JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Article

pubs.acs.org/JAFC

Classification and Adulteration Detection of Vegetable Oils Based on Fatty Acid Profiles

Liangxiao Zhang,*[,†,∥,⊥] Peiwu Li,*[,†,§,∥,⊥] Xiaoman Sun,[†,⊥] Xuefang Wang,[†,⊥] Baocheng Xu,[†,⊥] Xiupin Wang,[†,⊥] Fei Ma,[†,‡,⊥] Qi Zhang,[†,‡,§] and Xiaoxia Ding[†,∥]

[†]Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan 430062, China

[‡]Key Laboratory of Biology and Genetic Improvement of Oil Crops, [§]Key Laboratory of Detection for Mycotoxins, [∥]Laboratory of Risk Assessment for Oilseeds Products (Wuhan), and [⊥]Quality Inspection and Test Center for Oilseeds Products, Ministry of Agriculture, Wuhan 430062, China

Supporting Information

ABSTRACT: The detection of adulteration of high priced oils is a particular concern in food quality and safety. Therefore, it is necessary to develop authenticity detection method for protecting the health of customers. In this study, fatty acid profiles of five edible oils were established by gas chromatography coupled with mass spectrometry (GC/MS) in selected ion monitoring mode. Using mass spectral characteristics of selected ions and equivalent chain length (ECL), 28 fatty acids were identified and employed to classify five kinds of edible oils by using unsupervised (principal component analysis and hierarchical clustering analysis), supervised (random forests) multivariate statistical methods. The results indicated that fatty acid profiles of these edible oils could classify five kinds of edible vegetable oils into five groups and are therefore employed to authenticity assessment. Moreover, adulterated oils were simulated by Monte Carlo method to establish simultaneous adulteration detection model for five kinds of edible oils by random forests. As a result, this model could identify five kinds of edible oils and sensitively detect adulteration of edible oil with other vegetable oils about the level of 10%.

KEYWORDS: *fatty acid profiles, adulteration identification, edible oil, GC/MS, chemometrics*

INTRODUCTION

Edible vegetable oils are a kind of important food in our daily life, which provide human being with energy and nutritional components including but not limited to essential fatty acids, phytosterols, tocopherols, phenolic compounds, vitamins, and volatile organic compounds (VOC).[1,2] All over the world, palm, soybean, rapeseed, sunflower, and peanut oils are major cooking oil. Among them, palm and soybean oils possess the largest market share in vegetable oil consumption.[3] However, owing to pleasant flavor and nontransgenic merit, the proportion of peanut, rapeseed and sunflower oils gradually increase though their prices are higher than soybean and palm oils. Moreover, sesame oil is a high price edible vegetable oil, which is often used as a flavor enhancer in Asian countries. Ensuring the authenticity of food has been a problem for millennia.[4] Edible vegetable oil is also not an exception. As the same as adulteration of olive oil in western countries, adulteration in other high-price oils is still the biggest source of agricultural fraud problems. Therefore, it is a great demand for a reliable method to detect such adulterations.

The main compound in vegetable oil is triglyceride (~95%), which is composed of fatty acids.[5,6] Fatty acid profiles can serve as a characteristic of edible vegetable oil. Traditionally, only high-abundance fatty acids were analyzed and employed to classify edible vegetable oils. Recently, however, some common fatty acids (especially oleic acid) become a target of oilseed breeding, and the distribution of common fatty acids in different vegetable oils become increasingly overlapped. On the contrary, the low-abundance fatty acids are more specific to edible vegetable oils, such as erucic acid in rapeseed oil. Though these low-abundance fatty acids were not so attractive to nutritionist, they possess high sensitivity to adulteration identification of edible oils. The entire family of fatty acids, which is referred to as "fatty acidomics",[7] was therefore taken as a key marker and quality parameter of different oilseeds and their products.[8]

Fatty acids are usually analyzed by gas chromatography mass spectrometry (GC-MS) with derivatization[9] and routinely identified by comparing retention index (equivalent chain length, ECL[10]) with those of the authentic FAME standards and/or doing similarity search in reference mass spectra library.[11−13]

As the same as metabolomics, the multivariate chemometric methods are important to the application of fatty acid profiles. One often resorts to multivariate chemometric methods for fatty acids in edible oils, mainly including principal component analysis (PCA), hierarchical clustering analysis (HCA), soft independent modeling of class analogy (SIMCA), partial least-squares discriminant analysis (PLS-DA) and support vector machine (SVM).[14−16] As described by Ai et al.,[5] machine learning techniques of an ensemble of classification models are of particular interest. Among them, Random Forests (RF) is powerful to deal with multivariate classification,[17,18] in which, a

Received: March 8, 2014
Revised: July 29, 2014
Accepted: July 31, 2014

dx.doi.org/10.1021/jf501097c | J. Agric. Food Chem. XXXX, XXX, XXX−XXX

visual exploration of the proximity matrix could be obtained using the multidimensional scaling (MDS) technique to intuitively measure similarity between different samples. Besides, it can also provide a measurement for variable importance.

In this study, we developed a simultaneous adulteration detection method based on fatty acid profiles for vegetable edible oils and selected soybean, rapeseed, sunflower, peanut and sesame oils as target oils to illustrate and validate this method. First, fatty acid profiles of five main edible oils were established by GC-MS in selected ion monitoring mode. Using mass spectral characteristics of selected ions and equivalent chain length (ECL), fatty acids were identified and employed to classify five kinds of edible oils by using unsupervised (principal component analysis and hierarchical clustering analysis) and supervised (random forests) multivariate statistical methods. Moreover, since triacylglycerols are relatively stable and no chemical reaction happens when physical blend of vegetable oils, adulterated oils were simulated by adulterating the one oil with at least 10% of other four oils or mixtures thereof and employed to establish simultaneous adulteration detection model for five types of edible oils by random forests.

2. EXPERIMENTAL SECTION

2.1. Materials and Regents. To ensure that the oil samples could represent the actual status of edible oils, we adhere to the following sampling rules: (a) with respect to oil seeds selected for laboratory pressing, the oilseed samples were collected from the main produce areas of China; (b) the oilseeds were collected from farmers and local markets without considering their varieties. Totally, the 17 soybean samples, 75 peanut samples, 57 sunflower seed samples, 76 rapeseed samples, and 73 sesame samples were collected from different product areas. The detailed information was shown in Supporting Information Table S1. The seed oils were prepared by oil mill machinery (TZC-0502, Brand of TEN GUARD, China). Supelco 37 component FAME mix (No. 47885-U) was purchased from Sigma (St. Louis, MO, USA). 11-Octadecenoic acid (C18:1n-7, >97.0 purity) and 7-hexadecenoic acid methyl ester were purchased from Sigma (St. Louis, MO, USA).

2.2. Experimental Procedure of Derivatization. As described in the previous study,[19] 0.06 g of vegetable oil sample was diluted with 2 mL of a solvent of diethyl ether and petroleum ether (v/v 1:1), and 1 mL of 0.4 M KOH−CH_3OH was added. Then, it was vortex-mixed for 30 s and placed at room temperature for 2.5 h, and then, 2 mL of redistilled water was added. It was then vortex-mixed and centrifuged at 4500 rpm for 2 min. The organic phase (200 μL) was collected and diluted by 800 μL of petroleum ether, prior to analysis by GC-MS.

2.3. GC-MS Analysis. The analyses were performed by Agilent GC-7890 gas chromatograph interfaced to a Agilent 5973 mass spectrometer. In the gas chromatography system, a fused silica capillary column DB-23 (30 m × 0.25 mm i.d. 0.15 μm film) (Agilent Technologies) was used. Helium (99.999% purity) was used as carrier gas at a flow-rate of 1.2 mL min^{-1}. The column was first set at 100 °C and held for 0.2 min, temperature was subsequently increased to 215 °C at the rate of 10 °C/min and held for 0.1 min, finally to 224 °C at the rate of 2 °C/min, which was held for an additional 0.2 min (total program time, 16.5 min). This is the optimum temperature programming for the conditions of both separation effect and run time. Mass spectrometric conditions were as follows: ionization mode: EI; electron energy 70 eV; temperatures of injector, ion-source and detector at 220, 250, and 150 °C, respectively. Solvent cut time was 3 min. Splitting ratio was 20:1. Selected ion monitoring (SIM) mode: m/z 55, 67, 74 and 79.

2.4. Identification of Fatty Acid Methyl Esters (FAME). Identification of fatty acids in SIM mode was proposed in our pervious study.[19] The procedures mainly include (1) an automatic mass spectral search was conducted to identify straight saturated FAMEs and (2) the retention times of straight saturated FAMEs were used to calculate the ECLs of unsaturated FAMEs according to eq 1. If the retention times of interest were out of straight saturated FAMEs, extrapolation will be employed using retention times of two nearest straight saturated FAMEs; and (3) ECL of FAMEs in the sample were compared with those in freely available databases to identify the unsaturated FAMEs.[12,13]

$$\mathrm{ECL}(x) = n + \frac{RT(x) - RT(n)}{RT(n+1) - RT(n)} \quad (1)$$

where n and $n + 1$ represent the numbers of carbons in the straight saturated FAME eluted immediately before and after the compound of interest, respectively.

Fatty acid percentage composition (percentage of peak area) was employed as quantitative results for edible oils.

2.5. Multivariate Analysis. Data matrix includes the relative content of fatty acids in edible oils. Data simulation of adulterated oils was conducted in Matlab 2011a for windows (The Mathworks, Natick, MA). Since the chemical properties of fatty acids and triacylglycerols are relatively stable, the fatty acid composition of blended oil equals to the summary of fatty acid composition of individual oils. To establish more precise adulteration detection model for five kinds of edible oils, two types of spurious edible oils were simulated by Monte Carlo method, including blending and adulterated oils. In this study, we defined the blended oil to mixing oil with random proportions of vegetable oils, while the adulterated oil was termed as spurious edible oil adulterated with low proportions of one or more than one vegetable oils. At first, 300 blended oils were simulated by blending five oils with random mixing proportions. In more details, for a blended oil, (1) random mixing proportions for five types of edible oils are created and then the sum of these proportions is normalized to unity; (2) one real edible oil of each type of edible oils is picked out at random; (3) fatty acid composition of this blended oil is calculated by weighted sum of five selected real oils; (4) repeat 1−3 for 300 times. For adulterated oils, one proportion of each kind of edible is set to 0.9 and other four proportions are created and the sum of these proportions is normalized to 0.1. Other procedures are the same with blended oils. For example, we could obtain an adulterated sesame oil sample by adding 10% content of soybean, peanut, sunflower, and rapeseed with random mixing proportions to 90% of sesame oil.

The data matrix was preprocessed through generalized log transformation and Pareto scaling (mean-centered and divided by the square root of standard deviation of each variable). As exploratory data analysis, principal component analysis (PCA) and hierarchical clustering analysis were employed to screen cluster of sampling and variable distribution in five groups. Then, the classification model for five kinds of edible oils was built by Random Forests. At last, the classification model for five types of edible oils and their blended and adulterated oils was also established by Random Forests. The model validation was conducted by OOB (out-of-bag) error in 3-fold cross validation.

Programs of automatic search tool for straight saturated FAMEs, and calculation of theoretical and experimental ECL were coded in Matlab 2011a for windows (The Mathworks, Natick, MA). Data preprocessing (Pareto scaling), Clustering (PCA and hierarchical clustering analysis) and classification (RF) were conducted by metabolomic data analysis tool of MetaboAnalyst 2.0.[20,21]

3. RESULTS AND DISCUSSION

3.1. Identification and Quantitation of Fatty Acids in Five Types of Edible Oils. In this study, fatty acidomics developed by GC/MS in SIM mode for classification and adulteration identification of edible vegetable oils. To detect more fatty acids in edible oils than the existing studies, selected ion monitoring mode (SIM) was used. The identification of FAMEs in samples was conducted using selected ions and retention indices (ECL).[13,19,22] At first, saturated FAMEs were automatically identified from background subtracted GC-MS data if the base peak of mass spectrum was at m/z 74. Totally,

dx.doi.org/10.1021/jf501097c | J. Agric. Food Chem. XXXX, XXX, XXX−XXX

Table 1. Identification and Relative Contents of Fatty Acids in Edible Vegetable Oils

fatty acids[a]	retention times	ECL	ECL in database[b]	soybean oil	peanut oil	sunflower oil	rapeseed oil	sesame oil
12:0	5.75	12.00	12.00	0.0034 ± 0.0005	0.0021 ± 0.0012	0.0017 ± 0.0006	0.0134 ± 0.0024	0.0013 ± 0.0005
14:0	7.48	14.00	14.00	0.0566 ± 0.005	0.0303 ± 0.0061	0.0432 ± 0.006	0.049 ± 0.0045	0.0141 ± 0.0019
15:0	8.32	15.00	15.00	0.0106 ± 0.0009	0.0071 ± 0.0009	0.0104 ± 0.0012	0.0197 ± 0.0031	0.0014 ± 0.0002
15:1 n-5c	8.66	15.41	15.42	0.0066 ± 0.0019	0.0036 ± 0.0033	0.0034 ± 0.001	0.0052 ± 0.0011	0.0038 ± 0.0024
16:0	9.14	16.00	16.00	12.5 ± 0.829	14.3 ± 0.703	5.93 ± 0.378	4.50 ± 0.430	10.5 ± 0.582
16:1 n-9c	9.31	16.22	16.22	0.0074 ± 0.002	0.0214 ± 0.0026	0.0081 ± 0.0012	0.0216 ± 0.0035	0.0153 ± 0.0016
16:1 n-7c	9.37	16.30	16.33	0.0352 ± 0.0053	0.0387 ± 0.0074	0.0286 ± 0.0079	0.1071 ± 0.0094	0.0621 ± 0.007
16:1 n-5c	9.47	16.44	**16.38**	0.0089 ± 0.0012	0.0071 ± 0.0013	0.0022 ± 0.0005	0.0111 ± 0.0015	0.0049 ± 0.0008
16:2 n-5c	9.68	16.70	**16.68**	0.0017 ± 0.0003	0.0013 ± 0.0003	0.0012 ± 0.0002	0.0459 ± 0.0059	0.0042 ± 0.0006
17:0	9.91	17.00	17.00	0.0739 ± 0.0063	0.063 ± 0.0107	0.0345 ± 0.0056	0.0291 ± 0.0037	0.0341 ± 0.0072
17:1 n-7c	10.13	17.28	17.34	0.0221 ± 0.002	0.0183 ± 0.0045	0.0117 ± 0.003	0.0733 ± 0.0175	0.0119 ± 0.0026
16:3 n-3c	10.54	17.88		0.0201 ± 0.0032	0.0025 ± 0.0006	0.0079 ± 0.0021	0.0018 ± 0.0005	0.0033 ± 0.0009
18:0	10.67	18.00	18.00	4.32 ± 0.576	4.21 ± 0.5298	5.14 ± 0.705	2.34 ± 0.333	5.58 ± 0.503
18:1 n-9c	10.87	18.29	18.27	16.2 ± 1.72	39.7 ± 3.15	16.4 ± 4.42	57.8 ± 10.1	33.2 ± 1.42
18:1 n-7c	10.91	18.34	18.36	0.887 ± 0.0868	0.347 ± 0.132	0.310 ± 0.0332	2.17 ± 0.354	0.636 ± 0.0564
18:1 n-x	11.15	18.68		0.0281 ± 0.0037	0.0413 ± 0.0078	0.0289 ± 0.0063	0.0384 ± 0.007	0.0366 ± 0.005
18:2 n-6c	11.23	18.80	18.76	55.6 ± 2.14	35.9 ± 3.17	70.9 ± 4.34	17.6 ± 2.16	48.9 ± 1.84
18:3 n-6c	11.50	19.18	19.09	0.0702 ± 0.0099	0.0258 ± 0.004	0.0515 ± 0.0071	0.034 ± 0.0028	0.0278 ± 0.003
19:1 n-8	11.54	19.24	**19.22**	0.0045 ± 0.0014	0.0083 ± 0.0028	0.0058 ± 0.0019	0.0098 ± 0.0027	0.0038 ± 0.001
18:3 n-3c	11.66	19.41	19.44	9.39 ± 1.37	0.0266 ± 0.0089	0.0547 ± 0.0105	7.81 ± 0.669	0.226 ± 0.0229
20:0	12.08	20.00	20.00	0.286 ± 0.0385	1.61 ± 0.158	0.241 ± 0.0207	0.5476 ± 0.1379	0.519 ± 0.0401
20:1 n-9c	12.28	20.27	20.28	0.0772 ± 0.0093	0.466 ± 0.106	0.0596 ± 0.0062	2.47 ± 3.38	0.0826 ± 0.0042
20:2 n-6c	12.69	20.81	20.84	0.0148 ± 0.0012	0.0096 ± 0.0024	0.0039 ± 0.0008	0.0653 ± 0.068	0.0097 ± 0.0017
21:0	12.84	21.00	21.00	0.0163 ± 0.0038	0.0122 ± 0.0014	0.0051 ± 0.0007	0.0104 ± 0.0024	0.0048 ± 0.0006
22:0	13.69	22.00	22.00	0.239 ± 0.0365	2.27 ± 0.402	0.582 ± 0.0587	0.163 ± 0.0724	0.0822 ± 0.0066
22:1 n-9c	13.96	22.29	22.30	0.0033 ± 0.0027	0.0184 ± 0.0105	0.001 ± 0.0014	4.042 ± 8.43	ND[c]
23:0	14.63	23.00	23.00	0.0150 ± 0.0012	0.019 ± 0.0028	0.0145 ± 0.0024	0.0111 ± 0.003	0.0086 ± 0.0008
24:0	15.69	24.00	24.00	0.0444 ± 0.0062	0.826 ± 0.139	0.107 ± 0.0137	0.0698 ± 0.0299	0.0336 ± 0.0049

[a]Shorthand annotation according to *A*:*B* n-*i* z, where *A* is the number of carbon atoms in the fatty acid chain, *B* is the number of double bonds, *n*-*i* is the location of the double bond on the *i*th carbon−carbon bond, counting from the terminal methyl carbon toward the carbonyl carbon and z is the geometrical configuration expressed as c for cis and t for trans.[12] [b]Normal: ECL values in database in ref 13. Bold: ECL values in Chrombox database. Possible structure is deduced by mass spectral characteristics. [c]Not detected.

11 saturated FAMEs, 12:0, 14:0, 15:0, 16:0, 17:0, 18:0, 20:0, 21:0, 22:0, 23:0, and 24:0 included, were identified. Their retention times and relative abundances (RA) at *m*/*z* 55, 67, and 79 were then employed to validate whether saturated FAMEs are straight chain. The retention times of 11 straight saturated FAMEs were used to calculate the ECL values for other FAMEs. Identification of unsaturated FAMEs in the edible oils was conducted by searching their ECL values in the customized database constructed in the previous studies[13,14] for the FAME with the nearest ECL value as the candidate.

The threshold of difference between the experimental ECL and reference ECL in the database is set to 0.1. Using this method, as shown in Table 1, 17 unsaturated FAMEs were identified. Among them, 12 unsaturated FAMEs were validated by comparing the retention time and mass spectra with the authentic standards, while 3 unsaturated FAMEs were exploratively identified by searching freely available Chrombox database.[12] The other 2 unsaturated FAMEs were inferred to the type using their mass spectral characteristics. Totally, 28 FAMEs were identified and quantified, which are significantly higher than full scan mode (23 FAMEs).[5] Fatty acid composition of vegetable oils is described by percentage content of total amount of all fatty acids and shown in Table 1.

3.2. Exploratory Data Analysis. After determination and quantification of fatty acids in five vegetable oils, the data matrix of relative contents was preprocessed through generalized log transformation and Pareto scaling. In exploratory data analysis, principal component analysis (PCA) and hierarchical clustering analysis (HCA) were employed to screen cluster of sampling and fatty acid distribution in five groups. From the score plot obtained from PCA in Figure 1a, five kinds of edible oils were clearly classified into five groups, among which rapeseed and peanut oils are far from sunflower, sesame and soybean oils, respectively.

To investigate variable distribution in five groups, heat map of fatty acid profiles of five vegetable oils was illustrated. In heat map, the Euclidean distance was used for measuring the similarities using Wards' linkage by clustering to minimize the sum of squares of any two clusters. As shown in Figure 1b, the similar cluster analysis results were obtained as PCA. More importantly, we can find the fatty acids distribution in five groups from this heat map as follows: (a) soybean oils, relatively high content of 16:3 n-3, 17:0, and 18:3 n-6c (γ linolenic acid); (b) peanut oils, relatively high content of 20:0, 22:0, and 24:0; (c) rapeseed oils, relatively high content of 12:0, 16:1 n-5, 16:1 n-7, 17:1 n-7, 16:2 n-5, 18:1 n-7, 20:1 n-9, and 22:1 n-9; and (d) sunflower oils, the highest content of 18:2 n-6 (linoleic acid).

3.3. Classification of Five Types of Edible Oils by Random Forests. After exploratory data analysis, we can find that five kinds of vegetable oils could be clearly classified into five groups. To build classification model for five types of edible oils, an effective supervised multivariate statistical method of random forests (RF) was used. Random forests are a

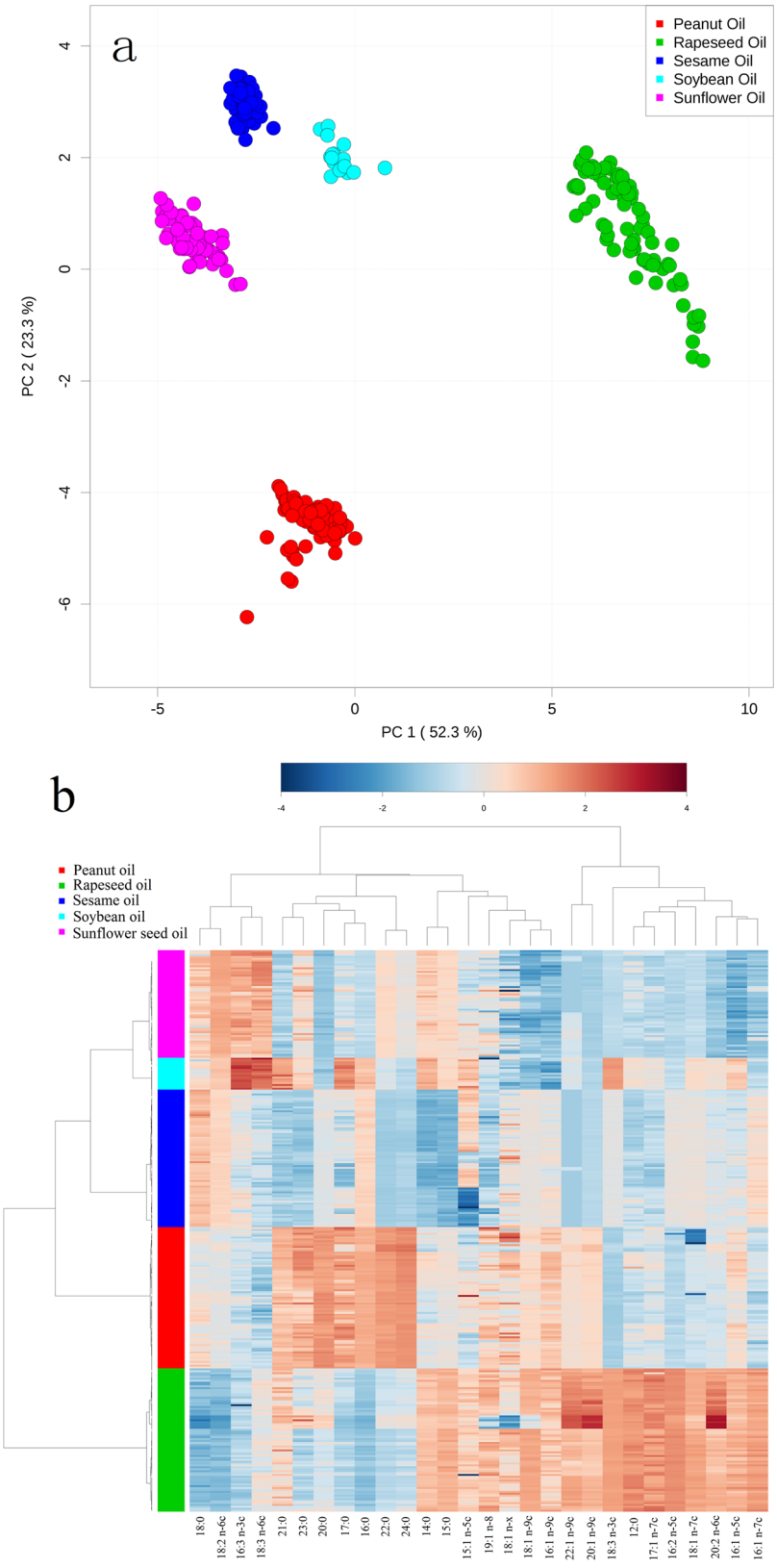

Figure 1. (a) Score plot between the first 2 PCs. (b) Heat map of five kinds of edible oils.

combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest.[23] The sample proximity matrix derived from these training trees is generated to collect the similarity information on the samples for classification of samples. 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), random forests could be directly employed to

dx.doi.org/10.1021/jf501097c | J. Agric. Food Chem. XXXX, XXX, XXX−XXX

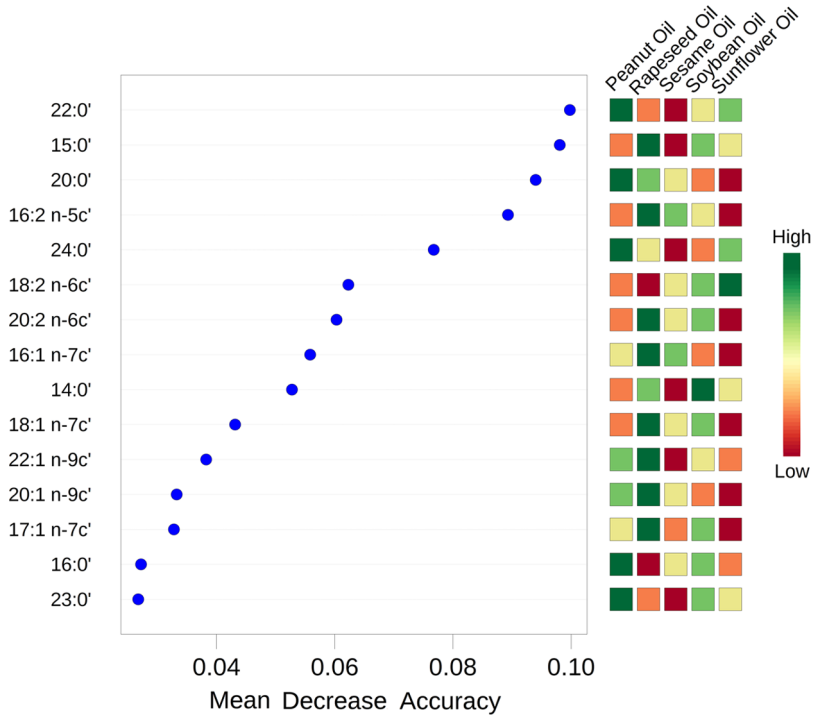

Figure 2. Significant features of five kinds of edible oils identified by Random Forest.

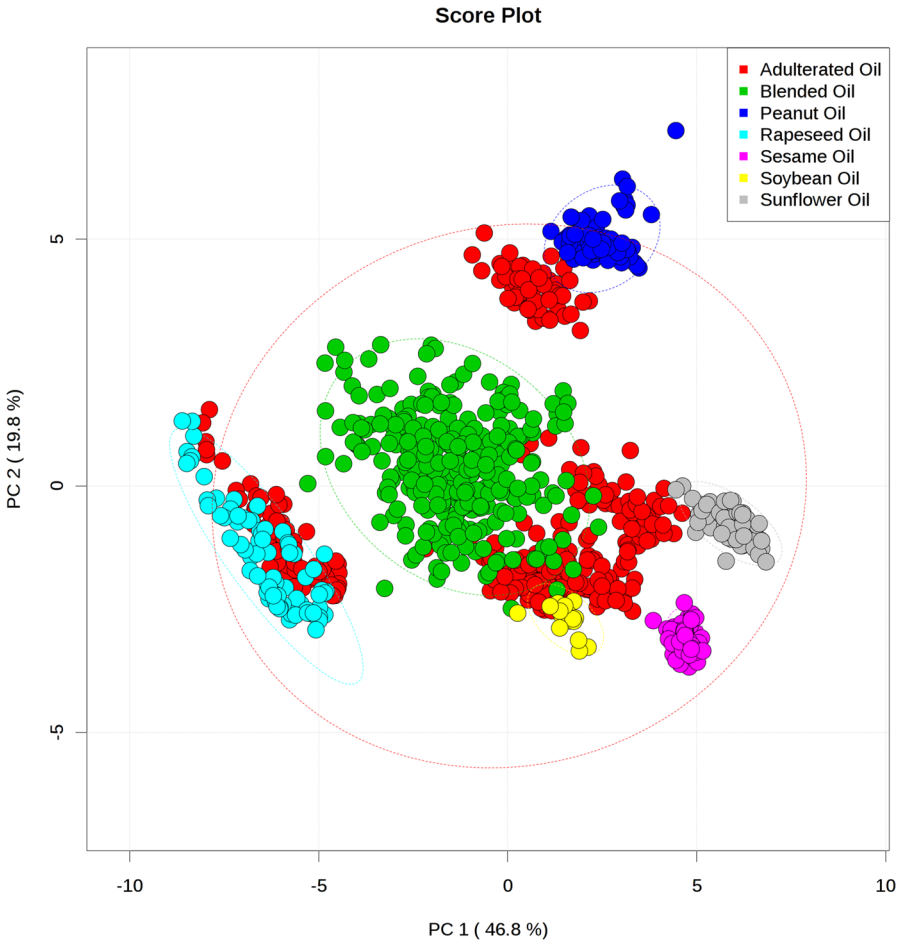

Figure 3. Score plot between the first 2 PCs for five kinds of edible oils and their blended/adulterated oils.

dx.doi.org/10.1021/jf501097c | J. Agric. Food Chem. XXXX, XXX, XXX–XXX

Table 2. Random Forest Classification Performance

	adulterated oil	blended oil	peanut oil	rapeseed oil	sesame oil	soybean oil	sunflower oil	class error
adulterated oil	299	0	0	1	0	0	0	0.00
blended oil	0	300	0	0	0	0	0	0.00
peanut oil	0	0	75	0	0	0	0	0.00
rapeseed Oil	1	0	0	75	0	0	0	0.01
sesame oil	0	0	0	0	73	0	0	0.00
soybean oil	3	0	0	0	0	14	0	0.18
sunflower oil	0	0	0	0	0	0	57	0.00

multiclass classification. Furthermore, the RF classifier needs optimize only one parameter of the number of classification trees, which is relatively insensitive to predictive effect. In study, the number of classification trees is set to 500. During tree construction, about one-third of the samples are left out of the bootstrap sample. This out-of-bag (OOB) data is then used as test sample to obtain an unbiased estimate of the classification error (OOB error). From Figure 2a, it is found that the five errors decrease to zero after less than 20 trees. OOB error equals to 0 in the final classification model.

Random forests could provide the measure for variable importance. Figure 2b shows the contribution of each variable to the classification. According to the mean decrease, all important fatty acids except 18:2 n-6 (linoleic acid) and 16:0 (palmitic acid) are of low content. These results indicate that low abundant fatty acids might be more important to classification of five types of edible oils. Therefore, fatty acid profiles could reflect more information on fatty acids in edible oils, which benefit classification and adulteration identification.

3.4. Adulteration Identification by Fatty Acid Profiles. From the above analysis, it seems it is very easy to classify these five kinds of edible oils. Indeed, classification of commonly used edible oils is not hard thing by fatty acid analysis. However, classification of edible oils is not enough to ensure identity of edible oils. The most important thing is to identify adulteration of expensive oils with low price oils. Therefore, to test whether fatty acid profiles could serve adulteration detection, 300 blended oils were simulated by mixing five oils with random proportions, while, 60 adulterated oils were simulated by blending 10% content of other four oils with random proportions to 90% soybean, peanut, sunflower, rapeseed, and sesame oil, respectively. Then, the discriminative model for five pure oils and their blend and adulterated oils was built by RF after generalized log transformation and Pareto scaling. From Figure 3a, blend oils appear in the center of PCA score plot. It is found that fatty acid profiles could detect adulteration of peanut, sunflower and sesame oils with other vegetable oils about the level of 10%. For soybean and rapeseed oils, pure and adulterated oils are heavily overlapped.

Therefore, RF has been used to obtain high effective model. As shown in Figure 3b, except soybean oils, the OOB errors decrease to 0 or near 0 ($\leq$0.01) and the total OOB error is 0.00557. As illustrated in Table 2, peanut, sesame, and sunflower oils could completely separate from other oils and their blend and adulterated oils, respectively. Among them, peanut and sesame are key target of oil fraud in China market. Moreover, only one adulterated oil and one rapeseed oil were misidentified by RF model. The main reason is that fatty acids possess broad range, such as high and low erucic acid rapeseed. Thus, the rapeseed oil with 10% adulterated with other oils still stay at the space of rapeseed oils. However, only one rapeseed and one adulterated oils were misidentified. Finally, 3 out of 14 soybean oils were misidentified to be blend oils, suggesting more other parameter like phytosterol should be employed to assess the authenticity of soybean oils.[24] Fortunately, the soybean oil is the cheapest one in five types of edible oils, which has low risk to be adulterated.

From the above analysis, the established RF model could serve the authenticity assessment for four expensive oils. To validate the RF model, 16 real adulterated oils were physically blended. The blend proportions of these 16 adulterated oils were shown in Supporting Information Table S2. The results indicate that all of 16 blend oils could be identified as adulterated oil. Therefore, fatty acid profiles could be a good strategy to identify adulterated oils for edible oils. Compared with the pervious studies,[5,6,14] the adulteration detection method possess the following advantages: (1) it can simultaneously detect adulteration of more than one edible oils; (2) it can also detect the adulteration with more than one edible oils.

In conclusion, fatty acid profiles of five kinds of edible oils were established by GC-MS in SIM mode and employed to classify five types of edible oils with the help of multivariate statistical methods. The results indicate that fatty acid profiles of edible oils could completely classify the five kinds of edible oils into five groups and are therefore taken as markers of these four edible oils. Moreover, simulated data test indicate that fatty acid profiles could be used to detect adulterated by 10% or little more than 10% other oils for peanut, sunflower, rapeseed and sesame oil. The strategy of fatty acid profiles might be also useful to simultaneously identify the origin of various edible oils and detect suspected economically motivated the adulteration with more than one vegetable oils.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Additional information on detailed information on five types oil seeds (Table S1) and the blend proportions of 20 real adulterated oils (Table S2), cumulative error rates by Random Forest classification for five types of edible oils (Figure S1), five types of edible oils and their blended/adulterated oils (Figure S2). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*Phone: +86 27 86812943. E-mail: liangxiao_zhang@hotmail.com.

*E-mail: peiwuli@oilcrops.cn.

Author Contributions

L.Z. and P.L. contributed equally to this study.

Funding

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

Project for Agro-product Quality & Safety Risk Assessment (GJFP2014006), the National Natural Science Foundation of China (21205118 and 31201447), the Special Fund for "Quality Inspection" Research in the Public Interest (2012104010-4), and the earmarked fund for China Agriculture research system (CARS-13).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank two anonymous reviewers for their insightful and detailed comments.

■ REFERENCES

(1) Cert, A.; Moreda, W.; Pérez, C. M. C. Chromatographic analysis of minor constituents in vegetable oils. *J. Chromatogr. A* **2000**, *881*, 131−148.
(2) Zhao, F.; Liu, J.; Wang, X.; Li, P.; Zhang, W.; Zhang, Q. Detection of adulteration of sesame and peanut oils via volatiles by GC×GC−TOF/MS coupled with principal components analysis and cluster analysis. *Eur. J. Lipid Sci. Technol.* **2013**, *115*, 337−347.
(3) Kabagambe, E. K.; Baylin, A.; Ascherio, A.; Campos, H. The Type of Oil Used for Cooking Is Associated with the Risk of Nonfatal Acute Myocardial Infarction in Costa Rica. *Nutr. Epidemiol.* **2005**, *135*, 2674−2679.
(4) Jeffrey, C. M.; Spink, J.; Lipp, M. Development and application of a database of food ingredient fraud and economically motivated adulteration from 1980 to 2010. *J. Food Sci.* **2012**, *77*, 118−126.
(5) Ai, F. F.; Bin, J.; Zhang, Z. M.; Huang, J. H.; Wang, J. B.; Liang, Y. Z.; Yu, L.; Yang, Z. Y. Application of random forests to select premium quality vegetable oils by their fatty acid composition. *Food Chem.* **2014**, *143*, 472−478.
(6) Yang, Y.; Ferro, M. D.; Cavaco, I.; Liang, Y. Z. Detection and identification of extra virgin olive oil adulteration by GC-MS combined with chemometrics. *J. Agric. Food Chem.* **2013**, *61*, 3693−3702.
(7) Wang, M.; Han, R. H.; Han, X. Fatty acidomics: Global analysis of lipid species containing a carboxyl group with a charge-remote fragmentation-assisted approach. *Anal. Chem.* **2013**, *85*, 9312−20.
(8) Bertrand, M. The new database seed oil fatty acids (SOFA). *Lipid Technol.* **2012**, *24*, 230−234.
(9) Drozd, J. Chemical derivatization in gas chromatography. *J. Chromatogr.* **1975**, *113*, 303−356.
(10) Miwa, T. K.; Mikolajczak, K. L.; Earle, F. R.; Wolff, L. A. Gas chromatographic characterization of fatty acids: Identification constants for mono- and dicarboxylic methyl esters. *Anal. Chem.* **1960**, *32*, 1739−1742.
(11) Härtig, C. Rapid identification of fatty acid methyl esters using a multidimensional gas chromatography-mass spectrometry database. *J. Chromatogr. A* **2008**, *1177*, 159−169.
(12) Wasta, Z.; Mjøs, S. A. A database of chromatographic properties and mass spectra of fatty acid methyl esters from omega-3 products. *J. Chromatogr. A* **2013**, *1299*, 94−102.
(13) Zhang, L. X.; Tan, B. B.; Zeng, M. M.; Lu, H. M.; Liang, Y. Z. Establishment of reliable mass spectra and retention indices library: Identification of fatty acids in human plasma without authentic standards. *Talanta* **2012**, *88*, 311−317.
(14) Diaz, T. G.; Meras, I. D.; Casas, J. S.; Franco, M. F. A. Characterization of virgin olive oils according to its triglycerides and sterols composition by chemometric methods. *Food Control* **2005**, *16*, 339−347.
(15) Li, C.; Yao, Y.; Zhao, G.; Cheng, W.; Liu, H.; Liu, C.; Shi, Z.; Chen, Y.; Wang, S. Comparison and Analysis of Fatty Acids, Sterols, and Tocopherols in Eight Vegetable Oils. *J. Agric. Food Chem.* **2011**, *59*, 12493−12498.
(16) Ruth, V. S. M.; Villegas, B.; Akkermans, W.; Rozijn, M.; Kamp, V. D. H.; Koot, A. Prediction of the identity of fats and oils by their fatty acid, triacylglycerol and volatile compositions using PLS-DA. *Food Chem.* **2010**, *118*, 948−955.
(17) Archer, K. J.; Kirnes, R. V. Empirical characterization of random forest variable importance measures. *Comput. Stat. Data Anal.* **2008**, *52*, 2249−2260.
(18) Rodriguez, G. V. F.; Ghimire, B.; Rogan, J.; Chica, O. M.; Rigol, S. J. P. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sens.* **2012**, *67*, 93−104.
(19) Zhang, L. X.; Li, P. W.; Sun, X. M.; Hu, W.; Wang, X. P.; Zhang, Q.; Ding, X. X. Untargeted fatty acid profiles based on the selected ion monitoring mode. *Anal. Chim. Acta* **2014**, *839*, 44−50.
(20) Xia, J.; Mandal, R.; Sinelnikov, I. V.; Broadhurst, D.; Wishart, D. S. MetaboAnalyst 2.0—A comprehensive server for metabolomic data analysis. *Nucleic Acids Res.* **2012**, *40*, 127−133.
(21) Xia, J.; Psychogios, N.; Young, N.; Wishart, D. S. MetaboAnalyst: A web server for metabolomic data analysis and interpretation. *Nucleic Acids Res.* **2009**, *37*, 652−660.
(22) Zhang, L. X.; Ji, B. B.; Tan, B. B.; Liang, Y. Z.; Liang, N. N.; Wang, X. L.; Dai, H. Identification of the composition of fatty acids in *Eucommia ulmoides* seed oil by fraction chain length and mass spectrometry. *Food Chem.* **2010**, *121*, 815−819.
(23) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5−32.
(24) Xu, B. C.; Zhang, L. X.; Wang, H.; Luo, D. L.; Li, P. W. Characterization and authentication of four important edible oils using free phytosterol profiles established by GC-GC-TOF/MS. *Anal. Methods* **2014**, *6*, 6860−6870.