs que apresentaram associação com a característica de produtividade NDF, identificados entre os 186 acessos de <i>Oryza sativa</i> L. spp. <i>japonica</i> . São apresentados: as marcas significativamente associadas (SNPs), o cromossomo, a posição do loco SNP no cromossomo (genoma de referência rice_v9 - phytosome assembly v9), o <i>p</i> valor, a mínima frequência alélica (MAF), a estimativa da variação fenotípica em % devido ao efeito do marcador (R^2), os alelos de cada loco SNP, a região gênica na qual o SNP está inserido e o transcrito com base no <i>international rice genome sequencing project</i> (IRGSP) (conclusão)	56

1. INTRODUÇÃO

O arroz (*Oryza sativa* L.) é uma planta cultivada pertencente ao gênero *Oryza*, tribo *Oryzae*, subfamília *Pooideae*, família *Poaceae*. O gênero *Oryza* é considerado o mais importante da tribo e compreende aproximadamente 22 espécies que estão distribuídas pelos continentes da Ásia, África, América do Sul e América Central. Dentre essas espécies apenas duas são cultivadas: *Oryza sativa* L. e *Oryza glaberrima* L., as quais são diplóides e constituídas pelo genoma AA, com número básico de cromossomos $2n=24$ (SIMMONDS, 1976) (CHANG, 1976).

A origem do arroz ocorreu provavelmente na Índia, e a sua propagação na China data de mais de 3.000 anos antes da era Cristã (FORNASIERI, 1993). A espécie selvagem *Oryza rufipogon* é o ancestral mais provável de *Oryza sativa*, por ser uma espécie perene e de fecundação cruzada e de onde deve ter originado o hábito semi-aquático dos genótipos de arroz cultivado (CHANG, 1976).

A espécie *Oryza sativa* L. divide-se em duas subespécies denominadas *indica* e *japonica*. Os cultivares da subespécie *indica* são predominantemente cultivados nos trópicos e subtrópicos e os da subespécie *japonica* em regiões temperadas. Existe ainda uma terceira subespécie denominada *javanica*, cujos genótipos são cultivados em planícies tropicais da Indonésia. Alguns autores classificam a subespécie *javanica* como pertencente à subespécie *japonica*, sendo considerada como *japonica* tropical (MACKILL; LEI, 1997)

Essas subespécies são diferenciadas através de um conjunto de caracteres que abrangem desde aspectos morfológicos até reações a estresses abióticos. A maioria dos cultivares da subespécie *indica* são geralmente sensíveis ao frio enquanto que os da subespécie *japonica* são classificados como tolerantes a baixas temperaturas. Outras características, como a relação comprimento/largura dos grãos, são usadas para identificar as duas subespécies. Os cultivares da subespécie *indica* apresentam grãos mais longos e finos e os da subespécie *japonica* geralmente apresentam grãos largos e curtos (MATSUO et al., 1993).

O genoma da espécie *Oryza sativa* L. é um genoma relativamente pequeno, de 466 Mb na subespécie *indica*, com aproximadamente 46.022 a 55.615 genes (YU et al., 2002) e 420 Mb na subespécie *japonica* com estimados 32.000 a 50.000 genes (GOFF et al., 2002).

A classificação do arroz nessas duas subespécies é de grande importância para os melhoristas, pois indica a adaptação a ambientes específicos e características de atratividade

dos consumidores, influenciando nas estratégias de melhoramento para utilização de germoplasma (MACKILL; LEI, 1997)

O arroz foi domesticado provavelmente no nordeste da Índia e no sul da China há cerca de 8000 anos (GOWDA et al., 2003). Devido à contínua seleção do arroz selvagem para o desenvolvimento de variedades de alto rendimento e bem adaptadas, a cultura foi se tornando cada vez mais homogênea, como resultado de um fenômeno denominado erosão genética, que por sua vez reduziu a variabilidade no arroz e pode acarretar sérios danos à agricultura mundial, aumentando a vulnerabilidade da cultura a doenças e insetos (GOWDA et al., 2003); (HOYT, 1992).

No Brasil, sugere-se que apenas sete genitores tenham contribuído com 70% do conjunto gênico das variedades cultivadas de arroz irrigado (RANGEL; GUIMARÃES; NEVES, 1996), dentro desta realidade, a conservação de acessos de arroz em bancos de germoplasma assume papel essencial (GILBERT et al., 1999).

A viabilidade genética inter e intra-específica representativa dos recursos genéticos vegetais é, em grande parte manejada e organizada nos bancos de germoplasma, que se constituem em estrutura física, onde as coleções são conservadas na forma de células, tecidos, sementes ou plantas (NASS, L. L.; NISHIKAWA, M. A. N.; FÁVERO, A. P.; LOPES, 2007).

A conservação dos recursos genéticos das plantas cultivadas, assim como de seus parentes silvestres, ao longo do último século se tornou uma das questões de grande importância entre melhoristas e cientistas de áreas afins (CALLOW; FORD-LLOYD; NEWBURY, 1997), pois os bancos de germoplasma são extremamente importantes como fontes de genes relacionados a características de interesse para incorporação em materiais domesticados ou geneticamente melhorados.

Os alelos desejáveis para o melhoramento estão disponíveis no genoma do arroz, mas dispersos nos acessos dos germoplasmas. O trabalho mais importante é conhecer sistematicamente onde todos esses genes estão presentes no genoma e atribuir suas funções (GOWDA et al., 2003).

Neste sentido, o “pré-melhoramento” tem sido uma das principais alternativas para o conhecimento e ampliação da base genética nas mais diversas culturas, tais como o arroz, pois busca conhecer a variabilidade genética que os acessos armazenados em bancos de germoplasma apresentam. Estudos nesta área são desenvolvidos para criação de linhagens provenientes de cruzamentos entre genitores de base genética mais ampla, as quais, posteriormente, são cruzadas com germoplasma elite, sem romper os blocos gênicos que conferem fenótipos favoráveis (BRONDANI et al., 2003).

Existem nos bancos de germoplasma algumas variedades tradicionais e também acessos menos conhecidos que são mantidos pelos pesquisadores ao longo dos anos, pois estabelecem um importante reservatório de diversidade genética e fonte de alelos de interesse para o melhoramento de arroz. E sua utilização, certamente, é uma estratégia importante para incorporação de variabilidade em programas de melhoramento, podendo gerar novas cultivares através do aumento da base genética e obtenção de novas combinações alélicas (MCCOUCH, 2004).

O melhoramento de plantas e o desenvolvimento de variedades envolvem a manipulação e seleção de combinações de características agrônômicas que produzem cultivares superiores. Tradicionalmente, o melhoramento de plantas é responsável por grandes aumentos na produtividade das culturas e para isso necessitam comparar milhares de plantas geneticamente distintas em ambientes múltiplos para selecionar plantas superiores em um processo que requer muito tempo e esforço (GOWDA et al., 2003).

Muitos caracteres de importância agrícola, como rendimento, qualidade e algumas formas de resistência a doenças, são controlados por muitos genes e são conhecidos como caracteres quantitativos (também caracteres "poligênicos", "multifatoriais" ou "complexos"). As regiões dentro dos genomas que contêm genes associados a uma característica quantitativa particular são conhecidas como loco de característica quantitativa (QTLs). A identificação desses QTLs baseada apenas na avaliação fenotípica convencional não é possível. Dessa forma, um grande avanço na caracterização de características quantitativas que criaram oportunidades para selecionar QTLs de interesse foi iniciado pelo desenvolvimento dos marcadores de DNA (ou moleculares) na década de 1980 (COLLARD et al., 2005).

Desde então, o uso de marcadores moleculares na cultura do arroz vem possibilitando o estudo da diversidade genética dentro da espécie (ZHU et al., 1998), a caracterização molecular de linhagens (GARLAND et al., 1999), a identificação de subespécies (XU et al., 1998), a construção de mapas altamente saturados (CAUSSE et al., 1994) e a identificação de características agronomicamente importantes (PANTALIO et al., 2016) (BISCARINI et al., 2016) (MWADZINGENI et al., 2017).

Os marcadores encontrados nesses estudos e que estão estreitamente ligados a genes de importância agrônômica podem ser posteriormente utilizados como ferramentas para a seleção assistida por marcadores (MAS) no melhoramento de plantas (RIBAUT; HOISINGTON, 1998). Estratégia que visa auxiliar a seleção fenotípica tornando-a mais eficiente quando comparado com as metodologias de melhoramento mais convencionais.

As marcas no DNA representam as diferenças genéticas entre organismos ou espécies, no entanto, não representam em sua totalidade os genes alvos, servindo como “sinais” da existência destes. Os marcadores que estão localizados próximos aos genes podem ser chamados de “etiqueta de genes” e não afetam o fenótipo da característica de interesse, pois estão localizados apenas perto do gene ou dos genes que controlam a característica de interesse (COLLARD et al., 2005).

Diversos marcadores moleculares foram desenvolvidos ao longo da história e têm sido aplicados no mapeamento genético e nos estudos de biodiversidade de arroz cultivado (ZHU et al., 1998). Em 2008, uma variante da tecnologia DArT foi desenvolvida baseada no mesmo princípio de redução da complexidade do genoma aliado à tecnologia de genotipagem por sequenciamento de última geração ou NGS (Next Generation Sequencing), o DArT-seq (KILIAN et al., 2005). Em comparação à primeira tecnologia DArT, a DarT-seq possui algumas vantagens como por exemplo, a geração de uma maior densidade de marcadores (dezenas de milhares de marcadores) (CRUZ; KILIAN; DIERIG, 2013) e adequação para trabalhos de mapeamento de alta resolução e de genética associativa. Em relação a outras metodologias de genotipagem por sequenciamento, a principal diferença está em como é realizada a redução da complexidade do DNA genômico. A partir do alinhamento, são detectados dois tipos de polimorfismos: o primeiro, denominado *silico DArT*, que é gerado pela variabilidade da distribuição dos sítios de restrição entre as amostras dos indivíduos analisados, observado na forma de presença ou ausência de um determinado fragmento (*tag*) de DNA codificado de forma binária: a presença de fragmento comum entre as amostras (1) e ausência de fragmento em comum entre as amostras (0); e o segundo: chamado de SNPs (*single-nucleotide polymorphisms*) que são polimorfismos resultantes das variações de bases únicas ao longo das sequências. Os SNPs são detectados a partir do alinhamento dos *reads* de cada *tag* em comum entre as amostras, formando *clusters* de *reads*, de forma que é selecionada a variante alélica mais abundante dentro de cada *cluster*, a qual é considerada referência, e a variante menos abundante é considerada um SNP. Sendo estas variantes classificadas pela presença/ausência das mesmas dentro de cada *cluster* formado, referente às *tags* analisadas (SANSALONI; GRATAPAGLIA, 2012).

Com a disponibilidade pública da informação da sequência do genoma do arroz, há um crescente interesse em identificar e caracterizar genes associados com formas qualitativas e quantitativas das variações fenotípicas. De particular interesse para os melhoristas de arroz, é a possibilidade de usar recursos de germoplasma existentes para a descoberta de genes e alelos com base em estratégias de mapeamento associativo (LU et al., 2005).

Dentro deste contexto, o Departamento de Genética da Escola Superior de Agricultura “Luiz de Queiroz” ao longo dos anos armazenou germoplasma de arroz oriundo de algumas instituições de pesquisa do Brasil e do exterior, como acessos filipinos, goianos e japoneses do Instituto de Arroz de Sequeiro do Japão (atualmente extinto). Ao todo estão armazenados 460 acessos disponíveis para estudos de diversidade genética e para inclusão em programas de melhoramento do arroz. O laboratório de diversidade genética e melhoramento de plantas vem desenvolvendo projetos voltados ao conhecimento aprofundado dos 192 acessos da subespécie *japonica* armazenados nesse banco.

O presente trabalho teve como objetivo realizar um estudo da diversidade, estruturação genética e mapeamento por associação nos 192 acessos da subespécie *japonica* de arroz, através de dados genotípicos da tecnologia de genotipagem por sequenciamento DArT-seq e dados fenotípicos de produtividade. Com o objetivo de melhor entendimento da diversidade e estrutura genética desses acessos, bem como o conhecimento de regiões genômicas, ou locos, responsáveis pelo controle de características relacionadas a produtividade e que poderiam ser utilizadas em programas de melhoramento.

Referências

- BISCARINI, F. et al. Genome-wide association study for traits related to plant and grain morphology, and root architecture in temperate rice accessions. **PLoS ONE**, v. 11, n. 5, p. 1–28, 2016.
- BRONDANI, C. et al. Transferability of microsatellite and sequence tagged site markers in *Oryza* species. **Hereditas**, v. 138, n. 3, p. 187–192, 2003.
- CALLOW, J. A.; FORD-LLOYD, B. V.; NEWBURY, H. J. Biotechnology and plant genetic resources: conservation and use. In: **Biotechnology in Agriculture series, No 19**. [s.l.] CAB International, 1997. p. 308.
- CAUSSE, M. A. et al. Saturated molecular map of the rice genome based on an interspecific backcross population. **Genetics**, v. 138, n. 4, p. 1251–1274, 1994.
- CHANG, T. T. The origin, evolution, cultivation, dissemination, and diversification of Asian and African rices. **Euphytica**, v. 25, n. 1, p. 425–441, 1976.
- COLLARD, B. C. Y. et al. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. **Euphytica**, v. 142, n. 1–2, p. 169–196, 2005.

- CRUZ, V. M. V.; KILIAN, A.; DIERIG, D. A. Development of DArT Marker Platforms and Genetic Diversity Assessment of the U.S. Collection of the New Oilseed Crop *Lesquerella* and Related Species. **PLoS ONE**, v. 8, n. 5, p. 1–13, 2013.
- FORNASIERI, D. F. **Manual da Cultura do Arroz**. [s.l.] FUNEP, 1993.
- GARLAND, S. H. et al. The use of microsatellite polymorphisms for the identification of Australian breeding lines of rice (*Oryza sativa* L.). **Euphytica**, v. 108, n. 1, p. 53–63, 1999.
- GILBERT, J. E. et al. Developing an appropriate strategy to assess genetic variability in plant germplasm collections. **Theoretical and Applied Genetics**, v. 98, n. 6–7, p. 1125–1131, 11 maio 1999.
- GOFF, S. A. et al. A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *japonica*). **Science**, v. 296, n. 5565, p. 92–100, 5 abr. 2002.
- GOWDA, M. et al. Advances in rice breeding, genetics and genomics. **Molecular Breeding**, v. 11, n. 5, p. 337–352, 2003.
- HOYT, E. **Conservação dos parentes silvestres das plantas cultivadas**. [s.l.] Addison-Wesley Iberoamericana; Embrapa/Cenargen, 1992.
- KILIAN, A. et al. The fast and the cheap SNP and DArT-based whole genome profiling for crop improvement. **Proceedings of the International Congress**, v. 1, n. 5, p. 443–461, 2005.
- LU, H. et al. Population Structure and Breeding Patterns of 145 U. S. Rice Cultivars Based marker analysis. **Crop Science**, v. 45, n. i, p. 66, 2005.
- MACKILL, D. J.; LEI, X. Genetic Variation for Traits Related to Temperate Adaptation of Rice Cultivars. **Crop Science**, v. 37, n. 4, p. 1340, 1997.
- MATSUO, T. et al. **Science of the Rice Plant**. [s.l.] Food and Agriculture Policy Research Center, 1993.
- MCCOUCH, S. **Diversifying selection in plant breeding** **PLoS Biology**, 2004. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC521731/pdf/pbio.0020347.pdf>>. Acesso em: 26 jun. 2017
- MWADZINGENI, L. et al. Genome-wide association analysis of agronomic traits in wheat under drought-stressed and non-stressed conditions. **PLoS ONE**, v. 12, n. 2, p. 1–13, 2017.
- NASS, L. L.; NISHIKAWA, M. A. N.; FÁVERO, A. P.; LOPES, M. A. Pré-melhoramento de germoplasma vegetal. In: **Recursos genéticos vegetais**. [s.l.] Embrapa Recursos Genéticos e Biotecnologia (CENARGEN), 2007. p. 683–716.

- PANTALIO, G. F. et al. Genome wide association study (GWAS) for grain yield in rice cultivated under water deficit. **Genetica**, v. 144, n. 6, p. 651–664, 2016.
- RANGEL, P. H. N.; GUIMARÃES, E. P.; NEVES, P. C. F. **Base genética das cultivares de arroz (Oryza sativa L.) irrigado do Brasil** Pesquisa Agropecuaria Brasileira, 1996.
- RIBAUT, J.-M.; HOISINGTON, D. Marker-assisted selection: new tools and strategies. **Trends in Plant Science**, v. 3, n. 6, p. 236–239, 1998.
- SANSALONI, C. P.; GRATTAPAGLIA, O. D. **Desenvolvimento e aplicações de DArT (Diversity Arrays Technology) e genotipagem por sequenciamento (Genotyping-by-Sequencing) para análise genética em Eucalyptus**. [s.l.] Universidade de Brasília, 2012.
- SIMMONDS, N. W. **Evolution of crop plants**. Longman ed. London & New York: Longman, 1976.
- XU, J. et al. Classification of rice germplasm. II. Discrimination of indica from japonica via analysis of amplicon length polymorphisms. **Plant Cell Reports**, v. 17, n. 8, p. 640–645, 5 maio 1998.
- YU, J. et al. A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. indica). **Science**, v. 296, n. 5565, p. 79–92, 2002.
- ZHU, J. et al. AFLP markers for the study of rice biodiversity.pdf. **Theoretical and Applied Genetics**, v. 96, n. 7, p. 602–611, 1998.

2. DIVERSIDADE E ESTRUTURA GENÉTICA EM ACESSOS JAPONÊS DE ARROZ

RESUMO

A avaliação dos recursos genéticos das espécies armazenadas em bancos de germoplasma é essencial para assegurar fontes de diversidade em trabalhos de melhoramento de plantas e também para a conservação dessas espécies. Um painel contendo 192 acessos de arroz da subespécie *japonica* pertencentes ao Banco de Germoplasma de Arroz do Departamento de Genética da ESALQ foi analisado com o objetivo de verificar a diversidade genética existente e sua estruturação. Para isso foi utilizado o método de genotipagem por sequenciamento DArT-seq que gerou, após a filtragem, 5.578 SNPs de alta qualidade. A diversidade genética nos acessos foi evidenciada pelo valor de H_S e a estruturação foi evidenciada pela formação de três subgrupos, identificados por meio da DAPC. Os valores encontrados para H_O e H_S foram de 0,0279 e 0,2553, respectivamente, e o valor de F_{IS} foi de 0,8906. Valores de H_O menores que de H_S são esperados para espécies como arroz, feijão e soja, pois se tratam de espécies com baixa taxa de fecundação cruzada e, portanto, um germoplasma com alta diversidade genética tende a ser uma mistura de linhas “puras”. Pela mesma razão encontramos o alto valor para F_{IS} . A estrutura genética encontrada entre os genótipos era esperada, visto que os acessos são oriundos de diferentes regiões do Japão e passaram por diferentes processos de seleção.

Palavras chave: *Oryza sativa* L. ssp. *japonica*; Germoplasma; SNP; DArT-seq; DAPC

ABSTRACT

The evaluation of genetic resources of species maintained in germplasm banks is essential to ensure sources of diversity in plant breeding works and also to species conservation. A panel containing 192 rice accessions from *japonica* subspecies belonging to the Rice Germplasm Bank of the Department of Genetics of ESALQ was analyzed in order to verify the genetic structuring and diversity. For this, the DArT-seq genotyping by sequencing method was used which generated, after filtering, 5,578 high-quality SNPs. Genetic diversity in the accessions was evidenced by the H_S value and the structure was evidenced by the formation of three clusters, identified using the DAPC. The values found for H_O and H_S were 0.0279 and 0.2553, respectively, and the F_{IS} value was 0.8906. H_O values lower than H_S are expected for species such as rice, beans and soybeans, since these are species with low cross-fertilization rate and, therefore, a germplasm with high genetic diversity tends to be a mixture of "pure" lines. For the same reason we find the high value for F_{IS} . The genetic structure found among genotypes was expected, as the hits are from different regions of Japan and have gone through different selection processes.

Keywords: *Oryza sativa* L. ssp. *japonica*; Germplasm; SNP; DArT-seq; DAPC

2.1. Introdução

O arroz (*Oryza sativa* L.) é um alimento básico para mais da metade da população mundial e sua produção gira em torno de 490 milhões de toneladas anualmente, deste montante, aproximadamente 90% é produzido na Ásia (NACHIMUTHU et al., 2015). Essa alta produção tem sido registrada anualmente devido a vários fatores, dentre eles, os ganhos genéticos registrados em programas de melhoramento. Entretanto, por consequência da seleção concentrada em algumas poucas características as variedades melhoradas têm se tornando cada vez mais homogêneas, como resultado de um fenômeno denominado erosão genética, que pode acarretar sérios danos à agricultura, dentre eles aumentar a vulnerabilidade das plantas.

Especula-se que esse processo de seleção e melhoramento ao longo das décadas para determinadas características agronômicas possa ter levado à perda ou ao esquecimento de genes que poderiam estar envolvidos com outras características (SWARUP et al., 1995) (FLINT-GARCIA; BODNAR; SCOTT, 2009). Nos programas de melhoramento genético e na agricultura, essa perda de genes está ocorrendo com diferentes intensidades, conforme a espécie. Em vários casos, todavia, a situação atual parece ser já bastante crítica (VENCOVSKY, 1987). É preciso lembrar, ainda, que a deriva tem efeito cumulativo, no sentido de que o desvio padrão das frequências alélicas vai crescendo com o decorrer das gerações (FALCONER, 1964).

Concomitantemente a esta perda de genes e, por conseguinte o estreitamento da base genética das variedades melhoradas surgiu uma grande necessidade de diversificação. Desta forma, a caracterização dos genótipos disponíveis é uma importante tarefa para a avaliação da diversidade genética. Estudos desta natureza podem auxiliar na revelação de importantes fontes de variabilidade genética para os melhoristas (SINGH, 1989).

A disponibilidade de variabilidade genética, o conhecimento sobre as características desejadas pelo mercado consumidor e produtores e, eficientes estratégias de seleção que permitam a exploração e utilização de recursos genéticos são as bases do sucesso em programas de melhoramento de plantas. E esses recursos genéticos, muitas vezes, encontram-se estocados em bancos de germoplasma em várias partes do mundo, abrigando uma ampla gama de alelos (HOISINGTON et al., 1999).

Os bancos de germoplasma são extremamente importantes, pois são a principal fonte de variabilidade em muitos programas de melhoramento de plantas. Eles servem como um reservatório para a identificação de alelos superiores que controlam as principais

características agronômicas (NACHIMUTHU et al., 2015). Porém, o uso dos materiais presentes nos bancos de germoplasma e, conseqüentemente, o desenvolvimento de estudos de diversidade genética por vezes é colocado em segundo plano, correndo-se o risco da perda de alelos pelo descarte de material não caracterizado e, como resultado, perda de variabilidade genética das espécies armazenadas. Além disso, a questão de coletar e preservar germoplasma não pode ser vista apenas sob a ótica de alelos individuais. Conjuntos ou blocos gênicos também têm grande importância para o melhoramento em que o objetivo é desenvolver genótipos com capacidade de adaptação a condições extremas ou atípicas para a espécie (VENCOVSKY, 1987).

De maneira geral, o nível de uso dos bancos genéticos é muito baixo, girando em torno de 5% (NASS, 1993). A utilização destes acessos contidos em bancos de germoplasma só deverá aumentar quando esse material disponibilizar níveis expressivos de informações genéticas e utilitárias, relacionadas à adaptação ambiental, potencial genético, potencial industrial, conhecimento etnobiológico e potencial socioeconômico. Neste sentido, o estudo e conservação de genótipos disponíveis em bancos de germoplasma tem sido uma das principais alternativas para a tentativa de ampliação da base genética nas mais diversas culturas.

O arroz é considerado uma espécie que apresenta grande diversidade genética em nível inter e intra-específico, de tal maneira o conhecimento a respeito da extensão da variação genética e a relação genética entre genótipos são vitais para a concepção de estratégias que afetam o melhoramento e sua conservação (ROY et al., 2015); (SOW et al., 2014).

Sendo os bancos de germoplasma os lugares mais apropriados para iniciar esse processo de investigação é necessário definir, dentre as várias metodologias existentes, qual utilizar. A diversidade genética pode ser analisada com o auxílio de marcadores agromorfológicos e moleculares. Os marcadores moleculares podem ajudar na estimativa da variabilidade genética global (BRONDANI et al., 2003) e nos últimos anos têm sido muito utilizados para explorar a diversidade genética conservada nos bancos de germoplasma, devido sua praticidade e qualidade nos resultados apresentados.

Um exemplo é a aplicação de genotipagem por sequenciamento (GBS) no melhoramento, conservação e pesquisas populacionais, que permite aos melhoristas de plantas conduzirem pesquisas em germoplasma ou espécies sem ter que desenvolver quaisquer ferramentas moleculares anteriormente ou mesmo, permite que conservacionistas possam

determinar a estrutura da população sem o conhecimento prévio do genoma ou da diversidade das espécies (ELSHIRE et al., 2011).

Neste contexto, para estudo de acessos contidos em banco de germoplasma, optamos pela utilização de marcadores SNP gerados a partir da tecnologia DArT-seq de genotipagem por sequenciamento (GBS), que tem apresentado ótimos resultados em trabalhos relacionados.

O Departamento de Genética da Escola Superior de Agricultura “Luiz de Queiroz” – USP ao longo dos anos armazenou germoplasma de arroz oriundo de algumas instituições de pesquisa do Brasil e do exterior e entre esses materiais encontram-se diversos acessos da subespécie *japonica*, oriundos do Instituto de Arroz de Sequeiro do Japão, atualmente extinto. Tais acessos estão passando por uma extensa pesquisa científica que tem a intenção de caracterização para se obter informações a respeito de sua origem, diversidade e estruturação genética.

O estudo da diversidade e estruturação genética em um painel contendo 192 acessos de arroz da subespécie *japonica*, pertencentes ao Banco de Germoplasma de Arroz do Departamento de Genética da ESALQ, foi realizado utilizando a genotipagem com a tecnologia por sequenciamento DArT-seq.

2.2. Material e métodos

2.2.1. Material vegetal

O painel utilizado neste estudo foi composto por 192 acessos de arroz da subespécie *japonica* (Apêndice A) que fazem parte do banco de germoplasma do Departamento de Genética e Melhoramento de Plantas da Escola Superior de Agricultura “Luiz de Queiroz” – Campus USP de Piracicaba, São Paulo. Esses acessos têm a característica de materiais de sequeiro, também conhecidos como arroz de terras altas. Além disso, 10 variedades comerciais da subespécie *indica* foram utilizadas como controle nas análises de agrupamento e estruturação para verificar a existência de possível mistura no painel, sendo 5 variedades da Embrapa Arroz e Feijão e 5 variedades da Epagri. Para cada acesso as sementes foram produzidas a partir de uma única espiguetta em casa de vegetação, na Escola Superior de Agricultura “Luiz de Queiroz”.

2.2.2. Extração de DNA

O DNA genômico foi extraído em duplicatas a partir de folhas jovens coletadas de uma única planta de cada um dos acessos utilizando o método de extração CTAB (DOYLE; DOYLE, 1987). Posteriormente, para verificação da qualidade e quantidade o DNA foi diluído em TE (Tris-HCl 10 μ L, EDTA 1 μ L, com pH 8,0) em volume entre 50 μ L e 100 μ L e as amostras foram aplicadas em gel de agarose 0,8%. A integridade do DNA foi atestada pelo aparecimento de apenas uma banda de elevado peso molecular sobre o gel e a concentração foi estimada com base no marcador de peso molecular DNA de fago λ .

2.2.3. Genotipagem

A genotipagem foi conduzida para 192 acessos japoneses de arroz mais as 10 testemunhas pela empresa Diversity Arrays Technology Pty Ltd. (DArT P/L) da Austrália, através da tecnologia de genotipagem por sequenciamento DArT-seq™, previamente descrita por (COURTOIS et al., 2013), que combina GBS com Diversity Arrays Technology (DArT), envolvendo a redução da complexidade do genoma usando enzimas de restrição PstI/TaqI seguido pelo sequenciamento de sequências curtas através do Illumina HiSeq 2500 (<http://www.diversityarrays.com/dart-application-dartseq-data-types>).

2.2.4. Filtragem dos dados genotípicos

Para obtenção de estimativas mais fidedignas, julgou-se necessária a seleção dos SNPs com a melhor qualidade através da aplicação dos seguintes filtros: profundidade de cobertura (*Read Depth*) ≥ 3 ; *Call Rate* $\geq 0,8$; Conteúdo de informação polimórfica (PIC) $\geq 0,1$; e, remoção de indivíduos com mais de 40% de dados perdidos.

2.2.5. Prova de genotipagem

A reprodutibilidade dos resultados de qualquer análise é um dos pontos a ser considerado antes de se iniciar um projeto. Com os dados do sequenciamento de nova geração não é diferente. Assim, após uma pesquisa sobre os métodos mais utilizados para sequenciamento e também, aqueles com histórico de sucesso para a espécie em estudo, optamos por utilizar a tecnologia de sequenciamento de nova geração DArTseq™. Para

avaliar a reprodutibilidade do método, fizemos uma prova de genotipagem que consistiu em genotipar por duas vezes seguidas, mas em tempos diferentes, 8 acessos do nosso painel de estudo. Os acessos genotipados pela segunda vez foram chamados de Prova de Genotipagem (PG) e estão identificados no trabalho com o mesmo número do BGA em questão: PG 001, PG 002, PG 003, PG 004, PG 005, PG 006, PG 007 e PG 008. O número de cada PG é correspondente ao número do BGA, ou seja, o BGA 001 é igual ao PG 001. Todas as PGs foram incluídas em todas as análises desse trabalho.

2.2.6. Diversidade genética da população

Utilizando o pacote *hierfstat* (JEROME; JOMBART; GOUDET, 2015) desenvolvido para o programa R (R DEVELOPMENT CORE TEAM, 2016) foram estimados: número de alelos por loco, porcentagem de locos polimórficos, heterozigosidade média observada (H_O) e heterozigosidade média esperada (H_S) (NEI, 1978) e índice de fixação de Wright (F_{IS}).

2.2.7. Estrutura genética da população

A estrutura do painel de acessos foi estudada utilizando uma Análise Discriminante de Componentes Principais (DAPC) (JOMBART; DEVILLARD; BALLOUX, 2010) através do pacote *adegenet* (JOMBART, 2008) no R (R DEVELOPMENT CORE TEAM, 2016). A DAPC é uma análise exploratória que ajuda a determinar o número mais provável de subpopulações dentro da população em estudo a partir de um subconjunto de marcadores SNPs bem distribuídos no genoma. Primeiramente é executada uma análise de componentes principais nos genótipos das amostras e em seguida é realizada uma análise discriminante a partir dos scores obtidos na análise dos componentes principais. A DAPC encontra um número de grupos que evidencia a estruturação ao minimizar a variação dentro de grupos e maximizar a variação entre grupos (JOMBART; DEVILLARD; BALLOUX, 2010). Tem um diferencial em relação ao programa *Structure* por não assumir que os subgrupos da população atendam a pressupostos como, por exemplo, equilíbrio de *Hardy-Weinberg* ou ausência de ligação entre os locos (PRITCHARD; STEPHENS; DONNELLY, 2000). Similar ao *Structure*, a DAPC também faz o agrupamento através da média K e o *Bayesian Information Criterion* (BIC) para identificar o número ótimo de grupos em uma população em que não há informação *a priori* sobre existência de grupos. Apenas para fins de comparação de resultados

procedemos também uma análise bayesiana estruturada no programa *Structure* (PRITCHARD; STEPHENS; DONNELLY, 2000).

2.2.8. Análise de agrupamento

Para a análise de agrupamento utilizando marcadores SNPs primeiramente foi calculada a matriz de distâncias de (ROGERS, 1972) modificada por (WRIGHT, 1978), posteriormente, com base nas distâncias calculadas, realizou-se a construção de dendrogramas pelo método UPGMA (*Unweighted Pair Group Method with Arithmetic Mean*) (SOKAL; MICHENER, 1958), método hierárquico que faz uso da média das distâncias de todos os genótipos para formação de cada grupo. A consistência dos dendrogramas foi verificada através do valor do coeficiente de correlação cofenética (CCC) (SOKAL; ROHLF, 1962) e, a consistência de cada nó foi verificada com base em 10.000 reamostragens *bootstrap* a fim de verificar a estabilidade dos agrupamentos formados. Todas as análises foram realizadas no programa R (R DEVELOPMENT CORE TEAM, 2016).

2.3. Resultados

2.3.1. Marcadores

Um conjunto de 210 acessos (192 acessos, 10 controles e 8 PGs) foram genotipados utilizando o método de genotipagem por sequenciamento DArT-seq, produzindo 16.000 marcadores SNPs, que foram filtrados e posteriormente utilizados nas análises deste projeto.

De todos os acessos genotipados apenas 1, o acesso BGA 181, apresentou mais de 40% de dados perdidos sendo eliminado. Após a filtragem, restaram 5.578 SNPs de alta qualidade que foram utilizados nas análises seguintes.

2.3.2. Diversidade genética

Os resultados da diversidade genética encontrada no painel de acessos em estudo estão apresentados na Tabela 1. O valor encontrado para heterozigosidade observada (H_O) foi de 0,0279 e para a heterozigosidade esperada (H_S) foi de 0,2553, enquanto que o valor de F_{IS} foi de 0,8906.

Tabela 1. Índices de diversidade genética

Parâmetros	Resultados
H _O	0,0279
H _S	0,2553
F _{IS}	0,8906

2.3.3. Estruturação genética

O resultado da Análise Discriminante dos Componentes Principais (DAPC), Figura 1, sugeriu a formação 3 subgrupos genéticos principais na população, o subgrupo 1 (S₁), o subgrupo 2 (S₂) e o subgrupo 3 (S₃). Entre os 3 subgrupos, o S₁ e S₂ chamam a atenção pela proximidade e sobreposição parcial, enquanto o S₃ encontra-se isolado.

O número de grupos (K) testados variou de 1 a 10, e o K ótimo foi determinado com base no *Bayesian Information Criterion* (BIC) (JOMBART; DEVILLARD; BALLOUX, 2010). O resultado do programa *Structure* proporcionou a formação de 4 subgrupos genéticos principais na população, com uma distribuição entre as subpopulações muito próxima ao encontrado com a DAPC, porém com maior sobreposição (dados não mostrados).

Fazendo uma aproximação dos resultados da DAPC (Figura 1) identificou-se que o S₃ é formado pelas testemunhas da subespécie *indica* com mais cinco acessos da subespécie *japonica* que com elas agruparam, sendo eles: BGA 43, BGA 45, BGA 56, BGA 78 e BGA 182.

Figura 1. Análise Discriminante dos Componentes Principais (DAPC) com 1 função discriminante nos 192 acessos do germoplasma de arroz, mais 10 controles e 8 PGs estruturados em 3 agrupamentos. Os agrupamentos foram nomeados como S₁, S₂ e S₃

Figura 2. Scatterplot da Análise Discriminante dos Componentes Principais (DAPC) com 2 funções discriminantes nos 192 acessos do germoplasma de arroz, mais 10 controles e 8 PGs estruturados em 3 agrupamentos. Os agrupamentos foram nomeados como S₁, S₂ e S₃. Inserido na parte superior direita o Bayesian information criterion (BIC) e na parte inferior o Discriminant Analysis Eigenvalues

Figura 3. Barplot da Análise Discriminante dos Componentes Principais (DAPC) nos 192 acessos do germoplasma de arroz, mais 10 controles e 8 PGs estruturados em 3 agrupamentos e identificados em cada um dos subgrupos

2.3.4. Agrupamento

O dendrograma obtido no agrupamento dos genótipos (Figura 4) é o resultado da ordenação em grupos dos indivíduos da população pelo critério UPGMA com base na matriz de distâncias de (ROGERS, 1972) modificada por (WRIGHT, 1978) que sugere a formação de 3 grupos genéticos principais. O S_1 que se destacou por ser o maior grupo formado por 135 indivíduos, o S_2 formado por 64 indivíduos e o S_3 , menor agrupamento, que teve sua formação com 11 indivíduos. Todos os *bootstraps*, ou a consistência das bifurcações em porcentagem, do dendrograma foram maiores que 70 e todos os acessos ficaram muito bem posicionados devido seus ótimos valores.

Figura 4. Dendrograma gerado pela análise de agrupamento de 202 genótipos obtidos através do método UPGMA com base na matriz de distâncias de (ROGERS, 1972) modificada por (WRIGHT, 1978)

2.3.5. Prova de genotipagem

Os acessos identificados como PG (Prova de Genotipagem) agruparam com os respectivos BGAs no *barplot* e no dendrograma, Figuras 3 e 4, respectivamente. O PG 001 agrupou com o BGA 001, o PG 002 agrupou com o BGA 002, o PG 003 agrupou com o BGA 003, o PG 004 agrupou com o BGA 004, o PG 005 agrupou com o BGA 005, o PG 006 agrupou com o BGA 006, o PG 007 agrupou com o BGA 007 e o PG 008 agrupou com o BGA 008.

2.4. Discussão

2.4.1. Diversidade genética

Os valores apresentados na Tabela 1 dizem respeito aos parâmetros de diversidade genética apresentados pelo painel em estudo. Sendo, heterozigosidade observada (H_O) = 0,0279, heterozigosidade esperada (H_S) = 0,2553, e, índice de fixação (F_{IS}) = 0,8906.

O valor de H_S = 0,2553 indica que existe diversidade genética entre os acessos avaliados. Já o valor de H_O = 0,0279, menor que valor de H_S , mostra que sendo o arroz uma espécie com baixa taxa de fecundação cruzada, os locos tendem a homozigose. Como outra evidência desse sistema de reprodutivo temos o alto valor do índice de fixação (F_{IS}) = 0,8906. O arroz é uma espécie autógama, portanto, o fluxo de genes é restrito, observando-se baixos níveis de heterozigosidade. Deve-se tomar cuidado com a alta taxa de fixação de alelos, principalmente quando se trata de bancos de germoplasma, que são o reservatório genético e, quando falamos de espécies autógamas, que naturalmente possuem fluxo de genes menor, pois a perda de alelos ou blocos gênicos é irreversível.

Outro resultado do baixo fluxo de genes é a ocorrência, geograficamente ou ecologicamente, de grupos distintos de arroz mostrando uma maior diferenciação genética dentro da espécie do que seria o caso em uma espécie de cruzamento. Como por exemplo, temos as duas subespécies principais de arroz, *indica* e *japonica*, que se especula serem produtos de eventos separados de domesticação de espécies ancestrais (GARRIS et al., 2005).

O conhecimento do sistema reprodutivo das espécies é importante para implementar o manejo e as práticas de conservação *in situ* e *ex situ* (KARASAWA et al., 2007). As variações no potencial reprodutivo dos genitores também afetam a frequência gênica da geração subsequente, o que é chamado de tamanho efetivo populacional (N_e). Na preservação de amostras de germoplasma, a questão de se manter a representatividade genética, ou de cuidar do tamanho efetivo, é igualmente importante (VENCOVSKY, 1987). Dessa forma, sabendo-se que o arroz é uma espécie com autofecundação sua conservação em bancos de germoplasma depende que a aplicação do tamanho efetivo populacional seja feita com cautela para evitar a perda irreversível de alelos.

Um estudo investigando o efeito da perda parcial da viabilidade em amostras de espécies alógamas na fase de sua preservação, contudo, facilmente aplicáveis às espécies autógamas como o arroz foi realizado por (VENCOVSKY, 1987). Esse trabalho mostrou a impossibilidade de recuperação do tamanho efetivo inicial, planejado, quando as perdas forem de 50%, ou mais. Sendo que, uma maneira de compensar perdas desse tipo é aumentando o

tamanho da amostra, logo após sua ocorrência. Entretanto, essa estratégia tem limitações e somente em certas circunstâncias, com perdas não pronunciadas, é que se pode recuperar o tamanho efetivo inicial. Por isso, manter tamanhos efetivos maiores, dentro das possibilidades, não deixa de ser estratégia saudável. Tanto na coleta como na preservação de amostras, é fundamental que se procure sempre reunir descendentes, como sementes, de um maior número possível de plantas genitoras e de preferência em número igual, ou aproximadamente igual, de cada uma. O que se deve tomar aleatoriamente é o conjunto de plantas genitoras e não o das sementes dessas.

Como o arroz é uma espécie com baixa taxa de fecundação cruzada, tendendo a possuir locos em homozigose, a diversidade genética presente no germoplasma pode por isso estar estruturada.

2.4.2. Estrutura da população

Os dados genotípicos da população foram analisados para inferir a existência de estrutura genética. Esta análise estima a probabilidade de cada genótipo ter sido originado de cada subpopulação a partir das frequências alélicas do conjunto de acessos.

Os resultados da Análise Discriminante dos Componentes Principais (DAPC) apresentados nas Figuras 1, 2 e 3 forneceram evidências para significativa estrutura populacional no painel de acessos de arroz da subespécie *japonica* estudado. Com a formação 3 subgrupos genéticos principais na população, o subgrupo 1 (S_1) que agrupou o maior número de indivíduos, o subgrupo 2 (S_2) e o subgrupo 3 (S_3) com o menor número de indivíduos agrupados. Entre os 3 subgrupos, o S_1 e S_2 chamam a atenção pela proximidade e sobreposição parcial, enquanto o S_3 encontra-se isolado. A sobreposição do S_1 e S_2 , claramente observável na Figura 1 e na Figura 3, muito provavelmente se deve pelo compartilhamento de características semelhantes. O S_3 se configurou como um subgrupo isolado, com menos indivíduos que o S_1 e o S_2 e, curiosamente formado por acessos do banco de germoplasma (os BGAs) mais os controles da subespécie *indica* (identificados como Test.).

Os cinco acessos do banco de germoplasma, BGA 43, BGA 45, BGA 56, BGA 78 e BGA 182, que agruparam com os controles da subespécie *indica* sugerem que tais acessos possam ter maior semelhança com genótipos da subespécie *indica* do que com a subespécie *japonica*. Questiona-se sobre uma possível mistura durante o processo de formação do banco de germoplasma, podendo esta ter ocorrido no processo de coleta e catalogação ou mesmo no

processo de multiplicação dos materiais em campo. A DAPC é um método multivariado desenhado para identificar e descrever agrupamentos de indivíduos geneticamente relacionados em uma população. Quando a informação de grupos *a priori* é desconhecida. É considerada melhor que o programa *Structure* para caracterizar a subdivisão populacional quando não é necessário assumir pressupostos de equilíbrio de *Hardy-Weinberg* e ausência de ligação, bem como quando se tem disponível uma grande quantidade de dados. Além disso, as ferramentas implementadas na DAPC para a identificação de agrupamentos e representação gráfica de estruturas entre grupos permite desvendar a complexa estrutura de uma população (JOMBART; DEVILLARD; BALLOUX, 2010). Pela confiabilidade da análise em questão, utilizada em diversos trabalhos recentes (RICE et al., 2016) (TURCHETTO et al., 2016) (D'ÁVILA et al., 2016) e, sabendo que o banco de germoplasma do Departamento de Genética possui acessos de ambas subespécies, acredita-se que uma possível mistura entre elas possa ter ocorrido.

A subespécie *japonica* faz parte de uma subdivisão da espécie *Oryza sativa* L. que possui duas subespécies, a *japonica* e a *indica*. Os cultivares da subespécie *japonica* são predominantemente cultivados em regiões temperadas nos trópicos e os da subespécie *indica* nos subtrópicos. Essas subespécies são diferenciadas através de um conjunto de caracteres que abrangem desde aspectos morfológicos até reações a estresses abióticos. A maioria dos cultivares da subespécie *indica* são geralmente sensíveis ao frio enquanto que os da subespécie *japonica* são classificados como tolerantes a baixas temperaturas. Outras características, como a relação comprimento/largura dos grãos, são usadas para identificar as duas subespécies. Os cultivares da subespécie *indica* apresentam grãos mais longos e finos e os da subespécie *japonica* geralmente apresentam grãos largos e curtos (MATSUO et al., 1993). Essas informações podem ajudar na definição dos acessos que agruparam com os indivíduos da subespécie *indica*.

Essa estrutura genética encontrada entre os genótipos era esperada, visto que os acessos são oriundos de diferentes regiões do Japão e passaram por diferentes processos de seleção. A presença de estrutura genética em arroz tem sido documentada em alguns trabalhos e através dessa estrutura de populações geneticamente definida, pode ser possível explorar os *pools* de genes de arroz de forma mais eficaz com abordagens baseadas em genética de populações utilizando as extensas coleções de recursos genéticos de arroz (GARRIS et al., 2005).

2.4.3. Prova de genotipagem

O *Barplot* (Figura 3) e o dendrograma (Figura 4) mostram as PGs agrupadas com seus respectivos acessos (BGA), resultado que era esperado visto que são os mesmos materiais genotipados em tempos diferentes. Essa prova de genotipagem foi propositalmente realizada para afirmar que a tecnologia de genotipagem por sequenciamento DArT-seq reproduz seus resultados com exatidão, ou seja, têm alta reprodutibilidade, sendo portanto, confiável para utilização em estudos genéticos que demandam genotipagem.

Além de ter reprodutibilidade, os SNPs derivados da metodologia DArT-seq proporcionam uma vantagem adicional pela inclusão de marcadores co-dominantes. O desenvolvimento de metodologias de genotipagem por sequenciamento (GBS) via enzimas de restrição, como DArT-seq, representa uma aplicação adicional das tecnologias de *next generation sequencing* além do sequenciamento, potencializando a análise genética com marcadores moleculares (SANSALONI; GRATTAPAGLIA, 2012).

Alguns trabalhos têm demonstrado a viabilidade do método DArT-seq em várias culturas, como em arroz (COURTOIS et al., 2013) (THI et al., 2014), cevada (IGARTUA et al., 2015), eucalipto (STEANE et al., 2011) e recentemente em trigo (VIKRAM et al., 2016) e melancia (YANG et al., 2016).

2.5. Conclusão

A tecnologia de genotipagem por sequenciamento DarT-seq é um método eficiente para o estudo de diversidade genética entre acessos de banco de germoplasma de diferentes espécies.

Existe diversidade genética no banco de germoplasma de arroz da ESALQ-USP e a mesma está estruturada em 3 subgrupos. O estudo nos mostra que o banco de germoplasma de arroz em questão é uma importante fonte de acessos para utilização em programas de melhoramento genético.

REFERÊNCIAS

- BRONDANI, C. et al. Transferability of microsatellite and sequence tagged site markers in *Oryza* species. **Hereditas**, v. 138, n. 3, p. 187–192, 2003.
- COURTOIS, B. et al. Genome-wide association mapping of root traits in a japonica rice panel. **PLoS ONE**, v. 8, n. 11, p. 1–18, 2013.
- D'ÁVILA, L. S. et al. Genetic structure and mating type analysis of the *Pyricularia oryzae* population causing widespread epidemics in southern Brazil. **Tropical Plant Pathology**, 2016.
- DOYLE, J. J.; DOYLE, J. L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. **Phytochemical Bulletin of the Botanical Society of America**, v. 1, n. 19, p. 11–15, 1987.
- ELSHIRE, R. J. et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PLoS ONE**, v. 6, n. 5, p. 1–10, 2011.
- FALCONER, D. **Quantitative Genetics**. Ronald Pre ed. New York: Ronald Press, 1964.
- FLINT-GARCIA, S. A.; BODNAR, A. L.; SCOTT, M. P. Wide variability in kernel composition, seed characteristics, and zein profiles among diverse maize inbreds, landraces, and teosinte. **Theoretical and Applied Genetics**, v. 119, n. 6, p. 1129–1142, 2009.
- GARRIS, A. J. et al. Genetic Structure and Diversity in *Oryza sativa* L. **Genetics Society of America**, v. 1638, n. March, p. 1631–1638, 2005.
- HOISINGTON, D. et al. Plant genetic resources: what can they contribute toward increased crop productivity? **Proceedings of the National Academy of Sciences of the United States of America**, v. 96, n. 11, p. 5937–5943, 1999.
- IGARTUA, E. et al. Selection footprints in barley breeding lines detected by combining genotyping-by-sequencing with reference genome information. **Molecular Breeding**, v. 35, n. 1, p. 14, 2015.
- JEROME, A.; JOMBART, T.; GOUDET, M. J. **Package “hierfstat”**, 2015.
- JOMBART, T. Adegenet: A R package for the multivariate analysis of genetic markers. **Bioinformatics**, v. 24, n. 11, p. 1403–1405, 2008.
- JOMBART, T.; DEVILLARD, S.; BALLOUX, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. **BMC Genetics**, v. 11, n. 1, p. 94, 2010.

- KARASAWA, M. M. G. et al. Mating system of Brazilian *Oryza glumaepatula* populations studied with microsatellite markers. **Annals of Botany**, v. 99, n. 2, p. 245–253, 2007.
- MATSUO, T. et al. **Science of the Rice Plant**. [s.l.] Food and Agriculture Policy Research Center, 1993.
- NACHIMUTHU, V. V. et al. Analysis of Population Structure and Genetic Diversity in Rice Germplasm Using SSR Markers: An Initiative Towards Association Mapping of Agronomic Traits in *Oryza Sativa*. **Rice (New York, N.Y.)**, v. 8, n. 1, p. 30, 2015.
- NASS, L. Utilization of genetic resources for maize and soybean breeding in Brazil. **Rev. Brasil. Genet**, v. 16, n. 4, p. 983–988, 1993.
- NEI, M. Estimation of average heterozygosity and genetic distance from a small number of individuals. **Genetics**, v. 89, n. 3, p. 583–590, 1978.
- PRITCHARD, J. K.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, v. 155, n. 2, p. 945–959, 2000.
- R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing** Vienna, Austria R Foundation for Statistical Computing, , 2016.
- RICE, A. M. et al. Population differentiation at a regional scale in spadefoot toads: Contributions of distance and divergent selective environments. **Current Zoology**, v. 62, n. 2, p. 193–206, 2016.
- ROGERS, J. S. Measures of genetic similarity and genetic distance. **University of Texas Publication**, v. 7, n. 1, p. 145–153, 1972.
- ROY, S. et al. Genetic Diversity and Population Structure in Aromatic and Quality Rice (*Oryza sativa* L.) Landraces from North-Eastern India. **PLoS ONE**, v. 10, n. 6, p. 1–13, 2015.
- SANSALONI, C. P.; GRATTAPAGLIA, O. D. **Desenvolvimento e aplicações de DArT (Diversity Arrays Technology) e genotipagem por sequenciamento (Genotyping-by-Sequencing) para análise genética em Eucalyptus**. [s.l.] Universidade de Brasília, 2012.
- SINGH, S. P. Patterns of variation in cultivated common bean (*Phaseolus vulgaris*, Fabaceae). **New York**, v. 43, n. December 1986, p. 39–57, 1989.
- SOKAL, R. .; MICHENER, C. A statistical method for evaluating systematic relationships. . 1958, p. 1409 – 1438.
- SOKAL, R. R.; ROHLF, F. J. The Comparison of Dendrograms by Objective Methods. **Taxon**, v. 11, n. 2, p. 33–40, fev. 1962.

- SOW, M. et al. Genetic diversity , population structure and differentiation of rice species from Niger and their potential for rice genetic resources conservation and enhancement. **Genet Resour Crop Evol**, v. 61, p. 199–213, 2014.
- STEANE, D. A. et al. Molecular Phylogenetics and Evolution Population genetic analysis and phylogeny reconstruction in Eucalyptus (Myrtaceae) using high-throughput , genome-wide genotyping. **Molecular Phylogenetics and Evolution**, v. 59, n. 1, p. 206–224, 2011.
- SWARUP, S. et al. Determinants of the high-methionine trait in wild and exotic germplasm may have escaped selection during early cultivation of maize. **The Plant journal : for cell and molecular biology**, v. 8, n. 3, p. 359–368, 1995.
- THI, N. et al. Characterization of a panel of Vietnamese rice varieties using DArT and SNP markers for association mapping purposes. **BioMed Central**, v. 14, n. 371, p. 1–16, 2014.
- TURCHETTO, C. et al. High levels of genetic diversity and population structure in an endemic and rare species: implications for conservation. **AoB PLANTS**, v. 8, p. 1–17, 2016.
- VENCOVSKY, R. **Tamanho efetivo populacional na coleta e preservação de germoplasma de espécies alógamas**. IPEF. **Anais...**1987
- VIKRAM, P. et al. Unlocking the genetic diversity of Creole wheats. **Nature Publishing Group**, n. August 2015, p. 1–12, 2016.
- WRIGHT, S. **Evolution and the genetics of populations**. Chicago: University of Chicago Press, 1978.
- YANG, X. et al. Genetic diversity and population structure of core watermelon (Citrullus lanatus) genotypes using DArTseq-based SNPs. **Plant Genetic Resources**, v. 14, n. 3, p. 226–233, 2016.

3. MAPEAMENTO ASSOCIATIVO DE CARACTERÍSTICAS AGRONÔMICAS EM ACESSOS JAPONESES DE ARROZ UTILIZANDO MARCADORES DART-SEQ (IN SILICO E SNP)

RESUMO

Este estudo determinou a associação genômica de caracteres agronômicos em acessos japoneses de arroz para melhoramento de produção. Cento e oitenta e seis genótipos de arroz pertencentes ao Banco de Germoplasma de Arroz do Departamento de Genética da ESALQ/USP foram genotipados utilizando o protocolo de sequenciamento da Tecnologia Diversidade Arrays (DART-seq). O número de dias para florescimento (NDF), a estatura de planta (EP), o comprimento da panícula (CP), peso de parcela (PP), a massa de mil grãos (MMG) e CICLO foram considerados para o estudo. O mapeamento associativo do genoma foi realizado com base em 24.266 marcadores silico DArTs e 1.965 marcadores SNPs. O p-valor <0,001 foi utilizado para determinação do grau de ligação entre loco e marcador e para a quantificação da magnitude da influência da região no fenótipo observado. O desequilíbrio de ligação (DL) mostrou a existência da deterioração da ligação com o aumento da distância física. Observou-se um total de 113 associações significativas genótipo-fenótipo quando utilizado marcadores silico DArTs e um total de 21 associações significativas genótipo-fenótipo quando utilizado marcadores SNPs. As associações relatadas nessa população podem ser úteis para iniciar a seleção assistida por marcadores (MAS), introgressão de caracteres direcionados no arroz para produtividade, para mapeamento fino ou clonagem dos genes subjacentes e QTLs.

Palavras-chave: *Oryza sativa* L. ssp. *japonica*; Germoplasma; Genotipagem por sequenciamento; GWAS

ABSTRACT

This study determined the genomic association of agronomic traits in Japanese accessions of rice for breeding production. One hundred and eighty-six rice genotypes belonging to the Rice Germplasm Bank of the Department of Genetics of ESALQ / USP were genotyped using the Diversity Arrays Technology sequencing protocol (DARTseq). Number of days for flowering (NDF), plant height (EP), panicle length (CP), plot weight (PP), mass of one thousand grains (MMG) and CYCLE were considered for the study. The association mapping of the genome was performed based on 24.266 silico DArTs markers and 1965 SNPs marker. The p-value <0.001 was used to determine the degree of linking between locus and marker and to quantify the magnitude of the influence of the region on the observed phenotype. Linkage disequilibrium (LD) showed deterioration of the linkage with the increase in physical distance. We observed a total of 113 significant associations between genotype-phenotype when used silico DArTs markers and a total of 21 significant genotype-phenotype associations when used SNPs markers. Associations reported in this population can be useful for initiating marker-assisted selection (MAS), introgression of targeted traits in rice for productivity, for fine mapping or cloning of underlying genes and QTLs.

Keywords: *Oryza sativa* L.ssp. *japonica*; Germplasm; Genotyping by sequencing; GWAS

3.1. Introdução

O arroz, principal fonte de calorias dietéticas para metade da humanidade, é a primeira planta de cultivo para a qual foi produzida uma sequência de genoma de referência de alta qualidade (MCNALLY et al., 2009). Servindo como uma planta modelo para a pesquisa genômica e o melhoramento genético (SINGH et al., 2013).

O sequenciamento e a anotação do genoma do arroz tem revelado milhares de genes e possibilitado a análise genética comparativa do seu germoplasma para identificar a variação genética ou os alelos múltiplos existentes, que podem ser aplicados através do melhoramento convencional ou da engenharia genética (GOWDA et al., 2003).

Alguns fatores como reduções de terras aráveis adequadas, água, mão-de-obra, fertilizantes altamente dependentes de energia e degradação ambiental tornam particularmente importantes os trabalhos na área do melhoramento de plantas com objetivo de localizar genes que possibilitem o melhor desempenho da cultura como uma solução para esses problemas (IRGPS, 2005).

Uma melhoria no potencial e na estabilidade de rendimento do arroz possibilita, por exemplo, que a produção seja realizada em terras agrícolas já disponíveis, sem a necessidade de abrir mais campos para plantio. Contudo, as variedades mais adaptadas são produto de uma combinação de biotecnologia e melhoramento genético convencional, e essas por sua vez dependem de uma sequência genômica de arroz de alta qualidade (IRGPS, 2005), construída aos poucos através de extensas pesquisas envolvendo busca de alelos em bancos de germoplasma e introdução de características desejáveis em cultivares elite.

Quase todas as variantes possíveis estão presentes no germoplasma do arroz para o estudo da natureza dos genes que controlam uma característica em particular. O foco imediato nas pesquisas genéticas tem sido a identificação e estudo de alelos superiores que controlam as principais características agronômicas, fisiológicas e econômicas, que contribuem direta ou indiretamente na melhoria do rendimento (SINGH et al., 2013) (NACHIMUTHU et al., 2015).

Para esse propósito, os marcadores moleculares tem sido valiosos na construção de mapas genéticos de alta resolução, para o estudo da estrutura da população e para a descoberta de relações entre marcador-característica em experimentos de mapeamento associativo. Com a possibilidade do sequenciamento de vários genomas e a revelação em massa em estudos de diversidade (SINGH et al., 2013). Vários tipos de marcadores moleculares têm sido desenvolvidos e descritos (SEMAGN; BJORNSTAD; NDJIONDJOP, 2006) (AGARWAL;

SHRIVASTAVA; PADH, 2008) podemos citar, por exemplo, os marcadores silico DArTs e SNPs desenvolvidos pela metodologia DArT-seq baseada no princípio de redução da complexidade do genoma aliado à tecnologia de genotipagem por sequenciamento de última geração - NGS (*Next Generation Sequencing*) (KILIAN et al., 2005) e que têm sido utilizados em diversos trabalhos relacionados a diversidade, estruturação e mapeamento genético.

Neste âmbito, o mapeamento por associação emergiu como uma poderosa ferramenta para identificar locos que controlam características quantitativas em várias espécies de culturas economicamente importantes (RICHARDS; FRIESEN; BRUEGGEMAN, 2017). Possui diversas vantagens, como a capacidade de detectar QTLs em uma resolução mais alta, menor tempo necessário para realizar as análises genéticas e potencial para identificar um maior número possível de locos associados com o fenótipo (YU et al., 2006) (ZHU et al., 2008). Além disso, elimina a principal desvantagem da análise de ligação clássica, uma vez que não necessita de populações genéticas específicas e nem com parentesco conhecido (GUPTA; RUSTGI; KULWAL, 2005), podendo ser conduzido para vários tipos de populações, tais como acessos de banco de germoplasma, cultivares, coleções nucleares e genótipos selvagens, que não possuem parentesco conhecido *a priori*. A possibilidade de uso de coleções de germoplasma, em vez de populações conhecidas, permite a exploração de eventos de recombinação ancestral que ocorreram durante a história evolutiva das plantas (ZHU et al., 2008).

O Departamento de Genética da Escola Superior de Agricultura “Luiz de Queiroz” – USP ao longo dos anos armazenou germoplasma de arroz oriundo de algumas instituições de pesquisa do Brasil e do exterior e entre esses materiais encontram-se diversos acessos da subespécie *japonica*, oriundos do Instituto de Arroz de Sequeiro do Japão, atualmente extinto. Tais acessos estão passando por uma extensa pesquisa científica com a intenção de juntar informações a respeito de sua origem, diversidade, estruturação genética e QTLs responsáveis pela melhoria do rendimento e adaptabilidade.

Diante do exposto, realizamos um mapeamento por associação para características de produtividade em um painel contendo 186 acessos de arroz da subespécie *japonica*, pertencentes ao Banco de Germoplasma de Arroz do Departamento de Genética da ESALQ. Para esse estudo utilizamos marcadores SNPs e silico DArTs gerados a partir da tecnologia DArT-seq de genotipagem por sequenciamento (GBS). Uma análise comparativa dos resultados encontrados através de marcadores SNPs e silico DArTs em um conjunto de acessos pode ajudar no entendimento e decisão sobre qual marcador é mais apropriado para a caracterização do germoplasma de arroz e projetar estratégias de melhoramento eficientes.

3.2. Material e Métodos

3.2.1. Material vegetal

O painel utilizado neste estudo foi composto por 186 acessos de arroz da subespécie *japonica* que fazem parte do banco de germoplasma do Departamento de Genética e Melhoramento de Plantas da Escola Superior de Agricultura “Luiz de Queiroz” – Campus USP de Piracicaba, São Paulo. Esses acessos apresentam características de materiais de sequeiro, também conhecidos como arroz de terras altas. Para cada acesso as sementes foram obtidas a partir de uma única espiguetas em casa de vegetação, na Escola Superior de Agricultura “Luiz de Queiroz”.

3.2.2. Fenotipagem

Para a fenotipagem foram instalados dois experimentos em delineamento *alpha-látice* conduzidos em campo nos anos agrícolas de 2014/2015 e 2015/2016, com 3 repetições por tratamento e parcelas com duas linhas de 1,50 metros. Sendo que, os 186 genótipos foram fenotipados para seis características agrônômicas: número de dias para florescimento (NDF), estatura de planta (EP), comprimento da panícula (CP), peso de parcela (PP), massa de mil grãos (MMG) e CICLO.

3.2.3. Extração de DNA

O DNA genômico foi extraído em duplicatas a partir de folhas jovens coletadas de uma única planta de cada um dos acessos utilizando o método de extração CTAB descrito por (DOYLE; DOYLE, 1987). Posteriormente, para verificação da qualidade e quantidade o DNA foi diluído em TE (Tris-HCl 10 µL, EDTA 1 µL, com pH 8,0) em concentração entre 50 ng/µL e 100 ng/µL e as amostras foram aplicadas em gel de agarose 0,8%. A integridade do DNA foi atestada pelo aparecimento de apenas uma banda de elevado peso molecular sobre o gel.

3.2.4. Genotipagem DArTseq

A genotipagem foi conduzida para 186 acessos japoneses de arroz pela empresa Diversity Arrays Technology Pty Ltd. (DArT P/L) da Austrália, através da tecnologia de

genotipagem por sequenciamento DArT-seqTM, previamente descrito por (COURTOIS et al., 2013), que combina GBS com Diversity Arrays Technology (DArT), envolvendo a redução da complexidade do genoma usando enzimas de restrição *PstI/TaqI* seguido pelo sequenciamento de sequências curtas através do Illumina (<http://www.diversityarrays.com/dart-application-dartseq-data-types>) HiSeq 2500. Que consiste em dois conjuntos de marcadores: (i) marcador de polimorfismo de nucleotídeo único (SNP) de genotipagem por sequenciamento (GBS) sob sequenciamento dos fragmentos derivados da redução da complexidade do genoma chamados de SNP, e (ii) variação de presença/ausência marcadores silico DArT que se referem a se um fragmento de DNA sequenciado definido ou não foi obtido após a digestão enzimática do genoma (KILIAN et al., 2012). Além da informação dos marcadores, o conjunto de dados recebido pela empresa continha também a posição genética e o respectivo cromossomo em que cada marcador foi encontrado, informação esta baseada no alinhamento com o genoma de referência Rice_v9 (*Phytozome assembly v9*).

3.2.5. Filtragem dos dados

Os marcadores SNPs obtidos pela genotipagem com a tecnologia DArT-seq passaram por uma filtragem de dados seguindo os parâmetros profundidade de cobertura (*Read Depth*) ≥ 3 ; *Call Rate* $\geq 0,8$; conteúdo de informação polimórfica (PIC) $\geq 0,1$; e, indivíduos com mais de 40% de dados perdidos. Já os marcadores silico DArTs foram filtrados para menos de 10% de dados perdidos e menos de 20% de heterozigotos. Posteriormente, foram mantidas somente marcas com um MAF (*Minor Allele Frequency*) maior que 0,005, ou seja, marcas com frequência menor de 5% foram descartados. Em ambos os conjuntos de marcadores, os 10% de dados faltantes e os 20% dos marcadores heterozigotos foram codificados e procedeu-se a imputação destes dados pelo software BEAGLE v.4.1 (BROWNING; BROWNING, 2016). Após a imputação, ainda foram retirados os marcadores heterozigotos que restavam no conjunto de dados.

3.2.6. Estrutura da população, matriz de parentesco e desequilíbrio de ligação

A estrutura da população (matriz Q) foi inferida através da análise de componentes principais (PCA) (PRICE et al., 2006), a matriz de parentesco *kinship* (K) para os dois conjuntos de marcadores SNPs e silico DArTs foi construída utilizando o algoritmo *Loiselle* (LOISELLE et al., 2008) e o desequilíbrio de ligação (DL), também para os dois conjuntos de

marcadores SNPs e silico DArTs, foi calculado com o coeficiente de determinação (r^2) (HILL, ROBERTSON, 1968) entre pares de marcadores e distância física entre estes. A média móvel (linha em vermelho) de marcadores adjacentes foi calculada usando uma janela de dez marcadores. Todas essas análises foram realizadas no pacote GAPIT (LIPKA et al., 2012) (TANG et al., 2016) no programa R.

3.2.7. Mapeamento associativo

Para a análise do mapeamento associativo foram estimados os valores de BLUPs (*Best Linear Unbiased Predictions*) pela análise conjunta dos dois experimentos para cada genótipo via pacote do R *lme4* (BATES et al., 2015). Sendo que, efeito de genótipos, blocos e interação foram considerados como aleatórios no modelo misto. A análise de associação foi realizada para os dois conjuntos de marcadores, SNPs e silico DArTs, através das ferramentas de mapeamento associativo baseadas na abordagem MLM (Modelo Linear Misto) disponíveis no pacote GAPIT (*Genome Association and Prediction Integrated Tool*) (LIPKA et al., 2012) (TANG et al., 2016) do programa R. Foram utilizadas informações de parentesco (*kinship*) e estrutura da população (PCA) para redução do erro tipo I. Apenas associações marcador-característica com p valor <0,001 foram declaradas significativas. Os gráficos *Manhattan plots* foram gerados no pacote '*qqman*' do R (TURNER, 2014).

3.2.8. Anotação dos genes

Os SNPs significativamente associados foram posicionados em genes ou em regiões intergênicas, para os que foram posicionados dentro dos genes foi obtida a sequência transcrita dos genes de acordo com o genoma de referência da cultivar Nipponbare *Oryza sativa* L. ssp. *japonica* (Os-Nipponbare-Reference-IRGSP-1.0), no banco de dados *Gramene* (<http://www.gramene.org/>). A partir dos transcritos identificados, foi realizada uma busca sobre a anotação funcional baseada nas três classes de termos GO, depositados no banco de dados do *UniProt* (<http://www.uniprot.org/>).

3.3. Resultados

3.3.1. Genotipagem

A genotipagem dos 186 acessos pertencentes ao banco de germoplasma do Departamento de Genética e Melhoramento de Plantas da Escola Superior de Agricultura “Luiz de Queiroz” – Campus USP de Piracicaba foi conduzida pela empresa Diversity Arrays Technology Pty Ltd. através da tecnologia de genotipagem por sequenciamento DArT-seq, previamente descrita por (COURTOIS et al., 2013).

Os dados brutos apresentaram 16.605 marcadores SNPs que após a filtragem e imputação foram reduzidos para 1.965 SNPs; e, 42.267 marcadores silico DArTs que após filtragem e imputação resultaram em 24.266 silico DArTs, distribuídos através dos 12 cromossomos do arroz e que foram utilizados para as análises de mapeamento associativo, (Tabela 2).

Dentre os resultados, o cromossomo 1 mostrou o maior número de SNPs (420) e silico DArTs (3.006), enquanto o cromossomo 12 mostrou o menor número de SNPs (47) e o cromossomo 10 mostrou o menor número de silico DArTs (1336), com uma média de 163,75 SNPs por cromossomo e de 2.022,16 silico DArTs por cromossomo.

Tabela 2. Número de marcas SNPs e silico DArTs identificadas entre os 186 acessos de *Oryza sativa* L. spp. *japonica* em cada um dos cromossomos e total, imagem de cariótipo modificado de http://ensembl.griemene.org/Oryza_sativa/Location/Genome

	Cromossomos	SNPs	Silico DArTs
1		420	3006
2		167	2708
3		143	2338
4		238	2254
5		200	1535
6		112	1925
7		127	1986
8		114	1636
9		64	1413
10		198	1336
11		135	2173
12		47	1956
Total de marcas		1.965	24.266

3.3.2. Estrutura da população, matriz de parentesco e desequilíbrio de ligação

A matriz de parentesco foi calculada para os dois conjuntos de marcadores, SNPs e silico DArTs utilizando o algoritmo *Loiselle* (LOISELLE et al., 2008), como pode ser observado na Figura 5. A diagonal do gráfico representa a máxima similaridade apresentada entres os genótipos, sendo a comparação de um genótipo com o mesmo genótipo, assim representada pela cor vermelha. Quanto menos similares os genótipos estes apresentam cores mais claras, sendo branco a menor similaridade possível entre os genótipos.

Figura 5. Kinship calculado pelo algoritmo *Loiselle* utilizando marcadores SNPs e marcadores Silico DArTs

Na Figura 6 observa-se os resultados do desequilíbrio de ligação (LD) calculado a partir de marcadores SNPs e marcadores silico DArTs, pode-se perceber que em ambos o desequilíbrio decai à medida que a distância física aumenta. Cada ponto representa a distância entre dois pares de marcadores com base nas informações de posição dos marcadores e seu coeficiente de correlação ao quadrado - r^2 . A linha vermelha é a média móvel entre 10 marcadores adjacentes. Quanto mais próximo de 1 maior o desequilíbrio de ligação e quanto mais próximo de 0 menor é o desequilíbrio de ligação entre os marcadores.

Figura 6. Decaimento do desequilíbrio de ligação conforme a distância genética entre os pares de marcadores calculado pelo pacote GAPIT através de marcadores SNPs e marcadores silico DArTs

3.3.3. Mapeamento associativo

Para fins de GWAS, os marcadores com baixa frequência alélica (<5%) foram removidos da matriz e os dados faltantes foram imputados. As matrizes finais dos marcadores continham 1.965 marcadores SNPs e 24.266 marcadores silico DArTs, analisados contra um painel contendo a informação fenotípica conjunta (dois experimentos) de 186 acessos da subespécie *japonica*.

O mapeamento associativo foi baseado no modelo linear misto (MLM) utilizando a matriz de parentesco *Loiselle* (*kinship*) e a estrutura da população (PCA) para explicar o efeito da estratificação populacional e do parentesco, reduzindo assim as chances de ocorrência do erro tipo I ou falsos positivos.

Com a finalidade de se observar os SNPs e silico DArTs significativos oriundos da genotipagem dos 186 acessos de arroz da subespécie *japonica* foram obtidos gráficos *Manhattan Plot*, elaborados a partir das coordenadas genômicas das associações exibidas ao longo do eixo X e com o logaritmo negativo do p-valor da associação exibido no eixo Y. Foi delimitado um p-valor de 0,001 para se considerar as associações como significativas. Desse modo, apenas as associações acima do valor estabelecido, *threshold* 3, foram consideradas estatisticamente significativas na análise de associação.

3.3.4. Análise de associação utilizando marcadores SNPs

Entre os 1.965 SNPs identificados ao longo dos 12 cromossomos após a filtragem e imputação, 21 marcadores SNPs (aproximadamente 1%) foram identificados como significativamente associados às características fenotípicas avaliadas. Das seis características analisadas, apenas 4 obtiveram marcadores associados: EP, CICLO, MMG e PP, como pode ser observado na Figura 7.

Apenas uma associação relacionada à característica EP foi encontrada, localizada no cromossomo 1. Para a característica CICLO foram identificadas 3 associações nos cromossomos 1, 8 e 11. A MMG foi a característica que apresentou o maior número de associações, 16 no total, distribuídas nos cromossomos 6 e 10. Já a característica PP teve apenas 1 associação significativa detectada no cromossomo 2.

Os gráficos *Q-Q plot* (Figura 7) foram gerados para os p-valores observados contra os p-valores esperados usando o modelo MLM ($Q + K$), onde os p-valores observados foram obtidos a partir do mapeamento associativo e os p-valores esperados partindo do pressuposto

de que nenhuma associação ocorreu entre o marcador e a característica. A linha azul representa os p-valores esperados e a linha vermelha representa os p-valores observados.

Figura 7. *Manhattan plot* e *Q-Q plot* da análise de associação para as características agrônômicas EP, CICLO, MMG e PP utilizando 1.965 marcadores SNPs em 186 acessos de arroz da subespécie *japonica*

Encontra-se na Tabela 3 as informações pertinentes em relação a cada marcador significativamente associado com as características fenotípicas avaliadas. Temos o marcador SNP, o cromossomo, a posição genética dentro do cromossomo onde o marcador foi encontrado, o p-valor, a MAF, o R^2 (%), os alelos, a região e o transcrito para aqueles marcadores encontrados dentro de genes.

Considerando-se os 21 SNPs associados significativamente na análise, 11 SNPs foram localizados em regiões intergênicas e 10 SNPs foram localizados dentro de genes.

O R^2 representa a variação genética do fenótipo explicada pelo marcador, destaca-se o SNP 1694 posicionado no cromossomo 10 que representa 7,9726% da variação fenotípica da característica massa de mil grãos (MMG), o SNP 1487 posicionado no cromossomo 8 que representa 7,3913% da variação fenotípica da característica ciclo e o SNP 450 posicionado no cromossomo 2 que representa 7,0934% da variação fenotípica da característica peso de parcela (PP).

Tabela 3. Marcadores SNPs que apresentaram associação com as características de produtividade EP, CICLO, MMG e PP, identificados entre os 186 acessos de *Oryza Sativa spp. japonica*. São apresentados: as marcas significativamente associadas (SNPs), o cromossomo, a posição do loco SNP no cromossomo (genoma de referência Rice_v9 - Phytosome assembly v9), o p valor, a mínima frequência alélica (MAF), a estimativa da variação fenotípica em % devido ao efeito do marcador (R^2), os alelos de cada loco SNP, a região gênica na qual o SNP está inserido e o transcrito com base no International Rice Genome Sequencing Project (IRGSP)

SNP	Chr.	Position	p value	MAF	R^2 (%)	Alelos	Região	Transcrito
EP								
419	1	42646685	0.0004	0.1935	5.8270	A/T	íntron	OS01T0967100-01
CICLO								
1487	8	16812555	0.0003	0.3065	7.3913	T/A	éxon	OS08T0363200-01
26	1	1292249	0.0006	0.1452	6.5094	T/A	íntron	-
1805	11	3955406	0.0008	0.1828	6.1841	T/A	íntron	-
MMG								
1694	10	11392885	0.0001	0.0484	7.9726	T/A	íntron	-
1191	6	9856259	0.0003	0.2527	6.7590	T/A	éxon	OS06T0281400-01
1659	10	9065267	0.0006	0.0753	6.2519	T/A	íntron	-
1696	10	11626357	0.0006	0.0753	6.2310	G/C	éxon	OS10T0369900-01
1624	10	6172388	0.0008	0.0591	5.9131	A/T	íntron	OS10T0189100-01
1651	10	8288086	0.0008	0.0591	5.9131	C/G	íntron	-
1654	10	8523965	0.0008	0.0591	5.9131	A/T	íntron	-
1662	10	9217226	0.0008	0.0591	5.9131	T/A	íntron	-
1665	10	9772839	0.0008	0.0591	5.9131	T/A	íntron	OS10T0337500-01
1679	10	10417505	0.0008	0.0591	5.9131	T/A	íntron	OS10T0346600-00
1683	10	10728124	0.0008	0.0591	5.9131	G/C	íntron	OS10T0351700-01
1685	10	10749484	0.0008	0.0591	5.9131	T/A	íntron	OS10T0352000-01
1687	10	10995369	0.0008	0.0591	5.9131	C/G	éxon	OS10T0358700-01
1688	10	11019544	0.0008	0.0591	5.9131	C/G	íntron	-
1689	10	11058540	0.0008	0.0591	5.9131	T/A	íntron	-
1690	10	11077666	0.0008	0.0591	5.9131	A/T	íntron	-
PP								
450	2	8414406	0.0004	0.2473	7.0934	T/A	íntron	-

3.3.5. Análise de associação utilizando marcadores silico DArTs

Entre os 24.266 silico DArTs identificados ao longo dos 12 cromossomos após a filtragem e imputação, 113 marcadores silico DArTs (aproximadamente 0,46%) foram identificados como significativamente associados as características fenotípicas avaliadas. Todas as seis características fenotípicas estudadas apresentaram associações, como pode ser observado na Figura 8.

Foram encontradas 16 associações relacionadas à característica EP nos cromossomos 3, 7 e 11. Em relação à característica CICLO foram identificadas 20 associações nos cromossomos 3, 8, 9, 10 e 12. Quanto a MMG, 13 associações detectadas estão distribuídas nos cromossomos 11 e 12. A característica PP teve apenas uma associação significativa encontrada no cromossomo 5. O NDF foi a característica que apresentou o maior número de associações, ao todo foram 61 associações detectadas nos cromossomos 3, 4, 7, 10 e 11. E por fim, a característica PP que teve apenas duas associações significativas encontradas no cromossomo 11.

Os gráficos *Q-Q plot* (Figura 8) foram gerados para os p-valores observados contra os p-valores esperados usando o modelo MLM ($Q + K$), onde os p-valores observados foram obtidos a partir do mapeamento associativo e os p-valores esperados partindo do pressuposto de que nenhuma associação ocorreu entre o marcador e a característica. A linha azul representa os p-valores esperados e a linha vermelha representa os p-valores observados.

Figura 8. *Manhattan plot* e *Q-Q plot* da análise de associação para as características agrônômicas EP, CICLO, MMG utilizando 24.266 marcadores silico DArTs em 186 acessos de arroz da subespécie *japonica* (continua)

Figura 8. *Manhattan plot* e *Q-Q plot* da análise de associação para as características agrônômicas PP, NDF e CP utilizando 24.266 marcadores silico DArTs em 186 acessos de arroz da subespécie *japonica* (conclusão)

Nas Tabelas 4 e 5 pode-se verificar as informações pertinentes em relação a cada marcador significativamente associado com as características fenotípicas avaliadas. Há o marcador silico DArT, o cromossomo, a posição genética dentro do cromossomo onde o marcador foi encontrado, o p-valor, a MAF, o R^2 (%), os alelos, a região e o transcrito para aqueles marcadores encontrados dentro de genes.

Considerando-se os 113 silico DArTs associados significativamente na análise, 90 silico DArTs foram localizados em regiões intergênicas e 23 silico DArTs foram localizados dentro de genes.

Como resultado do R^2 nesta análise, destaca-se o silico DArT 7873 posicionado no cromossomo 3 que representa 9,4783% da variação fenotípica da característica número de dias para o florescimento (NDF), o silico DArT 20226 posicionado no cromossomo 8 que

representa 9,2269% da variação fenotípica da característica ciclo e o silico DArT 20245 posicionado no cromossomo 8 que representa 8,6182% da variação fenotípica da característica ciclo.

Alguns silico DArTs identificados como significativamente associados apresentaram sobreposição de efeito.

Tabela 4. Marcadores silico DArTs que apresentaram associação com as características de produtividade EP, CICLO, MMG, PP e CP, identificados entre os 186 acessos de *Oryza Sativa* L. spp. *japonica*. São apresentados: as marcas significativamente associadas (SNPs), o cromossomo, a posição do loco SNP no cromossomo (genoma de referência Rice_v9 - Phytosome assembly v9), o p-valor, a mínima frequência alélica (MAF), a estimativa da variação fenotípica em % devido ao efeito do marcador, os alelos de cada loco SNP, a região gênica na qual o SNP está inserido e o transcrito com base no *International Rice Genome Sequencing Project* (IRGSP) (continua)

Silico DArT	Chr.	Position	p value	MAF	R ² (%)	Alelos	Região	Transcrito
EP								
16959	7	4854091	0.00008	0.2043	7.3515	T/A	íntron	-
16960	7	4856793	0.00017	0.2473	6.7063	A/T	íntron	-
7873	3	16110485	0.00036	0.0914	6.0062	A/T	íntron	-
17869	7	19057750	0.00058	0.4301	5.5743	A/T	íntron	-
17870	7	19057816	0.00058	0.4301	5.5743	T/A	íntron	-
25342	11	6922999	0.00076	0.2151	5.3220	T/A	íntron	-
26676	11	23987310	0.00094	0.0914	5.1347	A/T	íntron	-
26677	11	23987381	0.00094	0.0914	5.1347	A/T	íntron	-
26686	11	24068294	0.00094	0.0914	5.1347	T/A	éxon	OS11T0618300-01
26706	11	24102957	0.00094	0.0914	5.1347	G/C	íntron	-
26736	11	24272139	0.00094	0.0914	5.1347	T/A	íntron	-
26660	11	23923116	0.00094	0.0914	5.1347	A/T	íntron	OS11T0615250-00
26661	11	23923143	0.00094	0.0914	5.1347	T/A	íntron	OS11T0615250-00
26673	11	23963060	0.00094	0.0914	5.1347	G/C	íntron	-
26708	11	24103023	0.00094	0.0914	5.1347	T/A	íntron	-
26712	11	24110032	0.00094	0.0914	5.1347	T/A	íntron	-
CICLO								
20226	8	16727756	0.00007	0.2957	9.2269	T/A	íntron	-
20245	8	16898671	0.00011	0.3065	8.6182	T/A	íntron	-
20267	8	17150071	0.00015	0.2849	8.2805	T/A	íntron	-
22622	9	21160024	0.00040	0.1129	7.1908	T/A	íntron	-
20259	8	17115805	0.00043	0.2796	7.1128	A/T	íntron	-
20260	8	17115871	0.00043	0.2796	7.1128	T/A	íntron	-
20269	8	17178956	0.00043	0.2796	7.1128	T/A	íntron	-
27835	12	4386200	0.00045	0.2849	7.0714	T/A	íntron	-
20176	8	16322037	0.00047	0.3011	7.0043	T/A	éxon	OS08T0356700-01
20156	8	16092052	0.00049	0.3065	6.9716	T/A	íntron	-
20161	8	16165407	0.00049	0.3065	6.9716	G/C	íntron	-
20164	8	16166096	0.00049	0.3065	6.9716	G/C	íntron	-
20165	8	16166162	0.00049	0.3065	6.9716	T/A	íntron	-
20270	8	17178957	0.00056	0.2849	6.8128	T/A	íntron	-
20243	8	16898605	0.00067	0.2742	6.6264	T/A	íntron	-
20246	8	16916155	0.00067	0.2742	6.6264	T/A	íntron	-
8088	3	20700016	0.00080	0.0699	6.4316	A/T	íntron	-
20292	8	17462129	0.00082	0.2742	6.3982	T/A	íntron	-
20293	8	17462962	0.00082	0.2742	6.3982	T/A	íntron	-
23353	10	6881791	0.00097	0.1290	6.2179	T/A	íntron	-
MMG								
28559	12	14600123	0.00064	0.2688	6.1662	T/A	íntron	-
28593	12	15161017	0.00064	0.2688	6.1662	G/C	íntron	-
28594	12	15161083	0.00064	0.2688	6.1662	T/A	íntron	-
28595	12	15167600	0.00064	0.2688	6.1662	G/C	íntron	-
28596	12	15168405	0.00064	0.2688	6.1662	A/T	íntron	-
26529	11	22906379	0.00073	0.1613	6.0339	T/A	íntron	OS11T0599000-00
26530	11	22907470	0.00073	0.1613	6.0339	T/A	íntron	OS11T0599000-00
28576	12	14953548	0.00092	0.2634	5.8004	T/A	íntron	-
28599	12	15176170	0.00092	0.2634	5.8004	T/A	íntron	-
28615	12	15338383	0.00093	0.2742	5.7919	T/A	íntron	OS12T0448900-01
28525	12	14094615	0.00094	0.2473	5.7749	C/G	íntron	-
26555	11	23057415	0.00095	0.2527	5.7674	T/A	íntron	-
28555	12	14541532	0.00099	0.2688	5.7291	T/A	éxon	OS12T0439201-01
PP								
12068	5	972769	0.00060	0.2204	6.7435	G/C	íntron	-
CP								
26652	11	23793422	0.00019	0.1075	6.7715	C/G	íntron	-
26746	11	24322995	0.00060	0.1022	5.6873	T/A	íntron	-

Tabela 5. Marcadores silico DArTs que apresentaram associação com a característica de produtividade NDF, identificados entre os 186 acessos de *Oryza Sativa* L. spp. *japonica*. São apresentados: as marcas significativamente associadas (SNPs), o cromossomo, a posição do loco SNP no cromossomo (genoma de referência Rice_v9 - Phytosome assembly v9), o p valor, a mínima frequência alélica (MAF), a estimativa da variação fenotípica em % devido ao efeito do marcador, os alelos de cada loco SNP, a região gênica na qual o SNP está inserido e o transcrito com base no *International Rice Genome Sequencing Project* (IRGSP) (conclusão)

Silico DArT	Chr.	Position	p value	MAF	R ² (%)	Alelos	Região	Transcrito
NDF								
7873	3	16110485	0.00002	0.0914	9.4783	A/T	íntron	-
11422	4	30780676	0.00009	0.0699	7.8814	T/A	íntron	-
7876	3	16163705	0.00011	0.1075	7.6783	T/A	íntron	-
16959	7	4854091	0.00017	0.2043	7.2128	T/A	íntron	-
26739	11	24292430	0.00041	0.0914	6.3682	A/T	íntron	-
26762	11	24411833	0.00042	0.0860	6.3351	T/A	íntron	-
26664	11	23926064	0.00059	0.0860	6.0108	G/C	éxon	OS11T0615300-00
26665	11	23926130	0.00059	0.0860	6.0108	T/A	éxon	OS11T0615300-00
26666	11	23937223	0.00059	0.0860	6.0108	T/A	íntron	-
26667	11	23938493	0.00059	0.0860	6.0108	A/T	éxon	OS11T0615700-01
26680	11	24002538	0.00059	0.0860	6.0108	A/T	íntron	OS11T0616900-01
26682	11	24048235	0.00059	0.0860	6.0108	A/T	íntron	-
26683	11	24066364	0.00059	0.0860	6.0108	C/G	íntron	-
26687	11	24068823	0.00059	0.0860	6.0108	T/A	íntron	OS11T0618300-01
26688	11	24068824	0.00059	0.0860	6.0108	T/A	íntron	OS11T0618300-01
26697	11	24085074	0.00059	0.0860	6.0108	A/T	íntron	OS11T0618600-00
26698	11	24090879	0.00059	0.0860	6.0108	C/G	éxon	OS11T0618700-01
26702	11	24095152	0.00059	0.0860	6.0108	C/G	éxon	OS11T0618700-01
26703	11	24097105	0.00059	0.0860	6.0108	A/T	íntron	-
26707	11	24102957	0.00059	0.0860	6.0108	G/C	íntron	-
26713	11	24126981	0.00059	0.0860	6.0108	A/T	íntron	-
26714	11	24127047	0.00059	0.0860	6.0108	T/A	íntron	-
26716	11	24133895	0.00059	0.0860	6.0108	T/A	íntron	-
26635	11	23729605	0.00061	0.0968	5.9773	C/G	íntron	-
26636	11	23729754	0.00061	0.0968	5.9773	T/A	íntron	-
26638	11	23733863	0.00061	0.0968	5.9773	T/A	íntron	-
26639	11	23733929	0.00061	0.0968	5.9773	T/A	íntron	-
26640	11	23738928	0.00061	0.0968	5.9773	T/A	íntron	-
26726	11	24260213	0.00061	0.0968	5.9773	A/T	íntron	-
26741	11	24295310	0.00061	0.0968	5.9773	T/A	íntron	-
26748	11	24336517	0.00061	0.0968	5.9773	A/T	íntron	-
26799	11	24654733	0.00061	0.0968	5.9773	T/A	íntron	-
26804	11	24666757	0.00061	0.0968	5.9773	T/A	éxon	OS11T0629550-00
26807	11	24667309	0.00061	0.0968	5.9773	T/A	éxon	OS11T0629582-00
							éxon	OS11T0629566-00
26810	11	24718778	0.00061	0.0968	5.9773	T/A	éxon	OS11T0630300-01
26811	11	24764746	0.00061	0.0968	5.9773	T/A	íntron	-
26814	11	24790260	0.00061	0.0968	5.9773	T/A	íntron	-
26817	11	24799860	0.00061	0.0968	5.9773	C/G	íntron	-
26822	11	24808840	0.00061	0.0968	5.9773	T/A	éxon	OS11T0631450-01
26827	11	24895289	0.00061	0.0968	5.9773	T/A	íntron	-
26829	11	24917837	0.00061	0.0968	5.9773	T/A	íntron	-
26831	11	24917903	0.00061	0.0968	5.9773	T/A	íntron	-
26833	11	24924035	0.00061	0.0968	5.9773	A/T	íntron	-
26834	11	24924101	0.00061	0.0968	5.9773	T/A	íntron	-
23602	10	11434507	0.00061	0.0968	5.9773	T/A	íntron	-
26637	11	23733830	0.00061	0.0968	5.9773	T/A	íntron	-
26644	11	23787689	0.00061	0.0968	5.9773	T/A	íntron	-
26729	11	24263997	0.00061	0.0968	5.9773	G/C	íntron	-
26770	11	24447442	0.00061	0.0968	5.9773	T/A	íntron	-
26774	11	24456240	0.00061	0.0968	5.9773	A/T	íntron	-
26798	11	24654733	0.00061	0.0968	5.9773	T/A	íntron	-
26803	11	24666757	0.00061	0.0968	5.9773	T/A	éxon	OS11T0629550-00
							éxon	OS11T0629566-00
26806	11	24667309	0.00061	0.0968	5.9773	T/A	éxon	OS11T0629582-00
							éxon	OS11T0629566-00
26816	11	24799860	0.00061	0.0968	5.9773	C/G	íntron	-
26819	11	24806616	0.00061	0.0968	5.9773	C/G	íntron	-
26828	11	24917098	0.00061	0.0968	5.9773	T/A	íntron	-
16960	7	4856793	0.00069	0.2473	5.8540	A/T	íntron	-
26738	11	24287335	0.00071	0.0968	5.8251	G/C	íntron	-
26642	11	23787304	0.00071	0.0968	5.8251	C/G	íntron	-
25342	11	6922999	0.00088	0.2151	5.6154	T/A	íntron	-
23199	10	4686323	0.00093	0.0753	5.5564	C/G	íntron	-

3.3.6. Anotação dos genes relacionados

A partir do posicionamento dos SNPs dentro de genes foi possível identificar transcritos candidatos que poderiam estar relacionados à produtividade em arroz, de acordo com o genoma de referência da cultivar Nipponbare *Oryza sativa* L. ssp. *japonica* (Os-Nipponbare-Reference-IRGSP-1.0).

A partir desses transcritos foi realizada uma busca sobre a anotação funcional baseada nas três classes de termos GO, depositados no banco de dados do *UniProt*. Para poucos transcritos identificados houve alguma anotação sugerida, a maioria dos transcritos são apresentados com função desconhecida.

Considerando-se os marcadores SNPs tivemos dois genes com anotação no *UniProt*: o SNP 1191 que foi localizado dentro do gene *OS06G0281400* identificado com funções moleculares relacionadas a ligação de DNA (GO:0003677) e a atividade da topoisomerase tipo I de DNA (GO:0003917) com processo biológico de alteração topológica do DNA (GO:0006265); e o SNP 1679 que foi localizado dentro do gene *OS10G0346600* com função molecular de ligação de íons de cálcio (GO:0005509) sendo componente celular integral da membrana (GO:0016021).

Em relação aos marcadores silico DArTs para cinco genes foram encontradas anotações: os silico DArTs 26686 e 26688 localizados dentro do gene *OS11G0618300* com funções moleculares relacionadas a ligação de ATP (GO:0005524) e atividade da proteína quinase (GO:0004672) sendo componente celular integral da membrana (GO:0016021); os silico DArTs 26660 e 26661 localizados dentro do gene *OS11G0615250* com função molecular relacionada a atividade de timidilato sintase (GO:0004799) com processo biológico relacionado ao processo biossintético de dTMP (GO:0006231); o silico DArT 26697 dentro do gene *OS11G0618600* com funções moleculares relacionadas a ligação de ATP (GO:0005524) e atividade da proteína quinase (GO:0004672); os silico DArTs 26698 e 26702 dentro do gene *OS11T0618700-01* como componente celular integral da membrana (GO:0016021); e, os silicos DArTs 26804 e 26803 dentro do gene *OS11G0629550* com função molecular de ligação de ADP (GO:0043531).

3.4. Discussão

3.4.1. Matriz de parentesco e desequilíbrio de ligação

Os marcadores moleculares SNPs e silico DArTs foram usados para calcular a matriz de parentesco (k) entre os pares de indivíduos utilizando o algoritmo *Loiselle* (LOISELLE et al., 2008). A matriz de parentesco reflete a similaridade entre dois indivíduos aleatórios, fornecendo informações úteis para estudos de herança quantitativa (YU et al., 2006); (ZHANG et al., 2014).

A similaridade entre os pares de indivíduos estão representados na linha central dos gráficos apresentados na Figura 5, quanto menor a similaridade entre os indivíduos mais próximo de 0 é o valor. Os resultados da matriz de parentesco obtidos neste trabalho indicam que os acessos do banco de germoplasma do departamento de genética da ESALQ/USP não apresentam relação forte entre si, o que pode ser devido ao fato de que esses materiais vieram de regiões diversificadas de cultivo de arroz do Japão.

Da mesma forma, para os dois conjuntos de marcadores, calculamos o desequilíbrio de ligação (DL) através do quadrado da correlação (r^2) (Figura 6). Os resultados mostram que o decaimento do DL em ambos os gráficos aumentou à medida que aumentou a distância genética.

A análise do desequilíbrio de ligação é de grande importância para o mapeamento associativo, uma vez que esta abordagem detecta associações entre o marcador e a característica de interesse indiretamente através do DL (WÜRSCHUM et al., 2013). O DL é definido como a diferença entre as proporções haplotípicas observadas e aquelas esperadas se os alelos segregassem independentemente (Weiss e Clark, 2002). Quanto mais próximo de 1 maior o desequilíbrio de ligação entre os marcadores, ou seja, entre o marcador A e o marcador B, e quanto mais próximo de 0 menor é o desequilíbrio de ligação, ou seja, os alelos desses dois marcadores tendem a segregar de forma independente na população.

O DL calculado com marcadores silico DArTs apresentou uma quantidade muito maior de marcadores em desequilíbrio de ligação e um decaimento mais lento quando comparado ao resultado obtido com os marcadores SNPs. Como o número de marcadores silico DArT genotipados é 12x maior que o número de marcadores SNPs pode-se sugerir que como a análise de DL conseguiu abordar muito mais marcas possibilitou gerar um gráfico mais denso e por isso a informação a respeito do desequilíbrio ficou mais bem explorada com

os marcadores silico DArTs. O lento decaimento do DL com o aumento da distância genética é esperado em estudos de autógamias, onde o DL se estende a centenas de pares de bases.

A escolha da estatística r^2 em detrimento ao D' para análise do DL foi devido ao fato de r^2 reunir tanto o histórico de recombinação quanto de mutação da população, enquanto que o D' mostra apenas o histórico de recombinação (FLINT-GARCIA; THORNSBERRY; BUCKLER, 2003). O DL histórico é muito menor em espécies de fecundação cruzada do que em espécies autógamias, como o arroz, por exemplo, pois a autofecundação causa e mantém a homozigotidade, tornando as recombinações ineficazes na redução da DL. Por um lado, são necessários menos marcadores para cobrir completamente o genoma de espécies autógamias do que alógamas; Por outro lado, a resolução de mapeamento possível é limitada (WÜRSCHUM et al., 2013).

3.4.2. Mapeamento associativo

Os acessos de arroz depositados em bancos de germoplasma representam uma fase intermédia entre a domesticação de materiais selvagens e cultivares elite (LONDO et al., 2006), possuindo alta diversidade genética e muitos genes diferentes que proporcionam recursos genéticos para aplicação em programas de melhoramento. Nesse contexto, o mapeamento por associação baseado em acessos de germoplasma é útil para capturar essa variação fenotípica disponível, pois permite aos pesquisadores através do uso de tecnologias genéticas modernas explorar a diversidade natural e localizar genes valiosos no genoma (ZHANG et al., 2014).

O mapeamento por associação identifica os locos de características quantitativas (QTLs) examinando as associações entre marcador-característica atribuídas à força do desequilíbrio de ligação entre marcadores e polimorfismos funcionais em um conjunto de germoplasma (ZHU et al., 2008).

Aqui o mapeamento associativo foi baseado no modelo linear misto (MLM) que incorpora informações sobre o grau de parentesco (*kinship*) e estrutura genética da população (PCA). Esse procedimento é essencial para controlar possíveis associações espúrias ou falsos positivos no mapeamento associativo, reduzindo as chances de ocorrência do erro tipo I e tipo II, que ocorrem devido a existência de populações estruturadas (PRITCHARD et al., 2000); (RICHARDS; FRIESEN; BRUEGGEMAN, 2017); (YU et al., 2006).

A escolha do PCA para controle da estrutura genética da população deveu-se ao menor tempo computacional requerido e também a boa qualidade das informações obtidas.

Em um trabalho realizado por (ZHAO et al., 2007) os resultados apontaram que o método PCA fornece resultados comparáveis com abordagens alternativas como o algoritmo *Structure*.

Dois conjuntos de marcadores foram utilizados para proceder as análises, os SNPs e os silico DArTs, obtidos a partir da genotipagem por sequenciamento DArT-seq. O mapeamento associativo foi realizado através das ferramentas disponíveis no pacote *Genome Association and Prediction Integrated Tool* (GAPIT) (LIPKA et al., 2012) (TANG et al., 2016) do programa R. O uso do pacote GAPIT possui vantagens em relação aos métodos tradicionais de mapeamento associativo, pois analisa grandes conjuntos de dados com tempo computacional mínimo e produz resultados abrangentes incluindo objetos R e gráficos de alta qualidade (LIPKA et al., 2012). Apenas associações marcador-característica com p valor $<0,001$ foram declaradas significativas.

Para observar os SNPs e silico DArTs significativos resultantes foram obtidos gráficos *Manhattan Plot* elaborados a partir das coordenadas genômicas das associações exibidas ao longo do eixo X e com o logaritmo negativo do p-valor da associação exibido no eixo Y. Na análise de associação realizada entre seis características de produtividade e os dados genotípicos obtidos da genotipagem de 186 acessos japoneses de arroz, foram detectados 21 SNPs e 113 silico DArTs significativamente associados.

Entre os 1.965 marcadores SNPs identificados ao longo dos 12 cromossomos 1% foi identificado como significativamente associado às características fenotípicas avaliadas e entre os 24.266 marcadores silico DArTs identificados ao longo dos 12 cromossomos 0,46% foi identificado como significativamente associado às características fenotípicas avaliadas. Esses resultados sugerem que os marcadores SNPs possuem uma eficiência maior para detecção de associações, apesar do número de marcadores silico DArTs ter sido 12x maior do que o número de marcadores SNPs encontrados.

Os p-valores observados versus os p-valores esperados para cada associação marcador-característica foram plotados em gráficos *Q-Q plots* para avaliar o controle de erros tipo I. Foram observadas distribuições uniformes entre os p-valores observados e esperados quando utilizado marcadores SNPs (Figura 7) indicando que o modelo utilizado MLM (Q + K) apresentou um bom desempenho no mapeamento associativo para esses marcadores. Entretanto, a distribuição dos p-valores no mapeamento associativo utilizando marcadores silico DArTs não seguiu uma uniformidade satisfatória (Figura 8). Quando o modelo aplicado tem um bom desempenho na análise ele tem o potencial para corrigir as associações falso

positivas, os desvios da expectativa demonstram que a análise estatística pode ter causado associações espúrias.

Utilizando marcadores SNPs foi encontrada uma associação relacionada à característica EP, localizada no cromossomo 1. Para a característica CICLO foram identificadas três associações nos cromossomos 1, 8 e 11. A MMG se destacou, apresentando o maior número de associações, 16 no total, distribuídas nos cromossomos 6 e 10. E, para a característica PP foi encontrada uma associação significativa detectada no cromossomo 2.

Utilizando marcadores silico DArTs foram encontradas 16 associações relacionadas à característica EP nos cromossomos 3, 7 e 11. Foram identificadas 20 associações relacionadas à característica CICLO nos cromossomos 3, 8, 9, 10 e 12. Para MMG 13 associações foram detectadas nos cromossomos 11 e 12. Apenas uma associação significativa foi encontrada para a característica PP no cromossomo 5. O NDF foi a característica que apresentou o maior número de associações, ao todo foram 61 associações detectadas nos cromossomos 3, 4, 7, 10 e 11. Por fim, a característica PP teve duas associações significativas encontradas no cromossomo 11.

O padrão de distribuição entre os marcadores SNPs e silico DArTs utilizados é diferente, SNPs apresentam característica co-dominante enquanto que silico DArTs são dominantes. Marcadores co-dominantes permitem identificar genótipos heterozigotos e homozigotos, enquanto que os marcadores dominantes possuem informação genética limitada por loco, que não permite distinguir genótipos heterozigotos dos homozigotos. Isto impede a estimativa direta de frequências alélicas a partir dos dados genotípicos e consequentemente dos diversos parâmetros que determinam a estrutura genética de populações (WILLIAMS J K F, KUBELIK A R, LIVAK K G, 1990) (FERREIRA; GRATTAPAGLIA, 1998).

A dominância apresentada pelos marcadores silico DArTs pode restringir a possibilidade de sua utilização em análises genéticas onde seja importante diferenciar homozigotos de heterozigotos. Isto é particularmente importante em populações de plantas alógamas, onde a porcentagem de locos em heterozigose é alta. Nos casos onde os indivíduos avaliados são todos homozigotos, a segregação dominante não representa limitação, como em seleções dentro de variedades compostas por uma mistura de linhas puras ou populações locais onde os indivíduos embora heterogêneos entre si, são homozigotos. Em estudos envolvendo diversidade ou distância genética em plantas autógamas, como as plantas a serem analisadas para a identificação e determinação da diversidade são homozigotas, a segregação dominante não produz menos informações do que os marcadores co-dominantes (HOFFMANN; BARROSO, 2006).

A maior limitação dos marcadores dominantes é o fato deles apresentarem baixa transferibilidade mesmo entre indivíduos da mesma espécie. Portanto, mapas genéticos construídos com esses marcadores têm uma utilidade limitada, que se restringe quase que exclusivamente ao próprio *pedigree* utilizado para a sua construção, impossibilitando, por exemplo, experimentos de mapeamento comparativo e validação de QTLs (SANSALONI; GRATAPAGLIA, 2012).

As principais vantagens dos marcadores SNPs são que eles ocorrem em genomas com frequência muito grande, com cerca de um SNP sendo observado por 140 pb no arroz (IRGPS, 2005), detectando uma grande quantidade de polimorfismos entre alelos de determinado gene. Sendo mais reveladores (ou seja, nenhuma ambiguidade alélica) do que os marcadores dominantes e, portanto, mais poderosos na estimativa da estrutura da população (Q) e da matriz de parentesco (K) (ZHU et al., 2008). Já suas principais desvantagens são a necessidade de conhecimento prévio da sequência do gene de interesse e o custo/infraestrutura da técnica envolvida nas etapas necessárias ao sequenciamento dos diferentes fragmentos do DNA de interesse (FALEIRO, 2007).

Devido à maior densidade no genoma, menor taxa de mutação e melhor capacidade de detecção para sistemas de alto rendimento, os SNPs estão se tornando o marcador de escolha para estudos complexos de dissecção de características (ZHU et al., 2008). Com a crescente utilização do sequenciamento por genotipagem, que garante um número cada vez maior de SNPs, o mapeamento associativo utilizando esse tipo de marcador tem sido considerado uma ferramenta promissora (GUPTA; RUSTGI; KULWAL, 2005),

No trabalho realizado por (SINGH et al., 2013) que comparou marcadores SSR e SNPs, ao nível da estrutura populacional, os marcadores SNPs mostraram melhor relação genética com a população, enquanto que em nível de diversidade o SSR apresentou melhores amostras de agrupamento. Por esta razão, os marcadores SNPs devem ser preferencialmente utilizados para a determinação da estrutura populacional nas culturas. Além disso, os marcadores SNPs são na sua maioria derivados de genes, como o resultado da diversidade genética avaliada utilizando esses marcadores revela variações funcionais eles podem ser potencialmente exploradas para estudos de associação entre marcador-característica. Adicionalmente, os marcadores SNP foram derivados de uma região genômica do arroz que apresenta sintonia com o genoma do trigo e, portanto, podem ser igualmente úteis para avaliar a diversidade genética, a estrutura da população e outros estudos na cultura do trigo.

Os SNPs são uma grande promessa para abordagens genômicas, como no mapeamento por associação e na seleção genômica. Todavia, um número maior de SNPs no

presente trabalho certamente levaria a um maior número de associações significativas e detalhadas dos acessos estudados, apesar da alta qualidade dos SNPs após a filtragem e imputação, 1.965 marcas é um número baixo para realização do mapeamento associativo. Comparado com uma população natural ou uma população melhorada com ampla base genética, o nível de LD em germoplasma pode ser baixo devido à sua origem diversa, portanto, mais marcadores são requeridos para o mapeamento associativo (ZHANG et al., 2014).

O mapeamento de genes candidatos através de associação requer a identificação de SNPs entre e dentro de genes específicos. Poucos indivíduos diversos no painel de estudo de SNPs são necessários para identificar SNPs comuns, no entanto muitos outros são necessários para identificar SNPs mais raros. As regiões de promotores, íntrons, éxons e 5' / 3' são todos alvos para a identificação de SNPs de genes candidatos. A taxa de decaimento de LD para um loco de gene candidato específico determina o número de SNPs por unidade de comprimento (por exemplo, kb) necessário para identificar associações significativas (WHITT; BUCKLER, 2003). Portanto, o número e o comprimento dos amplicons em pares de base durante o sequenciamento necessários para amostrar suficientemente um loco de um gene candidato dependem quase que totalmente da distribuição de LD e dos SNPs, com uma maior densidade de marcadores SNPs necessários em regiões de LD relativamente baixa e diversidade de nucleotídeos elevados (ZHU et al., 2008).

Um número total de 150.325 SNPs distribuídos pelos 12 cromossomos de arroz foi utilizado no mapeamento associativo para características de rendimento de grãos em arroz com um total de 26 associações detectadas no trabalho de (PANTALIO et al., 2016). No trabalho de (BISCARINI et al., 2016) que estudou acessos de arroz de clima temperado 57.000 SNPs foram mantidos no mapeamento associativo para estudo de características da morfologia da planta e do grão e arquitetura de raiz, os resultados apontaram um total de 42 associações significativas.

Em relação aos marcadores silico DArTs acreditamos ter alcançado neste trabalho um número adequado de marcas, 24.266. No trabalho de (MWADZINGENI et al., 2017) um total de 16.383 marcadores silico DArTs distribuídos pelos 21 cromossomos do trigo foram mantidos para análises de estrutura da população, desequilíbrio de ligação e de associação resultando em 334 associações significativas. No trabalho de (BELLUCCI et al., 2015) 1.429 SNPs e 4.675 silico DArTs foram utilizados identificando 12 associações no mapeamento para altura da planta, rendimento de grãos e potencial de biomassa para a produção de bioetanol em trigo. Outro estudo de mapeamento associativo em arroz, utilizando acessos

japoneses e 9.727 silico DArTs e 6.717 SNPs detectou 97 associações significativas relacionadas a características da raiz (COURTOIS et al., 2013).

Comparado com estudos de outras plantas modelo como *Arabidopsis*, uma grande vantagem do arroz é a informação muito mais extensa disponível para um conjunto diverso de importantes características agronomicamente conhecidas de milhares de variedades em muitos ambientes diferentes. O conhecimento detalhado dos fenótipos, juntamente com um banco de dados aprofundado de genótipos, cria uma poderosa plataforma para a associação genética e a descoberta de alelos que podem ser combinados para alcançar os objetivos de aumento da produção de arroz nos próximos anos (MCNALLY et al., 2009).

Uma vez que as funções de todos os genes do arroz forem compreendidas, o melhorista poderá selecionar todas as combinações de alelos no banco de dados integrado genótipo-fenotípico. Os genes desejáveis poderão ser trazidos ao genótipo individual através da piramidação ou através da engenharia genética. Os marcadores moleculares estreitamente ligados a um gene desejado poderão ser usados para o desenvolvimento de novos cultivares através da seleção assistida por marcadores (MAS) (ZHANG et al., 2014). O uso de seleção assistida das marcas encontradas pode ser particularmente interessante aos melhoristas que pretendem incrementar seus programas de melhoramento com materiais que apresentem potencial para aumento de produtividade.

Uma vez obtidas associações verdadeiras com maior grau de confiabilidade, o passo seguinte consiste em testar as associações encontradas em diferentes locais, anos e em populações distintas, com diferentes *backgrounds* genéticos. Aqui o mapeamento associativo foi realizado a partir de dados fenotípicos obtidos na mesma área em dois anos agrícolas, de modo que as associações encontradas deverão ser confirmadas em oportunidades futuras.

3.4.3. Anotação de genes

A fim de estudar com mais profundidade as associações encontradas, os marcadores identificados como significativamente associados foram investigados no banco de dados *Gramene*, especializado em informações genéticas de espécies gramíneas e que conta com uma base de dados específica para *Oryza sativa* ssp. *japonica* (IRGSP-1.0), esta base de dados reúne informações como o mapa genético integrado do arroz, sequências e genes previamente detectados. A partir das informações de posição e o cromossomo onde cada marcador está inserido conseguimos pesquisar alguns detalhes a respeito dos alelos de cada

loco SNP, da região gênica na qual o SNP está inserido e o transcrito com base no International Rice Genome Sequencing Project (IRGSP) (Tabelas 3, 4 e 5).

Algumas marcas foram localizadas dentro de genes, por sua vez, para os transcritos desses genes foi realizada uma busca sobre sua possível anotação funcional baseada nas três classes de termos GO, depositados no banco de dados *UniProt*. A anotação GO utiliza palavras-chave que resumizam o conhecimento sobre as atividades funcionais de proteínas, os processos ou funções biológicas mais gerais nos quais as atividades específicas das proteínas estão inseridas, e a localização celular onde os processos ocorrem (BLAKE, 2013). Desta forma, os termos GO foram utilizados neste estudo para avaliar as funções das proteínas codificadas pelos genes nos quais os SNPs e silico DArTs foram inseridos.

A partir da busca no banco de dados *Gramene* foi verificado que entre os 21 SNPs e 113 silico DArTs identificados, 10 SNPs e 23 silico DArTs estavam em regiões transcritas do genoma, sendo que, para dois destes SNPs e para cinco silico DArTs haviam anotações funcionais definidas por termos GO. A maioria dos transcritos foram identificados com função desconhecida não havendo, portanto, estudos relacionados para determinar, relacionar ou prever a função aos genes em questão.

De modo geral, foram encontradas anotações funcionais diversas para os produtos dos genes nos quais os SNPs estavam inseridos. Houveram algumas coincidências nas anotações associadas aos transcritos, anotações envolvendo o metabolismo e transporte de moléculas foram encontrados para transcritos nos SNPs e silico DArTs.

Nenhum dos locos encontrados e avaliados neste estudo estão localizados em genes associados diretamente à sinalização celular. Entretanto, os silico DArTs 26686, 26687 e 26688 que estão localizados no gene *OS11G0618300* e o silico DArT 26697 que está localizado dentro do gene *OS11G0618600* estão associados a atividade proteíno-quinase. As proteínas quinases são enzimas que catalisam a transferência de grupos fosfatos para resíduos de aminoácidos das proteínas (LUCHS, 2006) e, embora estejam associadas com uma grande variedade de processos celulares, comumente estão envolvidas com processo de sinalização celular (NELSON; COX, 2002).

O SNP 1679 que foi localizado dentro do gene *OS10G0346600* está associado a função molecular de ligação de íons de cálcio na membrana celular, podendo fazer parte dos receptores da membrana que atuam como sinalizadores intracelulares, os íons de cálcio atuam como mensageiros secundários durante o processo de sinalização celular (DE MOURA; PINTO VIDAL, 2011).

As funções anotadas na base de dados *UniProt* estão mais relacionadas a funções bioquímicas básicas que podem estar, direta ou indiretamente, ligadas a sinalização celular. Devido a uma série de fatores é difícil fazer uma relação direta entre os genes em que se localizam os SNPs e silico DArTs e a produtividade no arroz. Primeiramente, o conjunto de genes avaliados no *UniProt* foi determinado por métodos de predição gênica e as anotações funcionais se baseiam na similaridade dos genes preditos com genes de outras espécies. Além disso, os genes nos quais os marcadores estavam inseridos podem não estar diretamente relacionados com um fenótipo selecionado, mas em equilíbrio de ligação com genes relacionados às características diretamente selecionadas (MEYER; PURUGGANAN, 2013).

Entretanto, os genes identificados na população atual aumentam o conjunto de recursos genéticos e genes candidatos atualmente disponíveis, sendo que alguns desses locos podem estar localizados em QTLs relacionados às características em estudo. As informações geradas neste estudo podem ser de grande valia para avaliações mais aprofundadas, utilizando o conjunto de marcadores significativos como ponto de partida para determinação dos genes mais importantes para a produtividade em arroz.

3.5. Conclusão

Com base nos resultados obtidos pode-se concluir que os SNPs apresentaram maior eficiência na detecção de associações entre genótipo e fenótipo, no entanto há a necessidade de que o número de SNPs genotipados seja maior para abranger mais regiões do genoma. Ressaltamos que, os SNPs são uma ferramenta valiosa para o melhoramento de plantas já que possuem característica co-dominante. Em relação ao modelo estatístico aplicado no mapeamento associativo, o modelo linear misto (MLM), acredita-se que tenha conseguido controlar eficientemente os falsos positivos no mapeamento utilizando marcadores SNPs.

Referências

- AGARWAL, M.; SHRIVASTAVA, N.; PADH, H. **Advances in molecular marker techniques and their applications in plant sciences** Plant Cell Reports Springer-Verlag, , 2 abr. 2008. Disponível em: <<http://link.springer.com/10.1007/s00299-008-0507-z>>. Acesso em: 15 jun. 2017
- BATES, D. et al. Fitting Linear Mixed-Effects Models Using lme4. **Journal of Statistical Software**, v. 67, n. 1, p. 1–48, 2015.

- BELLUCCI, A. et al. Association Mapping in Scandinavian Winter Wheat for Yield, Plant Height, and Traits Important for Second-Generation Bioethanol Production. **Frontiers in Plant Science**, v. 6, n. November, p. 1–12, 2015.
- BISCARINI, F. et al. Genome-wide association study for traits related to plant and grain morphology, and root architecture in temperate rice accessions. **PLoS ONE**, v. 11, n. 5, p. 1–28, 2016.
- BLAKE, J. A. Ten Quick Tips for Using the Gene Ontology. **PLoS Computational Biology**, v. 9, n. 11, p. 9–10, 2013.
- BROWNING, B. L.; BROWNING, S. R. Genotype Imputation with Millions of Reference Samples. **American Journal of Human Genetics**, v. 98, n. 1, p. 116–126, 2016.
- CHO, Y.-G. et al. Identification of Quantitative Trait Loci in Rice for Yield, Yield Components, and Agronomic Traits across Years and Locations. **Crop Science**, v. 47, n. 6, p. 2403, 1 nov. 2007.
- COURTOIS, B. et al. Genome-wide association mapping of root traits in a japonica rice panel. **PLoS ONE**, v. 8, n. 11, p. 1–18, 2013.
- DE MOURA, P. R.; PINTO VIDAL, F. A. Transdução de sinais: Uma revisão sobre proteína G. **Scientia Medica**, v. 21, n. 1, p. 1–11, 2011.
- FALEIRO, G. **Marcadores genético-moleculares**. 1. ed. Planaltina, DF: Embrapa Cerrados, 2007.
- FERREIRA, M. E.; GRATTAPAGLIA, D. **Introducción al uso de marcadores moleculares en el análisis genético**. 1. ed. Brasília: EMBRAPA-CENARGEN, 1998.
- FLINT-GARCIA, S. A.; THORNSBERRY, J. M.; BUCKLER, E. S. Structure of Linkage Disequilibrium in Plants. **Annual Review of Plant Biology**, v. 54, n. 1, p. 357–374, 2003.
- GOWDA, M. et al. Advances in rice breeding, genetics and genomics. **Molecular Breeding**, v. 11, n. 5, p. 337–352, 2003.
- GUPTA, P. K.; RUSTGI, S.; KULWAL, P. L. **Linkage disequilibrium and association studies in higher plants: Present status and future prospects** *Plant Molecular Biology*, 2005. Disponível em:
<https://pdfs.semanticscholar.org/c07c/539d6b84b92a0aa8b7f4780a2b2566f00c76.pdf>.
 Acesso em: 13 jun. 2017
- HITTALMANI, S. et al. Fine mapping and DNA marker-assisted pyramiding of the three major genes for blast resistance in rice. **Theoretical and Applied Genetics**, v. 100, n. 7, p. 1121–1128, 12 maio 2000.

- HOFFMANN, L. V.; BARROSO, P. A. V. **Marcadores Moleculares como Ferramentas para Estudos de Genética de Plantas** Embrapa: 4. Campina Grande, PB: [s.n.]. Disponível em: <<https://ainfo.cnptia.embrapa.br/digital/bitstream/CNPA/18326/1/DOC147.pdf>>.
- HUA, J. P. et al. Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. **Genetics**, v. 162, n. 4, p. 1885–1895, 2002.
- IRGPS. The map-based sequence of the rice genome. **Nature**, v. 436, n. 7052, p. 793–800, 11 ago. 2005.
- KILIAN, A. et al. The fast and the cheap SNP and DArT-based whole genome profiling for crop improvement. **Proceedings of the International Congress**, v. 1, n. 5, p. 443–461, 2005.
- KILIAN, A. et al. Diversity arrays technology: a generic genome profiling technology on open platforms. **Methods in molecular biology (Clifton, N.J.)**, v. 888, p. 67–89, 2012.
- LIPKA, A. E. et al. GAPIT: Genome association and prediction integrated tool. **Bioinformatics**, v. 28, n. 18, p. 2397–2399, 2012.
- LOISELLE, B. A. et al. Spatial Genetic Structure of a Tropical Understory Shrub , *Psychotria officinalis* (Rubiaceae) Author (s): Bette A . Loiselelle , Victoria L . Sork , John Nason , Catherine Graham Published by: Botanical Society of America Stable URL : <http://www.jstor>. **America**, v. 82, n. 11, p. 1420–1425, 2008.
- LONDO, J. P. et al. Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. **Proceedings of the National Academy of Sciences**, v. 103, n. 25, p. 9578–9583, 2006.
- LU, C. et al. Comparative mapping of QTLs for agronomic traits of rice across environments using a doubled haploid population. **TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik**, v. 93, n. 8, p. 1211–7, dez. 1996.
- LUCHS, A. Transdução de sinal: um olhar sobre a insulina. **Revista do Instituto Adolfo Lutz (Impresso)**, v. 65, n. 3, p. 157–164, 2006.
- MCNALLY, K. L. et al. 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**, v. 106, n. 30, p. 12273–8, 28 jul. 2009.
- MEYER, R. S.; PURUGGANAN, M. D. Evolution of crop species: genetics of domestication and diversification. **Nature Reviews Genetics**, v. 14, n. 12, p. 840–852, 2013.

- MWADZINGENI, L. et al. Genome-wide association analysis of agronomic traits in wheat under drought-stressed and non-stressed conditions. **PLoS ONE**, v. 12, n. 2, p. 1–13, 2017.
- NACHIMUTHU, V. V. et al. Analysis of Population Structure and Genetic Diversity in Rice Germplasm Using SSR Markers: An Initiative Towards Association Mapping of Agronomic Traits in *Oryza Sativa*. **Rice (New York, N.Y.)**, v. 8, n. 1, p. 30, 2015.
- NELSON, D. L.; COX, M. **LEHNINGER, Princípios de Bioquímica**. 3. ed. São Paulo: Sarvier, 2002.
- PANTALIO, G. F. et al. Genome wide association study (GWAS) for grain yield in rice cultivated under water deficit. **Genetica**, v. 144, n. 6, p. 651–664, 2016.
- PRICE, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. **Nature Genetics**, v. 38, n. 8, p. 904–909, 23 ago. 2006.
- PRITCHARD, J. K. et al. Association Mapping in Structured Populations. **The American Journal of Human Genetics**, v. 67, n. 1, p. 170–181, 2000.
- RICHARDS, J. K.; FRIESEN, T. L.; BRUEGGEMAN, R. S. Association mapping utilizing diverse barley lines reveals net form net blotch seedling resistance/susceptibility loci. **Theoretical and Applied Genetics**, v. 130, n. 5, p. 915–927, 2017.
- RODRIGUES, D.; SILVA, A. **ANÁLISE DE QTL PARA PRODUTIVIDADE NO CRUZAMENTO DE ARROZ EPAGRI 108 (indica) x IRAT 122 (japônica) POR MARCADORES SNPs** [s.l: s.n.].
- SANSALONI, C. P.; GRATTAPAGLIA, O. D. **Desenvolvimento e aplicações de DArT (Diversity Arrays Technology) e genotipagem por sequenciamento (Genotyping-by-Sequencing) para análise genética em Eucalyptus**. [s.l.] Universidade de Brasília, 2012.
- SEMAGN, K.; BJORNSTAD, A.; NDJIONDJOP, M. N. An overview of molecular marker methods for plants. **African Journal of Biotechnology**, v. 5, n. 25, p. 2540–2568, 2006.
- SINGH, N. et al. Comparison of SSR and SNP Markers in Estimation of Genetic Diversity and Population Structure of Indian Rice Varieties. **PLoS ONE**, v. 8, n. 12, p. e84136, 19 dez. 2013.
- TANG, Y. et al. GAPIT Version 2: An Enhanced Integrated Tool for Genomic Association and Prediction. **The Plant Genome**, v. 9, n. 2, p. 0, 2016.
- TURNER, S. D. **qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots** bioRxiv. [s.l: s.n.]. Disponível em: <<http://biorxiv.org/content/early/2014/05/14/005165>>. Acesso em: 19 maio. 2017.

- WHITT, S. R.; BUCKLER, E. S. Using natural allelic diversity to evaluate gene function. **Methods in molecular biology (Clifton, N.J.)**, v. 236, n. 4, p. 123–140, 2003.
- WILLIAMS J K F, KUBELIK A R, LIVAK K G, R. J. A. E T. S. V. DNA polymorphism amplified by arbitrary primers are useful as genetic markers. **Nucl Acids Res**, v. 18, n. 22, p. 6531–6535, 1990.
- WÜRSCHUM, T. et al. Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. **Theoretical and Applied Genetics**, v. 126, n. 6, p. 1477–1486, 22 jun. 2013.
- XIAO, J. et al. Identification of QTLs affecting traits of agronomic importance in a recombinant inbred population derived from a subspecific rice cross. **Theoretical and Applied Genetics**, v. 92, n. 2, p. 230–244, fev. 1996.
- YU, J. et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. **Nature Genetics**, v. 38, n. 2, p. 203–208, 25 fev. 2006.
- ZHANG, P. et al. Association mapping for important agronomic traits in core collection of rice (*Oryza sativa* L.) with SSR markers. **PLoS ONE**, v. 9, n. 10, p. 16, 2014.
- ZHAO, K. et al. An Arabidopsis example of association mapping in structured samples. **PLoS Genetics**, v. 3, n. 1, p. 0071–0082, 2007.
- ZHU, C. et al. Status and Prospects of Association Mapping in Plants. **The Plant Genome Journal**, v. 1, n. 1, p. 5, 2008.

APÊNDICES

APÊNDICE A. Lista de acessos japoneses que compõem o painel do Banco de Germoplasma de arroz cultivado do Departamento de Genética da ESALQ/USP com status de domesticação e característica de cultivo. *Informações referentes à Status e Tipo foram pesquisadas no site do National Institute of Agrobiological Science - NIAS. As informações de código e nome dos acessos são de registros do banco de germoplasma de arroz do Departamento de Genética da ESALQ/USP

(continua)

Ordem	Acesso	Nome do Acesso/Genótipo	Status*	Tipo*
BGA.001	001J	Kunihikari Mochi	Landrace	Sequeiro
BGA.002	002J	Senshou	Landrace	Sequeiro
BGA.003	003J	Fukuton	Landrace	Sequeiro
BGA.004	006J	Ezo Wase	Landrace	Sequeiro
BGA.005	008J	Shin Hakaburi	Desconhecido	Desconhecido
BGA.006	12B07J	Japão 10	Desconhecido	Desconhecido
BGA.007	12B06J	Senshou Ibaragi 1	Desconhecido	Desconhecido
BGA.008	014J	Yamanoi	Melhorado	Sequeiro
BGA.009	016J	Namekata Mochi	Landrace	Sequeiro
BGA.010	023J	Sonobe Mochi	Landrace	Sequeiro
BGA.011	024J	Wase Mochi	Landrace	Sequeiro
BGA.012	025J	Seion Uruchi	Landrace	Sequeiro
BGA.013	027J	Gaisen Mochi	Landrace	Sequeiro
BGA.014	029J	Shiro Hige	Desconhecido	Desconhecido
BGA.015	035J	Kinkabou	Landrace	Sequeiro
BGA.016	036J	Nakaahara Mochi	Desconhecido	Desconhecido
BGA.017	038J	Nourin Mochi	Desconhecido	Desconhecido
BGA.018	040J	Toukyo Hirayama	Desconhecido	Desconhecido
BGA.019	041J	Iwata Hata Mochi	Desconhecido	Desconhecido
BGA.020	042J	Susono Mochi	Desconhecido	Desconhecido
BGA.021	045J	Mitsukasane	Landrace	Sequeiro
BGA.022	048J	Mie	Landrace	Sequeiro
BGA.023	050J	Wase Esoshima Mochi	Desconhecido	Desconhecido
BGA.024	052J	Mizuhoshi	Landrace	Irrigado
BGA.025	054J	Kyuushuu	Landrace	Sequeiro
BGA.026	055J	Oohata Mochi	Landrace	Sequeiro
BGA.027	056J	Miyako	Landrace	Irrigado
BGA.028	057J	Yoridashi	Landrace	Irrigado
BGA.029	068J	Nourin 24	Desconhecido	Desconhecido
BGA.030	069J	Saitama Senshou	Melhorado	Sequeiro
BGA.031	070J	Kirishima	Landrace	Sequeiro
BGA.032	074J	Aichi Rikutou 1	Melhorado	Sequeiro
BGA.033	078J	Yonoyuki Mochi	Landrace	Sequeiro
BGA.034	079J	Dango Mochi	Landrace	Sequeiro

(continuação)

Ordem	Acesso	Nome do Acesso/Genótipo	Status*	Tipo*
BGA.035	080J	Sangoku	Landrace	Sequeiro
BGA.036	082J	Terenzu	Landrace	Sequeiro
BGA.037	092J	Iwate Kurumi Wase1	Melhorado	Sequeiro
BGA.038	099J	Eika Ine	Landrace	Sequeiro
BGA.039	106J	Araki 2	Landrace	2Sequeiro/3Irrigado
BGA.040	107J	Oiran	Landrace	Sequeiro
BGA.041	110J	Mikuni No Homare	Landrace	Sequeiro
BGA.042	112J	Japão 1	Desconhecido	Desconhecido
BGA.043	113J	Kyuushuu	Landrace	Sequeiro
BGA.044	117J	Gaisen Ibaragi 1	Desconhecido	Desconhecido
BGA.045	118J	Kurombo	Desconhecido	Desconhecido
BGA.046	119J	Shirotsuka Wase	Landrace	Sequeiro
BGA.047	121J	Nourin 5	Desconhecido	Desconhecido
BGA.048	122J	Chiyoda Wase	Landrace	Sequeiro
BGA.049	125J	Touzan Mochi	Desconhecido	Desconhecido
BGA.050	127J	Hakamuri 20	Melhorado	Sequeiro
BGA.051	129J	Esojima Mochi	Landrace	Sequeiro
BGA.052	130J	Esojima	Desconhecido	Desconhecido
BGA.053	131J	Mino	Landrace/Melhorado	Sequeiro/Irrigado
BGA.054	133J	Nourin Mochi	Desconhecido	Desconhecido
BGA.055	134J	Suzume Shirazu	Desconhecido	Desconhecido
BGA.056	136J	Nourin Mochi	Desconhecido	Desconhecido
BGA.057	139J	Nourin Mochi 1	Desconhecido	Desconhecido
BGA.058	140J	Nourin Mochi 17	Desconhecido	Desconhecido
BGA.059	144J	Miyamae Okute	Landrace	Sequeiro
BGA.060	147J	Kukutomi	Desconhecido	Desconhecido
BGA.061	150J	Atoshirazu	Landrace	Sequeiro
BGA.062	153J	Col/Fuzui/1965	Desconhecido	Desconhecido
BGA.063	156J	Tachiminori	Melhorado	Sequeiro
BGA.064	159J	Minami Hata Mochi	Desconhecido	Desconhecido
BGA.065	160J	Gaisen Mochi	Landrace	Sequeiro
BGA.066	164J	No Mochi	Landrace	Sequeiro
BGA.067	165J	Takasago Wase	Landrace	Sequeiro
BGA.068	170J	Tanaka Yakan	Landrace	Sequeiro
BGA.069	171J	Oosumi	Melhorado	Sequeiro
BGA.070	172J	Oohataho	Landrace	Sequeiro
BGA.071	173J	Nourin 11	Desconhecido	Desconhecido
BGA.072	174J	Col/Miyazaki/1963	Desconhecido	Desconhecido
BGA.073	175J	Tomoe Mochi	Landrace	Sequeiro
BGA.074	176J	Shinshuu Wase	Landrace	Sequeiro
BGA.075	177J	Nourin 16	Desconhecido	Desconhecido

(continuação)

Ordem	Acesso	Nome do Acesso/Genótipo	Status*	Tipo*
BGA.076	179J	Urasan 1	Melhorado	Sequeiro
BGA.077	183J	Yashino Mochi	Desconhecido	Desconhecido
BGA.078	186J	Kirishima	Landrace	Sequeiro
BGA.079	188J	Chiba Senshou	Melhorado	Sequeiro
BGA.080	189J	Okabo	Landrace	Sequeiro/Irrigado
BGA.081	191J	Col/Fuzui/1965	Desconhecido	Desconhecido
BGA.082	197J	Toga	Landrace	Sequeiro
BGA.083	199J	Kahee	Landrace	Sequeiro
BGA.084	200J	Hitachi Nishiki	Landrace	Sequeiro
BGA.085	202J	Ooba Kirishima	Landrace	Sequeiro
BGA.086	203J	Horarin	Landrace	Sequeiro
BGA.087	205J	Matsuyama	Landrace	Sequeiro
BGA.088	206J	Taishou Mochi	Landrace	Sequeiro
BGA.089	209J	Kangyouho	Desconhecido	Sequeiro
BGA.090	213J	Japão 2	Desconhecido	Desconhecido
BGA.091	214J	Kaneko Mochi	Landrace	Sequeiro
BGA.092	216J	Iwate Kurumi Wase1	Melhorado	Sequeiro
BGA.093	217J	Col/Miyazaki/1963	Desconhecido	Desconhecido
BGA.094	218J	Toukyo Kaneko	Desconhecido	Desconhecido
BGA.095	219J	Gaisen (4x)	Desconhecido	Desconhecido
BGA.096	221J	Ishikawa	Desconhecido	Sequeiro
BGA.097	228J	Tariu Saku Mochi	Desconhecido	Desconhecido
BGA.098	229J	Col/Tokushima/1967	Desconhecido	Desconhecido
BGA.099	230J	Miyanishiki	Landrace/Melhorado	Sequeiro/Irrigado
BGA.100	235J	Hatamurasaki	Melhorado	Sequeiro
BGA.101	236J	Toga	Landrace	Sequeiro
BGA.102	237J	Rikutou Shinriki 1	Melhorado	Sequeiro
BGA.103	239J	Col/Miyazaki/1963	Desconhecido	Desconhecido
BGA.104	245J	Ishikawa Mochi	Landrace	Sequeiro
BGA.105	246J	Mizugirai Mochi	Landrace	Sequeiro
BGA.106	248J	Col/Tokushima/1967	Desconhecido	Desconhecido
BGA.107	250J	Jouon	Landrace	Sequeiro
BGA.108	252J	Oota Wase	Landrace	Sequeiro
BGA.109	256J	Shizuoka	Desconhecido	Desconhecido
BGA.110	257J	Hiderishirazu	Landrace	Sequeiro/Irrigado
BGA.111	259J	Iwate Ryoon 1	Desconhecido	Desconhecido
BGA.112	260J	Kozo	Landrace	Sequeiro
BGA.113	261J	Kahei	Landrace	Sequeiro
BGA.114	263J	Tamasari 3	Melhorado	Sequeiro
BGA.115	264J	Miyakonojoo Mochi	Landrace	Sequeiro
BGA.116	268J	Ookuma Nishiki	Desconhecido	Desconhecido
BGA.117	269J	Nourin Mochi 6	Desconhecido	Desconhecido

(continuação)

Ordem	Acesso	Nome do Acesso/Genótipo	Status*	Tipo*
BGA.118	271J	Taishou Nishiki	Landrace	Sequeiro
BGA.119	273J	Shina Mochi	Landrace	Sequeiro
BGA.120	274J	Nagae Wase	Melhorado	Sequeiro
BGA.121	276J	Arabiya Mochi	Desconhecido	Desconhecido
BGA.122	277J	Tozo Mochi	Desconhecido	Desconhecido
BGA.123	278J	Urasar	Desconhecido	Desconhecido
BGA.124	279J	Ootama	Landrace	Sequeiro/Irrigado
BGA.125	280J	Okabo Mochi	Landrace	Sequeiro
BGA.126	284J	Furuwase	Desconhecido	Desconhecido
BGA.127	285J	Hirakawa Okute	Landrace	Sequeiro
BGA.128	286J	Nourin 7	Desconhecido	Desconhecido
BGA.129	287J	Oiran	Landrace	Sequeiro
BGA.130	289J	Shinhoku Daiou Mochi	Desconhecido	Desconhecido
BGA.131	290J	Riku Araki	Landrace	Sequeiro
BGA.132	293J	Suzume Shirazu	Desconhecido	Desconhecido
BGA.133	294J	Col/Miyazaki/1963	Desconhecido	Desconhecido
BGA.134	295J	Hideshirazu Mochi	Desconhecido	Desconhecido
BGA.135	299J	Japão 3	Desconhecido	Desconhecido
BGA.136	301J	Nourin Mochi 4	Desconhecido	Desconhecido
BGA.137	302J	Kazusa Wase	Landrace	Sequeiro
BGA.138	304J	Shinkuko Mochi	Desconhecido	Desconhecido
BGA.139	308J	Hikouki Gome	Landrace	Sequeiro
BGA.140	310J	Senshou	Landrace	Sequeiro
BGA.141	311J	Shindai Okoshi	Landrace	Sequeiro
BGA.142	312J	Col/Ooita/1964	Desconhecido	Desconhecido
BGA.143	314J	Taiwan Mochi	Landrace	Sequeiro
BGA.144	315J	Urasan	Landrace	Sequeiro
BGA.145	317J	Owari Mochi	Landrace/Melhorado	Sequeiro
BGA.146	320J	Nourin Mochi 2	Desconhecido	Desconhecido
BGA.147	322J	Senshou	Landrace	Sequeiro
BGA.148	325J	Gaisen Mochi	Melhorado	Sequeiro
BGA.149	330J	Japão 4	Desconhecido	Desconhecido
BGA.150	331J	Rikuu	Desconhecido	Desconhecido
BGA.151	335J	Rikuu 23	Melhorado	Sequeiro
BGA.152	336J	Ohata Wase	Landrace	Sequeiro
BGA.153	337J	Japão 5	Desconhecido	Desconhecido
BGA.154	338J	Japão 6	Desconhecido	Desconhecido
BGA.155	339J	Owari 79	Melhorado	Sequeiro
BGA.156	346J	Ouu 22	Melhorado	Sequeiro/Irrigado
BGA.157	352J	Rikuu 15	Melhorado	Sequeiro/Irrigado
BGA.158	355J	Col/Miyazaki/1963	Desconhecido	Desconhecido

(continuação)

Ordem	Acesso	Nome do Acesso/Genótipo	Status*	Tipo*
BGA.159	356J	Rikuu 13	Melhorado	Sequeiro
BGA.160	357J	Rikuu 22	Melhorado	Sequeiro
BGA.161	364J	Fujimizu Bansei	Landrace	Sequeiro
BGA.162	365J	Iwate Kinsen 1	Melhorado	Sequeiro
BGA.163	367J	Bansei Tarou	Landrace	Sequeiro
BGA.164	368J	Shiro Uzura	Landrace	Sequeiro
BGA.165	369J	Japão 7	Desconhecido	Desconhecido
BGA.166	370J	Japão 8	Desconhecido	Desconhecido
BGA.167	373J	Mogami Uruchi 1	Melhorado	Sequeiro
BGA.168	374J	Gaisen	Landrace	Sequeiro
BGA.169	378J	Okka Modoshi	Landrace	Sequeiro
BGA.170	379J	Mino Senshutsu	Landrace	Sequeiro
BGA.171	380J	Japão 9	Desconhecido	Desconhecido
BGA.172	382J	Kurohige	Landrace	Sequeiro
BGA.173	383J	Mogami Chikanari 1	Landrace	Sequeiro
BGA.174	384J	Kounoso Rikutou 2	Desconhecido	Desconhecido
BGA.175	386J	Minami Hata Mochi	Desconhecido	Desconhecido
BGA.176	387J	Wase Shinshuu	Landrace	Sequeiro
BGA.177	389J	Igisu Mochi	Landrace	Sequeiro
BGA.178	391J	Kurumi Wase 43	Melhorado	Sequeiro
BGA.179	393J	Hiderishirazu	Landrace	Sequeiro/Irrigado
BGA.180	394J	Susono Mochi	Desconhecido	Desconhecido
BGA.181	395J	Seta Gaisen	Landrace	Sequeiro
BGA.182	397J	Korotou Mochi	Landrace	Sequeiro
BGA.183	400J	Owari Hata Mochi	Desconhecido	Desconhecido
BGA.184	401J	Ishiyakushi Mochi	Landrace	Sequeiro
BGA.185	402J	Shiro Hige	Desconhecido	Desconhecido
BGA.186	407J	Edogawa	Landrace	Sequeiro
BGA.187	408J	Gose Yonkoku	Landrace	Sequeiro
BGA.188	409J	Mie	Landrace	Sequeiro
BGA.189	410J	Shizouka	Desconhecido	Desconhecido
BGA.190	413J	Chousen	Landrace	Sequeiro/Irrigado
BGA.191	415J	Tosa Mochi	Landrace	Sequeiro
BGA.192	418J	Aogara	Landrace	Sequeiro/Irrigado
BGA.461	Controle	Primavera	Melhorado	Sequeiro
BGA.462	Controle	BRS Pampa	Melhorado	Irrigado
BGA.463	Controle	AN Cambará	Melhorado	Sequeiro
BGA.464	Controle	Esmeralda	Melhorado	Sequeiro
BGA.465	Controle	Sertaneja	Melhorado	Sequeiro
BGA.466	Controle	Ep109	Melhorado	Irrigado
BGA.467	Controle	Tio Taka	Melhorado	Irrigado
BGA.468	Controle	116 Satoru	Melhorado	Irrigado

(conclusão)				
Ordem	Acesso	Nome do Acesso/Genótipo	Status*	Tipo*
BGA.469	Controle	Andosan	Melhorado	Irrigado
BGA.470	Controle	Marques	Melhorado	Irrigado