TERAHERTZ SPECTROSCOPY DETECTION OF GENETICALLY MODIFIED SUGAR BEET CONTAIN XA21 GENE BASED ON CHEMOMETRICS METHODS

© 2016   **Jianjun Liu,** Doctor of Measurement technology and instruments

School of Electrical Engineering, Jiujiang University, Jiujiang Jiangxi 332005, China

E-mail: liujianjun8888@hotmail.com

Due to traditional genetically modified detection has the disadvantage of high cost, long time and awkward to handle, a novel detection method of genetically modified based on terahertz spectroscopy combines with chemometrics method is proposed on this paper. In this work terahertz spectrum data of genetically modified sugar beet and its parents are pretreated by using principal component analysis, then using the weighted discrimination analysis model which an improved discrimination analysis by applying weighted algorithm to detection genetically modified sugar beet and its parents. It is found from the experimental results that the samples are expressed by the zonation of saline minerals. According to this phenomenon, it is easy to identify these genetically modified sugar beets. Combined terahertz spectroscopy with chemometrics method, this paper provides a precise, fast, convenient and nondestructive detection method for genetically modified organism.

ОПРЕДЕЛЕНИЕ СОДЕРЖАНИЯ ГЕНА XA21 В ГЕНЕТИЧЕСКИ МОДИФИЦИРОВАННОЙ САХАРНОЙ СВЁКЛЕ МЕТОДОМ ТЕРАГЕРЦОВОЙ СПЕКТРОСКОПИИ В СОЧЕТАНИИ С ХЕМОМЕТРИЧЕСКИМИ МЕТОДАМИ

© 2016 г.   **Jianjun Liu,** Doctor of Measurement technology and instruments

Недостатками традиционных методов определения генетически модифицированных продуктов являются высокая стоимость, трудновыполнимость и длительное время операций. Предложен новый метод определения генетически модифицированных продуктов на основе терагерцовой спектроскопии в сочетании с хемометрическими методами. Терагерцовые спектральные данные о генетически модифицированной сахарной свёкле и её родительских особях предварялись анализом главных компонентов, затем использовался взвешенный дискриминационный анализ. Экспериментально обнаружено, что исследованные образцы обладают выраженным зонированием солевых минералов, что облегчает обнаружение генетически модифицированной сахарной свёклы. Комбинируя методы хемометрии и терагерцовой спектроскопии, возможно создание точной, быстрой и удобной в применении методики обнаружения генетически модифицированных организмов.

1. Introduction

Sugar beet is one of the important economic crops in the world, however, suffered insects damage, the production of sugar beet suffered huge economic losses. Although it can reduce the insect pests' damage in a certain degree by using chemical pesticides, it certainly adds to the cost of sugar beet production, even more important, a large amount of chemical pesticide residues will seriously destroy the ecological environment. With the development of genetically

modified technology planting insect-resistant transgenic plants may be a feasible way. The first genetically modified plant was developed in 1993 by inserting the bacillus thuringiensis (Bt) gene which an insect resistance gene [1]. Despite the fact that the insect-resistant transgenic plants can improve crop production, the security issues caused by genetically modified crops have drawn the attention of the countries all around the world. So many countries have regulations on genetically modified crops. Especially in the European Union there is strict regulatory regime for genetically modified [2–5]. Therefore, it is very important to research affordable and effective detection methods for genetically modified organisms.

At present the main detection methods of genetically modified are including protein detection, gene chip detection, polymerase chain reaction and so on. Although those methods can effectively detect genetically modified, due to the disadvantage of high-cost, long-time and awkward to handle, this does not facilitate the detection of genetically modified organisms [6–10]. As a new, fast and nondestructive detection tool, spectrum detection technology is widely discussed because of simple operation and easy preparation. In recent years although many researchers have successfully using Raman, near infrared and visible light to detect genetically modified [11–15]. Terahertz is an electromagnetic wave which local in between the microwave and far-infrared, it has a dual characteristic of electronics and optics. Research shows that rotation and vibration level of many biological molecular is in terahertz band. Due to its unique fingerprint spectrum characteristics, THz spectroscopy has become one of the most dynamic fields of scientific research [16–18]. In 2014 Jianjun Liu *et al.* successfully used the terahertz spectroscopy and pattern recognition method to identify different transgenic cotton seeds [19]. In 2015 Wendao Xu *et al.* used terahertz spectroscopy and chemometrics to discriminate the transgenic rice containing the Cr1Ab protein and its parent [20].

In this paper the genetically modified sugar beet that implants the insect-resistant genes Xa21 protein and its parent (zhongtian 31) are selected as samples. The principal component analysis and weighted discriminate analysis are applied for terahertz spectroscopy detection of genetically modified organisms.

2. Methods

2.1. Weighted discriminate analysis

The main disadvantage of traditional discriminate analysis is that fisher criteria are not directly related to the classification rate under the condition of more classification. Fisher criteria over emphasis on those categories that have larger class space with other categories, this will cause bigger overlap to those categories with smaller class space.

In order to reduce the leading role of the edge class for projection direction, the small weight is given for edge class and the bigger weight is given for the other in discrete degree matrix, and to facilitate solving simultaneously, the criterion function of new and traditional Fisher has the same form, and we can solve this optimization problem using simple Eigen decomposition. Therefore different ref. [21] the discrete degree matrix S_b'' is redefined by using weight function in subspace A in this paper.

$$S_b'' = \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} P(c_i)P(c_j)\omega(d_{ij})(M_i' - M_j')(M_i' - M_j')^{\mathrm{T}},$$

where M_i is the Means of *i-th* group samples in subspace A; d_{ij} expresses the Mahanalobis distance between *i-th* and *j*-th group samples; $\omega(d_{ij})$ describe a decreasing function of d_{ij} that the drop speed should be greater than d_{ij}^{-1}, $\omega(d_{ij})$ is defined as follow:

$$\omega(d_{ij}) = \left. \eta_{bij} d_{ij}^{-4} \middle/ \sum \eta_{bij} d_{ij}^{-4} \right. .$$

The discrete degree matrix S_ω'' of within class is redefined as follow:

$$S_\omega'' = \sum_{i=1}^{K} P(c_i)\eta_{\omega i} E\left\{ (x_i' - M_i')(x_i' - M_i')^{\mathrm{T}} \middle| x_i' \in c_i \right\},$$

where η_{bij} and $\eta_{\omega i}$ express the interest degree correction coefficient of *i-th* group samples and $\eta_{bij} = \eta_{bji}$ $(i, j = 1, 2..., K, i \neq j)$. In normal circumstances $\eta_{bij} = \eta_{\omega i} = 1$, but in special cases, if we want to improve the identification of the *i-th* group samples, we can increase the correction factor to make the projection direction toward this class.

If the discrete degree matrix is singular in subspace class A, then we can use total discrete degree matrix to replace the discrete degree ma-
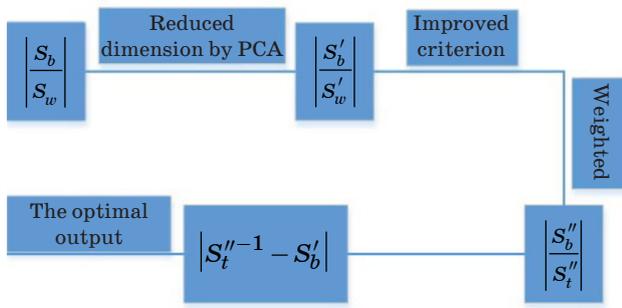
Fig. 1. The algorithm diagram of WDA based on PCA.

Fig. 2. Experimental setup of the THz-TDS system.

trix. The Improved Fisher criterion function is defined as follow:

$$J(\omega') = \left| \frac{\omega'^{\mathrm{T}} S_b'' \omega'}{\omega'^{\mathrm{T}} S_t'' \omega'} \right|,$$

where $S_t'' = S_\omega'' + S_b''$ and $S_\omega'' \geq 0, S_b'' \geq 0$.

2.2. Algorithm process of WDA combined with PCA

The traditional discriminate analysis is not directly related to the classification rate under the condition of more classification [22]. In order to increase the classification rate the improved weighted discriminate analysis (WDA) algorithm which introducing the weighted function in the traditional Fisher criteria using principal component analysis (PCA) to reduce the dimension of data. The algorithm diagram as shown in fig. 1.

3. Samples and equipment

Two different sets of sugar beet which genetically modified sugar beet (contain insect-resistant genes Xa21) and their parents are all supplied by the Sigma company in the USA. In this paper each type sample is made into 50 tablets, using the terahertz time-domain spectroscopy (TDM) system that the center wavelength of laser is 780 nm to obtain the information of each tablet. The THz-TDS system is used in this paper as shown in fig. 2.

In experiment in order to guarantee the experimental accuracy and repeatability, we inject the dry air into terahertz system to ensure the internal relative humidity is less than 2%. The relative humidity and indoor temperature is 25% and 292 K respectively.
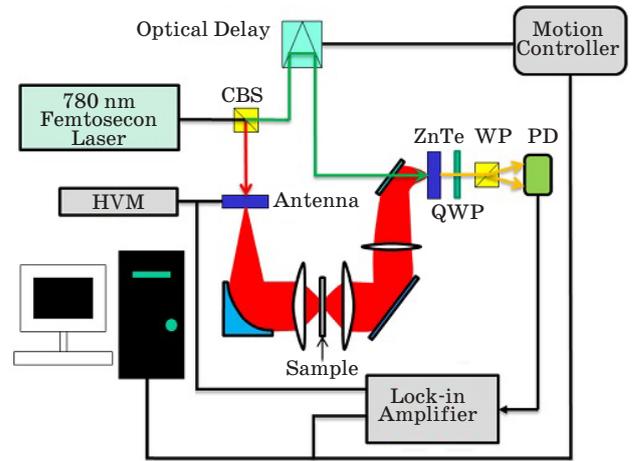
4. Results and discussion

4.1. The terahertz spectra of samples

The transmitted THz amplitude spectra of different set of sugar beet which obtained by FFT of frequency and time-domain signals are shown in fig. 3. Fig. 3(a) is the time-domain terahertz spectra, fig. 3(b) is the frequency-domain terahertz spectra. We can find that the average amplitude of genetically modified sugar beet (contain Xa21 gene) and its parent overlap and are hard to identify each other. Compared with the average amplitude of reference, the average amplitude of samples are much smaller. The result is caused by the refractive index and absorbance of samples, and it also shows that the frequency amplitude of sample is far less than the reference. The samples have a strong absorption in high frequency of terahertz.

Two average transmittance spectra of samples are shown in fig. 3(b). It can be seen from the fig. 3(b) that the light transmittance is less than 40%, overlap in 1.35 THz and the average transmittance of genetically modified sample is higher than its parents from 0.2 to 1.5 THz. In order to facilitate understanding the samples absorbance are obtained by using the follow formula:

$$absorbance(\omega) = -2\lg \left| \frac{E_s(\omega)}{E_r(\omega)} \right|,$$

where $E_S(\omega)$ and $E_r(\omega)$ is the frequency-domain spectroscopy of the reference and sample respectively. THz absorbance spectra of samples are shown in fig. 4. It can be seen from that the
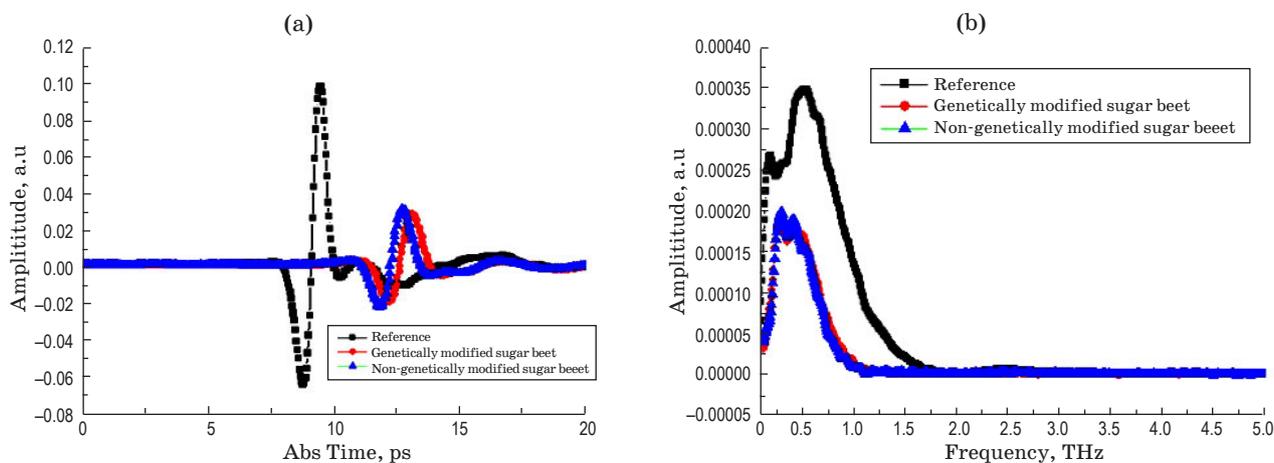
Fig. 3. Terahertz spectra of samples. **(a)** – the time-domain terahertz spectra of samples, **(b)** – the frequency-domain terahertz spectra of samples.
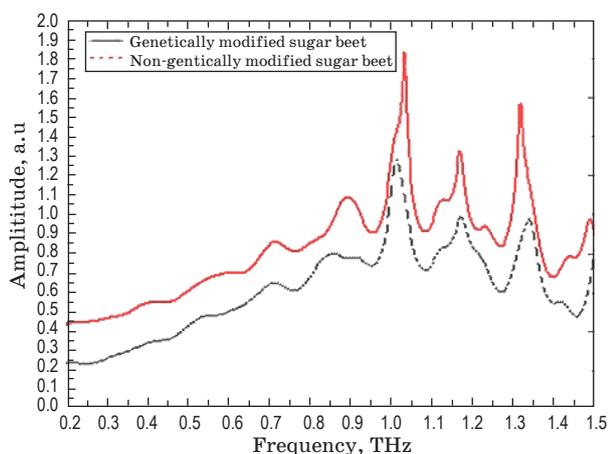

Fig. 4. THz absorbance spectra of samples.

Fig. 5. The scattered scores plot PC1 vs. PC2.

spectra of the genetically modified and non- genetically modified sugar beet are very similar. That is, it is very hard to identify genetically modified samples from non-genetically modified samples based on their absorbance curves only. Chemometrics methods can distinguish the tiny differences between samples. Therefore chemometrics methods are used to construct a qualitative detection model for genetically modified sugar beet and its parent in this paper.

4.2 Extracted the characteristic data of samples

The terahertz spectrum data of samples are collected in 0.2–1.5 THz band in this paper and 984 data points from the 984×984 spectrum matrix. Because of the large amount of raw spectral data, redundant information and large amount of calculation, it is necessary to extract the effec-
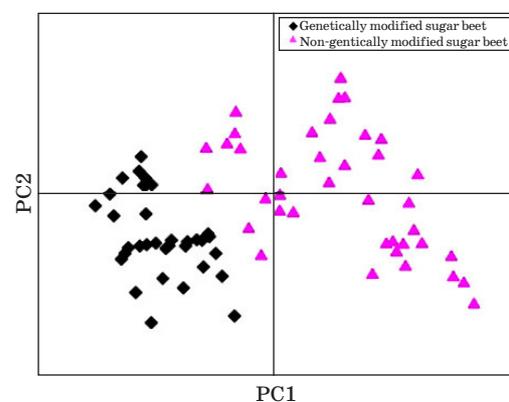
tive data and information for the data analysis. In order to save time and data calculation PCA has been performed to extract the characteristic data in this study. After performing PCA the first three principal components were proposed. The scattered scores which plot PC1 vs. PC2 is shown as fig. 5. It is found from the fig. 5 that the feature data of genetically modified sugar beet and its parent turn on zonal distribution, which means that the use of principal component for WDA analysis is feasible.

4.3. WDA identification analysis

Because of the first three principal components contains more than 99.5% of the original data information, so weighted discrimination analysis is performed on the first three principal components. The WDA model using the PCA data with the second derivative can obtain the better results than that using the original data.
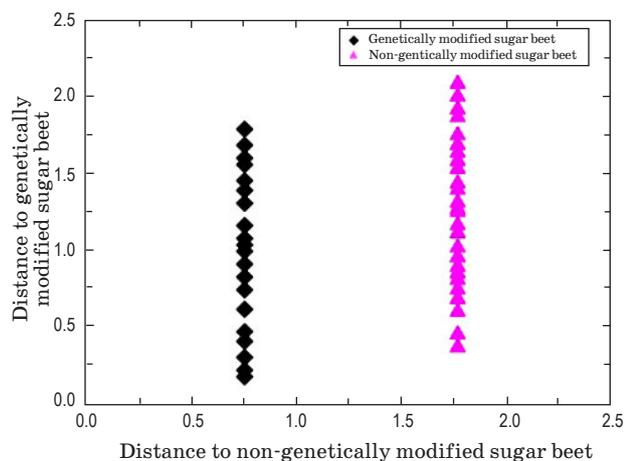
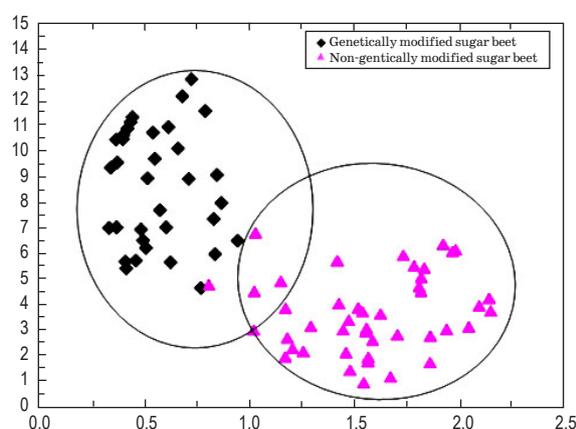**Fig. 6.** The Euclidean distance of samples.

Fig. 7. The classification result of samples by using WDA model.

For PCA data 5 samples are misidentified but for original data 8 samples are misidentified. The results from the principal component spectroscopy data have better accuracies than the original spectroscopy data.

The Euclidean distance of samples' THz spectroscopy is shown in fig. 6. It can be seen from the plot that the genetically modified sugar beet and its parents (non- genetically modified) are divided into two groups with few overlapping.

By applying PCA pretreatments to original data and combine THz spectroscopy with WDA model can achieve good recognition effect for genetically modified and non-genetically modified sugar beet. Fig. 7 shows the classification result of samples by using WDA model. The experimental results show that WDA model has a good ability to distinguish genetically modified and non-genetically modified sugar beet.

5. Discussions

The main work of this study is identifying genetically modified and non-genetically modified sugar beet based on terahertz spectroscopy conjunction with the WDA model. Combine terahertz spectroscopy technology with chemometrics methods is a relatively powerful tool for detecting genetically modified and non-genetically modified sugar beet. This method can avoid time-consuming and reduce costly chemical. It is valuable that using THz spectroscopy and chemometrics methods to determining genetically modified samples and the study shows the terahertz spectrum conjunction with chemometrics method have potential ability of detection genetically modified products. Further research is needed to create more valuable and robust model to identify other genetically modified organisms.

* * * * *

REFERENCES

1. *Fujimot H., Itoh K., Yamamoto M., Kyozuka J., Shimamoto K.* Insect resistant rice generated by introduction of a modified δ-endotoxin gene of Bacillus thuringiensis // Nat. Biotechnol. 1994. V. 4. № 4. P. 485–485.

2. *Aviron S., Sanvido O., Romeis J., Herzog F., Bigler F.* Case-specific monitoring of butterflies to determine potential effects of transgenic Bt-maize in Switzerland // Agr. Ecosyst. Environ. 2009. V. 131. № 3. P. 137–144.

3. *Zhang X.* Rapid isolation of single-chain antibodies from a human synthetic phage display library for detection of Bacillus thuringiensis (Bt) Cry1B toxin // Ecotoxicol. Environ. Safety. 2012. V. 81. № 1. P. 84–90.

4. *Giovannoli C., Anfossi L., Baggiani C., Giraudi G.* Binding properties of a monoclonal antibody against the Cry1Ab from Bacillus Thuringensis for the development of a capillary electrophoresis competitive immunoassay // Anal. Bioanal. Chem. 2008. V. 392. № 3. P. 385–393.

5. *Vergragt P.J., Brown H.S.* Genetic engineering in agriculture: new approaches for risk management through sustainability reporting // Technol. Forecast. Soc. Change. 2008. V. 75. P. 783–798.

6. *Borjigin M., Eskridge C., Niamat R.* Electrospun fiber membranes enable proliferation of genetically modified cells // International Journal of Nanomedicine. 2013. V. 8. P. 855–864.

7. *Milcamps A., Rabe S., Cade R.* Validity assessment of the detection method of maize event Bt10 through investigation of its molecular structure // Journal of Agricultural and Food Chemistry. 2009. V. 57. №. 8. P. 3156–3163.

8. *Fiehn O., Kopka J., Trethewey N.* Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry // Analytical Chemistry. 2000. V. 72. № 15. P. 3573–3580.

9. *Margarit E., Reggiardo M.I., Vallejos R.H.* Detection of BT Genetically Modified maize in foodstuffs // Food Research International. 2006. V. 39. P. 250–255.

10. *Zhu D., Liu J.F., Tang Y.B.* A reusable DNA biosensor for the detection of genetically modified organism using magnetic bead-based electrochemiluminescence // Sensors and Actuators B. 2010. V. 149. № 1. P. 221–225.

11. *Baranski R., Baranska M.* Discrimination between nongenetically modified (non-GM) and GM plant tissue expressing cysteine-rich polypeptide using FT-Raman spectroscopy // Agr. Food Chem. 2008. V. 56. № 12. P. 4491–4496.

12. *Roussel S.A., Hardy C.L., Hurburgh C.R., Rippke G.R.* Detection of Roundup Ready-soybeans by near-infrared spectroscopy // Appl. Spectrosc. 2001. V. 55. P. 1425–1430.

13. *Xie L., Ying Y., Ying T.* Combination and comparison of chemometrics methods for identification of transgenic tomatoes using visible and near-infrared diffuse transmittance technique // Food Eng. 2007. V. 82. P. 395–401.

14. *Xie L., Ying Y., Ying T., Yu H., Fu X.* Discrimination of transgenic tomatoes based on visible/near-infrared spectra // J. Anal. Chim. Acta. 2007. V. 584. № 2. P. 379–384.

15. *Xu W., Liu X., Xie L., Ying Y.* Comparison of Fourier Transform near-infrared, visible near-infrared, mid-infrared, and Raman spectroscopy as non-invasive tools for transgenic rice discrimination // Trans. ASABE. 2014. V. 57. № 1. P. 141–150.

16. *Jianjun Liu, Zhi Li, Fangrong Hu.* Hyper sausage neuron: recongnition of transgenic sugar-beet based on terahertz spectroscopy // Optics and Spectroscopy. 2015. V. 118. № 1. P. 182–187.

17. *Jianjun Liu, Zhi Li, Fangrong Hu.* Method for identifying transgenic cottons based on terahertz spectra and WLDA // Optik. 2015. V. 126. № 19. P. 1872–1877.

18. *Jianjun Liu, Zhi Li, Fangrong Hu.* A THz spectroscopy nondestructive identification method for transgenic cotton seed based on GA-SVM // Optical and Quantum Electronics. 2015. V. 47. № 2. P. 313–322.

19. *Jianjun Liu, Zhi Li.* Identification of GMOs by terahertz spectroscopy and ALAP-SVM // Optical and Quantum Electronics. 2014. V. 47. № 3. P. 685–695.

20. *Wendao Xu, Lijuan Xie, Zunzhong Ye.* Discrimination of transgenic rice containing the cry1ab protein using terahertz spectroscopy and chemometrics // Scientific Reports. 2015. V. 5. № 11115. P. 1–9.

21. *Zhang X.R., Liu F.* A patten classification method based on GA and SVM // 6th International Conference on Signal Processing. 2002. P. 110–113.

22. *Vaitilingom M., Pijnenburg H., Gendre F., Brignon P.* Real-time quantitative PCR detection of genetically modified Maximizer maize and Roundup Ready soybean in some representative foods // J. Agr. Food Chem. 1999. V. 47. P. 5261–5266.