

Análise preliminar de dados de sequenciamento por captura (*Capture-seq*) na identificação de marcadores SNPs associados a tolerância à seca em arroz

Ariadna Faria Vieira¹, Paula Arielle Mendes Ribeiro Valdisser², Anna Cristina Lanna³, Rosana Pereira Vianello⁴, Leandro Gomide Neves⁵, Claudio Brondani⁶

O arroz (*Oryza sativa* L.) é o segundo cereal mais produzido no planeta, e o Brasil é o nono maior produtor mundial e o maior produtor da cultura fora do continente asiático. A demanda por cereais economicamente importantes, como arroz, milho e trigo, aumentará significativamente nos próximos anos em um cenário que inclui a redução da disponibilidade de terras agricultáveis e diminuição da disponibilidade de água. Nesse momento, portanto, estudos que visem identificar genes e alelos relacionados à tolerância ao déficit hídrico, são fundamentais, pois são a base para o desenvolvimento de cultivares que possam garantir a sustentabilidade da cadeia produtiva do arroz e, conseqüentemente, atender à demanda de consumo. O sequenciamento de nova geração (NGS) tem contribuído para a identificação das variações no genoma, como os SNPs e InDels, as quais podem ser correlacionadas com o fenótipo desejável. A técnica de sequenciamento de captura (*Capture-seq*) se baseia na captura e sequenciamento de genes candidatos, previamente conhecidos, e tem o potencial de identificar variações estruturais do genoma de arroz que possam estar relacionadas com o fenótipo favorável para a tolerância à seca. O objetivo deste trabalho foi analisar um conjunto de dados de sequenciamento de captura de 300 genótipos de arroz de terras altas para 2.500 unigenes previamente identificados como relacionados a tolerância à seca na espécie. Para o sequenciamento NGS foi extraído o DNA das amostras por kit comercial e as mesmas foram quantificadas, diluídas e inseridas em placas de PCR a uma concentração de 40 ng/ μ L. As amostras foram enviadas à empresa Rapid Genomics (EUA) que executou a técnica de *Capture-seq*. Como resultado da análise preliminar dos dados de sequenciamento foi possível identificar 25.217 SNPs. A proporção de dados (bases nucleotídicas) não perdidos foi de aproximadamente 99,96% (7,5 milhões de pb identificadas). A proporção de homozigotos e heterozigotos foi, em média, 9,8% e 90,2% respectivamente. O cromossomo 1 apresentou o maior número de SNPs (4.038 SNPs), enquanto que os cromossomos 9, 10 e 12 apresentaram uma menor quantidade de SNPs identificados (1.135, 1.220 e 1.104 SNPs, respectivamente). Com a análise ainda em andamento, os próximos passos incluirão o estudo de associação entre SNPs e InDels com os dados de produtividade provenientes de dois ensaios de avaliação da tolerância à seca no campo e em casa de vegetação, por meio do programa Tassel. Espera-se identificar os melhores alelos de cada loco gênico associado à tolerância à seca e, a partir dessa informação, gerar marcadores para seleção assistida do programa de melhoramento de arroz e, eventualmente, clonar e transformar o arroz para a superexpressão desses genes.

¹ Estudante de Pós-Graduação em Genética e Melhoramento de Plantas na Universidade Federal de Goiás, estagiária da Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO, aryadnafv@hotmail.com

² Farmacêutica, analista da Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO, paula.valdisser@embrapa.br

³ Química, doutora em Ciências Agrárias, pesquisadora da Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO, anna.lanna@embrapa.br

⁴ Bióloga, doutora em Biologia Molecular, pesquisadora da Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO, rosana.vianello@embrapa.br

⁵ Engenheiro Florestal, doutor em Biologia Molecular, pesquisador da RAPID Genomics, Gainesville, Flórida, Ineves@rapids-genomics.com

⁶ Engenheiro-agrônomo, doutor em Ciências Biológicas, pesquisador da Embrapa Arroz e Feijão, Santo Antônio de Goiás, GO, claudio.brondani@embrapa.br