

**UNIVERSIDADE FEDERAL DO RECÔNCAVO DA BAHIA
CENTRO DE CIÊNCIAS AGRÁRIAS AMBIENTAIS E BIOLÓGICAS
EMBRAPA MANDIOCA E FRUTICULTURA
PROGRAMA DE PÓS-GRADUAÇÃO EM RECURSOS GENÉTICOS VEGETAIS
CURSO DE MESTRADO**

**MÉTODO PRECOCE PARA DIAGNÓSTICO DA
DOENÇA “COURO DE SAPO” DA MANDIOCA POR
INFRAVERMELHO PRÓXIMO**

Emile Lemos Freitas

**CRUZ DAS ALMAS-BAHIA
2018**

MÉTODO PRECOCE PARA DIAGNÓSTICO DA DOENÇA “COURO DE SAPO” DA MANDIOCA POR INFRAVERMELHO PRÓXIMO

Emile Lemos Freitas

Bacharelado em Ciências Biológicas
Universidade Estadual de Feira de Santana, 2013.

Dissertação submetida ao Colegiado do Programa de Pós-Graduação em Recursos Genéticos Vegetais da Universidade Federal do Recôncavo da Bahia, como requisito parcial para obtenção do título de Mestre em Recursos Genéticos Vegetais.

Orientador: DSc. Eder Jorge de Oliveira
Coorientadores: DSc. Saulo Alves Santo de Oliveira
DSc. Ana Carla Brito

**CRUZ DAS ALMAS – BAHIA
2018**

FICHA CATALOGRÁFICA

F866m	<p data-bbox="574 1209 1203 1355">Freitas, Emile Lemos. Método precoce para diagnóstico da doença “couro de sapo” da mandioca por infravermelho próximo / Emile Lemos Freitas._ Cruz das Almas, BA, 2018.</p> <p data-bbox="622 1361 694 1400">57f.; il.</p> <p data-bbox="622 1433 1109 1467">Orientador: Eder Jorge de Oliveira.</p> <p data-bbox="622 1473 1109 1507">Coorientador: Saulo Alves Santo de Oliveira.</p> <p data-bbox="574 1556 1203 1668">Dissertação (Mestrado) – Universidade Federal do Recôncavo da Bahia, Centro de Ciências Agrárias, Ambientais e Biológicas.</p> <p data-bbox="909 1736 1061 1769">CDD: 535.333</p>
-------	---

Ficha elaborada pela Biblioteca Universitária de Cruz das Almas - UFRB. Responsável pela Elaboração – Antonio Marcos Sarmiento das Chagas (Bibliotecário - CRB5 / 1615). Os dados para catalogação foram enviados pela usuária via formulário eletrônico.

**UNIVERSIDADE FEDERAL DO RECÔNCAVO DA BAHIA
CENTRO DE CIÊNCIAS AGRÁRIAS AMBIENTAIS E BIOLÓGICAS
EMBRAPA MANDIOCA E FRUTICULTURA
PROGRAMA DE PÓS-GRADUAÇÃO EM RECURSOS GENÉTICOS VEGETAIS
CURSO DE MESTRADO**

**MÉTODO PRECOCE PARA DIAGNÓSTICO DA
DOENÇA “COURO DE SAPO” DA MANDIOCA POR
INFRAVERMELHO PRÓXIMO**

Comissão Examinadora da Defesa de Dissertação
Emile Lemos Freitas

Aprovada em: 02 de fevereiro de 2018

Prof. Dr. Eder Jorge de Oliveira
Embrapa Mandioca e Fruticultura
Orientador

Dr. Diego Fernando Marmolejo Cortes
Embrapa Mandioca e Fruticultura
Examinador Externo

Dr. Onildo Nunes de Jesus
Embrapa Mandioca e Fruticultura
Examinador Externo

MÉTODO PRECOCE PARA DIAGNÓSTICO DA DOENÇA “COURO DE SAPO” DA MANDIOCA POR INFRAVERMELHO PRÓXIMO

Autora: Emile Lemos Freitas

Orientador: Eder Jorge de Oliveira

Coorientadores: Saulo Alves Santo de Oliveira e Ana Carla Brito

RESUMO: O couro de sapo causa perdas significativas na produtividade de raízes de mandioca (*Manihot esculenta* Crantz) e de modo geral, a diagnose é realizada de forma visual ou com uso de técnicas moleculares, geralmente de forma tardia. O objetivo deste trabalho foi analisar a viabilidade da espectrometria do infravermelho próximo (NIR) para detecção precoce da doença com diferentes modelos de calibração. A análise do NIR foi realizada em 238 acessos de mandioca (120 sadios e 118 infectados pelo couro de sapo). Os modelos de classificação utilizados foram *Bayesian Generalized Linear Model* (BGLM), *Extreme Learning Machine* (ELM), *High Dimensional Discriminant Analysis* (HDDA), *Partial Least Squares* (PLS), *Parallel Random Forest* (PRANDF) e *Support Vector Machines with Linear Kernel* (SVM). A capacidade preditiva dos modelos foi avaliada pela acurácia e índice de Kappa obtidas por validação cruzada. Os modelos foram testados utilizando todos os comprimentos de onda e seleção de comprimentos de onda por eliminação recursiva de características (RFE) com uso do algoritmo *random forest*. Os modelos preditivos demonstraram alta eficiência na distinção de acessos sadios e infectados, com acurácia acima de 80%. Quatro modelos (SVM, BGLM, PRANDF, PLS) apresentaram nível de concordância quase perfeita com base no índice de Kappa. Os modelos SVM e BGLM apresentaram elevada acurácia (99,07 e 98,92%, respectivamente) e reprodutibilidade (0,98) na classificação dos acessos. A seleção de comprimentos de onda NIR possibilitou a identificação de faixas mais informativas para construção dos modelos, porém reduziu o desempenho na classificação de todos os modelos, à exceção do ELM. O uso do NIR é uma alternativa viável para a detecção do couro de sapo, com as vantagens inerentes da detecção precoce e precisa, com elevada rapidez e baixo custo em comparação aos métodos tradicionais de diagnose.

Palavras Chave: BGML; Espectrometria; Seleção de variáveis; SVM.

EARLY METHOD FOR DIAGNOSIS OF CASSAVA FROG SKIN DISEASE USING NEAR INFRARED

Author: Emile Lemos Freitas

Advisor: Eder Jorge de Oliveira

Co-advisor: Saulo Alves Santo de Oliveira and Ana Carla Brito

ABSTRACT: The frogskin disease causes significant yield losses in cassava roots (*Manihot esculenta* Crantz) and in general, the diagnosis is carried out visually or using molecular techniques, usually diagnosed in old plants. The objective of this work was to analyze the feasibility of near infrared spectrometry (NIR) for early detection of the disease by using different calibration models. A NIR analysis was performed on 238 cassava accessions (120 healthy and 118 frogskin-infected). Six classification models were used: bayesian generalized linear model (BGLM), extreme learning machine (ELM), high dimensional discriminant analysis (HDDA), partial least squares (PLS), parallel random forest (PRANDF) and support vector machines with linear kernel (SVM). The predictive ability of the models was evaluated by the accuracy and Kappa index obtained by cross validation. The models are tested using all wavelength and wavelength selection by the recursive feature elimination (RFE) method based on random forest algorithm. The predictive models demonstrated high efficiency to distinguish healthy from frogskin-infected accessions (accuracy > 80%). Four models (SVM, BGLM, PRANDF, PLS) yielded almost perfect classification based on the Kappa index. The SVM and BGLM models presented high accuracy (99.07 and 98.92%, respectively) and reproducibility (0.98) to classify the cassava accession based on the frogskin symptoms. The selection of NIR wavelengths allowed the identification of high informative spectrum for building the classification models, but it reduced the classification performance of all models, except the ELM. The use of NIR is a viable alternative for the frogskin detection with inherent advantages of early and accurate detection, with high accuracy and low time-consuming compared to traditional methods of diagnosis

Keywords: BGML; Variable selection; Spectrometry; SVM.

SUMÁRIO

INTRODUÇÃO GERAL	1
REFERENCIAL TEÓRICO	3
1. Aspectos gerais da cultura	3
2. Doenças na mandioca	5
3. Fitoplasmas	7
4. Couro de sapo da mandioca	9
5. Métodos de detecção de fitoplasmas	10
6. Infravermelho próximo (NIR- Near Infrared).....	11
6.1. Interpretação dos espectros	12
6.2 Aplicações do NIR na detecção de doenças	16
7. Referências.....	18
MATERIAL E MÉTODOS	28
RESULTADOS.....	31
DISCUSSÃO	38
CONCLUSÃO	44
REFERÊNCIAS.....	46
CONSIDERAÇÕES FINAIS	52

INTRODUÇÃO GERAL

A mandioca (*Manihot esculenta* Crantz) é a base alimentar de milhões de pessoas na África e América Latina, sobretudo em regiões com baixa renda. Cultivada em 105 países, a cultura se destaca como a terceira principal fonte de carboidratos nos trópicos, depois do milho e do arroz (FAO, 2017). Seu cultivo se constitui como um meio de subsistência para muitas famílias que, além do consumo, utilizam suas raízes como fonte de renda (ANYANWU et al., 2015).

Apesar de ser conhecida por sua capacidade de adaptação a condições desfavoráveis de clima e solo, a cultura apresenta diversas limitações de produção. As principais adversidades são associadas a fatores abióticos como deterioração fisiológica pós-colheita e períodos prolongados de seca, e relacionadas a fatores bióticos como ataque de pragas e doenças e pouca disponibilidade de variedades resistentes e com alto teor de amido (REDDY, 2015).

Recentemente, a doença conhecida como couro de sapo da mandioca tem causado grandes prejuízos aos agricultores, sobretudo do Nordeste brasileiro. Esta doença atinge principalmente as raízes, causando perdas significativas na produtividade (REDDY, 2015). A etiologia do couro de sapo foi associada a um vírus da família Reoviridae (dsRNA) (CALVERT et al., 2008). Contudo, estudos desenvolvidos na Colômbia apontaram que um fitoplasma pertencente ao grupo III também está relacionado ao couro de sapo (ALVAREZ et al., 2009). Fitoplasmas são organismos procariontes, sem parede celular e não possuem todas suas rotas metabólicas para compostos vitais, sendo parasitas obrigatórios que habitam o floema das plantas (KAKIZAWA; YONEDA, 2015).

Os sintomas são característicos nas raízes, que apresentam aspecto lenhoso, sulcos na casca e entrecasca. Além disso, a entrecasca se torna quebradiça, espessa e opaca. A parte aérea praticamente não apresenta sintomas visíveis, sendo muitas vezes até mais vigorosas que as demais plantas, o que podem acentuar ainda mais o problema, considerando que isso faz com que os agricultores escolham materiais de plantio contaminados para formação das novas lavouras (ALVAREZ et al., 2009; SOUZA et al., 2014).

A forma mais efetiva de controlar doenças causadas por fitoplasma é a prevenção, por meio do uso de plantio de material vegetal sadio ou variedades

resistentes às doenças, controle de insetos vetores, e aplicação de práticas culturais para eliminar as fontes de fitoplasmas (LEE et al., 2000). No entanto, para o couro de sapo da mandioca os vetores de transmissão ou fontes de resistência não são conhecidos, sendo assim, o controle depende do diagnóstico das plantas infectadas.

Além da observação dos sintomas, as técnicas de diagnose do couro de sapo incluem teste ELISA para vírus e métodos moleculares. Estes últimos métodos são mais precisos que o primeiro e são realizados por PCR (*polymerase chain reaction*) tanto para diagnose de vírus quanto fitoplasmas (ANGEL et al., 1987; ALVAVEZ et al., 2009; SOUZA et al., 2014; CALVERT et al., 2008). No entanto, independente da técnica, deve-se considerar os custos elevados associados ao tempo de produção de anticorpos, como o ELISA, além dos custos de mão de obra, equipamentos e infraestrutura, bem como o tempo necessário na diagnose com base no DNA dos fitoplasmas ou RNA dos vírus (SOUZA et al., 2014).

Nos últimos anos, imagens hiper-espectrais têm sido utilizadas como uma alternativa para a detecção precoce de doenças, a exemplo da incidência de infecções causadas por *Fusarium* em trigo (BAURIEGEL et al., 2011; ROPELEWSKA; ZAPOTOCZNY, 2018). O uso da espectrometria por infravermelho próximo (*near infrared* – NIR) é um método indireto que tem a principal vantagem de não ser invasivo, com potencial de detecção de infecções em estágio inicial da doença. As análises são feitas sem necessidade de tratamento prévio das amostras e sem gerar resíduos, gerando espectros de refletância difusa. O deslocamento e alteração da intensidade das bandas de absorção podem gerar padrões específicos capazes de diagnosticar precocemente doenças e fornecer uma elevada precisão na classificação de diversos estresses (PASQUINI, 2003).

Nesse contexto, o objetivo do presente estudo foi avaliar a possibilidade de implementação do diagnóstico precoce do couro de sapo em acessos de mandioca oriundos do Banco Ativo de Germoplasma da Embrapa Mandioca e Fruticultura, por meio da espectrometria NIR, utilizando diferentes modelos de calibração e classificação.

REFERENCIAL TEÓRICO

1. Aspectos gerais da cultura

A mandioca (*Manihot esculenta* Crantz) é uma espécie nativa da América do Sul que pertence ao gênero *Manihot* da família Euphorbiaceae. Esse gênero possui 98 espécies, distribuídas desde o Arizona nos Estados Unidos, até a Argentina (CHACÓN et al., 2008). A América Central e o México representam um dos centros de distribuição com 17 espécies, já a América do Sul abriga cerca de 80% das espécies. Destas, 76 ocorrem no Brasil e 66 são nativas, com provável centro de origem e domesticação na região central do Brasil (OLSEN, 2004; CORDEIRO et al., 2015; CARRASCO et al., 2016).

Atualmente, a mandioca é cultivada em 103 países, destes 41 estão na África, que concentra mais da metade da produção mundial, e tem seu cultivo destinado principalmente para alimentação. A Ásia é o segundo maior produtor, cujos usos são para alimentação animal e para fins industriais com crescente utilização para geração de biocombustíveis. Na América, centro de diversidade, a mandioca é cultivada em 35 países, sendo o Brasil o maior produtor (BULL et al., 2011; FAO, 2017).

O Brasil ocupa o terceiro lugar na produção mundial de mandioca. Em nível mundial a produção de mandioca em 2016, foi de 21,08 milhões de toneladas, sendo 23,25 milhões de toneladas produzidas no Brasil. O país ficou atrás apenas da Nigéria (57,13 milhões de toneladas) e Tailândia (31,16 milhões de toneladas) (FAO, 2017).

A mandioca é uma das principais culturas permanentes no Brasil, em valor de produção agrícola, com importante contribuição para o consumo humano, alimentação animal e fins industriais. Cultivada de norte a sul, em todos os estados, a estimativa de produção nacional para o ano de 2017 foi de 20,8 milhões de toneladas. A região Norte é a maior produtora de mandioca com 39,5% da produção nacional, seguida do Nordeste e Sudeste com 23,6% e 20,6%, respectivamente. O Sul e Centro-Oeste contribuem juntos com 16,3%, sendo respectivamente 10,5 e 5,8% para cada região (IBGE, 2017).

No Nordeste brasileiro, a cultura possui importante contribuição para o consumo humano, alimentação animal e como matéria prima para inúmeros produtos, sendo muitas vezes fonte de subsistência e renda para os produtores. No Sul, com o destaque para o estado do Paraná, a maior produção é de amido para alimentação e uso industrial (HENRY; HERSHEY, 2002).

A raiz, rica em carboidratos, é a parte mais utilizada da mandioca na alimentação humana, que se dá a partir do cozimento das raízes e em formas derivadas, como a farinha e amido, e na produção de bolos, biscoitos e pães de queijo (HENRY; HERSHEY, 2002). A parte aérea também é usada para consumo humano como uma fonte rica em proteínas, minerais, vitaminas B1, B2 e C, e carotenoides (REDDY, 2015; AGRE et al., 2017). O consumo de folhas acontece principalmente em regiões africanas com escassez de alimentos (AGRE et al., 2017) e em pratos típicos no Norte e Nordeste, como a maniçoba.

Na alimentação animal, todas as partes da mandioca podem ser utilizadas. As raízes e a parte aérea são utilizadas *in natura* ou processadas para melhorar a conservação e ainda ajudar na redução dos compostos cianogênicos. Esses procedimentos podem ser simples, como trituração ou secagem, ou mais complexos como a fabricação de forragem, farinha, rações, pastilhas e silagem (RAVINDRAN, 1993; HENDERSON; 2015).

O uso industrial das raízes visa a exploração de dois principais produtos: a farinha e o amido. De modo geral a produção de farinha envolve o processamento em pequenas instalações que utilizam técnicas tradicionais, especialmente na África e no Nordeste do Brasil. Já para a produção do amido, o processamento geralmente é feito em instalações maiores e industriais (SÁNCHEZ et al., 2017). A produção de amido é destinada para a indústria alimentar, biocombustíveis, produtos farmacêuticos e têxteis, e outros produtos como adesivos, papéis e xarope de glicose (TONUKARI, 2004; ZHANG et al., 2016; AGRE et al., 2017; SÁNCHEZ et al., 2017).

O amido de mandioca tem grande expansão e absorção de água e suas pastas são relativamente macias, com maior viscosidade, melhor clareza e menores taxas de retrogradação do que amidos de cereais. Além disso, o amido ceroso, livre de amilose, também é encontrado na mandioca, resultado de mutação natural. Comparado com o normal, composto de 83% de amilopectina e 17% de amilose, o amido ceroso, que não possui amilose, aumenta a clareza,

estabilidade de géis e a temperatura de fusão dos grânulos. Entre os cerosos, o amido de mandioca é o único que não possui sinérese em temperaturas de – 20°C, o que demonstra a possibilidade de um grande uso industrial (SÁNCHEZ et al., 2010; FERRARO et al., 2016).

Apesar de sua importância econômica, o rendimento mundial de mandioca é de 11,24 t/ha (FAO, 2017), muito abaixo do seu potencial de 80 t/ha (HOWELER et al., 2013). Os principais problemas no cultivo que reduzem seu rendimento são deterioração fisiológica pós-colheita, ataque de pragas e doenças, e a falta de variedades resistentes e com alto teor de amido (REDDY, 2015). Na África, o baixo rendimento está associado, principalmente a fatores bióticos como doenças virais que afetam severamente a colheita (CHAVARRIAGA-AGUIRRE et al., 2016).

2. Doenças na mandioca

Pragas e doenças são responsáveis por uma redução na produção e na qualidade das safras. No caso da mandioca, apesar da ocorrência natural de sementes, a propagação é comumente feita por hastes, pois as plantas oriundas de sementes se estabelecem mais lentamente, são menos vigorosas e apresentam segregação genética devido à polinização cruzada (CALVERT; THRESH, 2002). No entanto, na propagação vegetativa, muitos patógenos endógenos e sistêmicos são transmitidos através do material de propagação usado em novos cultivos (HOKANSON et al., 2016; THOMAS-SHARMA, 2017).

A ocorrência de doenças é um fator limitante da cultura e ameaça a segurança alimentar. Infecções podem causar perdas de até 100% na produção das raízes, além de causar sintomas nas folhas e hastes que comprometem a qualidade. As folhas, que também participam da nutrição humana e alimentação animal, geralmente é a parte mais afetada por pragas e patógenos (REDDY, 2015; MCCALLUM; ANJANAPPA; GRUISSEM, 2017).

Os patógenos que atingem a mandioca têm etiologia associada a vírus, bactérias, fungos e fitoplasmas. Entre as doenças que esses organismos podem causar, a bacteriose causada por *Xanthomonas axonopodis* pv. *manihotis* possui maior ocorrência mundial. A doença se dissemina por hastes infectadas e pode se dispersar através de gotas de chuva e por ferramentas contaminadas (ALVAREZ et al., 2012). As perdas causadas por essa bactéria podem atingir 100% da

produção e os sintomas acometem tecidos foliares, hastes e pecíolos, desde pequenas manchas de aspecto aquoso até a necrose dos tecidos (CHEGE et al., 2017; MCCALLUM; ANJANAPPA; GRUISSEM, 2017).

As viroses também têm grande importância como componente de redução da produtividade da mandioca. Duas viroses transmitidas por insetos e hastes infectadas, causam grande preocupação no continente africano: o mosaico africano (*African cassava mosaic virus, Begomovirus*) e doença do castanho listrado da mandioca (*Cassava brown streak virus, Ipomovirus*) (LEGG; THRESH, 2003; LEGG, 2015). O castanho listrado da mandioca atinge as folhas, causando necrose e, em casos mais severos também pode ocorrer necrose nas hastes e perda de qualidade e quantidade de raiz, que pode ocasionar de 20 a 100% de perda da produção (KAWEESE et al., 2014; MCCALLUM; ANJANAPPA; GRUISSEM, 2017). Já os sintomas do mosaico africano ocorrem principalmente em folhas, que apresentam um mosaico amarelo e clorótico e deformação ligeiramente severa. Essa doença ocorre frequentemente na África e Ásia, e causa perdas de 30 a 80% (REDDY, 2015; MCCALLUM; ANJANAPPA; GRUISSEM, 2017). Na América, o mosaico comum (*Cassava common mosaic virus, Potexvirus*) e mosaico das nervuras (*Cassava vein mosaic virus, Cavemovirus*) possuem maior ocorrência. Outros vírus também encontrados pertencem às famílias Alphaflexiviridae, Reoviridae, Secoviridae, Caulimoviridae e Luteoviridae (CARVAJAL-YEPES et al., 2014).

Entre as doenças fúngicas, a podridão radicular (*Botryosphaeriaceae, Phytophthora sp.* e *Fusarium sp.*) causam danos que podem ocasionar perdas de até 80% de produção (BOAS et al., 2017). As maiores dificuldades para controle destas doenças estão no fato de que os patógenos habitam o solo e, assim, persistem por longos períodos. Os sintomas variam de acordo com o agente causal e atingem principalmente as raízes, sendo muitas vezes detectados apenas durante a colheita (REDDY, 2015). Outra doença causada por fungo, muito comum é antracnose (*Colletotrichum gloeosporioides*). Os sintomas da doença aparecem em folhas que apresentam manchas e desfolhação, e no caule, com cancrios profundos e corpos de frutificações visíveis (ALVAREZ et al., 2012).

Os fitoplasmas são responsáveis por duas doenças de grande impacto na produção de mandioca, o couro de sapo e o superbrotamento. Essas infecções são causadas por diferentes fitoplasmas que pertencem ao mesmo grupo 16SrIII,

(ALVAREZ et al., 2009; FLÔRES et al., 2013; SOUZA et al, 2014) classificados por similaridade das sequências de genes do rRNA 16S, diferindo apenas em seus subgrupos designados por letras que são inseridas após a nomenclatura dos grupos (PÉREZ-LÓPEZ et al, 2016). O couro de sapo (16SrIII-A e 16SrIII-L) está distribuído na América Central e Sul, com sintomas que atingem severamente as raízes com perdas na produção que podem ser totais. O superbrotamento (grupo 16SrIII-B) possui uma distribuição mais ampla e está presente nas Américas e na África. O principal sintoma refere-se à produção de brotações excessivas e atrofiadas, com entrenós encurtados e folhas pequenas. A grande produção da parte aérea consome grande parte das suas reservas de energia e ocasiona uma queda na produtividade de raízes, com perdas de produção de até 90% dos plantios (ALVAREZ et al., 2009; FLÔRES et al., 2013; REDDY, 2015). Outras doenças de ocorrência em algumas Ilhas do Pacífico com sintomas de brotações axilares, amarelamento e folhas pequenas estão sendo associadas a fitoplasmas (grupo 16SrI e 16SrII) (DAVIS et al., 2005). Na Costa do Marfim plantas com sintomas de amarelecimento e ondulação, e mosaico foliar foram associadas a fitoplasmas do grupo 16SrXXII-B (KRA et al., 2017).

3. Fitoplasmas

Os fitoplasmas são conhecidos como organismos do tipo micoplasmas (MLOs), cujo cultivo em meio axênico só foi possível recentemente e com grande dificuldade (CONTALDO et al., 2012). Na classificação atual os fitoplasmas pertencem ao Reino Monera, Domínio Bacteria, Filo Tenericutes, Classe Mollicutes, Ordem Acholeplasmatales, Família Acholeplasmataceae e Gênero "*Candidatus Phytoplasma*" (ZHAO et al., 2015). As espécies são classificadas de acordo com a análise filogenética por similaridade das sequências de genes oriundos do rRNA 16S. Existem 30 grupos de fitoplasmas propostos, nomeados 16SrI-16SrXXX e seus subgrupos designados por letras que são inseridas após a nomenclatura dos grupos (PÉREZ-LÓPEZ et al, 2016).

A primeira constatação da existência dos fitoplasmas ocorreu em 1967 por meio de microscopia eletrônica de transmissão. Foram descobertos organismos pleomórficos nos tubos crivados e em células do floema das plantas com infecções associadas a viroses. A etiologia dessas doenças era atribuída aos vírus, apesar destes não serem observados, devido à semelhança nos sintomas e

vias de transmissão. Esse novo grupo de agentes infecciosos possuíam características que os aproximavam das bactérias gram-positivas e assemelhava-se aos micoplasmas, já observados em animais (DOI et al., 1967; LEE et al., 2000).

Os fitoplasmas possuem de 0,2 a 0,8 μm de diâmetro, uma única membrana, sem parede celular e um cromossomo circular de cadeia dupla e pode conter até quatro plasmídeos diferentes. O seu genoma é o menor entre as bactérias patogênicas, contendo de 680 a 1,600 kb, com baixa porcentagem de citosina/guanina (C+G) e a ausência de algumas rotas metabólicas relacionada à sua evolução parasitária. Apesar disso, possuem mecanismos de sobrevivência em organismos tão diversos quanto o floema vegetal e a hemolinfa de insetos (BAI et al., 2009; BERTACCINI e DUDUK, 2010; KAKIZAWA e YONEDA, 2015).

Diversas alterações metabólicas são associadas à infecção por fitoplasma em plantas. Foram relatadas alterações nos níveis hormonais, redução da taxa de transporte de aminoácidos e de fotossíntese, além da concentração de seus pigmentos e nutrientes, não só restritos aos açúcares. Os sintomas nas plantas são variados, sendo que a resposta a um fitoplasma pode ser distinta em diferentes espécies vegetais ou a mesma sintomatologia pode estar associada a vários patógenos. As manifestações mais comuns são amarelecimento; proliferação de brotos por quebra de dominância apical, incluindo o crescimento das vassouras de bruxa; alongamento de entrenós e de outras estruturas; alterações florais como esterilidade, virescência e filodia (BERTACCINI e DUDUK, 2010; BERTACCINI et al., 2014; KAKIZAWA e YONEDA, 2015).

Os fitoplasmas atingem várias culturas comerciais com danos econômicos significativos e estão associados a doenças em centenas de plantas. Entre estas estão cultivos para alimentação, em arbóreas para sombra e uso de madeira e em plantas ornamentais (BERTACCINI et al., 2014). Entre as espécies comerciais estão a uva (ZAMBON et al., 2018), maçã (RID et al., 2016), morango (PÉREZ-LÓPEZ; DUMONCEAUX, 2016), soja (ACOSTA et al., 2015), mandioca (ALVAREZ et al., 2009; FLÔRES et al., 2013) coco (HARRISON et al., 2014), tomate (BUXA et al., 2015), batata (CHENG et al., 2015) e espécies de citrus (ALIZADEH, et al., 2017).

4. Couro de sapo da mandioca

A infecção por fitoplasmas, denominada popularmente como couro de sapo foi primeiramente relatada em mandioca na Colômbia em 1971. Entretanto, hoje está difundida no Brasil, Paraguai, Panamá, Venezuela, Costa Rica, Peru e na Argentina. No Brasil, o couro de sapo tem sido relatado nos estados do Amazonas, Pará, Bahia, Minas Gerais e Rio de Janeiro (MEISSNER FILHO; VELAME, 2006; ALVAREZ et al., 2007; CARDOZO TELLES et al., 2016; FEO et al., 2015). Apesar de não ter sido relatada para a Ásia e África, o couro de sapo é uma importante doença que atinge as Américas. As perdas atingem de 50 a 100% na produtividade de raízes (REDDY, 2015). Na Colômbia a doença já se expandiu por todas as áreas de cultivo e as perdas chegam a 90% na produtividade de raízes comerciais (ALVAREZ et al., 2007).

A etiologia da doença ainda não foi totalmente elucidada. O couro de sapo foi classificado como uma doença viral e atribuída à família Reoviridae (dsRNA), como demonstrado nos estudos de Calvert et al., (2008), que detectaram a presença dos vírus em plantas doentes. No entanto, Alvarez et al. (2009) identificaram na Colômbia um fitoplasma pertencente ao grupo III-L relacionado ao couro de sapo. No Brasil, a doença foi relacionada a dois subgrupos: grupo III-L (OLIVEIRA et al., 2014) e grupo III-A (SOUZA et al., 2014). Além dos fitoplasmas, a presença de um possível Reovírus foi também detectada demonstrando uma infecção pelos dois agentes. No entanto, apenas os fitoplasmas estavam presentes em todas as plantas sintomáticas (SOUZA et al., 2014). Adicionalmente, Carvajal-Yepes et al., (2014) descreveram um complexo viral associado à severidade do couro de sapo, fazendo com que a etiologia da doença seja bastante complexa e de difícil diagnose. Os três vírus descritos por Carvajal-Yepes et al. (2014), pertencentes às famílias Secoviridae (*Torradovirus*), Alphaflexiviridae e Luteoviridae, estão associados às manifestações incomuns de sintomas foliares. Entre eles, apenas o *Torradovirus* poderia isoladamente induzir experimentalmente os sintomas em folhas. No Brasil, apenas o Reovírus foi relatado em infecções por couro de sapo (SOUZA et al., 2014).

Os sintomas do couro de sapo atingem predominantemente as raízes das plantas, sendo incomuns na parte aérea. As raízes tornam-se fibrosas, com sulcos profundos e diâmetro reduzido. Já a casca apresenta um espessamento, opacidade e torna-se quebradiça. Dependendo da severidade da doença, as

raízes podem apresentar variações na coloração dos tecidos esquemático e parenquimático saindo do branco para creme, rosa ou marrom (ALVAREZ et al., 2009, SOUZA et al., 2014). Em contraste, com o pouco desenvolvimento e tuberização das raízes, as hastes costumam ser mais espessas, com parte aérea, muitas vezes mais desenvolvida e vistosa, favorecendo a seleção de material de plantio infectado (ALVAREZ et al., 2009). Sendo assim, a gravidade e incidência das raízes infectadas tendem a aumentar entre os ciclos de plantios.

Desse modo, para o controle da doença, é necessário que a parte aérea seja colhida junto com as raízes e o material vegetal de plantio seja selecionado de plantas saudáveis ou de variedades resistentes. As hastes de plantas infectadas devem ser isoladas e destruídas. Em relação ao uso de variedades resistentes ao couro de sapo como medida de controle, pouco progresso tem sido obtido, sobretudo no Brasil. Na Colômbia, apenas foram encontradas plantas com diferentes níveis de tolerância (ALVAREZ et al. 2012), mas ainda longe da imunidade necessária para o efetivo controle da doença.

Entre os mecanismos de transmissão naturais, a propagação das hastes é a forma mais conhecida de transmissão do couro de sapo da mandioca. No entanto, acredita-se que existam insetos vetores e transmissão por sementes. A partir de sementes, apesar de improvável, Calvert e Thresh (2002) já relatou sua ocorrência. Já a transmissão por meio dos insetos, embora não seja conhecido para fitoplasma associado ao couro de sapo (REDDY, 2015), sua transmissão já foi relatada para o Reovirus. As espécies de mosca branca relacionadas à transmissão foram *Bemisia tuberculata*, *Trialeurodes variabilis* e *Aleurotrachelus sociales* (ANGEL et al., 1987; MEISSNER FILHO; VELAME, 2006).

5. Métodos de detecção de fitoplasmas

A detecção de fitoplasmas pode ser realizada por diferentes métodos, que variam quanto à sensibilidade e precisão. Métodos biológicos e observações dos sintomas são importantes como diagnóstico preliminar de fitopatologias, porém podem induzir a erros. Outros métodos de detecção como a microscopia e a sorologia são ainda pouco eficazes. Técnicas baseadas na análise do DNA são formas mais sensíveis, rápidas e capazes de identificar especificamente o patógeno causador da doença (NEJAT; VADAMALI, 2013).

Atualmente, as técnicas mais utilizadas de diagnose de fitoplasmas são baseadas na reação em cadeia da polimerase (PCR). Os iniciadores universais específicos abrangem sequências altamente conservadas, como as regiões 16S e 23S do rDNA e, a região intermediária 16S-23S. A PCR para detecção de fitoplasmas é realizada em duas etapas, na primeira é utilizada iniciadores universais, seguido de outra PCR com iniciadores específicos. Um fragmento de aproximadamente 1,8 kb é amplificado pelos iniciadores universais P1/P7, e os iniciadores específicos R16F2n/R16R2, amplificam uma sequência de aproximadamente 1,2 Kb. Este último fragmento é utilizado para classificação. O chamado *Nested PCR* com R16mF2/mR1 como iniciadores externos e R16F2n/R16R2 como iniciadores internos também são usados para amplificação de fragmentos de 1,2 kb. Outras combinações de iniciadores também estão disponíveis para essa técnica (GUNDERSEN; LEE, 1996; ZHAO et al., 2010; NEJAT; VADAMALI, 2013).

As análises de polimorfismo no comprimento do fragmento de restrição (RFLP) são utilizadas para a classificação de fitoplasmas por meio de enzimas de restrição *in silico*. As 17 enzimas de restrição são: *AluI*, *BamHI*, *BfaI*, *Bsh1236* (*BstUI*), *DraI*, *EcoRI*, *HaeIII*, *HhaI*, *HinfI*, *HpaI*, *HpaII*, *KpnI*, *MboI*, *MseI*, *RsaI*, *SspI* e *TaqI*. As sequências obtidas são comparadas com os padrões descritos na literatura para identificar o grupo ao qual esse fitoplasma pertence. Para cada par de isolados, os coeficientes de similaridade podem ser calculados com uso do programa iPhyClassifier (SOUZA et al. 2014).

Uma nova técnica baseada em DNA, denominada de amplificação isotérmica mediada por loop (LAMP) surge como um método de diagnose de patógenos baseado na amplificação específica de sequências de ácidos nucleicos alvos sem ciclagem térmica. As reações em curto espaço de tempo e a diversidade de detecção permitem seu uso em campo e em laboratórios com infraestrutura relativamente limitada (SOUZA et al., 2014; DICKINSON, 2015; KOGOVSĚK et al, 2015). Ademais, técnicas espectroscópicas de ressonância magnética nuclear também têm sido utilizadas (CHOI et al, 2004).

6. Infravermelho próximo (NIR- Near Infrared)

A faixa do infravermelho posiciona-se entre a luz visível e as micro-ondas e compreende comprimentos de ondas de 750 a 1.000.000 nm e está dividida em

três regiões: próximo, médio e distante. A faixa posterior à região visível, em termos de comprimento de onda, correspondente ao infravermelho próximo (NIR), intervalo de 750 a 2.500 nm. O infravermelho médio localiza-se de 2.500 a 50.000 nm e o infravermelho distante, onde as ondas são as mais longas, de 50.000 a 1.000.000 nm (PASQUINI, 2003).

Cada região do infravermelho possui características de emissão de luz particular. As faixas do distante e médio estão intimamente relacionadas aos movimentos de rotação e vibrações das moléculas, respectivamente, enquanto que a faixa do infravermelho próximo se caracteriza por uma combinação de bandas harmônicas. As bandas harmônicas são resultantes da excitação do estado estacionário de energia para um estado mais elevado, constituindo-se de múltiplos de uma dada frequência fundamental. Assim, a região do NIR permite identificar principalmente compostos com ligações N-H, C-H e O-H. A técnica fundamenta-se na energia infravermelha de vibração, em que se monitoram as vibrações associadas diretamente com alterações na estrutura molecular. As moléculas ativadas pela região do infravermelho estabelecem modelos mecânicos com vibração dipolo que possuem frequência e amplitude específica (WEHRENS, 2011).

Os modos de medição dessas interações são por meio de transmitância, transfectância, reflectância difusa e interatância. A transmissão e transfectância são comumente empregadas para líquidos, enquanto que a reflectância difusa é principalmente usada para análise de sólidos, mas também pode ser usado para granulados e outros sólidos, géis e amostras pastosas (PASQUINI, 2003; WEHRENS, 2011). A partir dessas características físico-químicas, uma análise óptica pode ser realizada com as interações das radiações eletromagnéticas com elementos ou moléculas do material em análise. O resultado dessa interação é traduzido graficamente em um espectro em que cada faixa correspondente às suas características e composição (PASQUINI, 2003). No entanto, quanto mais complexa é a amostra, mais fácil encontrar regiões do espectro de NIR com picos de vibrações e regiões sobrepostas (TÜRKER-KAYA; HUCK, 2017).

6.1. Interpretação dos espectros

Os espectros NIR possuem uma grande quantidade de informações que podem ser utilizados para avaliar as amostras. No entanto, a complexidade

dessas informações, com presença de bandas amplas e sobrepostas, não permite uma análise direta dos dados (PASQUINI, 2003). No entanto, a calibração multivariada contorna esse problema com a aplicação de uma análise indireta, por meio de correlações entre os dados espectrais e suas propriedades. Essa análise é realizada por modelos matemáticos utilizando uma matriz, em que cada linha corresponde a uma amostra e as colunas correspondem às características da amostra, sendo que cada valor é uma variável (PASQUINI, 2003; VARMUZA; FILZMOSER, 2009).

6.1.1 Pré-processamento dos dados

Para a construção de modelos, a adequação dos dados é tão importante quanto a escolha do método de calibração. Os espectros contêm naturalmente ruídos, que são flutuações aleatórias e discretas no espectro, como erros de medição, aumento de picos e dispersão da linha. Nesse sentido, os dados podem ser melhorados por um pré-tratamento apropriado (WEHRENS, 2011).

Alguns pré-tratamentos visam à redução de problemas relacionados ao espalhamento de radiação por um sólido, que modificam a distribuição espacial das variáveis na matriz, grande número de variáveis não correlacionadas com a propriedade estudada e ruídos gerados na análise. Esses métodos de pré-processamento são realizados antes da construção dos modelos e baseiam-se em conceitos matemáticos, na dinâmica químico-física ou distúrbio dos dados. Os mais utilizados em dados espectroscópicos são: centralização na média, escalonamento de variância, variação de padrão normal e derivadas (PASQUINI, 2003; VARMUZA; FILZMOSER, 2009; WEHRENS, 2011).

Um dos pré-processamentos consiste na remoção da linha de base dos espectros, que influencia na altura e na área do pico. Essas alterações são causadas por efeitos de dispersão da superfície da amostra. Com essa técnica, a constante passa a ser igual a zero, pois os efeitos aditivos são removidos, baseando-se nas diferenças entre as intensidades nos comprimentos de onda sequenciais, em vez das intensidades entre si. Com a primeira derivada, além das correções de linha de base, as intensidades de bandas são acentuadas, porém os ruídos também são acentuados. Para corrigir esses ruídos é necessária a utilização de alguns filtros de suavização, entre esses métodos o mais comum é o

Savitzky-Golay, que é uma regressão polinomial (VARMUZA; FILZMOSER, 2009; WEHRENS, 2011).

Outra forma de tratar os dados é aplicar correções na matriz. A transformação por *Standard Normal Variate* (SNV) trata cada espectro separadamente por meio da autenticação dos valores de reflexão nas linhas, ou seja, nas amostras. Por esse método todo o espectro tem média igual a zero e desvio padrão igual a um, eliminando do conjunto de dados de valores arbitrários e fatores de multiplicação. Após as transformações dos dados, geralmente a centralização e o bloco de escala são aplicados com o objetivo de que todas as colunas sejam centradas com média zero e possuam a mesma escala de variância (VARMUZA; FILZMOSER, 2009; WEHRENS, 2011).

A seleção de variáveis é um dos pré-processamentos que tem se desenvolvido muito nos últimos anos. Sem a seleção, os modelos lineares podem ser penalizados com variáveis não informativas que podem prejudicar a avaliação entre as distâncias das amostras. Já no caso dos não-lineares existe o risco de correlação casual com variáveis irrelevantes e redução da predição dos modelos (WEHRENS, 2011). Seu objetivo é identificar e selecionar variáveis relevantes de um grande conjunto de dados multivariados e eliminar as irrelevantes. Esse procedimento geralmente é aplicado para melhorar os modelos e a sua interpretação ou mesmo para determinar um conjunto suficiente de predição (PASQUINI, 2003).

Essa etapa tem sido crucial na construção de modelos de classificação, por simplificar problemas estatísticos, acelerar o processamento de dados e ajudar no reconhecimento de classes. Um dos métodos de seleção de variáveis mais sofisticados é baseado no método chamado de *Randon Forest*. Este método é importante sobretudo quando há um pequeno grupo de dados, pois considera o efeito da eliminação ou inclusão de uma variável no desempenho do modelo. Portanto, o limite do número de variáveis excluídas é calculado de acordo com a diminuição de classificações erradas, observada em cada inclusão de uma variável (GENUER; POGGI; TULEAU-MALOT, 2015).

6.1.2 Métodos de classificação

A análise qualitativa é realizada por modelos de classificação, em que os dados são distribuídos em duas ou mais classes. Nesse caso, deve-se saber a

qual classe as amostras pertencem e a tarefa dos modelos é descrever os grupos a partir dos dados multivariados. Essa estrutura deve ser capaz de classificar novos dados e ser o mais confiável possível, obtendo o menor número de amostras erroneamente classificadas (VARMUZA; FILZMOSER, 2009).

Há uma grande variedade de sistemas de classificação de dados multivariados disponível. Os mais utilizados na quimiometria, em que há mais variáveis do que amostras, com elevada correlação são: *Linear Discriminant Analysis* (LDA) com suas variações, *Partial Least Squares* (PLS) em várias versões e *Support Vector Machine* (SVMs) (VARMUZA; FILZMOSER, 2009).

Os modelos são escolhidos de acordo com os objetivos da pesquisa e podem ser lineares ou não lineares. Os lineares são elaborados com equações relativamente mais simples, em que as amostras podem ser separadas por uma superfície de decisão em dois grupos, como acontece com o LDA e o PLS. Os não-lineares são elaborados com uma estrutura algorítmica menos evidente, como as *Classification trees*, SVMs ou redes neurais. A comparação entre diferentes modelos é realizada pela estimativa de medidas de desempenho para classificação, como taxas de classificação errônea (% erros) ou habilidades preditivas (% correta) para novos casos na validação (VARMUZA; FILZMOSER, 2009; WEHRENS, 2011).

6.1.3 Validação

A validação é a avaliação da qualidade de um modelo preditivo, a capacidade de realizar previsões precisas ou corretas da realidade. No entanto, em sistemas biológicos é um processo difícil, pois é frequente a existência de muitas variações significativas que não são relacionadas com os modelos. Por isso, quanto maior for essas variações, maior deve ser o número de amostras na validação para se obter estimativas confiáveis. Outro aspecto a ser avaliado é a representatividade dos modelos para o sistema de estudo, pois havendo diferença entre os dados utilizados para predição e os dados para os quais as previsões são necessárias, essas previsões serão abaixo do esperado e, em alguns casos, até mesmo errôneas (VARMUZA; FILZMOSER, 2009; WEHRENS, 2011).

Quando o estudo é realizado com um número elevado de amostras a validação mais usada é a validação externa, que é feita com a divisão em um conjunto de treinamento e um conjunto de testes. Para isso, é importante que os

dados sejam independentes e que essa divisão não seja realizada de forma tendenciosa. Assim as divisões aleatórias podem ser consideradas. Nesse tipo de validação, o primeiro grupo de amostras é utilizado para a construção do modelo e o segundo na avaliação desse modelo (WEHRENS, 2011).

Outro tipo de avaliação utilizada é a validação cruzada. Essa avaliação tem o mesmo princípio da avaliação externa, cujo conjunto de dados é utilizado para a construção dos modelos e o outro serve para avaliar a precisão do modelo em estudo. No entanto, diferentemente da validação externa, o conjunto usado para avaliar e construir os modelos é o mesmo. Para a construção do modelo, o conjunto de dados é dividido aleatoriamente em partes iguais, sendo um grupo deixado de fora. Assim, um grupo, ou amostra, diferente é usado na validação. As divisões e avaliações são feitas até que todos os dados possam ser previstos (WEHRENS, 2011).

6.2 Aplicações do NIR na detecção de doenças

A técnica do NIR tem a principal vantagem de ser um método não invasivo capaz de fornecer informações qualitativas e quantitativas. Suas aplicações em vegetais são amplas, sendo possível caracterizar a composição química dos tecidos, estádios fisiológicos e propriedades físicas (PASQUINI, 2003). Além disso, o NIR pode ser uma alternativa para a detecção precoce de doenças. Métodos espectrométricos têm como vantagem o fato de serem realizados em tempo real, de forma rápida, confiável e podem detectar doenças antes do aparecimento dos sintomas. Além disso, é uma valiosa fonte de informação para prevenir o desenvolvimento e a propagação das mesmas (SANKARAN et al., 2010).

Em folhas, parte da luz é refletida em direção ao observador, sendo que cada faixa espectral possui informações que dependem das características da superfície da folha, constituição interna e composição bioquímica. Na faixa espectral do infravermelho próximo há uma grande absorção na folha devido aos pigmentos fotossintéticos, sendo refletidos apenas padrões diferentes. Destes é possível inferir o estado fisiológico, a constituição da folha, e portanto, é possível mensurar diferentes tipos de estresses considerando a análise espectral do NIR (PENUELAS; FILELLA, 1998; PASQUINI, 2003).

Devido ao estresse, alterações de reflectância podem ser encontradas em plantas infectadas. Os distúrbios metabólicos gerados pelo patógeno afetam a capacidade da clorofila para absorver a luz e com a deterioração metabólica, aumentos de reflectância visível são observados conforme o avanço da infecção (XU et al., 2007). Na investigação de doenças, a técnica tem sido muito usada por ser um método menos oneroso, não destrutivo e rápido (ABU-KHALAF; SALMAN, 2013), podendo ser uma solução para o diagnóstico precoce das doenças e fornecer uma exatidão na classificação.

Com essa técnica, Rumpf et al. (2010) obtiveram taxas de detecção da cercosporiose em beterraba de 97%, sendo que em 86% dos casos foi possível identificar o agente infeccioso (*Cercospora beticola*, *Uromyces betae* ou *Erysiphe betae*), além de ser possível a detecção pré-sintomática da doença. A detecção também foi possível para infecções causadas por vírus em sementes de melancia (LEE et al., 2016), arroz (JINENDRA et al., 2010); fungos em maçã (SPÁCILOVÁ et al., 2011), feijão (MALTHUS; MADEIRA, 1993), berinjela (WU et al., 2008), milho (BERARDO et al., 2005); além de bactérias em tomate (JONES et al., 2010).

6.2.1 A espectrometria na detecção de doenças em mandioca

Em mandioca, diferentes técnicas de imagens já foram usadas para classificação e avaliação da severidade causadas por pragas e doenças. Em um recente estudo, Sims et al. (2018) avaliaram os danos causados por infecções transmitidas pela mosca branca (*Bemisia tabaci*), bem como desenvolveram modelos de classificação da severidade causada pela praga com uso da região do VIS-NIR. Os resultados destes autores mostraram uma perfeita classificação de folhas infectadas severamente pela fuligem causada por fungos saprófitas, além de ser demonstrarem a relação existente entre as alterações específicas de componentes da planta com diferentes infecções, o que pode ser utilizado em estudos posteriores para diagnóstico.

Abdullakasim (2014) utilizou índices principais de RGB (*red green blue*), para analisar a presença da mancha marrom na mandioca em quatro diferentes níveis de severidade. Este método demonstrou 100% de concordância com a avaliação visual do número de pontos de infecção e $R^2=0,90$ na avaliação da severidade da infecção. Raji et al. (2015) avaliaram o uso da fluorescência induzida

pela luz solar na detecção precoce de danos causados pelo mosaico da mandioca por sensoriamento remoto. Neste estudo foi avaliada a classificação e a severidade de plantas infectadas, sendo obtidos elevadas correlações com a imagem de fluorescência ($R^2= 0,85$), e além disso, a doença foi facilmente detectada em estádios iniciais. A classificação do mosaico africano da mandioca também foi possível utilizando imagens multiespectrais por transmissão, reflexão e dispersão (SANGARE et al., 2015).

7. Referências

ABDULLAKASIM, W. Quantification of the severity of brown leaf spot disease in cassava using image analysis. **Thai Society of Agricultural Engineering Journal**, v. 20, n.2, p.24-32, 2014.

ABU-KHALAF, N.; SALMAN, M. Detecting plant diseases using visible/near infrared spectroscopy. **NIR news**, v. 24, n. 4, p. 12-25, 2013.

ACOSTA, K.; SILVA, F. N.; ZAMORA, L.; QUIÑONES, M.; MARTINEZ, R.; PINOL, B.; URQUIZA, G. P. C.; CARVALHO, C. M.; ZERBINI, F. M. Detection and molecular characterization of an aster yellows group Phytoplasma associated with soybean and evidence of two new 16SRI subgroups in Cuba. **Journal of Plant Pathology**, v. 97, n. 2, p. 339-344, 2015.

AGRE, A. P.; BHATTACHARJEE, R.; DANSI, A.; BECERRA LÓPEZ-LAVALLE, L. A.; DANSI, M.; SANNI, A. Assessment of cassava (*Manihot esculenta* Crantz) diversity, loss of landraces and farmers preference criteria in southern Benin using farmers' participatory approach. **Genetic Resources and Crop Evolution**, v. 64, n. 2, p. 307-320, 2017.

ALIZADEH, H.; QUAGLINO, F.; AZADVAR, M.; KUMAR, S.; ALIZADEH, A.; BOLBOLI, F.; CASATI, P.; BIANCO, P. A. First report of a new citrus decline disease (CDD) in association with double and single infection by '*Candidatus Liberibacter asiaticus*' and '*Candidatus Phytoplasma aurantifolia*' related strains in Iran. **Plant Disease**, v.101, p.2145, 2017.

ALVAREZ, E.; LLANO, G. A.; MEJÍA, J.F. Cassava diseases in Latin America, Africa and Asia. **The cassava handbook**, p. 258, 2012.

ALVAREZ, E.; MEJÍA, J.F.; LLANO, G. A.; LOKE, J. B. Detection and characterization of a phytoplasma associated with frog skin disease in cassava. **Bulletin of Insectology**, v. 60, n. 2, p. 273-274, 2007.

ALVAREZ, E.; MEJÍA, J.F.; LLANO, G. A.; LOKE, J. B.; CALARI, A.; DUDUK, B.; BERTACCINI, A. Characterization of a phytoplasma associated with frogskin disease in cassava. **Plant Disease**, v. 93, p.1139-1145, 2009.

ANGEL, J. C.; NOLT, B.; PINEDA, L. B. Estudios sobre la transmisión por moscas blancas (Homoptera: *Aleyrodidae*) de virus asociados con el cuero de sapo en yuca (*Manihot esculenta* Crantz). **Acta Agronomica**, v. 37, p. 27-39, 1987.

ANYANWU, C. N.; IBETO, C. N.; EZEHOA, S. L.; OGBUAGU, N. J. Sustainability of cassava (*Manihot esculenta* Crantz) as industrial feedstock, energy and food crop in Nigeria. **Renewable Energy**, v. 81, p. 745-752, 2015.

BAI, X.; CORREA, V. R.; TORUÑO, T. Y.; AMMAR, E. D.; KAMOUN, S.; HOGENHOUT, S. A. AY-WB phytoplasma secretes a protein that targets plant cell nuclei. **Molecular Plant-Microbe Interactions**, v. 22, n. 1, p. 18-30, 2009.

BAURIEGEL, E.; GIEBEL, A.; GEYER, M.; SCHMIDT, U.; HERPPICH, W. B. Early detection of *Fusarium* infection in wheat using hyper-spectral imaging. **Computers and Electronics in Agriculture**, v. 75, n. 2, p. 304-312, 2011.

BERARDO, N.; PISACANE, V.; BATTILANI, P.; SCANDOLARA, A.; PIETRI, A.; MAROCCO, A. Rapid detection of kernel rots and mycotoxins in maize by near-infrared reflectance spectroscopy. **Journal of Agricultural and Food Chemistry**, v. 53, n. 21, p. 8128-8134, 2005.

BERTACCINI, A.; DUDUK, B. Phytoplasma and phytoplasma diseases: a review of recent research. **Phytopathologia Mediterranea**, v. 48, n. 3, p. 355-378, 2010.

BERTACCINI, A.; DUDUK, B.; PALTRINIERI, S.; CONTALDO, N. Phytoplasmas and phytoplasma diseases: a severe threat to agriculture. **American Journal of Plant Sciences**, v. 5, n. 12, p. 1763-1788, 2014.

BOAS, S. A. V., OLIVEIRA, S. A. S., BRAGANÇA, C. A. D., RAMOS, J. B.; OLIVEIRA, E. J. Survey of fungi associated with cassava root rot from different producing regions in Brazil. **Scientia Agricola**, v. 74, n. 1, p. 60-67, 2017.

BULL, S. E.; NDUNGURU, J.; GRUISSEM, W.; BEECHING, J. R.; VANDERSCHUREN, H. Cassava: constraints to production and the transfer of biotechnology to African laboratories. **Plant cell reports**, v. 30, n. 5, p. 779-787, 2011.

BUXA, S. V.; DEGOLA, F.; POLIZZOTTO, R.; DE MARCO, F.; LOSCHI, A.; KOGEL, K. H.; TOPPI, L. S.; VAN BEL, A. J. E.; MUSETTI, R. Phytoplasma

infection in tomato is associated with re-organization of plasma membrane, ER stacks, and actin filaments in sieve elements. **Frontiers in plant science**, v. 6, p.650, 2015.

CALVERT, L. A.; CUERVO, M.; LOZANO, I.; VILLAREAL, N.; ARROYAVE, J. Identification of three strains of a virus associated with cassava plants affected by frogskin disease. **Journal of Phytopathology**, v.156, p. 647-653, 2008.

CALVERT, L. A.; THRESH, J. M.; The viruses and virus diseases of cassava. In: HILLOCKS, R.J, THRESH, J.M, BELLOLLI, A.C, eds. **Cassava: Biology, Production and Utilization**, CAB International., Wallingford, 2002.

CARDOZO TÉLLEZ, L.; PARDO, J. M.; ZACHER, M.; TORRES, A.; ÁLVAREZ, E. First Report of a 16SrIII Phytoplasma associated with frogskin disease in cassava (*Manihot esculenta*) in Paraguay. **Plant Disease**, v. 100, n. 7, p. 1492-1492, 2016.

CARRASCO, N. F.; OLER, J. R. L.; MARCHETTI, F. F.; CARNIELLO, M. A.; AMOROZO, M. C. M.; VALLE, T. L.; VEASEY, E. A. Growing Cassava (*Manihot esculenta*) in Mato Grosso, Brazil: genetic diversity conservation in small-scale agriculture. **Economic botany**, v. 70, n. 1, p. 15-28, 2016.

CARVAJAL-YEPES, M.; OLAYA, C.; LOZANO, I.; CUERVO, M.; CASTANO, M.; CUELLAR, W. J. Unraveling complex viral infections in cassava (*Manihot esculenta* Crantz) from Colombia. **Virus research**, v. 186, p. 76-86, 2014.

CHACÓN, J.; MADRIÑÁN, S.; DEBOUCK, D.; RODRIGUEZ, F.; TOHME, J. Phylogenetic patterns in the genus *Manihot* (Euphorbiaceae) inferred from analyses of nuclear and chloroplast DNA regions. **Molecular Phylogenetics and Evolution**, v. 49, n. 1, p. 260-267, 2008.

CHAVARRIAGA-AGUIRRE, P.; BRAND, A.; MEDINA, A.; PRÍAS, M.; ESCOBAR, R.; MARTINEZ, J.; DÍAZ, P.; LÓPEZ, C.; ROCA, W. M.; TOHME, J. The potential of using biotechnology to improve cassava: a review. **In Vitro Cellular & Developmental Biology-Plant**, v. 52, n. 5, p. 461-478, 2016.

CHEGE, M. N.; WAMUNYOKOLI, F.; KAMAU, J.; NYABOGA, E. N. Phenotypic and genotypic diversity of *Xanthomonas axonopodis* pv. *manihotis* causing bacterial blight disease of cassava in Kenya. **Journal of Applied Biology & Biotechnology** v.5, n. 02, p.38-44, 2017.

CHENG, M.; DONG, J.; LEE, M.; BOTTNER-PARKER, K. D.; ZHAO, Y.; DAVIS, R. E.; LASKI, P. J.; ZHANG, Z.; MCBEATH, J. H. Group 16SrXII phytoplasma strains, including subgroup 16SrXII-E ('*Candidatus* Phytoplasma fragariae') and a new subgroup, 16SrXII-I, are associated with diseased potatoes (*Solanum*

tuberosum) in the Yunnan and Inner Mongolia regions of China. **European journal of plant pathology**, v. 142, n. 2, p. 305-318, 2015.

CHOI, Y. H., TAPIAS, E. C., KIM, H. K., LEFEBER, A. W., ERKELENS, C., VERHOEVEN, J. T. J. Metabolic discrimination of *Catharanthus roseus* leaves infected by phytoplasma using ¹H-NMR spectroscopy and multivariate data analysis. **Plant Physiology**, v. 135, n. 4, p. 2398-2410, 2004.

CONTALDO, N.; BERTACCINI, A.; PALTONIERI, S.; WINDSOR, H. M.; WINDSOR, G. D. Axenic culture of plant pathogenic phytoplasmas. **Phytopathologia Mediterranea**, v.51, p. 607-617, 2012.

CORDEIRO, I.; SECCO, R.; SILVA, M. D.; SODRÉ, R. C.; MARTINS, M. L. L. *Manihot* in lista de espécies da flora do Brasil. Jardim Botânico do Rio de Janeiro, 2015. Disponível em: <<http://floradobrasil.jbrj.gov.br/jabot/floradobrasil/FB17591>>.

DAVIS, R. I.; AROCHA, Y.; JONES, P.; MALAU, A. First report of the association of phytoplasmas with plant diseases in the territory of Wallis and Futuna. **Australasian Plant Pathology**, v. 34, n. 3, p. 417-418, 2005.

DICKINSON, M. Loop-mediated isothermal amplification (LAMP) for detection of phytoplasmas in the field. **Plant Pathology: Techniques and Protocols**, v. 1302, p. 99-111, 2015.

DOI, Y.; TERANAKA, M.; YORA, K.; ASUYAMA, H. Mycoplasma-or PLT group-like microorganisms found in the phloem elements of plants infected with mulberry dwarf, potato witches' broom, aster yellows, or paulownia witches' broom. **Japanese Journal of Phytopathology**, v. 33, n. 4, p. 259-266, 1967.

FAO - Food and Agriculture Organization of the United Nations. FAOSTAT database. Disponível em <<http://faostat3.fao.org/analysis/Q/QC/E>>. Acesso em: 20 de setembro de 2017.

FEO, L. D.; ZANINI, A.; RODRÍGUEZ PARDINA, P.; CUERVO, M.; CARVAJAL-YEPES, M.; CUELLAR, W. J. First report of *Cassava common mosaic virus* and *cassava frogskin-associated virus* infecting cassava in Argentina. **Plant Disease**, v. 99, n. 5, p. 733-734, 2015.

FERRARO, V.; PICCIRILLO, C.; TOMLINS, K.; PINTADO, M. E. Cassava (*Manihot esculenta* Crantz) and yam (*Dioscorea* spp.) crops and their derived foodstuffs: safety, security and nutritional value. **Critical reviews in food science and nutrition**, v. 56, n. 16, p. 2714-2727, 2016.

FLÔRES, D.; HAAS, I. C.; CANALE, M. C.; BEDENDO, I. P. Molecular identification of a 16SrIII-B phytoplasma associated with cassava witches' broom disease. **European journal of plant pathology**, v. 137, n. 2, p. 237-242, 2013.

GENUER, R.; POGGI, J. M.; TULEAU-MALOT, C. VSURF: An R package for variable selection using random forests. **R Journal**, v. 7, n. 2, p.19-33, 2015.

GUNDERSEN, D. E.; LEE, I.-M. Ultrasensitive detection of phytoplasmas by nested-PCR assays using two universal primer pairs. **Phytopathologia mediterranea**, v.35, p. 144-151, 1996.

HARRISON, N. A.; DAVIS, R. E.; OROPEZA, C.; HELMICK, E. E.; NARVAEZ, M.; EDEN-GREEN, S.; DOLLET, M.; DICKINSON, M. '*Candidatus* Phytoplasma palmicola', associated with a lethal yellowing-type disease of coconut (*Cocos nucifera* L.) in Mozambique. **International journal of systematic and evolutionary microbiology**, v. 64, n. 6, p. 1890-1899, 2014

HENDERSON, L. Cassava: an economically viable animal feed?: feeds & grazing. **Stockfarm**, v. 5, n. 11, p. 32-33, 2015.

HENRY, G.; HERSHEY, C. Cassava in South America and the Caribbean. **Cassava: Biology, production and utilization**, p. 17-40, 2002.

HOKANSON, K. E., ELLSTRAND, N. C., DIXON, A. G., KULEMBEKA, H. P., OLSEN, K. M.; RAYBOULD, A. Risk assessment of gene flow from genetically engineered virus resistant cassava to wild relatives in Africa: an expert panel report. **Transgenic research**, v. 25, n. 1, p. 71-81, 2016.

HOWELER, R. L.; THOMAS, N.; CALDERÓN CEDEÑO, C.; ARAUJO, Q. R.; ARAUJO, R.K.P.; NEY, J. R. M.; BONSU, O.; AMOAH, F. M.; OPPONG, F. K. **Save and grow: cassava. A guide to sustainable production intensification Produire plus avec moins Ahorrer para crecer**. FAO, Roma (Italia), 2013.

IBGE – Instituto Brasileiro de Geografia e Estatística. Produção Agrícola. Disponível em: [ftp://ftp.ibge.gov.br/Producao_Agricola/Levantamento_Sistematico_da_Producao_Agricola_\[mensal\]/Fasciculo/lspa_201509.pdf](ftp://ftp.ibge.gov.br/Producao_Agricola/Levantamento_Sistematico_da_Producao_Agricola_[mensal]/Fasciculo/lspa_201509.pdf) > Acesso em: 10 de novembro de 2017.

JINENDRA, B.; TAMAKI, K.; KUROKI, S.; VASSILEVA, M.; YOSHIDA, S.; TSENKOVA, R. Near infrared spectroscopy and aquaphotomics: Novel approach for rapid in vivo diagnosis of virus infected soybean. **Biochemical and biophysical research communications**, v. 397, n. 4, p. 685-690, 2010.

JONES, C. D.; JONES, J. B.; LEE, W. S. Diagnosis of bacterial spot of tomato using spectral signatures. **Computers and Electronics in Agriculture**, v. 74, n. 2, p. 329-335, 2010.

KAKIZAWA, S.; YONEDA, Y. The role of genome sequencing in phytoplasma research. **Phytopathogenic Mollicutes**, v. 5, n. 1, p. 19-24, 2015.

KAWEESI, T., KAWUKI, R., KYALIGONZA, V., BAGUMA, Y., TUSIIME, G.; FERGUSON, M. E. Field evaluation of selected cassava genotypes for cassava brown streak disease based on symptom expression and virus load. **Virology Journal**, v. 11, n. 1, p. 216, 2014.

KOGOVSĚK, P.; HODGETTS, J.; HALL, J.; PREZELJ, N.; NIKOLIC, P.; MEHLE, N.; LENARCIC, R.; ROTTER, A.; DICKINSON, M.; BOONHAM, N.; DERMASTIA, N. LAMP assay and rapid sample preparation method for on-site detection of flavescence dorée phytoplasma in grapevine. **Plant Pathology**, v. 64, n. 2, p. 286-296, 2015.

KRA, K. D.; TOUALY, Y. M. N.; KOUAMÉ, A. C.; DIALLO, H. A.; AROCHA ROSETE, Y. First report of a phytoplasma affecting cassava orchards in Cote d'Ivoire. **New Disease Reports**, v. 35, p. 21-21, 2017.

LEE, I.; DAVIS, R. E.; GUNDERSEN-RINDAL, D. E. Phytoplasma: phytopathogenic mollicutes. **Annual Reviews in Microbiology**, v. 54, n. 1, p. 221-255, 2000.

LEE, H.; KIM, M. S.; LIM, H. S.; PARK, E.; LEE, W. H.; CHO, B. K. Detection of cucumber green mottle mosaic virus-infected watermelon seeds using a near-infrared (NIR) hyperspectral imaging system: Application to seeds of the "Sambok Honey" cultivar. **Biosystems Engineering**, v. 148, p. 138-147, 2016.

LEGG, J. P.; KUMAR, P. L.; MAKESHKUMAR, T.; TRIPATHI, L.; FERGUSON, M.; KANJU, E.; NTAWURUHUNGA, P.; CUELLAR, W. Cassava virus diseases: biology, epidemiology, and management. **Advances in Virus Research**, v. 91, p. 85-142, 2015.

LEGG, J. P.; THRESH, J. M. Cassava virus diseases in Africa. In: **Proceedings of the First International Conference on Plant Virology in Sub-Saharan Africa** (4–8 June 2001, Ibadan, Nigeria), IITA, Ibadan, Nigeria. p. 517-522. 2003.

MALTHUS, T. J.; MADEIRA, A. C. High resolution spectroradiometry: spectral reflectance of field bean leaves infected by *Botrytis fabae*. **Remote Sensing of Environment**, v. 45, n. 1, p. 107-116, 1993.

MCCALLUM, E. J.; ANJANAPPA, R. B.; GRUISSEM, W. Tackling agriculturally relevant diseases in the staple crop cassava (*Manihot esculenta*). **Current Opinion in Plant Biology**, v. 38, p. 50-58, 2017.

MEISSNER FILHO, P. E.; VELAME, K. V. C. O vírus do couro do sapo da mandioca, 2006. Artigo em Hipertexto. Disponível em: <http://www.infobibos.com/Artigos/2006_3/CouroSapo/index.htm>. Acesso em: 26 setembro de 2015.

NEJAT, N.; VADAMALAI, G. Diagnostic techniques for detection of phytoplasma diseases: past and present. **Journal of Plant Diseases and Protection**, v.120, p. 16-25, 2013.

OLIVEIRA, S. A. S.; ABREU, E. F. M.; ARAÚJO, T. S.; OLIVEIRA, E. J.; ANDRADE, E. C.; GARCIA, J. M. P.; ALVAREZ, E. First report of a 16SrIII-L phytoplasma associated with frogskin disease in cassava (*Manihot esculenta* Crantz) in Brazil. **Plant disease**, v. 98, n. 1, p. 153-153, 2014.

OLSEN, K. M. SNPs, SSRs and inferences on cassava's origin. **Plant Molecular Biology**, v. 56, n. 4, p. 517-526, 2004.

PASQUINI, C. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. **Journal of the Brazilian Chemical Society**, v. 14, n. 2, p. 198-219, 2003.

PEÑUELAS, J.; FILELLA, I. Visible and near-infrared reflectance techniques for diagnosing plant physiological status. **Trends in Plant Science**, v. 3, n. 4, p. 151-156, 1998.

PÉREZ-LÓPEZ, E.; DUMONCEAUX, T. J. Detection and identification of the heterogeneous novel subgroup 16SrXIII-(A/I) I phytoplasma associated with strawberry green petal disease and Mexican periwinkle virescence. **International Journal of Systematic and Evolutionary Microbiology**, v. 66, n. 11, p. 4406-4415, 2016.

PÉREZ-LÓPEZ, E.; OLIVIER, C. Y.; LUNA-RODRÍGUEZ, M.; DUMONCEAUX, T. J. Phytoplasma classification and phylogeny based on *in silico* and *in vitro* RFLP analysis of cpn60 universal target sequences. **International Journal of Systematic and Evolutionary Microbiology**, v. 66, n. 12, p. 5600-5613, 2016.

RAJI, S. N., SUBHASH, N., RAVI, V., SARAVANAN, R., MOHANAN, C. N., NITA, S., & KUMAR, T. M. Detection of mosaic virus disease in cassava plants by sunlight-induced fluorescence imaging: a pilot study for proximal

sensing. **International Journal of Remote Sensing**, v. 36, n. 11, p. 2880-2897, 2015.

RAVINDRAN, V. Cassava leaves as animal feed: potential and limitations. **Journal of the Science of Food and Agriculture**, v. 61, n. 2, p. 141-150, 1993.

REDDY, P. P. **Plant protection in tropical root and tuber crops**. Springer India, 2015.

RID, M.; MESCA, C.; AYASSE, M.; GROSS, J. Apple proliferation phytoplasma influences the pattern of plant volatiles emitted depending on pathogen virulence. **Frontiers in Ecology and Evolution**, v. 3, p. 152, 2016.

ROPELEWSKA, E.; ZAPOTOCZNY, P. Classification of *Fusarium*-infected and healthy wheat kernels based on features from hyperspectral images and flatbed scanner images: a comparative analysis. **European Food Research and Technology**, p.1-10, 2018.

RUMPF, T., M, A. K., S, U., OERKE, E. C., DEHNE, H. W., e PLÜMER, L. Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. **Computers and Electronics in Agriculture**, v.74, n. 1, p. 91-99, 2010.

SÁNCHEZ, A. S.; SILVA, Y. L.; KALID, R. A.; COHIM, E.; TORRES, E. A.; Waste bio-refineries for the cassava starch industry: New trends and review of alternatives. **Renewable and Sustainable Energy Reviews**, v. 73, p. 1265-1275, 2017.

SÁNCHEZ, T.; DUFOUR, D.; MORENO, I. X.; CEBALLOS, H. Comparison of pasting and gel stabilities of waxy and normal starches from potato, maize, and rice with those of a novel waxy cassava starch under thermal, chemical, and mechanical stress. **Journal of agricultural and food chemistry**, v. 58, n. 8, p. 5093-5099, 2010.

SANGARE; M.; AGNEROH; T. A.; BAGUI; O. K.; TRAORE; I.; BA; A.; ZOUEU; J. T. Classification of African mosaic virus infected cassava leaves by the use of multi-spectral imaging. **Optics and Photonics Journal**, v. 5, n.8, p. 261, 2015.

SANKARAN, S.; MISHRA, A.; EHSANI, R.; DAVIS, C. A review of advanced techniques for detecting plant diseases. **Computers and Electronics in Agriculture**, v. 72, n.1, p.1-13, 2010.

SIMS, N. C.; BARRO, P.; NEWNHAM, G. J.; KALYEBI, A.; MACFADYEN, S.; MALTHUS, T. J. Spectral separability and mapping potential of cassava leaf

damage symptoms caused by whiteflies (*Bemisia tabaci*). **Pest management science**, v. 74, n. 1, p. 246-255, 2018.

SOUZA, A. N.; SILVA, F. N.; BEDENDO, I. P.; CARVALHO, C. M. A Phytoplasma belonging to a 16SrIII-A subgroup and dsRNA virus associated with cassava frogskin disease in Brazil. **Plant Disease**, v.98, p.771-779, 2014.

SPÁCILOVÁ, V.; SAFRÁNKOVÁ, I. Potential of chlorophyll fluorescence and VIS/NIR spectroscopy measurement use for the detection of nitrogen content and disease infection of apple leaves. **Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis (Czech Republic)**, v.9, n.6, p317-328, 2011.

THOMAS-SHARMA, S.; ANDRADE-PIEDRA, J.; CARVAJAL YEPES, M.; HERNANDEZ NOPSA, J. F.; JEGER, M. J.; JONES, R. C. A.; GARRETT, K. A. KROMANN, P.; LEGG, J. P.; YUEN, J.; FORBES, G. A. A Risk assessment framework for seed degeneration: informing an integrated seed health strategy for vegetatively propagated crops. **Phytopathology**, v.107, n.10, p1123-1135, 2017.

TONUKARI, N. J. Cassava and the future of starch. **Electronic journal of biotechnology**, v. 7, n. 1, p. 5-8, 2004.

TÜRKER-KAYA, S.; HUCK, C. W. A Review of mid-infrared and near-infrared imaging: principles, concepts and applications in plant tissue analysis. **Molecules**, v. 22, n. 1, p.168, 2017.

VARMUZA, K.; FILZMOSER, P. **Introduction to multivariate statistical analysis in chemometrics**. CRC press, 2009.

WEHRENS, R. **Chemometrics with R: multivariate data analysis in the natural sciences and life sciences**. Springer Science & Business Media, 2011.

WU, D.; FENG, L.; ZHANG, Z.; HE, Y. Early detection of *Botrytis cinerea* on eggplant leaves based on visible and near-infrared spectroscopy. **Transactions of the ASABE**, v. 51, n. 3, p. 1133-1139, 2008.

XU, H. R.; YING, Y. B.; FU, X. P.; ZHU, S. P. Near-infrared spectroscopy in detecting leaf miner damage on tomato leaf. **Biosystems Engineering**, v. 96, n. 4, p. 447-454, 2007.

ZAMBON, Y.; CANEL, A.; BERTACCINI, A.; CONTALDO, N. Molecular diversity of phytoplasmas associated with grapevine yellows disease in north-eastern Italy. **Phytopathology**, v.108, n.2, p.206-214, 2018.

ZHANG, M.; XIE, L.; YIN, Z.; KHANAL, S. K.; ZHOU, K. Biorefinery approach for cassava-based industrial wastes: current status and opportunities. **Bioresource technology**, v. 215, p. 50-62, 2016.

ZHAO, Y.; WEI, W.; DAVIS, R. E.; LEE, I. M. Recent advances in 16S rRNA gene-based phytoplasma differentiation, classification and taxonomy. **Phytoplasmas: Genomes, Plant Hosts and Vectors**, p. 64-92, 2010.

ZHAO, Y.; WEI, W.; DAVIS, R. E.; LEE. Should '*Candidatus* Phytoplasma' be retained within the order Acholeplasmatales?. **International journal of systematic and evolutionary microbiology**, v. 65, n. 3, p. 1075-1082, 2015.

MATERIAL E MÉTODOS

Material Vegetal

Amostras de folhas de mandioca foram coletadas de acessos do Banco Ativo de Germoplasma de Mandioca (BAG-Mandioca) da Embrapa Mandioca e Fruticultura, localizada em Cruz das Almas-BA-Brasil (2°40'19"S, 39°06'22"W, altitude 220 m). As amostras foram constituídas por plantas sadias e infectadas com idade entre 12 a 18 meses pertencentes a 238 acessos de mandioca, sendo 120 acessos sadios e 118 infectados pelo couro de sapo, classificadas com base no diagnóstico visual da sintomatologia da doença nas raízes. Em cada acesso foram coletadas cinco folhas jovens de três plantas que possuíam sintomas típicos do couro de sapo nas raízes e de plantas sem infecção por couro de sapo ou outras doenças da cultura da mandioca. Portanto, no total foram analisadas 714 amostras.

Tratamento das amostras

As folhas coletadas foram acondicionadas em sacos de papel e desidratadas em estufa, a 60°C até peso constante. Depois de desidratadas, as folhas foram maceradas, padronizadas e mantidas em frascos fechados hermeticamente, em temperatura ambiente, até o momento da leitura no NIR. A padronização da granulometria das amostras foi realizada por peneiração em malha de "voil" a fim de reduzir desvios de reflexão durante as leituras.

Aquisição dos espectros

Os espectros de reflectância com transformação de Fourier na região do infravermelho próximo foram adquiridos no espectrofotômetro NIRFlex N500 FT NIR (Büchi, Labortechnik AG, Suíça) com o auxílio do software NIRWare 1.5, na faixa espectral de 1000 a 2500 nm, com resolução espectral de 8 cm⁻¹ e média de 32 varreduras. As medidas foram realizadas utilizando-se *vials*, contendo aproximadamente 100 mg de folhas secas e maceradas, oriundas de plantas sadias ou infectadas. Estes recipientes foram acondicionados diretamente no espectrômetro utilizando uma célula de medição *Vial holder* e, os espectros foram gravados no modo de reflexão difusa, com leituras em duplicata das 714 amostras, totalizando 1428 espectros.

Análise dos dados espectrais

O primeiro pré-processamento utilizado foi a derivação para remover deslocamentos sistemáticos da linha de base ("offsets") e com isso remover efeitos aditivos e multiplicativos dos espectros. Para isso, utilizou-se o algoritmo Savitzky-Golay com a primeira derivada, janela de análise com 11 pontos e polinômio de segunda ordem. Após esse tratamento, foi aplicada a variação normal padrão (SVN - *standard normal variate*) para remover efeitos de dispersão dos espectros, e em seguida os dados foram normalizados para uma mesma escala a fim de reduzir os problemas de efeitos multiplicativos nas análises. As funções *savitzkyGolay*, *padrãoNormalVariate* e *scale* do pacote *prospectr* do software R versão 3.4.0 (R Core Team, 2017), foram utilizadas para implementação do pré-processamento Savitzky-Golay, SNV e normalização, respectivamente.

Após o pré-processamento dos espectros foi avaliada a possibilidade de implementar procedimentos de seleção de variáveis com uso de eliminação recursiva de características (RFE) com uso do algoritmo *random forest*, para selecionar um pequeno conjunto (possivelmente não colineares) de espectros preditores ótimos para ajuste do modelo. O procedimento RFE implementado no pacote *Vsurf* do software R versão 3.4.0 (R Core Team, 2017), realiza a seleção *backward* de variáveis, classificando sua importância para o modelo inicial considerando todos os preditores. O algoritmo constrói vários modelos de calibração com base nos preditores mais importantes, onde p_i é um elemento de uma sequência predeterminada $\{p_1, p_2, \dots, p_n\}$ dos possíveis números de preditores, e com isso o conjunto de preditores que produz o melhor modelo entre os modelos candidatos é selecionado. Assim, a construção dos modelos de classificação foi realizada em duas condições: a) com toda a informação espectral e, b) com seleção de regiões com maior relevância para as diferenças amostrais entre genótipos infectados e sadios.

Métodos de classificação

Inicialmente cada amostra de folhas dos acessos coletados foi classificada em dois conjuntos de dados: amostras "infectadas" e "sadias". Em seguida, seis métodos de classificação foram testados: *Bayesian Generalized Linear Models*

(BGLM, método "bayesglm"), *Extreme Learning Machine* (ELM, método "elm"), *High Dimensional Discriminant Analysis* (HDDA, método "hdda"), *Partial Least Squares* (PLS, método "pls"), *Parallel Random Forest* (PRANDF, método "parRF") e *Support Vector Machines with Linear Kernel* (SVM, método "svmLinear").

A importância dos preditores para prever as classes é realizada para os modelos de classificação BGML, ELM, HDDA e SVM, pela análise da curva ROC (*Receiver Operating Characteristic Curve*) de cada preditor. Para isso, uma série de pontos de corte é aplicada aos dados do preditor, sendo a curva calculada a partir da sensibilidade e a especificidade de cada um dos pontos de corte. A área acima da curva, calculada pela regra trapezoidal, é usada como medida de importância da variável. No PLS, a medida de importância dos preditores é baseada nas somas ponderadas dos coeficientes de regressão absolutos. Os pesos são uma função da redução das somas dos quadrados em relação ao número de componentes PLS e são computados separadamente para cada resultado. Portanto, a contribuição dos coeficientes é ponderada proporcionalmente à redução das somas dos quadrados. No modelo PRANDF, essa medida é calculada com a permutação de cada preditor de dados do modelo. A diferença entre as duas acurácias é então calculada em média e normalizada pelo erro padrão. Se o erro padrão for igual a 0 para uma variável, a inclusão do preditor no modelo não é realizada.

A avaliação da eficiência dos métodos de classificação foi realizada por meio de validação cruzada (10-folds com cinco repetições). Neste caso, o método separa aleatoriamente o conjunto de dados em 10 subconjuntos, dentre os quais nove foram utilizadas para treinamento do modelo e um deles para validação do modelo de treinamento. Estas análises foram realizadas em cinco repetições para cálculo da acurácia média (porcentagem de amostras corretamente classificadas dentre todas as amostras) e coeficiente Kappa de Cohen (k) para avaliar o desempenho dos modelos. O coeficiente k baseia-se na concordância das classificações e no número de vezes que as amostras foram distribuídas na mesma classe. Os valores variam entre 0 e 1, sendo que "0" significa não haver concordância, além do puro acaso, e "1" significa concordância perfeita. Todas estas análises foram realizadas no pacote *caret* do software R versão 3.4.0 (R Core Team, 2017).

RESULTADOS

Curvas espectrais das amostras

A assinatura espectral dos acessos de mandioca sem infecção e infectados com couro de sapo demonstrou a existência de diversos pontos de diferenças entre estas duas classes de classificação (Figura 1). Por meio da análise visual do espectro, verificou-se que as principais faixas espectrais que revelam diferenças entre as amostras foram: 1370-1420, 1450-1550, 1675-1700, 1750-1850, 1950-2000, 2025-2050, 2100-2155, 2250-2300 e 2325-2375 nm, sendo, portanto, potencialmente úteis para discriminar os acessos de mandioca infectados com couro de sapo. Nas demais regiões, as assinaturas espectrais são muito semelhantes ao controle (acessos não infectados) e, portanto, são mais difíceis de discriminar os materiais.

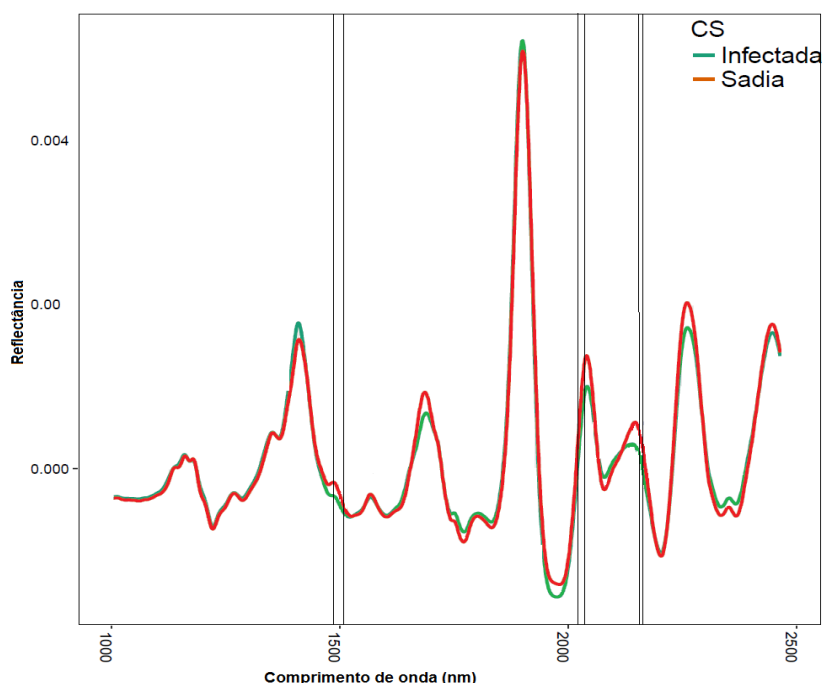


Figura 1. Assinatura espectral média, após pré-tratamento, dos 238 acessos de mandioca (120 acessos sadios e 118 infectados com couro de sapo). As regiões em destaque referem-se às faixas (1490-1516nm; 2029 a 2042nm; 2151 e 2153nm) que apresentaram maior importância nos modelos de classificação dos acessos quanto à infecção pelo couro de sapo.

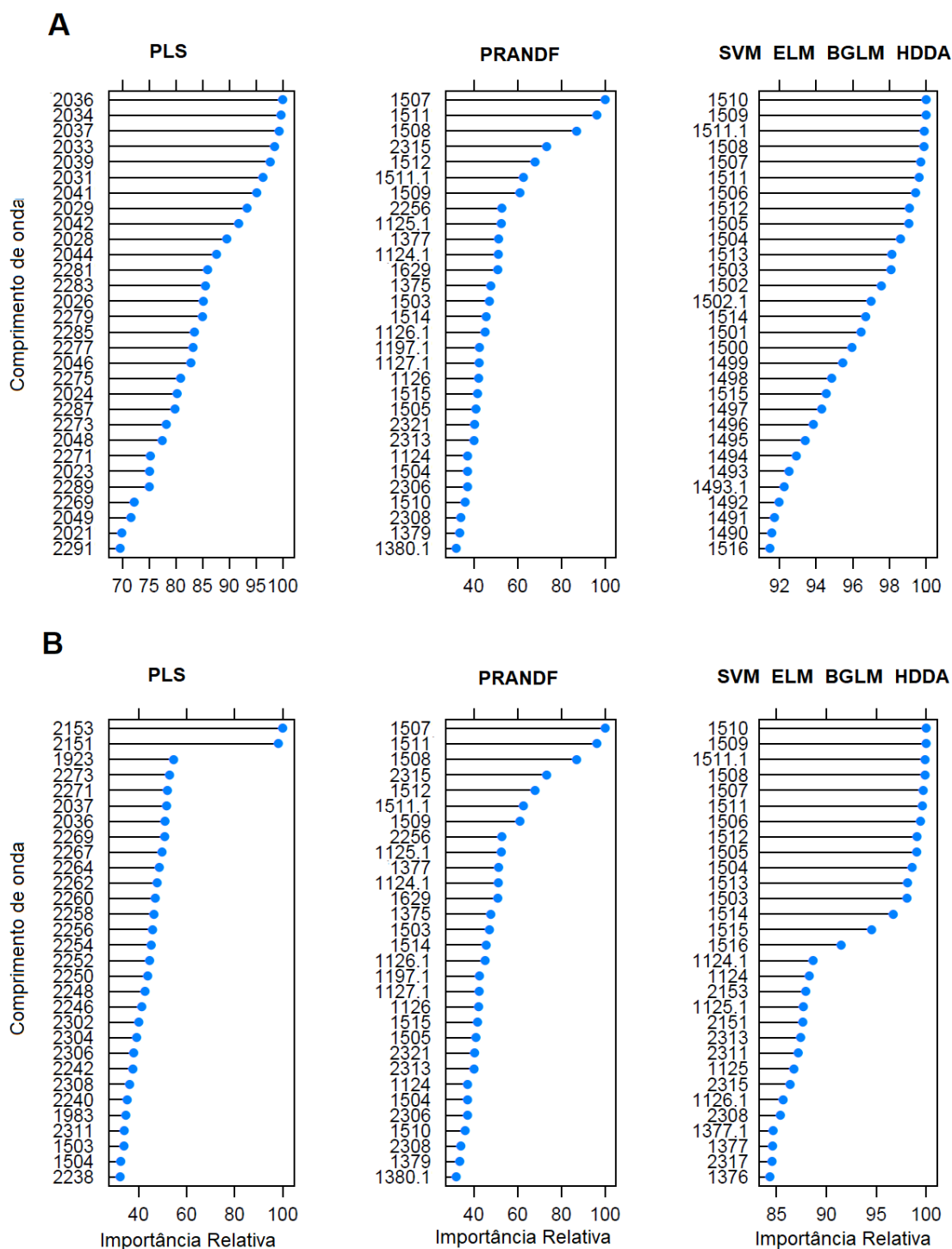


Figura 2. Importância dos 30 comprimentos de onda mais relevantes na predição dos modelos, sem seleção (A) e com seleção de comprimentos de onda do infravermelho próximo (B).

Por outro lado, após as análises de seleção de variáveis (comprimentos de onda) mais relevantes para explicar a diferença entre acessos infectados e não infectados com couro de sapo, via algoritmo *random forest*, três regiões mais específicas foram destacadas por serem mais importantes para os modelos de classificação (Figura 1). Segundo Workman Jr e Weyer (2012) a primeira região (1490-1516 nm) corresponde a vibrações de moléculas que contém ligações N-H e O-H, como carboidratos e aminoácidos. A segunda região (2029 a 2042 nm) também foi relacionada com ligações N-H de aminoácidos e a terceira (2151 e 2153 nm) com ligações C-H de compostos aromáticos.

As regiões destacadas na Figura 1 representam as faixas espectrais com correspondência com as classes infectadas e sadias, com relevância acima de 90% para a construção dos diferentes modelos de classificação (Figura 2). Comprimentos de onda individuais nas faixas de 1490 a 1516 nm e 2029 a 2042 nm apresentaram potencial máximo de discriminação entre as classes infectadas e sadias, utilizando todo o conjunto de dados. A primeira região (1490 a 1516 nm) foi mais importante na construção dos modelos de predição ELM, BGLM, HDDA, SVM e PRANDF, sendo que para o PRANDF uma menor faixa espectral (entre 1507 e 1511 nm) foi mais relevante. A segunda (2029 a 2042 nm) e a terceira (2151 e 2153 nm.) região foram importantes somente para o modelo de classificação baseado em PLS no modelo completo e reduzido, respectivamente. Por outro lado, a faixa espectral entre 1503 a 1516 apresentou maior relevância para os modelos de predição ELM, BGLM, HDDA, SVM.

A importância de cada preditor realizada pela curva ROC, pelos modelos: SVM, BGLM, ELM e HDDA, atribuiu uma importância maior a primeira área apresentada na Figura 1, permanecendo todos os comprimentos de onda nessa faixa. A distribuição do peso (Figura 2) entre os comprimentos decresceu suavemente, mantendo-se em uma variação de 92 a 100% de importância. O modelo PLS, que utiliza a regressão, também manteve esse comportamento de distribuição de peso, no entanto a variação foi maior e abrangeu de 70 a 100%. Pode-se observar que, além da região mencionada anteriormente com importância acima de 90%, o modelo apresentou uma segunda área que compreende os comprimentos de onda de 2261 a 2293 nm. O modelo PRANDF foi o que mais diferiu na distribuição de peso dos comprimentos de onda (de 40 a 100%) e faixa selecionada (1124-1197 nm; 1375-1379 nm; 1503-1515 nm e 2306-

2321 nm). Apesar disso, a maior importância situa-se na mesma faixa atribuída pela curva de ROC, entretanto, com um número reduzido de comprimentos de ondas individuais. Esse comportamento deve-se ao PRANDF atribuir uma seleção de variáveis na etapa de construção, igual ao modelo reduzido, desta forma, as variáveis que possuíam a mesma informação foram excluídas ainda na construção dos modelos.

Comportamento semelhante foi observado pelos modelos que utilizam a curva de ROC no modelo reduzido, como a seleção de variáveis por *random forest*. Para SVM, BGLM, ELM e HDDA, a faixa dos 30 comprimentos com maior importância teve uma distribuição de pesos de 85 a 100% e abrangeu os comprimentos de onda, além da previamente mencionada, 1124-1125 nm, 1376-1377 nm, 2151-2153 nm e 2308-2317 nm. O modelo PLS, também apresentou uma maior variação na distribuição de peso (40 a 100%) e a região que apresentou maior importância foi diferente do modelo completo, os comprimentos 2151 e 2153 nm. Os demais comprimentos foram 1503-1504 nm, 1923 nm, 2036-2037 nm, 2240-2271 nm (utilizada no modelo completo) e 2302-231 nm.

Métodos de classificação

A avaliação dos modelos propostos para a classificação de plantas sadias e infectadas com o couro de sapo foi realizada de acordo com os valores de acurácia e índice de Kappa (Tabela 1). A acurácia de classificação variou de 81,17% a 99,07% para o modelo completo com os comprimentos de onda, e de 82,27% a 96,06% para o modelo reduzido com seleção de comprimentos de onda via algoritmo *random forest*. Portanto, independente do método utilizado a acurácia de classificação dos acessos infectados e não infectados foi acima de 81%, o que indica uma confiabilidade muito elevada para uso desta ferramenta como método de diagnose do couro de sapo. Outra observação importante é que as médias de acurácia foram melhores com a utilização de toda a região espectral para a maioria dos métodos de classificação. Apenas o método de classificação ELM apresentou um ligeiro aumento na acurácia ao se utilizar o modelo reduzido (de 81,17 para 83,41%).

Tabela 1. Acurácia e coeficiente de Kappa com os dados espectrais completos e com a seleção de variáveis.

Método	Acurácia (%)		Kappa	
	Modelo completo	Modelo reduzido	Modelo completo	Modelo reduzido
SVM	99,07	96,06	0,98	0,96
BGLM	98,92	96,18	0,98	0,92
PRANDF	93,39	93,39	0,87	0,87
PLS	91,64	88,78	0,83	0,78
HDDA	85,37	82,27	0,71	0,64
EML	81,17	83,41	0,62	0,67
Média	91,59	90,02	0,83	0,81

Os modelos SVM e BGLM apresentaram as melhores acurácias 99,07 e 98,92%, respectivamente, com o uso de toda região espectral. Com a seleção dos comprimentos de onda a acurácia foi de 96,06 e 96,18% para SVM e BGLM, respectivamente. Apenas o modelo PRANDF não apresentou nenhuma variação na acurácia de classificação quando se utilizou os modelos completos e reduzidos.

Assim, como na acurácia, a seleção de comprimentos de onda interferiu nos valores de Kappa, havendo redução no nível de concordância para a maioria dos métodos, à exceção do PRANDF, que não foi alterado e o ELM que apresentou um aumento neste valor (de 0,62 para 0,67). Portanto, a seleção de comprimentos de onda do NIR para a classificação dos acessos de mandioca quanto à infecção causada pelo couro de sapo parece não ter efeitos positivos na melhoria dos parâmetros de qualidade da técnica.

O nível de concordância dos modelos com relação ao índice de Kappa foi avaliado por meio de uma classificação proposta por Landis e Koch (1977), apresentada na Tabela 2. Segundo essa classificação, a concordância foi quase perfeita para BGLM, SVM, PRANDF e PLS, variando entre 0,81 até 1,0 e; substancial para os modelos EML, HDDA, com níveis entre 0,61 e 0,80.

Tabela 2. Interpretação do coeficiente de Kappa

Coeficiente Kappa	Nível de concordância
< 0,00	Não existe
0,00 – 0,20	Mínima
0,21 – 0,40	Razoável
0,41 – 0,60	Moderada
0,61 – 0,80	Substancial
0,81 – 1,00	Quase perfeita

Adaptada de Landis e Koch (1977).

De forma geral, observou-se que os desvios padrão da acurácia e índice Kappa foram mais elevados para a maioria dos métodos de classificação quando se procedeu a seleção dos comprimentos de onda do infravermelho (Figura 3). Com isso, além da menor acurácia este procedimento pode incorrer em maiores imprecisões na diagnose do couro de sapo.

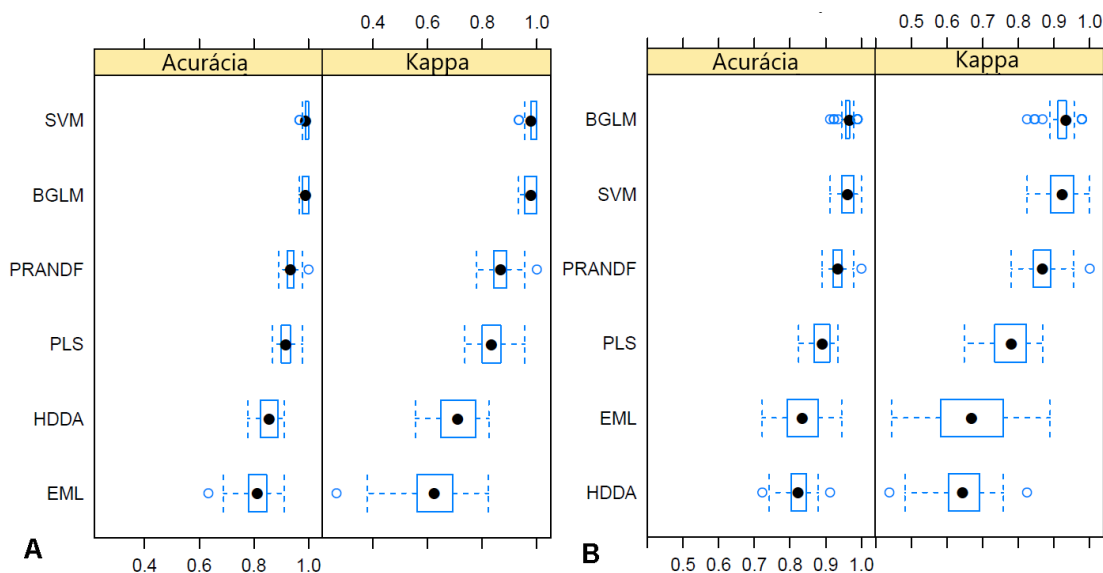


Figura 3. Acurácia e coeficiente de Kappa dos seis modelos testados para predição do couro de sapo em mandioca, sem seleção (A) e com seleção de comprimentos de onda do infravermelho próximo (B).

Na comparação entre as acurácias dos modelos de predição (Figura 4), verificou-se que embora o desvio padrão tenha aumentado, a seleção de

comprimentos de onda reduziu as diferenças entre os modelos e não alterou substancialmente as relações entre eles. Quando comparados, os modelos SVM e BGLM não apresentaram diferenças significativas na acurácia, independentemente da seleção ou não dos comprimentos de onda do NIR. Por outro lado, os modelos EML e HDDA não apresentaram diferenças na média das acurácias apenas quando se praticou a seleção de comprimentos de onda do NIR. As maiores diferenças estão entre os que obtiveram menor acurácia, ELM e HDDA com ambos os modelos SVM e BGLM, que tiveram melhor desempenho.

A maioria dos métodos de classificação utilizados para separar os acessos de mandioca infectados e não infectados com couro de sapo apresentou diferenças significativas entre eles e, portanto, a escolha dos métodos mais acurados para a diagnose deste e de outras doenças da mandioca deve levar em consideração as diferenças metodológicas impostas por cada método de análise a fim de selecionar aquele que proporciona maior acurácia e menor desvio padrão nos parâmetros de avaliação dos modelos.

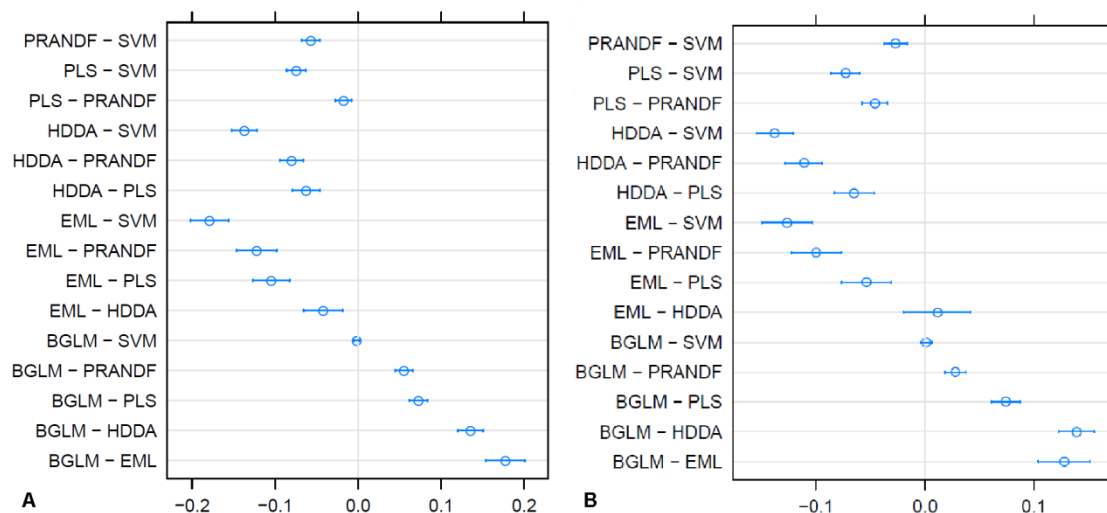


Figura 4. Comparação da acurácia entre os modelos de predição pelo teste t, ao nível de confiança de 0,997 sem seleção (A) e com seleção de comprimentos de onda do infravermelho próximo (B).

DISCUSSÃO

Viabilidade do uso do NIR para diagnose do couro de sapo

Este trabalho demonstrou a viabilidade do uso do NIR para diagnose indireta da presença do couro de sapo da mandioca, pois alguns métodos de classificação como SVM e BGLM apresentaram cerca de 99% de acurácia e 0,98 de índice de Kappa, o que reflete sua capacidade de separar as amostras infectadas com couro de sapo das amostras sadias. Mesmo os piores métodos de classificação apresentaram acurácias acima de 81% e índice de Kappa de 0,62, que ainda são bastante interessantes para a diagnose do couro de sapo, sobretudo por serem técnicas bastante robustas e rápidas.

A elevada capacidade de classificação dos acessos infectados com couro de sapo da mandioca pode estar associada ao fato de que de modo geral, os métodos de reflectância espectral dependem da radiação refletida ou emitida de diferentes corpos, de modo que cada material possui uma assinatura espectral específica (HUNSCHE et al., 2011), isso faz com que mesmo analisando acessos de mandioca geneticamente diferentes, o NIR permite identificar características que sejam comuns a acessos infectados com couro de sapo. Especificamente no caso da assinatura de plantas, a reflectância espectral, na faixa do infravermelho próximo entre 760 e 2500 nm, aumenta ou diminui dependendo de fatores físicos externos, bem como pela condição fisiológica o teor de água, a rugosidade da superfície e atividade estomática. Estas alterações, que afetam a penetração da luz através da estrutura da folha, fornecem uma assinatura espectral que pode ser utilizada para indicar modificações esses fatores (PINTER et al., 2003; XU et al., 2007; SANKARAN et al., 2010; KHALED et al., 2018a).

Até o momento não existem relatos das alterações morfológicas e bioquímicas causadas pelo couro de sapo da mandioca. Entretanto, de modo geral, fitoplasmas causadores de doenças em plantas tendem a alterar o metabolismo energético, sendo este um componente chave para a compreensão da sua biologia e patogenicidade (BERTACCINI et al., 2014). Em algumas espécies ornamentais as concentrações de carboidratos no floema, raízes e folhas de plantas infectadas com fitoplasma diferiram das plantas saudáveis e, além disso, houve uma forte associação entre os níveis alterados de oxigênio e dióxido de carbono e a abundância do fitoplasma (SEARS et al., 1997; LEPKA et

al., 1999; MAUST et al., 2003). Também foram relatados a redução na concentração de pigmentos fotossintéticos e de proteínas solúveis totais, bem como alterações no equilíbrio hormonal, transporte de aminoácidos e ocorrência de folato e presença de genes homólogos de endopetidase em genomas de fitoplasma nas plantas hospedeiras infectadas com diferentes fitoplasmas (LEPKA et al., 1999; JAGOUEIX-EVEILLARD et al., 2001; BERTAMINI et al., 2002; DAVIS et al., 2003; MAUST et al., 2003; MUSETTI et al., 2005). Possivelmente, algum tipo de alteração desta natureza possa estar ocorrendo nas plantas de mandioca infectadas com couro de sapo, permitindo que o NIR possa captar estas diferenças fisiológicas e bioquímicas de modo a permitir uma acurácia de predição bastante elevada na separação dos grupos infectados e sadios.

De fato, alguns autores já relataram o uso de comprimentos de ondas refletidas do visível ao infravermelho próximo (Visible-Near Infrared, V-NIR) para distinguir plantas sadias e infectadas com fungos causadores de doenças foliares como *Cercospora beticola*, *Erysiphe betae* e *Uromyces betae* em beterraba açucareira (MAHLEIN et al., 2010). Outro estudo conduzido por Abdulridha et al. (2016) demonstrou a potencialidade de uso do V-NIR para discriminar abacateiros infectados dos não infectados com a murcha do louro (*laurel wilt disease*), cancrios de *Phytophthora* e danos causados por salinidade do solo, tendo em vista a elevada acurácia de classificação (>98%) com uso do método perceptron multicamada, baseado em redes neurais.

Análise e seleção dos comprimentos de onda

A redução de variáveis (comprimentos de onda) nos modelos de classificação visa manter um menor número de variáveis, porém de maior importância relativa. Essa redução ocorre principalmente pela eliminação de variáveis não informativas e que contêm informações redundantes que podem ser omitidas para fins de predição (HAPFELMEIER; ULM, 2013). Entretanto, em relação à classificação dos acessos de mandioca infectados com o couro de sapo ou sadios, a seleção de comprimentos de onda não apresentou nenhum impacto positivo significativo nos valores médios de acurácia e índice de Kappa.

A seleção de variáveis apresentou um ligeiro aumento nos parâmetros de qualidade da classificação apenas para o modelo baseado em redes neurais, (ELM). Possivelmente, o aprimoramento do modelo baseado em redes neurais

ocorre pela eliminação de variáveis irrelevantes, reduzindo o número de comprimentos de onda altamente correlacionados e que trazem ruído para os modelos de classificação (WEHRENS, 2011; HAPFELMEIER; ULM, 2013). Nesse caso, a seleção de variáveis pode ser um método vantajoso por identificar as variáveis que são mais relevantes para o procedimento de classificação, tornando seu processamento e execução mais rápidos, principalmente quando se utiliza um grande conjunto de dados (PIIROINEN et al., 2017).

Para os modelos SVM, BGLM, PLS e HDDAM, a classificação baseada no modelo reduzido (apenas comprimentos de onda mais importantes) não promoveu melhoria na acurácia, pois afetou o desempenho da classificação, com o aumento do número de classificações errôneas e redução da reprodutibilidade. Na seleção de variáveis pelo algoritmo *random forest* o conjunto de variáveis é reduzido a cada etapa, pela eliminação das variáveis irrelevantes e redundantes. Essas etapas têm como objetivo, além de identificar as variáveis relevantes, melhorar ou mesmo manter a acurácia dos modelos sem qualquer tipo de seleção de dados. Porém, quando muitas variáveis são excluídas pode-se perder informações que contribuam para a classificação, prejudicando o desempenho dos modelos (HAPFELMEIER; ULM, 2013).

Os modelos estatísticos utilizados para a diferenciação entre acessos infectados e não infectados com couro de sapo foram bastante eficientes para encontrar uma relação entre os comprimentos de onda do NIR com a classe correspondente sob análise, e de modo geral, algumas faixas espectrais foram mais importantes nesta diagnose. A faixa de 1490 a 1516 nm foi a mais importante para a maioria dos modelos (cinco dos seis modelos testados) sendo que, dois deles apresentaram acurácia de 99% utilizando o conjunto completo de comprimentos de onda. Atribuída ao estiramento das ligações OH ou NH, esta faixa é sensível às variações de concentração de carboidratos e aminoácidos, que possivelmente estão relacionadas com a resposta fisiológica ao estresse (WORKMAN JR; WEYER, 2012). Esses resultados são concordantes com o estudo de Xu et al. (2007), em que a severidade da infestação causada por insetos nas folhas do tomateiro foi altamente correlacionada com comprimentos de onda individuais na faixa de 1450 e 1900 nm. Essas bandas contribuíram para uma boa performance de predição, realizada por meio de análise de regressão ($R^2 = 0,98$).

Ao utilizarem imagens hiperespectrais para análise da ferrugem amarela do trigo, Huang et al. (2007) também observaram que a região do NIR, de 1050 a 2500 nm também foi efetiva para detectar a doença ($R^2 = 0,91$ e $0,97$, nas avaliações realizadas em 2002 e 2003). Além da detecção de doenças, o modelo PLS já foi empregado com sucesso na identificação da resistência à doença de Fiji em cana de açúcar (PURCELL et al. 2009), considerando $R^2 = 0,98$ para a correta predição da doença, tendo a faixa espectral entre 1854 a 1921nm, como a mais correlacionada a resistência à doença.

Comparação entre os modelos de classificação

A acurácia da classificação dos acessos de mandioca quanto à infecção do couro de sapo foi de 99% com uso dos modelos SVM e BGLM. Este resultado é bastante positivo considerando a complexidade da diagnose desta doença da mandioca, que está associada à presença de vírus da família Reoviridae (dsRNA) (CALVERT et al. 2008) e fitoplasmas do grupo III-L (OLIVEIRA et al., 2014) e grupo III-A (SOUZA et al., 2014), ou mesmo a co-infecção por vírus e fitoplasmas (SOUZA et al., 2014). Assim, a detecção precoce por meio da identificação dos agentes causais é bastante complexa, envolvendo mais de uma análise, por ser atribuída a diferentes patógenos.

As taxas de detecção do couro de sapo em plantas sintomáticas com o uso de *nested PCR* para análise de fitoplasma foi de 89,7% em folhas (ALVAREZ et al., 2009). Por outro lado, Souza et al. (2014) relataram 100% de detecção do fitoplasma quando a análise foi realizada em diferentes tecidos infectados da planta. No entanto, quando foi realizada em tecidos individuais, folha, raiz ou haste, as taxas de detecção de fitoplasma foram de 21,9 a 87,5% e de 50% para o vírus. A detecção por técnicas baseadas em DNA, apesar de precisa e usual, apresenta a grande desvantagem de ser mais cara e demorada. Uma vez que o modelo de classificação está bem estabelecido, as técnicas com o uso do NIR tornam-se muito rápidas e diretas, além do potencial para uso em campo.

Mesmo em técnicas mais precisas como o PCR em tempo real, a técnica de infravermelho médio obteve taxa de detecção semelhante como no estudo de Hawkins et al. (2010) que conseguiram classificar com 95% de precisão o Huanglongbing em *Citrus*, cujo agente causal é um fitoplasma.

O método SVM tem sido aplicado com bastante sucesso em outras culturas para diagnose e classificação de diversas doenças de plantas. Rumpf et al. (2010) demonstraram a potencialidade de uso do V-NIR para: i) discriminar folhas de beterraba infectadas e não infectadas, ii) diferenciar doenças causadas por diferentes agentes patogênicos como *Cercospora beticola*, *Erysiphe betae* e *Uromyces betae*, e iii) identificar a presença de doenças mesmo antes do aparecimento de sintomas específicos, com base no método SVM. De acordo com Rumpf et al. (2010) o SVM provou ser uma poderosa ferramenta para classificação automática dos acessos infectados de beterraba, sendo que as suas principais vantagens estão relacionadas à sua capacidade de generalização, que é obtida buscando o hiperplano máximo para separação e aplicação de funções discriminantes não-lineares.

Por outro lado, apesar do BGML ser menos usual do que a SVM na classificação com o uso de imagens, estes dois modelos não diferiram na classificação dos acessos infectados com couro de sapo da mandioca e alcançaram uma acurácia elevada e alta reprodutibilidade. O modelo BGLM é um método linear generalizado com uma abordagem bayesiana, que difere da inferência tradicional ao assumir que os dados são fixos e os parâmetros do modelo são aleatórios (SMITH, 1973). Para o método linear generalizado, a classificação de plantas infectadas com dois vírus diferentes e um patógeno fúngico em cucurbitáceas foi possível com 85 a 100% de acurácia, com o uso de uma combinação de imagem térmica, fluorescência de clorofila e imagem hiperespectral (BERDUGO et al., 2014) demonstrando a eficiência da técnica.

O ELM apesar do sucesso de classificação em outros estudos (LORENTE et al., 2011; XIE et al, 2015), para a predição do couro de sapo, seu desempenho foi o menor entre os modelos testados e obteve uma taxa de 81,17% de acurácia para o modelo completo. O ELM é um método de redes neurais artificiais de uma única camada oculta, que escolhe os pesos de entrada e a tendência da camada oculta aleatoriamente e o peso de saída é escolhido analiticamente (HUANG; ZHU; SIEW, 2006; XIE et al, 2015). Este método possui uma boa generalização e é extremamente rápido, porém apresenta alguns inconvenientes, por se basear na minimização do risco empírico e calcular diretamente as soluções de mínimos quadrados da norma mínima. Estes inconvenientes resultam em estimativas

menos robustas, devido a *outliers* ou quando existe heteroscedasticidade nos dados (DENG; ZHENG; CHEN, 2009).

No caso do modelo HDDA, seu objetivo é corrigir problemas de multidimensionalidade dos dados, comuns aos dados espectrais (BOUVEYRON; GIRARD; SCHMID, 2007). Diferente dos métodos frequentemente utilizados para a redução da dimensionalidade, o HDDA não reduz a dimensão ao modelar os dados, dessa forma, é possível utilizar toda a informação de alta dimensão na classificação (JACQUES et al. 2010). No entanto, apesar do modelo ser adequado a dados de alta dimensão, o HDDA obteve um dos menores desempenhos na predição do couro de sapo. Resultado semelhante foi encontrado com o uso de três conjuntos de dados hiperespectrais em comparação com outros oito diferentes modelos de classificação (ZULLO; FAUVEL; FERRATY, 2017).

O PLS não foi o método com uma melhor classificação para esse conjunto de dados, porém sua acurácia foi apurada (91,64%), assim como o PRANDF (93,39%). O ótimo desempenho de classificação já foi demonstrado em diversos estudos com detecção de doenças para o PLS (PURCELL, et al. 2009; JONES et al, 2010; SERRANT; CESARE; BONIFAZI, 2013) e para o método de *random forest* (ISMAIL; MUTANGA, 2011; DHAU et al., 2018; KNAUER et al., 2017). O PLS é o método mais utilizado na análise de dados multivariados na quimiometria, e consiste em um modelo de regressão linear, insensível às variáveis colineares, que aceita um grande número de variáveis (WEHRENS, 2011). Já o PRAND, é um método não-linear, baseado uma combinação de preditores de árvores, de modo que cada divisão depende dos valores de um vetor aleatório. Neste modelo, usa uma seleção aleatória de recursos para dividir cada nó, reduzindo as taxas de erro (BREIMAN, 2001).

A performance dos modelos é dependente do conjunto de dados e da natureza do problema de classificação, dessa forma, o desempenho dos modelos difere de cada conjunto de dados e suas características (JACQUES et al., 2010; WEHRENS, 2011). Portanto, para cada tipo de classificação e composição espectral, diferentes metodologias podem ser necessárias para encontrar o melhor modelo possível para um resultado de classificação confiável.

Perspectivas de aplicação do NIR para diagnose do couro de sapo da mandioca

As dificuldades de identificação visual das plantas sintomáticas, bem como o elevado custo e a generalizada ocorrência de falsos positivos das análises moleculares demandam o desenvolvimento de técnicas mais avançadas, rápidas e eficientes para a diagnose do couro de sapo. De fato, alguns estudos demonstraram que os métodos espectrométricos são mais eficazes, mais precisos e menos demorados em comparação aos métodos de análise de DNA (MENESATTI et al., 2010; TOMKIEWICZ; PISKIER, 2012).

Como agravante, o couro de sapo da mandioca vem se espalhando com grande rapidez nas principais regiões produtoras da mandioca no Brasil e tem causado sérios prejuízos aos agricultores. Portanto, os resultados da classificação dos acessos infectados e não infectados com couro de sapo, baseado nos espectros NIR, trazem um impacto bastante positivo nesta direção, considerando acurácia de mais de cerca de 99% para os métodos SVM e BGLM. A rápida diagnose da presença do couro de sapo da mandioca, sobretudo no início do desenvolvimento da lavoura ou mesmo na seleção das plantas matrizes para serem utilizadas nos novos plantios, pode ajudar a reduzir a disseminação e os impactos negativos da doença.

CONCLUSÃO

Neste estudo, o uso da espectrometria baseada em NIR apresentou-se acurada na classificação de acessos infectados por couro de sapo em mandioca. Os modelos preditivos demonstraram alta eficiência na distinção de acessos saudáveis e infectados, destacando-se o SVM e o BGLM com o conjunto completo de comprimentos de onda do NIR. Esses modelos tiveram desempenho de 99,07 e 98,92% de acurácia e sua reprodutibilidade foi de 0,98 com concordância quase perfeita.

A seleção de comprimentos de onda do NIR não mostrou ser eficiente para melhorar a qualidade de classificação dos acessos de mandioca infectados com couro de sapo, pois reduziu os valores de acurácia e de Kappa, sendo, portanto, não recomendados para diagnose desta doença.

Assim, a método proposta neste estudo, mostrou ser uma alternativa com grande potencial de aplicação prática na classificação do couro de sapo em mandioca, de modo rápido, com custo baixo e com menor tempo de preparo e avaliação das amostras, em comparação aos métodos tradicionais.

REFERÊNCIAS

ABDULRIDHA, J.; EHSANI, R.; CASTRO, A. Detection and differentiation between laurel wilt disease, *Phytophthora* disease, and salinity damage using a hyperspectral sensing technique. **Agriculture**, v. 6, n. 4, p. 56, 2016.

ALLEM, A. C. the origin of *Manihot esculenta* crantz (Euphorbiaceae). **Genetic resources and crop Evolution**, v. 41, n. 3, p. 133-150, 1994.

ALVAREZ, E.; MEJÍA, J.F.; LLANO, G. A.; LOKE, J. B.; CALARI, A.; DUDUK, B.; BERTACCINI, A. Characterization of a phytoplasma associated with frogskin disease in cassava. **Plant Disease**, v. 93, p.1139-1145, 2009.

BERDUGO, C. A.; ZITO, R.; PAULUS, S.; MAHLEIN, A. K. Fusion of sensor data for the detection and differentiation of plant diseases in cucumber. **Plant pathology**, v. 63, n. 6, p. 1344-1356, 2014.

BERTACCINI, A.; DUDUK, B.; PALTRINIERI, S.; CONTALDO, N. Phytoplasmas and phytoplasma diseases: a severe threat to agriculture. **American Journal of Plant Sciences**, v. 5, n. 12, p. 1763, 2014.

BERTAMINI, M.; NEDUNCHEZHIAN, N.; TOMASI, F.; GRANDO, M.S. Phytoplasma [Stolbur-subgroup (Bois Noir-BN)] infection inhibits photosynthetic pigments, ribulose-1, 5-bisphosphate carboxylase and photosynthetic activities in field grown grapevine (*Vitis vinifera* L. cv. Chardonnay) leaves. **Physiological and Molecular Plant Pathology**, v. 61, n. 6, p. 357-366, 2002.

BOUVEYRON, C.; GIRARD, S.; SCHMID, C. High-dimensional discriminant analysis. **Communications in Statistics—Theory and Methods**, v. 36, n. 14, p. 2607-2623, 2007.

BREIMAN, L. Random forests. **Machine learning**, v. 45, n. 1, p. 5-32, 2001.

CALVERT, L. A.; CUERVO, M.; LOZANO, I.; VILLAREAL, N.; ARROYAVE, J. Identification of three strains of a virus associated with cassava plants affected by frogskin disease. **Journal of Phytopathology**, v.156, p. 647-653, 2008.

CALVERT, L. A.; THRESH, J. M. The viruses and virus diseases of cassava. In: HILLOCKS, R.J, THRESH, J.M, BELLOLLI, A.C, eds. **Cassava: Biology, Production and Utilization**, CAB International., Wallingford, 2002.

CARDOZO TÉLLEZ, L.; PARDO, J. M.; ZACHER, M.; TORRES, A.; ÁLVAREZ, E. First Report of a 16SrIII phytoplasma associated with frog skin disease in cassava (*Manihot esculenta*) in Paraguay. **Plant Disease**, v. 100, n. 7, p. 1492-1492, 2016.

CARVAJAL-YEPES, M.; OLAYA, C.; LOZANO, I.; CUERVO, M.; CASTANO, M.; CUELLAR, W. J. Unraveling complex viral infections in cassava (*Manihot esculenta* Crantz) from Colombia. **Virus research**, v. 186, p. 76-86, 2014.

DAVIS, R.E.; JOMANTIENE, R.; ZHAO, Y.; DALLY, E.L. folate biosynthesis pseudogenes, ψ folP and ψ folK , and an O-Sialoglycoprotein endopeptidase gene homolog in the phytoplasma genome. **DNA and Cell Biology**, v. 22, n.11, p. 697-706, 2003.

DENG, W.; ZHENG, Q.; CHEN, L. Regularized extreme learning machine. In: **Computational Intelligence and Data Mining, 2009. CIDM'09. IEEE Symposium on**. IEEE, 2009. p. 389-395.

DHAU, I.; ADAM, E.; MUTANGA, O.; AYISI, K. K. Detecting the severity of maize streak virus infestations in maize crop using in situ hyperspectral data. **Transactions of the Royal Society of South Africa**, v. 73, n. 1, p. 8-15, 2018.

FAO - Food and Agriculture Organization of the United Nations. FAOSTAT database. Disponível em <<http://faostat3.fao.org/analysis/Q/QC/E>>. Acesso em: 20 de setembro de 2017.

FEO, L. D.; ZANINI, A.; RODRÍGUEZ PARDINA, P.; CUERVO, M.; CARVAJAL-YEPES, M.; CUELLAR, W. J. First report of *Cassava common mosaic virus* and *Cassava frog skin*-associated virus infecting cassava in Argentina. **Plant Disease**, v. 99, n. 5, p. 733, 2015.

HAPFELMEIER, Alexander; ULM, Kurt. A new variable selection approach using random forests. **Computational Statistics & Data Analysis**, v. 60, p. 50-69, 2013.

HAWKINS, S. A.; PARK, B.; POOLE, G. H.; GOTTWALD, T.; WINDHAM, W. R.; LAWRENCE, K. C. Detection of citrus huanglongbing by Fourier transform infrared–attenuated total reflection spectroscopy. **Applied spectroscopy**, v. 64, n. 1, p. 100-103, 2010.

HUANG, G. B.; ZHU, Q. Y.; SIEW, C. K. Extreme learning machine: theory and applications. **Neurocomputing**, v. 70, n. 1, p. 489-501, 2006.

HUANG, W.; LAMB, D. W.; NIU, Z.; ZHANG, Y.; LIU, L.; WANG, J. Identification of yellow rust in wheat using in-situ spectral reflectance measurements and airborne hyperspectral imaging. **Precision Agriculture**, v. 8, n. 4-5, p. 187-197, 2007.

HUNSCHE, M.; BUERLING, K.; NOGA, G. Spectral and time-resolved fluorescence signature of four weed species as affected by selected herbicides. **Pestic. Biochem. Physiol**, v. 101, n. 1, p. 39-47, 2011.

ISMAIL, R.; MUTANGA, O. Discriminating the early stages of *Sirex noctilio* infestation using classification tree ensembles and shortwave infrared bands. **International journal of remote sensing**, v. 32, n. 15, p. 4249-4266, 2011.

JACQUES, J.; BOUYEYRON, C.; GIRARD, S.; DEVOS, O.; DUPONCHEL, L.; RUCKEBUSCH, C. Gaussian mixture models for the classification of high-dimensional vibrational spectroscopy data. **Journal of Chemometrics**, v. 24, n. 11-12, p. 719-727, 2010.

JAGOUÉIX-EVEILLARD, S.; TARENDAU, F.; GUOLTER, K.; DANET, J.L.; BOVÉ, J.M. GARNIER, M. *Catharanthus roseus* genes regulated differentially by mollicute infections. **Molecular Plant-Microbe Interactions**, v. 14, n. 2, p. 225-233, 2001.

JONES, C. D.; JONES, J. B.; LEE, W. S. Diagnosis of bacterial spot of tomato using spectral signatures. **Computers and Electronics in Agriculture**, v. 74, n. 2, p. 329-335, 2010.

KHALED, A. Y.; AZIZ, S. A.; BEJO, S. K.; NAWI, N. M.; SEMAN, I. A. Spectral features selection and classification of oil palm leaves infected by Basal stem rot (BSR) disease using dielectric spectroscopy. **Computers and Electronics in Agriculture**, v. 144, p. 297-309, 2018.

KHALED, A. Y.; AZIZ, S.; BEJO, S. K.; NAWI, N. M.; SEMAN, I. A.; ONWUDE, D. I. Early detection of diseases in plant tissue using spectroscopy—applications and limitations. **Applied Spectroscopy Reviews**, v. 53, n.1 p. 36-64, 2018.

KNAUER, U.; MATROS, A.; PETROVIC, T.; ZANKER, T.; SCOTT, E. S.; SEIFFERT, U. Improved classification accuracy of powdery mildew infection levels of wine grapes by spatial-spectral analysis of hyperspectral images. **Plant Methods**, v. 13, n. 1, p. 47, 2017.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, p. 159-174, 1977.

LEPKA, P., ATITT, M., MOLL, E. AND SEEMÜLLER, E. (1999) Effect of Phytoplasmal infection on concentration and translocation of carbohydrates and amino acids in periwinkle and tobacco. **Physiological and Molecular Plant Pathology**, 55, 59-68.

LORENTE, D.; ALEIXOS, N.; GÓMEZ-SANCHIS, J.; CUBERO, S.; BLASCO, J. Selection of optimal wavelength features for decay detection in citrus fruit using the ROC curve and neural networks. **Food and Bioprocess Technology**, v. 6, n. 2, p. 530-541, 2013.

MAHLEIN, A.K.; STEINER, U.; DEHNE, H.W.; OERKE, E.C. Spectral signatures of sugar beet leaves for the detection and differentiation of diseases. **Precision Agriculture**, v. 11, n. 4, p. 413-431, 2010.

MARTINELLI, F., SCALENGHE, R., DAVINO, S., PANNO, S., SCUDERI, G., RUISI, P; DAVIS, C. E. Advanced methods of plant disease detection. A review. **Agronomy for Sustainable Development**, v. 35, n. 1, p. 1-25, 2015.

MAUST, B.E.; ESPADAS , F.; TALAVERA, C.; AGUILAR , M.; SANTAMARÍA, J.M; OROPEZA, C. (2003) Changes in carbohydrate metabolism in coconut palms infected with the Lethal Yellowing Phytoplasma. **Phytopathology**, v. 93, n.8, p. 976 -981, 2003.

MENESATTI, P.; ANTONUCCI, F.; PALLOTTINO, F.; ROCCUZZO, G.; ALLEGRA, M.; STAGNO, F.; INTRIGLIOLO, F. Estimation of plant nutritional status by Vis–NIR spectrophotometric analysis on orange leaves *Citrus sinensis* (L) Osbeck cv Tarocco. **Biosystems Engineering**, v. 105, n. 4, p. 448–454, 2010.

MOSHOU, D., BRAVO, C., WEST, J., WAHLEN, S., MCCARTNEY, A.; RAMON, H. Automatic detection of 'yellow rust'in wheat using reflectance measurements and neural networks. **Computers and electronics in agriculture**, v. 44, n. 3, p. 173-188, 2004.

MOSHOU, Dimitrios et al. Automatic detection of 'yellow rust'in wheat using reflectance measurements and neural networks. **Computers and electronics in agriculture**, v. 44, n. 3, p. 173-188, 2004.

MUSETTI, R.; SANITÀ DI TOPPI, L.; MARTINI, M.; FERRINI, F.; LOSCHI, A.; FAVALI, M.A.; OSLER, R. Hydrogen Peroxide Localization and Antioxidant Status in the recovery of Apricot Plants from European Stone Fruit Yellows. **European Journal of Plant Pathology**, v. 112, n.1, p. 53-61, 2005.

OLIVEIRA, S. A. S.; ABREU, E. F. M.; ARAÚJO, T. S.; OLIVEIRA, E. J.; ANDRADE, E. C.; GARCIA, J. M. P.; ALVAREZ, E. First report of a 16SrIII-L phytoplasma associated with frogskin disease in cassava (*Manihot esculenta* Crantz) in Brazil. **Plant disease**, v. 98, n. 1, p. 153-153, 2014.

PIIROINEN, R.; HEISKANEM, J.; MAEDA, E.; VIINIKKA, A.; PELLIKKA, P. Classification of tree species in a diverse african agroforestry landscape using imaging spectroscopy and laser scanning. **Remote Sensing**, v. 9, n. 9, p. 875, 2017.

PINTER, P.J.; HATFIELD, J.L.; SCHEPERS, J.S.; BARNES, E.M.; MORAN, M.S.; DAUGHTRY, C.S.T.; UPCHURCH, D.R. Remote sensing for crop management. **Photogrammetric Engineering & Remote Sensing**, v. 69, n. 6, p. 647-664, 2003.

PURCELL, D. E.; O'SHEA, M. G.; JOHNSON, R. A.; KOKOT, S. Near-infrared spectroscopy for the prediction of disease ratings for Fiji leaf gall in sugarcane clones. **Applied spectroscopy**, v. 63, n. 4, p. 450-457, 2009.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>. 2017.

REDDY, P. P. **Plant protection in tropical root and tuber crops**. Springer India, 2015.

RUMPF, T.; MAHLEIN, A.K.; STEINER, U.; OERKE, E.C.; DEHNE, H.W.; PLUEMER, L. Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance. **Computers and Electronics in Agriculture**, v. 74, n. 1, p. 91-99, 2010.

SANKARAN, S., EHSANI, R., INCH, S. A.; PLOETZ, R. C. Evaluation of visible-near infrared reflectance spectra of avocado leaves as a non-destructive sensing tool for detection of laurel wilt. **Plant disease**, v. 96, n. 11, p. 1683-1689, 2012.

SANKARAN, S.; MISHRA, A.; EHSANI, R.; DAVIS, C. A review of advanced techniques for detecting plant diseases. **Computers and Electronics in Agriculture**, v. 72, n.1, 2010.

SEARS, B. B.; KLOMPARENS, K. L.; WOOD, J. I.; SCHEWE, G. Effect of altered levels of oxygen and carbon dioxide on phytoplasma abundance in *Oenothera* leaf tip cultures. **Physiological and molecular plant pathology**, v. 50, n. 4, p. 275-287, 1997.

SERRANTI, S.; CESARE, D.; BONIFAZI, G. The development of a hyperspectral imaging method for the detection of *Fusarium*-damaged, yellow berry and vitreous Italian durum wheat kernels. **Biosystems engineering**, v. 115, n. 1, p. 20-30, 2013.

SMITH, A. F. M. A general Bayesian linear model. **Journal of the Royal Statistical Society. Series B (Methodological)**, p. 67-75, 1973.

SOUZA, A. N.; SILVA, F. N.; BEDENDO, I. P.; CARVALHO, C. M. A Phytoplasma belonging to a 16SrIII-A subgroup and dsRNA virus associated with cassava frogskin disease in Brazil. **Plant Disease**, v. 98, n. 6, p. 771-779, 2014.

SPILLANE, C.; CURTIS, M. D.; GROSSNIKLAUS, U. Apomixis technology development—virgin births in farmers' fields?. **Nature biotechnology**, v. 22, n. 6, p. 687-691, 2004.

TOMKIEWICZ, D.; PISKIER, T. A plant based sensing method for nutrition stress monitoring. **Precision agriculture**, v. 13, n. 3, p. 370-383, 2012.

WORKMAN JR, J.; WEYER, L. **Practical guide and spectral atlas for interpretive near-infrared spectroscopy**. CRC Press, 2012.

XIE, C.; SHAO, Y.; LI, X.; HE, Y. Detection of early blight and late blight diseases on tomato leaves using hyperspectral imaging. **Scientific reports**, v. 5, p. 16564, 2015.

XU, H. R.; YING, Y. B.; FU, X. P.; ZHU, S. P. Near-infrared spectroscopy in detecting leaf miner damage on tomato leaf. **Biosystems Engineering**, v. 96, n. 4, p. 447-454, 2007.

ZULLO, A.; FAUVEL, M.; FERRATY, F. Experimental comparison of functional and multivariate spectral-based supervised classification methods in hyperspectral image. **Journal of Applied Statistics**, p. 1-19, 2017.

CONSIDERAÇÕES FINAIS

A espectrometria do infravermelho próximo mostrou-se um método eficiente para a classificação de acessos infectados com o couro de sapo da mandioca. O desenvolvimento dessa técnica representa um grande avanço em relação aos procedimentos de diagnose utilizados até o momento, sobretudo por ser uma ferramenta precisa, reprodutível, rápida e de baixo custo.

Os modelos de classificação analisados foram capazes de correlacionar com precisão as respostas espectrais do NIR com diferenças fisiológicas e bioquímicas (ainda não conhecidas) que fazem com que haja uma boa capacidade de diferenciação dos acessos infectados com couro de sapo, dos acessos sadios (não infectados).

Entre os modelos de classificação avaliados, o SVM e o BGLM se destacaram por apresentar a melhor acurácia (cerca de 99%) e concordância (0,98 índice Kappa) na classificação dos acessos infectados e sadios. Por outro lado, a utilização da seleção de comprimentos de onda no NIR mais informativos não resultou em melhoria ou manutenção da qualidade da classificação em comparação com o modelo completo (todos os comprimentos de onda), não sendo indicado para uso de forma rotineira em atividades de diagnose.

Além de todas as vantagens supracitadas a técnica de NIR pode ser estendida para equipamentos portáteis, permitindo um monitoramento prático em campo e identificação de plantas infectadas. Além disso, a espectrometria de infravermelho é uma ferramenta versátil e outras análises podem ser desenvolvidas, como mensuração de compostos e deficiências nutricionais, e serem avaliadas concomitantemente com o diagnóstico do couro de sapo.