Food Chemistry 378 (2022) 132104

Contents lists available at ScienceDirect

Food Chemistry

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

Geographical authentication of virgin olive oil by GC–MS sesquiterpene hydrocarbon fingerprint: Verifying EU and single country label-declaration

Beatriz Quintanilla-Casas[a,b], Berta Torres-Cobos[a,b], Francesc Guardiola[a,b], Maurizio Servili[c], Rosa Maria Alonso-Salces[d], Enrico Valli[e], Alessandra Bendini[e], Tullia Gallina Toschi[e], Stefania Vichi[a,b,*], Alba Tres[a,b]

[a] Departament de Nutrició, Ciències de l'Alimentació i Gastronomia, Campus de l'Alimentació Torribera, Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona. Av Prat de la Riba, 171. 08921 Santa Coloma de Gramenet, Spain

[b] Institut de Recerca en Nutrició i Seguretat Alimentària (INSA-UB), Universitat de Barcelona. Av Prat de la Riba, 171. 08921 Santa Coloma de Gramenet, Spain

[c] Dipartimento di Scienze Agrarie, Alimentari ed Ambientali, Università di Perugia, Via San Costanzo S.n.c., 06126 Perugia, Italy

[d] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Departamento de Biología, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Mar del Plata (UNMdP), Funes 3350, 7600 Mar del Plata, Argentina

[e] Dipartimento di Scienze e Tecnologie Agro-alimentari, Alma Mater Studiorum - Università di Bologna, Piazza Goidanich, 60, I-47521, Cesena, Italy

ARTICLE INFO

Keywords:
Virgin Olive Oil
Authenticity
Geographical origin
Sesquiterpene hydrocarbons
Fingerprint
SPME-GC–MS
PLS-DA

ABSTRACT

According to the last report from the European Union (EU) Food Fraud Network, olive oil tops the list of the most notified products. Current EU regulation states geographical origin as mandatory for virgin olive oils, even though an official analytical method is still lacking. Verifying the compliance of label-declared EU oils should be addressed with the highest priority level. Hence, the present work tackles this issue by developing a classification model (PLS-DA) based on the sesquiterpene hydrocarbon fingerprint of 400 samples obtained by HS-SPME-GC–MS to discriminate between EU and non-EU olive oils, obtaining an 89.6% of correct classification for the external validation (three iterations), with a sensitivity of 0.81 and a specificity of 0.95. Subsequently, multi-class discrimination models for EU and non-EU countries were developed and externally validated (with three different validation sets) with successful results (average of 92.2% of correct classification for EU and 96.0% for non-EU countries).

1. Introduction

Current global market of olive oil, in particular extra virgin (EVOO) and virgin olive oil (VOO) categories, is threatened by fraudulent practices due to its high nutritional, sensory and, therefore, economic value. As a proof of fact, the 2020 annual report from the Agri-Food Fraud Network (FFN) from the European Union (EU) placed olive oil and other edible oils as the most notified food category with respect to non-compliances (European Union, 2021). This report also revealed that almost the 40% of the total non-compliances were due to mislabelling issues. Geographical origin is known to play a key role for consumers' choice during olive oil purchasing process (Conte et al., 2020). Moreover, a recent study (Carzedda et al., 2021) highlighted the general preference for local products, not only due to sustainability awareness but also because of a greater perception of safety. For these reasons, the EU Regulation N° 29/2012 aims to protect consumers from misleading information, as well as to maintain the competitiveness of the sector, through a mandatory label-declaration of the geographical origin for both EVOO and VOO. According to its fourth article, the declaration of origin shall consist of a reference to the EU, to the EU member state or to the third country, as appropriate. In the case of olive oils produced in more than one EU or non-EU countries, or in a mixture of both EU and non-EU countries, the corresponding blend should be mentioned. Finally, according to the EU certification, the geographical provenance can also be stated as Protected Designation of Origin (PDO) or Protected Geographical Indication (PGI) (Regulation (EU) No 1151/2012). The declared provenance of virgin olive oils is currently only assessed through documental review by the corresponding control bodies because an official analytical method is not available yet. Disposing of an instrumental method for verifying the geographical origin of virgin

* Corresponding author at: Departament de Nutrició, Ciències de l'Alimentació i Gastronomia, Campus de l'Alimentació Torribera, Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Av Prat de la Riba, 171. Edifici Gaudí, 08921 Santa Coloma de Gramenet, Spain.
E-mail address: stefaniavichi@ub.edu (S. Vichi).

https://doi.org/10.1016/j.foodchem.2022.132104
Received 22 October 2021; Received in revised form 31 December 2021; Accepted 5 January 2022
Available online 10 January 2022

olive oil would help to detect counterfeiting cases. Thus, the opportunity of fraud would be reduced in this regard.

Numerous studies were performed for years aiming at virgin olive oil geographical authentication, most of which have been compiled in recent comprehensive reviews about olive oil authentication (Conte et al., 2020; Valli et al., 2016), even specifically regarding geographical assessment (Tahir et al., 2022). Tahir et al. (2022) evidenced that a wide variety of analytical techniques ranging from the popular separation tools (gas and liquid chromatography, GC and LC) to rapid spectroscopic methods have been combined with chemometrics for olive oil geographical authentication. The latter ones have been broadly applied as independent techniques (Near Infrared, NIR/ Mid-Infrared, MIR (Woodcock et al., 2008); Nuclear Magnetic Resonance, NMR (Alonso-Salces et al., 2015; Winkelmann & Küchler, 2019); Raman (Sánchez-López et al., 2016); Fluorescence (Lia et al., 2020); Fourier Transform Infrared, FTIR (Hennessy et al., 2009)) or combining the corresponding spectra through data-fusion techniques (Bevilacqua et al., 2013).

One of the first works that addressed geographical authentication of olive oil was based on the fatty acid (FA) profile (Forina & Tiscornia, 1982). Nowadays, FA are generally chosen together with other valuable analytical markers, such as volatile organic compounds (VOC) (Kosma et al., 2017) and phenols (Ben Hlima et al., 2017), to obtain complementary information. In this sense, the triacylglycerol composition (Vera et al., 2019) and its stereospecific distribution of FA (Vichi et al., 2007) have been used for this purpose as well. Phenols (Bakhouche et al., 2013) and VOC have been some of the traditional markers for origin discrimination of olive oil, without forgetting that these compounds might be heavily influenced by storage and processing factors. The analysis of the latter ones was not only carried out through chromatographic techniques (Cecchi et al., 2020; Lukić et al., 2019), but also by rapid tools based on electronic devices (e-nose and e-tongue) (Palagano et al., 2020; Souayah et al., 2017) or Proton Transfer Reaction-Mass Spectrometry (PTR-MS) (Araghipour et al., 2008). In the last years, trace elements analysis by Inductively Coupled Plasma (ICP)-MS (Damak et al., 2019) and isotopic fingerprint assessment (Bontempo et al., 2019; Portarena et al., 2017) have become popular due to the high correlation with the soil properties where the olive trees are grown.

Casadei et al. (2021) surveyed the olive oil stakeholders and the EU FFN national contact points about olive oil fraudulent practices. Regarding EVOO and VOO geographical origin counterfeiting, the results of the survey addressed to the EU FFN national contact points pointed out that verifying the compliance with declared EU and non-EU origin are the cases which need more control activities, including the provenance from a given country, followed by the authentication of geographical certifications (PDO and PGI). Conversely, some studies considered samples from two different countries, one EU and one non-EU (Ben Mohamed et al., 2018; Borges et al., 2017), while others proposed a country approach regardless of the EU membership including the olive oil traditional producing countries within the Mediterranean basin (Alonso-Salces et al., 2015; Bajoub et al., 2018; Cecchi et al., 2020). Just a few of them faced the EU vs non-EU discrimination (Bontempo et al., 2019; Palagano et al., 2020), evidencing the need for a fit-for-purpose analytical tool to verify the label-declared EU provenance as the foremost aim, which would also be suitable to assess the country of origin.

Therefore, the present research pursued to achieve a reliable instrumental method that allows geographical authentication of EVOO and VOO produced in very specific and homogeneous areas as well as in wider regions with higher heterogeneity in terms of the traditional cultivars and pedoclimatic conditions, according to the aforementioned priority standards. Hence, this study focused on developing and validating classification models (PLS-DA) for the first geographical authentication priority, EU vs. non-EU and for single countries (EU and non-EU members) (Fig. 1). Our proposal is based on a previous work (Quintanilla-Casas, Bertin, Leik, Bustamante, Guardiola, Valli, Bendini, Gallina Toschi, Tres, & Vichi, 2020), where preliminary models to verify the country of origin were developed based on sesquiterpene hydrocarbons (SH) fingerprint analysed by Headspace – Solid Phase Micro Extraction (HS-SPME) and GC–MS. These semi-volatile compounds are known to be related with genetic and pedoclimatic factors, while scarcely influenced by processing and storage conditions (Damascelli & Palmisano, 2013; Vichi et al., 2018). Both target and fingerprinting approaches were investigated and compared with regards to classification ability by PLS-DA, concluding that chromatographic fingerprint allowed for better geographical classification performance of virgin olive oils from different countries (Quintanilla-Casas, Bertin, et al., 2020). The present study involves a big dataset with great diversity, including different productive regions, olive cultivars, crop years and even analytical batches, in order to evaluate the performance of the authentication approach in a more realistic scenario where the natural variability is highly represented.

2. Material and methods

2.1. Sampling

The sample set consisted of 400 traceable virgin olive oils from different EU member states and third countries – here so-called non-EU, including 246 oils produced in 6 EU member states (Croatia, HRV; Greece, GRC; Italy, ITA; Portugal, POR; Slovenia, SVN; Spain, ESP) and 154 oils from 4 non-EU countries (Argentina, ARG; Morocco, MAR; Tunisia, TUN; Turkey, TUR) (Table 1). They were obtained in the framework of the projects OLEUM (EC H2020 Programme 2014–2020) and Autenfood (ACCIÓ- Programa Operatiu FEDER Catalunya 2014–2020) or directly purchased from producers. These samples were produced at real industrial conditions during different campaigns (harvests from 2015/16 to 2019/20) and were graded as EVOO or VOO according to the panel test assessment. Additional information about

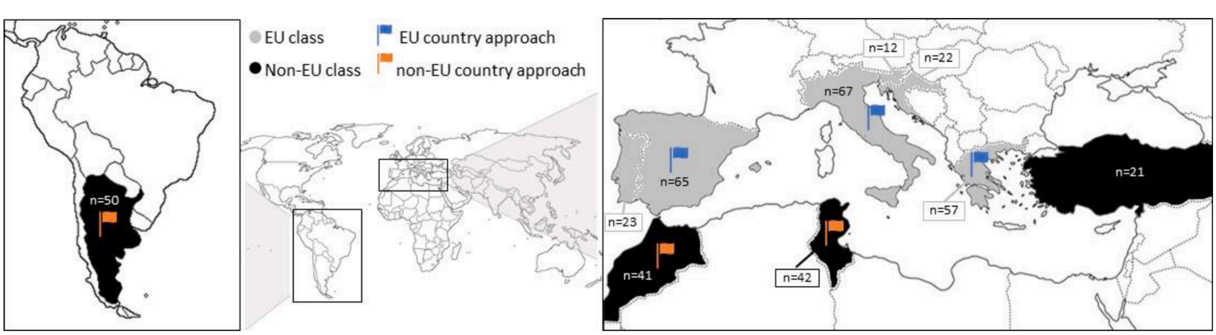

Fig. 1. Extra virgin and virgin olive oil sampling, according to the three authentication levels: EU (in grey) vs. non-EU (in black) oils, EU members (blue flag) and non-EU countries (orange flag). *EU: European Union.* (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Number of virgin olive oil samples (virgin and extra virgin categories) in the training and validation sets used to develop the EU vs non-EU binary discrimination models and the multi-class models for the country of origin.

EU member states	Training sets	Validation sets	Total	Non-EU countries	Training sets	Validation sets	Total
ESP	52	13	65	ARG	40	10	50
GRC	46	11	57	MAR	33	8	41
HRV[a]	17	5	22	TUN	34	8	42
ITA	54	13	67	TUR[a]	16	5	21
POR[a]	18	5	23				
SVN[a]	10	2	12				
Total	197	49	246	Total	123	31	154

ARG: Argentina; CHL: Chile; ESP: Spain; GRC: Greece; HRV: Croatia; ITA: Italy; MAR: Morocco; POR: Portugal; SVN: Slovenia; TUN: Tunisia; TUR: Turkey.
[a] Countries with<40 samples were not included in the corresponding multi-class models (country approach).

olive oil samples is available in Table S1 (Supplementary material). Samples were stored under N_2 atmosphere at −20 ℃ until analysis, which was performed over a period of seven months in 3 different analytical batches. These batches corresponded mainly to the harvesting season, and they included samples from several geographical origins that were randomly measured.

2.2. HS-SPME-GC–MS

SHs were extracted from samples by HS-SPME and subsequently analysed by GC–MS according to Torres-Cobos et al. (2021), based on the original protocol from Vichi et al. (2006). In order to evaluate the performance of the analytical system, qualitative and quantitative representative virgin olive oil samples (quality control samples) were analysed periodically. These quality control samples allowed to detect signal drift as well as magnitude changes between analytical batches. For this, the relative standard deviation (RSD%) was evaluated for the analytical signal as described in Quintanilla-Casas, Marin, Guardiola, García-González, Barbieri, Bendini, Toschi, Vichi, & Tres (2020).

2.3. Chemometrics

2.3.1. Data extraction and pre-processing

Seven data matrices were built, one for each extracted ion chromatogram (EIC) acquired in single ion monitoring (SIM): *m/z* 93, 119, 157, 159, 161, 189 and 204. These ions were selected because they are specific ions of SHs as reported by Vichi et al. (2006). Intensity of scans, from minute 21 to 42, were the matrix columns while the rows corresponded to the samples (400 × 3197). Retention time shifting among samples was corrected by means of Correlation Optimized Shifting (coshift) (Larsen et al., 2006) followed by Correlation Optimized Warping (COW) algorithm (Nielsen et al., 1998), both performed in Matlab R2020b®. Data was then normalized to the maximum intensity (row wise) to handle magnitude changes that can occur when a large sample set is analysed by GC–MS over a long period, as it was the case. Baseline correction (automated weighted least squares) was applied for the same reason. Once individually aligned and normalized, the 7 matrices were concatenated conforming a two-way unfolded matrix of 400 samples and 22,379 variables.

2.3.2. Calibration and validation of EU vs. non-EU discriminant model

In order to discriminate between the EU and the non-EU virgin olive oils, PLS-DA was applied to the unfolded matrix described above. The whole sample set (n = 400) was split into a training set that included 80% of the samples (n = 320, being 197 from EU and 123 from outside the EU) and a validation set conformed by the remaining 20% of the samples (n = 80, being 49 from the EU and 31 from outside the EU) (Table 1). It is important to highlight that samples were distributed in the training and validation sets at random after the analysis, but keeping the balance between the different olive harvests, regions, countries, and analytical batches. In order to evaluate the stability of the classification approach when different samples were used to build the model, this split was done three times (3 iterations). Therefore, three different PLS-DA models were built and cross-validated with three different training sets, and later, each of them was externally validated with the corresponding validation set that consisted of samples previously unseen by that model. Unit variance scaling and mean centering (column wise) were selected as preprocessing techniques, as there were variables in the fingerprint with very different intensity values. Software used was SIMCA v13.0© (Umetrics AB, Sweden).

In this EU vs. non-EU binary PLS-DA, classes were expressed as PLS dummy variables (being 0 for EU class, and 1 for non-EU class). The PLS predicted value (PV) obtained for each sample was used to classify it into the class with the highest PV, provided it was above the classification threshold (here, PV = 0.5). The three calibration models were internally validated by leave 10%-out cross-validation. The optimal number of latent variables (LV) of each PLS-DA model was selected according to the lowest root mean squared error of cross validation (RMSEcv), to the highest cumulative Q^2 - defined as the estimated total variation of the discriminant categories that can be predicted by the model - and to the misclassification results. Hotelling's T^2 and Q-residuals were used to detect outliers. Model overfitting was evaluated by examining various parameters such as the RMSEcv, the Q^2 obtained by the permutation test (Q^2 of 20 models developed after randomly permuting the sample's class) and the ANOVA of cross-validated residuals.

For each of the three iterations, cross-validation and external validation results were expressed as the % of correct classification for each category and for the total sampling, diagnostic sensitivity (eq. (1)) and diagnostic specificity (eq. (2)) (Magnusson & Örnemark, 2014). Results from the three iterations were averaged and the standard deviation was calculated (n = 3).

$$\textit{Diagnostic sensitivity} = \frac{\textit{Non-EU samples correctly classified}}{\textit{Total non-EU samples}} \quad (1)$$

$$\textit{Diagnostic specificity} = \frac{\textit{EU samples correctly classified}}{\textit{Total EU samples}} \quad (2)$$

2.3.3. Calibration and validation of discriminant models by country of origin

To develop classification models to verify the geographical origin of virgin olive oils by country, only the samples from the countries that were represented by at least 40 samples were included in the data set (n = 322; Table 1). They agreed with the main EU and to some of the main non-EU olive oil producing countries according to the Food and Agriculture Organization database (FAOSTAT) for the crop year 2018. Two independent multi-class PLS-DAs were developed: one to classify EU samples according to their EU member state (ITA, ESP and GRC, n = 189) and another for the non-EU ones (ARG, TUN and MAR, n = 133). Both the EU and the non-EU subsets were split in three different training sets (80% of the samples, n = 152 for the EU subsets and n = 107 for the non-EU subsets) and the three corresponding validation sets (20% of the samples, n = 37 for the EU subsets and n = 26 for the non-EU subsets) (3 iterations). As explained in section 2.3.2, samples were distributed in the training and validation sets at random after the analysis, but keeping the balance between the different olive harvests, regions, countries, and analytical batches. Unit variance scaling and mean centering (column wise) were selected as preprocessing techniques, as there were variables in the fingerprint with very different intensity values.

In a multi-class PLS-DA, a dummy Y matrix holding as many classification vectors as classes is used in the PLS regression, each vector having values of 1 for one class and 0 for all the other classes. Here, each sample was classified into the class corresponding to the vector leading to the highest PV; but samples whose PV did not reach the classification threshold (PV < 0.5) for any vector were not assigned to any country. That is because multi-class models indeed work as multiple binary models of each class against the rest of samples. All calibration models (three for the EU approach and three for the non-EU approach) were internally validated by leave 10%-out cross-validation and externally validated by predicting the class of samples in the corresponding validation sets, as those samples were out of the model's calibration step. The optimal number of LVs, outliers' assessment and models' overfitting evaluation were carried out as explained in section 2.3.2.

In this case, model's reliability for both internal and external validation, was expressed by: i) % of not assigned samples (samples with PV < 0.5) considering all countries (eq. (3)) and for each country (eq. (4)), and by ii) % of correctly classified samples (PV > 0.5) considering all countries (eq. (5)) and for each country (eq. (6)). Results were averaged and the standard deviation was calculated (n = 3).

$$\underset{(all\ countries)}{Not\ assigned\ samples\ (\%)} = \frac{Total\ unassigned\ samples}{Total\ samples} \times 100 \tag{3}$$

$$\underset{(per\ country)}{Not\ assigned\ samples\ (\%)} = \frac{Unassigned\ samples\ from\ a\ country}{Samples\ from\ a\ country} \times 100 \tag{4}$$

$$\underset{(all\ countries)}{Correctly\ classified\ samples\ (\%)} = \frac{Total\ correctly\ classified\ samples}{Total\ assigned\ samples} \times 100 \tag{5}$$

$$\underset{(per\ country)}{Correctly\ classified\ samples\ (\%)} = \frac{Correctly\ classified\ samples\ from\ a\ country}{Assigned\ samples\ from\ a\ country} \times 100 \tag{6}$$

2.3.4. Exploration of PLS-DA coefficients

As explained above, two different geographical authentication approaches were developed based on the SH fingerprint and PLS-DA: at the EU vs non-EU level, and at a country level (for EU and for non-EU countries). In order to explore which variables of the SH fingerprint were more relevant for each model and how they varied depending on whether they were grouped into a higher (e.g., EU class) or a lower (e.g. country) authentication level, both types of models were rebuilt using the same samples (n = 322). Thus, all samples from well represented EU and non-EU countries were employed to build a binary (EU, n = 189, vs. non-EU, n = 133) and a multi-class PLS-DA (ESP, n = 65; GRC, n = 57; ITA, n = 67; ARG, n = 50; MAR, n = 41; TUN, n = 42), cross-validated by leave 10%-out. Significant regression coefficients - positive values greater than the corresponding standard error – were extracted for the EU category from the binary model and for each of the EU countries (ESP, GRC and ITA) and compared.

3. Results and discussion

Geographical authentication models have been calibrated and validated to fulfil two complementary purposes: i) to discriminate between EU and non-EU virgin olive oils, and ii) to verify the country of provenance, among EU member states and non-EU countries. Results obtained in both approaches are shown and discussed in the following subsections.

3.1. Geographical authentication approach I: Discrimination between EU vs. non-EU virgin olive oils

Three binary PLS-DAs were calibrated with 320 different virgin olive oil samples produced inside (n = 197) and outside (n = 123) the EU. These models needed between 8 and 9 LVs, showed a minimum prediction power (Q^2) of 0.651 and a maximum cross-validation error (RMSEcv) of 0.255. Individual contingency tables for each model are available at Table S2 (Supplementary material). As shown in Table 2, successful discrimination between EU and non-EU virgin olive oils was achieved, providing an average correct classification of 99.6% in cross-validation and 89.6% in external validation. In both cases, the models performed slightly better for the EU class, providing the maximum specificity (equals to 1) in cross-validation, which slightly decreased to 0.95 in the external validation. This could be explained due to the higher number of samples considered in the EU compared to the non-EU. This fact was observed by Palagano et al. (2020), who developed a EU vs non-EU model using the whole VOC fingerprint obtained by Fast GC (FGC) and also achieved better classification results for the EU class in the external validation. While the classification rates that they obtained through Artificial Neural Network (ANN) on the FGC fingerprint of VOC (EU: 92.3% for EU and 88.4% for non-EU classes) were comparable to the ones in the present study (Table 2), the discrimination efficiency when models were built by PLS-DA slightly dropped. On the contrary, other previous studies (Bontempo et al., 2019) that also attempted to differentiate EU from non-EU oils but by stable isotope ratio analysis, were not capable of completely separating EU from non-EU oils, obtaining better discrimination results between oils from single countries. They suggested that the discrimination between EU and non-EU virgin olive oils was not attributable to different characteristics of the whole EU or non-EU categories, but to specific isotopic profiles of the individual countries that conformed both the EU and the non-EU categories resulting from the particular pedoclimatic conditions of each production area. In the present study, PLS discriminant models based on SH fingerprint seem to overcome the reported issue, providing successful results for the discrimination of EU and non-EU virgin olive oils. This will be shown in section 3.3, where models will be investigated and discussed through their PLS-DA coefficients.

3.2. Geographical authentication approach II: Verification of the country of provenance.

3.2.1. EU member states

Although 6 EU countries (ESP, GRC, HRV, ITA, POR and SVN) conformed the EU class of the wider EU vs. non-EU authentication model, only those with at least 40 samples were included in the country discrimination approach. Therefore, virgin olive oils from ESP, GRC and ITA, the main EU producer members, were considered to develop (n =

Table 2
EU vs. non-EU discrimination approach: internal (leave 10%-out cross-validation) and external validation results of the binary PLS-DA models. Average values ± standard deviation of the three sample sets (3 iterations) for each category are provided.

	Cross-validation of training sets				External validation			
	n[a]	Correct classification (%)[b]	Sensitivity[c]	Specificity[d]	n[e]	Correct classification (%)[b]	Sensitivity[c]	Specificity[d]
EU	197	100 ± 0.00		1 ± 0.00	49	95.2 ± 1.18		0.95 ± 0.01
Non-EU	123	98.9 ± 1.24	0.99 ± 0.01		31	80.7 ± 8.53	0.81 ± 0.09	
Total	320	99.6 ± 0.48			80	89.6 ± 2.89		

Binary PLS-DAs with 8–9 latent variables, RMSEcv < 0.255, $Q^2 > 0.651$, ANOVAcv p value < 0.05.
[a] Number of samples in each of the three training sets.
[b] Correctly classified samples × 100/Total samples.
[c] Non-EU samples correctly classified as non-EU/total non-EU samples.
[d] EU samples correctly classified as EU samples/total EU samples.
[e] Number of samples in each of the three validation sets.

Table 3
EU single member classification approach: internal (leave 10%-out cross-validation) and external validation of the PLS-DA models. Mean and standard deviation of the three sample sets (3 iterations), for each category.

	Cross-validation of training sets			External validation		
	n[a]	Not assigned[b] (% of total sampling)	Correctly classified[c,d] (% of assigned samples)	n[e]	Not assigned[b] (% of total sampling)	Correctly classified[c,d] (% of assigned samples)
ESP	52	1.9 ± 0.00	100 ± 0.00	13	7.7 ± 0.00	91.7 ± 8.34
GRC	46	0.0 ± 0.00	100 ± 0.00	11	12.1 ± 10.50	93.1 ± 4.70
ITA	54	0.6 ± 1.07	98.8 ± 0.87	13	5.1 ± 4.44	91.9 ± 6.32
Total	152	0.9 ± 0.38	99.6 ± 0.31	37	8.1 ± 2.70	92.2 ± 3.84

ESP: Spain; GRC: Greece; ITA: Italy. PLS-DAs with 8–10 latent variables, RMSEcv ESP < 0.264, RMSEcv GRC < 0.275, RMSEcv ITA < 0.291, $Q^2 > 0.639$, ANOVAcv p value < 0.05.
[a] Number of samples in each of the three training sets.
[b] Not assigned samples (%) (PV < 0.5) per country (unassigned samples from a country × 100/samples from a country) and for all countries (total unassigned samples × 100/total samples).
[c] Correctly classified samples (%) (PV > 0.5) per country (correctly classified samples from a country × 100/assigned samples from a country) and for all countries (total correctly classified samples × 100/total assigned samples).
[d] Weighted mean and standard deviation, given that the number of assigned samples was different for each of the three sets.
[e] Number of samples in each of the three validation set.

Table 4
Non-EU country classification approach: internal (leave 10%-out cross-validation) and external validation of the PLS-DA models. Mean and standard deviation of the three sample sets (iterations), for each category.

	Cross-validation of training sets			External validation		
	n[a]	Not assigned[b] (% of total sampling)	Correctly classified[c,d] (% of assigned samples)	n[e]	Not assigned[b] (% of total sampling)	Correctly classified[c,d] (% of assigned samples)
MAR	33	0.0 ± 0.00	100 ± 0.00	8	4.2 ± 7.22	95.7 ± 6.58
ARG	40	0.0 ± 0.00	100 ± 0.00	10	6.7 ± 5.77	96.4 ± 4.79
TUN	34	0.0 ± 0.00	100 ± 0.00	8	4.2 ± 7.22	95.7 ± 5.95
Total	107	0.0 ± 0.00	100 ± 0.00	26	5.2 ± 2.22	96.0 ± 3.25

MAR: Morocco; ARG: Argentina; TUN: Tunisia. Multi-class PLS-DAs with 7 latent variables, RMSEcv ARG < 0.243, RMSEcv MAR < 0.252, RMSEcv TUN < 0.227, $Q^2 > 0.755$, ANOVAcv p value < 0.05.
[a] Number of samples in each of the three training sets.
[b] Not assigned samples (%) (PV < 0.5) per country (unassigned samples from a country × 100/samples from a country) and for all countries (total unassigned samples × 100/total samples).
[c] Correctly classified samples (%) (PV > 0.5) per country (correctly classified samples from a country × 100/assigned samples from a country) and for all countries (total correctly classified samples × 100/total assigned samples).
[d] Weighted mean and standard deviation, given that the number of assigned samples was different for each of the three sets
[e] Number of samples in each of the three validation set.

152, three times) and validate (n = 37, three times) the EU member authentication models. Individual contingency tables for each model are available at Table S3 (Supplementary material). Bearing in mind this is a more challenging purpose due to the closeness of the productive areas, the overall cross-validation results were successful, with an average of 99.6% of samples correctly classified and<2% of unassigned samples, without remarkable differences among classes (Table 3). The same trend was observed for the external validation, where a 92.2% (mean value) of the 37 validation samples was correctly assigned to the corresponding country. As exposed in Table 3, the classification model not only showed an excellent performance for all countries, but also left a small number of samples unassigned (average 8.1% of total sampling).

Previous works shared the aim of discriminating olive oils produced in different EU countries. Those that included samples from the main producer countries (ESP, ITA and GRC) also reported good classification rates, but either applied less affordable analytical techniques (Winkelmann & Küchler, 2019) or required more than one instrumental method even for a reduced sample set (Schwolow et al., 2019). Therefore, our results show that the models based on the SH fingerprint were useful to verify the geographical origin at a country level for these three EU countries (ESP, ITA and GRC), even if a high natural variability was considered in the sampling. Applying the model to oils from other EU countries would require collecting a suitable number of samples from a country, and to develop and externally validate the models, but the results obtained for ESP, ITA and GRC point to successful models.

3.2.2. Countries outside the EU

The steps and criteria explained in the section above for the EU country authentication model, were also followed to build the classification model for the non-EU countries. Consequently, three well represented non-EU countries with at least 40 samples each (ARG, MAR and TUN) were selected to calibrate (n = 107, three times) and validate (n = 26, three times) the multi-class PLS-DA. Individual contingency tables for each model are available at Table S4 (Supplementary material). In this context, models' performance resulted successful, obtaining an overall correct classification of 100% and 96.0% for internal and external validation, respectively (Table 4). All samples were assigned to a given country in cross-validation and few samples were left unassigned at the external validation (average of 5.1%). Since there is a high distance between ARG and the two other non-EU countries (TUN and MAR), the good performance of model for the identification of ARG samples (96.4% correct classification) was somehow expected. However, the same model also showed an excellent efficiency of classification for samples from MAR (95.7%) and TUN (95.7%) that are closer countries.

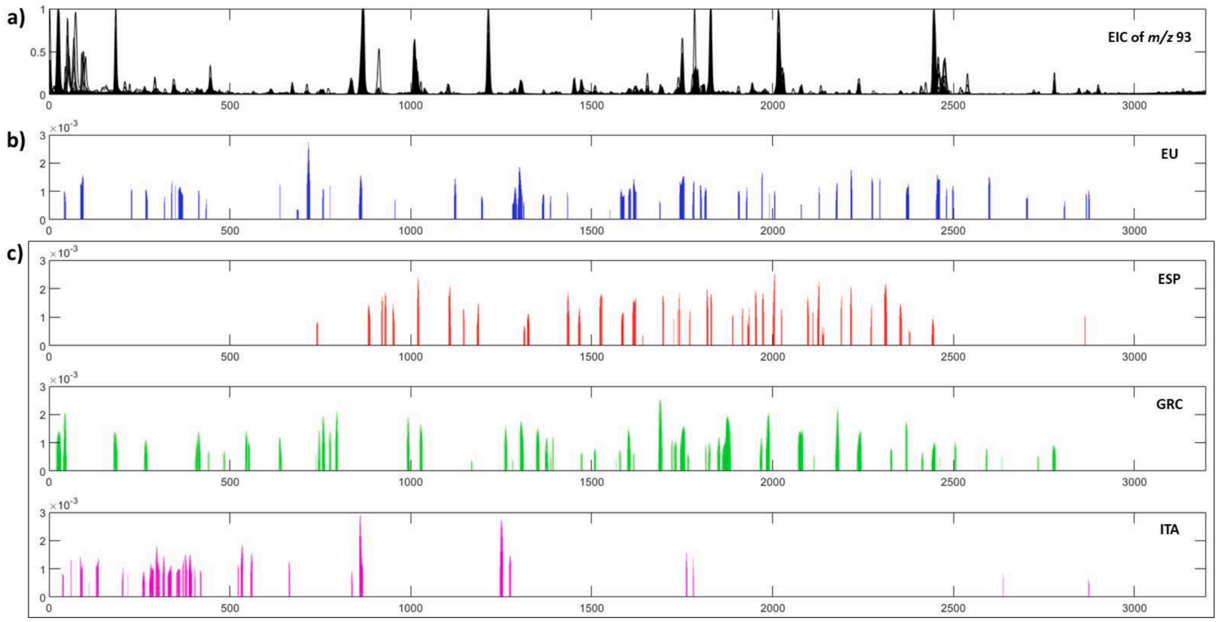

Fig. 2. PLS-DA coefficients exploration for EU virgin olive oils: a) EIC of *m/z* 93; b) Significant PLS regression coefficients for the EU class (binary PLS-DA: EU, n = 189, vs. non-EU, n = 133); c) Significant PLS regression coefficients for EU countries (obtained from the multi-class PLS-DA with 6 countries: ESP, n = 65, GRC, n = 57, ITA, n = 67, ARG, n = 50, MAR, n = 41, and TUN, n = 42). *EU: European Union; ESP: Spain; GRC: Greece; ITA: Italy.*

3.3. Exploration of discriminant models by PLS coefficients

One of the purposes of investigating PLS-DA coefficients was knowing whether variables with higher prediction influence for a given country remained or not when the model intended to discriminate on a higher authentication level (group of countries). In order to achieve that, as explained in section 2.3.4., a binary EU vs. non-EU model and a multi-class model by countries built with the very same samples were needed. Information about the corresponding discriminant models are available in Table S5 (Supplementary information), which indicated that the classification approach by countries resulted less efficient when EU and non-EU countries were considered together (multi-class PLS-DA with 6 countries in Table S5 of Supplementary Material) than when models only included the EU (Table 3) or the non-EU countries (Table 4).

By way of example, Fig. 2 shows the EIC of the *m/z* 93 (Fig. 2a) and the extracted PLS coefficients from the binary (EU vs non-EU, n = 322) (Fig. 2b) and the multi-class (6 countries (n = 322), three EU member states and three non-EU countries) (Fig. 2c) discriminant models. It evidenced that most of variables with high discriminant power for the EU category (Fig. 2b) also had a high discriminant power for some of the EU countries (Fig. 2c). This fact would confirm that the EU category is actually explained by SH information from different production areas within the EU, as previous studies suggested for the stable isotopic profile (Bontempo et al., 2019). However, it is worth to point out that some of the SH variables with high coefficients for each EU country (Fig. 2c) did not show high coefficients in the discrimination of the EU class (Fig. 2b), because different SH analytical information is extracted for each authentication purpose. Besides, it is worth to point out that significant coefficients corresponded to both major and minor SH (Fig. 2a), as observed in Quintanilla-Casas, Bertin, et al. (2020).

4. Conclusions

The present work showed that the SH fingerprint by HS-SPME-GC–MS together with the proposed chemometric approach could be the fit-for-purpose tool for virgin olive oil geographical authentication, fulfilling the needs of control activities previously highlighted by some of the EU FFN contact points. The developed model efficiently discriminated EU vs non-EU samples, classifying correctly a 89.6% of samples in external validation. Therefore, it overcame the challenge that the EU and non-EU classes are highly heterogeneous categories, conformed by groups of countries with particular pedoclimatic conditions. The SH fingerprint provided a large amount of information, but the PLS-DA allowed to consider the most relevant variables according to the categories driving the analysis. Successful results were also obtained when the classification models were performed by countries: the EU member states model correctly classified a 92.2% of assigned samples (8.1% unassigned samples) and the non-EU countries model correctly classified a 96.0% of assigned samples (5.2% of unassigned samples). It is remarkable that these high % of correct assignments were obtained with a data set that considered a high virgin olive oil natural heterogeneity and analytical variability as it included samples from the main cultivars for each production area, from different crop-years and they were analysed in several analytical batches. Bearing this information in mind, we can hypothesize that the proposed approach could be scaled down to authenticate the origin of oils obtained from smaller and highly close areas. Thus, it could be useful to face the second geographical authentication priority that aims to verify the provenance of PDO/PGI virgin olive oils.

Funding

This work was developed in the context of the project OLEUM “Advanced solutions for assuring authenticity and quality of olive oil at global scale”, funded by the European Commission within the Horizon 2020 Program (2014–2020, grant agreement no. 635690). The information and views set out in this article are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein. This study was also funded by the project AUTENFOOD, from *ACCIÓ-Generalitat de Catalunya* and the European Union through the ERDF programme for Catalonia 2014–2020 (Ref COMRDI-15-1-0035).

B. Quintanilla-Casas and A. Tres thanks the Spanish Ministry of Science, Innovation and Universities for a predoctoral fellowship (FPU16/01744) and for a “Ramón y Cajal” postdoctoral fellowship

(RYC-2017-23601), respectively. B. Torres-Cobos thanks Generalitat de Catalunya (Spain) for a predoctoral fellowship (2020FI_B00595).

CRediT authorship contribution statement

Beatriz Quintanilla-Casas: Formal analysis, Investigation, Methodology, Validation, Data curation, Visualization, Writing – original draft. **Berta Torres-Cobos:** Formal analysis, Investigation, Data curation. **Francesc Guardiola:** Supervision, Writing – review & editing. **Maurizio Servili:** Resources, Writing – review & editing. **Rosa Maria Alonso-Salces:** Resources, Writing – review & editing. **Enrico Valli:** Resources, Project administration, Writing – review & editing. **Alessandra Bendini:** Resources, Project administration, Funding acquisition, Writing – review & editing. **Tullia Gallina Toschi:** Resources, Project administration, Funding acquisition, Writing – review & editing. **Stefania Vichi:** Conceptualization, Methodology, Resources, Supervision, Project administration, Funding acquisition, Writing – review & editing. **Alba Tres:** Methodology, Conceptualization, Resources, Supervision, Project administration, Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

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

Acknowledgments

The authors thank the sensory panels involved in the OLEUM project and the Catalan cooperatives that provided traceable virgin olive oil samples.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2022.132104.

References

Alonso-Salces, R. M., Segebarth, N., Garmón-Lobato, S., Holland, M. V., Moreno-Rojas, J. M., Fernández-Pierna, J. A., … Héberger, K. (2015). 1H-NMR and isotopic fingerprinting of olive oil and its unsaponifiable fraction: Geographical origin of virgin olive oils by pattern recognition. *European Journal of Lipid Science and Technology, 117*(12), 1991–2006. https://doi.org/10.1002/ejlt.v117.1210.1002/ejlt.201400243

Araghipour, N., Colineau, J., Koot, A., Akkermans, W., Rojas, J. M. M., Beauchamp, J., … Ruth, S. V. (2008). Analytical Methods Geographical origin classification of olive oils by PTR-MS. *Food Chemistry, 108*(1), 374–383. https://doi.org/10.1016/j.foodchem.2007.10.056

Bajoub, A., Medina-Rodríguez, S., Ajal, E. A., Cuadros-Rodríguez, L., Monasterio, R. P., Vercammen, J., … Carrasco-Pancorbo, A. (2018). A metabolic fingerprinting approach based on selected ion flow tube mass spectrometry (SIFT-MS) and chemometrics: A reliable tool for Mediterranean origin-labeled olive oils authentication. *Food Research International, 106*, 233–242. https://doi.org/10.1016/j.foodres.2017.12.027

Bakhouche, A., Lozano-Sánchez, J., Beltrán-Debón, R., Joven, J., Segura-Carretero, A., & Fernández-Gutiérrez, A. (2013). Phenolic characterization and geographical classification of commercial Arbequina extra-virgin olive oils produced in southern Catalonia. *Food Research International, 50*(1), 401–408. https://doi.org/10.1016/J.FOODRES.2012.11.001

Ben Hlima, H., Ben Ayed, R., Ennouri, K., & Smaoui, S. (2017). Geographical discrimination of virgin olive oils from the tunisian coasts by combining fatty acids and phenolic acids profiles within a multivariate analysis. *Journal of Oleo Science, 66*(9), 963–971. https://doi.org/10.5650/JOS.ESS17080

Ben Mohamed, M., Rocchetti, G., Montesano, D., Ben Ali, S., Guasmi, F., Grati-Kamoun, N., & Lucini, L. (2018). Discrimination of Tunisian and Italian extra-virgin olive oils according to their phenolic and sterolic fingerprints. *Food Research International, 106*, 920–927. https://doi.org/10.1016/J.FOODRES.2018.02.010

Bevilacqua, M., Bucci, R., Magrì, A. D., Magrì, A. L., & Marini, F. (2013). Data Fusion for Food Authentication. Combining near and Mid Infrared to Trace the Origin of Extra Virgin Olive Oils. *NIR News, 24*(2), 12–15. https://doi.org/10.1255/NIRN.1355

Bontempo, L., Paolini, M., Franceschi, P., Ziller, L., García-González, D. L., & Camin, F. (2019). Characterisation and attempted differentiation of European and extra-European olive oils using stable isotope ratio analysis. *Food Chemistry, 276*, 782–789. https://doi.org/10.1016/j.foodchem.2018.10.077

Borges, T. H., Pereira, J. A., Cabrera-Vique, C., Lara, L., Oliveira, A. F., & Seiquer, I. (2017). Characterization of Arbequina virgin olive oils produced in different regions of Brazil and Spain: Physicochemical properties, oxidative stability and fatty acid profile. *Food Chemistry, 215*, 454–462. https://doi.org/10.1016/J.FOODCHEM.2016.07.162

Carzedda, M., Gallenti, G., Troiano, S., Cosmina, M., Marangon, F., de Luca, P., … Nassivera, F. (2021). Consumer Preferences for Origin and Organic Attributes of Extra Virgin Olive Oil: A Choice Experiment in the Italian Market. *Foods, 10*(5), 994. https://doi.org/10.3390/foods10050994

Casadei, E., Valli, E., Panni, F., Donarski, J., Farrús Gubern, J., Lucci, P., … Gallina Toschi, T. (2021). Emerging trends in olive oil fraud and possible countermeasures. *Food Control, 124*, 107902. https://doi.org/10.1016/j.foodcont.2021.107902

Cecchi, L., Migliorini, M., Giambanelli, E., Rossetti, A., Cane, A., Mulinacci, N., & Melani, F. (2020). Authentication of the geographical origin of virgin olive oils from the main worldwide producing countries: A new combination of HS-SPME-GC-MS analysis of volatile compounds and chemometrics applied to 1217 samples. *Food Control, 112*, 107156. https://doi.org/10.1016/j.foodcont.2020.107156

Conte, L., Bendini, A., Valli, E., Lucci, P., Moret, S., Maquet, A., … Gallina Toschi, T. (2020). Olive oil quality and authenticity: A review of current EU legislation, standards, relevant methods of analyses, their drawbacks and recommendations for the future. *Trends in Food Science & Technology, 105*, 483–493. https://doi.org/10.1016/j.tifs.2019.02.025

Damak, F., Asano, M., Baba, K., Suda, A., Araoka, D., Wali, A., … Tamura, K. (2019). Interregional traceability of Tunisian olive oils to the provenance soil by multielemental fingerprinting and chemometrics. *Food Chemistry, 283*, 656–664. https://doi.org/10.1016/j.foodchem.2019.01.082

Damascelli, A., & Palmisano, F. (2013). Sesquiterpene Fingerprinting by Headspace SPME-GC-MS: Preliminary Study for a Simple and Powerful Analytical Tool for Traceability of Olive Oils. *Food Analytical Methods, 6*(3), 900–905. https://doi.org/10.1007/s12161-012-9500-9

European Union. (2021). 2020 Annual Report. The EU Agri-Food Fraud Network and the Administrative Assistance and Cooperation System. https://doi.org/10.2875/20163.

Forina, M., & Tiscornia, E. (1982). Pattern-recognition methods in the prediction of Italian olive oil origin by their fatty-acid content. *Annali Di Chimica, 3*(4), 143–155.

Hennessy, S., Downey, G., & O′Donnell, C. P. (2009). Confirmation of Food Origin Claims by Fourier Transform Infrared Spectroscopy and Chemometrics: Extra Virgin Olive Oil from Liguria. *Journal of Agricultural and Food Chemistry, 57*(5), 1735–1741. https://doi.org/10.1021/JF803714G

Kosma, I., Vatavali, K., Kontakos, S., Kontominas, M., Kiritsakis, A., & Badeka, A. (2017). Geographical Differentiation of Greek Extra Virgin Olive Oil from Late-Harvested Koroneiki Cultivar Fruits. *JAOCS, Journal of the American Oil Chemists' Society, 94*(11), 1373–1384. https://doi.org/10.1007/S11746-017-3036-5

Larsen, F. H., van den Berg, F., & Engelsen, S. B. (2006). An exploratory chemometric study of 1H NMR spectra of table wines. *Journal of Chemometrics, 20*(5), 198–208. https://doi.org/10.1002/(ISSN)1099-128X10.1002/cem.v20:510.1002/cem.991

Lia, F., Formosa, J. P., Zammit-Mangion, M., & Farrugia, C. (2020). The first identification of the uniqueness and authentication of Maltese extra virgin olive oil using 3D-fluorescence spectroscopy coupled with multi-way data analysis. *Foods, 9*(4), 498. https://doi.org/10.3390/foods9040498

Lukić, I., Carlin, S., Horvat, I., & Vrhovsek, U. (2019). Combined targeted and untargeted profiling of volatile aroma compounds with comprehensive two-dimensional gas chromatography for differentiation of virgin olive oils according to variety and geographical origin. *Food Chemistry, 270*, 403–414. https://doi.org/10.1016/J.FOODCHEM.2018.07.133

Magnusson, B., & Örnemark, U. (2014). *Eurachem guide: The Fitness for Purpose of Analytical Methods - A Laboratory Guide to Method Validation and Related Topics* (Eurachem, Ed.; 2nd ed.). www.eurachem.org.

Nielsen, N. P. V., Carstensen, J. M., & Smedsgaard, J. (1998). Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *Journal of Chromatography A, 805*(1–2), 17–35. https://doi.org/10.1016/S0021-9673(98)00021-1

Palagano, R., Valli, E., Cevoli, C., Bendini, A., & Gallina Toschi, T. (2020). *Compliance with EU vs. extra-EU labelled geographical provenance in virgin olive oils: A rapid untargeted chromatographic approach based on volatile compounds*. https://doi.org/10.1016/j.lwt.2020.109566.

Portarena, S., Baldacchini, C., & Brugnoli, E. (2017). Geographical discrimination of extra-virgin olive oils from the Italian coasts by combining stable isotope data and carotenoid content within a multivariate analysis. *Food Chemistry, 215*, 1–6. https://doi.org/10.1016/J.FOODCHEM.2016.07.135

Quintanilla-Casas, B., Bertin, S., Leik, K., Bustamante, J., Guardiola, F., Valli, E., … Vichi, S. (2020). Profiling versus fingerprinting analysis of sesquiterpene hydrocarbons for the geographical authentication of extra virgin olive oils. *Food Chemistry, 307*, 125556. https://doi.org/10.1016/j.foodchem.2019.125556

Quintanilla-Casas, B., Marin, M., Guardiola, F., García-González, D. L., Barbieri, S., Bendini, A., … Tres, A. (2020). Supporting the sensory panel to grade virgin olive oils: An in-house-validated screening tool by volatile fingerprinting and chemometrics. *Foods, 9*(10), 1509. https://doi.org/10.3390/foods9101509

Sánchez-López, E., Sánchez-Rodríguez, M. I., Marinas, A., Marinas, J. M., Urbano, F. J., Caridad, J. M., & Moalem, M. (2016). Chemometric study of Andalusian extra virgin olive oils Raman spectra: Qualitative and quantitative information. *Talanta, 156–157*, 180–190. https://doi.org/10.1016/J.TALANTA.2016.05.014

Schwolow, S., Gerhardt, N., Rohn, S., & Weller, P. (2019). Data fusion of GC-IMS data and FT-MIR spectra for the authentication of olive oils and honeys—is it worth to go the extra mile? *Analytical and Bioanalytical Chemistry, 411*(23), 6005–6019. https://doi.org/10.1007/S00216-019-01978-W

Souayah, F., Rodrigues, N., Veloso, A. C. A., Dias, L. G., Pereira, J. A., Oueslati, S., & Peres, A. M. (2017). Discrimination of Olive Oil by Cultivar, Geographical Origin and Quality Using Potentiometric Electronic Tongue Fingerprints. *Journal of the American Oil Chemists' Society, 94*(12), 1417–1429. https://doi.org/10.1007/S11746-017-3051-6

Tahir, H. E., Arslan, M., Komla Mahunu, G., Adam Mariod, A., Hashim, S. B. H., Xiaobo, Z., … Musa, T. H. (2022). The use of analytical techniques coupled with chemometrics for tracing the geographical origin of oils: A systematic review (2013–2020). *Food Chemistry, 366*, 130633. https://doi.org/10.1016/j.foodchem.2021.130633

Torres-Cobos, B., Quintanilla-Casas, B., Romero, A., Ninot, A., Alonso-Salces, R. M., Toschi, T. G., … Vichi, S. (2021). Varietal authentication of virgin olive oil: Proving the efficiency of sesquiterpene fingerprinting for Mediterranean Arbequina oils. *Food Control, 128*, 108200. https://doi.org/10.1016/j.foodcont.2021.108200

Valli, E., Bendini, A., Berardinelli, A., Ragni, L., Riccò, B., Grossi, M., & Gallina Toschi, T. (2016). Rapid and innovative instrumental approaches for quality and authenticity of olive oils. *European Journal of Lipid Science and Technology, 118*(11), 1601–1619. https://doi.org/10.1002/ejlt.v118.1110.1002/ejlt.201600065

Vera, D. N., Jiménez-Carvelo, A. M., Cuadros-Rodríguez, L., Ruisánchez, I., & Callao, M. P. (2019). *Authentication of the geographical origin of extra-virgin olive oil of the Arbequina cultivar by chromatographic fingerprinting and chemometrics*. https://doi.org/10.1016/j.talanta.2019.05.064.

Vichi, S., Guadayol, J. M., Caixach, J., López-Tamames, E., & Buxaderas, S. (2006). Monoterpene and sesquiterpene hydrocarbons of virgin olive oil by headspace solid-phase microextraction coupled to gas chromatography/mass spectrometry. *Journal of Chromatography A, 1125*(1), 117–123. https://doi.org/10.1016/j.chroma.2006.05.029

Vichi, S., Pizzale, L., & Conte, L. S. (2007). Stereospecific distribution of fatty acids in triacylglycerols of olive oils. *European Journal of Lipid Science and Technology, 109*(1), 72–78. https://doi.org/10.1002/ejlt.v109:110.1002/ejlt.200600199

Vichi, S., Tres, A., Quintanilla-Casas, B., Bustamante, J., & Guardiola, F. (2018). Sesquiterpene hydrocarbons, a promising tool for virgin olive oil geographical authentication. In M. Kontominas (Ed.), *Authentication and detection of adulteration of virgin olive oil*. Nova Science Publishers Inc.

Winkelmann, O., & Küchler, T. (2019). Reliable Classification of Olive Oil Origin Based on Minor Component Profile Using 1H-NMR and Multivariate Analysis. *European Journal of Lipid Science and Technology, 121*(12), 1900027. https://doi.org/10.1002/ejlt.201900027

Woodcock, T., Downey, G., & O'donnell, C. P. (2008). Confirmation of Declared Provenance of European Extra Virgin Olive Oil Samples by NIR Spectroscopy. *Journal of Agricultural and Food Chemistry, 56*, 11520–11525. https://doi.org/10.1021/jf802792d