



Identificação de sementes de mandioca *waxy* com uso de espectrofotometria do infravermelho próximo

Úrsula Maria Gonzaga Santos de Oliveira¹, Juraci Souza Sampaio Filho², Massaine Bandeira e Sousa³ e Eder Jorge de Oliveira⁴

¹ Estudante de Biologia da Universidade Federal do Recôncavo da Bahia, estagiária da Embrapa Mandioca e Fruticultura, bolsista Fapesb, Cruz das Almas, BA; ² Engenheiro-agrônomo, doutorando do programa de pós graduação em ciências agrárias da Universidade Federal do Recôncavo da Bahia, bolsista Fapesb, Cruz das Almas, BA; ³ Bióloga, doutora em Genética e Melhoramento de Plantas, pós-doutoranda da Embrapa Mandioca e Fruticultura, Cruz das Almas, BA; ⁴ Engenheiro-agrônomo, doutor Genética e Melhoramento de Plantas, pesquisador da Embrapa Mandioca e Fruticultura, Cruz das Almas, BA.

Introdução: Novas propriedades do amido de mandioca têm sido demandadas pelas indústrias de alimentos. O amido *waxy* reduz a liberação de água da pasta durante o processo de retrogradação, evitando assim modificações na qualidade sensorial e no período de armazenamento dos alimentos. Metodologias rápidas para seleção indireta do gene *waxy* podem otimizar a seleção dos genótipos desejáveis e permitir uma redução no ciclo de melhoramento da mandioca. A espectrofotometria do infravermelho próximo (NIRS) é uma alternativa robusta e rápida para a identificação de genótipos contrastantes, considerando a correlação entre as propriedades de vibração diferencial das moléculas e suas implicações com a absorção de luz em determinadas frequências. Embora mudanças na molécula de DNA não sejam detectáveis, mudanças sutis na expressão dos produtos gênicos podem ser detectadas.

Objetivo: Avaliar o potencial da espectrofotometria NIRS para a identificação de genótipos com amido *waxy* em sementes de mandioca e validar modelos de classificação.

Material e Métodos: Foram obtidas 1.127 sementes F_2 oriundas de diferentes progênies de mandioca por meio de polinização controlada. As polinizações foram realizadas utilizando parentais com amido *waxy* (*wxwx*) vs não-*waxy* (*wild-type*, $Wx_$) e entre parentais *waxy* F_1 ($Wxwx$). Cada semente foi identificada individualmente e os espectros foram obtidos no espectrômetro NIRFlex N-500. Os espectros das amostras foram obtidos em quadruplicata, com wavelength variando de 1000–2500 nm. Após a coleta dos dados espectrais, as sementes foram semeadas em tubetes em casa de vegetação (mistura 1:1:2, areia, vermiculita e solo), com germinação de 100%. Passados 45 dias da germinação, os *seedlings* foram transplantados a campo, em sequeiro, no espaçamento de 0,90 x 0,60 m entre linhas e plantas, respectivamente. A colheita foi realizada aos oito meses após o plantio e realizada a avaliação através do teste de coloração de iodo a 2% (2 g KI e 0.2 g I_2 em água destilada), na seção transversal das raízes dos *seedlings*. A cor azul escura indica genótipo não-*waxy* e a cor marrom avermelhada indica genótipo *waxy*. Para a classificação dos clones foram analisados quatro modelos: *k-nearest neighbor algorithm* (KNN); *eXtreme Gradient Boosting* (xgbLinear); *C5.0 decision tree* (CDT); e *parallel random forest* (parRF). A acurácia da classificação e o índice Kappa foram utilizados como parâmetros para avaliar a eficiência dos modelos com melhor ajuste na classificação dos clones *waxy* e não-*waxy*, considerando validação cruzada (80% dos dados), 10 repetições com 5-folds cada e um conjunto de validação externa (20% dos dados).

Resultados: Cerca de 20% dos clones avaliados foram classificados como *waxy*. A acurácia dos modelos de classificação, com base na validação cruzada foi elevada e variou entre 0,86 (parRF) a 0,92 (xgbLinear). Já o índice Kappa, que indica o quanto os modelos conseguem classificar corretamente as classes analisadas, apresentou tendência bastante similar à acurácia, com variação de 0,39 (parRF) a 0,71 (xgbLinear). A matriz de confusão foi obtida com intuito de avaliar mais claramente os resultados da classificação no conjunto de validação externa. Os resultados mostram que os modelos são eficientes em classificar os clones não-*waxy*, com porcentagens de acerto variando entre 96 e 100%, enquanto que para a classe *waxy* a porcentagem de acerto variou entre 30% (parRF) e 63% (KNN).

Conclusão: O modelo de classificação xgbLinear apresentou alta capacidade de classificação e seleção precoce dos genótipos de mandioca para tipo de amido *waxy* já na fase de sementes utilizando os espectros NIRS.

Significado e impacto do trabalho: Métodos precoces para detecção de genótipos *waxy* de mandioca via análise de folhas e NIRS já foram reportados; entretanto, nossa pesquisa demonstrou a possibilidade de detecção do fenótipo *waxy* na fase de sementes, trazendo o processo de seleção para um patamar mais precoce do que a seleção em casa-de-vegetação.