



**UNIVERSIDADE FEDERAL DE GOIÁS
ESCOLA DE AGRONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E
MELHORAMENTO DE PLANTAS**

**ANÁLISE COMPARATIVA DOS MÉTODOS DE AVANÇO POR BULK
E SSD NA IDENTIFICAÇÃO DE QTLs PARA PRODUTIVIDADE DE
GRÃO DE ARROZ NO CRUZAMENTO EPAGRI 108 X IRAT 122**

MARIANA RODRIGUES FEITOSA RAMOS

Orientador:
Dr. Claudio Brondani

Goiânia-GO

Brasil

Dezembro-2019

**TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR
VERSÕES ELETRÔNICAS DE TESES E DISSERTAÇÕES
NA BIBLIOTECA DIGITAL DA UFG**

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

1. Identificação do material bibliográfico: Dissertação Tese

2. Identificação da Tese ou Dissertação:

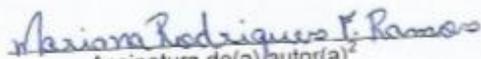
Nome completo do autor: Mariana Rodrigues Feitosa Ramos

Título do trabalho: ANÁLISE COMPARATIVA DOS MÉTODOS DE AVANÇO POR BULK E SSD NA IDENTIFICAÇÃO DE QTLs PARA PRODUTIVIDADE DE GRÃO DE ARROZ NO CRUZAMENTO EPAGRI 108 X IRAT 122

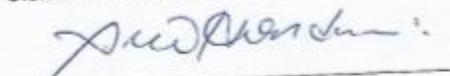
3. Informações de acesso ao documento:

Concorda com a liberação total do documento SIM NÃO¹

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF da tese ou dissertação.


Assinatura do(a) autor(a)²

Ciente e de acordo:


Assinatura do(a) orientador(a)²

Data: 10 / 01 / 2020

¹ Neste caso o documento será embargado por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Os dados do documento não serão disponibilizados durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro;
- Publicação da dissertação/tese em livro.

² A assinatura deve ser escaneada.

MARIANA RODRIGUES FEITOSA RAMOS

**ANÁLISE COMPARATIVA DOS MÉTODOS DE AVANÇO POR BULK E SSD NA
IDENTIFICAÇÃO DE QTLs PARA PRODUTIVIDADE DE GRÃO DE ARROZ NO
CRUZAMENTO EPAGRI 108 X IRAT 122**

Tese apresentada à Coordenação do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, da Universidade Federal de Goiás, como exigência para obtenção do título de doutor em Genética e Melhoramento de Plantas.

Orientador:
Dr. Claudio Brondani

**Goiânia, GO – Brasil.
2019**

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Rodrigues Feitosa Ramos, Mariana

Análise comparativa dos métodos de avanço por Bulk e SSD na identificação de QTLs para produtividade de grão de arroz no cruzamento Epagri 108 X Irat 122 [manuscrito] / Mariana Rodrigues Feitosa Ramos. - 2019.

CXLV, 145 f.: il.

Orientador: Prof. Dr. Claudio Brondani.

Tese (Doutorado) - Universidade Federal de Goiás, Escola de Agronomia (EA), Programa de Pós-graduação em Genética e Melhoramento de Plantas, Goiânia, 2019.

Bibliografia. Apêndice.

Inclui fotografias, gráfico, tabelas.

1. Dialelo. 2. RILs. 3. Marcadores SNPs. 4. Mapeamento por Intervalo. 5. Produtividade de grão. I. Brondani, Claudio, orient. II. Título.

CDU 631/635



UNIVERSIDADE FEDERAL DE GOIÁS

ESCOLA DE AGRONOMIA

ATA DE DEFESA DE TESE

Ata Nº 0062/2019 da sessão de Defesa de Tese de **Mariana Rodrigues Feitosa Ramos** que confere o título de Doutor(a) em **PPG em Genética e Melhoramento de Plantas** na área de concentração em **Genética e Melhoramento de Plantas**.

Ao/s onze dias do mês de dezembro do ano de dois mil e dezenove, a partir da(s) 08:15hrs, no(a) sala da **Embrapa Arroz e Feijão**, realizou-se a sessão pública de Defesa de Tese intitulada **"ANÁLISE COMPARATIVA DOS MÉTODOS DE AVANÇO POR BULK E SSD NA IDENTIFICAÇÃO DE QTLs PARA PRODUTIVIDADE DE GRÃO DE ARROZ NO CRUZAMENTO EPAGRI 108 X IRAT 122"**. Os trabalhos foram instalados pelo(a) Orientador(a), Professor(a) **Dr. Claudio Brondani- Orientador (Presidente) - (Embrapa Arroz e Feijão)** com a participação dos demais membros da Banca Examinadora: Professor(a) **Dra. Rosana Pereira Vianello (Embrapa Arroz e Feijão)**, membro titular externo; Professor(a) **Dra. Tereza Cristina de Oliveira Borba (Embrapa Arroz e Feijão)**, membro titular externo, Profa. **Dra. Gesimária Ribeiro Costa Coelho (Embrapa Arroz e Feijão)**; Professor(a) **Dr. Ivanildo Ramalho –Emater Go**, membro titular externo. Durante a arguição os membros da banca não fizeram sugestão de alteração do título do trabalho [se for o caso inserir: conforme explicitado abaixo]. A Banca Examinadora reuniu-se em sessão secreta a fim de concluir o julgamento da Tese tendo sido(a) o(a) candidato(a) **aprovado(a)** pelos seus membros. Proclamados os resultados pelo(a) Professor(a) **Dr. Claudio Brondani- Orientador (Presidente) - (Embrapa Arroz e Feijão)**, Presidente da Banca Examinadora, foram encerrados os trabalhos e, para constar, lavrou-se a presente ata que é assinada pelos Membros da Banca Examinadora, ao(s) **onze dias do mês de dezembro do ano de dois mil e dezenove**.

TÍTULO SUGERIDO PELA BANCA



Documento assinado eletronicamente por **MARIANA RODRIGUES FEITOSA RAMOS, Discente**, em 11/12/2019, às 13:45, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Claudio Brondani, Usuário Externo**, em 11/12/2019, às 13:47, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Tereza Cristina de Oliveira Borba, Usuário Externo**, em 11/12/2019, às 14:55, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Ivanildo Ramalho do Nascimento Júnior, Usuário Externo**, em 12/12/2019, às 07:02, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **GESIMARIA RIBEIRO COSTA COELHO, Usuário Externo**, em 17/12/2019, às 09:05, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Rosana Pereira Vianello, Usuário Externo**, em 18/12/2019, às 09:57, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1053122** e o código CRC **FB7C52BB**.

AGRADECIMENTOS

Agradeço a Deus em primeiro lugar por ter permitido que tantas pessoas especiais participassem da realização deste trabalho. Meus amados pais que tanto contribuíram para que eu alcançasse os meus objetivos, por suas críticas e elogios.

Aos amigos de hoje e sempre, em especial, a Ariadna Faria, Helma Cronemberger, Daniany Adorno, João Augusto e João Antônio, que contribuíram ricamente no desenvolvimento deste trabalho.

A meu orientador Claudio Brondani pela paciência, sabedoria e amizade.

A todos os funcionários, colegas e amigos da Embrapa Arroz e Feijão pelo empenho e por terem contribuído para a conclusão desta etapa.

Aos professores e funcionários do programa de Pós-Graduação em Genética e Melhoramento de Plantas pelo conhecimento compartilhado e pelo empenho na administração deste curso.

À Capes pela concessão da bolsa de doutorado e pelo financiamento desta pesquisa.

Muito Obrigada!

SUMÁRIO

RESMUMO		8
ABSTRACT		10
1	INTRODUÇÃO GERAL	12
2	REFERENCIAL TEÓRICO	17
2.1	INTRODUÇÃO	17
2.2	IMPORTÂNCIA ECONOMICA DO ARROZ (<i>Oryza sativa</i> L.): SEU CULTIVO, PRODUÇÃO E CONSUMO	19
2.2.1	Classificação Botânica e Origem	20
2.3	O GENOMA DO ARROZ <i>Oryza sativa</i> L.	21
2.4	VARIABILIDADE GENÉTICA DO ARROZ	22
2.5	MELHORAMENTO GENÉTICO DE AUTÓGAMAS	23
2.5.1	Escolha de Genitores para o Melhoramento Genético	23
2.5.2	Métodos de Melhoramento para Autógamas	25
2.6	PARÂMETROS GENÉTICOS POPULACIONAIS NO MELHORAMENTO	28
2.7	PRINCÍPIOS DA METODOLOGIA DE MODELOS MISTOS	31
2.8	MAPEAMENTO DE QTL (<i>Quantitative Trait Loci</i>)	34
2.9	COMPONENTES DE PRODUÇÃO E ALTURA DE PLANTAS	37
2.10	SEQUENCIAMENTO DE NOVA GERAÇÃO (NGS)	39
2.11	GENOTIPAGEM DArT (<i>Diversity Arrays Technology</i>)	40
2.12	REFERÊNCIAS	42
3	ARTIGO CIENTÍFICO: ANÁLISE COMPARATIVA DOS MÉTODOS DE AVANÇO POR BULK E SSD NA IDENTIFICAÇÃO DE QTLs PARA PRODUTIVIDADE DE GRÃO DE ARROZ NO CRUZAMENTO EPAGRI 108 X IRAT 122	49
3.1	INTRODUÇÃO	49
3.2	MATERIAL E MÉTODOS	53
3.2.1	Avanço das Populações	53
3.2.2	Avaliação das Famílias F₇ e F₈	54
3.2.3	Análises Estatísticas Via Modelos Mistos	54
3.2.4	Estimativas dos Parâmetros Genéticos	55
3.2.5	Análises Moleculares	56
3.2.5.1	Extração e Quantificação do DNA Genômico	56
3.2.5.2	Mapa de Ligação e Análise de QTL	56
3.2.5.3	Desequilíbrio de ligação (DL) e blocos haplotípicos	56
3.2.5.4	Análises de componentes principais (PcA) e Estruturação DArWin	57
3.2.5.5	Anotação dos genes	57
3.3	RESULTADOS	57
3.3.1	Análise de Deviance e Significância dos Efeitos do Modelo	57
3.3.2	Estimativa dos Parâmetros Genéticos e Estatísticos	58
3.3.3	Valores Genotípicos das Linhagens (RILs) Bulk e SSD	59

3.3.3.1	Caráter Produtividade de Grãos (PG)	59
3.3.3.2	Caráter Altura de Plantas (ALT)	62
3.3.4	Identificação das RILs com Melhor Desempenho Produtivo	64
3.3.5	Mapeamento Genético	68
3.3.6	Desequilíbrio de Ligação	69
3.3.7	Análise de QTLs das RILs SSD e Bulk	69
3.3.8	Análise de Componentes Principais (PcA) e DArWin	75
3.3.9	Anotação Putativa dos Genes de Blocos Haplotípicos	78
3.4	DISCUSSÃO	83
3.4.1	Dados Fenótipos	83
3.4.2	Dados Genotípicos	90
3.5	CONCLUSÕES	97
3.6	REFERÊNCIAS	97
APÊNDICES	111

RESUMO

RAMOS, M. R. F. **Análise comparativa dos métodos de avanço por Bulk e SSI identificação de QTLs para produtividade de grão de arroz no cruzamento Ep 108 X Irat 122**. 2019. 145 f. Doutorado (Doutorado em Genética e Melhoramento de Plantas) – Escola de Agronomia, Universidade Federal de Goiás, Goiânia, 2019.¹

Um aspecto relevante de todos os programas de melhoramento genético de arroz é a extensa variabilidade genética disponível e armazenada em bancos de germoplasma. Um grande desafio é justamente o modo de como selecionar os genótipos mais adequados para atender os objetivos desses programas. Uma alternativa interessante é a montagem de coleções nucleares. Além da caracterização *per se*, os acessos que se destacaram por sua variabilidade genética ou desempenho produtivo foram cruzados entre si em esquema de dialelo. Os híbridos resultantes foram autofecundados para obtenção da geração F₂, que foi avançada por Bulk e SSD até F₇. Dos cruzamentos mais produtivos, um em particular chamou a atenção, inicialmente pela distância genética entre os genitores (RW= 0,91), e posteriormente, pelo alto valor de capacidade específica de combinação - o Epagri 108 (*Oryza sativa* spp. indica) x Irat 122 (*Oryza sativa* spp. japonica). Esse trabalho objetivou realizar análise de QTLs para produtividade e altura de plantas utilizando duas populações do cruzamento Epagri 108 x Irat 122, avançadas pelos métodos de SSD (geração F₈) e Bulk (geração F_{7,8}). As 158 linhagens (RILs) de cada método (SSD e Bulk) foram avaliadas por dois anos (safras 2016/2017 e 2017/2018), no delineamento látice duplo 18x18 com duas repetições, compostas por parcelas de quatro linhas de três metros, na Fazenda Palmital (Goianira, GO). As RILs foram genotipadas pela metodologia DArTseq®, que gerou cerca de 6 mil SNPs. O modelo estatístico adotado para a análise dos dados de produtividade foi modelo linear misto (MLM) por meio do programa R. Para as avaliações do primeiro e segundo ano (safras 2016/2017 e 2017/2018) e análise conjunta (dois anos/safras), o grupo de RILs-Bulk apresentou maiores médias de produtividade quando comparado ao grupo RILs-SSD e Testemunhas. Porém, quanto ao componente de variância genética, o grupo SSD apresentou a maior estimativa seguido por Bulk e Testemunhas. As produtividades das RILs-Bulk variaram de 4.010,75 kg ha⁻¹ a 5.815,42 kg ha⁻¹, enquanto que as RILs-SSD variaram de 3.321,76 kg ha⁻¹ a 8.096,27 kg ha⁻¹, ambas superando o grupo das Testemunhas, que variaram de 2.754,30 kg ha⁻¹ a 3.643,73 kg ha⁻¹. Para o caráter altura de plantas (ALT), no primeiro ano, as plantas variaram de 116 cm a 165 cm para as RILs- Bulk. Já as RILs-SSD apresentaram variação de 91 cm a 177 cm, enquanto que as Testemunhas variaram de 100 cm a 104 cm. No segundo ano, as RILs-Bulk tiveram variação de 101 cm a 130 cm, enquanto as RILs-SSD variaram de 81 cm a 132 cm, já as Testemunhas apresentaram alturas de 96 cm a 117 cm. Na análise conjunta, as Testemunhas apresentaram as menores alturas. Para a análise de QTLs foi utilizado o mapeamento por intervalo múltiplo, com um total de 2.115 SNPs, e foram identificados 3 QTLs nas RILs-SSD para o caráter produtividade de grãos (PG), dos quais 2 QTLs foram localizados no cromossomo 6 (qGYLD6.1 e qGYLD6.2), um para o segundo ano de avaliação, com variação fenotípica de 23,56%, e o outro para a análise conjunta, explicando 9,45% da variação fenotípica. O outro QTL foi identificado no cromossomo 9 (qGYLD9), para o segundo ano, com variação fenotípica de 7,45%. Para o caráter altura (ALT) foi identificado um QTL no cromossomo 1 (qPHT1), com a variação fenotípica de 14,01%. Para as RILs- Bulk, com total de 2.354 marcadores, 3 QTLs foram identificados para o caráter PG, sendo dois QTLs mapeados nos cromossomos 6 e 9 (qGYLD6 e qGYLD9), referentes ao segundo ano de avaliação, apresentando variação fenotípica de

21,65% e 3,71%, respectivamente. Na análise conjunta um QTL foi mapeado no cromossomo 7 (qGYLD7), com variação fenotípica de 12,9%. Para ALT nenhum QTL foi encontrado nas RILs-Bulk. A partir da identificação desses QTLs nos blocos haplotípicos, a próxima etapa será a validação dos marcadores em acessos do banco de germoplasma da Embrapa antes de serem incorporados a rotina de seleção assistida, com a finalidade de identificar materiais com maior potencial produtivo. Para o cruzamento Epagri 108 x Irat 122 o método SSD foi o mais eficiente na geração de linhagens superiores de arroz para produtividade de grãos, porém a um custo operacional maior em relação ao método Bulk. RILs derivadas tanto de Bulk quanto SSD identificaram QTLs para o caráter PG, entretanto, SSD identificou maior número de QTLs com maior efeito na variação do caráter.

Palavras-chave: Dialelo, RILs, Marcadores SNPs, Mapeamento por Intervalo, produtividade de grão.

¹Orientador: Prof. Dr. Claudio Brondani. Embrapa Arroz e Feijão.

ABSTRACT

RAMOS, M. R. F. **Comparative analysis of Bulk and SSD generation advancement methods in the identification of QTLs for rice grain yield at crossbreeding Epagri 108 X Irat 122**. 2019. 145 f. Doutorado (Doutorado em Genética e Melhoramento de Plantas) – Escola de Agronomia, Universidade Federal de Goiás, Goiânia, 2019.¹

A relevant aspect of all rice breeding programs is the extensive genetic variability available and stored in germplasm banks. A major challenge is precisely how to select the most appropriate genotypes to meet the objectives of these programs. An interesting alternative is the assembly of core collections. Besides the characterization per se, the accessions that stood out for their genetic variability or productive performance were crossed in a diallel scheme. The resulting hybrids were self-fertilized to obtain generation F₂, which was advanced by Bulk and SSD until F₇. Among the most productive crosses, one in particular was interesting due to the genetic distance between the parents (RW = 0.91), and the high value of specific combining ability - Epagri 108 (*Oryza sativa* spp. Indica) x Irat 122 (*Oryza sativa* spp. Japonica). This study aimed to perform QTL analysis for plant yield and height using two populations of Epagri 108 x Irat 122 cross, advanced by SSD (generation F₈) and Bulk (generation F_{7:8}) methods. The 158 recombinant inbred lines of each method (SSD and Bulk) were evaluated for two years (2016/2017 and 2017/2018 seasons), in a 18x18 double lattice design with two replications, consisting of four-line plots of three meters in Palmital Farm (Goianira, GO). The RILs were genotyped by the DArTseq® methodology, which generated about 6,000 SNPs. The statistical model adopted for the grain yield data analysis was mixed linear model (MLM) through the R program. For the first and second year evaluations (2016/2017 and 2017/2018 seasons) and joint analysis (two years/seasons), the RILs-Bulk group presented higher grain yield averages when compared to the RILs-SSD and testers group. However, regarding the genetic variance component, the SSD group presented the highest estimate followed by Bulk and testers. Bulk-RIL yields ranged from 4,010.75 kg ha⁻¹ to 5,815.42 kg ha⁻¹, while SSD-RILs ranged from 3,321.76 kg ha⁻¹ to 8,096.27 kg ha⁻¹, both exceeding the testers group, which ranged from 2,754.30 kg ha⁻¹ to 3,643.73 kg ha⁻¹. For the plant height trait (ALT), in the first year, the plants ranged from 116 cm to 165 cm for RILs-Bulk. On the other hand, RILs-SSD ranged from 91 cm to 177 cm, while the testers ranged from 100 cm to 104 cm. In the second year, RILs-Bulk ranged from 101 cm to 130 cm, while RILs-SSD ranged from 81 cm to 132 cm, while the testers presented heights from 96 cm to 117 cm. In the joint analysis, the testers presented the lowest heights. For QTL analysis, multiple interval mapping was used, with a total of 2,115 SNPs, and 3 QTLs were identified in the SSIL-RILs for the grain yield (PG) trait, of which 2 QTLs were located on chromosome 6 (qGYLD6.1 and qGYLD6.2), one for the second year of experiment, with a phenotypic variation of 23.56%, and the other for the joint analysis, explaining 9.45% of the phenotypic variation. The other QTL was identified on chromosome 9 (qGYLD9) for the second year, with a phenotypic variation of 7.45%. For the trait height (ALT) a QTL on chromosome 1 (qPHT1) was identified, with a phenotypic variation of 14.01%. For RILs-Bulk, with a total of 2,354 markers, 3 QTLs were identified for the PG character, two QTLs mapped on chromosomes 6 and 9 (qGYLD6 and qGYLD9), referring to the second year of evaluation, presenting a phenotypic variation of 21.65. % and 3.71%, respectively. In the joint analysis a QTL was mapped on chromosome 7 (qGYLD7), with phenotypic variation of 12.9%. For ALT no QTL was found in the RILs-Bulk. From the identification

of these QTLs in haplotypic blocks, the next step will be the validation of markers in Embrapa germplasm bank accesses before being incorporated into the assisted selection routine, in order to identify materials with higher grain yield potential. For the Epagri 108 x Irat 122 cross, the SSD method was the most efficient in the generation of superior rice lines for grain yield, but at a higher operating cost than the Bulk method. RILs derived from both Bulk and SSD identified QTLs for the PG character; however, SSD identified a higher number of QTLs with greater effect on trait variation.

Keywords: Diallel, RILs, SNP Markers, Interval Mapping, grain yield.

¹Orientador: Prof. Dr. Claudio Brondani. Embrapa Arroz e Feijão.

1 INTRODUÇÃO GERAL

O arroz (*Oryza sativa* L.) é um dos principais alimentos básicos da maior parte da população mundial, sendo o terceiro cereal mais produzido e cultivado, principalmente na Ásia, cujo continente responde por cerca de 90% da produção mundial. O Brasil é o nono maior produtor mundial, e na safra 2018/2019 sua produção anual foi de 10,7 milhões de toneladas aproximadamente (Conab, 2019).

No cenário mundial, segundo dados gerados pelo Departamento de Agricultura dos Estados Unidos - USDA, a cultura do arroz acumulou a produção, até fevereiro de 2019, de 772,5 milhões de toneladas (Conab, 2019). Considerando o comércio internacional do arroz brasileiro, as exportações são direcionadas, principalmente, para Cuba, Venezuela e Senegal, que importam juntos cerca de 40% desse arroz. No âmbito das importações, Paraguai, Uruguai e Argentina comercializam cerca de 93% do volume total. Isso devido às características similares dos grãos produzidos em tais países (Agrianual, 2019).

No Brasil, a produção de arroz é proveniente de dois sistemas de cultivo: irrigado por inundação e de sequeiro ou terras altas. Na safra 2018, foram produzidos 11,3 milhões de toneladas de arroz irrigado e 1,3 milhões toneladas de arroz de sequeiro (Conab, 2019). A orizicultura irrigada e de sequeiro estão presentes em quase todas as regiões do território brasileiro, contudo, a maior parte da produção está concentrada em cinco estados. Na Região Sul, responsável por mais de 80% da oferta nacional, destaca-se o estado do Rio Grande do Sul, maior produtor do país com cerca de 71% da produção nacional e 8,7 milhões de toneladas por ano; Santa Catarina concentrou 9,2% da produção, com 1,1 milhões de toneladas; Tocantins, com 677 mil toneladas (5,6%); Mato Grosso – 650 mil toneladas (3,9%) e Maranhão – 246 mil toneladas (1,8%). Juntos, estes estados, em 2018, responderam por 93% da produção nacional (Agrianual, 2019; Planeta Arroz, 2019).

O grande número de acessos armazenados em bancos de germoplasma de arroz torna viável a introdução de variabilidade genética adicional às populações de melhoramento. A diversificação proveniente da ampliação da base genética permite o surgimento de novas combinações alélicas e adaptações a ambientes específicos, podendo proporcionar, por exemplo, uma redução da vulnerabilidade a doenças e insetos, e maior estabilidade da produção (Wang et al., 2014). O Brasil possui uma grande coleção de

genótipos de arroz adaptados ao cultivo em condições tropicais. A incorporação de variabilidade genética ao programa de melhoramento do arroz brasileiro pode permitir que o Brasil aumente a produtividade e qualidade de grão desta cultura, atendendo tanto a demanda interna, quanto as exportações. A Embrapa Arroz e Feijão, juntamente com a Embrapa Recursos Genéticos e Biotecnologia, elaborou a Coleção Nuclear de Arroz da Embrapa (CNAE) (Abadie et al., 2005), que reúne 550 acessos, representativos da variabilidade genética de aproximadamente 10.000 acessos de arroz armazenados no Banco Ativo de Germoplasma à época.

Normalmente em um programa de melhoramento genético os cruzamentos são realizados entre genitores com fenótipos favoráveis para características complementares entre si, com o intuito de gerar populações segregantes com variabilidade suficiente para realização de seleções de linhagens superiores para os caracteres de interesse. A decisão mais importante é a escolha criteriosa dos genitores para realizar as hibridações, pois permite que os esforços dos melhoristas concentrem-se naquelas populações segregantes potencialmente capazes de fornecer famílias superiores, traduzindo-se em maior eficiência do programa (Ramalho et al., 2012). Entre outros fatores, essa escolha depende dos caracteres a serem melhorados, do tipo de herança do caráter e da fonte de germoplasma disponível. Se o caráter a ser melhorado for de herança qualitativa, isto é, controlado por poucos genes e pouco influenciado pelo ambiente, a escolha dos genitores é mais fácil. Normalmente é realizada a hibridação de uma cultivar portadora do alelo de interesse com outra que apresente boas características agrônômicas (Abreu, 1997). Entretanto, quando se trata de caracteres quantitativos, como produtividade de grão, Baezinger & Peterson (1991) classificam os métodos de escolha dos genitores em duas categorias: a) os que incluem apenas as informações dos parentais, como, comportamento per se, coeficiente de parentesco e análise multivariada para estimar divergência genética; b) os que utilizam o comportamento de suas progênes, como os cruzamentos dialélicos, a estimativa de $m + a$ (média + valor aditivo).

A análise dialélica é uma metodologia que pode auxiliar na escolha de genitores com base nos seus valores genéticos, principalmente considerando a capacidade de se combinarem em híbridos que produzam populações segregantes para seleção de linhagens promissoras. Por meio desse método é possível conhecer também o controle genético dos caracteres, que orienta na condução das populações segregantes e a seleção. Todavia, a divergência genética está relacionada ao grau de distanciamento entre

populações quanto ao conjunto de caracteres específicos. Sobretudo a divergência genética entre genitores tem sido avaliada com a intenção de identificar prováveis combinações com maiores efeitos heteróticos, de forma que, os descendentes segregantes tenham maiores chances de recuperação de genótipos superiores (Cruz & Ragazzi, 1994; Ramalho et al., 2012).

A formação da população base é uma fase crítica, pois dela dependerá todo o sucesso futuro do programa de melhoramento. A população base deve apresentar média adequada para os caracteres de interesse e suficiente variabilidade genética para possibilitar a seleção. No que se refere à média alta, está implícito, no caso de produtividade de grãos, que na sua formação devem estar presentes cultivares e linhagens adaptadas. A inclusão de material exótico, com pouca adaptação, resultará em redução da média populacional (Ramalho et al., 1999).

Diferentes métodos de melhoramento podem ser utilizados, em espécies autógamas, como o arroz, tendo como objetivo obter linhas homozigóticas superiores às já existentes em cultivo. Nas autógamas, os métodos tradicionais de condução das populações segregantes são: o método genealógico; método do descendente de uma única semente (SSD - *Single Seed Descent*); método da população (Bulk); método massal; seleção recorrente e retrocruzamento (Borém et al., 1997; Ramalho et al., 2012). Para que a geração de populações segregantes seja eficiente, é fundamental, para que seja avaliada a eficiência relativa dos métodos de melhoramento genético para os caracteres de interesse aos diferentes programas, uma vez que a variabilidade é função da diversidade genética dos genitores (Castanheira & Santos, 2004).

O arroz (*Oryza sativa* L.), tem sido utilizado como organismo modelo para estudo de plantas cultivadas. A importância do estudo do genoma do arroz não está relacionada apenas ao seu aspecto econômico e social, mas fundamentalmente à sua relevância em estudos genômicos como planta modelo (Sohrabi et al., 2012). Os cereais evoluíram independentemente de um mesmo ancestral há cerca de 50 a 70 milhões de anos, mas apesar deste longo período de evolução, os genomas apresentam uma alta conservação (Goff et al., 2002). O arroz é o cereal com menor genoma, apresentando em torno de 373 Mpb (Milhões de pares de base), quando comparado aos do sorgo, milho, cevada e trigo, com cerca de 750, 2500, 4800 e 16000 Mpb, respectivamente (Kawahara et al., 2013).

Além da sequência do genoma do arroz existem disponíveis atualmente 1.249.965 ESTs (*Expression Sequence Tags*). Esta disponibilidade de ESTs, oriundos de bibliotecas de cDNA diversas, permitiu a obtenção de informações sobre a expressão de genes importantes para a cultura do arroz. Permitiu também o desenvolvimento de uma gama de marcadores baseados em DNA para a utilização em variados estudos genéticos. Todas estas informações serão necessárias para melhor caracterização funcional do genoma do arroz (Paterson et al., 2010).

O conhecimento dos padrões fenotípicos e genotípicos dos acessos presentes nos bancos de germoplasma é uma ferramenta útil nos programas de melhoramento auxiliando na busca de acessos desejáveis. O detalhamento do perfil gênico de acessos com a utilização de ferramentas moleculares torna-se uma tecnologia importante por permitir a detecção direta de variações no genoma, que promovem variações fenotípicas (Sakai et al., 2013). Assim, possibilita que a seleção de genitores para programas de melhoramento passe a ser baseado na variação no genoma de interesse. O uso de marcadores moleculares tem sido a tecnologia utilizada para a construção de mapas genéticos de vários organismos (Resende et al., 2014).

O mapeamento genético permite identificar regiões cromossômicas responsáveis por alguma variação observada em características quantitativas, bem como determinar a forma de ação dos genes ou grupos de genes associados a essas regiões. Os QTL (*quantitative trait loci*) são locos ou segmentos cromossômicos que governam as características quantitativas, mas essa definição refere-se apenas a uma associação estatística entre uma região do genoma e um caráter fenotípico, que possuem distribuição contínua, tais como, altura e peso de plantas e de animais; produção de grãos; teor de óleo etc. Com o surgimento dos marcadores moleculares tornou-se possível mapear QTLs que afetam esses caracteres quantitativos (Guo et al. 2014). Marcadores genéticos em ligação próxima com QTL são usados para mapeá-los, e, também, para a seleção assistida por marcadores (MAS), por exemplo, em conjunto com informações fenotípicas. A disponibilidade de marcadores moleculares foi aumentada recentemente com o advento dos microssatélites, dos SNPs e dos DArTs e os genes a eles ligados podem ser mapeados em grupos de ligação (Resende, et al., 2014).

O arroz cultivado tem sido adotado como um importante modelo para a pesquisa científica em plantas o que tem ajudado bastante na compreensão dos controles genéticos e moleculares e processos biológicos associados aos caracteres quantitativos (Bai

et al., 2011). A produtividade de grãos é resultante da ação de diversos fatores genéticos e fisiológicos da planta, bem como das interações com o ambiente e as práticas agronômicas adotadas. Por ser um caráter quantitativo, regulado pela interação de vários genes de pequeno efeito e pela ação ambiental sobre o genótipo, a produtividade tende a ser de baixa ou moderada herdabilidade (Xing & Zhang, 2010). Estudos com mapeamento de QTLs e clonagem têm sido importantes na identificação de genes que regulam o peso de grãos, incluídos tanto os genes para o tamanho quanto para o enchimento de grão, por exemplo (Qui et al., 2017; Dixit et al., 2019). Portanto, tem sido crescente a busca pelo entendimento desses caracteres componentes (Bhati et al., 2015).

O objetivo desse trabalho foi comparar os métodos de condução de população segregante SSD e Bulk quanto a capacidade de identificar QTLs para os caracteres produtividade de grãos e altura de plantas das linhagens puras recombinantes (RILs) derivadas do cruzamento Epagri 108 x Irat 122.

2 REFERENCIAL TEÓRICO

2.1 INTRODUÇÃO

O arroz é consumido em todos os continentes e desempenha papel importante na economia e na sociedade. É a terceira maior cultura cerealífera do mundo, ficando atrás apenas da cultura do milho e do trigo. No Brasil o arroz é cultivado em todo o território, e detém posição privilegiada do ponto de vista social e econômico dentre as culturas anuais (Agrianual, 2019).

O melhoramento de plantas foi o grande responsável pelo aumento da produtividade nas espécies cultivadas, e tem permitido obter novos genótipos produtivos e adaptados às condições mais adversas de cultivo. Apesar da extensa variabilidade genética disponível, os programas de melhoramento genético do mundo todo têm priorizado a utilização de um grupo restrito de genitores elite, com a finalidade de serem preservadas as melhores combinações de genes. Estas combinações de genitores elites tem promovido o estreitamento da base genética da cultura do arroz, quanto aos patamares de produtividade e na redução de ganhos genéticos com a seleção (Ramalho et al., 2012).

Com o decorrer do tempo, os aumentos nos ganhos por seleção só são possíveis se a variabilidade genética adicional for introduzida nas populações de melhoramento (Uphoff et al., 2015). A diversificação proveniente da ampliação da base genética permite o surgimento de novas combinações alélicas e adaptações a ambientes específicos, podendo proporcionar, por exemplo, uma redução da vulnerabilidade a doenças e insetos, e maior estabilidade da produção (Huang et al., 2015).

Para o desenvolvimento de novas cultivares de plantas autógamas, podem ser utilizados diferentes métodos de condução de populações de melhoramento, que variam de acordo com as características a serem melhoradas e os genitores utilizados. Métodos tradicionais de condução das populações segregantes até sua fixação são diversos, e suas descrições e implicações relacionadas à informação obtida sobre as linhagens, alterações nas frequências gênicas, tamanho populacional adequado, inferência ou não da seleção natural, dentre outras são discutidos por Borém et al. (1997) e Ramalho et al. (2012). Todavia, é importante que seja avaliada a eficiência relativa dos métodos de melhoramento genético para os caracteres de interesse para os programas, uma vez que a variabilidade é função da diversidade genética dos genitores e, também, dos métodos de condução das populações segregantes (Castanheira & Santos, 2004).

Dos tipos de populações que são geradas pelos programas de melhoramento, as linhas puras recombinantes (RILs), que detêm elevado grau de homozigose são utilizadas para estudos de genética e genômica bem como para desenvolver novas cultivares. A população passa por processos de caracterização e avaliação a campo, estimação de média e parâmetros genéticos, avaliação da interação com ambiente e também análise de diversidade genética por marcadores moleculares (Ramalho et al., 2012).

Para a estimação de parâmetros genéticos, a metodologia de modelos mistos tem se mostrado eficiente, uma vez que tem a capacidade de reduzir significativamente os erros associados de natureza aleatória. O método REML/BLUP é comumente utilizado para estimar os componentes de variância e prever as médias genótípicas. A estimativa da interação genótipo por ambiente pode ser realizada por vários métodos disponíveis, sendo que a escolha do método depende do objetivo de cada estudo. Aqueles baseados nos valores genóticos têm se mostrado eficientes na seleção de linhagens mais adaptadas e mais estáveis. A média harmônica da performance relativa dos valores genóticos (MHPRVG) é um exemplo e, tem a capacidade de classificar um genótipo pela produtividade, adaptabilidade e estabilidade, simultaneamente (Resende et al., 2014).

O uso de marcadores moleculares tem sido a tecnologia utilizada para a construção de mapas genéticos de vários organismos (Ma et al., 2016). A construção de mapas genéticos cobrindo todo o genoma permite chegar cada vez mais rápido a genes específicos, e em todo o conhecimento gênico de uma espécie. Mapas de ligação cobrindo todo o genoma do arroz foram desenvolvidos ao longo dos anos (McCouch et al., 1997), fornecendo novas oportunidades para a identificação de regiões genômica relacionadas à caracteres de importância agrônômica.

A importância de estudos visando o melhoramento da cultura do arroz, não está voltada simplesmente por ser considerado alimento básico, mas também por ser considerado uma planta modelo para estudos genômicos, apresentando o menor genoma em comparação aos demais cereais, em torno de 373 Mpb (Milhões de pares de base) (Kawahara et al., 2013).

Nas últimas décadas a automatização do sequenciamento permitiu um aumento significativo na rapidez dos processos para a obtenção de dados, elucidando genomas de diversas espécies. As plataformas denominadas NGS (Sequenciamento de Nova Geração- *Next Generation Sequencing*) foram disponibilizadas para o mercado no ano de 2005. As técnicas de nova geração de sequenciamento utilizam a construção de bibliotecas de DNA,

dispensando o uso da clonagem via vetores, ocasionando ampla utilização dessas técnicas para estudos genéticos (Shendure & Ji, 2008).

Com a automatização das plataformas para o sequenciamento genômico utilizando o NGS, novas técnicas de genotipagem foram desenvolvidas para a busca de marcadores SNPs através da diminuição da complexidade genômica pela análise de regiões-alvo. Uma dessas técnicas é a genotipagem por sequenciamento é a variante da tecnologia DArT que foi desenvolvida baseada no mesmo princípio de redução da complexidade do genoma aliado a tecnologia de genotipagem por sequenciamento de última geração, o DArT-seq (Kilian, 2005).

2.2 IMPORTÂNCIA ECONOMICA DO ARROZ (*Oryza sativa* L.): SEU CULTIVO, PRODUÇÃO E CONSUMO

O arroz é tradicionalmente um dos produtos alimentícios mais consumidos no país, sendo o seu consumo anual estimado em torno de 12 milhões de toneladas. Adicionalmente, a cadeia produtiva do arroz apresenta destaque na criação de trabalho e renda para a economia interna, com um parque industrial nacional de beneficiamento altamente desenvolvido (Conab, 2019).

O arroz (*Oryza sativa* L.) é cultivado em cinco continentes, tanto em regiões tropicais como temperadas, e aproximadamente 90% da produção e do consumo mundial estão concentrados na Ásia. Os principais países produtores da cultura são China, Índia e Indonésia (Agrianual, 2019). Entre os produtos destinados à alimentação humana, o arroz é o segundo em importância, ficando atrás apenas do trigo, e, em algumas partes do mundo, especialmente na Ásia, é base da alimentação de sua população.

No mercado internacional, segundo dados do *United States Department of Agriculture* (USDA), a produção mundial de arroz base beneficiado ficará em 487,8 milhões de toneladas na Safra 2018/19, retração de aproximadamente 800 mil toneladas em relação à safra 2017/18. Esse resultado é reflexo principalmente da expectativa de retração na produção chinesa em função de uma redução de área, apesar da retomada da produção norte-americana (Conab, 2019).

Nos EUA, o incremento de produção foi resultado principalmente de uma expansão de área, apesar de ser identificado também um pequeno aumento da produtividade. Sobre o consumo mundial este está estimado em 487,9 milhões de

toneladas, o que representa crescimento de 6,4 milhões de toneladas (1,33%). É importante destacar que é a primeira vez nos últimos anos que o consumo mundial será superior a produção mundial (Conab, 2019).

Na Safra 2017/18, que está sendo comercializada atualmente, apesar do cenário de aumento da oferta e dos estoques de passagem mundiais, espera-se que o fluxo no mercado internacional continue elevado, fato que colabora com a manutenção da demanda do produto de importantes países produtores, como a Tailândia e a Índia. A China (principal país produtor, consumidor e importador) mantém a tendência dos últimos anos de aumento dos estoques de passagem, por meio, da aquisição de produto de terceiros países (Conab, 2019). A China teve estimativa de 495.066 mil toneladas de produção de arroz na safra 2017/2018 e o consumo de 142.487 mil toneladas, enquanto na safra 2018/2019 a estimativa de produção foi de 491.139 mil toneladas e o consumo de 142.782.

No Brasil, a Safra 2015/16 foi também significativamente influenciada pelo fenômeno El Niño, sendo o Rio Grande do Sul (RS), principal estado produtor, o mais afetado. O excesso de chuva prejudicou extensas áreas de cultivo e reduziu a produtividade média brasileira, que refletiu na produção de 10,6 milhões de toneladas base casca no Brasil (Agrianual, 2019).

2.2.1 Classificação Botânica e Origem

O arroz pertence à divisão Angiosperma, classe das magnoliopsea (monocotiledôneas), ordem *Glumiflorae*, família *Poaceae* (gramíneas), subfamília *Oryzoideae* e gênero *Oryza* (Guimarães, 2014). O gênero *Oryza* possui duas espécies cultivadas: *O. glaberrima Steud.* (arroz de origem africana), e *O. sativa L.* (arroz de origem asiática), que é a mais conhecida por sua importância na alimentação humana, e amplamente cultivado em regiões tropicais e temperadas do mundo (Rangel, 1999).

O. sativa L. é uma espécie autógama diploide, com flores hermafroditas e possuindo um número de cromossomo $2n=24$. O ancestral silvestre *O. rufipogon*, tem distribuição ampla compreendida entre a Ásia subtropical e tropical, com formas anuais, perenes e intermediárias. Não há consenso sobre o local de origem do arroz cultivado, e acredita-se que tenha sido na Índia, China ou regiões do sudeste da Ásia (Guimarães et al., 2006).

A evolução da espécie *O. sativa* resultou na formação de duas subespécies, indica e japônica. O arroz Indica adaptou-se às regiões do Himalaia ao norte da Índia, Afeganistão e Irã. A subespécie Japônica foi domesticada no sudeste da Ásia. A partir dessa região o arroz foi introduzido para Filipinas, Indonésia, Malásia e Taiwan. Posteriormente os portugueses introduziram o arroz no Continente Europeu e Americano. No Brasil as variedades do sistema de cultivo irrigado são consideradas da subespécie Indica, enquanto que as variedades de sequeiro são Japônicas (Khush, 1997).

2.3 O GENOMA DO ARROZ

O arroz (*Oryza sativa* L.) tem sido utilizado como organismo modelo para estudo de plantas cultivadas. A importância do estudo do genoma do arroz não está relacionada apenas ao seu aspecto econômico e social, mas fundamentalmente à sua relevância em estudos genômicos como planta modelo (Sohrabi et al., 2012). Os cereais evoluíram independentemente de um mesmo ancestral à cerca de 50 a 70 milhões de anos, mas apesar deste longo período de evolução, os genomas apresentam uma alta conservação (Goff et al., 2002). O arroz é o cereal com menor genoma, apresentando em torno de 337 Mpb (Milhões de pares de base), quando comparado aos do sorgo, milho, cevada e trigo, com cerca de 750, 2500, 4800 e 16000 Mbp, respectivamente (Kawahara et al., 2013).

Dos 37.544 genes identificados no genoma do arroz, 2.859 não foram encontrados quando comparados ao genoma de *Arabidopsis*, sugerindo serem únicos de arroz e possivelmente outros cereais. Ainda nesse trabalho, mais de 80.000 locos polimórficos foram encontrados entre as subespécies indica e japônica, dentre os quais, os marcadores SNPs, que ocorrem na frequência entre 0,53 e 0,78% no genoma (Matsumoto et al., 2005). Além da sequência do genoma do arroz existem disponíveis atualmente 1.249.965 ESTs (*Expression Sequence Tags*), e clones de cDNA completos. Esta disponibilidade de ESTs, oriundos de bibliotecas de cDNA diversas, permitiu a obtenção de informações sobre a expressão de genes importantes para a cultura do arroz. Permitiu também o desenvolvimento de uma gama de marcadores baseados em DNA para a utilização em variados estudos genéticos. Todas estas informações serão necessárias para melhor caracterização funcional do genoma do arroz (Paterson et al., 2010).

Outras espécies e subespécies de arroz foram sequenciadas, como *Oryza glaberrima* (Wang et al., 2014), *Oryza sativa* spp. *aus* (Sakai et al., 2013) e espécies

silvestres. Devido à baixa divergência genética entre os genitores utilizados na maioria dos programas de melhoramento genético do arroz, os genes de caracteres de interesse presentes no material exótico e que não foram selecionados ao longo do processo de domesticação da espécie *Oryza sativa* L. podem representar uma alternativa importante no desenvolvimento de linhagens e cultivares superiores. Esses materiais podem contribuir com genes que aumentem a estabilidade da produção de grão, e isso inclui caracteres que conferem tolerância a estresses abióticos, resistência a pragas e doenças, e incremento da produtividade (Wang et al., 2014).

2.4 VARIABILIDADE GENÉTICA DO ARROZ

Por meio do processo evolutivo e da domesticação a qual foi submetida *O. sativa*, surgiram inúmeras variedades geneticamente divergentes, as quais foram se adaptando às mais distintas condições de cultivo. Durante a Revolução Verde na década de 1960, surgiu o conceito de variedade moderna de arroz que possuía como principais características o porte baixo, alto perfilhamento, responsiva à adubação e com alto potencial produtivo. A primeira cultivar considerada moderna foi a IR-8, lançada pelo IRRI (*International Rice Research Institute*), semi-anã, altamente produtiva, e utilizada como genitora em programas de melhoramento do mundo todo (Vieira, 2007).

A introdução das cultivares de porte baixo nos plantios das várzeas irrigadas, em substituição às variedades tradicionais, produziu um aumento na produtividade do arroz ao longo das décadas de 1970 e 1980. A partir desse período, os ganhos em produtividade têm sido pequenos e, de um modo geral, resultantes mais do aumento do uso de insumos e de tecnologia, do que do crescimento do potencial produtivo das cultivares (Vieira, 2007). Uma das principais causas apontadas para a estagnação dos patamares de produtividade é a estreita base genética existente nas cultivares comerciais, resultante do melhoramento intensivo das variedades modernas, através da utilização de um grupo restrito de genitores elite, com a finalidade de serem preservadas as melhores combinações de genes.

Considerando as cultivares mais plantadas nos principais estados produtores de arroz irrigado do Brasil, sete ancestrais foram os responsáveis por 70% da composição gênica da cultura (Costa et al., 2006). No Rio Grande do Sul, que é o maior produtor de arroz irrigado, seis ancestrais contribuem com 86% dos alelos das variedades mais plantadas. O aumento da base genética das novas cultivares, provenientes da utilização de

parentais mais divergentes geneticamente, é uma das principais alternativas para quebrar o atual patamar de produtividade do arroz (Brondani et al., 2006). O grande número de acessos armazenados em bancos de germoplasma de arroz torna viável a introdução de variabilidade genética adicional às populações de melhoramento. A diversificação proveniente da ampliação da base genética permite o surgimento de novas combinações alélicas e adaptações a ambientes específicos, podendo proporcionar, por exemplo, uma redução da vulnerabilidade a doenças e insetos, e maior estabilidade da produção.

2.5 MELHORAMENTO GENÉTICO DE AUTÓGAMAS

2.5.1 Escolha de Genitores para o Melhoramento Genético

Os programas de melhoramento são dependentes da variabilidade genética presente no grupo de genitores elite para poder aumentar as chances de obtenção de combinações alélicas favoráveis. A variabilidade genética reduzida resulta em ganhos genéticos pequenos, diminuindo a eficiência no desenvolvimento de linhagens e cultivares superiores em um programa de melhoramento. Portanto, quanto maior o grau de divergência entre os genitores de um programa, maiores serão as probabilidades de reagrupar alelos em novas combinações que sejam favoráveis (Nass et al., 2001).

Estudos sobre a divergência genética entre indivíduos ou populações para os programas de melhoramento que visam hibridações, permitem a obtenção de informações para a identificação de genitores que possibilitem efeitos heteróticos maiores nas progênes, e obter genótipos superiores em gerações segregantes. A divergência genética deveria ser critério decisivo nas seleções de genitores, pois genitores contrastantes aumentam as possibilidades de surgirem combinações híbridas superiores (Nass et al., 2001; Barili et al., 2011).

Normalmente em um programa de melhoramento genético os cruzamentos são realizados entre genitores com fenótipos favoráveis para características complementares entre si, com o intuito de gerar populações segregantes com variabilidade suficiente para realização de seleções de linhagens superiores para os caracteres de interesse. A decisão mais importante é a escolha criteriosa dos genitores para realizar as hibridações, pois permite que os esforços dos melhoristas concentrem-se naquelas populações segregantes potencialmente capazes de fornecer famílias superiores, traduzindo-se em maior eficiência

do programa. Entre outros fatores, essa escolha depende dos caracteres a serem melhorados, do tipo de herança do caráter e da fonte de germoplasma disponível.

A formação da população base é uma fase crítica, pois dela dependerá todo o sucesso futuro do programa de melhoramento. A população base deve apresentar média adequada para o caractere de interesse e suficiente variabilidade genética para possibilitar a seleção. No que se refere à média alta, está implícito, no caso de produtividade de grãos, que na sua formação devem estar presentes cultivares e linhagens adaptadas (Ramalho et al., 2012).

A decisão sobre o número de parentais envolvidos na formação da população base é outro questionamento importante. Ramalho et al. (1999) citam que se o número de parentais for muito grande, a probabilidade de encontrar todos os genitores com boa expressão para o caráter é muito pequena e mesmo que isso fosse possível, a contribuição dos alelos de cada parental seria tão pequena que a maioria seria perdida após os primeiros ciclos seletivos. Por outro lado, se o número for muito pequeno, a chance de associar a maioria dos alelos favoráveis para determinado caráter é também pequena. Outro ponto que surge é como inter cruzar os parentais para formar a população base. Em princípio, os autores citam que 10 a 20 parentais é um número satisfatório.

Na avaliação dos indivíduos ou famílias, a seleção na população base pode ser fenotípica massal ou através de teste de famílias. A primeira opção é aconselhável quando o caráter apresenta herdabilidade alta e pode ser selecionado visualmente com eficiência. Para os caracteres com menor herdabilidade, a seleção deve ser efetuada a partir de gerações avançadas em experimentos com repetição. Em qualquer um dos casos, é considerada a fase mais importante, visto que, se as melhores famílias ou indivíduos forem identificados, certamente a recombinação contribuirá para aumentar a frequência de alelos favoráveis na população (Ramalho, 1997). Com relação ao número de famílias a serem avaliadas, Ferreira (1998), trabalhando com diferentes populações de feijoeiro, concluiu que devem ser usadas, no mínimo, 100 a 200 famílias para representar a variabilidade das populações.

O uso de um número adequado de indivíduos que irão formar a população melhorada é de grande importância no melhoramento populacional para evitar a diminuição da variabilidade genética. Assim, o melhorista deve levar em consideração não só a seleção de indivíduos superiores baseando-se nos objetivos do seu programa, como também, deve evitar a perda de alelos que contribuem positivamente para um melhor

comportamento geral da população. Para a seleção de um determinado caráter, deve-se utilizar uma amostra da população cujo tamanho efetivo seja suficientemente grande para garantir os progressos na direção que se deseja, como também para assegurar a presença de alelos favoráveis para todos os demais caracteres de interesse (Rangel et al., 1999).

2.5.2 Métodos de Melhoramento para Autógamas

Nas espécies autógamas, uma cultivar é representada por um genótipo homozigótico (linhagem), ou uma mistura de linhagens fenotipicamente semelhantes. Essa linhagem ou linhagens podem ter sua constituição genotípica mantida indefinidamente, a menos que ocorra mistura mecânica ou algum cruzamento com outras cultivares (Nass et al., 2001).

Segundo Nass et al. (2001), existem várias opções de métodos de melhoramento aplicáveis às espécies autógamas. Contudo, eles podem ser incluídos em três categorias: introdução de linhagens, seleção de linhas puras e hibridação.

A introdução de linhagens é considerada um método de melhoramento, pois contribui efetivamente para a melhoria do potencial genético em uma dada região. A introdução pode ser visualizada sob dois enfoques; o primeiro é a introdução de germoplasma para ser utilizado como fonte de variabilidade em hibridações. O segundo enfoque da introdução de linhagens refere-se ao uso direto em uma dada região. Nesse caso, as linhagens são introduzidas e submetidas a avaliações em experimentos com repetições em vários locais representativos da região, antes de sua recomendação (Nass et al., 2001).

Em algumas espécies autógamas, especialmente aquelas utilizadas por agricultores de subsistência, tais como arroz e feijão, os agricultores não possuem o hábito de adquirir sementes anualmente, e reutilizam os grãos colhidos como sementes por várias gerações. Essa variabilidade é decorrente de mistura mecânica de linhagens diferentes, seguida muitas vezes de fecundação cruzada entre elas no campo, e da ocorrência de mutação. Assim, é esperado que o germoplasma usado por vários anos pelos agricultores sejam uma mistura de linhas puras. Considerando o número de gerações de cultivo sucessivo e a área semeada anualmente, o número de linhagens genotipicamente diferente dentro da população é enorme. Além do mais, com a atuação da seleção natural, é esperado que persistam apenas aquelas combinações gênicas que sejam mais adaptadas, ou seja,

mais favoráveis às condições de cultivo prevalentes na região. Do exposto, é esperado para essas espécies, que a seleção dentro do germoplasma em uso pelos agricultores, seja um método eficiente. O método é denominado seleção de linhas puras pelas razões já expostas (Ramalho et al., 2012).

A hibridação deve ser utilizada quando o objetivo do programa de melhoramento é reunir em um único indivíduo – linhagem – os alelos desejáveis que se encontram em linhagens distintas. Para a aplicação desse método existem três etapas fundamentais: a escolha de genitores a serem cruzados; a obtenção da população segregante; e a escolha do método de condução da população segregante (Nass et al., 2001).

A primeira etapa refere-se à escolha de genitores, uma vez que em qualquer espécie há infinito número de combinações possíveis. No entanto, para ter sucesso na escolha dos genitores de um programa de melhoramento, os objetivos devem ser claros, pois a decisão depende dos caracteres a serem melhorados, do tipo de controle genético dos caracteres e da fonte de germoplasma disponível (Nass et al., 2001).

Se o caráter a ser melhorado for de herança qualitativa, isto é controlada por poucos genes e pouco influenciada pelo ambiente, a decisão sobre os genitores é mais fácil. Nesse caso, normalmente é realizada a hibridação de uma linhagem portadora do alelo de interesse com outra que apresente boas características agronômicas. Em se tratando de caracteres quantitativos, controlado por muitos genes e muito influenciado pelo ambiente, a escolha dos genitores já não é tão fácil. Esses devem ser tais que possibilitem a obtenção de populações segregantes com média alta, associada a grande variabilidade para o caráter sob seleção.

Os métodos de escolha de genitores, visando ao melhoramento de caráter quantitativo, são classificados de duas formas: a primeira delas envolvendo os procedimentos que utilizam apenas informações dos pais; e a segunda utilizando o desempenho de suas progênies. Das alternativas para a escolha de genitores utilizando seu próprio desempenho, o mais empregado é a média do caráter em questão e a divergência entre os genitores. Dos procedimentos para escolha de pais visando o desempenho das progênies, o cruzamento em dialelo tem sido o mais utilizado (Nass et al., 2001).

Após a escolha dos genitores e realizando os cruzamentos para a obtenção da população segregante, lembrando que é essencial que a população tenha alelos favoráveis no maior número de locos e variabilidade genética para ser explorada. Assim, a população

formada deve ser conduzida por um método de melhoramento. Dos métodos de condução da população segregante estão os métodos que não separam a fase de endogamia e seleção como: Método Massal; Método Genealógico ou Pedigree (nos quais a partir da geração F_2 , quando já existe variabilidade, inicia-se a seleção que continua no decorrer das sucessivas gerações de endogamia). Já os métodos que separam a fase de endogamia e seleção são: Métodos da População (Bulk); Método Descendência de uma única Semente (SSD); Método Bulk dentro de famílias; Retrocruzamentos; Seleção Recorrente. Alguns dos métodos serão descritos abaixo como forma de exemplificar (Ramalho et al., 2012).

Método Massal - a seleção fenotípica de indivíduos superiores é efetuada a partir da geração F_2 e continua nas sucessivas gerações (Ramalho et al., 2012).

Método Genealógico ou Pedigree – o princípio do método é a seleção individual de plantas na população segregante, bem como a avaliação de cada progênie separadamente. O mérito dos indivíduos selecionados é avaliado pelo teste de progênies. Dessa forma, a seleção é praticada com base no genótipo dos indivíduos. A seleção de plantas individuais a partir da geração F_2 , as quais são colhidas isoladas e semeadas em linhas na geração F_3 , quando então é realizada a seleção das melhores famílias (melhores linhas) e dos melhores indivíduos dentro destas famílias (melhores indivíduos dentro das linhas). O processo se repete até que a maioria dos locos esteja em homozigose, quando as melhores linhagens identificadas irão participar de experimentos regionais de competição de cultivares (Ramalho et al., 2012).

Método da População (Bulk) – a partir da geração F_2 , as plantas são colhidas em conjunto, massalmente, e tomada uma amostra de sementes para a obtenção da população F_3 . O processo se repete por algumas gerações, quando o bulk é “aberto”, ou seja, são colhidas plantas individuais, que darão origem às famílias para serem avaliadas em experimentos com repetição, até serem identificadas as melhores linhas puras. Essas deverão, então, comprovar a sua superioridade nos experimentos regionais de competição de linhagens. Na utilização desse método há alguns questionamentos, o primeiro é quanto ao número de indivíduos que deve estar presente em cada geração. Na prática tem sido utilizado de 1000 a 2000. Esse aspecto deve ser considerado devido à perda de alelos favoráveis por amostragem, problema este que pode ser contornado utilizando um número adequado de indivíduos a cada geração. O segundo questionamento é quanto ao momento da abertura do bulk, já que o método tem como princípio avançar a população sem nenhuma seleção artificial, até que a maioria dos locos esteja em homozigose, para só

então iniciar o processo de seleção. A frequência de heterozigotos reduz-se a metade a cada geração e a homozigose completa irá sendo atingida rapidamente. Na geração F_5 , por exemplo, é esperado que 93,75% de todos os locos segregantes já estejam em homozigose. A partir de então o acréscimo na proporção de homozigotos é menor e não há vantagem em se protelar mais a abertura do bulk (Ramalho et al., 2012).

Método Descendência de uma única Semente (SSD) – consiste em avançar as gerações segregantes, tomando uma única semente de cada indivíduo, já a partir da geração F_2 , para obter a geração seguinte. Tal procedimento é repetido até que o nível de homozigose desejado seja atingido. Dessa forma, cada linhagem corresponde a uma planta F_2 diferente e, portanto, reduz-se a perda devido à amostragem deficiente. Entre as vantagens do método está à redução do espaço necessário para o avanço das gerações. Uma segunda vantagem é que não há perda por amostragem dos indivíduos F_2 , haja vista que todos eles estarão representados em todas as gerações. Além do mais, não ocorre à ação da seleção natural, que poderia eliminar alguns fenótipos desejáveis. Uma desvantagem é a perda de variabilidade dentro das famílias, uma vez que, a variabilidade liberada dentro das famílias, com o aumento da endogamia, é equivalente a existente entre plantas na geração F_2 . Para reduzir esta perda é necessário partir de uma população F_2 a maior possível. Outra desvantagem é que a perda de um indivíduo perde-se toda a representatividade da planta F_2 (Ramalho et al., 2012).

2.6 PARÂMETROS GENÉTICOS POPULACIONAIS NO MELHORAMENTO

A utilização do termo parâmetro genético populacional refere-se as características de uma população, em particular a média e variâncias. As estimativas dessas constantes é um dos objetivos principais da inferência estatística, que por sua vez, é conseguida por meio de amostras da população que fornecem as estimativas dos parâmetros. As populações utilizadas em programas de melhoramento apresentam parâmetros de interesse classificados em natureza genética e não genética. A estimação dos parâmetros genéticos é necessário para obter informações sobre a natureza da ação dos genes envolvidos na herança dos caracteres sob investigação; estabelecer a base para a escolha dos métodos de melhoramento á população. Portanto, quando há a pretensão em estimar, experimentalmente, as variâncias genética, tanto o ambiente quanto os genótipos devem constituir amostras apropriadas da população (Gardner, 1963).

A teoria da genética quantitativa apresenta os conceitos de média e variância de populações, os quais são baseados nas frequências e efeitos de genes, isto é, na estrutura genética da população em estudo, que, também, depende de vários outros fatores, como o nível de ploidia, ligação gênica, sistema de acasalamento, dentre outros. Esta teoria foi fundamentada, nos estudos realizados independentemente por Hardy (1908) e Weinberg (1908), do princípio de que as frequências gênicas e genotípicas de uma população de acasalamento ao acaso se mantêm constantes de geração em geração, desde que a população seja suficientemente grande para evitar oscilação genética e que não haja migração, seleção e mutação. Assim, existindo essas condições de equilíbrio, a frequência genotípica corresponde ao quadrado das frequências gênicas (Wang et al., 2014).

As questões primárias, para os caracteres métricos, são formuladas em termos de variância, tendo como ideia básica a variação sendo parcelada em componentes atribuíveis a diferentes causas (Falconer, 1987). A partição da variância genotípica foi realizada por Fisher (1918), mostrando que ocorre em três componentes: a variância genética aditiva, devido aos efeitos médios dos genes; a variância dominante, devido as interações intra-alélicas; e a variação epistática, devido as interações inter-alélicas entre pares de locos. Fisher além de propor o fracionamento da variância genotípica, estabeleceu também cálculos de correlações entre parentes, enfatizando que a correlação entre pai e filho e entre irmãos completos diferem em relação a dominância (Falconer, 1987).

Segundo Falconer (1987), o estudo da variação se fundamenta no desmembramento em componentes atribuídos a diferentes causas. A variância total é denominada de variância fenotípica e inclui a variância genotípica e a variância causada pelo ambiente, admitindo-se que a covariância entre o genótipo e o ambiente seja nula.

A variância promovida pelos desvios atribuídos ao ambiente não pode ser removida, pois contém, por definição, toda a variância não genética, e grande parte desta está fora de controle experimental. Contudo, com a utilização de linhagens altamente endogâmicas ou a F_1 do cruzamento entre duas linhagens dessa natureza, é possível obter genótipos idênticos, eliminando-se, deste modo, a variância genotípica dentro de linhas e maximizando-a entre linhas. Se um grupo de tais indivíduos é cultivado sob uma amplitude normal de ambiente, a sua variância fenotípica fornece uma estimativa da variância ambiental. Ao se subtrair a variância fenotípica de uma população geneticamente heterogênea, por exemplo uma F_2 , da variância ambiental proveniente de uma população F_1 , obtém-se uma estimativa da variância genotípica da população (Falconer, 1987).

A variância genotípica manifesta a diversidade entre indivíduos de uma população. Se os genes expressam efeitos aditivos, o valor genotípico do caráter em questão pode ser aumentado ou reduzido pela substituição de um alelo. Assim, alguns genes apresentam o efeito dominante de forma que a presença de um alelo resulta no incremento do valor genotípico. Contudo, pode ocorrer também interação entre diferentes genes, como no caso de epistasia (Cruz & Carneiro, 2003).

A herdabilidade pode ser determinada como a proporção herdável da variabilidade total, isto é, a razão entre variância genotípica e variância fenotípica. Assim, seu papel preditivo do valor genético por meio do valor fenotípico é muito expressivo, pois somente o valor fenotípico de um indivíduo pode ser mensurado, mas é o valor genético que influenciará a próxima geração. Assim, a herdabilidade quantifica a confiabilidade do valor fenotípico como guia para o valor genético (Falconer, 1987).

A herdabilidade é estimada por meio de uma análise de variância, sendo possível estimar tanto a herdabilidade no sentido amplo, que pode ser definida como a razão da variância genotípica pela variância fenotípica; quanto a no sentido restrito, que é a razão da variância aditiva pela variância fenotípica. A herdabilidade no sentido amplo considera toda a variância genotípica; sendo importante no caso de espécies de plantas que são propagadas assexuadamente, já que o genótipo é herdado integralmente. A herdabilidade no sentido restrito considera apenas o componente de variância aditiva que, na maioria dos casos, tem maior utilidade para o melhoramento, pois quantifica a importância da proporção genética que pode ser transmitida para a próxima geração (Ramalho et al., 2000).

A estimativa da herdabilidade é de grande utilidade para um programa de melhoramento de plantas, seja na escolha dos métodos de melhoramento apropriados, dos locais para a condução dos testes de rendimento e do número de repetições, como na predição do ganho antes de se realizar a seleção, servindo inclusive de apoio para a definição da estratégia de seleção. Todavia, muitos dos procedimentos a serem adotados no melhoramento estão condicionados à magnitude da herdabilidade (Borém, 1997).

A herdabilidade não é uma propriedade imutável de uma característica, portanto, depende da população na qual é estimada e do ambiente onde a população se encontra. Todavia, sempre que um valor para a estimativa da herdabilidade de um dado caráter é mencionado, ele se refere a uma população particular, sob condições particulares Segundo (Falconer, 1987).

Guimarães et al. (2006), avaliaram o potencial de progênies $F_{2:4}$ de arroz irrigado por meio das estimativas de seus parâmetros genéticos e fenotípicos. Foram avaliadas 163 progênies $F_{2:4}$ de arroz irrigado e Testemunhas, para os caracteres produção de grãos, dias até o florescimento (DF) e altura de planta. Os caracteres AP, PG e DF apresentaram altas estimativas de herdabilidade, podendo obter maiores progressos por seleção. Com a seleção praticada apenas para PG, houve ganho genético para AP e DF. Apesar da baixa magnitude do ganho indireto, que foi considerado satisfatório, pois ocorreu no sentido desejado e a população quando comparada com as Testemunhas.

Fernandes et al. (2015), obtiveram informações sobre caracteres agronômicos e caracteres de qualidade de grão de uma população indica x japonica e estimaram diferentes parâmetros genéticos para posterior recomendação como fonte de variabilidade. Assim, realizaram avaliação fenotípica de 147 linhagens endogâmicas recombinantes, juntamente com três Testemunhas, avaliadas em ensaios de campo conduzidos na Estação do Instituto Rio Grandense de Arroz (IRGA), em Cachoeirinha, na safra 2013/14. Realizaram também estimativas de correlações entre os caracteres e uma análise de trilha. Houve alta variabilidade para caracteres relacionados à qualidade de grão e características agronômicas, o que era esperado por ser uma população proveniente de um cruzamento bastante divergente (subespécies indica e japonica) e, ainda assim, há linhagens que apresentaram caracteres desejáveis para o desenvolvimento de novas variedades, tanto para qualidade de grãos como para características agronômicas. As estimativas de correlação e a análise de trilha sugeriram que durante a seleção deveria ser dada maior ênfase para o caráter índice de centro branco, devido a este caráter apresentar valores médios de herdabilidade, ganho esperado com seleção moderado e alta correlação genética e efeito direto negativo com a característica rendimento de engenho, facilitando a seleção e a obtenção de progressos nestes caracteres.

2.7 PRINCÍPIOS DA METODOLOGIA DE MODELOS MISTOS

O modelo linear matemático é considerado como misto quando é composto por efeitos fixos, além da média geral, e efeitos aleatórios, além do erro experimental. A definição de um efeito como fixo ou aleatório depende dos objetivos do estudo. Contudo, um efeito é fixo quando ele é atribuído a um número finito de níveis, portanto, pode ser estimável. Por outro lado, efeito aleatório consiste em uma amostra aleatória de uma

população, isto é, um número infinito de níveis de onde se retira uma amostra. Os efeitos aleatórios são preditos, uma vez que não possuem um valor real fixo *per se* (Searle, 1987).

Handerson (1973) propôs o uso dos modelos mistos. Estes modelos são úteis principalmente em análise de dados desbalanceados. Detêm como vantagens em detrimento à análise de variância tradicional (ANOVA): produzem estimativas ou predições mais acuradas de efeitos de tratamentos quando existem dados perdidos nos experimentos; permitem o ajuste de diferentes variâncias para cada grupo de tratamentos, ou seja, permite considerar variâncias heterogêneas; a modelagem da estrutura de correlação em experimentos com dependência espacial, com medidas repetidas e em múltiplos experimentos, resulta em estimativas mais precisas. Por outro lado, como desvantagem apresenta o efeito *shrinkage*, em que as predições dos efeitos aleatórios são forçadas em direção à média geral, penalizando estimativas baseadas em pequenas amostras (Resende, 2004).

Segundo Resende (2007), a metodologia de modelos mistos, o BLUP (*Best Linear Unbiased Prediction*) é o melhor preditor linear não viesado, que estima o valor genotípico, associado aos efeitos aleatórios, enquanto o BLUE (*Best Linear Unbiased Estimation*) é o melhor estimador não viciado dos efeitos fixos do modelo. Além disso, o procedimento REML (*Restricted Maximum Likelihood*, ou, máxima verossimilhança restrita), permite estimar os componentes de variância para o modelo.

O estimador dos efeitos fixos do modelo (BLUEs) é obtido através do método dos quadrados mínimos generalizados (GLS), enquanto o preditor dos efeitos aleatórios (BLUPs) é obtido pela maximização da função densidade de probabilidade conjunta dos vetores de observações e dos valores genéticos em relação aos efeitos fixos e aleatórios. Na prática, a rigor, não se tem BLUP, mas sim EBLUP, ou seja, BLUP empírico, já que os valores paramétricos são substituídos por suas estimativas. Os testes de comparações múltiplas não devem ser utilizados em modelos mistos, pois eles são derivados sob o enfoque de efeitos fixos de tratamento. Com os BLUPs, obtém-se um ordenamento dos genótipos, pela ordem decrescente de seus valores genotípicos. É importante ressaltar que em alguns casos, especialmente, em programas de melhoramento de plantas anuais, BLUE e BLUP fornecem os mesmos resultados; isso pode acontecer quando os dados são balanceados (Resende, 2007).

A predição de valores por meio de BLUPs requer um conhecimento prévio dos componentes de variância do modelo. Como isso não é possível, as estimativas desses

valores são utilizadas. Dos métodos para se estimar os componentes de variância temos a ANOVA, proposto por Fisher e derivado do método dos quadrados mínimos; o método da Máxima Verossimilhança ou ML, desenvolvido por Hartley & Rao (1967) e o método da Máxima Verossimilhança Residual (Maximum likelihood restricted) ou REML, de Patterson & Thompson (1971), derivados da função de verossimilhança.

Os métodos baseados em função de verossimilhança são recomendados para estimar componentes de variância em avaliações genéticas com dados desbalanceados, pois fornece estimativas não negativas dos componentes de variância e considera as perdas de graus de liberdade resultantes dos efeitos fixos. Esse último fato é responsável por gerar estimativas não viciadas dos componentes de variância. O método REML divide os dados em duas partes: os efeitos fixos do modelo e os contrastes dos erros, sendo que apenas o último é utilizado para estimar os componentes de variância, uma vez que ele contém todas as informações inerentes aos parâmetros de variância, e esse é o motivo do nome “máxima verossimilhança residual”, ou seja, considera apenas os contrastes dos resíduos pelos efeitos fixos, durante sua estimação. O método REML exige normalidade dos resíduos, porém a heterogeneidade de variâncias pode ser contornada, além de permitir relaxar a suposição de independência dos erros, permitindo maior flexibilidade na modelagem (Resende et al., 2014).

A estimação dos componentes de variância via REML proporciona a maximização da função de verossimilhança, que é uma função de densidade de probabilidade, fixada sobre o conjunto de dados (y), em que se procura o parâmetro θ mais verossímil ou plausível a esse conjunto de dados. Como o método REML divide a função de verossimilhança em duas partes independentes – uma parte referente aos efeitos fixos e a outra, referente aos resíduos – os efeitos fixos são removidos do modelo. Portanto, somente a porção da verossimilhança dos efeitos aleatórios é maximizada, ou seja, consideram-se apenas os resíduos (Resende et al., 2014). Os métodos REML tem como vantagens a possibilidade de analisar dados desbalanceados; o ajuste de modelos e delineamentos que não podem ser acomodados à ANOVA tradicional; a possibilidade lidar com uma estrutura complexa de dados; a comparação entre indivíduos através do tempo e do espaço é permitida, ao possibilitar o uso de um grande número de informações, o que também geram estimativas e predições mais precisas; a flexibilização da modelagem dos dados (Resende et al., 2014).

A metodologia dos modelos mistos além de estimar os componentes de variância e fazer previsões e estimativas em torno dos valores genotípicos, também permite testar se o modelo que está sendo empregado é adequado. É possível testar a significância dos efeitos aleatórios do modelo através da análise de *deviance* (ANADEV), baseada no teste de razão de verossimilhança (LRT), com a finalidade de verificar se o modelo possui uma estrutura hierárquica ou aninhada. Essa análise permite a comparação de modelos que possuem diferentes efeitos aleatórios, porém a estrutura de efeitos fixos é mantida. A *deviance* consiste na razão entre as verossimilhanças do modelo completo em relação ao modelo sem o efeito que se deseja testar (Resende et al., 2014).

2.8 MAPEAMENTO DE QTL (*Quantitative Trait Loci*)

O mapeamento genético permite identificar regiões cromossômicas responsáveis por alguma variação observada em características quantitativas, bem como determinar a forma de ação dos genes ou grupos de genes associados a essas regiões. QTLs são regiões do genoma responsáveis pela expressão de caracteres fenotípicos, que possuem distribuição contínua, tais como, altura e peso de plantas e de animais; produção de grãos; teor de óleo etc. Com o surgimento dos marcadores moleculares tornou-se possível mapear QTLs que afetam esses caracteres quantitativos (Guo et al. 2014).

Mapear um QTL significa identificar sua posição no genoma e estimar o efeito aditivo, efeito de dominância e outros efeitos presentes no modelo adotado. Cada gene tem um pequeno efeito na expressão do fenótipo, portanto, são ditos como complementares ou aditivos. As interações entre alelos podem ser de dominância completa, incompleta ou codominância. Por outro lado, as interações que ocorrem entre genes de cromossomos diferentes, ou genes que não estão ligados, são epistáticas (Resende, 2012).

Ao considerar que o efeito do QTL não é observável, a ideia, a princípio, baseada na ligação genética, seria identificá-los por meio de genes de manifestação observável, isto é, de características qualitativas, com localização no mesmo cromossomo, posições próximos ou ligados. Assim, a identificação dos QTLs é feita por meio de análises de segregação conjunta entre o QTL e os genes que são denominados de locos marcadores (Ferreira & Grattapaglia, 1998). Contudo, se um loco marcador qualquer, de fácil identificação fenotípica, estiver fisicamente ligado, a pequena distância, a um gene que controla um caráter de interesse, a seleção deste marcador pode resultar indiretamente

na seleção do gene de interesse, a não ser que ocorram recombinações (Ferreira & Grattapaglia, 1998).

Em uma determinada população obtida de um cruzamento fechado, o passo inicial é a identificação de marcadores polimórficos entre os genitores desse cruzamento. A genotipagem da progênie desse cruzamento permite o desenvolvimento de mapas genéticos podem ser construídos, mostrando a ordem dos locos nos respectivos cromossomos e a distância relativa entre os marcadores, resultante da frequência de recombinação entre eles. Alguns pré-requisitos são importantes para o mapeamento de QTLs, tais como: o delineamento genético adequado (populações F_2 , retrocruzamentos, linhas puras recombinantes); número de marcadores polimórficos entre os genitores, e dados fenotípicos obtidos por experimentos com baixo coeficiente de variação (Lynch & Walsh, 1998).

O primeiro procedimento no mapeamento é a escolha dos genitores de modo a maximizar a probabilidade de encontrar polimorfismo. A determinação de ligação gênica entre um QTL e um marcador molecular depende da existência de desequilíbrio de ligação entre seus alelos. Por essa razão, para a construção de mapa molecular e identificação de QTLs são escolhidas duas linhagens genética e fenotipicamente divergentes em relação aos caracteres de interesse. Essas linhagens são hibridizadas e delas obtidas populações segregantes, que poderão ser F_2 , F_3 , etc., retrocruzamentos (RC), linhagens dupla-haplóides e linhas puras recombinantes (RIL – *Recombinant Inbreed Lines*). Em seguida, os indivíduos dessa população são avaliados quanto aos caracteres de interesse. Utilizando-se famílias F_2 , diferentes modelos podem ser ajustados de forma a estimar separadamente contribuições de efeitos aditivos e dominantes. Quando são usados RC, somente é estimado o efeito aditivo de uma substituição alélica. Deve ser considerado que com a utilização de marcadores altamente polimórficos baseados em SSR, por exemplo, elimina-se a necessidade de cruzamentos distantes para maximizar o polimorfismo molecular. Porém, no contexto de mapeamento de características quantitativas, a seleção de genitores baseia-se na utilização de pais fenotipicamente contrastantes para as características de interesse, visando maximizar o efeito da substituição alélica nos locos associados a caracteres quantitativos (Ferreira & Grattapaglia, 1998). Os marcadores SNP são uma alternativa para gerar mapas de alta resolução para análise de QTLs. Os SNPs são a forma mais abundante de variação no DNA e possuem baixa taxa de mutação, facilidade de genotipagem e natureza bialélica.

Marathi et al. (2012) avaliaram 310 populações recombinantes (RILs) de cultivares de arroz, em três locais diferentes, e como resultado identificaram 44 QTLs para características associadas à produtividade, porém apenas quatro se expressaram nos três locais avaliados.

Ikeda et al. (2013) avaliaram população F_2 de duas variedades de arroz (Koshihikari e Habataki) e identificaram QTLs relacionado ao número de grãos, tamanho de grãos e componentes de produtividade. Bai et al. (2011) utilizaram linhagens isogênicas de arroz e encontram QTLs relacionados aos componentes de produtividade. Liu et al. (2015) avaliaram 63 linhagens recombinantes de arroz e encontraram QTLs relacionados a fertilidade de espiguetas, peso de mil grãos, número de espiguetas por panículas e rendimento de grãos.

O método mais simples de mapeamento de QTL é conhecido por análise de marcas individuais (*single point analysis*). De acordo com este método, independentemente da quantidade de marcadores disponíveis, a associação entre o QTL e um único marcador é testada separadamente. A identificação de um QTL ligado a um marcador baseia-se na comparação entre as médias das classes genotípicas dos marcadores. Havendo diferença significativa entre médias, existe possibilidade de um QTL estar associado ao marcador em questão. A análise de marcas individuais é muito utilizada quando se deseja apenas a detecção de um QTL ligado à marca, ou quando o objetivo é descartar marcas não informativas, como um passo preliminar para uma análise mais elaborada (Lynch & Walsh, 1998). Uma desvantagem desse método é que o efeito de um QTL fica confundido com sua posição, que não pode ser precisamente determinada.

Outro método é o mapeamento por intervalo, o qual considera um par de marcas adjacentes e infere sobre a existência de um QTL em qualquer posição entre essas duas marcas. O mapeamento por intervalo foi proposto por Lander & Bostein (1989) como uma alternativa para eliminar alguns dos principais problemas apresentados pela análise de marcas individuais. Esse método permite obter estimativas mais precisas de efeito e posição do QTL, além de aumentar o poder de detecção em relação à análise de marcas individuais. Outra vantagem desse método é que o número de indivíduos necessários para a análise é menor que no método de marcas individuais. Um modelo linear é usado para associar o fenótipo ao genótipo do QTL, e as estimativas dos parâmetros são obtidas pelo método da máxima verossimilhança. O teste de razão de verossimilhança, ou LOD Score, é

usado para verificar a existência de ligação entre os marcadores e o QTL, de forma similar à análise de marcas individuais (Lynch & Walsh, 1998).

Quando se avalia a presença de um QTL em um intervalo, deseja-se que este teste seja independente dos efeitos de possíveis QTLs em outras regiões do cromossomo. Este teste pode ser construído utilizando-se uma combinação do mapeamento por intervalo com o modelo de regressão múltipla (Zeng, 1994). De acordo com este método, os efeitos dos QTLs fora do intervalo considerado são inseridos no modelo sob a forma de cofatores, sendo eliminados como resíduos.

O terceiro método é o de mapeamento de múltiplos intervalos, proposto por Kao et al. (1999), e consiste no mapeamento de vários QTLs simultaneamente, em intervalos distintos, obtendo-se um aumento na precisão e no poder de detecção. O mapeamento de múltiplos intervalos permite a estimação dos parâmetros já existentes nos métodos anteriores, como posição e efeitos do QTL, além da possibilidade de estimar o número de QTLs no genoma e os efeitos de interação entre eles, denominados efeitos epistáticos (Kao et al., 1999).

2.9 COMPONENTES DE PRODUÇÃO E ALTURA DE PLANTAS

A produção de uma planta de arroz é uma característica agrônômica complexa, determinada pelos três caracteres componentes: número de panículas, número de grãos por panícula e peso de grãos. O número de panículas é dependente da capacidade da planta para a produção de perfilhos, incluindo perfilhos primários, secundários e perfilhos superiores. O número de grãos por panícula também pode ser atribuído a dois subcomponentes: número de espiguetas e taxa de fecundação de espiguetas. O peso de grãos é largamente determinado pelo tamanho do grão, que é especificado por suas três dimensões (comprimento, largura e espessura), e o grau de enchimento (Bai et al., 2011).

As cultivares de arroz diferem enormemente nos níveis de produtividade de grão, existindo diversas combinações de genes e alelos relacionados a cada caráter componente da produção devido à grande variabilidade genética inerente da espécie. Além disso, o nível de produtividade do arroz é uma característica altamente influenciada pelas condições ambientais e as práticas de cultivo adotadas. Outro fator responsável por essa inconstância nos níveis de produtividade é a interação genótipo e ambiente, proporcionando cultivares adaptadas a condições ambientais específicas (Wang et al., 2015).

A herança de caracteres quantitativos envolve múltiplos genes, cada um tendo um pequeno efeito que é sensível às mudanças ambientais. Esses caracteres são conhecidos como tendo geralmente baixa herdabilidade e sendo assim, são considerados de difícil investigação. No entanto, o desenvolvimento de marcadores moleculares, o mapeamento genômico e as análises envolvendo QTL têm facilitado a investigação das bases genéticas desses caracteres quantitativos. A análise a ser adotada para a identificação e compreensão destes QTLs vai depender da população de mapeamento a ser utilizada (Ikeda et al., 2013).

Os genes que controlam os caracteres relacionados à produtividade em arroz podem ser divididos em duas classes: I) genes dos processos biológicos básicos, sendo que tais genes podem causar efeitos deletérios sobre as plantas como redução do desenvolvimento, anormalidades na morfologia e taxas muito baixas de produtividade, por exemplo, os genes relacionados ao perfilhamento e ao desenvolvimento de panículas. II) genes de alterações quantitativas dos caracteres, em que diferentes alelos de determinado loco, incluindo alelos de perda de função, podem causar variações nas medidas dos caracteres, como os genes relacionados ao número de grãos por panícula, tamanho de panícula e dimensão do grão (Wang et al., 2015). Muitos estudos básicos procuraram identificar regiões genômicas relacionadas à produtividade em arroz. Entre os QTLs que controlam os caracteres produtivos, e têm sido clonados e caracterizados. A herança de caracteres quantitativos complexos, como é o caso da produtividade, envolve múltiplos genes, cada um com um pequeno efeito que é sensível a alterações ambientais. Esses caracteres são conhecidos em geral como tendo baixa herdabilidade e, por esse motivo, são difíceis de investigar (Xing & Zhang, 2010).

A altura da planta tem correlação negativa com a produtividade, pois afeta diretamente o número de perfilhos produtivos (Zhou et al., 2016). Em plantas altas, a energia que deveria ser destinada à formação dos grãos é direcionada para a manutenção e sobrevivência da planta. Além disso, plantas altas tendem a acamar, acarretando em perdas de grãos e resultando em dificuldades na colheita.

Na Revolução Verde, em 1960, foi introduzido em linhagens de arroz o gene *sd1*, para a produção de cultivares semi-anãs, reduzindo o porte da planta e, conseqüentemente, as perdas por acamamento (Bai et al., 2011). O número de perfilhos, bem como o ângulo em que eles estão inseridos; o número, a densidade e tamanho das panículas; a quantidade e o peso dos grãos, este último correlacionado ao comprimento e largura dos grãos e; a largura e o ângulo de inserção da folha bandeira são caracteres

correlacionados positivamente à produtividade da cultura. Já que possuem alta herdabilidade, a seleção baseada nesses caracteres tende a resultar em um incremento da produtividade.

2.10 SEQUENCIAMENTO DE NOVA GERAÇÃO (NGS)

O sequenciamento de DNA baseado no método de Sanger obteve grande sucesso e esteve envolvido em todos os projetos desse tipo durante aproximadamente três décadas, desde o seu desenvolvimento, em 1977. No entanto, apesar de sua robustez, o método apresenta desvantagens, dentre as quais podem ser citadas a necessidade de uso de bibliotecas de clones para a preparação das amostras, processo este caro e demorado, e a baixa velocidade de produção alcançada com esse tipo de tecnologia, a qual é, em muitas ordens de magnitude, inferior àquela desejável em projetos de genômica. Assim sendo, com o intuito de superar as deficiências inerentes ao método de Sanger, novas técnicas alternativas surgiram. Tais técnicas visavam um processo de sequenciamento livre de clonagem baseada em vetores e a redução significativa da duração e custo do processo como um todo (Chaisson et al., 2004).

Das alternativas criadas em busca de menor custo e maior rapidez para o processo, as tecnologias NGS foram as primeiras a se consolidarem como soluções realmente eficientes, em resposta à limitada produtividade dos métodos de "primeira geração", pelo poder de gerarem informação muitas vezes maior e com grande economia de tempo e custo por base sequenciada. Como exemplo, em 2005 a empresa Roche lançou o sequenciador 454 (Droege, 2008), com tecnologia baseada no método de pirosequenciamento (Leamon, 2003).

As tecnologias NGS são capazes de gerar dados na ordem de milhões a bilhões de pares de bases por experimento (ou corrida) (Mardis, 2011; Pareek et al., 2011), contra as relativamente poucas milhares de bases proporcionadas pelo método de Sanger automatizado. Diferem deste último, também, por gerarem produtos do sequenciamento (leituras) de pequena extensão tipicamente da ordem de 30 a 400 pares de bases em contraposição aos comprimentos de 700 a 1000 típicos da tecnologia baseada em Sanger. O tamanho da leitura, nas plataformas NGS, é menor, porém como uma amostragem maior do genoma, permitindo uma grande cobertura de sequenciamento (de 30 vezes ou mais) a um baixo custo relativo, se comparada com a cobertura típica obtida com as plataformas de

eletroforese capilar. Assim, com relação a esta última, conseguem promover um sequenciamento muito mais rápido, produtivo e bem menos oneroso (Chaisson et al., 2004).

2.11 GENOTIPAGEM DArT (*Diversity Arrays Technology*)

DArT (*Diversity Arrays Technology*) refere-se a técnica que permite a detecção simultânea de centenas de locos polimórficos espalhados pelo genoma. Esta técnica não necessita de informações da sequência de DNA ou de *primers* específicos e tem baixo custo de genotipagem por loco. Os marcadores moleculares SNPs derivados da técnica DArT são bialélicos, comportam-se de maneira co-dominante (dose dupla, dose única ou ausência) ou dominante (presença ou ausência) .

A técnica DArT tem sido amplamente utilizada em várias espécies de importância agrônômica, arroz (Xie et al., 2006), sorgo (Mace et al., 2008), cana-de-açúcar (Heller-Uszynska et al., 2010), dentre outras. Muitos métodos têm sido desenvolvidos para reduzir a complexidade do genoma, como por exemplo, GBS (*Genotyping by Sequencing*), no entanto, a técnica DArT proporciona uma vantagem significativa por meio da seleção de uma fração correspondente a genes predominantemente ativos. Esta seleção é conseguida através do uso de uma combinação de enzimas de restrição que separam sequências de número de cópias baixas a partir da fração repetitiva do genoma (Jaccoud et al., 2001).

A tecnologia DArT pode resumidamente ser dividida em duas etapas: desenvolvimento dos painéis de diversidade e genotipagem. Na primeira etapa o DNA das amostras a serem estudadas é agrupado e clivado por enzimas de restrição, sendo uma de corte raro e outra de corte frequente, e, posteriormente, aos fragmentos digeridos são ligados adaptadores às extremidades de corte raro. Posteriormente, uma reação da polimerase em cadeia (PCR) é realizada para amplificação dos fragmentos gerados utilizando-se *primers* complementares à sequência dos adaptadores. A complexidade do genoma é reduzida, pois somente fragmentos com as duas extremidades ligadas aos adaptadores são amplificados (Jaccoud et al., 2001). A etapa de redução da complexidade do genoma determina a eficiência da identificação dos polimorfismos na técnica de DArT, especialmente em espécies que apresentam genomas extensos e complexos (Helleur-Uszynska et al., 2011). Ainda na primeira etapa, os fragmentos amplificados via PCR são

clonados em um vetor. Os insertos são novamente amplificados, purificados e arranjados em um suporte sólido (microarray), gerando assim um painel de diversidade através da impressão dos microarrays contendo as representações genômicas em lâminas de vidro, também denominadas slides (Jaccoud et al., 2001).

Embora a tecnologia DArT seja muito eficiente, existe disponível hoje a chamada DArTseq, tecnologia que também se baseia na redução de complexidade genômica e detecção de polimorfismos, porém baseada em NGS. A principal diferença entre a tecnologia DArTseq e outras metodologias de genotipagem por sequenciamento trata-se de como a redução da complexidade do genoma é obtida. A metodologia utilizada na redução da complexidade pela tecnologia DArTseq é simples e também eficiente na seleção de regiões de baixo número de cópias. Estima-se que 90% dos marcadores identificados via DArTseq alinham-se uma única vez no genoma de referência (quando esta facilidade está disponível) (Cruz et al., 2013).

A alta cobertura praticada pela tecnologia DArTseq assegura alta qualidade dos marcadores identificados, tanto baseada na reprodutibilidade (> 97%) quanto *Call rate*, e além disto, esta tecnologia permite a exploração de dois tipos de dados: dados dominantes (marcas presentes e ausentes) análogos aos obtidos via DArT, porém extraídos de forma *in silico*; e marcadores do tipo SNP (*Single Nucleotide Polymorphism*). Sendo assim, a tecnologia DArTseq caracteriza-se como uma técnica simples, altamente multiplexável, adequada para estudos populacionais, caracterização de germoplasmas, melhoramento genético, dentre outros. Adicionalmente, DArTseq representa uma alternativa a protocolos caros e complexos como a produção “chip” de alta densidade, o que requer um investimento substancial de recursos. Adicionalmente marcadores disponibilizados via “chip” são específicos para a população em que foram desenvolvidos, representando, assim, um grave problema quando se considera o nível de polimorfismo necessário para a condução do estudo desejado (Cruz et al., 2013). A tecnologia é realizada pela empresa DArT Pty/Ltd. (Australia), e utilizada com sucesso para a construção de mapas genéticos, estudos de caracterização de coleções de germoplasma e de diversidade genética, e é um sistema rápido, robusto e que requer quantidades mínimas de DNA.

2.12 REFERÊNCIAS

ABADIE, T.; CORDEIRO, C. M. T.; FONSECA, J. R.; ALVES, R. B. N.; BURLE, M. L.; BRONDANI, C.; RANGEL, P. H. N.; CASTRO, E. M.; SILVA, H. T.; FREIRE, M. S.; ZIMMERMANN, F. J. P.; MAGALHÃES, J. R. **Construção de uma coleção nuclear de arroz para o Brasil. Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 40, n. 2, p. 129136, fev. 2005.

ABREU, A. de F.B. **Predição do potencial genético de populações segregantes do feijoeiro utilizando genitores inter-raciais.** (Tese – Doutorado em Genética e Melhoramento de Plantas). Lavras: UFLA, p.79, 1997.

AGRIANUAL. **Anuário da Agricultura Brasileira.** Arroz. São Paulo: FNP, p. 161-167. 2019.

ANTONINI, S. R. C.; MENEGHIN, S. P.; URASHIMA, A. S. **Técnicas Básicas de Biologia Molecular.** Universidade Federal de São Carlos. Centro de Ciências Agrárias-Campus de Araras. Araras- SP, 2004.

BAEZINGER, P. S.; PETERSON, C. J. Genetic variation: Its origin and use for breeding self-pollinated species. Raleigh: **C.A.B. International**, Cap.4, p.69-92, 1991.

BAI, X. F.; LUO, L. J.; YAN W. H.; KOVI, M. R.; XING, Y. Z. Quantitative trait loci for rice yield-related traits using recombinant inbred lines derived from two diverse cultivars. **Journal of Genetics**, v. 90, n. 2, p. 209–215, 2011.

BANDEIRA, L. F. **Caracterização genética de populações naturais de araticum (*Annona crassiflora* Mart.)**, Dissertação (Mestrado em Genética) - Biologia, Universidade Federal de Goiás. Goiânia, P. 78, 2003.

BARILI, L.D.; VALE, N. M.; ARRUDA, B.; TOALDO, D.; ROCHA, F.; COIMBRA, J. L. M.; BERTOLDO, J. G.; GUIDOLIN, A. F. Escolha de Genitores Contrastantes para Compor o Bloco de Cruzamentos de Genótipos de Feijão. **Revista Brasileira Agrociência**, Pelotas, v.17, n. 3-4, p. 303-310, Jul-set, 2011.

BHATI, M.; BABU, G. S.; RAJPUT, A. S. Genetic variability, correlation and path coefficient for grain yield and quantitative traits of elite rice (*Oryza sativa* L.) genotypes at Uttar Pradesh. **Electronic Journal of Plant Breeding**, v. 6, n. 7, p. 586-591, jun. 2015.

BORÉM, A. **Melhoramento de plantas.** Viçosa: UFV, 547 p. 1997.

BRONDANI, C.; BORBA, T.C.O.; RANGEL, P.H.N.; BRONDANI, R.P.V. Determination of genetic variability of traditional varieties of Brazilian rice using microsatellite markers. **Genetics and Molecular Biology**, v.29, p.676-684, 2006.

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

CASTANHEIRA, A.L.M.; SANTOS, J.B. RAPD marker assessment of self-pollinated inbreeding methods for common bean segregant populations. **Crop Breeding and Applied Biotechnology**, v.4, p.16, 2004.

CHAISSON M, PEVZNER P, TANG H. Fragment assembly with short reads. **Bioinformatics**; 20(13):2067-74, 2004.

CONAB- COMPANHIA NACIONAL DE ABASTECIMENTO. **Produção brasileira de grãos**. Brasília: Companhia Nacional de Abastecimento. 2019. Disponível em: <<http://www.agricultura.gov.br>>. Acesso em: 05 de jul. de 2019.

COSTA, R. B.; RESENDE, M. D. V. de; GONÇALVES, O. S.; REGO, F. L. H.; CONTINI, A. Z.; ROA, R. A. R.; MARTINS, W. J. Genetic evaluation of *Hevea brasiliensis* [(Wild ex Adr. De Juss) Muell. Arg.] for juvenile vigour using the Reml/Blup method. **Floresta e Ambiente**, Rio de Janeiro, n. 12, p. 08-14, 2006.

CRUZ, C. D.; REGAZZI, A. J. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa, Minas Gerais: Ed. UFV, p. 390, 1994.

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

CRUZ, V. M. V.; KILLIAN, 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, Issue 5, e64062, May 2013.

DIXT, S.; SINGH, M.U.; ABBAI, R.; RAM, T.; SINGH, A. P.; VIRK, P. S.; KUMAR, A. Identification of genomic region(s) responsible for high iron and zinc content in rice. **Nature**, Scientific RepoRts 9:8136, 1038/s41598-019-43888, 2019.

DOVALE, J.C.; SOAREWS, P.C.; CORNÉLIO, V. M. O.; REIS, M. S.; BORGES, V.; BISI, R. B.; SOARES, A. A.; FRITSCHÉ-NETO, R. Contribuição genética na produtividade do arroz irrigado em Minas Gerais no período de 1998 a 2010. **Bragantia**, Campinas, v. 71, n. 4, p.460-466, 2012.

DROEGE M, Hill B. The Genome Sequencer FLX™ System: longer reads, more applications, straight forward bioinformatics and more complete data sets. **J. Biotech**;136:3-10, 2008.

FALCONER, D.S. **Introdução à genética quantitativa**. Tradução de Martinho de Almeida e Silva e José Carlos Silva. Viçosa: UFV, 279p, 1987.

FERNANDES, R. H.; BARBOSA NETO, J. F.; MARSCHALEK, R.; WICKERT, E. **Estimativa de parâmetros genéticos para caracteres de qualidade de grão e características agrônômicas de linhagens endogâmicas recombinantes de arroz (*Oryza sativa* L.)**. (Dissertação de Mestrado), 2015.

FERREIRA, M. E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. 3.ed. Brasília: Embrapa-Cenargen, (EMBRAPA-CENARGEN. Documentos, 20), p. 220, 1998.

GARDNER, O.G.; Estimates of genetic parametes in cross fertilizing plants and their implications in plant breeding. In: Hason, W. D. & ROBINSON, H. S. (eds). **Statistical Genetics and plant breeding**. Washington, National academy of Science-National Research Council, p. 52-225, 1963.

GOFF, S.A.; RICKE, D.; LAN, T.-H.; PRESTING, G.; WANG, R.; DUNN, M.; GLAZEBROOK, J.; SESSIONS, A.; OELLER, P.; VARMA, H. A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *japonica*). **Science**, Washington, v.296, n.5565, p.92-100, 2002.

GUIMARÃES, C. M; SANTOS, A. B.; MAGALHÃES JÚNIOR, A. M.; STONE, L. F. In: SANTOS, A.B. dos; STONE, L.F.; VIEIRA, N.R. de A. (Ed.). **A cultura do arroz no Brasil**. Santo Antônio de Goiás: Embrapa Arroz e Feijão, P. 53-96, 2006.

GUIMARÃES, C. M; SANTOS, A. B.; MAGALHÃES JÚNIOR, A. M.; STONE, L. F. In: SANTOS, A.B. dos; STONE, L.F.; VIEIRA, N.R. de A. (Ed.). **A cultura do arroz no Brasil**. Santo Antônio de Goiás: Embrapa Arroz e Feijão, P. 53-96, 2006.

GUO, L.; GAO, Z.; QIAN, Q. Application of resequencing to rice genomics, functional genomics and evolutionary analysis. **Rice**, New York, v. 7, n. 1, p. 4, 2014.

HALLAUER, A. R.; MIRANDA FILHO, J. B. **Quantitative Genetics in Maize Breeding**. The Iowa State University Press. 2 ed. 1982.

HARTLEY, H. O.; RAO, J. N. K. Maximum-likelihood estimation for the mixed analysis of variance model. **Biometrika**, Londres, v. 54, n. 1-2, p. 93-108, 1967.

HAYASHI, K.; HASHIMOTO, N.; DAIGEN, M.; ASHIKAWA, I. Development of PCRbased SNP markers for rice blast resistance genes at the Piz locus. **Theoretical and Applied Genetics**, Berlin, v. 108, n. 7, p. 1212-1220, 2004.

HELLER-USZYNSKA, K.; USZYNSKI, G.; HUTTNER, E.; EVERS, M.; CARLIG, J.; CAIG, V.; AITKEN, K.; JACKSON, P.; PIPERIDIS, G.; COX, M.; GILMOUR, R.; D'HONT, A.; BUTTERFIELD, M.; GLASZMANN, J. C.; KILIAN A. Diversity Arrays Technology effectively reveals DNA polymorphism in a large and complex genome of sugarcane. **Mol Breeding, Canberra**, DOI 10.1007/s11032-010-9460-y, Maio 2010.

HUANG, M.; CHEN, L.; CHEN Z. Diallel analysis of combining ability and heterosis for yield and yield components in rice by using positive loci. **Euphytica**, 205 (1): 37-50, 2015.

IKEDA, M.; MIURA, K.; AYA, K.; KITANO, H.; MATSUOKA, M. Genes offering the potential for designing yield-related traits in rice. **Current opinion in plant biology**, v. 16, n. 2, p. 213-220, 2013.

JACKSON, S. A. Rice: The First Crop Genome. **Rice**, Nova York, v. 9, n. 1, p. 1-3, 2016.

KAO, C. H.; ZENG, Z. B.; TEASDALE, R. D. Multiple interval mapping for quantitative trait loci. **Genetics**, Baltimore, v. 152, n. 3, p. 1203-1216, July 1999.

KAWAHARA, Y.; BASTIDE, M.; HAMILTON, J. P.; KANAMORI, H.; MCCOMBIE, W. R.; OUYANG, S. et al. Improvement of the *Oryza Sativa* Nipponbare reference genome using next generation sequence and optical map data. **Rice**, New York, v. 6, n. 1, p. 1-10, 2013.

KHUSH, G.S. Origin, dispersal, cultivation and variation of rice. **Plant Molecular Biology**, Zurich, v. 35, n. 1/2, p. 25-34, 1997.

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.

LANDER, E.S.; BOTSTEIN, D. Mapping medelian factors underlying quantitative traits using RFLP linkage maps. **Genetics**, Baltimore, v. 121, n. 1, p. 185-199, Jan. 1989.

LEAMON JH, LEE WL, TARTARO KR, LANZA JR, SARKIS GJ, DE WINTER AD, et al. A massively parallel PicoTiterPlate™ based platform for discrete picoliter-scale polymerase chain reactions. **Electrophoresis**; 24:3769-77, 2003.

LI, J.-Y.; WANG, J.; ZEIGLER, R. S. The 3,000 rice genomes project: new opportunities and challenges for future rice research. **GigaScience**, Londres, v. 3, n. 1, p. 8, 2014.

LIU, D.; KANG, M.; WANG, F.; LIU, W.; FU, C.; LI, J.; ZHU, M.; ZENG, X.; LIAO, Y.; LIU, Z.; HUANG, H. Mapping of the genetic determinant for grain size in rice using a recombinant inbred line (RIL) population generated from two elite indica parents. **Euphytica**, Dordrecht, v. 206, n. 1, p. 159-173, 2015.

LYNCH, M.; WALSH, B. **Genetics and analysis of quantitative traits**. Sunderland: Sinauer Associates, p. 980, 1998.

MA, X.; FU, Y.; ZHAO, X.; JIANG, L.; ZHU, Z.; GU, P.; XU, W.; SU, Z.; SUN, C.; TAN, L. Genomic structure analysis of a set of *Oryza nivara* introgression lines and identification of yield-associated QTLs using whole-genome resequencing. **Scientific RepoRts**. | 6:27425 | DOI: 10.1038/srep27425, 2016.

MARATHI, B.; GULERIA, S.; MOHAPATRA, T.; PARSAD, R.; MARIAPPAN, N.; KURUNGARA, V. K.; ATWAL, S. S.; PRABHU, K. V.; SINGH, N. K.; SINGH, A. K. QTL analysis of novel genomic regions associated with yield and yield related traits in new plant type based recombinant inbred lines of rice (*Oryza sativa* L.). **BMC Plant Biology**, v. 12, n. 1, p. 137, 2012.

MARDIS ER. A decade's perspective on DNA sequencing technology. **Nature**;470:198203, 2011.

MATSUMOTO, T.; WU, J.; ITOH, T.; NUMA, H.; ANTONIO, B.; SASAKI, T. THE map-based sequence of the rice genome. **Nature**, v. 436, n. 7052, p. 793–800, 2005.

MCCOUCH, S. R.; CHEN, X.; PANAUD, O.; TEMNYKH, S.; XU, Y.; CHO, Y. G.; HUANG, N.; ISHII, T.; BLAIR, M. Microsatellite marker development, mapping and applications in rice genetics and breeding. **Plant molecular biology**, The Hague, v. 35, n. 1-2, p. 89-99, 1997.

NASS, L. L.; VALOIS, A. C.C.; MELO, I. S.; VALADARES-INGLIS, M. C. **Recursos Genéticos e Melhoramento de Plantas**. Fundação MT, Rondonópolis, 2001.

PAREEK CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. *J. Appl. Genetics*; 52:413-35, 2011.

PARIDA, S. K.; MUKERJI, M.; SINGH, A. K.; SINGH, N. K.; MOHAPATRA, T. SNPs in stress-responsive rice genes: validation, genotyping, functional relevance and population structure. **BMC Genomics**, London, v. 13, n. 1, p. 426, 2012.

PATERSON, A. H.; FREELING, M.; SASAKI, T. Grains of knowledge: Genomics of model cereals. **Genome Research**, Woodbury, v. 15, p. 1643–1650, 2010.

PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, Londres, v. 58, n. 3, p. 545-554, 1971.

QIU, X.; CHEN, K. ; LV, W; OU, X.; ZHU, Z.; XING, D. YANG, L.; FAN, F.; YANG, J.; XU, J.; ZHENG, T.; LI, Z. Examining two sets of introgression lines reveals backgroundindependent and stably expressed QTL that improve grain appearance quality in rice (*Oryza sativa* L.). **Theor Appl Genet**. V.130, p. 951–967, 2017.

RAMALHO, M.A.P. Melhoramento do feijoeiro. In: ABREU, A. de F.B.; GONÇALVES, F.M.A.; MARQUES JR., O.G.; RIBEIRO, P.H.E. (ed.). SIMPÓSIO SOBRE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS, 1997, Lavras. **Anais...** Lavras: UFLA/GEN, p.167-196, 1997.

RAMALHO, M.A.P., SANTOS, J.B., PINTO, C.A.B.P. **Genética na Agropecuária**. Lavras: UFLA, Imprensa Universitária, 472p, 2000.

RAMALHO, M.A.P.; ABREU, A.F.B.; SANTOS, J.B.; NUNES, J.A.R. **Aplicações da genética quantitativa no melhoramento de plantas autógamas**. Lavras: Editora UFLA, Ed. 1, 522p, 2012.

RAMALHO, M.A.P.; GONÇALVES, F.M.A.; SOUZA SOBRINHO, F. de S. Seleção recorrente no melhoramento do cafeeiro. In: SIMPÓSIO SOBRE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS, 3: Genética e melhoramento do cafeeiro, 1999, Lavras. **Anais...** Lavras: UFLA/GEN, p.66-81, 1999.

RANGEL, P.H.; GUIMARÃES, E.P.; NEVES, P.C.F. Base genética das cultivares de arroz (*Oryza sativa* L.) irrigado do Brasil. **Pesquisa agropecuária brasileira** v.31, p.349-357, 1996.

RANGEL, P.H.N.; FERREIRA, M.E.; FAGUNDES, P.R.R.; TERRES, A.L.; MAGALHÃES JR., A.M. de; FRANCO, D.F. Ampliação da base genética das populações

utilizadas no melhoramento do arroz de várzea no Brasil. In: 1º CONGRESSO BRASILEIRO DE ARROZ IRRIGADO, 1., REUNIÃO DA CULTURA DO ARROZ IRRIGADO, 23., 1999, Pelotas.. **Anais...** Pelotas: EMBRAPA Clima Temperado, p. 15, 1999.

RANGEL, P.H.N.; FERREIRA, M.E.; FAGUNDES, P.R.R.; TERRES, A.L.; MAGALHÃES JR., A.M. de; FRANCO, D.F. Ampliação da base genética das populações utilizadas no melhoramento do arroz de várzea no Brasil. In: 1º CONGRESSO BRASILEIRO DE ARROZ IRRIGADO, 1., REUNIÃO DA CULTURA DO ARROZ IRRIGADO, 23., 1999, Pelotas.. **Anais...** Pelotas: EMBRAPA Clima Temperado, p. 15, 1999.

RESENDE, D. V.; SILVA, F. F.; AZEVEDO, C. F. **Estatística Matemática, Biométrica e Computacional: Modelos Mistos, Multivariados, Categóricos e Generalizados (REML/BLUP), Inferência Baysiana, Regressão Aleatória, Seleção Genômica, QTLGWAS, Espacial e Temporal, Competição, Sobrevivência.** Viçosa, MG, Suprema, 2014.

RESENDE, M. D. V. D.; SILVA, F. F.; AZEVEDO, C. F. **Estatística Matemática, Biométrica e computacional: Modelos Mistos, Multivariados, Categóricos e Generalizados (REML/BLUP), Inferência Bayesiana, Regressão Aleatória, Seleção Genômica, QTL-GWAS, Estatística Espacial e Temporal, Competição, Sobrevivência.** Viçosa: Suprema, 881 p, 2014.

RESENDE, M. D. V. **Matemática e estatística na análise de experimentos e no melhoramento genético.** Colombo: Embrapa Florestas, 561 p, 2007.

RESENDE, M. D. V. **Métodos estatísticos ótimos na análise de experimentos de campo.** Colombo: Embrapa Florestas 65 p, , 2004.

SAKAI, H. LEE, S. S.; TANAKA, T.; NUMA, H.; KIM, J.; KAWAHARA, Y.; WAKIMOTO, H.; YANG, C. C.; IWAMOTO, M.; ABE, T.; YAMADA, Y.; MUTO, A.; INOKUCHI, H.; IKEMURA, T.; MATSUMOTO, T.; SASAKI, T.; ITOH, T. Rice annotation project database (RAP-DB): An integrative and interactive database for rice genomics. **Plant and Cell Physiology**, Kyoto, v. 54, n. 2, p. 1-11, 2013.

SAKI, T. Development in Rice Genome Research Based on Accurate Genome Sequence. **International Journal Plant Genomics**, Hindawi, v., p. 1-9, 2008.

SEARLE, S. R. **Linear models for unbalanced data.** Nova York: John Wiley & Sons, Inc., 1987.

SHENDURE, J.; JI, H. Next- generation DNA sequencing. **Nature Biotechnology**, New York, v.26, n.10, p. 1135-1145, 2008.

SOHRABI, M.; RAFII, M. Y.; HANAFI, M. M.; AKMAR, A. S. N.; LATIF, M. A. Genetic Diversity of Upland Rice Germplasm in Malaysia Based on Quantitative Traits. **The Scientific World Journal**, v. 2012, p. 1-9, 2012.

UPHOFF, N.; FASOULA, V.; ISWANDI, A.; KASSAM, A.; THAKUR, A. K. Improving the phenotypic expression of rice genotypes: Rethinking “intensification” for production systems and selection practices for rice breeding. **The Crop Journal**, Amsterdam, v. 3, n. 3, p. 174-189, 2015.

VIEIRA, J. **Caracterização morfológica e molecular do banco de germoplasma de arroz irrigado (Oryza sativa L.) da EPAGRI**. 2007. 115 f. Dissertação (Mestrado em Recursos Genéticos Vegetais) – Centro de Ciências Agrárias, Universidade Federal de Santa Catarina, 2007.

WANG, M.; WANG, M.; YU, Y.; HABERER, G.; MARRI, P. R.; FAN, C.; GOICOECHEA, J. L. ZUCCOLO, A.; SONG, X.; KUDRNA, D.; AMMIRAJU, J. S. S.; COSSU, R. M.; MALDONADO, C.; CHEN, J.; LEE, S.; SISNEROS, N.; BAYNAST, K.; GOLSER, W.; WISSOTSKI, M.; KIM, W.; SANCHEZ, P.; NDJIONDJOP, M-N.; SANNI, K.; LONG, M.; CARNEY, J.; PANAUD, O. The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. **Nature Genetics**, Nova York, v. 46, n. 9, p. 982-988, 2014.

WANG, Y.; XIONG, G.; HU, J.; JIANG, L.; YU, H.; XU, J.; FANG, Y.; ZENG, L.; XU, E.; XU, J.; YE, W.; MENG, X.; LIU, R.; CHEN, H.; JING, Y.; WANG, Y.; ZHU, X.; LI, J.; QIAN, Q. Copy Number Variation at the GL7 Locus Contributes to Grain Size Diversity in Rice. **Nature Genetics**, v. 47, n. 8, p. 944–948, ago. 2015.

XIE, Y.; MCNALLY, K.; LI, C. Y.; LEUNG, H.; ZHU, Y. Y. A High-throughput Genomic Tool: Diversity Array Technology Complementary for Rice Genotyping. **Journal of Integrative Plant Biology**, v. 48, n. 9, p. 1069-1076, Set. 2006.

XING, Y.; ZHANG, Q. Genetic and Molecular Bases of Rice Yield. **Annual Review of Plant Biology**, v. 61, p. 421–442, 2010.

YU, J.; BUCKLER, E. S. Genetic Association mapping and genome organization of maize. **Current Opinion in Biotechnology**, Amsterdam, v. 17, p. 155–160, 2006.

ZENG, Z.B. Precision mapping of quantitative trait loci. **Genetics** Baltimore, v. 136, n. 4, p. 1457-1468, Apr. 1994.

ZHOU, L.; LIU, S.; WU, W.; CHEN, D.; ZHAN, X.; ZHU, A.; ZHANG, Y.; CHENG, S.; CAO, L.; LOU, X.; XU, H. Dissection of genetic architecture of rice plant height and heading date by multiple-strategy-based association studies. **Scientific Reports**, Londres, v. 6, n. 7, p. 29718, 2016.

3 ARTIGO CIENTÍFICO: ANÁLISE COMPARATIVA DOS MÉTODOS DE AVANÇO BULK E SSD NA IDENTIFICAÇÃO DE QTLs PARA PRODUTIVIDADE DE GRÃO DE ARROZ NO CRUZAMENTO EPAGRI 108 X IRAT 122

3.1 INTRODUÇÃO

O arroz (*Oryza sativa* L.) está entre as mais importantes espécies cultivadas, e para que sua produção consiga acompanhar o contínuo crescimento da demanda mundial devem-se intensificar pesquisas em melhoramento genético visando ao aumento, da produtividade de grãos (Neves et al., 2017).

As cultivares modernas de arroz são obtidas por cruzamentos com genitores próximos geneticamente, com o objetivo de conservar as combinações favoráveis acumuladas por muitas gerações de melhoramento, e realizar a seleção das progênes superiores. Contudo, o cruzamento entre genitores aparentados tem limitado o ganho genético anual para a produtividade de linhagens elites dos programas de melhoramento genético, além de aumentar a suscetibilidade aos ataques de pragas e doenças (Huang et al., 2015).

Como estratégia para incrementar a produtividade na cultura do arroz, a exploração da variabilidade genética proveniente de variedades silvestres ou de materiais introduzidos é uma boa resposta para mudanças nesse cenário (Uphoff et al., 2015). Uma alternativa que se apresenta como importante solução dessa crescente demanda é a obtenção de cultivares híbridas. Programas de melhoramento têm utilizado tal estratégia para atingir patamares cada vez maiores de produtividade, explorando o vigor híbrido ou heterose (Uphoff et al., 2015).

Diante de muitos cruzamentos que podem ser realizados para ampliar a base genética dos programas de melhoramento, o uso da análise de cruzamento dialélico tem sido a melhor e mais completa informação sobre o comportamento das linhagens em combinações híbridas, por permitir a estimação de diferentes componentes da variância genotípica, podendo obter a capacidade combinatória dos genitores em uma série de combinações e o desempenho das melhores ou piores combinações híbridas, e a verificação da ação gênica predominante em um grupo de genótipos, proporcionando informações sobre o controle genético das características avaliadas, auxiliando o melhorista na escolha da melhor estratégia de melhoramento (Ramalho et al., 2012).

A capacidade combinatória permite inferir que os parentais diferem em relação à capacidade de se combinarem, além de ser um indicativo de que no controle do caráter ocorre a predominância dos efeitos aditivos, enquanto a capacidade específica mostra que os cruzamentos são heterogêneos, por apresentarem componentes diferentes do esperado com base na capacidade combinatória de seus parentais, avaliando a contribuição dos efeitos não aditivos, de dominância ou epistasia, na expressão dos caracteres (Ramalho et al., 2012).

Para o desenvolvimento de novas cultivares em espécies autógamas, como o arroz, diferentes métodos de condução de populações de melhoramento podem ser utilizados, os quais variam de acordo com as características a serem melhoradas e os genitores utilizados, tendo como um dos principais objetivos obter linhas homozigóticas com desempenho produtivo superior às cultivares comerciais (Castanheira & Santos, 2004).

Os métodos de condução de populações segregantes mais utilizados são o genealógico; o descendente de uma única semente (SSD - *Single Seed Descent*); e o de população (Bulk). As respectivas descrições e implicações relacionadas à informação obtida sobre as linhagens, alterações nas frequências gênicas, tamanho populacional adequado, influência ou não da seleção natural, dentre outras, são discutidos por Borém et al. (1997) e Ramalho et al. (2012). Considerando os tipos de populações que são geradas pelos programas de melhoramento têm-se as linhas puras recombinantes (RILs), que detêm elevado grau de homozigose e são utilizadas para estudos de genética e genômica, bem como para desenvolver novas cultivares. Esse tipo de população de melhoramento passa por processos de caracterização e avaliação a campo, estimação de média e parâmetros genéticos, avaliação da interação com ambiente e também análise de diversidade genética por marcadores moleculares (Ramalho et al., 2012).

Os detalhes do perfil gênico de acessos com a utilização de ferramentas moleculares torna-se uma tecnologia importante por permitir a detecção direta de variações no genoma, que promovem variações fenotípicas. Assim, possibilita que a seleção de genitores para programas de melhoramento passe a ser baseado na variação no genoma de interesse, como em trabalhos de Rizzi (2017) e Monzón (2015), na cultura do arroz.

A produtividade em arroz foi incrementada pelo desenvolvimento de cultivares melhoradas para o índice de colheita e a arquitetura da planta através do uso de genes de

plantas de porte baixo, além do desenvolvimento de cultivares híbridas através da exploração da heterose (Xing & Zhang, 2010).

O arroz consiste em importante planta modelo para estudos genômicos, pois possui o menor genoma entre as espécies de cereais (Kawahara et al., 2013). Esforços na pesquisa genômica funcional em arroz também têm sido coordenados com um enorme progresso (Han & Zhang, 2008), possibilitando a identificação de um grande número de genes, e sua posterior clonagem e caracterização funcional. Esse progresso tem ajudado bastante na compreensão dos controles genéticos e moleculares e processos biológicos associados aos caracteres de produtividade em arroz (Valluru et al., 2014).

Os genes que controlam os caracteres relacionados à produtividade em arroz podem ser divididos em duas classes: (a) genes para os processos biológicos básicos, sendo que tais genes podem causar efeitos deletérios sobre as plantas como redução do desenvolvimento, anormalidades na morfologia e taxas muito baixas de produtividade, como exemplo, os genes relacionados ao perfilhamento e ao desenvolvimento de panículas; e (b) genes diretamente envolvidos em caracteres de herança quantitativa, incluindo alelos de perda de função. Exemplos dessa classe são aqueles genes relacionados ao número de grãos por panícula, tamanho de panícula e dimensão do grão (Valluru et al., 2014).

A produtividade de grãos de uma planta de arroz é determinada por três caracteres componentes: número de panículas por metro, número de grãos por panícula e peso de grãos. Número de panículas é dependente da capacidade da planta para produzir perfilhos, a produtividade é geralmente contribuída pelos perfilhos primários e alguns perfilhos secundários iniciais, enquanto perfilhos secundários tardios e perfilhos terciários exercem pouca contribuição, embora eles também consumam nutrientes e água. Em programas de melhoramento, há uma tendência na busca por menor número de panículas, porém com maior tamanho, a fim de alcançar uma melhor produtividade (Xing & Zhang, 2010).

Muitos estudos básicos procuraram identificar regiões genômicas relacionadas à produtividade em arroz. Entre os QTLs que controlam os caracteres produtivos e têm sido clonados e caracterizados, podem ser destacados o LAX1, APO1, GN1a, DEP1, Ghd7 e Ghd8, GS3 e GL3. Quanto ao número de grãos em arroz, o gene LAX1 foi identificado codificando o fator de transcrição bHLH, que está envolvido na formação de todos os tipos de meristemas axilares (Komatsu et al., 2003); o gene APO1 também foi identificado

codificando proteínas F-box, regulando positivamente o número de espiguetas (Ikeda et al., 2007); e o gene GN1a demonstrou o papel importante no aumento da produtividade (Xie et al., 2015).

Relacionado ao número de grãos em arroz, no trabalho de Huang et al. (2009), o gene DEP1 foi identificado codificando uma proteína contendo um domínio PEBP, evidenciando três caracteres fenotípicos: panícula densa, grande número de grãos por panícula e panícula ereta. O gene Ghd8 foi identificado como um QTL relacionado à produtividade de grãos e envolvido com a ramificação da panícula (Yan et al., 2011). Um exame mais detalhado nas panículas revelou que o gene Ghd7 altera o número de ramificações primárias e secundárias, codificando uma proteína com o domínio HD1 (Xue et al., 2008). Em estudos mais recentes, o gene Ghd7 foi identificado em análise GWAS em arroz relacionado ao potencial produtivo (Lu et al., 2012; Yonemaru et al., 2014), e de acordo com Wang et al. (2015), essa mesma região genômica também foi responsável pela variação do conteúdo de clorofila, evidenciando sua importância em relação à produtividade.

O primeiro gene de grande importância identificado como relacionado ao peso de grão foi o GS3. Este QTL foi detectado em torno da região centromérica do cromossomo 3, contribuindo para o aumento da produtividade (Fan et al., 2006). Mao et al. (2010) identificaram mais dois alelos desse gene, classificando, assim, quatro tipos de alelos GS3, sendo que a proteína putativa é composta por quatro tipos de domínios: domínio OSR, domínio de transmembrana, domínio TNFR/NGFR rico em cisteína e o domínio VWFC. Análises revelaram que estes domínios apresentaram diferentes funções na regulação do peso do grão. Estudos mais recentes também demonstraram a associação dessa região genômica com a produtividade, como por exemplo na análise GWAS em arroz de Yang et al. (2014), no mapeamento de QTL de Xu et al. (2015) e também no trabalho de Liang et al. (2016) que identificou 10 genes associados à produtividade, com grande destaque para o GS3 que foi significativo para mais de um ambiente testado.

O tamanho de grãos também está relacionado ao aumento da produtividade em arroz, e o QTL GLW7 foi identificado codificando o fator de transcrição OsSPL13 (Si et al., 2016), enquanto que o QTL GS2 foi identificado codificando o fator de regulação do crescimento OsGRF4 (Hu et al., 2015), e o QTL WFP foi identificado codificando o fator de transcrição OsSPL14, que também promove as ramificações da panícula (Miura et al., 2010). Nos trabalhos de Qi et al. (2012), Hu et al. (2012), Zhang et al. (2012), Xu et al.

(2015) e Zeng et al. (2016), o QTL GL3 foi relacionado com a fosforilação de proteínas nas espiguetas para acelerar a divisão celular e, assim, promover maiores produtividades.

O objetivo desse trabalho foi comparar os métodos de condução de população segregante SSD e Bulk quanto a capacidade de identificar QTLs para os caracteres produtividade de grãos e altura de plantas das linhagens puras recombinantes (RILs) derivadas do cruzamento Epagri 108 x Irat 122.

3.2 MATERIAL E MÉTODOS

3.2.1 Avanço das Populações

O cruzamento (Epagri 108 x Irat 122) advém de um dialelo realizado com a finalidade de obter cruzamentos que derivariam progênie com alto potencial produtivo e geneticamente divergentes, superiores em uma ou mais características aos genótipos elites do programa de melhoramento de arroz da Embrapa. Esse cruzamento foi escolhido dentre as 66 combinações do dialelo, por possuir plantas com maior produtividade média, e ser um cruzamento inter-subespecífico, envolvendo uma cultivar indica (Epagri 108) e uma cultivar japônica (Irat 122). Os 12 genótipos selecionados como genitores do dialelo foram escolhidos a partir da caracterização agrônômica e molecular dos 550 acessos constituintes da Coleção Nuclear de Arroz da Embrapa (CNAE), e cujos critérios foram a alta produtividade, e pertencentes a diferentes estratos (variedades tradicionais, melhorados no Brasil e melhorados no exterior). Os 12 genótipos foram cruzados no esquema de dialelo completo, sem recíprocos, no ano de 2005, e foram então obtidas 66 combinações F_1 , que foram autofecundados para obtenção da geração F_2 e avançados até F_7 .

A população Epagri 108 x Irat 122 foi avançada pelo método Bulk a partir da geração F_2 , no ano de 2006. Para o avanço de geração eram retiradas amostras de 500 sementes, e plantadas novamente no ano seguinte, até a geração F_7 . Por fim, foi realizada a abertura das famílias em $F_{7;8}$; as plantas foram dispostas em linhas de cinco metros, sendo todas colhidas individualmente.

A população Epagri 108 x Irat 122 também foi avançada pelo método SSD (descendente de semente única) no ano de 2006. Partindo de uma população de 500 sementes em F_2 , dez sementes de cada planta F_2 foram semeadas em copos plásticos individuais, e após 20 dias da semeadura, cinco mudas foram transplantadas no campo.

Destas cinco plantas, apenas uma foi escolhida aleatoriamente para avançar à geração seguinte. Todo o processo foi repetido até a geração F₇.

Para a condução dos dois anos de experimentos (2017 e 2018) na fazenda Palmital em Goianira-GO foram retiradas aleatoriamente 158 sementes das linhagens Bulk e SSD da população Epagri 108 x Irat 122, para comporem os experimentos.

3.2.2 Avaliação das Famílias F₇ e F₈

As populações SSD e Bulk do cruzamento Epagri 108 x Irat 122 foram avaliadas em experimento de campo, em duas safras agrícolas (2016/2017 e 2017/2018). As 158 RILs F₇ e F₈ de cada método, juntamente com 8 Testemunhas (dois genitores e as cultivares Araguaia, Maninjau, Irga 417, BRS Catiana, BRS Tropical, ITA Puitá CL) foram avaliadas no delineamento experimental de látice quadrado 18x18, com parcelas de 4 linhas de 3 metros, com duas repetições.

Os caracteres avaliados foram produtividade de grão (PG) e altura de plantas (ALT). Para produtividade de grão, foram colhidas todas as plantas das duas linhas centrais da parcela, que foram pesadas em balança de precisão, e os valores em gramas foram convertidos para kg ha⁻¹. O caráter altura das plantas em (cm) foram obtidos pela tomada das alturas de cinco plantas dentro das parcelas por meio de uma fita métrica, sendo obtido então o valor médio de altura para a parcela.

3.2.3 Análises Estatísticas Via Modelos Mistos

O modelo estatístico adotado para análise dos dados foi tipo misto:

$$Y_{ijkmn} = \mu + g_{i(m)} + r_j + b_{k(j)} + p_m + t_n + \varepsilon_{ijkmn}$$

em que:

Y_{ijkmn} é o valor observado do genótipo i , na repetição j , dentro do bloco k , e μ é a constante inerente a todas às observações, que representa a média geral do ensaio. Como efeitos aleatórios no modelo tem-se: $g_{i/m}$ é o efeito do genótipo i dentro do grupo m , r_j é o efeito da repetição j , $b_{k/j}$ é o efeito do bloco k dentro da repetição j , e ε_{ijkmn} é o efeito dos erros associados. Da mesma forma, como efeitos fixos do modelo tem-se: p_m que é o efeito de grupos, que contém o genótipo i e t_n que é o efeito de Testemunhas n .

Os efeitos do modelo foram estimados com o pacote Sommer do software R (Bates et al., 2015; R Core Team, 2016). Os componentes de variância foram estimados pelo método REML (*restricted maximum likelihood*). Isso permitiu obter a melhor estimativa não viesada para os efeitos fixos, ou seja, os eBLUES (*empirical best linear unbiased estimator*) e a melhor predição linear não viesada para os efeitos aleatórios, os eBLUPs (*empirical best linear unbiased predictor*). Foi testada a significância de cada um dos efeitos dentro do modelo, tanto nas análises individuais, como na análise conjunta, através da análise de deviance. A análise foi realizada desconsiderando cada um dos efeitos em teste (repetição, bloco, genótipo) e subtraindo a deviance do modelo sem o referido efeito da deviance do modelo completo. A significância de cada efeito foi verificada a partir do teste de razão de verossimilhança (LRT), usando a distribuição de χ^2 a 1% e 5% de probabilidade, com um grau de liberdade.

3.2.4 Estimativas dos Parâmetros Genéticos

A variância fenotípica foi estimada por:

$$\hat{\sigma}_p^2 = \hat{\sigma}_g^2 + \left(\frac{\hat{\sigma}_e^2}{k}\right)$$

A herdabilidade, no sentido amplo ao nível de médias de progênies, na análise individual por ano foi estimada por:

$$\hat{h}^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2 + \left(\frac{\hat{\sigma}_e^2}{k1}\right)}$$

E a análise conjunta foi estimada por:

$$\hat{h}_2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2 + \left(\frac{\hat{\sigma}_e^2}{k2}\right)}$$

em que: k1: representa o número de repetições no primeiro ano; k2: representa o produto do número de repetições, dos dois anos. Para avaliação da precisão e da qualidade dos experimentos foram considerados:

- Coeficiente de variação experimental (CVe)

$$\widehat{CV}_e = \frac{\hat{\sigma}_e^2}{m} \times 100$$

- Acurácia seletiva:

$$\hat{r}_{gg} = [1 - \frac{1}{1 + bCV_r^2}]^{1/2}$$

3.2.5 Análises Moleculares

3.2.5.1 Extração e Quantificação do DNA Genômico

O DNA genômico total foi obtido pela extração realizada a partir da coleta de folhas jovens de cada genótipo 2 genitores, 158 RILs avançadas em Bulk, e 158 RILs avançadas por SSD) utilizando o kit comercial de extração Biopur (DNeasy 96 Plant Kit (Qiagen). As amostras foram quantificadas, diluídas, distribuídas em placas e enviadas para a empresa *Diversity Arrays* (Austrália), onde foi realizada a análise DArT-seq.

3.2.5.2 Mapa de Ligação e Análise de QTL

O mapa de ligação de cada população RILs (SSD e Bulk) foi obtido pelo programa MapDisto versão 1.7.7 (Lorieux, 2014), utilizando o LOD ≥ 3 e frequência máxima de recombinação $\theta = 0,3$. As frações de recombinação foram transformadas em distâncias genéticas (cM) pela função de Kosambi (1943).

A análise de QTL foi realizada pelo software estatístico R utilizando o pacote R/qtl, considerando os dados de 2.115 marcadores SNPs nas 146 RILs SSD e 2.354 marcadores SNPs e DArTs nas 108 RILs Bulk, juntamente com as médias ajustadas das RILs(Blups) para os caracteres produtividade grão e altura de plantas.

O mapeamento de QTL considerado foi o mapeamento múltiplo, que permite identificar mais de um QTL no intervalo considerado e ainda inclui o teste de epestasias entre QTLs. O valor de LOD considerado foi igual ou maior a três.

A nomenclatura dos QTLs seguiu as diretrizes descritas por McCouch et al.(1997).

3.2.5.3 Desequilíbrio de ligação (DL) e blocos haplotípicos

As estimativas da extensão de DL foram calculadas pelo método usual (r^2) utilizando o pacote LDcorSV do R (Desrousseaux et al., 2013). Em contrapartida, a Matriz

de Relacionamento Genético (GRM) foi estimada pelo algoritmo proposto por Yang et al., (2010), utilizando o pacote GCTA (Yang et al., 2011). O decaimento DL foi explicado pelo modelo não linear (Hill & Weir, 1988) e ajustado para a função “nls” no programa R.

Os blocos haplotípicos foram identificados utilizando o software Haploview (Barret et al., 2005), com porcentagem mínima de genotipagem dos indivíduos $\geq 75\%$ e $R^2 \geq 0,8$.

3.2.5.4 Análises de componentes principais (PcA) e Estruturação DArWin

Para a análise de PcA foram utilizados 2.115 marcadores para as 146 RILs SSD, e para as Bulk foram utilizados 2.354 marcadores e 108 RILs, considerando também os dois parentais em ambas populações de RILs.

Para a análise DArWin foi considerada a matriz de distância Dice, que levou em consideração todos os indivíduos (146 RILs SSD e 108 RILs Bulk), juntamente com os marcadores, sendo retirados 50% dos dados de genotipagem faltantes.

3.2.5.5 Anotação dos genes

Os SNPs identificados foram estudados quanto ao efeito teórico devido à sua posição no genoma, utilizando o software SnpEff 4.3 (Cingolani et al., 2012). A identificação de genes candidatos subjacentes aos QTL de interesse foi obtida através de busca em dois bancos de dados do genoma do arroz (Rice Genome Annotation Project e Rice FRENDA) e na literatura existente, visando à procura de regiões que estavam dentro ou perto dos picos dos QTL significativos.

Com a utilização do banco de dados QTARO (<http://qtaro.abr.affrc.go.jp/>) foram identificados os QTLs já referenciados na literatura para a cultura do arroz.

3.3 RESULTADOS

3.3.1 Análise de Deviance e Significância dos Efeitos do Modelo

Os resultados da análise de deviance para os efeitos de genótipos dentro dos grupos (método Bulk, método SSD e Testemunhas) e seus respectivos componentes de variância, para o primeiro ano (2017) e segundo ano (2018) de experimento e análise

conjunta dos dois anos, para os caracteres produtividade de grãos (PG) e altura de plantas (ALT) estão descritos na (Tabela 1).

O modelo estatístico adotado mostrou-se adequado pela análise de deviance, tanto para as análises individuais, como para a análise conjunta de ambos os caracteres avaliados. Verificou-se pela análise que para o caráter PG os efeitos de genótipos do primeiro ano (2017), somente o método SSD foi significativo ($p < 0,05$) e, conseqüentemente, seu componente de variância genética, enquanto que o método Bulk e as Testemunhas não apresentaram significância. Já para o segundo ano (2018) os três grupos de genótipos (método Bulk, método SSD e Testemunhas) foram significativos ($p < 0,01$) bem como os seus componentes de variância genética. Para a análise conjunta dos dois anos, apenas o SSD foi significativo ($p < 0,01$). A análise de deviance evidenciou, assim, a presença de variabilidade genética entre as linhagens testadas (Tabela 1).

Para o caráter ALT, no primeiro ano, os efeitos foram significativos para Bulk e SSD ($p < 0,05$ e $p < 0,01$, respectivamente). Já para o segundo ano todos os grupos de RILs apresentaram efeitos significativos ($p < 0,05$), bem como para a análise conjunta, significativa a ($p < 0,05$) para todos os grupos. Isto indica uma grande variabilidade genética para o caráter, refletindo ampla variação de alturas para as RILs de cada grupo (Tabela 1).

Tabela 1. Resumo da análise de deviance para os componentes de variância do caráter Produtividade de Grãos (PG) em kg ha^{-1} e Altura de Plantas (ALT) em cm, dos grupos de genótipos de RILs Bulk, SSD e Testemunhas, para o 1º ano (2017), 2º ano (2018) de experimento e a análise conjunta (dois anos).

	Produtividade de Grão - 1º Ano (2017)			Produtividade de Grão - 2º Ano (2018)			Produtividade de Grão - Análise Conjunta		
	LRT	p.value	σ^2_g	LRT	p.value	σ^2_g	LRT	p.value	σ^2_g
Efeitos									
RILs	2.520526	0.1123731 ^{ns}	901724.8	39.68001	2.991714e-10*	281726.1	2.375834	0.1232258 ^{ns}	221473.2
Bulk									
RILs	3.923106	0.0476268*	1144089	116.8518	0*	697206.8	9.166177	0.002465291*	459966.8
SSD									
Test	0.06855229	0.7934564 ^{ns}	603289.2	16.79586	4.162406e-05*	911485	1.252687	0.2630399 ^{ns}	806471.6
	Altura de plantas - 1º Ano (2017)			Altura de plantas - 2º Ano (2018)			Altura de Plantas - Análise Conjunta		
	LRT	p.value	σ^2_g	LRT	p.value	σ^2_g	LRT	p.value	σ^2_g
Efeitos									
RILs	7.253815	0.007075054**	62.33939	18.98905	1.31471e-05**	43.51756	194.7631	0*	156.4166
Bulk									
RILs	40.12756	2.379092e-10**	188.3271	132.7922	0*	195.6609	369.3151	0*	279.8536
SSD									
Test	1.092872e-09	0.9999736 ^{ns}	0	5.174348	0.02292277*	101.3849	3.984716	0.04591484*	81.72965

*Test=Testemunhas; σ^2_g = Variância Genética; * significativo a 5%; ** significativo a 1% e ^{ns} não significativo pelo teste LRT.

3.3.2 Estimativa dos Parâmetros Genéticos e Estatísticos

Os resultados das médias ajustadas pelos BLUEs (efeitos fixos) para o caráter produtividade de grãos, o grupo Bulk se destacou com as maiores médias ($4.953,25 \text{ kg.ha}^{-1}$; $3.225,59 \text{ kg.ha}^{-1}$ e $4.189,42 \text{ kg.ha}^{-1}$), respectivamente para o 1º e 2º ano e análise conjunta. O grupo SSD teve o segundo melhor desempenho produtivo, com médias ($4.284,43 \text{ kg.ha}^{-1}$;

2.735,15 kg.ha⁻¹ e 3.614,56 kg.ha⁻¹) para 1º e 2º ano e análise conjunta. As Testemunhas apresentaram médias abaixo dos grupos Bulk e SSD, sendo (3.158,01 kg.ha⁻¹; 3.369,85 kg.ha⁻¹ e 3.456,70 kg.ha⁻¹) para 1º e 2º ano e análise conjunta (Tabela 2).

O caráter ALT o grupo Bulk apresentou maiores médias para ambos os anos avaliados e análise conjunta, posteriormente o grupo SSD, e em seguida as Testemunhas (Tabela 2).

Tabela 2. Médias ajustadas (Blue - efeito fixo) dos grupos Bulk, SSD e Testemunhas, para 1º ano (2017); 2º ano (2018) de experimento e análise conjunta (dois anos), para os caracteres Produtividade de Grão (PG) em kg.ha⁻¹ e Altura de Plantas (ALT) em cm.

Grupos	Produtividade de Grão (PG)			Altura de Plantas (ALT)		
	1º Ano (2017)	2º Ano (2018)	Análise Conjunta	1º Ano (2017)	2º Ano (2018)	Análise Conjunta
Bulk	4.953,25	3.225,59	4.189,42	125,37	121,27	128,48
SSD	4.284,34	2.735,15	3.614,56	114,94	111,59	118,13
Test	3.158,01	3.369,85	3.456,70	104,26	104,33	108,42

*Test: grupo de Testemunhas. *Médias ajustadas (Blue - efeito fixo) foram Significativas a 5% probabilidade pelo teste F.

O parâmetro variância fenotípica (σ^2_f) para o caráter PG apresentou a maior variação dentro do grupo das Testemunhas em ambos os anos e análise conjunta. Para o primeiro ano o grupo SSD apresentou variação maior que Bulk, o que também persistiu no segundo ano e análise conjunta. Para ALT o grupo das Testemunhas foi o com maior variação no 1º ano, seguindo do grupo SSD e por último Bulk. O segundo ano o grupo SSD teve a maior variação, posteriormente as Testemunhas e Bulk. Na análise conjunta SSD continuou com maior variação, seguido de Bulk e Testemunhas (Tabela 3).

A grande magnitude de (σ^2_g) indica que existe variabilidade genética dentro da população de RILs avaliadas nos métodos Bulk e SSD e nas Testemunhas. O LRT na análise de deviance comprova a significância de σ^2_g nos experimentos de avaliação das RILs (Tabelas 1 e 3). No primeiro ano de experimento para o caráter PG, apenas o grupo SSD apresentou o componente de σ^2_g significativo ($p < 0,05$) na análise LRT, indicando uma maior variabilidade genética dentro do grupo de 158 RILs de SSD. No segundo ano de experimento, para o caráter PG as RILs dos dois métodos e as Testemunhas foram significativos pelo teste LRT, porém as Testemunhas apresentam maiores σ^2_g , seguida de RILs-SSD e por último, RILs-Bulk. Para a análise conjunta apenas o grupo SSD foi significativo. Para o caráter ALT, no primeiro ano de ensaio, ambos Bulk e SSD apresentaram efeito significativo de σ^2_g . No segundo ano todos os grupos foram significativos, assim como para a análise conjunta (Tabela 3).

Nesse trabalho observou-se que a variância genética entre as RILs do método SSD foi maior do que no método Bulk para ambos os anos avaliados e análise conjunta do

caráter PG. Coerentemente, também se observou maior h^2 em SSD de (0,23; 0,77; 0,29) do que em Bulk (0,19; 0,57; 0,17) respectivamente para 1º ano e 2º ano e análise conjunta. A h^2 para o caráter ALT também foi maior em SSD (0,57; 0,78; 0,85) para 1º ano e 2º ano, e análise conjunta quando comparado ao Bulk e as Testemunhas (Tabela 3).

A precisão experimental tem grande efeito na estimação dos parâmetros genéticos em geral de uma população. Os coeficientes de variação experimental (CVe) do trabalho indicam que os dois anos de experimento tiveram de boa a moderada precisão experimental para ambos os caracteres avaliados (Tabela 3). O CVe do 1º ano de experimento foi alto (59,46), já no 2º ano foi dentro do esperado com valor de (21,30). A acurácia do presente trabalho foi de moderada a alta, para ambos os caracteres avaliados, sendo os experimentos considerados como de boa precisão (Tabela 3).

Tabela 3. Parâmetros Genéticos: variância fenotípica (σ^2_f); variância genotípica (σ^2_g), herdabilidade (h^2), acurácia seletiva (r_{gg}); variância ambiental (σ^2_e) e coeficiente de variação ambiental (CVe); dos grupos Bulk, SSD e Testemunhas (Test) para 1º ano de experimento (2017), 2º ano de experimento (2018) e análise conjunta dos dois anos/safras; para os caracteres Produtividade de Grão (PG) e Altura de Plantas (ALT).

PRODUTIVIDADE DE GRÃO (PG)									
Parâmetros	1º Ano(2017)			2º Ano (2018)			Análise Conjunta		
	Bulk	SSD	Test	Bulk	SSD	Test	Bulk	SSD	Test
σ^2_f	487.115,6	902.596,3	111.687,4	459.895,33	484.131,75	430.051,77	422.991,1	153.889,18	188.539,66
σ^2_g	901.724,8	1.144.089,0	603.289,2	2.817.261,0	6.972.068,0	911.485,0	2.214.732,0	4.599.668,0	806.471,6
h^2	0,19	0,23	0,14	0,57	0,77	0,81	0,17	0,29	0,42
r_{gg}		0,48			0,85			0,49	
σ^2_e		739.445,7			410.779,0			431.570,0	
CVe (%)		59,46			21,30			52,25	

ALTURA DE PLANTAS (ALT)									
Parâmetros	1º Ano (2017)			2º Ano (2018)			Análise Conjunta		
	Bulk	SSD	Test	Bulk	SSD	Test	Bulk	SSD	Test
σ^2_f	201.607,2	327.595,0	139.267,90	974.892,6	249.632,6	155.356,6	248.828,6	372.265,6	174.141,7
σ^2_g	623.393,9	1.883.271,0	0	435.175,6	1.956.609,0	1.013.849,0	1.564.166,0	2.798.536,0	817.296,5
h^2	0,30	0,57	0	0,45	0,78	0,65	0,77	0,85	0,63
r_{gg}		0,73			0,85			0,87	
σ^2_e		278.535,8			107.943,4			184.824,0	
CVe (%)		14,09			9,09			14,09	

Teste: REML/BLUP

3.3.3 Valores Genotípicos das Linhagens (RILs) Bulk e SSD

3.3.3.1 Caráter Produtividade de Grão (PG)

Com o método REML foi possível prever tanto os componentes de variância quanto os valores genotípicos de cada linhagem pelos BLUPs, livres de interação ambiental. A produtividade de grãos é um dos caracteres de maior importância para o melhoramento da cultura do arroz, além de ser um dos mais desafiadores em função da sua maior complexidade.

O método Bulk apresentou uma variação de 4.010,80 a 5.815,40 kg.ha⁻¹ no primeiro ano para as 158 linhagens (Figuras 1 e 2). As 158 RILs de SSD tiveram variação

de produtividade de 3.321,80 a 8.096,30 kg.ha⁻¹, e mostram a ocorrência de indivíduos transgressivos, que superaram a média dos próprios genitores (Figuras 1 e 3). O grupo Testemunhas variou de 2.754,30 kg.ha⁻¹ a 3.643,73 kg.ha⁻¹, com o genitor Epagri 108 apresentando produtividade de 3.256,40 kg.ha⁻¹, enquanto que o genitor Irat 122 foi a única parcela perdida no primeiro ano de ensaio (Figura 1).

No 2º ano de avaliação, o grupo SSD teve variação de 1.089,81 a 4.790,07 kg.ha⁻¹. Já o grupo Bulk variou de 2.277,18 kg.ha⁻¹ a 4.304,34 kg.ha⁻¹ (Figuras 1, 2 e 3), enquanto que as Testemunhas variaram de 1.888,69 kg.ha⁻¹ a 4.534,5 kg.ha⁻¹. O genitor Epagri 108 teve produtividade de 4.322,35 kg.ha⁻¹ e o Irat122 teve 3.313,89 kg.ha⁻¹. Epagri108 foi o único genitor que superou algumas das linhagens mais produtivas, tanto do grupo SSD quanto do Bulk. Porém as maiores produtividades do experimento foram encontradas nas RILs (Figura 1). No 2º ano houve redução da produtividade média das RILs dos dois grupos quando comparada ao 1º ano (Figuras 2 e 3).

Considerando a análise conjunta, a produtividade de grãos variou de 2.855,16 kg.ha⁻¹ a 5.985,11 kg.ha⁻¹ para grupo SSD, de 3.706,96 a 4.610,25 kg.ha⁻¹ para o grupo Bulk e de 2.677,21 kg.ha⁻¹ a 4.304,64 kg.ha⁻¹ para as Testemunhas. Já os genitores apresentaram a produtividade de 3.456,70 kg.ha⁻¹ para Epagri 108 e de 3.813,80 kg.ha⁻¹ para Irat 122 (Figuras 1, 2 e 3).

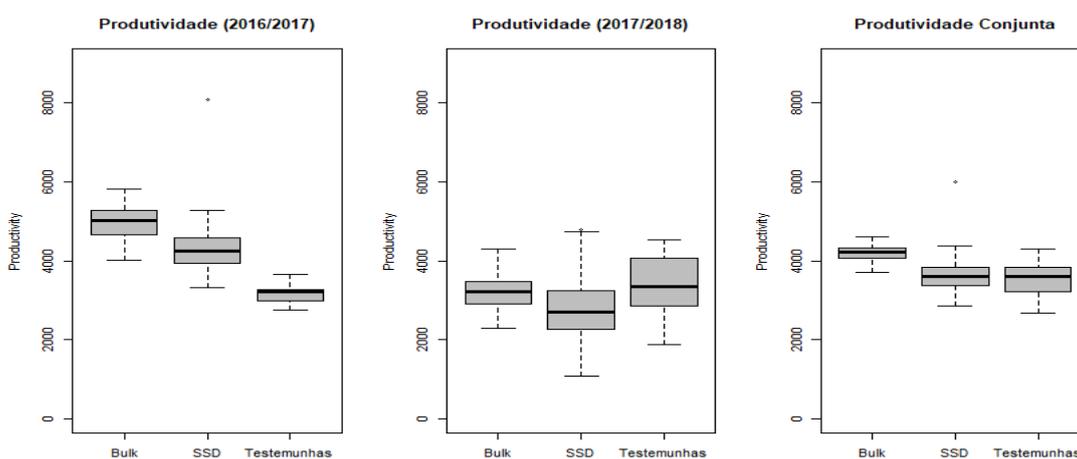


Figura 1. Boxplots para o caráter produtividade de grão (PG) em kg.ha⁻¹ para as RILs (BULK, SSD e TESTEMUNHAS) para o 1º ano de experimento (safra 2016/2017), para o 2º ano de experimento (safra 2017/2018) e a análise conjunta dos dois anos.

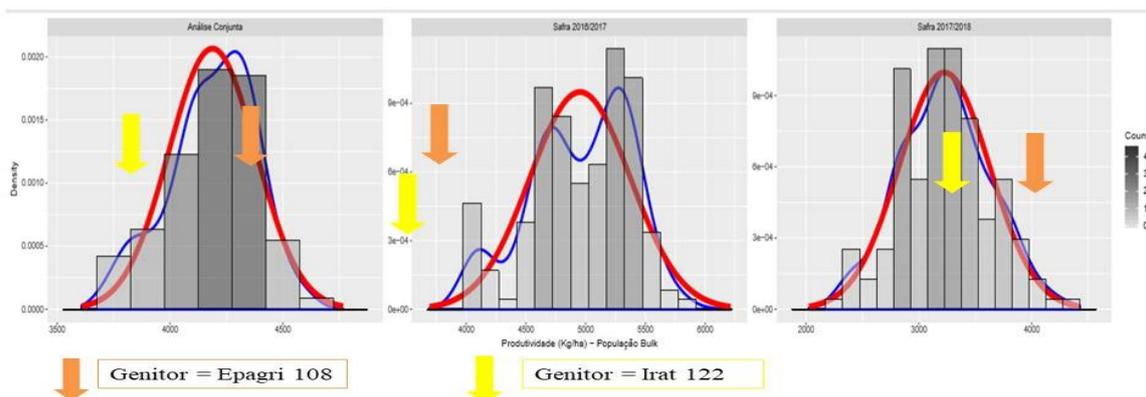


Figura 2. Histogramas do grupo Bulk para o caráter produtividade de grão em $\text{kg}\cdot\text{ha}^{-1}$ no primeiro ano de ensaio (safra 2016/2017): a linha vermelha indica a curva normal; linha azul indica curva de densidade. Valor genético dos genitores: Epagri 108= $3.256,35 \text{ kg}\cdot\text{ha}^{-1}$; Irat 122= $0,00 \text{ kg}\cdot\text{ha}^{-1}$. Segundo ano de ensaio (safra 2017/2018), valor genético dos genitores: Epagri 108= $4.322,35 \text{ kg}\cdot\text{ha}^{-1}$; Irat 122= $3.313,89 \text{ kg}\cdot\text{ha}^{-1}$. Análise conjunta (duas safras), valor genético dos genitores: Epagri 108= $3.456,70 \text{ kg}\cdot\text{ha}^{-1}$; Irat 122= $3.813,79 \text{ kg}\cdot\text{ha}^{-1}$.

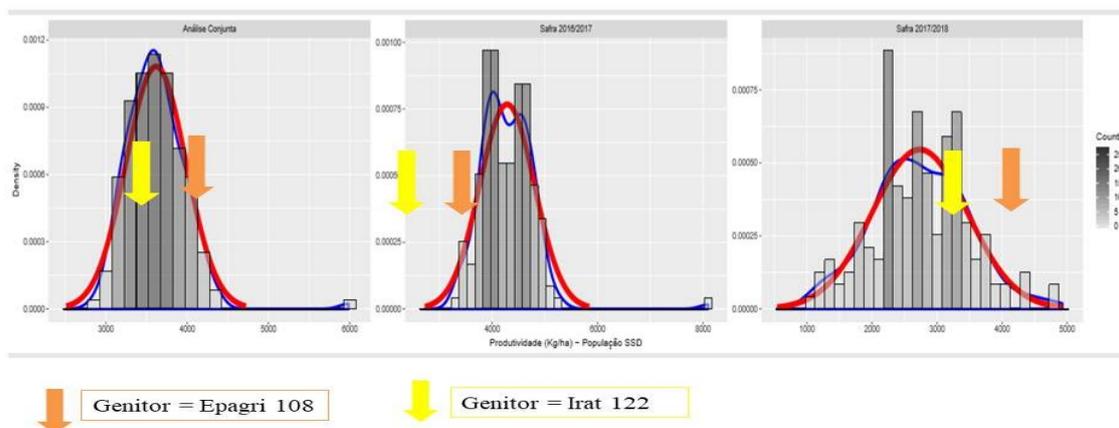


Figura 3. Histogramas do grupo SSD para o caráter produtividade de grão em $\text{kg}\cdot\text{ha}^{-1}$ no primeiro ano de ensaio (safra 2016/2017): a linha vermelha indica a curva normal; linha azul indica curva de densidade. Valor genético dos genitores: Epagri 108= $3.256,35 \text{ kg}\cdot\text{ha}^{-1}$; Irat 122= $0,00 \text{ kg}\cdot\text{ha}^{-1}$. Segundo ano de ensaio (safra 2017/2018), valor genético dos genitores: Epagri 108= $4.322,35 \text{ kg}\cdot\text{ha}^{-1}$; Irat 122= $3.313,89 \text{ kg}\cdot\text{ha}^{-1}$. Análise conjunta (duas safras), valor genético dos genitores: Epagri 108= $4.322,35 \text{ kg}\cdot\text{ha}^{-1}$; Irat 122= $3.313,89 \text{ kg}\cdot\text{ha}^{-1}$.

3.3.3.2 Caráter Altura de Plantas (ALT)

No 1º ano a altura das plantas variou de 116 cm a 165 cm para as RILs derivadas de Bulk. Já as RILs de SSD apresentaram variação de 91 cm a 177 cm, enquanto que as Testemunhas variaram de 100 cm a 104 cm. No 2º ano as RILs derivadas de Bulk tiveram variação de 101 cm a 130 cm, as RILs de SSD variaram de 81 cm a 132 cm, e as Testemunhas variaram de 96 cm a 117 cm. A análise conjunta das RILs derivadas de Bulk apresentaram alturas de 104 cm a 143 cm, as RILs de SSD variaram de 82 cm a 142 cm, e as Testemunhas variaram de 101 cm a 121 cm (Figuras 4, 5 e 6).

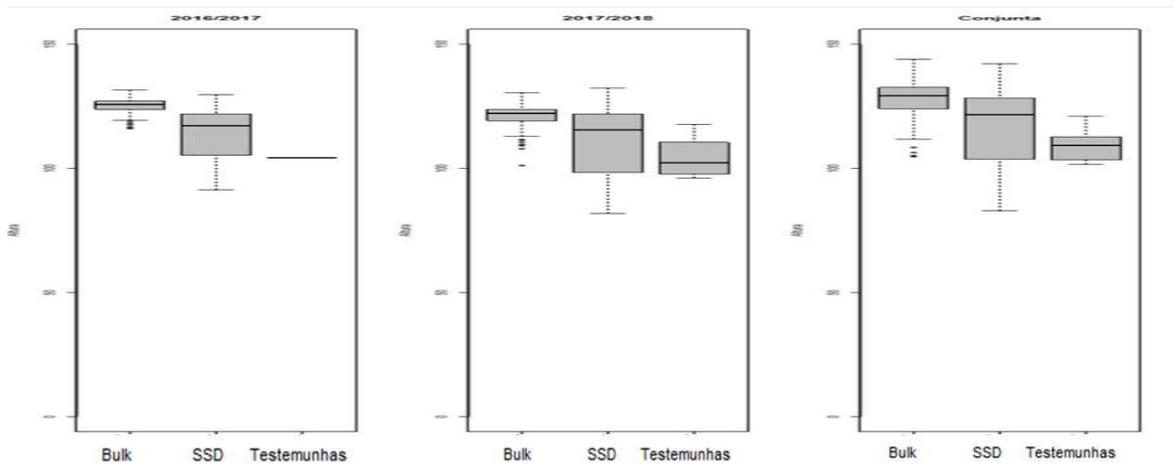


Figura 4. Boxplots para o caráter Altura de Plantas (ALT) em centímetros (cm) para RILs (BULK, SSD e Testemunhas) para o 1º ano de experimento (2017); 2º ano de experimento (2018) e análise conjunta dos dois anos.

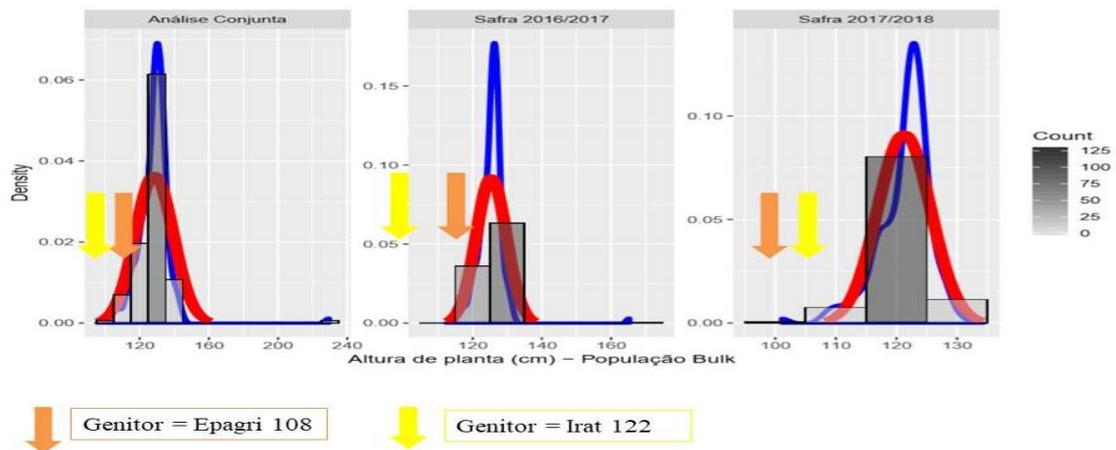


Figura 5. Histogramas dos grupos Bulk para o caráter altura de (cm) para primeiro ano de ensaio (safra 2016/2017); segundo ano de ensaio (safra 2017/2018) e análise conjunta (duas safras). A linha vermelha indica a curva normal; linha azul indica curva de densidade. O valor genético do genitor (Epagri 108=104 cm); o valor genético do genitor (Irat 122=0). Segundo ano de ensaio (safra 2017/2018): o valor genético do genitor (Epagri 108= 100 cm); o valor genético do genitor (Irat 122=106 cm). Análise conjunta (duas safras) o valor genético do genitor (Epagri 108=108 cm); e o valor genético do genitor (Irat 122=103 cm).

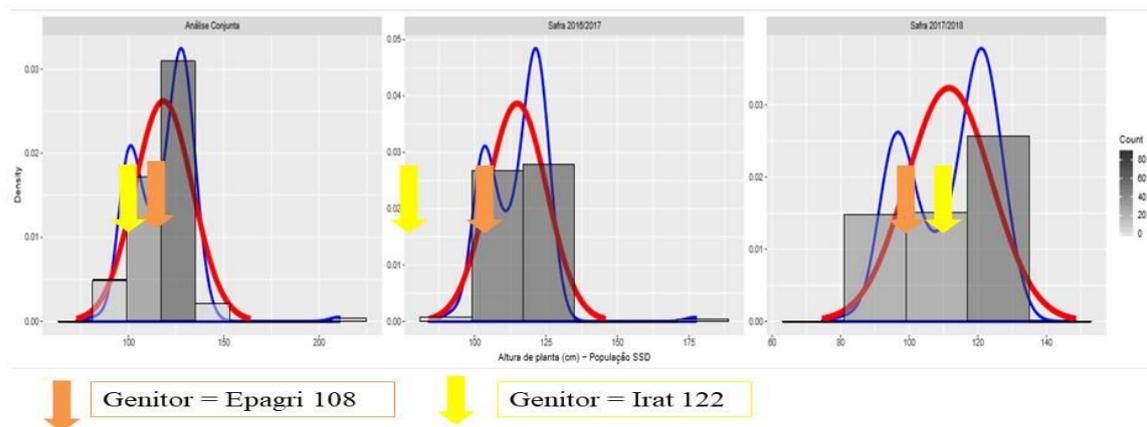


Figura 6. Histogramas dos grupos SSD para o caráter altura de (cm) para primeiro ano de ensaio (safra 2016/2017); segundo ano de ensaio (safra 2017/2018) e análise conjunta (duas safras). A linha vermelha indica a curva normal; linha azul indica curva de densidade. O valor genético do genitor (Epagri 108=104 cm); o valor genético do genitor (Irat 122=0). Segundo ano de ensaio (safra 2017/2018): o valor genético do genitor (Epagri 108= 100 cm); o valor genético do genitor (Irat 122=106 cm). Análise conjunta (duas safras) o valor genético do genitor (Epagri 108=108 cm); e o valor genético do genitor (Irat 122=103 cm).

3.3.4 Identificação das RILs com Melhor Desempenho Produtivo

Foram selecionadas as dez linhagens com maior produtividade e as dez com menor produtividade dentre as 158 linhagens de cada método, a fim de verificar qual método originaria as melhores linhagens para o caráter nessa amostragem. Foram utilizados como parâmetros as RILs superiores ao valor genotípico dos genitores, número de RILs superiores ao melhor genitor e o número de RILs superiores entre as 10 melhores (mais produtivas) (Tabela 4).

As dez linhagens com maior valor genotípico no grupo Bulk em 2017 foram superiores ao genitor Epagri 108 = 3.256,35 kg.ha⁻¹ (Irat 122 teve as duas parcelas perdidas), sendo que a melhor linhagem (RIL 93) se destacou com 5.815,43 kg.ha⁻¹. O grupo SSD também apresentou todas as dez linhagens com valores superiores aos valores genotípicos dos genitores, com destaque para a RIL 592 com produtividade de 8.096,26 kg.ha⁻¹. Entretanto, de forma geral, quando comparadas as 10 melhores RILs dos dois métodos, Bulk apresentou todas as linhagens superiores as linhagens de SSD para o primeiro ano (Tabela 4).

No experimento de 2018 cinco RILs de Bulk foram superiores ao genitor Irat 122 (3.313,88 kg.ha⁻¹), entretanto nenhuma das dez mais produtivas superou o genitor Epagri 108 (4.322,34 kg.ha⁻¹), sendo que a melhor RIL teve produtividade de 4.304,34 kg.ha⁻¹. Considerando o grupo SSD em 2018, quatro RILs foram superiores apenas ao genitor Irat 122, e seis superaram ambos os genitores.

Considerando a análise conjunta do grupo Bulk, nenhuma RIL foi coincidente das dez mais produtivas nos dois anos de experimento, e todas superaram os genitores. Já para SSD das RILs menos produtivas todas foram inferiores aos genitores, entretanto das mais produtivas todas superaram os genitores (Tabela 4).

Considerando o caráter ALT para o grupo Bulk nos experimentos de 2017 e 2018, todas as RILs apresentaram maiores valores que os próprios genitores. Já para o grupo SSD foram identificadas RILs com alturas menores às dos parentais nos dois experimentos, o que é desejável nos programas de melhoramento, uma vez que, plantas menores apresentam menor acamamento, e conseqüentemente menor perda de grãos durante a colheita (Tabela 5 e Apêndices IV). Na análise conjunta do grupo Bulk, as RILs com menores alturas variaram de 104 a 114 cm, enquanto que as mais altas variaram de

138 a 149 cm. Já as RILs SSD estimadas com menor altura variaram de 82 e 98 cm, enquanto as mais altas variaram de 134 a 210 cm (Tabela 5).

Tabela 4. Valores genotípicos (Blups) das dez RILs de Bulk e SSD selecionadas como mais produtivas e menos produtivas para o 1º ano (2017), 2º ano (2018) e análise conjunta.

RILs BULK MAIS PRODUTIVAS						
1º ano (2017)		2º ano (2018)		Análise Conjunta		
RILs	PG	RILs	PG	RILs	PG	
	93	5.815,42	132	4.304,34	51	4.610,25
	5	5.708,63	61	4.128,45	178	4.591,08
	27	5.707,32	58	4.109,26	183	4.565,07
10 +	38	5.618,18	22	4.025,41	15	4.522,51
	94	5.609,80	65	4.024,31	73	4.511,73
	69	5.589,09	56	3.925,38	76	4.486,00
	74	5.558,12	96	3.885,78	52	4.478,59
	40	5.515,55	11	3.884,67	250	4.471,52
	70	5.512,78	27	3.873,74	186	4.463,54
	82	5.481,62	85	3.858,83	158	4.445,97
RILs BULK MENOS PRODUTIVAS						
1º ano (2017)		2º ano (2018)		Análise Conjunta		
RILs	PG	RILs	PG	RILs	PG	
	3	4.116,91	123	2.615,32	77	3.816,09
	291	4.110,40	148	2.571,14	117	3.813,35
	198	4.107,03	25	2.505,40	168	3.812,86
10 -	164	4.104,72	124	2.474,24	69	3.801,88
	68	4.101,79	146	2.430,98	294	3.798,41
	115	4.075,62	126	2.419,91	170	3.789,45
	293	4.062,74	108	2.418,05	272	3.784,34
	76	4.059,28	105	2.401,81	100	3.737,53
	97	4.048,02	95	2.368,36	184	3.708,43
	21	4.010,75	137	2.277,17	241	3.706,96
RILs SSD MAIS PRODUTIVAS						
1º ano (2017)		2º ano (2018)		Análise Conjunta		
RILs	PG	RILs	PG	RILs	PG	
	592	8.096,26	172	4.790,07	594	5.985,11
	502	5.283,16	300	4.738,04	753	4.372,13
	508	5.157,77	276	4.383,83	566	4.345,12
10 +	565	5.148,78	306	4.338,72	503	4.274,95
	573	4.980,30	260	4.328,77	510	4.255,84
	528	4.949,63	233	4.123,70	551	4.224,30
	550	4.942,80	293	3.989,69	574	4.168,42
	618	4.924,42	171	3.848,34	569	4.160,35
	717	4.920,56	231	3.834,03	629	4.126,16
	594	4.902,69	178	3.797,58	596	4.099,44
RILs SSD MENOS PRODUTIVAS						
1º ano (2017)		2º ano (2018)		Análise Conjunta		
RILs	PG	RILs	PG	RILs	PG	
	793	3.590,79	192	1.548,18	675	3.134,46
	725	3.551,63	183	1.544,35	643	3.133,75
	677	3.527,55	182	1.414,17	678	3.100,46
10 -	608	3.443,81	290	1.380,26	716	3.096,64
	578	3.415,60	313	1.348,89	543	3.091,68
	510	3.414,05	266	1.305,94	671	3.051,79
	521	3.410,96	230	1.258,19	644	3.036,79
	527	3.407,40	316	1.158,27	612	2.982,07
	769	3.391,96	315	1.156,44	770	2.965,25
	621	3.321,76	244	1.089,81	794	2.855,16

Tabela 5. Valores genotípicos (Blups) das dez RILs de Bulk SSD selecionadas como as com maior altura e menor altura (ALT) para o 1º ano (2017), 2º ano (2018) e análise conjunta.

		RILs BULK ALTAS					
		1º ano (2017)		2º ano (2018)		Análise Conjunta	
		RILs	ALT	RILs	ALT	RILs	ALT
10 +		136	165,22	219	130,47	141	149,33
		219	131,55	94	128,74	221	143,92
		64	130,92	145	128,61	65	142,10
		94	130,76	43	128,51	95	142,03
		145	130,45	242	127,91	146	141,44
		43	130,15	64	127,83	46	140,43
		207	130,01	250	127,42	209	140,24
		242	129,68	207	127,40	243	139,39
		250	129,45	177	127,07	178	138,41
		177	129,38	16	126,62	251	138,30
		RILs BULK BAIXAS					
		1º ano (2017)		2º ano (2018)		Análise Conjunta	
		RILs	ALT	RILs	ALT	RILs	ALT
10 -		266	119,82	285	113,55	241	114,65
		285	119,74	95	112,96	290	114,62
		238	119,68	266	112,90	268	114,57
		95	119,35	200	111,54	97	113,38
		107	118,85	164	111,34	109	111,98
		164	118,08	21	110,43	167	111,72
		21	117,52	241	109,53	22	108,42
		118	116,63	118	109,17	123	106,32
		241	116,19	283	107,89	285	105,05
		283	116,06	150	101,14	242	104,74
		RILs SSD ALTAS					
		1º ano (2017)		2º ano (2018)		Análise Conjunta	
		RILs	ALT	RILs	ALT	RILs	ALT
10 +		542	177,42	522	132,33	543	145,71
		511	129,67	511	130,32	513	142,10
		655	128,43	655	129,77	656	138,57
		720	128,02	573	129,27	725	138,34
		573	127,85	770	128,66	574	137,58
		770	126,62	720	128,50	771	136,07
		612	126,40	621	128,03	613	135,08
		501	126,37	612	127,76	748	134,80
		545	125,88	587	127,63	700	134,71
		604	125,69	501	127,59	502	134,60
		RILs SSD BAIXAS					
		1º ano (2017)		2º ano (2018)		Análise Conjunta	
		RILs	ALT	RILs	ALT	RILs	ALT
10 -		762	99,85	692	93,54	694	98,18
		624	99,31	742	92,79	749	96,85
		652	99,29	514	92,61	607	96,48
		794	93,67	762	92,06	515	96,15
		793	91,35	748	91,87	714	95,98
		748	100,87	713	91,64	773	95,88
		605	100,32	624	91,02	764	95,73
		713	100,27	652	91,00	653	94,63
		514	100,25	794	90,48	628	94,52
		772	100,00	793	81,82	794	82,92

Para os grupos Bulk e SSD, no geral, das 158 RILs, houve correlações de baixa magnitude entre os caracteres produtividade de grãos e altura de plantas, dos anos avaliados e a análise conjunta (Figura 7). Estes resultados corroboram com os valores baixos de herdabilidade para os caracteres. A interação genótipo x ambiente

provavelmente, foi um fator de extrema relevância para não ocorrer coincidência entre as RILs dentro dos anos avaliados ambos os caracteres (Tabelas 4 e 5). Para o caráter PG, das RILs mais produtivas, apenas uma coincidência foi identificada para Bulk sendo a (RIL 27) no 1º ano e 2º ano, o mesmo ocorreu para SSD (RIL 594) no 1º ano e análise conjunta. Das menos produtivas não houve repetibilidade entre as RILs, entretanto foi observado que as RILs mais produtivas (RILs 69 e 76) de Bulk também foram as menos produtivas dentre a seleção das 10 mais e 10 menos produtivas (Tabela 4), evidenciando forte interação do caráter como ambiente.

Para o caráter altura de plantas obtivemos maiores números de coincidências de RILs nos anos avaliados, tanto para Bulk como para SSD. Para Bulk, das dez mais altas nove coincidiram nos anos 1º e 2º. Das mais baixas, oito se repetiram no 1º e 2º ano. Para SSD também foram encontradas coincidências, sendo que sete RILs mais altas no 1º ano também foram as mais altas no 2º ano, e das mais baixas oito repetiram o desempenho (Figura 7).

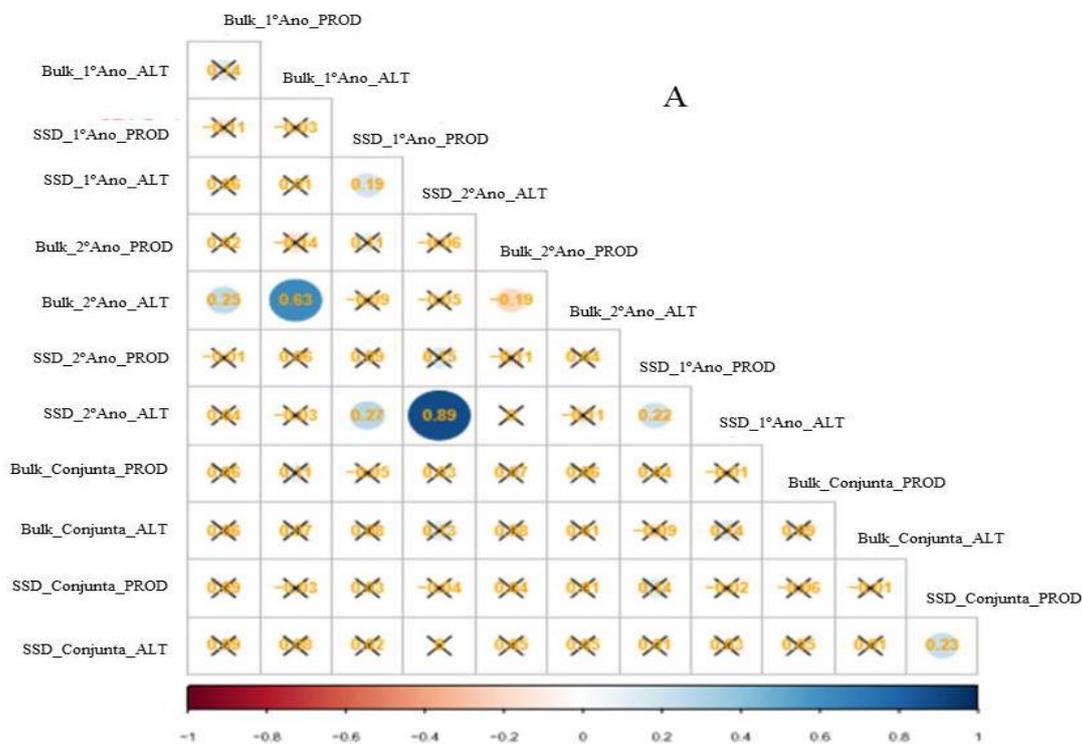


Figura 7. Estimativas de correlação de Pearson (r) a 5% de probabilidade, dos valores genotípicos (BLUPs) de Bulk e SSD para os caracteres Produtividade de Grãos (PROD) e Altura de Plantas (ATL). Apenas os círculos com os valores foram significativos.

3.3.5 Mapeamento Genético

A genotipagem das 146 RILs do grupo SSD e 108 RILs do grupo Bulk, mais os dois genitores, Irat 122 e Epagri 108, pela plataforma DArTseq, resultou na identificação de marcadores SNPs (codominantes) e DArTs (dominantes). O mapeamento genético, realizado no programa MapDisto, distribuiu os marcadores nos 12 cromossomos de arroz. O número de marcadores polimórficos variou entre as populações provenientes dos dois métodos de obtenção de RILs, em consequência de algumas filtragens realizadas (Tabelas 6 e 7). Os marcadores não distorcidos pelo teste de χ^2 , que se ajustaram à segregação esperada em F_7 , de 1:1, resultaram no total de 2.115 marcadores SNPs para RILs SSD. Já para as RILs Bulk foram considerados marcadores SNPs e DArTs. O acréscimo dos marcadores DArT para o Grupo Bulk foi justamente com o intuito de aumentar a densidade de marcas nos cromossomos, a fim de obter números aproximados de marcadores entre as RILs Bulk e SSD. Desse modo foram identificados 1.272 DArT e 1.082 SNPs, totalizando 2.354 marcadores para Bulk (Tabela 6).

Tabela 6. Filtragens utilizadas para obtenção do conjunto de marcadores finais para as análises.

	Marcadores SNPs RILs SSD	Marcadores SNPs e DArTs RILs Bulk
Número de marcadores iniciais	11.279	35.976
Filtro 1: eliminação de marcadores em duplicata	4.914	5.482
Filtro 2: eliminação de marcadores monomórficos; dados faltantes; locos em heterozigozidade e locos distorcidos	768	28.139
Total de marcadores após as filtragens 1 e 2	5.597	2.354
Filtro 3: Seleção de marcadores a cada 50.000 pb por cromossomos para RILs SSD	2.115	-
Total final de marcadores considerados para as análises	2.115	2.354

Tabela 7: Descrição de marcadores distribuídos nos grupos de ligação nas RILs SSD e Bulk.

Marcadores SNPs RILs SSD				Marcadores SNPs e DArTs RILs BULK				
Grupos de Ligação	Total de Marcadores	Distância Total (Mbp)	Distância Total (cM)	Grupos de Ligação	Total de Marcadores DArTs	Total de Marcadores SNPs	Distância Total ~ (Mbp)	Distância Total (cM)
1	135	43	203	1	173	105	42	203
2	68	34	169	2	97	94	34	169
3	40	35	190	3	47	59	35	189
4	274	34	130	4	137	107	34	130
5	63	27	120	5	83	81	27	120
6	249	27	114	6	99	112	29	126
7	327	29	155	7	59	75	29	155
8	309	27	115	8	108	98	28	117
9	229	22	92	9	69	98	22	92
10	286	23	92	10	82	73	20	75
11	73	28	93	11	174	98	28	94
12	62	27	99	12	144	82	27	100
Total	2.115	356	1.572	Total	1.272	1.082	355	1.570

3.3.6 Desequilíbrio de Ligação

O padrão de decaimento do desequilíbrio de ligação (DL) foi estimado com relação à distância entre os marcadores para cada cromossomo, considerando as 158 RILs de cada método separadamente (Figura 8).

Considerando todos os marcadores, a distribuição do r^2 mostrou decaimento com o aumento da distância física entre os marcadores, apresentando alto DL tanto para as RILs SSD quanto para as RILs Bulk. Em média o DL não decaiu rapidamente com o aumento da distância física dos marcadores para as RILs SSD, sendo o ponto médio de decaimento de 11,58 Mb, já para Bulk o decaimento foi mais rápido, com 0,84 Mb.

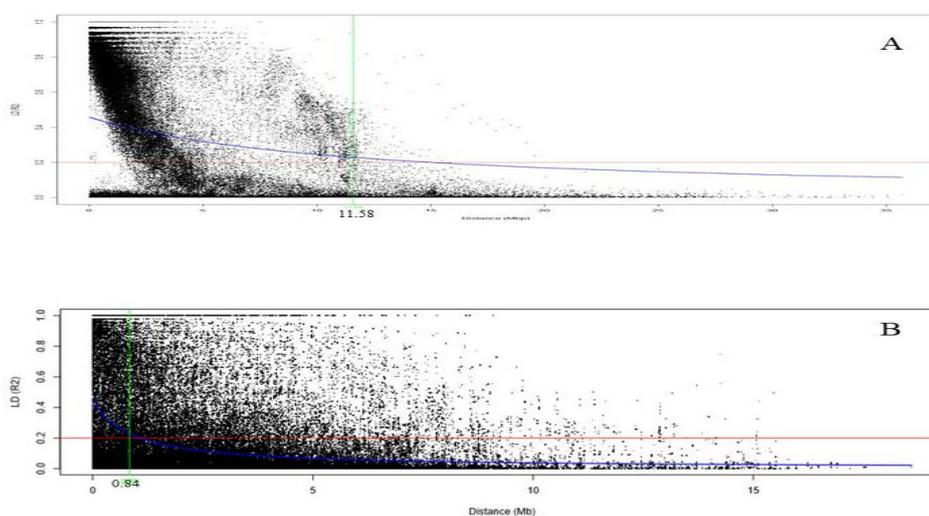


Figura 8: Padrão da curva de decaimento do DL com relação à distância entre os marcadores em todos os cromossomos. (A) RILs SSD, marcadores SNPs e (B) RILs Bulk, marcadores SNPs e DArTs. Linha azul refere-se ao decaimento; linha verde é o ponto de inflexão (ponto médio) e a linha vermelha é a estatística r^2 .

3.3.7 Análise de QTLs das RILs SSD e Bulk

A análise de QTL, por mapeamento múltiplo, identificou sete QTLs (LOD mínimo de 3,0), sendo quatro QTLs identificados para as RILs SSD e três QTLs para RILs Bulk. Nas RILs SSD três QTLs estavam envolvidos na produtividade de grãos e um na altura de planta. Já para as RILs Bulk os QTLs foram identificados apenas para produtividade de grãos.

Na análise de QTL por mapeamento múltiplo foram identificados sete QTLs, utilizando LOD mínimo de 3,0, sendo quatro QTLs identificados para as RILs SSD (3 para a produtividade e um para altura) e três QTLs para RILs Bulk todos para produtividade de grãos.

Considerando as RILs SSD, dois QTLs para PG foram encontrados para o segundo ano de avaliação, localizados nos cromossomos 6 e 9 (qGYLD6.1 e qGYLD9), com valores da variação fenotípica explicada de 23,56% e 7,45%, respectivamente, e um QTL foi identificado na análise conjunta, também localizado no cromossomo 6 (qGYLD6.2) com a variação fenotípica explicada de 9,45%. Para o caráter ALT, apenas um QTL foi identificado, no cromossomo 1 (qPTHT1) com a variação fenotípica de 14,01% (Tabela 8 e Figura 9).

Considerando as RILs Bulk, dois QTLs foram localizados nos cromossomos 6 e 9 (qGYLD6 e qGYLD9) no segundo ano, tendo variação fenotípica de 21,65% e 3,71%, respectivamente. Na análise conjunta, um QTL foi identificado no cromossomo 7 (qGYLD7) com variação fenotípica de 12,90% (Tabela 8 e Figura 10).

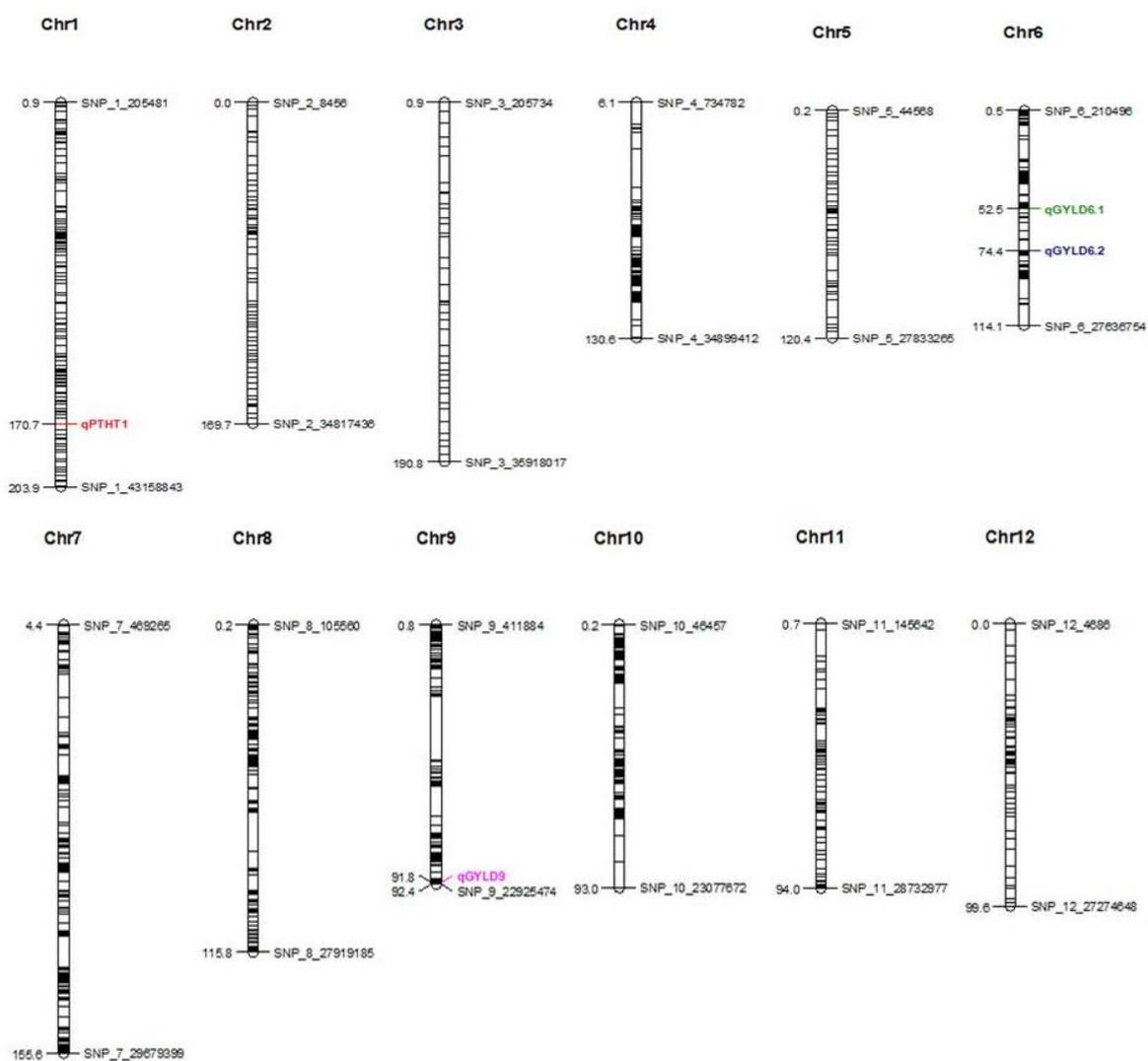


Figura 9. Localização dos QTLs nas RILs SSD para o caráter produtividade de grão (destacados em azul, verde e rosa) e Altura de Plantas (destacado em vermelho) no mapa genético.

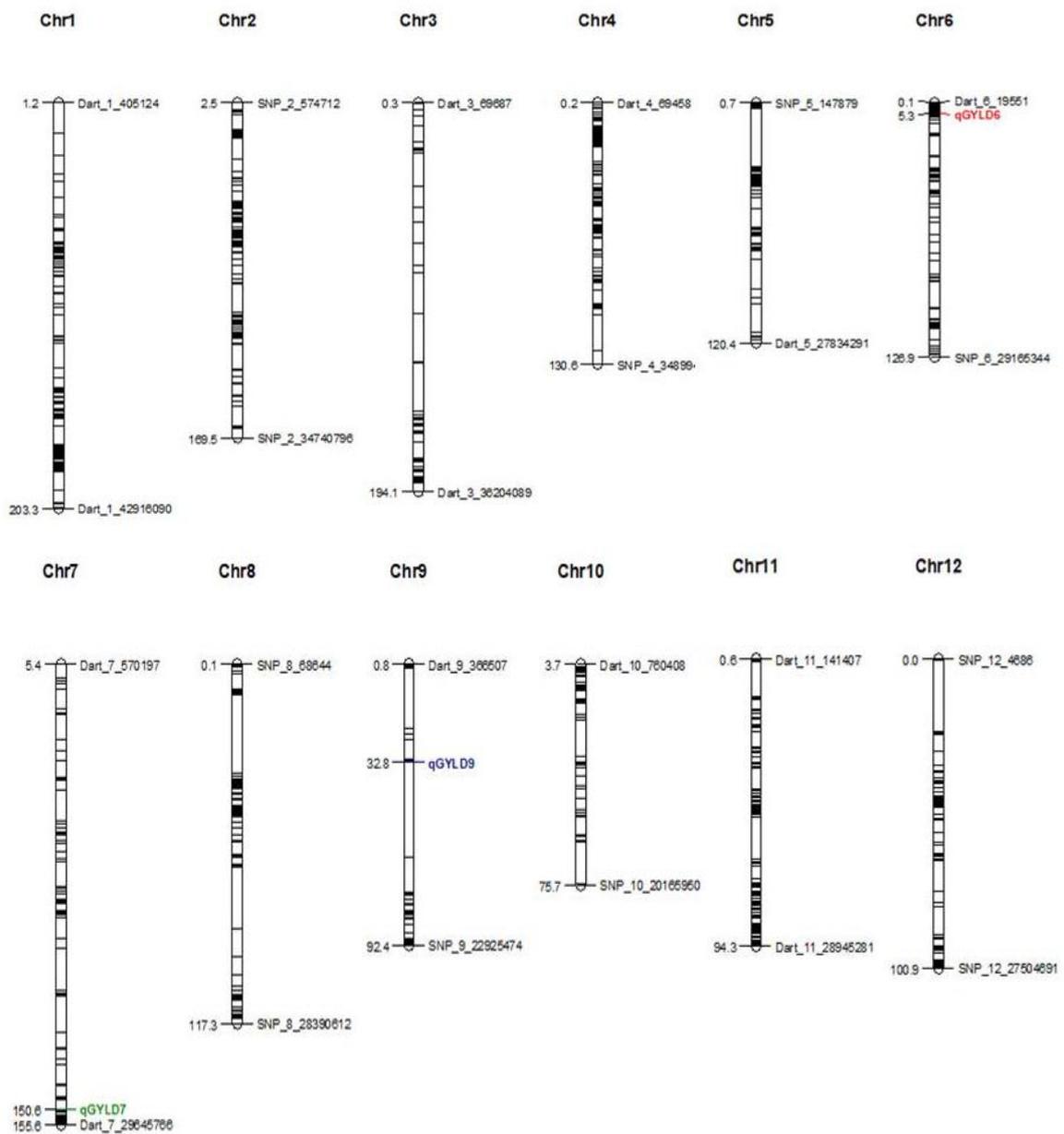


Figura 10. Localização dos QTLs nas RILs Bulk para o caráter produtividade de grão (destacados em azul e verde) no mapa genético.

Tabela 8. QTLs identificados nas RILs SSD e Bulk para o primeiro ano (2017), segundo ano (2018) e análise conjunta (dois anos).

Anos	QTL	Chr.	Intervalo Marcador	Distância (cM)	Bloco (DL)	Posição (cM)	Marcador Significativo	LOD	% Variação Fenotípica	ADD	ALELO FAV.	Troca de Alelo
RILs SSD												
PG 2ºano	qGYLD6.1	6	SNP_6_7843142 a SNP_6_11219760	46 a 55	SNP_6_9563117 a SNP_6_9114303 (329 Kb)	53	SNP_6_9538767	9,38	23,56	389,51	EPAGRI 108	C/T
PG 2ºano	qGYLD9	9	SNP_9_15081610 a SNP_9_22925474	46 a 92	SNP_9_22489199 a SNP_9_22925474 (436 Kb)	91,7	SNP_9_22544728	3,27	7,45	205,76	EPAGRI 108	A/G
PG Conj	qGYLD6.2	6	SNP_6_17061727 a SNP_6_22527564	68 a 90	SNP_6_17931455 a SNP_6_25390861 (100 Kb)	74	SNP_6_17931455	3,14	9,45	132,43	EPAGRI 108	G/A
ALT 1ºano	qPHT1	1	SNP_1_38310141 a SNP_1_38700162	169.8991 a 170.7682	SNP_1_38700162 a SNP_1_38710502 (0 Kb)	170	SNP_1_38700162	14,01	35,72	-6,01	IRAT 122	C/A
RILs BULK												
PG 2ºano	qGYLD6	6	Dart_6_1855213 a Dart_6_2172757	5.08 a 7.43	Dart_6_18552013 a Dart_6_1901482 (46 Kb)	5,33	Dart_6_1901482	6,61	21,65	199,55	EPAGRI 108	(DART)
PG 2ºano	qGYLD9	9	c9.loc31 a SNP_9_22925474	31.77 a 92.37	SNP_9_12730461 a SNP_9_12730961 (0 Kb)	32,77	SNP_9_12730961	3,71	11,39	144,26	EPAGRI 108	C/T
PG Conj	qGYLD7	7	Dart_7_1138309 a Dart_7_29645766	10.74 a 155.56	Dart_7_28888504 (49 Kb)	150	Dart_7_28888504	3,23	12,90	-70,15	IRAT 122	(DART)

Chr = cromossomos; Posição (cM) = Posição do QTL no mapa genético em cM; Bloco (DL) = Bloco em desequilíbrio de ligação; ADD= Efeito Aditivo; Porcentagem da variância fenotípica explicada pelo QTL; ALELO FAV. = alelo favorável vindo do genitor EPAGRI 108 ou IRAT 122.

Foram obtidos gráficos boxplot a partir dos valores genéticos (Blups) para PG das dez RILs mais produtivas e dez menos produtivas do grupo SSD, considerando os marcadores mais próximos que flanqueiam os QTLs qGYLD6.1, qGYLD9 e qGYLD6.2. Observou-se que todos os alelos favoráveis foram provenientes do parental Epagri 108 (cultivar brasileira da subespécie japônic). Considerando o caráter ALT, para o QTL (qPTHT1), os alelos das RILs com menor altura foram provenientes do parental Irat 122 (Figuras 11 e 12). Considerando as RILs Bulk, os QTLs qGYLD9 e qGYLD6 também apresentaram alelos favoráveis advindos do genitor Epagri 108, que elevaram a produtividade no segundo ano do experimento, enquanto que na análise conjunta, o QTL (qGYLD7) apresentou alelo do genitor Irat 122, o qual reduziu a produtividade. Tais resultados são mostrados nos gráficos boxplot dos efeitos dos alelos na produtividade de grãos, evidenciando que das dez RILs mais produtivas o alelo favorável foi oriundo do genitor Epagri 108, já para o caráter ALT o alelo do genitor Irat 122 foi o que predominou (Figuras 13 e 14).

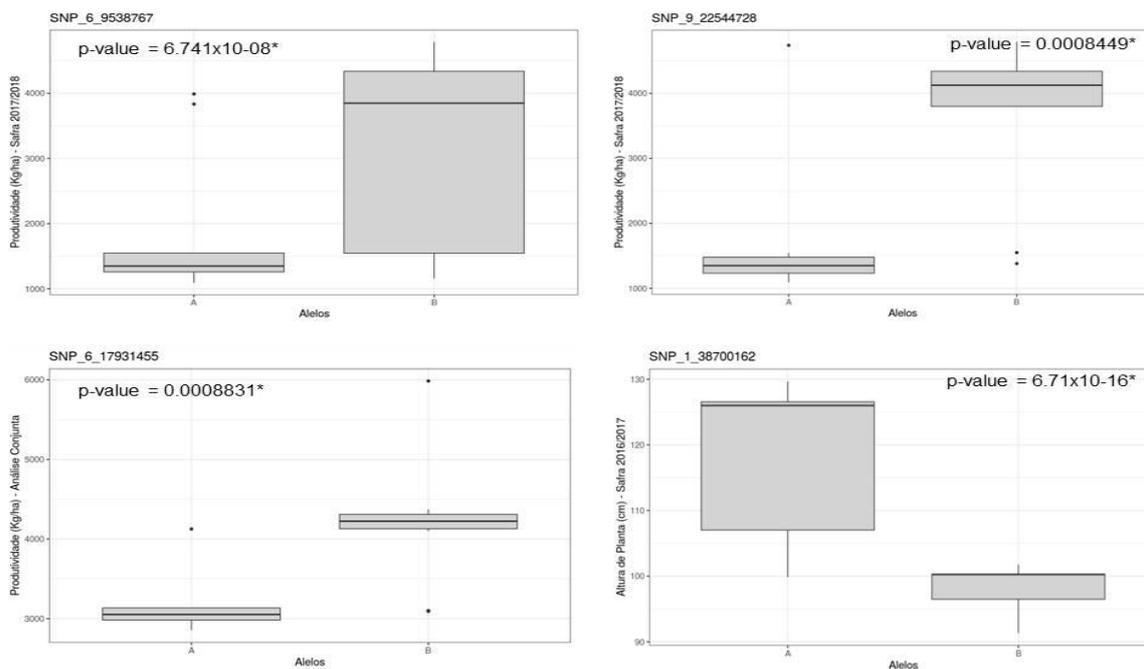


Figura 11. Gráfico boxplot dos efeitos genotípicos (Blups) das RILs SSD, para o caráter Produtividade de grãos (das dez RILs mais produtivas e dez RILs menos produtivas), e para o caráter altura de plantas (dez RILs mais altas e dez mais baixas). Alelos (A) referem-se ao genitor Irat 122 e alelo (B) genitor Epagri 108. Significância pelo Teste-Welch.

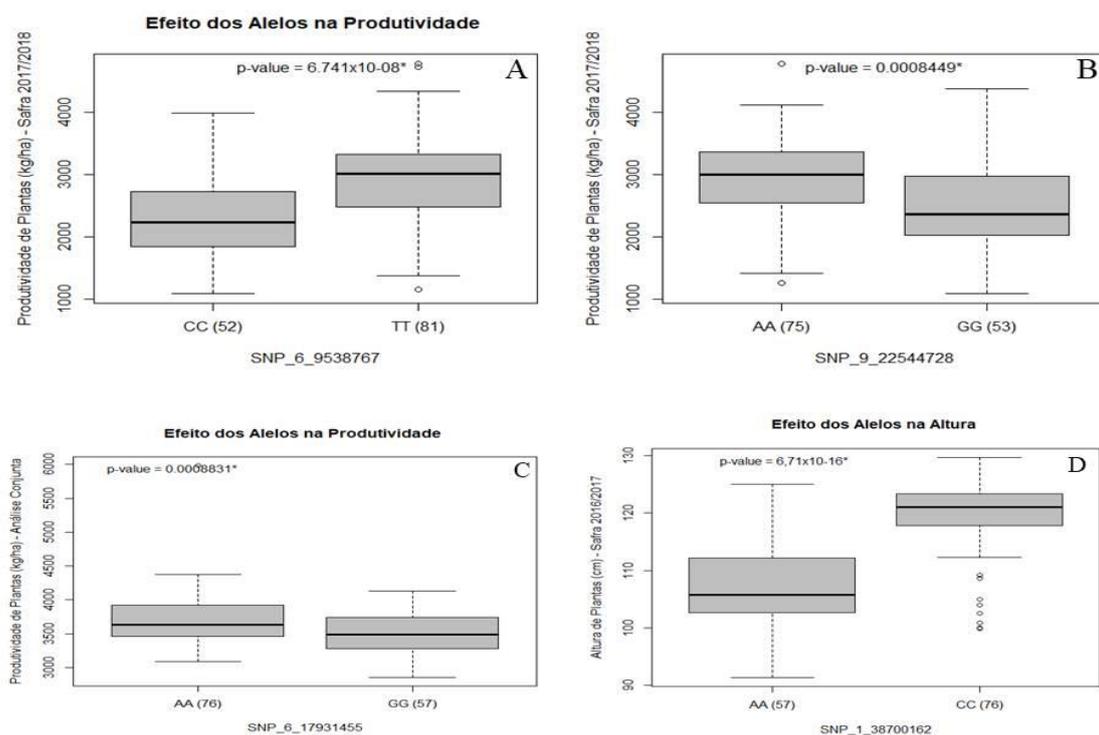


Figura 12. Gráfico boxplot dos efeitos genotípicos das 146 RILs SSD, para o caráter produtividade de grãos, e para altura de plantas. Gráfico (A): corresponde 52 RILs com SNP (CC) do genitor Irat 122, e 81 com (TT) do genitor Epagri 108, para o marcador SNP_6_9538767. Gráfico (B): 75 RILs com SNP (AA) do Epagri 108, e 53 (GG) do Irat 122, marcador SNP_9_22544728. Gráfico (C): 76 RILs com SNP (AA) do Epagri 108, e 57 (GG) do Irat 122. Gráfico (D): 57 RILs com SNP (AA) do genitor Irat 122, e 76 (CC) do Epagri 108. Significância pelo Teste-Welch.

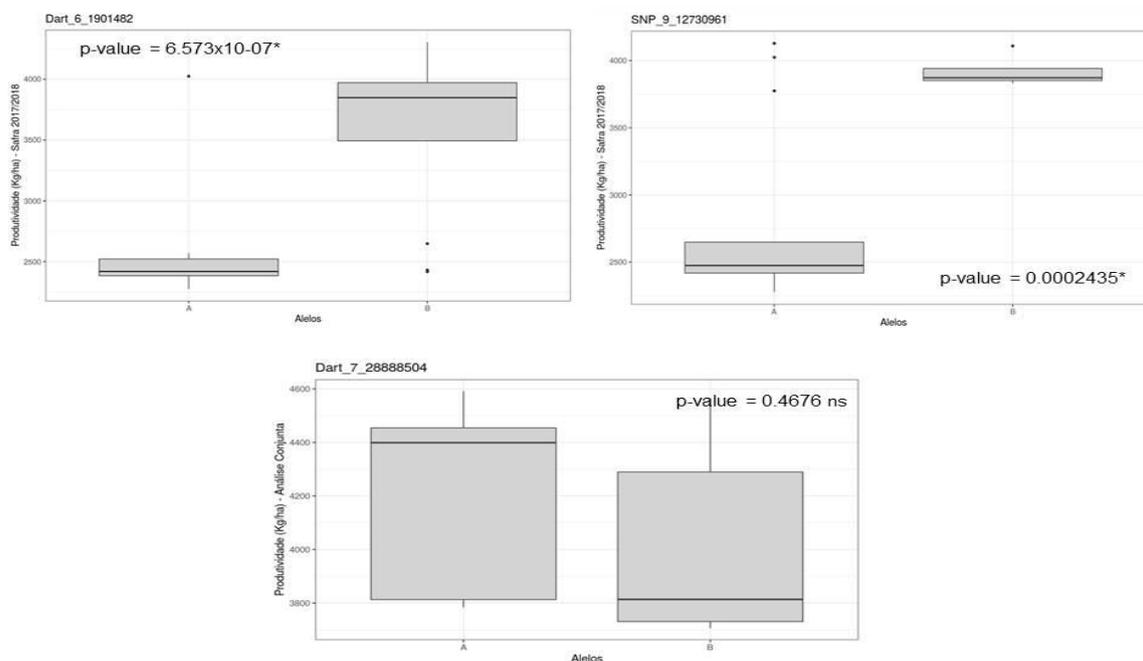


Figura 13. Gráfico boxplot dos efeitos genotípicos (Blups) das RILs Bulk, para o caráter Produtividade de grãos (dez RILs mais produtivas e dez RILs menos produtivas). Alelos (A) referem-se ao genitor Irat 122 e alelo (B) genitor Epagri 108. Significância pelo Teste-Welch's.

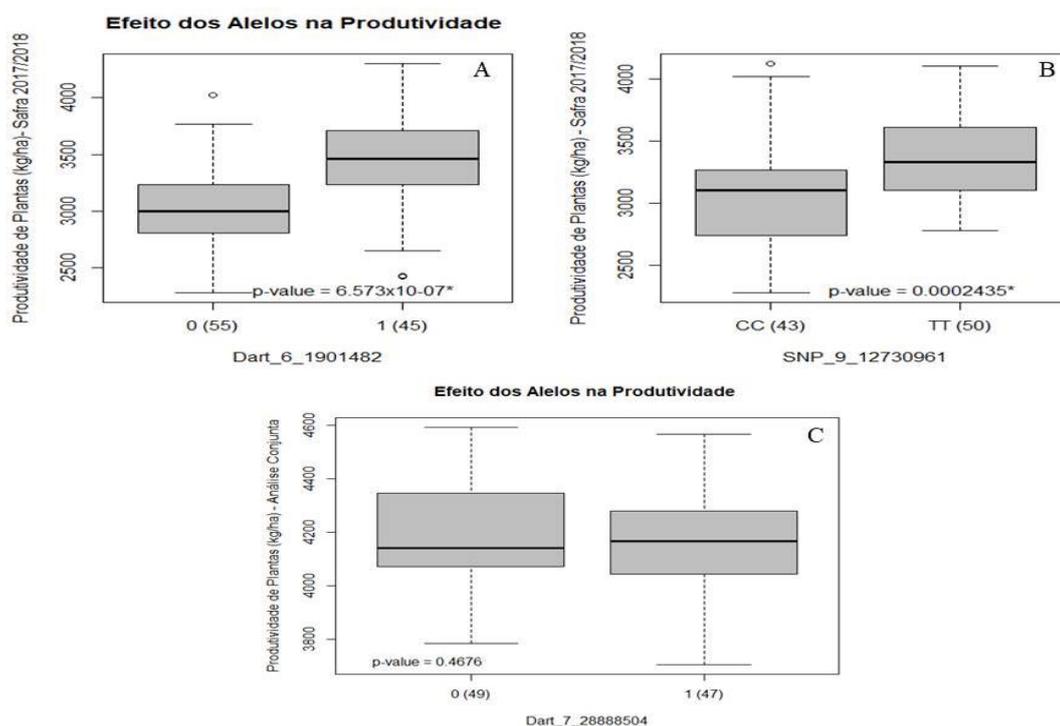


Figura 14. Gráfico boxplot dos efeitos genotípicos das 108 RILs Bulk, para o caráter produtividade de grãos. Gráfico (A): corresponde 55 RILs com marcador DArT (0 = ausência) do genitor Irat 122, e 45 com marcador (1=presença) do genitor Epagri 108, para o marcador Dart_6_1901482. Gráfico (B): 43 RILs com SNP (CC) do Irat 122, e 50 (TT) do Epagri 108, marcador SNP_9_12730961. Gráfico (C): 49 RILs com marcador Dart (0=ausência) genitor Epagri 108, e 47 (1=presença) genitor Irat 122. Significância pelo Teste-Welch's.

3.3.8 Análise de Componentes Principais (PcA) e DarWin

Considerando todos os marcadores para as RILs SSD e Bulk, a análise de componentes principais revelou que o primeiro componente principal (PC1) juntamente com o segundo componentes (PC2) foram capazes de explicar a variação total dos caracteres PG e ALT, isto para ambas as RILs (Figuras 15 e 16). Os PcAs não apresentaram uma clara divisão das RILs, ou seja, elas não se agruparam em torno de um ou outro genitor tanto para Bulk quanto SSD. Isso indica que não houve uma predominância do background genético específico para um ou outro genitor nas linhagens avançadas. Esse resultado também é coerente com o observado na estruturação observada no dendrograma DarWin (Figura 17).

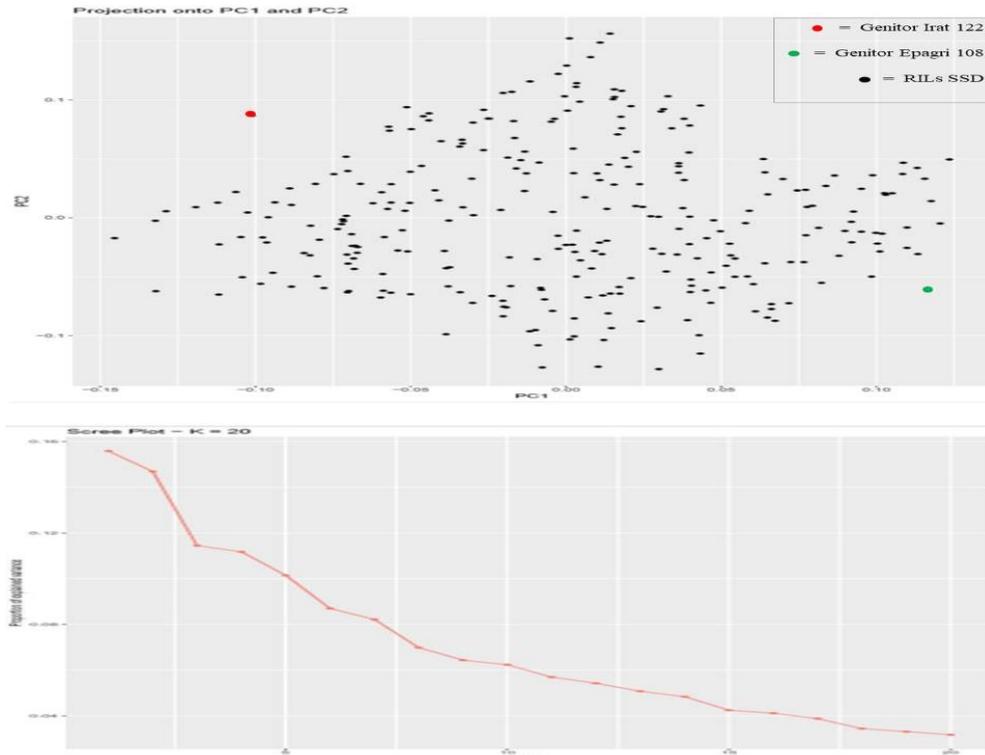


Figura 15. PcAs dos 2.115 marcadores SNPs das 146 RILs SSD (ponto em preto) e os parentais Epagri 108 (ponto em verde) e Irat 122 (ponto em vermelho).

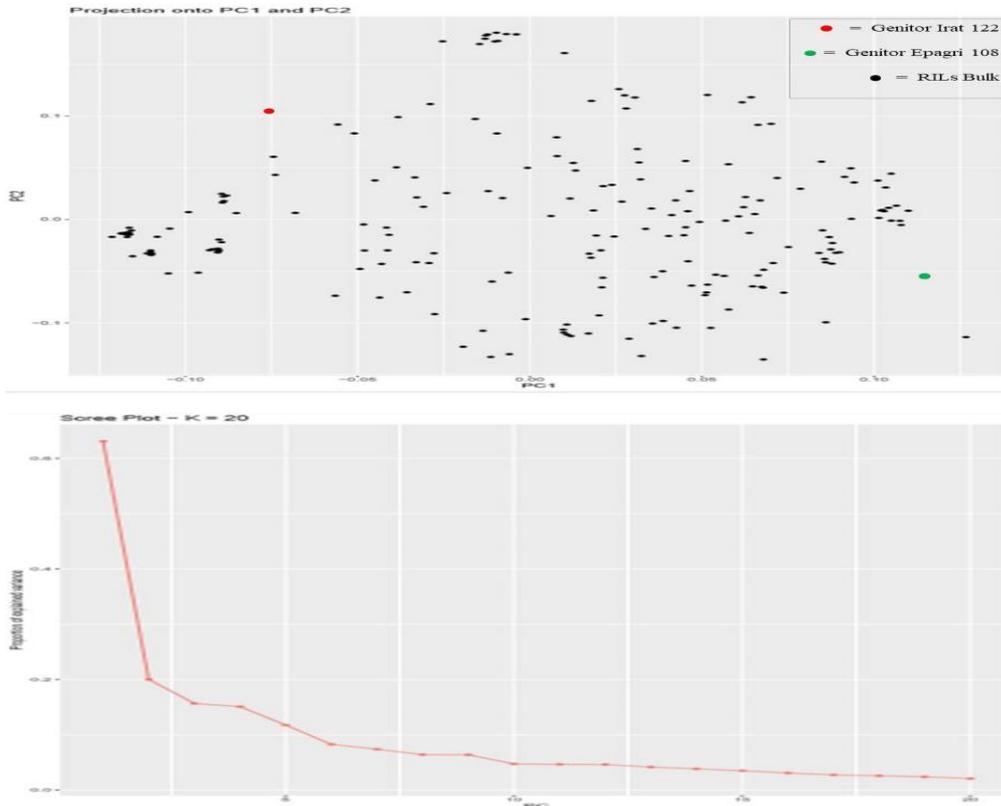


Figura 16. PcAs dos 2.354 marcadores SNPs e DArTs das 108 RILs Bulk (ponto em preto) e os parentais Epagri 108 (ponto em verde) e Irat 122 (ponto em vermelho).

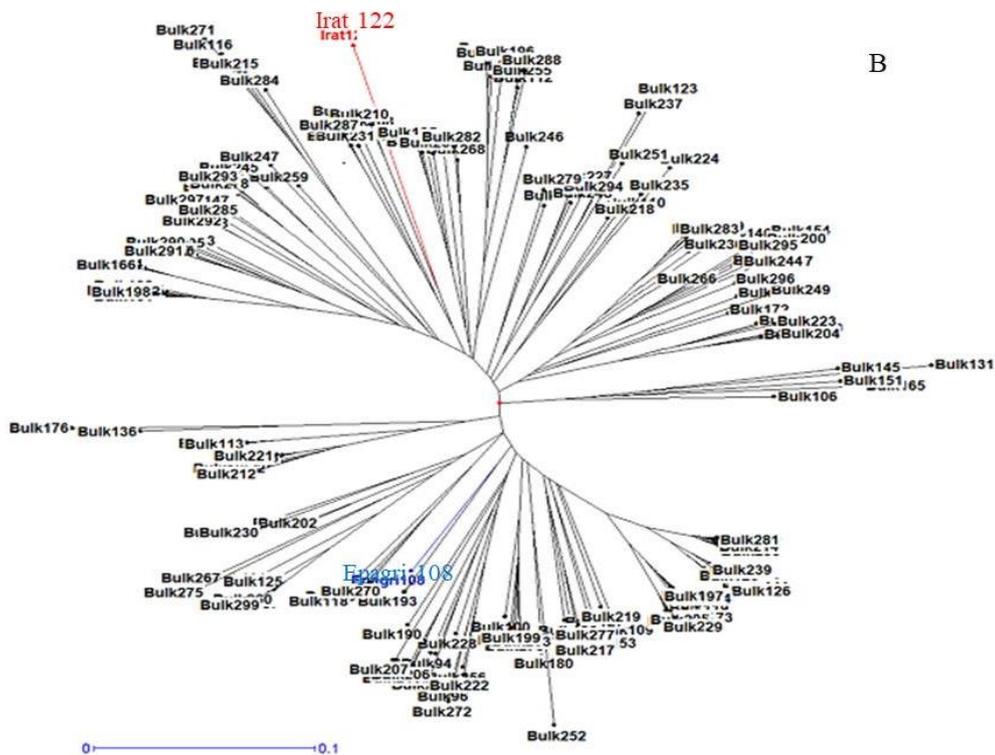
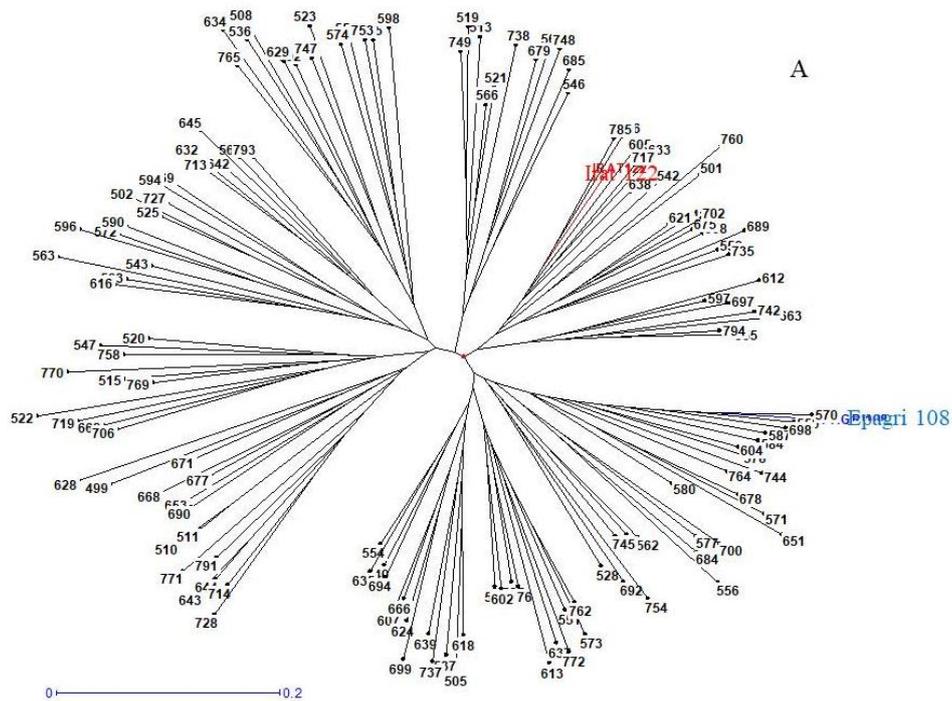


Figura 17. (A) Dendrograma DArWin das 146 RILs SSD e os parentais Epagri 108 e Irat122, com total de 2.115 marcadores. (B) Dendrograma DArWin das 108 RILs Bulk e os parentais Epagri 108 e Irat 122, com total de 2.354 marcadores.

3.3.9 Anotação Putativa dos Genes de Blocos Haplotípicos

A anotação funcional de genes presentes nos blocos haplotípicos de QTLs significativamente associados aos caracteres foi baseada em termos GO, depositados no banco de dados do RGAP (*Rice Genome Annotation Project*).

Para as RILs SSD, considerando o caráter PG, o pico do QTL foi identificado pelo SNP_6_9538767 (no experimento de 2018), localizado no cromossomo 6, está dentro de um bloco haplotípico de 329 Kb que contém 47 genes, e está em uma região intergênica dos (LOC_Os06g16590 e LOC_Os06g16600). Considerando os dados para a análise conjunta, o marcador SNP_6_17931455 está em um bloco de 100 kb com 10 genes, porém não está posicionado dentro de gene (Apêndice I). Outro QTL relacionado a produtividade foi identificado no cromossomo 9, este QTL teve como pico de significância o marcador Dart_9_22544728, que foi localizado no gene LOC_Os09g39240, está em um bloco haplotípico com extensão de 436 kb.

Considerando o caráter ALT, foi identificado um QTL no cromossomo 1, com pico de significância no marcador o SNP_1_38700162, localizado no gene LOC_Os01g66630. Esse QTL foi localizado em um bloco de 175 Kb com 25 genes (Apêndice I).

Para as RILs Bulk, para o caráter PG o pico de significância foi no Dart_6_1901782 no cromossomo 6, este marcador está localizado na região do LOC_Os06g04450 dentro de um bloco de 46 Kb. Outro QTL foi identificado, no cromossomo 7, e o pico de significância significativo foi no Dart_7_28888504, porém não está localizado em um gene. Já para o cromossomo 9, o marcador SNP_9_12730961 está posicionado no gene LOC_Os09g21110 (Apêndice II).

A análise de *Gene Ontology* (GO) revelou os genes dos blocos haplotípicos associados a processos biológicos, por exemplo, genes associados a respostas a estímulos abióticos; função molecular, como atividade de quinases; e componentes celulares, membrana e núcleo (Figuras 18 e 19).

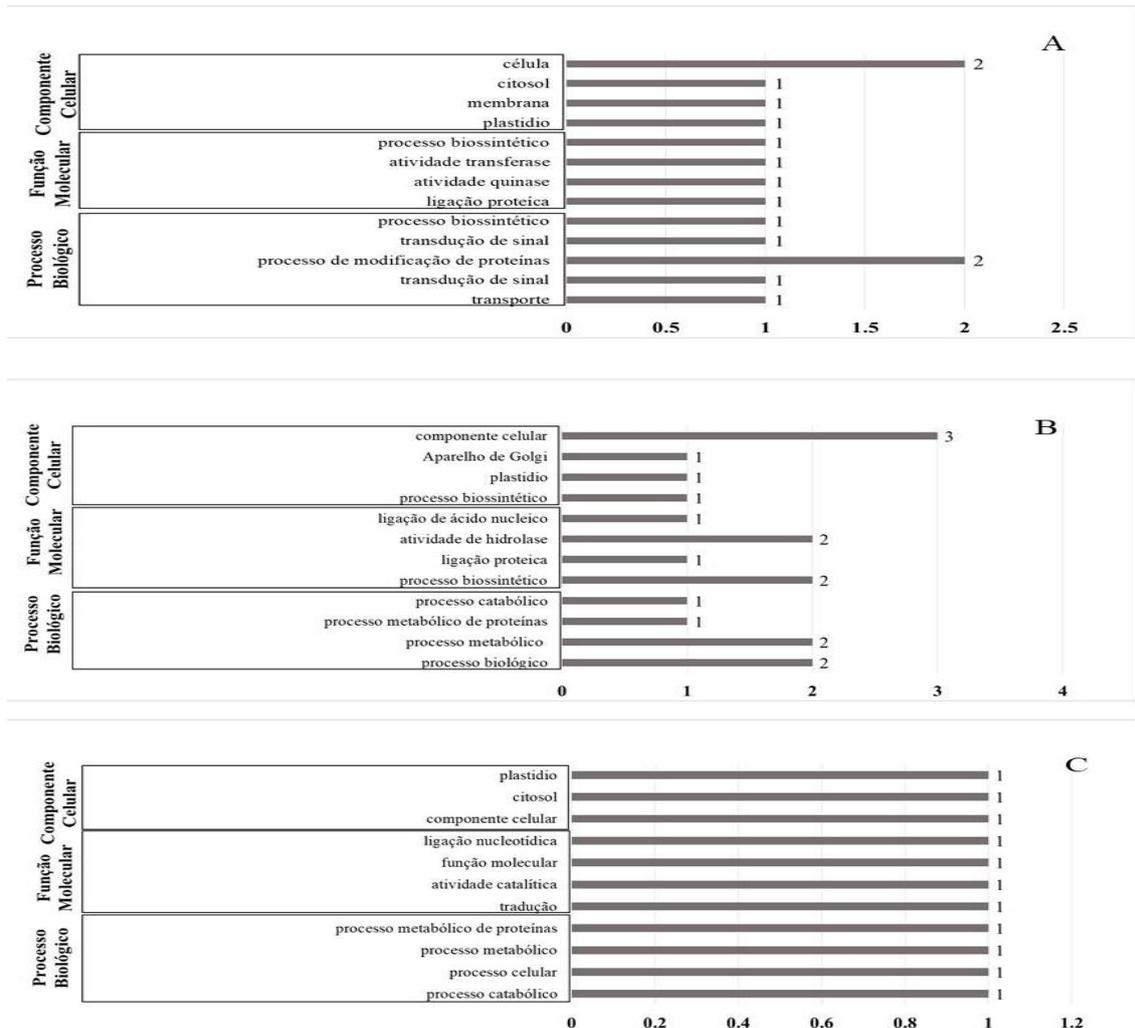


Figura 19. Classificação funcional de dentro do bloco haplotípico com pico de QTL significativo, anotados com os termos GO que incluem número de genes relacionados aos atributos de componente celular, função molecular e processo biológico, para os caracteres produtividade de grãos. **(A):** no cromossomo 6 considerando os 11 marcadores dentro do bloco haplotípico para o segundo ano de avaliação das RILs Bulk. **(B):** no cromossomo 7 considerando 7 marcadores dentro do bloco com o pico de QTL significativo para o segundo ano de avaliação das RILs Bulk. **(C):** no cromossomo 9 considerando 2 marcadores do bloco com pico de QTL significativo para a análise conjunta das RILs Bulk.

Adicionalmente, foi realizada uma busca na base de dados do QTARO (<http://qtaro.abr.affrc.go.jp/>) para identificar QTLs já descritos na literatura para os mesmos cromossomos nos quais identificamos QTLs nesse trabalho, e baseados também nos mesmos intervalos dos blocos haplotípicos que foram identificados (Tabela 9).

Tabela 9. QTLs identificados na base de dados QTARO, para o intervalo dos blocos haplotípicos contendo os QTL significativos.

QTLs	CHR	Caráter	Referências
yd1	1	Produção por plantas	Hua, J.P., Xing, Y.Z., Xu, C.G., Sun, X.L., Yu, S.B., and Zhang, Q. (2002).
tns6, sd6, qGY6-1 e gw-6	6	Número total de espiguetas / panícula; densidade de espiguetas (número de espiguetas por panícula); rendimento de grãos (t / ha); peso de 1000 grãos	Lin, H.X., Qian, H.R., Zhuang, J.Y., Lu, J., Min, S.K., Xiong, Z.M., Huang, N., and Zheng, K.L. (1996). Cui, K.H., Peng, S.B., Xing, Y.Z., Yu, S.B., Xu, C.G., and Zhang, Q. (2003). Lu, C., Shen, L., Tan, Z., Xu, Y., He, P., Chen, Y., and Zhu, L. (1996).
yldp 9.1	9	Produção por planta	Marri, P.R., Sarla, N., Reddy, L.V., and Siddiq, E.A. (2005).
ng1; sf1	1	Número de grãos/planta; Fertilidade das espiguetas	Lin, H.X., Qian, H.R., Zhuang, J.Y., Lu, J., Min, S.K., Xiong, Z.M., Huang, N., and Zheng, K.L. (1996).
yldp 9.1 e gy9	9	Produção por planta; Rendimento de grãos	Marri, P.R., Sarla, N., Reddy, L.V., and Siddiq, E.A. (2005). Cho, Y.I., Jiang, W., Chin, J.H., Piao, Z., Cho, Y.G., McCouch, S., and Koh, H.J. (2007).

Através da análise de impacto do SNP, considerando 2.115 marcadores SNPs para as RILs SSD, quatro deles foram os picos de significância para a identificação dos QTLs e estão dentro de blocos haplotípicos. Com relação ao impacto, foram identificados 0,46% de variantes com alto impacto, 2,07% com baixo impacto, 1,27% com impacto moderado e 96,19% com impacto modificador (Tabela 10).

Considerando a variante de alto impacto (variantes que têm impacto disruptivo na proteína, com a perda da função), uma variante está relacionada com códon de parada prematura (variante do SNP_1_38700162 do gene LOC_Os01g66630). Além disso, foram localizadas variantes de impacto com efeito modificador (variante SNP_9_22544728 do gene LOC_Os09g39240), e variantes de região intergênica (variante do SNP_6_9538767 e do SNP_6_17931455).

A análise de impacto do SNP para as RILs Bulk, foi realizada com 2.354 marcadores, dos quais três SNPs identificaram os picos de QTLs. Foram identificados 0,43% de variantes de alto impacto, 2,46% de variantes de baixo impacto, 1,33% de variantes de impacto moderado, e 95,76% de impacto modificador (Tabela 11).

As variantes de impacto modificador (geralmente variantes não codantes ou variantes que afetam genes não codificadores, em que as previsões são difíceis ou não há evidência de impacto), estiveram relacionadas a variante 3'UTR (variantes dos marcadores Dart_6_1901482 do gene LOC_Os06g04450, e do SNP_9_12730961 do gene LOC_Os09g21110), e para a variante *upstream* do gene (variante do marcador Dart_7_28888504 localizado fora do gene).

Tabela 10. Classificação e número de efeitos dos 2.115 SNPs identificados pela genotipagem DArT-seq das RILs SSD.

RILs	Efeitos	Tipo de Efeitos	Contagem	Percentual (%)	
SSD	Região	<i>Downstream</i>	3.222	37,06	
		Éxon	291	3,34	
		Intergênico	1.370	15,76	
		Íntron	449	5,16	
		Splice_Site_Acceptor	31	0,35	
		Splice_Site_Donor	1	0,01	
		Splice_Site_Region	31	0,35	
		<i>Upstream</i>	3.089	35,53	
		3'UTR	138	1,58	
		5'UTR	70	0,80	
		Impacto	Alto	40	0,46
			Baixo	180	2,07
			Moderado	111	1,27
	Modificador		8.361	96,19	
	Tipo	Variante 3'UTR	138	1,57	
		Variante 5'UTR prematura	20	0,22	
		Variante 5'UTR	50	0,57	
		Variante <i>downstream</i> do gene	3.222	36,79	
		Região intergênica	1.370	15,64	
		Variante íntron	505	5,76	
		missense_variant	111	1,26	
		Variante não codificante	43	0,49	
		splice_acceptor_variant	31	0,35	
		splice_donor_variant	1	0,01	
		splice_region_variant	36	0,41	
		Códon de parada	8	0,09	
		Variante sinônima	133	1,51	
Variante <i>upstream</i> do gene	3.089	35,27			

Tabela 11. Classificação e número de efeitos dos 2.354 marcadores identificados pela genotipagem DArT-seq das RILs Bulk.

RILs	Efeitos	Tipo de Efeitos	Contagem	Percentual (%)	
Bulk	Região	<i>Downstream</i>	3.592	37,10	
		Éxon	339	3,50	
		Intergênico	1.515	15,65	
		Íntron	470	4,85	
		Splice_Site_Acceptor	35	0,36	
		Splice_Site_Donor	1	0,01	
		Splice_Site_Region	37	0,38	
		<i>Upstream</i>	3.449	35,63	
		UTR_3	142	1,46	
		UTR_5	100	1,03	
		Impacto	Alto	42	0,43
			Baixo	239	2,46
			Moderado	129	1,33
	Modificador		9.270	95,76	
	Tipo	Variante 3'UTR	142	1,45	
		Variante 5'UTR prematura	29	0,29	
		Variante 5'UTR	71	0,72	
		Variante <i>downstream</i> do gene	3.592	36,83	
		Região intergênica	1.515	15,53	
		Variante íntron	536	5,49	
		missense_variant	129	1,32	
		Variante não codificante	32	0,32	
		splice_acceptor_variant	35	0,35	
		splice_donor_variant	1	0,01	
		splice_region_variant	41	0,42	
		Códon de parada	6	0,06	
		Variante sinônima	174	1,78	
Variante <i>upstream</i> do gene	3.449	35,36			

3.4 DISCUSSÃO

3.4.1 Dados Fenotípicos

Normalmente em um programa de melhoramento genético os cruzamentos são realizados entre genitores com fenótipos favoráveis para características complementares entre si, com o objetivo de gerar populações segregantes com variabilidade genética suficiente para realização de seleções de linhagens superiores para os caracteres de interesse. É indispensável o conhecimento do componente da capacidade combinatória, para a escolha de genitores geneticamente divergentes envolvidos em esquemas de cruzamentos nos programas de melhoramento, sobretudo quando se deseja identificar híbridos promissores e/ou desenvolver linhagens superiores a partir deles (Marchesan, 2008).

O cruzamento Epagri 108 x Irat 122 foi escolhido por possuir plantas com maior produtividade média, os genitores deterem alta capacidade combinatória e ser um cruzamento inter-subespecífico, envolvendo uma cultivar indica (Epagri 108) e uma cultivar japônica (Irat 122), além de ter sido conduzido por dois métodos de melhoramento populacional (Bulk e SSD), derivando RILs Bulk e RILs SSD (Ramos et al., 2019, Apêndice VII).

O grupo de RILs Bulk se destacou com as maiores médias (4.953,25 kg.ha⁻¹; 3.225,59 kg.ha⁻¹ e 4.189,42 kg.ha⁻¹) para o caráter PG, respectivamente para o 1º e 2º ano e análise conjunta. O grupo de RILs SSD teve o segundo melhor desempenho produtivo, com médias (4.284,43kg.ha⁻¹; 2.735,15 kg.ha⁻¹ e 3.614,56 kg.ha⁻¹) para 1º e 2º ano e análise conjunta. As Testemunhas apresentaram médias abaixo dos grupos Bulk e SSD, sendo (3.158,01 kg.ha⁻¹; 3.369,85 kg.ha⁻¹ e 3.456,70 kg.ha⁻¹) para 1º e 2º ano e análise conjunta. Para o caráter ALT o grupo Bulk apresentou maiores médias, seguido por SSD e as Testemunhas (Tabela 2).

Esses resultados refletiram a alta e significativa capacidade combinatória dos genitores previamente observada por Ramos et al. (2019). Segundo Bernardo et al. (2010) são preferíveis genitores com maiores efeitos de capacidade de combinação para constituírem as novas populações, favorecendo a seleção de linhagens homozigóticas, em plantas autógamas. A capacidade de combinação reflete o efeito de aditividade dos genes e tem importância fundamental em prever as trocas alélicas na média da população. Os resultados também podem ser explicados devido a forma de condução da população

segregante, já que em gerações iniciais (exemplo de gerações F_2 , F_3 , F_4), apenas as plantas mais produtivas e competitivas que superaram a seleção realizada pelo ambiente (natural) conseguiram se destacar (Ramalho et al., 2012). Resultado semelhante foi encontrado por Kambar et al. (2011), também na cultura do arroz, onde o método Bulk foi superior ao método pedigree e SSD na obtenção de linhagens com maior produtividade média de grãos, e também para a cultura do feijoeiro, (Silva et al., 2009). Isso indica que as plantas mais produtivas deixaram o maior número de sementes e a cada geração de avanço da população essas plantas tiveram maior representatividade.

Para o método Bulk, os resultados indicam que toda a diversidade útil das linhagens foi liberada, pois não houve processo de seleção durante a condução da população para a obtenção das linhagens ($F_{7:8}$). A atuação da seleção natural ocorreu como um fator que favoreceu a média do caráter produtividade de grão no método Bulk, o que não ocorre no método SSD. Esse resultado também foi obtido por Fahim et al. (1998), onde o método SSD proporcionou as piores médias de produtividade em progênies de arroz. Silva et al. (2013) também encontraram maiores médias no método Bulk quando comparado com SSD e Bulk dentro de F_2 para os caracteres de produtividade de grãos e teores de fibra bruta no feijoeiro comum. Entretanto, para o feijoeiro (Raposso et al., 2000) e para a soja (Funada et al., 2013), o método SSD foi mais eficiente em obter linhagens mais produtivas.

Trabalhos anteriores realizados por Wiebe (1942) com cevada e Allard (1991) com feijão, citado por Ramalho et al. (2012), verificaram a eficácia do método Bulk e a eficiência da seleção natural quando utilizaram mistura de linhas puras. Entretanto, o que interessava era o efeito da seleção natural em uma população segregante, pois era o fator predominante do método. Os autores indicam que a seleção natural atuava preferencialmente sobre indivíduos com maior estabilidade de produção, isto é, aqueles indivíduos que apresentavam a mesma produtividade tanto em condições favoráveis quanto em condições menos favoráveis (Ramalho et al., 2012). No presente trabalho, ocorreu a competição entre as plantas, independente do método, contudo, para Bulk as plantas dentro das parcelas ainda segregavam, principalmente para o caráter altura, ocorrendo plantas altas e baixas dentro da mesma parcela, assim como a presença e ausência de aristas. Nessas condições, as plantas mais altas são favorecidas quando em competição por espaço e disseminação do grão de pólen. Assim, as gerações segregantes conduzidas pelo método Bulk são submetidas a uma pressão da seleção natural. A adaptabilidade dos indivíduos

relaciona-se com a habilidade de sobrevivência de cada genótipo, e esta depende do número de descendentes que cada indivíduo deixa para a geração seguinte, o que provavelmente ocorreu com as RILs Bulk. A capacidade de sobrevivência em competição deve estar correlacionada positivamente com a adaptabilidade e produtividade. O ambiente deve ser escolhido de forma criteriosa para que esta correlação seja maximizada, isto é, aquele em que a seleção natural favoreça os indivíduos desejáveis (Ramalho et al., 2012; Borém, 2009).

Os coeficientes de variação experimental do trabalho indicam que os dois anos de avaliação tiveram de boa a moderada precisão experimental para ambos os caracteres avaliados (Tabela 3). A acurácia foi de moderada a alta. Resende & Duarte (2007) recomendam o uso da acurácia seletiva na avaliação da precisão experimental, já que o CVe não contempla toda a informação genética. A acurácia seletiva consiste na correlação entre o valor genotípico verdadeiro do genótipo e aquele estimado ou predito a partir das informações dos experimentos.

Nesse trabalho observou-se a maior variância genética nas RILs SSD do que nas RILs Bulk (Tabela 3). No primeiro ano de experimento para o caráter PG, apenas o grupo SSD apresentou a σ^2g significativa ($p < 0,05$) na análise LRT, indicando uma grande variabilidade genética dentro das 158 RILs de SSD. Uma explicação para o Bulk não ter apresentado significância, pode ser devido linhagens serem geneticamente muito próximas, e terem uma variação genética pequena entre si, quando comparado as linhagens SSD, este fato também é explicado pela própria estrutura do método de condução das linhagens Bulk, já que alguns indivíduos podem ter deixado o maior número de descendentes devido a maior capacidade de produção de pólen ou a maior altura, e pela interferência ambiental no ano de avaliação experimental.

No segundo ano de experimento, as RILs dos dois grupos e as Testemunhas foram significativas pelo teste LRT para PG, porém as Testemunhas apresentam maiores estimativas de σ^2g seguida de RILs SSD e por último, RILs Bulk. Para a análise conjunta apenas o grupo SSD foi significativo. O caráter ALT, no primeiro ano de ensaio, Bulk e SSD apresentaram o efeito significativo de σ^2g . No segundo ano todos os grupos foram significativos, assim como para a análise conjunta. A grande magnitude de σ^2g infere que existe variabilidade genética dentro da população de RILs avaliadas nos métodos Bulk e SSD e nas Testemunhas. O teste LRT na análise de deviance comprova a significância de σ^2g nos anos dos ensaios nas RILs avaliadas nos experimentos (Tabelas 1 e 3). Resultados

semelhantes foram encontrados por Miladinovic et al. (2011), Raposo et al. (2000) Fahim et al. (1998), com maiores valores para de variância genética e herdabilidade para o método SSD quando comparado ao método Bulk.

O parâmetro variância fenotípica (σ^2_f) para o caráter PG apresentou a maior variação dentro do grupo das Testemunhas em ambos os anos e na análise conjunta. Para o primeiro ano o grupo SSD apresentou variação maior que Bulk, o que também persistiu no segundo ano e análise conjunta. Para o caráter ALT o grupo das Testemunhas apresentou maior variação no 1º ano, seguido do grupo SSD e por último Bulk. No segundo ano o grupo SSD teve a maior variação, posteriormente as Testemunhas e Bulk. Na análise conjunta o SSD continuou com maior variação, seguido de Bulk e Testemunhas.

A herdabilidade pode ser definida como a proporção herdável da variabilidade total, ou seja, a razão entre variância genotípica e variância fenotípica (Falconer, 1987). Seu papel preditivo do valor genético por meio do valor fenotípico é muito relevante, pois somente o valor fenotípico de um indivíduo pode ser mensurado, mas é o valor genético que influenciará a próxima geração. A h^2 estimada foi em nível de parcelas, no sentido amplo, porque em linhas puras, devido à homozigose, apenas a variância aditiva se manifesta na porção genética dos componentes da variância. Nesse caso, a herdabilidade no sentido amplo e no sentido restrito é igual (Ramalho et al., 1993). Neste trabalho observou-se que a variância genética entre as RILs do método SSD foi maior do que no método Bulk para ambos os anos avaliados e análise conjunta do caráter PG (Tabela 3). Coerentemente, também a h^2 foi maior em SSD de (0,23; 0,77; 0,29) do que em Bulk (0,19; 0,57; 0,17), respectivamente para 1º ano, e 2º ano e análise conjunta. Deve-se ressaltar que a produtividade de grão é um caráter complexo e influenciado por diversos fatores, de modo que a estimativa de valores baixos de herdabilidade para esse caráter é comum (Pontes Júnior et al., 2016)

A h^2 para o caráter ALT também foi maior em SSD (0,57; 0,78; 0,85) para 1º ano, 2º ano, e análise conjunta em relação aos grupos Bulk e as Testemunhas (Tabela 3). Uma observação muito importante é que a herdabilidade não é apenas uma propriedade do caráter, mas também da população e das condições ambientais a que foram submetidas os indivíduos da população (Ramalho et al., 1993). Portanto, o valor da estimativa de herdabilidade de dado caráter sempre se refere a uma população específica e sob condições particulares (Hallauer et al., 2010). Na realidade a h^2 pode ser aumentada não somente pela introdução de mais variação genética na população, mas também estabilizando mais o

ambiente no qual as plantas irão se desenvolver. Dessa forma, o uso de ambientes desuniformes e populações segregantes provenientes de cultivares semelhante contribuem para reduzir a herdabilidade.

Comumente, espera-se que os métodos Bulk e SSD proporcionem maior variabilidade entre as progênies, em relação aos outros métodos para autógamias, pois a variabilidade disponível que é liberada a partir da geração F_7 é 1,97 da variância genética aditiva (σ^2_A) (Ramalho et al., 2012). No presente trabalho ambos os métodos indicaram variabilidade genética entre as 158 RILs avaliadas. Esses resultados são esperados durante a condução de uma população segregante, porém, vários fatores podem afetar a eficiência desses métodos. Entre estes fatores estão a atuação da seleção natural, as perdas decorrentes do processo de amostragem, além dos efeitos da interação genótipos com ambientes. A produtividade de grãos pode ser afetada por esses fatores, pois está relacionada com vantagens adaptativas, pelas quais atua a seleção natural (Corte et al., 2002).

A condução de populações segregantes pelo método Bulk é mais simples e fácil, mas tem como principal desvantagem a perda de alelos favoráveis por amostragem. Essa desvantagem pode ser contornada com o aumento do número de indivíduos (tamanho efetivo populacional - N_e) a cada geração. Tomando como referência a primeira geração segregante F_2 , o incremento da variância aditiva é expressivo nas gerações seguintes. Com a endogamia, permite explorar a variabilidade genética liberada dentro e entre linhagens de um cruzamento. Em Bulk com o decorrer das gerações na população ocorre a manutenção dos indivíduos mais adaptados devido a seleção natural (Ramalho et al., 2012). Assim, o conhecimento da ação da seleção natural é estratégico no manejo de populações, e fazendo uso desse conhecimento de maneira favorável para o melhorista, por exemplo, utilizando o número amostral da população o maior possível, de forma que represente todos as plantas, permite que a variabilidade genética de caracteres com valor adaptativo possa garantir a manutenção dos indivíduos face às alterações ambientais (Kremer et al., 2014; Templeton, 2006). No presente trabalho, o tamanho amostral para o avanço das gerações dos métodos Bulk e SSD foi a preconizada por programas de melhoramento de espécies (Borém & Miranda, 2009).

Funada et al. (2011), ressaltaram trabalhos realizados por Muehlbauer et al. (1981) em estudo de simulação, que no método SSD 82% das plantas em F_2 estão representadas na geração F_6 , e Kervella & Fouillou (1992), em estudo semelhante

utilizando o método Bulk, indicaram que apenas 67% das plantas em F₂ estão representadas em F₅, ou seja, indicando perdas por amostragem do método. Também foi levado em consideração, na comparação entre SSD e Bulk, que a variação das linhas em Bulk que podem ser perdidas no decorrer das gerações, podem ser parcialmente compensadas pela expressão da variação dentro da linha, enquanto no método SSD isso não ocorre. Outro ponto importante para se destacar, em relação ao método Bulk, é que durante a condução das populações segregantes, a seleção natural atua preservando os indivíduos mais adaptados para o ambiente em que ocorreu o avanço das populações. Assim, há necessidade dos melhoristas avançarem as populações segregantes em ambientes/e anos diferentes, como no presente trabalho, visando à obtenção de linhagens que associem bom desempenho e maior adaptação aos ambientes de cultivo em que se destacaram (Ramalho et al., 2012).

As 158 RILs de Bulk e SSD apresentaram variações de produtividade de grão nos dois anos avaliados (Figuras 1, 2 e 3). Entretanto, as RILs SSD tiveram maior variação de produtividade no 1º ano (3.321,80 kg.ha⁻¹ a 8.096,30 kg.ha⁻¹), isto também refletiu em indivíduos transgressivos aos genitores. Já para Bulk não foram identificadas RILs com produtividade superior aos genitores. No 2º ano de avaliação, o grupo SSD teve variação de (1.089,81 kg.ha⁻¹ a 4.790,07 kg.ha⁻¹), enquanto o grupo Bulk variou de (2.277,18 kg.ha⁻¹ a 4.304,34 kg.ha⁻¹). Neste ano houve redução da produtividade das RILs tanto de Bulk como de SSD, em relação ao 1º ano. Esse fato foi devido a maior pluviosidade no período de colheita ocorrida no segundo ano o que levou a protelar todo o período de colheita das parcelas, proporcionando maior tempo de exposição das plantas às chuvas e aos ventos, levando a perda de grãos pela degrana e acamamento das plantas, os resultados corroboram com os dados de pluviosidade (Apêndice IV). Para o caráter ALT, o grupo Bulk apresentou maiores valores quando comparados ao grupo SSD para os anos avaliados, enquanto que as Testemunhas apresentaram as menores alturas (Figuras 4, 5, e 6).

Na geração avançada, as RILs SSD superaram as RILs de Bulk pelo valor genotípico individual. Entretanto, o método Bulk pode ter sido penalizado devido ao fato de que em nenhum momento do avanço das gerações ter ocorrido com seleção direcionada das melhores plantas com as características fenotípicas desejáveis, já que em geração avançada não havia a representatividade de todas as plantas da geração inicial, como ocorre em SSD. A seleção no decorrer das gerações pelo método Bulk resultaram em

maior probabilidade de obtenção de linhagens mais produtivas para o arroz (Kambar et al., 2011), para aveia (Royce et al., 1947) e trigo (Pawar et al., 1986).

Nas 158 RILs de cada método foram selecionadas dez linhagens com maior produtividade e dez com menor produtividade, e também as dez plantas mais altas e mais baixas (Tabelas 4 e 5), a fim de verificar qual método originaria as melhores linhagens. Do exposto, levando em consideração os critérios, número de linhagens superiores das dez melhores e os valores genotípicos superiores aos genitores, para o caráter produtividade de grão, o melhor método de condução das RILs foi o SSD.

Para o caráter altura de plantas, SSD também foi o melhor para obter linhagens com altura reduzida, dentro dos dois anos avaliados e na análise conjunta. Resultado semelhante foi encontrado por Medimagh et al. (2012) na cultura da cevada indicando o método SSD como melhor para obter linhagens mais produtivas. Entretanto Júnior et al. (2014) encontraram melhores resultados para o método Bulk ao selecionarem as melhores linhagens, além de levarem em consideração as estimativas de parâmetros genéticos como herdabilidade, variância genética e a interação com ambiente na cultura do feijoeiro.

O método de melhoramento mais eficiente para a condução das populações segregantes depende de variáveis como a herdabilidade do caráter, os recursos disponíveis de cada programa de melhoramento, a própria habilidade dos melhoristas, o caráter a ser melhorado e os genitores utilizados (Ramalho et al., 2012). Das limitações que um programa de melhoramento enfrenta, o fator dos custos gerados para experimentação e a mão de obra necessária para a condução dos experimentos a campo são os mais impactantes.

O método Bulk permite a condução de maior número de populações segregantes de interesse para o melhoramento, com maiores médias, e menor custo envolvido para avaliação de experimentos a campo, como citado por Mendes et al. (2011) e Gonçalves et al. (2001). Também permite ter maiores cuidados com a fenotipagem e manejo das áreas de experimentação, aplicando por exemplo, o cultivo mínimo, que é a realização de plantios de outras culturas, antes de realizar os experimentos, como plantio de uma leguminosa, para melhorar a liberação de nutrientes no solo e principalmente evitar misturas de sementes de experimentos anteriores.

A população obtida do cruzamento Epagri 108 x Irat 122 foi conduzida por Bulk até a geração F₇ devido a facilidade do método e o menor custo envolvido, podendo ser mais amplamente utilizada em programas de melhoramento com diversos propósitos

(resistência à doenças, tolerância à seca, grãos especiais, etc.), como o da Embrapa, e aliando a eficiência do processo de seleção natural, para obter maior variabilidade nas progênes em geração avançada e estabilidade na produção (Faria et al., 2013; Corte et al., 2002; Del Peloso & Melo, 2005). Em contrapartida o método SSD permite a obtenção de linhagens homozigotas com maior eficiência para os caracteres desejáveis, com a condução do experimentos em vasos em casa de vegetação, otimizando espaço, tempo e manejo, o que viabiliza a condução de maior número de linhagens, tomada de dados mais acurados, mas com número reduzido de cruzamentos avaliados devido ao tamanho mínimo de progênes.

Diante dos resultados apresentados no presente trabalho, os genitores divergentes e de interesse para o melhoramento, proporcionaram populações segregantes com variabilidade genética e conseqüentemente linhagens homozigotas superiores aos próprios genitores e cultivares comerciais lançadas pela Embrapa.

3.4.2 Dados Genotípicos

O mapeamento utilizando populações derivadas de genitores geneticamente divergentes pode produzir bons resultados no mapeamento de QTL, permitindo identificar os efeitos de locos relacionados a caracteres de interesse (Zhu et al., 2017). Nesse trabalho foram utilizadas linhagens puras recombinantes derivadas do cruzamento Epagri 108 x Irat 122 pertencentes as subespécies indica e japonica, respectivamente. Esses genitores são fenotipicamente contrastantes em vários caracteres agrônômicos, como produtividade, arquitetura da planta e número de perfilhos, por exemplo.

A obtenção dos marcadores polimórficos e informações fenotípicas das populações de RILs Bulk e SSD, permitiu a realização da análise de QTL. O genitor Epagri 108 é uma cultivar brasileira, amplamente utilizada em cruzamentos do programa de melhoramento de arroz da Embrapa, e apresenta alta capacidade geral de combinação. Já o genitor Irat 122 é uma cultivar introduzida, não apresenta alta capacidade combinatória (Apêndice VII). O cruzamento entre esses dois genitores permitiu uma boa complementariedade genética, a qual proporcionou RILs com produtividades superiores aos seus genitores, resultante da segregação transgressiva.

Quanto maior o número de marcadores utilizados na construção de um mapa molecular, maior a chance de ampla cobertura no genoma, aumentando a possibilidade de

encontrar alelos de determinado marcador que estejam em desequilíbrio de ligação (DL) de um QTL, podendo estabelecer uma associação entre genótipo e o fenótipo do caráter de interesse. Os marcadores das RILs-SSD e RILs-Bulk foram posicionados nos 12 cromossomos e garantiram uma distribuição ampla no genoma. Qualquer fator que altere as frequências alélicas pode interferir na dinâmica do DL observado nas populações de RILs do estudo, como por exemplo, os processos de melhoramento genético, como a seleção natural ou artificial, os processos de amostragem dentre outros (Liu, 1998; Schaper et al., 2012). Isso pode ter resultado nas diferenças encontradas nas medidas de DL para as populações RILs SSD e RILs Bulk, visto que as populações são distintas devido aos métodos de melhoramento pela qual foram conduzidas, refletindo assim, em características específicas em cada uma delas.

Na análise de DL houve um decaimento gradual com o aumento da distância física entre os marcadores. As RILs Bulk e SSD apresentaram alto desequilíbrio de ligação, sugerindo a formação de blocos haplotípicos grandes. Para as RILs Bulk, entre as distâncias de 5 Mpb a 10 Mpb o decaimento foi progressivo, enquanto que para as RILs SSD no ponto de 5 Mpb o DL era extremamente alto e contínuo. Em arroz os valores de desequilíbrio de ligação são bastante variáveis, como por exemplo, em linhas endogâmicas, o DL têm alcance de 100-500 Kpb, mas estes valores não são comuns e resultam dos processos específicos de melhoramento ao qual a população tenha sido submetida (Liu et al., 2013). É esperado que em gerações mais avançadas de melhoramento o DL seja menor, devido as recombinações que ocorreram ao longo das gerações de autofecundações, proporcionando a formação de grandes blocos de haplótipos.

Forças evolutivas como recombinação, seleção natural, deriva genética, mutação, padrões de acasalamento ou migração, tem importante impacto no padrão do desequilíbrio de ligação em uma população (Remington et al., 2001). Quando resultantes da ação da seleção natural, observa-se que algumas combinações de alelos de locos distintos e independentes conferem vantagem adaptativa aos indivíduos que as possuem, isto é, funcionam melhor juntas do que independentemente, fato este que pode ter ocorrido com as RILs Bulk, uma vez que a população provavelmente passou por um processo da seleção natural contínuo, ao longo do avanço de geração. Chen et al. (2013) observaram haplótipos de tamanho de 100 kb em *O. sativa*, e ao compararem esse resultado com outras variedades e a espécie silvestre *O. rufipogon*, encontraram um DL muito mais alto nas espécies cultivadas, sendo subespécie Japonica de regiões temperadas a que apresentou

maior DL (> 500 Kb), seguida por Japonica de regiões tropicais (aproximadamente 150 Kb) e por Indica (aproximadamente 75 kb).

As populações de RILs de ambos os métodos foram estruturadas devido aos avanços das gerações, e essa estruturação foi revelada nos gráficos de neighbor-joining e análise de coordenadas principais (PcAs). Os genitores foram posicionados espacialmente distantes entre si, como esperado. Para ambas populações de RILs SSD e RILs Bulk houve um padrão de associação em torno de um ou outro genitor, indicando que esses eram geneticamente divergentes, mantendo-se um nível de estruturação, e que a diversidade genética foi mantida nas linhagens, independentemente do método de avanço de geração, sem privilegiar um ou outro genitor como o mais adaptado, como também observado por (Ma et al., 2016).

O mapa de ligação gerado para os caracteres quantitativos (PG e ALT) nas RILs de arroz foi construído com 2.115 marcadores SNPs polimórficos para RILs SSD, e 2.354 marcadores para RILs Bulk (SNPs e DArTs). Conforme estudos anteriores como de Yu et al. (2011), Yonemaru et al. (2014) e Quin et al. (2017), o mapeamento genético acima de mil marcadores proporciona uma distribuição ampla no genoma. O aumento da densidade de marcadores melhora consideravelmente a resolução e precisão do mapa para análise de QTL (Crowell et al., 2015; Yonemaru et al., 2014, Sakai et al., 2013). O mapa obtido no estudo possui 1.5072 cM para RILs SSD e 1.5070 cM para RILs Bulk (Tabela 6), cujas distâncias são relativamente maiores em relação ao mapa de Ma et al. (2016), que apresentou 1.070 cM.

Um número maior de indivíduos permite amostrar uma maior quantidade de recombinações entre os marcadores. Mapas que tenham maior saturação de marcadores resultam em ligações físicas mais robustas com QTLs, diminuindo a chance de ocorrerem recombinações entre marcador e QTL. Zhang et al. (2017) utilizaram 213 RILs e 3.016 marcadores SNPs para o mapeamento, resultando em 26 QTLs em três caracteres relacionados a produtividade. No presente trabalho, o tamanho populacional utilizado foi 146 RILs SSD, 108 RILs Bulk e dois parentais para a construção dos mapas genéticos, o que também permitiu a identificação de QTLs, porém foram 7 QTLs (4 para as RILs SSD, e 3 para as RILs Bulk).

Não foram encontrados os mesmos QTLs nos anos avaliados. Contudo, os valores de LODs foram altos o suficiente para determinar que os resultados foram bastante consistentes, ou seja, a cada ano, diferentes QTLs de produtividade se sobressaem, o que

pode dificultar a utilização de marcadores associados a QTLs para seleção assistida. Isso é um ponto em comum quando se estudam caracteres de herança complexa: não tem sido possível transferir os resultados da análise de QTLs para outros ambientes, anos ou outras populações, justamente devido à herança quantitativa de caracteres como a produtividade (Cobb et al., 2019).

Em nosso estudo, 6 QTLs associados à produtividade foram detectados. Tanto para as RILs SSD quanto para as RILs Bulk, os QTLs não foram os mesmos identificados nos anos distintos, porém esses QTLs foram identificados nos mesmos cromossomos, como no caso dos QTLs no cromossomo 6, (qGYLD6.1 e qGYLD6.2) das RILs SSD para o segundo ano e análise conjunta, e para as RILs Bulk, o QTL (qGYLD6) identificado no 2º ano.

Uma explicação para o ocorrido, é que os QTLs de efeitos mais pronunciados foram os detectados, principalmente devido ao método de análise utilizada, que foi o mapeamento múltiplo, considerando o valor de LOD maior ou igual a 3. Esses QTLs não foram identificados na mesma região genômica do cromossomo 6, e os alelos do marcador (SNP_6_9538767; SNP_6_17931455) não co-segregam nos mesmos blocos de haplótipos. Raghavan et al. (2017) também encontraram QTLs no cromossomo 6 relacionados ao tempo de florescimento e a produtividade de grãos sendo explicados por 22,97% da variação fenotípica. Quanto ao QTL obtido para o cromossomo 9 (qGYLD9) nas RILs SSD e RILs Bulk, ambos foram identificados apenas para o 2º ano, representando variação fenotípica de apenas 7,45% (RILs SSD) e 11,39% (RILs Bulk), podendo ser um indicativo que esta região do cromossomo 9 tenha de alguma forma associação com o caráter, como citado também por Swamy et al. (2013), que encontraram quatro QTLs associados a produtividade de grãos sob seca, explicando de 6 a 19% da variância fenotípica e foram consistentes nas avaliações. Um desses QTL foi identificado no cromossomo 9 que apresentou alta variação fenotípica explicada não só para produtividade, mas também para fertilidade de espiguetas e altura de plantas sob condições de déficit hídrico.

O QTL qGYLD7 (cromossomo 7) das RILs Bulk foi detectado apenas pela análise conjunta, e como na análise conjunta os valores de produtividade são resultado da média dos valores genotípicos dos dois anos individuais, esperaria-se que uma marca detectada na análise conjunta, também o seria em pelo menos um dos anos de avaliação. Porém, os contrastes do fenótipo entre as RILs é modificado pelo valor da média dos anos, encontrando um valor distinto, resultante da interação GxA, então, isso acaba por gerar

uma matriz de dados que equivaleria a uma terceira população. Contudo, o dado da análise conjunta, apesar de ser uma estimativa, pode refletir a identificação de locos de QTLs mais estáveis. De acordo com Wang et al. (2019), análises conjuntas ao longo de vários anos e locais podem ser usadas para determinar a estabilidade do QTL e estimar a interação entre o QTL aditivo e o ambiente .

A repetição das associações ao longo dos ambientes e/ou anos é tomada como evidência da ausência de interação QTLs x ambiente/anos, enquanto a detecção da associação em apenas um ou alguns dos ambientes ou anos é considerada como evidência de interação (Lynch & Walsh, 1998). No estudo de Zhu et al. (2017) na cultura do arroz, em uma população de RILs advindas de cruzamento biparental, cinco QTLs estiveram associados ao caráter produtividade de grãos e quatro destes QTLs foram consistentes dentro dos dois anos avaliados e distribuídos nos mesmos cromossomos, além do que todos os alelos de efeito positivo para o caráter foram oriundo apenas dos genitores.

Não surpreendentemente, ambos os QTLs associados a PG, tanto para RILs SSD quanto para RILs Bulk, produziram efeitos aditivos, sugerindo efeito positivo pelo genitor Epagri 108 sobre o caráter, uma vez que este genitor se adapta melhor as condições de produção por ser uma cultivar brasileira, além de ser o genitor mais produtivo. Entretanto, para as RILs Bulk, no QTL do cromossomo 7, o genitor Irat 122 foi o que contribuiu com o alelo favorável.

Para altura de plantas o QTL foi identificado apenas nas RILs SSD, e o genitor Epagri 108 foi o que contribuiu com o alelo favorável, reduziu a altura das RILs que possuíam esses alelos. Este fato é desejado pelo programa de melhoramento, pois a altura reduzida implica em menores condições de ocorrer acamamento e tombamento das plantas no campo, viabilizando melhores condições de competição por luz e espaço entre elas, e reduzindo as perdas de grãos na colheita (Alvarez et al., 2014). Raghavan et al. (2017) encontraram 15 QTLs para o caráter altura de plantas nos cromossomos 1, 3, 4, 5, 6, 7, 11 e 12, entretanto 23% da variância fenotípica foi explicada pelos QTLs presentes nos cromossomos 1, 3, 5 e 11, sendo que somente o cromossomo 1 era responsável por 9,8% da variação. No presente estudo, o QTL encontrado para altura também foi no cromossomo 1, e foi responsável por 35,72% da variação fenotípica e LOD de 14,01 sendo considerado alto, entretanto não foi localizado no mesmo bloco do estudo dos autores.

Nas RILs SSD e nas RILs Bulk foram identificados os blocos haplotípicos contendo os QTLs significativos para PG e ALT. Foram determinadas as suas regiões

gênicas, suas funções putativas e os termos GO (em atributos de componente celular, função molecular e processo biológico). Para as RILs SSD, o QTL (qGYLD6.1) identificado no ano 2018 está dentro de um bloco que contém 32 genes descritos, porém este pico de QTL não se encontra dentro de nenhum gene específico. Também o QTL qGYLD6.2 (identificado na análise conjunta), não está localizado dentro de gene, mas em região intergênica.

Já o QTL qGYLD9 está localizado dentro do gene (LOC_Os09g39240), em região de éxon, e é traduzido em uma proteína com função ainda não determinada. A variante do marcador SNP_9_22544728 do gene LOC_Os09g39240 tem impacto de efeito modificador no gene. Quanto ao caráter ALT, o QTL qPTHT1 está posicionado no gene (LOC_Os01g66630) que tem função relacionada a proteína quinase semelhante a receptor do domínio S (SRK). SRK compreende uma sub-família altamente polimórfica de quinases semelhantes a receptores (RLKs) e foi previamente relacionada com aumento da produtividade e da altura em arroz (Zou et al., 2015). A variante do marcador SNP_1_38700162 do gene LOC_Os01g66630 é considerado de alto impacto, tendo a função disruptiva na proteína ou a perda de função, pois está relacionada com o códon de parada prematura.

Nas RILs Bulk dois QTLs foram identificados para o caráter PG no 2º ano. O qGYLD6 foi identificado no cromossomo 6, em que o marcador significativo esteve posicionado dentro do gene LOC_Os06g04450 com função putativa de proteína de transporte da família Sec1, cuja função ainda não foi descrita em plantas. O QTL qGYLD9 no cromossomo 9 que está localizado no gene LOC_Os09g21110 responsável pela síntese da enzima leucyl-tRNA synthetase, e que apesar de já identificada em Arabidopsis, milho, sorgo e uva, sua função tem sido mais estudada em humanos, e relacionada com a regulação do crescimento celular via rota do complexo TOR (Son et al., 2019). Um dos genes do complexo TOR foi superexpresso em arroz na Embrapa Arroz e Feijão, e deu origem a plantas mais produtivas (TOR_Evento 7, dados não publicados). As variantes dos marcadores que identificaram os genes (LOC_Os06g04450 e LOC_Os09g21110) são de impacto modificador e estão relacionadas a (3'UTR) que são variantes não codificantes ou que não afetam os efeitos dos genes. O QTL qGYLD7 identificado na análise conjunta, não está localizado dentro de nenhum gene. Entretanto, na literatura há relatos do QTL (Ghd7) que regula o tempo de florescimento e o potencial produtivo (Gw7) no cromossomo 7. O efeito desse QTL afeta diretamente a produtividade, já que está fortemente

relacionado ao florescimento tardio das plantas proporcionando baixos índices de colheita (Yonemaru et al., 2017).

Para cada cromossomo em que foram identificados os QTLs posicionados dentro de blocos haplotípicos, foram determinadas as posições iniciais e finais de cada bloco, com a finalidade de verificar no banco de dados do Q-TARO os QTLs descritos na literatura dentro dessas mesmas posições para os caracteres PG e ALT. Foram identificados os QTLs relacionados à produtividade de grão referentes a panícula (tns6), número de espiguetas por panícula (sd6), rendimento de grão (qGY6-1 e yldp 9.1) e peso de 1000 grãos (gw-6). Para o caráter ALT os QTL encontrados têm relação com a produção da planta (yd1), número de grãos/planta (ng1) e fertilidade das espiguetas (sf1). Estudos realizados por Liao et al. (2001), Zhu et al. (2017) e Deshmunk et al. (2018), também identificaram QTLs em populações RILs de arroz nos mesmos cromossomos e para os mesmos caracteres avaliados, porém não nas mesmas posições físicas.

Métodos de mapeamento de populações derivadas de cruzamentos biparentais amostram apenas uma pequena fração de todos os possíveis alelos em uma determinada espécie. Por esse motivo, marcadores moleculares associados a QTLs geralmente só podem ser usados nas populações em que foram desenvolvidos (Yang et al., 2019, Ogawa et al., 2018).

Associações entre alelos de marcadores e alelos de QTLs podem ser usadas para selecionar, indiretamente, alelos favoráveis de QTLs. Para caracteres de baixa herdabilidade, a seleção fenotípica é realizada em gerações mais avançadas, pois a herdabilidade e a acurácia estatística da estimação das médias das progênes aumentam com o aumento do número de repetições, gerações, locais e anos de teste, o que leva a um aumento muito grande do número de plantas a serem avaliadas (Dixit et al., 2019). Assim, os melhoristas devem optar entre produzir estimativas mais precisas da média de um número menor de progênes ou amostrar um grande número de progênes, com estimativas menos precisas. A SAM apresenta-se como uma alternativa eficiente para permitir a seleção precoce de características de baixa herdabilidade, diminuindo o número de plantas a ser analisado por família, o que permite a avaliação de um número muito maior de progênes. A SAM é mais efetiva em gerações iniciais de seleção entre progênes de cruzamentos entre linhas endogâmicas, pois nessas gerações a herdabilidade é baixa e o desequilíbrio de ligação é alto (Guimarães et al., 2009).

Foram identificados nesse trabalho, nas RILs advindas de SSD e Bulk, blocos de haplótipos (alelos em diferentes locos que tendem a serem herdados conjuntamente) que indicam a associação da região genômica aos caracteres avaliados. As próximas etapas do trabalho serão o desenvolvimento de ensaios TaqMan a partir dos SNPs identificados, e posteriormente utilizados na validação utilizando um grupo de acessos da CNAE já caracterizados quanto a produtividade de grão. Espera-se, desse modo, aplicar esses ensaios na rotina de seleção assistida por marcadores do programa de melhoramento genético do arroz para a seleção de linhagens com maior potencial produtivo para uso em blocos de cruzamentos, e na seleção das potenciais melhores progênies nas gerações segregantes iniciais (F_2 e F_3).

3.5 CONCLUSÕES

11) O cruzamento Epagri 108 x Irat 122 foi capaz de gerar RILs fenotipicamente superiores aos genitores para os caracteres avaliados, além de possuírem boa distribuição dos alelos de ambos os genitores.

2) O método SSD foi o mais eficiente na geração de RILs mais produtivas e as com menor altura. Porém, devido ao maior custo de desenvolvimento dessas RILs por SSD, o método Bulk também pode ser uma opção interessante, já que também foram identificadas linhagens transgressivas por esse método.

3) Ambas as RILs derivadas de Bulk e SSD permitiram a identificação de QTLs para o caráter produtividade de grão, enquanto que para altura de plantas o QTL identificado foi somente no grupo de RILs derivadas de SSD.

4) O marcador SNP_6_9538767 para o caráter produtividade de grãos, e o marcador SNP_1_38700162 para o caráter altura de plantas, são os indicados para a etapa de validação para uso na seleção assistida.

3.6 REFERÊNCIAS

ABADIE, T.; CORDEIRO, C. M. T.; FONSECA, J. R.; ALVES, R. B. N.; BURLE, M. L.; BRONDANI, C.; RANGEL, P. H. N.; CASTRO, E. M.; SILVA, H. T.; FREIRE, M. S.;

ZIMMERMANN, F. J. P.; MAGALHÃES, J. R. **Construção de uma coleção nuclear de arroz para o Brasil. Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 40, n. 2, p. 129-136, fev. 2005.

ABDOLLAHI-ARPAHAHI, R.; PAKDEL, A.; NEJATI-JAVAREMI, A.; SHAHRBABA, M. M.; GHAFOURI-KESBI, F. The relation between the genetic architecture of quantitative traits and long-term genetic response. **Journal of Applied Genetics**, Poznań, v. 55, n. 3, p. 373-381, 2014.

ABREU, A. de F.B. **Predição do potencial genético de populações segregantes do feijoeiro utilizando genitores inter-raciais.** (Tese – Doutorado em Genética e Melhoramento de Plantas). Lavras: UFLA, p.79, 1997.

ALLENDORF, F. W. Genetic drift and the loss of alleles versus heterozygosity. **Zoo Biology**, New York, v. 5, n. 2, p. 181-190, 1986.

ALVAREZ, R. C. F.; CRUSCIOL, C. A. C.; NASCENTE, A. S. Produtividade de arroz de terras altas em função de reguladores de crescimento. **Revista Ceres**, Viçosa, v. 61, n.1, p. 042-049, jan/fev, 2014.

BAEZINGER, P. S.; PETERSON, C. J. Genetic variation: Its origin and use for breeding self-pollinated species. **Raleigh: C.A.B. International**, Cap.4, p.69-92, 1991.

BAGHERI, N.; JELODAR, N. B. Heterosis and Combining Ability Analysis for Yield and Related-Yield Traits in Hybrid Rice. **International Journal of Biology**, v. 2, No. 2; July, 2010.

BARRETT, J.C.; FRY, B.; MALLER, J.; DALY, M.J. Haploview: analysis and visualization of LD and haplotype maps. **Bioinformatics**, Oxford, v. 21, n. 2, p. 263-265, 2005.

BERNARDO, R. **Breeding for quantitative traits in plants.** Woodbury; Stemma Press, 2ª Ed., p. 400, 2010.

BERTOLDO, J.G.; COIMBRA, J.L.M.; GUIDOLIN, A.F.; NODARI, R.O.; ELIAS, H.T.; BARILLI, L.D.; VALE, N.M.do; ROZZETTO, D.S. Rendimento de grãos em feijão preto: o componente que mais interfere no valor fenotípico é o ambiente. **Ciência Rural**, v.39, p.1974-1982, DOI: 10.1590/ S0103-84782009005000166, 2009.

BORÉM, A. **Melhoramento de plantas.** Viçosa: UFV, 547 p. 1997.

BORÉM, A.; MIRANDA, G.V. **Melhoramento de Plantas.** 5.ed. Viçosa: Editora UFV, 529p, 2009.

BORGES, V.; SOARES, A. A.; REIS, M. S.; RESENDE, M. D. V.; CORNÉLIO, V. M. O.; LEITE, N. A.; VIEIRA, A. R. Desempenho genotípico de linhagens de arroz de terras altas utilizando metodologia de modelos mistos. **Bragantia**, Campinas, v. 69, n. 4, p. 833841, dez. 2010.

CASTANHEIRA, A.L.M.; SANTOS, J.B. RAPD marker assessment of self-pollinated inbreeding methods for common bean segregant populations. **Crop Breeding and Applied Biotechnology**, v.4, p.16, 2004.

CHEN, H.; HE, H.; ZHOU, F.; YU, H.; DENG, X. W. Development of genomics-based genotyping platforms and their applications in rice breeding. **Current Opinion in Plant Biology**, v. 16, n. 2, p. 247-254, 2013.

CHIORATO, A.F.; CARBONELL, S.A.M.; DIAS, L.A. dos S.;RESENDE, M.D.V. Prediction of genotypic values and estimation of genetic parameters in common bean. CINGOLANI, P.; PLATTS, A.; WANG, L.L.; COON, M.; NGUYEN, T.; WANG, L.; LAND, S. J.; LU, X.; RUDEN, D. M. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. **Fly**, Austin, v. 6, n. 2, p. 1-13, 2012.

COBB, J.N.; BISWAS, P.S.; PLATTEN, J.D. Back to the future: revisiting MAS as a tool for modern plant breeding. **Theoretical and Applied Genetics**, v.132(3), p.647-667, 2019.

COELHO, A.D.F., CARDOSO, A.A., CRUZ, C.D., ARAÚJO, G.A.A., FURTADO, M.R., AMARAL, C.L.F. Herdabilidades e correlações da produção do feijão e dos seus componentes primários, nas épocas de cultivo da primavera-verão e do verão-outono. **Ciência Rural**, Santa Maria, 32 (2):211-216, 2002.

COIMBRA, J.L.M.; KOPP, M.M.; SOUZA, V.Q. de; BENIN, G.;MARCHIORO, V.S;CARVALHO, F.I.F. de; OLIVEIRA, A.C. de. Prediction of genetic value in F3populations of *Avena sativa* L. using REML/BLUP. **Crop Breeding and Applied Biotechnology**, v.5, p.265-271, DOI: 10.12702/1984-7033, v. 05 n 03 a 02, 2005.

CORTE HR, RAMALHO MAP, GONÇALVES FMA; ABREU AFB. Natural selection for grain yield in dry bean populations bred by the Bulk method. **Euphytica** 123: 387-393, 2002.

CRAN. **Rproject.org/package=LDcorSV**. Acesso em setembro de 2019.

CROW, J. F. Dominance and overdominance. In: COORS, J. G.; PANDEY, S. **The genetics and exploitation of heterosis in crops**. Cidade do México: CIMMYT, p. 49-58, 1998.

CROWELL, S.; KORNILIEV, A.; FALCÃO, A.; ISMAIL, A.; GREGORIO, G.; MEZEY, J.; McCOUCH, S. Genome-wide association and high-resolution phenotyping link *Oryza sativa* panicle traits to numerous trait-specific QTL clusters. **Nature Communications** |7:10527|DOI: 10.1038/ncomms10527, 2015.

CRUZ, C. D.; RAGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. 4 ed. Viçosa, MG: UFV, v. 2. 514 p, 2012.

CRUZ, C. D.; REGAZZI, A. J. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa, Minas Gerais: Ed. UFV, p. 390, 1994.

DEL PELOSO, M. J.; MELO L.C. **Potencial de rendimento da cultura do feijoeiro comum**. Santo Antônio de Goiás: Embrapa Arroz e Feijão, 2005.

DESHMUKH, V.; MANKAR, S.P.; MUTHUKUMAR, C.; . DIVAHAR, P.; . BHARATHI, A.; THOMAS, H.B.; RAJURKAR, SELLAMUTHU , R.; A.; POORNIMA, R.; SENTHIVEL, S.; BABU, R. C. Genome-wide consistent molecular markers associated with phenology, plant production and root traits in diverse rice (*Oryza sativa* L.) accessions under drought in rainfed target populations of the environment. **Current Science**, VOL. 114, NO. 2, 25 January, 2018.

DESROUSSEAUX, D.; SANDRON. F.; SIBERCHICOT, A.; CIERCO-AYROLLES, C.; MANGIN, B. LDcorSV: **Linkage disequilibrium corrected by the structure and the relatedness**. R package version 1.3.1., 2013.

DIXT, S.; SINGH, M.U.; ABBAI, R.; RAM, T.; SINGH, A. P.; VIRK, P. S.; KUMAR, A. Identification of genomic region(s) responsible for high iron and zinc content in rice. **Nature**, Scientific RepoRts 9:8136, 1038/s41598-019-43888, 2019.

DUARTE, J. B.; VENCOVSKY, R. Estimaco e predico por modelo linear misto com ênfase na ordenaco de mdias de tratamentos genticos. **Scientia Agricola**, Piracicaba, v. 58, n. 1, p. 109-117, 2001.

ENCOVSKY, R. Herana quantitativa. In: PATERNIANI, E. (Coord.). **Melhoramento e a produo de milho no Brasil**. Piracicaba: Fundaco Cargill, p.137-214, 1987.

FAHIN, M.; DHANAPALA, M.P.; SENADHIRA, D.; LAWRENCE, M. J. Quantitative genetics of rice II. A comparison of the efficiency of four breeding methods. **Elsevier**, Field Crops Research 55: 257-266, 1998.

FALCONER, D. S. **Introduo à gentica quantitativa**. Viosa: UFV, 279 p, 1987.

FALCONER, D. S.; MACKAY, T. F. **Introduction to quantitative genetics**. Londres: Longman, 464 p, 1996.

FAN, C.; XING, Y.; MAO, H.; LU, T.; HAN, B.; XU, C.; LI, X.; ZHANG, Q. GS3, a Major QTL for Grain Length and Weight and Minor QTL for Grain Width and Thickness in Rice, Encodes a Putative Transmembrane Protein. TAG. **Theoretical and Applied Genetics**, v. 112, n. 6, p. 1164–1171, abr. 2006.

FARIA LC, MELO PGS, PEREIRA HS, DEL PELOSO MJ, et al.. Genetic progress during 22 years of improvement of carioca-type common bean in Brazil. **Field Crops Res.** 142: 68-74, <http://dx.doi.org/10.1016/j.fcr.2012.11.016>, 2013.

FERREIRA, M. E.; GRATTAPAGLIA, D. **Introduo ao uso de marcadores moleculares em anlise gentica**. 3 ed., Braslia: Embrapa – Cernagen, 220 p, 1998.

FOUILLOX, G.; BANNERROT, H. Selection methods in the common bean (*Phaseolus vulgaris* L.). In: GEPTS, P. **Genetic resources of Phaseolus beans**. Dordrecht: Kluwer Academic, P. 503-541, 1988.

- FREITAS, I.L. de J.; AMARAL JÚNIOR, A.T. do; VIANA, A.P.; PENA, G.F.; CAB A,P.da S.; VITTORAZZI, C.; SILVA, T.R. da C. Ganho genético avaliado com índices de seleção e com REML/Blup em milho-pipoca. **Pesquisa Agropecuária Brasileira**, v.48, p.1464-1471, DOI: 10.1590/S0100-204X2013001100007, 2013.
- FUNADA, M.; HELMS, T.C.; HAMMOND, J.J.; HOSSAIN, K.; DOETKOTT, C. Single-seed descent, single-pod, and Bulk sampling methods for soybean. **Euphytica**, P. 192:217–226, 2013.
- GOFF, S.A.; RICKE, D.; LAN, T.-H.; PRESTING, G.; WANG, R.; DUNN, M.; GLAZEBROOK, J.; SESSIONS, A.; OELLER, P.; VARMA, H. A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *japonica*). **Science**, Washington, v.296, n.5565, p.92-100, 2002.
- GONÇALVES FMA, RAMALHO MAP AND ABREU FB. Natural selection in four common bean traits. **Crop Breed. Appl. Biotechnol.** 1: 213-220, <http://dx.doi.org/10.13082/1984-7033.v01n03a01>, 2001.
- GUIMARÃES, C. T.; MAGALHÃES, J. V.; LANZA, M. A.; SCHUSTER, I. Marcadores moleculares e suas aplicações no melhoramento genético. **Informe Agropecuário**, Belo Horizonte, v.30, n.253, nov./dez., 2009.
- GUO, L.; GAO, Z.; QIAN, Q. Application of resequencing to rice genomics, functional genomics and evolutionary analysis. **Rice**, New York, v. 7, n. 1, p. 4, 2014.
- GUPTA, P. K.; RUSTGI, S.; KULWAL, P. L. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. **Plant molecular Biology**, Amsterdam, n. 57, p. 461- 485, jan., 2005.
- HALLAUER, A. R; CARENA, M. J.; MIRANDA FILHO, J. B. Quantitative Genetics in Maize Breeding. **Springer**, New York, p. 680, 2010.
- HAN, B.; HUANG, X. Sequencing-based genome-wide association study in rice. **Current Opinion in Plant Biology**, London, v. 16, n. 2, p. 133-138, 2013.
- HAN, B.; ZHANG, Q. Rice genome research: current status and future perspectives. **The Plant Genome Journal**, New York, v. 1, n. 2, p. 71-76, 2008.
- HAN, B.; ZHANG, Q. Rice Genome Research: Current Status and Future Perspectives. **The Plant Genome Journal**, v. 1, n. 2, p. 71-76, 2008.
- HASSEN, M. B.; CAO, T.V.; BARTHOLOMÉ, J.; ORASEN, G.; COLOMBI, C.; RAKOTOMALALA, J.; RAZAFINIMPIASA, L.; BERTONE, C.; BISELLI, C.; VOLANTE, A.; DESIDERIO, F.; JACQUIN, L.; VALÈ, G.; AHMADI, A. Rice diversity panel provides accurate genomic predictions for complex traits in the progenies of biparental crosses involving members of the panel. **Theoretical and Applied Genetics**, 131:417–435, 2018.

HAYASHI, K.; HASHIMOTO, N.; DAIGEN, M.; ASHIKAWA, I. Development of PCRbased SNP markers for rice blast resistance genes at the Piz locus. **Theoretical and Applied Genetics**, Berlin, v. 108, n. 7, p. 1212-1220, 2004.

HILL, W. G.; WEIR, B. S. Variances and covariances of squared linkage disequilibria in finite populations. **Theor. Popul. Biol.**, v. 33, p. 54–78, 1988.

HU, J.; WANG, Y.; FANG, Y.; ZENG, L.; XU, J.; YU, H.; SHI, Z.; PAN, J.; ZHANG, D.; KANG, S.; ZHU, L.; DONG, G.; GUO, L.; ZENG, D.; ZHANG, G.; XIE, L.; XIONG, G.; LI, J.; QIAN, Q. A Rare Allele of GS2 Enhances Grain Size and Grain Yield in Rice. **Molecular Plant**, v. 8, n. 10, p. 1455–1465, 5 out. 2015.

HUANG, M.; CHEN, L.; CHEN Z. Diallel analysis of combining ability and heterosis for yield and yield components in rice by using positive loci. **Euphytica**, 205 (1): 37-50, 2015.

HUANG, X.; FENG, Q.; QIAN, Q.; ZHAO, Q.; WANG, L.; WANG, A.; GUAN, J.; FAN, D.; WENG, Q.; HUANG, T.; DONG, G.; SANG, T.; HAN, B. High-throughput genotyping by whole-genome resequencing. **Genome Research**, v. 19, n. 6, p. 1068–1076, jun. 2009.

IKEDA, K.; ITO, M.; NAGASAWA, N.; KYOZUKA, J.; NAGATO, Y. Rice ABERRANT PANICLE ORGANIZATION 1, Encoding an F-Box Protein, Regulates Meristem Fate. **The Plant Journal: For Cell and Molecular Biology**, v. 51, n. 6, p. 1030–1040, set. 2007.

JÚNIOR, V. A. P.; MELO, P. G. S.; PEREIRA, H. S.; BASSINELLO, P. Z.; MELO, L. C. Genetic potential of common bean progenies selected for crude fiber content obtained through different breeding methods. **Genetics and Molecular Research**, 14 (2): 5763-5774, 2015.

KANBAR, A.; KONDO, K.; SHASHIDHAR, H. E. Comparative efficiency of pedigree, modified Bulk and single seed descent breeding methods of selection for developing high-yielding lines in rice (*Oryza sativa* L.) under aerobic condition. **Electronic Journal of Plant Breeding**, 2(2):184-193 . June, 2011.

KAWAHARA, Y.; BASTIDE, M.; HAMILTON, J. P.; KANAMORI, H.; MCCOMBIE, W. R.; OUYANG, S. et al. Improvement of the *Oryza Sativa* Nipponbare reference genome using next generation sequence and optical map data. **Rice**, New York, v. 6, n. 1, p. 1-10, 2013.

KERVELLA, J. & FOUILLOUX, G. A theoretical study of the Bulk breeding method I. Importance and consequences of losses due to sampling. **Euphytica**, 60: 185-195, 1992.

KHUSH, G.S. Origin, dispersal, cultivation and variation of rice. **Plant Molecular Biology**, Zurich, v. 35, n. 1/2, p. 25-34, 1997.

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.; WENZL, P.; HUTTER, E.; CARLING, J.; XIA, L.; BLOIS, H.; CAIG, V.; HELLER-USZYNSKA, K.; JACCOUD, D.; HOPPER, C.; ASCHENBRENNERKILIAN, M.; EVERS, M.; PENG, K.; CAYLA, C.; HOK, P.; USZYNSKI, G. Diversity arrays

technology: a genetic genome profiling technology on open platforms. In: POMPANON, F.; BONIN, A. **Data Production and Analysis in Population Genomics: Methods and Protocols**. 1^a ed. Totowa: Humana Press, cap 5. p. 6789, 2012.

KOMATSU, K.; MAEKAWA, M.; UJIIE, S.; SATAKE, Y.; FURUTANI, I.; OKAMOTO, H.; SHIMAMOTO, K.; KYOZUKA, J. LAX and SPA: Major Regulators of Shoot Branching in Rice. **Proceedings of the National Academy of Sciences**, v. 100, n. 20, p. 11765–11770, 30 set. 2003.

KREMER, A., POTTS, B. M., DELZON, S. Genetic divergence in forest trees: understanding the consequences of climate change. **Functional Ecology**, v. 28, p. 22–36, 2014.

KUMAR, S.; RAULIER, P.; CHAGNÉ, D.; WHITWORTH, C. Molecular-level and traitlevel differentiation between the cultivated apple (*Malus× domestica* Borkh.) and its main progenitor *Malus sieversii*. **Plant Genetic Resources**, p. 1-11. Disponível em <<http://dx.doi.org/10.1017/S1479262114000136>>. Acesso em: 9 jun., 2014.

LI, J.-Y.; WANG, J.; ZEIGLER, R. S. The 3,000 rice genomes project: new opportunities and challenges for future rice research. **GigaScience**, Londres, v. 3, n. 1, p. 8, 2014.

LI, L.; LU, K.; CHEN, Z.; MU, T.; HU, Z.; LI, X. Dominance, overdominance and epistasis condition the heterosis in two heterotic rice hybrids. **Genetics** 180: 1725–1742. DOI: 10.1534/genetics.108.091942, 2008.

LIANG, S.; SUN, C.; REN, G.; ZHAO, X.; ZHOU, M.; MCNEIL, D.; YE, G. Usefulness of the cloned and fine-mapped genes/QTL for grain yield and related traits in indica rice breeding for irrigated ecosystems. **Field Crops Research**, v. 187, p. 58–73, 15 fev. 2016.

LIAO, C. Y., WU, P., HU, B. & YI, K. K. Effects of genetic background and environment on QTLs and epistasis for rice (*Oryza sativa* L.) panicle number. **Theor Appl Genet** 103, 104–111, 2001.

LIU, B.H. **Statistical genomics: linkage, mapping and QTL analysis**. Boca Raton: CRC Press, p. 611, 1998.

LIU, T.; YU, T.; XING, Y. Identification and validation of a yield-enhancing QTL cluster in rice (*Oryza sativa* L.). **Euphytica**, v. 192, n. 1, p. 145-153, 2013.

LORIEUX, M. **Tutorial MapDisto**, version 1.7, 2014.

LU, L.; YAN, W.; XUE, W.; SHAO, D.; XING, Y. Evolution and Association Analysis of *Ghd7* in Rice. **PLoS One**, v. 7, n. 5, p. e34021, 30 maio 2012.

LYNCH, M.; WALSH, B. **Genetics And Analysis Of Quantitative Traits**. Sunderland: Sinauer Associates, p. 980, 1988.

MA, X.; FU, Y.; ZHAO, X.; JIANG, L.; ZHU, Z.; GU, P.; XU, W.; SU, Z.; SUN, C.; TAN, L. Genomic structure analysis of a set of *Oryza nivara* introgression lines and

identification of yield-associated QTLs using whole-genome resequencing. **Scientific RepoRts.** | 6:27425 | DOI: 10.1038/srep27425, 2016.

MAO, H.; SUN, S.; YAO, J.; WANG, C.; YU, S.; XU, C.; LI, X.; ZHANG, Q. Linking Differential Domain Functions of the GS3 Protein to Natural Variation of Grain Size in Rice. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 45, p. 19579–19584, 9 nov. 2010.

MARCHESAN, C. B. **Análise de um cruzamento dialélico parcial em pimentão visando caracteres agrônômicos e resistência ao Oídio** [Leveillula taurica (Lév.) Arn.]. 60 f. Dissertação (Mestre em Agricultura Tropical e Subtropical. Área de Concentração em Genética, Melhoramento Vegetal e Biotecnologia.) - Instituto Agrônômico, Campinas, São Paulo, 2008.

MATSUMOTO, T.; WU, J.; ITOH, T.; NUMA, H.; ANTONIO, B.; SASAKI, T. THE map-based sequence of the rice genome. **Nature**, v. 436, n. 7052, p. 793–800, 2005.

MCCOUCH, S. R.; CHEN, X.; PANAUD, O.; TEMNYKH, S.; XU, Y.; CHO, Y. G.; HUANG, N.; ISHII, T.; BLAIR, M. Microsatellite marker development, mapping and applications in rice genetics and breeding. **Plant molecular biology**, The Hague, v. 35, n. 1-2, p. 89-99, 1997.

MENDES MP, RAMALHO MAP ; ABREU AFB. Strategies in identifying individuals in a segregant population of common bean and implications of genotype x environment interaction in the success of selection. **Genet. Mol. Res.** 11: 872-880. PubMed <http://dx.doi.org/10.4238/2012.April.10.3>, 2011.

MENDES, M.P.; RAMALHO, M.A.P.; ABREU, Â. de F.B. Strategies for selecting individuals in common bean breeding programs. **BeanImprovementCooperative**, v.54, p.68-69, 2011.

MENDIMAGH, S.; FELAH, M.; GAQZZAH, M. Barley breeding for quality improvement in Tunisia. **African Journal of Biotechnology**, Vol. 11(89), pp. 15516-15522, 2012.

MILADINOVIĆ, J. BURTON, J. W.; BALEŠEVIĆ TUBIĆ¹, S.; MILADINOVIĆ, D. DJORDJEVIĆ, V.; DJUKIC, V. Soybean breeding: comparison of the efficiency of different selection methods. **Tübitak**, Turk J Agric For 35: 469-480, 2011.

MIURA, K.; IKEDA, M.; MATSUBARA, A.; SONG, X.-J.; ITO, M.; ASANO, K.; MATSUOKA, M.; KITANO, H.; ASHIKARI, M. OsSPL14 Promotes Panicle Branching and Higher Grain Productivity in Rice. **Nature Genetics**, v. 42, n. 6, p. 545–549, jun. 2010.

MONZÓN, D. L.R. **Variabilidade genética de germoplasma de arroz (Oryza sativa L.) para tolerância ao frio na germinação**. Tese, 2015.

- MUEHLBAUER, F. J; BURNELL, D. G; BOGYO T. P, BOGYO M. T. Simulated comparisons of single seed descent and bulk population breeding methods. **Crop Sci** 21:572–577, 1981.
- NASS, L. L.; VALOIS, A. C.C.; MELO, I. S.; VALADARES-INGLIS, M. C. **Recursos Genéticos e Melhoramento de Plantas**. Fundação MT, Rondonópolis, 2001.
- NEVES, P. C. F.; TAILLEBOIS, J.; MOURA NETO, F. P. Seleção para aptidão à produção de sementes em cruzamento e outros atributos em população recorrente de arroz (*Oryza sativa* L.). In: CONGRESSO BRASILEIRO DE ARROZ IRRIGADO, 10, 2017, Gramado. **Anais...** Porto Alegre: SOSBAI, p. 1-4, 2017.
- NUNES, J.A.R.; RAMALHO, M.A.P.; FERREIRA, D.F. Inclusion of genetic relationship information in the pedigree selection method using mixed models. **Genetics and Molecular Biology**, v.31, p.73-78, DOI: 10.1590/ S1415-47572008000100015, 2008.
- OGUNBAYO, S. A.; SIÉ, M.; OJO, D. K.; SANNI, K. A.; AKINWALE, M. G.; TOULOU, B.; SHITTU, A.; IDEHEN, E. O.; POPOOLA, A. R.; DANIEL, I. O.; GREGORIO, G. B. Genetic variation and heritability of yield and related traits in promising rice genotypes (*Oryza sativa* L.). **Journal of Plant Breeding and Crop Science**, Nairobi, v. 6, n. November, p. 153-159, 2014.
- PARIDA, S. K.; MUKERJI, M.; SINGH, A. K.; SINGH, N. K.; MOHAPATRA, T. SNPs in stress-responsive rice genes: validation, genotyping, functional relevance and population structure. **BMC Genomics**, London, v. 13, n. 1, p. 426, 2012.
- PATERSON, A. H.; FREELING, M.; SASAKI, T. Grains of knowledge: Genomics of model cereals. *Genome Research*, Woodbury, v. 15, p. 1643–1650, 2010.
- PAWAR, I.S.; PARODA, R.S.; SINGH, S. A comparison of pedigree, SSD and bulk methods in two wheat crosses. **Crop improve.**, 13: 34-37, 1986.
- PINHEIRO, L.C. de M.; GOD, P.I.V.G.; FARIA, V.R.; OLIVEIRA, A.G.; HASUI, A. A.; PINTO, E.H.G.; ARRUDA, K.M.A.; PIOVESAN, N.D.; MOREIRA, M.A. Parentesco na seleção paraproductividade e teores de óleo e proteína de soja via modelos mistos. **Pesquisa Agropecuária Brasileira**, v.48, p.1246-1253, DOI: 10.1590/S0100-204X2013000900008, 2013.
- PLANETA ARROZ. **Plano safra 2019/2020 entra em vigor**. Cachoeira do Sul: Planeta Arroz. 2019. Disponível em: <<http://www.planetaarroz.com.br>>. Acesso em 10 set., 2019.
- PONTES JÚNIOR, V. A.; MELO, P. G. S.; PEREIRA, H. S.; MELO, L. C. Genetic potential of common bean progenies obtained by different breeding methods evaluated in various environments. **Genetics and Molecular Research**, 2016.
- QI, P.; LIN, Y.-S.; SONG, X.-J.; SHEN, J.-B.; HUANG, W.; SHAN, J.-X.; ZHU, M.-Z.; JIANG, L.; GAO, J.-P.; LIN, H.-X. The Novel Quantitative Trait Locus GL3.1 Controls Rice Grain Size and Yield by Regulating Cyclin-T1;3. **Cell Research**, v. 22, n. 12, p. 1666– 1680, dez. 2012.

QIU, X.; CHEN, K. ; LV, W; OU, X.; ZHU, Z.; XING, D. YANG, L.; FAN, F.; YANG, J.; XU, J.; ZHENG, T.; LI, Z. Examining two sets of introgression lines reveals backgroundindependent and stably expressed QTL that improve grain appearance quality in rice (*Oryza sativa* L.). **Theor Appl Genet.** V.130, p. 951–967, 2017.

RAGHAVAN, C.; MAULEON, R.; LACORTE, V.; JUBAY, M.; ZAW, H.; BONIFACIO, J.; SINGH, R. K.; HUANG, B. E.; LEUNG, H. Approaches in Characterizing Genetic Structure and Mapping in a Rice Multiparental Population. **G3: Genes, Genomes, Genetics**, v. 7, June 2017, 1721, 2017.

RAMALHO, M. A. P.; ABREU, Â. F. B.; DOS SANTOS, J. B. **Melhoramento de espécies autógamas.** In: NASS, L. L.; VALOIS, A. C. C.; MELO, I. S. D.; VALADARES-INGLIS, M. C. (Ed.). **Recursos genéticos e melhoramento - plantas.** Rondonópolis: Fundação MT, p. 201-230, 2001.

RAMALHO, M. A. P.; SANTOS, J.B. dos; ZIMMERMANN, M. J. **Genética quantitativa em plantas autógamas: aplicações ao melhoramento de feijoeiro.** Goinânia: UFG, p. 271, 1993.

RAMALHO, M.A.P. **Melhoramento do feijoeiro.** In: ABREU, A. de F.B.; GONÇALVES, F.M.A.; MARQUES JR., O.G.; RIBEIRO, P.H.E. (ed.). SIMPÓSIO SOBRE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS, 1997, Lavras. **Anais...** Lavras: UFLA/GEN, p.167-196, 1997.

RAMALHO, M.A.P.; ABREU, A.F.B.; SANTOS, J.B.; NUNES, J.A.R. **Aplicações da genética quantitativa no melhoramento de plantas autógamas.** Lavras: Editora UFLA, Ed. 1, 522p, 2012.

RAMALHO, M.A.P.; CARVALHO, B.L.; NUNES, J.A.R. Perspectives for the use of quantitative genetics in breeding of autogamous plants. **ISRNG enetics**, v.2013, article ID 71812, DOI: 10.5402/2013/718127, 2013.

RAMALHO, M.A.P.; GONÇALVES, F.M.A.; SOUZA SOBRINHO, F. de S. **Seleção recorrente no melhoramento do cafeeiro.** In: SIMPÓSIO SOBRE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS, 3: Genética e melhoramento do cafeeiro, 1999, Lavras. **Anais...** Lavras: UFLA/GEN, p.66-81, 1999.

RAMALHO, M.A.P.; ARAÚJO, L.C. de A. Breeding self-pollinated plants. **Crop Breeding and Applied Biotechnology**, v.11, p.1-7, DOI: 10.1590/S1984-70332011000500002, 2011.

RAMOS, M. R. F.; MENDONÇA, J. A.; VIANELLO, R. P.; MORAIS JÚNIOR, O. P.; COLOMBARI, J. M.; BORBA, T. C. O.; CASTRO, A. P.; BRONDANI, C. Heterosis And Combining Ability For Grain Yield And Earliness In Accessions Ofa Rice Core Collection. **Functional Plant Breeding Journal**, v. 1 n 1, 2019.

RAPOSO, F.V.; RAMALHO, M. A.P.; ABREU, A. F. B. COMPARAÇÃO DE MÉTODOS DE CONDUÇÃO DE POPULAÇÕES SEGREGANTES DO FEIJOEIRO. **Pesq. agropec. bras.**, Brasília, v.35, n.10, p.1991-1997, out., 2000.

REIS, G. G.; FRITSCHÉ-NETO, R.; SOARES, P. C.; CORNÉLIO, V. M. O.; REIS, M. S.; MORAIS, O. P.; MARQUES, T. S. Accuracy and genetic progress of agronomic traits in irrigated rice program in Brazil. **African Journal of Agricultural Research**, Quênia, v. 10, n. 43, p. 4032-4038, 2015.

REMINGTON, D. L.; THORNSBERRY, J. M.; MATSUOKA, Y.; WILSON, L. M.; WHITT, S. R.; DOEBLEY, J.; KRESOVICH, S.; GOODMAN, M. M.; BUCKLER, E. S. Structure of linkage disequilibrium and phenotypic associations in the maize genome. **Proceedings of the National Academy of Sciences of the United States of America**, Boston, v. 98, n. 20, p. 11479–11484, 2001.

RESENDE, M. D. V. D.; SILVA, F. F.; AZEVEDO, C. F. **Estatística Matemática, Biométrica e computacional: Modelos Mistos, Multivariados, Categóricos e Generalizados (REML/BLUP), Inferência Bayesiana, Regressão Aleatória, Seleção Genômica, QTL-GWAS, Estatística Espacial e Temporal, Competição, Sobrevivência**. Viçosa, MG, Suprema, 881 p, 2014.

RESENDE, M. D. V. **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília: Embrapa Florestas, 975 p. 2002.

RESENDE, M. D. V. **SELEGEN–REML/BLUP: sistema estatístico e seleção genética computadorizada via modelos lineares mistos**. Colombo: Embrapa Florestas, 15p. (Documentos 7), 2007.

RESENDE, M. D. V.; DUARTE, J. B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical**, Goiânia, v. 37, n. 3, p. 182-194, 2007.

RIZZI, V. **Diversidade, estruturação genética e mapeamento associativo em germoplasma japonês de arroz utilizando marcadores DArT-seq**. Tese, 2017.

ROYCE, S.W.; COPP, L.G.L.; FRENKEL, O. H.. The effectiveness of selection for yield in wheat. **Heredity**, 1: 232-233, 1947.

SAKAI, H.; LEE, S. S.; TANAKA, T.; NUMA, H.; KIM, J.; KAWAHARA, Y. et al. Rice Annotation project database (RAP-DB): an integrative and interactive database for rice genomics. **Plant Cell Physiol**, Tokyo, v. 54, n. 2, p. 1-11, e6, 2013.

SANGHERA, G. S.; KASHYAP, S. C.; PARRAY, G. A. Genetic Variation for Grain Yield and Related Traits in Temperate Red Rice (*Oryza sativa* L.) Ecotypes. **Notulae Scientia Biologicae**, v. 5, n. 3, p. 400-406, 2013.

SCHAPER, E.; ERIKSSON, A.; RAFAJLOVIC, M.; SAGITOV, S.; MEHLIG, B. Linkage disequilibrium under recurrent bottlenecks. **Genetics**, Heidelberg, v. 190, n. 1, p. 217-229, 2012.

SI, L.; CHEN, J.; HUANG, X.; GONG, H.; LUO, J.; HOU, Q.; ZHOU, T.; LU, T.; ZHU, J.; SHANGGUAN, Y.; CHEN, E.; GONG, C.; ZHAO, Q.; JING, Y.; ZHAO, Y.; LI, Y.; CUI, L.; FAN, D.; LU, Y.; WENG, Q.; WANG, Y.; ZHAN, Q.; LIU, K.; WEI, X.; AN, K.; AN, G.; HAN, B. OsSPL13 Controls Grain Size in Cultivated Rice. **Nature Genetics**, v.

advance online publication, 7 mar. 2016. Disponível em:
<<http://www.nature.com/ng/journal/vaop/ncurrent/full/ng.3518.html>>. Acesso em: 22 nov. 2019.

SILVA, A. C.; MELO, P. G. S.; MELO, L. C.; BASSINELO, P. Z.; PEREIRA, H. S. Eficiência de métodos de melhoramento para teor de fibra e produtividade de grãos em progênies de feijoeiro comum. **Bragantia**, Campinas, v. 72, n. 4, p.326-331, 2013.

SOARES, E. R.; BASEGGIO, E. A.; DA SILVA LONDRERO, L.; CORREA, S. C. S.; ROSSINI, V. P.; et al. Componentes de produção e produtividade de arroz híbrido de sequeiro comparado a três cultivares convencionais. **Acta Agronômica**, Palmira, v. 59, n. 4, p. 435-441, 2017.

SOHRABI, M.; RAFII, M. Y.; HANAFI, M. M.; AKMAR, A. S. N.; LATIF, M. A. Genetic Diversity of Upland Rice Germplasm in Malaysia Based on Quantitative Traits. **The Scientific World Journal**, v. 2012, p. 1-9, 2012.

SON, K.; YOU, J.; YOON, M. et al. Nontranslational function of leucyl-tRNA synthetase regulates myogenic differentiation and skeletal muscle regeneration. **J Clin Invest**, 129(5):2088-2093, 2019.

SWAMY, M B. P.; AHMED, H.U; HENRY, A.; MAULEON, R.; DIXIT1, S.; VIKRAM, S.; TILATTO, R.; VERULKAR, S.B; PERRAJU, P.; MANDAL, N. P; VARIAR, M.; ROBIN S.; CHANDRABABU, R.; SINGH, O. N.; DWIVEDI, J. L.; PRASAD DAS, P.; MISHRA, M. K; YADAW, R. B.; ADITYA, T. L.; KARMAKAR, B.; SATOH, K.; MOUMENI, A.; KIKUCHI, S.; LEUNG, H.; KUMAR, A. Genetic, Physiological, and Gene Expression Analyses Reveal That Multiple QTL Enhance Yield of Rice MegaVariety IR64 under Drought. **PLOS ONE**, May 2013, v. 8, Issue 5, e62795, 2013.

TEMPLETON, A. R. **Population genetics and microevolutionary theory**. Hoboken: John Wiley & Sons, 512 p. 2006.

VALLURU, R.; REYNOLDS, M. P.; SALSE, J. Genetic and Molecular Bases of YieldAssociated Traits: A Translational Biology Approach between Rice and Wheat. **Theoretical and Applied Genetics**, v. 127, n. 7, p. 1463–1489, 10 jun. 2014.

WANG, L.; YANG, X.; CUI, S.; MU, G.; SUN, X.; LIU, L.; LI, Z. QTL mapping and QTL × environment interaction analysis of multi-seed pod in cultivated peanut (*Arachis hypogaea* L.). **The Crop Journal**, 7:249–260, 2019.

WANG, M.; WANG, M.; YU, Y.; HABERER, G.; MARRI, P. R.; FAN, C.; GOICOECHEA, J. L. ZUCCOLO, A.; SONG, X.; KUDRNA, D.; AMMIRAJU, J. S. S.; COSSU, R. M.; MALDONADO, C.; CHEN, J.; LEE, S.; SISNEROS, N.; BAYNAST, K.; GOLSER, W.; WISSOTSKI, M.; KIM, W.; SANCHEZ, P.; NDJIONDJOP, M-N.; SANNI, K.; LONG, M.; CARNEY, J.; PANAUD, O. The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. **Nature Genetics**, Nova York, v. 46, n. 9, p. 982-988, 2014.

WANG, Y.; XIONG, G.; HU, J.; JIANG, L.; YU, H.; XU, J.; FANG, Y.; ZENG, L.; XU, E.; XU, J.; YE, W.; MENG, X.; LIU, R.; CHEN, H.; JING, Y.; WANG, Y.; ZHU, X.; LI, J.; QIAN, Q. Copy Number Variation at the GL7 Locus Contributes to Grain Size Diversity in Rice. **Nature Genetics**, v. 47, n. 8, p. 944–948, ago. 2015.

XIE, W.; WANG, G.; YUAN, M.; YAO, W.; LYU, K.; ZHAO, H.; YANG, M.; LI, P.; ZHANG, X.; YUAN, J.; WANG, Q.; LIU, F.; DONG, H.; ZHANG, L.; LI, X.; MENG, X.; ZHANG, W.; XIONG, L.; HE, Y.; WANG, S.; YU, S.; XU, C.; LUO, J.; LI, X.; XIAO, J.; LIAN, X.; ZHANG, Q. Breeding Signatures of Rice Improvement Revealed by a Genomic Variation Map from a Large Germplasm Collection. **Proceedings of the National Academy of Sciences**, v. 112, n. 39, p. E5411–E5419, 29 set. 2015.

XING, Y.; ZHANG, Q. Genetic and Molecular Bases of Rice Yield. **Annual Review of Plant Biology**, v. 61, p. 421–442, 2010.

XU, F.; SUN, X.; CHEN, Y.; HUANG, Y.; TONG, C.; BAO, J. Rapid Identification of Major QTLs Associated with Rice Grain Weight and Their Utilization. **PLoS One**, v. 10, n. 3, p. e0122206, 2015.

XUE, W.; XING, Y.; WENG, X.; ZHAO, Y.; TANG, W.; WANG, L.; ZHOU, H.; YU, S.; XU, C.; LI, X.; ZHANG, Q. Natural Variation in Ghd7 Is an Important Regulator of Heading Date and Yield Potential in Rice. **Nature Genetics**, v. 40, n. 6, p. 761–767, jun. 2008.

YAN, W.-H.; WANG, P.; CHEN, H.-X.; ZHOU, H.-J.; LI, Q.-P.; WANG, C.-R.; DING, Z.H.; ZHANG, Y.-S.; YU, S.-B.; XING, Y.-Z.; ZHANG, Q.-F. A Major QTL, Ghd8, Plays Pleiotropic Roles in Regulating Grain Productivity, Plant Height, and Heading Date in Rice. **Molecular Plant**, v. 4, n. 2, p. 319–330, mar. 2011.

YANG, J.; BENYAMIN, B.; MCEVOY, B. P.; GORDON, S.; HENDERS, A. K.; NYHOLT, D. R.; MADDEN, P. A.; HEATH, A. C.; MARTIN, N. G.; MONTGOMERY, G. W.; GODDARD, M. E.; VISSCHER, P. M. Common SNPs explain a large proportion of the heritability for human height. **Nat Gen**, v. 42, p. 565–569, 2010.

YANG, J.; LEE, S. H.; GODDARD, M. E.; VISSCHER, P. M. GCTA: A tool for genomewide complex trait analysis. **Am J Hum Genet**, v. 88, p. 76–82, 2011.

YANG, W.; GUO, Z.; HUANG, C.; DUAN, L.; CHEN, G.; JIANG, N.; FANG, W.; FENG, H.; XIE, W.; LIAN, X.; WANG, G.; LUO, Q.; ZHANG, Q.; LIU, Q.; XIONG, L. Combining High-Throughput Phenotyping and Genome-Wide Association Studies to Reveal Natural Genetic Variation in Rice. **Nature Communications**, v. 5, p. 5087, 8 out. 2014.

YANG, Y.; GUO, M.; SUN, S.; ZOU, Y.; YIN, S.; LIU, Y.; TANG, S.; GU, M.; YANG, Z.; YAN, C. Natural variation of OsGluA2 is involved in grain protein content regulation in rice. **Nature Communications**, 10.1038/s41467-019-09919-y, 2019.

YONEMARU, J.; MIZOBUCHI, R.; KATO, H.; YAMAMOTO, T.; YAMAMOTO, E.; MATSUBARA, K.; HIRABAYASHI, H.; TAKEUCHI, Y.; TSUNEMATSU, H.; ISHII, T.; OHTA, H.; MAEDA, H.; EBANA, K.; YANA, M. Genomic regions involved in yield potential detected by genome-wide association analysis in Japanese high-yielding rice cultivars. **BMC Genomics**, 15:346, 2014.

YU, H. ET AL. GAINS in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. **PLoS One** 6, e17595, 2011.

YU, J.; HU, S.; WANG, J.; WONG, G. K.; LI, S.; LIU, B. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). **Science**, New York, v. 296, n. 5565, p. 79-92, 2002.

YU, J.; WANG, J.; LIN, W.; LI, S. G.; LI, H.; ZHOU, J. et al. The genomes of *Oryza sativa*: a history of duplications. **PLoS Biology**, San Francisco, v. 3, n. 2, p. e38, 2005.

ZENG, Y.; JI, Z.; WEN, Z.; LIANG, Y.; YANG, C. Combination of Eight Alleles at Four Quantitative Trait Loci Determines Grain Length in Rice. **PLoS One**, v. 11, n. 3, p. e0150832, 2016.

ZHANG, S.; HE, X.; ZHAO, J.; CHENG, Y.; XIE, Z.; CHEN, Y.; YANG, T.; DONG, J.; LIU, Q.; LIU, W.; MAO, X.; FU, H.; CHEN, Z.; LIAO, Y.; LIU, B. Identification and validation of a novel major QTL for harvest index in rice (*Oryza sativa* L.). **Springer Open**. DOI 10.1186/s12284-017-0183-0, 2017.

ZHANG, X.; WANG, J.; HUANG, J.; LAN, H.; WANG, C.; YIN, C.; WU, Y.; TANG, H.; QIAN, Q.; LI, J.; ZHANG, H. Rare Allele of OsPPKL1 Associated with Grain Length Causes Extra-Large Grain and a Significant Yield Increase in Rice. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 52, p. 21534–21539, 26 dez. 2012.

ZHU, M.; LIU, D.L.; LIU, W.; LI, D.; LIAO, Y.; LI, J.; FU, C.; FU, F.; HUANG, H.; ZENG, X.; MA, X.; WANG, F. QTL mapping using an ultra-highdensity SNP map reveals a major locus for grain yield in an elite rice restorer R998. **Scientific REPORTS** | 7: 10914 | DOI:10.1038/s41598-017-10666-7, 2017.

ZOU, X.; QIN, Z.; ZHANG, C.; LIU, B.; LIU, J.; ZHANG, C.; LIN, C.; LI, H.; ZHAO, T. Over-expression of an S-domain receptor-like kinase extracellular domain improves panicle architecture and grain yield in rice. **Journal of Experimental Botany**, 66(22):7197–7209, 2015.

APÊNDICES

APÊNDICE I – Funções putativas dos genes para os cromossomos 1, 6, e 9 das RILs SSD.

Apêndice A. Função Putativa dos Genes do Cromossomo 1 para RILs SSD para o primeiro ano

Gene	Função putativa do gene
LOC_Os01g66330	ATP-dependent Clp protease ATP-binding subunit clpX, putative, expressed
LOC_Os01g66340	expressed protein
LOC_Os01g66350	DUF647 domain containing protein, putative, expressed
LOC_Os01g66360	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase, putative, expressed
LOC_Os01g66379	expressed protein
LOC_Os01g66400	expressed protein
LOC_Os01g66410	expressed protein
LOC_Os01g66420	PHD finger protein, putative, expressed
LOC_Os01g66440	expressed protein
LOC_Os01g66450	retrotransposon protein, putative, unclassified, expressed
LOC_Os01g66480	expressed protein
LOC_Os01g66490	no apical meristem protein, putative, expressed
LOC_Os01g66500	phosphoribosylformylglycinamide synthase, putative, expressed
LOC_Os01g66510	MLO domain containing protein, putative, expressed
LOC_Os01g66520	serine/threonine-protein kinase RIO-like, putative, expressed
LOC_Os01g66530	ARGOS, putative, expressed
LOC_Os01g66544	expressed protein
LOC_Os01g66560	signal recognition particle 72 kDa protein, putative, expressed
LOC_Os01g66570	ZOS1-19 - C2H2 zinc finger protein, expressed
LOC_Os01g66580	RNA polymerase III RPC4 domain containing protein, expressed
LOC_Os01g66590	DUF260 domain containing protein, putative, expressed
LOC_Os01g66600	rhodanese-like, putative, expressed
LOC_Os01g66610	serine/threonine-protein kinase receptor precursor, putative, expressed
LOC_Os01g66620	glucosyltransferase, putative, expressed
LOC_Os01g66630	S-domain receptor-like protein kinase, putative, expressed

Apêndice B . Função Putativa dos Genes no Cromossomo 6 para as RILs SSD para o segundo ano de avaliação (2018).

Genes	Função Putativa do gene
LOC_Os06g16170	expressed protein
LOC_Os06g16190	expressed protein
LOC_Os06g16200	proteins of unknown function domain containing protein, expressed
LOC_Os06g16210	expressed protein
LOC_Os06g16240	expressed protein
LOC_Os06g16250	expressed protein
LOC_Os06g16260	OsSigP5 - Putative Type I Signal Peptidase homologue; employs a putative Ser/His catalytic dyad, expressed
LOC_Os06g16270	heat shock factor binding protein 2, putative, expressed
LOC_Os06g16280	expressed protein
LOC_Os06g16290	ribosomal protein L7Ae, putative, expressed
LOC_Os06g16300	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1 precursor, putative, expressed
LOC_Os06g16330	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1 precursor, putative, expressed
LOC_Os06g16340	expressed protein
LOC_Os06g16350	peroxidase precursor, putative, expressed
LOC_Os06g16360	expressed protein
LOC_Os06g16370	CCT/B-box zinc finger protein, putative, expressed
LOC_Os06g16380	expressed protein
LOC_Os06g16390	SET domain containing protein, expressed
LOC_Os06g16400	helix-loop-helix DNA-binding domain containing protein, expressed
LOC_Os06g16410	armadillo/beta-catenin repeat family protein, putative, expressed
LOC_Os06g16420	amino acid transporter, putative, expressed
LOC_Os06g16430	FHA domain containing protein, putative, expressed
LOC_Os06g16440	expressed protein
LOC_Os06g16450	Leucine Rich Repeat family protein, expressed
LOC_Os06g16460	expressed protein
LOC_Os06g16480	expressed protein
LOC_Os06g16500	expressed protein
LOC_Os06g16540	transposon protein, putative, CACTA, En/Spm sub-class, expressed
LOC_Os06g16550	hypothetical protein
LOC_Os06g16560	expressed protein
LOC_Os06g16580	transposon protein, putative, CACTA, En/Spm sub-class, expressed
LOC_Os06g16590	hypothetical protein

Apêndice C. Função Putativa dos Genes no Cromossomo 6 para as RILs SSD para a análise conjunta dos dois anos de avaliações.

Gene	Função Putativa do Gene
LOC_Os06g30830	OsMADS76 - MADS-box family gene with M-alpha type-box, expressed
LOC_Os06g30850	expressed protein
LOC_Os06g30860	WRKY31, expressed
LOC_Os06g30870	expressed protein
LOC_Os06g30880	expressed protein
LOC_Os06g30894	expressed protein
LOC_Os06g30901	HEAT repeat family protein, putative, expressed
LOC_Os06g30910	CAAX amino terminal protease family protein, putative, expressed
LOC_Os06g30920	expressed protein
LOC_Os06g30930	

Apêndice D. Função Putativa dos Genes no Cromossomo 9 para RILs SSD para o segundo ano de avaliação (2018).

Gene	Função putativa do gene
LOC_Os09g39160	thiol protease SEN102 precursor, putative, expressed
LOC_Os09g39170	cysteine proteinase COT44 precursor, putative, expressed
LOC_Os09g39180	RNA recognition motif containing protein, putative, expressed
LOC_Os09g39190	OsFBO22 - F-box and other domain containing protein, expressed
LOC_Os09g39200	expressed protein
LOC_Os09g39220	delta14-sterol reductase, putative, expressed
LOC_Os09g39230	prephenate dehydratase domain containing protein, expressed
LOC_Os09g39240	expressed protein
LOC_Os09g39250	expressed protein
LOC_Os09g39260	prephenate dehydratase domain containing protein, expressed
LOC_Os09g39270	ZOS9-20 - C2H2 zinc finger protein, expressed
LOC_Os09g39280	expressed protein
LOC_Os09g39310	expressed protein
LOC_Os09g39320	CTR1-like protein kinase, putative, expressed
LOC_Os09g39334	expressed protein
LOC_Os09g39340	expressed protein
LOC_Os09g39350	expressed protein
LOC_Os09g39360	expressed protein
LOC_Os09g39370	expressed protein
LOC_Os09g39380	monodehydroascorbate reductase, putative, expressed
LOC_Os09g39390	oxidoreductase, aldo/keto reductase family protein, putative, expressed
LOC_Os09g39400	histidine-containing phosphotransfer protein, putative, expressed
LOC_Os09g39410	male sterility protein, putative, expressed
LOC_Os09g39420	WD domain, G-beta repeat domain containing protein, expressed
LOC_Os09g39430	GDSL-like lipase/acylhydrolase, putative, expressed
LOC_Os09g39440	inosine-uridine preferring nucleoside hydrolase family protein, putative, expressed
LOC_Os09g39450	expressed protein
LOC_Os09g39462	nucleic acid binding protein, putative, expressed

LOC_Os09g39474	expressed protein
LOC_Os09g39490	histone-like transcription factor and archaeal histone, putative, expressed
LOC_Os09g39500	ubiquitin fusion protein, putative, expressed
LOC_Os09g39510	Cupin domain containing protein, expressed
LOC_Os09g39520	Cupin domain containing protein, expressed
LOC_Os09g39530	Cupin domain containing protein, expressed
LOC_Os09g39540	40S ribosomal protein S25, putative, expressed
LOC_Os09g39550	expressed protein
LOC_Os09g39560	genetic modifier, putative, expressed
LOC_Os09g39570	beta-amylase, putative, expressed
LOC_Os09g39580	calmodulin-binding heat-shock protein, putative, expressed
LOC_Os09g39590	metallo-beta-lactamase, putative, expressed
LOC_Os09g39594	expressed protein
LOC_Os09g39600	expressed protein
LOC_Os09g39610	expressed protein
LOC_Os09g39620	protein kinase family protein, putative, expressed
LOC_Os09g39630	conserved hypothetical protein
LOC_Os09g39640	protein kinase, putative, expressed
LOC_Os09g39650	protein kinase family protein, putative, expressed
LOC_Os09g39660	ZOS9-21 - C2H2 zinc finger protein, expressed
LOC_Os09g39670	oxidoreductase, short chain dehydrogenase/reductase family domain containing family, expressed
LOC_Os09g39680	transporter, major facilitator family, putative, expressed
LOC_Os09g39690	protein binding protein, putative, expressed
LOC_Os09g39700	expressed protein
LOC_Os09g39720	1-aminocyclopropane-1-carboxylate oxidase, putative, expressed
LOC_Os09g39730	Core histone H2A/H2B/H3/H4 domain containing protein, putative, expressed
LOC_Os09g39740	ACT domain containing protein, expressed
LOC_Os09g39750	domain of unknown function DUF966 domain containing protein, expressed
LOC_Os09g39754	expressed protein
LOC_Os09g39760	pectinesterase, putative, expressed
LOC_Os09g39770	C2 domain containing protein, putative, expressed
LOC_Os09g39780	peptidyl-prolyl cis-trans isomerase, putative, expressed
LOC_Os09g39790	lipase, putative, expressed
LOC_Os09g39800	hypothetical protein
LOC_Os09g39810	AP2 domain containing protein, expressed
LOC_Os09g39850	AP2 domain containing protein, expressed
LOC_Os09g39860	protein kinase family protein, putative, expressed
LOC_Os09g39870	2-phosphoglycerate kinase-related, putative, expressed
LOC_Os09g39880	hypothetical protein
LOC_Os09g39910	ABC transporter, ATP-binding protein, putative, expressed
LOC_Os09g39930	tyrosine protein kinase domain containing protein, putative, expressed
LOC_Os09g39940	plastocyanin-like domain containing protein, putative, expressed
LOC_Os09g39950	POEI23 - Pollen Ole e I allergen and extensin family protein precursor, expressed
LOC_Os09g39960	dynamamin family protein, putative, expressed
LOC_Os09g39970	expressed protein

APÊNDICE II - Funções putativas dos genes para os cromossomos 6, 7 e 9 das RILs Bulk.

Apêndice E. Função Putativa dos Genes no Cromossomo 6 para RILs Bulk para o segundo ano de avaliação (2018).

Gene	Função Putativa do Gene
LOC_Os06g04370	expressed protein
LOC_Os06g04380	aminomethyltransferase, putative, expressed
LOC_Os06g04390	expressed protein
LOC_Os06g04390	expressed protein
LOC_Os06g04399	hypersensitive-induced reaction protein 4, putative, expressed
LOC_Os06g04399	hypersensitive-induced reaction protein 4, putative, expressed
LOC_Os06g04399	hypersensitive-induced reaction protein 4, putative, expressed
LOC_Os06g04399	hypersensitive-induced reaction protein 4, putative, expressed
LOC_Os06g04399	hypersensitive-induced reaction protein 4, putative, expressed
LOC_Os06g04399	hypersensitive-induced reaction protein 4, putative, expressed
LOC_Os06g04420	lipoxygenase 4, putative
LOC_Os06g04450	Sec1 family transport protein, putative, expressed
LOC_Os06g04450	Sec1 family transport protein, putative, expressed
LOC_Os06g04450	Sec1 family transport protein, putative, expressed
LOC_Os06g04450	Sec1 family transport protein, putative, expressed
LOC_Os06g04450	Sec1 family transport protein, putative, expressed
LOC_Os06g04450	Sec1 family transport protein, putative, expressed
LOC_Os06g04450	Sec1 family transport protein, putative, expressed
LOC_Os06g04450	Sec1 family transport protein, putative, expressed
LOC_Os06g04440	-
LOC_Os06g04430	-
LOC_Os06g04410	-
LOC_Os06g04350	-
LOC_Os06g04360	-

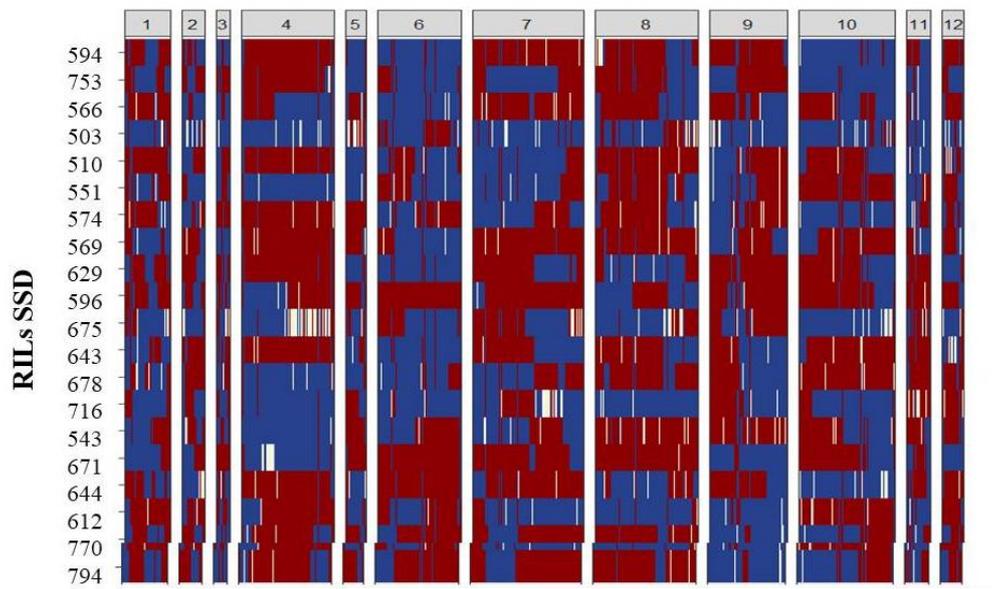
Apêndice F. Função Putativa dos Genes no Cromossomo 7 para RILs Bulk para a análise conjunta (duas anos)

Gene	Putative Function
LOC_Os07g48410	RNA-binding zinc finger protein, putative, expressed
LOC_Os07g48410	RNA-binding zinc finger protein, putative, expressed
LOC_Os07g48410	RNA-binding zinc finger protein, putative, expressed
LOC_Os07g48410	RNA-binding zinc finger protein, putative, expressed
LOC_Os07g48410	RNA-binding zinc finger protein, putative, expressed
LOC_Os07g48410	RNA-binding zinc finger protein, putative, expressed
LOC_Os07g48430	nucleoside-triphosphatase, putative, expressed
LOC_Os07g48430	nucleoside-triphosphatase, putative, expressed
LOC_Os07g48430	nucleoside-triphosphatase, putative, expressed
LOC_Os07g48430	nucleoside-triphosphatase, putative, expressed
LOC_Os07g48430	nucleoside-triphosphatase, putative, expressed
LOC_Os07g48430	nucleoside-triphosphatase, putative, expressed
LOC_Os07g48440	expressed protein
LOC_Os07g48440	expressed protein
LOC_Os07g48440	expressed protein
LOC_Os07g48450	no apical meristem protein, putative, expressed
LOC_Os07g48450	no apical meristem protein, putative, expressed
LOC_Os07g48450	no apical meristem protein, putative, expressed
LOC_Os07g48450	no apical meristem protein, putative, expressed
LOC_Os07g48450	no apical meristem protein, putative, expressed
LOC_Os07g48460	stress responsive protein, putative, expressed
LOC_Os07g48460	stress responsive protein, putative, expressed
LOC_Os07g48460	stress responsive protein, putative, expressed
LOC_Os07g48480	-
LOC_Os07g48420	-

Apêndice G. Função Putativa dos Genes no Cromossomo 9 para RILs Bulk para a análise segundo ano (2018).

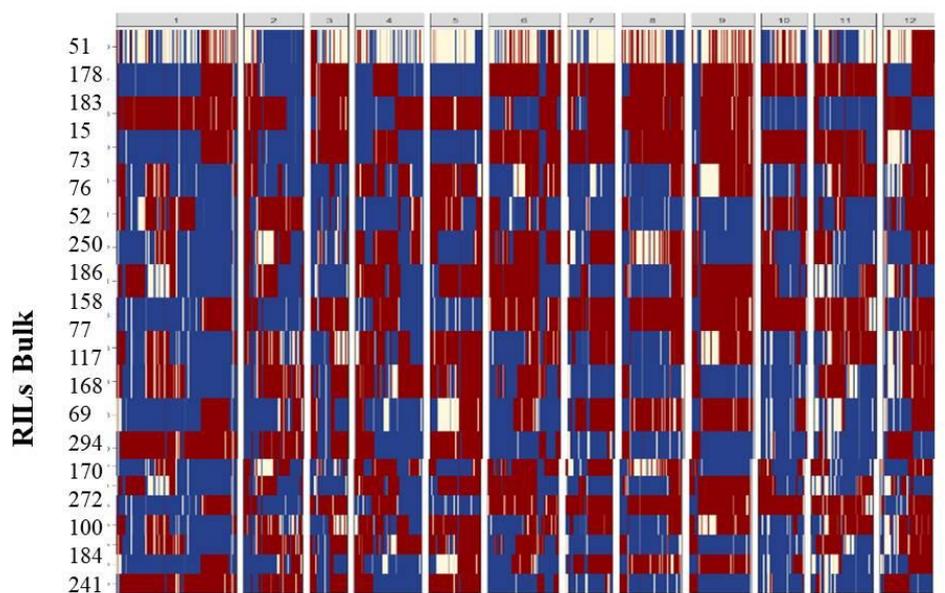
Locus	Putative Function
LOC_Os09g21100	mov34/MPN/PAD-1 family protein, expressed
LOC_Os09g21110	leucyl-tRNA synthetase, cytoplasmic, putative, expressed
LOC_Os09g21110	leucyl-tRNA synthetase, cytoplasmic, putative, expressed
LOC_Os09g21110	leucyl-tRNA synthetase, cytoplasmic, putative, expressed
LOC_Os09g21110	leucyl-tRNA synthetase, cytoplasmic, putative, expressed
LOC_Os09g21110	leucyl-tRNA synthetase, cytoplasmic, putative, expressed
LOC_Os09g21110	leucyl-tRNA synthetase, cytoplasmic, putative, expressed
LOC_Os09g21110	leucyl-tRNA synthetase, cytoplasmic, putative, expressed
LOC_Os09g21110	leucyl-tRNA synthetase, cytoplasmic, putative, expressed

APÊNDICE III - Genótipos Gráficos dos 2.115 marcadores das RILs SSD e dos 2.354 marcadores Bulk, considerandoas dez RILs mais produtivas e as dez RILs menos produtivas pela análise conjunta (dois anos).



Marcadores

- Epagri 108
- Irat 122
- Faltantes



Marcadores

- Epagri 108
- Irat 122
- Faltantes

APÊNDICE IV – Blups das 158 RILs SSD e Bulk, e as Testemunhas.

Tabela 12. Blups das 158 RILs Bulk para os caracteres produtividade de grãos (PG) e altura de plantas (ALT) para 1º ano, 2º ano e análise conjuntas (dois anos).

RILS Bulk	PG 1º ano	PG 2º ano	PG Conjunta	ALT 1º ano	ALT 2º ano	ALT Conjunta
2	5400,11	3374,34	4189,42	122,34	117,38	128,48
3	4116,91	3476,22	4406,89	121,46	116,08	120,92
6	4578,73	2686,50	3866,69	128,67	123,66	118,75
13	5273,59	2938,87	3938,15	127,17	123,62	137,03
14	5708,63	3295,18	4299,79	127,98	125,28	132,90
15	4822,46	3128,18	4522,51	127,36	124,64	134,85
16	4607,19	3806,66	4115,37	128,41	126,62	133,55
17	5152,97	3451,67	4119,96	126,05	123,54	136,15
18	4697,51	3743,88	4305,60	123,94	119,90	129,98
20	5253,63	2995,89	4160,28	124,10	121,28	125,16
21	4010,75	3884,68	4278,17	117,53	110,43	125,29
22	5441,23	2914,42	3886,78	126,50	124,48	108,42
24	5259,75	3082,68	4351,24	126,68	123,66	131,12
26	5003,61	3242,24	4289,99	125,40	119,60	131,81
29	4704,51	3772,40	4216,27	122,89	118,29	128,72
30	4924,19	3523,18	4168,02	126,18	123,58	122,17
31	5223,09	2780,41	4229,24	128,04	125,46	130,75
34	5128,07	3512,39	4237,24	122,51	116,76	135,54
35	5012,29	3317,56	4316,43	127,55	124,32	121,44
36	5337,82	2829,97	4232,07	125,30	121,03	135,27
37	4470,27	3719,89	4297,16	125,15	121,95	127,88
38	5267,05	4025,41	4060,85	121,78	116,71	127,77
40	5406,05	3015,93	4442,39	125,21	121,77	119,40
41	5127,55	3391,37	4353,75	126,65	123,74	128,50
43	5220,29	2505,40	4290,07	130,15	128,52	131,77
46	5080,17	2937,58	4205,82	127,15	123,19	140,43
48	5707,32	3873,75	4198,04	126,84	123,87	132,51
51	5428,29	3741,13	4610,25	123,54	119,17	131,87
52	5442,36	3246,35	4478,59	126,51	122,69	123,43
54	5348,7	3125,63	4408,31	125,05	121,69	131,43
55	5391,55	3208,75	4345,10	127,78	125,86	127,84
57	5185,1	3393,37	4386,23	126,03	120,85	134,41
62	4432,53	3193,44	4313,22	126,05	122,31	129,67
63	5353,86	3302,48	3959,88	126,94	124,41	129,66
64	5282,77	3508,64	4368,11	130,92	127,83	132,25
65	5342,24	3177,62	4364,88	125,51	121,82	142,10
68	4101,8	3196,06	4356,55	125,13	122,34	129,02
69	5618,18	3479,82	3801,88	127,03	122,15	127,71
73	5367,1	3012,84	4511,73	125,88	121,90	132,48
74	5515,55	3190,58	4328,84	129,03	124,92	129,44

75	5476,9	3685,25	4434,05	127,20	123,19	137,19
76	4059,28	3279,74	4486,00	126,94	123,77	132,98
77	5364,88	2877,17	3816,09	125,76	122,66	132,33
79	4999,00	3734,80	4309,15	122,77	116,60	128,94
81	4763,83	3090,06	4287,31	122,86	117,69	122,23
82	4750,61	3215,86	4078,37	127,59	124,19	121,83
83	4672,22	2882,91	4104,20	126,49	122,48	133,73
89	5274,33	2809,48	4011,86	127,63	124,82	131,14
94	4796,09	2919,07	4261,33	130,76	128,75	133,88
95	4805,08	3180,26	4082,91	119,36	112,97	142,03
97	4048,03	2794,73	4131,87	125,00	119,07	113,38
100	5466,69	3358,24	3737,53	127,88	126,48	127,38
102	4840,62	3180,49	4443,17	126,61	124,86	134,99
103	4738,30	2696,89	4141,25	127,47	126,16	131,26
104	5375,71	2837,35	4015,28	125,22	121,25	134,54
105	4531,8	3925,39	4324,25	120,68	115,57	127,63
106	4777,48	3463,11	4101,93	127,05	123,71	116,63
107	4428,16	4109,27	4137,27	118,86	116,96	132,24
109	5144,44	3131,29	4105,07	124,65	121,23	111,98
111	4688,94	2894,65	4255,58	126,06	122,69	125,92
113	5058,83	4128,45	4018,04	122,78	118,95	129,91
115	4075,62	3277,13	4364,47	125,58	122,25	121,80
117	4567,54	2902,94	3813,35	127,13	121,96	128,62
118	4549,09	2941,18	3970,07	116,64	109,17	133,18
123	4559,77	4024,31	3968,21	123,21	119,50	106,32
129	4857,28	3367,99	4134,69	127,97	126,19	123,05
130	5023,59	3281,85	4164,78	121,13	116,07	134,68
131	4610,88	2648,65	4227,21	126,75	124,28	119,07
132	5589,09	3103,18	3945,22	122,68	118,12	131,61
135	5512,78	3234,57	4440,99	126,11	122,90	121,77
136	4614,77	3412,05	4433,92	165,23	122,56	129,80
141	4750,49	3652,39	4080,49	128,77	121,56	129,33
144	4782,85	2851,48	4166,67	125,39	122,95	136,29
145	5558,12	2799,17	4063,68	130,45	128,62	128,49
146	4712,15	3482,56	4392,08	123,89	120,05	141,44
147	5079,57	3484,98	4120,87	120,43	114,40	125,45
148	4582,44	2812,49	4273,08	120,94	118,27	116,02
149	4779,11	3218,03	3973,99	125,05	122,16	117,67
150	4725,89	3145,31	4120,08	123,76	101,14	127,51
153	5128,44	2827,39	4071,82	123,95	122,84	124,97
154	5189,65	3724,56	4207,75	123,18	118,32	124,69
156	5481,63	3436,81	4375,31	125,65	124,78	122,58
158	5263,21	3336,92	4445,97	125,66	123,38	129,28
161	5226,58	3549,28	4345,51	124,86	121,13	130,03
162	4934,12	3858,83	4358,68	126,54	124,03	127,44

163	5302,88	3366,95	4274,74	127,45	124,33	131,47
164	4104,73	3610,04	4363,70	118,09	111,35	133,59
167	4146,28	3102,60	3880,79	126,26	124,02	111,72
168	4123,19	3048,12	3812,86	126,06	123,28	130,84
170	4601,86	3162,42	3789,45	127,22	123,74	129,65
172	4652,91	3604,34	4037,30	124,62	124,82	134,44
175	5249,78	3170,91	4125,50	128,54	126,39	126,64
177	5815,43	3378,63	4313,41	129,39	127,08	135,90
178	5609,80	3838,70	4591,08	124,10	119,92	138,41
183	4176,33	2368,37	4565,07	124,01	119,53	125,48
184	5364,43	3885,79	3708,43	122,93	116,84	124,76
186	5281,10	3458,95	4463,54	125,84	121,59	121,54
187	5232,75	3553,65	4361,20	125,90	121,67	129,53
188	5176,02	2770,99	4352,66	126,53	122,77	129,62
190	4791,42	3351,54	4217,33	125,18	121,16	131,26
191	4893,5	3742,95	4130,81	120,09	114,08	127,65
192	5023,46	3252,08	4246,37	125,17	120,65	115,39
193	4795,09	2781,42	4218,82	125,81	122,17	128,00
194	4651,57	3249,84	4048,33	126,92	123,80	129,03
197	4579,56	2401,81	4058,81	127,32	124,02	132,44
198	4107,04	3588,75	3893,82	125,81	122,61	132,71
199	4761,88	2812,08	3873,15	126,87	122,90	129,44
200	5175,45	2418,05	4047,00	120,04	111,54	132,07
204	5350,49	3448,93	4161,69	125,52	122,06	114,83
207	4602,24	3234,13	4389,95	130,01	127,41	128,64
209	4999,75	3527,62	4039,74	124,48	119,65	140,24
211	5318,09	3317,94	4255,45	126,77	122,45	126,43
212	4738,98	3349,64	4366,02	124,02	118,86	131,94
213	5349,47	2829,16	4113,92	126,64	123,76	125,09
216	4666,11	2881,32	4301,19	127,13	123,72	132,57
217	5326,44	3261,25	4011,13	127,23	123,75	132,73
218	5063,68	2812,49	4356,72	123,95	119,48	133,53
219	4736,29	2797,56	4177,92	131,55	130,48	125,15
221	4906,1	3679,24	4034,78	123,81	118,47	143,92
222	5396,5	2706,53	4250,75	125,69	122,02	124,01
225	4914,5	3438,97	4297,00	120,87	114,58	129,59
226	4650,46	3774,87	4206,61	122,81	118,03	116,52
229	5310,97	2615,32	4133,63	126,02	123,07	122,02
233	5473,43	2474,25	4263,70	127,34	124,76	130,37
235	5028,69	3186,83	4308,14	121,84	117,61	133,51
238	4125,36	2419,91	4218,40	119,68	115,60	119,24
241	4661,70	3711,37	3706,96	116,19	109,53	114,65
242	4829,53	3233,73	4139,82	129,68	127,91	104,74
243	4449,61	3072,19	4130,79	125,89	122,11	139,39
244	5358,30	3240,88	3949,62	122,10	116,32	129,19

246	5281,23	3060,51	4364,75	125,73	122,17	120,33
247	5200,6	4304,34	4302,27	122,16	121,31	128,85
250	4710,94	3829,32	4471,52	129,46	127,42	119,90
251	4637,51	3765,03	4163,69	124,06	120,89	138,30
252	5272,58	2933,05	4131,63	128,06	125,70	125,48
254	4721,69	2858,75	4286,38	122,48	116,85	135,20
256	4514,86	2277,18	4044,77	125,94	122,07	121,30
257	5267,04	3291,47	3859,31	125,55	122,74	129,79
260	5404,01	2997,09	4340,82	125,48	122,42	129,33
263	5103,98	2985,69	4353,81	125,79	122,39	129,18
264	5250,32	2648,50	4214,64	123,21	119,58	129,60
266	5249,17	3083,26	4230,49	119,82	112,91	123,09
268	5239,14	3077,95	4294,61	125,11	120,84	114,57
269	4774,97	2906,30	4296,28	124,61	121,24	129,27
270	5157,54	2809,03	4076,90	126,91	122,84	127,00
271	4272,64	2430,98	4217,09	122,04	116,55	133,41
272	4922,05	3167,64	3784,34	125,93	122,78	119,78
278	4927,48	2571,14	4165,98	126,96	123,97	129,66
280	5378,58	3423,21	4083,12	126,68	123,67	132,88
281	5060,13	3281,03	4399,01	127,39	124,93	131,94
283	4633,69	3629,26	4243,03	116,06	107,90	133,19
285	5139,20	3143,69	4115,20	119,75	113,56	105,05
290	4634,36	3611,56	4257,64	124,04	121,14	114,62
291	4110,41	3251,32	4096,71	120,02	117,18	125,36
293	4062,74	3200,75	3829,43	122,00	117,02	114,91
294	5104,23	3166,05	3798,41	127,45	124,24	119,85
300	4411,76	2946,31	4236,76	125,39	121,88	133,18
301	5335,69	3488,23	3912,66	126,96	123,88	127,81

Tabela 13. Blups 158 RILs SSD para os caracteres produtividade de grãos (PG) e altura de plantas (ALT) para 1º ano, 2º ano e análise conjuntas (dois anos).

RILs SSD	PG 1º ano	PG 2º ano	PG Conjunta	ALT 1º ano	ALT 2º ano	ALT Conjunta
499	4199,13	2961,57	3614,57	118,67	122,90	118,13
501	4063,29	3273,96	3607,14	126,38	127,60	123,57
502	5283,17	2844,78	3596,29	117,84	122,82	134,60
503	4474,78	2968,14	4274,95	116,11	116,76	122,17
505	4511,84	3078,18	3791,69	115,04	110,58	120,21
508	5157,78	3274,67	3830,69	118,52	117,95	117,40
510	3414,05	3171,66	4255,84	121,36	120,81	123,32
511	4474,80	3501,99	3158,05	129,67	130,33	127,62
513	4687,38	3256,33	3896,42	107,73	100,25	142,10
514	4337,19	2780,14	3969,34	100,26	92,61	107,68
515	3972,81	3272,13	3658,05	112,28	104,76	96,15
519	4798,32	3124,40	3522,16	116,90	114,79	114,40
520	4046,06	3848,34	4018,95	116,53	114,72	121,09
521	3410,97	4790,07	3675,71	102,25	95,25	120,63
522	4110,63	2800,08	3439,40	124,86	132,34	98,56
523	4448,33	2646,63	3523,15	103,12	96,62	132,90
525	3968,74	2005,99	3702,83	102,29	95,47	100,89
527	3407,41	3289,24	3279,38	102,84	96,03	98,82
528	4949,64	2199,11	3169,17	118,82	117,74	99,92
535	4708,13	3797,58	3928,58	125,42	122,78	124,24
536	4397,99	2636,62	4091,04	108,38	102,48	133,14
537	4747,39	2740,32	3680,33	109,55	104,73	108,04
540	3730,79	2314,79	3918,56	102,73	96,91	110,40
541	4253,49	1414,18	3189,75	102,65	95,58	99,98
542	3800,03	1544,36	3358,80	177,43	120,71	99,95
543	4601,76	2711,24	3091,68	123,03	121,57	110,71
545	4058,14	3485,16	3821,74	125,88	126,28	130,07
546	4144,42	2725,10	3610,81	124,21	124,45	134,47
547	3907,22	2187,03	3521,43	121,86	120,04	132,19
550	4942,84	3756,87	3282,36	121,34	127,45	127,73
551	3780,42	2439,99	4224,30	112,04	107,82	127,21
552	4080,89	1862,88	3249,09	121,64	121,28	112,96
554	4662,71	3624,26	3307,24	115,52	113,32	127,99
555	4448,68	1548,19	4019,45	114,14	113,14	119,48
556	4320,12	3016,58	3513,13	102,63	93,66	116,79
558	4101,69	2633,07	3683,41	117,15	115,99	99,52
559	4531,15	3105,35	3487,23	120,11	117,92	122,00
562	4556,14	3358,61	3839,23	114,76	108,02	125,65
563	4024,56	3690,35	3919,83	124,00	122,19	117,81
565	5148,79	3654,18	3631,04	121,40	118,22	131,86
566	4721,54	3779,29	4345,12	111,67	108,20	127,49
568	4847,39	3674,32	4087,32	119,62	117,87	113,56

569	3939,39	3450,14	4160,35	119,80	114,35	125,07
570	4100,08	3405,21	3530,31	104,83	96,11	125,57
571	4634,85	3100,77	3630,83	123,71	123,66	102,72
572	3905,86	3456,65	3912,56	114,89	113,06	131,24
573	4980,30	3275,12	3506,14	127,86	129,28	117,90
574	4593,82	2468,15	4168,42	123,78	125,22	137,58
576	3845,73	3370,86	3745,17	106,16	97,36	130,93
577	4105,33	2778,04	3449,42	105,41	99,68	104,78
578	3415,60	3742,96	3505,42	110,39	102,51	104,39
580	4734,50	2484,37	3249,55	121,19	120,79	111,38
584	4574,00	2699,85	3849,06	104,52	97,84	127,18
585	3861,23	2624,41	3806,38	105,73	100,55	101,79
587	4754,25	3487,20	3310,30	125,51	127,63	103,81
588	4509,93	2019,86	4064,40	121,97	121,60	134,42
590	3937,66	2907,51	3591,43	102,40	96,20	128,34
592	8096,27	2452,66	3424,78	120,99	122,14	99,32
594	4902,69	3154,31	5985,11	120,11	120,99	127,11
596	4451,42	2786,20	4099,44	121,41	121,18	125,63
597	4832,11	2155,85	3745,55	119,03	120,35	127,84
598	4297,91	1593,37	3834,99	101,99	96,14	124,57
602	4584,68	3739,44	3410,60	101,91	96,03	98,87
604	4213,17	2914,27	3988,49	125,69	126,60	100,17
605	4047,53	2167,23	3605,32	100,33	93,87	133,88
607	4350,43	3301,85	3358,20	115,73	114,83	96,48
608	3443,82	2255,79	3762,80	116,24	114,78	119,35
612	4367,25	2659,17	2982,07	126,40	127,77	120,03
613	4106,06	3089,45	3644,22	121,25	120,87	135,08
616	4144,37	2275,49	3572,20	116,48	115,03	127,62
618	4924,43	3146,91	3417,12	120,19	119,45	120,20
619	4836,27	1258,20	4092,80	116,89	115,27	126,15
621	3321,76	3834,03	3696,03	125,33	128,03	120,53
624	3930,82	2689,93	3217,18	99,32	91,02	134,06
628	4653,27	4123,70	3377,21	119,04	116,65	94,52
629	4088,20	2021,33	4126,16	103,85	98,80	124,07
632	4324,87	3156,41	3363,18	123,51	124,22	101,52
633	3943,25	3005,33	3716,72	123,14	120,60	130,97
634	4147,14	2235,45	3426,28	125,26	125,02	130,29
635	4523,22	2481,94	3435,50	120,73	116,49	133,36
636	4882,12	2026,94	3715,94	102,39	94,54	126,47
637	4578,85	3021,19	3863,69	108,63	101,22	100,92
638	3835,56	2268,29	3862,45	121,59	124,76	110,06
639	4073,37	2466,90	3239,37	122,75	118,57	128,18
642	3774,02	1772,08	3440,99	123,15	123,12	130,21
643	3855,01	1089,81	3133,75	103,39	102,41	129,89
644	3807,71	2250,96	3036,79	107,81	98,34	100,92

645	4517,93	2514,15	3207,21	123,68	123,06	107,15
647	4214,11	1800,27	3703,94	112,67	107,66	130,59
649	3981,16	2833,57	3373,57	119,28	124,64	114,83
651	4849,37	3252,91	3454,09	123,25	121,04	124,47
652	4053,00	3214,78	4056,57	99,30	91,01	130,11
653	3762,09	2267,88	3552,66	109,55	102,05	94,63
655	4508,76	2478,50	3171,49	128,43	129,77	109,74
656	4612,40	3333,14	3712,78	105,35	98,73	138,57
662	4635,14	2660,45	3929,47	124,02	124,36	103,48
663	4108,59	2753,38	3823,54	121,15	118,91	133,21
668	3735,04	1643,79	3505,13	101,76	93,83	126,77
671	3775,05	1849,67	3051,79	108,57	103,71	98,34
675	4071,02	3236,86	3134,46	116,51	110,70	109,15
677	3527,55	2606,51	3571,22	103,51	95,87	120,50
678	4444,36	4328,78	3100,46	107,67	99,26	100,50
679	3917,16	2382,11	4020,60	103,11	95,24	106,81
684	3694,84	2978,81	3293,96	121,88	115,70	100,14
685	3935,97	2358,44	3280,38	109,12	103,37	128,07
687	4644,63	2185,22	3320,62	103,94	96,01	109,24
688	3973,37	2590,24	3720,07	120,30	121,88	101,23
689	4612,41	1305,94	3390,65	120,49	119,26	126,44
690	3697,26	2178,25	3551,23	105,29	97,57	127,17
692	4540,07	1839,18	3139,99	101,80	93,54	103,68
694	3984,85	3333,24	3612,79	121,93	121,03	98,18
696	3802,74	1869,76	3554,37	121,90	119,12	129,09
697	4388,39	1995,27	3141,57	103,11	96,11	128,84
698	4039,54	3167,21	3537,96	105,90	99,07	102,55
699	4805,90	3180,36	3530,29	125,63	127,49	104,07
700	4320,43	3324,36	4031,52	122,34	122,69	134,71
702	4565,61	1642,69	3738,84	120,52	117,60	128,87
705	4222,22	4383,84	3580,29	120,19	119,34	126,58
706	4003,01	3509,46	3888,06	120,84	122,77	126,04
707	4449,80	2294,64	3585,81	103,26	93,62	126,84
711	4715,41	2197,76	3639,59	123,33	123,86	100,15
713	3967,52	2370,07	3778,19	100,27	91,64	130,53
714	3644,00	2231,38	3337,23	102,98	94,67	95,98
716	3899,65	1831,10	3096,64	119,43	119,78	100,02
717	4920,56	2281,29	3190,93	121,58	121,99	124,55
719	4719,45	3619,37	3916,15	112,81	109,41	127,57
720	4575,13	2679,25	4057,36	128,03	128,51	114,66
725	3551,64	3107,03	3783,00	102,91	98,03	138,34
727	4470,05	2637,97	3236,95	123,21	124,60	99,47
728	3829,69	2186,50	3708,10	112,11	107,93	130,19
730	3821,24	2531,85	3236,59	110,52	105,16	113,06
735	4668,63	1380,26	3287,07	119,04	117,10	110,90

737	3840,71	2290,56	3594,65	124,88	122,06	124,32
738	4178,21	2182,19	3244,61	114,24	109,04	132,91
740	4120,86	3989,70	3465,96	120,84	119,85	117,14
742	3911,25	2402,78	3757,34	102,52	92,80	127,45
743	3926,91	2667,12	3306,23	120,39	119,36	99,39
744	4339,91	3218,36	3369,29	104,00	97,50	126,02
745	4406,86	2695,86	3742,55	117,87	116,08	101,74
747	4851,62	3251,81	3683,85	125,03	126,50	122,63
748	3879,70	2806,03	4082,29	100,87	91,88	134,80
749	4859,44	4738,05	3359,00	123,58	123,32	96,85
753	4613,73	2406,85	4372,13	106,74	99,17	130,54
754	4079,85	2272,09	3765,88	104,92	96,75	107,17
758	3926,40	2838,46	3396,86	124,63	123,74	103,03
760	4899,63	2578,53	3409,34	120,80	120,34	133,89
762	4224,69	2115,67	3957,47	99,86	92,06	127,25
764	4544,64	4338,73	3459,71	121,91	122,67	95,73
765	4380,45	2460,74	4078,86	122,03	121,67	129,13
769	3391,97	2287,91	3623,66	119,89	120,38	128,59
770	4233,25	2759,95	2965,25	126,62	128,67	125,21
771	4265,45	2916,79	3588,79	117,18	114,66	136,07
772	4526,45	3326,77	3647,31	100,01	97,81	121,23
773	4250,04	2187,60	3893,94	103,03	95,01	95,88
785	4700,02	1348,89	3485,59	115,81	114,25	100,01
791	4288,73	1956,22	3603,29	104,81	98,09	119,61
793	3590,80	1156,44	3460,71	91,35	81,83	102,58
794	4054,23	1158,28	2855,16	93,68	90,48	82,92

Tabela 14. Blups das Testemunhas para os caracteres produtividade de grãos (PG) e altura de plantas (ALT) para 1º ano, 2º ano e análise conjuntas (dois anos).

Testemunhas	PG 1º ano	PG 2º ano	PG Conjunta	ALT 1º ano	ALT 2º ano	ALT Conjunta
Epagri 108	3256,35	4322,35	3456,70	104,27	100,59	108,42
Irat 122	0,00	3313,89	3813,80	0,00	106,90	103,89
Araguaia	3096,12	1888,69	3662,98	104,27	114,16	113,44
Maninjau	3226,63	4534,50	2970,35	104,27	117,71	111,99
IRGA 417	2754,31	2932,83	3839,87	104,27	96,07	121,02
BRS Catiana	3643,73	3808,89	2677,21	104,27	98,74	101,59
BRS Tropical	3254,45	3355,25	4304,65	104,27	103,90	102,99
Ita Puitá CL	2874,48	2802,36	3561,79	104,27	96,52	110,10

APÊNDICE V – Fotos das RILs SSD e Bulk (planta e sementes) mais produtivas e os genitores dos experimentos de 2017 e 2018, realizados na Fazenda Palmital-GO.



RILs Bulk 93

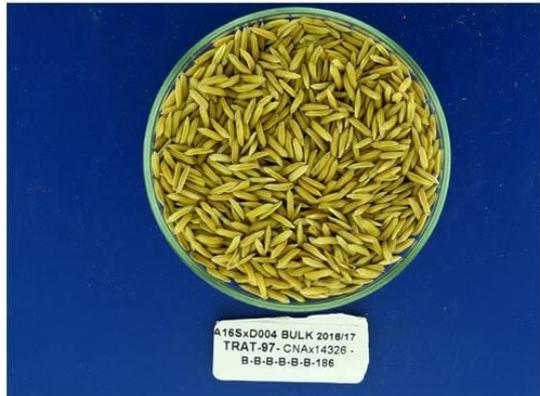


RILs Bulk 132



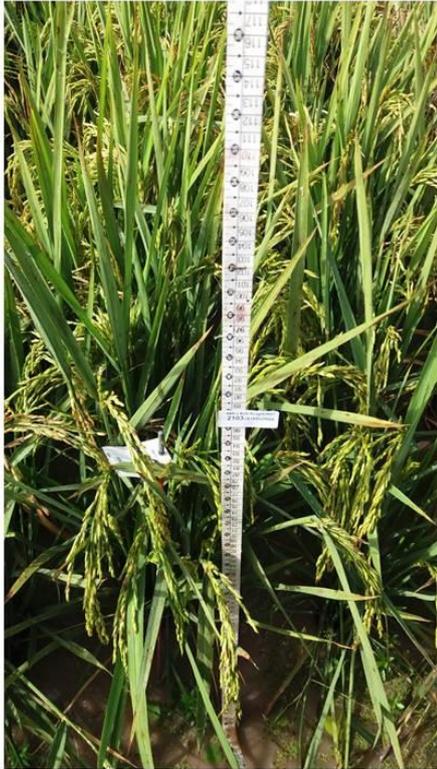


RILs Bulk 51

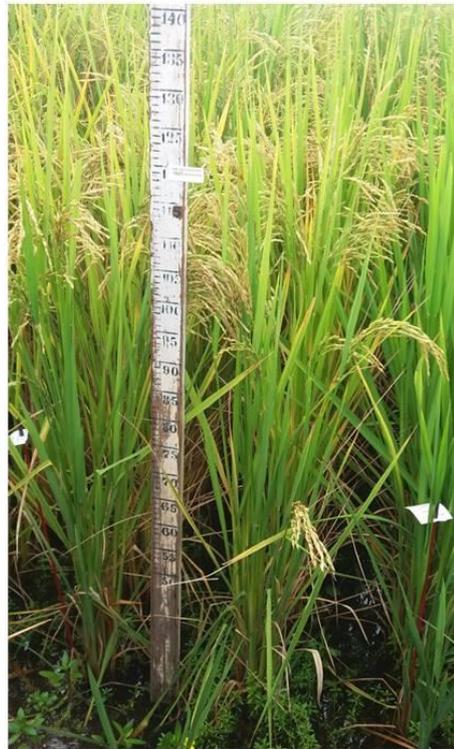


RILs SSD 532





RILs SSD 172



RILs SSD 594

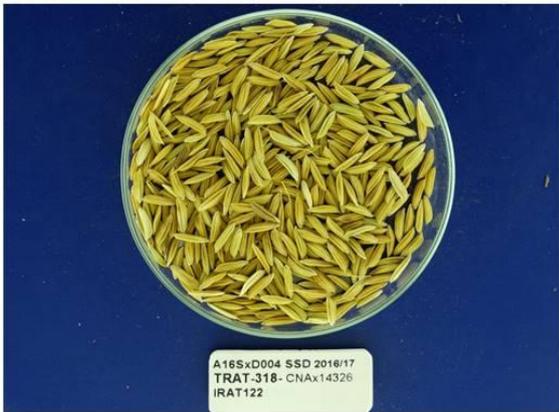




EPAGRI 108



IRAT 122



APÊNDICE VI – Dados Climáticos médios durante o período experimental

Dados Climáticos médios na Fazenda Palmital, Embrapa Arroz e Feijão em Goianira-GO

	Goianira-GO			
	V ¹	R ²	M ³	V ¹
Nebulosidade (décimos)	7	7	7	7
Radiação Solar (h)	197	187	168	180
Precipitação Total (mm)	161	275	163	203
Temp. Máxima (°C)	36	32	32	33
Temp. Mínima (°C)	22	21	21	22
Umidade Relativa (%)	61	78	80	69
Comprimento Dia (h)	12:14			

Fonte: Inmet 2019. ¹Período Vegetativo (out-nov). ²Período Reprodutivo (dez-fev). ³Período de Maturação (mar-abr).



**FUNCTIONAL PLANT
BREEDING JOURNAL**

Scientific Journal. ISSN: XXXX-XXXX
Volume 1, Number 1, Article nº 1, January/March :
D.O.I: <http://dx.doi.org/XX.XXXX/2526-4117/v1n1a>
Accepted: 12/07/2018 Published: 05/03/2019

HETEROSIS AND COMBINING ABILITY FOR GRAIN YIELD AND EARLINESS IN ACCESSIONS OF A RICE CORE COLLECTION

Mariana Rodrigues Feitosa Ramos¹, João Antônio Mendonça², Rosana Pereira Vianello², Odilon Peixoto de Morais Júnior¹, José Manoel Colombari², Tereza Cristina de Oliveira Borba², Adriano Pereira de Castro², Claudio Brondani²

¹ Universidade Federal de Goiás (UFG), 74.690-900, Goiânia, GO, Brazil

² Embrapa Arroz e Feijão, 75.375-000, Santo Antônio de Goiás, GO, Brazil Corresponding author: Claudio Brondani (claudio.brondani@embrapa.br).

Abstract: Rice has extensive collections of germplasm, but a small fraction of genotypes are used in breeding programs worldwide. This aimed to assess the heterosis and the combining ability for higher grain yield and earliness, and their correlations to genetic distance among accessions belonging to the Embrapa's Rice Core Collection. Crosses were made in complete diallel, without the reciprocals, being phenotyped in generations F_2 and F_7 for grain yield (GY) and days to flowering (DTF). It were estimated the varietal effect, mean heterosis, varietal heterosis, specific combining ability and general combining ability (GCA) of the parents in each generation. For the selection of new lines, the combinations involving the genitors with the highest magnitudes and positive effects for GCA were considered more promising for the GY, of which Canela Curta, Maninjau, Epagri 108 and Diamante stood out, and were also the most productive. For the DTF, the genitors with the largest earliness estimates were Araguaia, CT11632, Irat 122, Pratinha Branco and Tox 503. The recommended parameters for estimating GY in F_7 generation from the data of the F_2 generation were the mean of GY, GCA and the mean Rogers-W genetic distance coefficient. Crosses with Rogers-W distance coefficients above 0.9 showed higher GY and earliness in F_7 generation, whereas in F_2 the superior crosses were those that showed Rogers-W between 0.7 and 0.89. The above genotypes and the specific combinations identified may lead to new rice cultivars.

Keywords: *Oryza sativa*, diallel analysis, genetic divergence, SSR markers, molecular breeding

Introduction

Most breeding programs of self-pollinating species, such as rice, generate segregating populations of biparental crosses between elite genitors, usually genetically related, for the development of cultivars. This strategy aims to ensure success in obtaining new superior inbred lines through recombination among parents who already accumulate many favorable alleles, since successive generations of self-fertilization from the F_2 generation do not promote sufficient recombination to increase the transgressive frequency of favorable gene combinations (Dennis et al., 2008). However, this uniformity, in addition to limiting the annual genetic gain in grain yield, makes the elite germplasm of these programs genetically uniform, and more likely to be susceptible to pests and diseases (Huang et al., 2014).

An alternative to increase the exploration of the potential variability to be worked by breeding programs is the better organization of the genetic resources available in genebanks, through core collections formed by a smaller set of accessions representing the largest fraction of genetic variability of the total collection (Brown, 1989). For rice, core collections have already been developed for the germplasm stored in countries such as Brazil (Abadie et al., 2005), the U.S.A. (Yan et al., 2007) and China (Zhang et al., 2011). However, there would be a multitude of possible crossings to be made with such accessions with intent to broaden the genetic base useful to the breeding program. The most widely used method for selecting parents is diallel analysis, which allows for estimating combining ability, a term introduced by Sprague and Tatum (1942).

Diallel analysis permits to infer the potential of each genitor to provide superior hybrid combinations, the gene action that controls the traits involved, and the existence of heterosis, providing the opportunity to identify and develop superior lines (Huang et al., 2015). The general combining ability generates information regarding a set of genes of predominantly additive effects, and the specific combining ability is related to a set of genes of dominance and epistatic effects (Asfaliza et al., 2012; Fasahat et al., 2016).

The use of molecular markers for genetic characterization of germplasm offers real benefits to increase efficiency at various stages of breeding programs and can be adopted with high accuracy and at reduced costs for, as examples, cross validation, certification of cultivars and determination of seed lot genetic uniformity. The objective of this study was to assess the heterosis and the combining ability for higher grain yield and earliness, and their correlations to genetic distance among accessions belonging to the Embrapa Rice Core Collection.

Material and Methods

Among the 550 accessions of the Embrapa Rice Core Collection, the 12 genotypes selected were the most productive in a previous experiment (Bueno et al., 2012) carried out at Palmital Farm (Goianira, GO), which belongs to Embrapa Arroz e Feijão. Those included five landraces, two Brazilian cultivars and five genotypes from abroad (Table 1). The 12 genotypes constituted the group of parents of the complete diallelic crosses without reciprocals, which resulted in 66 combinations of crosses. During the experiment in generation F_2 , and during the process of advance of generations, five of the 66 combinations (Bico Roxo x Lageado, Canarinho x Lageado, Canarinho x Bico Roxo, Diamante x Bico Roxo and Diamante x Irat 122) were lost. The F_1 hybrids were self-

fertilized to obtain generation F₂, which was evaluated through field experiment, at Palmital Farm, using a randomized completeblock design, with three replications in the 2006/2007 crop season. The plot consisted of 4 rows, each 4 meters in length, and a plot useful area of 2.40 m² (2 central rows). Afterwards, the diallelic cross group was advanced in *bulk*, from generation F₂ to F₇. The F₇ generation was evaluated in the 2013/2014 crop season, using the randomized complete block design with two replications, with the same size of the plot and in the same location of the F₂ generation.

The traits evaluated in the experiments of generations F₂ and F₇ were number of days to flowering (DTF), corresponding to the number of days between sowing and mean flowering of the plot (50% of flowered panicles), and grain yield (GY), corresponding to the dried grain weight, in kg ha⁻¹. The phenotypic means of the parents and the crosses in the two generations, adjusted for the year and the block effects, were estimated by the joint analysis of the two experiments using the following fixed linear model:

$$y_{ijk} = \mu + \frac{(v_i + v_j)}{2} + \theta(\bar{h}_k + h_{ki} + h_{kj} + s_{ijk}) + \bar{e}_{ijk} \quad (\text{Eq. 1})$$

where Y_{ijk} is the observation of the treatment i , in block k , in year j ; μ is the constant, inherent to all observations; a_j is the effect of the year j ; $b_{k(j)}$ is the effect of block k within the year j ; t_i is the effect of the treatment i , being i equal to parents or crosses; and e_{ijk} is the experimental error associated with the observation ijk , assuming $NID \cap (0, \sigma^2)$, where NID stands for normally and independently distributed. The interaction treatments *versus* years were considered as a component of the experimental error, a necessary condition for the estimation of the adjusted means of the treatments (Searle, 1971). In addition, treatments and years were treated as fixed effects for the following reasons: (i) both locations, represented by the years, consisted of environments with their own characteristics that were targets of interest for the rice breeding program; (ii) treatments included lines or parental cultivars selected for desirable traits in the breeding program; thus, they were not randomly selected. For each generation (year) was obtained the vector estimation \hat{Y} , corresponding to the means of treatments adjusted for effects described in the linear model, and also the covariance matrix [$\hat{V}(\hat{Y})$] of the adjusted means for the joint analysis of the genotype group, in accordance with Morais Júnior et al. (2017a).

The combining ability of the accessions was estimated using the Gardner and Eberhart (1966) method, based on the generational model proposed by Pereira et al. (2008), with necessary adaptation to the generalized method of diallelic analysis of unbalanced data, as proposed by Silva et al. (2000). The justification for using this method is based on the existence of unbalance among the experimental data, due to the different number of repetitions between some treatments. The Gardner and Eberhart method (1966) was used in the diallelic analysis involving the crosses and their parents in the generations F₂ and F₇, thus allowing for the decomposition of the total heterosis effects in mean heterosis, heterosis of parents and specific heterosis (or specific combining ability – SCA). In addition, Cruz and Regazzi (1994) point out that this method is recommended when there are significant differences between diallelic mean values due to heterogeneity among parents, and also in the presence of heterosis. Such conditions are present in the set of crosses due to the different origins of the parents. To estimate the effects of genitors and heterosis components for each generation, the following linear model was used:

$$y_{ijk} = \mu + \frac{(v_i + v_j)}{2} + \theta(\bar{h}_k + h_{ki} + h_{kj} + s_{ijk}) + \bar{e}_{ijk}$$

where y_{ijk} is the estimated mean of crosses between i and j genitors, in the F_k generation; μ is the general mean; v_i and v_j are the effects of the i and j genitors, respectively; q is the indicator variable, where $q = 0$ if $i=j$, otherwise $q = 1$; h_k is the mean heterosis in the F_k generation; h_{ki} and h_{kj} represent the heterosis attributed to the i or j genitors, respectively, in generation F_k ; S_{ijk} is the specific heterosis (or SCA) of the cross between the i and j genitors in the F_k generation; and e_{ijk} is the mean experimental error associated with the mean y_{ijk} (non-homogeneous and non-orthogonal). In order to perform a decomposition of the sum of squares (SS) of treatments, which are non-orthogonal necessarily in this case, adjustments were made in SS (v), SS (h_k), SS (h_{ki}), SS (h_{kj}), SS (S_{ijk}). For each k , the adapted equation of the general sum of squares hypothesis in linear models (Searle, 1971) was used for each variation component:

$$SS - (C\beta)'(CQC')^{-1}C\beta$$

where C it is a matrix of estimate contrasts; β is the solution vector of the diallelic equation; Q is the covariance matrix of β [$V(\beta)$], divided by the mean square of residue from the joint variance analysis (s^2), i.e., ... $Q = V(\beta) s^2$. Information on the structure of $V(\beta)$ are detailed in Silva et al. (2000) and Morais Júnior et al. (2017b). The constraint matrix was structured, admitting the following restrictions in the solution space: $\sum_i v_i = 0$; $\sum_j h_{kj} = 0$, for each k ; and $\sum_j s_{ijk}$, for each i and k . From the effects of genitor (v_i or v_j) and of genitor heterosis h_{kj} in the mean of the generations, estimates of the general combining ability (GCA) of each genitor (g_i) were obtained by the expression:

$$g_i = \frac{1}{2} v_i + h_i; \left(\sum_i g_i = 0 \right).$$

The genitors and diallelic crosses for each generation (F2 and F7) were compared with the purpose of obtaining the varietal effect, in addition to the estimates of mean heterosis for all crosses, heterosis of crosses and specific heterosis in F2. All statistical analyses to obtain adjusted means of treatments and effects were performed using platform R (R Core Team, 2013). Pearson coefficient correlation analyses between the Rogers genetic distance modified by Wright (RW) (Wright, 1978), specific heterosis (s_{ij}) and the means of the two characters evaluated were obtained by the Genes program (Cruz, 2013). The RW genetic distance estimates were determined from the genotyping data of the parents through 24 SSR markers described by Borba et al. (2009).

Results and discussion

The coefficients of variation (CV%) obtained by the joint analysis of the experiments, in generations F2 and F7, were 16.98% and 3.65% for GY and DTF, respectively. These values are considered satisfactory for the experimental design and plot size adopted for the evaluated traits, in accordance with Silva et al. (2000); thus indicating a high experimental precision. The control group displayed the highest GY mean, followed by the group of crosses in generation F7, that were superior to the respective genitors and the crosses in generation F2, also surpassing the general mean for the trait (Tables 1 and 2). In the generation advances by the bulk method, the average of crosses in F7 should have been equal to the average of the genitors, since there is no dominance effects. In addition, the GY F2 average should have been higher than that of the parents due to the presence of half of heterosis. According to Charlesworth et al. (2017) these results can be explained by

the fact that over the course of the inbreeding generations, the most adapted individuals can generate more descendants due to greater grain yield. For the GY trait, the mean heterosis referring to generations F2 and F7 was not significant ($p > 0.05$), which indicates that crosses in both F2 and F7 generation do not differ from the mean of the genitors. However, specific heterosis (or SCA) was highly significant, indicating the occurrence of certain crosses that stood out in relation to the respective parents (positive heterosis) (Table 3). For DTF, there was significance for mean heterosis ($p < 0.01$) in both generations F2 and F7, indicating that the set of crosses in both generations differs from the mean of the parents. Although the magnitude of the mean specific heterosis (or SCA), has been reduced from F2 to F7, it remained significant, indicating the presence of epistasis between loci involving additive effects (Bhullar et al., 1979).

Grain yield (GY)

The parameters varietal effect (v), varietal heterosis (h_v) and the combining ability of the parents (g_i) in generations F2 and F7 were estimated (Table 4). The varietal heterosis of a genitor is the difference between the mean of the heterosises of all their crosses and the mean heterosis of the diallel in the same generation (Cruz and Regazzi, 1994). Therefore, with the significance of varietal heterosis, it can be inferred that the parents differ statistically from each other in terms of their mean allele frequencies. Canela Curta, CT11632, Irat 122, Lageado, Maninjou and Pratinha Branco genitors showed significant estimates of varietal heterosis in generation F2 (h_vF2), indicating that these genitors have a greater divergence in relation to the group of genitors of the diallel for the GY trait, while CT11632, Irat 122 and Pratinha Branco showed negative effects for this trait. Varietal heterosis in generation F7 (h_vF7) was positive and significant for Lageado and Maninjou (repeating the result of generation F2), along with the Canarinho, and was negative and significant for CT11632 and Irat 122 (repeating the result of generation F2), together with Epagri 108.

Epagri 108, Diamante, Irat 122, Maninjou and Canela Curta showed the highest estimates ($\alpha = 0.05$) of varietal effect (i.e. the *per se* effect) for GY, while the genitors Lageado, Canarinho, Bico Roxo and Tox 503 showed negative estimates (Table 4). The Maninjou genitor, in addition to having a prominent varietal effect, increased, on average, the GY of the crosses in which it participated, considering the two generations evaluated. On the other hand, despite having a negative varietal effect, on average, Lageado also increased the GY in both generations. This demonstrates the importance of conducting studies to select potential genitors for breeding programs through diallelic crosses, because, despite the negative and significant varietal effect, Lageado participated in specific crosses with potential to generate lines with high GY. The genitors that showed the highest varietal effect also showed the highest values for general combining ability in F2 and F7, with the exception of Irat 122 (Table 4).

The specific heterosis or specific combining ability - SCA (S_{ij}) is an important parameter for the identification of the best combinations of pair-on-pair parents (Asfaliza et al., 2012). Among the 61 combinations evaluated for GY, eight were positive and significant in generation F2 and two in F7, while the negative and significant estimates were eight in F2 and three in F7 (Table 5). The trend of reduction of specific heterosis, resulting from the advances in inbreeding, was already expected, because in F7 generation the dominance effects are practically null. Thus, contributions in medium and specific heterosis in combinations in advanced generations, as in F7, are the result of epistatic effects involving additive effects that have been accumulated with generations of

self-fertilization. In addition, the importance of specific heterosis will be manifested if there are considerable differences between the average allele frequencies of at least part of the parents. Therefore, differences between degrees of complementation of frequencies may also be responsible for the manifestation of this type of combining ability.

Considering the mean SCA values of the cross groups classified as Indica x Indica (acronym I x I), Indica x Japonica (I x J), and Japonica x Japonica (J x J), in F2 the highest mean SCA was between I x I crossings, and the lowest I x J, while in F7, the largest was between J x J, and the smallest was, once again, for I x J crosses. It is expected that crosses of type I x J show higher heterosis, as reported by Najeeb et al. (2013). However, the broad genetic diversity of the diallelic genitors of the present study resulted in the greater heterosis within the Indica and Japonica groups.

Estimates of varietal heterosis in the early F2 generation were used to predict SCA of crosses in advanced F7 generation of the diallel. Considering the three genitors that showed positive and significant varietal heterosis in F2, Maninjau participated in a cross with positive and significant SCA (with Pratinha Branco) and one with negative and significant SCA (with Araguaia) in F7, while Canela Curta participated in a cross with negative SCA (with Canarinho) (Table 5). Considering the three genitors with negative and significant varietal heterosis in F2, CT11632 and Pratinha Branco were present in a cross with positive SCA in F7 (with Araguaia and Maninjau, respectively), whereas CT11632 and Irat 122 were genitors of a cross with a negative value of SCA in F7, as result of depression by endogamy. These results indicate that estimates of varietal heterosis were not a good predictor for SCA in F7, and this has implications for the use of generation F2 to predict the potential crosses with higher GY in F7. According to Gärtner et al. (2009), specifically regarding GY, heterosis can vary very strongly depending on the specific combination, and in this case the selection of the best parental combination may be erroneously determined. Additionally, one of the ways to improve the heterosis prediction would be the use of molecular markers to infer the complementarity between the genitors. The genitors with the highest general combining ability (GCA) values (g_i) can be included in breeding programs of self-pollinating crops in order to develop genotypes that exceed their genitors in advanced generations (Fasahat et al., 2016). In the present study, of the 12 genitors evaluated, four showed high and positive in both generations (Canela Curta, Diamante, Epagri 108 and Maninjau). At least one of these four genitors was involved in the 20 most productive combinations, and in only four of the 20 least productive combinations in F2, while in F7 they were present in 19 and four combinations, respectively. In addition, these genitors also had outstanding per se productivity, as they were among the five most productive in the joint analysis of generations F2 and F7, demonstrating the importance of the contribution of its additive effects (Table 1). Huang et al. (2015) also observed that parents with higher GCA showed higher productivity performance. Of the four parents with high g_i , only Maninjau was positive and significant in the five estimates evaluated for GY (Table 4). In self-pollinating species, defining the combining ability of a given genotype in very early generations may lead to an erroneous conclusion about their capacity to derive promising inbred lines, because for these species is not important the effects of dominance, but the additive effects of homozygous loci (Lorencetti et al., 2006; Bertan et al., 2009). GCA is usually related to genes in additivity (Huang et al., 2015), suggesting that selection for the GY trait could be performed in early generations (Asfaliza et al., 2012). However, non-additive interactions are markedly present

in early generations, such as dominance, over-dominance, and epistasis, which are difficult to work with in the progeny due to low heritability and interactions with the environment (Liu et al., 2015; Morais Júnior et al., 2017b), which may induce an error in the choice of genitors with greater productive potential.

Earliness analysis

As described in this study, the Epagri 108, Diamante, Maninjau and CanelaCurta accessions were recommended as potential sources for obtaining lines with greater productive potential. With the exception of CanelaCurta, these genitors had a varietal effect (*per se*) that was favorable and significant for earliness in flowering. However, for the other estimated parameters, only Maninjau showed earliness, and only as GCA in generation F7. In addition, it was observed that GY specific heterosis correlated positively with DTF specific heterosis, indicating that the most productive combinations were also the latter ones ($p < 0.01$). Late flowering in rice allows higher biomass accumulation, and consequently higher productivity (Jung and Müller, 2009; Gao et al., 2014), while very early varieties have comparatively lower productivity (Shahriar et al., 2014). On the other hand, the parents that generated earlier specific combinations, in both F2 and F7, were Bico Roxo, Canarinho, Irat 122 and Lageado. Of these, only Irat 122 had a significant GCA for earliness (in F2 and F7), along with Araguaia, CT11632, Pratinha Branco and Tox503 (Table 6). For DTF in generation F2, all the parents were present in at least one cross that showed positive and significant SCA (increase in the cycle), with particular emphasis on Pratinha Branco (present in four crossings). The Maninjau accession, on the other hand, was present in three crosses that showed GCA which significantly decreased the cycle (Table 7). In F7, however, Canela Curta participated in three crosses with negative and significant SCA, while Bico Roxo, Lageado, Araguaia, Irat 122 and Maninjau did not participate in crosses with negative and significant SCA for DTF. In the case of Maninjau, a prominent genitor for GY, none of the three crosses in which it participated showed SCA for earliness in F7 (Table 7).

SSR markers to predict GY and earliness

One of the main applications of combining ability estimation is the possibility of defining heterotic groups which, for example, form the basis of hybrid programs of rice (Li et al., 2008) and corn (Fasahat et al., 2016). Molecular markers with high PIC (Polymorphism Information Content) can be used to predict the performance of hybrids (Soni et al., 2018), but the correlation is highly dependent on the predictive capacity of the model and the genetic structure of individuals (Dan et al., 2016). In the present study, no significant correlation was found between the genetic distance of the parents, obtained by the Rogers-W coefficient (RW acronym), and the specific heterosis in F2 and SCA in F7 generations (data not shown). The mean RW distance of the group of genitors, obtained by the panel of 24 SSR markers, was high (0.88; Table 8), and is usually reported in groups of genotypes with great genetic variability, which is indicative of unadapted accessions (Borba et al., 2009).

From the mean genetic distance data of the parents at each crossing, the overall mean for each parent was obtained. In accordance with this criteria, the most divergent were Epagri 108 and Irat 122 (both with RW = 0.93), and the least divergent was Lageado landrace (0.71). The diallelic combination with the lowest RW distance occurred between Pratinha Branco x Canela Curta landraces (0.45), while 21 combinations showed maximum distance (1.0). Considering the groups formed in accordance with the origin of the materials, the largest mean distance was in crosses between improved varieties from Brazil (IVB x IVB) and between improved varieties from other countries (IVC x IVC) (RW = 0.92). Crossings between landraces (LRC x LRC) showed smaller mean distance (RW = 0.74) in relation to those two groups. This result indicates that the materials developed by rice breeding programs analyzed in this study have good genetic diversity. With the mean RW values (RW_m), calculated from the RW of each genitor of a given cross, classes comprising high genetic distance (classes 1 and 2, RW_m 0.9 to 1), intermediate (classes 3 and 4, RW_m 0.7 to 0.89) and low (class 5, RW_m less than 0.7) were obtained (Table 9). Class 1, with 21 crosses, resulted in the highest GY average in generation F7, while class 3, with 16 crosses, showed the highest GY average in F2. Class 4, with 11 crosses, showed the highest earliness average in generation F2, while class 2, with 9 crosses, had the highest earliness average in generation F7. Crosses belonging to the two classes with the greatest RW_m (classes 1 and 2) resulted in the combinations with higher GY average and earliness in F7, while in F2, the combinations of genitors with the highest mean GY and earliness belonged to the classes with intermediate RW_m genetic distances (classes 3 and 4). Based on these results, in theory, in a breeding program that uses routinely the germplasm genotyping with molecular markers, pairs of genitors with intermediate RW_m values should be chosen, in order to generate superior individuals early in F2 generation. If in such programs the objective would be to proceed the generation advancement of progenies (e.g., by bulk method), for selection of lines in advanced generations, pairs of genitors resulting in high RW_m genetic distance between them should be chosen.

Table 1. Adjusted means (considering F₂ and F₇) for grain yield (GY, in kg ha⁻¹) and days to flowering (DTF) of the diallel genitors and the check cultivars used in the experiments. I: *Oryza sativa* subspecies Indica, J: *O. sativa* subspecies Japonica.

Genotypes	Code	I/J	Country	Group*	Grain Yield**	DTF
Genitors						
Bico Roxo	BGA012924	I	Brazil	LRC	4441.0 e	128.53 a
Canarinho	BGA013974	I	Brazil	LRC	3054.0 f	125.93 b
Canela Curta	BGA013191	I	Brazil	LRC	6645.0 b	107.22 c
Lageado	BGA013263	I	Brazi.	LRC	2864.33 f	128.70 a
Pratinha Branco	BGA013619	J	Brazil	LRC	5348.50 d	90.72 e
Araguaia	BGA004206	J	Brazil	IVB	4814.50 e	85.39 g
Diamante	BGA004899	I	Brazil	LRC	7523.83 a	99.55 d
Epagri108	BGA010688	I	Brazil	IVB	7772.66 a	100.55 d
CT11632	BGA009115	J	Colombia	IVC	5881.83 c	89.22 f
Irat 122	BGA004752	J	France	IVC	7067.66 a	101.89 d
Maninjau	BGA006910	I	Indonesia	IVC	6991.66 a	92.89 e
Tox 503	BGA004788	J	Nigeria	IVC	4579.66 e	89.55 f
Check cultivars						
BRS Alvorada	-	I	Brazil		7337.87 a	101.33 a
BRS Irga 409	-	I	Brazil		7204.10 a	94.20 b
BRS Taim	-	I	Brazil		6887.32 b	99.33 a
BRS Irga 424	-	I	Brazil		6720.32 b	87.33 c

* LRC: landrace; IVB: improved varieties from Brazil; IVC: improved varieties from other countries.

** Means followed by equal letters in the column do not differ from each other by Scott-Knott test at 5% probability.

Table 2. Joint analysis of variance involving the two generations (F₂ and F₇) for the traits of grain yield (GY) and days to flowering (DTF).

FV	GY			DTF		
	DF	SS	MS	DF	SS	MS
Generation (K)	1	37,192,043	37,192,043**	1	12	12 ^{ns}
Block/K	4	6,620,976	1,655,244 ^{ns}	4	98	24 ^{ns}
Group (G)	3	55,990,222	18,663,407**	3	1,400	466**
Treatment/G	134	1,072,307,950	8,002,298**	134	32,311	241**
Diallel/G (D)	131	1,070,999,237	8,175,567**	131	31,592	241**
Genitors/G (P)	11	172,678,855	15,698,078**	11	14,924	1,356**
F ₂ Crossings/G (C ₂)	60	700,049,229	11,667,487**	60	12,664	211**
F ₇ Crossings/G (C ₇)	60	198,266,541	3,304,442**	60	4,007	66**
Checks/G (T)	3	1297,684	432,561 ^{ns}	3	717	239 ^{ns}
Residues	307	341,332,399	1,111,832	313	4,196	13

General Means	6.210,95			100,41		
Genitors	5.617,66 c #			102,36 a #		
F ₂ Crossings	4.834,02 d			94,77 b		
F ₇ Crossings	6.674,49 b			91,23 c		
Checkers	7.030,24 a			95,56 b		
F ₂ Generation	4.972,29 b			95,91 a		
F ₇ Generation	6.726,97 a			93,25 b		

CV%	16,98			3,65		

** , ns: significant at 1% probability and not significant by the F test, respectively.

#Means followed by equal letters in the column do not differ from each other by Scott-Knott test at 5% probability.

Table 3. Variance analysis of the non-orthogonal decomposition of the sum of squares for grain yield (GY) in kg ha⁻¹ and days to flowering (DTF).

FV	GY			DTF		
	DF	SS	MS	DF	SS	MS
Treatments	131	1.07	817,556,669 **	131	315.92	241.2**
Varieties	11	1.79	163,162,693 **	11	148.66	1351.4**
Het. in F ₂	60	4.24	706,088,028 **	60	191.02	318.4**
Mean Het. F ₂	1	2.76	2,759,158 ns	1	288.30	2883.0**
$\hat{h}_v F_2$	11	7.41	673,308,136 **	11	339	308.2**
$\hat{s}_y F_2$	48	3.56	742,306,008 **	48	295.45	61.6**
Het. in F ₇	60	144,001,440	2,400,024 **	60	825.35	137.6**
Mean SCA F ₇	1	3,299,152	3,299,152 ns	1	214.28	2142.8**
$\hat{h}_v F_7$	11	5,388,1976	489,836,145 **	11	261.53	237.8**
$\hat{s}_y F_7$	48	982,276,130	20,464,086 **	48	1,546,116	32.2**
Residue	307	3,41	111.183.192	313	4,196,924	134,057,508
CV%	18.20			3.54		

** , ns: significant at 1% probability and not significant by the F test, respectively.

Table 4. Estimates of the effects of varieties (\hat{v}), varietal heterosis in the generations F₂ ($\hat{h}_v F_2$) and F₇ ($\hat{h}_v F_7$), general combining ability (\hat{g}_{F_2} and \hat{g}_{F_7}) and mean heterosis, for the grain yield among twelve genitors, in kg ha⁻¹.

Genitors	\hat{v}	$\hat{h}_v F_2$	$\hat{h}_v F_7$	\hat{g}_{F_2}	\hat{g}_{F_7}
Canela Curta	1.053.22 *	1,148.21 **	302.58 ns	1,674.82	829.19
Araguaia	-777.27 ns	-75.54 ns	-178.71 ns	-464.19	-567.36
Bico Roxo	-873.70 *	-536.19 ns	-8.42 ns	-973.05	-445.27
Canarinho	-2,265.12 **	556.16 ns	932.54 **	-576.39	-200.01
CT11632	290.05 ns	-782.37 **	-859.18 **	-637.35	-714.16
Diamante	1.932.05 *	-494.11 ns	-322.66 ns	471.91	643.36
Irat 122	1.475.88 **	-886.56 ns	-1,002.4 **	-148.62	-264.46
Lageado	-3160.5 **	886.744 **	1,055.79 *	-693.51	-524.46
Maninjuau	1,399.88 *	944.74 **	1,000.34 *	1,644.69	1,700.29
Pratinha Branco	-243.27 ns	-577.90 *	-204.41 ns	-699.55	-326.06
Tox 503	-1,012.11 *	20.13 ns	-724.70 ns	-485.92	-496.81
Epagri 108	2,180.88 **	203.29 ns	3,57 *	887.15	365.74
Mean Heterosis		-287.88 ns	148.93 ns		
General Mean	5,791.9				
CV (%)	18.20				

ns: not significant. * and **: significant by the t test, 5% and 1% probability, respectively.

Table 5. Estimates of specific heterosis (\hat{s}_y) of the crosses in F₂ (above the diagonal) and specific combining ability (in \hat{s}_y) F₇ (below the diagonal) for the grain yield, in kg ha⁻¹.

F ₂	C. Curta	Araguaia	B. Roxo	Canarinho	CT11632	Diamante	Irat 122	Lageado	Maninjuau	P. Branco	Tox 503	Epagri 108
Canela Curta	-	3,600.2**	-263.30	175.11	-636.92	-1,928.17**	46.84	1,615.15	-1,121.22	-384.97	145.89	-1,248.68
Araguaia	3.56	-	662.78	-1,451.88**	1,071.83	-1,734.93**	-421.89	-1,025.01	-1,573.7**	950.28	711.40	-789.16
Bico Roxo	267.48	-497.46	-	NA	-883.80	NA	-473.03	NA	1,677.25**	-1,284.11	-632.48	1,196.69
Canarinho	-1,932.28**	434.77	NA	-	-1,724.47**	3,265.04**	-277.93	NA	9,005.7	-2,175.52**	616.60	667.28
CT11632	79.36	1,618.41*	403.83	202.07	-	-329.267	-148.73	-1,168.1	-153.29	2,866.19**	1,470.81*	-364.25**
Diamante	180.84	208.89	NA	-33.94	413.69	-	NA	2,121.15	-584.05	-544.06	-766.69	500.98
Irat 122	1,639.66	379.71	144.63	-627.12	-2,208.98**	NA	-	-775.32	-257.51	452.46	185.59	1,669.52**
Lageado	206.66	793.21	NA	NA	-26.48	1,042.0	41.81	-	1,839.38**	-2,202.65**	-1,107.77	703.15
Maninjuau	-69.58	-2,598.03**	-67.11	1,571.62	-400.23	-633.25	-964.43	-111.93	-	1,116.90	-1,290.47	-559.04
Prat. Branco	-295.73	-246.18	304.51	-745.02	-164.38	-288.40	-100.08	-1,061.08	2,592.66**	-	1,824.51**	-619.05
Tox 503	-535.48	204.06	-27.51	531.72	137.86	-31.65	1,178.16	-1,016.83	249.91	-148.23	-	-1,157.43
Epagri 108	455.46	-300.98	-528.56	598.17	-55.18	-858.20	516.61	132.61	430.36	151.71	-542.03	-

NA: missing crosses. * and **: significant by the t-test, 5% and 1% probability, respectively.

Table 6. Estimates of the effects of varieties (\hat{v}), varietal heterosis in the generations F₂ (\hat{h}_{F_2}) and F₇ (\hat{h}_{F_7}), overall combining ability (\hat{g}_{F_2} and \hat{g}_{F_7}) and mean heterosis, for the trait days to flowering among twelve rice genitors.

GENITORS	\hat{v}	\hat{h}_{F_2}	\hat{h}_{F_7}	\hat{g}_{F_2}	\hat{g}_{F_7}
Canela Curta	3.87**	2.96**	1.74**	4.90	3.68
Araguaia	-17.95**	2.98**	3.96**	-5.99	-5.02
Bico Roxo	25.18**	-7.61**	-8.31**	4.98	4.28
Canarinho	22.59**	-4.48**	-7.13**	6.81	4.16
CT11632	-14.12**	1.64**	3.49**	-5.42	-3.57
Diamante	-3.79**	2.59**	4.48**	0.70	2.59
Irat 122	-1.45**	-5.46**	-2.69*	-6.19	-3.43
Lageado	25.35**	-6.69**	-8.51**	5.98	4.16
Maninjabu	-10.45**	8.28**	2.36*	3.06	-2.87
Pratinha Branco	-12.62**	2.17*	3.79**	-4.14	-2.52
Tox 503	-13.79**	0.95**	3.53**	-5.94	-3.37
Epagri 108	-2.79**	2.65**	3.28*	1.26	1.88
Mean Heterosis	-	-8.25**	-11.18**	-	-
General Mean	103.35				
CV(%)	3.54				

ns; not significant. * and **: significant by the t-test, 5% and 1% probability, respectively.

Table 7. Estimates of specific heterosis (\hat{s}_{ij}) of crosses in F₂ (above the diagonal) and specific combining ability (\hat{c}_{ij}) in F₇ (below the diagonal) for the trait days to flowering.

F ₂ F ₇	C. Curta	Araguaia	B. Roxo	Canarinho	CT11632	Diamante	Irat 122	Lageado	Maninjabu	P. Branco	Tox 503	Epagri 108
Canela Curta	-	7,87**	2,54	5,07**	-2,95	-2,56	-1,17	5,65**	-2,42	-2,47	-3,92**	-5,62**
Araguaia	2,09	-	-0,96	-3,28	2,93	-0,42	-1,28	-2,26	-6,28**	0,16	2,71	0,76
Bico Roxo	3,79	-1,00	-	NA	-3,28	-2,25	5,74	-2,00	-4,01**	4,29	-0,61**	3,77**
Canarinho	1,41	-1,88	NA	-	-0,61	3,77**	-2,58	NA	-0,33	-4,69**	-0,08	2,71
CT11632	-4,86**	0,84	-2,95	0,16	-	-1,49	1,14	-3,71**	-3,61*	6,33**	7,38**	-2,06
Diamante	8,48**	-1,31	NA	1,51	0,73	-	NA	4,60**	-0,47	-0,52	-0,72	-2,17
Irat 122	-6,00	-0,80	1,40	6,02**	-2,25	NA	-	-3,75**	0,41	1,86**	1,16	6,46**
Lageado	3,91	-3,38	NA	NA	2,66	1,01	-2,97	-	4,99**	-5,80	-3,00	3,29
Maninjabu	-1,56	-1,86	-1,65	-3,53	0,19	-3,46	1,05	-1,03	-	5,11**	-0,58	-2,53
Prat. Branco	-6,41**	2,79	-1,50	-3,88	4,84**	-2,31	1,20	-1,88	8,14**	-	4,11**	-2,08
Tox 503	6,44**	1,14	-0,15	-4,53**	4,69**	-6,96**	0,05	-0,53	4,49	-0,36**	-	-3,03
Epagri 108	-7,31**	3,39	2,09	4,71**	-4,06	2,28	2,30	2,21	-0,76	-0,61	-4,26	-

NA: missing crosses. **: significant by the t-test, 5% and 1% probability, respectively.

Table 8. Genetic distance (RW coefficient) among pairs of diallel parents, (*) - distances of the crosses that were not evaluated for grain yield and days to flowering. Mean RW: Mean of the distances between a genitor and all other genitors.

Mean RW	Genitors	Lageado	BicoRoxo	Canarinho	CanelaCurta	PratinhaBranco	Epagri 108	Diamante	Araguaia	Maninjabu	Irat122	CT11632
0.72	Lageado	0.00										
0.91	BicoRoxo	0.66 (*)	0.00									
0.95	Canarinho	0.61 (*)	0.91 (*)	0.00								
0.84	CanelaCurta	0.59	0.82	1.00	0.00							
0.87	PratinhaBranco	0.60	0.77	1.00	0.45	0.00						
0.93	Epagri 108	0.84	1.00	1.00	1.00	1.00	0.00					
0.90	Diamante	0.79	0.89 (*)	0.89	1.00	1.00	0.80	0.00				
0.88	Araguaia	0.73	0.89	0.89	0.89	0.87	0.97	1.00	0.00			
0.91	Maninjabu	0.81	1.00	0.89	1.00	1.00	0.79	0.71	0.95	0.00		
0.93	Irat 122	0.79	0.89	1.00	1.00	1.00	0.91	0.91 (*)	0.94	0.90	0.00	
0.83	CT11632	0.60	0.87	0.87	0.74	0.83	0.97	0.94	0.71	0.94	0.87	0.00
0.92	Tox 503	0.75	1.00	1.00	0.77	1.00	1.00	0.96	0.88	1.00	1.00	0.79

Table 9. Mean grain yield (GY, in kg ha⁻¹) data and number of days to flowering (DTF) in RW genetic distance classes.

Class	RW range*	Number of Crosses	GY average F2	GY average F7	DTF average F2	DTF average F7
1	1 (1)	21	5602.1	6346.5	95.5	91.8
2	0.9-0.99 (0.94)	9	5650.3	5864.1	90.3	86.9
3	0.8-0.89 (0.87)	16	5901.4	5874.9	94.4	90.7
4	0.7-0.79 (0.76)	11	5053.6	5480.6	89.6	88.0
5	<0.7 (0.56)	4	4777.0	5326.3	91.1	89.3

* Between parenthesis: Average of mean RW within each class

Conclusions

On average, the Lageado and Maninjabu genitors contributed to increase the grain yield of the combinations that they were part of, in both generations, while the CT11632 and Irat 122 genitors showed the opposite behavior. The genitors that showed the highest

varietal effect and general combining ability for GY in generations F2 and F7 were Epagri 108, Diamante, Maninjau and Canela Curta; thus they can be recommended for the breeding program as important genetic assets to obtain lines with greater grain yield potential. Regarding days to flowering, the genitors with the largest earliness estimates were Araguaia, CT11632, Irat 122, Pratinha Branco and Tox 503. The most suitable parameters for estimating grain yield in generation F7, from the data attained from F2, were the average grain yield, general combining ability and the RW genetic distance between pairs of genitors, obtained through genotyping using SSR markers.

References

ABADIE, T.; CORDEIRO, C.M.T.; FONSECA, J.R.; ALVES, R.B.N.; BURLE, M.L.; BRONDANI, C.; RANGEL, P.H.N.; CASTRO, E.M.; SILVA, H.T.; FREIRE, M.S.; ZIMMERMANN, F.J.P.; MAGALHÃES, J.R. 2005. Construção de uma coleção nuclear de arroz para o Brasil. **Pesquisa Agropecuária Brasileira**, 40:129–136.

ASFALIZA, R.; RAFII, M.Y.; SALEH, G.; OMAR, O.; PUTEH, A. 2012. Combining ability and heritability of selected rice varieties for grain quality traits. **Australian Journal of Crop Sciences**, 6:1718–1723.

BERTAN, I.; CARVALHO, I.F.I.F.; OLIVEIRA, A.C.; SILVA, J.A.G.; BENIN, G.; HARTWIG, I.; SCHMIDT, D.M.; VALÉRIO, I.P.; FONSECA, D.R.; SILVEIRA, G. 2009. Efeitos da heterose e endogamia em caracteres de importância agrônômica em trigo. **Revista Ceres**, 56:753–763.

BHULLAR, G.S.; GILL, K.S.; KHEHRA, A.S. 1979. Combining ability analysis over F1-F5 generations in diallel crosses of bread wheat. **Theoretical and Applied Genetics**, 55:77–80.

BORBA, T.C.O.; BRONDANI, R.P.V.; RANGEL, P.H.N.; BRONDANI, C. 2009. Microsatellite marker-mediated analysis of the Embrapa Rice Core Collection genetic diversity. **Genetica**, 137:293-304.

BROWN, A.H.D. 1989. Core collections: A practical approach to genetic resources management. **Genome**, 31:818–824.

BUENO, L.G.; VIANELLO, R.P.; RANGEL, P.N.; UTUMI, M.M.; CENTENO, A.C.; PEREIRA, J.A.; FRANCO, D.F.; NETO, F.M.; MENDONÇA, J.A.; COELHO, A.S.; OLIVEIRA, J.P.; BRONDANI, C. 2012. Adaptabilidade e estabilidade de acessos de uma coleção nuclear de arroz. **Pesquisa Agropecuária Brasileira**, 47:216–226.

CHARLESWORTH, D.; BARTON, N.H.; CHARLESWORTH, B. 2017. The sources of adaptive variation. **Proceedings of the Royal Society B**, 284:20162864.

CRUZ, C.D.; REGAZZI, A.J. 1994. **Modelos biométricos aplicados ao melhoramento genético**. UFV, Viçosa, 390p.

CRUZ, C.D. 2013. GENES - a software package for analysis in experimental statistics and quantitative genetics. *Acta Scientiarum*, 35:271–276.

DAN, Z.; HU, J.; ZHOU, W.; YAO, G.; ZHU, R.; ZHU, Y.; HUANG, W. 2016. Metabolic prediction of important agronomic traits in hybrid rice (*Oryza sativa* L.). *Scientific Reports*, 6:21732.

DENNIS, E.S.; ELLIS, J.; GREEN, A.; LLEWELLYN, D.; MORELL, M.; TABELL, L.; PEACOCK, W.J. 2008. Genetic contributions to agricultural sustainability. *Philosophical Transactions of the Royal Society B*, 363: 591–609.

FASAHAT, P.; RAJABI, A.; RAD, J.M.; DERERA, J. 2016. Principles and utilization of combining ability in plant breeding. *Biometrics & Biostatistics International Journal*, 4:00085

GAO, H.; JINA, M.; ZHENG, X.M.; CHEN, J.; YUAN, D.; XIN, Y.; WANG, W.; HUANG, D.; ZHANG, Z.; ZHANG, Z.; ZHOU, K.; SHENG, P.; MA, J.; MA, W.; DENG, H.; JIANG, L.; LIU, S.; WANG, H.; WU, C.; YUAN, L.; WAN, J. 2014. Days to heading 7, a major quantitative locus determining photoperiod sensitivity and regional adaptation in rice. *Proceedings of the National Academy of Sciences of the United States of America*, 111:16337–16342.

GARDNER, C.O.; EBERHART, S.A. 1966. Analysis and interpretation of the variety cross diallel and related populations. *Biometrics*, 22:439–452.

GÄRTNER, T.; STEINFATH, M.; ANDORF, S.; LISEC, J.; MEYER, R.C. 2009. Improved heterosis prediction by combining information on DNA and metabolic markers. *PLoS ONE*, 4:e5220.

HUANG, M.; CHEN, L.Y.; CHEN, Z.Q. 2015. Diallel analysis of combining ability and heterosis for yield and yield components in rice by using positive loci. *Euphytica*, 205:37–50.

HUANG, S., WANG, L.; LIU, L.; FU, Q.; ZHU, D. 2014. Nonchemical pest control in China rice: a review. *Agronomy for Sustainable Development*, 34:275–291

JUNG, C.; MÜLLER, A.E. 2009. Flowering time control and applications in plant breeding. *Trends in Plant Science*, 14:563–573.

LI, L.; LU, K.; CHEN, Z.; MU, T.; HU, Z.; LI, X. 2008. Dominance, overdominance and epistasis condition the heterosis in two heterotic rice hybrids. *Genetics*, 180:1725–1742.

LIU, C.; SONG, G.;ZHOU, Y.;QU, X.;GUO, Z.; LIU, Z.; JIANG, D.; YANG, D. 2015. OsPRR37 and Ghd7 are the major genes for general combining ability of DTH, PH and SPP in rice. **Scientific Reports**, 5:12803.

LORENCETTI, C.; CARVALHO, F.I.F.; OLIVEIRA, A.C.; VALÉRIO, I.P.; BENIN, G.; ZIMMER, P.D.; VIEIRA, E.A. 2006. Distância genética e sua associação com heterose e desempenho de híbridos em aveia. **Pesquisa Agropecuária Brasileira**,41:591–598.

MORAIS JÚNIOR, O.P.; BRESEGHELLO, F.; DUARTE, J.B.; MORAIS, O.P.; RANGEL, P.H.N.; COELHO, A.S.G. 2017a. Effectiveness of recurrent selection in irrigated rice breeding. **Crop Science**, 57:3043–3058.

MORAIS JÚNIOR, O.P.; PEREIRA, J.A.;MELO, P.G.S. GUIMARÃES, P.H.R.; MORAIS, O.P. 2017b. Gene action and combining ability for certain agronomic traits in red rice lines and commercial cultivars. **Crop Science**, 57:1295–1307.

NAJEEB, S.; AHANGAR, A.; DAR, S.H. 2013. An analysis of hybrid sterility in rice (*Oryza sativa* L.) using genetically diverse germplasm under temperate ecosystem. **African Journal Agricultural Research**, 8:3820–3827.

PEREIRA, J.A.; MORAIS, P.O.; BRESEGHELLO, F. 2008. Análise da heterose de cruzamentos entre variedades de arroz-vermelho. **Pesquisa Agropecuária Brasileira**, 43:1135–1142.

R CORE TEAM. 2013. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0.

SEARLE, S.R. 1971. **Linear models**. John Wiley & Sons, New York, 532p.

SHAHRIAR, M.D.; ROBIN, A.H.K.; HOQUE, A. 2014. Diversity assessment of yield, yield contributing traits, and earliness of advanced T-aman rice (*Oryza sativa* L.) lines. **Journal of Bioscience and Agriculture Research**, 1:102–112.

SILVA, S.A.G.; MORAIS, O.P.;RAVA, C.A.; COSTA, J.G.C. 2000. Método generalizado de análise de dialelos desbalanceados. **Pesquisa Agropecuária Brasileira**, 35:1999–2005.

SONI, S.K.; TIWARI, S.; NEWMAN, J.T.; DOSSOU-AMINON, I.; SUNDARAM, R.M. 2018. Prediction of hybrid performance in crop plants: molecular and recent approaches. **International Journal of Current Microbiology and Applied Sciences**, 7:98–108.

SPRAGUE, G.F.; TATUM, L.A. 1942. General versus specific combining ability in single crosses of corn. **Journal of American Society of Agronomy**, 34:923–932.

WRIGHT, S. 1978. Evolution and the genetics of populations. Variability within and among natural populations. **University of Chicago Press**, Chicago, 590p.

YAN, W.; RUTGER, J.N.; BRYANT, R.J.; BOCKELMAN, H.E.; FJELLSTROM, R.G.; CHEN, M.H.; TAI, T.H.; MCCLUNG, A.M. 2007. Development and evaluation of a core subset of the USDA rice (*Oryza sativa* L.) germplasm collection. **Crop Science**, 47:869–878.

ZHANG, P.; LI, J.; LI, X.; LIU, X.; ZHAO, X.; LU, Y. 2011. Population structure and genetic diversity in a rice core collection (*Oryza sativa* L.) investigated with SSR markers. **PLoS ONE**, 6:e27565.