Acta Scientiarum
http://periodicos.uem.br/ojs
ISSN on-line: 1807-8664
Doi: 10.4025/actascitechnol.v45i1.61270
CC BY

Near infrared spectroscopy as a tool for agricultural expertise: identification of tomato seedlings

Amanda Rodrigues Vinhandelli[1], Annelisa Arruda De Brito[1], Raquel Cintra de Faria[1], Luiz Fernandes Cardoso Campos[1], Gilberto Alessandre Soares Goulart[1], Gustavo Henrique de Almeida Teixeira[2], Abadia dos Reis Nascimento[1] and Luís Carlos Cunha Junior[1*]

[1]Escola de Agronomia, Universidade Federal de Goiás, Avenida Esperança, s/n., 74690-900, Goiânia, Goiás, Brazil. [2]Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal, São Paulo, Brazil. *Author for correspondence. E-mail: cunhajunior.l.c@ufg.br

ABSTRACT. Tomatoes are one of the most prominent vegetables globally, with significant cultural and economic relevance in various nations, including Brazil. The term ‘safe food’ is becoming more popular as consumer preferences and supply chain dynamics become evolved in these processes. In light of these issues, the use of safety and quality management methods for fruits and vegetables have increased dramatically, with traceability being one of these solutions worth highlighting. When it comes to traceability, evaluation of tomato seedlings, plants, and fruits to identify groups or hybrids becomes particularly crucial throughout the marketing process, since the consumer of seedlings or fruit has difficulties recognizing whether that product truly belongs to the group indicated by the merchant. Thus, the potential of near infrared spectroscopy (NIRS) combined with the PC-LDA and PLS-DA algorithms was tested for the discrimination of two significant commercial groups, Salada and Saladete, as well as eleven cultivars belonging to these groups, which were tested for this purpose. The results show that, by using the PLS-DA model, the portable NIR equipment is capable of differentiating tomato seedlings in nurseries of the Salada and Saladete groups, with an accuracy of 99.7% and sensitivity of 100%. The technique showed to be efficient for individual models of tomato seedlings in the Salada group, with accuracy over 90% and sensitivity above 93% for all models. For the Saladete group's individual models, the technique proved effectiveness for the hybrids Parma, BS-110012, Giácomo, Guara, and Tyna.

Keywords: hybrids; PC-LDA; PLS-DA; traceability; *Solanum lycopersycum* L.

Received on October 20, 2021.
Accepted on May 17, 2022.

Introduction

Tomato cultivation for fresh consumption has a high production cost, of which the inputs, labor and seeds correspond, on average, for 60% of the total cost (Pagliuca, Deleo, Boteon, Mueller, & Valmorbida, 2017). Thus, profitability begins with the choice of the tomato and hybrid group to be implemented, as well as the production and delivery of healthy seedlings (Diniz, Guimarães, & Luz, 2006). However, it is not uncommon for loads or seedlings to be exchanged due to lack of inputs, such as lack of seeds of a certain hybrid, resulting in the delivery of tomato seedlings from the same group, but not of the same genetic material, causing economic damage that normally occurs at the time of production. These cases end up being prosecuted, and consequently subjected to expertise.

Regarding the performance of judicial expertise, the elucidation is technically effective, but ineffective for reducing damages to the producer. This ambiguity occurs because such identification process is carried out in the fruit production stage, a period of easy distinction between genetic materials. However, at this point, the economic damage has already been done, as the implantation and input values were already spent at the time of the inspection. According to Pagliuca et al. (2017), these stages present values that can reach up to 50% of the total cost of production. Thus, it is necessary to study methods and techniques that allow an expert to identify plant material quickly and before planting.

In this context, near infrared spectroscopy (NIRS) can be an alternative technology. In the field of non-destructive testing, the technique of spectroscopy in the near infrared range presents itself as a fast tool, which allows real-time analysis, thus demonstrating reliable results, which reduces the cost and time spent on routine analysis in laboratories (Brimmer & Hall, 2001; Muñiz, Magalhães, Carneiro, & Viana, 2012). Dale and Klatt (1999) showed the ability to use NIR diffuse reflectance to identify the quality standard of paper money stock. Silva et al. (2013) differentiated types of inks and pen brands through the NIR spectrum that came from circles produced by these pens with results between 94.9 and 100% of correctness, thus contributing to increase the technological level in criminal laboratories, which allow forensic specialists a faster analysis and unbiased interpretation of the evidence. Oliveira, Honorato, Honorato, and Pereira (2018) demonstrated that portable NIR spectroscopy is 100% viable to identify counterfeit Brazilian banknotes from the original ones.

In vegetable applications, Soares et al. (2017) demonstrated that portable NIR spectroscopy can be useful for monitoring illegal wood trade, since the models developed for six species presented efficiency rates above 90% for wood discrimination in the field. Carvalho et al. (2018) demonstrated the efficiency of NIR spectroscopy to classify intact macadamia nut cultivars with an accuracy of 94.4%. 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 to control the timber trade, considering the technique demonstrated that can separate six different but visually confusing species of *Dalbergia* from various countries.

Moreover, this work aimed to evaluate the use of portable NIR spectroscopy associated with techniques of linear discriminant analysis of principal components (PC-LDA) and Partial Least Squares for Discriminant Analysis (PLS-DA), in order to obtain a fast method to assist in agricultural expertise for the authentication of two groups of tomatoes seedlings (Salada and Saladete) and 11 tomato hybrids, in order to reduce economic damages arising from the implantation and phytotechnical treatments of non-acquired plant genetic material.

Material and methods

The experiment was carried out in the Horticulture Sector of the School of Agronomy of the Federal University of Goiás (latitude 16º 35' 12" S, longitude 49º 21' 14" W Gr, at 730 m altitude), Goiânia, state of Goiás. The tomatoes hybrids (Table 1) were manually planted in polypropylene trays containing 162 cells. However, sowing was done in the 98 central cells of each tray. The substrate used was composed of coconut fiber (Amafibra 11), peat and Bioplant nature (coconut fiber, rice husk, pine bark and nutrients). After planting, the substrate was covered with vermiculite. Irrigation and fertilization were carried out with the aid of automatic bars following the commercial standard.

Hybrids of greater commercial use were used as plant material for the treatments, as described in Table 1. The treatments were subdivided into two large groups of tomato with indeterminate growth habits: Salada and Saladete (Alvarenga, 2013).

NIR spectral collections were carried out 35 days after sowing with Felix Intruments portable near infrared spectrophotometer, model F-750 (Camas, Washington, United States), which uses interactance as optical geometry and 300 range at 1200 nm. Readings were performed randomly on 30 plants of each variety tray , and the NIR spectrum was collected on the abaxial face of the trefoil at the end of each branch. The branch was chosen as the second to emerge after the cotyledon. Since temperature is one of the factors that affect performance in predicting models (Golic & Walsh, 2006), in order to reduce temperature interference in spectral collection, the temperature measurement of the trefoils was performed using an infrared thermometer (Benetech GM-32). Before the collection of NIR spectra, only those trefoils that presented a temperature between 27-28ºC were considered valid samples.

Table 1. Commercial hybrids of indeterminate growth cultivated.

Group	Hybrid	Company	Sample number
Salada	Dominador	Agristar	30
	Dylla	Syngenta	30
	ParonNTY		30
Saladete	BS II0012	Blue Seeds	29
	Giácomo	RijkZwaan	30
	Guará	HM Clause	28
	Parma	SuperSeed	30
	Ravena	Sakata	30
	Totalle	Nunhens	29
	Tyna	Sakata	29
	Helena	Feltrin	30

For data analysis, the following techniques were used: Principal Component Analysis (PCA), Discriminant Linear Analysis of Principal Components (PC-LDA) and Partial Least Squares for Discriminant Analysis (PLS-DA) (Naes, Isaksson, Fearn, & Davies, 2002). To optimize models in order to reduce light scattering and increase signal, the following spectral processing were tested: Standard Normal Variate (SNV) (Ozaki, McClure, & Christy, 2006), Multiplicative Scatter Correction (MSC) (Souza & Poppi, 2012) and the first (1SG) and second (2SG) derivatives of Savitzky-Golay (Brown, Vega-Montoto, & Wentzell, 2000).

Data were processed using The Unscrambler software version 10.0.3. The validation of the discrimination models was according the following calculations of the figures of merit: Accuracy (AC) (Cunha Júnior, Nardini, Khatiwada, Teixeira, & Walsh, 2015), false positive rate (FPR), false rate negative (FRN), specificity (SPEC) and sensitivity (SEN) (Botelho, Reis, Oliveira, & Sena, 2015). In the case of the PLS-DA models, the data used for the calculations were extracted from the result of the total cross validation of the calibration models. The flowchart for the procedure used in the data analysis is shown in Figure 1.

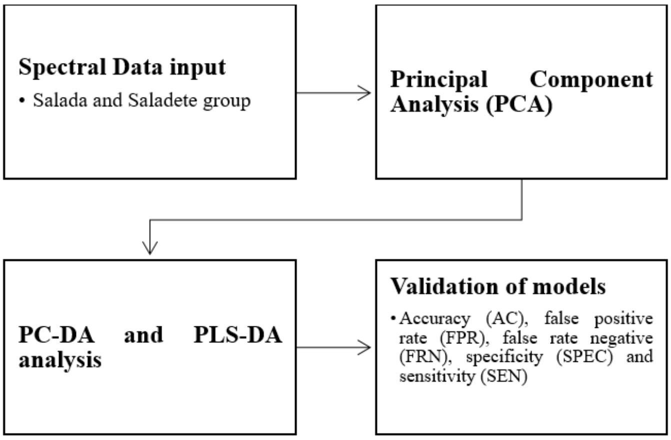

Figure 1. The flowchart for the analysis procedure.

Results and discussion

Spectrums, principal component analysis and global model

By observing the behavior of the Salada and Saladette absorbance spectra, there is a peak at the wavelength of 660 nm (Figure 2A) due to the presence of chlorophyll, which gives a characteristically green color to tomato seedlings (Gómez, Wang, & Pereira, 2006). In the region from 730 nm, a low absorbance can be seen for both hybrids, indicating that the seedlings have a high water content (Zhang, Li, & Zhang, 2012). For the spectra of each cultivar, it is observed that the cultivar Dominador had a higher absorbance in the region from 490 to 1080 nm, while the hybrid ParonNTY had the lowest peak (Figure 2B) in relation to the cultivars of the Salada group. As for the cultivars of the Saladete group, there was also greater absorbance at the wavelength of 660 mm and a decrease from the 730 nm region. The cultivar with the highest absorbance was BS 110025, followed by Ravena, Tyna, Helena, Parma, Totalle, Guará and Giácomo, respectively (Figure 2C). After pre-processing in the first derivative between the cultivars of the Saladete group, a negative peak was formed in the spectral range between 630 and 730 nm, demonstrating the low absorption in the region related to the red pigment (Beghi, Giovenzana, Tugnolo, & Guidetti, 2018). This means that, the chlorophyll present in the leaves was not degraded yet, and the smallest variation was detected in the cultivar BS 110025, followed by the cultivars Giácomo, Ravena, Parma, Totalle, Tyna, Guará and Helena respectively (Figure 2D). There was also an ascending behavior in the region from 1,080 to 1,130 nm (Figure 2D) for both hybrids, with peaks in this region being characteristic of water-rich samples (Nicolaï et al., 2017).

It was possible to explain about 98% of the total variability of the absorbance spectra for the Salada and Saladete groups in the first two main components, however, there was no evident separation of the groups (Figure 3A). Regarding the PCA analyzes for each isolated group, the first two main components explained 83 and 70% of the total variability of the data from the hybrids of the Saladete (Figure 3B) and Salada (Figure 3C) groups, respectively, demonstrating the difficulty of unsupervised methods associated with the NIRS technique in separating these types of material.

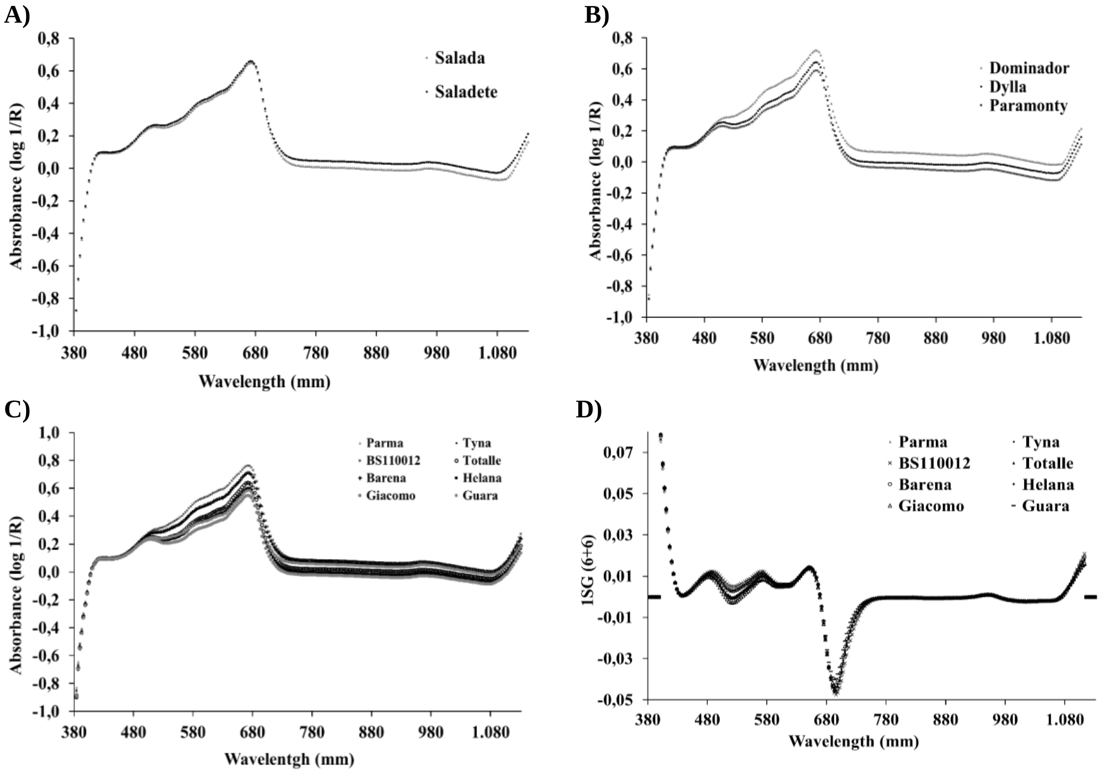

Figure 2. Spectra collected in the form of absorbance for the Salada and Saladete groups (A); Spectra collected in the form of absorbance for cultivars in the Salada group (B); Spectra collected in the form of absorbance for cultivars in the Saladete group (C). Spectra collected in the form of absorbance after pre-processing the first (1SG 6+6) derivative Savitzky-Golay for cultivars of the Saladete group (D).

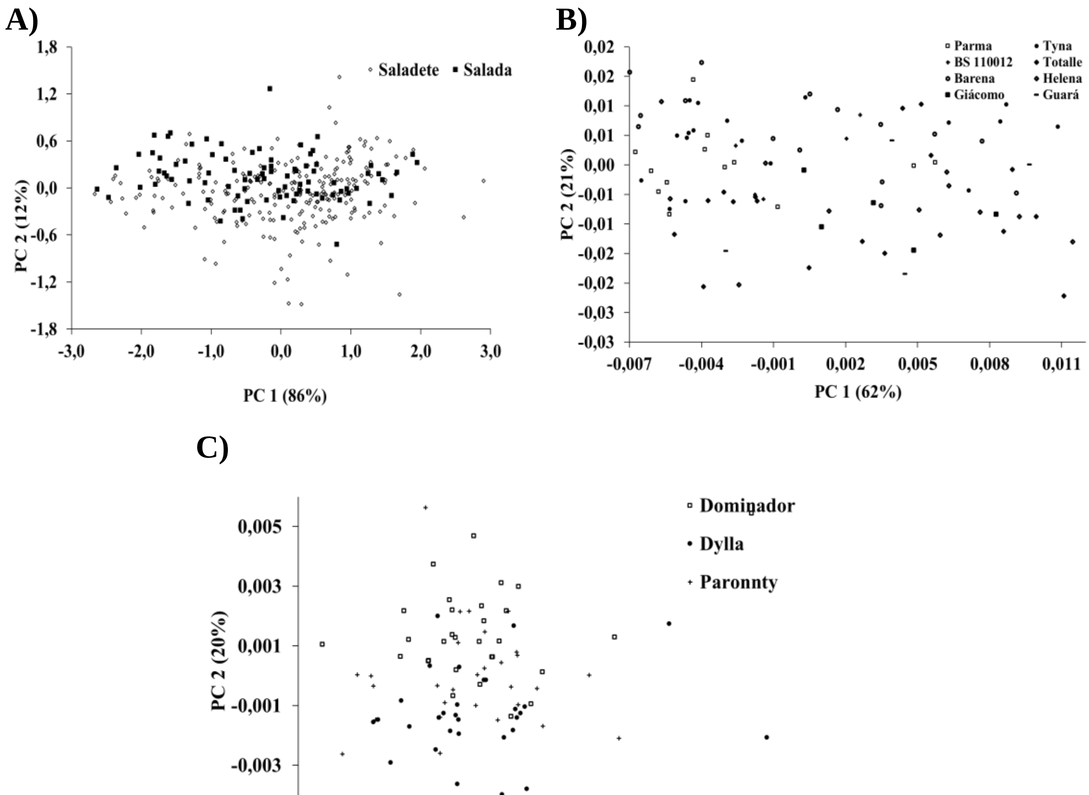

Figure 3. Analysis of principal components (PC) with near infrared absorbance spectra (381-1131 nm) for Salada and Saladete groups (A); cultivars from the Saladete group (B); and cultivars from the Salada group (C).

Supervised models such as PC-LDA with 10 principal components using a spectral range of 384-1131 nm were constructed to discriminate tomato seedling hybrids (Table 2), in which the probability of Salada and Saladette groups and their hybrids have the desired characteristic returns the results of prediction of positive values (PPV); on the other hand, the specificity for each hybrid and group of hybrids, returns results of prediction of negative values (PNV) (Amodio, Ceglie, Chaudhry, Piazzolla, & Colelli, 2017). The results obtained by NIR spectra associated with the PC-LDA technique are promising, especially when associated with spectral treatments with a window of 4+4 in the first derivative and a window of 6+6 in the second derivative of Savitzky-Golay, as they presented an accuracy of 88, 6% in both calibration models (Table 2). The accuracy values were calculated according to the Equation 1, as follows:

$$Accuracy: \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (1)$$

where:
TP is for true positive values, TN for true negative values, FP for false positive values and FN is for false negative values.

From this result, calibration models were performed to separate the groups (Salada and Saladete) and specific models for each hybrid within each group.

Table 2. Results of the Linear Discriminant Analysis of Principal Components (PC-LDA) models in the spectral length between 384 and 1131 nm among all cultivars of the Salada and Saladete groups, using 10 principal components.

	Salada/Saladete											
n =	30	30	30	30	29	29	29	30	30	30	28	AC (%)
Treat.	Dominador	Dylla	ParonNTY	Parma	Tyna	BS 110012	Totalle	Ravena	Helena	Giácomo	Guará	
no treat.	24	26	29	24	23	27	15	23	23	27	27	82.5
NSS	22	25	28	17	23	25	15	22	21	26	28	77.5
1SG(4+4)	29	26	28	26	27	26	22	27	25	25	27	88.6
1SG(6+6)	26	28	29	26	27	25	18	28	23	26	27	87.1
2SG(4+4)	28	27	25	18	14	22	11	17	18	20	21	68.0
2SG (6+6)	30	27	27	26	24	29	21	25	25	28	26	88.6

no treat.: no treatment, NSS: normal signal standardization; 1SG: first Savitzky-Golay derivative; 2SG: Savitzky-Golay second derivative, AC: accuracy.

Specific models for seedlings of the Salada and Saladete groups

The calibration models that used the PC-LDA algorithms presented good performance with accuracy greater than 96% for all. The models in which the spectra were treated by the Savitzky-Golay first derivative in the windows of 4 and 6 derivation points showed an accuracy of 99.7% in the separation of groups of seedlings (Table 3). The results obtained by these models are promising when compared with the models applied to other vegetables, such as the proposed by Canneddu, Cunha Júnior, and Teixeira (2016), for the separation of marketable macadamia, which obtained the best PC-LDA calibration model with an accuracy of 88.3%. Also Carvalho et al. (2018), who likewise worked with macadamia variety separation, and obtained models with accuracy of approximately 60%.

Regarding the PLS-DA calibration models, accuracy values greater than 99% were obtained (Table 3), showing that this method is the most appropriate. This is because PLS-DA allows variability in the limit line between the studied groups (Figure 4), using limits of 0.4 or 0.5 (Table 3). The PLS-DA model with a limit of 0.4 without pre-processing is highlighted, which presented 100% sensitivity for the seedlings in the Salada group (Table 3), and it was calculated according to Equation 2. Moreover, when used, it can identify the Saladete group with 100% certainty, as seen in Figure 4A.

$$Sensitivity: \frac{TP}{TP+FN} \times 100 \quad (2)$$

where:
TP is for true positive values and FN is for false negative values.

A suitable calibration model is the PLS-DA after the first Savitzky-Golay derivative in the 4-point window with a limit of 0.5, with an accuracy of 99% (Table 3). According to the prediction, this model can guarantee 100% accuracy of the Salada group (Figure 4B).

Soares et al. (2017), when studying the identification of six wood species from the Amazon with the use of NIRS and PLS-DA, obtained specificity among the species greater than 90%, as well as the sensitivity, with three of the six species reaching 100% of sensitivity. This corroborates with the result obtained in the present study, because demonstrates that the methodology is suitable for the proposal.

In Figure 4, a reference value equal to 1was adopted for the Salada group and equal to 0 for the Saladete group. Then, the separation limits equal to 0.4 and 0.5 were tested. Consequently, the best result for the separation limit was 0.4, as this value was the one that best discriminated the two groups.

Table 3. Results of the calibration models in the spectral length between 384 and 1131nm among all hybrids of the Salada and Saladete groups.

Salada/ Saladete/ PCA-LDA									
PP	SL	MC	Salada (n = 90)	Saladete (n = 235)	AC	Sens	Spec	PPV	PNV
no treat.	_	10	89	233	99.08	98.89	95.88	97.8	99.57
NSS	_	10	89	232	98.77	98.89	95.87	96.74	99.57
1SG (4+4)	_	10	89	235	99.69	98.89	95.92	100	99.58
1SG (6+6)	_	10	90	234	99.69	100	95.90	98.9	100
2SG (4+4)	_	10	89	223	96.00	98.89	95.71	88.12	99.55
Salada/ Saladete/ PLS-DA									
PP	SL	MC	Salada (90)	Saladete (235)	AC	Sens	Spec	PPV	PNV
no treat.	0.5	9	89	234	99.38	98.89	95.90	98.89	99.57
no treat.	0.4	9	90	234	99.69	100	95.90	98.9	100
NSS	0.5	8	88	234	99.08	97.78	95.90	98.88	99.15
NSS	0.4	8	90	233	99.38	100	95.88	97.83	100
1SG (4+4)	0.5	5	87	235	99.08	96.67	95.92	100	98.74
1SG (4+4)	0.4	5	89	234	99.38	98.89	95.90	98.89	99.57
1SG (6+6)	0.5	6	88	234	99.08	97.78	95.90	98.88	99.15
1SG (6+6)	0.4	6	90	232	99.08	100	95.87	96.77	100
2SG (4+4)	0.5	7	85	235	98.46	94.44	95.92	100	97.92
2SG (4+4)	0.4	7	89	234	99.38	98.89	95.9	98.89	99.57

PP: pre-processing; SL: separation limit; MC: main components; AC: accuracy; Sens: sensitivity; Spec: specificity; PPV: prediction of positive values; PNV: prediction of negative values; NSS - normal signal standardization; no treat.: no treatment; 1SG: Savitzky-Golay first derivative with 4 and 6 derivation points; 2SG: Savitzky-Golay second derivative with 4 and 6 derivation points.

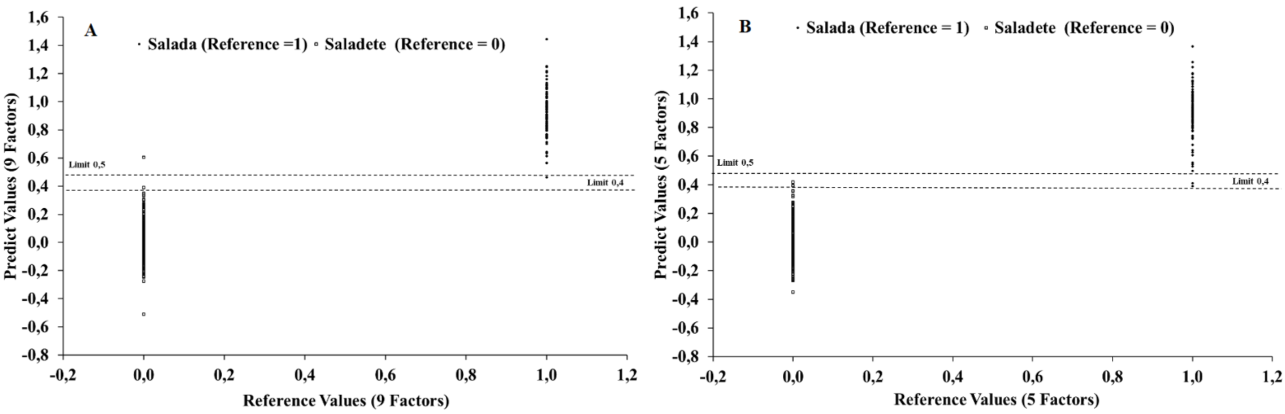

Figure 4. Representation of the total cross-validation results of the Partial Least Squares calibration model for Discriminant Analysis with a spectrum in the range of 384-1131 nm for seedlings of the Salada (n = 90) and Saladete (n = 235) groups. A) Non-processed absorbance spectral models; B) Pre-processed spectrum model with first Savitzky-Golay derivative with 4-point window.

Separation of hybrids from the Salada group

The best individual PLS-DA models for seedlings of tomato hybrids in the Salada group presented accuracy greater than or equal to 90% (Table 3). For the cultivar Dominador, it was possible to identify 29 of the 30 analyzed data, resulting in an accuracy of 95.6 and a false negative value of 98.3%. The PLS-DA model for the ParonNTY hybrid performed an accuracy of 93.3 and a false negative rate of 96.6%. The models for both hybrids were elaborated after pre-processing with second derivative of Savitzky-Golay and presented sensitivity and specificity of 96.7 and 85.3% for Dominator, respectively, and of 93.3 and 84.9% for ParonNTY, respectively (Table 4). The specificity values were obtained according to the Equation 3.

$$Specificity: \frac{TN}{((TN+FP)\times 100)} \quad (3)$$

where:

TN means true negative and FP means false positive.

For the Dylla cultivar (Table 4), the calibration model with the best performance was after spectral pre-processing with the Savitzky-Golay first derivative, showing accuracy of 90.0, sensitivity of 93.33 and specificity of 84.1% (Table 4).

These results are satisfactory when compared to the results for separating cultivars in macadamia fruit, which can be explained by the cross-pollination of the fruits (Carvalho et al., 2018). Consequently, this shows that regarding the identification of variety, hybrids or species, the use of NIRS technology tends to be more interesting for vegetative parts, as in this case, seedlings.

Table 4. Results of the calibration models in the spectral length between 384 and 1131nm for the cultivars Dominador, Dylla and ParonNTY.

				Dominador						
Model	PP	SL	MC	Dominador (n = 30)	Others (n = 60)	AC	Sens	Spec	PPV	PNV
PLS-DA	no treat.	0.5	9	23	55	86.67	76.67	84.62	82.14	88.71
PLS-DA	1SG(4+4)	0.4	6	29	52	90.00	96.67	83.87	78.38	98.11
PLS-DA	2SG(6+6)	0.5	15	29	58	96.67	96.67	85.29	93.55	98.31
				Dylla						
Model	PP	SL	MC	Dylla (n = 30)	Others (n = 60)	AC	Sens	Spec	PPV	PNV
PLS-DA	no treat.	0.5	4	24	55	87.78	80.00	84.62	82.76	90.62
PLS-DA	1SG(4+4)	0.4	10	28	53	90.00	93.33	84.13	80.00	96.36
PLS-DA	2SG(6+6)	0.5	11	26	54	88.89	86.67	84.38	81.25	93.10
				Paronnty						
Model	PP	SL	MC	ParonNTY (n = 30)	Others (n = 60)	AC	Sens	Spec	PPV	PNV
PLS-DA	no treat.	0.5	14	27	57	93.33	90.00	85.07	90.00	95.00
PLS-DA	1SG(4+4)	0.4	11	28	54	91.11	93.33	84.38	82.35	96.43
PLS-DA	2SG(6+6)	0.4	12	28	56	93.33	93.33	84.85	87.50	96.55

PP: pre-processing; SL: separation limit; MC: main components; AC: accuracy; Sens: sensitivity; Spec: specificity; PPV: prediction of positive values; PNV: prediction of negative values; no treat.: no treatment; 1SG: Savitzky-Golay first derivative with 4 and 6 derivation points; 2SG: Savitzky-Golay second derivative with 4 and 6 derivation points.

Separation of the Saladete group hybrids

The individual PLS-DA calibration models for seedlings of tomato hybrids of the Saladete group presented a general accuracy above 86% (Table 3). However, in specific models and with unbalanced sample numbers, accuracy is not a very relevant parameter to be observed in isolation. Thus, the sensitivity parameter must be examined instead of the accuracy, which allows us to assess with confidence the result obtained for a sample of the labeled class (the hybrid of the specific model) (Morais & Lima, 2018).

Taking into account the sensitivity and accuracy parameters, two models are listed; PLS-DA for the hybrids Parma (94.5 Accuracy and 90% Sensitivity) and BS-110012 (98.3 Accuracy and 93.1% Sensitivity), both constructed after spectral processing with the second derivative of Savitzky-Golay (Table 5).

By evaluating three performance parameters of accuracy, sensitivity and specificity models (Morais & Lima, 2018), we can list as satisfactory the results of the PLS-DA models for Giácomo hybrids (90.6, 70.0 and 95.1% accuracy, sensitivity and specificity, respectively), and Guará (95.7, 83.3 and 95.2% accuracy, sensitivity and specificity, respectively), both after processing with the second derivative of Savitzky-Golay. Also, the Tyna hybrid (94.0, 89.7 and 95.1% accuracy, sensitivity and specificity, respectively) after application of the first Savitzky-Golay derivative (Table 5).

PLS-DA models for the calibration of tomato seedlings from hybrids Totalle (88.9, 34.5 and 95.2% accuracy, sensitivity and specificity, respectively), Ravena (86.8, 10.0 and 95.3% accuracy, sensitivity and specificity, respectively) and Helena (90.2, 33.3 and 95.3% accuracy, sensitivity and specificity, respectively) showed unsatisfactory results, especially regarding sensitivity, which is the parameter that allows the identification of the hybrid object of the model (Table 5). This result may have occurred due to the similarity between the hybrids, which was also observed by Carvalho et al. (2018) in their study with macadamia.

PLS-DA models combined with pre-processing, mainly the derivatives proposed by Savitzky-Golay, presented promising results for classification and identification of most of the hybrids of tested tomato plants, as proposed by Soares et al. (2017) for wood from species originating in the Amazon. However, the fact that the hybrids present genetic materials more similar than the species made it difficult to create the spectral identity by the NIR of some hybrids (Totalle, Ravena, and Helena, in Table 5), using the tools proposed in this study. Thus, in order to obtain robust models, there is a need of a greater number of hybrids and the use of chemometric tools for the selection of spectral variables, such as a genetic algorithm (Carvalho et al., 2018), regression by partial least quadratic interval (Nørgaard et al., 2000), or optimization of the PLS wavelength window (Guthrie, Walsh, Reid, & Liebenberg, 2005; Cunha Júnior, Teixeira, Nardini, & Walsh, 2016).

Table 5. Results of the calibration models in the spectral length between 384 and 1131 nm for cultivars Parma, Tyna, BS110012, Totalle, Ravena, Helena, Giácomo and Guará.

				Parma						
Model	PP	SL	MC	Parma (= 30)	others (n = 205)	AC	Sens	Spec	PPV	PNV
PLS-DA	no treat.	0.4	7	19	200	93.19	63.33	95.24	79.17	94.79
PLS-DA	1SG(4+4)	0.4	5	21	199	93.62	70	95.22	77.78	95.67
PLS-DA	1SG(6+6)	0.4	5	21	199	93.62	70	95.22	77.78	95.67
PLS-DA	2SG(6+6)	0.4	16	27	195	94.47	90	95.12	72.97	98.48
				Tyna						
Model	PP	SL	MC	Tyna (= 29)	others (n = 206)	AC	Sens	Spec	PPV	PNV
PLS-DA	no treat.	0.4	10	15	193	88.51	51.72	95.07	53.57	93.24
PLS-DA	1SG(4+4)	0.4	15	26	195	94.04	89.66	95.12	70.27	98.48
PLS-DA	1SG(6+6)	0.4	6	10	201	89.79	34.48	95.26	66.67	91.36
PLS-DA	2SG(6+6)	0.4	9	18	192	89.36	62.07	95.05	56.25	94.58
				BS 110012						
Model	PP	SL	MC	BS 110012 (= 29)	others (n = 206)	AC	Sens	Spec	PPV	PNV
PLS-DA	no treat.	0.4	10	22	199	94.04	75.86	95.22	75.86	96.6
PLS-DA	1SG(4+4)	0.4	6	21	198	93.19	72.41	95.19	72.41	96.12
PLS-DA	1SG(6+6)	0.4	6	22	198	93.62	75.86	95.19	73.33	96.59
PLS-DA	2SG(6+6)	0.5	6	27	204	98.3	93.1	95.33	93.1	99.03
				Totalle						
Model	PP	SL	MC	Totalle (= 29)	others (n = 206)	AC	Sens	Spec	PPV	PNV
PLS-DA	no treat.	0.4	10	7	199	87.66	24.14	95.22	50	90.05
PLS-DA	1SG(4+4)	0.4	7	8	200	88.51	27.59	95.24	57.14	90.5
PLS-DA	1SG(6+6)	0.4	7	10	199	88.94	34.48	95.22	58.82	91.28
PLS-DA	2SG(6+6)	0.4	5	4	200	86.81	13.79	95.24	40	88.89
				Ravena						
Model	PP	SL	MC	Ravena (= 30)	others (n = 205)	AC	Sens	Spec	PPV	PNV
PLS-DA	no treat.	0.4	1	0	205	87.23	0	95.35	_	87.23
PLS-DA	1SG(4+4)	0.4	3	0	205	87.23	0	95.35	_	87.23
PLS-DA	1SG(6+6)	0.4	5	3	201	86.81	10	95.26	42.86	88.16
PLS-DA	2SG(6+6)	0.4	2	0	205	87.23	0	95.35	_	87.23
				Helena						
Model	PP	SL	MC	Helena (= 30)	others (n = 205)	AC	Sens	Spec	PPV	PNV
PLS-DA	no treat.	0.4	9	4	199	86.38	13.33	95.22	40	88.44
PLS-DA	1SG(4+4)	0.5	7	0	203	86.38	0	95.31	0	87.12
PLS-DA	1SG(6+6)	0.4	6	5	199	86.81	16.67	95.22	45.45	88.84
PLS-DA	2SG(6+6)	0.4	9	10	202	90.21	33.33	95.28	76.92	90.99
				Giácomo						
Model	PP	SL	MC	Giácomo (= 30)	others (n = 205)	AC	Sens	Spec	PPV	PNV
PLS-DA	no treat.	0.4	7	18	193	89.79	60	95.07	60	94.15
PLS-DA	1SG(4+4)	0.4	5	19	192	89.79	63.33	95.05	59.38	94.58
PLS-DA	1SG(6+6)	0.4	4	18	193	89.79	60	95.07	60	94.15
PLS-DA	2SG(6+6)	0.4	5	21	192	90.64	70	95.05	61.76	95.52
				Guará						
Model	PP	SL	MC	Guará (= 28)	others (n = 207)	AC	Sens	Spec	PPV	PNV
PLS-DA	no treat.	0.4	17	26	200	96.17	86.67	95.24	83.87	98.04
PLS-DA	1SG(4+4)	0.4	13	26	203	97.45	86.67	95.31	92.86	98.07
PLS-DA	1SG(6+6)	0.4	14	24	202	96.17	80	95.28	88.89	97.12
PLS-DA	2SG(6+6)	0.4	10	25	200	95.74	83.33	95.24	83.33	97.56

PP: pre-processing; SL: separation limit; MC: main components; AC: accuracy; Sens: sensitivity; Spec: specificity; PPV: prediction of positive values; PNV: prediction of negative values; no treat.: no treatment; 1SG: Savitzky-Golay first derivative with 4 and 6 derivation points; 2SG: Savitzky-Golay second derivative with 4 and 6 derivation points.

Conclusion

The portable NIR equipment he was capable of discriminating tomato seedlings, of Salada and Saladete groups with the PLS-DA model, with an accuracy of 99.7 and a sensitivity of 100%.

For individual models of tomato seedlings in the Salada group, the methodology proved to be efficient, with accuracy above 90 and sensitivity above 93% for all models. As for the individual models of the Saladete group, the methodology proved to be efficient for the hybrids Parma, BS-110012, Giácomo, Guara and Tyna.

Portable NIR spectroscopy proved to be a tool to verify fraud in tomato seedling trade through. Considering the ability to discriminate samples, future works should be focused on identifying the original location of product to ensure the traceability.

Acknowledgements

We are grateful for the financial support of the Ministry of Science, Technology, Innovations and Communications 'TED No. 14.0030.00/2018', and National Council for Scientific and Technological Development (Process No. 406617/2018-0).

References

Alvarenga, M. A. R. (2013). *Tomate: produção em campo, em casa-de-vegetação e em Hidroponia* (2. ed.). Lavras, MG: UFLA.

Amodio, M. L., Ceglie, F., Chaudhry, M. M. A., Piazzolla, F., & Colelli, G. (2017). Potential of NIR spectroscopy for predicting internal quality and discriminating among strawberry fruits from different production systems. *Postharvest Biology and Technology, 125*, 112-121. DOI: https://doi.org/10.1016/j.postharvbio.2016.11.013

Beghi, R., Giovenzana, V., Tugnolo, A., & Guidetti, R. (2018). Application of visible/near infrared spectroscopy to quality control of fresh fruits and vegetables in large-scale mass distribution channels: a preliminary test on carrots and tomatoes. *Journal of the Science of Food and Agriculture, 98*(7), 2729-2734. DOI: https://doi.org/10.1002/jsfa.8768

Botelho, B. G., Reis, N., Oliveira, L. S., & Sena, M. M. (2015). 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, 181*, 31-37. DOI: https://doi.org/10.1016/j.foodchem.2015.02.077

Brimmer, P. J., & Hall, J. W. (2001). Method development and implementation of near-infrared spectroscopy in industrial manufacturing support laboratories. In P. Williams, & K. Norris (Eds.)*, Near-infrared technology: in the agricultural and food industries* (2nd ed., p. 187-199). St. Paul, MN: American Association of Cereal Chemists.

Brown, C. D., Vega-Montoto, L., & Wentzell, P. D. (2000). Derivative preprocessing and optimal corrections for baseline drift in multivariate calibration. *Applied Spectroscopy, 54*(7), 1055-1068. DOI: https://doi.org/10.1366/0003702001950571

Canneddu, G., Cunha Júnior, L. C., & Teixeira, G. H. A. (2016). Quality evaluation of shelled and unshelled macadamia nuts by means of near-infrared spectroscopy (NIR). *Journal of Food Science, 81*(7), C1613-C1621. DOI: https://doi.org/10.1111/1750-3841.13343

Carvalho, L. C., Morais, C. L. M., Lima, K. M. G., Leite, G. W. P., Oliveira, G. S., Casagrande, I. P., ... Teixeira, G. H. A. (2018). Using intact nuts and near infrared spectroscopy to classify macadamia cultivars. *Food Analytical Methods, 11*(7), 1857-1866. DOI: http://dx.doi.org/10.1007/s12161-017-1078-9

Cunha Júnior, L. C., Nardini, V., Khatiwada, B. P., Teixeira, G. H. A., & Walsh, K. B. (2015). Classification of intact açaí (*Euterpe oleracea* Mart.) and juçara (*Euterpe edulis* Mart) fruits based on dry matter content by means of near infrared spectroscopy. *Food Control, 50*, 630-636. DOI: https://doi.org/10.1016/j.foodcont.2014.09.046

Cunha Júnior, L. C., Teixeira, G. H. A., Nardini, V., & Walsh, K. B. (2016). Quality evaluation of intact açaí and juçara fruit by means of near infrared spectroscopy. *Postharvest Biology and Technology, 112*, 64-74. DOI: https://doi.org/10.1016/j.postharvbio.2015.10.001

Dale, J. M., & Klatt, L. N. (1999). Principal component analysis of diffuse near-infrared reflectance data from paper currency. *International Journal of Forensic Document Examiners, 5*, 123-129.

Diniz, K. A., Guimarães, S. T. M. R., & Luz, J. M. Q. (2006). Húmus como substrato para a produção de mudas de tomate, pimentão e alface. *Bioscience Journal, 22*(3), 63-70.

Golic, M., & Walsh, K. B. (2006). Robustness of calibration models based on near infrared spectroscopy for the in-line grading of stonefruit for total soluble solids content. *Analytica Chimica Acta, 555*, 286-291. DOI: https://doi.org/10.1016/j.aca.2005.09.014

Gómez, A. H., Wang, J., Hu, G., & Pereira, A. G. (2006). Electronic nose technique potential monitoring mandarin maturity. *Sensors and Actuators B: Chemical, 113*(1), 347-353. DOI: https://doi.org/10.1016/j.snb.2005.03.090

Guthrie, J. A., Walsh, K. B., Reid, D. J., & Liebenberg, C. J. (2005). Assessment of internal quality attributes of mandarin fruit. 1. NIR calibration model development. *Australian Journal of Agricultural Research, 56*(4), 405-416. DOI: https://doi.org/10.1071/AR04257

Morais, C. L. M., & Lima, K. M. G. (2018). Principal component analysis with linear and quadratic discriminant analysis for identification of cancer samples based on mass spectrometry. *Journal of the Brazilian Chemical Society, 29*(3), 472-481. DOI: https://doi.org/10.21577/0103-5053.20170159

Muñiz, G. I. B., Magalhães, W. L. E., Carneiro, M. E., & Viana, L. C. (2012). Fundamentos e estado da arte da espectroscopia no infravermelho próximo no setor de base florestal. *Ciência Florestal, 22*(4), 865-875. DOI: https://doi.org/10.5902/198050987567

Naes, T., Isaksson, T., Fearn, T., & Davies, T. (2002). *A user-friendly guide to multivariate calibration and classification.* Chichester, GB: NIR Publications.

Nicolaï, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biology and Technology, 46*(2), 99-118. DOI: https://doi.org/10.1016/j.postharvbio.2007.06.024

Nørgaard, L., Saudland, A., Wagner, J., Nielsen, J. P., Munck, L., & Engelsen, S. B. (2000). Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy. *Applied Spectroscopy, 54*(3), 413-419. DOI: https://doi.org/10.1366/0003702001949500

Oliveira, V. S., Honorato, R. S., Honorato, F. A., & Pereira, C. F. (2018). Authenticity assessment of banknotes using portable near infrared spectrometer and chemometrics. *Forensic Science International, 286*, 121-127. DOI: https://doi.org/10.1016/j.forsciint.2018.03.001

Ozaki, Y., McClure, W. F., & Christy, A. A. (2006). *Near-infrared spectroscopy in food science and technology*. Hoboken, NJ: John Wiley & Sons.

Pagliuca, L. G., Deleo, J. P. B., Boteon, M., Mueller, S., & Valmorbida, J. (2017). Análise da economicidade da produção de tomate de mesa em diferentes escalas de produção na região de Caçador/SC. *Custos e @gronegócio on line, 13*, 227-243.

Silva, C. S., Borba, F. S. L., Pimentel, M. F., Pontes, M. J. C., Honorato, R. S., & Pasquini, C. (2013). Classification of blue pen ink using infrared spectroscopy and linear discriminant analysis. *Microchemical Journal, 109*, 122-127. DOI: https://doi.org/10.1016/j.microc.2012.03.025

Snel, F. A., Braga, J. W. B., Silva, D., Wiedenhoeft, A. C., Costa, A., Soares, R., ... Pastore, T. C. M. (2018). Potential field-deployable NIRS identification of seven *Dalbergia* species listed by CITES. *Wood Science and Technology, 52*(5), 1411-1427. DOI: https://doi.org/10.1007/s00226-018-1027-9

Soares, L. F., Silva, D. C., Bergo, M. C. J., Coradin, V. T. R., Braga, J. W. B., & Pastore, T. C. M. (2017). Avaliação de espectrômetro NIR portátil e PLS-DA para a discriminação de seis espécies similares de madeiras amazônicas. *Química Nova, 40*(4), 418-426. DOI: https://doi.org/10.21577/0100-4042.20170014

Souza, A. M., & Poppi, R. J. (2012). Experimento didático de quimiometria para análise exploratória de óleos vegetais comestíveis por espectroscopia no infravermelho médio e análise de componentes principais: um tutorial, parte I. *Química Nova, 35*(1), 223-229. DOI: https://doi.org/10.1590/S0100-40422012000100039

Zhang, Q., Li, Q., & Zhang, G. (2012). Rapid determination of leaf water content using VIS/NIR spectroscopy analysis with wavelength selection. *Journal of Spectroscopy, 27*(2), 93-105. DOI: https://doi.org/10.1155/2012/276795