



## Classificação de genótipos de mandioca quanto à presença de compostos cianogênicos nas raízes via espectroscopia na região do visível (VIS) e infravermelho próximo (NIR)

Vinícius Ribeiro de Souza Bispo<sup>1</sup>; Massaine Bandeira e Sousa<sup>2</sup>; Gilvanara Damasceno de Souza<sup>3</sup>; Ravena Rocha Bessa de Carvalho<sup>3</sup>; Luciana Alves de Oliveira<sup>4</sup>; Eder Jorge de Oliveira<sup>4</sup>

<sup>1</sup>Estudante de Licenciatura em Biologia da Universidade Federal do Recôncavo da Bahia, vini19902016@outlook.com; <sup>2</sup>Pós-doutora da Embrapa Mandioca e Fruticultura, massainebandeira@hotmail.com; <sup>3</sup>Mestranda da Universidade Federal do Recôncavo da Bahia, gilvanaradamascenoo@gmail.com; ravenarochabc@yahoo.com; <sup>4</sup>Pesquisador A da Embrapa Mandioca e Fruticultura, luciana.oliveira@embrapa.br; eder.oliveira@embrapa.br.

A busca por ferramentas de fenotipagem em larga escala que permitam avaliação precoce e acurada das características de importância agrônômica tem ganhado espaço em programas de melhoramento genético. A utilização da espectroscopia na região do visível (VIS) e infravermelho próximo (NIR) surge como alternativa robusta e rápida para avaliação de atributos de importância agrônômica cujas análises laboratoriais são de alto custo e de difícil mensuração. O VIS-NIR apresenta como principais vantagens, a elevada capacidade de obtenção dos dados analíticos, de forma não destrutiva ou invasiva, aliada à alta penetração do feixe de radiação para análise de diferentes componentes. Devido à alta confiabilidade dos resultados, o NIR tem sido utilizado para análises de atributos relacionados à qualidade da raiz de mandioca, como exemplo, o teor de compostos cianogênicos (TCC). O teor destes compostos é utilizado para classificar as variedades de mandioca em: mansas (baixos teores), bravas (teores elevados) e intermediárias, em algumas situações. O presente trabalho teve como objetivo avaliar modelos de classificação quanto ao TCC em genótipos de mandioca, utilizando espectros VIS-NIR. Para isso, o TCC foi determinado em 278 clones avaliados em delineamento de blocos casualizados, com três repetições. O TCC foi analisado com base no teste picrato, onde três a quatro amostras de diferentes raízes foram inseridas em um tubo de ensaio, com solução de tolueno, picrato alcalino, usado como reagente de cor, e papel filtro. Após 24 horas foi realizada a leitura da coloração do papel filtro, seguindo uma escala de cores com variação de 1 a 9. A média das notas quanto à TCC foi utilizada para a classificação em mansa (notas de 1 a 4), intermediária (nota igual a 5) e brava (nota > 6). As leituras dos espectros do VIS-NIR portátil ASD QualitySpec Trek (comprimento de ondas variando de 350 a 2500 nm) foi realizada em 12 amostras de raízes frescas (porção central da raiz com aproximadamente 1 cm de espessura) e 4 amostras de raízes processadas com auxílio de um processador de alimentos. Os dados, obtidos das leituras das amostras foram utilizados para análise dos seguintes modelos de classificação: *High Dimensional Discriminant Analysis* – HDDA, *Parallel random forest* – PRANDF, *Support vector machines with linear kernel* – SVM, *Least Squares Support Vector Machine* – LSSVM e *Penalized Discriminant Analysis* – PDA. Os algoritmos de classificação foram implementados no ambiente de programação R com auxílio do pacote *caret*. O poder de discriminação dos modelos na validação cruzada foi avaliado a partir do número de amostras corretamente classificadas (acurácia) e a concordância da classificação (índice Kappa). Os modelos de classificação SVM (acurácia de 0,68 – amostra fresca) e PRANDF (acurácia de 0,67 e 0,65 para amostras frescas e processadas) foram os mais acurados, embora tenham apresentado concordância Kappa de magnitude mediana (variando entre 0,45 a 0,49). Os modelos de classificação também foram avaliados pela probabilidade em acertar corretamente a classificação dos indivíduos em mandioca mansa, intermediária e brava (sensibilidade) e a probabilidade em errar a sua classificação (especificidade). Os modelos SVM e PRANDF, apresentam altos valores de sensibilidade (0,63 e 0,61, respectivamente) e especificidade (média de 0,82). Portanto, conclui-se que os modelos SVM e PRANDF foram os mais acurados ao classificar os genótipos de mandioca quanto ao TCC e que as leituras do VIS-NIR em amostras frescas mostraram-se mais eficiente quando comparado ao uso de amostras processadas.

**Significado e impacto do trabalho:** O uso da tecnologia do VIS-NIR está aumentando devido à praticidade na análise de elementos de difícil análise ou com alto custo operacional. A tecnologia VIS-NIR pode acelerar a execução de análises que convencionalmente exigiriam maior tempo e cuidados, a exemplo da quantificação do teor de compostos cianogênicos. No entanto, para o uso rotineiro do VIS-NIR, ainda será necessário aumentar o número de genótipos com informação do teor de compostos cianogênicos para incrementar o tamanho da população de treinamento e validar os modelos de classificação, a fim de garantir maior acurácia do processo.