Available online at www.sciencedirect.com
ScienceDirect
Journal of Food Engineering 82 (2007) 395–401
JOURNAL OF FOOD ENGINEERING
www.elsevier.com/locate/jfoodeng

Combination and comparison of chemometrics methods for identification of transgenic tomatoes using visible and near-infrared diffuse transmittance technique

Lijuan Xie, Yibin Ying *, Tiejin Ying

College of Biosystems Engineering and Food Science, Zhejiang University, 310029 Hangzhou, PR China

Received 2 February 2007; received in revised form 25 February 2007; accepted 27 February 2007
Available online 12 March 2007

Abstract

The potential of visible/near-infrared (Vis/NIR) diffuse transmittance spectroscopy, in combination with a variety of chemometrics techniques, was investigated to examine the feasibility to non-destructively distinguish transgenic tomatoes from non-transgenic tomatoes. One hundred tomatoes inserted with antisense ethylene receptor gene *LeETR*1 and the same number of their parents were scanned in the Vis/NIR regions. Principal component analysis (PCA), soft independent modelling of class analogy (SIMCA) and discriminant partial least squares (DPLS) regression based on PCA scores were applied to classify tomatoes with different genes into two groups. The results show that differences between transgenic and non-transgenic tomatoes do exist and excellent classification can be obtained after optimizing spectral pre-treatment. The correct classifications of the calibration as well as the validation data set for transgenic and non-transgenic tomatoes were 100% using DPLS after second derivative spectral pre-treatment. The results in the present study show that Vis/NIR spectroscopy together with chemometrics techniques could be a rapid tool to be used for differentiating transgenic tomatoes from conventional tomatoes.

Keywords: Vis/NIR spectroscopy; Transgenic tomato; Identification; Chemometrics

1. Introduction

Tomato (Lycopersicon esculentum Mill) plays a significant role in the agricultural market both for its large consumption and for its richness in health-related food components such as folate, vitamin C, and various carotenoids and phytochemicals. The amounts involved in its production and commercialization are impressive (Campbell, Canene-Adams, Lindshield, & Boileau, 2004; Jha & Matsuoka, 2004; Pedro & Ferreira, 2005). But tomato quality decreases rapidly after harvest (Van Dijk, Boeriu, Stolle-Smits, & Tijskens, 2006). A large annual loss due to spoilage makes a means to control ripening of great economic importance (Hoeberichts, Van Der Plas, & Woltering, 2002). An antisense ethylene receptor gene *LeETR*1 is introduced into tomato through *Agrobacterium tumefaciens*-mediated method to delay tomato ripening. However, there are global issues that arise from the use of genetic modification techniques (Zafar & Asif, 2004; Perry, 2002). To monitor and verify the presence and the amount of genetically modified organisms (GMOs) in agricultural crops and in products derived, it has generated a demand for analytical methods capable of detecting, identifying and quantifying either the DNA introduced or the protein(s) expressed in transgenic plants (Anklam, Gadani, Heinze, Pijnenburg, & Guy, 2002; Bonfini et al., 2001). Ideally an identification technique should be rapid, easy to use and of low cost (Belton et al., 1995).

Near-infrared (NIR) spectroscopy is a procedure that can detect and measure the chemical composition of

* Corresponding author. Tel.: +86 571 86971140; fax: +86 571 86971885.
E-mail address: ybying@zju.edu.cn (Y. Ying).

0260-8774/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jfoodeng.2007.02.062

biological materials based on the absorption of NIR radiation by bonds between light atoms (Tigabu & Odén, 2002). The advantage of NIR spectroscopy over other analytical techniques which have been developed, such as polymerase chain reaction (PCR), enzyme linked immunosorbent assays (ELISAs), biosenso, microarray, chip, electrophoresis, X-ray fluorescence, mass spectrometry, etc. (Zafar & Asif, 2004) in that it is nondestructive, simple to apply, fast, requires no sample pre-treatment and can easily be automated (Chen, Arnold, & Small, 2004; Gestal et al., 2004). Though NIR spectrometers are not precise enough to detect compounds at the DNA concentration level (parts per trillion), spectral differences caused by larger structural changes (if any) accompanying the modification might be measurable. Recently, this technique has been used to distinguish transgenic products from conventional ones. Roussel, Hardy, Hurburgh, and Rippke (2001) detected and segregated Roundup Ready™ soybeans from conventional soybeans using partial least-squares (PLS), locally weighted regression (LWR) and artificial neural networks (ANN) models by NIR spectroscopy. 93% accurate classification was obtained using a database of approximately 8000 samples with LWR method. Rui, Luo, Huang, Wang, and Zhang (2005) applied a back propagation (BP) algorithm to discriminate transgenic corns and their parents by continuous wave of NIR diffuse reflectance spectroscopy ranged from 830 to 2500 nm. Xie, Ying, Ying, Yu, and Fu (2007) used Vis/NIR diffuse reflectance spectroscopy combined with multivariate analysis to differentiate 70 transgenic tomatoes and 94 of their parents. Partial least-squares discriminant analysis (PLSDA) model with the leave-one-out cross-validation technique after the second derivative pre-treatment was proved to have the best satisfactory calibration and prediction ability.

The work undertaken in this current study was designed to examine the feasibility of using the visible and near infrared (Vis/NIR) spectroscopic techniques, in combination with a variety of chemometrics techniques, to distinguish transgenic tomatoes with antisense *LeETR*1 from non-transgenic tomatoes.

2. Materials and methods

2.1. Samples

The tomato samples were grown in a standard greenhouse and marked at the fruit color-breaking time. Two sets of samples with similar sizes for both varieties ($n = 200$) were obtained. One set had 98 samples (50 transgenic tomatoes with antisense *LeETR*1 and 48 of their parents, non-transgenic ones) and the other had 102 samples (50 transgenic and 52 non-transgenic ones). The experiments were conducted in different days. The samples in each set were picked and measured on the same day and they were all at the sixth stage (USDA, 1976) with the same interval from color-breaking to ripe. All the samples were cleaned and laid at room temperature (25 °C, 60% relative humidity) for two hours to equilibrate to experiment environment before Vis/NIR diffuse transmittance spectra collection. One hundred and forty samples (70 transgenic tomatoes and the same number of non-transgenic ones) were used for calibration, and the remaining 60 ones for validation. The samples for calibration and validation sets were chosen randomly.

2.2. Spectral measurement

Diffuse transmission spectra of tomato samples were collected by a fiber spectrometer system with the range from 350 to 1289 nm which consisted of a light source (Ocean Optics Inc., USA), a sample holder, an adjustable collimating len, a fiber spectrometer (Ocean Optics Inc., USA), a bifurcated optic fiber cable (Ocean Optics Inc., USA), and a computer. The light source installed in a special arc lamp-chimney was four 50 W tungsten halogen lamps which were placed at four locations, 90° apart. The sample holder acted as both a position fitting and as a flexible support to accommodate samples with different sizes, as well as a light seal against light source. The spectrometer was equipped with a 2048-element linear silicon CCD array detector.

Spectrometer parameters setting, spectra data obtaining and storing were carried out via software OOIBsae32 (Ocean Optics Inc., USA). The Integration time was 150 ms, and the resolution was 1 cm^{-1} in this work. Samples were analyzed at room temperature. The measurement was expressed as percent transmission (%).Three replicates of each sample were taken and their mean value was calculated using OMINIC 6.1a (Thermo Electron Corp., Madison, WI, USA). Reference spectrum for air and dark spectrum were measured and stored prior to sample spectra measurement.

2.3. Spectral data pre-treatment

Vis/NIR spectra are often pre-processed to reduce undesirable systematic noise, such as baseline variation, light scattering, path length differences and so on, and enhance the contribution of the chemical composition (Tigabu & Odén, 2002). In this study, three types of pre-processing were employed: multiplicative scattering correction (MSC), Savitzky–Golay first and second derivative (Geladi, Macdougall, & Martens, 1985). In our case, the possible sources of systematic variation could be due to the path length slight difference arising from the positioning of individual tomato with slight different sizes during scanning.

2.4. Chemometrics methods

Chemometrics methods can obtain direct information from the NIR spectra, highlight the chemical differences between samples and reduce variation due to physical effects. The combination of spectroscopy and chemomet-

rics methods was investigated for quantitative and qualitative analysis. Principal component analysis (PCA), soft independent modelling of class analogy (SIMCA) and discriminant partial least squares (DPLS) have proven to be effective in many applications (Andre, 2003; Chen, Zhao, Zhang, Liu, & Fang, 2005; Cozzolino, Smuth, & Gishen, 2003; Sáiz-Abajo, González-Sáiz, & Pizarro, 2004), and were therefore used in the present study to classify tomatoes with different genes. All the chemometrics analysis was carried out using the commercial software package, TQ Analyst v6.2.1 (Thermo Nicolet Corporation, Madison, WI, USA).

2.4.1. PCA

PCA is a method applied to get an overview of the data by extracting the main information in the Vis/NIR spectra recorded on tomato samples, reduce the number of variables, build up new variables, known as principal components (PCs) and detect possible spectral outliers. Each spectrum will have its own unique set of scores; therefore, a spectrum can be represented by its PCA scores in the factor space instead of intensities in the wavelength space (Park, Abbott, Lee, Choi, & Choi, 2003). The PCA transforms the original independent variables (wavelengths) into new axes, or PCs. The PCs and thus the axes are uncorrelated, but the data presented on these axes will still be correlated (Martens & Naes, 1989). By plotting the PCs, one can view interrelationships between different variables, detect and interpret sample patterns, groupings, similarities or differences (Mouazen, Karoui, De Baerdemaeker, & Ramon, 2006).

PCA either on the pure spectral data or the pre-treated data can provide very important information regarding the potential capability of separation of objects. As PCA is fitting a subspace with respect to the optimized maximum variance of the data structure, a further analysis has been applied for an improved separation of the groups (Andre, 2003).

2.4.2. SIMCA

SIMCA is an elaborate method based on PCA (Sáiz-Abajo et al., 2004). With SIMCA, each sample type is modeled independently of the others and the process does not attempt to maximize inter-cluster distances. New samples are treated separately by each cluster model and an assessment is made of cluster membership on the basis of the distance to the cluster centroid. Additionally, an evaluation of the likelihood of an unknown to samples in each cluster concerned may be made on the basis of an *F*-test (Blanco et al., 2001); this compares the sum of squares of a residual spectrum (the difference between the actual spectrum of a sample and the spectrum produced by the model) to the variance within a class, thus providing a measure of certainty which may be attached to each identification. SIMCA was reported to produce very high correct classification rate in the separation of very similar materials (Downey, McIntyre, & Davies, 2002).

2.4.3. DPLS

Leverage diagnostic was applied to detect the outlier sample and provide the information that how much influence each sample had on the method standards. The diagnostic routine calculates the mean spectrum for all the standards and then measures the distance between the mean spectrum and the spectrum of each standard. If a sample had leverage value that was noticeably different from the leverage values of other samples, it was examined closely in order to know whether it provided any useful information or it must be removed. The information can help to identify samples that might be outliers and select useful standards.

Once the outlier detection was performed, DPLS together with Vis/NIR spectra was used to establish models for quantifying transgenic and non-transgenic tomatoes. It is a PLS application for the optimum separation of classes and each sample was assigned a dummy variable as a reference value, which is an arbitrary number or letter indicating whether the sample belongs to a particular group or not (Cozzolino et al., 2003). In this case, samples of transgenic tomato were assigned a numeric value of 2, and those of non-transgenic tomato assigned 1. The DPLS model was then developed by assigning the reference value (dummy variable) for each sample. A sample was considered to be correctly categorized if the predicted value lay on the two sides of the assigned values, 0.5 is the cutoff criteria which is similar to those reported by others (Cozzolino et al., 2003; Andre, 2003). The optimum number of factors used in DPLS was determined by minimum value of predicted residual error sum of squares (PRESS). Statistics calculated for the DPLS calibration models included root mean square error of calibration (RMSEC), root mean square error of prediction (RMSEP) and correlation coefficient *r*. It is expected to have ideal models with the lower RMSEC and RMSEP as well as the higher *r*.

3. Results and discussion

3.1. The diffuse transmittance spectra of tomatoes

Fig. 1 shows the diffuse transmittance raw spectra and second derivative of average spectra on two varieties of tomatoes. There are many crossovers and overlapping among these samples in the Fig. 1a which means the spectra are very similar. Therefore, it is difficult to discriminate the varieties directly based on diffuse transmittance raw spectra. Fig. 1b indicates the variations in the second derivative of the average spectra and the transmission percent for averaged spectrum of transgenic tomatoes at 750–820 nm, which is related to the third and forth overtone of C–H stretching vibrations and the first overtone of C=C stretching vibration, is remarkable lower than that of non-transgenic tomatoes. Differences in composition and/or structure among samples with different genes can be reflected in the Vis/NIR spectrum, which can theoretically be used to develop a classification method.

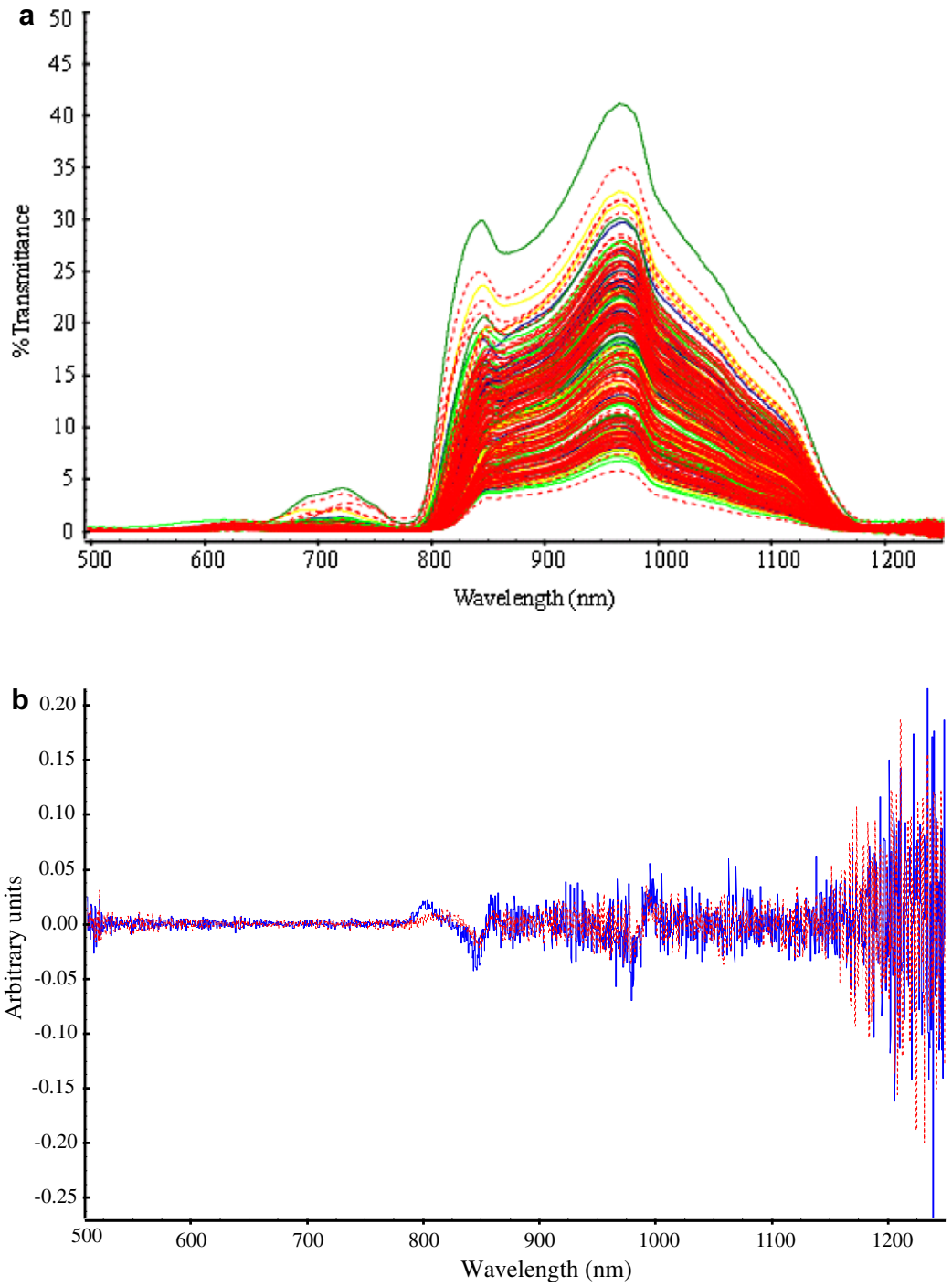

Fig. 1. Vis/NIR raw spectra (a) and second derivative of average spectra (b) recorded on transgenic (---) and non-transgenic (—) tomatoes.

3.2. PCA

PCA is an effective data mining technique and it can find an alternative set of coordinate axes. PCA was performed as the first attempt to extract and visualize the main information in multivariate data to examine qualitative differences between the two kinds of samples. MSC, Savitzky–Golay first and second derivative pre-processing were employed and the second derivative of the spectra which has turned out to be the best data pretreatment for optimum separation of both groups in this study was used. The number of smoothing point is seven. All the spectra of samples were used for PCA. The initial 10 PCs, which account for the most spectral variations 90.410% and indicated as positive or negative, are used to make differentiation clearer. Vis/NIR spectra in derivative model in the range from 600 to 1200 nm were found to be the most important and the spectral data with low signal to noise ratio on both sides were removed.

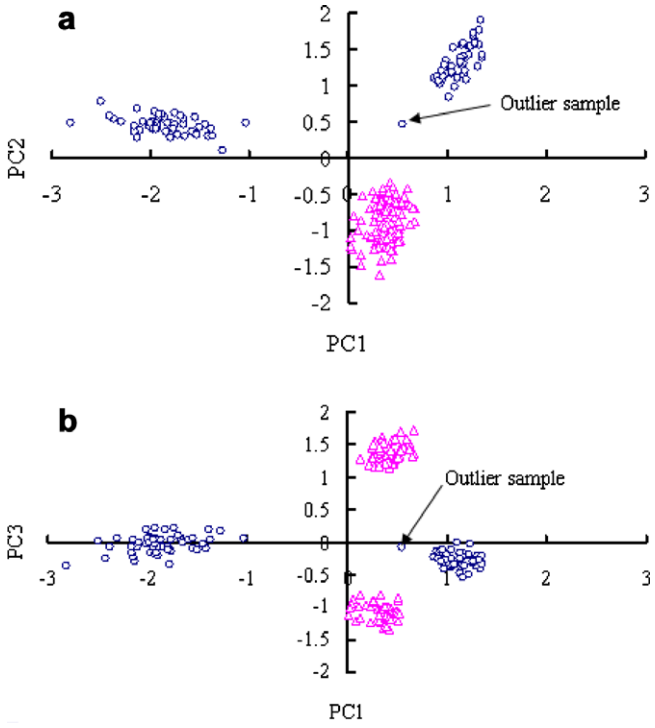

Fig. 2. PCA score plots for the first three principal components. Triangular marks represent transgenic samples and circles represent non-transgenic samples.

Fig. 2 shows the two-dimensional (2D) principal component score plot using first three score vectors, PC1, PC2, PC3 derived from the second derivative spectra of the samples and indicates how these principal components are able to separate the objects. From this figure, we can find that the samples are divided clearly into three or four groups without overlapping each other. It can also be found that samples originating from the same gene in two picked sets appear to separate between themselves into two corresponding groups because they might not be picked on the same days and had biological variation and different chemical ingredients. Fig. 2 also shows that non-transgenic samples have scores between −2.81–1.04 to 0.88–1.36 except a sample with 0.55 score in the first component, between 0.09–1.89 and −0.55–0.20 in the second and third components, respectively, however, transgenic samples from 0.02 to 0.67, −1.61 to −0.35, −1.36–0.81 to 1.16–1.73 in the first three components, respectively. Fig. 2a also shows that non-transgenic tomatoes forming two groups are located on the top half of the plot while transgenic ones also composing of one cluster are observed on the reverse half, which indicates that non-transgenic samples have positive scores in the second component and transgenic ones have the negative ones. The non-transgenic sample appearing among the transgenic ones in the first component could have been an outlier detected through leverage diagnostic that will be discussed later. In general, the figure clearly indicated a potential for distinguishing between the groups and that no more components than the first three principal components were needed to differentiate more clearly between groups due to no overlapping of the samples.

The result suggests the discrimination between transgenic and non-transgenic tomatoes is possible and that different spectral attributes of samples are associated with characteristics of the sample. As PCA only indicates the visualizing dimension spaces, a further quantitative analysis was utilized for an improved separation.

3.3. SIMCA

A summary of SIMCA results obtained is shown in Table 1. It can be seen in this table that models discriminating tomatoes with a degree of accuracy range from 83.5% to 97%, which demonstrated the good discriminatory power to differentiate the two varieties. The most successful model, using the first three principal components, was derived from raw spectral data. In this case, 99 of the non-transgenic tomatoes (99% accurate classification) were correctly identified as were 95 out of the 100 transgenic tomatoes (95% accurate classification). MSC process can not improve accuracy but deteriorate it for non-transgenic tomatoes but do yield a better classification of transgenic ones compared with raw spectra. The identification of transgenic tomatoes was more constant, ranging from a low of 95 to a high of 100 out of 100 samples. That is to say, SIMCA method may be sufficiently sensitive to be of practical utility. However, compared with PCA (Fig. 2), SIMCA method couldn't achieve 100% accurate classification although it is based on PCA.

3.4. DPLS

Leverage diagnostic result shows that there are several samples with leverage values that are noticeably different from those for the other samples. After reexamination, these samples would be removed as outliers according to various models.

Table 2 shows the calibration and validation statistics for the number of incorrectly classified samples with MSC and derivative spectra pre-treatment using region from 600 to 1200 nm. The derivative can increase classification accuracy compared with raw spectra and MSC process. The optimum models involved the use of spectral

Table 1
Performance of SIMCA models

Data pre-treatment	No. of incorrectly classified samples	
	Non-transgenic tomatoes ($n = 100$)	Transgenic tomatoes ($n = 100$)
None	1	5
1st derivative	5	4
2nd derivative	5	2
MSC	23	1
MSC + 1st derivative	33	0
MSC + 2nd derivative	27	0

Table 2
Tomato classification results of DPLS models

Data pre-treatment	No. of incorrectly classified samples				Correct percent
	Non-transgenic tomatoes		Transgenic tomatoes		
	Calibration set	Validation set	Calibration set	Validation set	
None	9 (1)	2 (2)	8 (1)	3	88.8
1st derivative	9 (2)	5 (1)	6	3	88.3
2nd derivative	0	0 (1)	0	0	100
MSC	7	3	13	8	84.5
MSC + 1st derivative	10	3	4 (1)	2	90.5
MSC + 2nd derivative	0 (1)	0	0 (1)	0 (3)	100

The numbers in brackets represent the number of outlier samples.

data after the second derivative data pre-treatment step and the percent correct classifications of the calibration as well as the validation data set for transgenic and non-transgenic tomatoes were 100%.

Table 3 is the statistics for tomato samples using DPLS models on the raw spectra and spectra with various pre-treatment. The results indicate that the DPLS models developed on the derivative spectra showed the better statistics compared with raw spectra. During development of PLS regression models, some samples were removed as outliers from all further data analysis. The best prediction results were obtained using the raw spectra with the second derivative pre-treatment, with $r = 0.98412$, RMSEC value of 0.0887, RMSEP value of 0.0961 with one outlier removed.

Table 3
The statistics for tomato samples using DPLS models on the raw spectra and spectra with various pre-treatment

Data pre-treatment	r	RMSEC	RMSEP	Factors
None	0.73584	0.339	0.317	1
1st derivative	0.78161	0.312	0.308	1
2nd derivative	0.98412	0.0887	0.0961	1
MSC	0.70509	0.355	0.355	1
MSC + 1st derivative	0.76590	0.321	0.342	1
MSC + 2nd derivative	0.95389	0.150	0.148	1

Fig. 3 shows the Vis/NIR predictions of tomato varieties in the validation set using the DPLS model with the spectra after second derivative treatment. It reveals two distinct groups. The diagonal line represents ideal results (actual = calculated value) and so the closer the points are to this, the better is the model. The transgenic sample was classified correctly if the value was between 1.5 and 2.5, else the sample was classified wrong. And it was non-transgenic sample if the value was between 0.5 and 1.5. That is to say, the samples using a predicted value of ±0.5 as a cut-off were all considered to be correctly classified by the

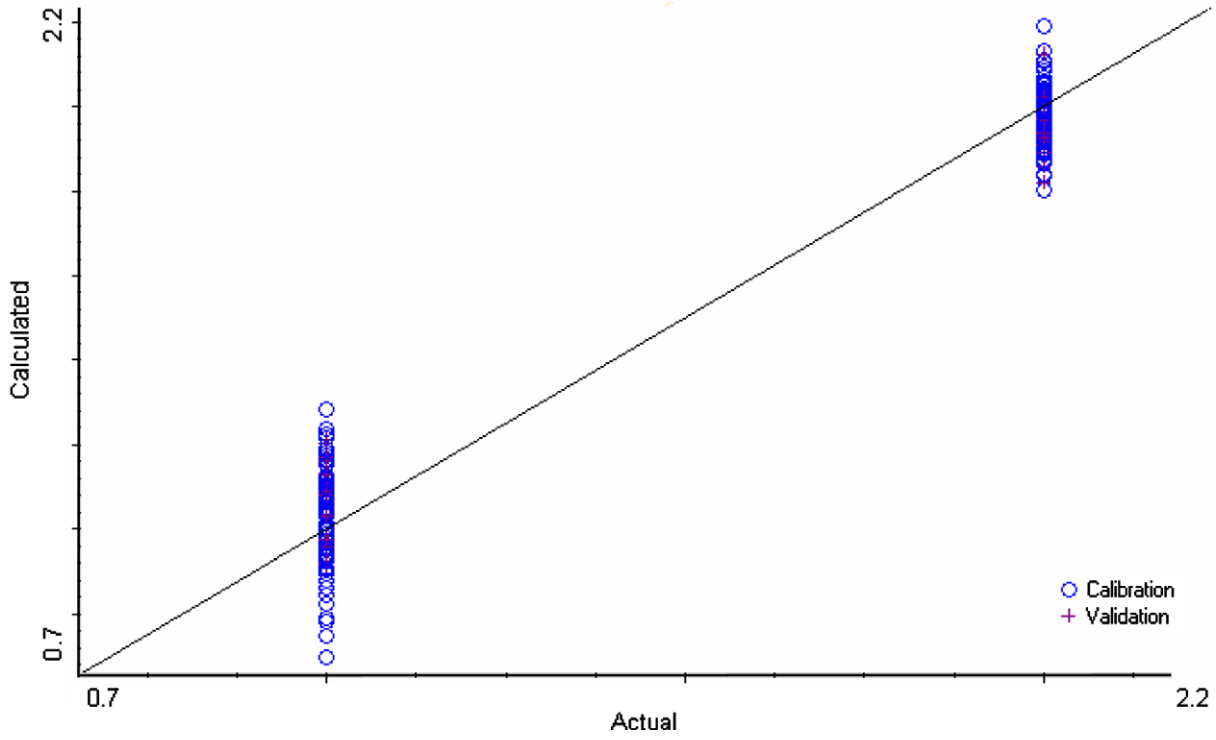

Fig. 3. Plot of actual vs calculated values of transgenic and non-transgenic tomatoes using DPLS model and second derivative (calibration and validation set).

model. Based on the vibrational responses of chemical bonds to Vis/NIR radiation, the model can discriminant or identify varieties (Fig. 3). It is probable that the higher the variability between sample types in those chemical entities, which respond in these regions of the spectrum, the better the accuracy of the model (Cozzolino et al., 2003). This suggests that DPLS models with spectra after second derivative treatment contain enough information for discriminating the samples.

4. Conclusion

In conclusion, initial attempts to detect and classify transgenic and non-transgenic tomatoes using Vis/NIR spectroscopy were successful. The correct identification of 100% of tomatoes is a highly encouraging result using the DPLS model with the spectra after second derivative treatment. These results suggest that the methods may have commercial and regulatory potential to avoid time-consuming recalibration work for each sample and costly and laborious chemical and sensory analysis.

Acknowledgements

The authors gratefully acknowledge the financial support provided by the National Natural Science Foundation of China (Grant No. 30671197) and the Program for New Century Excellent Talents in University (No. NCET-04-0524).

References

Andre, M. (2003). Multivariate analysis and classification of the chemical quality of 7-aminocephalsporanic acid using near-infrared reflectance spectroscopy. *Analytical Chemistry, 75*, 3460–3467.

Anklam, E., Gadani, F., Heinze, P., Pijnenburg, H., & Guy, V. D. E. (2002). Analytical methods for detection and determination of genetically modified organisms in agricultural crops and plant-derived food products. *European Food Research and Technology, 214*, 3–26.

Belton, P. S., Kemsley, E. K., McCann, M. C., Ttofis, S., Wilson, R. H., & Delgadillo, I. (1995). The identification of vegetable matter using Fourier transform infrared spectroscopy. *Food Chemistry, 54*(4), 437–441.

Blanco, M., Maspoch, S., Villarroya, I., Peralta, X., González, J. M., & Torres, J. (2001). Geographical origin classification of petroleum crudes from near-infrared spectra of bitumen. *Applied Spectroscopy, 55*(7), 834–839.

Bonfini, L., Heinze P., Kay S., & Van Den Eede, G. (2001). *Review of GMO detection and quantification techniques*. Ispra, Italy.

Campbell, J. K., Canene-Adams, K., Lindshield, B. L., & Boileau, T. W. M. (2004). Tomato phytochemicals and prostate cancer risk. *Journal of Nutrition, 134*, 3486S–3492S.

Chen, J., Arnold, M. A., & Small, G. W. (2004). Comparison of combination and first overtone spectral regions of near-infrared calibration models for glucose and other biomolecules in aqueous solutions. *Analytical Chemistry, 76*(18), 5405–5413.

Chen, Q., Zhao, J., Zhang, H., Liu, M., & Fang, M. (2005). Qualitative identification of tea by near infrared spectroscopy based on soft independent modeling of class analogy pattern recognition. *Journal of Near Infrared Spectroscopy, 13*, 327–332.

Cozzolino, D., Smuth, H. E., & Gishen, M. (2003). Feasibility study on the use of visible and near-infrared spectroscopy together with chemometrics to discriminate between commercial white wines of different varietal origins. *Journal of Agricultural and Food Chemistry, 51*, 7703–7708.

Downey, G., McIntyre, P., & Davies, A. N. (2002). Detecting and quantifying sunflower oil adulteration in extra virgin olive oils from the eastern Mediterranean by visible and near infrared spectroscopy. *Journal of Food and Agricultural Chemistry, 50*(20), 5520–5525.

Geladi, P., Macdougall, D., & Martens, H. (1985). Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Applied Spectroscopy, 39*(3), 491–500.

Gestal, M., Gómez-Carracedo, M. P., Andrade, J. M., Dorado, J., Fernández, E., Prada, D., et al. (2004). Classification of apple beverages using artificial neural networks with previous variable selection. *Analytica Chimica Acta, 524*, 225–234.

Hoeberichts, F. A., Van Der Plas, L. H. W., & Woltering, E. J. (2002). Ethylene perception is required for the expression of tomato ripening-related genes and associated physiological changes even at advanced stages of ripening. *Postharvest Biology and Technology, 26*(2), 125–133.

Jha, S. N., & Matsuoka, T. (2004). Non-destructive determination of acid-brix ratio of tomato juice using near infrared spectroscopy. *International Journal of Food Science and Technology, 39*(4), 425–430.

Martens, H., & Naes, T. (1989). *Multivariate Calibration* (2nd ed.). United Kingdom: Wiley, Ltd. Chichester.

Mouazen, A. M., Karoui, R., De Baerdemaeker, J., & Ramon, H. (2006). Classification of soils into different moisture content levels based on VIS-NIR spectra, Written for presentation at the 2006 ASABE Annual International Meeting Sponsored by ASABE, Oregon Convention Center, Portland, Oregon.

Park, B., Abbott, J. A., Lee, K. J., Choi, C. H., & Choi, K. H. (2003). Near-infrared diffuse reflectance for quantitative and qualitative measurement of soluble solids and firmness of delicious and gala apples. *Transactions of the ASAE, 46*(6), 1721–1731.

Pedro, A. M. K., & Ferreira, M. M. C. (2005). Nondestructive determination of solids and carotenoids in tomato products by near-infrared spectroscopy and multivariate calibration. *Analytical Chemistry, 77*, 2505–2511.

Perry, J. N. (2002). GM crops and the environment, Proc. Cambridge Soc. Application.

Roussel, S. A., Hardy, C. L., Hurburgh, C. R., & Rippke, G. R. (2001). Detection of roundup ready™ soybeans by near-infrared spectroscopy. *Applied Spectroscopy, 55*, 1425–1430.

Rui, Y. K., Luo, Y. B., Huang, K. L., Wang, W. M., & Zhang, L. D. (2005). Application of near-infrared diffuse reflectance spectroscopy to the detection and identification of transgenic corn. *Spectroscopy and Spectral Analysis, 25*, 1581–1583.

Sáiz-Abajo, M. J., González-Sáiz, J. M., & Pizarro, C. (2004). Near infrared spectroscopy and pattern recognition methods applied to the classification of vinegar according to raw material and elaboration process. *Journal of Near Infrared Spectroscopy, 12*, 207–219.

Tigabu, M., & Odén, P. C. (2002). Multivariate classification of sound and insect-infested seeds of a tropical multipurpose tree, Cordia Africana, with near infrared reflectance spectroscopy. *Journal of Near Infrared Spectroscopy, 10*, 45–51.

USDA (1976). *United States Standards for Grades of Fresh Martet Tomatoes*. Washington DC: Agricultural Marketing Service.

Van Dijk, C., Boeriu, C., Stolle-Smits, T., & Tijskens, L. M. M. (2006). The firmness of stored tomatoes (cv. Tradiro). 2. kinetic and near infrared models to describe pectin degrading enzymes and firmness loss. *Journal of Food Engineering, 77*, 585–593.

Xie, L., Ying, Y., Ying, T., Yu, H., & Fu, X. (2007). Distrimination of transgenic tomatoes based on visible/near-infrared spectra. *Analytica Chimica Acta, 584*, 379–384.

Zafar, Y., & Asif, M. (2004). *Capacity Building in Biosafety of Genetically Modified Crops: GMOs (Genetically Modified Organisms) Detection*. Faisalabad, Pakistan: National Institute for Biotechnology and Genetic Engineering.