Microchemical Journal 202 (2024) 110837

Contents lists available at ScienceDirect

Microchemical Journal

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

The phospholipid chromatographic fingerprint: An analytical cutting-edge strategy in the distinguished characterization of olive oil

Rosalía López-Ruiz [a,b], Ana M. Jimenez-Carvelo [a,*], Luis Cuadros-Rodríguez [a,b]

[a] Department of Analytical Chemistry, University of Granada, Av. Fuentenueva s/n, E-18071 Granada, Spain
[b] Department of Chemistry and Physics, University of Almería, Agri-food International Campus of Excellence (CeiA3), E-04120 Almeria, Spain

ARTICLE INFO

Keywords:
Phospholipids
Edible vegetable oils
Liquid chromatography
Instrument-agnostic fingerprinting
Ultraviolet absorption and charged aerosol detectors
Mass spectrometry

ABSTRACT

In this study, two basics are addressed to achieve the characterization of edible vegetable oils from a universal perspective. Firstly, the use of a very specific chemical fraction scarcely studied, such as the phospholipids, is proposed to tackle vegetable oil characterization. For this, a new analytical method for phospholipid fraction is developed, which is based on reverse phase liquid chromatography coupled to universal detector such as charged aerosol detector (LC-CAD). In addition, a additional method using LC-(Q-Orbitrap)MS has been developed for the chemical identification of the compounds present in the phospholipid fraction. Secondly, it is proved that the instrument-agnostizing methodology is suitable to obtain a unique and time-consistent chromatographic fingerprint for each vegetable oil, which is independent of the instrument used. This could lead for the setting up of universal databases and the development of a single global multivariate model enabling edible vegetable oils discrimination by any laboratory at any time. This ultimately leads to resource and time reduction, generating lower analysis costs. The main results have been to be able to unequivocally characterise the different edible vegetable oils under study using data mining/machine learning methods such as partial least squares-discriminant analysis, support vector machine and classification a regression trees. In addition, more than 60 chemical compounds have been characterised in samples of olive oil of different categories and other edible vegetable oils respectively. This resulted in the proposal of tentative chemical markers which could be used to identify a particular edible vegetable oil.

1. Introduction

The saponifiable fraction of edible vegetable oils consists of various families of chemical compounds, including phospholipids (PhLs). PhLs belong to a specific category of membrane components that possess an alcohol skeleton. The phosphate group can be esterified with different polar organic molecules, particularly glycerol, amino alcohols, ethanolamine, choline, or inositol. Two class of PhLs can be considered: glycerophospholipids (GPhLs) and sphingophospholipid (SPhLs). In GPhL structure, two hydroxyl groups are esterified with fatty acids, while the third hydroxyl group is phosphorylated. On the other hand, SPhL structure is based on a sphingosine skeleton instead of a glycerol. Sphingosine is an amino alcohol (2-amino-4- octadecene-1,3-diol) linking a phosphate group to a fatty acid chain. In general, GPhLs are much more abundant than SPhLs, as the latter are only found in in the membrane of nerve cells [1].

PhLs are transferred to the oil phase during the extraction process and their concentration in vegetable oils typically ranges from 1 to 18 g/kg [2]. Despite their presence, there is limited research on characterizing PhLs in vegetable oils. However, some evidence suggests that PhLs exhibit biological activity, including their influence on antioxidant capacity and oxidative stability [2]. They seem to play a synergistic role in maintaining vegetable oil stability by aiding in the regeneration of other antioxidants, such as phenols or tocopherols, through the donation of hydrogen atoms from amino groups. As a result, the primary antioxidant effect is observed in PhLs with amino alcohols as polar groups (Table 1), specifically phosphatidylcholines (PCs), phosphatidylethanolamines (PEs), and phosphatidylserines (PSs) [2].

In general, two main approaches could be considered for ensuring the authenticity of a food product, such as olive oil, one of them is based on scientific analysis and the other one based on traceability verification [3–5]. Note that, since even the simplest food is also a complex multicompositional matrix, the way to ensure its quality/authenticity should be carried out by applying a multivariate approach to evaluate

* Corresponding author.
E-mail address: amariajc@ugr.es (A.M. Jimenez-Carvelo).

https://doi.org/10.1016/j.microc.2024.110837
Received 14 April 2024; Received in revised form 10 May 2024; Accepted 20 May 2024
Available online 21 May 2024

Table 1
Characteristics of the chemical compounds that constitute the EMS and that have been employed for the determination of the SRS.

Chemical compound	Molar mass (Da)	Molecular formula	Elution order	Retention time (min)	Detector	Assigned SRS
Triphenylphosphate	326.28	$C_{18}H_{15}O_4P$	(1)	1.35	UVD / CAD	1.00
Heptaphenone	190.28	$C_{13}H_{18}O$	(2)	2.41	UVD	1.93
Deltamethrin	502.20	$C_{22}H_{19}Br_2NO_3$	(3)	4.24	UVD / CAD	3.54
Methyl linoleate	294.50	$C_{19}H_{34}O_2$	(4)	5.45	UVD / CAD	4.61
Methyl oleate	296.50	$C_{19}H_{36}O_2$	(5)	5.82	UVD / CAD	4.93
Vitamin D3	384.60	$C_{27}H_{44}O$	(6)	6.22	UVD / CAD	5.29
Methyl arachidate	326.60	$C_{21}H_{42}O_2$	(7)	6.79	CAD	5.78
Squalene	410.70	$C_{30}H_{50}$	(8)	8.20	UVD / CAD	7.03
Dodecyl arachidate	480.80	$C_{32}H_{64}O_2$	(9)	9.36	CAD	8.05
Glyceryl trioleate	885.40	$C_{57}H_{104}O_6$	(10)	10.60	UVD / CAD	9.14
Glyceryl tristearate	891.50	$C_{57}H_{110}O_6$	(11)	11.18	CAD	9.66

UVD: ultraviolet detector; CAD: charged aerosol detector

the food as a whole [6]. In the field of the food analytical chemistry, the application of the multivariate approach involves changing the conventional schemes of the analytical methods one by one, focused on the identification and quantification of chemical compounds (targeted approach), and replacing them with new analytical methods based on the application of broad-based chemical information (untargeted approach) from which a sample of the product is analysed obtaining a non-specific but characteristic analytical signal of each one. In this context, the potential of the application of the instrumental fingerprinting methodology has been demonstrated, and proof of this are the studies found in literature in this regard [7–9]. Briefly, instrumental fingerprints are defined as non-specific analytical signals, which contain all the information of interest of the analysed product, allowing to unequivocally authenticate it.

Note that, the application of instrumental fingerprinting-based analytical methods is still limited. In fact, to date there is no official analytical method based on the multivariate approach to assess the authenticity of olive oil. But, in this line, a first approximation has been performed by the AOAC International, which has proposed the creation of a working group in order to promote the development of non-targeted analytical methods for the control of foodstuff, such as olive oil, milk (liquid and powder) and honey [10]. But even then, several challenges remain to be addressed when implementing multivariate analytical methods based on instrumental fingerprints in routine analysis, especially, when chromatographic techniques, such as gas and liquid chromatography, are applied. The central challenge for their implementation lies in minimizing or removing the dependence on the analytical instrument used to obtain the chromatographic signal, because it is well known that some non-negligible variations in retention times (RT) or even peak intensities may occur when repeated chromatographic analyses are performed.

In this regard, the solution to this issue would be to obtain an instrument-independent chromatographic signal which would lead to the creation of a global chromatographic fingerprints database. From this, it would be possible to generate a single multivariate model that would combine any chromatographic instrumental fingerprint obtained by applying the same analytical method in similar chromatographic equipment but in different laboratories or in the same equipment but on a different date. In order to achieve this objective, Cuadros et al. [11,12] have recently proposed an innovative methodology to be followed in order to obtain standardised instrumental fingerprints when the gas and liquid chromatography are employed, which has been named 'instrument-agnostizing'. Both intensity values and RT of the chromatographic signal are standardised using first a reference intensity signal from a suitable internal standard (IS) and then an external standard mixture (ESM), respectively. Essentially, this last step is based on first establishing a set of system-independent constant values, named as 'standard retention scores' (SRS), from the analysis of the ESM. A comprehensive description about the process of 'instrument-agnostizing' of chromatographic signals as well as, for instance, the comparison with traditional chromatographic signal alignment algorithms, such as 'icoshift' [13], can be read in the following papers published by the authors [11,12,14].

In this contextual framework, a unique and characteristic instrument-agnostic chromatographic fingerprint PhL fraction was successfully obtained for olive oil and other edible vegetable oils samples, using ultra-high-performance liquid chromatography coupled to two measuring devices in series: a non-destructive ultraviolet absorption detector and a charged aerosol detector (UHPLC-UVD-CAD) as analytical technique. First, a multivariate model for differentiating the vegetable oils concerned was built, using partial least squares-discriminant analysis (PLS-DA), support vector machine (SVM) and classification and regression trees (CART) as the data mining/machine learning chemometric tools employed for the establishment of a single classification multivariate method. After that, the phospholipid fraction using a Vanquish Flex Quaternary LC, and a Q–Exactive™ Orbitrap mass spectrometer (UHPLC-(Q-Orbitrap)MS) was characterized.

2. Material and methods

2.1. Reagents and materials

PhL pure standards of phosphatidylethanolamine (PE), phosphatidylcholine (PC), phosphatidylinositol (PI), phosphatidylserine (PS), phosphatidylglycerol (PG) and 2-oleoyl-1-stearoylsn-san-*glycero*-3-phosphocholine (SOPC) were provided by LGC standards (Teddington, United Kingdom). From these, PhL standard solutions were prepared at 1000 mg/L in ethanol/water 90:10 (v/v).

Chemical pure standards: triphenylphosphate (TPP), heptaphenone, deltamethrin, methyl linoleate, methyl oleate, vitamin D3, methyl arachidate, squalene, dodecyl arachidate, glyceryl trioleate and glyceryl trisestearate were acquired in Sigma-Aldrich (St. Louis, MO, USA) From each, a 6000 mg/L standard solution was prepared, also in ethanol: water 90:10 (v/v). Lastly, from these 11 individual solutions, a working external standard mixture (EMS) solution was prepared at concentrations of each ranging from 20-100 mg/L. In addition, TPP was also used as IS.

HPLC-grade solvents, such as n-hexane was purchased from Panreac Quimica S.L.U. (Barcelona, Spain), 2-propanol from Honeywell (Deutschland, Germany) and ethanol from VWR (Darmstadt, Germany). Deionized water was obtained using a Milli-Q system (Millipore, Bedford, MA).

2.2. Samples and sample preparation

104 samples of edible vegetables oils were used in this study: 70 olive oil of different categories (43 extra-virgin olive oils, 15 virgin olive oils, 12 lampante olive oils), 4 olive–pomace oils and 30 seed oils different of olive such as avocado, almond, rapeseed, sunflower, soya, and grapeseed.

Vegetable oil samples was extracted according to Pérez-Beltrán et al. [14]. 1 g of vegetable oils was weighted in a 10 mL centrifuge tube and 4 mL of n-hexane was added. The mixture was shaken in a vortex 30 s and

1 mL of ethanol:water 90:10 (v/v) was then added. Later the tube was again vortexed 30 s and the ethanol:water fraction (bottom phase) was collected. This step was repeated two times and the solution resulting from merging the three extracts was filtered with a 0.22 μm nylon filter.

Finally for UHPLC-UVD-CAD analysis, 750 μL of filtered extract was introduced in a LC vial with 125 μL of ethanol:water 90:10 (v/v) and 125 μL of TPP-IS at 6000 mg/L. For UHPLC–(Q–Orbitrap)MS, 100 μL of filtered extract was introduced in a LC vial with 850 μL of ethanol:water 90:10 (v/v) and 50 μL of TPP-IS at 10 mg/L.

2.3. UHPLC −UVD-CAD conditions

PhL fingerprints were obtained using an ultra-high-performance liquid chromatography (UHPLC) system Dionex Ultimate 3000 UHPLC + (Thermo Scientific, Waltham, MA, USA) equipped with a RS autosampler and column compartment. The analytical column was a Hypersil Gold (50 × 2.1 mm, 1.9 μm particle size). A gradient elution was performed using a mobile phase composed of an aqueous phase, water:acetonitrile, 80:20 (v/v) and an organic phase, isopropanol: acetonitrile (90:10 (v/v), both containing 10 mM ammonium formate and 0.1 % formic acid. The gradient elution started at 70 % of A and linearly decreased to 45 % in 2 min. From 2-10 min, A was decreased to 0 % (hold 2 min). Finally, it was returned to the initial condition (70 % A) in 2 min and kept constant for 3 min. Flow rate was 0.35 mL/min, column temperature was set at 45 °C, injection volume 10 μL. The total running time was 16 min.

UHPLC was coupled to a rapid separation (RS) variable wavelength UV detector (UVD) (Thermo Scientific, Waltham, MA, USA) operating at 200 nm. In addition, HPLC was also coupled in series to a charged aerosol detector (CAD) (Thermo Scientific, Waltham, MA, USA) with a power function of 1.00.

Chromeleon™ version 7.0 software (Thermo Scientific, Waltham, MA, USA) was used to display and export the chromatographic data.

2.4. UHPLC-(Q-Orbitrap)MS conditions

The characterization of PhLs was carried out using a Scientific Vanquish Flex Quaternary LC, and a Q-Exactive™ Orbitrap mass spectrometer (Thermo Scientific, Waltham, MA, USA). The chromatographic conditions were the same as for UHPLC-UVD-CAD.

Regarding analyte detection, the Q-Exactive Orbitrap analyser operated in Full Scan MS and data dependent mass fragmentation (dd)-MSMS (positive and negative ionization modes) were used. Electrospray ionization (ESI) conditions included capillary temperature of 300 °C, heater temperature of 305 °C, spray voltage of 4 kV, S-lens radio frequency (RF) amplitude of 50 V, and use of 95 % purity N_2 as auxiliary and sheath gas. Full Scan MS data was acquired in the *m/z* range of 60–900, at a resolution of 70,000 at *m/z* 200, and an AGC target of 10^6 for both positive and negative modes. Furthermore, dd-MSMS acquisition was performed at a resolution of 35,000 at *m/z* 200, an isolation window of *m/z* 3.0, an AGC target value of 10^5, multiplexing 2 and loop count 5. The software Xcalibur™ 4.3 (Thermo Scientific, Waltham, MA, USA) and Compound Discoverer™ 3.3 (Thermo Scientific, Waltham, MA, USA) were used for data acquisition and processing.

2.5. Chromatographic analysis

In order to have the required data for the study, two chromatographic experiments were carried out. In the first one, ten independent ESM aliquots were analysed in duplicate over one week under reproducibility conditions. The 104 vegetable oil solutions were analysed consecutively in a single batch. In addition, analyses of the ESM were inserted at the beginning and at the end of the batch, as well as every 10 samples.

The chromatograms acquired by both detectors in each of the chromatographic runs were always recorded.

2.6. Instrument-agnostizing of chromatographic signals

As mentioned in the introduction section, to reach a successful harmonised database of instrument-agnostic chromatographic fingerprints, which could be used to build a single and universal multivariate model, it is necessary to standardise the signals (chromatograms) acquired using similar chromatographic equipment in order to mitigate the effects of instrument variability as well as those due to time (date) of sample analysis. In this concern, the instrument-agnostizing methodology proposed by Cuadros *et al.* [11,12] was applied on raw UHPLC–CAD chromatograms. Chromatographic raw data files, embedded in a data vector composed of 9459 intensity elements, were exported in 'comma separated value' (CSV) format, and then converted to MATLAB format (version 2017, The Mathworks Inc. MA, USA).

Firstly, the relationship between RT and SRS values was established from both UVD and CAD chromatograms obtained from ten initial EMS analyses. Subsequently, the intensity values were normalised using the peak intensity of the internal standard TPP obtained from the CAD as reference. Then, the CAD chromatograms were converted from the RT domain to the SRS domain in order to unify the scale on the x-axis.

2.7. Multivariate analysis

All multivariate analysis was carried out using PLS_Toolbox (version 8.6.1, Eigenvector Research Inc., Manson, WA, USA) working under Matlab (version R2017b, The Mathworks Inc. MA, USA) environment.

Each chromatogram was embedded in a data vector of the acquired intensity values corresponding to each retention time. Previously, in order to fix into the same number of variables in all chromatograms, a re-sampling process was first applied. In this way, each of the instrument-agnostic CAD chromatographic vector was converted into a new data vector consisting of 1139 intensity elements.

Partial least squares-discriminant analysis (PLS-DA), support vector machine (SVM) and classification and regression tree (CART) were then applied as classification methods for the development of the multivariate models. The data extracted from the UHPLC-CAD chromatograms were allocated into two classes: virgin olive oil class (both virgin and extra virgin olive oils) and non-olive oil class (all other edible vegetable oils, including olive pomace oils). In addition, the selected samples for the training and external validation sets were split using the CADEX algorithm (Kennard-Stone method) [15].

2.8. Data treatment by unknown analysis

Unknown analysis was carried out using Compound Discoverer 3.3, employing lipidomic strategies searching in on-line libraries (*m/z* Vault and Chemspider) and mass list search (mass list provided by the software) that contains compounds of the family of lipids as fatty acids, phospholipids, between others (around 1.000 compounds). In particular, the selected libraries were *LipidMAPS* for ChemSpider, *LipidBlast in-silico database* for *m/z* Vault [16] and *LIPID MAPS Structure Database* [17] as mass list database. The parameters introduced in the workflow analysis were: align RT using a reference file of the PhL analytical standards injected; detect compounds with mass tolerance of 5 ppm and min peak area intensity 10^6; merge features with mass tolerance of 5 ppm and RT tolerance of ± 0.2 min; group compounds with mass tolerance of 5 ppm and RT tolerance of ± 0.2 min, align peaks true and preferred ions $[M + H]^+$, $[M + NH_4]^+$, $[M-H]^-$, $[M + COO]^-$; search in databases with mass tolerance of 5 ppm; fill gaps and mark background compounds with S/N threshold of 1.5 and max sample/blank ratio of 5 (relation between a signal of a suspect compound present in the sample/blank).

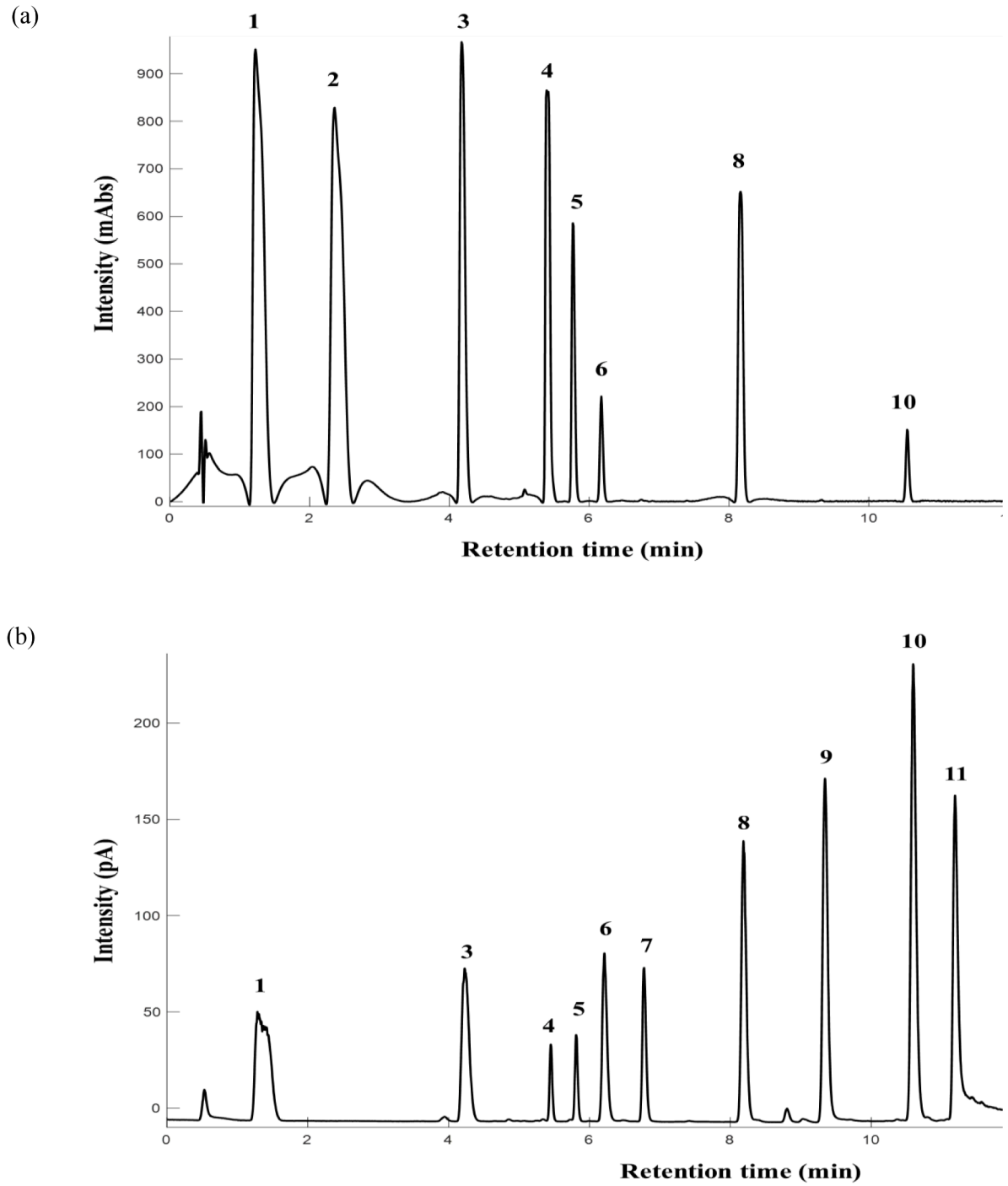

Fig. 1. Chromatograms of the external standard mix (ESM) composed by different chemical compounds obtained using: (a) ultraviolet detector (UVD) and (b) charged aerosol detector (CAD). Note that the numbering in the figures corresponds to the order of elution of the different compounds, which can be found in Table 1.

3. Results and discussion

3.1. Development of UHPLC analytical method

The ESM mixture of PhLs described in section 2.1 was used to optimise the chromatographic method for the analysis of PhLs. As a starting point, two chromatographic studies reported in the literature [18,19] were tested. They used the same composition of mobile phases, acetonitrile:water (60:40, v/v) as aqueous phase isopropanol:acetonitrile (90:10, v/v) as organic phase, both containing 10 mM of ammonium formate and 0.1 % of formic acid. Run times and elution gradients were similar and flow rate was 0.35 mL/min for Narvaez-Rivas et al. [19] and 0.26 mL/min for Kiyonami et al. [18]. With this conditions, three different analytical columns were tested: Hypersil Gold aQ (100 x 2.1 mm, 1.9 um), Hypersil Gold (100 × 2.1 mm, 1.9 um) and Hypersil Gold (50 × 2.1 mm, 1.9 um). To obtain a quick elution of the compounds and with the best peak shape, the best results were obtained using Hypersil Gold (50 × 2.1 mm, 1.9 um) and the conditions of Narvaez et al. [19].

However, the analysis time was too long, and the compounds eluted too late (between 14–20 min). Therefore, it was decided to reduce the elution gradient time from 31 min to 26 min and to 16 min, maintaining the percentages of mobile phase but reducing the times of each step. Employing the gradient of 16 min, compounds eluted adequately, but the compounds eluted late (between 10 to 14 min). So, it was decided to

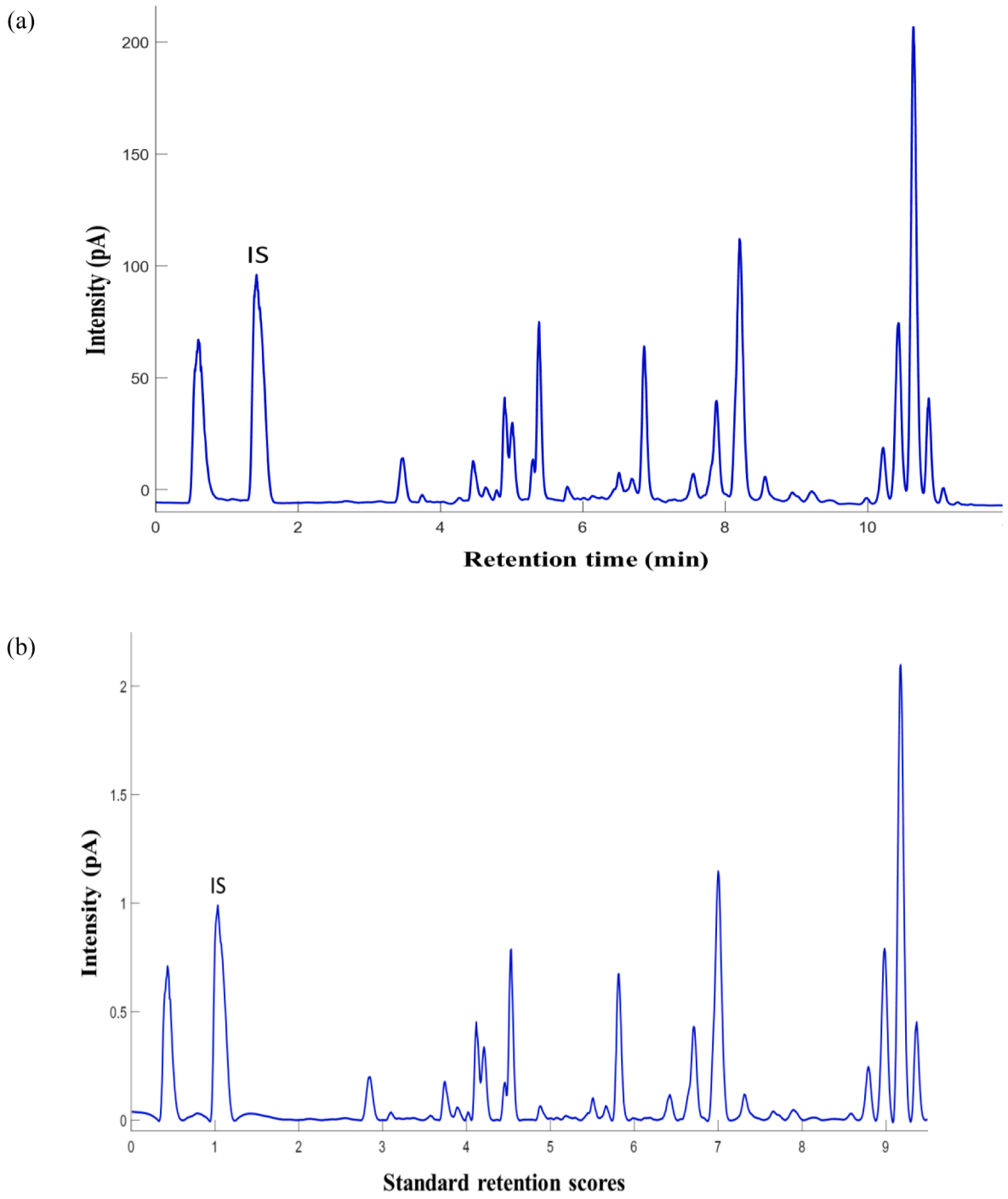

Fig. 2. Chromatograms of a randomly chosen olive oil sample obtained using the charged aerosol detector: (a) raw chromatogram, and (b) instrument-agnostic chromatographic fingerprint. (IS: Internal standard).

increase the polarity of the aqueous phase from 40 % of water to 80 % of it and improve the elution of the compounds between 4–9 min. Fig. 2 shows a raw CAD chromatogram of a randomly chosen olive oil sample obtained when the optimised method for UHPLC is applied.

3.2. Instrument-agnostic chromatographic fingerprints: UHPLC-CAD analysis

For this purpose, the RT of each of the 11 compounds constituting the EMS were taken on the 20 chromatograms (UVD and CAD) obtained from the 10 replicates initially analysed, and then the corresponding median values were calculated. A display of the chromatograms obtained using UVD (200 nm) and CAD are shown in Fig. 1a and 1b, respectively. Note that not all mixture components are "seen" by both detectors, UVD and CAD. Hence the need to use both detectors to obtain a set of SRS values evenly spaced over the full elution time (12 min) of the vegetable oil samples.

For more details, Table 1 reports the identity, the elution order, the retention times, the detector with which the corresponding peak was acquired and, finally, the SRS value assigned to each EMS component.

The synergistic use of both detectors, UVD and CAD, for the estimation of the sequence of SRS values does not necessarily imply that both must be used to carry out the agnostizing of the chromatograms of the samples. Once the SRS values corresponding to each RT in each

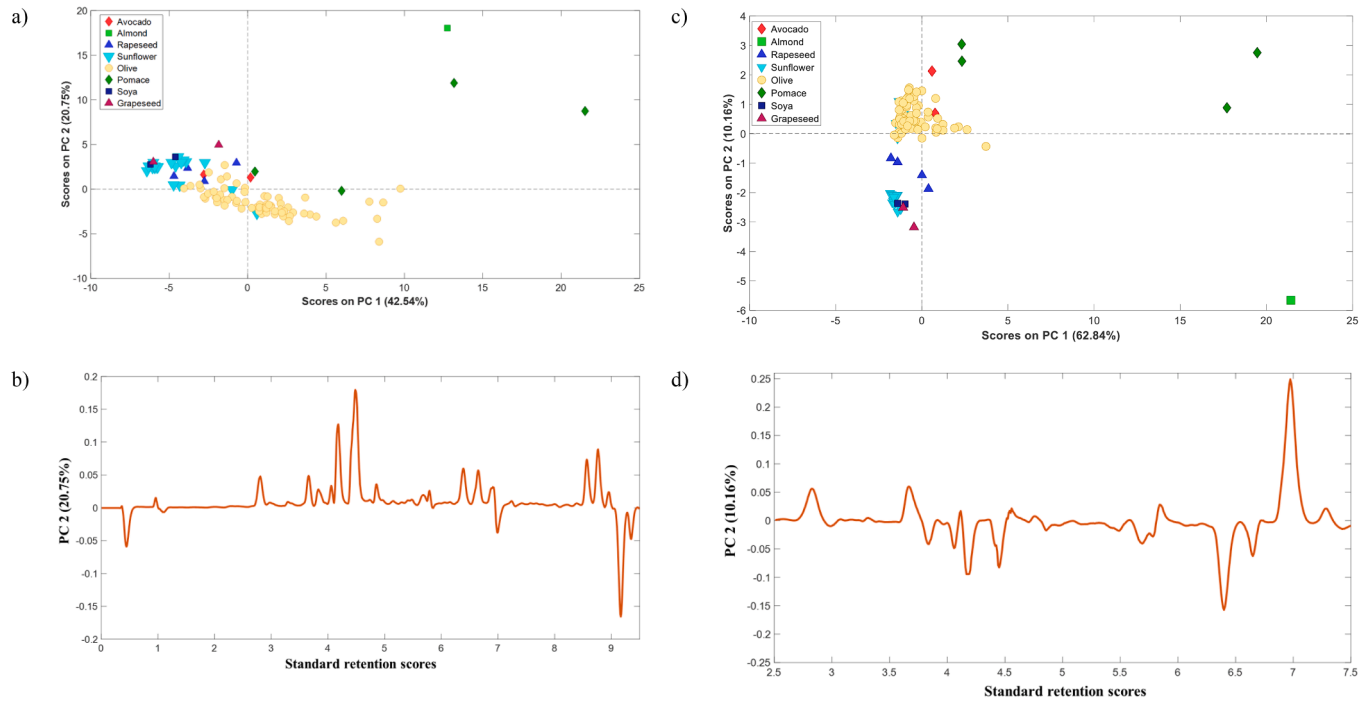

Fig. 3. PCA scores (a) and loadings (b) plots showing all olive and non-olive samples considering the whole instrument-agnostic chromatographic fingerprints. PCA scores (c) and loadings (d) plots showing all olive and non-olive samples only considering the reduced instrument-agnostic chromatographic fingerprints comprised between the SRS values of 2.5 and 7.5 (PhL fraction).

chromatographic run have been established, the position standardisation (alignment), which involves the transfer from the RT domain to the SRS domain, can be performed using only the raw chromatogram acquired by one of the detectors. This is an additional advantage of the proposed methodology. In fact, in this study, results are presented using only the data obtained from the CAD.

Once the chromatograms have been aligned and defined on the new SRS x-axis, following the protocol established by Cuadros *et al.* [11,12], the intensity standardisation can be carried out. For this, all the intensity values constituting each chromatogram are divided by the intensity value (peak height) corresponding to the internal standard (TPP), i.e., the second peak appearing at around 1.4 min. In this way, an instrument-agnostic chromatogram is finally obtained for each sample, resulting from plotting the relative intensity values (y-axis) versus the new SRS values (x-axis). Note that this chromatogram is independent of the type and state of maintenance of the instrument and can be considered as a true chromatographic fingerprint representative of the analysed fraction of each sample. Fig. 2 shows both the raw chromatogram (a) and the final instrument-agnostic chromatogram (or chromatographic fingerprint) (b for the same olive oil sample.

3.3. Single multivariate analytical method from instrument-agnostic chromatograms (chromatographic fingerprints)

Data vectors representing the chromatographic fingerprints, each consisting of 1139 variables, were used to perform the successive steps of this multivariate study.

3.4. Preliminary screening study

First, an exploratory principal component analysis (PCA) was performed in order to screen the intrinsic data structure. PCA model was built considering 6 principal components (PCs) which explained 89.66 % of the total variance. Fig. 3a and 3b present, respectively, the resulting PC2/PC1 score and loading plots.

There is a slight clustering trend that differentiates olive oils other edible vegetable oils. It is also noted that the instrument-agnostic chromatographic fingerprints of olive pomace oils, as well as almond oil, appears to be quite dissimilar to those of other non-olive vegetable oils. When inspecting the loading plot, it was observed that the final region of the fingerprint, comprised between the SRS values of 7.50 and 9.49 (approx. 10 to 12 min), had a significant influence on the observed groupings.

Since PhLs do not elute in region of the chromatographic fingerprint (see section 3.4), it was decided to remove it. In order to strictly assess the potential of the phospholipid fraction, the initial region corresponding to the solvent front and the IS peak was also removed. Therefore, a new PCA model was developed, considering just the data included in the range of the SRS values of 2.5 and 7.5, which significantly reduced the number of variables from 1139 variables to 600. Thus, it also reduced computational resources. This new PCA model was developed by selecting 6 PCs, which explained 93.5 % of the total variance.

In Fig. 3c and 3d, both the scores and loadings plot of PC1 *vs* PC2 are displayed. Scores distribution shows the clear grouping of the two main sample types of edible vegetable oils. Olive oils cluster have PC2 positive and PC1 negative scores, while the cluster consisting of the rest of the edible vegetable oils have negative scores for both PCs. Similar to the first PCA model, olive pomace oils exhibit a PhL fingerprint dissimilar to other edible vegetable oils, as does almond oil. Note that, within the non-olive vegetable oils group, there is a small subgroup for rapeseed oils emerges, indicating a particular profile of the PhL fraction.

3.5. Analytical multivariate characterization of edible vegetable oils

Based on previous screening results, it was decided to carry out the characterization of the vegetable oils (olive and non-olive oils) considering only the chromatographic fingerprint data between 2.5 and 7.5 SRS values. Three data mining tools, namely PLS-DA, SVM, and CART, were applied for this purpose. For this, two classes were considered to build the classification models: virgin olive oil class (target class) and non-olive oil class. The training was set of 73 samples (47 olive oils and

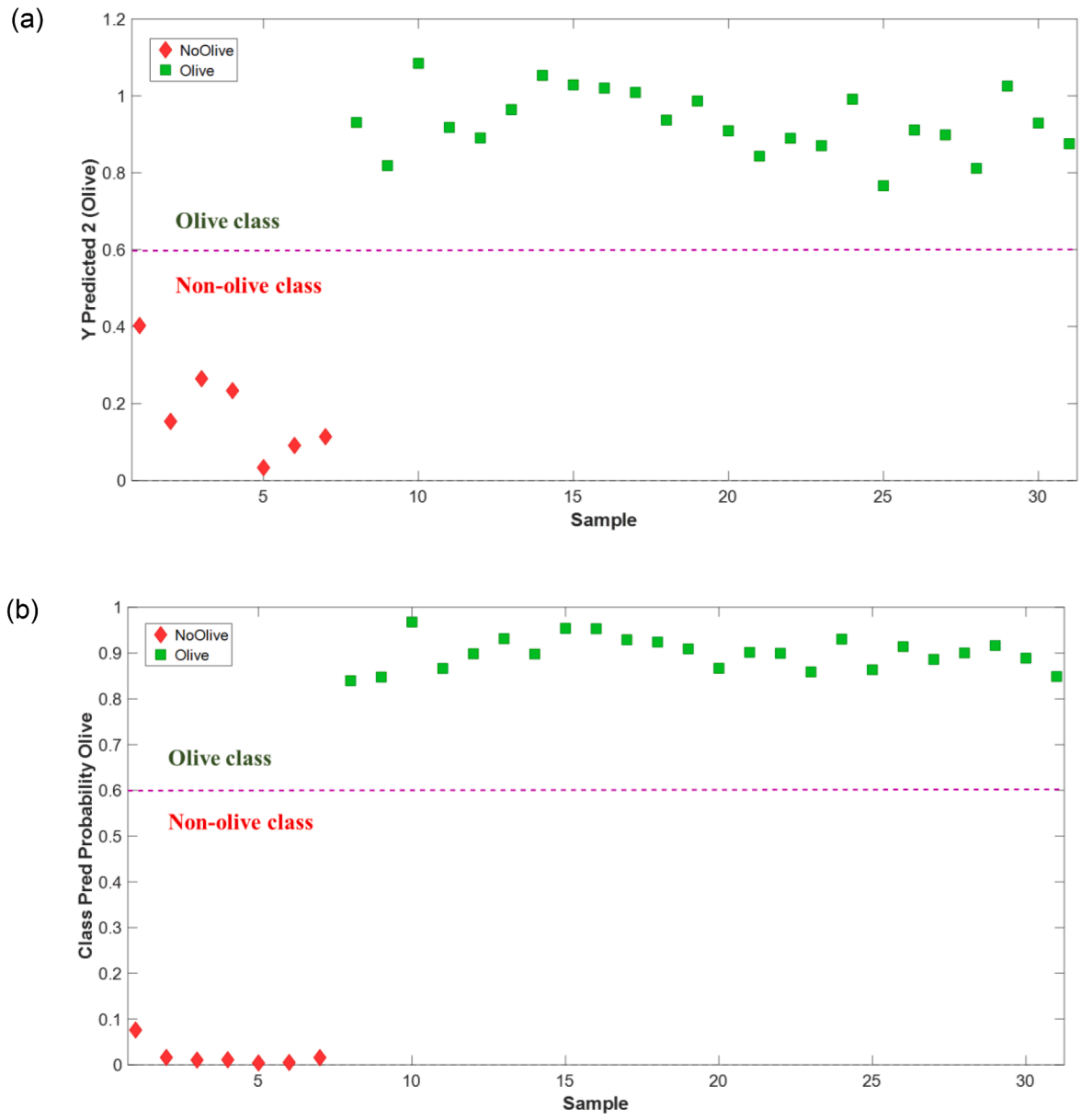

Fig. 4. Classification plots of the validation set samples from: (a) PLS–DA model, and (b) SVM model. *(The purple dashed lines denote the classification threshold).*

26 non-olive oils) and further validated with a data set of 31 samples (24 olive oils and 7 non-olive oils), as outlined in subsection 2.7.

PLS-DA was performed with 7 latent variables (LVs) that could explain 93.95 % of the total variance. The classification results are showed in Fig. 4a in which all vegetable oils samples were correctly classified according to a classification threshold of 0.6. Thus, the classification metrics of sensitivity (SENS), specificity (SPEC), precision (PREC), and efficiency (EFFI) were equal to 1.00. Note that, these metrics were calculated from the success/error contingencies from the classification results of the external validation set (see Fig. S1 in supplementary material). Moreover, more details on the specific features of the meaning of the classification metrics can be found in the tutorial published by Cuadros *et al.* [20].

A SVM classification model, considering the same two classes, was then built. The Kernel type algorithm radial basis function (RBF) with gamma and cost values, established by default in the PLS Toolbox software, was applied. As observed in Fig. 4b, all validation samples were classified within their corresponding classes. Thus, the classification metrics values were equal to 1.0 similar to PLS-DA. It is noteworthy that, unlike PLS-DA, the SVM model exhibited a larger separation between the olive and non-olive classes, resulting in lower classification uncertainty.

Finally, CART method was applied. For this, the Matlab statistics and machine learning toolbox package, specifically 'classretree' function, was employed. Such function creates a binary decision tree for predicting the class from the value of the variable response Y (normalised chromatographic intensities) as a function of each value of the predictor variable X (SRS). The classification tree is constituted by a set of nodes, where each non-terminal node, involving a decision rule, is divided according to a threshold value of Y for a given X. Fig. 5 shows the built classification tree with this method.

A key finding arising from the application of the three independent classifiers is that when highly informative instrumental fingerprints are available, showing significant between-object variability, any of the three classifiers could be chosen. An optimal selection of a classifier will depend on a number of factors, most notably the required computational resources. Despite PLS-DA and SVM yielding similar results and SVM exhibiting a higher efficiency of class separation, perhaps the use of PLS-

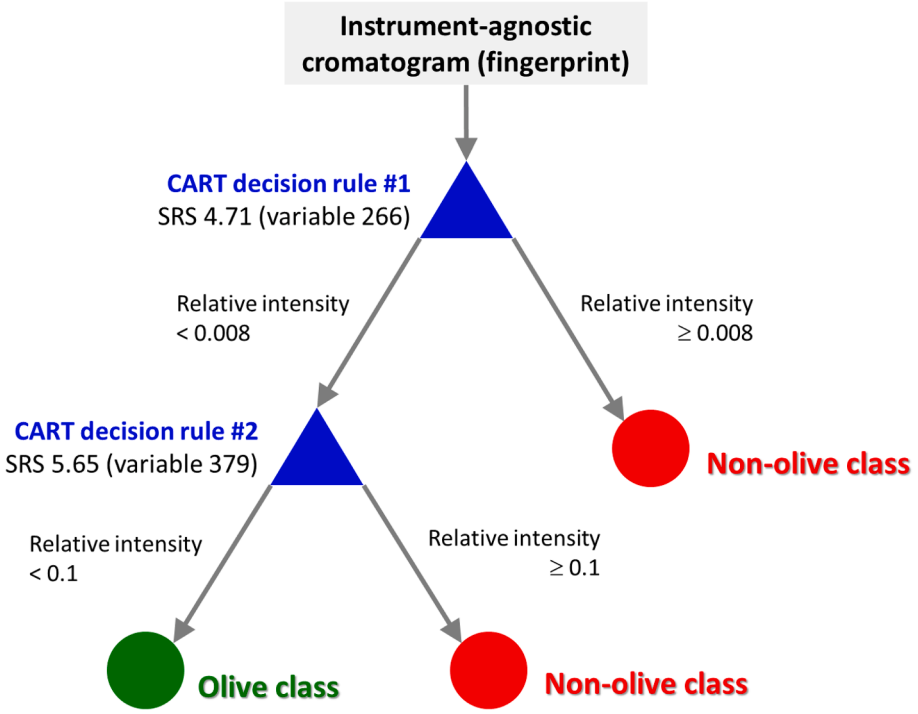

Fig. 5. Classification tree built from a CART model having only two consecutive decision rules.

Table 2
HRMS parameters for the PhLs analytical standards.

PhL (molecular formula)	Theoretical *m/z*	Mass error (ppm)	Fragment ion	Mass error (ppm)	Molecular formula	Adduct	RT (min)	SRS
PI (16:0/18:2) $C_{43}H_{79}O_{13}P$	833.519	−2.702	279.239	−0.263	$C_{18}H_{31}O_2$	$[M-H]^{-1}$	6.84	6.85
PG (16:0/18:1(9Z)) $C_{40}H_{77}O_{10}P$	747.518	−3.382	281.248	−3.604	$C_{18}H_{33}O_2$	$[M-H]^{-1}$	7.19	7.15
PE (16:0/18:2) $C_{39}H_{74}O_8NP$	716.522	−0.448	575.503	−0.151	$C_{37}H_{67}O_4$	$[M+H]^{+1}$	7.35	7.28
PS (18:1/18:0) $C_{42}H_{80}NO_{10}P$	790.559	1.239	605.552	−1.069	$C_{42}H_{71}ON$	$[M+H]^{+1}$	7.63	7.52
Lecithine $C_{39}H_{76}NO_8P$	718.538	−0.906	577.519	−1.259	$C_{37}H_{69}O_4$	$[M+H]^{+1}$	7.68	7.56
SOPC $C_{44}H_{86}NO_8P$	788.616	−0.446	184.073	−0.82	$C_{15}H_{15}NO_4P$	$[M+H]^{+1}$	8.01	7.84

Abbreviations: PI: phosphatidylinositol; PG: phosphatidylglycerol; PE: phosphatidylethanolamine; SOPC: 2-oleoyl-1-stearoylsn-san-glycero-3-phosphocholine

DA requires fewer computational resources, making it a very viable option. However, when any of the above is compared to CART, this one holds an advantage as it not only performs classification but also enables the selection of potential chemical markers characteristic of the object under study.

3.6. Phospholipids characterisation: UHPLC-(Q-Orbitrap)MS analysis

3.6.1. PhL characterisation in analytical standards

Two characteristic ions were selected in order to carry out the mass-spectrometric characterisation of each PhL standard by LC–HRMS: the precursor ion (molecular ion as the quantification ion) and a fragment ion (as confirmation ion). For this purpose, spectrometric conditions were adjusted by injecting individual standard solutions at 1 mg/L in ethanol:water 90:10 (v/v) using ESI + and ESI–. The precursor ions ($[M+H]^+$ or $[M-H]^-$) were acquired from full scan mass spectra considering that the mass error must be lower than 5 ppm. After that, (dd)-MSMS spectra of each precursor ion were studied to select a fragment ion for each PhLs. The criteria used to choose the fragments was based on the relative abundance of the ion and the information provided by databases as ChemSpider. The MS conditions for the 6 PhLs, i.e., the precursor and fragment ion, their exact mass, molecular formula, their associated mass errors, and RT (min), are shown in Table 2.

3.6.2. PhL characterisation in vegetable oil samples

The first step in the PhLs characterisation in vegetable oil samples was the determination and confirmation of the 6 PhLs available as analytical standards in oil samples. PG, PS and SOPC were the PhLs less detected in the samples. These were detected only in non-olive oil samples, and in particular, PG was detected only in almond oil, PS in almond, soya and in oleic high sunflower oils, and SOPC in almond and oleic high sunflower oils. Lecithin was detected in in the major samples (olive and non-olive oils) except in soya and sunflower oils. PE was detected in all samples, except in soja oils and finally PI was detected in all vegetable oil samples.

The next step is the characterisation of the discriminant families of PhLs to identified which PhLs could be from each type of edible vegetable oil. The samples were processed and a total of 103 PhLs were tentatively detected, where 65 out them were linked to only one type of oil. Most of these discriminants PhLs were from PCs family (lecithin) (25

Table 3
PhLs found in each type of vegetable oils samples and their spectrometric characterization.

Vegetable oil type	PhL name	PhL Formula	Mass error (ppm)	Molecular mass	*m/z*	RT (min)	SRS	Adduct	Family
Non-olive	sphinganine-1-phosphocholine	$C_{23}H_{51}N_2O_5P$	−5.13	466.351	467.358	2.80	3.42	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphocholines)
Pomace	PI (O-18:0/21:0)	$C_{48}H_{95}O_{12}P$	4.63	894.66	448.337	3.01	3.60	$[M+2H]^{+2}$	Glycerophospholipids (Glycerophosphoinositols)
Pomace	PG (20:0/0:0)	$C_{26}H_{53}O_9P$	0.36	540.343	539.336	3.18	3.75	$[M-H]^{-1}$	Glycerophospholipids (Glycerophosphoglycerols)
Lampante	PA (O-16:0/13:0)	$C_{32}H_{65}O_7P$	2.95	592.448	593.456	3.58	4.09	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphates)
Lampante	PC (O-16:1(11Z)/2:0)	$C_{26}H_{52}NO_7P$	1.86	521.349	522.357	3.75	4.23	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphocholines)
Lampante	PA (18:1(9Z)/0:0)	$C_{21}H_{41}O_7P$	0.04	436.259	419.256	3.78	4.26	$[M+H-H_2O]^{+1}$	Glycerophospholipids (Glycerophosphates)
Non-olive	4-azaniumyl-4-deoxy-α-L-arabinopyranosyl undecaprenyl phosphate	$C_{60}H_{100}NO_7P$	−1.29	977.722	978.73	3.84	4.31	$[M+H]^{+1}$	Phenol lipids (Polyprenols)
Pomace	PG (17:0/0:0)	$C_{23}H_{47}O_9P$	1.46	498.297	497.29	4.16	4.58	$[M-H]^{-1}$	Glycerophospholipids (Glycerophosphoglycerols)
Pomace	1-Octadecyl-*sn*-glycero-3-phosphocholine	$C_{26}H_{57}NO_6P$	0.07	510.392	528.426	4.46	4.83	$[M+NH_4]^{+1}$	Glycerophospholipids (Glycerophosphocholines)
Non-olive	1-octadecyl-2-methyl-*sn*-glycero-3-phosphocholine	$C_{27}H_{59}NO_6P$	−0.03	524.408	542.442	4.62	4.97	$[M+NH_4]^{+1}$	Glycerophospholipids (Glycerophosphocholines)
Lampante	PI (O-16:0/14:1(9Z))	$C_{39}H_{75}O_{12}P$	2.2	766.501	767.507	4.88	5.19	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphoinositols)
Lampante	PI (P-16:0/14:1(9Z))	$C_{39}H_{73}O_{12}P$	2.36	764.486	782.519	5.03	5.32	$[M+NH_4]^{+1}$	Glycerophospholipids (Glycerophosphoinositols)
Lampante	PI-Cer(d18:1/14:0)	$C_{38}H_{74}NO_{11}P$	3.03	751.502	752.509	5.05	5.33	$[M+H]^{+1}$	Sphingolipids (Phosphosphingolipids)
Lampante	PI-Cer(d20:1/14:0)	$C_{40}H_{78}NO_{11}P$	2.88	779.533	780.541	5.05	5.33	$[M+H]^{+1}$	Sphingolipids (Phosphosphingolipids)
Lampante	PI(P-16:0/17:2(9Z,12Z))	$C_{42}H_{77}O_{12}P$	2.72	804.517	822.551	5.10	5.38	$[M+NH_4]^{+1}$	Glycerophospholipids (Glycerophosphoinositols)
Lampante	PE-GlcDG(P-16:0/16:1(9Z))	$C_{43}H_{82}NO_{12}P$	1.98	835.559	836.566	5.10	5.38	$[M+H]^{+1}$	Glycerolipids (Glycosyldiradylglycerols)
Lampante	PG (O-18:0/14:0)	$C_{38}H_{77}O_9P$	4.36	708.534	709.541	5.13	5.40	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphoglycerols)
Lampante	PG(12:0/22:2(13Z,16Z))	$C_{40}H_{75}O_{10}P$	2.96	746.512	764.549	5.15	5.42	$[M+NH_4]^{+1}$	Glycerophospholipids (Glycerophosphoglycerols)
Lampante	PS(22:0/12:0)	$C_{40}H_{78}NO_{10}P$	2.4	763.538	764.545	5.15	5.42	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphoserines)
Lampante	PI(P-16:0/16:1(9Z))	$C_{41}H_{77}O_{12}P$	3.07	792.518	791.51	5.15	5.42	$[M-H]^{-1}$	Glycerophospholipids (Glycerophosphoinositols)
Non-olive	PG (13:0/22:2(13Z,16Z))	$C_{41}H_{77}O_{10}P$	1.55	760.52661	761.534	5.17	5.44	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphoglycerols)
Non-olive	SM (d16:1/18:1)	$C_{39}H_{77}N_2O_6P$	−3.92	700.549	699.542	5.33	5.57	$[M-H]^{-1}$	Sphingolipids (Phosphosphingolipids)
Lampante	PA (12:0/15:1(9Z))	$C_{30}H_{57}O_8P$	4.5	576.381	577.389	5.34	5.58	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphates)
Pomace	**PC (14:0/24:1(15Z))***	$C_{46}H_{90}NO_8P$	2.92	815.643	816.65	5.48	5.70	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphocholines)
Pomace	1,2-dioleoyl-L-α-phosphatidylcholine	$C_{44}H_{85}NO_8P$	−2.03	786.6	787.607	5.80	5.97	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphocholines)
Pomace	PC (18:0/20:2(11Z,14Z))	$C_{46}H_{88}NO_8P$	2.54	813.627	814.634	5.82	5.99	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphocholines)
Pomace	PC (O-16:0/22:2(13Z,16Z))	$C_{46}H_{90}NO_7P$	2.78	799.648	800.655	5.83	6.00	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphocholines)
Lampante	PA (16:0/18:2(9Z,12Z))	$C_{37}H_{69}O_8P$	0.35	672.473	671.466	5.94	6.01	$[M-H]^{-1}$	Glycerophospholipids (Glycerophosphates)
Lampante	PC (18:0/11:1(10E))	$C_{37}H_{72}NO_8P$	0.66	689.5	690.507	5.95	6.01	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphocholines)
Lampante	PA (14:1(9Z)/22:2(13Z,16Z))	$C_{39}H_{71}O_8P$	0.01	698.489	716.522	6.00	6.14	$[M+NH_4]^{+1}$	Glycerophospholipids (Glycerophosphates)
Lampante	PC (13:0/18:2(9Z,12Z))	$C_{39}H_{74}NO_8P$	−0.22	715.515	716.522	6.02	6.16	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphocholines)
Lampante	MIPC (d18:0/26:0)	$C_{56}H_{110}NO_{16}P$	2.32	1083.759	1084.766	6.09	6.22	$[M+H]^{+1}$	Sphingolipids (Phosphosphingolipids)
Lampante	PA (12:0/15:0)	$C_{30}H_{59}O_8P$	4.68	578.397	579.405	6.12	6.24	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphates)
Pomace	PA (O-16:0/19:1(9Z))	$C_{38}H_{75}O_7P$	3.82	674.528	675.535	6.13	6.25	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphates)
Pomace	PG (P-18:0/17:2(9Z,12Z))	$C_{41}H_{77}O_9P$	1.6	744.532	745.539	6.14	6.26	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphoglycerols)

(*continued on next page*)

Table 3 (*continued*)

Vegetable oil type	PhL name	PhL Formula	Mass error (ppm)	Molecular mass	*m/z*	RT (min)	SRS	Adduct	Family
Pomace	PS (O-20:0/17:2(9Z,12Z))	$C_{43}H_{82}NO_9P$	4.28	787.576	788.584	6.14	6.26	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphoserines)
Pomace	PG (O-18:0/20:3(8Z,11Z,14Z))	$C_{44}H_{83}O_9P$	0.84	786.578	787.586	6.14	6.26	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphoglycerols)
Pomace	PI (16:1(9Z)/22:4 (7Z,10Z,13Z,16Z))	$C_{47}H_{81}O_{13}P$	4.64	884.546	885.553	6.18	6.29	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphoinositols)
Lampante	PA (16:0/18:1(11Z))	$C_{37}H_{71}O_8P$	−0.17	674.489	692.522	6.28	6.38	$[M+NH_4]^{+1}$	Glycerophospholipids (Glycerophosphates)
Lampante	PA (18:0/18:2(9Z,12Z))	$C_{39}H_{73}O_8P$	0.13	700.504	699.497	6.28	6.38	$[M-H]^{-1}$	Glycerophospholipids (Glycerophosphates)
Lampante	PC (16:0/15:1(14))	$C_{39}H_{76}NO_8P$	−0.42	717.531	718.538	6.31	6.40	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphocholines)
Lampante	PC (18:0/18:2(10Z,12Z))	$C_{44}H_{84}NO_8P$	−0.83	785.593	786.6	6.36	6.44	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphocholines)
Non-olive	PA (O-18:0/17:0)	$C_{38}H_{77}O_7P$	3.97	676.543	677.551	6.43	6.50	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphates)
Non-olive	PG (O-18:0/17:2(9Z,12Z))	$C_{41}H_{79}O_9P$	2.52	746.54805	747.555	6.43	6.50	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphoglycerols)
Non-olive	PA (O-20:0/18:0)	$C_{41}H_{83}O_7P$	3.05	718.59	719.597	6.43	6.50	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphates)
Non-olive	PS (O-18:0/19:1(9Z))	$C_{43}H_{84}NO_9P$	3.05	789.591	790.598	6.44	6.51	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphoserines)
Non-olive	PG (O-16:0/22:2(13Z,16Z))	$C_{44}H_{85}O_9P$	1.47	788.5941	789.602	6.44	6.51	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphoglycerols)
Pomace	PG (O-18:0/21:0)	$C_{45}H_{91}O_9P$	1.37	806.6417	807.648	6.67	6.71	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphoglycerols)
Pomace	1-tridecanoyl-2-heptadecanoyl-*sn*-glycero-3-phosphocholine	$C_{38}H_{77}NO_8P$	−0.86	706.538	724.571	6.74	6.77	$[M+NH_4]^{+1}$	Glycerophospholipids (Glycerophosphocholines)
Pomace	PS (O-20:0/14:0)	$C_{40}H_{80}NO_9P$	4.07	749.56	750.567	6.74	6.77	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphoserines)
Pomace	PG (O-16:0/19:1(9Z))	$C_{41}H_{81}O_9P$	1.72	748.563	749.57	6.74	6.77	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphoglycerols)
Pomace	PI (14:0/22:2(13Z,16Z))	$C_{45}H_{83}O_{13}P$	4.58	862.561	861.554	6.74	6.77	$[M-H]^{-1}$	Glycerophospholipids (Glycerophosphoinositols)
Pomace	Man-Cer (d16:0(15Me(3OH))/ 14:0(13Me))	$C_{38}H_{76}NO_{11}P$	2.88	753.518	754.525	6.75	6.78	$[M+H]^{+1}$	Sphingolipids (Phosphosphingolipids)
Pomace	PG (P-20:0/21:0)	$C_{47}H_{93}O_9P$	1.08	832.65662	833.664	6.75	6.78	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphoglycerols)
Pomace	PI (18:0/20:4(5Z,8Z,11Z,14Z))	$C_{47}H_{83}O_{13}P$	3.69	886.56	887.568	6.94	6.94	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphoinositols)
Pomace	PA (22:1(11Z)/22:2(13Z,16Z))	$C_{47}H_{87}O_8P$	4.21	810.617	811.624	7.00	7.00	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphates)
Pomace	PA (O-16:0/20:2(11Z,14Z))	$C_{39}H_{75}O_7P$	2.5	686.527	687.534	7.06	7.01	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphates)
Pomace	PG (O-18:0/17:0)	$C_{41}H_{83}O_9P$	1.63	750.579	751.586	7.06	7.01	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphoglycerols)
Pomace	PI (14:0/22:1(11Z))	$C_{45}H_{85}O_{13}P$	3.88	864.576	863.569	7.06	7.01	$[M-H]^{-1}$	Glycerophospholipids (Glycerophosphoinositols)
Pomace	PI (17:0/22:1(11Z))	$C_{48}H_{91}O_{13}P$	3.74	906.623	905.616	7.06	7.01	$[M-H]^{-1}$	Glycerophospholipids (Glycerophosphoinositols)
Pomace	PI (O–18:0/20:5 (5Z,8Z,11Z,14Z,17Z))	$C_{47}H_{83}O_{12}P$	4.01	870.566	871.573	7.07	7.05	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphoinositols)
Pomace	PG (O-16:0/21:0)	$C_{43}H_{87}O_9P$	2.86	778.611	779.618	7.45	7.37	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphoglycerols)
Pomace	PG (20:0/22:1(11Z))	$C_{48}H_{93}O_{10}P$	−0.16	860.65	861.658	7.48	7.39	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphoglycerols)
Pomace	PC (O-24:0/20:4 (8Z,11Z,14Z,17Z))	$C_{52}H_{98}NO_7P$	1.19	879.709	880.716	8.52	8.28	$[M+H]^{+1}$	Glycerophospholipids (Glycerophosphocholines)

Abbreviations: Cer: Ceramide; GlcDG: glucopyranosyl-sn-glycerol; Man: mannose; MIPC: mannosylinositol phosphorylceramide; PA: phosphatidic acid; PC: phosphatidylcoline; PI: phosphatidylinositol; PG: phosphatidylglycerol; PS: phosphatidylserine; SM: Sphingomyelin; SOPC: 2-oleoyl-1-stearoylsn-san-glycero-3-phosphocholine.
**Chemical marker according to CART*

compounds) followed by PIs and phosphates (19 compounds). Also, phosphosphingolipids were detected (5 compounds) and one polyprenol. More detailed information can be found in table S1 of the supplementary material.

These compounds were specific of each type of oil evaluated and were detected in lampante olive oil (24 compounds), pomace olive oils (30 compounds) and other edible vegetable oils (30 compounds), as it can be seen in Table 3. Each compound PhLs was characterized by HRMS according to molecular weight, *m/z*, mass error lower than 5 ppm and RT in min (Table 3). Note that the chromatographic peaks of compounds that elute from 10 min to the end were removed from the fingerprint because they were ceramides and triacylglicerols and not PhLs.

Comparing these results with the scarce bibliography found, Boukhchina et al. [21] identified 5 types of PhLs in olive oil, and non-olive oil. Showing that PG, PE, and PC were the PhLs found in a major relative abundance, similar to the results obtained as indicated in Table S1, PCs the PhLs found in a high detection rate. Finally, Antonelli et al. [22] determined PhLs in olive oil samples, they found that PGs were the main PhLs found in olive oil samples, followed by phosphatidic acid (PA) and PIs. In our study, results are close similar where the second

and third specie found were PI and PA also. So, to conclude and due to the scarce and non-recently bibliography found, we can indicate that results are in concordance with previous ones. In addition, it is important to note that our work was the only one in which HRMS was employed for the characterization of PhLs, giving a high advance in the field of PhLs characterization with the power of mass resolution.

In addition, as an application example, a tentative identification of the PhL present in the same sample showed in Fig. 2 is given in the supplementary material (Table S2).

3.6.3. Tentative chemical markers

The CART classification model, as opposed to the other two (PLS-DA and SVM), misclassifies one sample of the validation set. Specifically, it is a sample of extra virgin olive oil, which is classified as non-olive. So chemical markers can be associated with olive and non-olive samples, according to the decision rule that was at SRS of 4.71 and 5.65. At SRS of 4.83, the compound 1-octadecyl-*sn*-glycero-3-phosphocholine was found and at 5.70, the compound PC (14:0/24:1(15Z)) both found in pomace samples. The second compound is a typical compound of olive oil samples and is in accordance with the CART law. However, 1-octadecyl-*sn*-glycero-3-phosphocholine was a tentative chemical marker and may not be reliably assigned as such because the difference between the values provided by the CART decision rule (4.71) and the SRS assigned to the compound (4.83) was too large (0.12).

4. Future trends and conclusions

In conclusion, this study successfully generated a distinctive UHPLC-UVD-CAD chromatographic fingerprint PhL fraction for both olive oils and other edible vegetable oils. Through the application of chemometrics such as PLS-DA, SVM, and CART, a robust multivariate model for authenticating olive oil was established. Additionally, the PhL fraction was characterised using UHPLC-(Q-Orbitrap)MS, providing valuable information and identifying two chemical markers of olive oils, one of which is fully reliable while the other could not be ensured. Furthermore, a comparative analysis among various edible vegetable oils based on this methodology was conducted, enhancing our understanding of their composition and differentiation. Lastly, the added value of using instrument-agnostic chromatograms of the PhL fraction for the first time should be highlighted. This opens the way to the generation of instrument-agnostic fingerprint databases leading to unique classification models that can be used by any laboratory without the need to generate a particular model each time an instrumental fingerprint is acquired. The proposal provides a solution to the major challenge of the widespread and universal use of the multivariate approach in the field of food quality and authenticity control.

CRediT authorship contribution statement

Rosalía López-Ruiz: Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis. **Ana M. Jimenez-Carvelo:** Writing – review & editing, Validation, Supervision, Software, Methodology, Investigation, Data curation, Conceptualization. **Luis Cuadros-Rodríguez:** Writing – review & editing, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

Grant CPP2021-008672 funded by MICIU/AEI/10.13039/501100011033 and, by the "European Union NextGenerationEU/PRTR". RLR acknowledges to the Andalusia Ministry of Economic Transformation, Industry, Knowledge and Universities for financial support from "Ayudas para Captación, Incorporación y Movilidad de Capital Humano de I + D + i (PAIDI 2020)". AMJC acknowledges the Grant (RYC2021-031993-I) funded by MCIN/AEI/501100011033 and "European Union NextGeneration EU/PRTR". Funding for open access charge: Universidad de Gra-nada / CBUA.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.microc.2024.110837.

References

[1] A. Gani, B.A. Ashwar, Food biopolymers: Structural, functional and nutraceutical properties, Springer International Publishing, 2021. Doi: 10.1007/978-3-030-27061-2.
[2] I. Criado-Navarro, A. Mena-Bravo, M. Calderón-Santiago, F. Priego-Capote, Determination of glycerophospholipids in vegetable edible oils: Proof of concept to discriminate olive oil categories, Food Chem. 299 (2019), https://doi.org/10.1016/j.foodchem.2019.125136.
[3] A.F. El Sheikha, Food authentication: Introduction, techniques, and prospects, in: C. Galanakis (Ed.), Food Authentication and Traceability, Elsevier, 2020: pp. 1–34. Doi: 10.1016/B978-0-12-821104-5.00006-4.
[4] Y. Lu, P. Li, H. Xu, A Food anti-counterfeiting traceability system based on blockchain and internet of things, Procedia Comput. Sci. 199 (2022) 629–636, https://doi.org/10.1016/j.procs.2022.01.077.
[5] J. Qian, L. Ruiz-Garcia, B. Fan, J.I. Robla Villalba, U. McCarthy, B. Zhang, Q. Yu, W. Wu, Food traceability system from governmental, corporate, and consumer perspectives in the European Union and China: A comparative review, Trends Food Sci. Technol. 99 (2020) 402–412, https://doi.org/10.1016/j.tifs.2020.03.025.
[6] A. de Araújo Gomes, S.M. Azcarate, I. Špánik, L. Khvalbota, H.C. Goicoechea, Pattern recognition techniques in food quality and authenticity: A guide on how to process multivariate data in food analysis, TrAC Trends Anal. Chem. 164 (2023) 117105, https://doi.org/10.1016/j.trac.2023.117105.
[7] L. Cuadros-Rodríguez, C. Ruiz-Samblás, L. Valverde-Som, E. Pérez-Castaño, A. González-Casado, Chromatographic fingerprinting: An innovative approach for food "identitation" and food authentication - A tutorial, Anal. Chim. Acta 909 (2016) 9–23, https://doi.org/10.1016/j.aca.2015.12.042.
[8] S. Medina, J.A. Pereira, P. Silva, R. Perestrelo, J.S. Câmara, Food fingerprints – A valuable tool to monitor food authenticity and safety, Food Chem. 278 (2019) 144–162, https://doi.org/10.1016/j.foodchem.2018.11.046.
[9] A.M. Jiménez-Carvelo, S. Martin-Torres, L. Cuadros-Rodríguez, A. González-Casado, Nontargeted fingerprinting approaches, in: Food Authentication and Traceability, Elsevier, 2020: pp. 163–193. Doi: 10.1016/B978-0-12-821104-5.00010-6.
[10] AOAC international, Food Authenticity, (2019). https://www.aoac.org/wp-content/uploads/2019/08/FAM_Program_Update-_July_2019-1.pdf (accessed April 10, 2024).
[11] L. Cuadros-Rodríguez, F. Ortega-Gavilán, S. Martín-Torres, S. Medina-Rodríguez, A.M. Jimenez-Carvelo, A. González-Casado, M.G. Bagur-González, Standardization of chromatographic signals – Part I: Towards obtaining instrument-agnostic fingerprints in gas chromatography, J. Chromatogr. A 1641 (2021), https://doi.org/10.1016/j.chroma.2021.461983.
[12] L. Cuadros-Rodríguez, S. Martín-Torres, F. Ortega-Gavilán, A.M. Jiménez-Carvelo, R. López-Ruiz, A. Garrido-Frenich, M.G. Bagur-González, A. González-Casado, Standardization of chromatographic signals – Part II: Expanding instrument-agnostic fingerprints to reverse phase liquid chromatography, J. Chromatogr. A 1641 (2021), https://doi.org/10.1016/j.chroma.2021.461973.
[13] G. Tomasi, F. Savorani, S.B. Engelsen, Icoshift: An effective tool for the alignment of chromatographic data, J. Chromatogr. A 1218 (2011) 7832–7840, https://doi.org/10.1016/j.chroma.2011.08.086.
[14] C.H. Pérez-Beltrán, A.M. Jiménez-Carvelo, S. Martín-Torres, F. Ortega-Gavilán, L. Cuadros-Rodríguez, Instrument-agnostic multivariate models from normal phase liquid chromatographic fingerprinting. A case study: Authentication of olive oil, Food Control 137 (2022), https://doi.org/10.1016/j.foodcont.2022.108957.
[15] R.W. Kennard, L.A. Stone, Computer Aided Design of Experiments, Technometrics 11 (1969) 137–148. Doi: 10.1080/00401706.1969.10490666.
[16] T. Kind, K.H. Liu, D.Y. Lee, B. Defelice, J.K. Meissen, O. Fiehn, LipidBlast in silico tandem mass spectrometry database for lipid identification, Nat. Methods 10 (2013) 755–758, https://doi.org/10.1038/nmeth.2551.
[17] M. Sud, E. Fahy, D. Cotter, A. Brown, E.A. Dennis, C.K. Glass, A.H. Merrill, R. C. Murphy, C.R.H. Raetz, D.W. Russell, S. Subramaniam, LMSD: LIPID MAPS structure database, Nucleic Acids Res. 35 (2007), https://doi.org/10.1093/nar/gkl838.

[18] R. Kiyonami, D.A. Peake, Y. Yokoi, K. Miller, Rapid Phospholipid Characterization Using a Novel Intelligent Workflow on a Tribrid Orbitrap Mass Spectrometer, 2015. http://tools.thermofisher.com/content/sfs/posters/PN-64451-LC-MS-Phospholipid-Characterization-ASMS2015-PN64451-EN.pdf (accessed July 7, 2023).
[19] M. Narváez-Rivas, Q. Zhang, Comprehensive untargeted lipidomic analysis using core-shell C30 particle column and high field orbitrap mass spectrometer, J. Chromatogr. A 1440 (2016) 123–134, https://doi.org/10.1016/j.chroma.2016.02.054.
[20] L. Cuadros-Rodríguez, E. Pérez-Castaño, C. Ruiz-Samblás, Quality performance metrics in multivariate classification methods for qualitative analysis, TrAC - Trends Anal. Chem. 80 (2016) 612–624, https://doi.org/10.1016/j.trac.2016.04.021.
[21] S. Boukhchina, K. Sebai, A. Cherif, H. Kallel, P.M. Mayer, Identification of glycerophospholipids in rapeseed, olive, almond, sunflower oil by LC-MS and LC-MS-MS, Can. J. Chem. 82 (2004) 1210–1215, https://doi.org/10.1139/v04-094.
[22] M. Antonelli, B. Benedetti, C. Cavaliere, A. Cerrato, C.M. Montone, S. Piovesana, A. Lagana, A.L. Capriotti, Phospholipidome of extra virgin olive oil: Development of a solid phase extraction protocol followed by liquid chromatography–high resolution mass spectrometry for its software-assisted identification, Food Chem. 310 (2020) 125860, https://doi.org/10.1016/j.foodchem.2019.125860.