Metabolome Comparison of Transgenic and Non-transgenic Rice by Statistical Analysis of FTIR and NMR Spectra

Keykhosrow KEYMANESH, Mohammad Hassan DARVISHI, Soroush SARDARI

(*Biotechnology Department, Pasteur Institute of Iran, #69 Pasteur Ave., Tehran, Iran*)

Abstract: Modern biotechnology, based on recombinant DNA techniques, has made it possible to introduce new traits with great potential for crop improvement. However, concerns about unintended effects of gene transformation that possibly threaten environment or consumer health have persuaded scientists to set up pre-release tests on genetically modified organisms. Assessment of 'substantial equivalence' concept that established by comparison of genetically modified organism with a comparator with a history of safe use could be the first step of a comprehensive risk assessment. Metabolite level is the richest in performance of changes which stem from genetic or environmental factors. Since assessment of all metabolites in detail is very costly and practically impossible, statistical evaluation of processed data of grain spectroscopic values could be a time and cost effective substitution for complex chemical analysis. To investigate the ability of multivariate statistical techniques in comparison of metabolomes as well as testing a method for such comparisons with available tools, a transgenic rice in combination with its traditionally bred parent were used as test material, and the discriminant analysis were applied as supervised method and principal component analysis as unsupervised classification method on the processed data which were extracted from Fourier transform infrared spectroscopy and nuclear magnetic resonance spectral data of powdered rice and rice extraction and barley grain samples, of which the latter was considered as control. The results confirmed the capability of statistics, even with initial data processing applications in metabolome studies. Meanwhile, this study confirms that the supervised method results in more distinctive results.

Key words: rice; principal component analysis; discriminant analysis; nuclear magnetic resonance; Fourier transform infrared spectroscopy; transgene; safety assessment; metabolome analysis

The genetically modified (GM) plants are made through introduction of new genes possibly from distant phylogenic species. These plants have shown improved performance in facing abiotic stress [1] and pest and pathogen attacks [2-3], or production of more qualified or nutritionally valuable products [4]. As the area under cultivation of GM plants and their products' portion in the market is increasing around the world, more consumers are exposed to the outputs of recombinant DNA techniques [5]. This tendency has led to debates about the possible unintended effects of new products on environment as well as the consumer health. As a result, scientific bodies suggest safety assessment procedures through which some of them have acquired international acceptance [6]. The 'substantial equivalence' concept which is based on the comparison of a GM product with a traditionally bred parent with a long record of safe usage is usually the first step in safety supervision [7].

Fingerprinting approaches at metabolite level could be very informative and reflect the changes which stem from genetic manipulation, but investigation of all chemical components of the target organism is technically very difficult. On the other hand, statistical analysis of spectra results from spectroscopy techniques could be a time and cost effective and accurate way for metabolome comparison studies [8].

There are many reports about metabolome studies based on nuclear magnetic resonance (NMR) and Fourier transform infrared spectroscopy (FTIR). These studies cover a vast range from medical investigations to safety-intended studies of GM products [9-11]. Reaching a convincing result by these methods depends on efficient exploitation of data from the spectra and then analysis of the data through a plausible classification method. Generally, there are two main classification methods. First, supervised method, in which the predefined classes are the starting point for the analysis process that leads to models based on the input clusters, and the second one

is unsupervised method that in a multivariate manner assesses how similar a set of samples are and the samples with more similarities are included to a group. Eventually, every statistical procedure for classification belongs to one of these two major methods [12].

Usually, experiments with similar aims as ours are done by use of a few specialized computer applications. However, in this study, in addition to assessment of the transgenic rice in comparison with its parental variety, we intended to examine the power of less specialized and more available softwares for such experiments. We used transgenic rice in combination with its traditionally bred parent as test materials and exerted linear discriminant analysis as supervised and principal component analysis as unsupervised classification methods on the processed data extracted from NMR and FTIR spectra of transgenic and non-transgenic rice and barley grain samples which were used as control.

MATERIALS AND METHODS

Plant materials

The seed samples were derived from the rice variety Tarom mulaiee and its counterpart variety which is transformed with *cry1Ab* gene from *Bacillus thuringinesis* (*Bt*). The samples have been provided from transgenic plants and their parental variety which was grown similarly as control. For further processing of collected data and spectral analysis, the different samples taken from the same non-transgenic and transgenic rice seeds samples numbered 1–10 and 11–20, respectively. A barley sample of an unknown variety was used merely as control. The treatment procedures on this sample were like the rice samples.

DNA was extracted from rice samples and the existence of the transgene was confirmed by use of relevant primers and observation of concerned band (Fig. 1). The DNA was extracted as described by Wang et al [13], and subjected to one cycle of 94ºC for 5 min, 35 cycles of three steps each (94ºC, 1 min; 60ºC, 1 min; and 72ºC, 3 min) in 25 µL of PCR buffer (10 mmol/L Tris-HCl, pH 8.4, 50 mmol/L KCl and 1.5 mmol/L MgCl$_2$) containing 0.2 mmol/L of each dNTP, 20 ng each of RG100 primers, and 40 ng of each of *hpt* or *cry1Ab Bt* primers and 1 U of *Taq*
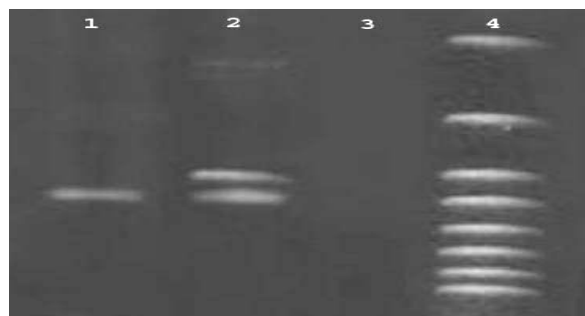

Fig. 1. PCR analysis of seeds from putative transgenic rice.
    PCR primers for rice locus RG100 (1.0 kb product) and *cry1Ab* gene (1.2 kb product) as described by Ghareyazie et al [14]. Sample DNA: Lane 1, Non-transgenic rice; Lane 2, Transgenic rice; Lane 3, Water control (no DNA); Lane 4, Molecular weight markers. The primers for locus RG100 are 5′-GCTGGACGTGCCAAAGAGAG-3′ (forward) and 5′-CGAACCACAGCCACAGCATG-3′ (reverse). The primers for the *cry1Ab* gene are 5′-GGCGGCGAGAGGATCGAGAC-3′ (forward) and 5′-GGCGGGACGTTGTTGTTC-3′ (reverse).

polymerase. PCR products were then analyzed by agarose gel electrophoresis [14].

H-NMR sample preparation

For each sample, 2.5 g of complete rice grain was powdered. Sample extraction was carried out in three steps. In the first and second steps, 10 mL methanol was added to the rice powder in a flask. The third step was addition of 15 mL of 2:1 mixture of methanol and dichloromethane. In each step, the powdered rice grain and solvent were stirred for 45 min and the resulting solution was transmitted to a glass vessel through filter paper. After solvent volatilization, a 13 mg per 700 µL solution was prepared as sample for H-NMR spectroscopy by addition of d6-DMSO to the residue. The NMR instrument was a 400 MHz Brucker.

FTIR sample preparation

For each sample, 2.5 g of complete rice grain was fine powdered. The thin tablets used as samples in FTIR were prepared by mixing the fine powdered rice with KBr with 2% ratio. The FTIR instrument was a Brucker Tensor 27.

Statistical analysis

Linear discriminant analysis (LDA) is a frequently used supervised classification method. Discriminant analysis finds a set of prediction equations based on independent variables that are used to classify individuals into groups [15]. The main target of principal

component analysis (PCA) is to reduce the number of variables in such a way that most of the variation would be expressed by new uncorrelated variables or principal components. Based on principal components, classification of samples is possible in a way that between-class variance is maximized and individuals in the same group have the least difference [16].

There are many commercial computer applications that are used successfully in similar experiences. In this study, we used NCSS (2006 release) and MESTREC (version 4.9.9, 2006) to implement the analysis process from the first phase which is transformation of spectroscopy diagram in a manner that can be used as an input material for the future multivariate analysis.

Preparing data for analysis of H-NMR and FTIR spectra

Using the MESTREC software, H-NMR spectra of samples were transformed in such a way, in which every spectrum was redefined as a point in a multi-dimensional space. In fact, it was the key operation in which its results were used as input for the oncoming analysis phase. For the FTIR spectra, according to the above-mentioned process, for first stage operation, we used area under the curve option of NCSS application

to calculate the area under the curve for 187 limited areas which were defined for all spectra. In this way, the results for each spectrum mirrored the shape of it; therefore, these data were used as raw inputs for the statistical analysis stage leading to classification. Fig. 2 shows the way of data preparation for H-NMR and FTIR experiments in a graphical manner.

RESULTS

As shown in Fig. 3, principal component analysis indicated no definite clusters. The points that represent the samples are dispersed in a non-distinctive manner in the PCA diagram of FTIR. The first and second scores represent 69.89% and 17.41% of total variations, respectively. The PCA diagram of H-NMR is better clustered though the accumulation of samples has no rational meaning. In this case, the first and second scores cover 23.33% and 19.97% of variations, respectively. The difference between diagrams could be due to the fact that in the latter case only 43.3% of the total variations has been involved in contrast to 87.3% of FTIR case, so the less variation among individuals has led to less disseminated points.

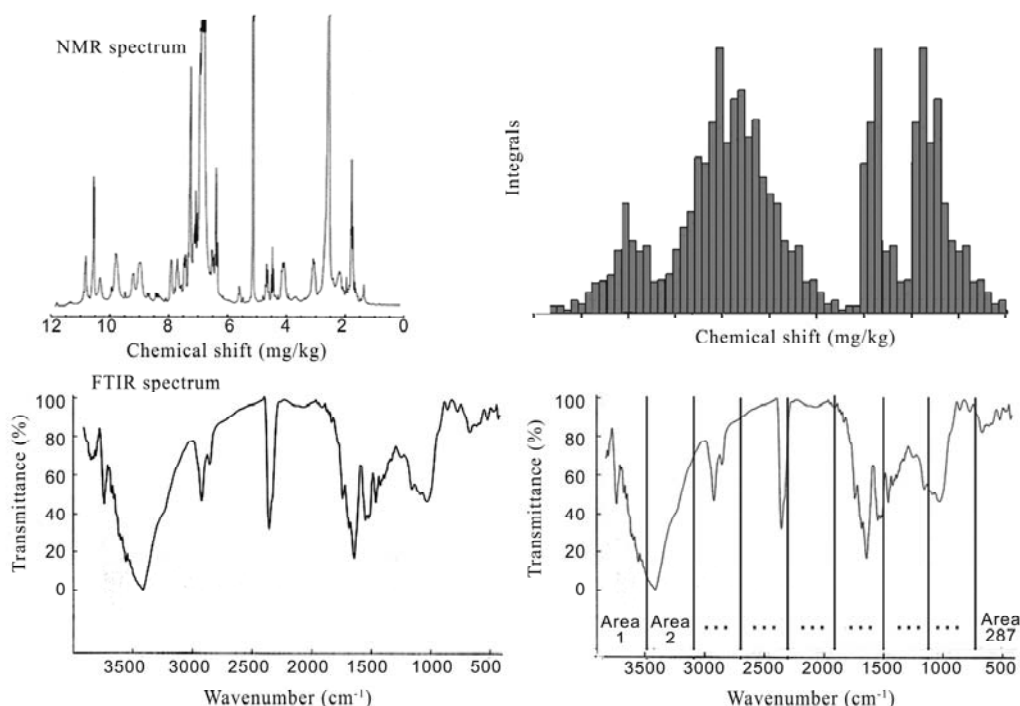Classification of samples by LDA was carried out

Fig. 2. The way that the H-NMR and FTIR spectra were processed for use in statistical analysis.
The images show graphically how the spectroscopic graphs were prepared for data extraction. The images are merely figurative and for better understanding of the process.
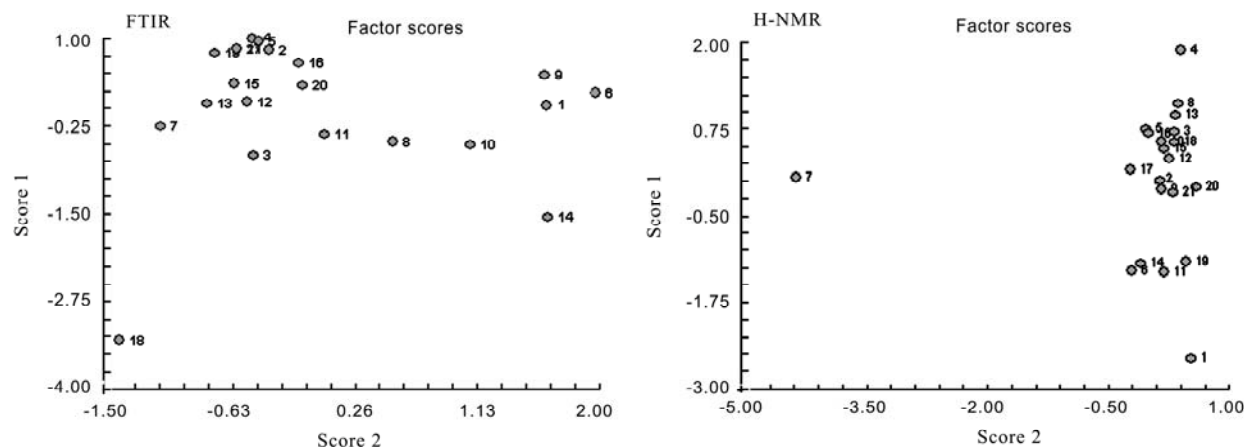
Fig. 3. Diagrams resulted from principal component analysis of FTIR and H-NMR spectra.
Numbers 1 to 10 are common rice samples, and numbers 11 to 20 are transgenic rice samples and number 21 is a barley sample.

through stepwise selection in which retention or removal of variables depended on PIN and POUT values in NCSS application, respectively. For PIN, 0.5 and for POUT, 0.99 are known as suitable values since these amounts would permit the most number of efficient variables. As indicated in Fig. 4, for FTIR spectra, discriminant functions enabled the classification with 95.23% accuracy. H-NMR spectra of samples were classified with 100% accuracy.

DISCUSSION

The results showed the advantage of supervised methods over unsupervised one in case of classification. Similar conclusion has been acquired in other studies[17]. However, we should have in mind that unsupervised methods would be more reliable in the case that there is no record of the samples in hand.

At this level, statistical comparison of extracts

from different plant materials only informed us whether the compared materials have any significant differences and based on which the materials could be distinguished and be classified in different groups.

From the statistical view point, abundance of variables which are calculated integral of limited counterpart areas of spectra obliged us to use data reduction procedures such as PCA and LDA that not only classify the samples but also implement the classification by use of the variables which represent a major portion of the variation.

The differences between transgenic and non-transgenic materials could have different reasons. These differences may be resulted from metabolites that the genetic manipulation is done for their production. On the other hand, the differences may be the result of increase or decrease in production of metabolites other than the targeted one. These changes may include some unpredicted events which probably
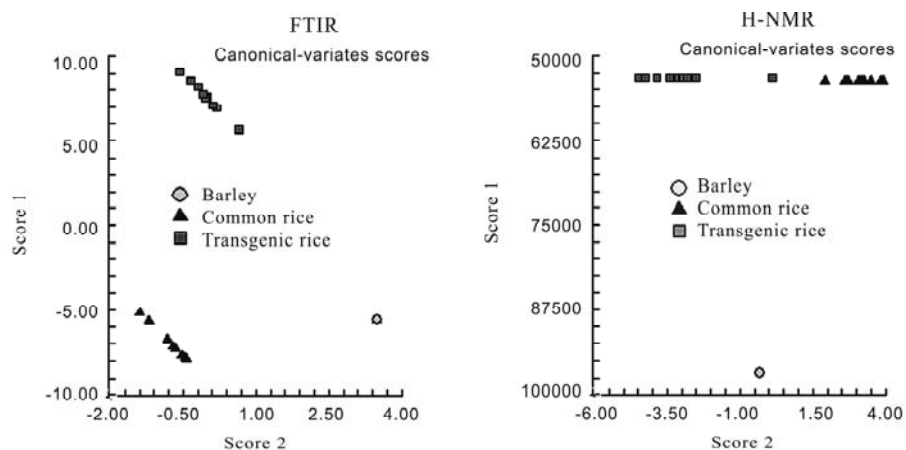

Fig. 4. Diagrams resulted from linear discriminant analysis of FTIR and H-NMR spectra.

lead to unwanted effects on consumer health. Therefore, identifying the reason of differences between transgenic and their non-transgenic counterparts could be the next step in such investigations. In addition, the statistical comparison of plant materials has further advantages than a primitive stage in regulations related to transgenic or genetically modified (GM) products; in fact, it can be used as a quick method for identification of transgenic products in cases that there is no reliable information about the plant material, e.g. in an imported consignment. Due to increasing share of GM products, the improvement and facilitation of identification methods play a vital role in accurate implementation of regulations which are approved in many countries for control and assessment of GM products [18].

REFERENCES

1   Thomsona J A, Mundreea S G, Farrant J M. The development of genetically modified maize for abiotic stress tolerance. *S Afr J Bot*, 2007, **73**: 494–495.

2   Rafiq M, Fatima T, Husnain T, Bashir K, Khan M A, Riazuddin S. Regeneration and transformation of an elite inbred line of maize (*Zea mays* L.), with a gene from *Bacillus thuringiensis*. *S Afr J Bot*, 2006, **72**: 460–466.

3   Sharma A, Sharma R, Imamura M, Yamakawa M, Machii H. Transgenic expression of cecropin B, an antibacterial peptide from *Bombyx mori*, confers enhanced resistance to bacterial leaf blight in rice. *FEBS Lett*, 2000, **484**: 7–11.

4   Alexander J S, Sachdev H P S, Matin Q. Genetic engineering for the poor: Golden rice and public health in India. *World Dev*, 2008, **36**: 144–158.

5   Thomson J. Genetically modified food crops for improving agricultural practice and their effects on human health. *Trends Food Sci Tech*, 2003, **14**: 210–228.

6   Miraglia M, Berdal K G, Brera C, Corbisier P, Holst-Jensen A, Kok E J, Marvin H J P, Schimmel H, Rentsch J, van Rie J P P F, Zagon J. Detection and traceability of genetically modified organisms in the food production chain. *Food Chem Toxicol*, 2004, **42**: 1157–1180.

7   Konig A, Cockburn A, Crevel R W R, Debruyne E, Grafstroem R, Hammerling U, Kimber I, Knudsen I, Kuiper H A, Peijnenburg A A C M, Penninks A H, Poulsen M, Schauzu M, Wal J M. Assessment of the safety of foods derived from genetically modified (GM) crops. *Food Chem Toxicol*, 2004, **42**: 1047–1088.

8   Rischer H, Oksman-Caldentey K M. Unintended effects in genetically modified crops: Revealed by metabolomics? *Trends Biotechnol*, 2006, **24**: 102–104.

9   Charlton A, Allnutt T, Holmes S, Chisholm J, Bean S, Ellis N, Mullineaux P, Oehlschlager S. NMR profiling of transgenic peas. *Plant Biotechnol J*, 2004, **2**: 27–35.

10   Serkova N J, Spratlin J L, Eckhardt S G. NMR-based metabolomics: Translational application and treatment of cancer. *Curr Opin Mol Ther,* 2007, **9**: 572–585.

11   Emura K, Yamanaka S, Isoda H, Watanabe K N. Estimation for different genotypes of plants based on DNA analysis using near-infrared (NIR) and Fourier-transform infrared (FT-IR) spectroscopy. *Breeding Sci*, 2006, **56**: 399–403.

12   Colquhoun I J, Le Gall G, Elliott K A, Mellon F A, Michael A J. Shall I compare thee to a GM potato? *Trends Genet,* 2006, **22**: 525–528.

13   Wang K H, Cho Y G, Yoon U H, Eun M Y. A rapid DNA extraction methods for RFLP and PCR analysis from a single dry seed. *Plant Mol Biol Rep*, 1998, **16**: 1–9.

14   Gharayazie B, Alinia F, Menguito C A, Rubia L G, de Palma J M, Liwanag E A, Cohen M B, Khush G S, Bennett J. Enhance resistance to two stem borers in an aromatic rice containing a synthetic *cry1A(b)* gene. *Mol Breeding*, 1997, **3**: 401–414.

15   Park C H, Park H. A comparison of generalized linear discriminant analysis algorithms. *Pattern Recogn*, 2008, **41**: 1083–1097.

16   Shih F Y, Zhang K. A distance-based separator representation for pattern classification. *Image Vis Comput*, 2008, **26**: 667–672.

17   Rezzi S, Axelson D E, Héberger K, Reniero F, Mariani C, Guillou C. Classification of olive oils using high throughput flow ^{1}H NMR fingerprinting with principal component analysis, linear discriminant analysis and probabilistic neural networks. *Anal Chim Acta*, 2005, **552**: 13–24.

18   Fontes E M G. Legal and regulatory concerns about transgenic plants in Brazil. *J Invertebr Pathol*, 2003, **83**: 100–103.