



# COLEÇÃO NUCLEAR DE ARROZ DE TERRAS ALTAS DIRECIONADA PARA TOLERÂNCIA À SECA: GERAÇÃO, VALIDAÇÃO, E COMPARAÇÃO DE METODOLOGIAS PARA AMOSTRAGEM DE COLEÇÃO NUCLEAR BASEADAS EM DADOS MOLECULARES

Marco Pessoa-Filho<sup>1,2</sup>, Thiago Gledson Rios Terra<sup>3</sup>, Tarcísio Castro Alves de Barro Leal<sup>3</sup>, Cléber Moraes Guimarães<sup>4</sup>, Paulo Hideo Nakano Rangel<sup>4</sup>, Márcio Elias Ferreira<sup>1, 2, 5</sup>

<sup>1</sup>Universidade de Brasília – Instituto de Biologia, Depto. de Biologia Celular, Pós-graduação em Biologia Molecular marcopessoaf@gmail.com, <sup>2</sup>EMBRAPA Recursos Genéticos e Biotecnologia ferreira@cenargen.embrapa.br, <sup>3</sup>Universidade Federal de Tocantins tarcisio@uft.edu.br, <sup>4</sup>EMBRAPA Arroz e Feijão cleber@cnpaf.embrapa.br, phrangel@cnpaf.embrapa.br, <sup>5</sup>Universidade Católica de Brasília – UCB

**Palavras-chave:** *Oryza*, coleção nuclear, microssatélites, seca

Uma coleção nuclear é definida como uma sub-amostra de acessos que representam, com o menor índice de redundância possível, a diversidade genética de uma espécie cultivada. Seu objetivo é facilitar o acesso de usuários a uma coleção de menor tamanho que represente a diversidade contida no *pool* gênico da cultura, para fins de melhoramento e pesquisa. Este trabalho propõe um algoritmo que utiliza dados de distância genética, obtidos a partir de genotipagem com marcadores moleculares, para selecionar uma coleção nuclear que atenda aos requisitos mínimos de sua definição. Perfis multi-loco de marcadores microssatélites em acessos de arroz de terras altas (*Oryza sativa* spp. *japonica*) do banco de germoplasma da EMBRAPA foram utilizados. Uma matriz de distâncias genéticas (coeficiente *Shared Allele Distance*) foi utilizada como entrada para as análises do programa COREX, desenvolvido para este trabalho. Estatísticas descritivas para as coleções obtidas foram calculadas com o programa PowerMarker 3.25. A estruturação da variabilidade genética da coleção foi avaliada com o programa STRUCTURE 2.1. Dois outros programas – Mstrat e PowerCore – para geração de coleções nucleares foram utilizados, para fins comparativos. Avaliou-se a maximização de variabilidade fenotípica, no campo, para tolerância à seca, a partir da metodologia de geração de coleções nucleares proposta. Como resultado, foi definido um grupo de amostras contendo 84,3% do total de alelos na coleção base em 18% do total de acessos. Índices de diversidade gênica tiveram um valor médio de 0,764 e foram significativamente maiores ( $p \leq 6,309 \times 10^{-5}$ ; teste  $\lambda$ ) que a média de 0,667 obtida para a coleção base. A estimativa média de *FST* diminuiu, havendo uma diferenciação moderada entre os grupos inferidos. Houve aumento significativo da variabilidade fenotípica em uma sub-coleção gerada, demonstrando que a metodologia é capaz de maximizar, também, diversidade fenotípica. De fato, a coleção apresentou variabilidade fenotípica quando avaliada no campo para caracteres morfo-agronômicos e para tolerância à seca. Portanto, foi demonstrada a possibilidade de se selecionar uma coleção nuclear baseando-se somente em dados genético-moleculares, para se extrair uma sub-amostra que reflete a riqueza alélica e fenotípica da coleção base em um tamanho amostral mínimo.

**Fontes financiadoras:** CAPES, CNPq, EMBRAPA Macroprograma I