Study on Rapid Identification Methods of Transgenic Rapeseed Oil Based on Near Infrared Spectroscopy

Shiping Zhu, Jing Liang, Lin Yan

Study on Rapid Identification methods of Transgenic Rapeseed Oil Based on Near Infrared Spectroscopy

Shiping Zhu[1], Jing Liang, Ling Yan

College of Engineering and Technology, Southwest University, Chongqing, China
zhu_s_p@sina.com, liangjing_lj@163.com, ylaxx_forever@sina.com

Abstract: A rapid identification method of transgenic rapeseed oil based on near infrared spectroscopy was proposed. As transgenic food attracted more and more attention and in order to meet the requirement about labeling transgenic food, an accurate, fast, easy, efficient and low cost detection technology is needed. In this study, 117 rapeseed oil samples (including 64 non-transgenic rapeseed oil and 53 transgenic rapeseed oil samples) were used as test material. Principal component analysis (PCA) and discriminant partial least squares (DPLS) were applied to classify transgenic and non- transgenic rapeseed oil. Firstly, the paper studied the plot of principal component scores for the NIR spectra of transgenic and non- transgenic rapeseed oil. The result was that Non-transgenic and transgenic samples can be broadly distinguished in principal component space. Secondly, 35 samples were used to build DPLS identification model, and the other 82 samples as prediction set samples were identified by the model. The overall correct identification rate acquired was 96.34%, in which the correct identification rate of non-transgenic rapeseed oil was 95.56% and the one of transgenic rapeseed oil was 97.30%. The results showed that the attempt to discriminate transgenic rapeseed oil using NIR is feasible and excellent classification can be obtained by combining DPLS method.

Keywords: Near infrared spectroscopy, DPLS, Transgenic food, Rapeseed oil, Rapid identification

1. Introduction

Transgene means that implant the gene which extracting from another organism into an organism via transgenic biotechnology. Transgene involves animal and plant, whereas transgenic plant is discussed more at present. The food adopted transgenic plant as raw material is called Transgenic food [1] or Genetically Engineered food [2].

With the rapidly development of biotechnology, transgenic plant have been cultivated in abundance and transgenic food, such as soybean oil, rapeseed oil and maize oil, have entered the daily life of people. According to report, 90 percent of the edible oil sold on markets is transgenic oil at present and the consumption of oil per capita every year is up to 13kg in some metropolises. Consequently the security of transgenic food has increasingly captured the highly attention from governments, non-governmental organizations, academia and even ordinary people [3].

Since 1900, European Union issued over 10 laws about safety administration of GMO (short for genetically modified organism), and required labeling GMO as well as its production. Beginning in 1998 EU prescribed the mandatory label for GM food. In China, for protecting lawful rights of consumer, State Council issued *Biosafety Administration Regulations on Agricultural GMO* in 2001 and Ministry of Agriculture issued *Labeling Administration Regulations on Agricultural GMO* in 2002[4].

With the increasing demand of labeling transgenic food, the techniques for detection and identification of transgenic food are improving constantly and the methods are various. At present, the qualitative and quantitative detection for transgenic elements in most of food are available.

The detection methods of transgenic food include two broad categories, viz. the qualitative detection for exogenous gene and foreign protein. The detection methods for exogenous gene include Polymerase Chain Reaction PCR, semi nested PCR, nested PCR, Multip-lex PCR, DNA blotting etc. The quantitative detection methods for exogenous gene include quantitative PCR, PCR-ELISA method, gene chip technique. The detection methods for foreign protein include Western blotting, enzyme-linked immunosorbent assay (ELISA), Latera Flow, Lateral Flow Strip etc. [5-6]

The various methods above have different traits. But with the intensified research for transgenic produce at home and abroad as well as the increasing detection demand for transgenic produce, a strong demand for detection technique which is more accurate, rapid, easy to use, efficient and of low cost, has generated.

In the field of quality testing of agricultural products, near infrared spectroscopy analytical technology, as a new method, can make rapid and accurate analysis or detection for the element or quality of chemical products, medicament, food, agricultural products etc. Compared with traditional chemical analysis methods, it has advantages such as rapid (analytic time <1min), accurate, no chemical pretreatment for samples, nondestructive and so on.

At present, introducing NIR analytical technology into rapid detection of transgenic food is the cutting edge topic on detection method study of transgenic food. [6.10.11]

S. A. Roussel et al in Grain Quality Lab of Lovwa Sate University distinguished Roundup Ready soybeans from conventional soybeans using NIR. 95% accurate classification for 20 samples of conventional soybeans and 84% accurate classification for 19 samples of Roundup Ready soybeans were obtained [12]. Ronald J. F. J. Oomen et al studied the discrimination of transgenic potato using NIR principal component space [13].

Interiorly, Rui Yukun et al identified transgenic maize and their parents using back propagation (BP) algorithm by NIR [14]. Xie Lijuan, Ying Yibin et al applied NIR spectrometer to discriminate transgenic tomatoes and their parents [15-17]. In 2008, Xie Lijuan, Ying Yibin et al used Vis/NIR spectrum diffuse reflection method to perform rapid and nondestructive qualitative analysis for transgenic tomato leaves and non -transgenic ones. The accurate classification was 89.7% [18].

In spite of Rui Yukun et al had measured the erucic acid and Glucosinolate content of transgenic rapeseed and their parents by NIR spectrometer [19], they hadn't identified transgenic rapeseed and their parents. Besides, at home and board there isn't any report on the research that whether rapeseed oil edible is transgenic. Therefore, it has theoretical significance and application value that developing rapid discrimination method for transgenic rapeseed oil based on NIR.

The objective of this study is to apply NIR spectroscopic techniques for detection of transgenic rapeseed oil. The specific goals were to (1) investigate the feasibility of near infrared spectroscopy in detecting transgenic rapeseed oil, (2) combine the PCA for classifying NIR rapeseed oil spectra into two groups, non-transgenic ones and transgenic ones and (3) establish the DPLS model to discriminate the transgenic samples. This research can provide discrimination of transgenic rapeseed oil as well as other transgenic farm produce with new method and technical support.

2. Materials

2.1. Samples

117 rapeseed oil samples were collected from seven brands such as Jin Long Yue, Li Yu, Hong Qingting, Xiao Mifeng, Lu Hua, including 64 non-transgenic rapeseed oil samples and 53 transgenic ones. The volume of every sample was 100ml. 35 samples (19 non-transgenic rapeseed oil samples and 16 transgenic ones) were used for calibration set, and the rest 82 ones for validation set. All the samples chose for calibration and validation sets were randomly.

2.2. Spectral measurement

The NIR spectra of every rapeseed oil sample were collected by MATRIX-F portable Fourier transform infrared spectrometer of BRUKER in the spectral regions of $11997\sim3999cm^{-1}$. Spectrometer parameters setting, spectral data collecting were performed via software OPUS. The number of scanning was 32, the resolution was $4cm^{-1}$ and the spectral points were 4148. Fig 1 showed the NIR spectra of 117 rapeseed oil samples.
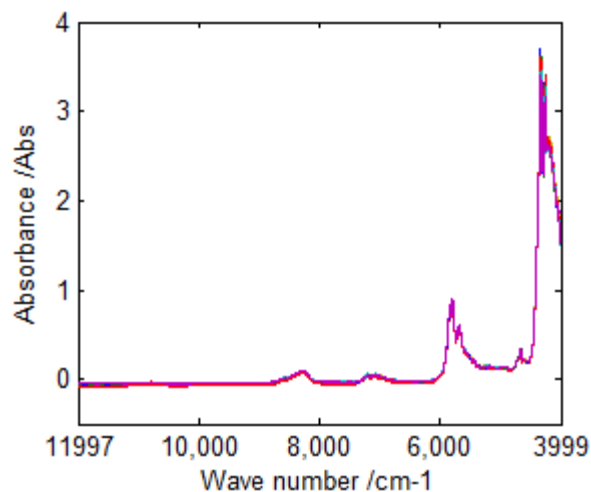
Fig . 1. NIR spectra of 117 rapeseed oil samples

3. Methods

3.1 PCA

Principal component analysis (PCA) is a multivariate modelling and analysis technique commonly used in chemometric studies. It can reduce the number of dimensions without much loss of information based on their similarities and differences to emphasise their similarities and differences. It can define the ''principal components'' (PCs) which describe independent variation structures in the data. These PCs are orthogonal, so that the data set presented on these axes are uncorrelated with each other. Visualizing the PCs by various plotting systems can view interrelationships between different variables. [20]

3.2 DPLS

The PLS algorithm has been widely used as a chemometric tool for near-infrared spectral analysis. Ii is a regression method based on characteristic variable. Its advantage is that it decomposes not only matrix X, but also matrix Y. While decomposing matrix Y it uses the information of matrix X; while decomposing matrix X it uses the information of matrix Y. Through that the regressive results become better. The DPLS method applies partial least squares (PLS) regression to binary classification problems, in which the dependent variabley codifies the class of each sample. In this study, the integer 1 codifies the sample as belonging to non-transgenic rapeseed oil and the integer 2 codifies the sample as belonging to transgenic ones. [21-22]

4. Results and discussion

4.1. PCA

PCA can indicate relationships among groups of variables in a data set and show relationships that might exist between objects. All the spectra of 64 non-transgenic rapeseed oil and 53 transgenic ones are used for PCA. The spectra were original, without any optimizing or pre-treatment. According to MATLAB software, the score plots of principal components can be obtained.

Fig 2 showed the three-dimensional score plot of NIR principal components (PC2、PC3、PC4) for 117 rapeseed oil samples. Fig 3 showed two-dimensional score plot of principal components PC2 and PC4. In this figure, non-transgenic rapeseed oil were represented by "△", and transgenic ones by "○". From this figure, it

can be found that non-transgenic rapeseed oil and transgenic ones can be discriminated roughly in their score plot of principal component, but the dividing line was not obvious.
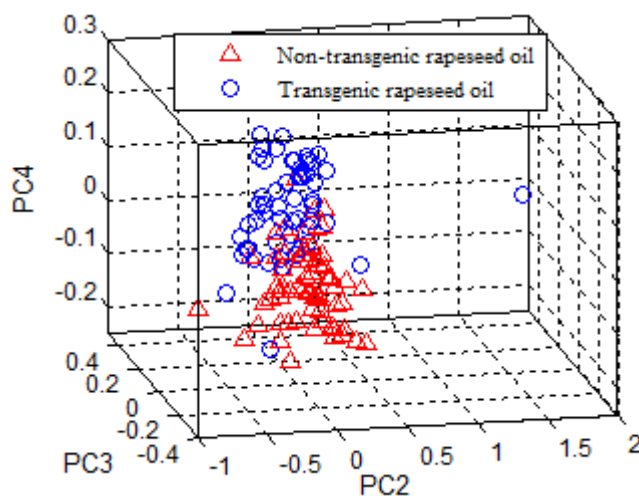

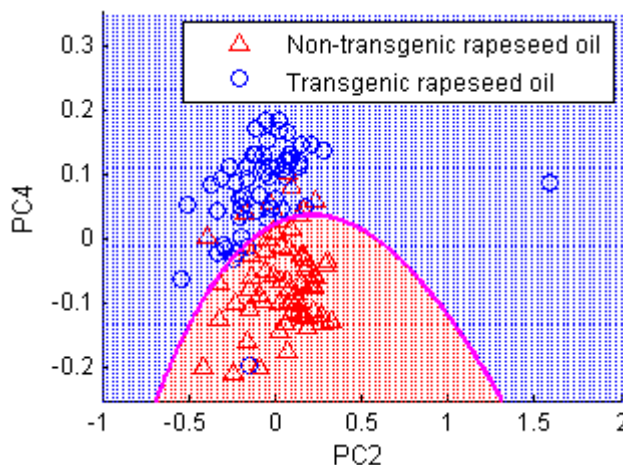**Fig.2.** Three-dimensional Score plot of three principal components (PC2、PC3、PC4)

Fig.3. Two-dimensional Score plot of two principal components (PC2 and PC4)

4.2. DPLS

The software OPUS was used to establish DPLS (Discriminant partial least squares) model on original NIR spectra of calibration set. Then the model was used to predict validation set. The prediction result was that $R^2=0.7487$ and RMSEP value of 0.249.

Fig 4 showed the predictions of rapeseed oil varieties in the calibration set using the DPLS model with the original NIR spectra. Non-transgenic rapeseed oil were represented by "△", and transgenic ones by "○". Non-transgenic rapeseed oil with predicted values ranged form 0.5 to 1.5 and transgenic samples from 1.5 to 2.5 were all considered to be broadly correctly classified by the model. Fig 5 showed the true value of 82 prediction categories vectors .The prediction categories vectors within the range from 0.5 to 1.5 were identified as "1" and the ones from 1.5 to 2.5 were identified as "2". The discriminate functions acquired were showed in fig 6.
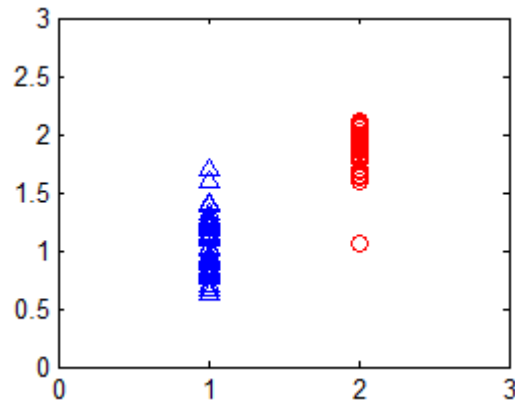
Fig.4. Plot of actual vs calculated values of transgenic and non-transgenic rapeseed oil
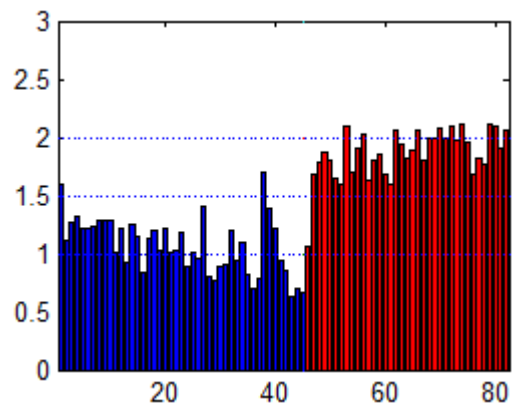

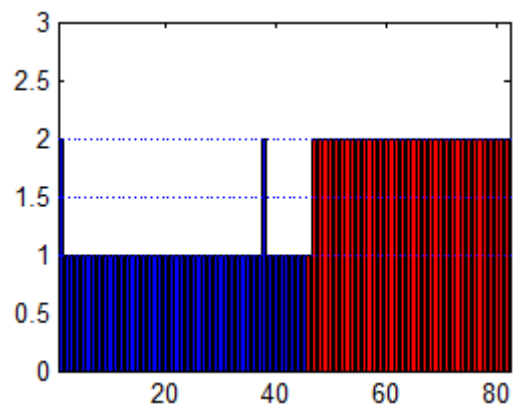**Fig.5.** True value of 82 prediction categories vectors

Fig.6. Discriminate functions of 82 prediction categories vectors

From fig 6, it can be found that there were 3 inaccurate decision samples. Two non-transgenic samples that labeled as "1" were discriminated as "2" and one transgenic sample that labeled as "2" was discriminated as "1". So the general discrimination accurate rate was (82-3)/82=96.34%; the discrimination accurate rate of non-transgenic samples was (45-2)/45=95.56%; the one of transgenic samples was (37-1)/37= 97.30%.

5. Conclusion

The study showed that the attempt to discriminate transgenic rapeseed oil using NIR is feasible. The DPLS model obtained the prediction result was that R^2=0.7487, RMSEP value of 0.249 and the discrimination accurate rate of 96.34% which is broadly satisfactory result using DPLS model on the original spectra. The results showed that excellent classification can be obtained by combining NIR and DPLS method. Introducing NIR into

rapidly discrimination of transgenic food can provide new method and technical support for the detection of transgenic food. However it is still in the infant and exploration stage. There are some limitations consisting in the methods that used to establish models. Thus the research about optimizing model should be concerned in the filed of rapid detection for transgenic food using NIR.

Acknowledgments

References

1. Soon-Chun Jeong, In Soon Pack, Eun-Young Cho, et al.: Molecular analysis and quantitative detection of a transgenic rice line expressing a bifunctional fusion TPSP.J. Food Control.18 (11):1434-1442(2007)

2. Rita Batista, Maria Margarida Oliveira.: Facts and fiction of genetically engineered food. J. Trends in Biotechnology. 27(5):277-286(2009)

3. Chen Yourong, Wang Hua.: The safety of Gentic engineering and genetically food .J. Food Science. 23(12):145-149(2002) (In Chinese)

4. Liu Jinhua, Wei Chunyan, Luo Yanfei, et al.: Detection of transgenic ingredient in food .J. Journal of Northeast Agricultural University. 35(05): 634-638(2004) (In Chinese)

5. Farid E. Ahmed.: Detection of genetically modified organisms in food .J. Trends in Biotechnology. 20(5): 215-223(2002)

6. Zhou Shaoyun.: Progress in Study of Genetically modified food detection .J. Fujian Science and Technology of Rice and Wheat. 25 (1): 43-45(2007) (In Chinese)

7. Lu Wanzhen, Yuan Hongfu, Xu Guangtong, et al.: Modern Near Infrared Spectroscopy Analytical Technology. Beijing (2000)

8. Musleh Uddin, Emiko Okazaki, Hideto Fukushima, et al.: Nondestructive determination of water and protein in surimi by near-infrared spectroscopy .J. Food Chemistry. 96(3): 491-495(2006)

9. Zhu Shiping, Wang Yiming, Zhang Xiaochao.: Design and implementation of near infrared spectroscopy analysis software system for agriculture product quality detection .J. Transaction of the Chinese Society of Agriculture Engineering. 19(04): 175-179(2003) (In Chinese)

10. Zhou Pingping, Wu Ronning, Zhang Jianzhong.: The advances in the detection method of GMO foods .J. Chinese Journal of Food Hygiene. 16(3): 254-258(2004) (In Chinese)

11. Liu Zhenquan, Xia Dong, Lu Lixia, et al.: Expectationof methods of detection geneetically modified foods .J. Food Research and Development. 28(5): 161-164(2007) (In Chinese)

12. S. A. Roussel, C. L. Hardy, C. R. Hurburgh, et al.: Detection of Roundup Ready™ Soybeans by Near-Infrared Spectroscopy .J. Applied Spectroscopy.55(10): 1425-1430(2001)

13. Ronald J. F. J. Oomen, Emmanouil N. Tzitzikas, Edwin J. Bakx, et al.: Modulation of the cellulose content of tuber cell walls by antisense expression of different potato (Solanum tuberosum L.) CesA clones .J. Phytochemistry. 65(5):535-546(2004)

14. Rui Rukun, Luo Yunbo, Huang Kunlun, et al.: Application of near-infraraed diffuse reflectance spectroscopy to the detection and identification of transgenic corn .J. Spectroscopy and Spectral Analysis. 25(10): 1581-1583(2005) (In Chinese)

15. Lijuan Xie, Yibin Ying, Tiejin Ying.: Combination and comparison of chemometrics methods for identification of transgenic tomatoes using visible and near-infrared diffuse transmittance technique J. Journal of Food Engineering. 82(3): 395-401(2007)

16. Lijuan Xie, Yibin Ying, Tiejin Ying, et al.: Discrimination of transgenic tomatoes based on visible/near-infrared spectra .J. Analytica Chimica Acta.584 (2): 379-384(2007)

17. Lijuan Xie, Yibin Ying, Tie jinYing.: Classification of tomatoes with different genotypes by visible and short-wave near-infrared spectroscopy with least-squares support vector machines and other chemometrics .J. Journal of Food

Engineering. 94, (1):34-39(2009)

18. Xie Lijuan, Ying Yibin, Ying Tiejin, et al.: Application of Vis / NIR Diffuse Reflectance Spectroscopy to the Detection and Identification of Transgenic Tomato Leaf .J. Spectroscopy and spectral analysis. 28(5): 1062-1066(2008) (In Chinese)

19. Rui Rukun, Huang Kunlun, Wang Weiming, et al.: Detection of Erucic Acid and Glucosinolate in Intact Rapeseed by Near-Infrared Diffuse Renectance Spectroscopy .J. Spectroscopy and Spectral Analysis. 26(12):2190-2192(2006) (In Chinese)

20. Eui-Cheol Shin a, Brian D. Craft a, Ronald B. Pegg et al.: Chemometric approach to fatty acid profiles in Runner-type peanut cultivars by principal component analysis (PCA) .J. Food Chemistry.119: 1262–1270(2010)

21. Ruy L. Milidiu, Raul P. Renteria.: DPLS and PPLS: two PLS algorithms for large data sets .J. Computational Statistics & Data Analysis. 48:125-138(2005)

22. Cristina Botella, Joan Ferré, Ricard Boqué.: Classification from microarray data using probabilistic discriminant partial least squares with reject option .J. Talanta. 80:321–328(2009).