

# SCIENTIFIC REPORTS



OPEN

## Discrimination of transgenic soybean seeds by terahertz spectroscopy

Wei Liu<sup>1,2,\*</sup>, Changhong Liu<sup>1,\*</sup>, Feng Chen<sup>3</sup>, Jianbo Yang<sup>4</sup> & Lei Zheng<sup>1,5</sup>

Received: 08 February 2016  
Accepted: 30 September 2016  
Published: 26 October 2016

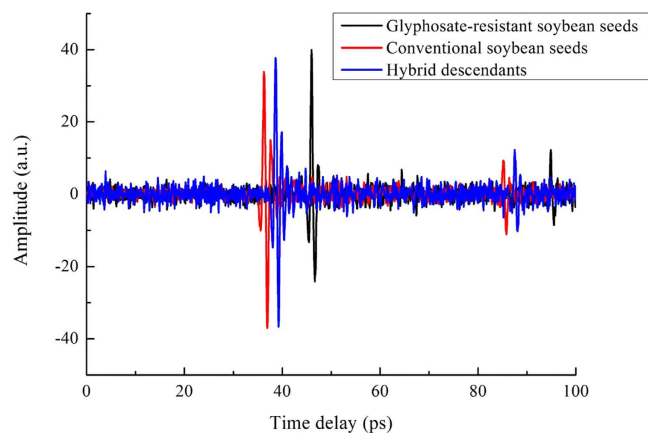
Discrimination of genetically modified organisms is increasingly demanded by legislation and consumers worldwide. The feasibility of a non-destructive discrimination of glyphosate-resistant and conventional soybean seeds and their hybrid descendants was examined by terahertz time-domain spectroscopy system combined with chemometrics. Principal component analysis (PCA), least squares-support vector machines (LS-SVM) and PCA-back propagation neural network (PCA-BPNN) models with the first and second derivative and standard normal variate (SNV) transformation pre-treatments were applied to classify soybean seeds based on genotype. Results demonstrated clear differences among glyphosate-resistant, hybrid descendants and conventional non-transformed soybean seeds could easily be visualized with an excellent classification (accuracy was 88.33% in validation set) using the LS-SVM and the spectra with SNV pre-treatment. The results indicated that THz spectroscopy techniques together with chemometrics would be a promising technique to distinguish transgenic soybean seeds from non-transformed seeds with high efficiency and without any major sample preparation.

Soybean (*Glycine max* L.) is one of the most important crops in the world that provides high amounts of oils, proteins, and biologically active components. Genetically modified (GM) soybean has made rapid strides in the past decades and the GM cultivation area is increasing year by year. The ROUNDUP READY (RR) soybean which incorporates a gene from the bacteria *Agrobacterium tumefaciens* conferring resistance to the glyphosate-based ROUNDUP herbicide, now rapidly displaced conventional soybeans because of advantages for crop management and yields. Of the global 90 million hectares of soybean grown in 2010, an impressive 73.3 million hectares were RR soybean<sup>1</sup>. Despite the widely recognized benefits of GM organisms, governmental organizations and/or the general public are still hesitant or opposed to the cultivation and application of GM organisms for a variety of reasons, such as their impact on human and animal health, effects on the environment, and socioeconomic effects. Particularly in the case of crop-to-crop gene flow, as more transgenic cultivars become available, the contamination of conventional cultivars with transgenes and the unintended combination of transgenes through natural hybridization will become increasingly probable.

Gene flow is a natural phenomenon that normally occurs via pollen but can also be caused by natural or human movement during production or commercialization of seeds and vegetative propagules. Soybean is an annual self-pollinating species where pollination occurs either in the bud stage or before flowers completely open. Soybean pollen is too heavy for wind transport but pollination by honey bees, thrips species and predatory Hemiptera has been shown to increase the possibility of out-crossing<sup>2,3</sup>. Consequently, out-crossing with pollen from adjacent GM crops would affect the seed purity of a non-GM crop variety<sup>4</sup>. Therefore, it is necessary to research on the reliable detection of GM organisms, which is one of the most significant consumer concerns regarding food quality and safety.

Numerous techniques have been proposed for identification and classification of GM organisms, such as polymerase chain reaction (PCR)<sup>5,6</sup>, enzyme-linked immune sorbent assay (ELISA)<sup>7,8</sup>, microarrays<sup>9,10</sup>, electrophoresis<sup>11,12</sup>, biosensors<sup>13</sup>, chips<sup>14</sup>, mass spectrometry<sup>15</sup> and chromatographic protein profiles<sup>16</sup>. However, most of these methods are destructive, time consuming, costly, difficult operations, and impossible for rapid on-site

<sup>1</sup>School of Biotechnology and Food Engineering, Hefei University of Technology, Hefei 230009, China. <sup>2</sup>Intelligent Control and Compute Vision Lab, Hefei University, Hefei 230601, China. <sup>3</sup>Department of Food, Nutrition and Packaging Sciences, Clemson University, Clemson, SC 29634, United States. <sup>4</sup>Rice Research Institute, Anhui Academy of Agricultural Sciences, Hefei 230031, China. <sup>5</sup>School of Medical Engineering, Hefei University of Technology, Hefei 230009, China. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to J.Y. (email: yjianbo@263.net) or L.Z. (email: lzheng@hfut.edu.cn)



**Figure 1. Average THz time-domain spectra of glyphosate-resistant (DP4546RR) and conventional (Wandou 28) soybean seeds and their hybrid descendants (DP4546RR × Wandou 28).**

measurements. As non-destructive technologies, spectroscopic techniques are rapid and easy to operate without complicated sample preparations. Infrared spectroscopy<sup>17</sup>, near infrared (NIR)<sup>18–21</sup>, visible/near infrared (VIS-NIR)<sup>22</sup>, and multispectral imaging<sup>23,24</sup> techniques combined with chemometric methods have shown their success in the rapid identification of GM organisms. Although many of the spectroscopic techniques mentioned above have been used to identify GM organisms, little attention has been paid to the use of terahertz (THz) spectroscopy for the detection of GM organisms.

THz spectroscopy usually refers to electromagnetic wave with the frequency of 0.1–10 THz (wavelength 30  $\mu\text{m}$ –3 mm), and the band between microwave and infrared<sup>25</sup>. THz radiation is low-energy, non-ionizing and can penetrate a wide variety of non-conducting materials such as paper, wood, and plastic but it does not penetrate metal or water<sup>26</sup>. More importantly, theoretical studies show that most biological molecules, such as DNA components, protein and amino acids exhibit fingerprint spectra in the THz region<sup>27</sup>. Recently, THz spectroscopy together with chemometric methods is increasingly used in the fields of agricultural and food industry<sup>27–29</sup>. Moreover, it has also recently emerged as a powerful approach for the discrimination of different transgenic cottons and rice<sup>30–32</sup>. However, the detection of GM organisms using THz spectroscopy in previous studies needed to press samples into sheets or pellets, and up till now, there are only a few research reports about the application of THz spectroscopy technique for rapid and non-destructive detection of GM organisms without any sample preparation<sup>33</sup>.

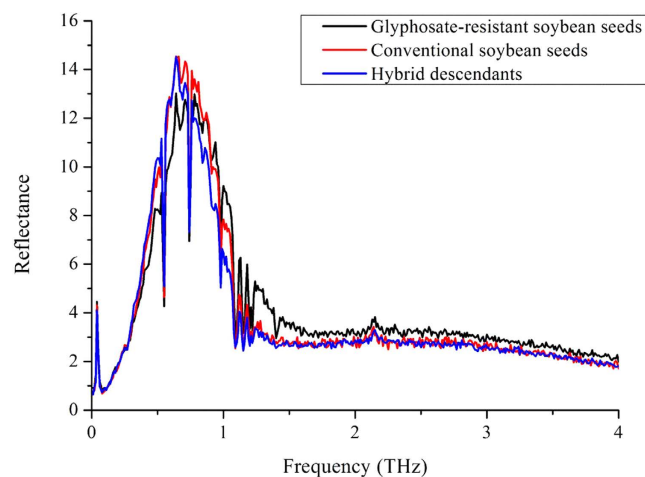
Thus, the objective of the current study was to develop a non-disruptive method to discriminate among glyphosate-resistant, conventional and hybrid descendant seeds without sample preparation using THz spectroscopy combined with different chemometric methods including principal component analysis (PCA), least squares-support vector machines (LS-SVM) and PCA-back propagation neural network (PCA-BPNN).

## Results

**Spectra analysis.** The average amplitudes of glyphosate-resistant (DP4546RR) and conventional (Wandou 28) soybean seeds and their hybrid descendants (DP4546RR × Wandou 28) in the time-domain spectra were shown in Fig. 1. Clearly, the general trend of all spectra was very similar, except for some subtle alterations. Although there was apparent time delay in time-domain spectra, no significant difference in the waveforms and the pulse amplitudes was observed among glyphosate-resistant and conventional soybean seeds and their hybrid descendants. Therefore, the frequency-domain spectra were obtained by transformation from time-domain spectra with fast Fourier transform (FFT) which was shown in Fig. 2. As shown in Fig. 2, these three classes of soybean seeds all had distinct absorption peaks at around 0.6 THz and 0.8 THz. The most obvious difference between 0.5 THz and 1.5 THz could be attributed to the chemical differences among glyphosate-resistant and conventional soybean seeds and their hybrid descendants. However, the spectra of these three classes of soybean seeds overlapped and were hard to differ from each other. Thus, it is difficult to discriminate these three classes of soybean seeds based on frequency-domain spectra with simple methods.

**Surface reflection intensity and penetration depth of soybean seeds.** Surface reflection intensity and penetration depth of glyphosate-resistant and conventional soybean seeds and their hybrid descendants were shown in Figs 3 and 4. Figure 3 showed that the surface reflection intensity of glyphosate-resistant soybean seed was higher than the non-transformed variety and the F1 progeny of the cross DP4546RR × Wandou 28. In addition, the penetration depth of glyphosate-resistant soybean seeds was relatively thinner than for seeds obtained from conventional variety Wandou 28 and progeny from crosses of DP4546RR × Wandou 28 (Fig. 4).

**Varieties discrimination.** PCA was performed initially to examine the qualitative difference of glyphosate-resistant and conventional soybean seeds and their hybrid descendants in principal component (PC) space. All the raw reflectance spectra obtained from the 60 glyphosate-resistant (DP4546RR), 60 conventional (Wandou 28) and 60 progeny seeds (DP4546RR × Wandou 28) were used for PCA. Figure 5 shows the three dimensional PC score plot of the samples. The results indicated that the initial three principal components (PCs),



**Figure 2.** Average reflectance spectra of glyphosate-resistant (DP4546RR) and conventional (Wandou 28) soybean seeds and their hybrid descendants (DP4546RR × Wandou 28).

which account for 93% of the spectral variations (65.68%, 24.97%, and 3.7% for PC1, PC2, and PC3, respectively), allows for a clear differentiation among the three seed samples. Furthermore, these results suggested that discrimination among glyphosate-resistant, conventional and their hybrid descendant seeds was possible and that the different spectral attributes of these samples were associated with the characteristics of the seeds.

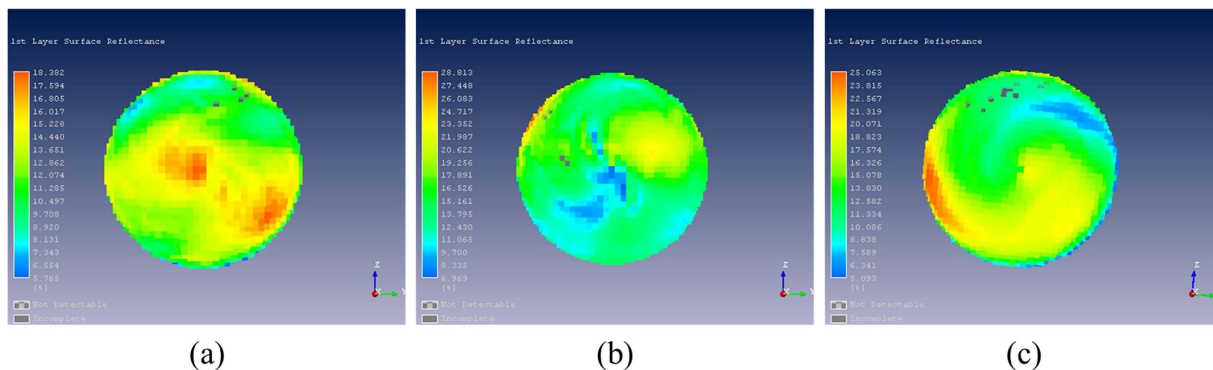
Back propagation neural network (BPNN) was performed on the first 20 PCs that contain more than 99.5% of the variation in the raw spectra data, which is defined as PCA-BPNN. In the PCA-BPNN method, to produce a high accuracy, the optimum values of the hidden nodes, the goal error and iteration times were determined to be 20,  $1 \times 10^{-8}$  and 800, respectively. With these optimal parameters, the soybean seeds could be classified and the discrimination results were listed in Table 1. Compared with the discrimination accuracies from the models using the raw spectra and the spectra with the first and second derivative pre-treatments, the one using the spectra with the standard normal variate (SNV) transformation pre-treatment achieved the best discrimination accuracies (the accuracies were 89.17% and 76.67% in calibration and validation set, respectively).

Least squares-support vector machine (LS-SVM) with the spectra range between 0.5 and 1.5 THz was used to build the calibration models. In the stage of model development using LS-SVM with radial basis function (RBF) kernel, the crucial step was the optimization of the meta-parameters  $\gamma$  and  $\sigma^2$  because their values determined the boundary complexity and the prediction performance. The optimal combination of  $(\gamma, \sigma^2)$  for soybean seeds discrimination were found at the value of (0.25, 1) on the raw spectra, (6.9644, 0.015625) on the spectra with the first derivative, (2.82843, 0.0625) on the spectra with the second derivative, and (2, 0.0625) on the spectra with SNV, respectively. The discrimination results were also listed in Table 1. Similar to BPNN, the LS-SVM method using the spectra with SNV pre-treatment produced the best discrimination accuracies as compared with the raw spectra and the spectra with the first and second derivative pre-treatments. The discrimination accuracies in the calibration and validation set were 97.5% and 88.33%, respectively.

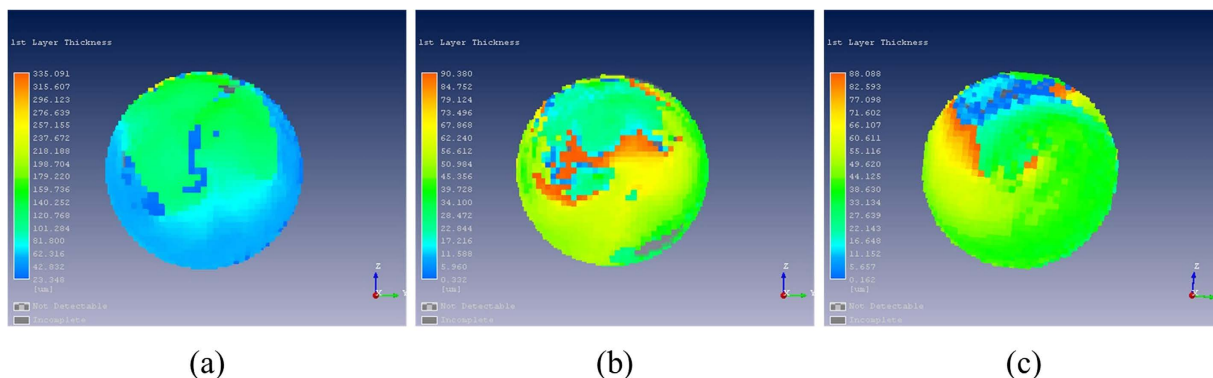
In order to analyse which seeds were more difficult to differentiate from the others, the best model using LS-SVM and the spectra with SNV pre-treatment were selected to assess its classification performance for differentiating the two parental varieties (DP4546RR and Wandou 28) and their offspring (DP4546RR × Wandou 28) of soybean seeds by using a confusion matrix. Table 2 lists the observation numbers, sensitivity and specificity of each set of seeds in validation data set. The highest number of misclassified seeds was observed between glyphosate-resistant soybean seeds and hybrid progeny, where three samples of glyphosate-resistant soybean seeds were wrongly classified as samples of hybrid descendants.

## Discussion

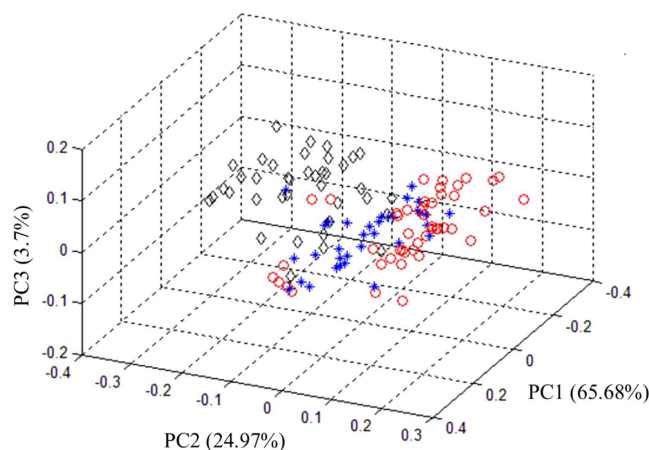
Different varieties of soybean seeds may have different chemical components and seed coat structure<sup>34–36</sup>. THz waves have the ability to penetrate a wide variety of materials, and vibration and rotational energy levels of most biological molecules (DNA, protein) are in the THz band. Thus, there was an apparent time delay in the time-domain spectra and different reflectance in the frequency-domain spectra of the glyphosate-resistant variety DP4546RR, the conventional variety Wandou 28 and their hybrid progeny seeds which were shown in Figs 1 and 2. In Fig. 1, there were significant differences in time delay among seeds from the two parental varieties and their progeny, which might be attributed to differences in chemical composition and coat structure of the seeds. Furthermore, as shown in Fig. 2, these three sets of soybean seeds showed a similar trend in absorption spectra between 0.1 THz and 4 THz, especially from 0.5 THz to 1.5 THz. Apparent absorption peaks at around 0.6 THz and 0.8 THz were observed which may be due to the water absorption of soybean seed samples in the THz region<sup>37</sup>. In Figs 3 and 4, when comparing the surface reflection intensity and the penetration depth of these three types of seeds, the hardness or thickness of the seed coat of the glyphosate-resistant soybean variety seem to be lower/finer than the Wandou 28 variety with offspring of the two varieties showing intermediate hardness/thickness as expected. From the Table 2, the highest frequency of misclassification of seeds was observed between



**Figure 3.** Surface reflection intensity of soybean seeds. (a) Glyphosate-resistant soybean seed; (b) Conventional soybean seed; (c) Hybrid descendants.



**Figure 4.** Penetration depth of soybean seeds. (a) Glyphosate-resistant soybean seed; (b) Conventional soybean seed; (c) Hybrid descendants.



**Figure 5.** Three dimensional score plot of the first three principal components for the glyphosate-resistant (DP4546RR,  $\diamond$ ) and conventional (Wandou 28,  $*$ ) soybean seeds and their hybrid descendants (DP4546RR  $\times$  Wandou 28,  $\circ$ ).

glyphosate-resistant soybean seeds and hybrid descendants, which might be due to the similar chemical components of these two varieties.

Chemometrics can improve the understanding of chemical information and correlate quality parameters or physical properties using mathematical and statistical methods. The advanced methods can enable a more thorough scientific characterization of samples than the information obtained from simple analysis. In this study, the classification of the transgenic and conventional soybean seeds and their hybrid descendants with THz spectroscopy can be seen as a high-dimensions and non-linear pattern recognition problem. Each chemometric

Chemometric methods	Data pre-treatment	Number of misclassified samples in calibration set	Number of misclassified samples in validation set	Accuracy in calibration set (%)	Accuracy in validation set (%)
LS-SVM	no	4	10	96.67	83.33
	1 <sup>st</sup> derivative	2	12	98.33	80
	2 <sup>nd</sup> derivative	2	13	98.33	78.33
	SNV	3	7	97.5	88.33
PCA-BPNN	no	12	15	90	75
	1 <sup>st</sup> derivative	16	17	86.67	71.67
	2 <sup>nd</sup> derivative	19	20	84.17	66.67
	SNV	13	14	89.17	76.67

**Table 1. Comparison of discrimination performance obtained with PCA-BPNN and LS-SVM methods and the THz spectra with different pre-treatments.**

Actual class	Predicted class		
	Glyphosate-resistant	Conventional	Hybrid descendants
Glyphosate-resistant	16	1	3
Conventional	0	18	2
Hybrid descendants	1	0	19
Sensitivity (%)	80	90	95
Specificity (%)	100	95	87.5

**Table 2. Matrix detailing the classification results of seeds from glyphosate-resistant (DP4546RR), conventional (Wandou 28) and the progeny (DP4546RR × Wandou 28) using the LS-SVM method and the THz spectra with SNV pre-treatment.**

method has its own advantages and applications. As seen from Table 1, the best discrimination accuracies were obtained using LS-SVM method combined with the spectra with SNV pre-treatment for discrimination of glyphosate-resistant and conventional soybean seeds and their hybrid descendant. The reasons may be as follow: (1) SVM is specifically designed to operate in high-dimensional characteristic space with fewer training variables or samples; (2) It has effective performance for multivariate function estimation especially for non-linear classification.

## Methods

**Soybean seed samples.** Glyphosate-resistant GM soybean seeds (ROUNDUP READY soybean DP4546RR, Monsanto Canada Inc.), conventional soybean seeds (Wandou 28), and their hybrid descendants (DP4546RR × Wandou 28) were provided. The F1 progeny of the cross DP4546RR × Wandou 28 was resistant to glyphosate, but was not genetically homogenous. In the current study, three varieties (glyphosate-resistant and conventional soybean seeds and their hybrid descendants) were defined. A total of 180 soybean seed samples were randomly divided into a calibration set (2/3, including 40 seeds of each sample) and a validation set (1/3, including 20 seed samples of each plant variety).

**THz spectroscopy imaging system.** THz time-domain spectra of all the intact samples, including 60 glyphosate-resistant (DP4546RR) and 60 conventional (Wandou 28) soybean seeds and 60 their hybrid descendants (DP4546RR × Wandou 28), were collected by a Fiber-Coupled (FiCO) terahertz spectroscopy imaging system (Zomega Terahertz Corporation, East Greenbush, NY, USA) with independent fiber-coupled emitter and receiver heads. The system is designed to perform both transmission and reflection spectroscopy and imaging measurements in the range from 0.1 THz to 4 THz, with a waveform acquisition speed up to 500 Hz. For this study, the measurements were conducted in a reflection mode. All of the spectra were collected after the spectroscopy system warmed up for half an hour to reach a stable state. The measurements were carried out under dry air. Firstly, a measure of the noise level was obtained by blocking the path of the THz beam with a metallic plate. The measured signal provides the spectral characteristics of the noise of the system. The noise level allows determining the dynamic range of the system as a function of frequency as the difference (in dB) between the maximum signal available and the noise level. For FiCO, the peak dynamic range (in power) is around 50 dB. And then, each soybean seed sample was measured.

**Spectra data pre-treatment.** For each sample, the acquired time-domain spectra data were transformed into the frequency-domain spectra using the FFT. The shape of the time-domain data is a pulse with duration of the order of the picosecond, which is equivalent to terahertz in the frequency-domain. In this study, derivative processes, including the first and second derivative, and SNV were selected as spectra pre-treatment methods.

**Chemometric methods.** Multivariate analysis including PCA, LS-SVM, and PCA-BPNN were selected as the chemometric methods in present study to classify and screen these three varieties. All the selected chemometric methods were compared and the accuracy results obtained from the calibration and validation set were summarized accordingly. All of these chemometrics analysis and statistics were performed using the commercial software Matlab 2009 (The Mathworks Inc., Natick, MA, USA), Image J 1.48 and Origin 8.5.

## References

- JAMES, C. Executive Summary. Global Status of Commercialized Biotech/GM Crops: 2010. ISAAA Brief 42–2010. Ithaca, NY: ISAAA. Available from: <http://www.isaaa.org/resources/publications/briefs/42/executivesummary/> (2010).
- Erickson, E. H. Effect of honey bees on yield of three soybean cultivars. *Crop Sci.* **15**, 84–86 (1975).
- Yoshimura, Y., Matsuo, K. & Yasuda, K. Gene flow from GM glyphosate-tolerant to conventional soybeans under field conditions in Japan. *Environ. Biosafety Res.* **5**, 169–173 (2006).
- Rong, J. *et al.* Low frequency of transgene flow from Bt/CpTI rice to its nontransgenic counterparts planted at close spacing. *New Phytol.* **168**, 559–566 (2005).
- Abdullah, T., Radu, S., Hssan, Z. & Hashim, J. K. Detection of genetically modified soy in processed foods sold commercially in Malaysia by PCR-based methods. *Food Chem.* **98**, 575–579 (2006).
- Dörries, H. H., Remus, I., Grönwald, A., Grönwald, C. & Berghof-Jäger, K. Development of a qualitative, multiplex real-time PCR kit for screening of genetically modified organisms (GMOs). *Anal. Bioanal. Chem.* **396**, 2043–2054 (2010).
- Liu, G. *et al.* Liquid-phase hybridization based PCR-ELISA for detection of genetically modified organisms in food. *Food Control* **15**, 303–306 (2004).
- Shan, G., Embrey, S. K. & Schafer, B. W. A highly specific enzyme-linked immunosorbent assay for the detection of Cry1Ac insecticidal crystal protein in transgenic WideStrike cotton. *J. Agric. Food Chem.* **55**, 5974–5979 (2007).
- Wang, J. From DNA biosensors to gene chips. *Nucleic Acids Res.* **28**, 3011–3016 (2000).
- Leimani, S. *et al.* A microarray-based detection system for genetically modified (GM) food ingredients. *Plant Mol. Biol.* **61**, 123–139 (2006).
- Obeid, P. J., Christopoulos, T. K. & Ionno, P. C. Rapid analysis of genetically modified organisms by in-house developed capillary electrophoresis chip and laser-induced fluorescence system. *Electrophoresis* **25**, 922–930 (2004).
- Levandi, T., Leon, C., Kaljurand, M., Garcia-Cañas, V. & Cifuentes, A. Capillary electrophoresis time-of-flight mass spectrometry for comparative metabolomics of transgenic versus conventional maize. *Anal. Chem.* **80**, 6329–6335 (2008).
- Mannelli, I., Minunni, M., Tombelli, S. & Mascini, M. Quartz crystal microbalance (QCM) affinity biosensor for genetically modified organisms (GMOs) detection. *Biosens. Bioelectron.* **18**, 129–140 (2003).
- Nesvold, H., Kristoffersen, A. B., Holst-Jensen, A. & Berdal, K. G. Design of a DNA chip for detection of unknown genetically modified organisms (GMOs). *Bioinformatics* **21**, 1917–1926 (2005).
- Ocaña, M. E., Fraser, P. D., Patel, R. K. P., Halket, J. M. & Bramley, P. M. Mass spectrometry detection of CP4 EPSPS in genetically modified soya and maize. *Rapid Commun. Mass Spectrom.* **21**, 319–328 (2007).
- García, M. C. *et al.* Rapid characterisation of (glyphosate tolerant) transgenic and non-transgenic soybeans using chromatographic protein profiles. *Food Chem.* **113**, 1212–1217 (2009).
- Alcantara, G. B. *et al.* Assessment of genetically modified soybean crops and different cultivars by Fourier transform infrared spectroscopy and chemometric analysis. *Orbital-Electron. J. Chem.* **2**, 41–52 (2010).
- Roussel, S. A., Hardy, C. L., Hurburgh, C. R. & Rippke, G. R. Detection of Roundup Ready™ soybeans by near-infrared spectroscopy. *Appl. Spectrosc.* **55**, 1425–1430 (2001).
- Lee, J. H. & Choung, M. G. Nondestructive determination of herbicide-resistant genetically modified soybean seeds using near-infrared reflectance spectroscopy. *Food Chem.* **126**, 368–373 (2011).
- Agelet, L. E., Gowen, A. A., Hurburgh, C. R. Jr. & O'Donnell, C. P. Feasibility of conventional and Roundup Ready® soybeans discrimination by different near infrared reflectance technologies. *Food Chem.* **134**, 1165–1172 (2012).
- Luna, A. S., da Silva, A. P., Pinho, J. S. A., Ferré, J. & Boqué, R. Rapid characterization of transgenic and non-transgenic soybean oils by chemometric methods using NIR spectroscopy. *Spectrochim. Acta A* **100**, 115–119 (2013).
- Xie, L., Ying, Y., Ying, T., Yu, H. & Fu, X. Discrimination of transgenic tomatoes based on visible/near-infrared spectra. *Anal. Chim. Acta* **584**, 379–384 (2007).
- Liu, C. *et al.* Nondestructive determination of transgenic *Bacillus thuringiensis* rice seeds (*Oryza sativa* L.) using multispectral imaging and chemometric methods. *Food Chem.* **153**, 87–93 (2014).
- Liu, C. *et al.* Non-destructive discrimination of conventional and glyphosate-resistant soybean seeds and their hybrid descendants using multispectral imaging and chemometric methods. *J. Agric. Sci.* **154**, 1–12 (2016).
- Qin, J., Xie, L. & Ying, Y. Determination of tetracycline hydrochloride by terahertz spectroscopy with PLSR model. *Food Chem.* **170**, 415–422 (2015).
- Redo-Sanchez, A. *et al.* Assessment of terahertz spectroscopy to detect antibiotic residues in food and feed matrices. *Analyst* **136**, 1733–1738 (2011).
- Qin, J., Ying, Y. & Xie, L. The Detection of Agricultural Products and Food Using Terahertz Spectroscopy: A Review. *Appl. Spectrosc. Rev.* **48**, 439–457 (2013).
- Gowen, A. A., O'Sullivan, C. & O'Donnell, C. P. Terahertz time domain spectroscopy and imaging: Emerging techniques for food process monitoring and quality control. *Trends Food Sci. Technol.* **25**, 40–46 (2012).
- Redo-Sanchez, A., Laman, N., Schulkin, B. & Tongue, T. Review of terahertz technology readiness assessment and applications. *J. Infrared Milli. Terahz. Waves* **34**, 500–518 (2013).
- Liu, J., Mao, L., Ku, J., He, J. & Fan, L. A THz spectroscopy non-destructive identification method for transgenic cotton seed based on GA-SVM. *Opt. Quant. Electron.* **47**, 313–322 (2014).
- Liu, J. *et al.* Identification of GMOs by terahertz spectroscopy and ALAP-SVM. *Opt. Quant. Electron.* **47**, 685–695 (2015).
- Xu, W. *et al.* Discrimination of transgenic rice containing the Cry1Ab Protein using terahertz spectroscopy and chemometrics. *Sci. Rep.* **5**, 11115 (2015).
- Liu, W., Liu, C., Hu, X., Yang, J. & Zheng, L. Application of terahertz spectroscopy imaging for discrimination of transgenic rice seeds with chemometrics. *Food Chem.* **210**, 415–421 (2016).
- Kumar, V., Rani, A., Solanki, S. & Hussain, S. M. Influence of growing environment on the biochemical composition and physical characteristics of soybean seed. *J. Food Compos. Anal.* **19**, 188–195 (2006).
- Mullin, W. J. & Xu, W. Study of soybean seed coat components and their relationship to water absorption. *J. Agric. Food Chem.* **49**, 5331–5335 (2001).
- Redondo-Cuenca, A., Villanueva-Suárez, M. J., Rodríguez-Sevilla, M. D. & Mateos-Aparicio, I. Chemical composition and dietary fibre of yellow and green commercial soybeans (*Glycine max*). *Food Chem.* **101**, 1216–1222 (2006).
- Exter, M. V., Fattinger, C. & Grischkowsky, D. Terahertz time-domain spectroscopy of water vapor. *Opt. Lett.* **14**, 1128–1130 (1989).

## Acknowledgements

This study is supported by the National Key Research and Development Plan of China (2016YFD0401104), the Key Science & Technology Specific Projects of Anhui Province (15czz03117), the National Natural Science Foundation of China (31401544), the Key Special Foundation for Young Scientists of Anhui Province (2013SQRL077ZD), the Funds for Huangshan Professorship of Hefei University of Technology (407-037019), the Talent Research Funds of Hefei University(15RC09), and the Fundamental Research Funds for the Central Universities (JZ2016HG TB0712).

## Author Contributions

J.B.Y. and L.Z. designed the project; W.L. and C.H.L. carried out the experiments; W.L. simulated the model and analysed the data; W.L., F.C., J.B.Y. and L.Z. wrote the manuscript. All authors discussed the results and commented on the manuscript.

## Additional Information

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Liu, W. *et al.* Discrimination of transgenic soybean seeds by terahertz spectroscopy. *Sci. Rep.* **6**, 35799; doi: 10.1038/srep35799 (2016).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016