Botanical discrimination and classification of honey samples applying gas chromatography/mass spectrometry fingerprinting of headspace volatile compounds

Konstantinos A. Aliferis a,⁎, Petros A. Tarantilis b, Paschalis C. Harizanis c, Eleftherios Alissandrakis c

aDepartment of Plant Science, McGill University, 2111 Lakeshore Rd., Sainte-Anne-de-Bellevue, Quebec, Canada H9X 3V9
bLaboratory of Chemistry, Department of Science, Agricultural University of Athens, 75 Iera Odos St., 118 55 Athens, Greece
cLaboratory of Sericulture and Apiculture, Department of Crop Science, Agricultural University of Athens, 75 Iera Odos St., 118 55 Athens, Greece

A R T I C L E   I N F O
Article history:
Received 14 September 2009
Received in revised form 3 November 2009
Accepted 31 December 2009

Keywords:
Biomarkers
Botanical discrimination
Botanical origin
Chemometrics
Headspace GC/MS
Honey
Solid-phase microextraction (SPME)
Volatiles

A B S T R A C T
A validated method for the discrimination and classification of honey samples performing GC/MS fingerprinting of headspace volatile compounds was developed. Combined mass spectra of honey samples originated from different plants and geographical regions of Greece were subjected to orthogonal partial least squares-discriminant analysis™ (OPLS™-DA), soft independent modelling of class analogy (SIMCA), and OPLS™-hierarchical cluster analysis (OPLS™-HCA). Analyses revealed an excellent separation between honey samples according to their botanical origin with the percentage of misclassification to be as low as 1.3% applying OPLS™-HCA. Fragments (m/z) responsible for the observed separation were assigned to phenolic, terpenoid, and aliphatic compounds present in the headspace of unifloral honeys. On the other hand, a variable classification for citrus and thyme honeys according to their geographical origin could be achieved. Results suggested that the developed methodology is robust and reliable for the botanical classification of honey samples, and the study of differences in their chemical composition.

1. Introduction
Honey is a very important product because of its nutritional and therapeutic values (Molan, Mizrahi, & Lensky, 1997; Nasuti, Gabbianelli, Falcioni, & Cantalamessa, 2006). Because of the high price of certain honey types, adulteration with low cost and nutritional value substances (Arvanitoyannis et al., 2005; Tzouros & Arvanitoyannis, 2001) is possible. GC/MS analyses enabled the authentication and classification of food products (Arvanitoyannis et al., 2005; Tzouros & Arvanitoyannis, 2001). GC/MS fingerprinting has been proved to be a powerful platform for the discrimination between honey samples (Baroni et al., 2006; Bertelli, Plessi, Sabatini, & Restani, 2008), infrared spectroscopy (IR) (Ruoff et al., 2006; Woodcock, Downey, Kelly, & O'Donnell, 2007), front face fluorescence spectroscopy (Karoui, Dufour, Bosset, & De Baerdemaeker, 2007), atomic emission spectroscopy (AES), and inductively coupled plasma atomic emission spectrometry (ICP-AES) (Nozal Nalda, Bernal Yaguee, Diego Calva, & Martin Gomez, 2005) have been used for the chemical analyses of honeys and their classification. Also, microscopic (Dimou, Katsaros, Tzavela-Klonari, & Thrasyvoulou, 2006) and physiochemical characteristics (Corbella & Cozzolino, 2006; Serrano, Villarrejo, Espejo, & Jodral, 2004) have been employed for the botanical and geographical determination of honey samples.

Regarding the isolation of honey volatile compounds, several techniques have been applied, including the Likens–Nickerson methodology, dynamic headspace, purge-and-trap systems, ultrasound-assisted extraction and more recently the solid-phase microextraction (Anklam, 1998; Cuevas-Glorya, Pino, Santiago, & Sauri-Duch, 2007). The recent developments in chemometrics analyses have enabled the authentication and classification of food products (Arvanitoyannis et al., 2005; Tzouros & Arvanitoyannis, 2001). GC/MS fingerprinting has been proved to be a powerful platform for the discrimination between honey samples (Baroni et al., 2006;
2. Materials and methods

2.1. Chemicals and reagents

Benzenophene was purchased from Fluka Chemika (Buchs, Switzerland) and methanol from Merck (Darmstadt, Germany).

2.2. Botanical and geographical origins of honey samples

A total of 77 unifloral honey samples of different botanical origins namely, chestnut (*Castanea sativa*, six samples), cotton (*Gossypium hirsutum*, six samples), fir (*Abies* spp., six samples), heather (*Erica manipuliflora*, six samples), pine (*Pinus* spp., six samples), thyme (*Corydathymus capitatus*, 20 samples) and citrus (*Citrus* spp., 27 samples), were analysed (Table 1).

Citrus honey samples had been collected from several provinces of north (Arta) and south Greece (Argolida, Crete, Lakonia) and Leros, and Crete. Samples were kept at −20 °C in glass vials. Prior to analyses, samples were left to thaw for 24 h at room temperature.

2.3. Isolation and gas chromatography–mass spectrometry analysis of headspace volatile honey compounds

The isolation of the aroma compounds was performed using the SPME procedure. An SPME holder (Supelco, Bellefonte, PA, USA) and the divinylbenzene/carbonbox/polydimethylsiloxane fibre (Supelco, Bellefonte, PA, USA) were used to extract headspace volatiles from honey. This type of fibre provides an excellent sorption capacity and the broadest range of extracted volatiles from the headspace of honey (Čajka, Hajšlová, Cochran, Holadová, & Klimánková, 2007). The honey samples (6 mL of deionised water solution of 3 g honey mL⁻¹) were placed in 15 mL screw-top vials with polytetrafluoroethylene/silicone septa. Benzenophene was used as the internal standard and 20 μL of a solution of 10 μg mL⁻¹ in methanol were added in the samples prior to extraction. The vials were maintained in a water bath at 60 °C under constant stirring during the whole procedure. Equilibration time was set at 30 min, followed by 60 min headspace sampling time. The extraction conditions had been previously optimised based on data of previous study (Alissandrakis, Tarantilis, Harizanis, & Polissiou, 2007b).

The analysis of the isolated compounds was performed using a Hewlett Packard 5890 II GC, equipped with a Hewlett Packard 5972 MS detector. The column used was an HP-5MS (Crosslinked 5% PH Me Siloxane) capillary column (30 m × 0.25 mm i.d., 0.25 μm film thickness) and the gas carrier was Helium at 1 ml min⁻¹ rate. The injector and MS-transfer line temperatures were maintained at 220 °C and 290 °C, respectively. Oven temperature was held at 40 °C for 3 min, raised to 160 °C at 3 °C min⁻¹ and then to 200 °C at 10 °C min⁻¹. Electron impact mass spectra were recorded at a mass range of 40–500 Da. An electron ionisation system was used with ionisation energy of 70 eV.

2.4. Mass spectral data processing

Acquisition of total ion chromatograms, collection of combined MS spectra, and automated peak deconvolution were performed using the Hewlett Packard ChemStation G1701AA, Version A.03.00. The assignment of m/z fragments to the corresponding headspace volatile compounds was based on fragmentation pattern data of our previous works (Alissandrakis et al., 2007a, 2007b). Additionally, mass spectra searches were performed using the libraries of the National Institute of Standards and Technology (NIST) and on-line databases ([The Golm Metabolome Database](http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/msri/gmd_sspq.html) and the Spectral Database for Organic Compounds SDBS ([http://riodb01.ibase.aist.go.jp/sdbs/cgi-bin/direct_frame_top.cgi](http://riodb01.ibase.aist.go.jp/sdbs/cgi-bin/direct_frame_top.cgi)).

Data were expressed as ratios of the abundance of each m/z fragment against the abundance of the fragment m/z = 182, which is the molecular ion of the internal standard (benzophenone) that was used. Extraction of the ion chromatogram for this fragment (m/z = 182) revealed that benzenophene does not naturally occur in any of the isolated compounds. Data were then exported to MS Excel® and fragments not related to honey volatiles, such as m/z = 44 which corresponds to CO₂, were excluded from further analyses. Additionally, major fragments corresponding to naphthalene (m/z = 128) and p-dichlorobenzene (m/z = 146 and 148), compounds that were formerly used in the common apiculture practise (Bogdanov et al., 2004; Harizanis, Alissandrakis, Tarantilis, & Polissiou, 2008; Tzanaki, Thrasyvoulou, Karazafiris, & Zotou, 2006), were excluded. Finally, after the removal of the fragments non-related to honey headspace volatiles, the data matrix was composed of 77 columns (honey samples) and 154 variables (m/z fragments). Prior to pattern recognition analysis, data were further normalised by dividing the relative abundances by their sum.

2.5. Multivariate data analyses and visualisation

For chemometrics analyses the pre-processed data were imported into the SIMCA-P + 12.0 software (Umetrics, MKS Instruments Inc., Sweden). A free demo version of the software is available ([http://www.umetrics.com/default.asp?pagename=downloads_software/c/1](http://www.umetrics.com/default.asp?pagename=downloads_software/c/1)). Data were Pareto scaled (1/sqrt(SD)) which is a compromise between unit variance scaling (UV) and no scaling (Eriksson, Johansson, Kettanen-Wold, & Wold, 2001). PCA, OPLS™-DA, and SIMCA were performed for the whole dataset in order to discriminate and classify honey samples according to their botanical origin and detect corresponding biomarkers. The classification of honey samples according to their geographical origin was a secondary task of the research. Performing PCA in a set of

---

Table 1

<table>
<thead>
<tr>
<th>Botanical origin</th>
<th>Number of samples</th>
<th>Time of collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chestnut</td>
<td>6</td>
<td>June</td>
</tr>
<tr>
<td>Citrus</td>
<td>27</td>
<td>April</td>
</tr>
<tr>
<td>Cotton</td>
<td>6</td>
<td>August</td>
</tr>
<tr>
<td>Fir</td>
<td>6</td>
<td>June</td>
</tr>
<tr>
<td>Heather</td>
<td>6</td>
<td>November</td>
</tr>
<tr>
<td>Pine</td>
<td>6</td>
<td>October</td>
</tr>
<tr>
<td>Thyme</td>
<td>20</td>
<td>July</td>
</tr>
</tbody>
</table>

---

Senyuva et al., 2009). Therefore, in the present study a new methodology for the botanical discrimination and classification of honeys was developed, with their classification based on the geographical origin to be a secondary task. Headspace solid-phase microextraction (HS-SPME) was performed for the extraction of volatile compounds followed by GC/MS fingerprinting. Chemometrics analyses were performed for the combined MS spectra. For the classification and discrimination between the different types of honey and the detection of the corresponding biomarkers, principal components analysis (PCA), orthogonal partial least squares-discriminant analysis™ (OPLS™-DA), soft independent modelling of class analogy (SIMCA), and OPLS™-hierarchical cluster analysis (OPLS™-HCA) were carried out. Results of the analyses highlight the potential of the use of combined mass spectra for the discrimination and classification of honey samples performing GC/MS fingerprinting of headspace volatile compounds.
observations representing one or more groups, knowledge related to
group membership is not used to find the location of the principal
components. Thus, for the detection of differences between the
honey samples, OPLS™-DA was chosen because it is a modification
of the PLS-DA that has been designed to handle the variation in X
that is orthogonal to Y. The partitioning of the X-data applying
OPLS™-DA, provides improved model transparency and interpret-
ability, without modifying its predictive power (Trygg & Wold,
2002). The detection of the influential fragments \((m/z)\) was based
on scaled and centred PLS regression coefficients (CoeffCS) with
95% jack-knifed confidence intervals (Efron & Gong, 1983).

The multivariate analysis was based on the methodology of Aliferis
and Chrysayi-Tokousbalides (2006). Additionally, OPLS™-HCA
was applied for selected datasets in order to further evaluate the
ability of the developed models to classify honey samples accord-
ing to their botanical and geographical origins using the Ward’s

2.6. Performance statistics

For the evaluation of the developed models, the cumulative
fraction of the total variation of the \(X\)'s that could be predicted
by the extracted components [predictive ability, \(Q^2_{(cum)}\)] and the
fraction of the sum of squares of all \(X\)’s \((R^2_X)\) and \(Y\)’s \((R^2_Y)\) ex-
plained by the current component (explained variation) were used.

3. Results and discussion

3.1. Chemometrics analysis for the discrimination and classification of
honey samples according to their botanical origin and the detection of
biomarkers

3.1.1. Principal components analysis (PCA)

The principal task of the present study was the botanical dis-
crimination between honey samples, their classification, and the
detection of corresponding biomarkers. For the initial overview
of the dataset and the detection of trends and outliers, PCA was
carried out. Analysis showed no samples being outside the Hotell-
ing \(T^2\) 95% confidence ellipse that could influence the analyses, and
high values of explained variation \((R^2_X = 0.77)\) and predictive abil-
ity \([Q^2_{(cum)} = 0.67, 4 \text{ PCs}]\).

3.1.2. Orthogonal partial least squares™-discriminant analysis
(OPLS™-DA)

Performing OPLS™-DA a model was developed from the train-
ing sets of observations of known membership class. Cross valida-
tion was performed for the training sets using the default software
options giving high values for the \(R^2_X = 0.82, R^2_Y = 0.79,\) and
\(Q^2_{(cum)} = 0.73\) \((6 \text{ PCs})\). A very satisfactory grouping of honey sam-
ples could be achieved according to their botanical origin (Fig. 1a).
In order to further examine the predictive ability of the model and detect the corresponding biomarkers for the observed
separation, OPLS™-DA was performed comparing all possible pairs
of groups of honey samples. Analyses revealed a very strong dis-
crimination between the different types of honey compared in
pairs, as it is indicated by the high values of \(Q^2_{(cum)}\) with the excep-
tion of the pair fir/pine for which a moderate separation could be
achieved \([Q^2_{(cum)} = 0.68]\) (Table S1 in Supplementary data).

3.1.3. Orthogonal partial least squares™-hierarchical cluster analysis
(OPLS™-HCA)

In agreement to results of OPLS™-DA, application of OPLS™-
HCA for the whole dataset resulted to a very good classification
of the samples according to their botanical origin (Fig. 1b). Out of
the 77 samples that were simultaneously analysed, only one sam-
ples (heather honey) was incorrectly classified as fir honey (Fig. 1b),
a percentage of misclassification of 1.3%. The percentage of the
right classification of samples based on their botanical origin was
raised from 98.7% to 100% when citrus and thyme groups were ex-
cluded from the dataset (Fig. 2a and b).

![Fig. 1. (a) OPLS™-DA PC1/PC2 score plot for the whole dataset. The ellipse represents the Hotelling \(T^2\) with 95% confidence, and (b) OPLS™-dendrogram illustrating the OPLS™-HCA using the Ward’s method. Honey samples are grouped according to their botanical origin [chestnut (♣), citrus (■), cotton (▲), fir (●), heather (□), pine (●), and thyme (◇)].](image-url)
3.1.4. Soft independent modelling of class analogy (SIMCA)

SIMCA was additionally carried out in order to examine the hypothesis that following the applied protocols, a simultaneous discrimination of citrus and thyme samples could be achieved. For each of those two types of honey, a separate PCA model was computed which was used to infer membership class of all treatments. The results are summarised in the Coomans’ plot (Fig. 3) where the co-ordinates are the distances to the model values (Moderate outliers, DModX) for the prediction set observations down onto the thyme model (X-axis) and the citrus model (Y-axis) (Eriksson et al., 2001). As Fig. 3 shows, there are no samples in the lower left part of the plot which means that there are no honey samples with fragmentation pattern that fit to both citrus and thyme models. Samples in the lower right and the upper left part belong to thyme and citrus models, respectively. All other samples are clustered into the upper right area in which honey samples that fit to none of the citrus and thyme models can be found. The results cross-validated those obtained from OPLS™-DA and OPLS™-HCA, indicating an excellent classification and discrimination for citrus and thyme honey samples following the applied protocols.

3.1.5. Detection of biomarkers for the observed discriminations

Further analyses of the data using scaled and centred PLS regression coefficients with 95% jack-knifed confidence intervals, unravelled the fragments (m/z) responsible for the observed discrimination and classification of honey samples (Table S2 in Supplementary data). For the assignment of the fragments to the corresponding headspace volatiles, results of previous works (Alissandrakis et al., 2007a, 2007b) were combined with MS database searches. Fragments were assigned to phenolic (m/z 43, 65, 91, and 92), terpenoid (m/z 41, 43, 55, 67, 71, and 93), and aliphatic compounds such as acids (m/z 74 and 87), esters (m/z 41, 43, 55, 56, and 57), alcohols (m/z 41, 43, 55, 56, and 57), and aldehydes (m/z 41, 43, 55, 56, 57, 69, 70, 81, 82, 84, 95, 96, and 110). Additionally, m/z fragments influential for the separation of citrus and thyme honeys were assigned to the major compounds previously isolated from the headspace of Greek unifloral citrus (Alissandrakis et al., 2007b) and thyme (Alissandrakis et al., 2007a) honeys (Table S3 in Supplementary data). In Fig. 4 representative chromatograms of citrus and thyme honey samples and corresponding biomarkers for their discrimination are presented. Fragments of the linalool and the 1-p-menthen-9-al isomers, limonene and methyl anthranilate were influential for the classification of citrus honey samples whereas fragments of the phenylacetaldehyde, 1-phenyl-2,3-butanedione, 3-hydroxy-4-phenyl-2-butanone, 3-hydroxy-1-phenyl-2-butanone, and 3-hydroxy-4-phenyl-3-buten-2-one were influential for the classification of thyme honey samples.

3.2. Discrimination and classification of citrus and thyme honeys according to their geographical origin

3.2.1. Geographical discrimination between citrus honey samples

The application of OPLS™-DA for the citrus dataset revealed that a variable discrimination could be achieved between samples collected from different geographical regions. A moderate discrimination could be achieved between citrus honey samples collected from Argolida and Chania [Q²(cum) = 0.74], Argolida and Lakonia [Q²(cum) = 0.64], Arta and Lakonia [Q²(cum) = 0.88], and Chania and Lakonia [Q²(cum) = 0.87]. On the other hand, a poor discrimination was obtained comparing in pairs samples collected from Argolida and Arta [Q²(cum) = 0.21], and Arta and Chania [Q²(cum) = 0.20] (Table S4 in Supplementary data and Fig. 5a). OPLS™-HCA confirmed the unsatisfactory classification of citrus honeys based on their geographical origin, with samples collected from different regions to be clustered in the same clusters. For example, samples...
collected from Argolida, Arta, and Chania are clustered in clusters II, III and IV in Fig. 5b. On the other hand, all samples collected from Lakonia were cluster in cluster I (Fig. 5b).

### 3.2.2. Geographical discrimination between thyme honey samples.

For the thyme honey dataset, the application of OPLS™-DA revealed a satisfactory discrimination between samples collected...
Fig. 5. (a) OPLS™-DA PC1/PC2 score plot for citrus honeys collected in Argolida (■), Arta (●), Chania (▲), and Lakonia (♦). The ellipse represents the Hotelling $T^2$ with 95% confidence, and (b) OPLS™-dendrogram illustrating the OPLS™-HCA using the Ward’s method.

Fig. 6. (a) OPLS™-DA PC1/PC2 score plot for thyme honeys collected in Calymnos (○), Cos (▲), Crete (◇), and Leros (■). The ellipse represents the Hotelling $T^2$ with 95% confidence, and (b) OPLS™-dendrogram illustrating the OPLS™-HCA using the Ward’s method.
from Calimnos and Cos \(Q^2_{\text{cum}} = 0.76\), and Crete and Calimnos \(Q^2_{\text{cum}} = 0.73\). A moderate discrimination was observed between samples collected from Crete and Cos \(Q^2_{\text{cum}} = 0.51\), and Leros and Calimnos \(Q^2_{\text{cum}} = 0.51\), whereas a poor discrimination was observed between honey samples collected from Crete and Leros \(Q^2_{\text{cum}} = 0.21\), and Leros and Cos \(Q^2_{\text{cum}} = 0.05\) (Table S5 in Supplementary data and Fig. 6a). OPLS™-HCA confirmed the results of the OPLS™-DA (Fig. 6b). The only satisfactory classification could be achieved between honey samples collected from broader geographical regions (i.e. between Crete/Leros [clusters I–IV] and Cos/Calimnos [clusters V–VII]).

4. Conclusion

The developed chemometrics model based on combined MS spectra of headspace volatiles of 77 samples of honey from seven of the most common botanical origins provided a strong discrimination applying OPLS™-DA and a percentage of correct classification of samples higher than 98% performing OPLS™-HCA. Additionally, compounds belonging to several chemical groups such as phenolics, terpenoids, and aliphatics (i.e. acids, esters, alcohols, and aldehydes) were detected as biomarkers for the observed discrimination. The results clearly showed that HS-SPME–GC/MS fingerprinting of honey volatiles combined with state-of-the-art chemometrics analyses is a robust, reliable, rapid, and of high potential method for the discrimination and classification of honey samples based on their botanical origin. Therefore, it can be considered as a non-time consuming and cost effective methodology suitable for routine analyses of honeys for their botanical classification.

On the other hand, the developed models showed a variable discriminative ability for citrus and thyme honey samples according to their geographical origin. This is not surprising since the regions from which the samples had been collected share a very similar climate and composition of plant populations.

To our knowledge, the present work is the first report highlighting the use and potential of combined MS spectra in GC/MS fingerprinting studies. The use of combined MS spectra provides a significant advantage to such approach since the identification of the isolated compounds is not a priori required. The results revealed the applicability of this approach in targeted chemometrics studies of volatile compounds of honey and possibly other food products and/or biological samples.

Appendix A. Supplementary data


References


