

## Mapeamento Associativo para Produtividade em Arroz sob Déficit Hídrico

Gabriel Feresin Pantalhão<sup>1,2</sup>, Tereza Cristina de Oliveira Borba<sup>3</sup>, Cleber Morais Guimarães<sup>4</sup>, Marcelo Gonçalves Narciso<sup>5</sup>, Rosana Pereira Vianello<sup>6</sup>, Claudio Brondani<sup>7</sup>

### Resumo

Esse trabalho objetivou detectar, via Genotipagem por Sequenciamento (GBS), o polimorfismo de marcadores SNPs em 283 acessos de arroz de terras altas componentes da CNAE (Coleção Nuclear de Arroz da Embrapa) e associá-los à produtividade sob déficit hídrico. Após a filtragem dos dados brutos, foram contabilizados 285.379 SNPs distribuídos ao longo dos 12 cromossomos do arroz. As informações moleculares foram integradas aos dados fenotípicos derivados do experimento de avaliação de produtividade e Índice de Suscetibilidade à Seca (ISS), conduzido no ano de 2010 em Porangatu (GO) em ambiente com e sem deficiência hídrica, para possibilitar a análise de mapeamento associativo e, com isso, detectar marcadores SNPs relacionados à tolerância à seca e oportunizar o desenvolvimento de um conjunto de marcadores úteis para a seleção assistida para esse caráter, assim como genes para estudos de engenharia genética do arroz. Através da análise de associação, foram detectados 48 SNPs relacionados com os caracteres avaliados, dentre os quais 13 foram relacionados ao ISS e 35 à produtividade em condição de déficit hídrico. Dentre os 48 SNPs, foram identificados 35 SNPs ancorados em 31 genes de arroz. Dentre os genes identificados, sete deles continham SNPs associados ao ISS, enquanto que os restantes 24 genes continham SNPs associados à produtividade dos acessos em ambiente com deficiência hídrica. Esses genes podem ser avaliados para serem efetivamente utilizados na seleção assistida por marcadores. Adicionalmente, esses genes podem ser superexpressos para avaliar sua capacidade de aumentar a tolerância à seca, e em caso positivo, gerar cultivares comerciais de arroz geneticamente modificadas mais tolerantes a esse estresse.

### Introdução

O arroz é uma das espécies mais importantes para a alimentação humana, representando a principal fonte de carboidrato para mais da metade da população mundial (Khush 2005). Estima-se que a população mundial chegará a nove bilhões dentro dos próximos 40 anos, o que irá demandar um aumento expressivo na produção de alimentos em relação aos níveis atuais (Godfray et al. 2010), em um cenário de redução da área cultivável e escassez de recursos hídricos. A seca é um fator ambiental que limita a produção das culturas, como a do arroz de terras altas (*Oryza sativa* L.), sendo assim, os programas de melhoramento de arroz devem priorizar a busca por novas estratégias que visem ao aumento da tolerância e da produtividade em condições de déficit hídrico (Rao 2004).

O conhecimento de fatores envolvidos na tolerância à seca e das respostas das plantas de arroz ao estresse fornece as informações que servem de base para a obtenção de cultivares tolerantes e, conseqüentemente, com uma maior produtividade sob essas condições. As novas tecnologias de genotipagem e sequenciamento oferecem a oportunidade de detecção de regiões do genoma associadas a caracteres quantitativos, abrindo caminho para o melhor entendimento do controle genético desses caracteres e para a possibilidade de utilização das informações do genoma na seleção assistida por marcadores (SAM) (Tuberosa and Salvi 2006).

O mapeamento associativo, ou análise de associação, tem sido aplicado com sucesso em plantas, sendo utilizado primeiramente na identificação de genes associados a caracteres de importância econômica. O sequenciamento de nova geração (NGS) está expandindo os conhecimentos sobre a variabilidade genética do arroz. A alta qualidade dos genomas de referência publicamente disponíveis para o arroz e o resequenciamento de diversos acessos de bancos de germoplasma têm formado uma base sólida para os estudos genômicos, estruturais e funcionais. A tecnologia de genotipagem por sequenciamento (GBS) tem sido utilizada na

<sup>1</sup> Parte da dissertação de mestrado do primeiro autor

<sup>2</sup> Doutorando do Programa de Pós-graduação em Genética e Melhoramento de Plantas – UFG. e-mail: gabrielferesin@hotmail.com

<sup>3</sup> Pesquisadora da Embrapa Arroz e Feijão – CNPAF - EMBRAPA/Santo Antônio de Goiás - GO. e-mail: tereza.borba@embrapa.br

<sup>4</sup> Pesquisador da Embrapa Arroz e Feijão – CNPAF - EMBRAPA/Santo Antônio de Goiás - GO. e-mail: cleber.guimaraes@embrapa.br

<sup>5</sup> Pesquisador da Embrapa Arroz e Feijão – CNPAF - EMBRAPA/Santo Antônio de Goiás - GO. e-mail: marcelo.narciso@embrapa.br

<sup>6</sup> Pesquisadora da Embrapa Arroz e Feijão – CNPAF - EMBRAPA/Santo Antônio de Goiás - GO. e-mail: rosana.vianello@embrapa.br

<sup>7</sup> Pesquisador da Embrapa Arroz e Feijão – CNPAF - EMBRAPA/Santo Antônio de Goiás – GO. e-mail: claudio.brondani@embrapa.br

identificação de inúmeros SNPs, a um custo bastante reduzido (Elshire et al. 2011). Dessa forma, o objetivo deste trabalho foi identificar SNPs pela tecnologia GBS e associar esses SNPs com a produtividade e o índice de suscetibilidade à seca avaliados no experimento de deficiência hídrica através da análise de mapeamento associativo.

### Material e Métodos

O mapeamento associativo foi conduzido a partir de um painel composto por 283 acessos da Coleção Nuclear de Arroz da Embrapa (CNAE) pertencentes ao sistema de cultivo de sequeiro. O experimento em ambiente com e sem deficiência hídrica foi conduzido no Sítio de Fenotipagem para Seca da Embrapa Arroz e Feijão, na Estação Experimental da Agência Rural (SEAGRO), localizada no município de Porangatu – Goiás. Os 283 acessos componentes da CNAE e pertencentes ao sistema de cultivo sequeiro foram avaliados quanto à produtividade e ao índice de suscetibilidade à seca (ISS) através de um delineamento experimental do tipo Blocos Aumentados de Federer durante o ano de 2010.

O DNA genômico dos 283 acessos componentes do painel de associação foi obtido a partir de folhas jovens utilizando-se kit comercial DNeasy 96 Plant Kit (QIAGEN). Para a genotipagem em larga escala de SNP, utilizou-se a tecnologia de sequenciamento de nova geração denominada de genotipagem por sequenciamento (GBS, *Genotyping by sequencing*). O DNA genômico foi enviado e analisado pelo Instituto de Diversidade Genômica da Universidade de Cornell (*Buckler Laboratory*). Neste instituto foram construídas as bibliotecas genômicas e conduzido o resequenciamento dos indivíduos.

A análise de associação entre os marcadores SNPs e os caracteres de interesse, potencial produtivo em ambiente com e sem deficiência hídrica e índice de suscetibilidade à seca (ISS) no ano de 2010 avaliado foi realizada pelo programa computacional TASSEL versão 3.0 standalone (Bradbury et al. 2007), utilizando-se o módulo *Mixed Linear Model* (MLM). A matriz de coancestralidade (matriz K ou kinship) foi obtida a partir desse mesmo *software*. Os marcadores SNPs identificados e os dados de estruturação foram considerados como fatores de efeito fixo, enquanto que a matriz de parentesco foi considerada como fator de efeito aleatório. Para confirmar a significância das associações entre locos e caracteres fenotípicos, foi utilizado o método FDR (*false discovery rate*), obtido pelo *software* Qvalue versão 1.0 (Storey 2002).

Depois que os SNPs foram identificados para a associação com os caracteres de interesse através do Modelo Linear Misto, foram utilizadas as posições fornecidas de cada marcador SNP com a finalidade de se encontrar transcritos conhecidos do genoma do arroz publicamente disponíveis (cultivar Nipponbare; MSU Rice Genome versão 7.0).

### Resultados e Discussão

Esse experimento de deficiência hídrica conduzido em Porangatu (GO) indicou que a produtividade média dos acessos em condição com disponibilidade de água no solo mostrou-se superior à condição sob deficiência hídrica (2385,47 kg.ha<sup>-1</sup> e 1081,4 kg.ha<sup>-1</sup>, respectivamente), equivalendo a uma redução de aproximadamente 55% na produtividade, que representa um padrão severo de deficiência hídrica para materiais de ciclo curto, de acordo com Heinemann (2010).

Analisando os gráficos Boxplot para o caráter produtividade em ambos os ambientes (Figura 1), verificou-se uma grande diferença na dispersão dos dados. Em ambiente com irrigação normal, observou-se uma maior dispersão dos dados variando de 614,82 kg.ha<sup>-1</sup> a 4731,13 kg.ha<sup>-1</sup>, enquanto que em ambiente com deficiência hídrica, houve uma redução da produtividade média e da dispersão dos dados, causando um “achatamento” verificado pela variação de 519,26 kg.ha<sup>-1</sup> a 1624,08 kg.ha<sup>-1</sup> na produtividade dos acessos. Nesse caso, a resposta ao déficit hídrico reduziu a dispersão dos dados de produtividade, ou seja, os acessos tiveram um comportamento bastante semelhante frente ao estresse. Em contrapartida, no ambiente com fornecimento adequado de água, os acessos tendem a expressar de modo diferenciado seu potencial produtivo.

A dispersão dos dados de índice de suscetibilidade à seca (ISS) foi observada pelo gráfico Boxplot (Figura 2), que variou de -0,1534 a 1,5739, ou seja, houve materiais que reduziram em mais de 50% sua produtividade, mostrando-se extremamente sensíveis a essas condições (por exemplo, o acesso CT11632-3-3-M). Por outro lado, houve materiais que não alteraram sua produtividade com a redução na disponibilidade de água, como o IRAT 13.

Dentre as cultivares mais produtivas em ambiente com e sem deficiência hídrica, três apresentaram maiores produtividades em ambas as condições: IRAT 141 (linhagem melhorada no exterior), CARISMA (cultivar brasileira lançada em 1999) e MATO GROSSO (variedade tradicional coletada em Santa Catarina em 2001), indicando, portanto, materiais de alto poder exploratório em programas de melhoramento.

Nesse trabalho, a tecnologia de genotipagem por sequenciamento (GBS) foi utilizada para identificar SNPs nos 283 acessos componentes do painel de associação. Os comandos (*Pipeline*) utilizados para a identificação e posterior validação desses marcadores resultaram em um total de 516.240 SNPs distribuídos nos 12 cromossomos do arroz. Para aumentar a confiabilidade da análise baseada nos marcadores SNPs, foi realizada uma “filtragem” dos dados com uma frequência alélica de 0,05, resultando em 285.379 SNPs (Tabela 1).

O mapeamento associativo realizado com 285.379 SNPs distribuídos nos 12 cromossomos de arroz, através da abordagem de modelos mistos, identificou 48 marcadores associados de forma significativa aos caracteres avaliados (produtividade e índice de suscetibilidade à seca).

Dentre os 48 SNPs estatisticamente significativos, 13 foram associados ao ISS (índice de suscetibilidade à seca) e 35 associados à produtividade em ambiente com deficiência hídrica. De acordo, com os gráficos do tipo Manhattan Plot (Figuras 3 e 4), que evidenciam a distribuição dos SNPs significativos, as cinco associações com os maiores valores de significância foram entre os SNPs S1\_31622748, S1\_42199857, S8\_27223483, S3\_5310087 e S9\_18281732 e o caráter produtividade em ambiente sob déficit hídrico. Dos SNPs associados no mapeamento associativo, somente os cromossomos 11 e 12 não apresentaram associações.

Desses 48 SNPs, aproximadamente 73% deles (35 SNPs) estavam contidos em regiões codantes do genoma de arroz, disponível no *Rice Genome Annotation Project* (cultivar Nipponbare; MSU Rice Genome versão 7.0). Dessas regiões genômicas codantes, alguns SNPs estavam presentes no mesmo gene, resultando em 31 genes identificados. Desse total, sete genes apresentaram SNPs relacionados com o Índice de Suscetibilidade à Seca (ISS) e 24 genes apresentaram SNPs relacionados com a produtividade dos acessos sob deficiência hídrica (Tabela 2).

Esses genes significativamente associados à seca serão utilizados no desenvolvimento de ferramentas para uso na seleção assistida por marcadores. Adicionalmente, esses genes poderão ser utilizados na geração de plantas de arroz geneticamente modificadas mais tolerantes à seca.

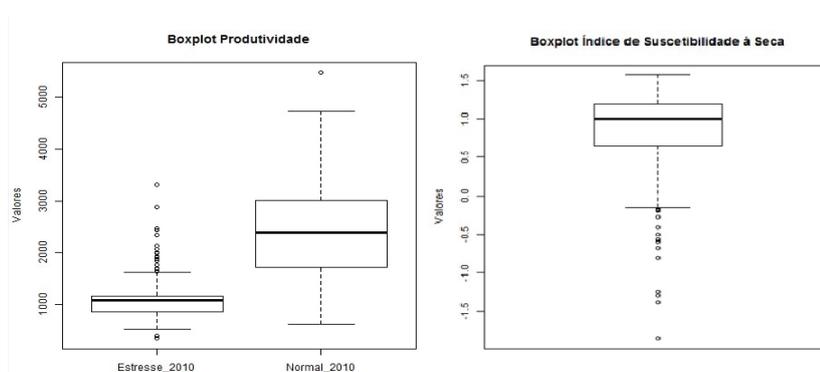


Figura 1. Gráfico boxplot para a produtividade nos ambientes com e sem deficiência hídrica.

Figura 2. Gráfico boxplot para o índice de suscetibilidade à seca.

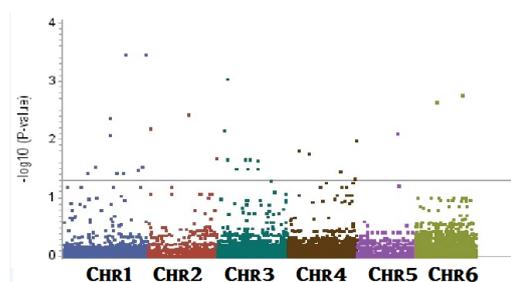


Figura 3. Gráfico Manhattan Plot dos valores de significância das associações nos cromossomos 1 a 6.

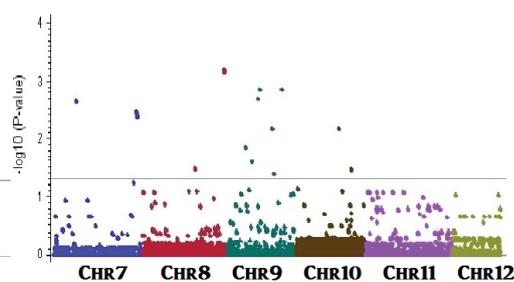


Figura 4. Gráfico Manhattan Plot dos valores de significância das associações nos cromossomos 7 a 12.

Tabela 1. Número de marcadores SNPs obtidos pela genotipagem de 283 acessos de arroz de terras altas da CNAE por GBS.

CROMOSSOMO	FMA = 0,01	FMA = 0,05
1	68866	37233
2	53217	29301
3	55374	28999
4	49842	27552
5	39588	20469
6	44376	25017
7	40913	22078
8	39421	22217
9	31866	17468
10	33344	19559
11	41780	25110
12	17653	10376
Total	516.240	285.379

Tabela 2. Genes que apresentaram os SNPs associados significativamente em cada cromossomo do arroz cultivar Nipponbare.

Cromossomo	Genes
1	5
2	2
3	7
4	3
5	1
6	1
7	3
8	1
9	6
10	2
11	0
12	0
Total	31

### Apoio

Embrapa Arroz e Feijão, Universidade Federal de Goiás (UFG) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

### Referências

- Bradbury PJ et al. (2007) TASSEL: software for association mapping of complex traits in diverse samples. **Bioinformatics**: 2633–2635.
- Elshire RJ et al. (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. **PLoS ONE**.
- Godfray HCJ et al. (2010) Food Security: The Challenge of Feeding 9 Billion People. **Science**: 812-818.
- Heinemann AB (2010) Caracterização dos padrões de estresse hídrico para a cultura do arroz (ciclo curto e médio) no estado de Goiás e suas consequências para o melhoramento genético. **Ciência e Agrotecnologia**: 29-36.
- Khush GS (2005) What it will take to feed 5.0 billion rice consumers in 2030. **Plant Molecular Biology**: 1-6.
- Rao NK (2004) Plant genetic resources: advancing conservation and use through biotechnology. **African Journal of Biotechnology**: 136-145.
- Storey JD (2002) A direct approach to false discovery rates. **Journal of the Royal Statistical Society**: 479-498.
- Tuberosa R and Salvi S (2006) Genomics-based approaches to improve drought tolerance of crops. **Trends in Plant Science**: 405-412.