Talanta 129 (2014) 629–635

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Short communication

Characterization of volatile components in four vegetable oils by headspace two-dimensional comprehensive chromatography time-of-flight mass spectrometry

Wei Hu [a,b,1], Liangxiao Zhang [a,d,e,*,1], Peiwu Li [a,c,d,e,*,1], Xiupin Wang [a,e,*], Qi Zhang [a,b,c], Baocheng Xu [a,e], Xiaoman Sun [a,e], Fei Ma [a,c,e], Xiaoxia Ding [a,d]

[a] Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan 430062, China
[b] Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Wuhan 430062, China
[c] Key laboratory of Detection for Mycotoxins, Ministry of Agriculture, Wuhan 430062, China
[d] Laboratory of Risk Assessment for Oilseeds Products (Wuhan), Ministry of Agriculture, Wuhan 430062, China
[e] Quality Inspection and Test Center for Oilseeds Products, Ministry of Agriculture, Wuhan 430062, China

ARTICLE INFO

Article history:
Received 19 April 2014
Received in revised form
6 June 2014
Accepted 10 June 2014
Available online 17 June 2014

Keywords:
Volatile organic compounds
Classification
Edible vegetable oil
Headspace GC × GC–TOFMS
Chemometrics

ABSTRACT

Edible oil adulteration is the biggest source of food fraud all over the world. Since characteristic aroma is an important quality criterion for edible oils, we analyzed volatile organic compounds (VOCs) in four edible vegetable oils (soybean, peanut, rapeseed, and sunflower seed oils) by headspace comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (Headspace–GC × GC-TOFMS) in this study. After qualitative and quantitative analysis of VOCs, we used unsupervised (PCA) and supervised (Random forests) multivariate statistical methods to build a classification model for the four edible oils. The results indicated that the four edible oils had their own characteristic VOCs, which could be used as markers to completely classify these four edible oils into four groups.

1. Introduction

Edible vegetable oils are important to our daily life by providing energies, nutritional components, and pleasant flavors [1–3]. Soybean, rapeseed, sunflower seed, and peanut oils are major cooking oils in the world. As the same as olive oil in western countries, these oils have become a target of adulteration, while oil adulteration is the biggest source of food fraud [4]. To keep consumers from adulterated oils, a reliable method for detecting such adulteration is in great demand. Aroma is an important quality criterion for edible vegetable oils as a characteristic parameter [3,5]. Intuitively, edible vegetable oils possess their respective characteristic aroma. Volatile organic compounds (VOCs) have low molecular weights (less than 300 Da), produce an odor sensation, and are easily vaporized at room temperature [5]. VOCs play a significant role in wine aroma, and the presence/absence of VOCs in different proportions can be taken as a marker for identifying adulteration.

Many analytical methods have been proposed to isolate, identify, and quantify the volatile components that characterize aroma of oil [5]. Among these techniques, solid-phase micro-extraction (SPME) is a simple and fast technique for the extraction of volatile and non-volatile compounds without any solvent preparation [6–7]. Headspace is a fast, universal, sensitive, solvent-free, and economical method for isolation of volatile analytes from complex matrices [8]. Recently, the headspace solid-phase micro-extraction-gas chromatography (Headspace SPME-GC) method was utilized to determine solvent residues and aldehydes in edible oils [9–10].

Chemometrics is a multivariate data analysis tool often coupled with metabolomics and non-destructive testing methods such as near infrared spectroscopy (NIR). In respect to oil fraud, chemometrics is a powerful tool used qualitatively for classifying unknown samples with similar characteristics and quantitatively for determining adulterant analytes in samples [11]. Recent reports demonstrated that the chemometric methods of principal component analysis (PCA), self-organizing maps based on chaotic parameters, and cluster discriminant analysis (CDA) were used to distinguish edible oils from refined recycled cooking oils, identify edible oils from different regions, and detect adulteration of extra virgin olive oil with inferior edible oils, respectively [12–14].

* Corresponding authors at: Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan 430062, China. Tel.: +86 27 86812943; fax: +86 27 86812862.
E-mail addresses: liangxiao_zhang@hotmail.com (L. Zhang), peiwuli@oilcrops.cn (P. Li), wangxiupin@caas.cn (X. Wang).
[1] These authors contributed equally to this study.

http://dx.doi.org/10.1016/j.talanta.2014.06.010
0039-9140/

Generally, chemometric data analyses for oil fraud detection are increasingly and extensively used in routine quality assurance (QA) testing [12].

The aim of this study is to analyze volatile compounds in four vegetable oils by headspace comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry (Headspace GC × GC–TOF/MS) and to develop a classification and adulteration identification model for the four vegetable oils. At first, we developed the Headspace GC × GC–TOF/MS method to analyze volatile organic components in soybean, peanut, rapeseed, and sunflower seed oil samples. After qualitative and quantitative analysis of the volatile components, unsupervised multivariate statistical methods including the PCA and Hierarchical clustering analysis (HCA) and the supervised multivariate statistical method of random forests (RF) were used to build a classification model for the four edible oils.

2. Materials and methods

2.1. Samples and materials

To ensure that the selected oilseed samples could represent the actual status of peanut, soybean, rapeseed, and sunflower seed oils, we collected four types of oilseeds from the main producing areas of China but not edible oils from supermarkets. In total, 18, 25, 24, and 27 samples of soybeans, peanuts, sunflower seeds, and rapeseeds were collected for the study, respectively. The detailed information about these oilseeds is shown in Supplementary Material Table S1. All the samples were serially numbered and stored in a room at a constant temperature before extraction. A quality control (QC) sample was prepared by mixing 2 mL of each oil sample and used to check the repeatability of experiments and reliability of analytical results.

TenGuard oil presser, was from Foshan Taidian Intelligent Technology Co. Ltd., and ultra-pure Milli-Q water was manufactured by Millipore Elix Advantage 5 (Billerica, MA, USA).

2.2. Preparation of oil samples

A moderate quantity of seeds was cleaned, triturated, and squashed by a grinding mill, and then the fragments were put into Taigu oil presser. The exudation was collected into the centrifuge tube and then centrifuged at 4500 rpm in 5 min. The supernatant liquid was cold drawn into the oil we needed. Each oil sample was numbered, and the same type of seed, which had been differently labeled, was numbered in the same manner.

2.3. Chromatographic conditions

The oil extracts and QC sample were analyzed using a LECO Pegasus 4D GC × GC–TOF/MS instrument (LECO Corporation, St. Joseph, MI, USA) equipped with Agilent 6890N. A non-polar/mid-polar column set was optimized for GC × GC separation. The first dimension (1D) column was 60 m × 250 μm × 0.1 μm DB-5 MS (Agilent J&W GC Columns, USA) and the second dimension (2D) column was 2 m × 150 μm × 0.15 μm Rxi-17Sil MS (Restek, USA).

The temperatures of the GC inlet and transfer line were set at 200 °C and 225 °C, respectively. High purity helium (99.9995%) was used as the carrier gas at a constant flow of 1 mL min^{-1}. Cryogenic modulation was performed with a 4 s modulation period (PM). A CombiPAL autosampler (CTC Analytics, Zwingen, Switzerland) was used with an injection volume of 1 mL and split ratio of 10:1. The incubation temperature was 150 °C, incubation time was 30 min, agitator speed was 500 rpm, agitator speed during extraction was 100 rpm, agitator-on time was 5 s, agitator-off time was 2 s, syringe temperature was 150 °C, syringe flush time was 30 s, and GC cycle time including the oven cooling time was 55 min.

The oven temperature for the first column was held at 40 °C for 2 min, and then ramped to 200 °C at the rate of 5 °C/min and held for 1 min. The second oven was operated at 10 °C higher than the first oven throughout the process. The modulation period was 4 s with the heat pulse of 1 s.

A Pegasus® IV time-of-flight mass spectrometer (LECO Corp.) was used as a detector. The detector voltage was set to −1650 V, and MS was operated in electron impact ionization mode (70 eV). The acquisition delay was 30 s, and ions were collected in the mass range of 35–600 amu at an acquisition rate of 100 spectra s^{-1}. The ion source temperature was 230 °C.

2.4. Data processing

The raw data were pre-processed by the LECO ChromaTOFTM workstation. The task was selected as below: The baseline was computed; the peaks were found above the baseline; a library search was performed on all identified peaks; the area and height of the peaks were calculated without a calibration. The peaks were extracted with a signal-to-noise ratio (S/N) of above 10. The minimum similarity match is 700 before the names are assigned, and the allowed molecular weights are from 50 to 1000. All the collected masses were searched for in the selected library. Each peak was automatically determined using the software after background correction and resolution. Tentative identification was conducted by searching the National Institute of Standards and Technology (NIST) library. Eventually, we obtained a table containing the information of the peak number, name, similarity, R.T. (s) and area % of each oil sample.

A Matlab program was used to collect all chemical compounds with the similarity above 700. The same volatile compound was selected to match the name of each oil sample in the same class. At last, we acquired a csv table with oil sample serial numbers in the first column and chemical compounds in the first row.

2.5. Statistical analysis

Our data matrix includes peak areas of edible VOCs. Before multivariate analysis, the data matrix was preprocessed by generalized log2 transformation and Pareto scaling (mean-centered and divided by the square root of the standard deviation of each variable). In exploratory data analysis, PCA and HCA were employed to screen sampling clusters and variable distributions in the four groups of edible oils. To build a classification model for the four edible oils, an effective supervised multivariate statistical method of random forests (RF) was used.

Data were processed on a Pentium 4 personal computer. Data simulation was implemented for adulterated oils in Matlab 2011a for windows (The Mathworks, Natick, MA). Data preprocessing (transformation and scaling), clustering (PCA and HCA), and classification (RF) were conducted using the metabolomics data analysis tool MetaboAnalyst 2.0 [15].

3. Results and discussion

3.1. Compositions of volatile components in four kinds of edible oils

Headspace volatiles from the four kinds of edible oils were analyzed by Headspace GC × GC–TOF/MS. Table 1 lists the volatile organic components tentatively identified in each oil sample. A total of 114 VOCs were tentatively identified based on the mass spectral similarity search, including 83, 64, 74, and 88 VOCs from

Table 1
Tentative identification and presence of volatile organic components in four classes of edible oils.

No.	Compound	CAS no. [a]	Similarity	RT1	RT2	Soybean oil	Peanut oil	Sunflower seed oil	Rapeseed oil
1	Formaldehyde, dimethylhydrazone	2035-89-4	805	6.70	1.71	+[b]	+	+	+
2	Butanal	123-72-8	933	5.77	3.83	+	+	+	+
3	Hexane	110-54-3	929	5.83	3.61	+	+	+	+
4	Trichloromethane	67-66-3	979	5.97	3.81	+	+	+	+
5	Cyclopentane, methyl-	96-37-7	893	6.03	3.66	+	+	+	+
6	1-Penten-3-ol	616-25-1	918	6.70	1.86	+	+	+	+
7	Heptane	142-82-5	921	6.83	3.69	+	+	+	+
8	Furan, 2-ethyl-	3208-16-0	965	6.90	3.92	+	+	+	+
9	1-Pentanol	71-41-0	920	8.03	0.13	+	+	+	+
10	Heptane, 2,4-dimethyl-	2213-23-2	930	8.63	3.87	+	+	+	+
11	Hexanal	66-25-1	903	8.77	0.37	+	+	+	+
12	Cyclotrisiloxane, hexamethyl-	541-05-9	935	9.17	3.79	+	+	+	+
13	2-Hexenal	505-57-7	948	10.03	2.5	+	+	+	+
14	1-Hexanol	111-27-3	925	10.50	0.35	+	+	+	+
15	Pentanoic acid	109-52-4	865	10.97	1.26	+	+	+	+
16	2-Heptanone	110-43-0	943	11.03	1.31	+	+	+	+
17	Heptanal	111-71-7	926	11.37	2.39	+	+	+	+
18	Trichloroacetic acid, pentyl ester	33972-81-5	808	11.43	1.35	+	+	+	+
19	Butyrolactone	96-48-0	959	11.90	0.12	+	+	+	+
20	Dimethyl sulfone	67-71-0	915	12.17	2.64	+	+	+	+
21	2-Heptenal, (Z)-	57266-86-1	923	12.97	2.67	+	+	+	+
22	1-Heptanol	111-70-6	935	13.37	1.33	+	+	+	+
23	1-Octen-3-one	585-25-1	938	13.77	0.56	+	+	+	+
24	1-Octen-3-ol	3391-86-4	886	13.70	0.46	+	+	+	+
25	Benzonitrile	100-47-0	919	13.90	1.64	+	+	+	+
26	2-Octanone	111-13-7	958	13.97	2.53	+	+	+	+
27	Furan, 2-pentyl-	3777-69-3	895	13.97	2.38	+	+	+	+
28	Cyclotetrasiloxane, octamethyl-	556-67-2	944	14.23	0.95	+	+	+	+
29	Octanal	124-13-0	924	14.37	2.51	+	+	+	+
30	1-Nonen-4-ol	35192-73-5	910	14.50	2.8	+	+	+	+
31	2(3H)-Furanone, 5-ethyldihydro-	695-06-7	932	15.97	3.96	+	+	+	+
32	2-Octenal, (E)-	2548-87-0	941	15.97	2.78	+	+	+	+
33	4-Nonenal, (E)-	2277-16-9	849	17.10	2.66	+	+	+	+
34	Nonanal	124-19-6	933	17.37	2.6	+	+	+	+
35	Cyclopentasiloxane, decamethyl-	541-02-6	794	18.83	1.78	+	+	+	+
36	2,4-Nonadienal, (E,E)-	5910-87-2	957	20.57	1.24	+	+	+	+
37	Benzothiazole	95-16-9	956	21.03	0.43	+	+	+	+
38	2-Ethoxyethanol	110-80-5	804	5.23	1.56	+	−	+	+
39	Propanal, 2-methyl-	78-84-2	900	5.63	1.7	+	−	+	+
40	1-Propanol	78-83-1	915	6.03	3.76	+	+	+	−
41	Methacrolein	78-85-3	907	5.63	3.73	+	-	+	+
42	Butanal, 3-methyl-	590-86-3	933	6.30	3.87	+	−	+	+
43	2-Butenal	4170-30-3	971	6.37	0.05	+	+	+	+
44	Butanal, 2-methyl-	96-17-3	922	6.43	3.88	+	−	+	+
45	1-Butanol	71-36-3	921	6.43	1.8	+	+	+	−
46	Benzene	71-43-2	885	6.37	3.92	+	−	+	+
47	Diethylene glycol	111-46-6	909	5.77	1.58	+	−	+	+
48	2-Penten-1-ol, (E)-	1576-96-1	834	8.10	0.18	+	+	−[c]	+
49	4-Heptenal, (Z)-	6728-31-0	857	11.23	2.49	+	+	+	−
50	Pyrazine, 2,5-dimethyl-	123-32-0	901	11.63	0.95	+	+	−	+
51	Cyclohexanecarboxaldehyde	2043-61-0	834	11.97	2.76	+	+	+	−
52	2-Heptenal, (E)-	18829-55-5	854	12.63	2.67	+	+	+	−
53	3-Ethylcyclopentanone	10264-55-8	883	12.97	1.18	+	+	+	−
54	Decane	124-18-5	871	17.17	2.04	+	+	+	−
55	2,5-Furandione, 3,4-dimethyl-	766-39-2	920	15.50	3.83	+	+	+	−
56	1-Octanol	111-87-5	919	16.37	2.49	−	+	+	+
57	Phenylethyl Alcohol	60-12-8	945	17.70	3.58	+	+	−	+
58	3-Nonen-2-one	14309-57-0	908	18.43	0.93	+	+	+	−
59	3-Hepten-1-ol	10606-47-0	733	20.77	3.14	+	−	+	+

Table 1 (*continued*)

No.	Compound	CAS no. [a]	Similarity	RT1	RT2	Soybean oil	Peanut oil	Sunflower seed oil	Rapeseed oil
60	4-Oxononanal	–	912	21.50	1.86	+	+	+	−
61	2-Decenal, (Z)-	2497-25-8	933	21.83	2.9	+	+	+	+
62	Carbon disulfide	75-15-0	911	5.50	3.68	−	−	+	+
63	Cyclohexane	110-82-7	866	6.37	3.75	−	−	+	+
64	2-Pentanone	107-87-9	910	6.77	0.01	+	+	−	−
65	1-Penten-3-one	1629-58-9	930	6.77	0.03	+	−	−	+
66	2,3-Pentanedione	600-14-6	779	6.83	0.26	+	−	−	+
67	Propanoic acid	79-09-4	877	6.77	3.87	+	−	−	+
68	Furan, 2-propyl-	4229-91-8	953	8.57	0.14	+	−	+	−
69	2-Pentenal, (E)-	1576-87-0	920	7.83	2.26	+	−	+	+
70	2-Butenal, 3-methyl-	107-86-8	908	8.43	0.55	−	−	+	+
71	Formamide, N,N-dimethyl-	68-12-2	913	8.63	0.95	+	−	−	+
72	Pyrazine, methyl-	109-08-0	942	9.37	0.94	+	−	−	+
73	Formic acid, pentyl ester	638-49-3	924	9.37	0.34	+	−	+	−
74	8-Chloro-1-octanol	23144-52-7	782	9.57	2.23	+	−	−	+
75	Furfural	98-01-1	941	9.63	2.92	+	−	−	+
76	Cyclopentanone, 2-methyl-	1120-72-5	743	9.70	1.44	+	−	+	−
77	2-n-Butyl furan	4466-24-4	878	13.97	2.36	−	−	+	+
78	6-Hepten-3-ol	19781-77-2	745	12.77	3.18	+	−	−	+
79	Camphene	79-92-5	918	12.77	0.54	−	−	+	+
80	5-Hepten-2-one, 6-methyl-	110-93-0	911	13.90	0.72	+	−	−	+
81	3-Octanol	589-98-0	948	14.10	2.35	+	+	−	−
82	Pyrazine, 2-ethyl-5-methyl-	13360-64-0	908	14.30	2.96	+	−	−	+
83	Pyrazine, trimethyl-	68-12-2	913	8.63	0.95	+	−	−	+
84	Benzyl Alcohol	100-51-6	906	15.37	3.44	+	−	+	−
85	2(5H)-Furanone, 5-ethyl-	2407-43-4	908	15.50	3.94	+	+	−	+
86	n-Caproic acid vinyl ester	3050-69-9	925	16.10	1.1	−	+	+	−
87	2,2-Dimethylpropanoic acid, tridec-2-ynyl ester	–	778	16.10	3.43	+	−	−	+
88	Pyrazine, 3-ethyl-2,5-dimethyl-	13360-65-1	926	16.63	2.97	+	−	−	+
89	7-Oxabicyclo[2.2.1]hept-5-en-2-one	95530-78-2	766	16.63	3.35	+	−	−	+
90	Furan, 2-butyltetrahydro-	1004-29-1	884	14.10	0.9	−	+	+	−
91	Undecane	1120-21-4	952	17.23	0.14	−	+	+	−
92	2-Nonenal, (E)-	18829-56-6	906	19.03	0.94	−	+	+	−
93	Oxalic acid, allyl octyl ester	–	868	20.23	0.77	−	+	+	−
94	Glyphosate	1071-83-6	999	4.97	0.96	−	+	−	−
95	Thiocyanic acid, methyl ester	556-64-9	954	7.17	0.53	−	−	−	+
96	Ethenamine, N-methylene-	38239-27-9	902	7.30	0.24	−	−	−	+
97	Disulfide, dimethyl	624-92-0	933	7.63	0.26	−	−	−	+
98	3-Pentenenitrile	4635-87-4	901	7.97	0.67	−	−	−	+
99	Propanoic acid, 2-hydroxy-2-methyl-	594-61-6	824	8.83	0.49	−	−	−	+
100	Cyclopentanol, 2-methyl-	24070-77-7	831	9.77	1.3	−	−	+	−
101	2-Furanmethanol	98-00-0	950	10.17	0.82	−	−	−	+
102	5-Cyano-1-pentene	5048-19-1	947	10.30	0.95	−	−	−	+
103	2(5H)-Furanone	497-23-4	932	12.03	0.18	−	−	−	+
104	Hexanenitrile, 5-methyl-	19424-34-1	917	12.63	2.79	−	−	−	+
105	1,5-Hexadien-3-ol	924-41-4	839	5.83	3.78	−	−	−	+
106	Hexanenitrile	628-73-9	891	10.77	2.71	−	−	−	+
107	2(3H)-Furanone, dihydro-4-methyl-	1679-49-8	921	13.17	3.9	+	−	−	-
108	1-Butene, 4-isothiocyanato-	3386-97-8	792	13.83	1.29	−	−	−	+
109	2,5-Furandione, dihydro-3-methyl-	4100-80-5	834	14.97	3.68	−	−	−	+
110	4H-Pyran-4-one, 2,3-dihydro-3,5-dihydroxy-6-methyl-	28564-83-2	937	18.57	3.69	−	−	−	+
111	Benzenepropanenitrile	645-59-0	946	21.43	0.41	−	−	−	+
112	2-Undecenal	53448-07-0	924	24.57	2.92	−	+	−	−
113	2-Methoxy-4-vinylphenol	7786-61-0	922	23.37	3.77	−	−	−	+
114	2-Cyclohexen-1-ol, 1-butyl-	88116-46-5	813	19.37	1.62	−	+	−	−

[a] '–' means there is no CAS no. for this compound.
[b] '+' represents this component appear in this edible oil.
[c] '−' denotes this component does not appear in this edible oil.

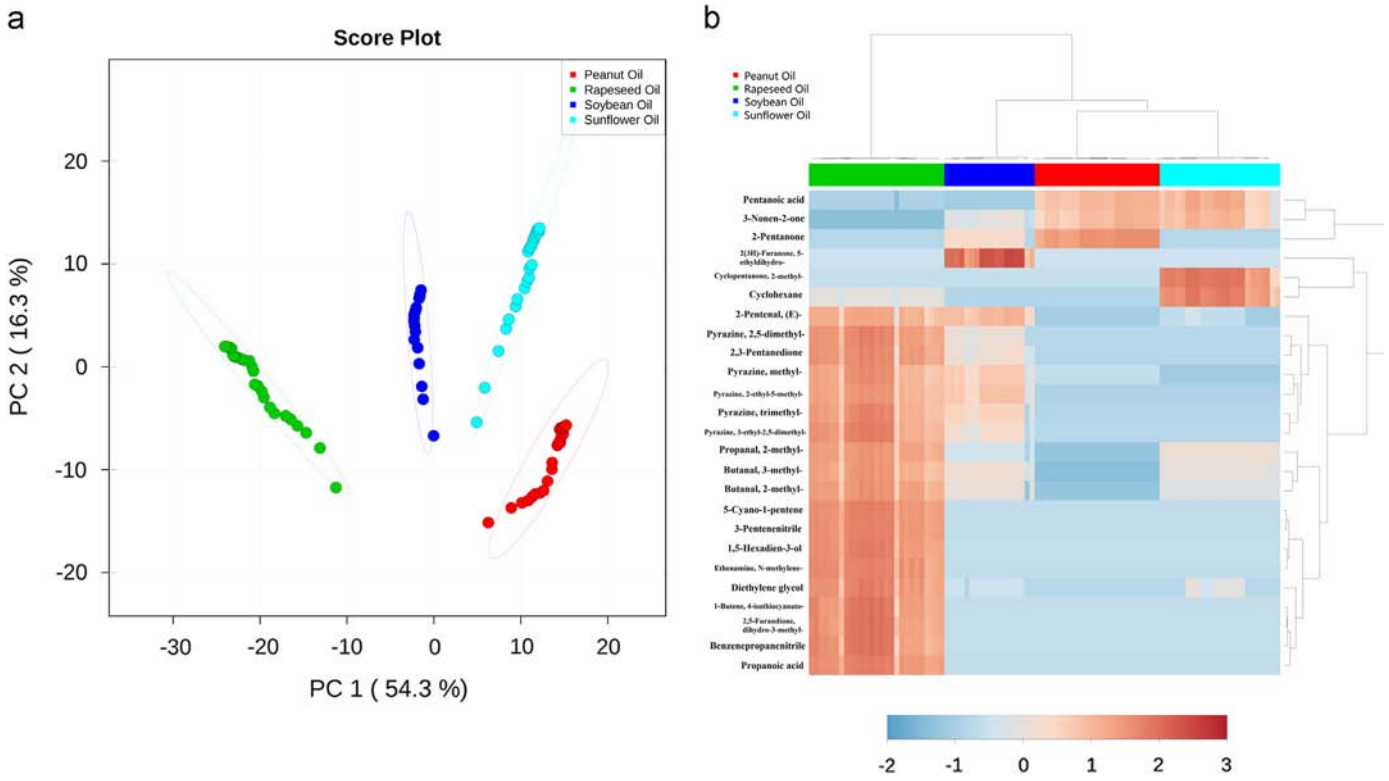

Fig. 1. (a) Score plot obtained from PCA of data about four kinds of edible oils; (b) heat map of volatile components of four kinds of edible vegetable oils.

soybean, peanut, sunflower seed, and rapeseed oils, respectively. No. 1–37 VOCs are common ones in the four types of edible oils. From the view of adulteration identification, selective components are more interesting to us. As shown in Table 1, rapeseed oil has 16 selective volatile compounds, including Thiocyanic acid methyl ester, N-methylene-ethenamine, Dimethyl disulfide, 3-Pentenenitrile, 2-Hydroxy-2-methyl-propanoic acid, 2-Furanmethanol, 5-Cyano-1-pentene, 2(5H)-Furanone, 5-Methyl-hexanenitrile, 1,5-Hexadien-3-ol, Hexanenitrile, 4-Isothiocyanato-1-butene, Dihydro-3-methyl-2,5-furandione, 2,3-Dihydro-3,5-dihydroxy-6-methyl-4H-pyran-4-one, Benzenepropanenitrile and 2-Methoxy-4- vinylphenol. In contrast, 2-Methyl-cyclopentanol is the selective component in sunflower seed oil, while 2-Undecenal, 2-Methoxy-4-vinylphenol, and 1-butyl-2-Cyclohexen-1-ol are selective components in peanut oil and Dihydro-4-methyl-2(3H)-Furanone is in soybean oil. Though the above components are tentatively identified by library search, their mass spectra and retention times could be employed to detect them in GC–MS data of edible oils. It is well known that selective components are important to adulteration detection because they could be employed to test whether an edible oil is adulterated into other oils. However, some selective components are of low concentrations. For example, the relative content of 2-Methyl-cyclopentanol, the selective component in sunflower seed oil, ranges from 0.043% to 1.07% in sunflower seed oil in this study. If an oil sample is adulterated with a low content of sunflower oil, it is hard to discover this adulteration by this selective component. In comparison, the multivariate model is a more robust method to detect adulteration of edible oils.

3.2. Exploratory data analysis

After determination and quantification of volatile components in the four edible oils, the data matrix of peak areas was preprocessed by generalized log2 transformation and Pareto scaling. At first, PCA and HCA were used to screen sample clusters and variable distributions in the four groups. As seen from the score plot obtained by PCA in Fig. 1a, four types of edible oils were clearly classified into four groups. Meanwhile, the four groups were mainly differentiated from each other in the first principal component (PC), and the first two PCs of each group showed narrow ranges, which indicated that the first two PCs (especially the first PC) contained adequate information for classifying these types of edible oils.

To investigate variable distributions in the four groups, a heat map was illustrated for the VOC profiles of the four edible oils. In the heat map, the similarity measure was Euclidean distance, while the clustering algorithm was Ward's linkage minimizing the sum of squares of any two clusters. To clearly demonstrate important variables, only 25 out of the 115 variables were employed to build the heat map. As shown in Fig. 1b, the similar results of cluster analysis were obtained. More importantly, we could find variable distributions in the four groups from this heat map, which showed the same variable distributions as in Table 1.

3.3. Classification of four kinds of edible oils by random forests

After exploratory data analysis, we found that the four edible oils could be clearly classified into four groups. To build a classification model for the four edible oils, an effective supervised multivariate statistical method of random forests (RF) was used. Random forests are a multitude of tree predictors combined in such a way that each tree depends on the values of a random vector sampled independently, with the same distributions for all the trees in the forest [16]. The sample proximity matrix derived from these training trees is generated to collect similarity information of the samples for classification. Class prediction is based on the majority vote of the ensemble. Compared with other supervised multivariate statistical methods such as partial least squares-discriminant analysis (PLS-DA) and support vector machine (SVM), RF is directly available for multi-class classification. In this study, the number of classification trees were set to

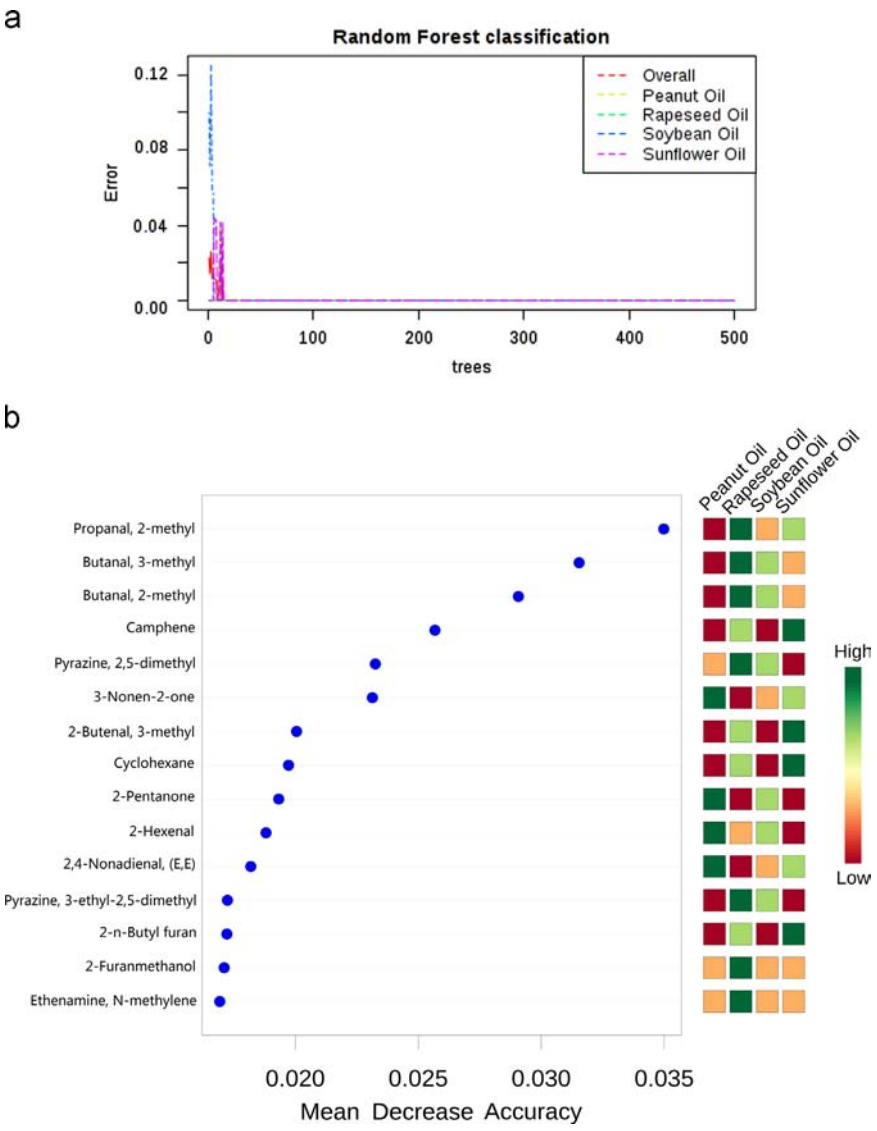

Fig. 2. (a) Cumulative error rates by random forest classification; (b) significant features identified by random forests.

500. During tree construction, one third of the samples were taken as the test set. The out-of-bag (OOB) data were then used as the test sample to obtain an unbiased estimate of the classification error (OOB error). From Fig. 2a, it was found that five errors decreased to zero after less than 20 trees, and the OOB error equaled 0 in the final classification model.

Random forests could provide a measure for variable importance. Fig. 2b illustrates the contribution of variables to classification. According to the mean decrease, 15 most important volatile compounds including 2-methyl-Propanal, 3-methyl-Butanal, 2-methyl-Butanal, camphene, 2,5-dimethyl-Pyrazine, 3-Nonen-2-one, 3-methyl-2-Butenal, Cyclohexane, 2-Pentanone, 2-Hexenal, (E,E)-2,4-Nonadienal, 3-ethyl-2,5-dimethyl-Pyrazine, 2-n-Butyl furan, 2-Furanmethanol, and N-methylene-Ethenamine were selected for classifying the four edible oils. This result is the same as the invariable analysis results. In the future, these important volatile compounds could be essential markers for edible oils and would be employed to detect adulteration of edible oils. Especially, the selective components selected in this study could correctly classify four edible oils into four groups with the help of random forests. Moreover, since volatile components are important aroma components, these selective volatile components could also be utilized to research the formation mechanism of the characteristic aromas and evaluate the quality and grade of edible oils.

4. Conclusion

Adulteration of edible oils is the largest source of food fraud all over the world. In this study, volatile components in four edible vegetable oils were analyzed by Headspace GC × GC–TOF/MS and applied to classifying edible oils with the help of multivariate statistical methods. The results indicate that each type of edible oil has its own selective volatile components and the VOC profiles of the edible oils can completely classify the oils into four groups. Therefore, the VOCs can be taken as the markers of these four edible oils. In the future, these important volatile compounds can be employed to detect adulteration of edible oils.

Acknowledgments

This work was supported by the Project of National Science & Technology Pillar Plan (2012BAK08B03), the National Natural

Science Foundation of China (21205118 and 31000787), the Special Fund for “Quality Inspection” Research in the Public Interest (2012104010-4), and the earmarked fund for China Agriculture research system (CARS-13).

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.talanta.2014.06.010.

References

[1] A. Cert, W. Moreda, C.M.C. Pérez, J. Chromatogr. A 881 (2000) 131–148.
[2] S.O. Yeboah, Y.C. Mitei, J.C. Ngila, L. Wessjohann, J. Schmidt, Food Res. Int. 47 (2012) 259–266.
[3] F.F. Zhao, J.Q. Liu, X.P. Wang, P.W. Li, W. Zhang, Q. Zhang, Eur. J. Lipid Sci. Tech. 115 (2013) 337–347.
[4] C.M. Jeffrey, J. Spink, M. Lipp, J. Food Sci. 77 (2012) 118–126.
[5] B.T. Sonia, C. Elisabetta, L.C. Pier, D. Douja, Z. Mokhtar, Food Chem. 99 (2006) 315–325.
[6] M.H. Faouzia, M. Hedia, D. Douja, F. Xavier, L.C. Louisette, Z. Mokhhtar, Food Chem. 103 (2013) 467–476.
[7] C. Teresa, A. Barbara, Food Chem. 141 (2013) 2025–2035.
[8] J.F. Cavalli, X. Fernandez, L. Lizzani-Cuvelier, A.M. Loiseau, J. Agric. Food Chem. 51 (2013) 7709–7716.
[9] M. Michulec, W. Wardencki, J. Chromatogr. A 1071 (2005) 119–124.
[10] C.H. Ma, J.J. Ji, C. Tan, D.M. Chen, F. Luo, Y.R. Wang, X. Chen, Talanta 120 (2014) 94–99.
[11] J.C. Moore, M. Lipp, J.C. Griffiths, Food Tech. 62 (2011) 46–50.
[12] F. Mümtaz, H. Dıraman, D. Özdemir, J. Am. Oil Chem. Soc. 90 (2013) 1661–1671.
[13] J.S. Torrecilla, J.C. Cancilla, G. Matute, P. Díaz-Rodríguez, A.I. Flores, J. Food Process Eng. 118 (2013) 400–405.
[14] T.B. Liu, L.J. Zhou, Z.W. Chen, B.B. Li, Y. Shi, J. Am. Oil Chem. Soc. 90 (2013) 1599–1606.
[15] J. Xia, N. Psychogios, N. Young, D.S. Wishart, Nucleic Acids Res. 37 (2009) 652–660.
[16] L. Breiman, Mach. Learn. 45 (2001) 5–32.