ISSN 1984-6428
ONLINE
www.orbital.ufms.br

| Vol 2 | | No. 1 | | January-March 2010 |

Full Paper

Assessment of genetically modified soybean crops and different cultivars by Fourier transform infrared spectroscopy and chemometric analysis

Glaucia Braz Alcantara[a], Andersson Barison[b], Maiara da Silva Santos[c], Luís P. S. Santos[d], José F. F. de Toledo[e] and Antonio Gilberto Ferreira[c]*

[a]Instituto de Química, Universidade Federal de Goiás, P.O. Box 131, CEP 74001-970, Goiânia/GO, Brazil.
[b]Departamento de Química, Universidade Federal do Paraná, CEP 81531-990, Curitiba/PR, Brazil.
[c]Departamento de Química, Universidade Federal de São Carlos, P.O. Box 676, CEP 13565-905, São Carlos/SP, Brazil.
[d]Centro Federal de Educação Tecnológica do Maranhão, Av. Getúlio Vargas, 04, CEP 65030-000, São Luis/MA, Brazil.
[e]Embrapa Soja/Embrapa, Rod. Carlos João Strass, P.O. Box 231, CEP 86001-970, Londrina/PR, Brazil.

Received: 04 February 2010; revised: 29 March 2010; accepted: 29 March 2010. Available online: 03 June 2010.

ABSTRACT: *This paper describes the potentiality of Fourier transform infrared (FT-IR) spectroscopy associated to chemometric analysis for assessment of conventional and genetically modified soybean crops. Recently, genetically modified organisms have been queried about their influence on the environment and their safety as food/feed. In this regard, chemical investigations are ever more required. Thus three different soybean cultivars distributed in transgenic Roundup Ready™ soybean and theirs conventional counterparts were directly investigated by FT-IR spectroscopy and chemometric analysis. The application of PCA and KNN methods permitted the discrimination and classification of the genetically modified samples from conventional ones when they were separately analysed. The analyses showed the chemical variation according to genetic modification. Furthermore, this methodology was efficient for cultivar grouping and highlights cultivar dependence for discrimination between transgenic and non-transgenic samples. According to this study, FT-IR and chemometrics could be used as a quick, easy and low cost tool to assess the chemical composition variation in genetically modified organisms.*

Keywords: *soybean; FT-IR; chemometrics; transgenic; genetically modified organisms*

Introduction

* Corresponding author. E-mail: glabraz@quimica.ufg.br

Modified organisms, such as agricultural crops, have usually been developed for various purposes, including resistance to pests, herbicides or harsh environmental conditions, improved product shelf life and increased nutritional value [1, 2]. Whereas the use of genetically modified (GM) organisms has seen a great increase in agriculture and food science, their cultivation and commercialization have caused an enormous public debate [3]. The acceptance of new food technologies, especially of genetically modified foods, has recently attracted much attention in research, principally because food authentication is a constant concern of consumers and the food industry [4].

Due to improved production and lower agrochemical use, the supporters defending transgenic cultivars claim that they are more beneficial than the original ones and are substantially equivalent in chemical composition to non-transgenic cultivars. On the other hand, opponents to transgenic cultivars maintain that their cultivation involves many risks, such as potential allergenicity due to novel protein expression and transference of antibiotic resistance from marker genes, etc [5]. Moreover, although transgenic products are almost identical to the former ones, there are studies demonstrating that some variations could cause nutritional and toxicological consequences [6].

In this regard, many studies have shown that the general attitude towards the application of genetic modification in food production is negative [7-9]. GM products are commonly less well accepted than their conventional counterparts and the insurance of food composition influences the quality certification.

On the other hand, the herbicide tolerant Roundup Ready™ soybean (RR soybean) developed to improve resistance to glyphosate herbicide became commercially available in 1996 and is a successful example in commercialization of a genetically modified organism [4, 10]. Since then, several studies have described its assessment, although few papers have approached the chemical content differences between the transgenic soybean and its conventional counterparts, such as the detection and separation of RR soybeans from conventional soybeans using some chemometric models and near-infrared spectroscopy (NIR) [11, 12] or Fourier transform infrared photoacoustic spectroscopy (FT-IR-PAS) [13].

GM products contain an additional trait encoded by an introduced gene(s), which generally produces additional protein(s) that confers the trait of interest. Raw material (e.g. grains) and processed products (e.g. foods) derived from GM crops might thus be identified by testing for the presence of introduced DNA, or by detecting expressed novel proteins encoded by the genetic material [14]. However, other methods have been developed about chemical evaluation and detection of genetically modified organisms in food crops, such as corn/maize [15, 16] and tomatoes [17-20].

Amongst the diverse techniques applied for food evaluation, direct investigations applied to intact samples can be presented as a great alternative. In this context, Fourier Transform Infrared (FT-IR) spectroscopy can be a useful analytical technique which reduces the drawbacks connected to sample pretreatment, such as extraction, purification and the time rate for the analysis. Moreover, FT-IR spectroscopy and chemometric analysis have been demonstrated as a relevant application in highly similar matrices [21]. Both have been much used in food quality control, mainly to detect adulteration and/or geographical origin in manufactured products including wine, honey and apple juices [22]. On the other hand, FT-IR spectroscopy is a well established technique that provides highly specific molecular information of a wide range of compounds used in different fields. Recently, FT-IR technique has also been used to distinguish transgenic from conventional products [18]. FT-IR spectrometers are not precise enough to detect compounds at the DNA concentration level, though spectral differences caused by structural changes accompanying the genetic modification might be measurable [11]. Furthermore, a combination of spectroscopy and chemometric methods, adopted in order to handle the overwhelming size and complexity of data, can highlight the chemical differences between samples and provide quantitative and qualitative information [23].

Therefore, the objective of this study is to use FT-IR spectroscopy and chemometrics to distinction between different cultivars of genetically modified and conventional soybean crops. Furthermore, this report describes a simple and efficient method to provide information on the differences in chemical composition between conventional and genetically modified soybeans.

Material and Methods

Samples

Three different soybean cultivars A (BRS 133), B (BRS 134) and C (EMBRAPA 59), comprising the genetically modified (GM) version and its respective conventional version (CV), were analyzed. Both GM and CV versions of each cultivar were obtained from Embrapa Soybean (Brazilian Corporation for Agricultural Research – Soybean Center). The genetic modification of all transgenic samples was executed to induce the improvement in resistance to Roundup^{TM} herbicide (transgenic soybean variety tolerant to glyphosate).

The soybean plants were cultivated in pairs in an experimental greenhouse, with exactly the same environmental conditions to ensure that all alterations observed were related to the genetic modification and not to any other factors. Besides, it has been demonstrated that plants cultivated at different conditions could be discriminated by

chemometric analyses [24]. The dry seeds were harvested and powdered in a mortar after the removing the pericarp. No other sample manipulation was carried out for spectroscopic analysis.

FT-IR Analysis

All FT-IR spectra were acquired with 32 scans in a Bruker Equinox/55 spectrometer in the region of 4000 to 400 cm^{-1} with a spectral resolution of 4 cm^{-1}. KBr disks were prepared using 1:99 mg powdered soybean samples and dry potassium bromide, respectively, which produced translucent pellets. Fifteen replicates from different KBr pellets, which they were arisen from the same powdered amount of seeds, were analyzed for each kind of sample.

Chemometric Analysis

Principal components analysis (PCA) is used to obtain a lower dimensional graphical representation which describes the variation in a data set. In PCA, a new set of axes are defined and constructed so that a maximum amount of variation is described in a minimum number of axes. Since it reduces the dimensions required to visualize the data, PCA is a powerful method for studying multidimensional data sets and is an excellent tool for preliminary data exploration [25]. In this study, PCA was used to examine data sets for expected or unexpected clusters, including the presence of outliers.

To perform PCA analysis, all FT-IR spectra were converted into ASCII files and were then exported to Pirouette™ (Infometrix, USA) and the bands related to water absorption (4000-3000 cm^{-1}) and noisy regions (2800-1800 cm^{-1} and 900-400 cm^{-1}) were removed from FT-IR spectra, before statistical analysis, in order to ensure that chemometric models are not based on spectral differences in these regions. Finally, PCA was performed for data exploration. After PCA, k-Nearest Neighbor (k-NN) models were constructed and applied to classify ten new samples consisting of GM and CV soybeans, using the same wavenumbers as the PCA.

Samples were analysed in two ways: differences between GM and CV soybeans (for each cultivar and all cultivars together) and between cultivars (only CV and GM, separately) were evaluated. Therefore, six data matrices Xm,n (m=lines and n=columns) were generated for chemometric analysis: i) matrix 30x613 of "cultivar A" samples (GM versus CV); ii) matrix 30x613 of "cultivar B" samples (GM versus CV); iii) matrix 30x613 of "cultivar C" samples (GM versus CV); iv) matrix of all GM and CV cultivars 90x613; v) matrix 45x613 of different CV cultivars and finally, vi) matrix 45x613of different GM cultivars.

Results and Discussion

Spectral data for multivariate analysis

The FT-IR spectra of genetically modified soybean crops (GM) and their unmodified controls (conventional, CV) presented high similarity without any pronounced differences. An example of FT-IR spectra of the GM and CV samples is reported in Fig. 1. FT-IR spectral data essentially displayed signals of soybean oil, phospholipids and proteins (Fig. 1). According to the literature, phospholipids and triglycerides have shown a high spectral overlapping in FT-IR spectra due to their high molecular similarity [26]. Therefore, as FT-IR analyses were performed from the intact material (without any sample pre-treatment), the compound derivatives of soybean oil (triglycerides and phospholipids) and proteins were detected.

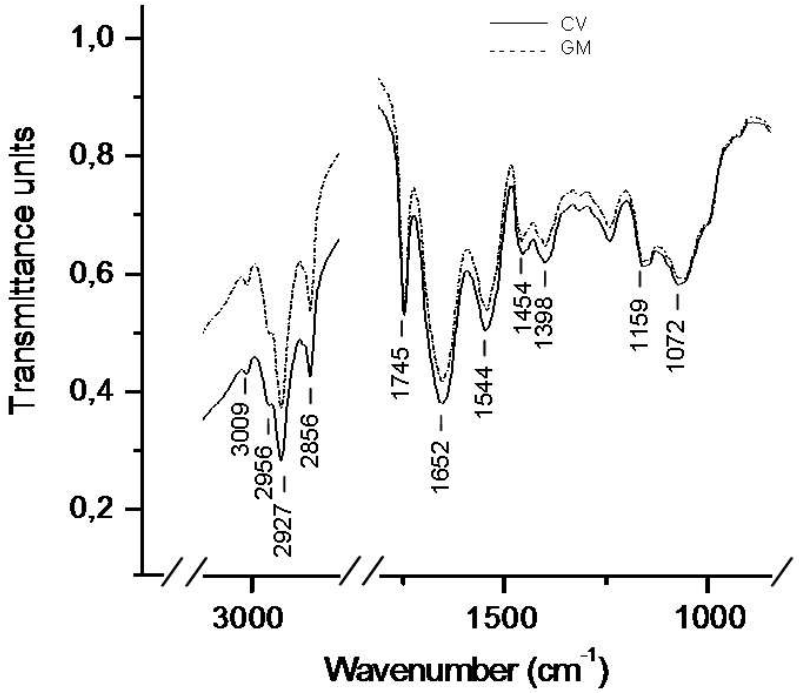

Figure 1. FT-IR spectral profile of conventional (CV) and genetically modified (GM) soybean samples with assignment of absorption bands.

FT-IR spectra showed intense bands which correspond to phospholipids: alkane bands corresponding to symmetric CH_2, asymmetric CH_2, asymmetric CH_3 stretching and CH_2 scissoring vibrational modes at 2856, 2927, 2956 and 1454 cm^{-1}, respectively; carbonyl stretching vibration, located at 1745 cm^{-1}; and the highly overlapped PO_2^- and P–O–C infrared active vibrations centered around 1544 and 1072 cm^{-1}. On the other hand, several bands were attributed as soybean oil spectrum (triglycerides) including the carbonyl C=O stretching band at 1745 cm^{-1}, the C=C stretching vibration at 1652 cm^{-1} and the CH_2 and CH_3 scissoring vibrations at 1454 and 1398 cm^{-1}, and the intense C–O stretching bands at 1159 and 1072 cm^{-1}. Absorption bands located at 2854, 2956 and 3009 cm^{-1} arose from symmetric CH_2 stretching and asymmetric CH_3 and CH_2 stretching,

respectively. The OH stretching vibration possibly overlapped the alkenes C-H stretching vibrations. While lipid components may show response at 1544 and 1652 cm^{-1}, the two bands in question also correspond to the well-known amide II and amide I bands, which are very strong in virtually all proteins.

All these above regions of FT-IR spectra were used in chemometric analysis.

Chemometric analysis of GM v CV soybean

Chemometric analyses of the FT-IR spectral data set from GM and CV soybean seeds were carried out to evaluate the genetic modification and its effect on the metabolic profile.

Thus PCA was performed on the four data sets of FT-IR: "cultivar A", "cultivar B", "cultivar C" and "all cultivars together". For all PCA, some preprocessing were tested and the best results were obtained when log_{10}, multiplication by -1 and first derivative were applied to the samples. Log_{10} followed by multiplication by -1 were used to transform transmittance in absorbance spectra, and the first derivative was applied to correct some remaining baseline imperfections [27]. The variables were then mean centered and the cross validation was applied in the chemometric analysis. These conditions were used in all PCA performed in this study.

PCA of the FT-IR spectra of the "cultivar A" data set showed the separation between GM and CV seeds (Fig. 2). This analysis presented only one GM soybean replicate was allocated out of its group, it was considered an outlier and after excluding the outlier, the data were obtained with 87.48% total variance. CV soybean samples of "cultivar A" were located on the positive side of the PC2 axis while GM samples were located on the negative side of the PC2 axis. Examination of PC1 and PC2 loadings of FT-IR data indicated that this separation occurred due to spectral domains situated at about 1544 and 1652 cm^{-1}, corresponding to two bands of amide (amide II and amide I), or P-O-C and alkene C=C vibrations, respectively, suggesting a chemical variation in protein/oil soybean contents. According to the literature [1], new proteins are expected to be found in GM organisms, consequently, in Fig. 2, it was shown by this chemical variation.

PCA of "Cultivar B" showed a good distinction between GM and CV species (Fig. 3). No outliers were found in PCA analysis and the total variance was 81.45%. CV samples of "cultivar B" were located on the more positive side of the PC1 and PC2 axis, while GM soybean samples were located on the more negative side. The loadings of these FT-IR data suggested that bands of CH_3 and CH_2 stretching vibrations (at 2856, 2927, 2956 and 1454 cm^{-1}) were decisive for allocation of GM samples. Some contribution of loadings corresponding to the P-O-C bond (1544 cm^{-1}) was observed in GM and CV

separation of "cultivar B" samples. Furthermore, these vibrations may be also applied to amide of proteins. Therefore, according to this result, it was confirmed that the aliphatic chain of phospholipids/oil and protein contents were again relevant for distinction between GM and CV soybean.

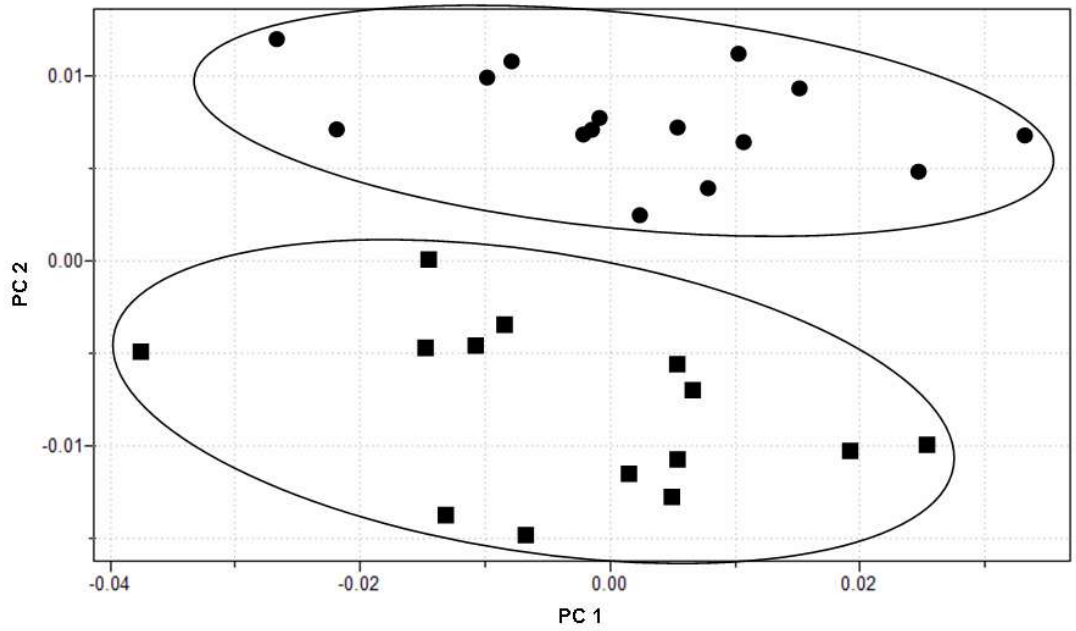

Figure 2. PCA score plot of FT-IR data from "cultivar A", showing the distinction between CV (●) and GM (■) soybean. PC1vPC2, 66.88 and 20.60%, respectively.

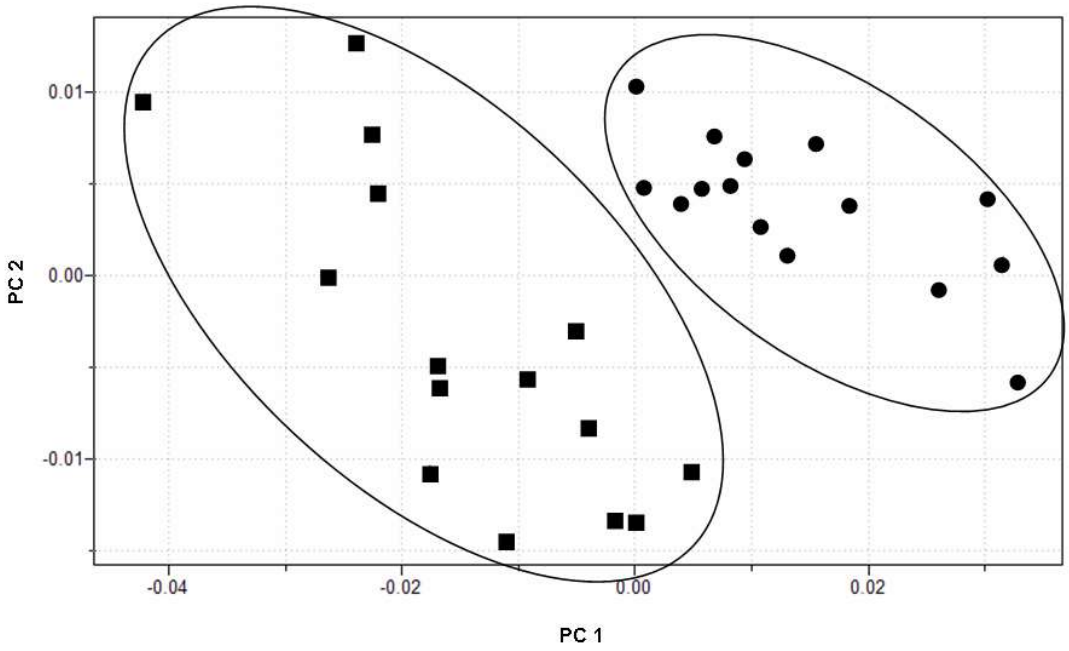

Figure 3. PCA score plot of FT-IR data from "cultivar B", showing the distinction between CV (●) and GM (■) soybean. PC1vPC2, 69.53 and 11.92%, respectively.

PCA of "Cultivar C" demonstred a satisfactory separation between GM and CV

species (Fig. 4). This analysis showed four outliers (two CV and two GM) and after excluding the outliers, the data were obtained with 86.46% variance. Fig. 4 shows that the CV samples of "cultivar C" were located on the more positive side of the PC1 and PC2 axes, while those from GM samples were located on the more negative side of the PC1 and PC2 axes. The same discrimination profile was observed for "cultivar B". The assessment of PC1 and PC2 loadings of "cultivar C" showed the influence of absorptions at 1652 cm^{-1} (amide I) for CV samples and absorptions at 1745 and 1544 cm^{-1} (C=O and amide II, respectively) for GM samples. These results corroborated the relevant influence of protein contents for transgenic soybean distinction.

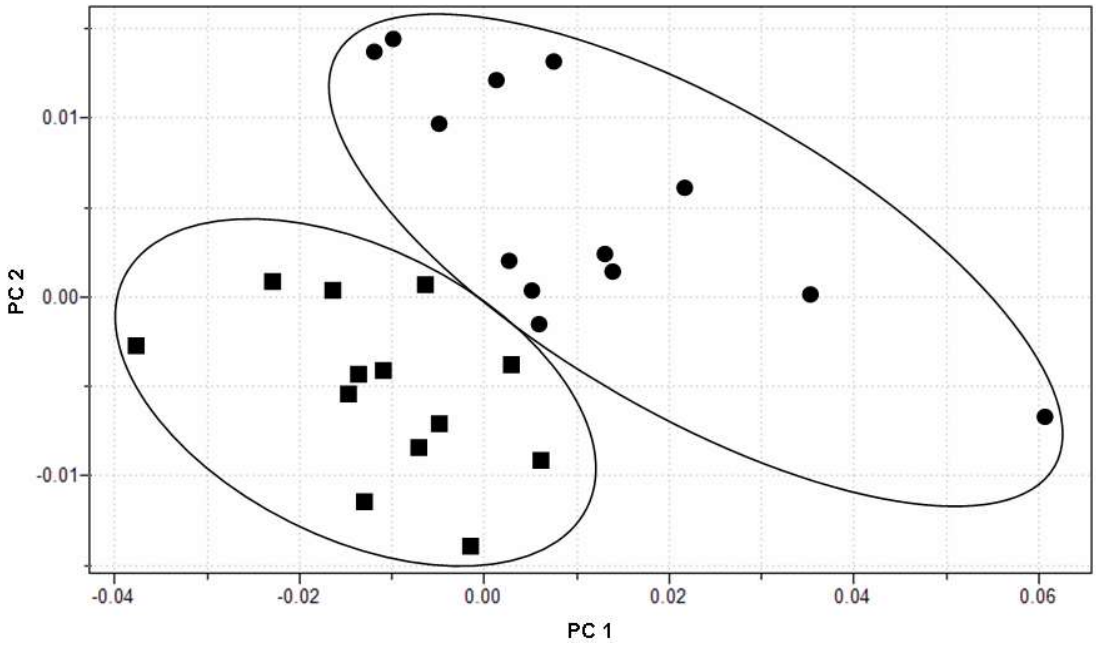

Figure 4. PCA score plot of FT-IR data from "cultivar C", showing the distinction between CV (●) and GM (■) soybean. PC1vPC2, 74.05 and 12.41%, respectively.

The three data sets described above were extremely satisfactory for discrimination between conventional (CV) and genetically modified (GM) soybean species. These interpretations could not be carried out from visual analysis of FT-IR spectra alone. Therefore, chemometric analyses were essential to this assessment.

These results are in agreement the study published by Mounts and coworkers (1996) [28], where the oils obtained from GM soybeans and non-GM soybeans showed some difference in chemical content. Nevertheless, without DNA detection, we have described the differences of protein content in GM soybean by means of FT-IR and chemometric analysis. Besides, this information can be accessed directly from intact seeds, without any sample pre-treatment.

Although soybean genetic modification was evaluated according to cultivars, a

combined analysis was also carried out. The data of all cultivars were pooled and a new PCA was run. PC1 until PC4 were investigated, better results were found to PC1xPC2 combination though. The PCA of "all cultivars together" showed some differentiation between CV and GM samples on the PC2 axis (Fig. 5). The preponderance of GM samples was observed on the more negative side of PC2, while the CV samples were observed on the more positive PC2 side. Therefore, after analysis of all cultivars together, it was inferred that the genetic modification in soybean grains was more easily detected by FT-IR and chemometric analysis when the cultivars were analyzed separately.

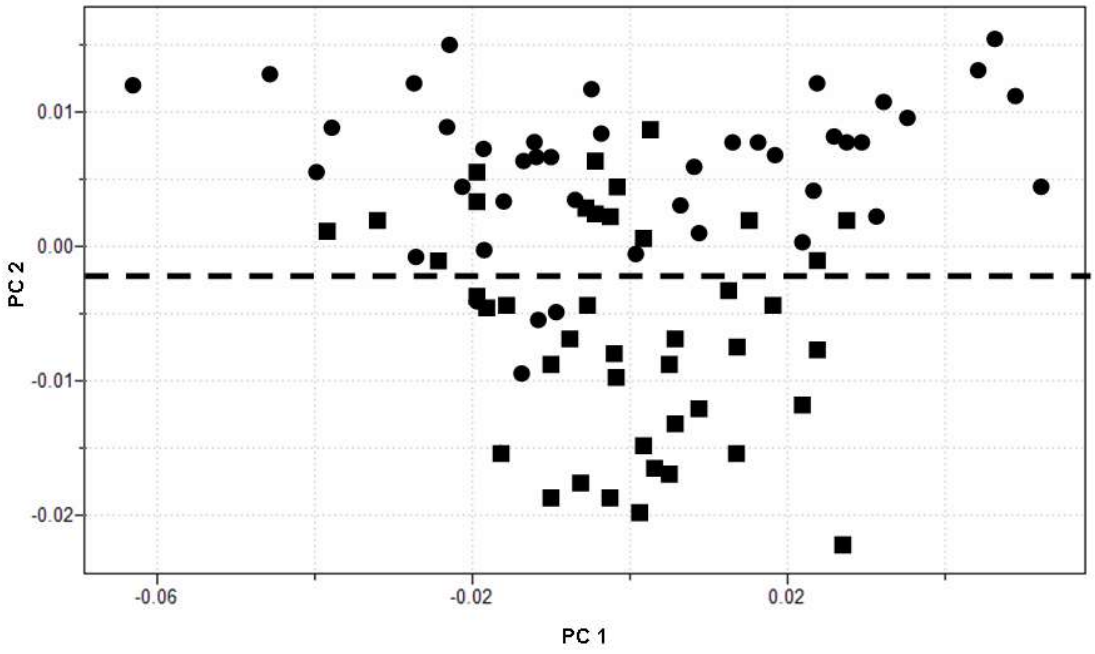

Figure 5. PCA score plot of FT-IR data from "all cultivars together", showing the distinction between CV (●) and GM (■) soybean. PC1vPC2, 71.46 and 12.58%, respectively.

Prediction of GM and CV samples was performed by the KNN method, in which an unknown pattern was classified according to the majority of the votes of its K^{th} nearest neighbors in the *n*-space [29]. The same preprocessing as in PCA was applied in the KNN analysis. Thus, for KNN classification, new FT-IR spectra of twenty replicates of each cultivar (ten GM and ten CV) were acquired and submitted to the KNN chemometric method. When six nearest neighbors (6-NN) were used, no prediction misses could be detected for "cultivar A" and "cultivar B". However, for "cultivar C" a single miss was found which corresponded to the GM sample. Therefore, KNN was shown to be an excellent method to predict/identify GM soybean crops, since 100% samples from cultivars A and B and 95% from cultivar C were correctly classified. However, when the different cultivars are evaluated together or in mixture, the KNN method can not be efficient to predict with exaction, according to results showed in Figure 5.

In this regard, all results demonstrated the potentiality of medium FT-IR and chemometrics for genetically modified soybean analysis, as well as the NIR spectroscopy [11, 12] and FT-IR-PAS [13].

Chemometric analysis of different soybean cultivars

A parallel investigation was performed evaluating only the soybean cultivars. Thus, FT-IR spectra from the three CV cultivars (A, B and C), and those from the three GM cultivars were separately subjected to chemometric analysis by PCA.

Comparatively, the three different CV cultivars showed a better separation in PCA scores plot (Fig. 6a) than GM cultivars (Fig. 6b). Absorptions corresponding to the phospholipid and aliphatic chain were again highlighted as loadings for this discrimination. Therefore, the genetic modification in soybean samples allowed the closeness between cultivars suggesting that cultivar discrimination through FT-IR and chemometrics may be more difficult.

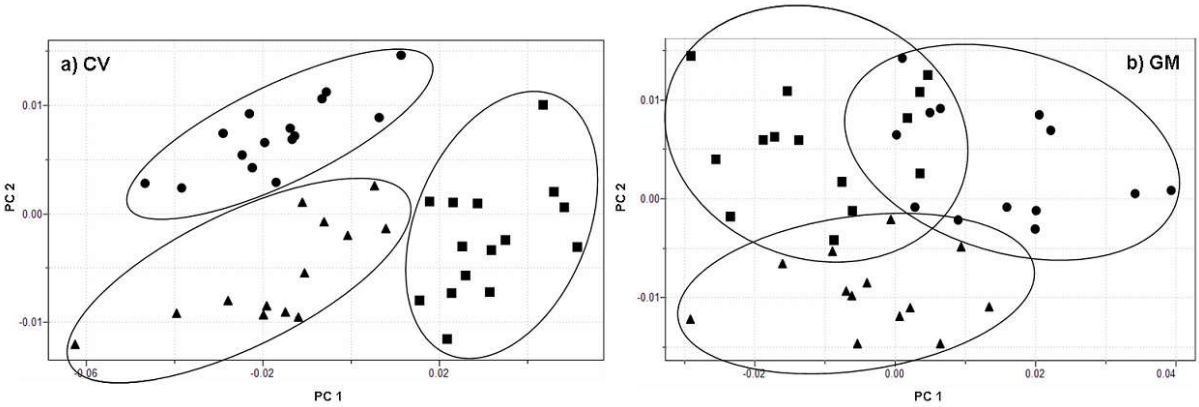

Figure 6. PCA scores plot of FT-IR data from a) CV soybean samples (PC1vPC2, 85.45 and 5.64%, respectively); and b) GM soybean samples (PC1vPC2, 60.94 and 16.76%, respectively), showing the distinction between cultivars A (●), B (■) and C (▲).

Considering de fact that transgenic and non-transgenic cultivars are different soybean varieties, the results showed above (Figure 6) were in concordance to those results about the distinction between GM and CV samples (Figures 2 to 4), when each cultivar were separately evaluated, which may be considered two different varieties of soybean.

Conclusion

In summary, FT-IR spectroscopy associated to chemometric analysis was shown to be an excellent tool to assess the chemical composition variation and to distinguish between GM and CV soybean crops. From PCA, the variations in phospholipids, triglyceride and protein content were highlighted as essential for metabolic distinction

between soybean samples. The KNN method was able to recognize GM or CV samples from their FT-IR spectra, when different cultivars were separately analyzed. In addition, the simple PCA analysis of FT-IR spectra also permitted sample discrimination according to different cultivars. Therefore, due to the easy of use and low cost of this technology, FT-IR analysis and chemometric methods are very attractive tools available to be applied in analysis to get rapid answers on the evaluation and detection of genetically modified organisms, especially GM soybean.

Acknowledgements

The authors thank the Brazilian Institutions CNPq, CAPES and FAPESP, for financial support and fellowships.

References and Notes

[1] Gachet, E.; Martin, G. G.; Vigneau, F.; Meyer, G. *Trends Food Sci. Technol*. **1999**, *9*, 380.
[2] García-Villalba, R.; Leon, C.; Dinelli, G.; Segura-Carretero, A.; Fernández-Gutiérrez, A.; Garcia-Cañas, V.; Cifuentes, A. *J. Chromatogr*. A **2008**, *1191*, 164.
[3] König, A.; Cockburn, A.; Crevel, R. W. R.; Debruyne, E.; Grafstroem, R.; Hammerling, U.; Kinber, I.; Knudsen, I.; Kuiper, H. A.; Peijnenburg, A. A. C. M.; Penninks A. H.; Poulsen, M., Schauzu, M.; Wal, J. M. *Food Chem. Toxicol*. **2004**, *42*, 1047.
[4] James, D.; Schmidt, A.; Wall, E.; Green, M.; Masri, S. *J. Agric. Food Chem*. **2003**, *51*, 5829.
[5] Phipps, R. H.; Beever, D. E. *J. Anim. Feed Sci.* **2000**, *9*, 543.
[6] Aumaitre, A.; Aulrich, K.; Chesson, A.; Flachowsky, G.; Piva, G. *Livest. Prod. Sci.* **2002**, *74*, 223.
[7] Tenbült, P.; De Vries, N. K.; van Breukelen, G.; Dreezens, E.; Martijn, C. *Appetite* **2008**, *51*, 129.
[8] Saba, A.; Vassallo, M. *Food Qual. Pref*. **2002**, *13*, 13.
[9] Bredahl, L. *J. Consum. Policy* **2001**, *24*, 23.
[10] Eapen S. *Biotechnol. Adv*. **2008**, *26*, 162.
[11] Roussel, S. A.; Hardy, C. L.; Hurburgh, C. R.; Rippke, G. R. *Appl. Spectrosc*. **2001**, *55*, 1425.
[12] Alishahi, A.; Farahmand, H., Prieto, N.; Cozzolino, D. *Spectrochim. Acta, Part A* **2010**, *75*, 1.
[13] Caires, A. R. L.; Teixeira, M. R. O.; Súarez, Y. R.; Andrade, L. H. C.; Lima, S. M. *Applied Spectroscopy* **2008**, *62*, 1044.
[14] Ahmed, F. E. *Trends Biotechnol*. **2002**, *20*, 215.
[15] Rui, Y. K.; Luo, Y. B.; Huang, K. L.; Wang, W. M.; Zhang, L. D. *Spectrosc. Spect. Anal*. **2005**, *25*, 1581.
[16] Manetti, C.; Bianchetti, C.; Bizzarri, M.; Casciani, L.; Castro, C.; D'Ascenzo, G.;

Delfini, M.; Di Cocco, M. E.; Laganà, A.; Miccheli, A.; Motto, M.; Conti, F. *Phytochemistry* **2004**, *65*, 3187.

[17] Le Gall, G.; Colquhoun, I.; Davis, A. L.; Collins, G. J.; Verhoeyen, M. E. *J. Agric. Food Chem. 2003*, *51*, 2447.

[18] Xie, L.; Ying, Y.; Ying, T. *J. Food Eng.* **2007**, *82*, 395.

[19] Xie, L.; Ying, Y.; Ying, T. *J. Agric. Food Chem*. **2007**, *55*, 4645.

[20] Xie, L.; Ying, Y.; Ying, T.; Yu, H.; Fu, X. *Anal. Chim. Acta* **2007**, *584*, 379.

[21] Alcantara, G. B.; Honda, N. K.; Ferreira, M. M. C.; Ferreira, A. G. *Anal. Chim. Acta* **2007**, *595*, 3.

[22] Reid, L. M.; O'Donnell, C. P.; Downey G. *Trends Food Sci. Technol*. **2006**, *17*, 344.

[23] Michelini, E.; Simoni, P.; Cevenini, L.; Mezzanotte, L.; Roda, A. *Anal. Bioanal. Chem.* **2008**, *392*, 355.

[24] Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. Chemometrics: A Practical Guide, 1st ed. New York: John Wiley & Sons, 1998.

[25] Bailey, N. J. C.; Oven, M.; Holmes, E.; Nicholson, J. K.; Zenk, M. H. *Phytochemistry* **2003**, *62*, 851.

[26] Kuligowski, J.; Quintás, G.; Esteve-Turrillas, F. A.; Garrigues, S.; de la Guardia, M. *J. Chromatogr. A* **2008**, *1185*, 71.

[27] Pirouette™, Multivariate Data Analysis, Pirouette User Guide, version 3.11. Infometrix Inc., Woodinville, 2003.

[28] Mounts, T. L.; Abidi S. L.; Rennick, K. A. *J. Am. Oil Chem. Soc*. **1996**, *73*, 581.

[29] Kowalski, B. R.; Bender, C. F. *Anal. Chem*. **1972**, *44*, 1405.