https://doi.org/10.1016/j.foodchem.2018.01.125

Establishment of a Spectral Database for Classification of Edible Oils using Matrix-assisted Laser Desorption/Ionization Mass Spectrometry

Tsz-Tsun Ng[a,b], Suying Li[a,b], Cheuk-Chi A. Ng[a,b], Pui-Kin So[a,b], Tsz-Fung Wong[a], Zhen-Yan Li[a], Shu-Ting Chan[a] and Zhong-Ping Yao[a,b,c,*]

[a]State Key Laboratory of Chirosciences, Food Safety and Technology Research Centre and Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong Special Administrative Region, China

[b]State Key Laboratory of Chinese Medicine and Molecular Pharmacology (Incubation) and Shenzhen Key Laboratory of Food Biological Safety Control, Shenzhen Research Institute of Hong Kong Polytechnic University, Shenzhen 518057, China

[c]Key Laboratory of Natural Resources of Changbai Mountain and Functional Molecules (Yanbian University), Ministry of Education, Yanji, Jilin 133002, China

*Corresponding author:
Dr. Zhong-Ping Yao
Department of Applied Biology and Chemical Technology, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
Tel: (852) 34008792. Fax: (852)23649932.
E-mail: zhongping.yao@polyu.edu.hk

Abstract

In this study, we aim to establish a comprehensive spectral database for analysis of edible oils using matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS). More than 900 edible oil samples, including 30 types of edible oils, were analyzed and compared, and the characteristic peaks and spectral features of each edible oil were obtained. Edible oils were divided into eight groups based on their characteristic spectral patterns and principal component analysis results. An overall correct rate of 97.2% (98.1% for testing set) was obtained for classification of 435 edible oil products using partial least square-discriminant analysis, with nearly 100% correct rate for commonly used edible oils. Differentiation of counterfeit edible oils, repeatedly cooked edible oils and gutter oils from normal edible oils could also be achieved based on the MALDI-MS spectra. The establishment of this spectral database provides reference spectra for spectral comparison and allows rapid classification of edible oils by MALDI-MS.

Keywords: Edible oils; gutter oils; MALDI-MS; classification; characteristic peaks; principal component analysis; partial least square-discriminant analysis.

1. Introduction

Edible oils are one of the major sources of fat for human being. Most of the edible oils contain two essential fatty acids, omega-3 and omega-6 fatty acids, which cannot be synthesized by human body and must be absorbed to maintain human health (Gunstone, 2004). Edible oils are thus essential to our daily life. However, the quality and safety of edible oil products in the market are not secured with the widespread of edible oil frauds in recent years (Hong et al., 2017; Peng et al., 2017). Edible oil frauds typically include edible oil counterfeit, which involves using cheaper edible oils to replace more expensive ones (Hong, et al., 2017; Zhu, Wang, & Chen, 2017), edible oil adulteration, which involves mixing edible oils with cheaper ones (Calvano, De Ceglie, D'Accolti, & Zambonin, 2012; Jimenez-Carvelo, Gonzalez-Casado, & Cuadros-Rodriguez, 2017), and the use of gutter oils (i.e., recycled cooking oils), which involves refining wasted or used cooking oils and labeling them as normal edible oils (Peng et al., 2017). Emergence of gutter oils that have potential health problems has raised many public concerns. However, identification of gutter oils is a challenging analytical task, since gutter oils have very diverse preparation processes and sources and it may not be effective to rely on detection of targeted toxic substances (Centre for Food Safety Hong Kong, 2013) or specific food marker compounds (Jin et al., 2013) for determination of gutter oils.

Current practice relies on standard gas chromatography-flame ionization detection (GC-FID) method (International Organization for Standardization, 2014) for analysis of edible oils. In this method, the fatty acid content in an edible oil sample is determined using the GC-FID-based method and matched with the Codex standard (Codex Alimentarius, 2015a, 2015b). The disadvantages of this conventional technique, such as the need of chemical derivatization and time-consuming column separation (Cozzolino & Giulio, 2011), induced the development of simple and rapid analytical techniques to meet the high analytical demand of edible oils. Various techniques, such as Raman spectroscopy (Li et al., 2017) laser-induced fluorescence (Mu et al., 2016), infrared spectroscopy (Rohman, 2016), electronic nose (Xu, Yu, Liu, & Zhang, 2016), low field nuclear magnetic resonance (Zhu, et al., 2017), ion mobility spectrometry (L. Zhang et al., 2016), and mass spectrometric techniques, including direct infusion electrospray ionization (Catharino et al., 2005), direct analysis in real time (DART) (Vaclavik, Cajka, Hrbek, & Hajslova, 2009) and matrix-assisted laser desorption/ionization (MALDI) (Ayorinde, Elhilo, & Hlongwane, 1999; Jakab, Nagy, Heberger, Vekey, & Forgacs, 2002; Lay, Liyanage, Durham, & Brooks, 2006; Ng, So, Zheng, & Yao, 2015), have been employed for analysis of edible oils. In particular, MALDI-MS has distinctive features over the other techniques, such as short analysis time, high sensitivity and high tolerance to impurities. In MALDI-MS, several hundreds of samples can be loaded onto different spots of

one target plate and introduced into the mass spectrometer for analysis, and spectrum for each sample can be obtained within seconds. Very recently, a simple protocol was developed to allow direct analysis of edible oils, with no memory effect and clear mass spectra for spectral comparison of edible oils (Ng et al., 2015). Studies have proved that MALDI-MS spectra of different edible oil species, especially the triacylglycerides (TAGs) region, were specific and could be used for characterization of edible oils (Lay et al., 2006; Ng et al., 2015; Picariello, Paduano, Sacchi, & Addeo, 2009; Picariello, Sacchi, & Addeo, 2007). Recent studies also demonstrated differentiation of geographical origins of olive oils and detection of olive oil adulteration using MALDI-MS and statistical analysis (Peršurić, Osuga, Galinac Grbac, Peter-Katalinić, & Kraljević Pavelić, 2017; Jergović, Peršurić, Saftić, & Kraljević Pavelić, 2017). Ayorinde and co-workers calculated and compared the fatty acid contents of different edible oils based on their MALDI-MS spectra and GC chromatograms, and concluded that the results obtained from the MALDI-MS and GC methods agreed with each other (Ayorinde, Garvin, & Saeed, 2000; Hlongwane, Delves, Wan, & Ayorinde, 2001).

Although MALDI-MS has been extensively used to analyze edible oils (Ayorinde, Elhilo et al., 1999; Lisa, Holcapek, & Bohac, 2009; Picariello et al., 2007; Picariello et al., 2009; Q. Zhang et al., 2016), so far there has been no systematic investigation and establishment of a

97 comprehensive MALDI spectral database of various edible oils, which will definitely benefit
98 the technique and the analysis of edible oils. In this study, more than nine hundred edible oil
99 samples were collected and analyzed using the MALDI-MS protocol we developed previously
100 (Ng et al., 2015). The MALDI-MS spectra obtained were stored in an in-house developed
101 system, and a comprehensive MALDI-MS spectral database including more than thirty edible
102 oil species was then established to allow rapid classification of edible oils by comparing the
103 MALDI-MS spectra of the samples with the reference spectra in the database.
104

105 ## 2. Experimental section

106 ### 2.1 Chemicals and materials

107 Thirty types of edible oil products were collected from collaborators and authentic stores in
108 Hong Kong, mainland China and Taiwan, with origins from various countries or regions (see
109 Table S-1 for the complete list of the oil samples and Table S-2 for their origins). Three bottles
110 were collected for each commercially available oil product. Standards of olive oil, castor oil,
111 corn oil, soybean oil, sesame oil, coconut oil, peanut oil, linseed oil, palm oil, canola oil,
112 cottonseed oil, sunflower seed oil, safflower oil and lard oil were purchased from Supelco and
113 Sigma-Aldrich (St. Louis, USA). The heated oil samples were prepared in the laboratory. The
114 gutter oil samples were provided by Syscan Technology Holdings Limited (Wuhan, China).

All samples were sealed and stored in a dry and dark environment before analysis. MALDI matrices 2, 5-dihydroxybenzoic acid (DHB) and α-cyano-4-hydroxycinnanic acid (CHCA) were purchased from Aldrich (St. Louis, MO, USA). HPLC grade acetone and HPLC grade tert-butyl methyl ether (TBME) were purchased from Acros Organic (Waltham, MA, USA). HPLC grade acetonitrile (ACN) was purchased from Anaqua Chemical Supply (Houston, TX, USA) and HPLC grade dichloromethane (DCM) was purchased from RCI Labscan (Bangkok, Thailand). Polyethylene glycol standards were purchased from Fluka (St. Louis, MO, USA). Sodium iodide (NaI) was purchased from Panreac Química (Barcelona, Spain) and sodium chloride (NaCl) was purchased from Riedel-de Haën (Seelze, Germany). Derivatization reagent, trimethylsulfonium hydroxide (TMSH, 0.25 M solution in methanol) was purchased from Acros Organic, and a mixture of 37 fatty acid methyl standards (FAME) was purchased from Supelco. All chemicals were used directly without further purification.

2.2 MALDI-MS analysis

The edible oil samples were analyzed using the protocol we reported previously (Ng et al., 2015) with slight modification. Briefly, each edible oil sample was directly transferred by a medical cotton tip onto the spot on the MALDI plate, which was pre-deposited with DHB matrix layer. The MALDI plate was then inserted into an ultrafleXtreme MALDI-TOF/TOF

133 mass spectrometer (Bruker, Billerica, MA, USA) equipped with a 355 nm smartbeam-II laser
134 for the MALDI-MS analysis. The mass spectrometer was operated at a frequency of 2000 Hz
135 and in positive and reflectron mode. The matrix suppression cut-off was set to *m/z* 500 and the
136 spectra were acquired with an *m/z* range of 500-2000 Da. The ion pulse excitation was set to
137 140 ns. The ion source voltage 1, ion source voltage 2, lens voltage, reflector voltage 1 and
138 reflector voltage 2 were set to 20 kV, 17.75 kV, 7 kV, 21.1 kV and 10.85 kV, respectively. The
139 reflector base detector voltage was set to 2.32 kV and the sample rate was set to 5.0 GS/s. The
140 mass spectrometer was calibrated with the PEG solution mixture
141 (PEG600/PEG1000/PEG2000/NaI = 1/2/2/5 (v/v)). Each mass spectrum was acquired
142 manually by irradiating at least 5 random positions (1000 laser pulse per one position) on the
143 sample spot. The sub-spectra were accepted if the signals in the TAGs region (typically at the
144 region of *m/z* 570-750 or *m/z* 850-920) were observed and the absolute intensity of the top peak
145 of the final spectra obtained by combining at least 5 positive sub-spectra reached $1x10^4$. The
146 mass spectra were processed in flexAnalysis 1.4 program (Bruker, Billerica, MA, USA) under
147 "centroid" peak detection algorithm, "SavitzkyGolay" peak smoothing algorithm (the width
148 equal to 0.2 *m/z* and the cycles equal to 1) and baseline subtraction algorithm "TopHat". Each
149 edible oil sample was analyzed in triplicate. The sodium adducted TAGs were observed in the
150 MALDI-MS mass spectra and their identification were based on the results of the previous

151 studies (Ayorinde, Elhilo et al., 1999; Lay et al., 2006; Picariello et al., 2009) and further
152 confirmed with MS/MS analysis in this study (see the Supporting Information). For the TAGs
153 in rapeseed oil, the peak assignments were based on the theoretical calculation of fatty acid
154 contents from the GC-FID results.
155
156 **2.3 Statistical analysis**
157 The data obtained from three samples of the same edible oil products were averaged. The
158 normalized intensities of monoisotopic peaks of TAGs and cyclolinopeptides in the spectra
159 (absolute intensity of the peak observed / total absolute intensity of all the peaks observed in
160 the mass spectrum) were input into the statistics software (Umetrics Simca 14.0, Andover, MA,
161 USA) for principal component analysis (PCA) and partial least square-discriminant analysis
162 (PLS-DA). The PCA and PLS-DA models were established using “simple mode” and “autofit”
163 (automatic cross-validation) with the default settings. In the PLS-DA, 198 edible oil products,
164 including the products with different brands, collected from different locations and possessed
165 different TAGs patterns, were selected as the training set for establishment of the PLS-DA
166 models. One “grouping” model using the groups as the dummy Y variables and seven sub-
167 models with the species as the dummy Y variables were established. The remaining products
168 were used as testing set to test the performance of the established PLS-DA models.

3. Results and discussion

3.1 MALDI-MS spectra and spectral database of edible oils

The rapid and simple protocol we previously developed for MALDI-MS analysis of edible oils (Ng et al., 2015) was employed in this study. This protocol allows direct loading of edible oil samples on the predeposited DHB matrix layers that can be stable over one week, and offers mass spectra of edible oils with high quality and high reproducibility. The typical spectra of all the 30 edible oils can be found in Figure S-1. To our knowledge, MALDI-MS spectra of some edible oils, e.g., perilla oil, pine nut oil and rice bran oil, have not been reported yet. For those edible oils previously investigated by MALDI-MS, the spectra obtained in the study were consistent with those in the literatures (Ayorinde, Elhilo et al., 1999; Ayorinde, Eribo, Balan, Johnson, & Wan, 1999; Jakab et al., 2002; Lay et al., 2006; Picariello et al., 2007).

For each oil product from the market, i.e., a specific edible oil of a specific brand, typically three bottles of edible oils were collected and analyzed in this study. The mass spectra obtained from different bottles of the same oil products were highly similar to each other, with an example of the spectra for peanut oil as shown in Figure S-2. However, for the same kind of edible oils of different brands, variations were observed for the spectra obtained. For example,

as shown in Figure S-3, similar ion peaks were detected for the mass spectra obtained for the peanut oil samples from three different brands, but their intensity ratios varied, e.g., different intensity ratios between *m/z* 903.7 (LLO), *m/z* 905.8 (LOO) and *m/z* 907.8 (OOO) obtained from the three different brands of samples. This was consistent with the varied range of C18:1 (35.0-69.0%) and C18:2 (12.0-43.0%) contents of peanut oils as described in the Codex standards (Codex Alimentarius, 2015a). Variations of TAG patterns were also observed for other edible oils but at different extents. Some of the edible oils such as peanut oil and olive oil could have larger variations, while some of the edible oils such as canola oil could have smaller variations. However, as illustrated below, such variations within the same oil species were normally smaller than the differences between different oil species, allowing differentiation of different oil species.

Special consideration was given to canola oil (i.e., low erucic acid rapeseed oil) and rapeseed oil that are two closely related edible oil species. Their fatty acid contents were very similar, except that canola oil contains less than 2% C22:1 (see Figures S-1.5 and S-1.23 for their typical mass spectra). The TAGs of C22:1 were observed at *m/z* 935.8-1019.9 and their suggested compositions are listed in Table S-7. The collected rapeseed oil samples were confirmed by conventional GC-FID analysis, and it was found that out of the 21 collected

205 “rapeseed oil” products, only 6 contained more than 2% C22:1 and confirmed as rapeseed oil.
206 The remaining samples were counted as canola oil in this study.
207
208 As shown in Table S-1, more than 900 edible oil samples, including 30 species of edible oils
209 with different brands for each species, were analyzed to construct the spectral database. For the
210 collected edible oil samples that showed MALDI-MS spectra obviously different from those
211 of the same species, confirmatory analysis using the conventional GC-FID method for FAME
212 analysis was performed, and the edible oil samples with fatty acid contents failed to match the
213 Codex standards were rejected and were not included in the database. Some edible oil samples
214 with high levels of oxidized/thermal products were also rejected. The sample information (oil
215 type, origin of manufacture, origin of raw materials, and source of collection), full spectra and
216 the spectra of TAGs region of the authentic samples were then exported and stored into a
217 database on an Apache server. As shown in Figure S-4, an in-house developed user interface
218 (HTML/JavaScript at front-end and PHP at back-end) allows users to retrieve individual
219 spectra grouped by the type of oils.
220
221 **3.2 Characteristic peaks and PCA of the MALDI-MS spectra**

Spectra of different edible oils were compared and characteristic peaks have been found for various edible oils. The mass spectra of castor oil (Figure S-1.6) and fish oil (Figure S-1.10) were easily distinguished from those of other edible oils. Castor oil showed a very simple mass spectrum with a major peak at *m/z* 955.7 that corresponds to TAG of 12-OH-C18:1 (Ayorinde, Elhilo et al., 1999). Fish oil possessed a complex mass spectrum which contained abundant signals for both TAGs commonly observed in vegetable oils (i.e. TAGs from *m/z* 850 to *m/z* 920) and TAGs of C20:5 and C22:6 (Codex Alimentarius, 2014) such as *m/z* 981.8 and *m/z* 1009.8, which were only observed in fish oil. These specific MALDI-MS patterns allowed castor oil and fish oil to be easily distinguished based on their spectra.

The TAGs regions of butter (Figure S-1.3) and coconut oil (Figure S-1.7) were mainly the peaks lower than *m/z* 850 with abundant peaks such as *m/z* 661.5 and *m/z* 689.6 which corresponded to TAGs with C12:0 and C14:0, respectively (Saraiva, Cabral, Eberlin, & Catharino, 2009). These TAGs were only found in butter and coconut oil and could be used for differentiation of butter and coconut oil from other edible oils. Comparison between the spectra of butter and coconut oil showed that the spectral pattern of these two species were different. Coconut oil had higher intensities of peaks at *m/z* 605.5 and *m/z* 633.5 while butter had higher intensities of peaks at *m/z* 771.6 and *m/z* 827.6, i.e., coconut oil contained more short chain

240 fatty acids such as C12:0. Margarine is a mixture of vegetable oils but is physically similar to
241 butter. It is not easy to differentiate them by naked eyes. As shown in Figure S-5, the mass
242 spectra of margarine showed strong signals of higher mass TAGs such as *m/z* 879.8 and *m/z*
243 903.7 (i.e. higher level of C18:1 and C18:2), while the mass spectrum of butter contained
244 significant lower mass TAGs peaks as discussed previously. The remarkably different mass
245 spectra of margarine and butter allowed them to be easily distinguished from each other.
246
247 As shown in the MALDI-MS spectra of flaxseed (Figure S-1.11) and perilla oil (Figure S-1.20),
248 two specific peaks at *m/z* 895.7 (LnLnLn) and *m/z* 897.7 (LnLnL) were observed and could
249 serve as the markers of these two oils. Peak at *m/z* 899.7 for flaxseed oil showed obviously
250 higher abundance than for perilla oil, allowing differentiation between these two oils. In
251 addition, some characteristic peaks such as *m/z* 999.5 and *m/z* 1062.6 were observed only in
252 flaxseed oil (see Figure S-6a and S-6b). The peaks found in such high mass region for edible
253 oils usually corresponded to thermal/oxidation products which were formed after prolonged
254 heating (Picariello et al., 2009; Q. Zhang et al., 2016). However, accurate mass measurements,
255 database searching and MS/MS analysis (see Figure S-6c and S-6d for the results) indicated
256 that the most abundant peak (*m/z* 1062.6) in the spectrum corresponded to sodium adducted
257 ion of cyclolinopeptide A, a natural product found in flaxseed oil extract (Gaymes, Cebrat,

Siemion, & Kay, 1997), and the other peaks at *m/z* 1080.7, *m/z* 1096.7, *m/z* 999.6 and *m/z* 983.5 corresponded to the sodium adducted ions of cyclolinopeptides B, C, E and J (Stefanowicz, 2004). For the 28 flaxseed oil products collected for the analysis, the cyclolinopeptide peaks were detected in 25 products, but not in the other 3 products that were confirmed also as flaxseed oil based on their TAGs distribution. Although the cyclolinopeptides were not observed in all the flaxseed oil products, the presence of such peaks could provide additional information for characterization of flaxseed oil.

Palm oil is one of the most popular edible oil species used in the food industry (Gunstone, 2011). There are seven types of palm oil derivatives that possess different fatty acid contents (Codex Alimentarius, 2015a). However, only two types of palm oil derivatives, i.e., palm oil (Figure S-1.17) and palm superolein (Figure S-1.18), were found in the local markets. Palm oil showed higher abundances of peaks at *m/z* 855.7 (PPO) and *m/z* 881.8 (POO) than other common vegetable oils such as canola oil and olive oil. Compared with palm oil, palm superolein had weaker peak at *m/z* 855.7 but stronger peak at *m/z* 881.8, as it contains less C16:0 but more C18:1. Lard shared similar MALDI-MS spectral pattern with that of palm superolein.

Rapeseed oil (Figure S-1.23), as discussed in the previous section, shared similar MALDI-MS spectrum with canola oil, except that peaks at higher mass region such as *m/z* 991.9 and *m/z* 1019.9, which corresponded to TAGs of C20:1 and C22:1, were observed only in rapeseed oil. However, the intensities of these additional peaks varied at different rapeseed oil products, consistent with the reported C22:1 contents from 2% to 60% in rapeseed oil (Codex Alimentarius, 2015a). The permitted C22:1 content in edible oil products has been regulated in some regions. For example, the threshold level is 5% by weight of fatty acid content in Hong Kong (Harmful Substances in Food Regulations, 1983). This MALDI-MS approach is potentially useful for rapid screening of the illegal edible oil products by detecting the TAGs peaks of C22:1 (Table S-7).

Except for the above edible oil species that showed very obvious characteristic peaks, MALDI-MS spectra of the other edible oil species mainly consisted of similar peaks such as *m/z* 877.7 (PLL), *m/z* 879.7 (PLO), *m/z* 901.7 (LLL), *m/z* 903.7 (LLO), *m/z* 905.8 (LOO) and *m/z* 907.8 (OOO), but with different intensity ratios. Among these edible oil species, highest peak abundances were observed for *m/z* 877.7 in cottonseed oil (Figure S-1.9), *m/z* 879.7 in rice bran oil (Figure S-1.24), *m/z* 901.7 in safflower oil (Figure S-1.25), soybean oil (Figure S-1.27), walnut oil (Figure S-1.29) and wheat germ oil (Figure S-1.30), *m/z* 903.7 in corn oil (Figure S-

1.8), pine nut oil (Figure S-1.21), pumpkin seed oil (Figure S-1.22) and sunflower oil (Figure S-1.28), *m/z* 905.8 in sesame oil (Figure S-1.26), *m/z* 907.8 in almond oil (Figure S-1.1), avocado oil (Figure S-1.2), camellia oil (Figure S-1.4), canola oil (Figure S-1.5), hazelnut oil (Figure S-1.13), high oleic acid sunflower oil (Figure S-1.14) and olive oil (Figure S-1.16). The most abundant peaks in peanut oil varied as discussed before (Figure S-3), which could be either *m/z* 903.7, *m/z* 905.8 or *m/z* 907.8. These edible oil species could be differentiated by comparing the peak intensity ratios in their mass spectra. For example, the mass spectra of both olive oil (Figure S-1.16) and canola oil (Figure S-1.5) had abundant peaks at *m/z* 907.8, but olive oil showed higher abundance of peak at *m/z* 881.8, while canola oil showed higher abundance of peak at *m/z* 905.8 and additional peaks at *m/z* 899.7, *m/z* 901.7 and *m/z* 903.7). Different intensity ratios of *m/z* 881.8 and *m/z* 907.8 were observed for olive oil, avocado oil, hazelnut oil and high oleic acid sunflower oil, allowing differentiation among them.

PCA, a commonly used tool for statistical analysis, is used to facilitate the differentiation and classification of edible oils. Excluding castor oil and fish oil samples, which showed spectra that were observably different from those of other oil species, spectra of 435 edible oil products were analyzed using PCA. As shown in Figure 1a, butter and coconut oil were clustered very separately from each other and other edible oils, in consistent with their observably different

312 TAG patterns. PCA after excluding coconut oil and butter revealed two groups of edible oils,
313 with one group containing flaxseed oil and perilla oil and the other group containing palm oil,
314 lard and palm superolein, which were clustered separately from the remaining edible oils (see
315 Figure 1b). A new PCA score plot (Figure 2a) was obtained for PCA of the remaining edible
316 oils, which could be divided into 4 regions. Based on these PCA results, the edible oils were
317 divided into seven different groups and one additional group of castor oil and fish oil that
318 showed very specific peaks, which are listed with their characteristic peaks in Table 1. The
319 edible oils in groups 2, 3, 4, 5, 6 and 7 could be further differentiated with further PCA as
320 shown in Figures 1c-d and 2b-e. It is noticed that although the MALDI-MS spectra of some
321 edible oils such as peanut oil showed obvious variations within the same species as discussed
322 previously, such variations were normally still smaller than the differences between different
323 oil species, allowing them to be clearly differentiated from other oil species. However, 3 pairs
324 of edible oils, i.e., olive oil and avocado oil, canola oil and almond oil, and corn oil and
325 pumpkin seed oil, were found difficult to be differentiated based on their TAGs patterns (see
326 Figures S-1.2 and S-1.16; S-1.1 and S-1.5; and S-1.8 and S-1.22, and PCA results in Figures
327 2b, 2c and 2e).
328
329 **3.3 Classification of edible oils using PLS-DA**

PLS-DA is a supervised technique that correlates the *class* (i.e., the edible oil species in this study) and *X-variables* (i.e., the relative intensities of peaks in the MALDI-MS spectra) through the scores calculated by the statistical software (Berrueta, Alonso-Salces, & Heberger, 2007). The *class* of future samples could be predicted after the establishment of PLS-DA models using training samples. This cannot be achieved using the PCA method (Ballabio & Consonni, 2013; Berrueta et al., 2007).

Samples from 198 representative edible oil products were used as the training set for establishment of PLS-DA models in this study. The training set included edible oil products of different species, different collection sources and different MALDI-MS spectral patterns. Based on the PCA results, a two-step approach was employed for the PLS-DA analysis. The samples were firstly classified into the groups (groups 1-7 as shown in Table 1) using the "grouping" PLS-DA model and then further classified into single edible oil species using 7 sub-models (model groups 1-7). The data of samples from the training set were input into the SIMCA software for PLS-DA processing to establish a "grouping" model with the groups as the dummy Y variables. Seven sub-models with the species as the dummy Y variables were established using the training data of the edible oil samples in each group.

348 The data of 198 edible oil products in the training set were used to establish and cross-validate
349 the model automatically by the software, and 237 edible oil products in the testing set were
350 input into to the established models for the external validation. The predicted scores (PS) for
351 the matching of each oil species was calculated for each edible oil sample, and the edible oil
352 sample was then classified as the edible oil species which gave the highest score. For example,
353 the PLS-DA analysis of a sample gave predicted scores of 0.01, -0.02, -0.01, 0.06, 0.18, -0.15
354 and 0.92 for groups 1, 2, 3, 4, 5, 6 and 7, respectively, indicating that this sample belonged to
355 group 7. Further analysis using the model of group 7 gave predicted scores of 1.16, 0.03, 0.04,
356 -0.04, and -0.19 for the five edible oils in group 7, i.e., olive oil, camellia oil, high oleic acid
357 sunflower oil, avocado oil, and hazelnut oil, respectively. The oil sample was then determined
358 as olive oil.
359
360 As shown in Table S-3, for the training set, a correct classification rate of 96.0% was obtained,
361 indicating the effectiveness of the established PLS-DA models and approach for classification
362 of edible oils. It was noticed that for the 8 edible oil samples that were misclassified, they were
363 mainly the oil species with sample sizes equal to or less than 3 (e.g. almond oil and high oleic
364 acid sunflower oil), indicating that further sample collection for these oil species might be
365 needed to improve the model establishment and the classification. The misclassification of

grapeseed oil was due to the high similarity of the MALDI-MS spectra among grapeseed oil, safflower oil and sunflower oil. As shown in Table 2, for the testing set, the correct classification rate was 98.3%, with only 4 tested samples misclassified, indicating the established models could predict external samples as good as the training samples. All the misclassifications occurred in the grouping step, with 1 sample from group 4 misclassified as group 5 and 2 samples from group 5 and 1 sample from group 8 misclassified as group 4. The spectra of the misclassified samples were re-examined. It was found that although the mass spectra of the two misclassified samples from group 5 (both were sunflower oils) were similar to those of the others in the group, it showed slightly higher peaks at *m/z* 879.7 and *m/z* 905.8 than the others (see Figures S-7a and S-7b for the spectral comparison). As shown in the PLS-DA loading plot (Figure S-7c), these two peaks were the important factors for classification of samples in group 4, thus causing the misclassification. Similar explanation could be applied to the other misclassified samples. Such re-inspecting the spectra of the problematic PLS-DA results might help to reduce the chance of misclassification. Overall, including both training set and testing set, 97.2% of the tested samples were classified correctly using PLS-DA and only 12 out of the 435 tested samples were misclassified. For commonly used edible oils such as olive oil, peanut oil and canola oil, correct classification rates of 100% were achieved. Such classification could be considered as highly accurate with the high correct classification rate.

Most misclassifications were related to the species with small sample sizes, such as pine nut oil and avocado oil in the training set. These could be improved by collecting and analyze more samples for those oil species.

The predicted scores of the tested samples indicated how well the samples were matched with the training samples. Predicted scores closer to 1 indicated the better fitting between the models and the tested samples. The ranges, means and the standard deviations of the correctly classified edible oil products are listed in Table S-4. The predicted scores of butter, coconut oil, flaxseed oil and perilla oil which possessed characteristic peaks in their MALDI-MS spectra were very close to 1, indicating the high agreement between the established models and the tested samples. The predicted scores of most edible oil species were higher than 0.7 except for grapeseed oil, camellia oil and hazelnut oil. The low predicted scores of some edible oil species were normally due to the lack of characteristic spectral patterns of the species, increasing the difficulty in classification of those species. The range of the predicted scores reflected the variations of the mass spectral patterns of the edible oil species. As shown in Table S-4, peanut oil, walnut oil and soybean oil showed larger variations among all the oil species.

Five collected edible oil samples (1 flaxseed oil, 2 sesame oils, 1 sunflower oil and 1 safflower oil), which failed to pass the confirmation using the GC-FID analysis as discussed previously (Figure S-9), were analyzed by the established PLS-DA models. As shown in Table S-5, all these 5 samples were not classified as the claimed edible oil species. These results were consistent with those obtained using the GC-FID analysis, demonstrating the reliability of the PLS-DA models.

3.4 Analysis of repeatedly cooked edible oils and gutter oils

Spectra of the repeatedly cooked edible oils and gutter oil samples are shown in Figure S-8. In these samples, peaks corresponding to oxidized products, such as peaks at *m/z* 915-955 (Picariello et al., 2009; Q. Zhang et al., 2016), were observed, and the TAGs contents in some of the gutter oil samples were relatively low. Gutter oils have diverse compositions and correspondingly there is no universal spectral pattern for their identification, but the differences in the TAGs contents and the presence of extra peaks of the oxidized products could provide clues for distinguishing gutter oils from normal edible oils. As gutter oils have to be labeled as normal edible oils in order to be sold in the market, identification of mislabeling is a reasonable strategy to screen out gutter oils. In this study, the spectra of the repeatedly cooked edible oils and gutter oil samples were exanimated and analyzed using the established PLS-DA models,

and the results are shown in Table S-6. Compared with the predicted scores obtained from the authentic edible oil samples (see Table S-4), some of the tested gutter oil samples such as s470 and s471 gave either too high or too low predicted scores for the classification, and some of the tested gutter oil samples such as s472 and s473, although classified normally, showed low TAGs content or high level of oxidized products in their spectra (see Figures S-8d and S-8e), which were obviously different from those of the authentic edible oil samples. Similar results were obtained for the repeatedly cooked edible oils, which showed additional oxidized products and TAGs patterns that no longer matched with soybean oil.

4. Conclusions

In this study, 30 types of edible oils with more than 900 oil samples were analyzed using MALDI-MS. Spectral patterns for various edible oils were obtained, and a comprehensive MALDI-MS spectral database was built for analysis of edible oils. Based on the spectral patterns and PCA results, edible oils were divided into 8 groups with characteristic peaks for each group. PLS-DA models were established for classification of edible oils, with overall 97.2% (96.0% for training set and 98.1% for testing set) of edible oil products correctly classified from 435 edible oil products and nearly 100% correct classification rates for commonly used edible oils. Counterfeit edible oils and gutter oils could also be differentiated

from normal edible oils using the models. Improved analysis could be achieved with incorporation of more edible oil products, particularly for those species with small sample sizes, into the database. The results of this study showed that MALDI-MS analysis combined with the established spectral database and statistical analysis could be an effective approach for rapid classification of edible oils.

Acknowledgments

Thanks are given to the Hong Kong Innovation and Technology Fund (Grant No. ITS/070/15) and The Hong Kong Polytechnic University for the financial supports, and to Nestlé R&D (China) Limited, Wuhan Syscan Technology Holdings Limited, Hong Kong Hop Hing Oil Group Limited and Prof. Jentaie Shiea (National Sun Yat-sen University) for the assistance in collection of the edible oil samples used in this study. We would like to acknowledge the University Research Facility in Life Sciences and University Research Facility in Chemical and Environmental Analysis for providing the instruments used in this study. The technical supports from Dr. Melody Wong and Mr. Koon-Hung Wan are also appreciated.

References

Ayorinde, F. O., Elhilo, E., & Hlongwane, C. (1999). Matrix-assisted laser desorption/ionization tme-of-flight mass spectrometry of canola, castor and olive oils. *Rapid Commun. Mass Spectrom.*, *13*(10), 737-739.

Ayorinde, F. O., Eribo, B. E., Balan, K. V., Johnson, J. H., & Wan, L. W. (1999). Determination of major triacylglycerol components of polyunsaturated specialty oils using matrix-assisted laser desorption ionization time-of-fight mass spectrometry. *Rapid Commun. Mass Spectrom., 13*(10), 937-942.

Ayorinde, F. O., Garvin, K., & Saeed, K. (2000). Determination of the fatty acid composition of saponified vegetable oils using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom., 14*(7), 608-615.

Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: linear models. PLS-DA. *Anal. Methods, 5*, 3790-3798.

Berrueta, L. A., Alonso-Salces, R. M., & Heberger, K. (2007). Supervised pattern recognition in food analysis. *J. Chromatogr. A, 1158*(1-2), 196-214.

Calvano, C. D., De Ceglie, C., D'Accolti, L., & Zambonin, C. G. (2012). MALDI-TOF mass spectrometry detection of extra-virgin olive oil adulteration with hazelnut oil by analysis of phospholipids using an ionic liquid as matrix and extraction solvent. *Food Chem., 134*(2), 1192-1198.

Catharino, R. R., Haddad, R., Cabrini, L. G., Cunha, I. B. S., Sawaya, A. C. H. F., & Eberlin, M. N. (2005). Characterization of vegetable oils by electrospray ionization mass spectrometry fingerprinting: Classification, quality, adulteration, and aging. *Anal. Chem., 77*(22), 7429-7433.

Centre for Food Safety Hong Kong. (2013). *Food Safety Focus Issue 78*. Retrieved from: http://www.cfs.gov.hk/english/multimedia/multimedia_pub/multimedia_pub_fsf_78_01.html

Codex Alimentarius. (2014). *Proposed draft CODEX standard for fish oils*. Retrieved from: http://www.iffo.net/system/files/2014-10-16_Report_eWG_Fish_oil.docx

Codex Alimentarius. (2015a). *CODEX standard for named vegetable oils* (CODEX STAN 210-1999). Retrieved from: http://www.fao.org/input/download/standards/336/CXS_210e_2015.pdf

Codex Alimentarius. (2015b). *CODEX Standard for olive oils and olive pomace oils* (CODEX STAN 33-1981). Retrieved from: http://www.fao.org/input/download/standards/88/CXS_033e_2015.pdf

Gaymes, T. J., Cebrat, M., Siemion, I. Z., & Kay, J. E. (1997). Cyclolinopeptide A (CLA) mediates its immunosuppressive activity through cyclophilin-dependent calcineurin inactivation. *Febs Lett., 418*(1-2), 224-227.

Gunstone, F. D. (2004). *The chemistry of oils and fats : sources, composition, properties and uses*. Blackwell: CRC Press.

Gunstone, F. D. (2011). *Vegetable oils in food technology : composition, properties and uses*. (2nd ed.). Chichester: Wiley-Blackwell.

Harmful Substances in Food Regulations — Schedule 1 Maximum Concentration of Certain Substances Present in Specified Foods, Cap 132 § 55(1) (1983). Retrieved from: https://www.elegislation.gov.hk/hk/cap132AF

Hlongwane, C., Delves, I. G., Wan, L. W., & Ayorinde, F. O. (2001). Comparative quantitative fatty acid analysis of triacylglycerols using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and gas chromatography. *Rapid Commun. Mass Spectrom., 15*(21), 2027-2034.

Hong, E., Lee, S. Y., Jeong, J. Y., Park, J. M., Kim, B. H., Kwon, K., & Chun, H. S. (2017). Modern analytical methods for the detection of food fraud and adulteration by food category. *J. Sci. Food Agric.* DOI: 10.1002/jsfa.8364.

International Organization for Standardization. (2014). *Animal and vegetable fats and oils — Gas chromatography of fatty acid methyl esters* (ISO:12966).

Jakab, A., Nagy, K., Heberger, K., Vekey, K., & Forgacs, E. (2002). Differentiation of vegetable oils by mass spectrometry combined with statistical analysis. *Rapid Commun. Mass Spectrom., 16*(24), 2291-2297.

Jergović, A. M., Peršurić, Ž., Saftić, L. & Kraljević Pavelić, S. (2017). Evaluation of MALDI-TOF/MS Technology in Olive Oil Adulteration. *J. Am. Oil Chem. Soc., 94*, 749-757.

Jimenez-Carvelo, A. M., Gonzalez-Casado, A., & Cuadros-Rodriguez, L. (2017). A new analytical method for quantification of olive and palm oil in blends with other vegetable edible oils based on the chromatographic fingerprints from the methyl-transesterified fraction. *Talanta, 164*, 540-547.

Jin, J., Chen, J., Tian, Y., Zou, L., Wang, L., & Li, F. (2013). Determination of undecanoic acid and 13-methyl-tetradecanoic acid connected to the glyceride with internal standard

517 method and its application to the identification of adulteration of illegal cooking oil.
518 *Chin. J. Chromatogr., 31*(6), 556-560.

519 Lay, J. O., Jr., Liyanage, R., Durham, B., & Brooks, J. (2006). Rapid characterization of edible
520 oils by direct matrix-assisted laser desorption/ionization time-of-flight mass
521 spectrometry analysis using triacylglycerols. *Rapid Commun. Mass Spectrom., 20*(6),
522 952-958.

523 Li, Y., Fang, T., Zhu, S., Hunag, F., Chen, Z., & Wang, Y. (2017). Detection of olive oil
524 adulteration with waste cooking oil via Raman spectroscopy combined with iPLS and
525 SiPLS. *Spectrochim. Acta Mol. Biomol. Spectrosc.* DOI: 10.1016/j.saa.2017.06.049.

526 Lisa, M., Holcapek, M., & Bohac, M. (2009). Statistical Evaluation of Triacylglycerol
527 Composition in Plant Oils Based on High-Performance Liquid Chromatography-
528 Atmospheric Pressure Chemical Ionization Mass Spectrometry Data. *J. Agric. Food*
529 *Chem., 57*(15), 6888-6898.

530 Mu, T., Chen, S., Zhang, Y., Chen, H., Guo, P., & Meng, F. (2016). Portable Detection and
531 Quantification of Olive Oil Adulteration by 473-nm Laser-Induced Fluorescence. *Food*
532 *Anal. Methods, 9*(1), 275-279.

533 Ng, T. T., So, P. K., Zheng, B., & Yao, Z. P. (2015). Rapid screening of mixed edible oils and
534 gutter oils by matrix-assisted laser desorption/ionization mass spectrometry. *Anal.*
535 *Chim. Acta, 884,* 70-76.

536 Peng, G. J., Chang, M. H., Fang, M., Liao, C. D., Tsai, C. F., Tseng, S. H., Kao, Y. M., Chou,
537 H. K., & Cheng, H. F. (2017). Incidents of major food adulteration in Taiwan between
538 2011 and 2015. *Food Control, 72,* 145-152.

539 Peršurić, Z., Osuga, J., Galinac Grbac, T., Peter-Katalinić J. & Kraljević Pavelić, S. (2017).
540 MALDI-SpiralTOF technology for assessment of triacylglycerols in Croatian olive oils.
541 *Eur. J. Lipid Sci. Technol., 119,* 1500375. DOI: 10.1002/ejlt.201500375.

542 Picariello, G., Paduano, A., Sacchi, R., & Addeo, F. (2009). Maldi-tof mass spectrometry
543 profiling of polar and nonpolar fractions in heated vegetable oils. *J. Agric. Food Chem.,*
544 *57*(12), 5391-5400.

545 Picariello, G., Sacchi, R., & Addeo, F. (2007). One-step characterization of triacylglycerols
546 from animal fat by MALDI-TOF MS. *Eur. J. Lipid Sci. Technol., 109,* 511-524.

Rohman, A. (2016). The use of infrared spectroscopy in combination with chemometrics for quality control and authentication of edible fats and oils: A review. *Appl. Spectrosc. Rev.* DOI: 10.1080/05704928.2016.1266493.

Rosaria Cozzolino, & Giulio, B. D. (2011). Application of ESI and MALDI-TOF MS for triacylglycerols analysis in edible oil. *Eur. J. Lipid Sci. Technol., 113*, 180-187.

Saraiva, S. A., Cabral, E. C., Eberlin, M. N., & Catharino, R. R. (2009). Amazonian Vegetable Oils and Fats: Fast Typification and Quality Control via Triacylglycerol (TAG) Profiles from Dry Matrix-Assisted Laser Desorption/Ionization Time-of-Flight (MALDI-TOF) Mass Spectrometry Fingerprinting. *J. Agric. Food Chem., 57*(10), 4030-4034.

Stefanowicz, P. (2004). Electrospray mass spectrometry and tandem mass spectrometry of the natural mixture of cyclic peptides from linseed. *Eur. J. Mass Spectrom., 10*(5), 665-671.

Vaclavik, L., Cajka, T., Hrbek, V., & Hajslova, J. (2009). Ambient mass spectrometry employing direct analysis in real time (DART) ion source for olive oil quality and authenticity assessment. *Anal. Chim. Acta, 645*(1-2), 56-63.

Xu, L., Yu, X., Liu, L., & Zhang, R. (2016). A novel method for qualitative analysis of edible oil oxidation using an electronic nose. *Food Chem., 202*, 229-235.

Zhang, L., Shuai, Q., Li, P., Zhang, Q., Ma, F., Zhang, W., & Ding, X. (2016). Ion mobility spectrometry fingerprints: A rapid detection technology for adulteration of sesame oil. *Food Chem., 192*, 60-66.

Zhang, Q., Lin, S., Li, J., Liu, Y. T., Qin, W., & Shen, Q. (2016). Application of Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry for the Analysis of Compounds in Deep-Fat Frying Oil. *Food Anal. Methods, 9*(8), 2352-2363.

Zhu, W., Wang, X., & Chen, L. (2017). Rapid detection of peanut oil adulteration using low-field nuclear magnetic resonance and chemometrics. *Food Chem., 216*, 268-274.