Application of Ultra-high Performance Liquid Chromatography-Mass Spectrometry (UHPLC-MS) Metabolomic Fingerprinting to Characterise GM and Conventional Maize Varieties

Lukáš VÁCLAVÍK¹, Jaroslava OVESNÁ², Ladislav KUČERA², Jan HODEK², Kateřina DEMNEROVÁ³ and Jana HAJŠLOVÁ¹

¹Department of Food Analysis and Nutrition, Faculty of Food and Biochemical Technology, Institute of Chemical Technology Prague, Prague, Czech Republic, ²Department of Molecular Biology, Division of Plant Genetics, Breeding and Product Quality, Crop Research Institute, Prague, Czech Republic, ³Department of Biochemistry and Microbiology, Faculty of Food and Biochemical Technology, Institute of Chemical Technology Prague, Prague, Czech Republic

Abstract


The feasibility of metabolomic fingerprinting approach based on ultra-high performance liquid chromatography-quadrupole-time-of-flight mass spectrometry (UHPLC-QTOFMS) was studied to assess its ability to discriminate between maize varieties, and to show the associations between them on the metabolomic level. The non-targeted metabolomic analysis was applied to assess the variability within two varieties grown under different environmental conditions and to characterise the association within a sample set comprising both conventional and transgenic (MON-ØØ81Ø-6) maize varieties cultivated under the same environmental conditions (locality). Typical metabolomic fingerprints were established for individual plants. The plants representing two varieties formed well separated clusters. Metabolomic fingerprints of the second sample set enabled their unambiguous discrimination. The differences in metabolomic fingerprints between maize varieties were identified and documented by grouping in PCA and/or CA. The results indicate a similar genetic basis of transgenic maize varieties as they descend from a MON 810 event. The results explicitly showed that the variability of the metabolites in MON 810 did not exceed the ranges measured within the conventional varieties, thus supporting the concept of substantial equivalence.

Keywords: chemometric analysis; maize; metabolomics

The use of genetically modified (GM) crops in agriculture and agro-industry may offer a number of advantages over the conventional counterparts, such as the resistance to various diseases, pests, herbicides, or increased yields and nutritive value (Koziel et al. 1993; Gao et al. 2000; Barros et al. 2010). Among other GM crops, the commercial transgenic maize (Zea mays L.), containing various forms of the Cry gene (Cry1Ab, Cry1Ac, Cry1af, Cry2Ab2, Cry34ab1, Cry3Bb1, Cry9c or Cry1Fa) from the Bacillus thuringiensis (Bt), is grown on the largest scale (Clive 2012). The expression of the above genes allows the...
Bt maize to produce proteins that are highly toxic to some Coleopteran and/or Lepidopteran and to protect itself from major insect pests frequent in European and North American agriculture (e.g. European corn borer) (Clive 2012).

The commercial growth and use of GM crops for food and other purposes is regulated in many countries and comprehensive safety assessment procedures are required prior to their approval (Marmiroli et al. 2008). The safety assessment of the GM crops is based on the concept of “substantial equivalence” and involves comparative analyses of targeted chemical constituents representing nutrients, toxins, and antinutrients in a GM crop and the appropriate conventional counterpart (OECD 1993; FAO/WHO 2000). Relatively recently, non-targeted metabolomics-based approaches were applied to extend the coverage of analytes in substantial equivalence studies and to identify the compounds differing in abundance in GM and non-GM plants (Barros et al. 2010; Heinemann et al. 2011; Manneti et al. 2006; Leon et al. 2009; Frank 2012).

Although a large amount of data has been evaluated to assess the safety of the GM plant and products, metabolomics-based studies aiming at comprehensive analysis of as many low molecular compounds (metabolites) in a biological sample as possible, have not been extensively used up to now. This challenging task can be only completed with the use of advanced analytical platforms facilitating the analysis of a wide range of metabolites differing both by physico-chemical properties and abundance, such as nuclear magnetic resonance (NMR) and/or mass spectrometry (MS) coupled with the separation techniques employing liquid chromatography (LC), gas chromatography (GC), and capillary electrophoresis (CE) (Dunn & Ellis 2005; Wishart 2008; Cevallos-Cevallos et al. 2009).

As the data generated by these techniques are very complex, advanced processing software tools and multivariate chemometric methods are needed for their interpretation (Katajamaa & Dresic 2007).

In this feasibility study, the potential of non-targeted metabolomics workflow using ultra-high performance liquid chromatography-quadrupole-time-of-flight mass spectrometry (UHPLC-QTOFMS) instrumentation and chemometrics has been explored for the evaluation of the conventional and GM maize (MON 810 event). The samples analysed comprised: (i) GM and non-GM unrelated varieties coming from different environmental conditions and (ii) various commercial Bt maize and non-GM maize varieties grown under identical field conditions in a field trial. Here we report the possibility of using UHPLC-MS-based metabolomic fingerprinting for comprehensive evaluation of the metabolome of maize leaves.

**MATERIAL AND METHODS**

**Plant material.** The first set of leaf samples (approximately 5 g) used in this study was kindly provided by a private grower. The second set was sourced from Czech Institution for Testing in Agriculture (CISTA) in field trials. The plant leaves were collected in phenological phase on BCHCH-scale: Leaf development, code 19 (Lancashire et al. 1991). The first sample set consisted of transgenic Bt variety Kvalitas YG (event MON 810) and the non-GM unrelated maize, both grown under different environmental conditions. For each of the above mentioned groups (i.e., transgenic and conventional maize), 10 samples representing leaves were taken from individual plants. The second sample set represented both GM Bt maize varieties (n = 10) and non-GM maize varieties (n = 10), grown in the field trial as comparators under identical conditions in the southern part of the Czech Republic. Like in the case of the first sample set, leaves were collected from individual plants of each variety (Table 1). Until further processing, the leaf tissues were stored in the dark at −70°C.

**Verification of GM and non GM nature of samples.** Briefly, the collected leaf tissue samples were frozen and homogenised in liquid nitrogen. DNAs were isolated from each leaf sample in duplicates by CTAB based method (Ovesná et al. 2008). DNAs quality and quantities were checked by electrophoretic separation of sample aliquots on 0.7% agarose gel and spectrophotometry (NanoPhotometer; Implen, Los Angeles, USA) at 260 and 280 nm. The DNA (100 ng) was further used for real-time PCR-based event specific (MON-ØØ81Ø-6) assay (ABI 7900 HT) to verify the presence of MON 810 event. Specific primer sets were used to detect the presence of 3SS CaMV promoter and NOS terminator (Ovesná et al. 2006; Querci et al. 2010) and to exclude other GM modifications using PCR (MJ Research, Watertown, USA).

**Metabolite extraction procedure.** The leaves were homogenised in liquid nitrogen. The obtained powder (1 g) was extracted by shaking with 10 ml of methanol for 5 minutes. The mixture was cen-
trifuged for 5 min (10 000 rpm, 20°C), the supernatant was passed through a 0.22 µm PTFE filter and transferred into a 2 ml amber autosampler vial. Prior to UHPLC-MS analysis, the samples were diluted 10-fold with pure methanol to avoid the possible saturation of the MS detector. The sample extracts were stored at –22°C.

**Ultra high performance liquid chromatography-mass spectrometry (UHPLC-MS).** An Acquity Ultra Performance LC system coupled with Synapt G2 HD quadrupole-time-of-flight mass spectrometer (Waters, Milford, USA) was used in this study. Chromatographic separation was performed with an Acquity UPLC HSS T3 reversed phase analytical column (50 mm × 2.1 mm i.d., 1.7 µm particle size; Waters, Milford, USA) maintained at 40°C. The sample volume of 3 µl with the partial loop injection mode was used. The mobile phase consisted of 0.1% aqueous formic acid (A) and 0.1% formic acid in methanol (B). The gradient elution was carried out as follows: 0–7 min eluent B 5 – 100%; 7–9 min eluent B 100%; 9–10 min column equilibration – eluent B 5%. The mobile phase flow was 0.5 ml/minute.

The Synapt G2 HD instrument was operated in positive electrospray ionisation (ESI) mode; the parameter settings used during the measurements were as follows: capillary voltage (+2500 V), cone voltage (+30 V), source temperature (120°C), desolvation temperature (350°C). Nitrogen was used as both desolvation and cone gas at a flow rate of 800 and 10 l/h, respectively. Leucine-enkephaline was used as lock mass (m/z 556.2771) for internal calibration at a concentration of 2 ng/µl and a flow rate of 10 µl/minute. Both full MS and the MS/MS fragmentation mass spectra were acquired at a rate 2 spectra per second in the range m/z 50–1000. The mass resolving power of the instrument was approximately 20 000 FWHM (full width at half maximum) at m/z 200.

In order to diminish any possible time dependent changes in UHPLC-MS chromatographic fingerprints, the in-batch sequence of the samples was randomised. Two repeated analyses of each sample were performed. Within the measurement sequence, blank samples (pure methanol) were analysed.

---

### Table 1. List of varieties used in the study, breeding company and basic characteristic of plant material

<table>
<thead>
<tr>
<th>Variety</th>
<th>GM</th>
<th>Company</th>
<th>FAO number</th>
<th>Registration</th>
<th>Hybrid/Line</th>
<th>Early vigor (9–1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRAZI YG</td>
<td>MON 810</td>
<td>KWS SAAT AG</td>
<td>320</td>
<td>2009</td>
<td>Sc</td>
<td>8.3</td>
</tr>
<tr>
<td>DKC 2961 YG</td>
<td>MON 810</td>
<td>Monsanto Technology LLC</td>
<td>230</td>
<td>2010</td>
<td>Sc</td>
<td>5.7</td>
</tr>
<tr>
<td>DKC 3512 YG</td>
<td>MON 810</td>
<td>Monsanto Technology LLC</td>
<td>290</td>
<td>2008</td>
<td>Sc</td>
<td>7</td>
</tr>
<tr>
<td>DKC 3795</td>
<td>MON 810</td>
<td>Monsanto Technology LLC</td>
<td>260</td>
<td>2010</td>
<td>Sc</td>
<td>8.3</td>
</tr>
<tr>
<td>DKC 3872 YG</td>
<td>MON 810</td>
<td>Monsanto Technology LLC</td>
<td>270</td>
<td>2011</td>
<td>Sc</td>
<td>6.3</td>
</tr>
<tr>
<td>ES IMPERIAL YG</td>
<td>MON 810</td>
<td>Euratis Semences</td>
<td>300</td>
<td>2011</td>
<td>Sc</td>
<td>9</td>
</tr>
<tr>
<td>JOKARI CS</td>
<td></td>
<td>Caussade Semences</td>
<td>320</td>
<td>2007</td>
<td>Sc</td>
<td>8.3</td>
</tr>
<tr>
<td>JUXXIN</td>
<td></td>
<td>Société RAGT 2n</td>
<td>330</td>
<td>2008</td>
<td>Sc</td>
<td>9</td>
</tr>
<tr>
<td>KARAS YG</td>
<td>MON 810</td>
<td>KWS SAAT AG</td>
<td>400</td>
<td>2008</td>
<td>Sc</td>
<td>9</td>
</tr>
<tr>
<td>KRABAS YG</td>
<td>MON 810</td>
<td>KWS SAAT AG</td>
<td>300</td>
<td>2011</td>
<td>Sc</td>
<td>8.3</td>
</tr>
<tr>
<td>LG 30290</td>
<td></td>
<td>LIMAGRAIN GENETICS Grandes Cultures S.A.</td>
<td>280</td>
<td>2011</td>
<td>Sc</td>
<td>8.3</td>
</tr>
<tr>
<td>MERCURIO</td>
<td></td>
<td>KWS SAAT AG</td>
<td>250</td>
<td>2010</td>
<td>Tc</td>
<td>8.3</td>
</tr>
<tr>
<td>OXYGEN YG</td>
<td>MON 810</td>
<td>Société RAGT 2n</td>
<td>330</td>
<td>2011</td>
<td>Sc</td>
<td>7</td>
</tr>
<tr>
<td>P9494</td>
<td></td>
<td>Pioneer Hi-Bred International, Inc.</td>
<td>340</td>
<td>2011</td>
<td>Sc</td>
<td>7.7</td>
</tr>
<tr>
<td>PR38V12</td>
<td></td>
<td>Pioneer Hi-Bred International, Inc.</td>
<td>330</td>
<td>2007</td>
<td>Sc</td>
<td>7</td>
</tr>
<tr>
<td>PR39T47</td>
<td>MON 810</td>
<td>Pioneer Hi-Bred International, Inc.</td>
<td>270</td>
<td>2008</td>
<td>Sc</td>
<td>8.3</td>
</tr>
<tr>
<td>RH07103</td>
<td></td>
<td>Société RAGT 2n</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
<td>nd</td>
</tr>
<tr>
<td>RH11056</td>
<td>MON 810</td>
<td>Société RAGT 2n</td>
<td></td>
<td>Sc</td>
<td></td>
<td>7.7</td>
</tr>
<tr>
<td>SUBIANCA</td>
<td></td>
<td>Dow AgroSciences GmbH</td>
<td>340</td>
<td>2010</td>
<td>Sc</td>
<td>7</td>
</tr>
<tr>
<td>TExXUD</td>
<td></td>
<td>Société RAGT 2n</td>
<td>340</td>
<td>2005</td>
<td>Sc</td>
<td>5</td>
</tr>
</tbody>
</table>

nd – not detected
Data processing and statistical analysis. With regard to the complexity of UHPLC-MS records obtained by the analysis of the sample sets studied, an automated algorithm was employed for the data mining and aligning procedures. The software automatically organised the data into matrix of \((x \times y)\) containing \(y\) peaks (characterised by RT, \(m/z\) value, and intensity) in each of the \(x\) records. To ensure that the data comprise only the peaks originating from the maize extract, the peaks observed in the blank samples records were removed.

The processing of the LC-MS data was carried out with the use of MarkerLynx software, Version 4.1 (Waters, Milford, USA). Automated data mining was performed in the retention time (RT) window of 0.2–10.0 min and \(m/z\) range 50–1000 at an intensity threshold setting of 500 cps. To reduce the dimensionality of the data matrix obtained, the peaks representing isotopic ions were automatically excluded during the data mining based on isotope spacing. The detected peaks were aligned across the sample set using a mass tolerance window 0.01 Da and a retention time window 0.1 minutes. Data matrix comprising the list of peaks characterised by RT, \(m/z\), and intensity was obtained in this way. The data set No.1 was Pareto-scaled (square root of the standard deviation is used as the scaling factor) prior to the principal component analysis (PCA) facilitated by the Extended Statistics module. In the case of the data set No. 2, the peaks with zero intensity in at least one sample were excluded from the data matrix. The data matrices of \((20 \times 2613)\) containing 2613 peaks (characterised by RT, \(m/z\) value, and intensity) in each of the 20 records were obtained for each data set. The data set No. 1 was processed with PCA, while for processing the data set No. 2 both cluster analysis (CA) and PCA were used. In the case of the sample set No. 2, PCA and CA were conducted with the use of the software package STATISTICA, Version 10 (StatSoft, Tulsa, USA). Three dimensional PCA based on correlation matrix of \((20 \times 20)\) resulted from the factor analysis of UHPLC-MS data obtained by the analysis of the data set No. 2 (20 maize varieties, 2613 peaks).

RESULTS AND DISCUSSION

DNA analysis confirmed the expected nature of the plant material. The conventional varieties were free of contamination with GM maize and neither of the analysed varieties contained other GM event. Hence, the material for the preliminary investigation of the possible associations between the inserted GM event and the metabolomic fingerprint was available. If there is some general effect of the transgene MON 810 other than the pest resistance, a specific metabolic pattern has to be identified across the two sets.

LC-MS metabolomic fingerprinting

In this study, non-targeted analysis of maize methanolic leaf extracts was performed to obtain comprehensive metabolomic fingerprints containing information on as many metabolites as possible. Considering the non-targeted nature of the analyses to be performed, some generic settings had to be applied to both LC separation and MS detection. Since various metabolites differ significantly in terms of polarity, an analytical column with reversed stationary phase containing embedded polar groups was used instead of conventional \(C_{18}\) or \(C_{8}\) reversed stationary phases to enhance the retention of highly polar components present in the maize leaf metabolome. The use of stationary phase with sub-2 \(\mu m\) particles significantly increased the chromatographic resolution and enabled relatively rapid analysis with injection-to-injection run time of 11 minutes. Since the typical base width of the detected peaks was below 12 s, sufficient characterisation of the analytes peak shapes was possible at acquisition speed settings used.

The reproducibility of both \(m/z\) values and retention times (RT) of the detected metabolites has a major impact on the overall quality of UHPLC-MS metabolomic data. High fluctuations of these parameters during the measurement sequence result in a poor outcome of the data mining process and lead to misinterpretation of the results. To explore the possible RT fluctuation, RTs of three peaks (RT 0.7, 4.2, and 7.1 min) were monitored during the measurement sequence of 20 sample injections. Based on these data, typical peak RT variability was found to be below 5% (relative standard deviation, RSD) that can be considered as an acceptable value. QTOF mass analyser can be considered an effective tool for metabolomics-based studies of biological samples (Xie et al. 2008). It also enables accurate mass measurement in a broad \(m/z\) range. To obtain as low mass errors as possible,
regular external calibration of the instrument was performed. Additionally, a continuous correction of small mass drifts correction was carried out with the use of mass locking compound delivered to the instrument via secondary ESI sprayer. Under these conditions, the achieved mass accuracy was below 5 ppm. Despite that in the initial optimisation experiments both positive and negative ESI modes were used, the preliminary results have explicitly shown that the latter polarity setting yielded significantly fewer informative records. Therefore, only positive ESI was further used.

Data mining and data pre-treatment

It should be noted that the peaks extracted from the UHPLC-MS fingerprints could not be generally considered to be the individual metabolites because of the signal redundancy. This redundancy was, to some extent, decreased in the data mining step by automatic removal of the peaks of monoisotopic ions (i.e., signals with the lowest m/z value within the isotope pattern). It is well known, however, that in ESI, ionisation of a single metabolite may result in the formation of multiple signals, such as various adduct or fragment ions. For example, in positive ESI ionisation ammonium [M + NH₄]⁺, sodium [M + Na]⁺, or potassium adducts [M + K]⁺, are frequently formed.

Since metabolites typically occur in distinct concentrations, large differences in intensities (by up to several orders of magnitude) were observed with many peaks. Such differences can significantly affect the data variance determined treatment by the multivariate analysis. To modify the weights of the respective peaks (variables), the Pareto scaling of the data was performed (Ivosev et al. 2008). This type of scaling enabled the reduction but not complete elimination of the above abundance differences. It provides good results when applied to LC-MS data because it takes into account that larger peaks can be more reliable, but all variables are equivalent. The scaling of the data was not performed for the set No. 2, the log scaling when applied to LC-MS data (intensity) yielded less informative results. The detection of correlations between different features in a set of feature vectors is a very important data mining task because the correlation indicates a dependency between the features or some association of the cause and effect between them (Böhm et al. 2004). Our simplified approach was based on a combination correlation matrix (20 × 20) resulted from the factor analysis of LC-MS data (20 maize varieties, 2613 peaks, peak intensity) and PCA. Therefore, this multivariate analysis is not affected by large differences in the peaks intensities.

Chemometric analysis

In the next step, principal component analysis (PCA) was applied to the pre-processed experimental data. This unsupervised pattern recognition technique represents a highly useful and widely employed tool for the interpretation of complex data. PCA allows the dimensionality reduction and visualisation of the intrinsic patterns present in the original data in the form of a few principal components (PCs) while retaining the maximum possible variability (Beruetta et al. 2007). In this study, PCA was used to explore the differences in LC-MS metabolomic fingerprints of leaves originating from various varieties of both transgenic and conventional maize. The sample leaves extracts were characterised by LC-MS metabolomic fingerprints that were processed as described above. As can be seen in Figure 1, two relatively well-resolved sample clusters of the respective maize varieties were formed in the PCA space defined by the first and second principal components calculated as based on sample set No. 1. The separation of these samples clearly documents the differences in their metabolomic fingerprints. These differences are apparently linked to changes in the metabolites concentrations. The differences between the samples of the same variety are in-

![Figure 1. Two dimensional PCA scores plot showing association between metabolomic fingerprints of two maize varieties represented by 10 samples per cultivar (each cultivar forms a specific group)](image)
duced by varying micro-environmental and soil conditions. As shown in the two-dimensional PCA graph, the metabolomic profiles of ten different leaf samples representing identical varieties were closer to one another than to the set of profiles derived from the other varieties. The two sets did not overlap, which proved the potential of the technique to differentiate between cultivars, thus we proceeded to analyse the data obtained by analysis of the sample set No. 2.

As can be seen in the three-dimensional scores plot (Figure 2), the PCA grouped the examined varieties according to their origins, i.e. according to the respective breeding company by which the

Figure 3. Results of cluster analysis for LC-MS data (peak intensities) of 10 transgenic (marked with an asterisk) and 10 conventional maize varieties (20 maize varieties, 2613 peaks).
respective variety was released. This finding was also explicitly documented by the CA (Figure 3). The first, second, and third principal components explained for 55.1, 19.4, and 22.3% of the data variance, respectively. The MON 810 varieties were in the PCA scores plot closer to one another compared to conventional varieties, which indicated on their similar genetic basis. This finding was expected as all GM MON 810 descended from a single modified line. Also, the company may use different types of germplasm in their breeding program. Although the GM varieties seem to be closer to one another, they do not exceed the variability range that was calculated for the conventional varieties. The range of values representing maize leaves metabolome is documented by the distance of Jokari CS variety from the other GM and conventional varieties (Figure 3).

As two sets of varieties were used in this study, one representing two varieties GM and conventional, each represented by multiple samples, and the second several GM and conventional varieties, we were able to demonstrate the capability of metabolomics-based approach to differentiate efficiently between the maize varieties, and to trace their relationships on the metabolomic level. No significant correlation was found between the metabolic profiles measured at the given developmental stage and agronomically important traits, namely earliness, yield or use of the varieties (silage, grains).

We were not able to find any specific patterns/signal typical for GM varieties even though these were cultivated under the same environmental conditions as their conventional counterparts. Moreover, as the ranges of the metabolite intensities detected in individual GM and conventional varieties overlapped, it can be expected that in the case of equivalence the metabolite concentrations in GM varieties will not differ from those in the conventional cultivars and will not be shifted in PCA plot with respect to them (EFSA 2011). Our findings support the opinion of equivalence of MON 810 with conventional maize. However, studies under different environmental conditions with a higher number of replicates are necessary.

In general, the composition of the leaf tissue may have an impact on the safety of GM plants. Much more data for the composition analysis of maize kernels are available (Kuiper et al. 2001; Cellini et al. 2004). The presented approach allows large-scale evaluation of leaf metabolome and can provide extensive data for GMO safety assessment.

CONCLUSIONS

UHPLC-QTOFMS technique was demonstrated to hold a great potential in discriminative metabolomics of conventional and Bt transgenic maize varieties. The processing of metabolomic fingerprints (mass spectra) by advanced chemometric methods demonstrated the capability of this approach to classify various maize varieties in both GM and non-GM sample sets. On the other hand, distinguishing MON 810 event was not possible. This means, in fact, that 'substantial equivalence' principle has been confirmed.

References


Vol. 31, 2013, No. 4: 368–375 Czech J. Food Sci.


Received for publication April 10, 2013
Accepted after corrections May 3, 2013

Corresponding author:
Prof. Ing. Jana Hašlová, CSc., Vysoká škola chemicko technologická v Praze, Fakulta potravinářské a biochemické technologie, Ústav analýzy potravin a výživy, Technická 5, 166 28 Praha 6, Česká republika; E-mail: jana.haslova@vscht.cz