

**IDENTIFICAÇÃO DE MILHO TRANSGÊNICO POR ANÁLISE DISCRIMINANTE
LINEAR E ESPECTROSCOPIA RAMAN**

S. R. Dib^{1,*}, T. V. Silva², J. A. Gomes Neto¹, L. J. M. Guimarães³, E. J. Ferreira⁴, E. C. Ferreira¹

¹ Universidade Estadual Paulista, Av. Prof. Francisco Degni, 55, 14800-900, Araraquara, SP

² Universidade Estadual de Londrina, Rodovia Celso Garcia, Km 445, 86057-970 Londrina, PR

³ Embrapa Milho e Sorgo, Rod MG 424, Km 45, 35701-970, Sete Lagos, MG

⁴ Embrapa Instrumentação, Rua XV de novembro, 1452, 13560-970, São Carlos, SP

* Autor correspondente, e-mail: rdib.samia@gmail.com

Resumo: O milho é um dos principais ingredientes energéticos utilizado na nutrição animal e humana, por ser uma fonte barata de carboidratos, proteínas e óleos. Com a finalidade de produção de novas espécies com maior resistência às pragas, insetos e herbicidas, a matriz genética tem sido modificada em laboratório, resultando em produtos transgênicos. A biotecnologia moderna tem possibilitado grande número de genes passíveis de utilização na melhoria genética do milho. No entanto, ainda existem dúvidas de diversas naturezas com relação ao emprego e segurança dessa tecnologia. Nesse contexto, faz-se importante a implementação de métodos para identificação de sementes de milho produzidas e ofertadas no mercado. A maioria dos métodos para a detecção de organismos geneticamente modificados são baseados em reação em cadeia da polimerase (PCR), um método complexo e de alto custo que demanda a extração do material genético para análise. Técnicas espectrais não destrutivas têm demonstrado potencial para a referida finalidade. Neste sentido, a espectroscopia Raman e ferramentas quimiométricas foram utilizadas para o desenvolvimento de um método de classificação de milho transgênico e convencional. Os resultados obtidos mostraram que proteínas e carboidratos são importantes à discriminação de classes. O modelo baseado em LDA (Linear Discriminant Analysis), apresentou exatidão satisfatória, com baixo erro e alta precisão.

Palavras-chave: milho transgênico, classificação, espectroscopia raman

**IDENTIFICATION TRANGENIC CORN BY LINEAR DISCRIMINANT ANALYSIS AND
RAMAN SPECTROSCOPY**

Abstract: Corn is one of the main energy ingredients used in animal and human nutrition as it is a cheap source of carbohydrates, proteins and oils. In order to produce new species with greater resistance to pests, insects and herbicides, the genetic matrix has been modified in the laboratory, resulting in transgenic products. Modern biotechnology has enabled a large number of genes that can be used to improve maize genetics. However, there are still several doubts regarding the use and safety of this technology. In this context, it is important to implement methods to identify maize seeds produced and offered in the market. Most methods for detecting genetically modified organisms are based on polymerase chain reaction (PCR), a complex and costly method that requires extraction of genetic material for analysis. Non-destructive spectral techniques have shown potential for this purpose. In this sense, Raman spectroscopy and chemometric tools were used for the development of a transgenic and conventional corn classification method. The results showed that proteins and carbohydrates are important to class discrimination. The Linear Discriminant Analysis (LDA) model presented satisfactory accuracy, low error and high precision.

Keywords: transgenic corn, classification, Raman spectroscopy

1. Introdução

O milho é um dos principais ingredientes energéticos utilizado na nutrição animal e humana. Trata-se de uma fonte de baixo custo para obtenção de carboidratos, proteínas, ferro e outros minerais (NUSS; TANUMIHARDJO, 2010). Segundo o Departamento de Agricultura dos Estados

Unidos (USDA), a projeção de safra mundial de milho para o ciclo 2019/2020 será de 1,12 bilhão de toneladas (FIESP, 2019). Com a finalidade de produção de novas espécies com maior resistência a pragas, insetos e herbicidas específicos e também com alguma característica nutricional aprimorada ou mesmo ausência de compostos alergênicos, o milho pode ter sua composição genética modificada (CORUJO et al., 2019).

De acordo com o Serviço Internacional para Aquisição de Aplicações Agro-biotecnológicas (ISAA), a aplicação de Organismos Geneticamente Modificados (OGM) tem crescido na agricultura. Vinte e seis países plantaram 191,7 milhões de hectares de culturas biotecnológicas, o que acrescentou 1,9 milhão de hectares ao recorde de plantios em 2017. Os países com a maior área de culturas biotecnológicas plantadas são os EUA, Brasil, Argentina, Canadá e Índia e são responsáveis coletivamente, de 91% da área global total de culturas biotecnológicas (ISAA, 2019).

Apesar das vantagens, o uso de OGM em alimentos não é aceito em alguns países. A caracterização das amostras de milho é, portanto, um desafio científico e tecnológico premente na agroindústria de alimentos. Portanto, uma tecnologia de identificação de produtos geneticamente modificados (GM) e suas matérias-primas é considerada importante para garantir a segurança alimentar dos consumidores.

Atualmente, os métodos de detecção mais comuns são: reação em cadeia da polimerase (PCR) (WANG et al., 2014), transferência de ácido nucleico (manchas de Southern) (SOUTHERN, 1975), espectrometria de massa (OCAÑA et al., 2007C), perfis cromatográficos de proteínas, além da eletroforese (GARCÍA et al., 2009).

Em geral, os referidos métodos apresentam alta especificidade e sensibilidade. Contudo, geralmente possuem alto custo por análise e há morosidade de detecção. Recentemente, técnicas espectrais não destrutivas como a espectroscopia de infravermelho próximo (NIRS) (LUNA, et al., 2010), imagens multiespectrais (LIU et al., 2016), espectroscopia terahertz (LIU et al., 2016) e a espectrometria de emissão óptica com plasma induzido por laser (LIBS) (LIU et al., 2019) associadas a métodos quimiométricos têm demonstrado relativo sucesso na rápida identificação de culturas GM.

A espectroscopia Raman é uma alternativa promissora, capaz de irradiar uma amostra em um comprimento de onda específico (monocromático) com um laser, por exemplo. Uma fração dessa radiação sofre espalhamento inelástico e as pequenas diferenças de energia entre a radiação incidente e espalhada são transformadas em números de onda, dando origem ao denominado espectro Raman capaz de caracterizar algumas moléculas presentes no material (SMITH; DENT, 2005). Assim, a espectroscopia Raman fornece informações relevantes à detecção e diferenciação de amostras biológicas. Recentemente algumas aplicações da técnica associada a modelos multivariados em amostras de milho foram estudadas, tais como para avaliar qualidade de grãos de milho saudáveis ou com a presença de fungos *Aspergillus flavus* (FARBER, 2018) e avaliar a composição de milho amarelo e roxo por meio da presença da amilose/amilopectina (ILDIZ et al., 2019). Nesse escopo de pesquisa, o presente trabalho investigou o potencial da espectroscopia Raman para diferenciar sementes de milhos convencionais e transgênicos.

2. Materiais e Métodos

Amostras de sementes de híbridos experimentais simples da tecnologia VT PRO, possuindo três transgenes cry1A.105 + cry2Ab2 e c4epsps, responsáveis pela resistência a lagartas e tolerância ao herbicida glifosato foram fornecidas pela Embrapa Milho e Sorgo de Sete Lagoas - MG. O conjunto de amostras compreende 14 amostras convencionais e 14 geneticamente modificadas (pareadas).

Inicialmente, foram tomadas aleatoriamente 10 sementes de cada amostra, as quais foram individualmente homogeneizadas em um moinho criogênico. Após a moagem, as sementes foram convertidas em pastilhas e os espectros Raman foram adquiridos utilizando um espectrômetro (iRaman BWS415-785H (B&W Tek)) com fonte de excitação de 785 nm (potência < 350 mW) e resolução espectral de 3,5 cm⁻¹.

Antes da aquisição espectral foram otimizados os parâmetros instrumentais: tempo de integração, potência do laser (%) e o número de espectros a ser tomado em cada ponto de medida (n). Após a otimização, os espectros foram capturados em 5 diferentes regiões das pastilhas, sendo $n=50$, tempo de integração 10s e potência do laser 90%. Os espectros foram pré-processados individualmente no MATLAB® (R2010a) com PLS Toolbox. Aplicou-se o filtro Savitzky-Golay, utilizando um polinômio de 2ª ordem e uma janela de 15 pontos. Em seguida foi realizada a correção da linha de base utilizando a função *Whittaker filter* e por último os espectros foram normalizados pela banda 1456 cm^{-1} .

Para o propósito de comparação das médias, *teste t de student* foi aplicado nos valores de intensidades máximas de cada banda espectral identificada entre os pares transgênicos e convencionais. Para a construção do modelo de predição para sementes de milho transgênico e convencionais por LDA (Linear Discriminant Analysis), as amostras do conjunto de calibração foram compostas de 80% e de validação 20% do total das amostras.

Considerando apenas as amostras do conjunto de calibração e a partir das variáveis selecionadas pelo *teste t de student*, foi aplicada uma nova seleção de variáveis utilizando o algoritmo computacional Wrapper, auxiliado pela seleção sequencial para trás (Backward Selection) para selecionar as variáveis a partir do conjunto das 27 variáveis, previamente selecionadas pelo teste estatístico. Tanto o modelo de classificação, quanto a metaheurística utilizada para a seleção de variáveis foram empregadas por meio do software WEKA.

3. Resultados e Discussão

Os espectros das 280 pastilhas sendo estas 140 referentes a classe dos grãos transgênicos e 140 referentes aos convencionais, foram adquiridos utilizando as condições otimizadas, o pré-processamento descrito foi aplicado e as bandas presentes nos espectros Raman das amostras de milho foram identificadas.

Na referida região, os espectros das duas classes apresentaram aspectos semelhantes em relação aos compostos químicos presentes. As atribuições das bandas foram realizadas de acordo com a literatura (ILDIZ et al., 2019, FARBER, 2018) majoritariamente os espectros são caracterizados por bandas de espalhamento associadas a presença de carboidratos ($302, 357, 399, 476, 575, 712, 771, 852, 934, 1000, 1026, 1079, 1121, 1152, 1259, 1333, 1380, 1395$ e 1458 cm^{-1}), proteínas (523 e 1655 cm^{-1}), lignina (1593 e 1628 cm^{-1}) e carotenoides ($1000, 1518$ e 1550 cm^{-1}).

Após o estudo sistemático das atribuições das bandas, observou-se a presença das mesmas em ambas as classes, assim a busca de variáveis para a separação das classes foram feitas por meio da análise dos valores de intensidade máxima para cada banda identificada. Neste sentido, as intensidades foram selecionadas e foram feitos então o cálculo das médias e desvio padrão para cada amostra. Visando melhor avaliar as diferenças entre as médias aplicou-se o teste t de student com nível de significância de 5%. Os resultados obtidos mostraram que para todas as bandas identificadas pelo menos um par (transgênico e convencional) resultou em diferença estatística significativa. No entanto, algumas bandas como 302 cm^{-1} (amilose, amilopectina, modos esqueléticos do anel piranose), 476 cm^{-1} (deformações CCO e CCC), 1043 cm^{-1} ($-\text{C}=\text{C}-$ no plano) e 1121 cm^{-1} ($\nu(\text{C}-\text{O}) + \nu(\text{C}-\text{C}) + \delta(\text{C}-\text{O}-\text{H})$) foram as bandas que apresentaram maior probabilidade de distinção entre todos os pares (transgênicos e convencionais) de acordo com a diferença estatística entre as médias.

Em seguida, foi aplicada a abordagem Wrapper com *Backward Selection* (conjunto das amostras de calibração). A Tabela 1 descreve as variáveis selecionadas.

De acordo com a Tabela 1, apenas 10 variáveis foram empregadas para o desenvolvimento do modelo. Vale destacar que as atribuições das bandas selecionadas são referentes a presença de proteínas e carboidratos, sendo tais compostos importantes para diferenciar as classes de milho transgênico e milho convencional.

Assim, a partir das variáveis selecionadas, o modelo baseado na análise de discriminantes lineares (LDA) para predição das classes foi desenvolvido. O modelo proposto foi avaliado por meio

do conjunto de validação e as respectivas figuras de mérito encontram-se na Tabela 2.

Tabela 1. Variáveis selecionadas por *Wrapper* para a construção do modelo de predição de sementes de milho transgênico e convencional.

Variáveis selecionadas (cm ⁻¹)	Modo vibracional	Atribuições
357	amilose, amilopectina, modos esqueléticos do anel piranose	Carboidratos
476	deformações CCO e CCC	Carboidratos
523	S-S gauche-gauche-trans	Proteínas
619	δ(C-C-O)	Carboidratos
712	δ(C-C-O) deformações esqueléticas do anel glicosídico	Carboidratos
852	δ(C-C-H) + δ(C-O-C), ligação glicosídica, região anomérica	Carboidratos
1026	δ (C-OH)	Carboidratos
1043	δ (C-OH)	Carboidratos
1333	ν(C-O); δ(C-O-H)	Carboidratos
1395	δ(C-C-H)	Carboidratos
1655	ν(C=O) (Amida I)	Proteínas

δ = flexão ou deformação angular; ν = estiramento

Tabela 2. Resultados estatísticos do LDA obtido para o modelo de predição das sementes.

Modelo 1	RMSEC	RMSEP	Taxa de acerto
	0.3189	0.4246	75 %

RMSEC = raiz quadrada do erro médio quadrático de calibração; RMSEP = raiz quadrada do erro médio quadrático de predição

Analisando os resultados obtidos, observa-se que para o modelo desenvolvido o $R^2=0,9$, os valores de RMSEC e RMSEP foram relativamente baixos e a taxa de acerto do modelo para o conjunto de validação foi de 75 %. Dessa maneira o modelo ajustado para a predição de sementes de milho transgênicos e convencionais se enquadraram nos parâmetros de um modelo satisfatório.

4. Conclusões

Os estudos sobre a capacidade da espectroscopia Raman para distinção de sementes de milho transgênico e convencional indicou potencial para técnica. A composição orgânica das amostras de milho sugere que proteínas e carboidratos são importantes para a discriminação das classes. O modelo gerado para predição de sementes de milho transgênico e convencional apresentou parâmetros satisfatórios.

Agradecimentos

Cnpq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), Capes (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) e Embrapa Milho e Sorgo.

Referências

- CORUJO, M. et al. Use of omics analytical methods in the study of genetically modified maize varieties tested in 90 days feeding trials. *Food Chemistry*, v. 292, p.359 - 37, 2019.
- FARBER, C.; Kurouski, D. Detection and Identification of Plant Pathogens on Maize Kernels with a Hand-Held Raman Spectrometer. *Anal. Chem*, v. 90, p. 3009-3012, 2018
- FIESP - Federação das Indústrias do Estado de São Paulo - Nota: 5º levantamento USDA da safra 2019/20 - Setembro/19. Disponível em: < <https://www.fiesp.com.br/indices-pesquisas-e-publicacoes/safra-mundial-de-milho-2/attachment/file-20190917152537-boletimmilhosetembro2019/>>. Acesso em: 18 set. 2019.
- GARCÍA, C. et al. Rapid characterization of (glyphosate tolerant) transgenic and non-transgenic soybeans using chromatographic protein profiles. *Food Chem*, v.113, p.1212-1217, 2009.
- ILDIZ, G. O. et al. A comparative study of the yellow dent and purple flint maize kernel components by Raman spectroscopy and chemometrics. *Journal of Molecular Structure*, v.1184,

p. 246 - 253, 2019.

ISAA- International Service for the Acquisition of Agri-biotech Applications – Disponível em: <
<http://www.isaaa.org/resources/publications/briefs/54/>> Acesso em: 18 set. 2019.

LIU, C. et al. Non-destructive discrimination of conventional and glyphosate-resistant soybean seeds and their hybrid descendants using multispectral imaging and chemometric methods. *J. Agric. Sci.* v.154, p. 1–12, 2016.

LIU, X. et al. Rapid identification of genetically modified maize using laser-induced breakdown spectroscopy. *Food. Bioprocess. Tech.*, v.12, p. 347- 357, 2019.

LIU, W. et al. Application of terahertz spectroscopy imaging for discrimination of transgenic rice seeds with chemometrics. *Food Chemistry.* v. 210, p. 415-421, 2016.

LUNA, A. S. et al. Rapid characterization of transgenic and non-transgenic soybean oils by chemometric methods using NIR spectroscopy. *Spectrochim. Acta A.* v. 100, p. 115-119, 2013.

NUSS, E. T.; TANUMIHARDJO, S. S. Maize: A Paramount Staple Crop in the Context of Global Nutrition. *Comprehensive Reviews in Food Science and Food Safety.* v, 9, p. 417 - 436, 2010.

OCAÑA, M. F. et al. Mass spectrometric detection of CP4 EPSPS in genetically modified soya and maize. *Rapid Commun. Mass Spectrom.*, v. 21, p. 319-328, 2007C.

SMITH, E.; DENT, G. *Modern Raman Spectroscopy - A Practical Approach.* 2005.

SOUTHERN, E. M. Detection of Specific Sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.*, v. 98, p. 503-517, 1975.

WANG, X. et al. Multiplex event- specific qualitative polymerase chain reaction for detecting three trans- genic rice lines and application of a standard plasmid as a quantitative reference molecule. *Analytical Biochemistry*, v. 464, p.1-8, 2014.