Application Note

Food

Chemometric methods for botanical classification of Chinese honey based on the volatile compound profile

Solid-phase microextraction and gas chromatography-mass spectrometry

Authors

Hui Chen, Linghe Jin, and Chunlin Fan
Agro-Product Safety Research Center,
Chinese Academy of Inspection and Quarantine,
Beijing, China

Wenwen Wang
Agilent Technologies (China) Company, Ltd.
Beijing, China

Abstract

This study develops a method for the discrimination and prediction of honey samples from various botanical origins. The method was based on the nontargeted volatile profiles obtained by solid-phase microextraction and gas chromatography-mass spectrometry (SPME-GC/MS) combined with chemometrics. The blind analysis of nontargeted volatile profiles was carried out by SPME-GC/MS for 87 authentic honey samples from four botanical origins including acacia, linden, vitex, and rape honey. Quality control of the samples was performed by Principal Component Analysis (PCA). Then, sample class prediction models based on partial least squares discriminant analysis (PLS-DA), naïve Bayes (NB), and back-propagation artificial neural network (BP-ANN) were constructed. The 100 % accuracy results revealed a perfect classification among the different botanical origins. The results indicated that all could be predicted correctly. In addition, the reliability and practicability of the models were validated by an independent set of an additional 20 authentic honey samples. All 20 samples were accurately classified. Finally, the characteristic volatile compounds of linden honey were tentatively identified. It is suggested that the proposed method is reliable and accurate for the classification of honey from various botanical origins, as well as finding marker compounds.

Introduction

Honey is among the most appreciated natural products in the world for its nutritional and medicinal properties. Consumers are more concerned about the botanical and geographical origins of honey. The price of honey is usually based on its botanical/geographical origins. Therefore, it is essential to develop a fast and powerful method to identify honey from different origins. GC/MS has an advantage in the identification and quantification of organic compounds in complex samples. The combination of GC/MS and solid-phase microextraction (SPME) is useful for analyzing volatile compounds in honey[1].

This Application Note describes a recently published study[2] of nontargeted volatile profiles for classification of the botanical origin of Chinese honey by SPME and GC/MS combined with chemometrics. The study develops a procedure to classify and predict the botanical origin of honeys based on nonspecific volatile fingerprint and multivariate analysis. Blind raw data from honey were generated from the SPME-GC/MS analysis in full-scan mode. Multivariable optimization was conducted using various filter parameters. Then, prediction models were constructed based on partial least squares discriminant analysis (PLS-DA), naïve Bayes (NB), and back-propagation artificial neural network (BP-ANN). Finally, the statistically significant variables were tentatively identified.

Materials and methods

Honey samples

Eighty-seven authentic honey samples, including 19 acacia (*Robiniapseudoacacia* L.), 22 linden (*TiliaamurensisRupr*.), 22 vitex (*Vitexnegundo* var. *heterophyllaRehd*.) and 24 rape (*Brassica campestris* L.), were collected in Beijing, Jilin, Hebei, and Shaanxi, China, respectively. More than 500 g of each honey sample was directly collected from beekeepers, and stored in a refrigerator at 4 °C until analysis.

Figure 1. Parts of honey samples used in this study.

Chemicals

Hexane (MS grade) was purchased from Fisher Scientific (Shanghai, China). *n*-Hexane was purchased from Sigma-Aldrich, and a series of alkanes (C_8H_{18}–$C_{25}H_{52}$) was used to calculate the retention index. Methyl decanoate (Sigma-Aldrich) was used as an internal standard. Water was purified using a Milli-Q water purification system (Millipore, Bedford, MA, USA).

Solid-phase micro-extraction

The SPME procedure was performed using a CTC auto injection system with a 2 cm-50/30 µm divinyl-benzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS)-coated fiber (p/n SU57348U). Before analysis, the fiber was conditioned for 1 hour at 270 °C in the injection port of the GC. The SPME conditions were as follows: 3 g of honey was placed in a 20-mL headspace screwtop vial (p/n 5183-4474) with 1.5 mL of deionized water and 0.5 g of sodium chloride, and the vial was sealed with a poly-tetrafluoroethylene (PTFE)/silicone septum (p/n 5183-4477). The DVB/CAR/PDMS-coated fiber was then exposed to the headspace of the sample solution for 30 minutes at 80 °C. When the extraction step was completed, the fiber was removed from the vial and inserted into the injection port of the GC for 2 minutes of thermal desorption at 250 °C.

Data processing and statistical analyses

The Agilent ChemStation data were transformed to MassHunter data by Agilent MassHunter GC/MS translator Version B.07.00. The transformed GC/MS data were imported into MassHunter unknown analysis software (Version B.07.01), and deconvolution and identification were performed according to the optimized parameters of the unknown analysis. The processed honey sample profiles were exported by the script as a cef file, and imported into Agilent Mass Profiler Professional (Version 13.0).

Chemometric methods such as principal component analysis (PCA), one-way analysis of variance (ANOVA), and prediction models including PLS-DA, NB, and BP-ANN were used.

Results and discussion

Data mining

Data filtering and chemometric analyses were carried out using Agilent MPP software. All the cef files were subjected to data filtering. A total of 2,734 entities were obtained through data alignment across four sample groups. According to the MPP workflow, the first filter was *filter by flags*, and this step was used to eliminate unreliable compounds. The present and marginal flags were set to filter an entity from the entire set if it was above the threshold or saturated for each sample. Entities in which at least 2 out of 87 samples had acceptable values were retained. The next filter was *filter by frequency*, by which entities were filtered based on their frequency of occurrence across samples. In this step, the entity must be present to pass the filter. These filtering conditions retained entities that appeared in each sample, in at least one condition. The third filter, *significance analysis*, was based on the p-value calculated by one-way ANOVA. To ensure that only the entities with a significant difference were retained, in most cases, the selected p-value cut-off was 0.05. After the three filter steps, the number of entities was reduced from 2,734 to 114. To identify compounds with abundance ratios or differences between a treatment and a control outside of a given fold-change cut-off or threshold, fold-change analysis was used as the final filter step. There were 110 entities retained when the fold-change cut-off was 2, demonstrating that a series of filter steps significantly reduced the number of variables, and the dimensionality of the dataset.

Principle component analysis

PCA is the most commonly used unsupervised statistical method to reduce the dimensionality of large datasets, and identify differences and associations between variables and samples. Based on the previous filter results, PCA was applied to 87 honey samples of four botanical origins to analyze their natural grouping. Generally, if the first four PCs explain more than 75 % of the total variation, a reliable model may be obtained[3]. It is necessary to further select and reduce the previously mentioned variables, thus strongly influencing the reliability of the model. The fold change cut-off could be optimized based on the above-mentioned filter steps. With the total variation explained by PC1–PC4 considered as the evaluation criteria, a fold-change cut-off of 200 was chosen. Figure 2 shows the score plot of the honey samples with a fold-change cut-off of 200. Linden honey samples were on the right plot of the first principal components (PC1), while the three other types of honey were on the left part of PC1. Therefore, PC1 separated the honey samples into linden honey and non-linden honey clearly. The linden and

Instrument conditions

Table 1. GC and mass spectrometer conditions.

Parameter	Value
GC system	Agilent 7890A*
Column	Agilent HP-5MS, 30 m × 0.25 mm, 0.25 µm (p/n 19091S-433)
Oven program	50 °C hold 2 minutes, at 5 °C/min to 180 °C, hold 2 minutes, at 10 °C/min to 250 °C hold 5 minutes
Carrier gas	Helium
Flow rate	1.0 mL/min
Injection mode	CTC auto injection
Injection port temperature	250 °C
MS System	Agilent 5975C*
Ion source	EI, 70 eV
Ion source temperature	230 °C
Quadrupole temperature	150 °C
Spectral acquisition	Full scan, 40–600 *m/z*

* The Agilent 7890B GC system and Agilent 5977B MS system are available, and have demonstrated better results.

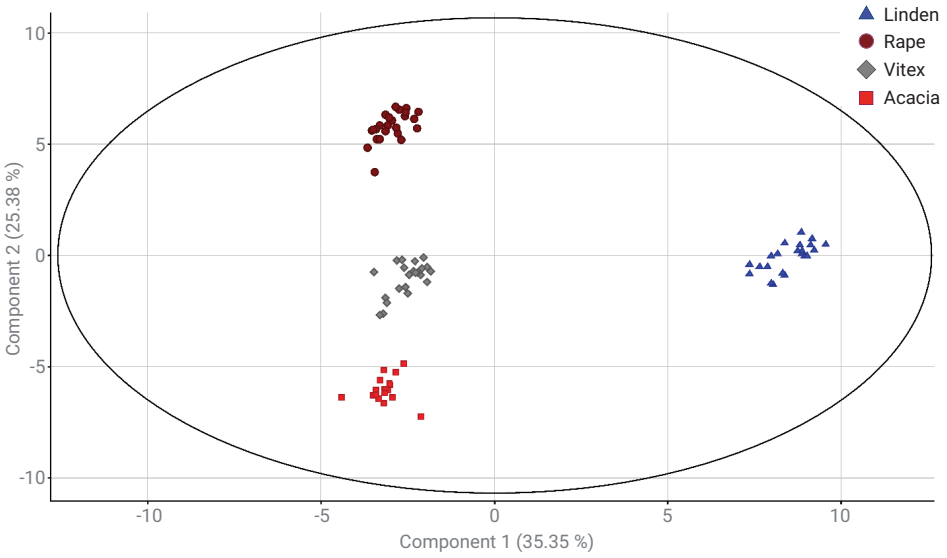

Figure 2. PCA scores of honey samples from different botanical origins with a fold change cut-off of 200.

rape honey samples both had positive PC2, while the vitex and acacia honey samples had negative PC2. PC1 and PC2 accounted for 35.35 % and 25.38 % of the variance, respectively, and the first four PCs accounted for 79.12 % of the total variability. Therefore, a fold-change cut-off of 200 was selected, and the 70 retained variables were prepared for model construction.

Model construction and prediction

The PCA results indicated that the filtered data could be used to discriminate the botanical origins of the honey samples. The 70 selected variables were used to develop a statistical model for the classification and prediction of honey of various origins. This study constructed three classification models, PLS-DA, NB, and BP-ANN, based on 87 authentic honey samples, including 19 acacia, 22 linden, 24 rape, and 22 vitex honey samples. The percentage of correctly classified honey samples during model training and validation demonstrated the recognition and prediction ability. Table 2 lists the accuracy of the three models; 100 % of the honey samples were accurately classified.

Although the recognition ability in model training and the prediction ability in model cross-validation was 100 %, validation of the constructed model using samples not included in the original 87 authentic honey samples was an essential step. Therefore, an additional 20 authentic honey samples not included in the development of the statistical model, including five honey samples of each botanical origin, were analyzed as the test set. Table 3 lists the prediction results with confidence measures. All 20 authentic honey samples were correctly predicted. Although the predicted group was consistent with the actual group, the values of the confidence measures varied for the PLS-DA, NB, and BP-ANN models.

Table 2. Model training and cross-validation results by PLS-DA, NB, and BP-ANN models.

	PLS-DA	NB	BP-ANN
Model training			
Recognition ability (%)	100	100	100
Model cross-validation			
Prediction ability (%)	100	100	100

The confidence measure is an essential indicator of the reliability of prediction results. A confidence measure greater than 0.7 indicates high reliability in the results. A confidence measure in the range of 0.5–0.7 indicates problematic sample classification, and a value less than 0.5 suggests incorrect information[4].

Table 3 shows that the confidence measures of all the NB prediction results were 1.00, whereas the confidence measures ranged 0.59–0.95 and 0.77–0.99 for the PLS-DA and BP-ANN models, respectively. Three vitex samples in the PLS-DA results had confidence measures ranging 0.59–0.69. The other two vitex samples had confidence measures of 0.71 and 0.76. The PLS-DA model produced slightly worse predictions for the vitex samples. In the BP-ANN results, except for vitex sample V03 and V05, which had confidence measures of 0.81 and 0.77, respectively, the samples had confidence measures greater than 0.98. In comparison, the NB and BP-ANN models predicted all the honey samples with satisfactory confidence measures.

Table 3. Prediction results by PLS-DA, NB, and BP-ANN models.

No.	Sample code	Actual name	Predicted name	Confidence measure		
				PLS-DA	NB	BP-ANN
1	A01	acacia	acacia	0.80	1.00	0.99
2	A02	acacia	acacia	0.85	1.00	0.99
3	A03	acacia	acacia	0.82	1.00	0.99
4	A04	acacia	acacia	0.93	1.00	0.99
5	A05	acacia	acacia	0.89	1.00	0.98
6	L01	linden	linden	0.81	1.00	0.98
7	L02	linden	linden	0.71	1.00	0.99
8	L03	linden	linden	0.95	1.00	0.99
9	L04	linden	linden	0.87	1.00	0.99
10	L05	linden	linden	0.82	1.00	0.99
11	R01	rape	rape	0.91	1.00	0.99
12	R02	rape	rape	0.89	1.00	0.99
13	R03	rape	rape	0.77	1.00	0.99
14	R04	rape	rape	0.84	1.00	0.99
15	R05	rape	rape	0.93	1.00	0.99
16	V01	vitex	vitex	0.66	1.00	0.99
17	V02	vitex	vitex	0.69	1.00	0.99
18	V03	vitex	vitex	0.59	1.00	0.81
19	V04	vitex	vitex	0.76	1.00	0.99
20	V05	vitex	vitex	0.71	1.00	0.77

Identification of volatile compound markers

A Venn diagram was used to identify the entities list in the prediction model for each botanical origin. Figure 3 shows the entity lists for each botanical origin and cross-section. The entity lists for acacia, vitex, linden, and rape honey were 50, 43, 52, and 47, respectively. Because linden honey is superior in quality and price to the other honeys, the marker identifications focused on the volatile compounds only found in linden honey samples. Figure 3 shows that eight of the initially selected volatile compounds only appeared in linden honey. They were tentatively identified by NIST. Table 4 lists the literature with their retention times, retention indexes, and ions. Except for pentanoic acid, 2-methyl-, anhydride, all the compounds were reported in the literature reference for linden honey[5-8]. Among them, *cis*-rose oxide was proposed as an indicator compound for linden honey by Blank; *et al*[6]. Blank also found the compound in the blossoms of the lime tree (*Tilia cordata*), but not in other types of honey[6]. Table 4 lists the compounds that could also be tentatively considered as marker compounds of linden honey despite the lack of literature on the volatile compounds in lime blossoms. These compounds had significant contributions in discriminating linden honey from other honeys.

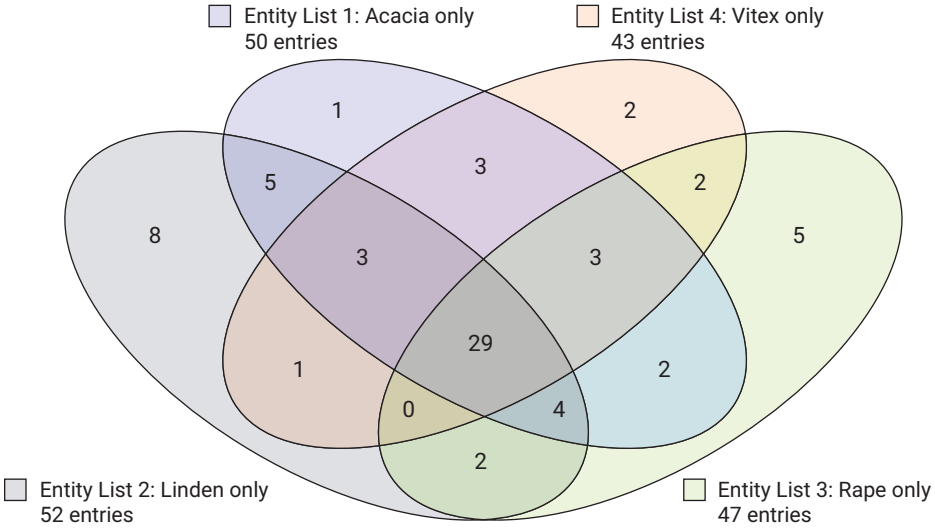

Figure 3. Venn diagram of honey samples from different botanical origins.

Table 4. Marker compounds tentatively identified in linden honey.

No.	Tentative compound identification	RT	RI	Ions	CAS	Ref.
1	Pulegone[a,b]	9.67	1008	109, 81, 152	89-82-7	[5]
2	*cis*-Rose oxide[a,b]	13.17	1126	139, 69, 83	876-17-5	[5-7]
3	Benzofuran, 4,5,6,7-tetrahydro-3,6-dimethyl-[a,b]	14.27	1164	108, 150, 79	494-90-6	[5]
4	1-methyl-4-(1-methylpropyl) benzene[a,b]	15.52	1206	119, 91, 117	1595-16-0	[5]
5	4,7-dimethyl-Benzofuran[a,b]	15.78	1215	145 146 148	28715-26-6	[5]
6	thymol[a,b]	18.1	1298	135, 150, 91	89-83-8	[5]
7	Carveol[a,b]	19.72	1359	119, 91, 134	99-48-9	[8]
8	Pentanoic acid, 2-methyl-, anhydride[a]	21.17	1414	99, 71, 41	63169-61-9	–

[a] Identified by NIST14.

[b] Identified by the literature reference.

Conclusion

To discriminate honey samples of various botanical origins, this study successfully applied the proposed classification methodology based on SPME and untargeted GC/MS analysis, combined with chemometrics. Three classification models, PLS-DA, NB, and BP-ANN, were constructed based on 87 authentic honey samples, with an accuracy of 100 %. The prediction results of 20 additional authentic honey samples indicated that the developed models were practical and reliable. Therefore, the combination of SPME-GC/MS nonspecific volatile compound profiles and chemometrics is an alternative, promising method for the classification and discrimination of honey samples of different botanical origins.

References

1. Cuevas-Glory, L. F.; *et al.* A review of volatile analytical methods for determining the botanical origin of honey. *Food Chem*. **2007**, *103*, 1032–1043.
2. Chen, H.; *et al.* Non-targeted volatile profiles for the classification of the botanical origin of Chinese honey by solid-phase microextraction and gas chromatography–mass spectrometry combined with chemometrics. *J. Sep. Sci.* **2017**, *40*, 4377–4384.
3. Chudzinska, M.; Baralkiewicz, D. Estimation of honey authenticity by multielements characteristics using inductively coupled plasma-mass spectrometry (ICP-MS) combined with chemometrics. *Food Chem. Toxicol*. **2010**, *48*, 284–290.
4. Wang, M.; *et al.* An integrated approach utilising chemometrics and GC/MS for classification of chamomile flowers, essential oils and commercial products. *Food Chem.* **2014**, *152*, 391–398.
5. Lušic, D.; *et al.* Volatile Profile of Croatian Lime Tree (*Tilia sp.*), Fir Honeydew (*Abies alba*) and Sage (*Salvia officinalis*) Honey. *Food Technol. Biotech.* **2007**, *45*, 156–165.
6. Blank, I.; Fischer, K. H.; Grosch, W. Intensive neutral odorants of linden honey Differences from honeys of other botanical origin. *Zeitschrift für Lebensmittel-Untersuchung und Forschung*. **1989**, *189*, 426–433.
7. Piasenzotto, L.; Gracco, L.; Conte, L. Solid phase microextraction (SPME) applied to honey quality control. *J. Sci. Food Agri*. **2003**, *83*, 1037–1044.
8. Špánik, I.; *et al.* Characterisation of VOC composition of Slovak monofloral honeys by GC × GC-TOF-MS. *Chem. Papers* **2013**, *67*, 127–134.

www.agilent.com/chem

Printed in the USA, May 22, 2018
5991-8967EN

Agilent
Trusted Answers