THIS IS AN OPEN ACCESS
ARTICLE UNDER THE CC BY
LICENSE CREATIVE COMMON

ISSN 2175-6813

Revista Engenharia na Agricultura

Viçosa, MG, DEA/UFV - DOI: 10.13083/reveng.v30i1.13333 v.30, p.175-194, 2022

AN ALTERNATIVE METHOD FOR IDENTIFICATION OF INDUSTRIAL TOMATO HYBRIDS USING NIRS

Igor Galvão Santana[1], Annelisa Arruda de Brito[1], Fernanda Campos de Oliveira Aguiar[1], Luiz Fernandes Cardoso Campos[1]*, Gilmarcos de Carvalho Corrêa[2], Abadia dos Reis Nascimento[2] & Luís Carlos Cunha Junior[2]

1 - Federal University of Goiás, Agronomy Department, Goiânia, Goiás, Brazil

2 - Federal University of Goiás, Horticulture Sector, Goiânia, Goiás, Brazil

Keywords:
Spectroscopy
PC-LDA
PLS-DA
Solanum lycopersicum L

ABSTRACT

The use of high productive-potential hybrids was one of the reasons for the increase in productivity in the agricultural segment of tomatoes for industrial processing. Thus, among the cultivars available on the market, in general, those that combine greater productivity with quality and that satisfy the needs of industries are chosen. In this context, the objective of this work was to evaluate the best time to implement near-infrared spectroscopy (NIRS) as an alternative method for identifying industrial tomato hybrids. Seeds from the hybrids: CRV8126, H9553, HMX4890, TPX28699 were used and 10 spectra were collected from a set of 20 seeds, from cotyledonary leaves of ten seedlings at 15 days after sowing (DAS) and from the true leaf of ten seedlings at 30 DAS. The results showed that the technique of spectroscopy in the near-infrared range, associated with multivariate analysis, allowed the discrimination of the studied hybrids. The phase in which the best results were obtained in the identification of each hybrid was in the seed, obtaining accuracy values above 90.00% and sensitivity of 100.00%, which proves the use of this instrumental technique on a portable scale for tomato hybrids with a high assertiveness rate.

Palavras-chave:
Espectroscopia
PC-LDA
PLS-DA
Solanum lycopersicum L

MÉTODO ALTERNATIVO PARA IDENTIFICAÇÃO DE HÍBRIDOS DE TOMATE INDUSTRIAL COM O USO DO NIRS

RESUMO

A utilização de híbridos de alto potencial produtivo, foi um dos motivos do aumento da produtividade, no segmento agrícola do tomate para processamento industrial. Assim, dentre os cultivares disponíveis no mercado, em geral, escolhe-se as que combinam maior produtividade com qualidade e que satisfaçam às necessidades das indústrias. Nesse contexto, objetivou-se com este trabalho avaliar o melhor momento de implementação da espectroscopia no infravermelho próximo (*Near Infrared Spectroscopy* – NIRS) como método alternativo de identificação de híbridos de tomate industrial. Foram utilizadas sementes dos híbridos: CRV8126, H9553, HMX4890, TPX28699 e coletados 10 espectros de um conjunto de 20 sementes, das folhas cotiledonares de dez mudas aos 15 dias após o semeio (DAS) e da folha verdadeira de dez mudas aos 30 DAS. Os resultados apresentados demonstram que a técnica da espectroscopia na faixa do infravermelho próximo, associada à análise multivariada, permitiram a discriminação dos híbridos estudados. A fase em que se obteve os melhores resultados na identificação de cada híbrido foi na semente, obtendo valores de acurácia acima de 90.00% e sensibilidade de 100.00%, o que comprova o uso dessa técnica instrumental em escala portátil para híbridos de tomate com uma alta taxa de assertividade.

Received 19 Oct, 2021 • Accepted 13 Dec, 2021 • Published 23 Jun, 2022

Email (corresponding author)*: luizfernandescampos@hotmail.com

SECTION EDITOR IN CHARGE

Gabriel Araújo e Silva Ferraz

INTRODUCTION

Among the states with the highest production of industrial tomatoes, Goiás is the national leader with 1.3 tons produced (IBGE, 2020) and about 80.00% of production is for processing (sauces, extracts, and pulps), by tomato sauce industry and this is one of the reasons that more than half of the tomato agribusinesses are set in the state (MELO & FONTE, 2011). In addition, edaphoclimatic conditions promote the cultivation of tomatoes for industrial processing in the Central-West region of Brazil (VILELA *et al*., 2012; SILVA-JÚNIOR *et al*., 2015).

Currently, the market has several cultivars with different agronomic and industrial characteristics. When choosing a cultivar, soluble solids content, color, leaf coverage, firmness, disease resistance, uniform maturation, retention of the stem in the plant, and mainly productivity are considered (LUZ *et al*., 2016). Thus, among the cultivars available on the market, as a general rule, those that combine higher productivity with quality and that meet the needs of industries are chosen (BOITEUX *et al*., 2012).

The cultivation of tomatoes for processing with a high technological level in the state of Goiás has a high production cost, with inputs, labor, and seeds corresponding on average for 70.00% of the total cost (IFAG, 2020). Thus, profitable production begins with choosing the hybrid to be implemented, as well as the production and delivery of healthy seedlings (DINIZ *et al*., 2006). Nevertheless, it is not uncommon the occurrence of exchange of loads or production of seedlings exchanged due to a lack of contracted input such as seed, generating the delivery of non-contracted hybrids, causing economic damage that is only observed at production. These cases end up being prosecuted and consequently assessed. Such cases are easy to be elucidated because the expertise takes place at the time of fruit production, where economic damage to the producer must have occurred. Thus, it is necessary to study methods and techniques that allow expertise in case of doubts to rapidly identify the seedling production site, since the initial phase of seeds and seedlings corresponds to 11.00% of the total cost of production (HF BRASIL, 2018).

In the field of non-destructive testing, the near-infrared range spectroscopy technique (NIRS) proves to be a fast tool, which allows real-time analysis and reliable results, which reduces the cost and time spent on routine analysis in laboratories (WILLIAM & NORRIS, 2001; MUÑIZ *et al*. 2012).

In vegetable applications, Downes *et al*. (2012), obtained potential results using spectroscopy as a tool to assess the characteristics of pulpwood, as well as to describe the radial variation of cellulose content and yield. Dudley *et al*. (1975) demonstrated the efficiency of NIRS in determining oil, protein, and starch contents. Snel *et al.* (2018) reported that the use of portable NIRS associated with the Partial Least Squares for Discriminant Analysis (PLS-DA) chemometric method is a tool for controlling the timber trade, as they demonstrated that the technique was capable of separating six different visually confused species of *Dalbergia* spp. from several countries.

The cost savings of NIRS measurements related to improving product control and quality are often achieved and can provide significantly faster results compared to traditional laboratory analysis. In discontinuous batch processes, NIRS allows multiple quality estimates to be performed within a production cycle as opposed to a single final batch of analysis (XIAOBO *et al*., 2010).

The objective of this study was to evaluate the best time to implement NIRS, as an alternative method for identifying industrial tomato hybrids, by determining the appropriate wavelength and spectral pre-processing for the creation of a linear discriminant model capable to identify and classify industrial tomato hybrids.

MATERIAL AND METHODS

The study was conducted at the Fruit and Vegetables Technology and Post-Harvest Laboratory and the Vegetable Garden (latitude 16º35’12” S, longitude 49º21’14” W Gr, at 730 m altitude), both located at the School of Agronomy of the Federal University of Goiás, Goiania, State of Goiás.

This experiment used the seeds of the following hybrids: H9553 from Heinz company, HMX4890 from Cargill company, CRV8126, and TPX28699 from CRV plant breeding seed segment from Vivati company.

The seeds were shipped to the Laboratory of Technology and Post-Harvest of Fruits and Vegetables at the School of Agronomy -UFG for spectra acquisition. Ten spectra were collected from a set of 20 seeds of each hybrid using a portable spectrometer F-750 (Felix Instruments, Washington, USA), with optical interactance geometry. For each spectrum, seed homogenization was carried out randomly so that all seeds contributed to the spectrum validation.

The seeds were manually sown in a 162-cell polystyrene tray, where only 72 cells were filled with a substrate made from coconut fiber, rice husk, peat, and vermiculite-coated. Then, they were wrapped with polyethylene film (Stretch) keeping the temperature and relative humidity constant. After seedling emergence, the tray was taken to a greenhouse with controlled micro-sprinkler irrigation, twice a day. Afterward, spectrum was collected from cotyledonary leaves of ten seedlings at 15 days after sowing (DAS) and from the true leaf of ten seedlings at 30 DAS.

Spectral data and reference values were analyzed using the Unscrambler Chemometric program version 10.0.3 (CAMO, Oslo, Norway). It was used in the first step, the principal component analysis (PCA) which aims to correlate a large number of variables using linear combinations to obtain a new set, facilitating its interpretation.

Pre-treatments applicable to spectra include the application of derivatives, multiplicative scattering correction (MSC), and standard normal variation transformation (NSS). The application of the Savitzky-Golay first or second derivative on the original spectral data is a procedure that can highlight spectral shoulders, as well as minimize the effect of spectral baseline slopes (MARETTO, 2011).

Principal Components Linear Discriminant Analysis (PC-LDA) and Partial Least Squares for Discriminant Analysis (PLS-DA) were also performed (NAES *et al*., 2002). The validation of discrimination models was evaluated according to the calculation of figures of merit, as described by Botelho *et al*. (2015) and defined below.

The false-positive rate (FPR) represents the percentage of samples that had false-positive errors and is calculated as the ratio between the absolute number of false positives (FP) and the sum of the absolute number of false-positive (FP) and true negative errors (TN) multiplied by 100, represented by Equation 1:

$$FPR = \left(\frac{FP}{(FP+TN)}\right) * 100 \qquad (1)$$

On the other hand, the false-negative rate (FNR) represents the percentage of samples that had false-negative errors, calculated as the ratio between the absolute number of false negatives (FN) and the sum of the absolute number of false-negative errors (FN) and true positives (TP) multiplied by 100, represented by Equation 2:

$$FNR = \left(\frac{FN}{(FN+TP)}\right) * 100 \qquad (2)$$

Specificity (SPEC) represents the percentage of samples belonging to other classes (y=0) that were identified as belonging to these classes. This figure of merit is calculated by the ratio between the absolute number of true negatives (TN) and the sum of the absolute number of true negatives (TN) and false-positive errors (FP) multiplied by 100, represented by Equation 3:

$$SPEC = \left(\frac{TN}{(TN+FP)}\right) * 100 \qquad (3)$$

Complementarily, sensitivity (SEN) represents the percentage of samples belonging to the discriminated class that was identified as belonging to that class. Therefore, it is calculated as the ratio between the absolute number of true positives (TP) and the sum of the absolute number of true positives (TP) and false-negative errors (FN) multiplied by 100, represented by Equation 4:

$$SEN = \left(\frac{TP}{(TP+FN)}\right) * 100 \qquad (4)$$

RESULTS AND DISCUSSIONS

Seeds

Figure 1 shows the behavior of the spectra of the seeds of the hybrids H9553, CRV8126, HMX4890, TPX28699. The spectra obtained from the seeds of the hybrids presented a similar spectral pattern concerning the absorbance peaks, which was expected, as it is the same species. The highest absorption was observed between the range of 500 to 700 nm, which was attributed to the combination bands of the functional groups -C=O,-NH,-CH, and C-C (MECOZZI *et al.*, 2011).

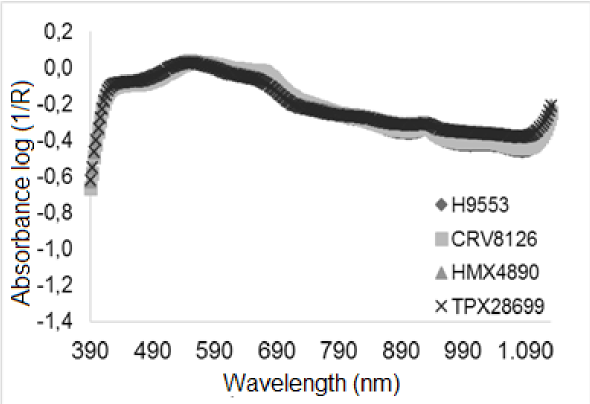

Figure 1. Spectra collected in the form of absorbance for the hybrids HMX4890, H9553, TPX28699, and CRV8126 in the seed phase

In order to eliminate interferences in the collected spectra, it is important to emphasize that the appropriate pre-treatment must be carefully chosen since, in addition to spectral signals, the noise will also be accentuated (BRAGA & POPPI, 2004). The analysis of the pre-processing MSC for the hybrid HMX 4890 (Figure 2C) with the absorbance curve of the hybrid H9553 (Figure 2A) without treatment showed a peak at a wavelength close to 600 nm.

Figure 2B shows the application of the pre-processing in the second derivative with polynomial (5+5) for hybrid CRV8126 and hybrid TPX28699 (Figure 2D) first derivative with polynomial (4+4). It is possible to observe a sharp drop in wavelength close to 700 mm, with the smallest variation in cultivar CRV8126.

It can be seen in Figure 3 that the principal component analysis (PCA) demonstrates the ability of the model to differentiate between hybrids. The PCA model for seeds showed good results in the discrimination of hybrids CRV8126 and H9553 for

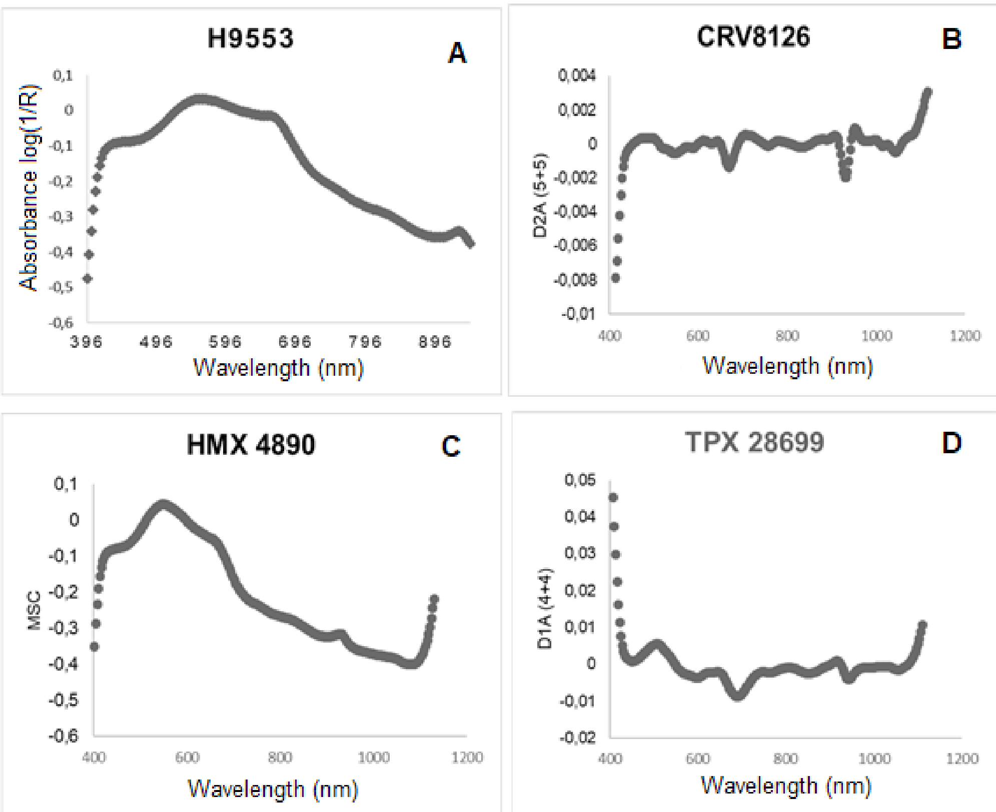

Figure 2. Spectra collected in the form of absorbance for the seeds of hybrids H9553, CRV8126, HMX4890, and TPX28699 after pre-processing

hybrids TPX28699 and HMX4890. Ranking with two major components 64.00% of the data and with one major component 20.00% of the data.

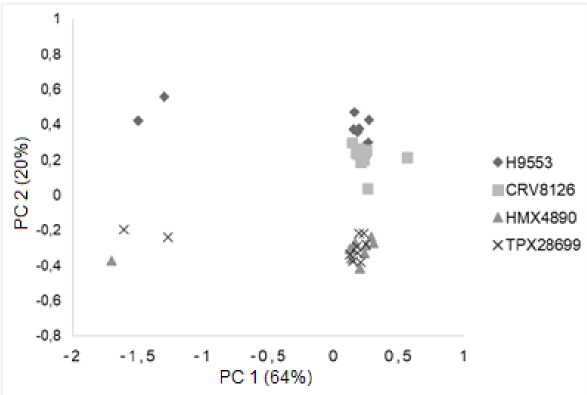

Figure 3. Principal component analysis for 360-1131 nm near-infrared spectra in absorbance for the hybrids HMX4890, H9553, TPX28699, and CRV8126 in the seed phase

Data analysis for seeds using the wavelength in the range 360 to 1,131 nm as a parameter and approaching all tomato hybrids studied in this work showed a better result in the model with absorbance spectrum (Table 1), resulting in a 90.00% accuracy. Thus, it was possible to differentiate 9 out of 10 data for hybrid H9553 and TPX28699, 8 out of 10 data for hybrid HMX4890, and 10 out of 10 data for hybrid CRV8126.

The use of PCA-LDA model with different pre-processing techniques, including NSS, first derivative (4+4), second derivative (7+7), MSC and with the absorbance spectrum, and also with the parameters separation threshold of 0.4; 0.5 and no threshold, good accuracy was found for the untreated model (absorbance spectrum) for the hybrid H9553 (Table 2). Thus, it was possible to predict how many of the 10 data from the hybrid H9553 were classified as hybrid 9553 as well as which of the 40 data from the other hybrids were classified as really too hybrid.

The model that excelled the most for the PCA-LDA was with the absorbance spectrum, presenting principal components equal to 15 and classifying 9 of the 10 data from the hybrid H9553 and 30 of the 30 data from the other hybrids, resulting in an accuracy of 97.50% (Table 2). The analysis of the PLS-DA model showed a value of 92.50% for accuracy in the D1A pre-processing (9+9) and without absorbance spectrum, both identifying 10 of the 10 data for the hybrid H9553 and 27 of the 30 data for the other hybrids. This introduced a model that guarantees good results regarding the discrimination of the hybrid H9553.

Similar to this work, Monferrere *et al.* (2012) found differences between sunflower seeds by oleic content levels: poor oleic (≤ 25.00% $w.w^{-1}$ oleic acid), medium oleic (between 26.00% and 76.00% $w.w^{-1}$), and high oleic (≥77.00% $w.w^{-1}$ oleic acid) using near-infrared diffuse reflectance spectroscopy (NIRDRS) and multivariate data analysis by PCA, LDA, and PLS-DA. The PLS-DA was the method that obtained the best graphical and discriminant analysis capability, due to its ability to recognize the three groups individually.

For data analysis of the hybrid CRV8126 (Table 3), the PCA-LDA model was used with NSS pre-processing, first derivative (7+7), second derivative (9+9), MSC, and with absorbance spectrum.

When evaluating the results, it is possible to observe good accuracy for NSS and D1A pre-processing (7+7), both with 92.50% accuracy (Table 3). It was possible to identify 7 for NSS at wavelength 390-1,053 nm and 8 for D1A (7+7) at wavelength 414-1,116 nm from the 10 data corresponding to the hybrid CRV8126.

For the PLS-DA model, the D2A (5+5) preprocessing resulted in 97.50% accuracy, predicting with high sensitivity and specificity the data that are actually from the CRV8126 hybrid (Table 3).

The pre-treatments or applied transformations MSC or NSS, together or not with the second derivative, corrected the effect of light scattering present in the spectra, mainly caused by the lack of homogeneity of the samples (SILEONI, 2011). This spreading is mainly caused by the particle size or water content. All this was important in terms of the analysis, allowing a good projection with good discrimination of species and stages.

Table 1. Results of PC-LDA calibration models using near-infrared spectroscopy (360-1131 nm) among seeds of each hybrid

Global seed					
Calibration (n)	10	10	10	10	
PP	H9553	CRV8126	HMX4890	TPX28699	AC
ABS	9	10	8	9	90.0
NSS Dt	10	9	4	9	80.0
D1A 9+9	9	8	6	10	82.5
D2A 4+4	9	5	10	10	85.0
MSC	9	10	7	9	87.5

P.P: pre-processing; AC: accuracy; LS: separation threshold; ABS: absorbance spectra; NSS: normal signal standardization; D1A: first derivative with 9-point derivation window; D2A: 4-derivation point second derivative; MSC: multiplicative scattering correction

Table 2. Results of calibration models using near-infrared spectroscopy for tomato seed using hybrid H9553 as the primary objective

H9553 / PC - LDA											
P.P	ST	PC	C.O	H9553 (n=10)	O.H. (n=40)	AC	Sens	Spec	PPV	NVP	AV
ABS	-	15	396-945	9	30	97.5	90.0	100.0	100.0	96.8	90.0
NSS	-	15	390-1,053	8	29	92.5	80.0	96.7	88.9	93.6	76.7
D1A 4+4	-	15	408-1,110	7	28	87.5	70.0	93.3	77.8	90.3	63.3
D2A 7+7	-	15	417-1,110	8	27	87.5	80.0	90.0	72.7	93.1	70.0
MSC	-	15	396-1,131	8	29	92.5	80.0	96.7	88.9	93.6	76.7
H9553 / PLS – DA											
P.P	ST	C.P	WL	H9553 (n=10)	O.H. (n=40)	AC	Sens	Spec	PPV	NVP	AV
ABS	0.5	8	396-945	10	28	95.0	100.0	93.3	83.3	100.0	93.3
ABS	0.4	8	396-945	10	27	92.5	100.0	90.0	76.9	100.0	90.0
NSS	0.5	3	390-1,053	8	28	90.0	80.0	93.3	80.0	93.3	73.3
NSS	0.4	3	390-1,053	9	24	82.5	90.0	80.0	60.0	96.0	70.0
D1A 9+9	0.5	7	417-1,011	10	27	92.5	100.0	90.0	76.9	100.0	90.0
D1A 9+9	0.4	7	417-1,011	10	26	90.0	100.0	86.7	71.4	100.0	86.7
D2A 4+4	0.5	6	411-1,113	8	28	90.0	80.0	93.3	80.0	93.3	73.3
D2A 4+4	0.4	6	411-1,113	9	27	90.0	90.0	90.0	75.0	96.4	80.0
MSC	0.5	3	396-1,131	7	27	85.0	70.0	90.0	70.0	90.0	60.0
MSC	0.4	3	396-1,131	8	25	82.5	80.0	83.3	61.5	92.6	63.3

P.P: pre-processing; LS: separation threshold; PC: principal components; ABS: absorbance spectra; WL: wavelength; OH: other hybrids; AC: accuracy; NSS: normal signal standardization; D1A: first derivative with 9-point derivation window; D2A: 4-derivation point second derivative; Sens: sensitivity; Spec: specificity; PPV: prediction of positive values; NVP: prediction of negative values; AF: ability to avoid failure; MSC: multiplicative scattering correction

For the HMX4890 hybrid, the use of the PCA-LDA model resulted in 90.00% accuracy when applying NSS pre-treatment, being able to distinguish 6 out of the 10 data referring to the HMX4890 hybrid. However, compared with the models of the other hybrids, it was the one with the lowest efficiency in discriminating with a low percentage of sensitivity and with only 60.00% of the ability to avoid failures (Table 4).

In the PLS-DA model, the MSC pre-processing resulted in 100.00% accuracy, predicting with high sensitivity, specificity, and prediction of positive values of data that are actually from the hybrid HMX4890 (Table 4). The PLS method has the advantage of not requiring chromatographic analysis to generate the concentration matrix, thus obtaining the infrared spectrum that may result in fast and economical classification of sunflower seed varieties (MONFERRERE *et al*., 2012).

For the hybrid. TPX28699, good accuracy was found in the PCA-LDA model for the NSS, D2A (4+4), and MSC pre-processing, respectively classifying 9, 7, and 9 of the 10 data of the hybrid TPX28699 and 28 of the 30 data of the others hybrids,

Table 3. Results of calibration models using near-infrared spectroscopy for tomato seed using the hybrid CRV8126 as primary objective

CRV8126 / PC – LDA											
P.P	ST.	PC	WL	CRV8126 (n=10)	D.H (n=40)	AC	Sens	Spec	PPV	NVP	AV
ABS	-	15	396-945	3	28	77.5	30	93.3	60.0	80.0	23.3
NSS	-	15	390-1,053	7	30	92.5	70	100.0	100.0	90.9	70.0
D1A 7+7	-	15	414-1,116	8	29	92.5	80	96.7	88.9	93.6	76.7
D2A 9+9	-	15	423-1,104	7	27	85.0	70	90.0	70.0	90.0	60.0
MSC	-	15	396-1,131	6	28	85.0	60	93.3	75.0	87.5	53.3
CRV8126 / PLS – DA											
P.P	ST.	PC	WL	CRV8126 (n=10)	OH (n=40)	AC	Sens	Spec	PPV	NVP	AV
ABS	0.5	8	396-945	9	29	95.0	90.0	96.7	90.0	96.7	86.7
ABS	0.4	8	396-945	9	28	92.5	90.0	93.3	81.8	96.6	83.3
NSS Dt	0.5	6	510-1,131	9	29	95.0	90.0	96.7	90.0	96.7	86.7
NSS Dt	0.4	6	510-1,131	9	26	87.5	90.0	86.7	69.2	96.3	76.7
D1A 7+7	0.5	2	414-1,116	10	28	95.0	100.0	93.3	83.3	100.0	93.3
D1A 7+7	0.4	2	414-1,116	8	27	87.5	80.0	90.0	72.7	93.1	70.0
D2A 5+5	0.5	4	414-1,116	10	29	97.5	100.0	96.7	90.9	100.0	96.7
D2A 5+5	0.4	4	414-1,116	10	26	90.0	100.0	86.7	71.4	100.0	86.7
MSC	0.5	7	396-1,131	10	27	92.5	100.0	90.0	76.9	100.0	90.0
MSC	0.4	7	396-1,131	9	26	87.5	90.0	86.7	69.2	96.3	76.7

P.P: pre-processing; ST: separation threshold; PC: principal components; ABS: absorbance spectra; WL: wavelength; OH: other hybrids; AC: accuracy; NSS: normal signal standardization; D1A: first derivative with 9-point derivation window; D2A: 4-derivation point second derivative; Sens: sensitivity; Spec: specificity; PPV: prediction of positive values; NVP: prediction of negative values; AV: ability to avoid failure; MSC: multiplicative scattering correction

Table 4. Results of calibration models using near-infrared spectroscopy for tomato seed using the hybrid HMX4890 as primary objective

HMX4890 / PC - LDA											
P.P	ST	PC	C.O	HMX4890 (n=10)	D.H (n=40)	AC	Sens	Spec	PPV	NVP	AV
ABS	-	15	396-945	5	30	87.5	50.0	100.0	100.0	85.7	50.0
NSS	-	15	390-1,053	6	30	90.0	60.0	100.0	100.0	88.2	60.0
D1A 7+7	-	15	414-1,116	5	30	87.5	50.0	100.0	100.0	85.7	50.0
D2A 4+4	-	15	411-1,113	5	30	87.5	50.0	100.0	100.0	85.7	50.0
MSC	-	15	396-1,131	4	30	85.0	40.0	100.0	100.0	83.3	40.0
HMX4890 / PLS - DA											
P.P	ST	C.P	WL	HMX4890 (n=10)	D.H (n=40)	AC	Sens	Spec	PPV	NVP	AV
ABS	0.5	10	396-945	10	24	85.0	100.0	80.0	62.5	100.0	80.0
ABS	0.4	10	396-945	9	28	92.5	90.0	93.3	81.8	96.6	83.3
NSS	0.5	11	390-1,053	10	22	80.0	100.0	73.3	55.6	100.0	73.3
NSS	0.4	11	390-1,053	9	30	97.5	90.0	100.0	100.0	96.8	90.0
D1A 4+4	0.5	8	408-1,110	10	22	80.0	100.0	73.3	55.6	100.0	73.3
D1A 4+4	0.4	8	408-1,110	9	29	95.0	90.0	96.7	90.0	96.7	86.7
D2A 7+7	0.5	10	417-1,110	10	23	82.5	100.0	76.7	58.8	100.0	76.7
D2A 7+7	0.4	10	417-1,110	10	27	92.5	100.0	90.0	76.9	100.0	90.0
MSC	0.5	10	396-1,131	10	21	77.5	100.0	70.0	52.6	100.0	70.0
MSC	0.4	10	396-1,131	10	30	100.0	100.0	100.0	100.0	100.0	100.0

P.P: pre-processing; LS: separation threshold; PC: principal components; ABS: absorbance spectra; WL: wavelength; D.H: other hybrids; AC: accuracy; NSS: normal signal standardization; D1A: first derivative with 9-point derivation window; D2A: 4-derivation point second derivative; Sens: sensitivity; Spec: specificity; PPV: prediction of positive values; NVP: prediction of negative values; AV: ability to avoid failure; MSC: multiplicative scattering correction resulting in an accuracy of 92.50% (Table 5).

In the PLS-DA analysis, the pre-processing with the first derivative with (4+4) polynomial and NSS, obtained the best accuracy values with 100.00%, correctly classifying the data from the hybrid TPX28699 and the other hybrids with high sensitivity and specificity in the 0.5 separation threshold as it was the one that best discriminated (Table 5).

Table 5. Results of calibration models using near-infrared spectroscopy for tomato seed using the hybrid TPX28699 as primary objective

TPX28699 / PC - LDA											
P.P	ST	PC	WL	TPX28699 (n=10)	OH (n=40)	AC	Sens	Spec	PPV	NVP	AV
ABS	-	15	396-945	7	28	87.5	70.0	93.3	77.8	90.3	63.3
NSS	-	15	390-1,053	9	28	92.5	90.0	93.3	81.8	96.6	83.3
D1A 4+4	-	15	408-1,110	8	25	82.5	80.0	83.3	61.5	92.6	63.3
D2A 4+4	-	15	411-1,113	7	30	92.5	70.0	100.0	100.0	90.9	70.0
MSC	-	15	396-1,131	9	28	92.5	90.0	93.3	81.8	96.6	83.3
TPX28699 / PLS - DA											
P.P	ST	C.P	WL	TPX28699 (n=10)	OH (n=40)	AC	Sens	Spec	PPV	NVP	AV
ABS	0.5	10	396-945	9	29	95.0	90.0	96.7	90.0	96.7	86.7
ABS	0.4	10	396-945	9	28	92.5	90.0	93.3	81.8	96.6	83.3
NSS	0.5	11	390-1,053	10	30	100.0	100.0	100.0	100.0	100.0	100.0
NSS	0.4	11	390-1,053	10	28	95.0	100.0	93.3	83.3	100.0	93.3
D1A 4+4	0.5	8	408-1,110	10	30	100.0	100.0	100.0	100.0	100.0	100.0
D1A 4+4	0.4	8	408-1,110	10	26	90.0	100.0	86.7	71.4	100.0	86.7
D2A 7+7	0.5	10	417-1,110	10	29	97.5	100.0	96.7	90.9	100.0	96.7
D2A 7+7	0.4	10	417-1,110	10	27	92.5	100.0	90.0	76.9	100.0	90.0
MSC	0.5	9	396-1,131	10	28	95.0	100.0	93.3	83.3	100.0	93.3
MSC	0.4	9	396-1,131	10	27	92.5	100.0	90.0	76.9	100.0	90.0

P.P: pre-processing; LS: separation threshold; PC: principal components; ABS: absorbance spectra; WL: wavelength; OH: other hybrids; AC: accuracy; NSS: normal signal standardization; D1A: first derivative with 9-point derivation window; D2A: 4-derivation point second derivative; Sens: sensitivity; Spec: specificity; PPV: prediction of positive values; NVP: prediction of negative values; AV: ability to avoid failure; MSC: multiplicative scattering correction

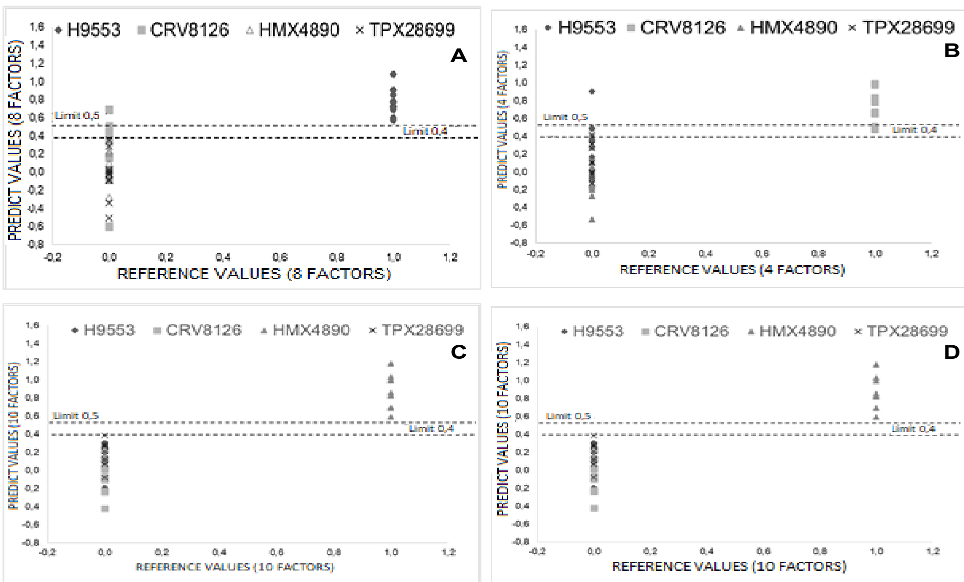

Figure 4. Prediction results of PLS-DA models built using NIR spectra for tomato hybrid seeds. A) hybrid H9953, ABS pre-processing (369-945 nm). B) hybrid CRV8126, pre-processing D2A (5+5) (414-1,116 nm). C) HMX4890 hybrid, MSC pre-processing (396-1,131 nm). C) TPX28699 hybrid, D1A pre-processing (4+4) (408-1110 nm)

It can be seen in Figure 4 that a reference value of 1 was adopted for the hybrid analyzed and 0 for the others. In Figure 4A, the hybrid H9553 was the analyzed hybrid, in Figure 4B the hybrid CRV8126, in Figure 4C the hybrid HMX4890, and Figure 4D the hybrid TPX28699. Then, the separation thresholds equal to 0.4 and 0.5 were tested. It can be observed that the best result for the separation threshold was equal to 0.4 because it was the one that best discriminated the analyzed hybrid from the others. And for the seed model, the hybrid CRV8126 (Figure 4B) stood out, as it discriminated the hybrid with the lowest number of factors in relation to the others.

Cotyledon leaf

Figure 5 shows a peak at wavelength 555 nm. The behavior of the cotyledonary leaf spectra is similar for all hybrids at different wavelengths. In a work with different castor bean cultivars, Santos *et al.* (2014) also found no discrepancy between the spectra, but they were able to identify differences using other tools, such as PCA and PLS-DA, to obtain more information about the spectrum.

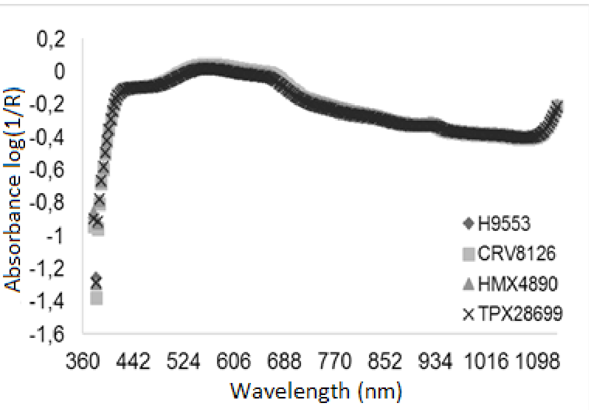

Figure 5. Near-infrared spectra collected as absorbance for the hybrids HMX4890, H9553, TPX28699, and CRV8126 in the cotyledonary leaf phase

The analysis of PCA for cotyledon leaf (Figure 6) in relation to discrimination shows that there was no separation of hybrids, and all indicators that represent each hybrid are grouped in the same region. A similar result was found in the study by Mendes (2014), who managed to group seven Eucalyptus hybrids in the same region in relation to three other species, but there was no differentiated distribution between the hybrids. Ranking with one major component 49.00% of the data and with two major components 20.00% of the data.

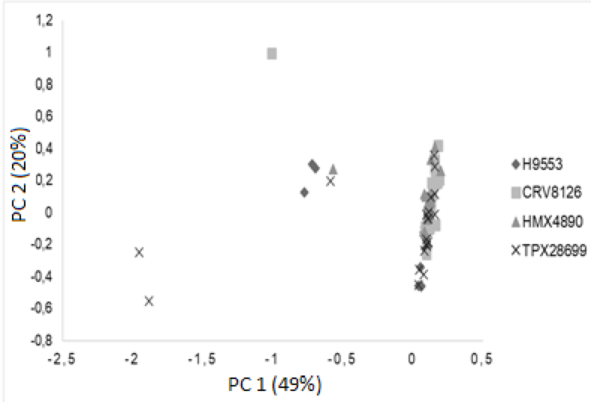

Figure 6. Principal component analysis for near-infrared spectra 384-1131 nm in absorbance for hybrids HMX4890, H9553, TPX28699, and CRV8126 in the cotyledonary leaf phase

In analyzing the cotyledonary leaf data using as a parameter the wavelength in the interval from 384 to 1,131 nm and approaching all tomato hybrids studied in this work, a better result was observed in the spectral treatment with a 5+5 window in the first derivative, resulting in an accuracy of 75.00% (Table 6).

For data analysis of the hybrid H9553 (Table 7) using the PCA-LDA model with an evaluated wavelength from 406 to 1116 nm, the best result was obtained for D1A (5+5) with an accuracy of 97.50%, identifying 19 of the 20 data for hybrid H9553 and 59 out of 60 data for the other hybrids and a false negative value of 98.33%.

For the PLS-DA model (Table 7) of hybrid H9553 cotyledonary leaf calibration, the preprocessing D2A (5+5) with separation threshold of 0.5 and D1A (4+4) with separation threshed 0.4 and 0.5 and without treatment, showed 100.00% accuracy, sensitivity, and specificity, respectively, thus presenting satisfactory results, especially in terms of sensitivity, which allows for the identification of the hybrid object of the model.

In a study carried out by Soares *et al.* (2017), in which six Amazonian wood species were evaluated, the specificity between species was all greater than 90.00% as well as the sensitivity, with three out of the six species reaching 100.00% sensitivity. This probably explains the effectiveness of using the

Table 6. Results of PC-LDA calibration models using near-infrared spectroscopy (384-1131 nm) between cotyledon leaves of each hybrid

Global cotyledon leaf					
Calibration (n)	20	20	20	20	AC
PP	H9553	CRV8126	HMX4890	TPX28699	
ABS	10	15	10	17	65.0
NSS Dt	18	15	6	13	65.0
D1A 5+5	19	13	12	16	75.0
D2A 7+7	14	16	10	16	70.0
MSC	13	14	5	12	55.0

P.P: pre-processing; AC: accuracy; ABS: absorbance spectra; NSS: normal signal standardization; D1A: first derivative with 9-point derivation window; D2A: 4-derivation point second derivative; MSC: multiplicative scattering correction

Table 7. Results of calibration models using near-infrared spectroscopy for tomato cotyledonary leaf using hybrid H9553 as primary objective

H9553 / PCA - LDA											
P.P	ST	PC	WL	H9553 (n=20)	OH (n=60)	AC	Sens	Spec	PPV	NVP	AV
ABS	-	15	387-1,131	8	60	85.0	40.0	100.0	100.0	83.3	40.0
NSS Dt	-	15	387-1,131	15	58	91.3	75.0	96.7	88.2	92.1	71.7
D1A 5+5	-	15	406-1,116	19	59	97.5	95.0	98.3	95.0	98.3	93.3
D2A 7+7	-	15	417-597	15	59	92.5	75.0	98.3	93.8	92.2	73.3
MSC	-	15	381-1,131	12	50	77.5	60.0	83.3	54.6	86.2	43.3
H9553 / PLS - DA											
P.P	ST	C.P	WL	H9553 (n=20)	OH (n=60)	AC	Sens	Spec	PPV	NVP	AV
ABS	0.5	9	387-1,131	20	60	100.0	100.0	100.0	100.0	100.0	100.0
ABS	0.4	9	387-1,131	20	59	98.8	100.0	98.3	95.2	100.0	98.3
NSS	0.5	7	387-780	18	60	97.5	90.0	100.0	100.0	96.8	90.0
NSS	0.4	7	387-780	20	59	98.8	100.0	98.3	95.2	100.0	98.3
D1A 4+4	0.5	6	402-1,119	20	60	100.0	100.0	100.0	100.0	100.0	100.0
D1A 4+4	0.4	6	402-1,119	20	60	100.0	100.0	100.0	100.0	100.0	100.0
D2A 5+5	0.5	7	402-1,005	20	60	100.0	100.0	100.0	100.0	100.0	100.0
D2A 5+5	0.4	7	402-1,005	20	59	98.8	100.0	98.3	95.2	100.0	98.3
MSC	0.5	9	381-1,131	19	60	98.8	95.0	100.0	100.0	98.4	95.0
MSC	0.4	9	381-1,131	19	58	96.3	95.0	96.7	90.5	98.3	91.7

P.P: pre-processing; LS: separation threshold; PC: principal components; ABS: absorbance spectra; WL: wavelength; OH: other hybrids; AC: accuracy; NSS: normal signal standardization; D1A: first derivative with 9-point derivation window; D2A: 4-derivation point second derivative; Sens: sensitivity; Spec: specificity; PPV: prediction of positive values; NVP: prediction of negative values; AV: ability to avoid failure; MSC: multiplicative scattering correction

NIRS as it also achieved results above 90.00% in individuals of the same species, which have more similar individuals and, therefore, there is a greater level of difficulty.

For hybrid CRV8126, good accuracy was found for the model with absorbance spectra, classifying 14 out of 20 data from the hybrid and 58 out of 60 data from the other hybrids, resulting in an accuracy of 90.00% (Table 8). For the PLS-DA model, the models with absorbance spectra and with the D2A pre-processing (9+9), presented an accuracy of 98.75% correctly classifying 20 out of the 20 data for the hybrid CRV8126 and 59 of the 60 data for the other hybrids (Table 8).

For the hybrid HMX4890 and TPX 28699, the use of PCA-LDA model achieved accuracy values less than 85.00% and values less than 80.00% of sensitivity and ability to avoid failures, showed unsatisfactory results, especially in terms of insensitivity, which allows the identification of the hybrid object of the model (Table 9). This result may have occurred due to the similarity between the hybrids, which was also observed by Carvalho *et al*. (2017) in macadamia.

For the PLS-DA model, the NSS and D1A (5+5) pre-processing with 0.5 and 0.4 separation threshold respectively, showed an accuracy of 97.50% correctly classifying 19 of the 20 data for the hybrid CRV8126 and 59 of the 60 data from the other hybrids (Table 9).

Table 8. Results of calibration models using near-infrared spectroscopy for tomato cotyledonary leaf using the hybrid CRV8126 as primary objective

CRV8126 / PCA - LDA											
P.P	ST	PC	WL	CRV8126 (n=20)	OH (n=60)	AC	Sens	Spec	PPV	NVP	AV
ABS	-	15	387-1,131	14	58	90.0	70.0	96.7	87.5	90.6	66.7
NSS Dt	-	15	387-1,131	12	50	77.5	60.0	83.3	54.6	86.2	43.3
D1A 7+7	-	15	402-1,110	11	59	87.5	55.0	98.3	91.7	86.8	53.3
MSC	-	15	381-1,131	12	46	72.5	60.0	76.7	46.2	85.2	36.7
CRV8126 / PLS - DA											
P.P	ST	C.P	WL	CRV8126 (n=20)	OH (n=60)	AC	Sens	Spec	PPV	NVP	AV
ABS	0.5	13	387-1,131	17	38	68.8	85.0	63.3	43.6	92.7	48.3
ABS	0.4	13	387-1,131	19	55	92.5	95.0	91.7	79.2	98.2	86.7
NSS	0.5	11	387-780	18	41	73.8	90.0	68.3	48.7	95.4	58.3
NSS	0.4	11	387-780	18	57	93.8	90.0	95.0	85.7	96.6	85.0
D1A 9+9	0.5	13	410-1,104	18	40	72.5	90.0	66.7	47.4	95.2	56.7
D1A 9+9	0.4	13	410-1,104	20	58	97.5	100,0	96.7	90.9	100.0	96.7
D2A 9+9	0.5	12	420-597	20	39	73.8	100,0	65.0	48.8	100.0	65.0
D2A 9+9	0.4	12	420-597	20	59	98.8	100,0	98.3	95.2	100.0	98.3
MSC	0.5	15	381-1,131	15	39	67.5	75,0	65.0	41.7	88.6	40.0
MSC	0.4	15	381-1,131	15	52	83.8	75,0	86.7	65.2	91.2	61.7

P.P: pre-processing; LS: separation threshold; PC: principal components; ABS: absorbance spectra; WL: wavelength; OH, other hybrids; AC: accuracy; NSS: normal signal standardization; D1A: first derivative with 9-point derivation window; D2A: 4-derivation point second derivative; Sens: sensitivity; Spec: specificity; PPV: prediction of positive values; NVP: prediction of negative values; AV: ability to avoid failure; MSC: multiplicative scattering correction

Table 9. Results of the calibration models using near-infrared spectroscopy for tomato cotyledonary leaf using the hybrid HMX4890 as the primary objective

HMX4890 / PCA - LDA											
P.P	ST	PC	WL	HMX4890 (n=20)	OH (n=60)	AC	Sens	Spec	PPV	NVP	AV
ABS	-	15	387-1,131	7	59	82.5	35.0	98.3	87.5	81.9	33.3
NSS Dt	-	15	387-1,131	9	58	83.8	45.0	96.7	81.8	84.1	41.7
D1A 9+9	-	15	410-1,104	10	57	83.8	50.0	95.0	76.9	85.1	45.0
D2A 7+7	-	15	417-597	8	56	80.0	40.0	93.3	66.7	82.4	33.3
MSC	-	15	381-1,131	7	58	81.3	35.0	96.7	77.8	81.7	31.7
HMX4890 / PLS - DA											
P.P	ST	C.P	WL	HMX4890 (n=20)	OH (n=60)	AC	Sens	Spec	PPV	NVP	AV
ABS	0.5	7	387-1,131	13	59	90.0	65.0	98.3	92.9	89.4	63.3
ABS	0.4	7	387-1,131	13	59	90.0	65.0	98.3	92.9	89.4	63.3
NSS	0.5	13	387-780	19	59	97.5	95.0	98.3	95.0	98.3	93.3
NSS	0.4	13	387-780	19	59	97.5	95.0	98.3	95.0	98.3	93.3
D1A 5+5	0.5	13	406-1,116	19	59	97.5	95.0	98.3	95.0	98.3	93.3
D1A 5+5	0.4	13	406-1,116	19	59	97.5	95.0	98.3	95.0	98.3	93.3
D2A 5+5	0.5	8	402-1,005	17	58	93.8	85.0	96.7	89.5	95.1	81.7
D2A 5+5	0.4	8	402-1,005	17	58	93.8	85.0	96.7	89.5	95.1	81.7
MSC	0.5	9	381-1,131	12	58	87.5	60.0	96.7	85.7	87.9	56.7
MSC	0.4	9	381-1,131	12	58	87.5	60.0	96.7	85.7	87.9	56.7

P.P: pre-processing; LS: separation threshold; PC: principal components; ABS: absorbance spectra; WL: wavelength; OH: other hybrids; AC: accuracy; NSS: normal signal standardization; D1A: first derivative with 9-point derivation window; D2A: 4-derivation point second derivative; Sens: sensitivity; Spec: specificity; PPV: prediction of positive values; NVP: prediction of negative values; AV: ability to avoid failure; MSC: multiplicative scattering correction

The individual PLS-DA calibration models for the TPX28699 hybrid generally showed values of accuracy greater than 86.00% (Table 10). However, in specific models and with unbalanced sample numbers, accuracy is not a very relevant parameter to be observed isolately. Thus, one must look at the sensitivity that allows us to observe the confidence that a positive result for a sample of the labeled class (the hybrid of the specific model) is obtained (MORAIS & LIMA, 2018).

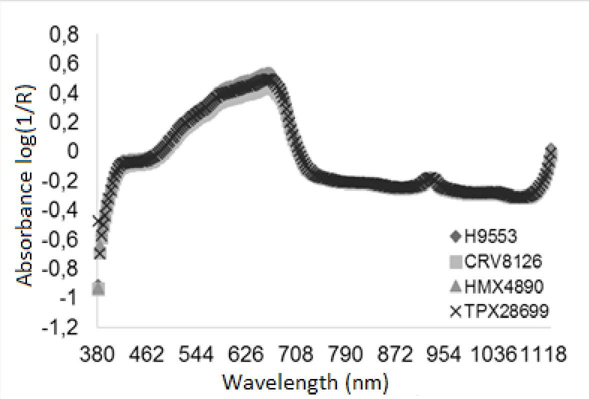

Figure 7. Near-infrared spectra collected as absorbance for hybrids HMX4890, H9553, TPX28699, and CRV8126 in true leaf phase

After the pre-processing in the first derivative with polynomial (9+9) for hybrid H9553, the second derivative with polynomial (9+9) for hybrid CRV8126 and HMX4890, and for hybrid TPX 28699 D2A (5+5), a reduction was observed in the wavelength close to the 700 mm with the greatest variation for the TPX 28699 hybrid (Figure 8).

It can be seen in Figure 9 that the PCA for the true leaf showed a better result than the previous ones since the distribution of indicators for each hybrid was better highlighted. It is also observed that the hybrid CRV8126 indicators are located more in the second quadrant of the graph, with some discrimination being possible. Ranking with two major components 64.00% of the data and with one major component 20.00% of the data.

For the calibration of the models for true leaf, 20 data were used as a reference value. The spectral range used was from 384 to 1,131 nanometers and the analyses were performed using the PC-LDA and PLS-DA models. The best result was obtained for the second derivative with polynomial (5+5) with 77.50% accuracy (Table 11).

Table 10. Results of calibration models using near-infrared spectroscopy for tomato cotyledonary leaf using the hybrid TPX28699 as primary objective

TPX28699 / PCA - LDA											
P.P	ST	PC	WL	TPX28699 (n=20)	OH (n=60)	AC	Sens	Spec	PPV	NVP	AV
ABS	-	15	387-1,131	14	48	77.5	70.0	80.0	53.9	88.9	50.0
NSS	-	15	387-780	15	49	80.0	75.0	81.7	57.7	90.7	56.7
D1A 4+4	-	15	402-1,119	16	50	82.5	80.0	83.3	61.5	92.6	63.3
D2A 9+9	-	15	420-597	14	51	81.3	70.0	85.0	60.9	89.5	55.0
MSC	-	15	381-1,131	17	45	77.5	85.0	75.0	53.1	93.8	60.0
TPX28699 / PLS – DA											
P.P	ST	C.P	WL	TPX28699 (n=20)	OH (n=60)	AC	Sens	Spec	PPV	NVP	AV
ABS	0,5	7	387-1,131	10	58	85.0	50.0	96.7	83.3	85.3	46.7
ABS	0,4	7	387-1,131	13	54	83.8	65.0	90.0	68.4	88.5	55.0
NSS	0,5	7	387-780	7	56	78.8	35.0	93.3	63.6	81.2	28.3
NSS	0,4	7	387-780	15	52	83.8	75.0	86.7	65.2	91.2	61.7
D1A 9+9	0,5	6	410-1,104	11	58	86.3	55.0	96.7	84.6	86.6	51.7
D1A 9+9	0,4	6	410-1,104	15	55	87.5	75.0	91.7	75.0	91.7	66.7
D2A 7+7	0,5	6	417-597	12	56	85.0	60.0	93.3	75.0	87.5	53.3
D2A 7+7	0,4	6	417-597	18	52	87.5	90.0	86.7	69.2	96.3	76.7
MSC	0,5	4	381-1,131	4	57	76.3	20.0	95.0	57.1	78.1	15.0
MSC	0,4	4	381-1,131	8	53	76.3	40.0	88.3	53.3	81.5	28.3

P.P: pre-processing; LS: separation threshold; PC: principal components; ABS: absorbance spectra; WL: wavelength; D.H: other hybrids; AC: accuracy; NSS: normal signal standardization; D1A: first derivative with 9-point derivation window; D2A: 4-derivation point second derivative; Sens: sensitivity; Spec: specificity; PPV: prediction of positive values; NVP: prediction of negative values; AV: ability to avoid failure; MSC: multiplicative scattering correction

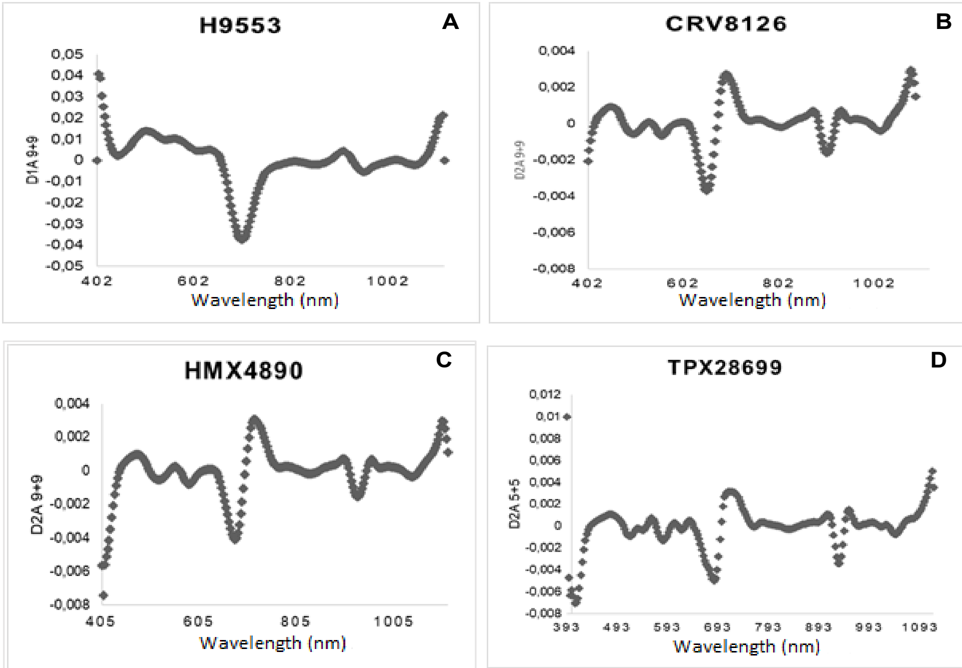

Figure 8. Spectra collected in the form of absorbance for true leaves of hybrids H9553, CRV8126, HMX4890, and TPX28699 after pre-processing

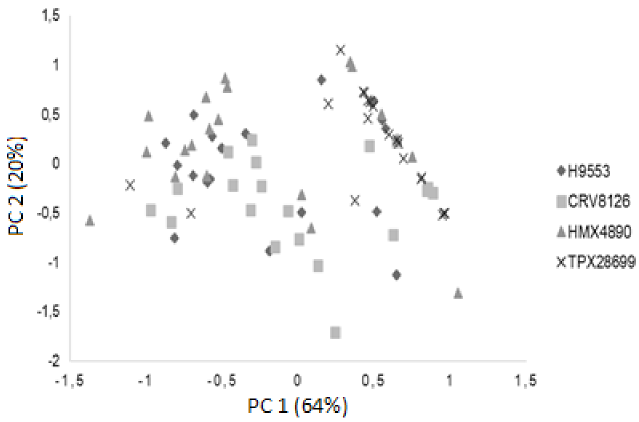

Figure 9. Principal component analysis for 384-1131-nm near-infrared spectra in absorbance for hybrids HMX4890, H9553, TPX28699, and CRV8126 in true leaf phase

Table 11. Results of PC-LDA calibration models using near-infrared spectroscopy (384-1131 nm) between true leaves of each hybrid

	True leaf global				
Calibration (n)	20	20	20	20	AC
PP	H9553	CRV8126	HMX4890	TPX28699	
ABS	8	14	17	16	68.8
NSS	13	15	14	15	71.3
NSS Dt	16	12	14	16	72.5
D1A 5+5	13	14	14	17	72.5
D2A 5+5	17	12	15	18	77.5
MSC	13	15	13	15	70.0

P.P: pre-processing; AC: accuracy; ABS: absorbance spectra; NSS: normal signal standardization; D1A: first derivative with 9-point derivation window; D2A: second derivative 4 points of derivation; MSC: multiplicative scattering correction

When evaluating the results in Table 12, it can be inferred that the best pre-processing for the PC-LDA model is D2A (5+5), with no value for the separation threshold and with the principal components equal to 15. It was identified 18 out of 20 data corresponding to the hybrid H9553 and 58 out of the 80 data corresponding to the other hybrids. Thus, it resulted in an accuracy of 95.00%. Sensitivity was 90.00%, specificity 72.50%.

For the PLS-DA model, the pre-processing D1A (9+9) with a value of 0.5 for the separation threshold, presented the best result with an accuracy of 98.75%, resulting in a sensitivity of 95.00 % and specificity of 75.00% (Table 12).

For the hybrid CRV8126 (Table 13), the spectral range used was from 378 to 1131 nanometers and the analyses were performed using the PCA-LDA and PLS-DA models with D1A (7+7), D2A (9+9), being also evaluated without any type of treatment, NSS and MSC.

Using the PCA-LDA model with the first polynomial derivative (7+7), with no value for the separation threshold and with the principal components equal to 15, it was possible to identify 14 of the 20 data referring to the hybrid CRV8126 and 58 data of the 80 referring to the other hybrids (Table 13).

For the hybrid HMX4890, good accuracy was found for NSS pre-processing in the PCA-LDA model, classifying 15 out of the 20 data from the hybrid and 57 out of the 80 data from the other hybrids, resulting in an accuracy of 90.00% for the PCA-LDA model (Table 14).

Regarding the PLS-DA model, the D1A (4+4) and D2A (9+9) preprocessing with values of 0.5 and 0.4 for the separation threshold, presented the best results with an accuracy of 91.25%, correctly classifying 17 of the 20 data for the hybrid HMX4890 and 56 of the 80 data for the other hybrids (Table 14).

For the hybrid TPX28699, the pre-processing D2A (7+7) showed the best result, with no separation threshold and principal component equal to 15, and obtained a classification of 17

Table 12. Results of calibration models using near-infrared spectroscopy for true tomato leaf using hybrid H9553 as primary objective

H9553 / PCA – LDA											
P.P	ST	PC	WL	H9553 (n=20)	OH (n=80)	AC	Sens	Spec	PPV	NVP	AV
ABS	-	15	393-1,131	6	56	77.5	30.0	70.0	20.0	80.0	0.0
NSS	-	15	378-1,143	15	56	88.8	75.0	70.0	38.5	91.8	45.0
D1A 4+4	-	15	378-1,143	16	58	92.5	80.0	72.5	42.1	93.6	52.5
D2A 5+5	-	15	378-1,143	18	58	95.0	90.0	72.5	45.0	96.7	62.5
MSC	-	15	378-1,143	16	56	90.0	80.0	70.0	40.0	93.3	50.0
H9553 / PLS – DA											
P.P	ST	PC	WL	H9553 (n=20)	OH (n=80)	AC	Sens	Spec	PPV	NVP	AV
ABS	0.5	10	393-1,131	19	59	97.5	95.0	73.8	47.5	98.3	68.8
ABS	0.4	10	393-1,131	20	55	93.8	100.0	68.8	44.4	100.0	68.8
NSS Dt	0.5	7	378-1,143	19	56	93.8	95.0	70.0	44.2	98.3	65.0
NSS Dt	0.4	7	378-1,143	19	59	97.5	95.0	73.8	47.5	98.3	68.8
D1A 9+9	0.5	11	378-1,143	19	60	98.8	95.0	75.0	48.7	98.4	70.0
D1A 9+9	0.4	11	378-1,143	20	55	93.8	100.0	68.8	44.4	100.0	68.8
D2A 5+5	0.5	7	378-1,143	19	57	95.0	95.0	71.3	45.2	98.3	66.3
D2A 5+5	0.4	7	378-1,143	19	57	95.0	95.0	71.3	45.2	98.3	66.3
MSC	0.5	6	378-1,143	18	57	93.8	90.0	71.3	43.9	96.6	61.3
MSC	0.4	6	378-1,143	19	56	93.8	95.0	70.0	44.2	98.3	65.0

P.P: pre-processing; LS: separation threshold; PC: principal components; ABS: absorbance spectra; WL: wavelength; OH: other hybrids; AC: accuracy; NSS: normal signal standardization; D1A: first derivative with 9-point derivation window; D2A: second derivative 4 points of derivation; Sens: sensitivity; Spec: specificity; PPV: prediction of positive values; NVP: prediction of negative values; AV: ability to avoid failure; MSC: multiplicative scattering correction

Table 13. Results of the calibration models using near-infrared spectroscopy for true tomato leaf using the hybrid CRV8126 as the primary objective

P.P	ST	PC	WL	CRV8126 (n=20)	OH (n=80)	AC	Sens	Spec	PPV	NVP	AV
CRV8126 / PCA - LDA											
ABS	-	15	393-1,131	17	54	88.8	85.0	67.5	39.5	94.7	52.5
NSS Dt	-	15	378-1,143	11	55	82.5	55.0	68.8	30.6	85.9	23.8
D1A 7+7	-	15	378-1,143	14	58	90.0	70.0	72.5	38.9	90.6	42.5
D2A 9+9	-	15	378-1,143	12	59	88.8	60.0	73.8	36.4	88.1	33.8
MSC	-	15	378-1,143	12	53	81.3	60.0	66.3	30.8	86.9	26.3
CRV8126 / PLS - DA											
P.P	ST	C.P	WL	CRV8126 (n=20)	OH (n=80)	AC	Sens	Spec	PPV	NVP	AV
ABS	0.5	3	393-1,131	9	39	60.0	45.0	48.8	18.0	78.0	-6.3
ABS	0.4	3	393-1,131	13	49	77.5	65.0	61.3	29.6	87.5	26.3
NSS Dt	0.5	5	378-1,143	11	38	61.3	55.0	47.5	20.8	80.9	2.5
NSS Dt	0.4	5	378-1,143	16	47	78.8	80.0	58.8	32.7	92.2	38.8
D1A 7+7	0.5	5	378-1,143	13	38	63.8	65.0	47.5	23.6	84.4	12.5
D1A 7+7	0.4	5	378-1,143	16	49	81.3	80.0	61.3	34.0	92.5	41.3
D2A 9+9	0.5	6	378-1,143	11	40	63.8	55.0	50.0	21.6	81.6	5.0
D2A 9+9	0.4	6	378-1,143	15	54	86.3	75.0	67.5	36.6	91.5	42.5
MSC	0.5	5	378-1,143	10	39	61.3	50.0	48.8	19.6	79.6	-1.3
MSC	0.4	5	378-1,143	13	49	77.5	65.0	61.3	29.6	87.5	26.3

P.P: pre-processing; LS: separation threshold; PC: principal components; ABS: absorbance spectra; WL: wavelength; OH: other hybrids; AC: accuracy; NSS: normal signal standardization; D1A: first derivative with 9-point derivation window; D2A: second derivative 4 points of derivation; Sens: sensitivity; Spec: specificity; PPV: prediction of positive values; NVP: prediction of negative values; AV: ability to avoid failure; MSC: multiplicative scattering correction

Table 14. Results of the calibration models using near-infrared spectroscopy for true tomato leaf using the hybrid HMX4890 as the primary objective

P.P	ST	PC	WL	HMX4890 (n=20)	OH (n=80)	AC	Sens	Spec	PPV	NVP	AV
HMX4890 / PCA - LDA											
ABS	-	15	393-1,131	12	36	60.0	60.0	45.0	21.4	81.8	5.0
NSS	-	15	378-1,143	15	57	90.0	75.0	71.3	39.5	91.9	46.3
D2A 9+9	-	15	378-1,143	16	55	88.8	80.0	68.8	39.0	93.2	48.8
D2A 9+9	-	15	378-1,143	14	52	82.5	70.0	65.0	33.3	89.7	35.0
MSC	-	15	378-1,143	14	57	88.8	70.0	71.3	37.8	90.5	41.3
HMX4890 / PLS - DA											
P.P	ST	C.P	WL	HMX4890 (n=20)	OH (n=80)	AC	Sens	Spec	PPV	NVP	AV
ABS	0.5	13	393-1,131	14	55	86.3	70.0	68.8	35.9	90.2	38.8
ABS	0.4	13	393-1,131	14	55	86.3	70.0	68.8	35.9	90.2	38.8
NSS	0.5	5	378-1,143	10	55	81.3	50.0	68.8	28.6	84.6	18.8
NSS	0.4	5	378-1,143	10	55	81.3	50.0	68.8	28.6	84.6	18.8
D1A 4+4	0.5	12	378-1,143	17	56	91.3	85.0	70.0	41.5	94.9	55.0
D1A 4+4	0.4	12	378-1,143	17	56	91.3	85.0	70.0	41.5	94.9	55.0
D2A 9+9	0.5	10	378-1,143	17	56	91.3	85.0	70.0	41.5	94.9	55.0
D2A 9+9	0.4	10	378-1,143	17	56	91.3	85.0	70.0	41.5	94.9	55.0
MSC	0.5	5	378-1,143	12	55	83.8	60.0	68.8	32.4	87.3	28.8
MSC	0.4	5	378-1,143	12	55	83.8	60.0	68.8	32.4	87.3	28.8

P.P: pre-processing; LS: separation threshold; PC: principal components; ABS: absorbance spectra; WL: wavelength; OH: other hybrids; AC: accuracy; NSS: normal signal standardization; D1A: first derivative with 9-point derivation window; D2A: second derivative 4 points of derivation; Sens: sensitivity; Spec: specificity; PPV: prediction of positive values; NVP: prediction of negative values; AV: ability to avoid failure; MSC: multiplicative scattering correction

of the 20 data of the hybrid TPX28699 and 53 of the 80 data of the other hybrids, resulting in an accuracy of 87.50% (Table 15).

In the PLS-DA analysis, the best result obtained was for the second derivative pre-processing with polynomial (5+5), with an accuracy of 95.00% (Table 15).

In Figure 10, a reference value equal to 1 was used for the individually analyzed hybrids H9553 (Figure 10A), CRV8126 (Figure 10B), HMX4890 (Figure 10C), TPX28699 (Figure 10D), and for the other hybrids, it was equal to 0. Then, the separation thresholds equal to 0.4 and 0.5 were tested. It can be observed that the best result for the separation threshold was equal to 0.4 because it was the one that best discriminated the analyzed hybrid from the others. Besides, for the true leaf model, the hybrid HMX4890 stood out as it discriminated the hybrid with fewer factors in relation to the others.

In the work carried out by Duarte *et al.* (2008), where oil and protein contents were evaluated for different maize hybrids, extremely positive results were obtained, capable of differentiating the contents of these components according to each hybrid. Analogously, when evaluating the differentiation capacity of the tomato hybrids in this work, extremely positive and efficient results were also found, with accuracies greater than 90.00%.

PLS-DA models combined with pre-processing, particularly derivatives proposed by Savitzky–Golay showed promising results for classification and identification of most hybrids of tested tomato plants, as well as proposed by Soares *et al.* (2017) for wood from species originating in the Amazon. However, because hybrids present genetic material more similar than species, it impaired the creation of spectral identity by NIR in the cotyledonary leaf. Thus, for robust models, a greater number of hybrids and the use of chemometric tools for the selection of spectral variables such as genetic algorithm is necessary for further works (CARVALHO *et al.*, 2017), regression by partial least square interval (NORGAARD *et al.*, 2000) or optimization of the

Table 15. Results of the calibration models using near-infrared spectroscopy for true tomato leaf using hybrid TPX28699 as primary objective

TPX28699 / PCA - LDA											
P.P	ST	PC	WL	TPX28699 (n=20)	OH (n=80)	AC	Sens	Spec	PPV	NVP	AV
ABS	-	15	393-1,131	18	41	73.8	90.0	51.3	31.6	95.4	41.3
NSS Dt	-	15	378-1,143	17	52	86.3	85.0	65.0	37.8	94.6	50.0
D1A 5+5	-	15	378-1,143	18	51	86.3	90.0	63.8	38.3	96.2	53.8
D2A 7+7	-	15	378-1,143	17	53	87.5	85.0	66.3	38.6	94.6	51.3
MSC	-	15	378-1,143	16	48	80.0	80.0	60.0	33.3	92.3	40.0
TPX28699 / PLS - DA											
P.P	ST	C.P	WL	TPX28699 (n=20)	D.H (n=80)	AC	Sens	Spec	PPV	NVP	AV
ABS	0.5	7	393-1,131	13	57	87.5	65.0	71.3	36.1	89.1	36.3
ABS	0.4	7	393-1,131	15	48	78.8	75.0	60.0	31.9	90.6	35.0
NSS Dt	0.5	6	378-1,143	13	54	83.8	65.0	67.5	33.3	88.5	32.5
NSS Dt	0.4	6	378-1,143	15	49	80.0	75.0	61.3	32.6	90.7	36.3
D1A 5+5	0.5	6	378-1,143	15	54	86.3	75.0	67.5	36.6	91.5	42.5
D1A 5+5	0.4	6	378-1,143	16	50	82.5	80.0	62.5	34.8	92.6	42.5
D2A 5+5	0.5	7	378-1,143	18	58	95.0	90.0	72.5	45.0	96.7	62.5
D2A 5+5	0.4	7	378-1,143	19	56	93.8	95.0	70.0	44.2	98.3	65.0
MSC	0.5	7	378-1,143	14	54	85.0	70.0	67.5	35.0	90.0	37.5
MSC	0.4	7	378-1,143	15	50	81.3	75.0	62.5	33.3	90.9	37.5

P.P: pre-processing; LS: separation threshold; PC: principal components; ABS: absorbance spectra; WL: wavelength; OH: other hybrids; AC: accuracy; NSS: normal signal standardization; D1A: first derivative with 9-point derivation window; D2A: second derivative 4 points of derivation; Sens: sensitivity; Spec: specificity; PPV: prediction of positive values; NVP: prediction of negative values; AV: ability to avoid failure; MSC: multiplicative scattering correction

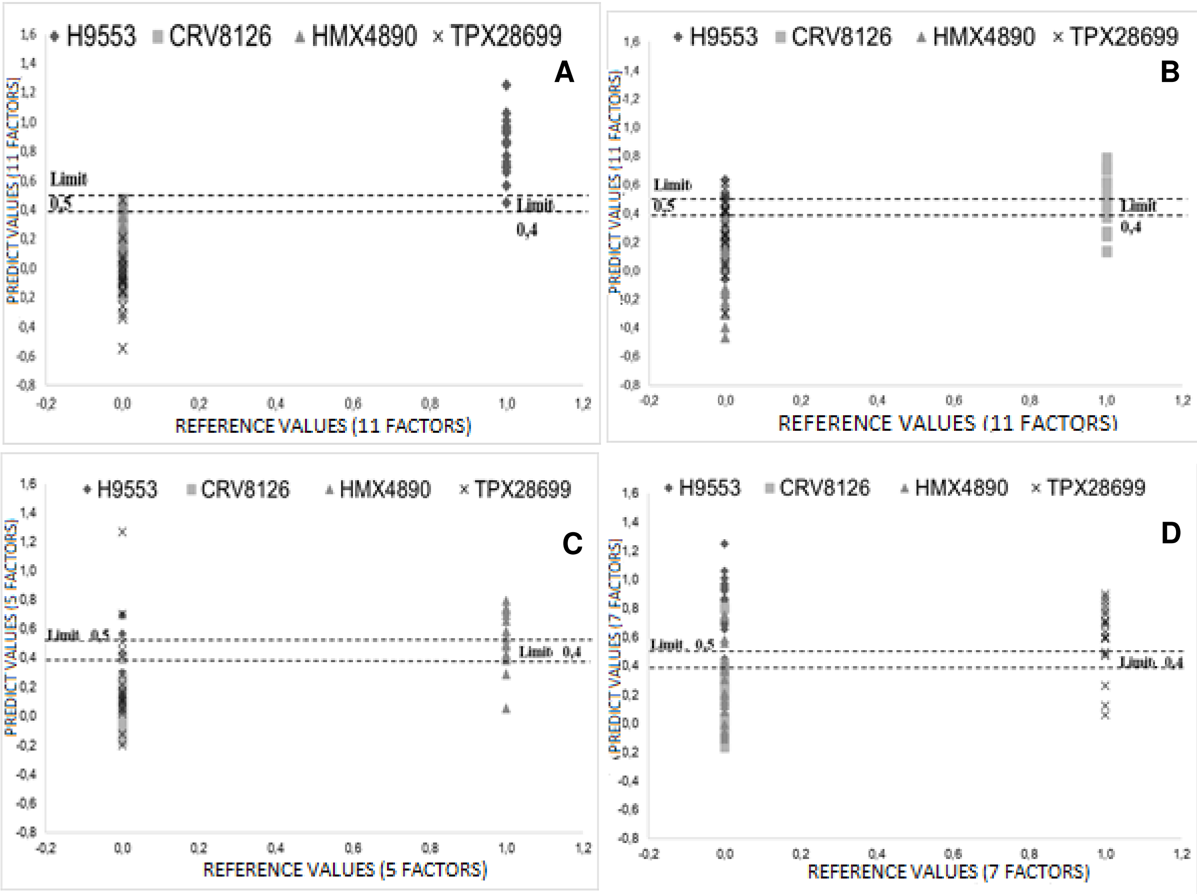

Figure 10. Prediction results of PLS-DA models constructed with NIR spectra for true leaf of tomato hybrids. A) hybrid H9953, pre-processing D1A (9+9) (378-1,143 nm). B) hybrid CRV8126, pre-processing D2A (9+9) (378-1,143 nm). C) HMX4890 hybrid, pre-processing D2A (9+9) (378-1,143 nm). C) TPX28699 hybrid, D2A (5+5) preprocessing (378-1,143 nm)

PLS wavelength window (GUTHRIE *et al.,* 2005; CUNHA JUNIOR *et al*., 2016).

CONCLUSION

- The results from this experiment demonstrate that the technique of spectroscopy in the near-infrared range (NIRS) associated with multivariate analysis allowed the discrimination of the assessed hybrids.
- The phase in which the best results were obtained in the identification of each hybrid was in the seed, allowing the use of the instrumental technique on a portable scale for tomato hybrids with a high rate of assertiveness.
- A relatively small set of samples was used in this work mainly in relation to the total number of tomato hybrids existing on the market, showing that further studies may be carried out to increase the database and the model's credibility.

AUTHORSHIP CONTRIBUTION STATEMENT

SANTANA, I.G.: Conceptualization, Data curation, Formal Analysis, Investigation, Writing – original draft; **BRITO, A.A.:** Data curation, Formal Analysis, Investigation, Methodology, Writing – original draft; **AGUIAR, F.C.O.:** Data curation, Formal Analysis, Investigation; **CAMPOS, L.F.C.:** Data curation, Methodology, Supervision, Writing – review & editing; **CORRÊA, G.C.:** Data curation, Supervision, Writing – review & editing; **NASCIMENTO, A.R.:** Data curation, Funding acquisition, Supervision, Writing –

review & editing; **CUNHA JUNIOR, L.C.:** Conceptualization, Methodology, Project administration, Supervision, Writing – review & editing.

DECLARATION OF INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGMENTS

We would like to thank the financial support granted by MCTI “TED no. 14.0030.00/2018” and the National Council for Scientific and Technological Development (Process no. 406617/2018-0).

REFERENCES

BOITEUX, L. S.; FONSECA, M. E. N.; GIORDANO, L. B.; MELO, P. C. T. Melhoramento genético. *In*: CLEMENTE F. M. V. T.; BOITEUX L. S. (ed.). **Produção de tomate para processamento industrial**. Brasília: Embrapa, 2012. cap. 2, p. 31-50.

BOTELHO, B. G.; REIS, N.; OLIVEIRA, L. S.; SENA, M. M. Development and analytical validation of a screening method for simultaneous detection of five adulterants in raw milk using mid-infrared spectroscopy and PLS-DA. **Food Chemistry**, v. 181, p. 31-37, 2015.

BRAGA, J. W. B.; POPPI, R. J. Validação de modelos de calibração multivariada: uma aplicação na determinação de pureza polimórfica de carbamazepina por espectroscopia no infravermelho próximo. **Química Nova**, v. 27, n. 6, p. 1004-1011, 2004.

CARVALHO, L. C.; MORAIS, C. L. M.; LIMA, K. M. G.; LEITE, G. W. P.; OLIVEIRA, G. S.; CASAGRANDE, I. P.; SANTOS NETO, J. P.; TEIXEIRA, G. H. A. Using Intact Nuts and Near Infrared Spectroscopy to Classify Macadamia Cultivars. **Food Analytical Methods**, 2017.

CUNHA JÚNIOR, L. C. S.; TEIXEIRA, G. H. D. A.; NARDINI, V.; WALSH, K. B. Quality evaluation of intact açaí and juçara fruit by means of near infrared spectroscopy. **Postharvest Biology and Technology**, v. 112, p. 64-74, 2016.

DINIZ, K. A.; GUIMARÃES, S. T. M. R.; LUZ, J. M. Q. Húmus como substrato para a produção de mudas de tomate, pimentão e alface. **Bioscience Journal**, v. 22, n. 3, p. 63-70, 2006.

DOWNES, G. M.; HARWOOD, C. E.; WIEDEMANN, J.; EBDON, N.; BOND, H.; MEDER, R. Radial variation in Kraft pulp yield and cellulose content in *Eucalyptus globulus* wood across three contrasting sites predicted by near infrared spectroscopy. **Canadian Journal of Forest Research**, v. 42, n. 8, p. 1577–1586, 2012.

DUARTE, A.; CARVALHO, C.; CAVICHIOLI, J. Densidade, teor de óleo e produtividade de grãos em híbridos de milho. **Bragantia**, v. 67, n. 3, p.759-767, 2008

DUDLEY, J. W.; HYMOWITZ, T.; COLLINS, F. I.; BROWN, C. M. Near-infrared reflectance measurement of protein and oil concentration in corn, soybean, and oat seed. *In*: HILL, L. D. (ed.). **Quality in world markets saddle river**: corn quality in world markets. Danville, Illinois: Interstate Printers and Publishers, 1975. cap. 3, p. 205-210.

GUTHRIE J. A.; WALSH, K. B.; REID, D. J.; LIEBENBERG, C. J. Assessment of internal quality attributes of mandarin fruit. NIR calibration model development. **Australian Journal of Agricultural Research**, v. 56, p. 405-416, 2005.

HF BRASIL. **Principais itens do Custo de Produção de Goiânia: indústria em (%)**, 2018. Available in: https://www.hfbrasil.org.br/br/especial-hortalicas-custo-de-producao-de-tomate-em-goiania-go.aspx Access in: Oct. 10, 2020.

IFAG -**Instituto para o Fortalecimento da Agropecuária de Goiás**, 2020. Available in: http://ifag.org.br/custo-de-producao-tomate/. Access in: Jun. 5, 2021.

IBGE. **Levantamento Sistemático da Produção Agrícola, Estatística da Produção Agrícola, julho 2020**. Brasília: IBGE, 2020. Available in: https://biblioteca.ibge.gov.br/visualizacao/periodicos/2415/epag_2020_jul.pdf. Access in: Sep. 2, 2020.

LUZ, J. M. Q.; BITTAR, C. A.; OLIVEIRA R. C.; NASCIMENTO, A. R.; NOGUEIRA, A. P. O. Desempenho e divergência genética de genótipos de tomate para processamento industrial. **Horticultura Brasileira**, v. 34, p. 483-490, 2016.

MARETTO, D. A. **Aplicação de máquinas de vetores de suporte para desenvolvimento de modelos de classificação e calibração multivariada em espectroscopia no infravermelho**. 2011. Tese (Doutorado em Ciências) - Instituto de Química, Universidade Estadual de Campinas, Campinas, 2011.

MECOZZI, M.; PIETROLETTI, M.; TORNAMBÈ, A. Molecular and structural characteristics in toxic algae cultures of Ostreopsis ovata and Ostreopsis spp. evidenced by FTIR and FTNIR spectroscopy. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 78, n. 5, p. 1572-1580, 2011.

MELO, P. C. T.; FONTE, L. C. Brazil processing tomato season 2010: results and future perspectives. **Tomato News**, v. 3, p. 15-19, 2011.

MENDES, R. A. G. **Utilização da espectroscopia em reflectância no infravermelho próximo para discriminação de espécies da família Myrtaceae**. 2014. Dissertação (Mestrado em Botânica) - Universidade de Brasília, Brasília, 2014.

MONFERRERE, G. L.; AZCARATE, S. M.; CANTARELLI, M. Á.; ALEMÃO, I. F.; CAMIÑA, J. M. Chemometric characterization of sunflower seeds. **Journal of Food Science**, v. 77, n. 9, p. 1018-1022, 2012.

MORAIS, C. L. M.; LIMA, K. M. G. Principal Component Analysis with Linear and Quadratic Discriminant Analysis for Identification of Cancer Samples Based on Mass Spectrometry. **Journal of the Brazilian Chemical Society**, v. 29, p. 472-481, 2018.

MUÑIZ, G. I. B.; MAGALHÃES, W. L. E.; CARNEIRO, M. E.; VIANA, L. C. Fundamentos e estado da arte da espectroscopia no infravermelho próximo no setor de base florestal. **Ciência Florestal**, v. 22, n. 4, p. 865-875, 2012.

NAES, T.; ISAKSSON, T.; FEARN, T.; DAVIES, T. **A user-friendly guide to multivariate calibration and classification.** Chichester: NIR Publications. 2002.

NORGAARD, L.; SAUDLAND, A.; WAGNER, J.; NIELSEN, J. P.; MUNCK, L.; ENGELSEN, S. B. Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy. **Applied Spectroscopy**, v. 54, p. 413-419, 2000.

SANTOS, M. B. H.; GOMES, A. A.; VILAR, W. T. S.; ALMEIDA, P. B. A.; MILANI, M.; NÓBREGA, M. B. M.; MEDEIROS, E. P.; GALVÃO, R. K. H.; ARAÚJO, M. C. U. Non-destructive NIR spectrometric cultivar discrimination of castor seeds resulting from breeding programs. **Journal of the Brazilian Chemical Society**, v. 25, n. 5, 969-974, 2014.

SILEONI, V.; VAN DEN BERG, F.; MARCONI, O.; PERRETTI, G.; FANTOZZI, P. Internal and external validation strategies for the evaluation of long-term effects in NIR calibration models. **Journal of Agricultural and Food Chemistry**, v. 59, p. 1541-1547, 2011.

SILVA-JÚNIOR, A. R.; RIBEIRO, W. M.; NASCIMENTO, A. R.; SOUZA, C. B. Cultivo do tomate industrial no estado de Goiás: evolução das áreas de plantio e produção. **Conjuntura Econômica Goiana**, v. 34, p. 97-109, 2015.

SNEL, F. A.; BRAGA, J. W. B.; SILVA, D.; WIEDENHOEFT, A. C.; COSTA, A.; SOARES, R.; CORADIN, V. T. R. PASTORE, T. C. M. Potential field-deployable NIRS identification of seven Dalbergia species listed by CITES. **Wood Science and Technology**, v. 52, n. 5, p. 1411-1427, 2018.

SOARES, L. F.; SILVA, D. C.; BERGO, M. C. J.; CORADIN, V. T. R.; BRAGA, J. W. B.; PASTORE, T. C. M. Avaliação de espectrômetro NIR portátil e PLS-DA para a discriminação de seis espécies similares de madeiras amazônicas. **Química Nova**, v. 40, n. 4, p. 418-426, 2017.

VILELA, N. J.; MELO, P. C. T.; BOITEUX, L. S.; CLEMENTE, F. M. V. T. Perfil socioeconômico da cadeia agroindustrial no Brasil. *In*: CLEMENTE, F. M. V.; BOITEUX, L. S. (ed.). **Produção de tomate para processamento industrial**. Brasília: Embrapa Hortaliças, 2012. cap. 1, p. 17-27.

WILLIAMS, P. C.; NORRIS, K. Variable affecting near-infrared spectroscopic analysis. *In*: WILLIAMS, P. C.; NORRIS, K. (eds.). **Near-Infrared Technology in Agricultural and Food Industries**, 2nd Ed. St. Paul, Minnesota: American Association of Cereal Chemists, Inc. 2001, cap. 4, p.171-189.

XIAOBO, Z.; JIEWEN Z.; POVEY, M. J. W.; HOLMES, M.; HANPIN, M. Métodos de seleção de variáveis em espectroscopia no infravermelho próximo. **Analytica Chimica Acta**, v. 667, n. 1, p. 14-32, 2010.