View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Repositori Institucional de la Universitat Jaume I

1 **Quality classification of Spanish olive oils by untargeted gas chromatography coupled**
2 **to hybrid quadrupole–time of flight mass spectrometry with atmospheric pressure**
3 **chemical ionization and metabolomics-based statistical approach**

4 C. Sales, M.I. Cervera, R. Gil, T. Portolés, E. Pitarch, J. Beltran*

5 Research Institute for Pesticides and Water (IUPA), University Jaume I, Avda. Sos Baynat,
6 E-12071 Castellón, Spain.

7 *Corresponding author: beltranj@uji.es

8 **ABSTRACT**

9 The novel atmospheric pressure chemical ionization (APCI) source has been used in
10 combination with gas chromatography (GC) coupled to hybrid quadrupole time-of-flight
11 (QTOF) mass spectrometry (MS) for determination of volatile components of olive oil,
12 enhancing its potential for classification of olive oil samples according to their quality
13 using a metabolomics-based approach. The full-spectrum acquisition has allowed the
14 detection of volatile organic compounds (VOCs) in olive oil samples, including extra
15 virgin, virgin and lampante qualities. A dynamic headspace extraction with cartridge
16 solvent elution was applied. The metabolomics strategy consisted of three different steps: a
17 full mass spectral alignment of GC-MS data using MzMine 2.0, a multivariate analysis
18 using Ez-Info and the creation of the statistical model with combinations of responses for
19 molecular fragments. The model was finally validated using blind samples, obtaining an
20 accuracy in oil classification of 70 % taking the official established method, "PANEL
21 TEST" as reference.

22

23

24

25

26 .

27 # 1. INTRODUCTION

28 Food quality is a matter of concern for consumers and producers. This quality is, in most
29 cases, connected to the origin of the products, as it is the case for French wine
30 (Arvanitoyannis, Katsota, Psarra, Soufleros, & Kallithraka, 1999) or Spanish ham
31 (Sánchez-Peña, Luna, García-González, & Aparicio, 2005), and contributes to rise their
32 prizes. For this reason and to avoid fraud, guarantee of the genuine quality is a critical step
33 from the economical point of view. Thus, the characterization of this kind of products is an
34 important aim where analytical chemists can be of great support.

35 In vegetable products, qualitative and quantitative analysis of volatile organic compounds
36 (VOCs) has been an important issue of scientific interest for their organoleptic
37 characterization. Thus, the determination of VOCs and characterization of volatile profiles
38 in agricultural or feeding products, among others, are interesting work fields involved in
39 assuring food quality (Dutta, Kashwan, Bhuyan, Hines, & Gardner, 2003; Ribes, Carrera,
40 Gallego, Roca, Berenguer & Guardino, 2007). One of the most interesting studies is related
41 to the determination of olive oil quality, as its classification is typically performed by
42 testers who establish if an olive oil must be labelled as extra virgin, virgin or lampante (not
43 recommended for consumption) (Council, 2016). This strategy is known as "PANEL
44 TEST", which classifies the oils according to two main properties: defects and goods. The
45 classification requirements are established by the International Council of Olive Oil
46 (Council, 2016), regarding their characteristics. These defects can be fusty, mouldy, sour
47 and woody, and the goods (positive factors) can be fruity (specifying green attribute), bitter
48 and spicy. According to the literature (Kalua, Allen, Bedgood Jr, Bishop, Prenzler &
49 Robards, 2007; Luna, Morales, & Aparicio, 2006), the organic compounds responsible of

50 these flavours are typically esters, ketones, aldehydes, alcohols, terpenes, phenols and their
51 derivatives, in concentrations ranging from few ng L^{-1} to hundreds of mg L^{-1} and with
52 different odour thresholds. PANEL TEST methodology could be considered slightly
53 subjective, as the opinion of testers may vary in such a significant way that the same
54 sample can sometimes be designated as belonging to two different qualities at a time. These
55 discrepancies can lead to a quality misclassification of an oil, causing considerable
56 economical loses, commercial problems and fraud. A more objective alternative could be
57 based in the use of chromatographic techniques coupled to mass spectrometry (MS), which
58 could allow the determination of chemical composition of the volatile fraction of olive oil
59 samples, even at really low concentration levels.

60 In this sense, the most adequate analysis technique is gas chromatography coupled to mass
61 spectrometry (GC-MS) (Angerosa, Servili, Selvaggini, Taticchi, Esposto, Montedoro,
62 2004; Flath, Forrey, & Guadagni, 1973).

63 According to the volatile characteristics of the compounds of interest, a specific extraction
64 technique is also an important issue in order to perform a suitable separation from the
65 matrix. From several studies on determination of VOCs in different matrices (Barco-
66 Bonilla, Plaza-bolaños, Fernández-Moreno, Romero-González, Garrido, Martínez, 2011;
67 Jiménez, Aguilera, Beltrán, & Uceda, 2006; Lam & Proctor, 2003; Salemi, Lacorte,
68 Bagheri, & Barceló, 2006; Serrano, Beltrán, & Hernández, 2009), there is an evidence of
69 the extended use of trapping processes of the compounds in some kind of sorbent; either by
70 forcing them to pass through the sorbent bed (P&T) (Barco-Bonilla et al., 2011; Salemi et
71 al., 2006) or letting them to establish an equilibrium between the vapour phase and the
72 adsorbent in a closed place (SPME) (Pouliarekou, Badeka, Tasioula-margari, Kontakos,

73 Longobardi, & Kontominas, 2011; Serrano et al., 2009). Other techniques also used for
74 VOCs extraction include: direct headspace (HS) injection (Hu et al., 2014), stir bar sorptive
75 extraction (SBSE) (Bicchi, Iori, Rubiolo, & Sandra, 2002), or liquid phase micro extraction
76 (LPME) (Lee, Lee, Rasmussen, & Pedersen-Bjergaard, 2008).

77 The determination of the chemical fingerprint of food samples is an interesting well-known
78 approach for characterization of food products, for example in oil (Reboredo-Rodríguez, et
79 al 2016, Reboredo-Rodríguez, González-Barreiro, Cancho-Grande, & Simal-Gándara,
80 2012) or tomato samples (Vaz-Freire, Da Silva, & Freitas, 2009). Chemical fingerprint,
81 normally obtained using targeted analyses, can be further used to classify unknown
82 samples, but due to the fact that a limited number of compounds must be selected a priori,
83 some information regarding the samples is lost. In this context, metabolomics, which is
84 defined as "the unbiased, global screening approach to classify samples based on metabolite
85 patterns or fingerprints that change in response to disease, environmental or genetic
86 perturbations with the ultimate goal to identify discriminating metabolites" (Cevallos-
87 Cevallos, Reyes-De-Corcuera, Etxeberria, Danyluk, & Rodrick, 2009), can be an
88 interesting approach to solve complex classification problems. One of the main drawbacks
89 of metabolomics occurs when the relations or discrepancies between samples are just
90 determined by a compound or group of compounds which are present at very low levels in
91 complex matrices and can be easily dismissed. High resolution (HR) MS systems, such as
92 time of flight (TOF)(Kind, Tolstikov, Fiehn, & Weiss, 2007) or magnetic sector (Kieken et
93 al., 2009), enhance the detection of molecules in complex matrices at very low levels
94 (Salihovic, Nilsson, Hagberg, & Lindström, 2013) due to their accurate mass measurements
95 and extraordinary sensitivity, providing good results for the determination of non-target

compounds. Furthermore, the use of atmospheric pressure chemical ionization (APCI), a very promising ionization source in GC, which is softer than the common electron ionization (EI), allows to obtain high intensity peaks for the molecular ion ($M^{+\cdot}$) and/or the protonated molecule ($[M+H]^+$), depending on the nature of the compounds (Portolés, Sancho, Hernández, Newton, & Hancock, 2010). This information is really convenient as in some cases the molecular peak is absent in EI due to its high fragmentation and in terms of sensitivity, the peak intensity can be reduced. GC coupled to hybrid quadrupole time of flight (QTOF) MS equipped with APCI allows the acquisition of the accurate-mass full scan spectra at low and at high collision energy (MS^E mode) being a very useful tool for elucidation purposes.

Data processing, together with data acquisition, are the mainstays of metabolomics. The direct observation of sample chromatograms does not give significant information about the difference between samples quality, and thus specialized software is required to obtain chromatographic peaks and masses from raw data. In literature, metabolomics studies use different software to get the information needed from the chromatograms, such as XCMS package of R (Díaz, Pozo, Sancho, & Hernández, 2014), MetAlign (Tikunov et al., 2005) or MzMine 2.0 (Kind et al., 2007).

The aim of this work has been the development of a GC-(APCI)QTOF MS methodology to obtain chemical profile/fingerprint of olive oil volatile compounds in order to establish differences between virgin olive oil qualities using P&T extraction and through the use of metabolomics techniques in order to give a more objective decision in their classification compared with that provided by the "PANEL TEST".

Con formato: Español (alfab. internacional)

Con formato: Español (alfab. internacional)

Código de campo cambiado

Con formato: Español (alfab. internacional)

Código de campo cambiado

Con formato: Español (alfab. internacional)

Código de campo cambiado

119 # 2. MATERIALS AND METHODS

120 ## 2.1. Chemicals and reagents

121 Internal standard triphenyl phosphate (TPP) ≥99% was purchased from Sigma Aldrich
122 (Germany). Diethyl ether (residue analysis quality GC) and hexane (trace analysis quality
123 (AT) GC) were provided by Scharlau (Barcelona, Spain).

124 Supelclean ENVI-Carb® SPE tubes 500 mg, volume 6 mL, 120-400 mesh, surface area 100
125 m^2 g^{-1}, used as traps were purchased from Supelco (Barcelona, Spain).

126

127 ## 2.2. Olive oil samples

128 A total of 425 olive oil samples were provided by the "Interprofesional del Aceite de Oliva
129 Español" Organization (INTERPRO, Spain), the "Agencia para el Aceite de oliva del
130 Ministerio de Agricultura, Alimentación y Medio Ambiente" and the oficial control
131 services from the "Consejería de Agricultura, Pesca y Desarrollo Rural de la Junta de
132 Andalucía".

133 Oil samples were taken from different regions of Spain and included 300 quality
134 characterized samples (120 extra virgin, 120 virgin and 60 lampante) and 125 blind samples
135 (the quality was unknown during analysis). Samples were stored in freezer at −22 °C until
136 their use. Samples were characterized by means of pH measurements and physicochemical
137 and organoleptic properties by the official participating laboratories (Laboratorio Arbitral
138 Agroalimentario del MAGRAMA, Laboratorios Agroalimentarios de Córdoba y Atarfe de
139 la Junta de Andalucía) and their corresponding certified "PANEL TESTs".

140

141 **2.3. Sample treatment**

142 Olive oil samples were allowed to defrost at room temperature before analysis. Then, they
143 were aliquoted in 4 different 10 mL vials. One aliquot was used to perform the extraction
144 and the remaining ones were stored at 4 ºC.

145 5 g of olive oil were weighted directly into a 150 mL flask before inserting a magnetic
146 stirrer. The flask was rapidly closed with a glass tap with a nitrogen entrance and the exit
147 connected to the sorbent trap (Envi-Carb cartridge). The cartridge was conditioned with 2 x
148 5 mL of a mixture of hexane:diethyl ether (50/50; v/v) and vacuum dried for 10 min.
149 Sample extraction was then carried out for 60 min at 40 °C with a nitrogen flow of 1 L
150 min^{-1} and with stirring at 300 rpm. After extraction, Envi-carb® cartridges were eluted by
151 gravity with 5 mL of the hexane:diethyl ether mixture (50/50; v/v), into a glass tube
152 previously weighted. 50 µL of TPP solution at 5 mg L^{-1} (in hexane) were added as internal
153 standard. The entire extract was then concentrated until total removal of diethyl ether, *i.e.*
154 until approximately a final volume of 0.5 mL, under vacuum conditions with a MiVac Duo
155 Concentrator (Genevac, Italy). Finally, in order to adjust the final volume of the extract (0.5
156 mL of hexane), several drops of hexane were added until adjusted mass (0.3274 g). An
157 aliquot of 100 µL of this extract was transferred to a 20 mL vial in order to generate a pull
158 of extracts to prepare the quality control (QC) sample. The remaining extract was divided
159 into two different vials with 200 µL inserts, sealed and stored in freezer at -20 ºC until their
160 analysis by GC-QTOF MS. In each extraction batch, 2 extra virgin, 2 virgin and 2 lampante
161 oil samples were processed simultaneously.

162

163

164 **2.4. GC-(APCI)QTOF MS**

165 The chromatographic analysis were performed using an Agilent 7890A gas chromatograph,
166 equipped with an Agilent 7693 autosampler, coupled to a quadrupole-TOF mass
167 spectrometer, Xevo G2 QTOF MS (Waters Corporation, Manchester, UK), with APCI
168 source. The GC separation was performed using a fused silica HP-5MS capillary column
169 with a length of 30 m x 0.25 mm ID and a film thickness of 0.25 µm (J&W Scientific,
170 Folson, CA, USA). The oven program was set as follows: 40 ºC (3 min); 5 ºC/min to 160
171 ºC (1 min); 50 ºC/min to 300 ºC (2.2 min) (total runtime 33 min). Injections of 1 µL of
172 sample extracts were performed using pulsed splitless mode (50 psi) at a temperature of
173 270 ºC with a pulse time of 0.90 min. Helium (99.999 %; Praxair, Valencia, Spain) was
174 used as the carrier gas at a constant flow rate of 3 mL min^{-1}.

175 The interface temperature was set to 310 ºC using N_2 as auxiliary gas at a flow rate of 150
176 L h^{-1}, the make-up gas flow rate was set at 300 mL min^{-1} and the cone gas flow rate at 16 L
177 h^{-1}. The APCI corona pin was operated at 1.7 µA with a cone voltage of 20 V. The
178 ionization process occurred within an enclosed ion volume, which enabled control over the
179 protonation/charge transfer processes. Water placed in an uncapped vial, which was located
180 within a special designed holder placed in the source door, was used as modifier. Xevo
181 QTOF MS was operated with a scan time of 0.4 s, acquiring the mass range *m/z* 50–650.
182 TOF MS resolution was approximately 18,000 (FWHM) at *m/z* 614. For MS^E
183 measurements, two alternating acquisition functions were used applying different collision
184 energies; 4 eV for low energy function (LE), and a collision energy ramp (10-40 eV) for
185 high energy function (HE). Heptacose (PFTBA) was used for the daily mass calibration.
186 Mass accuracy calibration was continuously performed using a background ion coming

187 from the GC-column bleed as lock mass (protonated molecule of
188 octamethylcyclotetrasiloxane, *m/z* 297.0830).

189 In order to avoid bias in the methodology, samples were analyzed in batches of 60 (20
190 pertaining to each class, extra virgin, virgin and lampante) randomly distributed. In
191 addition, in each analysis batch, 10 QCs were injected along the sequence (one QC every
192 10 samples and 2 QCs at the beginning and at the end of the sequence).

193

194 **2.5. Data processing**

195 GC-QTOF MS data were converted from ".raw" to ".cdf" format using Databridge, a
196 software provided by MassLynx v4.1 (Waters Corporation). Data mining was carried out
197 using MzMine 2.0 software with the conditions indicated in **Figure 1**. The first part of the
198 procedure (including cropping, mass detection, building chromatograms and baseline cut-
199 off) led to a data matrix of areas corresponding to the markers detected in the form of
200 combination of m/z value and retention time (*m(m/z in Da))t(time in sec)*). Data mining was
201 performed by MzMine 2.0 software following the procedure described in **Figure 1**.

202 Data were saved in a ".csv" file, exporting the row identification, the retention time, the
203 *m/z*, the number of detected peaks and the peak area.This matrix was then filtered to
204 remove too low intensity peaks (could be considered noise) and unpaired markers (which
205 were only present in a low proportion -less than 20%- of samples). The last step consisted
206 in a "forced integration" of the considered markers in all samples, in order to eliminate the
207 presence of zeroes in the final data matrix. The forced-integrated area was established as
208 the 1 % of the background absolute signal at the considered RT, as this was the

209 normalization step which led to best results. Finally a last filter was applied to eliminate
210 duplicated signals that corresponds to the same marker.

211 Data matrix obtained after applying MzMine 2.0 was exported to a text file (.csv). Using
212 Microsoft Excel absolute areas were converted to relative areas using the area of one of the
213 TPP m/z ions (m/z 327), which was quantified in each sample using TargetLynx®
214 software. The relative areas were corrected with the nearest QC and then pareto scaled to
215 minimize the impact of the peaks with higher standard deviation. In order to carry out the
216 statistical analysis the EzInfo software (U-Metrics) was applied as it is easy to use and,
217 from our previous experience, it allows to obtain well founded results (Díaz et al., 2014).

218

219

220

221 ## 3. RESULTS AND DISCUSSION

222 ### 3.1. GC-(APCI)QTOF MS

223 Chromatographic analysis of sample extracts was carried out by GC-(APCI)QTOF MS
224 working in MS^E mode in order to get both maximum sensitivity and full information. From
225 previous experiences in our laboratory (Cervera, Portolés, López, Beltrán, & Hernández,
226 2014) it was expected that the soft ionization provided by the APCI source combined with
227 low energy acquisition function, would allow to see the molecular ion or the protonated
228 molecule ($M^{+\cdot}$ or $[M+H]^+$), as only low fragmentation occurs. Additionally, the information
229 acquired at high energy function (collision energy ramp of 10-40 eV) allowed obtaining the
230 fragmentation pattern, which resulted very useful for elucidation purposes.

231 As an example, **Figure S12** shows the accurate-mass full spectrum obtained for a
232 sesquiterpene, the α-farnesene, acquired by GC-APCI-QTOF MS (at low and high energy)
233 comparing them with its corresponding EI (70 eV) spectrum from a commercial library
234 (NIST). The information about the protonated molecule in the low energy function together
235 with the fragmentation obtained in the high energy one was of great help in the elucidation
236 process, which followed the steps detailed in a previous work (Portolés, Pitarch, López,
237 Hernández, & Niessen, 2011). All instrumental parameters were selected according to
238 previous screening research works (Cervera et al., 2014; Portolés, Mol, Sancho, López, &
239 Hernández, 2014), which have been found to be reproducible and robust, basic
240 requirements of metabolomics. An example of the instrument performance is shown in **Fig**
241 **S2**, a total ion chromatogram obtained for the injection of a QC using the detailed
242 parameters.

Con formato: Fuente: Negrita

243

244 **3.2. Extraction procedure optimization**

245 Taking into account the properties of the potential analytes, especially their volatility and
246 considering the advantages of obtaining a final liquid extract (that can be injected several
247 times, even in different chromatographic instruments) purge and trap (P&T) with solvent
248 trap elution was selected as extraction technique. This approach requires low economical
249 investment and provides high reproducibility with the advantage of keeping a second
250 extract vial for further analysis. In previous works a similar P&T method with subsequent
251 GC-MS analysis was developed in our research group for the determination of volatiles in
252 tomato and melon samples(Beltran, Serrano, López, Peruga, Valcárcel, & Roselló, 2006;
253 Fredes, Sales, Barreda, Valcárcel, Roselló, Beltrán, 2016), and it has been used as a starting

254 point for method development, although some changes were considered to improve its
255 performance in application to olive oil samples.

256 Efficiency of two different sorbent materials was tested in the P&T step: Tenax TA® SPE
257 and non-porous carbon Envi-Carb cartridges. For this purpose, aliquots of an extra virgin
258 olive oil sample were extracted using Tenax (n=3) and Envi-Carb (n=3) with the same
259 extraction conditions. The comparison of peak areas of some selected analytes in these
260 chromatograms showed that both cartridges provided almost the same performance, as seen
261 previously in melon samples (Fredes et al, 2016). Although Tenax can be a bit more
262 efficient for some compounds (20% more area) both sorbents allow to extract quantitively
263 the majority of the volatiles from the sample. - Moreover, compared to Tenax, Envi-Carb
264 cartridges are commercially available (while Tenax cartridges are produced on demand,
265 with longer supply times) and have an analysis cost of almost 3 times lower. Finally, Envi-
266 Carb was selected to proceed with further development.

267

268 **3.3. Metabolomics optimization**

269 After optimization of extraction and chromatographic procedures, 425 oil samples
270 (including 300 samples for model generation and 125 samples for model validation) were
271 analyzed in several batches, both for extraction (6 samples each extraction batch) and
272 chromatographic analysis (including samples, 10 QC and 10 blank samples in each
273 sequence). Firstly considering a completely unbiased analysis, a Principal Component
274 Analysis (PCA) was applied over the training data matrix corresponding to 300 olive oil
275 samples. **Figure 2a** illustrates the PCA obtained.

Con formato: Fuente: Negrita

Con formato: Fuente: Negrita

Data mining was carried out using MzMine 2.0 software with the conditions indicated in **Figure 1**. The first part of the procedure (including cropping, mass detection, building chromatograms and baseline cut-off) led to a data matrix of areas corresponding to the markers detected in the form of combination of m/z value and retention time (*m(m/z in Da))t(time in sec)*). This matrix was then filtered to remove too low intensity peaks (could be considered noise) and unpaired markers (which were only present in a low proportion - less than 20% - of samples). The last step consisted in a "*forced integration*" of the considered markers in all samples, in order to eliminate the presence of zeroes in the final data matrix. The forced integrated area was established as the 1 % of the background absolute signal at the considered RT, as this was the normalization step which led to best results. Finally a last filter was applied to eliminate duplicated signals that corresponds to the same marker.

Data matrix obtained after applying MzMine 2.0 was exported to a *text* file (.csv). Using Microsoft Excel absolute areas were converted to relative areas using the area of one of the TPP *m/z* ions (m/z 327), which was quantified in each sample using TargetLynx® software. The relative areas were corrected with the nearest QC and then pareto scaled to minimize the impact of the peaks with higher standard deviation. In order to carry out the statistical analysis the EzInfo software (U-Metrics) was applied as it is easy to use and, from our previous experience, it allows to obtain well founded results (Díaz et al., 2014).

Firstly considering a completely unbiased analysis, a Principal Component Analysis (PCA) was applied over the training data matrix corresponding to 300 olive oil samples**. Figure 3a** illustrates an example of a PCA corresponding to analysis subset of 60 olive oil samples.

This PCA analysis seemed to be at least promising, as at first sight a rough distinction between “extra” and “lampante” can be seen, although “virgin” samples appeared in the middle and overlapped the two other groups. Additionally, this step allows to check the correct normalization of the samples by analyzing the behavior of the QC samples, ideally clustered in the center of the Score Plot. This type of analysis is employed for dimension reduction; as it is not guided, the QC samples joining in the center of the plot implies the correct normalization of the batch.

Con formato: Fuente: (Predeterminado) Times New Roman, 12 pto

Con formato: Fuente: Times New Roman, 12 pto, Inglés (Estados Unidos)

The next step consisted in performing a Partial Least Squares-Discriminant Analysis (PLS-DA), which considers additional information about groups during classification process. It is a biased PCA which provides a better separation between groups based on some parameters. Thus, a PLS-DA was applied to construct a statistical model to find differences in volatile compounds composition between samples (see **Figure 3b2b**). As it can be seen, the PLS-DA analysis separated the three groups of samples slightly better, though some group overlapping was still present.

Finally, in order to get information relative to the relevant compounds allowing sample classification, an Orthogonal Partial Least Squares-Discriminant Analysis (OPLS-DA) was applied to highlight the differences between groups facing two different groups at a time, regarding different compound concentration in samples. For this purpose, S-Plot graphics for all two-classes comparisons, Virgin vs Extra, Extra vs Lampante and Virgin vs Lampante, were obtained (**Figure 43**) and studied looking for endpoints. Theoretically, in an S-Plot the points with a p-corr value between 0.9 and 1 are closely related to the compounds characterizing the samples of one group, which constitute important class markers. Thus the most significant ions, *i.e.* the further from the center in the S-Plot and

with a p-corr close to 1, were selected as tentative markers. According to this, a total of 15 markers (corresponding to combination of an m/z ion and retention time) were selected and the corresponding chromatographic peaks, processed with TargetLynx® for a better precision of peak areas. **Table 1** shows the data for the 15 markers, together with the assigned elemental compositions for the protonated molecule and for the main fragment ions. In most cases they are labelled as *markers* as the high number of possible compounds with the same fragments made unfeasible to unequivocally identify them. We have been able to completely identify some of the markers even with the confirmation through injection of reference standards,

An example of the elucidation process of a marker can be found in **Figure ~~S1~~4**, with the experimental APCI accurate-mass spectra obtained by GC-QTOF MS for the high and low collision energy functions and with the possible chemical structures proposed for the most abundant fragment ions together with the experimental mass errors (in mDa). The structure was assigned to ethyl-2-methylbutanoate and was then confirmed by the injection of its standard. This compound has been already identified by some authors in olive oil samples (Dierkes, Bongartz, Guth, & Hayen, 2012; Kiralan, Ozkan, Koyluoglu, Ugurlu, Bayrak, & Kiritsakis, 2012) and has been reported as related to the fruity, sweet and green apple flavour.

Finally, considering the 15 markers selected, a series of linear combinations of these ions were studied in order to achieve the best separation possible between groups, and a model for the classification of the three groups was built. This model was based in a combined double classification, first differentiating between Extra/Virgin oils and Virgin/Lampante oils and then distinguishing among the three different classes.

345 As an efficiency pre-test, the samples used to build the model were classified using it, 346 comparing the model classification results with the classes assigned by the corresponding 347 "PANEL TEST", which was considered as reference. After the first classification, which 348 distinguishes between only two classes, "extra" and "lampante" oils, nearly 90 % of oils 349 were correctly classified, considering that "virgin" oils can be present in both groups. Then, 350 a second classification was applied to both groups, allowing a final distinction between the 351 three different classes. After the application of these series of binary classifications, around 352 70 % of the samples were correctly classified as extra, virgin or lampante. In all cases, the 353 worst classification efficiency corresponded to virgin samples, which are always the most 354 difficult to assign.

355 In order to take the maximum advantage possible from the samples, PLS-DA and OPLS-356 DA analysis considering the defects perceived by the PANEL TEST were created. As it can 357 be seen in **Fig S3a**, the samples cannot be easily separated by defects with a PLS-DA, as 358 the overlapping of samples with different defects is quite important. From **Fig S3b-S3e** 359 obtained facing samples with each kind of defect with the flawless extras in several OPLS-360 DA analyses, a clearer distinction is observed, which points out the fact that, from an 361 analytical point of view, the extra samples are easier to classify from the rest by the absence 362 of defect rather than the presence of a good attribute. This preliminary approach to the 363 study of the defects shows a promising potential for the refining of the classification 364 methodology, and will be further studied in depth with an appropriate design to maximize 365 the outputs of the research, but they won´t be analysed in more detail here.

366

367 **3.4. Model Validation**

368 Once the model was created with the 300 training samples, it was validated using the 125
369 blind samples to evaluate its efficiency. Results obtained for the ternary classification for
370 the blind samples were sent to the project coordinating centre in order to evaluate the
371 accuracy and the performance of the model. As it can be seen in **Figure 5,** extra samples
372 were always correctly classified as extra or, in some few cases, as virgin, but never as
373 lampante, which would lead them to be discarded with a high economic loss. In addition,
374 there was a low portion of virgin oils which got classified as lampantes and a small quantity
375 of lampantes which got overrated. It is important to highlight that the method is able to
376 distinguish the samples which have clear defects and poor fruity properties, as the match in
377 the first binary classification is around 90%. It is also remarkable that almost none of the
378 lampantes (a residual 5%) are overclassified as extras. This can be very useful in discarding
379 clear lampante samples and filtering the amount of samples which have to be tested by the
380 test panel, being an important complement to it. The distinction between lampantes and
381 virgins, as they usually share defects, has shown to be the most difficult, with a
382 considerable number of lampantes classified as virgin. This is also agreement with the
383 preliminary results of the PLSDA and OPLS-DAs obtained considering the defects found
384 in the samples **(Fig S3)**, and the efforts of future works will focus on the refining of the
385 methodology from those two perspectives.

386

387 ## 4. CONCLUSIONS

388 A simple and efficient methodology has been developed for the analysis of volatile
389 compounds in olive oil samples, with a prior extraction by P&T technique and subsequent
390 analysis by GC-(APCI)QTOF MS.

391 The use of GC-(APCI)QTOF MS has allowed to work in MS^E mode, which provides
392 information at low and high collision energies which has been proved to be an important
393 added benefit in order to elucidate the markers. The developed method has demonstrated to
394 be sensitive enough and capable of obtaining a considerable number of resolved
395 chromatographic peaks for volatile compounds. The use of APCI source with water as
396 modifier has favoured the formation of the protonated molecules ($[M+H]^+$) which, added to
397 the soft in source fragmentation, enhanced the sensitivity for most of the compounds when
398 compared to the traditional EI source.

399 The methodology was evaluated in terms of accuracy, obtaining an overall accuracy near to
400 70 % on 125 blind samples, discarding only lampante samples and a low percentage of
401 virgin samples, which makes it perfect for being a complementary analysis together with
402 the "PANEL TEST" method.

403 The metabolomics-based approach allowed the classification of the samples with a
404 considerable performance and a minimal data treatment after the method was built. Among
405 the different software used, the combination of MzMine 2.0 and Ez-Info dropped the best
406 results. MzMine 2.0 allowed the inclusion of a large number of samples in order to build
407 the model, and the addition of new samples to expand the population taken to create it. On
408 the same way, Ez-Info provided reliable results with a really simple user interface and was
409 preferred to other software due to the previous experience working with it.

410 In order to improve the classification method, a higher number of samples should be
411 required to include every possible combination of defects and goods. Considering the huge
412 variety of possible defects and goods than can give a specific classification to the samples,

the overall accuracy of 70 % achieved enable the developed methodology to be an important complement to the official PANEL TEST.

Acknowledgements

The authors acknowledge the financial support of Generalitat Valenciana, as research group of excellence (PROMETEO II/2014/023) and Collaborative Research on Environment and Food-Safety (ISIC/2012/016). This work and C. Sales grant has been supported by Universitat Jaume I research promotion plans (P1-1B2013-70).

The investigation has been performed within the frame of scientific collaboration between the "Ministerio de Agricultura, Alimentación y Medio Ambiente", the "Consejería de Agricultura, Pesca y Desarrollo Rural de la Junta de Andalucía" and the "Interprofesional del Aceite de Oliva Español".

FIGURE CAPTIONS

Figure 1: Workflow followed for chromatographic data treatment on MzMine 2.0.

Figure 2: PCA (a) and PLS-DA (b) figures corresponding to the samples analysed to create the model.

~~Comparison between the MS spectrum of α-farnesene obtained by (A) electron ionization (NIST library), and by APCI-QTOF (MS^E mode) (B) high energy function and (C) low energy function.~~ **Figure 3:** SPLOTs Extra vs Virgin, Virgin vs Lampante and Extra vs

Lampante. Tentative markers are enlightened in red squares and designated by their corresponding M(m/z)T(Rt). Example of a PCA (a) and PLS-DA (b) figures corresponding to one batch analysis of 60 samples.

Figure 4: GC-(APCI) QTOF MS narrow-window extracted ion chromatogram (mass window 0.02 Da) showing the detection of a possible marker, tentatively identified as ethyl-2-methylbutanoate- Experimental APCI accurate-mass spectra for the high collision energy function and for the low collision energy function are also shown with the chemical structures proposed for the most abundant fragment ion together with the experimental mass errors (in mDa).SPLOTs Extra vs Virgin, Virgin vs Lampante and Extra vs Lampante. Tentative markers are enlightened in red squares and designated by their corresponding M(m/z)T(Rt).

Figure 5: Confusion matrix showing the Comparison comparison between the objective (up) of getting every blind sample classified as "PANEL TEST" classification and the results (down) after processing these unknown samples through the entire developed procedure.

Con formato: Fuente: Sin Negrita

Figure S1: Comparison between the MS spectrum of α-farnesene obtained by (A) electron ionization (NIST library), and by APCI-QTOF (MS^E mode) (B) high energy function and (C) low energy function. GC-(APCI) QTOF MS narrow-window extracted ion chromatogram (mass window 0.02 Da) showing the detection of a possible marker, tentatively identified as ethyl-2-methylbutanoate- Experimental APCI accurate-mass spectra for the high collision energy function and for the low collision energy function are also shown with the chemical structures proposed for the most abundant fragment ion together with the experimental mass errors (in mDa).

457 **Figure S2:** Full-scan total ion chromatogram (TIC) obtained by the injection of a QC.

458

459 **Figure S3:** (A) PLS-DA analysis based on the defects determined by the PANEL TEST in
460 a subset of 150 samples used for the creation of the model. OPLS-DA analysis facing (B)
461 "Atrojado" defect versus no defect, (C) "Agrio" defect versus no defect, (D) "Moho" defect
462 versus no defect and (E) "Heno" defect versus no defect. Darkened area highlights the
463 region of major predominance of the defect.

464

465

466 **Conflict of Interest**

467 The authors declare no conflict of interest

468

469 **REFERENCES**

470 Angerosa, F., Servili, M., Selvaggini, R., Taticchi, A., Esposto, S., & Montedoro, G.
471 (2004). Volatile compounds in virgin olive oil: occurrence and their relationship with
472 the quality. *Journal of Chromatography A*, *1054*(2004), 17–31.

473 Arvanitoyannis, I. S., Katsota, M. N., Psarra, E. P., Soufleros, E. H., & Kallithraka, S.
474 (1999). Application of quality control methods for assessing wine authenticity : Use of
475 multivariate analysis (chemometrics). *Trends in Food Science & Technology*,
476 *10*(1999), 321–336.

477 Barco-Bonilla, N., Plaza-Bolaños, P., Fernández-Moreno, J. L., Romero-González, R.,
478 Garrido, A., & Martínez, J. L. (2011). Determination of 19 volatile organic
479 compounds in wastewater effluents from different treatments by purge and trap
480 followed by gas-chromatography coupled to mass spectrometry. *Analytical and*
481 *Bioanalytical Chemistry*, *400* (2011), 3537–3546.

482 Beltran, J., Serrano, E., López, F. J., Peruga, a, Valcarcel, M., & Rosello, S. (2006).
483 Comparison of two quantitative GC-MS methods for analysis of tomato aroma based
484 on purge-and-trap and on solid-phase microextraction. *Analytical and Bioanalytical*
485 *Chemistry*, *385*(2006), 1255–1264.

Con formato: Español (alfab. internacional)

Código de campo cambiado

Con formato: Español (alfab. internacional)

Con formato: Español (alfab. internacional)

486 Bicchi, C., Iori, C., Rubiolo, P., & Sandra, P. (2002). Headspace sorptive extraction
487 (HSSE), stir bar sorptive extraction (SBSE), and solid phase microextraction (SPME)
488 applied to the analysis of roasted Arabica coffee and coffee brew. *Journal of*
489 *Agricultural and Food Chemistry*, *50*(2002), 449–459.

490 Cervera, M. I., Portolés, T., López, F. J., Beltrán, J., & Hernández, F. (2014). Screening and
491 quantification of pesticide residues in fruits and vegetables making use of gas
492 chromatography-quadrupole time-of-flight mass spectrometry with atmospheric
493 pressure chemical ionization. *Analytical and Bioanalytical Chemistry*, 406(2014 ,
494 6843-6855

495 Cevallos-Cevallos, J. M., Reyes-De-Corcuera, J. I., Etxeberria, E., Danyluk, M. D., &
496 Rodrick, G. E. (2009). Metabolomic analysis in food science: a review. *Trends in*
497 *Food Science & Technology*, *20*(2009), 557–566.

498 Council, I. O. (2016). Designations and definitions of olive oils. URL
499 http://www.internationaloliveoil.org/estaticos/view/83-designations-and-definitions-
500 of-olive-oils?lang=en_US . Accessed 14.04.16

501 Díaz, R., Pozo, O. J., Sancho, J. V, & Hernández, F. (2014). Metabolomic approaches for
502 orange origin discrimination by ultra-high performance liquid chromatography
503 coupled to quadrupole time-of-flight mass spectrometry. *Food Chemistry*, *157(2014)*,
504 84–93.

505 Dierkes, G., Bongartz, A., Guth, H., & Hayen, H. (2012). Quality Evaluation of Olive Oil
506 by Statistical Analysis of Multicomponent Stable Isotope Dilution Assay Data of
507 Aroma Active Compounds. *Journal of Agricultural and Food Chemistry, 60(2012),*
508 394-401.

509 Dutta, R., Kashwan, K. R., Bhuyan, M., Hines, E. L., & Gardner, J. W. (2003). Electronic
510 nose based tea quality standardization. *Neural Networks 16*(2003), 847–853.

511 Flath, R. A., Forrey, R. R., & Guadagni, D. G. (1973). Aroma components of olive oil.
512 *Journal of Agricultural and Food Chemistry, 21 (1973), 948-952*

513 Fredes, A., Sales, C., Barreda, M., Valcárcel, M., Roselló, S., & Beltrán, J. (2016).
514 Quantification of prominent volatile compounds responsible for muskmelon and
515 watermelon aroma by purge and trap extraction followed by gas chromatography –
516 mass spectrometry determination. *Food Chemistry*, *190(2016)*, 689–700.

517 Hu, W., Zhang, L., Li, P., Wang, X., Zhang, Q., Xu, B., Ma, F., & Ding, X. (2014).
518 Characterization of volatile components in four vegetable oils by headspace two-
519 dimensional comprehensive chromatography time-of-flight mass spectrometry.
520 *Talanta*, *129*(2014) 629–635.

521 Jiménez, A., Aguilera, M. P., Beltrán, G., & Uceda, M. (2006). Application of solid-phase
522 microextraction to virgin olive oil quality control. *Journal of Chromatography. A*,
523 *1121*(2006), 140–144.

524 Kalua, C. M., Allen, M. S., Bedgood Jr, D.R., Bishop, A. G., Prenzler, P. D., & Robards, K.
525 (2007). Olive oil volatile compounds , flavour development and quality : A critical
526 review. *Food Chemistry, 100 (2007)*, 273–286.

527 Kieken, F., Pinel, G., Antignac, J.-P., Monteau, F., Paris, A. C., Popot, M.-A., Bonnaire,

Con formato: Español (alfab. internacional)

Con formato: Español (alfab. internacional)

Y., & Le Bizec, B. (2009). Development of a metabonomic approach based on LC-ESI-HRMS measurements for profiling of metabolic changes induced by recombinant equine growth hormone in horse urine. *Analytical and Bioanalytical Chemistry*, *394*(2009), 2119–2128.

Kind, T., Tolstikov, V., Fiehn, O., & Weiss, R. H. (2007). A comprehensive urinary metabolomic approach for identifying kidney cancer. *Analytical Biochemistry*, *363*(2007), 185–195.

Kiralan, M., Ozkan, G., Koyluoglu, F., Ugurlu, H. A., Bayrak, A., & Kiritsakis, A. (2012). Effect of cultivation area and climatic conditions on volatiles of virgin olive oil. *European Journal of Lipid Science and Technology*, *114*(2012), 552–557.

Lam, H. S., & Proctor, a. (2003). Milled Rice Oxidation Volatiles and Odor Development. *Journal of Food Science*, *68* (2003), 2676–2681.

Lee, J., Lee, H. K., Rasmussen, K. E., & Pedersen-Bjergaard, S. (2008). Environmental and bioanalytical applications of hollow fiber membrane liquid-phase microextraction: a review. *Analytica Chimica Acta*, *624*(2008), 253–268.

Luna, G., Morales, M. T., & Aparicio, R. (2006). Characterisation of 39 varietal virgin olive oils by their volatile compositions. *Food Chemistry*, *98*(2), 243–252.

Portolés, T., Mol, J. G. J., Sancho, J. V, López, F. J., & Hernández, F. (2014). Validation of a qualitative screening method for pesticides in fruits and vegetables by gas chromatography quadrupole-time of flight mass spectrometry with atmospheric pressure chemical ionization. *Analytica Chimica Acta*, *838(2014)*, 76–85.

Portolés, T., Pitarch, E., López, F. J., Hernández, F., & Niessen, W. M. A. (2011). Use of soft and hard ionization techniques for elucidation of unknown compounds by gas chromatography/time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry*, *25*(2011), 1589–1599.

Portolés, T., Sancho, J. V, Hernández, F., Newton, a, & Hancock, P. (2010). Potential of atmospheric pressure chemical ionization source in GC-QTOF MS for pesticide residue analysis. *Journal of Mass Spectrometry*, *45*(2010), 926–936.

Pouliarekou, E., Badeka, A., Tasioula-Margari, M., Kontakos, S., Longobardi, F., & Kontominas, M. G. (2011). Characterization and classification of Western Greek olive oils according to cultivar and geographical origin based on volatile compounds. *Journal of Chromatography. A*, *1218(2011)*, 7534–7542.

Reboredo-Rodríguez, P., González-Barreiro, C., Cancho-Grande, B., Valli, E., Bendini, A., Gallina Toschi, T., Simal-Gandara, T. (2016). Characterization of virgin olive oils produced with autochthonous Galician varieties. *Food Chemistry*, 212 (2016), 162-171.

Reboredo-Rodríguez, P., González-Barreiro, C., Cancho-Grande, B., & Simal-Gándara, J. (2012). Dynamic headspace/GC–MS to control the aroma fingerprint of extra-virgin olive oil from the same and different olive varieties. *Food Control*, *25*(2012), 684–695.

Ribes, A., Carrera, G., Gallego, E., Roca, X., Berenguer, M. J., & Guardino, X. (2007).

570 Development and validation of a method for air-quality and nuisance odors monitoring
571 of volatile organic compounds using multi-sorbent adsorption and gas
572 chromatography/mass spectrometry thermal desorption system. *Journal of*
573 *Chromatography. A*, *1140*(2007), 44–55.

574 Salemi, A., Lacorte, S., Bagheri, H., & Barceló, D. (2006). Automated trace determination
575 of earthy-musty odorous compounds in water samples by on-line purge-and-trap-gas
576 chromatography-mass spectrometry. *Journal of Chromatography. A*, *1136*(2006),
577 170–175.

578 Salihovic, S., Nilsson, H., Hagberg, J., & Lindström, G. (2013). Trends in the analysis of
579 persistent organic pollutants (POPs) in human blood. *TrAC Trends in Analytical*
580 *Chemistry*, *46 (2013)*, 129–138.

581 Sánchez-Peña, C. M., Luna, G., García-González, D. L., & Aparicio, R. (2005).
582 Characterization of French and Spanish dry-cured hams: influence of the volatiles
583 from the muscles and the subcutaneous fat quantified by SPME-GC. *Meat Science*,
584 *69*(2005), 635–645.

585 Serrano, E., Beltrán, J., & Hernández, F. (2009). Application of multiple headspace-solid-
586 phase microextraction followed by gas chromatography-mass spectrometry to
587 quantitative analysis of tomato aroma components. *Journal of Chromatography. A*,
588 *1216*(2009), 127–133.

589 Tikunov, Y., Lommen, A., Vos, C. H. R. De, Verhoeven, H. A., Bino, R. J., Hall, R. D., &
590 Bovy, A. G. (2005). A Novel Approach for Nontargeted Data Analysis for
591 Metabolomics . Large-Scale Profiling of Tomato Fruit Volatiles, Plant Physiology,
592 *139*(2005), 1125–1137.

593 Vaz-Freire, L. T., Da Silva, M. D., & Freitas, A. M. C. (2009). Comprehensive two-
594 dimensional gas chromatography for fingerprint pattern recognition in olive oils
595 produced by two different techniques in Portuguese olive varieties Galega Vulgar,
596 Cobrancosa e Carrasquenha. *Analytica Chimica Acta*, *633*(2009), 263–270.

597

Con formato: Español (alfab. internacional)

Con formato: Español (alfab. internacional)