CrossMark

RESEARCH PAPER

Olive oil authenticity studies by target and nontarget LC–QTOF-MS combined with advanced chemometric techniques

Natasa P. Kalogiouri[1] · Nikiforos A. Alygizakis[1] · Reza Aalizadeh[1] ·
Nikolaos S. Thomaidis[1]

Abstract Food analysis is continuously requiring the development of more robust, efficient, and cost-effective food authentication analytical methods to guarantee the safety, quality, and traceability of food commodities with respect to legislation and consumer demands. Hence, a novel reversed-phase ultra high performance liquid chromatography–electrospray ionization quadrupole time of flight tandem mass spectrometry analytical method was developed that uses target, suspect, and nontarget screening strategies coupled with advanced chemometric tools for the investigation of the authenticity of extra virgin olive oil. The proposed method was successfully applied in real olive oil samples for the identification of markers responsible for the sensory profile. The proposed target analytical method includes the determination of 14 phenolic compounds and demonstrated low limits of detection ranging from 0.015 μg mL^{-1} (apigenin) to 0.039 μg mL^{-1} (vanillin) and adequate recoveries (96–107 %). A suspect list of 60 relevant compounds was compiled, and suspect screening was then applied to all the samples. Semiquantitation of the suspect compounds was performed with the calibration curves of target compounds having similar structures. Then, a nontarget screening workflow was applied with the aim to identify additional compounds so as to differentiate extra virgin olive oils from defective olive oils. Robust classification-based models were built with the use of supervised discrimination techniques, partial least squares–discriminant analysis and counterpropagation artificial neural networks, for the classification of olive oils into extra virgin olive oils or defective olive oils. Variable importance in projection scores were calculated to select the most significant features that affect the discrimination. Overall, 51 compounds were identified and suggested as markers, among which 14, 26, and 11 compounds were identified by target, suspect, and nontarget screening respectively. Retrospective analysis was also performed and identified 19 free fatty acids.

Introduction

Food authenticity has become increasingly important in recent years because of the drive for more accurate and truthful labeling. A product is characterized as authentic as long as it firstly is described accurately by the label and secondly complies with the legislation in force in the country where it is marketed or sold [1, 2]. Authenticity is a multifaceted issue that covers many aspects, including characterization, adulteration, mislabeling, and misleading origin [3]. Thus, there is a growing necessity to develop advanced analytical methods to be used with appropriate data processing tools that could successfully guarantee the authenticity of various food matrices.

In that respect, there is emerging concern for the guarantee of the authenticity of olive oil because of its economic importance, as well its nutritional, sensory, and therapeutic properties, which have been extensively elaborated by science [4–6]. The main authenticity issues that are associated with olive oil quality are adulteration, misdescription of the geographical

✉ Nikolaos S. Thomaidis
ntho@chem.uoa.gr

[1] Laboratory of Analytical Chemistry, Department of Chemistry, National and Kapodistrian University of Athens, Panepistimioupolis Zografou, 15771 Athens, Greece

origin, the production type (conventional or organic), and the taste. The latter is the result of certain constituents present in olive oil that affect its sensory profile. According to the International Olive Council [7] and its trade standards (Sensory analysis of olive oil. Method for the organoleptic assessment of virgin olive oil. COI/T.20/Doc. No 15/Rev. 8, 2015), there are three positive attributes for extra virgin and virgin olive oils (fruity, bitter, and pungent) and 16 negative attributes (fusty/muddy, musty/humid/earthy, winey/vinegary, acid/sour, rancid, frostbitten olives, heated or burned, hay/wood, rough, greasy, vegetable water, brine, metallic, esparto, grubby, and cucumber). The presence of positive sensory characteristics is necessary for the classification of olive oils as "extra virgin," whereas those with negative attributes have objectionable taste and are characterized as "defective." The official method for sensory evaluation of olive oil is implemented in a panel test developed by the International Olive Oil Council [7] that is filled out by trained tasters (COI/T.20/ Doc. No 14/ Rev. 4) [8]. However, as Tena et al. [9] have recently reviewed, the official method is questioned by numerous olive oil sectors and it fails in cases where testers are not able to analyze defects at very low intensities. Moreover, it has several drawbacks since it is time-consuming, it lacks stable and standardized reference oils with different intensities of bitterness and pungency, and it also requires a group of 8 to 12 testers for statistically confirmed results [10, 11]. Therefore, as recently suggested by García-González and Aparicio [11] and Tena et al. [9], an objective measurement of virgin olive oil sensory quality should follow another strategy based on analytical chemistry.

Many analytical procedures have been used to identify and quantify the volatile components that characterize olive oil flavor/aroma in the past 30 years. Among them, gas chromatography is the main technique applied for this purpose, as reviewed by Escuderos [12]. Apart from the volatile fraction, however, another group of compounds widely known as "bioactive constituents," mainly consisting of phenolic compounds, has been reported as important for the flavor of olive oil [13]. Consequently, their detection and identification in olive oils constitutes a challenging field that should be further developed.

In this field, high-resolution mass spectrometry (HRMS) has proved its excellent analytical performance, allowing the analysis of a wide range of compounds in food, providing screening and tentative identification for both nontarget and target compounds [1, 14, 15]. In olive oil analysis, several studies have been published regarding the qualitative and quantitative analysis of bioactive constituents with liquid chromatography (LC)–HRMS, focusing in most cases on geographical origin and varietal discrimination [4, 6, 16–22]. Still, there is minor information for the sensory discrimination between extra virgin and defective olive oils. It has been suggested that certain phenolic compounds, and more specifically certain secoiridoid derivatives, such as oleuropein and lingstroside derivatives, are responsible for the bitter taste [23]. Nevertheless, the relationship between the

individual hydrophilic phenols and olive oil's sensory characteristics has not been clearly defined [11], and there is still controversy about which individual phenols are the main contributors to the taste attributes [11, 24]. Recent physicochemical and high-pressure LC (HPLC) methods for evaluation of the bitterness of olive oil have produced inconsistent results with respect to the influence of different phenols [23, 25–27]. Even though Andrewes et al. [28] have suggested that decarboxymethyl lingstroside aglycone is a pungent compound, correlation between quantitative and sensory data has not been found. Moreover, Dierkes et al. [10] developed a target HPLC–HRMS profiling method to identify several relevant bitter and pungent components, but no correlation between the total phenolic content and the bitterness/pungency ratio could be found. These gaps in the scientific literature concerning olive oil's organoleptic characteristics and their correlation with certain compounds (markers) could be fulfilled with the use of HRMS nontargeted analytical approaches.

Nontargeted methods combined with suitable chemometric tools increase the breadth of traditional targeted analysis and accelerate new prospects for novel applications [14, 15, 29]. The coupling of time-of-flight (TOF) mass spectrometry (MS) with LC has proved its excellent analytical performance and offers a good combination of selectivity and sensitivity at high resolution and subsecond scan speeds [30]. Since use of LC–HRMS could result in the generation and detection of a large number of features (m/z), it would be a challenging task to identify them to investigate the authenticity of olive oil. In addition, the coupling of LC–HRMS with chemometric tools could decrease remarkably the number of detected features and introduce the most meaningful m/z that could discriminate between extra virgin olive oils and defective olive oils [31].

Therefore, the primary purpose of the present work is the development and application of an integrated LC–HRMS workflow, including target, suspect, and nontarget screening approaches, coupled with supervised pattern recognition techniques, for olive oil fingerprinting. For that purpose, we developed a target quantitative method for the determination of 15 compounds and a suspect screening method with a list of 60 compounds coupled with a semiquantitative method for the identified compounds. The identification workflow included strict rule-based filtering steps, deep interpretation of MS/MS spectra, and retention time prediction. Then, a nontarget screening workflow was applied to establish extensive and reliable pattern recognition models for olive oil fingerprinting by classification of olive oil samples into extra virgin olive oils and defective olive oils. The variable importance in projection (VIP) score was calculated to select the most significant features that affect the discrimination.

Materials and methods

Chemicals and standards

All standards and reagents were of high purity (more than 95 %). Methanol of LC–MS grade and sodium hydroxide (purity greater than 99 %) were purchased from Merck (Darmstadt, Germany). Ammonium acetate (purity 99.0 % or greater) for HPLC and formic acid (LC–MS Ultra) were purchased from Fluka (Buchs, Switzerland). 2-Propanol was purchased from Fisher Scientific (Geel, Belgium). Distilled water was provided by a Milli-Q purification apparatus (Direct-Q UV, Millipore, Bedford, MA, USA). For the analytical method validation, the following reagents were used: syringic acid (purity 95 %) was purchased from Extrasynthèse (Genay, France), gallic acid (purity 98 %), ferulic acid (purity 98 %), epicatechin (purity 97 %), p-coumaric acid (4-hydroxycinnamic acid; purity 98 %), homovanillic acid (purity 97 %), quercetin (purity 98 %), oleuropein (purity 98 %), and pinoresinol (purity 95 %) were obtained from Sigma-Aldrich (Steinheim, Germany), hydroxytyrosol (purity 98 %) was purchased from Santa Cruz Biotechnologies, and caffeic acid (purity 99 %; internal standard), vanillin (purity 99 %), ethyl vanillin (purity 98 %), apigenin (4,5,7-trihydroxyflavone; purity 97 %), and tyrosol [2-(4-hydroxyphenyl) ethanol, purity 98 %] were acquired from Alfa Aesar (Karlsruhe, Germany). To confirm the identity of suspect and nontarget compounds, luteolin (purity 98 %) was acquired from Santa Cruz Biotechnologies. For the determination of free fatty acids, hexanoic acid (purity 99.5 %), octanoic acid (purity 99.5 %), dodecanoic acid (purity 99.5 %), myristic acid (purity 99.5 %), pentadecanoic acid (purity 99.5 %), palmitic acid (purity 99 %), palmitoleic acid (purity 98.5 %), heptadecanoic acid (purity 98 %), heptadecenoic acid (purity 99 %), stearic acid (purity 98.5 %), oleic acid (purity 99 %), α-linoleic acid (purity

99 %), α-linolenic acid (purity 99 %), arachidic acid (purity 99 %), cis-eicosenoic acid (purity 99 %), heneicosanoic acid (purity 99 %), docosanoic acid (purity 99 %), tricosanoic acid (purity 99 %), and lignoceric acid (purity 99 %) were purchased from Sigma-Aldrich. Stock standard solutions of individual compounds (1000 μg mL^{-1}) were solubilized in methanol and stored at -20 °C in dark brown glass bottles. All intermediate standard solutions containing the analytes were prepared by dilution of the stock solutions in methanol.

Olive oil samples

Three standard defective olive oil samples with a known score (Rancid, Fusty, and Musty) were acquired from the International Olive Council and 3 defective and 16 extra virgin olive oils of the Kolovi and Adramitiani varieties, both monovarietal and mixtures, were provided along with the sensory evaluation by ELGO-DIMITRA I.O.S.V. on Lesvos. These samples were produced from olives during the harvesting period in 2014–2015 and cultivated in different regions on Lesvos. To provide the sensory evaluation, the oils were subjected to an extended panel based on EU Regulation No. 1348/2013 [32] and International Olive Council instructions [8]. The results are expressed as the median of the rates reported by eight analysts. The highest mean coefficient of variation in all cases was less than 20 %. Figure 1 describes the sensory profile of the extra virgin olive oils and defective olive oils, represented as spider plots. More information about the samples concerning the exact organoleptic scores, the geographical origin, the variety, the production type, and the time of harvest can be found in the electronic supplementary material (Tables S1a, S1b).

All samples were protected from light and humidity and stored in dark glass bottles at the ideal temperature of 14–15 °C [33]. Moreover, to better preserve the quality of the
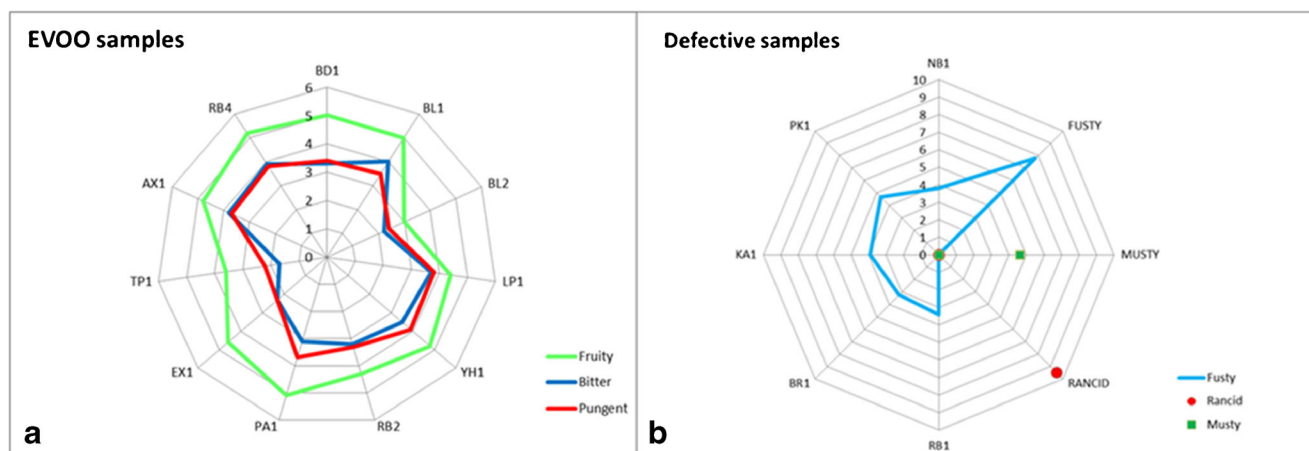

Fig. 1 Spider plots describing the organoleptic profile of olive oil samples with sensory attributes for **a** extra virgin olive oil (EVOO) samples (fruity, bitter, pungent) and **b** defective olive oil samples (fusty, rancid, musty)

olive oils and increase the resistance to autoxidation, nitrogen as an inert gas was added to the bottles [34].

Sample extraction and quality control

A liquid–liquid microextraction method as developed and validated by Becerra-Herrera et al. [35] was used to isolate the phenolic compounds from the olive oil samples. For this, 0.5 g of each sample was weighed and 0.5 mL of methanol–water (80:20, v/v) was added to 2-mL Eppendorf tubes. The mixture was then vortexed for 2 min, and centrifuged for 5 min at 13,400 rpm. Furthermore, the upper phase, which consisted of methanol, was collected and filtered through membrane syringe filters of regenerated cellulose (CHROMAFIL® RC) (15-mm diameter, 0.22-µm pore size, provided by Macherey-Nagel, Düren, Germany). Then, 200 µL of the methanolic phase was diluted with ultrapure water to 0.5 mL. Finally, 5 µL of this solution was injected into the chromatographic system. Procedural blanks were also prepared and processed in the chromatographic system to detect any potential contamination. Quality control samples were used to verify that the analytical system had been stabilized before analysis of the main batch of samples and to assess its performance. A typical quality control sample was prepared, as suggested by Want et al. [36], by our mixing all aliquots of the samples. At the beginning of the analysis, the quantity control sample was injected five times for conditioning and afterward it was injected at regular intervals (i.e., every ten sample injections) throughout the analytical run to provide a set of data from which repeatability can be assessed. The calculated relative standard deviations (RSDs) for the retention time (t_R) and the peak areas as well as Δm errors ($n = 7$) are presented in the electronic supplementary material (Table S2), proving the good performance of the analytical system.

Reversed-phase ultra high performance liquid chromatography–electrospray ionization quadrupole time-of-flight tandem mass spectrometry analysis

Reversed-phase (RP) chromatographic analysis was performed with an ultra high performance LC (UHPLC) system with an HPG-3400 pump (Dionex UltiMate 3000 RSLC, Thermo Fisher Scientific, Germany) interfaced with a quadrupole TOF (QTOF) mass spectrometer (Maxis Impact, Bruker Daltonics, Bremen, Germany) in negative electrospray ionization mode. Separation was performed with an Acclaim RSLC C_{18} column (2.1 mm × 100 mm, 2.2 µm) purchased from Thermo Fisher Scientific (Driesch, Germany) with an ACQUITY UPLC BEH C_{18} precolumn (1.7 µm, VanGuard precolumn, Waters, Ireland). The separation was performed at a column temperature of 30 °C. The solvents used consisted of 90 % water, 10 % methanol, and 5 mM ammonium acetate (solvent A), and 100 % methanol and 5 mM ammonium acetate (solvent B). The elution gradient adopted started with 1 % of organic phase B (flow rate 0.2 mL min⁻¹) for 1 min,

gradually increasing to 39 % in the next 2 min, and then increasing to 99.9 % (flow rate 0.4 mL min⁻¹) in the following 11 min. These almost pure organic conditions were kept constant for 2 min (flow rate 0.48 mL min⁻¹), and then the initial conditions (1 % solvent B, 99 % solvent A) were restored within 0.1 min (flow rate decreased to 0.2 mL min⁻¹) to reequilibrate the column for the next injection.

The QTOF MS system was equipped with an electrospray ionization (ESI) interface, operating in negative mode with the following settings: capillary voltage of 3500 V; end plate offset of 500 V; nebulizer pressure of 2 bar (N_2); drying gas flow rate of 8 L min⁻¹ (N_2); and drying temperature of 200 °C. A QTOF external calibration was performed daily with a sodium formate cluster solution, and a segment (0.1 − 0.25 min) in every chromatogram was used for internal calibration, with use of a calibrant injection at the beginning of each run. The sodium formate calibration mixture consisted of 10 mM sodium formate in a mixture of water and 2-propanol (1:1). Full-scan mass spectra were recorded in the range from 50 to 1000 m/z, with a scan rate of 2 Hz. MS/MS experiments were conducted with use of AutoMS data-dependent acquisition mode based on the fragmentation of the five most abundant precursor ions per scan. For certain masses of interest, if the intensity of the m/z was low, a second analysis including the list of the selected precursor ions was performed in AutoMS (data-dependent acquisition) mode. The instrument provided a typical resolving power (full width at half maximum) between 36,000 and 40,000 at m/z 226.1593, 430.9137, and 702.8636.

Screening strategies

A target list was created that included 15 significant phenolic compounds that have been identified in olive oil and have been reported in the literature [4–6, 10, 16–22]. The list consisted of different classes of phenolic compounds, such as phenolic acids, secoiridoids, flavonoids, and lignans, with commercially available standards. The initial target list can be found in the electronic supplementary material (Table S3a). A suspect list was also generated from the literature and included all the phenolic compounds that have already been identified in olive oil and in different organs of *Olea europaea* (stems, leaves, drupes) so we could scan the olive oil samples for their presence. The initial suspect list consisted of 60 bioactive constituents and is presented in the electronic supplementary material with the molecular formulas and the simplified molecular-input line-entry system (SMILES) formulas of the suspect compounds, as well as the references to the studies in which they have previously been reported (Table S3b).

Target, suspect, and nontarget screening workflows were followed as they have been suggested by Krauss et al. [37] and Gago-Ferrero et al. [38]. Target screening was performed with the software packages Target Analysis 1.3 and Data Analysis 4.1 (Bruker Daltonics, Bremen, Germany), as well as other

tools available in these packages (Bruker Compass Isotope Pattern and SmartFormula Manually). Extracted ion chromatograms (EICs) were obtained with use of the function Find Compounds-Chromatogram (in the Target Analysis software package), which creates the base peak chromatograms for the masses that achieve thresholds of intensity and accuracy according to the following parameters that were set: mass accuracy window of 2 mDa, a satisfactory isotopic fit was denoted only when mSigma (mSigma is a measure of the goodness of fit between the measured and the theoretical isotopic pattern) was below or equal to 50, signal to noise threshold of 3, minimum area threshold of 800, and minimum intensity threshold of 200. The relative tolerance of the retention time window was set lower than ±0.2 min. All the target compounds that were included in the database were identified on the basis of mass accuracy, isotopic pattern, retention time (t_R), and MS/MS fragments.

For the identification of the suspect compounds, the masses of the deprotonated ions were calculated on the basis of the molecular formula, and EICs were created in Target Analysis 1.3 with the following parameters: mass accuracy threshold of 2 mDa, isotopic fit below or equal to 50, ion intensity of more than 800, peak area threshold of 2000, and peak score (area/intensity ratio) of more than 4 (the peak score should preferably be between 4 and 38). These parameters have already been optimized for suspect identification by our group [38]. If one or more peaks were detected with use of EICs, the isotopic pattern and the MS/MS fragments were examined in Data Analysis 4.1 to confirm that the peak represents the suspect compound. The comparison and interpretation of the MS/MS fragments were performed with use of literature data and in silico fragmentation tools, mainly Metfrag [39] and spectral libraries such as MassBank [40]. Moreover, the possible retention time of each suspect compound was predicted and compared with the experimental retention time by a model developed in-house that was based on the quantitative structure–retention relationship (QSRR) [41] since reference standard solutions were not commercially available for most of the suspect compounds. More information on the development and optimization of the support vector machine QSRR model can be found in the electronic supplementary material (Section S6).

Following the suspect screening, nontarget screening was performed. Nontarget screening involves the detection of peaks and the identification of compounds without one having a priori information or available standards [42, 43]. Peak picking was performed as explained in detail in "Data processing and chemometrics." The selected peaks were tentatively identified according to the mass accuracy (less than 2 mDa) and isotopic pattern of the precursor ion (less than 50 mSigma), their fragmentation pattern, and the retention time of the extracted ion chromatographic peak. Elemental compositions of the precursor and fragment ions were suggested, and plausible

molecular formulas were proposed with use of the Smart Formula tool in Data Analysis 4.1. MS/MS spectra were examined and interpreted as discussed for suspect screening to determine tentative candidates. The QSRR prediction model was also used as a complementary tool for the identification of the nontarget compounds in cases where there were no standards available.

The level of confidence achieved in the identification of the detected compounds was established according to Schymanski et al. [44]. Level 1 corresponds to confirmed structures where a reference standard is available, level 2 corresponds to probable structures (level 2a, evidence by matching spectra with spectra from the literature or a library; level 2b, diagnostic evidence where no other structure fits the experimental MS/MS information), level 3 corresponds to a tentative candidate or candidates, level 4 corresponds to unequivocal molecular formulas, and level 5 corresponds to the exact mass(es) of interest. The detected compounds were labeled on the basis of this classification.

Data processing and chemometrics

To process LC–HRMS data, first, binary files of all the analyzed samples were converted to mzXML files with use of the software program Proteowizard [45]. Then, these files were processed with the R language and the XCMS package [46] to extract peaks that are present in the samples. This procedure involved peak picking by the CentWave algorithm [46], grouping of peaks representing the same analyte across the samples, and a step to correct the chromatographic drift of the retention time. The next step was to regroup the peaks for which the retention time was changed. In the final step, a step of filling in the missing peaks was implemented. This replaces the missing values of nondetected peaks with a small value of the intensity [47]. The CAMERA package was also used for deisotoping and removing the adduct peaks to avoid collinearity during the model construction [48]. The internal parameters of the algorithms used for peak peaking, grouping, and retention time alignment were optimized with the package IPO [49]. The optimal settings are presented in detail in the electronic supplementary material (Table S4). Overall, 304 molecular features were obtained and grouped for 19 olive oil samples. These samples were split into a training set and a test set randomly so as to generate the classification models and then evaluate the accuracy of the classification for the external set of samples. Multivariate classification methods such as partial least squares–discriminant analysis (PLS-DA) [50, 51] and self-organizing maps [52], which are supervised pattern recognition techniques, were used to classify the olive oils into extra virgin olive oil and defective olive oil samples and investigate, subsequently, the relationship between the samples. The VIP method [53] was applied to distinguish and to detect the most important compounds responsible for

discrimination. VIP scores estimate the importance of each variable (in this case m/z) in the projection used to build the PLS-DA model and could be useful criteria to select significant m/z [53, 54]. The VIP score is shows the contribution of a variable (m/z) in the final latent variables. To prioritize the peaks that caused greater variation in the discrimination between samples, VIP scores were calculated for the PLS-DA model. Those m/z with a VIP score greater than 0.83 were considered as the most important because they cause greater variation [52], and the nontarget identification workflow was applied for their identification, as described in "Screening strategies."

Results and discussion

Target screening results

A data-dependent method was used to scan real olive oil samples for the presence of target compounds, and the presence of 14 target compounds was determined. They were ferulic acid, gallic acid, homovanillic acid, p-coumaric acid and syringic acid from the phenolic acids, tyrosol and hydroxytyrosol from the phenolic alcohols, vanillin and ethyl vanillin from the phenolic aldehydes, apigenin, quercetin, and epicatechin from the flavonoids, pinoresinol, which is a lignan, and the secoiridoid oleuropein. The mass accuracies of the precursor ions as well as those of the qualifiers of the detected compounds were less than 2 mDa compared with standard solutions and the isotopic fit was less than 50 mSigma in all cases. The most abundant fragments provided by the MS/MS (AutoMS) spectra were confirmed with use of Metfrag [39] as well as literature records. Target screening results are summarized in Table 1.

After verification and confirmatory analysis, representative qualifier ions of the detected compounds were cross-verified with fragments presented in previous studies. The MS/MS fragments of the target compounds are presented in Table 1. The MS/MS spectrum of quercetin shows a fragment at m/z 151.0024, corresponding to $C_7H_3O_4$ [5, 55]. The qualifier ions of oleuropein, m/z 307.0823 and m/z 377.1241, corresponding to $C_{15}H_{15}O_7$ and $C_{19}H_{21}O_8$ respectively, have also been reported by Kanakis et al. [56]. Pinoresinol shows characteristic fragmentation at m/z 151.0399, corresponding to $C_8H_7O_3$ [35]. The EICs of the target compounds identified are presented in Fig. 2.

The accuracy of the proposed RP-UHPLC–ESI-MS target method was examined to ensure that it is suitable for identification and quantification purposes. All the analytical parameters, including precision (RSD), limits of detection and quantification [57], linearity (calibration curves and regression coefficient, r^2), recovery, and matrix effect [58] were calculated and are presented in Table 2. Method precision expressed as

intraday and interday precision was evaluated for each phenolic compound at a spiked concentration of 0.5 μg mL^{-1} in terms of the RSD. The linearity of the method and the linearity of the standard calibration curves were evaluated and established by injection of standard solutions at ten different concentrations between 0.05 and 5 μg mL^{-1} in olive oil extracts. Caffeic acid was used as an internal standard at a concentration of 0.5 μg mL^{-1} instead of syringic acid, which is used in the official method of the International Olive Council for the determination of biophenols [59], since caffeic acid was not detected in any of the samples, in contrast to syringic acid, which was detected in most of the samples.

Calibration curves were constructed with use of the peak area of the analyte divided by the peak area of the internal standard and were linear, with $r^2 > 0.99$ in all cases. The precision limit was 4.9 % or less RSD for intraday experiments and 6.4 % or less RSD for interday experiments, indicating the good precision of the method developed. The limits of detection and quantification were adequate and ranged between 0.015 μg mL^{-1} (apigenin) and 0.034 μg mL^{-1} (vanillin) and 0.046 and 0.091 μg mL^{-1} respectively. The analytes showed satisfying recovery efficiency (96–108 %). Low matrix suppression was observed for all the phenolic compounds (matrix effect 92–99 %).

In the next step, the target compounds detected were quantified in all samples, with our taking into consideration that quantitative analysis is crucial to provide a comprehensive overview of the phenolic composition of extra virgin olive oils. The concentrations of ferulic acid, gallic acid, homovanillic acid, p-coumaric acid, syringic acid, tyrosol, hydroxytyrosol, vanillin, ethyl vanillin, apigenin, luteolin, quercetin, epicatechin, pinoresinol, and oleuropein were quantified from their corresponding calibration curves with use of caffeic acid as the internal standard. Quantitative results for the target compounds can be found in the electronic supplementary material (Table S5a; the results are expressed in milligrams per kilogram as the mean values ± the standard deviation, $n = 3$).

Suspect screening

In suspect screening, 26 bioactive compounds out of the 60 on the initial suspect list were tentatively identified in real olive oil samples with ion intensities above 800 and peak areas of more than 2000 in all cases. The results showed high mass accuracy (less than 2 mDa) and acceptable isotopic fit values (less than 50 mSigma). The peak score describes the peak area/peak intensity ratio and was calculated to lie in the range from 4 to 19 for all the suspect compounds [38]. MS/MS spectra were examined and verified with MetFrag [39] as well as literature records. Table 3 summarizes the suspect screening results, providing information about the identification criteria and the level of identification of each compound. The QSRR

Table 1 Target screening results

Compound	Molecular formula	[M−H]⁻ m/z standard	[M−H]⁻ m/z experimental	t_R (min)	Δt_R (min)	Fragment m/z	Elemental formula
Vanillin	$C_8H_8O_3$	151.0400	151.0400	4.73	+0.02	71.0140	$C_3H_3O_2$
						95.0140	$C_5H_3O_2$
						108.0217	$C_6H_4O_2$
						136.0162	$C_7H_4O_3$
Apigenin	$C_{15}H_{10}O_5$	269.0455	269.0453	8.24	0	149.0248	$C_8H_5O_3$
						151.0037	$C_7H_3O_4$
Caffeic acid	$C_9H_8O_4$	ND	–	–	–	–	–
Epicatechin	$C_{15}H_{14}O_6$	289.0716	289.0716	4.37	+0.02	137.0248	$C_7H_5O_3$
						151.0416	$C_8H_7O_3$
Ethyl vanillin	$C_9H_{10}O_3$	165.0557	165.0558	5.44	-0.01	67.0190	C_4H_3O
						92.0266	C_6H_4O
						108.0215	$C_6H_4O_2$
						137.0245	$C_7H_5O_3$
Ferulic acid	$C_{10}H_{10}O_4$	193.0506	193.0506	1.40	+0.01	134.0370	$C_8H_6O_2$
						178.0271	$C_9H_6O_4$
Gallic acid	$C_7H_6O_5$	169.0142	169.0140	1.25	+0.01	125.0246	$C_6H_5O_3$
Homovanillic acid	$C_9H_{10}O_4$	181.0506	181.0506	1.50	+0.02	59.0134	$C_2H_3O_2$
						122.0369	$C_7H_6O_2$
						137.0610	$C_8H_9O_2$
						154.0266	$C_7H_6O_4$
Hydroxytyrosol	$C_8H_{10}O_3$	153.0557	153.0557	3.53	+0.01	123.0446	$C_7H_7O_2$
Oleuropein	$C_{25}H_{32}O_{13}$	539.1770	539.1770	5.96	0	59.0138	$C_2H_3O_2$
						89.0244	$C_3H_5O_3$
						101.0242	$C_4H_5O_3$
						111.0083	$C_5H_3O_3$
						121.0295	$C_7H_5O_2$
						307.0823	$C_{15}H_{15}O_7$
						377.1242	$C_{19}H_{21}O_8$
p-Coumaric acid	$C_9H_8O_3$	163.0400	163.0402	1.34	0	119.0506	C_8H_7O
						93.0349	C_6H_5O
Pinoresinol	$C_{20}H_{22}O_6$	357.1343	357.1340	6.49	+0.01	151.0399	$C_8H_7O_3$
Quercetin	$C_{15}H_{10}O_7$	301.0353	301.0350	7.42	-0.04	121.0294	$C_7H_5O_2$
						151.0024	$C_7H_3O_4$
Syringic acid	$C_9H_{10}O_5$	197.0455	197.0454	1.44	+0.01	182.0218	$C_8H_6O_5$
Tyrosol	$C_8H_{10}O_2$	137.0608	137.0610	4.07	-0.01	81.0262	C_5H_5O
						93.0345	C_6H_5O
						112.0530	$C_6H_8O_2$

ND not detected

model was used for the prediction of the possible retention time in cases where no reference standards were available (see the electronic supplementary material, Section S6). The difference between the experimental retention time and the predicted retention time was less than 1 min for all the suspect compounds, except for syringaresinol (2-min difference). More information about the QSRR model and the applicability domain for the suspect compounds can be found in the electronic supplementary material (Section S7).

The initial suspect list consisted mainly of all the possible secoiridoid derivatives of oleuropein because oleuropein is the major secoiridoid found in the pulp of olives and its concentration decreases during the maturation process as derivatives are formed. Oleuropein aglycone, 10-hydroxyoleuropein aglycone, methyl oleuropein aglycone, 10-hydroxy-10-methyl oleuropein aglycone, 10-hydroxydecarboxymethyl oleuropein aglycone, lingstroside aglycone, decarboxymethyl oleuropein aglycone (oleacein), and decarboxymethyl
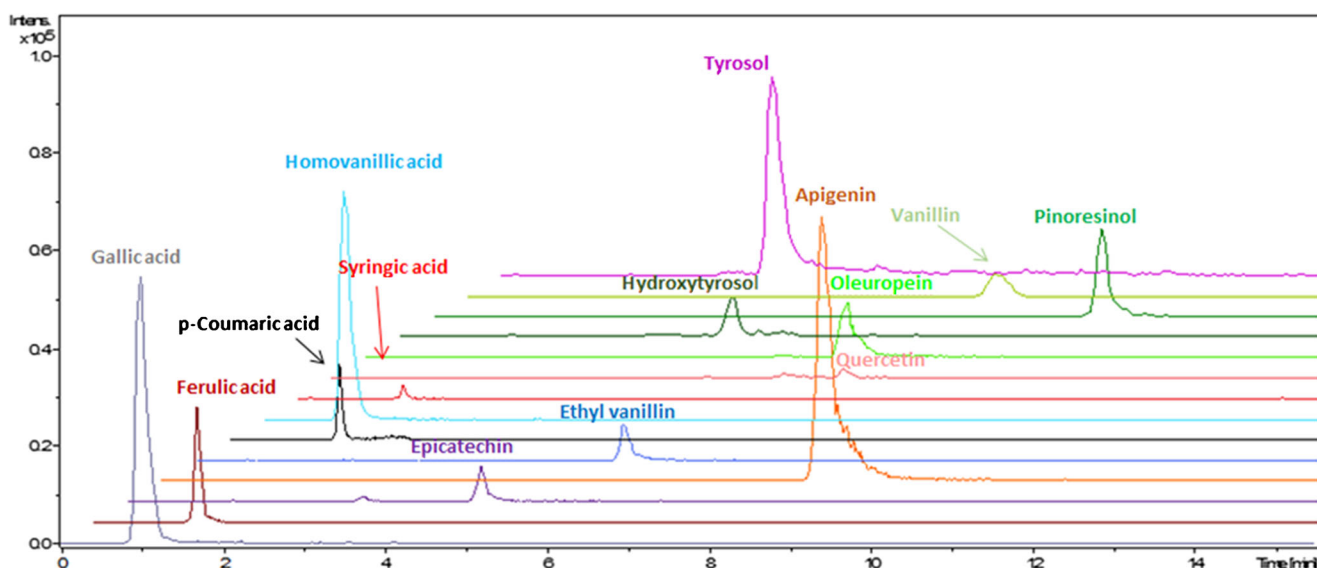
Fig. 2 Extracted ion chromatogram of the target analytes in an extra virgin olive oil sample

lingstroside aglycone (oleocanthal) were tentatively identified at level 2 (level 2a or level 2b as summarized in Table 2). One isomer of oleuropein aglycone was identified at level 3. The identification of oleuropein aglycone, lingstroside aglycone, oleacein, and oleocanthal is of high importance because they have been correlated with the positive attributes of bitter and pungent taste [10]. Moreover, oleacein and oleocanthal are both considered important because of their decisive role in health protection [60]. Studies have demonstrated that oleacein exhibits anti-inflammatory and antimicrobial activity and skin protection properties and reduces disorders due to

metabolic syndrome [4], whereas oleocanthal exhibits breast anticancer and potent antioxidant activity [61]. The precursor ions of oleacein and oleocanthal were detected at m/z 319.1185 and m/z 303.1237 respectively. Both compounds appear as a single broad peak in the EICs of the full-scan (AutoMS) spectrum. The MS/MS spectrum of oleacein is presented in Fig. 3. The qualifier ions detected at m/z 69.0342, 95.0502, 139.0608, 183.0660, and 195.0656 correspond to C_4H_5O, C_6H_7O, $C_8H_{11}O_2$, $C_9H_{11}O_4$, and $C_{10}H_{11}O_4$ respectively, and they have also been reported by Dierkes et al. [10]. The peak at m/z 165.0556 corresponds to $C_9H_9O_3$ [55, 56].

Table 2 Results of the validation of the target screening method

Compound	LOD (μg mL^{-1})	LOQ (μg mL^{-1})	Intraday precision RSD, (%) ($n = 10$)	Interday precision RSD (%) ($n = 2 \times 3$)	Equation	r^2	Recovery (%)	Matrix effect (%)
Gallic acid	0.029	0.091	4.5	5.2	$y = 0.1143x + 0.0242$	0.995	104	98
p-Coumaric acid	0.030	0.092	1.2	4.1	$y = 0.0307x + 0.0097$	0.993	97	95
Ferulic acid	0.034	0.086	2.2	2.7	$y = 0.0322x + 0.0261$	0.995	101	99
Syringic acid	0.029	0.090	2.1	2.4	$y = 0.0671x + 0.0064$	0.992	101	98
Homovanillic acid	0.027	0.089	2.4	6.3	$y = 0.0222x - 0.0008$	0.995	100	94
Tyrosol	0.030	0.091	4.1	3.2	$y = 0.0411x - 0.0055$	0.997	97	96
Hydroxytyrosol	0.034	0.077	2.9	4.3	$y = 0.0847x - 0.0225$	0.997	105	97
Pinoresinol	0.023	0.075	2.1	2.7	$y = 0.0303x - 0.0054$	0.997	107	98
Apigenin	0.015	0.046	3.5	4.1	$y = 0.5239x + 0.1027$	0.996	102	99
Oleuropein	0.022	0.068	4.9	2.9	$y = 0.1264x - 0.0192$	0.998	98	97
Vanillin	0.039	0.078	1.1	4.9	$y = 0.1134x - 0.0215$	0.996	96	92
Ethyl vanillin	0.028	0.084	3.7	6.3	$y = 0.1612x - 0.0267$	0.996	100	97
Epicatechin	0.029	0.088	4.8	5.5	$y = 0.0956x - 0.0154$	0.993	96	94
Quercetin	0.023	0.069	2.9	4.3	$y = 0.0303x - 0.0051$	0.995	98	95

LOD limit of detection, *LOQ* limit of quantification, *RSD* relative standard deviation

Table 3 Compounds identified through suspect screening, along with the identification criteria and the level of identification

Compound	Molecular formula	[M−H]$^-$ m/z calculated	[M−H]$^-$ m/z experimental	t_R (min)	t_R predicted (min)	Fragment m/z	Elemental formula	Peak score[c]	Level of identification[d]
Elenolic acid	$C_{11}H_{14}O_6$	241.0714	241.0714	4.51	4.26	59.0137	$C_2H_3O_2$	10	2a
						95.0496	C_6H_7O		
						127.0400	$C_6H_7O_3$		
						151.0402	$C_8H_7O_3$		
						171.0300	$C_7H_7O_5$		
Elenolic acid isomer	$C_{11}H_{14}O_6$	241.0722	241.0714	3.92	–[a]	95.0494	C_6H_7O	4	3
Elenolic acid isomer II	$C_{11}H_{14}O_6$	241.0722	241.0714	1.31	–[a]	95.0494	C_6H_7O	6	3
Elenolic acid methyl ester	$C_{12}H_{16}O_6$	255.0874	255.0874	4.58	4.53	67.0189	C_4H_3O	8	2a
						69.0346	C_4H_5O		
						185.0455	$C_8H_9O_5$		
Luteolin	$C_{15}H_{10}O_6$	285.0404	285.0404	7.55	Confirmed with a standard	133.0295	$C_8H_5O_2$	12	1
						151.0036	$C_7H_3O_4$		
Hydroxylated form of elenolic acid	$C_{11}H_{14}O_7$	257.0667	257.0667	1.36	–[a]	59.0104	$C_2H_3O_2$	11	2a
						137.0603	$C_8H_9O_2$		
						181.0535	$C_9H_9O_4$		
Hydroxylated form of elenolic acid isomer	$C_{11}H_{14}O_7$	257.0667	257.0665	1.41	–[a]	–	–	8	3
Hydroxytyrosol acetate	$C_{10}H_{12}O_4$	195.0660	195.0662	6.71	6.48	134.0373	$C_8H_6O_2$	15	2b
						149.0608	$C_9H_9O_2$		
						161.0246	$C_9H_5O_3$		
Hydroxytyrosol acetate isomer	$C_{10}H_{12}O_4$	195.0660	195.0662	5.74	6.48	59.0136	$C_2H_3O_2$	12	3
						134.0373	$C_8H_6O_2$		
						161.0246	$C_9H_5O_3$		
Decarboxymethyl oleuropein aglycone	$C_{17}H_{20}O_6$	319.1185	319.1185	5.61	6.14	69.0342	C_4H_5O	16	2a
						95.0502	C_6H_7O		
						123.0451	$C_7H_7O_2$		
						139.0608	$C_8H_{11}O_2$		
						165.0556	$C_9H_9O_3$		
						183.0660	$C_9H_{11}O_4$		
						195.0656	$C_{10}H_{11}O_4$		
Decarboxymethyl lingstroside aglycone	$C_{17}H_{20}O_5$	303.1237	303.1237	6.42	6.76	124.0531	$C_7H_8O_2$	15	2a
						137.0608	$C_8H_{10}O_2$		
						147.0453	$C_9H_7O_2$		
						165.0556	$C_9H_9O_3$		
						183.0662	$C_9H_{11}O_4$		
10-Hydroxyoleuropein aglycone	$C_{19}H_{22}O_9$	393.1193	393.1191	4.82	5.48	137.0244	$C_7H_5O_3$	8	2b
						181.0502	$C_9H_9O_4$		
Oleuropein aglycone isomer	$C_{19}H_{22}O_8$	377.1241	377.1241	5.88	6.88	111.0087	$C_5H_3O_3$	16	3
						149.0241	$C_8H_5O_3$		
						263.0926	$C_{14}H_{15}O_5$		
						275.0923	$C_{15}H_{15}O_5$		
Oleuropein aglycone	$C_{19}H_{22}O_8$	377.1241	377.1247	7.29	6.88	111.0088	$C_5H_3O_3$	19	2a
						149.0244	$C_8H_5O_3$		
						195.0644	$C_{10}H_{11}O_4$		
						275.0919	$C_{15}H_{15}O_5$		
						307.0823	$C_{15}H_{15}O_7$		
Lingstroside aglycone	$C_{19}H_{22}O_7$	361.1291	361.1293	6.63	6.83	259.0975	$C_{15}H_{15}O_4$	19	2a
						291.0875	$C_{15}H_{15}O_6$		
Syringaresinol	$C_{22}H_{26}O_8$	417.1554	417.1556	6.18	8.10	127.0408	$C_6H_7O_3$	9	2b
						181.0503	$C_9H_9O_4$		
Oleoside	$C_{16}H_{22}O_{11}$	389.1089	389.1087	7.91	–[a]	113.0244	$C_5H_5O_3$	12	2a
						139.0032	$C_6H_3O_4$		
						149.0240	$C_8H_5O_3$		
						165.0552	$C_9H_9O_3$		
						183.0666	$C_9H_{11}O_4$		
Oleoside isomer	$C_{16}H_{22}O_{11}$	389.1089	389.1087	5.7	–[a]	165.0552	$C_9H_9O_3$	11	3
1-Hydroxypinoresinol	$C_{20}H_{22}O_7$	373.1292	373.1292	6.39	6.39	121.0294	$C_7H_5O_2$	8	2b
						151.0401	$C_8H_7O_3$		

Table 3 (continued)

Compound	Molecular formula	$[M-H]^-$ m/z calculated	$[M-H]^-$ m/z experimental	t_R (min)	t_R predicted (min)	Fragment m/z	Elemental formula	Peak score[c]	Level of identification[d]
						163.0402	$C_9H_7O_3$		
1-Hydroxipinoresinol isomer	$C_{20}H_{22}O_7$	373.1292	373.1294	6.42	–[a]	–	–	9	3
Oleanolic acid	$C_{30}H_{48}O_3$	455.3535	455.3540	13.94	12.69[b]	44.9980	CHO_2	8	2a
						120.0940	C_9H_{12}		
						152.1202	$C_{10}H_{16}O$		
						407.3316	$C_{29}H_{43}O$		
Maslinic acid	$C_{30}H_{48}O_4$	471.3484	471.3485	12.82	12.58	44.9982	CHO_2	12	2a
						405.3156	$C_{29}H_{41}O$		
						423.3423	$C_{29}H_{43}O_2$		
1-Acetoxypinoresinol	$C_{22}H_{24}O_8$	415.1398	415.1399	6.42	7.20	151.0402	$C_8H_7O_3$	10	2b
						280.0951	$C_{14}H_{16}O_6$		
						343.1188	$C_{19}H_{19}O_6$		
Methyl oleuropein aglycone	$C_{20}H_{24}O_8$	391.1412	391.1418	7.51	7.37	59.0140	$C_2H_3O_2$	6	2b
						67.0192	C_4H_3O		
						99.0456	$C_5H_7O_2$		
						111.0087	$C_5H_3O_3$		
						137.0608	$C_8H_9O_2$		
						291.0875	$C_{16}H_{15}O_6$		
10-Hydroxy-10-methyl oleuropein aglycone	$C_{20}H_{24}O_9$	407.1347	407.1347	6.71	6.75	99.0453	$C_5H_7O_2$	19	2b
						111.0087	$C_5H_3O_3$		
						121.0295	$C_7H_5O_2$		
						135.0453	$C_8H_7O_2$		
						137.0243	$C_7H_5O_3$		
						149.0245	$C_8H_5O_3$		
						163.0402	$C_9H_7O_3$		
						179.0351	$C_9H_7O_4$		
						195.0665	$C_{10}H_{11}O_4$		
						241.0871	$C_{15}H_{13}O_3$		
10-Hydroxydecarboxymethyl oleuropein aglycone	$C_{17}H_{20}O_7$	335.1150	335.1151	4.28	5.52	59.0139	$C_2H_3O_2$	8	2a
						85.0296	$C_4H_5O_2$		
						121.0292	$C_7H_5O_2$		
						151.0401	$C_8H_7O_3$		
						153.0557	$C_8H_9O_3$		
						155.0716	$C_8H_{11}O_3$		
						199.0613	$C_9H_{11}O_5$		

[a] The retention time prediction results are not reliable and other methods of verification such as the tandem mass spectrometry fragmentation pattern should be applied

[b] The retention time prediction result is not reliable because oleanolic acid is outside the applicability domain of the model

[c] Ratio of peak area to peak intensity

[d] Level 1 corresponds to confirmed structures where a reference standard is available; level 2a corresponds to evidence obtained by matching the spectrum with spectra from the literature or a library; level 2b corresponds to diagnostic evidence where no other structure fits the experimental tandem mass spectrometry information; level 3 corresponds to a tentative candidate [44]

The MS/MS spectrum of oleocanthal shows a similar fragment as with oleacein at m/z 165.0556, corresponding to $C_9H_9O_3$ [10, 56]. Moreover, the peak at m/z 183.0662 corresponds to $C_9H_{11}O_4$ and has been reported in previous work [10]. Oleuropein aglycone (identification level 2a) was eluted as two different peaks (with retention times of 5.86 and 7.21 min), suggesting the existence of an isomer (identification level 3). Three qualifier ions were the same for both compounds (m/z 111.0087 and 111.0088 [17], m/z 149.0241 [17] and m/z 149.0244 [10], and m/z 275.0919 and 275.0923, which have been reported by Kanakis et al. [56] and Dierkes et al. [10], corresponding to $C_5H_3O_3$, $C_8H_5O_3$, and $C_{15}H_{15}O_5$ respectively). The fragment m/z 195.0644, corresponding to $C_{10}H_{11}O_4$, has been reported by Kanakis et al. [56] and Dierkes et al. [10]. Lingstroside aglycone (identification level 2a) shows two qualifier ions m/z 259.0975 [10, 56] and m/z 291.0875 [10, 56], corresponding to $C_{15}H_{15}O_4$ and $C_{15}H_{15}O_6$ respectively. For 10-hydroxydecarboxymethyl oleuropein aglycone (identification level 2a), the fragment at m/z 199.0614, corresponding to $C_9H_{11}O_5$, has already been reported by Kanakis et al. [56].
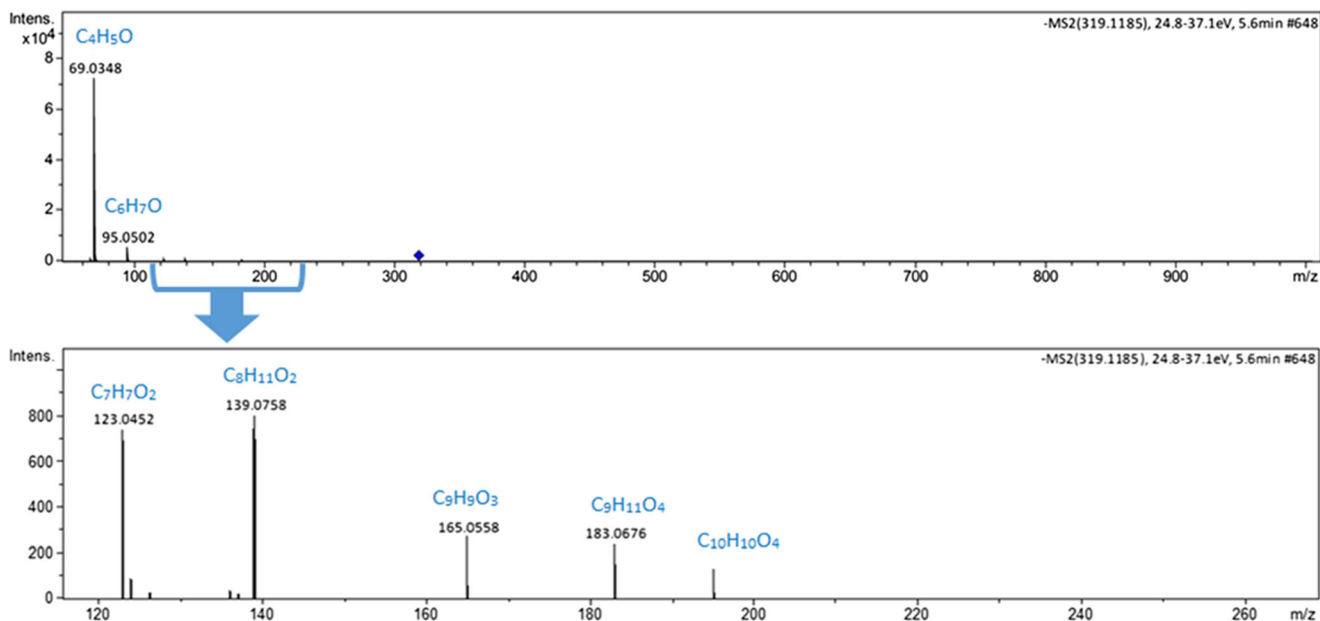
Fig. 3 Tandem mass spectrometry spectra of decarboxymethyl oleuropein aglycone (oleacein)

Elenolic acid, which is a nonphenolic compound and has been described as marker of maturation of olives [23], elenolic acid methyl ester, and the hydroxylated form of elenolic acid were tentatively identified at level 2a. For the hydroxylated form of elenolic acid, four fragments were recorded, and m/z 137.0603, corresponding to $C_8H_9O_2$, as well as m/z 181.0535, corresponding to $C_9H_9O_4$, have also been suggested by Capriotti et al. [17]. In addition, two isomers of elenolic acid and one isomer of the hydroxylated form of elenolic acid were identified at level 3.

From the hydroxytyrosol class, two hydroxytyrosol derivatives were detected, hydroxytyrosol acetate (identification level 2b) and an isomer of hydroxytyrosol acetate (identification level 3). Next, from the class of lignans, which have been suggested as varietal markers [20, 60] and have antiviral activities [60], 1-hydroxypinoresinol, 1-acetoxypinoresinol, and syringaresinol were identified at level 2b. One isomer of 1-hydroxypinoresinol was identified at level 3.

From the flavonoid class, the presence of luteolin was confirmed with a reference standard (identification level 1). Finally, the triterpenic acids oleanolic acid and maslinic acid were identified at level 2a. The secoiridoid oleoside was identified at level 2a and an isomer of oleoside was identified at level 3.

After the determination, the suspect compounds belonging to the classes of lignans, flavonoids, and secoiridoids were semiquantified on the basis of target compounds having similar structures, as suggested in previous work [5, 16, 19, 35]. 1-Acetoxypinoresinol, 1-hydroxypinoresinol, and syringaresinol were semiquantified with use of the pinoresinol calibration curve. Hydroxytyrosol acetate was semiquantified with use of the hydroxytyrosol calibration curve, and all the secoiridoids were semiquantified with use of the oleuropein calibration curve. Semiquantification results are presented in the electronic supplementary material (Table S5b; the concentrations are expressed in milligrams per kilogram as the mean values ± the standard deviation, $n = 3$).

A remarkable difference is observed in the total phenolic content between the extra virgin olive oil samples and the defective olive oil samples. The total phenolic content of the extra virgin olive oils ranged between 222 and 318 mg kg^{-1}, whereas the defective olive oils (RB1, BR1, PK1, KA1, NB1, Fusty, Musty, and Rancid) demonstrate impressively low values between 47 and 136 mg kg^{-1}. Therefore, it can be concluded that the taste of olive oil is directly related to the concentration of the phenolic compounds. The total phenolic content of each olive oil sample (sum of the concentrations of the quantified target compounds and the semiquantified concentrations of the suspect phenolic compounds) is presented in the electronic supplementary material (Fig. S5; the concentrations are expressed in milligrams per kilogram as the mean values ± the standard deviation, $n = 3$).

Nontarget screening

The nontarget screening workflow resulted in the generation of 304 features. The VIP method was then applied to distinguish the most important compounds responsible for discrimination. To prioritize the peaks that cause greater variation in the discrimination between samples, VIP scores were calculated for the PLS-DA model [53], and those with a VIP score greater than 0.83 were considered as the most important [51, 53, 54]. Of the 304 features, 151 were calculated to have a VIP score above 0.83 and were investigated in a subsequent

analysis. Fifteen of the 151 important compounds already existed in the target and suspect list [tyrosol, hydroxytyrosol, apigenin, oleuropein, ethyl vanillin, elenolic acid, decarboxymethyl oleuropein aglycone (oleacein), decarboxymethyl lingstroside aglycone (oleocanthal), hydroxytyrosol acetate, 10-hydroxyoleuropein aglycone, oleuropein aglycone, lingstroside aglycone, 10-hydroxy-10-methyl oleuropein aglycone, methyl oleuropein aglycone, and 10-hydroxydecarboxymethyl oleuropein aglycone] and were subsequently excluded from the nontarget list.

In an attempt to identify the remaining 136 masses, an inclusion list consisting of the m/z of their precursor ions was created, and the QTOF system was operated in Auto MS/MS mode to obtain the MS/MS spectra of the unknown analytes. From the 136 nontarget compounds, 7 compounds were successfully identified and confirmed with reference standards (identification level 1). These compounds were hexanoic acid, octanoic acid, palmitic acid, α-linolenic acid, α-linoleic acid, oleic acid, and arachidic acid. What is more interesting is that their EICs revealed great variations in the intensities of the free fatty acids among the samples. The most significant variations were observed in the EICs of hexanoic acid and octanoic acid (Fig. 4). Their presence in defective olive oil samples proves that they can be used as markers. The intensities of hexanoic acid and octanoic acid were high for Rancid and RB1 (which is a defective olive oil characterized as musty) and minimum for the extra virgin olive oils. In contrast, the peaks of the rest of the free fatty acids, palmitic acid, linolenic acid, linoleic acid, oleic acid, and arachidic acid, were most intense in the extra virgin olive oils.

Following the nontarget procedure, cinnamic acid and quinic acid were tentatively identified at level 2a. The predicted t_R was very close to the experimental value for both compounds (cinnamic acid, experimental t_R = 6.37 min and predicted t_R = 6.38 min; quinic acid, experimental t_R = 1.12 min and predicted t_R = 1.37 min), and the MS/MS fragments were matched and confirmed with use of MassBank [38] records. In addition, sinapic acid and acetosyringone were identified at level 2b. Finally, an unequivocal molecular formula was assigned (identified at level 4) for 24 compounds. In these cases, the MS/MS spectra were not informative enough for us to proceed with further identification. Identification attempts, including the identification levels, the experimental and predicted t_R, and the list of fragments and tentative structures are summarized in the electronic supplementary material (Table S8).

Retrospective analysis

On the successful nontarget identification of the free fatty acids reported in the previous section, retrospective analysis was performed to search for the possible presence of all the free fatty acids encountered in olive oil. Nineteen fatty acids were identified: hexanoic acid, octanoic acid, dodecanoic acid, myristic acid, pentadecanoic acid, palmitic acid, palmitoleic acid, heptadecanoic acid, heptadecanoic acid, stearic acid, oleic acid, α-linoleic acid, α-linolenic acid, arachidic acid, cis-eicosenoic acid, heneicosanoic acid, docosanoic acid, tricosanoic acid, and lignoceric acid. The distribution of the VIP scores
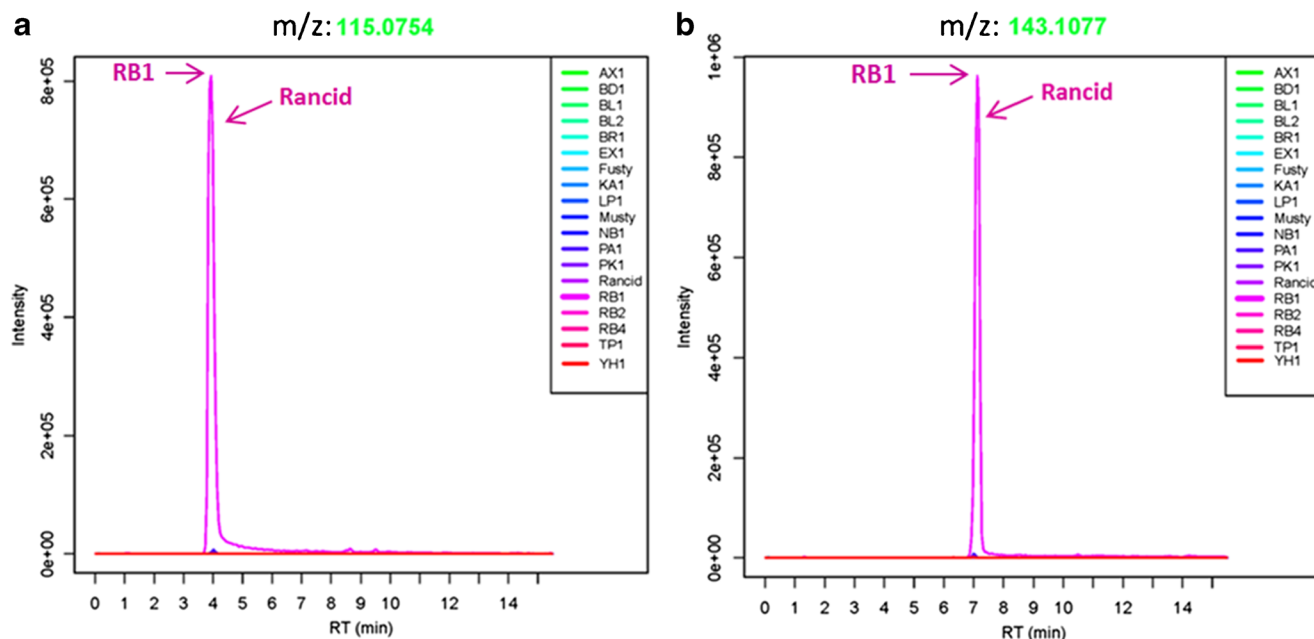
Fig. 4 Extracted ion chromatograms of **a** hexanoic acid and **b** octanoic acid in the analyzed samples. Both acids are markers for the defective olive oils. The intensities of hexanoic acid and octanoic acid were higher for RB1 and Rancid and were very low in the extra virgin olive oils. *RT* retention time

Fig. 5 Sample distribution based on the partial least squares–discriminant analysis model. The samples in *red* belonged to the test set. *EVOO* extra virgin olive oil
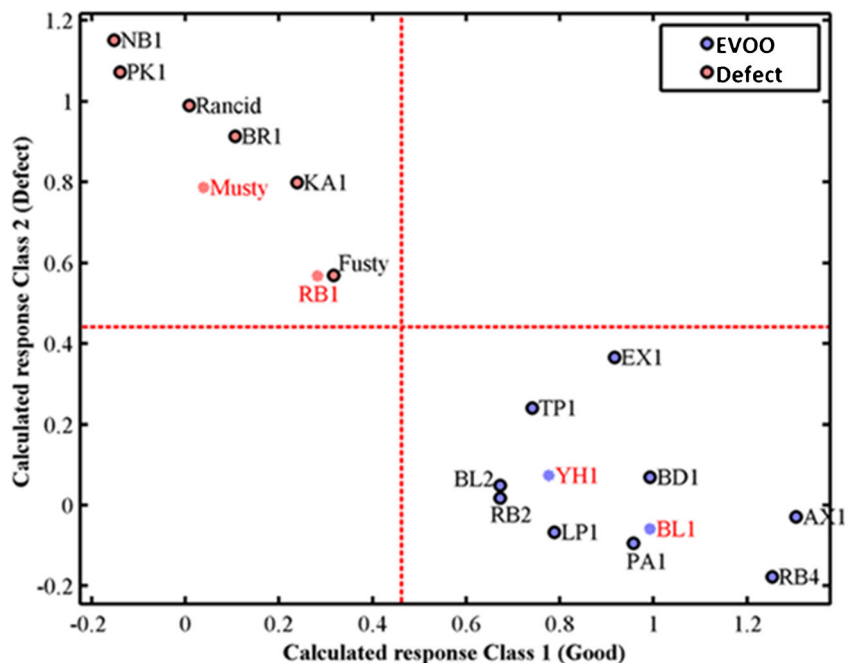
of the identified free fatty acids is presented in the electronic supplementary material (Table S9a). Their identification proves that they were extracted together with the phenolic constituents during the single extraction. Their recoveries were calculated to evaluate the performance of the proposed method for determination of free fatty acids and were in the range from 95 to 108 % ($n = 5$), as presented in the electronic supplementary material (Table S9b). The results were satisfactory, showing that

RP-UHPLC–ESI-QTOF-MS/MS could be used for the simultaneous determination of polyphenols and free fatty acids, both extracted from extra virgin olive oil samples in a single extraction.

Prediction models and classification

Nineteen samples were used to study the discrimination between samples. Fifteen samples were used to train the PLS-

Fig. 6 Mapping of samples with use of self-organizing maps of the counterpropagation artificial neural network model developed. The external test set samples are shown in *black*. Neurons in *blue* represent extra virgin olive oils and neurons in *red* represent defective olive oil samples
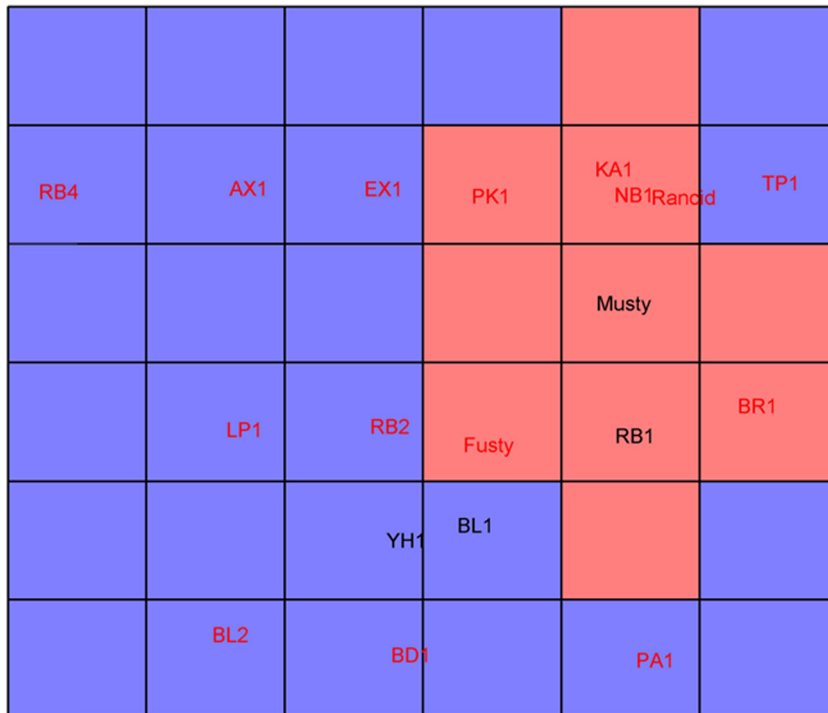
DA model [53, 54], consisting of nine and six samples belonging to good and defective samples of olive oil respectively. Four samples (Musty, RB1, YH1, and BL1) were used to evaluate the external accuracy of the PLS-DA model developed. The accuracy of the model was assessed internally and externally by the receiver operating characteristic curve. More information about the optimization of the PLS-DA model and the interpretation of the results can be found in the electronic supplementary material (Section S10). According to Fig. 5, all 19 samples were classified with high accuracy into two groups, extra virgin olive oils or defective olive oils. The model developed was evaluated with four samples, and their classes were calculated by the model. In conclusion, the model developed is robust and can be applied to unknown samples to understand their sensory profile with high accuracy.

In addition to PLS-DA, counterpropagation artificial neural networks (CP-ANNs) [51–53] were used to build a classification model. To develop this model without overfitting issues, the number of neurons (or size of the map) and the number of epochs were optimized with use of genetic algorithms. The procedures for the optimization of CP-ANNs can be found in the electronic supplementary material (Section S11). The final map, obtained with use of self-organizing maps, was proposed on the basis of a neuron size of 6×6, a frequency of 0.3, and 300 epochs, and is presented in Fig. 6, where the blue and red neurons represent the extra virgin olive oil samples and defective olive oil samples respectively. The external test set samples are shown in black, and their classes were predicted with high accuracy.

Conclusions

This study has made progress toward the organoleptic profiling of extra virgin olive oil, demonstrating the prospects of a novel RP-LC–ESI-QTOF-MS/MS analytical method. The use of target, suspect, and nontarget screening strategies in combination with supervised classification techniques, PLS-DA and CP-ANNs, constitutes a powerful tool that can be successfully applied in the investigation of the authenticity of extra virgin olive oil.

The use of target and suspect screening resulted in the determination of 14 target compounds and 26 suspect compounds. Using nontarget screening, we identified 11 compounds as responsible for the discrimination between extra virgin olive oils and defective olive oils after data processing with the R language and the XCMS package, following the VIP method for the suggestion of the most important features. Overall, 51 compounds are suggested as markers responsible for olive oil's organoleptic characteristics. There was a clear increase in the hexanoic acid and octanoic acid levels in defective olive oils. Similarly, a clear increase was observed in the concentrations of palmitic acid, linolenic acid, linoleic

acid, oleic acid, and arachidic acid in the extra virgin olive oils. Nevertheless, the detection of those free fatty acids demonstrates that the proposed method can be applied in the identification of phenolic compounds and free fatty acids simultaneously, since they are both extracted in a single liquid–liquid microextraction and are detected in the same analytical run.

Furthermore, two robust classification models using PLS-DA and CP-ANNs were built on the basis of all the features detected, and they can classify olive oils into two groups, defective olive oils and extra virgin olive oils, with high accuracy.

References

1. Romdhane K. Food authenticity and fraud. In: Picó Y, editor. Chemical analysis of food: techniques and applications. 1st ed. Waltham: Academic; 2012. p. 499–517.
2. Dennis J. Recent developments in food authentication. Analyst. 1998;123:151–6.
3. Aparicio R, Conte LS, Fiebig HJ. Olive oil authentication. In: Aparicio R, Harwood J, editors. Handbook of olive oil: analysis and properties. 2nd ed. New York: Springer; 2013. p. 589–653.
4. Bajoub A, Hurtado-Fernandez E, El Ajal A, Ouazzani N, Fernandez-Gutiérrez A, Carrasco-Pancorbo A. A comprehensive 3-years study of the phenolic profile of Moroccan monovarietal virgin olive oils from Meknès region. J Agric Food Chem. 2015;63:4376–85.
5. Sánchez de Medina V, Priego-Capote F, de Castro MDL. Characterization of monovarietal virgin olive oils by phenols profiling. Talanta. 2015;132:424–32.
6. Ballus CA, Quirantes-Piné R, Bakhouche A, de Oliveira da Silva LF, Coutinho EF, da Croce DM, et al. Profile of phenolic compounds of Brazilian virgin olive oils by rapid resolution liquid chromatography coupled to electrospray ionisation time-of-flight mass spectrometry (RRLC-ESI-TOF-MS). Food Chem. 2015;170:366–77.
7. International Olive Council. Sensory analysis of olive oil. Method for the organoleptic assessment of virgin olive oil. COI/T.20/Doc. No 15/Rev. 8. Madrid: International Olive Council; 2015.
8. International Olive Oil Council. Sensory analysis of olive oil standard guide for the selection, training and monitoring of skilled virgin olive oil tasters. COI/T.20/Doc. No 14 Rev. 4. Madrid: International Olive Oil Council; 2013.
9. Tena N, Wang SC, Aparicio-Ruiz R, Garcia-Gonzalez DL, Aparicio R. An in-depth assessment of analytical methods for olive oil purity, safety and quality characterization. J Agric Food Chem. 2015;63(18):4509–26.

10. Dierkes G, Krieger S, Duck R, Bongartz A, Schmitz OJ, Hayen H. High-performance liquid chromatography-mass spectrometry profiling of phenolic compounds for evaluation of olive oil bitterness and pungency. J Agric Food Chem. 2012;60:7597–606.

11. García-González DL, Aparicio R. Research in olive oil: challenges for the near future. J Agric Food Chem. 2010;58:12569–77.

12. Escuderos ME. Olive oil aroma evaluation by gas chromatographic method: a critical review. Crit Rev Anal Chem. 2011;41(1):70–80.

13. Montedoro G, Bertuccioli M, Anichini F. Aroma analysis of virgin olive oil by head space volatiles extraction techniques. In: Charalampous G, Inglet G, editors. Flavor of foods and beverages. 1st ed. New York: Academic; 1978. p. 247–81.

14. Cifuentes A. Food analysis: present, future, and foodomics. ISRN Anal Chem. 2012;2012:1–16.

15. Ibañez C, Simó C, García-Cañas V, Acunha T, Cifuentes A. The role of direct high-resolution mass spectrometry in foodomics. Anal Bioanal Chem. 2015;407:6275–87.

16. Bakhouche A, Lozano-Sánchez J, Beltrán-Debón R, Joven J, Segura-Carretero A, Fernández-Gutiérrez A. Phenolic characterization and geographical classification of commercial Arbequina extra-virgin olive oils produced in southern Catalonia. Food Res Int. 2013;50:401–8.

17. Capriotti AL, Cavaliere C, Crescenzi C, Foglia P, Nescatelli R, Samperi R, et al. Comparison of extraction methods for the identification and quantification of polyphenols in virgin olive oil by ultra-HPLC-QToF mass spectrometry. Food Chem. 2014;158: 392–400.

18. García-Villalba R, Carrasco-Pancorbo A, Oliveras-Ferraros C, Vazquez-Martín A, Menéndez JA, Segura-Carretero A, et al. Characterization and quantification of phenolic compounds of extra-virgin olive oils with anticancer properties by a rapid and resolutive LC-ESI-TOF MS method. J Pharm Biomed Anal. 2010;51:416–29.

19. Lozano-Sánchez J, Segura-Carretero A, Menendez JA, Oliveras-Ferraros C, Cerretani L, Fernández-Gutiérrez A. Prediction of extra virgin olive oil varieties through their phenolic profile. potential cytotoxic activity against human breast cancer cells. J Agric Food Chem. 2010;58:9942–55.

20. Ouni Y, Taamalli A, Gómez-Caravaca AM, Segura-Carretero A, Fernández-Gutiérrez A, Zarrouk M. Characterisation and quantification of phenolic compounds of extra-virgin olive oils according to their geographical origin by a rapid and resolutive LC-ESI-TOF MS method. Food Chem. 2011;127:1263–7.

21. Vichi S, Cortés-Francisco N, Caixach J. Insight into virgin olive oil secoiridoids characterization by high-resolution mass spectrometry and accurate mass measurements. J Chromatogr A. 2013;1301:48–59.

22. Taamalli A, Arráez Román D, Zarrouk M, Segura-Carretero A, Fernández-Gutiérrez A. Classification of 'Chemlali' accessions according to the geographical area using chemometric methods of phenolic profiles analysed by HPLC-ESI-TOF-MS. Food Chem. 2012;132:561–6.

23. Favati F, Condelli N, Galgano F, Caruso MC. Extra virgin olive oil bitterness evaluation by sensory and chemical analyses. Food Chem. 2013;139:949–54.

24. Bendini A, Cerretani L, Carrasco-Pancorbo A, Gómez-Caravaca AM, Segura-Carretero A, Fernández-Gutiérrez A, et al. Phenolic molecules in virgin olive oils: a survey of their sensory properties, health effects, antioxidant activity and analytical methods. an overview of the last decade. Molecules. 2007;12:1679–719.

25. Gutiérrez-Rosales F, Ríos JJ, Gómez-Rey ML. Main polyphenols in the bitter taste of virgin olive oil. structural confirmation by on-line high-performance liquid chromatography electrospray ionization mass spectrometry. J Agric Food Chem. 2003;51:6021–5.

26. Mateos R, Cert A, Carmen Pérez-Camino M, García JM. Evaluation of virgin olive oil bitterness by quantification of secoiridoid derivatives. J Am Oil Chem Soc. 2004;81:71–5.

27. Siliani S, Mattei A, Benevieri Innocenti L, Zanoni B. Bitter taste and phenolic compounds in extra virgin olive oil: an empirical relationship. J Food Qual. 2006;29:431–41.

28. Andrewes P, Busch JL, de Joode T, Groenewegen A, Alexandre H. Sensory properties of virgin olive oil polyphenols: identification of deacetoxy-ligstroside aglycon as a key contributor to pungency. J Agric Food Chem. 2003;51:1415–20.

29. García-Cañas V, Simó C, Herrero M, Ibáñez E, Cifuentes A. Present and future challenges in food analysis: foodomics. Anal Chem. 2012;84:10150–9.

30. Cuadros Rodríguez L, Ruiz-Samblás C, Valverde-Som L, Pérez-Castaño E, González-Casado A. Chromatographic fingerprinting: an innovative approach for food 'identitation' and food authentication - a tutorial. Anal Chim Acta. 2016;909:9–23.

31. Gómez-Caravaca AM, Maggio RM, Cerretani L. Chemometric applications to assess quality and critical parameters of virgin and extra-virgin olive oil. A review. Anal Chim Acta. 2016;913:1–21.

32. European Commission. Commission implementing Regulation (EU) No 1348/2013 of 16 December 2013 amending Regulation (EEC) No 2568/91 on the characteristics of olive oil and olive-residue oil and on the relevant methods of analysis. Off J Eur Union L. 338:31. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2013.338.01.0031.01.ENG&toc=OJ:L:2013:338:FULL. Accessed 18 Jan 2016.

33. Piscopo A, Poiana M. Packaging and storage of olive oil. In: Muzzalupo I, editor. Olive germplasm - the olive cultivation, table olive and olive oil industry in Italy. Rijeka: InTech; 2012. doi:10.5772/51827.

34. Stefanoudaki E, Williams M, Harwood J. Changes in virgin olive oil characteristics during different storage conditions. Eur J Lipid Sci Technol. 2010;112:906–14.

35. Becerra-Herrera M, Sánchez-Astudillo M, Beltrán R, Sayago A. Determination of phenolic compounds in olive oil: new method based on liquid–liquid micro extraction and ultra high performance liquid chromatography-triple–quadrupole mass spectrometry. LWT Food Sci Technol. 2014;57:49–57.

36. Want EJ, Wilson ID, Gika H, Theodoridis G, Plumb RS, Shockcor J, et al. Global metabolic profiling procedures for urine using UPLC–MS. Nat Protoc. 2010;5:1005–18.

37. Krauss M, Singer H, Hollender J. LC–high resolution MS in environmental analysis: from target screening to the identification of unknowns. Anal Bioanal Chem. 2010;397:943–51.

38. Gago-Ferrero P, Schymanski EL, Bletsou AA, Aalizadeh R, Hollender J, Thomaidis NS. Extended suspect and non-target strategies to characterize emerging polar organic contaminants in raw wastewater with LC-HRMS/MS. Environ Sci Technol. 2015;49(20):12333–41.

39. Wolf S, Schmidt S, Muller-Hannemann M, Neumann S. In silico fragmentation for computer assisted identification of metabolite mass spectra. BMC Bioinf. 2010;11:148.

40. Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, et al. MassBank: a public repository for sharing mass spectral data for life sciences. J Mass Spectrom. 2010;45:703–14.

41. Aalizadeh R, Thomaidis NS, Bletsou A, Gago-Ferrero P. Quantitative structure-retention relationships models to support non-target high resolution mass spectrometric screening of emerging contaminants in environmental samples. J Chem Inf Model. 2016;56(7):1384–98.

42. Schymanski EL, Singer HP, Longree P, Loos M, Ruff M, Stravs MA, et al. Strategies to characterize polar organic contamination in wastewater: exploring the capability of high resolution mass spectrometry. Environ Sci Technol. 2014;48:1811–8.

43. Hernández F, Sancho JV, Ibáñez M, Abad E, Portolés T, Mattioli L. Current use of high-resolution mass spectrometry in the environmental sciences. Anal Bioanal Chem. 2012;403:1251–64.

44. Schymanski EL, Jeon J, Gulde R, Fenner K, Ruff M, Singer HP, et al. Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ Sci Technol. 2014;48:2097–8.

45. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, et al. A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol. 2012;30:918–20.

46. Tautenhahn R, Böttcher C, Neumann S. Highly sensitive feature detection for high resolution LC/MS. BMC Bioinf. 2008;9:504.

47. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem. 2006;78(3):779–87.

48. Kuhl C, Tautenhahn R, Böttcher C, Larson TR, Neumann S. CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. Anal Chem. 2012;84:283–9.

49. Libiseller G, Dvorzak M, Kleb U, Gander E, Eisenberg T, Madeo F, et al. IPO: a tool for automated optimization of XCMS parameters. BMC Bioinf. 2015;16:118.

50. Ballabio D, Consonni V. Classification tools in chemistry. Part 1: linear models. PLS-DA. Anal Methods. 2013;5:3790–8.

51. Ballabio D, Consonni V, Todeschini R. The Kohonen and CP-ANN toolbox: a collection of MATLAB modules for self organizing maps and counterpropagation artificial neural networks. Chemom Intell Lab Syst. 2009;98(2):115–22.

52. Mehmood T, Liland KH, Snipen L, Sæbø S. A review of variable selection methods in partial least squares regression. Chemom Intell Lab Syst. 2012;118:62–9.

53. Chong I-G, Jun C-H. Performance of some variable selection methods when multicollinearity is present. Chemom Intell Lab Syst. 2005;78:103–12.

54. Gosselin R, Rodrigue D, Duchesne C. A bootstrap-VIP approach for selecting wavelength intervals in spectral imaging applications. Chemom Intell Lab Syst. 2010;100:12–21.

55. Michel T, Khlif I, Kanakis P, Termentzi A, Allouche N, Halabalaki M, et al. UHPLC-DAD-FLD and UHPLC-HRMS/MS based metabolic profiling and characterization of different *Olea europaea* organs of Koroneiki and Chetoui varieties. Phytochem Lett. 2015;11:424–39.

56. Kanakis P, Termentzi A, Michel T, Gikas E, Halabalaki M, Skaltsounis A-L. From olive drupes to olive oil. An HPLC-Orbitrap-based Qualitative and quantitative exploration of olive key metabolites. Planta Med. 2013;79:1576–87.

57. Olivieri AC. Practical guidelines for reporting results in single- and multi-component analytical calibration: a tutorial. Anal Chim Acta. 2015;868:10–22.

58. Matuszewski BK, Constanzer ML, Chavez-Eng CM. Strategies for the assessment of matrix effect in quantitative bioanalytical methods based on HPLC-MS/MS. Anal Chem. 2003;75:3019–30.

59. International Olive Council. Determination of biophenols in olive oils by HPLC. COI/T.20/Doc No 29. Madrid: International Olive Council; 2009.

60. Cicerale S, Lucas L, Keast R. Biological activities of phenolic compounds present in virgin olive oil. Int J Mol Sci. 2010;11(2):458–79.

61. Antonini E, Farina A, Leone A, Mazzara E, Urbani S, Selvaggini R, et al. Phenolic compounds and quality parameters of family farming versus protected designation of origin (PDO) extra-virgin olive oils. J Food Compos Anal. 2015;43:75–81.