Food Control 128 (2021) 108200

Contents lists available at ScienceDirect

Food Control

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

Varietal authentication of virgin olive oil: Proving the efficiency of sesquiterpene fingerprinting for Mediterranean Arbequina oils

Berta Torres-Cobos [a,b], Beatriz Quintanilla-Casas [a,b], Agustí Romero [c], Antonia Ninot [c], Rosa M. Alonso-Salces [d], Tullia Gallina Toschi [e], Alessandra Bendini [e], Francesc Guardiola [a,b], Alba Tres [a,b,*], Stefania Vichi [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] IRTA-Mas Bové Ctra, Reus-El Morell Km 3,8 43120-Constantí, Spain
[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 - Centro Interdipartimentale di Ricerca Industriale (CIRI Agroalimentare), Alma Mater Studiorum - Università di Bologna, Bologna, Italy

ARTICLE INFO

Keywords:
Virgin olive oil
Cultivar
Sesquiterpene
Fingerprint
HS-SPME-GC-MS
PLS-DA

ABSTRACT

Virgin olive oil (VOO) is a highly appreciated product fundamental in the Mediterranean diet. Since its sensory attributes are greatly influenced by the olive cultivar, the varietal authentication of VOOs is needed to protect consumers from misleading information. The present study aims to evaluate the suitability of sesquiterpene hydrocarbon (SH) fingerprint as VOO cultivar marker beyond geographical, agronomical and processing conditions. The study was mainly focused on Mediterranean Arbequina oils. SH profile of more than 400 VOOs from 6 counties and 38 different cultivars and coupages was analysed by Headspace Solid Phase Microextraction-Gas Chromatography-Mass Spectrometry (HS-SPME-GC-MS). Partial Least Square-Discriminant Analysis (PLS-DA) classification models were built with the aligned chromatograms. A binary PLS-DA model was built to distinguish 'Arbequina' oils from those of other cultivars (non-'Arbequina' class) and it was externally validated. The results of the external validation showed a 95.1% of overall correct classification confirming the suitability of SH fingerprint as a screening method for the authentication of Arbequina VOO. Also, the discrimination capacity of SH fingerprinting to authenticate VOOs from other cultivars was preliminary explored and promising results were obtained.

1. Introduction

Qualitative characteristics of virgin olive oil (VOO) are not only determined by the processing and storage conditions but also by the cultivar and geographical origin (Dias et al., 2014; Montealegre et al., 2010). Hence, oils within the same commercial category can present very different compositional and sensory characteristics depending on the olive cultivar. Monovarietal VOOs that include information about the olive cultivar on the label, as well as VOOs included into a given Protected Designation of Origin which are required to use specific traditional olive cultivars (Council Regulation (EC) 510/2006), increase consumers' perceived quality and lead them to pay a higher price (Cabrera et al., 2015; Cicerale et al., 2016). Therefore, verifying the label-declared cultivar in VOO has become relevant to protect consumers from misleading information (Bajoub et al., 2018) and, currently, it can only be achieved by auditing traceability documents.

The varietal characterization of VOOs has been widely studied by addressing several major and minor compounds and by applying multiple analytical techniques and chemometric approaches (Montealegre et al., 2010; Aparicio et al., 2013; Bajoub et al., 2018). However, except DNA based methods (Agrimonti et al., 2011), which are costly to be used for routine analysis (Bajoub et al., 2018), reliable markers for VOO cultivar authentication are still unavailable. This is largely due to the

* Corresponding author. 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: atres@ub.edu (A. Tres).

https://doi.org/10.1016/j.foodcont.2021.108200
Received 29 November 2020; Received in revised form 22 April 2021; Accepted 23 April 2021
Available online 28 April 2021

Abbreviations

COW	Correlation Optimized Warping
EIC	Extracted Ion Chromatogram
LV	Latent Variables
PC	Principal Component
PCA	Principal Component Analysis
PLS-DA	Partial Least Discriminant Analysis
RMSEcv	Root Mean Squared Error of Cross Validation
SEcv	Standard Error of Cross-Validation
SH	sesquiterpene hydrocarbon
SIM	Selected Ion Monitoring
SPME-GC-MS	Solid Phase Microextraction coupled to Gas Chromatography-Mass Spectrometry
VOO	Virgin olive oil

high concomitant influence of other factors on VOO composition, such as the degree of maturation, phytosanitary status, harvest year or processing and storage conditions (García-González & Aparicio, 2010; Montealegre et al., 2010; Vichi et al., 2018). In this regard, most of the available studies on VOO cultivar authentication are addressed to monovarietal VOOs produced within a limited geographical area (Osorio-Bueno et al., 2005; Papadia et al., 2011; Piccinonna et al., 2016; Sacco et al., 2000; Sayago et al., 2019). They have been fundamental to demonstrate the genetic influence on VOO composition with other factors being similar, but they do not take into account the effect of technological and, in particular, environmental conditions beyond the olive cultivar itself. This is not fully representative for the actual VOO production, in which the same olive cultivar may be cultivated in different regions or countries (Tous, 2017). Marini et al. (2004) developed authentication models relying on a representative number of samples ($n > 500$) from cultivars produced in different regions of Southern Italy, achieving a satisfactory classification by applying supervised chemometric methods to several official quality and purity parameters. However, the need for multiple analytical techniques or time-expensive procedures hinders its application for the screening of large number of samples, and moreover, its suitability should be confirmed also for VOOs from other major production areas. Hence, there is a need for screening tools capable of authenticating VOOs according to their cultivar beyond geographical and technological factors.

Previous research has evidenced that the presence of sesquiterpene hydrocarbons (SHs) in VOO is highly dependent the olive cultivar and the growing area (Bortolomeazzi et al., 2001; Damascelli & Palmisano, 2013; Quintanilla-Casas et al., 2020; Vichi et al., 2006, 2010, 2018) while it is barely affected by technological factors such as olive post-harvest processing and oil storage conditions (Vichi et al., 2018). Specifically, the effect of genetic factors on VOO SHs was demonstrated when oils from different cultivars, produced in the same geographical region, presented significant differences in the SH composition (Guinda et al., 1996; Osorio-Bueno et al., 2005; Vichi et al., 2010).

Also recently, a fingerprinting approach has been applied to SH chromatograms obtained by SPME-GC-MS (Quintanilla-Casas et al., 2020). In a fingerprinting approach, chemometric tools are applied to highly dimensional analytical data, such as a chromatogram, to find specific patterns of a certain quality characteristic that are known as fingerprints, not requiring peak identification or quantitation (Ballin & Laursen, 2019; Berrueta et al., 2007; Bosque-Sendra et al., 2012; Quintanilla-Casas et al., 2020). Compared to the target approach, chromatographic fingerprinting considers more information as the whole analytical signal is used, and overcomes drawbacks related to the SH identification or quantitation as they are a wide category of compounds with high structural diversity but very similar mass spectra (Degenhardt et al., 2009; Quintanilla-Casas et al., 2020). This was demonstrated in our previous study, in which Partial Least Discriminant Analysis (PLS-DA) was able to find features in the SH fingerprint that were common between samples from the same region even if they belonged to different cultivars, and thus, models to authenticate VOO geographical origin could be developed by selecting the country of origin as the grouping variable to supervise the PLS-DA (Quintanilla-Casas et al., 2020). But, since the SH composition in olive fruit is known to be driven by both genetic and environmental factors, our hypothesis was that if the cultivar type was selected as the variable to supervise the PLS-DA analysis, the PLS-DA model would find different features on the SH fingerprint of VOO that would be characteristic of the cultivar.

On this basis, the aim of the present work was to assess the suitability of SH fingerprint as VOO cultivar marker beyond geographical, agronomical and processing conditions. For this purpose, we developed and validated a varietal authentication model based on the SH fingerprint obtained by SPME-GC–MS, particularly focusing on the discrimination of 'Arbequina' VOOs from the rest of VOOs, using a sample set of more than 400 VOOs from different cultivars produced under real processing conditions in different harvest seasons and in various EU and non-EU countries and regions. Moreover, to explore whether analogous models could also be developed to distinguish VOOs from other cultivars, the discrimination capacity of SH fingerprinting was preliminarily assessed for the seven main cultivars included in the sample set.

2. Material and methods

2.1. Sampling

The sample set was composed by 404 traceable VOOs and EVOOs from different countries and geographical regions (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), and under the surveys implemented by the Institut de Recerca i Tecnologia Agroalimentària (IRTA). Of these samples, 178 were from 'Arbequina' cultivar and 226 were monovarietal oils from 37 other cultivars ($n = 144$) and coupages (which are blends of different cultivars) that did not contain 'Arbequina' oil ($n = 82$). The VOO and EVOO samples pertained to virgin and extra virgin olive oil categories according to the European Commission regulation (ECC) No 2591/91 of July 11, 1991 and its amendments and were produced at real industrial conditions during 4 different campaigns (harvests from 2015/16 to 2018/19). More information about VOO and EVOO samples is available in Table S1 (Supplementary material). Samples were stored under N_2 atmosphere at −20 °C until analysis. The full sample set was analysed in four main batches throughout 2017–2019.

2.2. Headspace-solid phase microextraction (HS-SPME)

The SH fingerprint of VOO samples was analysed using a Combi-pal autosampler (CTC Analytics, Zwingen, Switzerland) at the conditions reported by Vichi et al. (2006). An aliquot of 2 g of oil was weighed into

Table 1
Number and geographical origin of VOO and EVOO samples from 'Arbequina' and non-'Arbequina' cultivars.

Origin		'Arbequina' (n)	Other cultivars[a] (n)
Argentina		4	37
Chile		1	1
Italy		1	35
Morocco		9	29
Portugal		3	13
Spain	Catalonia	155	93
	Andalusia	2	15
	Other regions	3	3
total samples		178	226

[a] 37 different cultivars plus 20 coupages.

a 10 mL vial fitted with a PTFE/silicone septum and kept at 70 °C under constant agitation. After 10 min of sample conditioning, a divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS) fiber (2 cm length, 50/30 μm film thickness) provided by Supelco (Bellefonte, PA) was exposed to the sample headspace for 60 min. Then, it was desorbed in the gas chromatograph injection port at 260 °C for 10 min. During the desorption step, the injector was maintained in split-less mode during 5 min.

2.3. Gas chromatography-mass spectrometry (GC-MS)

The SHs fingerprint was acquired by an Agilent 6890N Network GC system coupled to a quadrupolar mass selective analyser Agilent 5975C Inert MSD (Agilent Technologies, Santa Clara, California, USA) using helium as carrier gas, at a flow of 1.5 mL/min. Analytes were separated on a Supelcowax-10 capillary column (60 m × 0.25 mm i.d., 0.25 μm film thickness) (Supelco, Bellefonte, PA). Column temperature was held at 40 °C for 3 min, increased to 100 °C at 4 °C/min, then to 200 °C at 5 °C/min and to 260 °C at 15 °C/min, holding the last temperature for 5 min. The temperatures of the ion source and the transfer line were 230 and 280 °C, respectively. Mass spectra were recorded at 5.1 scan/s and the electron energy was 70 eV. Acquisition was performed in the selected ion monitoring (SIM) mode, by analysing the Extracted Ion Chromatogram (EIC) of each of the following SH specific ions: *m/z* 93, 119, 157, 159, 161, 189 and 204, which had been reported to be specific for SHs (Vichi et al., 2010). Therefore, only the chromatographic data belonging to SH compounds was studied.

2.4. Fingerprinting approach

The intensities of scans between minutes 21 and 42 (3197 scans) were considered for each ion (3197 scans × 7 ions = 22379 variables per sample). A data matrix was built for each ion, with scans' intensities of each EIC (columns) from each sample (rows). For each selected ion, the EICs of the 404 samples were aligned by Correlation Optimized Warping (COW) algorithm in Matlab® (Nielsen et al., 1998). Then, the 7 matrices of the aligned chromatograms were concatenated conforming a two-way unfolded matrix (404 samples × 22379 variables).

2.5. Chemometrics

2.5.1. Data pre-processing and exploration

The pre-processing of the aligned data matrix was performed with SIMCA software v13.0© (Umetrics AB, Sweden). Multiple pre-processing treatments were tested (mean centring, scaling to unit variance, log10, derivatives) until finding the optimal one for each data set. A Principal Component Analysis (PCA) was performed for the exploration of data (n = 404) and to identify potential outliers (according to Hotelling's T^2 range and distance to the model parameters).

2.5.2. Partial least squares discriminant analysis (PLS-DA): Arbequina vs non-Arbequina oils

A binary PLS-DA model was built after applying a first derivative and log 10 pre-processing to classify the 404 samples into the 'Arbequina' and the non-'Arbequina' classes, the latter including coupages and monovarietal oils from other cultivars,.

The model was internally validated through leave 10%-out cross-validation. The optimal number of Latent Variables (LV) of the PLS-DA model were selected according to the lowest Root Mean Squared Error of Cross Validation (RMSEcv) criteria. To assess the model over-fitting, permutation test and ANOVA on the cross-validated predictive residuals (p-value) were carried out. The Q^2 values and the percentage of correct classifications were assessed to evaluate the suitability of each PLS-DA model.

2.5.3. External validation

The binary PLS-DA model was then externally validated by predicting the class of samples that had not been used to develop the model. For this, the full data set (n = 404) was randomly split into a training set (80% of the sample set, n = 323) and a validation set (20% of the sample set, n = 81), maintaining a balance in the proportions of 'Arbequina' and non-'Arbequina' samples and of samples from different analytical batches and geographical origins (Table S2, Supplementary material). This was carried out seven times, obtaining seven different training and validation sets. The efficiency of the classification was assessed as mean percentage of correct classification.

2.5.4. Evaluation of PLS-DA regression coefficients

The regression coefficients of the binary PLS-DA model developed with the full sample set were evaluated to explore the contribution of the variables obtained by each *m/z* ion. Regression coefficients were considered as significant when a jack-knife standard error of cross-validation (SEcv) was lower than the given coefficient value.

2.5.5. Multi-class PLS-DA model for 7 cultivars

Once the Arbequina vs non-Arbequina model was found suitable, it was explored whether SHs fingerprinting would be suitable to develop models to verify the identity of other cultivars. Thus, a multi-class PLS-DA model was developed as a preliminary model (after autoscaling, log10 and first derivative pre-processing) to evaluate the discrimination of the 7 cultivars included in the sample set that were represented by at least 10 samples (n = 256). The model was fitted and internally validated as described above.

3. Results and discussion

3.1. Data pre-processing and exploratory analysis

Seven ions (*m/z* 93, 119, 157, 159, 161, 189 and 204) were selected according to previous studies that indicated them to be specific of SH (Vichi et al., 2010). The data points of the EIC within the interval of elution of SH (21–42 min) were used as variables (22379 variables), following the fingerprinting approach applied by Quintanilla-Casas et al. (2020).

To solve the retention time shifting between samples, the EICs were aligned by Correlation Optimized Warping (COW) algorithm in Matlab®. This alignment algorithm was selected since it was specifically designed for chromatographic data. The COW method aligns the chromatographic profiles by piecewise linear stretching and compression, also known as warping, of the time axis of one of the profiles (Nielsen et al., 1998). Once aligned, the matrix obtained (22379 variables and 404 samples) was imported to SIMCA software v13.0c (Umetrics AB, Sweden) to develop and optimize the classification models.

After data pre-processing, a PCA was performed to explore the data and to detect potential outliers (3 Principal Components (PCs) accounted for 80.7% of the total variance explained). According to the Hotelling's T^2 range and distance to the model parameters, no outliers were detected. A first examination of the PCA score plot revealed that even under a non-supervised analysis, VOOs naturally clustered according to their cultivar. In fact, even if there was some overlap, 'Arbequina' samples tended to shape into a differentiated group at the upper part of the plot, while non-'Arbequina' samples located at the lower part (Fig. 1a), evidencing that the variability linked to the cultivar of origin was mainly explained by PC3 (2.0% of total explained variance). Certain clustering was also observed for other monovarietal VOOs (Fig. 1b), endorsing the behaviour observed for 'Arbequina' and non-'Arbequina' samples. To exclude that this natural clustering was only induced by the geographical origin of each cultivar, the score plot was coloured by sample's provenance (Fig. 1c). In this way, it revealed that samples from cultivars grown in different regions tended to group by cultivar rather than by geographic origin, leading to unclear clusters by country. This

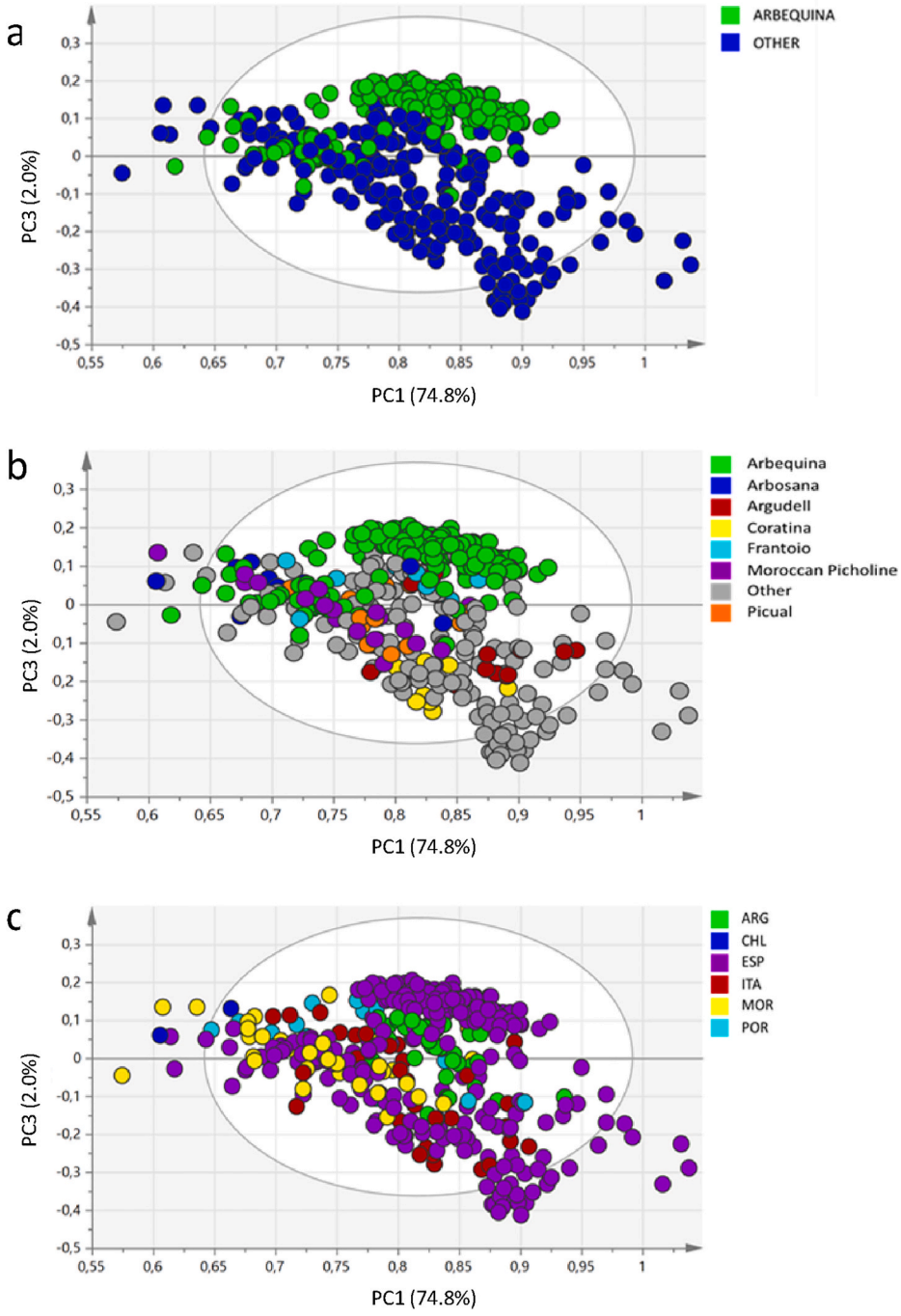

Fig. 1. First and third principal components of the PCA model (n = 404 samples, first derivative and log 10 pre-processing; variance explained by each factor is shown between parentheses), based on VOO sesquiterpene data, coloured by a) 'Arbequina'/non-'Arbequina' cultivars; b) Olive cultivar; c) Country of origin. (ARG: Argentina, CHL: Chile, ESP: Spain, ITA: Italy, MOR: Morocco and POR: Portugal).

suggested that the cultivar had a higher effect on the SH fingerprint than the geographical origin.

3.2. PLS-DA authentication models and internal validation

A binary PLS-DA classification model was built with the SH fingerprint of 404 samples to distinguish 'Arbequina' oils from those of other cultivars (non-'Arbequina' class). The model with 5 LV was internally validated through leave 10%-out cross-validation, resulting in a RMSEcv of 0.23 and a 99.8% of correct classification with all samples assigned to the correct class, except one (Table 2). Permutation tests, which display the prediction capacity of 20 random models, and ANOVA results ($p <$

Table 2
Results of the leave 10%-out cross-validation of the 'Arbequina' *vs* non-'Arbequina' PLS-DA classification model.

	N	'Arbequina'	Other	Correct class (%)
'Arbequina'	178	178	0	100
non-'Arbequina'	226	1	225	99.6
Total	404			**99.8**

N = 404, 5 LVs, $Q^2 = 0.82$, RMSEcv = 0.23, ANOVA p-value <0.05.

0.05) showed that the model had a high discrimination capacity and was not over-fitted. The results were very promising, especially considering that the 'Arbequina' class included samples from various geographical origins, and that the non-'Arbequina' class included a high number of other monovarietal and coupage oils.

In view of the encouraging results provided by the binary model based on the extensive 'Arbequina' and non-'Arbequina' sampling, and since the PCA exploratory analysis already revealed certain natural separation between other cultivars, we investigated whether analogous models could potentially be developed to authenticate VOOs from other cultivars. As a proof of concept for this scope, we performed a preliminary multi-class PLS-DA model to discriminate between the seven cultivars of the sample set that were represented by at least 10 samples. Only monovarietal samples were included in this PLS-DA model (n = 256). The results of the leave 10%-out cross-validation were promising (Table 3), displaying high percentages of correct classification (generally above 90%) for most of the cultivars and achieving a global 94.9% of correct classification. In addition, the ANOVA ($p < 0.05$) and the permutation test (Q^2 values of permuted models < 0) indicated the absence of a random classification and of model overfitting. The lowest percentage of correct classification was observed for the 'Frantoio' set (69.2%), probably due to the very different geographical origin of its samples (Italy, n = 3 and Argentina, n = 9). The low number of samples of this preliminary submodel might have not been enough to compensate such a source of confusion, but we can hypothesize that with a suitable sampling the authentication of this cultivar would not be precluded. Overall, based on these preliminary results, we could infer that developing future models to authenticate VOOs from any cultivar based on the SH fingerprint would be possible, providing that appropriate sampling was available.

3.3. External validation

Disposing of meaningful results is crucial for the relevant implementation of any authentication tool. In this case, verifying the real predictive ability of the developed model is necessary to exclude possible over-optimistic results and to prove the suitability of the method to assess VOO varietal origin. To confirm the reliability of the predictions obtained by internal validation, we carried out an external validation in which models were applied to samples that had not been included in their development to classify them as 'Arbequina' and non-'Arbequina' VOOs. For this, we randomly split the sample set into a training set (n = 323) and a validation set (n = 81). A PLS-DA model was developed with the training set and cross-validated by leave-10%-out, and it was then applied to predict the class of the 81 samples conforming the validation set. To increase the robustness of the validation, and to minimize the effect of the sample sets' composition, this process was run seven times and the results were expressed as mean values of the seven sets of external validation. The internal validation results of PLS-DA models built with the training sets were in agreement with those of the model developed with the global sample set (mean overall correct classification of 100%). The results of the external validation, summarized in Table 4, were extremely satisfactory. On average, 93.5% of 'Arbequina' samples and 96.4% of non-'Arbequina' samples were correctly classified, resulting in a 95.1% of overall correct classification.

On closer inspection, the external validation revealed that the model was not able to correctly classify 'Arbequina' oils produced in Argentina. Unfortunately, due to the reduced number of 'Arbequina' VOOs from Argentina, only 3 samples could be randomly included in each training set and only one in each validation set, while other non-'Arbequina' oils from Argentina were much more represented in both sets. This probably contributed to the fact that each Argentinian 'Arbequina' sample, randomly selected in each validation set, was incorrectly classified (Table 5). This seems to support our previous hypothesis formulated for the internal validation results obtained for the 'Frantoio' cultivar by the preliminary multi-class PLS-DA model. This outcome could be due to the extreme compositional differences reported between VOOs from the southern hemisphere and those from the Mediterranean area. Romero (2017) reported that 'Arbequina' VOO from Argentina and Australia has a very different fatty acid composition from other regions and discussed about the high temperature effect during the maturation season. In fact, the southern hemisphere 'Arbequina' oils may present compositions whose differences may even be outside of some of the limits set by the current trade standards (Aparicio et al., 2013). Apparently, for these samples, the geographical origin outperformed the cultivar differences (Rondanini et al., 2011; Torres et al., 2009), possibly leading to a better matching of the SH pattern of Argentinian 'Arbequina' oils with that of other Argentinian oils rather than with that of the rest of 'Arbequina' Mediterranean samples, even if the cultivar was the classification variable. Since only a scarce number of Argentinian 'Arbequina' samples was available (n = 4), the results are not conclusive in this aspect. Varietal misclassification of samples from non-Mediterranean region could be presumably resolved by widening the sampling with a representative number of VOOs from the region of interest. By this, it would be possible to determine if the model would be able to find proper common SH traits between Mediterranean and Argentinian 'Arbequina' oils, or if a specific model 'Arbequina' vs non-'Arbequina' would be required for Argentinian oils.

Even so, samples from different EU and non-EU Mediterranean

Table 4
Results of external validation of the 'Arbequina' *vs* non-'Arbequina' PLS-DA models. Results are mean values obtained from seven randomly selected validation sets.

	n	Arbequina	Other	Correct class (%)
'Arbequina'	36	34 ± 1	2 ± 2	93.5 ± 3.8
non-'Arbequina'	45	2 ± 1	43 ± 2	96.4 ± 1.3
Total	81			95.1 ± 2.4

N = 404, 5 LVs, $Q^2 > 0.80$, RMSEcv < 0.27, ANOVA p-value <0.05.

Table 3
Results of leave 10%-out cross-validation of the multi-class PLS-DA model for the classification of main cultivars in the sample set.

	N	'Arbequina'	'Picual'	'Arbosana'	'Moroccan Picholine'	'Coratina'	'Frantoio'	'Argudell'	No class[a]	Correct class (%)	RMSEcv
'Arbequina'	178	173	0	0	1	0	0	0	4	97.2	0.30
'Picual'	12	0	12	0	0	0	0	0	0	100	0.19
'Arbosana'	11	0	0	10	0	0	0	0	1	90.9	0.16
'Moroccan Picholine'	20	1	0	0	19	0	0	0	0	95.0	0.19
'Coratina'	10	0	0	0	0	10	0	0	0	100	0.15
'Frantoio'	13	0	0	0	0	0	9	0	4	69.2	0.20
'Argudell'	12	2	0	0	0	0	0	10	0	83.3	0.17
Total	256	176	12	10	20	10	9	10	9	**94.9**	

N = 256, 7 LVs, $Q^2 = 0.47$, ANOVA p-value <0.05.
[a] YPred<0.5.

Table 5
Number, country and cultivar of origin of samples misclassified in the external validation of the binary PLS-DA model, for each validation set. The number of misclassified samples is reported with respect to the total number of samples from the same country and class ('Arbequina' and non-'Arbequina') that were included in the validation set.

	Validation set	Number of misclassified samples	Country of origin of the misclassified samples	Cultivar of the misclassified samples
'Arbequina' samples classified as non-'Arbequina'	1	1/1	Argentina	'Arbequina'
	2	1/1	Argentina	'Arbequina'
		1/32	Spain	'Arbequina'
	3	1/1	Argentina	'Arbequina'
	4	1/1	Argentina	'Arbequina'
		1/32	Spain	'Arbequina'
	5	1/1	Argentina	'Arbequina'
		1/32	Spain	'Arbequina'
	6	1/1	Argentina	'Arbequina'
		1/32	Spain	'Arbequina'
	7	1/1	Argentina	'Arbequina'
Non-'Arbequina' samples classified as 'Arbequina'	1	1/22	Spain	'Empeltre'
	2	1/7	Italy	'Casaliva'
		1/22	Spain	'Empeltre'
	3	1/3	Portugal	Coupage
		1/22	Spain	'Empeltre'
	4	2/22	Spain	'Empeltre'
	5	1/7	Italy	'Casaliva'
		1/7	Italy	Coupage
		3/22	Spain	'Empeltre'
	6	1/6	Morocco	'Moroccan Picholine'
		3/22	Spain	'Empeltre'
	7	1/7	Italy	Coupage

regions were correctly assigned to their corresponding varietal class, even when their representativeness in the training set was low, such is the case of the Portuguese and Moroccan 'Arbequina' oils (Table 1). These results evidenced the suitability of the model for Mediterranean VOO cultivar authentication regardless of the oil's geographical origin. Although the present sampling was specially focused on Spanish 'Arbequina' oils, the results of the external validation proved the efficiency of the authentication model to distinguish 'Arbequina' oils produced also in the rest of the Mediterranean basin. To improve the predictive ability of the model for 'Arbequina' samples coming from other specific geographical origins, new samples from those regions should be included in the training set.

3.4. Exploration of PLS-DA regression coefficients

To study the variables that contributed the most to the discrimination between 'Arbequina' and other cultivar VOOs, we examined the significant regression coefficients of the 'Arbequina' vs non-'Arbequina' PLS-DA model. More than 200 variables resulted relevant to the model according to the significance of their regression coefficients (as defined according to their value and standard error), highlighting the key advantage of high-dimensional data approaches, such as fingerprinting, over other conventional approaches such as the target or even multi-target ones (Quintanilla-Casas et al., 2020).

The plotting of regression coefficients against the variables of the unfolded matrix (Fig. 2a) revealed that each EIC provided relevant variables to the model throughout different regions of the chromatogram. In particular, EICs of *m/z* 93, 119 and 204 provided the highest number of variables important for the discrimination of both the 'Arbequina' and the non-'Arbequina' classes. The same figure revealed that significant regression coefficients corresponded to sections of the unfolded matrix presenting either major and very minor variables. As an illustration of this, amplifying a section of the EIC of *m/z* 93 (34–38 min) (Fig. 2b), revealed that some of the highest regression coefficients corresponded to minor SHs or not well-resolved chromatographic peaks that could only be studied through a fingerprinting approach. This is in agreement with our previous findings dealing with the application of SH fingerprinting for VOO geographical authentication (Quintanilla-Casas et al., 2020).

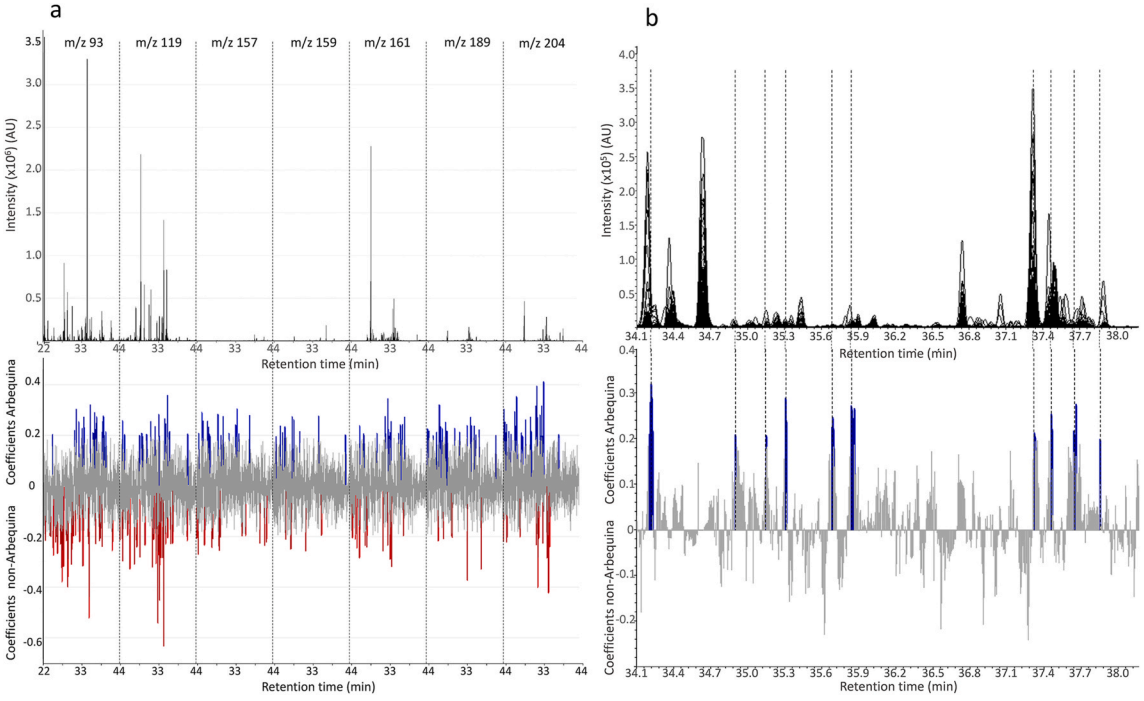

Fig. 2. PLS-DA regression coefficients of the Arbequina *vs* non-'Arbequina' model against the variables of the unfolded matrix. 'Arbequina' positive and negative relevant coefficients (selected from the significant ones according to an arbitrary threshold of 0.2) are highlighted in blue or in red, respectively. a) The sections of the unfolded matrix corresponding to each Extracted Ion Chromatogram are marked. b) Section of the ion *m/z* 93 chromatogram (34.059–38.126 min) plotted against its corresponding regression coefficients. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

4. Conclusions

In conclusion, SH fingerprint proved to be a suitable screening method for the varietal authentication of VOO, enabling an extremely satisfactory efficiency in the classification of 'Arbequina' and non-'Arbequina' oils produced under real industrial conditions in different regions of the Mediterranean basin, as assessed by external validation (95.1% of correct classification). Combining SH fingerprinting with PLS-DA allowed recognizing characteristic patterns relevant for each cultivar class minimizing those variables related with other factors such as the geographical origin. This confirmed our hypothesis that genetic and environmental factors exert distinct effects on the particularly complex SH fingerprint, which may provide suitable markers of both VOO geographical (Quintanilla-Casas et al., 2020) and varietal origin (as evidenced in this study), which would be revealed by PLS-DA depending on the variable selected for supervising the pattern recognition analysis. The VOOs produced in the Southern Hemisphere were not satisfactorily classified by the PLS-DA model built using mainly Mediterranean samples, evidencing the need to include more samples from this region to improve the predictive ability for 'Arbequina' oils from this geographical origin.

Furthermore, a preliminary multi-class PLS-DA model to discriminate between the other cultivars represented in the sample set resulted in a 94.9% of overall correct classification by leave 10%-out cross-validation, suggesting that successful classification models analogous to that developed and validated for 'Arbequina' samples could be potentially developed to authenticate other VOO cultivars.

Finally, the exploration of PLS-DA regression coefficients revealed that a high number of variables contributed to the discrimination model, several of which corresponded to minor SHs or not well-resolved chromatographic peaks that could only be studied through a fingerprinting approach. This confirmed the advantage of high-dimensional non-targeted data approaches like fingerprinting over other conventional approaches.

CRediT authorship contribution statement

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

Declaration of competing interest

None.

Acknowledgments

Authors are grateful to Catalan Cooperatives for providing traceable VOO samples in the frame of the Autenfood project.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodcont.2021.108200.

Funding

This work was developed in the context of the project AUTENFOOD, supported by ACCIÓ-Generalitat de Catalunya (Spain) and the European Union through the Programa Operatiu FEDER Catalunya 2014–2020 [grant number COMRDI-15-1-0035] and the project OLEUM, supported by the European Commission within the Horizon 2020 Program [grant agreement no. 635690]. This work was also supported by Generalitat de Catalunya (Spain) [grant number 2020FI_B00595]; and the Spanish Ministry of Science, Innovation and Universities and the FEDER Program [grant numbers FPU16/01744, RYC-2017-23601]. These funding sources had no involvement in the study design, collection, analysis or interpretation of data; or in the writing of the manuscript, or in the decision to submit the manuscript for publication. 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.

References

Agrimonti, C., Vietina, M., Pafundo, S., & Marmiroli, N. (2011). The use of food genomics to ensure the traceability of olive oil. *Trends in Food Science & Technology, 22*, 237–244. https://doi.org/10.1016/j.tifs.2011.02.002

Aparicio, R., Morales, M. T., Aparicio-Ruíz, R., Tena, N. D., & García-González, D. L. (2013). Authenticity of olive oil: Mapping and comparing official methods and promising alternatives. *Food Research International, 54*, 2025–2038. https://doi.org/10.1016/j.foodres.2013.07.039

Bajoub, A., Bendini, A., Fernández-Gutiérrez, A., & Carrasco-Pancorbo, A. (2018). Olive oil authentication: A comparative analysis of regulatory frameworks with especial emphasis on quality and authenticity indices, and recent analytical techniques developed for their assessment. A review. *Critical Reviews in Food Science and Nutrition, 58*, 832–857. https://doi.org/10.1080/10408398.2016.1225666

Ballin, N. Z., & Laursen, K. H. (2019). To target or not to target? Definitions and nomenclature for targeted versus non-targeted analytical food authentication. *Trends in Food Science & Technology, 86*, 537–543. https://doi.org/10.1016/j.tifs.2018.09.025

Berrueta, L. A., Alonso-Salces, R. M., & Héberger, K. (2007). Supervised pattern recognition in food analysis. *Journal of Chromatography A, 1158*, 196–214. https://doi.org/10.1016/j.chroma.2007.05.024

Bortolomeazzi, R., Berno, P., Pizzale, L., & Conte, L. (2001). Sesquiterpene, alkene, and alkane hydrocarbons in virgin olive oils of different varieties and geographical origins. *Journal of Agricultural and Food Chemistry, 49*, 3278–3283. https://doi.org/10.1021/jf001271w

Bosque-Sendra, J. M., Cuadros-Rodríguez, L., Ruiz-Samblás, C., & de la Mata, P. (2012). Combining chromatography and chemometrics for the characterization and authentication of fats and oils from triacylglycerol compositional data–a review. *Analytica Chimica Acta, 724*, 1–11. https://doi.org/10.1016/j.aca.2012.02.041

Cabrera, E. R., Arriaza, M., & Rodríguez-Entrena, M. (2015). Is the extra virgin olive oil market facing a process of differentiation? A hedonic approach to disentangle the effect of quality attributes. *Grasas Y Aceites, 66*, 4. https://doi.org/10.3989/gya.0253151

Cicerale, S., Liem, G., & Keast, R. S. J. (2016). Consumer perception, attitudes, liking and preferences for olive oil. In D. Boskou, & M. Clodoveo (Eds.), *Products from olive tree*. London, UK: IntechOpen. https://doi.org/10.5772/64554.

Council Regulation (EC). No 510/2006 of 20 March 2006 on the protection of geographical indications and designations of origin for agricultural products and foodstuffs. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32006R0510.

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*, 900–905. https://doi.org/10.1007/s12161-012-9500-9

Degenhardt, J., Köllner, T. G., & Gershenzon, J. (2009). Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry, 70*, 1621–1637. https://doi.org/10.1016/j.phytochem.2009.07.030

Dias, L. G., Fernandes, A., Veloso, A. C. A., Machado, A. A. S. C., Pereira, J. A., & Peres, A. M. (2014). Single-cultivar extra virgin olive oil classification using a potentiometric electronic tongue. *Food Chemistry, 160*, 321–329. https://doi.org/10.1016/j.foodchem.2014.03.072

García-González, D. L., & Aparicio, R. (2010). Research in olive oil: Challenges for the near future. *Journal of Agricultural and Food Chemistry, 24*, 12569–12577. https://doi.org/10.1021/jf102735n

Guinda, Á., Lanzón, A., & Albi, T. (1996). Differences in hydrocarbons of virgin olive oils obtained from several olive varieties. *Journal of Agricultural and Food Chemistry, 44*, 1723–1726. https://doi.org/10.1021/jf950571o

Marini, F., Balestrieri, F., Bucci, R., Magri, A. L., & Marini, D. (2004). Supervised pattern recognition to authenticate Italian olive oil varieties. *Chemometrics and Intelligent Laboratory Systems, 73*, 85–93. https://doi.org/10.1016/j.chemolab.2003.12.007

Montealegre, C., Alegre, M. L. M., & García-Ruíz, C. (2010). Traceability markers to the botanical origin in olive oils. *Journal of Agricultural and Food Chemistry, 58*, 28–38. https://doi.org/10.1021/jf902619z

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*, 17–35. https://doi.org/10.1016/S0021-9673(98)00021-1

Osorio-Bueno, E., Sánchez-Casas, J., Montaño-García, A., & Gallardo-González, L. (2005). Discriminating power of the hydrocarbon content from virgin olive oil of Extremadura cultivars. *Journal of the American Oil Chemists' Society, 82*, 1–6. https://doi.org/10.1007/s11746-005-1034-0

Papadia, P., Del Coco, L., Muzzalupo, I., Rizzi, M., Perri, E., Cesari, G., Simeone, V., Mondelli, D., Schena, F. P., & Fanizzi, F. P. (2011). Multivariate analysis of 1H-NMR spectra of genetically characterized extra virgin olive oils and growth soil correlations. *Journal of the American Oil Chemists' Society, 88*, 1463–1475. https://doi.org/10.1007/s11746-011-1812-1

Piccinonna, S., Ragone, R., Stocchero, M., Del Coco, L., De Pascali, S. A., Schena, F. P., & Fanizzi, F. P. (2016). Robustness of NMR-based metabolomics to generate comparable data sets for olive oil cultivar classification. An inter-laboratory study on Apulian olive oils. *Food Chemistry, 199*, 675–683. https://doi.org/10.1016/j.foodchem.2015.12.064

Quintanilla-Casas, B., Bertin, S., Leik, K., Bustamante, J., Guardiola, F., Valli, E., Bendini, A., Gallina-Toschi, T., Tres, A., & 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

Romero, A. (2017). Olive oil composition affected by environmental factors. In A. Kiritsakis, & F. Shahidi (Eds.), *Olives and olive oil as functional food. Bioactivity, chemistry and processing* (pp. 69–76). New York, NY: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119135340.ch4.

Rondanini, D. P., Castro, D. N., Searles, P. S., & Rousseaux, M. C. (2011). Fatty acid profiles of varietal virgin olive oils (Olea europaea L.) from mature orchards in warm arid valleys of Northwestern Argentina (La Rioja). *Grasas Y Aceites, 62*, 4. https://doi.org/10.3989/gya.125110

Sacco, A., Brescia, M. A., Liuzzi, V., Reniero, F., Guillou, C., Ghelli, S., & Van der Meer, P. (2000). Characterization of Italian olive oils based on analytical and nuclear magnetic resonance determinations. *Journal of the American Oil Chemists' Society, 77*, 619–625. https://doi.org/10.1007/s11746-000-0100-y

Sayago, A., González-Domínguez, R., Urbano, J., & Fernández-Recamales, A. (2019). Combination of vintage and new-fashioned analytical approaches for varietal and geographical traceability of olive oils. *Lebensmittel-Wissenschaft und -Technologie-Food Science and Technology, 111*, 99–104. https://doi.org/10.1016/j.lwt.2019.05.009

Torres, M. M., Pierantozzi, P., Cáceres, M. E., Labombarda, P., Fontanazza, G., & Maestri, D. M. (2009). Genetic and chemical assessment of Arbequina olive cultivar grown in Córdoba province, Argentina. *Journal of the Science of Food and Agriculture, 89*, 523–530. https://doi.org/10.1002/jsfa.3483

Tous, J. (2017). The influence of growing region and cultivar on olives and olive oil characteristics and on their functional constituents. In A. Kiritsakis, & F. Shahidi (Eds.), *Olives and olive oil as functional food. Bioactivity, chemistry and processing* (pp. 45–80). New York, NY: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119135340.ch4.

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*, 117–123. https://doi.org/10.1016/j.chroma.2006.05.029

Vichi, S., Lazzez, A., Grati-Kamoun, N., Lopez Tamames, E., & Buxaderas, S. (2010). Evolution of sesquiterpene hydrocarbons in virgin olive oil during fruit ripening. *Journal of Agricultural and Food Chemistry, 58*, 6972–6976. https://doi.org/10.1021/jf100497c

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 olive oil*. New York, NY: Nova Science Publishers, Inc., ISBN 978-1-53614-596-0