Mono- and sesquiterpenoid fingerprinting: a powerful and streamlined solution for pine nut authentication

B. Torres-Cobos[1,2], S. Nicotra[1,2], C. Asensio-Manzano[1], N. Aletà[3,4], A. Teixidó[4], M. Rovira[5], A. Romero[5], F. Guardiola[1,2], S. Vichi[1,2*], A. Tres[1,2]

[1] University of Barcelona, Department of Nutrition, Food Sciences and Gastronomy, Prat de la Riba 171, Santa Coloma de Gramenet, 08921, Spain.

[2] University of Barcelona, Institute of Research on Food Nutrition and Safety (INSA-UB), Prat de la Riba 171, Santa Coloma de Gramenet, 08921, Spain.

[3] Institute of Agrifood Research and Technology (IRTA). Fruit Tree Program. Torre Marimon. 08140 Caldes de Montbui, Spain

[4] Forest Science and Technology Centre (CTFC). Multifunctional Forest Management Program. 25280 Solsona, Lleida, Spain.

[5] Institute of Agrifood Research and Technology (IRTA), Ctra. de Reus – El Morell Km 3.8, Constantí, 43120, Spain.

*Corresponding author: S. Vichi, stefaniavichi@ub.edu

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.

Abstract

This study proposes a novel authentication method for pine nut geographical and botanical origin, using mono- and sesquiterpene fingerprints (extracted ion chromatograms from specific ions) analysed via solid-phase microextraction combined with gas chromatography-mass spectrometry, combined with chemometrics (partial least squares – discriminant analysis). It was tested on 253 samples from China, Russia (major producers of *Pinus koraiensis* and *Pinus sibirica*), Spain and Turkey (supplying *Pinus pinea*), across harvest years. The method achieved 100% accuracy in external validation when distinguishing Spanish from non-Spanish pine nuts, and 99% accuracy in differentiating *Pinus pinea* samples from two distinct Spanish regions. This simple, affordable, and automatable approach proves effective as a screening tool that could be applied to support official controls. Pine nuts are highly valued worldwide, with their sensory and nutritional characteristics influenced by their species and origin, which affect their price and make them vulnerable to counterfeiting.

Keywords: Pine nuts; Fraud; Terpenoids; Fingerprint; Authenticity; Geographical origin; PLS-DA.

1. Introduction

Pine nuts, popularly known as the “white gold”, are the most expensive nuts on the market. They account for only 1% of global tree nut production but have a supply value of more than 1.3 billion USD (INC, 2023). Asia stands as the primary global producer of pine nuts, with China, Russia, and North Korea leading the output, followed by the Mediterranean basin, with Turkey and the Iberian Peninsula as the main producers (INC, 2023).

While many agri-food products have cultivars farmed worldwide, the species of pine nuts are strictly tied to their geographical origins. The most common species of pine nuts among the 20 commercially available are *Pinus pinea* L. (*P. pinea*), predominantly growing in the Mediterranean region, and *Pinus koraiensis* Siebold & Zucc. (*P. koraiensis*) and *Pinus sibirica* (*P. sibirica*), primarily sourced from China and Russia, respectively (Awan and Pettenella, 2017; Moscetti et al., 2021). The sensory attributes, nutritional values and market price of pine nuts are highly dependent on the species and region of origin (Awan et al., 2017; Evaristo et al., 2013; Mutke, 2022). Mediterranean pine nuts are highly valued and appreciated by consumers, reaching prices higher than 100 EUR/kg (Mutke, 2022); however, they account for only 5% of the world average pine nut production (INC, 2023). In contrast, Chinese or Russian pine nuts are usually sold at much lower prices, often less than a third of the value of Mediterranean ones (Evaristo et al., 2013; Moscetti et al., 2021; Mutke, 2022).

Despite the differences among pine nuts from various origins and species, non-expert consumers often find them difficult to distinguish. Consequently, EU regulations and the International Organization for Standardization (ISO 7911:1991) mandate or acknowledge the declaration of the country of origin on pine nut packaging (Regulation (EU) No 2023/2429). Additionally, international commercial labelling standards recommend including the botanical species (UNECE, 2013). These label claims need to be verified by regulatory bodies to prevent fraud and protect consumers. Indeed, due to the significant price difference between pine nuts of different origins, they are highly vulnerable to economically motivated fraudulent practices

such as counterfeiting or adulteration. These practices can have serious consequences not only for the economy, impacting both the market and producers, but also for consumers' health, as they compromise the traceability chain of food products (Moscetti et al., 2021). In the particular case of pine nuts, misrepresenting their origin carries an added risk because Chinese *P. koraiensis* is sometimes marketed mixed with other pine seed species like *Pinus armandii* Franch., which has been linked to the dysgeusia called 'Pine Mouth Syndrome' (Mutke et al., 2013; Destillats et al., 2011).

For all these reasons, disposing of reliable methods for pine nut authentication is crucial to safeguard the interests of both producers and consumers. Traditional methods to authenticate pine nuts have been based on phenotypic observations of physical traits such as pine nut kernel morphology (Fardin-Kia et al., 2012; Loewe-Muñoz et al., 2018; Mikkelsen et al., 2014) but their susceptibility to external agents, and the fact that the evaluation is limited to whole kernels, hinder their effectiveness. Consequently, some studies have focused on genetic analysis to distinguish pine nuts species (Handy et al., 2011). Despite their reliability and accuracy, these methods are laborious, complex, destructive, and expensive (Fardin-Kia et al., 2012; Ríos-Reina et al., 2021), and thus, hardly applicable for routine analysis of large sample sets.

Alternatively, pine nut composition has been investigated using different analytical approaches, including targeted fatty acids analysis (Destaillats et al., 2010, 2011; Evaristo et al., 2013; Fardin-Kia et al., 2012) and comprehensive spectroscopic techniques such as near infrared spectroscopy (Loewe et al., 2017; Moscetti et al., 2021). Image analysis was also proposed for pine nut authentication (Ríos-Reina et al., 2021), although its application is restricted to entire kernels. While these methods showed promising results, there is still a need to fully evaluate their efficiency on sample sets that sufficiently represent the natural diversity of pine nut production, covering a wider range of origins, producers, harvest years, and species. Moreover, although pine nut species are strongly associated with specific geographical macro-areas, no study has yet focused on authenticating the origin of pine nuts from the same species within

these regions. Therefore, it is essential to develop a fit-for-purpose analytical method to verify pine nut authenticity.

In this context, previous research has demonstrated that mono- and sesquiterpenes could be reliable markers of varietal and geographical origin of different plant species and vegetable-derived products such as spices, alcoholic beverages and oils (Avula et al., 2015; Matsushita et al., 2018; Vichi et al., 2005; Marti et al., 2014; Quintanilla-Casas et al., 2022a; Torres-Cobos et al., 2021; Ugolini et al., 2024), but their potential has not yet been explored for the authentication of pine nuts. These terpenes are secondary plant metabolites whose presence and composition are highly dependent on the plant botanical and geographical origin. In fact, they are shaped by environmental and genetic factors, with minimal impact from other factors such as processing or storage conditions (Quintanilla-Casas et al., 2020; Vichi et al., 2010, 2018). When applied to virgin olive oils, sesquiterpene chromatographic fingerprint coupled with pattern recognition techniques, such as Partial Least Squares-Discriminant Analysis (PLS-DA), has been shown to be fast, robust and efficient for varietal and geographical authentication (Quintanilla-Casas et al., 2022A; Torres-Cobos et al., 2021). While other nut species typically lack appreciable amounts of terpenoids in their kernels, conifers produce an abundant amount of volatile and semi-volatile terpene metabolites and some of them have also been identified in pine nut kernels (Adelina et al., 2021; Rogachev et al., 2015). Recent reports documenting variations in the mono- and sesquiterpene composition among different pine species and origins (Arrabal et al., 2012; Faria et al., 2021; Kim et al., 2024) position the volatile terpene fingerprint as a promising marker for pine nut authentication. Moreover, monoterpenoids were reported as the main compounds in the essential oils of the pine bark, wood, needles, and cones, as well as in the volatile fraction of raw pine nut kernels (Adelina et al., 2021; Nikolic et al., 2022; Rogachev et al., 2015). The hypothesis of our study is that the mono- and sesquiterpene fingerprint of pine nuts can serve as a reliable and efficient marker to discriminate among pine nut kernel species and provenances.

The objective of the present research is to develop a fast, efficient, and reliable method to enhance the discrimination of pine nuts based on the volatile and semi-volatile terpene fingerprint obtained by Headspace Solid Phase Microextraction-Gas Chromatography-Mass Spectrometry (HS-SPME-GC–MS) from a wide sample set reflecting their natural variability. This involved the development and external validation of two PLS-DA classification models: (i) a multispecies geographical model to distinguish between Spanish *P. pinea* kernels vs. Asian kernels of other species, and (ii) a *P. pinea* geographical model to differentiate pine nuts of the same species from two distinct Spanish regions. This approach represents a novel application of the terpene fingerprinting method to pine nuts, addressing a critical gap in the authentication of this commodity and proposing a highly effective and scalable solution for food control.

2. Material and methods

2.1 Sampling

A set of 253 pine nut samples from different geographical regions and species was obtained from 2020 to 2023 in the frame of the Tracenuts project (PID2020-117701RB-I00) (**Table S1** of **Supplementary Information**). Among these samples, 83 were commercial samples from: China (CHN, n = 53), Russia (RUS, n = 22) and Turkey (TUR, n = 8). According to the natural distribution of pine nut species, it was assumed that CHN and RUS samples did not originate from *P. pinea* but primarily belonged to *P. koraiensis* and *P. sibirica*, the most commercially significant species from these countries (Awan and Pettenella, 2017). Commercial TUR samples may have belonged to both *P. pinea* and other local species (Bonari et al., 2020). All commercial samples were within their appropriate consumption date at the time of analysis. The remaining 170 samples were *P. pinea* kernels from Spain (ESP), sourced from two regions: Central Spain, Madrid and Castile and Leon (CS, n= 96) and Catalonia (CAT, n= 74). All the Spanish samples were traceable and were supplied by the Institute of Forest Science (ICIFOR-INIA, Madrid, Spain), the "*Centro de Servicios y Promocion Forestal y de su Industria de Castilla y León*" (CESEFOR foundation, Soria, Spain), and the Institute of Agrifood Research and Technology (IRTA-Torre Marimon, Caldes de

Montbui, Spain). Samples were directly harvested from forests and dried according to UNECE STANDARD DDP-12 (2013), similarly to commercial samples. Both commercial and non-commercial samples were stored at 4 °C until analysis.

2.2 Headspace-solid phase microextraction (HS-SPME)

Pine nut samples were analysed under conditions adapted from Vichi et al. (2006) using a Combi-pal autosampler (CTC Analytics, Zwingen, Switzerland). An aliquot of approximately 1 g of whole pine nuts was weighed into a 10 mL vial fitted with a PTFE/silicone septum. The sample was conditioned at 70 °C for 10 min, followed by exposing a divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS) fiber (2 cm length, 50/30 µm film thickness) from Supelco (Bellefonte, Pennsylvania, USA) to the sample headspace for 60 min, at the same temperature. Then, the fiber was desorbed at 260 ˚C for 10 min in the gas chromatograph injection port, the injector was maintained in split-less mode for the first 5 min. To monitor carryover, blank samples were alternated between injections.

2.3 Gas Chromatography-Mass Spectrometry (GC-MS)

The mono- and sesquiterpene 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). Helium was the 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, Pennsylvania, USA). 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 2.338 scan/s and the electron energy was 70 eV. Data acquisition was performed in the selected ion monitoring (SIM) mode between 0 and 42.7 min, by registering the Extracted Ion Chromatogram (EIC) of 7 ions which have been reported to be characteristic of the mono and sesquiterpene compounds and their oxygenated derivates: *m/z* 93, 95, 119, 159, 161, 189 and 204 (Maleknia et al., 2007; Reed, 1963;

171 Tani et al., 2003; Torres-Cobos et al., 2021; Vichi et al., 2010). Therefore, for each ion, the
172 intensities of a total of 6621 scans were acquired and used as fingerprinting data (section 2.3)
173 to build the authentication models (section 2.4).
174 After fingerprinting models were developed, we tentatively identified the compounds
175 corresponding to the scans leading to the most relevant regression coefficients (section 2.4.3).
176 To do so, acquisition was carried out in the full scan mode in the range *m/z* 35-350 and the MS
177 spectra at the retention times corresponding to the relevant scans were obtained. This tentative
178 molecular structure identification was a level 3 identification (tentative candidate, evidence
179 exists for possible structure, but insufficient information for one exact structure only) according
180 to Schymanski et al. (2014).

181 **2.4 Fingerprinting approach**

182 A fingerprinting approach was followed using the EICs of the 7 selected ions. Scan intensities
183 were considered from 0 to 47.2 min for each ion (6621 scans × 7 ions = 46347 variables per
184 sample). The acquisition interval has been extended from previous studies (Torres-Cobos et al.,
185 2021) to include the monoterpenoids that appear at the initial times of the chromatogram (from
186 0 to 30 minutes) due to their relevance and abundance in pine nuts. A data matrix was built for
187 each ion, with the scan intensities of each EIC (columns) for all the samples (rows) (7 matrices
188 of 6621 columns x 253 rows). Differences between injections were corrected by normalizing
189 each EIC, which consisted in dividing each scan intensity by the total sum of intensities (Nam et
190 al., 2020). Then, the EICs in each matrix were aligned by Correlation Optimized Warping (COW)
191 algorithm in Matlab® (Nielsen et al., 1998). Finally, the 7 aligned EIC matrices were concatenated
192 conforming a two-way unfolded matrix (253 samples × 46347 variables).

193 **2.5 Chemometrics**

194 **2.5.1. Data exploration and preliminary multi-class geographical model**

195 The data treatment and model building were performed with SIMCA software v13.0© (Sartorius,
196 Göttingen, Germany). A Principal Component Analysis (PCA) was performed for the exploration

of data and to identify potential outliers, according to Hotelling's T^2 range and Q-residuals model parameters. The exploratory analysis of the dataset showed no outliers.

A preliminary multi-class PLS-DA classification model was built to discriminate among the four countries of origin (CHN, TUR, RUS, ESP), to assess the potential of the terpene fingerprinting to distinguish pine nuts from different origins. Multi-class PLS-DA models operate as multiple binary models; each class being compared to the rest of the samples. A dummy Y matrix is used, containing as many classification vectors as classes. Each vector has values of 1 for a specific class and 0 for all other classes. The multi-class PLS-DA model was internally validated through leave 10%-out cross validation. The model's optimal pre-processing and number of latent variables (LV) were selected according to the lowest Root Mean Squared Error of Cross Validation (RMSEcv) criteria. The pre-processing was mean centring and scaling to the unit of variance. To evaluate model overfitting, permutation test (n = 20 permutations) and ANOVA on the cross-validated predictive residuals (p-value) were carried out (Eriksson et al., 2008; Quintanilla-Casas et al., 2020). The suitability PLS-DA model was evaluated by the Q^2 values and the percentage of correct classification of each class, and the resulting score plot was examined to identify any clustering among samples.

2.5.2. Partial Least Squares Discriminant Analysis (PLS-DA) binary classification models

After the initial data exploration and the exclusion of origins represented by fewer than 20 samples (TUR), two PLS-DA binary classification models were built: (i) a multi-species geographical model to discriminate between pine nuts from ESP (*P. pinea*) and non-ESP (other species from: CHN, and RUS,), and (ii) a *P. pinea* geographical model to classify the ESP *P. pinea* samples by their region of production: CAT and CS.

For each authentication model, the sample set was split following a stratified random sampling strategy into training (80 % of the samples of each class: ESP/non-ESP model, n = 196; CAT/CS model, n = 136) and validation set (20 % of the samples of each class: ESP/non-ESP model, n = 49; CAT/CS model, n = 34). This splitting was run seven times (7 iterations) to evaluate the effect

of the sample set composition and to increase the robustness of the external validation. The sample set splitting information, including validation and training sets, is summarized in **Table S1** of **Supplementary Information**.

In each iteration, a PLS-DA binary model (training model) was fitted and internally validated through leave 10%-out cross validation, using the samples in the corresponding training set. The model's optimal pre-processing and LV were selected according to the RMSEcv criteria. For all training models, the optimal pre-processing was mean centring and scaling to the unit of variance. To evaluate model overfitting, permutation test (n = 20 permutations) and ANOVA on the cross-validated predictive residuals (p-value) were carried out (Eriksson et al., 2008; Quintanilla-Casas et al., 2020). None of the training models was overfitted according to the permutation test and ANOVA p-value results. Subsequently, each training model was externally validated by predicting the class of the samples in the corresponding validation set, which had not been used to build the model. Therefore, for each type of model, seven training PLS-DA models and the corresponding seven external validations were obtained, to verify that results were not driven by specific influential samples and thus, to increase the robustness of the external validation.

In PLS-DA binary models, classes are expressed as PLS dummy variables (here, 1 for non-ESP, and CS classes, and 0 for ESP and CAT classes). The PLS predicted value (PV) of each sample was used for its classification into one class or the other according to a classification threshold (here, PV = 0.5). The performance of each PLS-DA model was evaluated by the Q^2 values and efficiency, which was expressed as the percentage of correct classification of each class, the sensitivity (the number of true positive results/ [the number of true positive results + the number of false negative results]) and specificity (the number of true negative results/ [the number of true negative results + the number of false positive results]) values. Wilson score intervals were calculated to establish confidence intervals for models' sensitivity and specificity (Wilson, 1927).

Non-ESP and CS samples were arbitrarily defined as the positive samples for the corresponding models.

2.5.3. Evaluation of PLS-DA regression coefficients

The regression coefficients of the PLS-DA models were studied to tentatively identify the key variables that mainly contributed to the discrimination between classes. The jack-knife standard error of cross-validation (SEcv) was used to evaluate the significance of the regression coefficients, considering as significant those with values higher than their corresponding SEcv (Torres-Cobos et al., 2024). Among the significant variables, only the ones with the highest absolute values (the 3 % higher regression coefficients) were considered. The corresponding compounds were tentatively identified based on their mass spectra and elution order from full scan injections as explained in section 2.4.

3. Results and discussion

3.1 Exploratory analysis and preliminary multi-class geographical model

The preliminary multi-class PLS-DA model to classify the samples according to their country of origin showed promising results. The inspection of the score plot (**Figure 1**) evidenced a clear clustering of samples by country of origin. LV1 was useful in discriminating TUR and ESP samples from CHN and RUS ones, whereas LV3 distinguished TUR from ESP and RUS from CHN pine nuts, achieving four groups quite differentiated from one another.

The leave 10%-out cross validation (**Table 1**) yielded a 100% correct classification for the pine nuts from ESP and TUR, and high correct classification rates for the CHN (92%) and RUS (95%) classes, with only 2 CHN samples misclassified as RUS, and one RUS misclassified as CHN. The misclassification may be attributed to the greater similarity between these two classes, as both CHN and RUS samples belong to species other than *P. pinea* and originate from regions geographically distant from the Mediterranean. Additionally, this similarity is evident in the scores plot (**Figure 1**), where the clusters for RUS and CHN samples show significant dispersion and partial overlap. This overlap suggests that the chemical fingerprinting approach may have

difficulty distinguishing between these two classes due to their more closely related terpene composition. Permutation test and ANOVA p-value allowed excluding model overfitting (**Table S2** of **Supplementary Information**).

Categories with n < 20, such as TUR, are not suitable for proper external validation. Therefore, constructing further binary models for broader and better-represented categories was the chosen option to yield reliable results. However, these preliminary findings indicate the potential for developing future models to authenticate pine nuts by country of origin based on mono- and sesquiterpene fingerprint, with the appropriate sampling.

Finally, although the sample set included various crop years, they did not significantly affect the PLS-DA models built with geographical origin as the classification variable (**Supplementary Figure S1**), highlighting the robustness of PLS-DA classification models in accounting for factors other than the variable of interest.

3.2 PLS-DA binary classification models

3.2.1 Multi-species geographical PLS-DA model

Leave 10%-out cross-validation of the 7 binary ESP/non-ESP PLS-DA training models (80% of the samples) provided a 100% of correct classification of both classes. To verify the reliability of these promising outcomes, models' performances were assessed through external validation. This involved predicting the class of samples from the respective validation sets. The external validation results were expressed as mean values ± standard deviation obtained from the 7 iterations (**Table 2**).

The external validation outcomes corroborated the leave 10-out cross validation, correctly classifying all samples of the validation sets into either ESP or non-ESP categories, with maximum sensitivity and specificity, and without deviation. These results evidenced the exceptional effectiveness of terpene fingerprinting in distinguishing ESP *P. pinea* kernels from those of other geographical and botanical origins potentially used for counterfeiting, regardless of the specific region, harvest year or commercial brands and producers.

3.2.2 *P. pinea* geographical PLS-DA model

To assess the capability and potential limitations of using volatile terpene fingerprint for authenticating the origin of pine nuts, a classification model was built in a more challenging scenario. The goal was to discriminate between samples from the same species, *P. Pinea,* produced in the same country, ESP, but in distinct regions, CAT and CS.

In this case as well, internal validation of the 7 binary PLS-DA models built using the training sets obtained from the 7 iterations of the sample-set splitting achieved 100% accuracy, correctly classifying all the training samples into their respective region of origin. These results were further corroborated by the external validation (**Table 3**), where all CAT pine nuts were correctly classified, providing a specificity of 1. Only one CS sample was misclassified as CAT, resulting in a sensitivity of 0.98 and an overall correct classification rate of 99%.

These findings confirm the ability of the volatile terpene fingerprint to distinguish pine nuts from different regions, even when they originate from the same species and relatively close geographical areas. These results align with previous studies demonstrating the influence of pedoclimatic factors on the sesquiterpene composition of olive oil, enabling the differentiation of olive oils from various Catalan Protected Designation of Origin (PDO) regions, even when derived from the same cultivar and nearby geographical areas (Quintanilla-Casas et al., 2022b). The developed model can be applied to verify the identity of pine nuts from these regions/species, but this opens future research for further expanding the model for other regions and species, by developing and validating the models with new samples from these regions and species. Similarly, further research might address open questions, as for instance, which would be the ability of the model in revealing mixed samples from various origins or species, or if the model would work for identifying pine nut identity in complex foods.

3.3 Exploration of PLS-DA regression coefficients

Unlike other food matrices where sesquiterpene fingerprinting has been previously explored for authentication such as virgin olive oil, pine nut terpene fingerprint contains a notable fraction

composed of monoterpenoids (Adelina et al.,2021; Rogachev et al., 2015). Given the interconnected biosynthetic pathways of mono- and sesquiterpenoids, they are likely equally influenced by genetic and environmental factors. Consequently, both could potentially contribute to the geographical and botanical differentiation of pine nuts, and for this reason the entire fraction was included for evaluation in this study. To confirm that class discrimination was consistently based on specific terpene patterns, and to ascertain whether both mono- and sesquiterpenoids contributed to the discrimination, we examined the highest significant regression coefficients from both PLS-DA models and tentatively identified the terpenoid structure of the corresponding chromatographic peaks. It is important to emphasize that the aim was not to conduct a comprehensive study of all discriminant variables or to move towards a targeted analysis. Instead, we focused on the most relevant variables to confirm their terpene nature and to gain insight into their general molecular structure, such as whether they were monoterpene or sesquiterpene hydrocarbons, or oxygenated derivatives.

To explore the variables that had the greatest impact on discriminating between pine nuts classes, we examined the highest significant regression coefficients of both of ESP/non-ESP, and CAT/CS PLS-DA models as explained in section 2.4.3. For both models, plotting the regression coefficients against the variables of the unfolded matrix (**Figure 2**) revealed that the relevant coefficients were distributed along the entire EICs of ions with *m/z* 93, 95, and 119, while being concentrated towards the end of EICs of ions with *m/z* 159, 161, 189, and 204. This is because the former fragment ions are common in both mono- and sesquiterpenoids eluting across the entire chromatogram (Maleknia et al., 2007; Tani et al., 2003; Vichi et al 2006; Vichi et al 2010), whereas the latter are specific of semi-volatile sesquiterpenes with higher retention times (Vichi et al., 2006; Vichi et al 2010). At first glance, these outcomes suggest that both mono- and sesquiterpene families could be regarded as valuable markers for authenticating the botanical and geographical origins of pine nuts, endorsing the hypothesis that both mono- and sesquiterpenes would contribute to discrimination. Specifically, the ESP samples' highest

352 regression coefficients were mostly found in the middle-final section of EICs, being slightly more
353 abundant in EICs *m/z* 119 and 204 (**Figure 2**), and thus, probably attributable to sesquiterpenes.
354 Conversely, non-ESP class was mainly distinguished by compounds detectable in across the
355 entire EICs *m/z* 93, 95, 119 (**Figure 2**), likely including several monoterpenoids. Likewise, both
356 mono- and sesquiterpene compounds appeared to drive the discrimination between CAT and
357 CS classes. As the most relevant CAT coefficients were in the middle-final section of most of EICs,
358 particularly in EICs *m/z* 95, 159, 204, they were probably attributable to sesquiterpenes. The
359 predominant coefficients distinguishing CS pine nuts, similarly to non-ESP samples, were
360 distributed along the whole EICs at EICs *m/z* 93, 95, 119, several of them probably corresponding
361 to monoterpenoids.
362 To gain more insight and further support these findings, we examined the compounds related
363 to the most relevant variables in each model. To exemplify some of the tentatively identified
364 compounds that mainly contribute to class distinction, **Figure 3** compares the EICs at *m/z* 93 and
365 204 corresponding to a non-ESP vs an ESP sample, and to a CS vs a CAT sample. Firstly, it is
366 remarkable that several of these significant variables corresponded to minor compounds or not
367 well-resolved chromatographic peaks, which might hinder their identification and quantification
368 using traditional target approaches. This underscores the fingerprinting approach as a more
369 suitable option for their analysis, confirming previous findings (Quintanilla-Casas et al., 2020).
370 Next, concerning the nature of compounds driving the discrimination between ESP and non-ESP
371 samples, the tentative identification of relevant compounds suggested that the relevant
372 compounds detected in EIC at m/z 93 included both monoterpene hydrocarbons, eluting in the
373 initial part of the chromatogram (e.g. compounds with mass spectra attributable to compounds
374 such as pinene, camphene, sabinene, carene, myrcene, and cymene, relevant for the non-ESP
375 class; limonene and mentha triene, for the ESP class; myrcene and cymene, for the CS class, and
376 pinene and carene, for the CAT class), and their oxygenated derivatives, with higher retention
377 times (e.g. compounds with mass spectra attributable to limonene oxide, relevant for non-ESP

and CAT classes; camphor, for non-ESP and CS classes; dihydrocarvone for CS; borneol for non-ESP). On the other hand, the tentative identification of relevant compounds in EIC at *m/z* 204 permitted to assume that several relevant compounds had a sesquiterpene structure (e.g. in the ESP/non-ESP model, compounds with mass spectra matching with that of copaene, junipene, cubebene, cadinene, muurolene and amorphene were relevant for the ESP class, and two not identified compounds whose spectra matched with those of various possible sesquiterpenes distinguished the non-ESP class. In the CAT/ CS model, compounds possibly corresponding to copaene, murolene and an unidentified sesquiterpene distinguished CS samples, while others, likely junipene and cubebene were relevant for CAT ones). The EIC at *m/z* 204 was selected as an example because it corresponds to the molecular ion of sesquiterpene hydrocarbons (Vichi et al., 2006). In summary, the examination of regression coefficients evidenced that numerous variables across all the acquired ions, corresponding to minor and major species, contributed significantly to class discrimination. Both monoterpene and sesquiterpene compounds, including hydrocarbons and their oxygenated derivatives, played a crucial role in the classification. Specifically, monoterpenes seemed to be more characteristic of the non-ESP and CS classes, whereas several compounds with sesquiterpene structure contributed equally to distinguish the origin of samples in both ESP/non-ESP and CAT/ CS models.

4. Conclusions

The volatile and semi-volatile terpene fingerprinting has proven to be a powerful method for authenticating pine nuts, with PLS-DA effectively identifying patterns related to their origin and species, while minimizing variables linked to factors such as harvest year or commercial producer, and therefore, our hypothesis was confirmed. This method provided a high efficiency (> 99%) in the discrimination of pine nuts of different species into ESP and non-ESP classes, and between pine nuts of the same species but from two nearby geographical regions, CAT and CS. Additionally, the preliminary multi-class PLS-DA origin model showed the potential of this

method to authenticate multiple geographical origins, provided a sufficiently comprehensive and diverse sample set is used.

Finally, the exploration of PLS-DA regression coefficients confirmed that the compounds related to the variables primarily contributing to the discrimination have a mono- and sesquiterpene structure, including both terpene hydrocarbons and some oxygenated derivatives.

In conclusion, volatile terpene fingerprinting proved to be fast, efficient and straightforward, making it easily to apply to large number of samples in routine laboratories. It could serve as a valuable supporting screening tool for official controls, enhancing their effectiveness and ensuring consumer protection.

Acknowledgments

INSA-UB Maria de Maeztu Unit of Excellence (Grant CEX2021-001234-M) funded by MICIU/AEI/ERDF, EU. INSA-UB authors are part of the LiBiFOOD Consolidated Research Group (2021-SGR-00854) recognised by AGAUR (Generalitat de Catalunya). The authors would like to express their gratitude to INIA-CIFOR and CESEFOR for providing the *P. pinea* samples from Madrid and Castile and Leon.

Research data:

Torres-Cobos, Berta; Nicotra, Soriana B; Asensio-Manzano, Cèlia; Aletà, Neus; Teixidó, Anna; Rovira, Mercè; Romero, Agustí; Guardiola Ibarz, Francesc; Vichi, Stefania; Tres Oliver, Alba, 2024, "Pine nut mono- and sesquiterpenoid fingerprints (Extracted Ion Chromatograms) obtained by HS-SPME-GC-MS ", https://doi.org/10.34810/data1730 CORA. Repositori de Dades de Recerca.

Funding

This work was developed in the context of the project TRACENUTS, PID2020-117701RB-I00 financed by MICIU/AEI/ 10.13039/501100011033. B. Torres-Cobos thanks the Spanish Ministry of Universities predoctoral fellowships FPU20/01454.

CRediT authorship contribution statement

B. Torres-Cobos: Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft. C. Asensio: Formal analysis, Data curation, Investigation. S. Nicotra: Data curation, Investigation. N. Aletà: Conceptualization, Resources, Writing – review & editing; A. Teixidó: Resources, Writing – review & editing. M. Rovira: Resources, Writing – review & editing. A. Romero: Resources, Writing – review & editing. F. Guardiola: Supervision, Writing – review & editing. S. Vichi: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing. A. Tres: Conceptualization, Methodology, Supervision, Writing – review & editing.

Declarations of 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.

References

Adelina, N.M., Wang, H., Zhang, L., & Zhao, Y. (2021). Comparative analysis of volatile profiles in two grafted pine nuts by headspace-SPME/GC–MS and electronic nose as responses to different roasting conditions. Food Research International, 140, Article 110026. https://doi.org/10.1016/j.foodres.2020.110026

Arrabal, C., García-Vallejo, M. C., Cadahia, E., Cortijo. M., & Fernández de Simón, B. (2012). Characterization of two chemotypes of *Pinus pinaster* by their terpene and acid patternsin needles. *Plant Systematics and Evolution*, *298*, 511–522. JSTOR, http://www.jstor.org/stable/43498310 (accessed Apr 16, 2024).

Avula, B., Smillie, T. J., Wang, Y. H., Zweigenbaum, J., & Khan, I. A. (2015). Authentication of true cinnamon (*Cinnamon verum*) utilising direct analysis in real time (DART)-QToF-MS. *Food Additives & Contaminants: Part A*, *32*, 1–8. https://doi.org/10.1080/19440049.2014.981763

Awan, H. U. M., & Pettenella, D. (2017). Pine nuts: a review of recent sanitary conditions and market development. *Forests*, *8*, Article 367. https://doi.org/10.3390/f8100367

Bonari, G., Chytrý, K., Çoban, S., & Chytrý, M. (2020). Natural forests of *Pinus pinea* in western Turkey: a priority for conservation. *Biodiversity and Conservation*, *29*, 3877–3898. https://doi.org /10.1007/s10531-020-02052-z

Destaillats, F., Cruz-Hernandez, C., Giuffrida, F., & Dionisi, F. (2010). Identification of the botanical origin of pine nuts found in food products by gas−liquid chromatography analysis of fatty acid profile. *Journal of Agricultural and Food Chemistry*, *58*, 2082-2087. https://doi.org/10.1021/jf9041722

Destaillats, F., Cruz-Hernandez, C., Giuffrida, F., Dionisi, F., Mostin, M., & Verstegen, G. (2011). Identification of the botanical origin of commercial pine nuts responsible for dysgeusia by gas-liquid chromatography analysis of fatty acid profile. *Journal of Toxicology*, *2011*, Article 316789. https://doi.org/10.1155/2011/316789

Eriksson, L., Trygg, J., & Wold, S. (2008). CV-ANOVA for significance testing of PLS and OPLS® models. *Journal of Chemometrics*, *22*, 594–600. https://doi.org/10.1002/cem.1187

Evaristo I., Batista D., Correia I., Correia P., & Costa R. (2013). Chemical profiling of Portuguese *Pinus pinea* L. nuts and comparative analysis with *Pinus koraiensis* Sieb. & Zucc. commercial kernels. *Options Méditerranéennes. Séries A: Mediterranean Seminars*, *105*, 99–104. http://om.ciheam.org/om/pdf/a105/00006787.pdf

Fardin-Kia, A. R., Handy, S. M., & Rader, J. I. (2012). Characterization of pine nuts in the U.S. market, including those associated with “pine mouth”, by GC-FID. *Journal of Agricultural and Food Chemistry*, *60*, 2701–2711. https://doi.org/10.1021/jf205188m

Faria, J.M.S., & Rodrigues, A.M. (2021). Metabolomic variability in the volatile composition of essential oils from *Pinus pinea* and *P. pinaster*. *Biology and Life Sciences Forum*, *2*, Article 14. https://doi.org/10.3390/BDEE2021-09428

Handy, S., Parks, M. B., Deeds, J. R., Liston, A., De Jager, L. S., Luccioli, S., Kwegyir-Afful, E., Fardin-Kia, A. R., Begley, T. H., Rader,J. I., & Diachenko, G. W. (2011). Use of the chloroplast gene ycf1 for the genetic differentiation of pine nuts obtained from consumers experiencing dysgeusia. *Journal of Agricultural and Food Chemistry*, *59*, 10995−11002. https://doi.org/10.1021/jf203215v

International Nut and Dried Fruits (INC), 2023. Nuts & Dried Fruits Statistical Yearbook 2022/20223. https://inc.nutfruit.org/publications/ (accessed Apr 10, 2024).

International Organization for Standardization (ISO). (1991). ISO 7911:1991 Unshelled pine nuts —Specification.

Kim, E., Yang, S., Jeon, B. B., Song, E., & Lee, H. (2024). Terpene compound composition and antioxidant activity of essential oils from needles of *Pinus densiflora*, *Pinus koraiensis*, *Abies holophylla*, and *Juniperus chinensis* by harvest period. *Forests*, *15*, Article 566. https://doi.org/10.3390/f15030566

Loewe-Muñoz, V., Álvarez, A., & Navarro-Cerrillo, R. M. (2018). Morphometric and chemical fruit variability of selected stone pine trees (*Pinus pinea* L.) grown in non-native environments. *Plant Biosystems - An International Journal Dealing with All Aspects of Plant Biology*, *152*, 547–555. https://doi.org/10.1080/11263504.2018.1435587

Loewe, V., Navarro-Cerrillo, R. M., García-Olmo, J., Riccioli, C., & Sánchez-Cuesta, R. (2017). Discriminant analysis of Mediterranean pine nuts (*Pinus pinea* L.) from Chilean plantations by near infrared spectroscopy (NIRS). *Food Control*, *73*, 634–643. https://doi.org/10.1016/j.foodcont.2016.09.012

Maleknia, S. D., Bell, T. L., & Adams, M. A. (2007). PTR-MS analysis of reference and plant-emitted volatile organic compounds. *International Journal of Mass Spectrometry*, *262*, 203-210. https://doi.org/10.1016/j.ijms.2006.11.010

Marti, G., Boccard, J., Mehl, F., Debrus, B., Marcourt, L., Merle, P., Delort, E., Baroux, L., Sommer, H., Rudaz, S., & Wolfender, J. (2014). Comprehensive profiling and marker identification in non-volatile citrus oil residues by mass spectrometry and nuclear magnetic resonance, *Food Chemistry*, *150*, 235–245. https://doi.org/10.1016/j.foodchem.2013.10.103.

Matsushita, T., Zhao, J. J., Igura, N., & Shimoda, M. (2018). Authentication of commercial spices based on the similarities between gas chromatographic fingerprints. *Journal of the Science of Food and Agriculture*, *98*, 2989–3000. https://doi.org/10.1002/jsfa.8797

Mikkelsen, A. Æ., Jessen, F., & Ballin, N. Z. (2014). Species determination of pine nuts in commercial samples causing pine nut syndrome. *Food Control*, *40*, 19–25. https://doi.org/10.1016/j.foodcont.2013.11.030

Moscetti, R., Berhe, D. H., Agrimi, M., Haff, R. P., Liang, P., Ferri, S., Monarca, D., & Massantini, R. (2021). Pine nut species recognition using NIR spectroscopy and image analysis. *Journal of Food Engineering*, *292*, Article 110357. https://doi.org/10.1016/j.jfoodeng.2020.110357

Mutke, S. (2022). Producción y comercio de piñones comestibles en el mundo. *Montes*, *147*, 40–43. https://www.revistamontes.net/Buscador.aspx?id=15180 (accessed Apr 14, 2024).

Mutke, S., Pastor, A., & Picardo, A. (2013). Toward a traceability of European pine nuts "from forest to fork". *Options Méditerranéennes. Séries A: Mediterranean Seminars*, *105*, 105–109. http://om.ciheam.org/om/pdf/a105/00006788.pdf

Nam, S. L., de la Mata, A. P., Dias, R. P., Harynuk, J. J. (2020). Towards standardization of data normalization strategies to improve urinary metabolomics studies by GC× GC-TOFMS. *Metabolites*, *10*(9), 376. https://doi.org/10.3390/metabo10090376

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

Nikolic, M. V., Jakovljevic, V. L., Bradic, J. V., Tomovic, M. T., Petrovic, B. P., & Petrovic, A. M. (2022). Korean and siberian pine: review of chemical composition and pharmacological profile. *Acta Poloniae Pharmaceutica*, *79*, 785-797. https://doi.org/10.32383/appdr/161040

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*, Article 125556. https://doi.org/10.1016/j.foodchem.2019.125556

Quintanilla-Casas, B., Torres-Cobos, B., Guardiola, F., Servili, M., Alonso-Salces, R. M., Valli, E., Bendini, A., Gallina Toschi, T. , Vichi, S., & Tres, A. (2022a). Geographical authentication of virgin olive oil by GCMS sesquiterpene hydrocarbon fingerprint: Verifying EU and single country label-declaration. *Food Chemistry*, *378*, Article 132104. https://doi.org/10.1016/j.foodchem.2022.132104

Quintanilla-Casas, B., Torres-Cobos, B., Guardiola, F., Romero, A., Tres, A., & Vichi, S. (2022b). Geographical authentication of virgin olive oil by GC-MS sesquiterpene hydrocarbon fingerprint: Scaling down to the verification of PDO compliance. *Food Control*, *139*, 109055. https://doi.org/10.1016/j.foodcont.2022.109055

Reed, R. I. (1963). 13 - Mass spectra of terpenes. In F. W. McLafferty (Ed.). Mass spectrometry of organic ions. New York, Academic Press, pp 637-699. ISBN: 9780323142779.

Regulation (EU) No 2023/2429 of the European Parliament and of the Council of 17 August 2023 as regards marketing standards for the fruit and vegetables sector, certain processed

fruit and vegetable products and the bananas sector, pp. 1–77. http://data.europa.eu/eli/reg_del/2023/2429/oj

Ríos-Reina, R., Callejón, R. M., & Amigo, J. M. (2021). Feasibility of a rapid and non-destructive methodology for the study and discrimination of pine nuts using near-infrared hyperspectral analysis and chemometrics, *Food Control*, *130*, Article 108365. https://doi.org/10.1016/j.foodcont.2021.108365

Rogachev, A. D., & Salakhutdinov, N. F. (2015). Chemical composition of *Pinus sibirica* (Pinaceae). *Chemistry & biodiversity*, *12*, 1-53. https://doi.org/10.1002/cbdv.201300195

Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H. P., & Hollender, J. (2014). Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environmental Science & Technology*, *48*, 2097–2098. https://doi.org/10.1021/es5002105

Tani, A., Hayward, S., & Hewitt, C. N. (2003). Measurement of monoterpenes and related compounds by proton transfer reaction-mass spectrometry (PTR-MS). *International Journal of Mass Spectrometry*, *223*, 561-578. https://doi.org/10.1016/S1387-3806(02)00880-1

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

Torres-Cobos, B., Quintanilla-Casas, B., Rovira, M., Romero, A., Guardiola, F., Vichi, S., & Tres, A. (2024). Prospective exploration of hazelnut's unsaponifiable fraction for geographical and varietal authentication: A comparative study of advanced fingerprinting and untargeted profiling techniques. *Food Chemistry*, *441*, Article 138294 https://doi.org/10.1016/j.foodchem.2023.138294

Ugolini, T., Mattagli, F., Melani, F., Zanoni, B., Migliorini, M., Trapani, S., Giambanelli, E., Parenti, A., Mulinacci, N., & Cecchi, L. (2024). HS-SPME-GC-MS and Chemometrics for the Quality Control and Clustering of Monovarietal Extra Virgin Olive Oil: A 3-Year Study on Terpenes and Pentene Dimers of Italian Cultivars. Journal of Agricultural and Food Chemistry, *72*(19), 11124-11139. https://doi.org/10.1021/acs.jafc.4c00610

581 UNECE (2013), UNECE Standard DDP-12 concerning the marketing and commercial quality
582 control of Pine Nut Kernels. 2013 Editions. United Nations, New York and Geneva.
583 https://unece.org/sites/default/files/2023-08/12_PineNutKernels_e.pdf

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

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

591 Vichi, S., Riu-Aumatell, M., Mora-Pons, M., Buxaderas, S., & López-Tamames, E. (2005)
592 Characterization of volatiles in different dry gins. *Journal of Agricultural and Food*
593 *Chemistry*, *53*, 10154–10160. https://doi.org/10.1021/jf058121b

594 Vichi, S., Tres, A., Quintanilla-Casas, B., Bustamante, J., & Guardiola, F. (2018). Sesquiterpene
595 hydrocarbons, a promising tool for virgin olive oil geographical authentication. In M.
596 Kontominas (Ed.). *Authentication and Detection of Adulteration of Olive Oil*. New York,
597 NY: Nova Science Publishers, Inc. ISBN: 978-1-53614-596-0.

598 Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal*
599 *of the American Statistical Association*, *22*(158), 209-212.
600 https://doi.org/10.1080/01621459.1927.10502953.

FIGURE CAPTIONS

Fig. 1. Score scatter plot (LV1 vs LV3) of the PLS-DA classification model developed by country of origin (n = 253, 9 LVs, Q^2 = 0.659, RMSEcv = 0.216, ANOVA p-value < 0.05), based on pine nut volatile and semi-volatile terpene fingerprinting data. CHN: China, TUR: Turkey, RUS: Russia, ESP: Spain.

Fig. 2. Regression coefficients of the PLS-DA ESP *vs* non-ESP and CAT *vs* CS models, plotted against the variables (acquisition points) of the unfolded matrix. For each model, the most relevant coefficients for the prediction of the ESP and CAT classes are highlighted in blue (negative coefficients) and those relevant for non-ESP and CS in red (positive coefficients).

Fig. 3. Extracted chromatograms of two representative ions (*m/z* 93, 204) and tentative identification of compounds corresponding to some of the most relevant variables. 1) α-pinene, 2) camphene, 3) β-pinene, 4) sabinene, 5) δ-carene, 6) myrcene, 7) limonene, 8) mentha triene isomer, 9) *p*-cymene, 10) limonene oxide isomer, 11) copaene isomer, 12) camphor, 13) non-identified sesquiterpene, 14) junipene, 15) cubebene isomer, 16) dihydrocarvone, 17) cadinene isomer, 18) murolene isomer, 19) amorphene, 20) borneol, 21) NI sesquiterpene. Blue: relevant for ESP, CAT; red: relevant for non-ESP, CS.

Table NO MARKS

Table 1. Leave 10%-out cross validation of the four-class PLS-DA model developed by country of origin, based on pine nut mono- and sesquiterpene fingerprinting data.

Multi species geographical model: CHN/TUR/RUS/ESP							
	n	CHN (n)	TUR (n)	RUS (n)	ESP (n)	Not assigned (n)	Correct classification (%)
CHN	53	49	0	2	0	2	92.5
TUR	8	0	8	0	0	0	100.0
RUS	22	1	0	21	0	0	95.5
ESP	170	0	0	0	170	0	100.0
Total	253						98.0

Model parameters (N = 253) : 9 LVs, Q^2 = 0.659, RMSEcv = 0.216, ANOVA p-value < 0.05. CHN: China; TUR: Turkey; RUS: Russia; ESP: Spain.

Table 2. External validation of the binary PLS-DA model to discriminate samples into ESP and non-ESP based on pine nut mono- and sesquiterpene fingerprinting data. Results are mean values (± standard deviation) obtained from seven iterations.

Multi species geographical model: ESP/non-ESP						
	n	ESP (n)	non-ESP (n)	Correct classification (%)	Sensitivity	Specificity
ESP	34	34.0 ± 0.0	0.0 ± 0.0	100.0 ± 0.0		
non-ESP	15	0.0 ± 0.0	15.0 ± 0.0	100.0 ± 0.0		
Total	49			100.0 ± 0.0	1.0 ± 0.0 (0.80—1.0)*	1.0 ± 0.0 (0.90—1.0)*

*Mean of the Wilson score intervals calculated for the sensitivity and specificity of each model.
Model parameters: mean values obtained with the training sets (N = 196) from 7 iterations: 5 LVs, Q^2 = 0.969, RMSEcv = 0.086, ANOVA p-value < 0.05.

Table 3. External validation of the Spanish PLS-DA model to discriminate samples into Catalonia and Central Spain based on pine nut mono- and sesquiterpene fingerprinting data. Results are mean values (± standard deviation) obtained from seven iterations.

***P. pinea* geographical model: CAT/CS**						
	n	CS (n)	CAT (n)	Correct classification (%)	Sensitivity	Specificity
CS	19	18.7 ± 0.5	0.3 ± 0.5	98.0 ± 3.0		
CAT	15	0.0 ± 0.0	15.0 ± 0.0	100.0 ± 0.0		
Total	34			99.0 ± 1.0	0.98 ± 0.03 (0.81—1.0)*	1.00 ± 0.00 (0.80—1.0)*

*Mean of the Wilson score intervals calculated for the sensitivity and specificity of each model.
Model parameters: mean values obtained with the training sets (N = 136) from 7 iterations: 5 LVs, Q^2 = 0.907, RMSEcv = 0.158, ANOVA p-value < 0.05. CAT: Catalonia; CS: Central Spain.

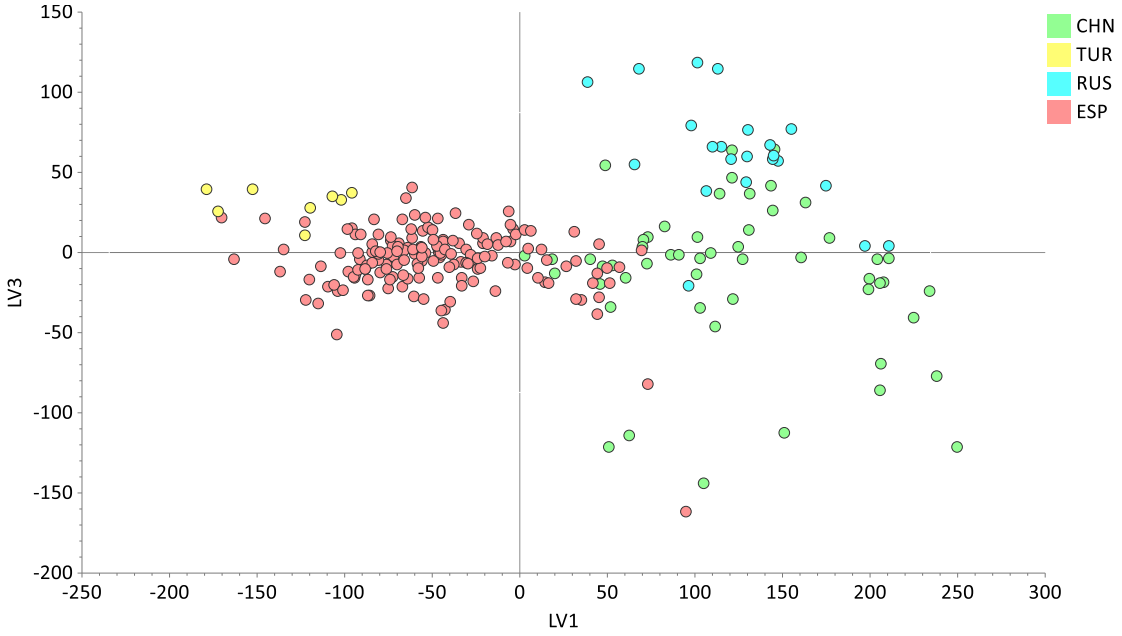

Fig. 1.

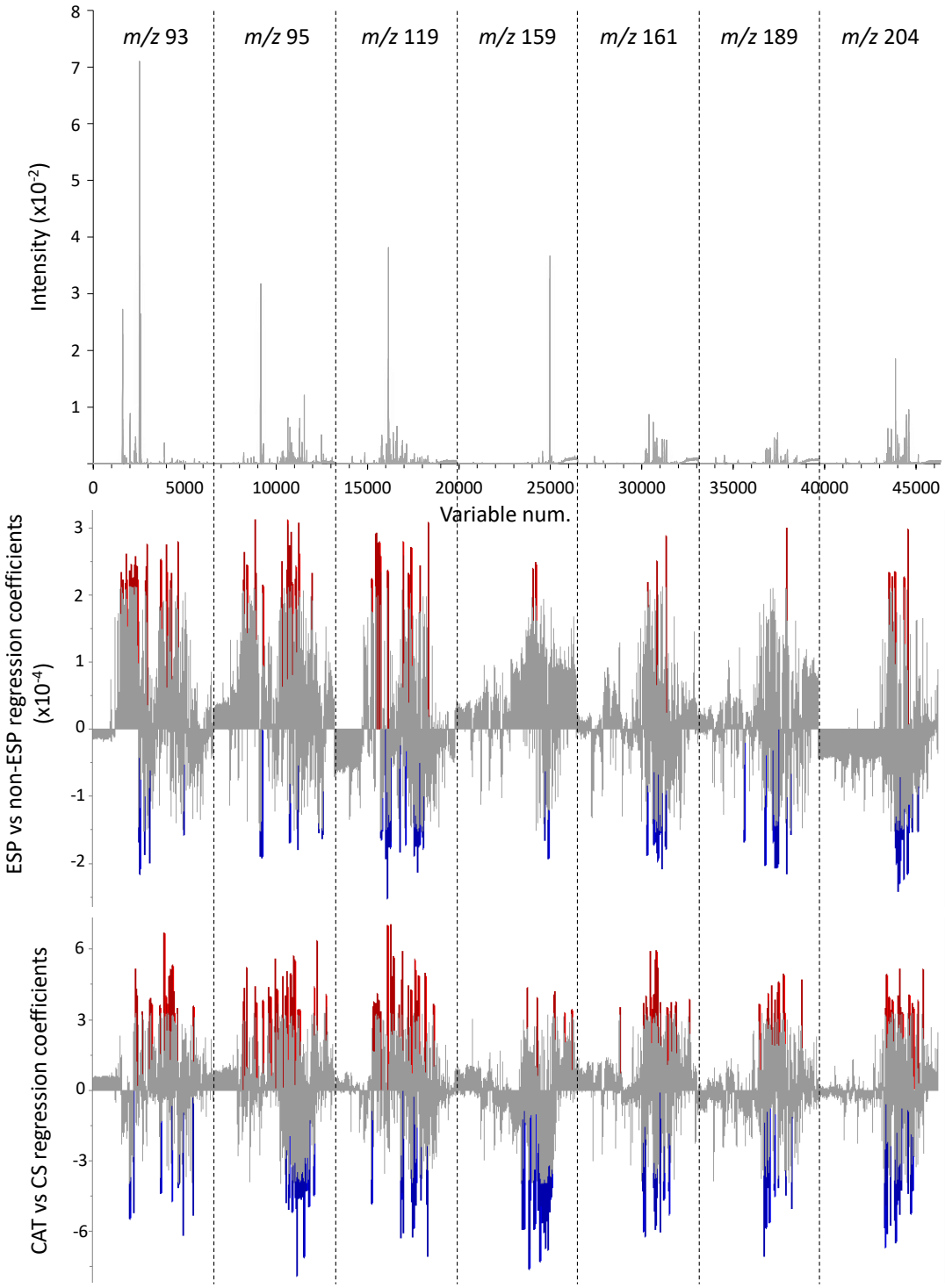

Fig. 2.

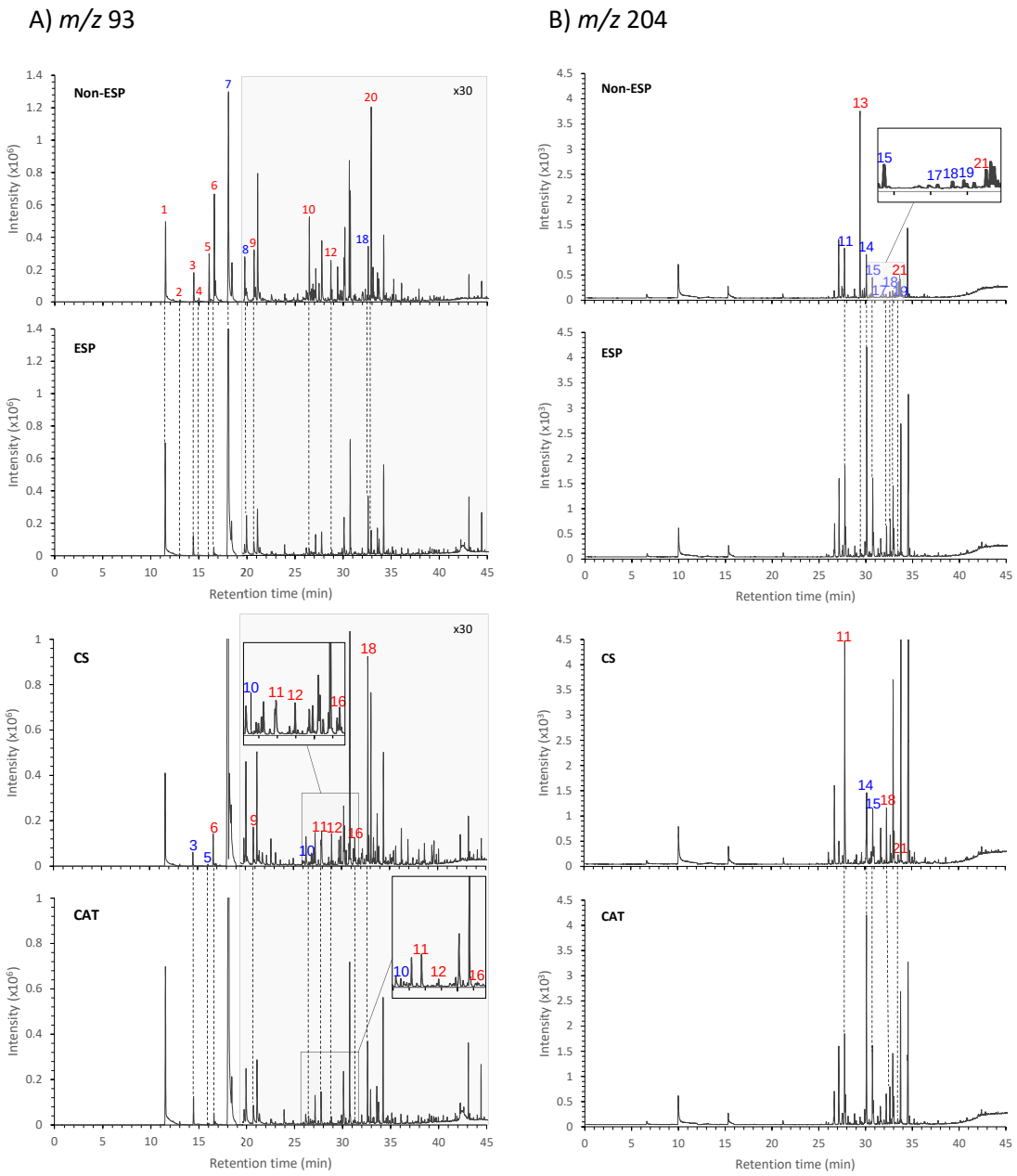

Fig. 3.