



Detecção e caracterização de genótipos de sorgo sob estresse hídrico usando PLS-DA, fusão de dados atômico-moleculares e seleção de variáveis

Camila Assis¹ (PG), Maria Lúcia F. Simeone² (PQ), Marco Aurélio G. Pimentel² (PQ), Marcelo M. Sena^{1*}

*marcsen2000@yahoo.com.br

¹Departamento de Química, ICEx, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brasil.

²Embrapa Milho e Sorgo, MG 424, Km 45, 35701-970 Sete Lagoas, MG, Brasil.

Palavras Chave: Sorgo, Fusão de Dados, Seleção de Variáveis.

Introdução

O sorgo, *Sorghum bicolor* (L.) Moench, é uma planta tropical considerada uma das culturas mais importantes da África, Ásia e América Latina. Com relação a composição nutricional, o sorgo é uma boa fonte de energia, proteínas, carboidratos, vitaminas e minerais. Nesse contexto, vários projetos de biofortificação têm surgido como uma alternativa para contribuir com a redução de deficiências minerais, principalmente de ferro e zinco, visando aumentar a densidade de nutrientes em culturas básicas, principalmente por meio de intervenção agrônômica e seleção genética. No Brasil, a Embrapa Milho e Sorgo conduz programas de melhoramento visando a seleção de genótipos de sorgo com melhor qualidade para consumo humano. O principal objetivo deste trabalho foi rastrear genótipos de sorgo quanto ao conteúdo mineral e molecular, investigando, o efeito do estresse hídrico na sua composição química, desenvolvendo e otimizando métodos analíticos baseados em NIRS e ICP-OES.

Resultados e Discussão

Para realização do trabalho, cem acessos de sorgo com alta variabilidade genética foram utilizados em dois ambientes: sem e com estresse hídrico. O objetivo foi avaliar o efeito do estresse hídrico no conteúdo mineral (usando ICP-OES) e molecular (usando NIRS). Os modelos de classificação (PLS-DA) foram construídos individualmente para cada técnica analítica e também com fusão de dados de baixo nível, a partir de 128 amostras

de treinamento e 64 de teste. Dois métodos de seleção de variáveis foram testados: GA e OPS. Os modelos de fusão de dados forneceram melhores resultados do que os modelos construídos com cada técnica individual. O modelo construído com todas as variáveis (1511) apresentou uma taxa de 1,6% de falso positivo e nenhum falso negativo no conjunto de treinamento, e nenhum erro de previsão no conjunto teste. Os modelos com seleção de variáveis não apresentaram nenhum erro de revisão, sendo que o OPS mostrou um melhor desempenho, pois selecionou menos variáveis (110, em comparação com 159 selecionadas pelo GA), com um gasto de tempo computacional bastante menor. O número de variáveis latentes em todos os modelos de fusão foi 7.

O OPS selecionou 3 elementos (Mn, Mg e P) e várias regiões espectrais relacionadas à água, a amido e a proteínas, permitindo uma interpretação atômico-molecular do modelo através da análise de seus vetores informativos (coeficientes de regressão e VIP scores). Dessa forma, foi possível associar elementos e moléculas ao efeito de estresse hídrico no sorgo.

Conclusões

A junção de NIR com informações do conteúdo mineral forneceu um modelo com excelente desempenho e uma ampla capacidade de interpretação. Em geral, elementos (e compostos) específicos presentes no sorgo foram fundamentais para classificar as amostras.