Food Control 152 (2023) 109834

Contents lists available at ScienceDirect

Food Control

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

Liquid chromatography coupled to high-resolution mass spectrometry for nut classification and marker identification

Guillem Campmajó [a,b,*], Javier Saurina [a,b], Oscar Núñez [a,b,c]

[a] Department of Chemical Engineering and Analytical Chemistry, Universitat de Barcelona, Martí i Franquès 1-11, E08028, Barcelona, Spain
[b] Research Institute in Food Nutrition and Food Safety, Universitat de Barcelona, Av. Prat de la Riba 171, Edifici Recerca (Gaudí), E08921, Santa Coloma de Gramenet, Spain
[c] Serra Húnter Fellow Programme, Generalitat de Catalunya, Via Laietana 2, E08003, Barcelona, Spain

ARTICLE INFO

Keywords:
Nuts
Non-targeted
LC–HRMS
Chemometrics
Adulteration
Food authentication

ABSTRACT

Fraud in nut and seed products poses an economic deception and a threat to human health because of their allergens. This study comprehensively evaluated the metabolomic diversity of ten different nut types through non-targeted liquid chromatography coupled to high-resolution mass spectrometry (LC–HRMS). First, LC–HRMS fingerprints were subjected to partial least squares regression-discriminant analysis (PLS-DA), and the developed multi-class model reached a classification accuracy of 100% after external validation. Then, variable importance in projection (VIP) scores obtained from two-input class PLS-DA models (*i.e.*, a specific nut type against all the other samples) allowed the selection of 136 discriminant compounds that were tentatively annotated/identified through HRMS data. Finally, as a case of study, successful detection and quantitation of almond-based products adulteration (with hazelnut or peanut) was achieved through a targeted LC–HRMS study, using some of the found markers and partial least squares (PLS) regression. In this context, new profiling approaches could be further implemented based on the reported markers using cheaper techniques.

1. Introduction

Analytical strategies based on *omics* approaches —*genomics*, *proteomics*, *metabolomics*, and *metallomics/isotopolomics*— have been widely proposed to solve food authenticity control. In particular, *metabolomics*, which is the closest omics discipline to the phenotype of biological systems, focuses on the analysis of small molecules (<1500 Da) (Creydt & Fischer, 2018). In this context, the use of *metabolomics* non-targeted methods, where instrumental responses (*i.e.*, mainly analytical signals obtained through chromatography and related techniques, spectroscopy, mass spectrometry, or electronic sensors) are analysed without assuming any previous knowledge, has proved its potential in this field (Medina, Perestrelo, Silva, Pereira, & Câmara, 2019).

Liquid chromatography coupled to high-resolution mass spectrometry (LC–HRMS) is a reliable instrumental platform to perform non-targeted analysis with high molecular coverage of the non-volatile metabolome. Using time-of-flight (TOF) or Orbitrap instruments, HRMS provides high resolving power, allowing accurate *m/z* measurements. Furthermore, structural information can be obtained through fragmentation data when using hybrid configurations like quadrupole-Orbitrap (Q-Orbitrap) or quadrupole-time-of-flight (Q-TOF). Besides, the hyphenation of LC with HRMS enhances both the selectivity and sensitivity of the analytical approach. Therefore, because of these instrumental capabilities and despite its higher cost and longer analysis time, it is usually preferred over spectroscopic techniques for tentative compound identification. In this line, LC–HRMS has been widely proposed in diverse applications such as the screening of chemical contaminants in food (Fu, Zhao, Lu, & Xu, 2017) or human biomonitoring (Caballero-Casero et al., 2021), the characterisation of natural plants (Alvarez-Rivera, Ballesteros-Vivas, Parada-Alfonso, Ibañez, & Cifuentes, 2019), or clinical research (Rochat, 2016). Particularly in the food fraud field, it has also been used to investigate markers related to specific authentication issues (Lacalle-Bergeron et al., 2021; Zhong et al., 2022).

Nuts and seeds are usually consumed as a snack, although they can also be added to salads, sausages, stews, or bakery products. It is well-known that their regular intake promotes beneficial health effects on humans (Bitok & Sabaté, 2018). However, according to the Food Fraud Risk Information database (Food Fraud Advisors, 2017), some nut-based products are at medium or high risk for fraud practices, such as adulterations or replacements with cheaper and lower-quality ingredients. In

* Corresponding author. Department of Chemical Engineering and Analytical Chemistry, University of Barcelona, Martí i Franquès 1-11, E08028, Barcelona, Spain.
E-mail address: gcampmajo@ub.edu (G. Campmajó).

https://doi.org/10.1016/j.foodcont.2023.109834
Received 10 March 2023; Received in revised form 21 April 2023; Accepted 29 April 2023
Available online 29 April 2023

this context, while fraud in raw nuts is unusual due to evident visual differences, it is more common in processed nut products such as flours or pastes, where its detection is more complicated. Moreover, fraud in these products implies an economic deception and a threat to human health because of their allergens (Luparelli et al., 2022), being a food authentication and safety issue. Thus, developing analytical methodologies to detect these practices is required.

Several approaches and analytical techniques have been proposed to detect nut species adulteration. Although some of the methods described in the literature rely on *genomics*, *proteomics*, or *metallomics* (Ding et al., 2020; Esteki, Vander Heyden, Farajmand, & Kolahderazi, 2017; Monaci, De Angelis, Bavaro, & Pilolli, 2015), most focus has been on *metabolomics*. In this line, spectroscopic fingerprinting approaches are commonly used for this aim. For instance, Taylan et al. employed Raman spectroscopy to detect green-pea adulteration in pistachio (Taylan et al., 2021), while Rovira et al. evaluated two infrared spectroscopic techniques —near-infrared (NIR) and attenuated total reflection-Fourier transform infrared (ATR-FTIR)— to assess cashew nut authenticity in front of four different adulterants (Rovira et al., 2022).

Instead, to date, most of the targeted and non-targeted methods based on chromatographic and related techniques have focused on the authentication of the cultivar or geographical origin of nut samples (Campmajó & Núñez, 2021; Suman, Cavanna, Sammarco, Lambertini, & Loffi, 2021). Indeed, to our knowledge, few studies using these techniques have dealt with nut species adulteration and classification. For instance, almond powder adulteration with apricot kernel was evaluated using the fatty acid profile obtained with gas chromatography with flame ionisation detection (GC-FID) (Esteki, Farajmand, Kolahderazi, & Simal-Gandara, 2017), while liquid chromatography with ultraviolet (LC-UV) and fluorescent detection (LC-FLD) fingerprinting assessed nut type classification (Campmajó et al., 2019; Campmajó, Saez-Vigo, Saurina, & Núñez, 2020). In the latter case, almond-based product adulteration was also studied. Moreover, only the adulteration of pistachio nut powder with green-pea has been evaluated by non-targeted LC−HRMS (Çavuş, Us, & Güzelsoy, 2018).

Therefore, in this study, non-targeted LC−HRMS was used to address the classification of ten nut-type samples (almonds, cashew nuts, hazelnuts, macadamia nuts, peanuts, pine nuts, pistachios, pumpkin seeds, sunflower seeds, and walnuts) through partial least squares regression-discriminant analysis (PLS-DA). Moreover, two-input class PLS-DA models (*i.e.*, a specific nut type in front of others) were built, and the evaluation of the corresponding variable importance in projection (VIP) scores allowed finding the most discriminant molecular features for each nut. After tentatively identifying these markers, targeted LC−HRMS (focusing on the found discriminant molecules) was proposed to detect and quantitate hazelnut and peanut adulterations of almond-based products by partial least squares (PLS) regression.

2. Materials and methods

2.1. Reagents and solutions

Purified water obtained with Elix® 3 coupled to a Milli-Q® system (Millipore Corporation, Bedford, MA, USA) and filtered through a 0.22-μm nylon membrane, and acetone and hexane purchased from Merck (Darmstadt, Germany), were used for the sample treatment. Regarding the LC−HRMS analysis, LC−MS grade water and methanol, as well as formic acid (96%), were provided by Merck.

The analytical reagent grade compound standards used for confirmation were: (−)-epicatechin and citric, malic, quinic, and tartaric acids from Merck, and (+)-catechin from Fluka (Steinheim, Germany).

2.2. Instrumentation

The chromatographic system consisted of an ultra-high-performance liquid chromatography (UHPLC) system equipped with an Accela 1250 quaternary pump and an Accela autosampler (Thermo Fisher Scientific, San Jose, CA, USA). The chromatographic separation was performed using a core-shell technology Kinetex C_{18} column (100 mm × 4.6 mm id., 2.6 μm particle size) and guard column (2 mm × 4.6 mm id., 2.6 μm particle size), from Phenomenex (Torrance, CA, USA), and using 0.1% (*v/v*) formic acid aqueous solution (solvent A) and methanol (solvent B) as the constituents of the mobile phase. Hence, the developed chromatographic method started with linear gradient elution from 5 to 75% solvent B in 30 min, continued with a 2.5 min-lineal increase up to 95%, and ended with an isocratic step at 95% for 2.5 min. Finally, 0.1 min-lineal decrease back to the initial conditions and 4.9 min of isocratic elution for column re-equilibration were set. The mobile phase flow rate was 400 μL min^{-1}, and the injection volume 10 μL (partial loop mode).

The UHPLC system was coupled to a hybrid Q-Orbitrap mass spectrometer (Q-Exactive Orbitrap, Thermo Fisher Scientific) equipped with a heated electrospray ionisation (H-ESI II) source operating in the negative ion mode. Nitrogen with a purity of 99.98%, purchased from Linde (Barcelona, Spain), was used for the ESI sheath, sweep, and auxiliary gas at flow rates of 60, 0, and 10 a.u. (arbitrary units), respectively. Other H-ESI parameters were established as follows: spray voltage, - 2.5 kV; probe heater temperature, 350 °C; capillary temperature, 320 °C; and S-lens RF level at 50 V. Full-scan HRMS data were acquired over an *m/z* range of 100–1500 at a mass resolution of 70,000 full width at half maximum (FWHM) at *m/z* 200. In addition, an automatic gain control (AGC) of 1.0×10^6 and a maximum injection time (IT) of 200 ms were established.

For the MS/HRMS experiments, targeted data-dependent scan mode, requiring accurate mass inclusion lists, was used to obtain the product ion scans of specific ions of interest. The acquisition was performed at a mass resolution of 17,500 FWHM at *m/z* 200. Precursor ions, isolated by the quadrupole with an isolation window of 0.5 *m/z*, were fragmented in the higher-energy collisional dissociation (HCD) cell using three-stepped normalised collision energies (NCE) ranging from 10 to 50%. Moreover, the targeted data-dependent acquisition was subordinated to an intensity threshold fixed at 1.0×10^5, and the AGC and IT values were established at 2.0×10^5 and 200 ms, respectively.

The Q-Orbitrap system was tuned and calibrated every three days, using commercially available calibration solutions for both negative and positive ion modes (Thermo Fisher Scientific). Moreover, the Xcalibur software v 4.1 (Thermo Fisher Scientific) was used to control the LC−HRMS system and acquire and process data.

2.3. Samples

2.3.1. Nut samples for the classification study

A set of 149 raw nut samples bought in Spanish commercial supermarkets were analysed for classification purposes. Samples encompassed various nut classes −30 almonds, 10 cashew nuts, 20 hazelnuts, 10 macadamia nuts, 20 peanuts, 10 pine nuts, 10 pistachios, 20 pumpkin seeds, 9 sunflower seeds, and 10 walnuts—, some of them processed with different thermal treatments —natural, fried, and toasted— (see Table S1 for nut sample details). Before the sample treatment, samples were crushed and homogenised. Moreover, a quality control (QC) sample consisting of a mix prepared by pooling 50 μL of each sample extract was employed.

2.3.2. Adulterated almond samples for the quantitative study

Two different almond adulteration scenarios (almond *vs.* hazelnut and almond *vs.* peanut) were considered to evaluate the suitability of the identified biomarkers to address its authentication in two matrices: natural almond flour and homemade almond custard cream. Both matrices were obtained from a random almond sample previously used in the classificatory study. Thus, the sample was crushed and homogenised to obtain the almond flour (as done in Section 2.3.1), whereas the custard cream was made from hen eggs, milk, sugar, and corn flour.

Thus, the adulterants were added in different proportions −0, 20, 40,

60, 80, and 100% for calibration; and 15, 25, 50, 75, and 85% for external validation— to both almond-based products. Five replicates of each blend were prepared, giving 55 samples for each case (adulteration scenario and matrix). Besides, a 50% adulterated sample was used in this case as the corresponding QC sample.

2.4. Sample treatment and analysis

A previously developed two-step sample treatment for phytochemical extraction from nut samples was carried out (Campmajó et al., 2019), consisting of ultrasound-assisted solid-liquid extraction (USLE) with acetone:water (70:30, *v/v*) followed by a defatting step with hexane.

Regarding the analysis procedure, samples were randomly injected along each sequence to minimise the influence of any instrumental drift in the chemometric results. Moreover, an extracting solvent blank and a QC sample were injected at the beginning and after every ten sample injections to control cross-contamination and avoid systematic errors.

2.5. Data treatment

LC−HRMS generates massive datasets, requiring software programs to properly reduce and handle the obtained data. Thus, aiming to obtain a matrix consisting of ion peak area values as a function of *m/z* and retention times, the application of MSConvertGUI from the ProteoWizard Toolkit (Chambers et al., 2012) and mzMine 2.53 (Pluskal, Castillo, Villar-Briones, & Orešič, 2010) software was required.

2.5.1. Non-targeted approach

A non-targeted approach was proposed for nut classification. First, raw data were reduced by establishing an absolute intensity threshold peak filter of 5.0×10^5 and transformed to mzXML format through MSConvertGUI. Then, the resulting LC−HRMS data were submitted to the mzMine 2.53 software for peak detection (exact mass detection, chromatogram detection, and chromatogram deconvolution), isotopic peak grouper, and retention time alignment. Briefly, exact mass lists for each scan in a sample were generated, establishing a noise level of 5.0×10^5. Afterwards, the ADAP chromatogram builder allowed the joining of the exact masses found in contiguous scans in a sample that fulfilled the following conditions: peak time range of 0–35 min, 15 minimum scans above an intensity threshold set at 1.0×10^6, and an *m/z* tolerance of 5 ppm. Next, individual chromatographic peaks were achieved through chromatogram deconvolution. Thus, for this purpose, the baseline cut-off algorithm was selected for peak recognition with a baseline level of 5.0×10^5, a minimum peak height of 1.0×10^6, and a peak duration range of 0.1–1 min. Subsequently, isotope removal was carried out considering that the most representative isotope was the most intense and setting an *m/z* and retention time tolerance of 5 ppm and 0.3 min, respectively. Finally, retention time alignment was carried out using the random sample consensus (RANSAC) peak list aligner method, following the following requisites: retention time tolerance before and after correction of 2.2 and 1.2 min, *m/z* tolerance of 5 ppm, 10^5 maximum RANSAC iterations to find the suitable model, and 80% as the minimum value to consider the model valid.

At the end of this workflow, a data matrix was exported to an Excel File containing the obtained LC−HRMS fingerprints: samples × variables. The 'samples' column included the 149 nut samples and the 16 QC samples, whereas the 'variables' row comprised all the detected molecular features (an exact *m/z* value at a specific retention time). Only the molecular features detected at least in 80% of the samples belonging to a nut class were selected to reduce the matrix dimensions. As a result, a 165 × 278 dimension data matrix was obtained, containing the chromatographic peak areas for each molecular feature in all samples.

2.5.2. Targeted approach

A targeted approach was applied to detect and quantify the adulteration in almond-based products, focusing on the discriminant markers identified for almond, hazelnut, and peanut samples encountered in the classificatory study.

First, raw data were processed with MSConvertGUI, applying an absolute intensity threshold peak filter of 1.0×10^5. Then, targeted peak detection was performed with mzMine 2.53 software using a list of targeted molecular features (an exact *m/z* value at a specific retention time) for each almond adulteration scenario: 28 for almond *vs.* hazelnut and 35 for almond *vs.* peanut. Besides, a noise level of 1.0×10^5, an intensity tolerance of 10% (maximum allowed deviation from the expected shape of a chromatographic peak), an *m/z* tolerance of 5 ppm, and a retention time tolerance of 0.5 min were established. Finally, in this case, the join aligner allowed matching of the detected molecular features across samples, setting a mass tolerance of 5 ppm, a retention time tolerance of 0.5 min, 80% of weight for *m/z*, and 20% of weight for retention time. Again, at the end of the workflow, a data matrix was exported to an Excel File containing the obtained LC−HRMS profiles.

2.6. Chemometric and statistical analysis

The chemometric analysis by principal component analysis (PCA), PLS-DA, and PLS regression was carried out using Solo 8.6 chemometric software from Eigenvector Research (Manson, WA, USA). Details of their theoretical background are addressed elsewhere (Massart et al., 1997).

LC−HRMS data matrices (normalised and autoscaled) were used as X-data matrices indistinctly of the chemometric method used, which depended on the aim of the study. In this line, PCA assessed a first exploratory analysis to check the absence of systematic errors through QC sample behaviour and allowed visualising sample trends. PLS-DA, particularly PLS1-DA (Brereton & Lloyd, 2014), was used in the classificatory study, requiring a Y-data matrix that defined the nut type of each sample. PLS was employed in the quantitation of almond-based product adulteration, demanding a Y-data matrix that expounded sample adulteration percentages. Moreover, the proper number of latent variables (LVs) for building PLS-DA and PLS models was selected, after Venetian blinds cross-validation (CV), at the first minimum of the cross-validation classification error (CVCE) and the root-mean-square error of cross-validation (RMSECV), respectively.

PLS-DA and PLS models' performance was checked by external validation. On the one hand, for the classificatory study, samples were stratified and randomly chosen: 60% were used as the calibration set, whereas the remaining 40% as the external validation set. Then, overall accuracy and each class sensitivity (capability to detect true positives) and specificity (capability to detect true negatives) were used to evaluate the classification models. While the former is calculated by dividing the number of well-classified samples by the total number of samples, class sensitivity (Eq. (1)) and specificity (Eq. (2)) are calculated as follows (Riedl, Esslinger, & Fauhl-Hassek, 2015):

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{1}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{2}$$

where TP is true positive samples, TN is true negative samples, FP is false positive samples, and FN is false negative samples.

On the other hand, PLS regression was done using some adulteration percentages for calibration and others for external validation, as detailed in Section 2.3.2. The model performance was evaluated through the root-mean-square error of calibration (RMSEC), RMSECV, and prediction (RMSEP), as well as the corresponding R^2 (determination coefficient) values. Eq. (3) shows how RMSEs are calculated. Moreover, relative error in each external validation to estimate the adulterant percentage was also assessed.

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(\hat{y}_i - y_i)^2}{n}} \quad (3)$$

where $\hat{y}_i$ is the predicted value, y_i is the actual value, n is the number of samples, and N is the number of predictions.

Finally, to find the most discriminant molecular features for each nut class, individual PLS-DA models of a specific nut class against all the others were built. Then, after external validation of the models, variables with the highest VIP scores were selected for further annotation and identification steps. Regression vector coefficients were also evaluated, with positive values indicating that variable contribution is related to the target class. Moreover, the significance of the differences in their peak area values between nut classes was evaluated statistically. Thus, after a Fisher test of variances, the student *t*-test for comparing the means of two classes was carried out. A confidence level of 0.99 was assumed, so when *p* (probability) values were lower than 0.01, differences in the molecular feature peak areas between the classes were considered significant.

2.7. Annotation and identification of the most discriminant compounds

The most discriminant molecular features for each nut class, selected through VIP scores, were putatively identified following Schymanski et al. HRMS identification levels (Schymanski et al., 2014). The established parameters to assess this identification step were: 5 ppm of exact mass tolerance, >85% of isotopic pattern fit, MS^2 data similarity, and retention time agreement. For the MS^2 comparison, public databases such as mzCloud (HighChem LLC, Bratislava, Slovakia), The Human Metabolome Database (Wishart et al., 2018), and LIPID MAPS Structure Database (Sud et al., 2006) were employed. Besides, Phenol-Explorer (Rothwell et al., 2013), a database including polyphenolic content in food, was also consulted. Finally, in some specific cases, MetFrag software (Wolf, Schmidt, Müller-Hannemann, & Neumann, 2010) was also used for tentative in-silico elucidation.

3. Results and discussion

3.1. Non-targeted LC–HRMS nut classification

As previously mentioned, the present study aimed to develop a non-targeted LC–HRMS method to classify nut samples according to their type and identify the most discriminant molecular features. Despite the non-targeted nature of the developed method, instrumental conditions were oriented to favour phenolic and polyphenolic compound detection since they have already been successfully proposed as potential markers in several food authentication issues (Lucci, Saurina, & Núñez, 2017; Proestos & Pesic, 2022). Thus, a total of 149 nut samples belonging to 10

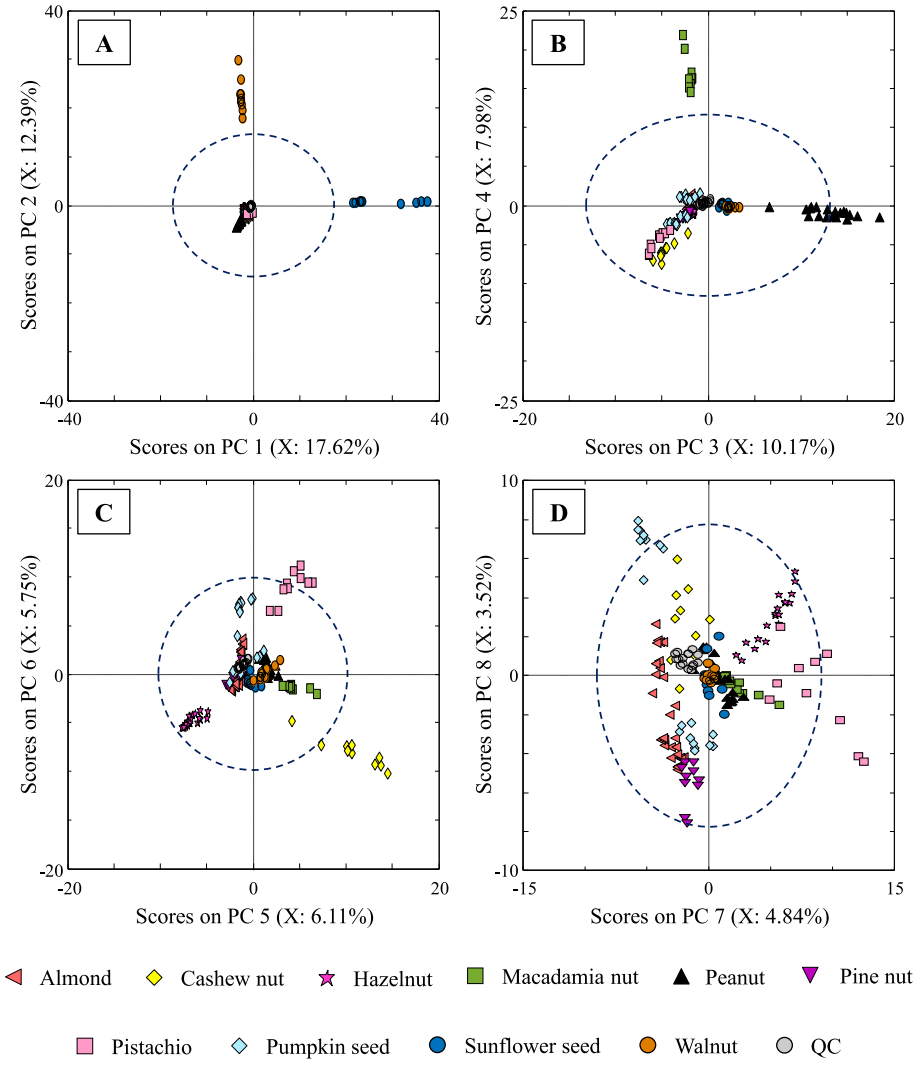

Fig. 1. PCA scores plot obtained for the analysed nut samples according to their type, using the non-targeted LC–HRMS data, of (A) PC1 *vs.* PC2, (B) PC3 *vs.* PC4, (C) PC5 *vs.* PC6, and (D) PC7 *vs.* PC8.

different nut classes were analysed following the proposed method. As an example, Fig. S1 depicts the total ion current (TIC) LC−HRMS chromatogram for a selected sample within each nut type. In this context, remarkable qualitative differences regarding peak distribution and signal intensity can be visually detected.

LC−HRMS data were subjected to PCA to appraise their discriminating capability. However, PCA was first employed to select the most appropriate data treatment, which is crucial for subsequent unequivocal results. In this case, normalisation (scaling each sample to the sum of the corresponding peak areas) and autoscaling (mean centring and variable scaling to unit standard deviation) were assessed to try to improve the data quality. As a result, it was found that performing a normalisation step before autoscaling provided a better sample grouping, reducing the effect of the HRMS detection variance.

In this context, Fig. 1 shows the PCA scores plot obtained after applying this data pretreatment and using the 165 × 278 dimension data matrix containing both nut and QC samples. A total of eight principal components (PCs), describing 68.37% of the variance, were chosen for the PCA analysis. As a result, the non-supervised chemometric plots showed QC samples grouped in the centre, indicating the lack of

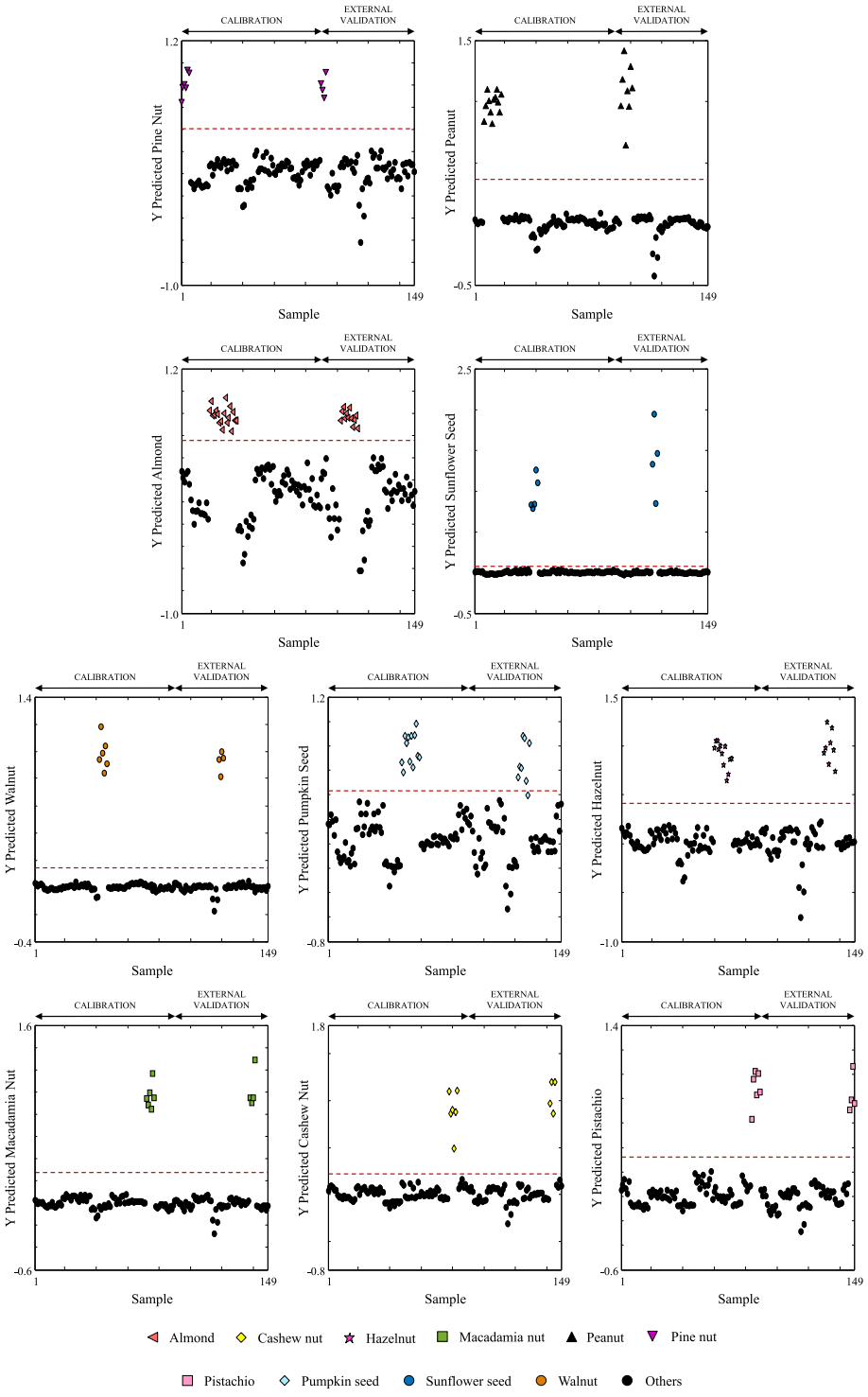

Fig. 2. Sample *vs.* Y Predicted plot for the two-input class PLS-DA models after external validation.

systematic errors affecting the reliability of the results. Moreover, the complexity of the studied issue —encompassing a significant number of sample classes (nut type) and factors (different geographical origins and thermal treatments)— was reflected in the low variance explained by the PCs (*e.g.*, 17.62% for PC1). Nevertheless, good sample distinction was achieved for almost all nut types. For instance, in Fig. 1A, where the plot of scores of PC1 *vs*. PC2 is depicted, sunflower seed and walnut samples were visibly separated through PC1 and PC2, respectively. PC3 and PC4 were highly related to peanut and macadamia nut samples, respectively (see Fig. 1B). Besides, the plot of scores of PC5 *vs*. PC6 (Fig. 1C) allowed the discrimination of cashew nut (displaying positive PC5 and negative PC6 values), hazelnut (displaying negative PC5 and PC6 values), and pistachio (displaying positive PC5 and PC6 values) samples. Finally, the scatter plot for scores of PC7 and PC8 (Fig. 1D) allowed a slight discrimination of pine nut and almond samples along the PC8, presenting negative PC7 values.

Given the excellent results observed in the PCA, with a remarkable separation of samples according to nut classes (except for pumpkin seed samples), PLS-DA was applied to the non-targeted LC−HRMS data. For this, QC samples were removed from the dataset, and a 149 × 278 dimension data matrix was subjected to the supervised classificatory analysis. In this case, nine LVs were selected to build the PLS-DA model, which described 71.35% and 95.53% of X-variance and Y-variance, respectively. As a result, visual sample classification was reached for all nut types investigated.

Therefore, to evaluate the classificatory ability of the non-targeted LC−HRMS data through PLS-DA, external validation was performed as described in Section 2.6. In this line, a PLS-DA calibration model, built with 60% of the analysed samples, was composed of nine LVs explaining 73.53% of X-variance and 95.62% of Y-variance. CV results —sensitivities of 100%, specificities above 97.6%, and classification accuracies above 98.8%, for each nut class under study— anticipated the excellent results obtained in the external validation —sensitivities, specificities, and classification accuracies of 100%, for each nut class under study—, proving the excellent discriminant capacity of the non-targeted LC−HRMS data. The external validation graphical results (Sample *vs*. Y predicted score plot) for each analysed nut type are shown in Fig. S2.

3.2. Annotation and identification of nut type markers

As previously mentioned, one of this study's main goals was to identify characteristic discriminant molecular features for each studied nut type. Thus, with this purpose, two-input class PLS-DA models were built: the first input corresponded to a specific nut class, while the second encompassed all the others. In this context, as shown in Fig. 2, the performance of each binary PLS-DA model was assessed through external validation, obtaining complete sample classification in all the cases, except for the Pumpkin seed *vs*. Others model, where 99.3% classification accuracy, 87.5% sensitivity, and 100% specificity were obtained.

Then, in each PLS-DA model, VIP loadings scores allowed the selection of the most discriminant molecular features for each nut under study. For instance, Fig. S3 presents the results obtained for the PLS-DA model of Walnut *vs*. Others: the classification plot depicting Sample *vs*. Y Predicted Walnut and the corresponding VIP scores plot. To obtain MS/HRMS data of the selected discriminant molecular features, an arbitrary sample of each nut type was analysed by LC−HRMS using a targeted data-dependent acquisition method, built with an inclusion list containing them.

Table S2 summarises the tentative annotation and identification of the discriminant markers found for each nut type. Most of the compounds were detected in their deprotonated form $[M-H]^-$, as expected considering that the HRMS acquisition was performed in the negative mode, although in some cases, their adduct with formic acid $[M+FA-H]^-$ or chlorine $[M+Cl]^-$, or even their deprotonated dimeric form $[2M-H]^-$ corresponded to the base peak. Some of the annotated/identified compounds are discussed below since some of them had been previously reported in the literature.

In the case of almond discriminant compounds, amygdalin —a cyanogenic diglucoside responsible for the bitterness of almonds— and amygdaloside were found (Lee, Zhang, Wood, Rogel Castillo, & Mitchell, 2013; Sang et al., 2003). Furthermore, sugars and derivatives such as *m/z* 341.1083 and 683.2243, annotated as disaccharide and tetrasaccharide + H_2O, respectively, were also detected (Gil Solsona, Boix, Ibáñez, & Sancho, 2018; Huang, Robinson, Dias, de Moura Bell, & Barile, 2022). Besides, Gil-Solsona et al. previously identified the first, annotating it as inulobiose, as a discriminant marker related to the Spanish almond variety.

Among the molecular features presenting high VIP scores in the cashew nut PLS-DA classification, the isomers with the molecular formula $C_{15}H_{14}O_6$, observed at the retention times of 12.80 and 15.50 min, were identified as the flavanols (+)-catechin and (−)-epicatechin that have been previously seen in cashew nut testa (Trox et al., 2011).

Several indoleacetic acid glycoside isomers were found to be discriminant markers for hazelnut classification. In this line, *m/z* 541.1458 (at a retention time of 20.91 min) and 540.1719 (at a retention time of 21.05 min) were assigned as isomers of 2-(3-hydroxy-2-oxoindolin-3-yl) acetic acid 3-O-6′-galactopyranosyl-2′′-(2″oxoindolin-3″yl) and hazelnutin D, respectively, which have been previously detected in hazelnut kernel (Shataer et al., 2021). Besides, two other indoleacetic acid glycoside chiral isomers with *m/z* 368.0984 —named 3-(O-β-D-glycosyl)dioxindole-3-acetic acid and hazelnutin E by Singldinger et al. (Singldinger et al., 2018) and Shataer et al. (Shataer et al., 2021), respectively— were found to be discriminant. These compounds presented chromatographic peaks at 7.31 and 8.39 min, although their retention time assignment was not possible.

Regarding macadamia nut markers, among others, various phenolic and polyphenolic compounds such as phenolic acid derivatives (hydroxybenzoic acid glucoside and apiosylglucosyl 4-hydroxybenzoate isomers), guaiacol hexose-pentose isomers, and oleoside dimethyl ester were tentatively identified.

In the case of peanut, several hydroxycinnamic acids (*i.e.*, *cis*- and *trans*-*p*-coumaroyl tartaric acids, feruloyl tartaric acid isomer, *p*-coumaric acid, coumaroyl-O-pentoside isomer, di-*p*-coumaroyl tartaric acid isomer, and *p*-coumaroylferuloyl tartaric acid isomer) and a derivative (such as *p*-coumaroylnicotinoyl tartaric acid), as well as an hydroxybenzoic acid isomer, were detected agreeing with literature and considered as discriminant (Ma et al., 2014).

In relation to pine nut molecular features, ascorbalamic acid isomer ($C_9H_{13}NO_8$) and vanillic acid glucoside isomers ($C_{14}H_{18}O_9$) were annotated as relevant for pine nut classification.

Moreover, accordingly to previous studies (Erşan, Güçlü-Üstündağ, Carle, & Schweiggert, 2016), protocatechuic acid and quercetin 3-O-glucoside were observed in their deprotonated form at 10.44 and 23.03 min, respectively, providing high VIP scores for pistachio classification. Organic acids, such as malic and isocitric acid, and nucleotides, as uridine monophosphate and adenosine 5′-monophosphate, also appeared to be discriminant for pistachio.

In the case of pumpkin seed, sugars —a trisaccharide in its $[M+Cl]^-$ form and sedoheptulose—, guanosine, and tyrosol diglycoside isomer, were tentatively identified, among others.

Finally, regarding sunflower seed discriminant compounds, as reported by Romani et al. (Romani, Pinelli, Moschini, & Heimler, 2017), several hydroxycinnamic acids were found and related to sunflower seed classification: 4-caffeoylquinic acid, 5-caffeoylquinic acid, 3,4-dicaffeoylquinic acid, and 4,5-dicaffeoylquinic acid. Besides, other phenolic compounds, such as flavonoid O-glycoside isomers and phenylacetic acid, were also annotated.

Instead, among the discriminant compounds presenting high VIP scores for walnut classification, several hydroxybenzoic acids and derivatives —galloyl-hexahydroxydiphenoyl-glucose isomer, digalloyl-

Table 1
Calibration, cross-validation, and external validation results obtained for each of the PLS regression models used to determine the almond products adulteration percentage.

Adulteration case	Matrix	Data matrix (samples × variables)	CALIBRATION			CROSS-VALIDATION		EXTERNAL VALIDATION						
			LVs	RMSEC (%)	R^2	RMSECV (%)	R^2	RMSEP (%)	R^2	Relative error in each adulterant percentage (%)				
										15%	25%	50%	75%	85%
Almond *vs.* Hazelnut	Flour	55 × 28	1	5.994	0.969	6.304	0.966	5.037	0.977	6.06	4.93	2.70	3.54	1.88
	Cream	55 × 28	2	7.773	0.946	10.154	0.909	8.887	0.933	6.07	7.14	9.24	7.77	7.55
Almond *vs.* Peanut	Flour	55 × 35	2	3.865	0.986	6.665	0.960	8.202	0.950	4.30	0.97	8.62	10.86	5.86
	Cream	55 × 35	3	5.237	0.977	8.648	0.937	9.852	0.893	12.89	8.54	8.15	8.20	5.61

LV: latent variable; R^2: determination coefficient; RMSEC: root-mean-square error of calibration; RMSECV: root-mean-square error of cross-validation; RMSEP: root-mean-square error of prediction.

hexahydroxydiphenoyl-glucose isomer, ellagic acid, and ellagic acid pentoside— and a saccharolipid as glansreginin A were found, in accordance to Regueiro et al. (Regueiro et al., 2014).

It should be mentioned that for a molecular feature to be discriminant, it is not necessary to be detected or not detected in an exclusive nut type. Sometimes it presents a higher or lower content compared to other matrices. Moreover, considering the complexity of the studied issue, where ten nut types were studied, some markers were discriminant just in front of some others. Therefore, to better understand the significance of the differences of each discriminant marker content in their corresponding nut type compared to each other, student *t*-tests (preceded by a Fisher test of variances) were performed. Thus, Table S3 summarises the obtained results. As an example, focusing on hazelnut markers, *m/z* 131.0462 (at a retention time of 2.27 min), tentatively identified as asparagine, showed lower content in hazelnut than in almond, cashew nut, macadamia nut, and peanut, whereas no differences were observed with the remaining types. Instead, *m/z* 368.0984 (at a retention time of 7.31 min), tentatively identified as a 3-(O-β-D-glycosyl)dioxindole-3-acetic acid isomer, presented higher values in hazelnut than in any other nut.

In addition, since some of the analysed nut types presented different processing treatments —natural, fried, and toasted for almond; natural and toasted for hazelnut and pumpkin seed; and fried and toasted for peanut—, differences in the identified markers because of the processing were studied. In this context, Table S4 presents the results obtained after performing univariate statistical analysis to evaluate the significance of differences. In general, differences were not significant except for some given cases (*e.g.*, quinic acid content was significantly higher in toasted almonds than natural and fried ones). Besides, differences between nut matrices prevailed over processing treatment ones.

3.3. Detection and quantitation of almond-based product adulterations through targeted LC−HRMS

To validate the applicability of the identified molecular features as a discriminant profile for nut authentication, the adulteration of almond-based products (natural almond flour and homemade almond custard cream) was evaluated through PLS regression. These products are at medium risk of adulteration with cheaper nuts (Food Fraud Advisors, 2017). Therefore, hazelnut and peanut were chosen as adulterants due to the difficulty of visually detecting them (*i.e.*, mainly due to physical similarities such as granulometry or colour) in the studied matrices and their lower price. Furthermore, peanut was especially selected because of their serious threat to food safety (*i.e.*, they can cause severe allergy episodes).

Hence, as detailed in Section 2.3.2, different blend percentages were prepared and analysed following the developed LC−HRMS method. Afterwards, LC−HRMS data were processed using the targeted approach described in Section 2.5.2, focusing on the discriminant markers identified in the supervised study (Table S2): 12, 16, and 23 molecular features for almond, hazelnut, and peanut, respectively. Therefore, 28 molecular features were monitored for the almond *vs.* hazelnut adulteration scenario, while 35 were for the almond *vs.* peanut one (the corresponding feature lists file is provided in the Supplementary Material as a CSV file).

Before PLS regression analysis, PCA was performed using the

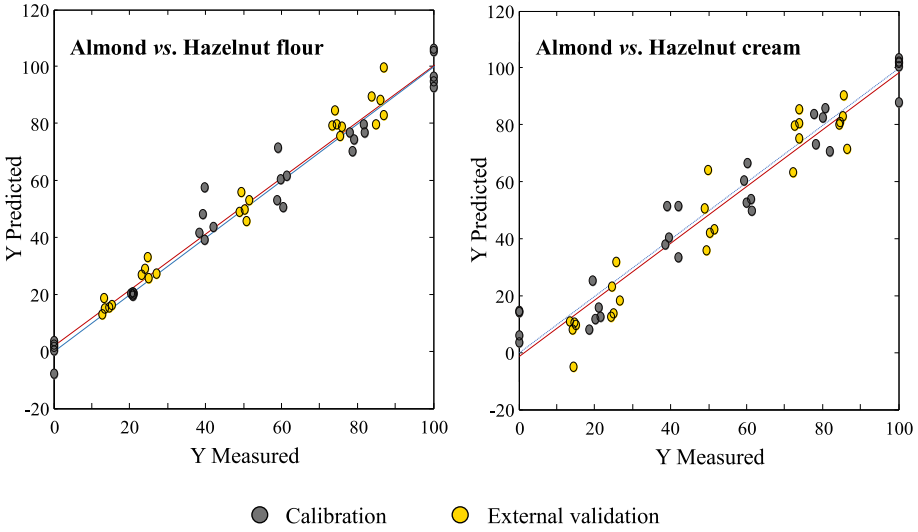

Fig. 3. External validation PLS results for the prediction of the percentage of adulteration of almond flour (on the left side) and almond custard cream (on the right side) with hazelnut. The blue line corresponds to the theoretical diagonal line, while the red line to the experimental adjusted one.

obtained targeted LC−HRMS data to observe non-supervised sample clustering and trends, as well as to check QC sample behaviour. For instance, Fig. S4 illustrates the scatter plot for scores of PC1 and PC2 (describing 70.8% of the variance) for the almond flour adulteration with peanut case. As observed, QC samples, which corresponded to a 50% adulterated sample, appeared in the centre of the plot, ensuring a good instrumental performance. Moreover, pure almond and peanut samples were distributed on opposite sides of the plot, displaying negative and positive PC1 scores, respectively. Intuitively, adulterated samples were ordered according to their adulterant percentage from the left (low adulterant percentages) to the right (high adulterant percentages) of the plot.

Then, PLS regression was applied to quantitate the adulteration level in each case under study. Thus, complementary to the X-data matrix, a Y-data matrix specifying the blend degree was required. Table 1 summarises the original data matrices used, the number of LVs employed to build each calibration PLS regression model, and the results obtained for the calibration, CV, and external validation. Good calibration models were built with low RMSEC ($\leq$7.773%) and $R^2 \geq 0.946$. Moreover, the similarity between calibration and CV parameter values indicated good internal consistency, preventing overfitting in the subsequent external validation. Regarding the external validation results, RMSEP and R^2 values ($\leq$9.852% and $\geq$0.893, respectively) indicated that the built PLS regression models showed a satisfactory ability to detect and quantitate almond adulterations. As an example, Fig. 3 shows the external validation PLS results for adulterating almond-based products with hazelnut (see Fig. S5 for the PLS results when adulterating with peanut). Results indicated that although more accurate quantitation was obtained of the almond flour matrix, no significant differences were observed between the studied matrices.

4. Conclusions

This study applied LC−HRMS, combined with chemometrics, to analyse nut product samples. In this line, 149 samples belonging to 10 nut types were analysed through non-targeted LC−HRMS aiming to find markers to prevent nut fraud (*e.g.*, adulteration, substitution, or replacement). PLS-DA allowed complete sample classification by the developed multi-class model (classification accuracy of 100% after external validation) and the identification of the most discriminant markers for each type of nut. In this regard, 136 molecular features were tentatively annotated/identified, taking benefit of the power of the MS/HRMS detection (*i.e.*, high sensitivity and selectivity, leading to good molecular coverage). For instance, organic acids, phenolic compounds, sugars, amino acids, and some derivatives were found among the compounds identified. Besides, although some of these markers' content varied due to thermal processing (*i.e.*, natural, toasted, or fried), differences between nut matrices prevailed. Moreover, to validate the use of the found markers for nut authentication, the adulteration of almond-based products (almond flour and homemade custard cream) with hazelnut or peanut was addressed. Thus, after targeted LC−HRMS analysis focusing on the corresponding markers, the obtained PLS results demonstrated their applicability to detect and quantitate the blend percentage.

Therefore, this study provides a set of nut and seed markers that could be further used in developing profiling approaches (for instance, using low-resolution mass spectrometers), which are more established in routine analysis, to detect adulteration in processed products.

Funding

This research was supported by the project PID2020-114401RB-C22 financed by the Agencia Estatal de Investigación (AEI/10.13039/501100011033) and by the Agency for Administration of University and Research Grants (Generalitat de Catalunya, Spain) under the project 2021SGR-00365.

CRediT authorship contribution statement

Guillem Campmajó: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Javier Saurina:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition. **Oscar Núñez:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

G. Campmajó thanks the University of Barcelona for the PhD APIF fellowship.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodcont.2023.109834.

References

Alvarez-Rivera, G., Ballesteros-Vivas, D., Parada-Alfonso, F., Ibañez, E., & Cifuentes, A. (2019). Recent applications of high resolution mass spectrometry for the characterization of plant natural products. *TrAC, Trends in Analytical Chemistry, 112*, 87–101. https://doi.org/10.1016/j.trac.2019.01.002

Bitok, E., & Sabaté, J. (2018). Nuts and cardiovascular disease. *Progress in Cardiovascular Diseases, 61*, 33–37. https://doi.org/10.1016/j.pcad.2018.05.003

Brereton, R. G., & Lloyd, G. R. (2014). Partial least squares discriminant analysis: Taking the magic away. *Journal of Chemometrics, 28*, 213–225. https://doi.org/10.1002/cem.2609

Caballero-Casero, N., Belova, L., Vervliet, P., Antignac, J. P., Castaño, A., Debrauwer, L., et al. (2021). Towards harmonised criteria in quality assurance and quality control of suspect and non-target LC-HRMS analytical workflows for screening of emerging contaminants in human biomonitoring. *TrAC, Trends in Analytical Chemistry, 136*, Article 116201. https://doi.org/10.1016/j.trac.2021.116201

Campmajó, G., Navarro, G. J., Núñez, N., Puignou, L., Saurina, J., & Núñez, O. (2019). Non-targeted HPLC-UV fingerprinting as chemical descriptors for the classification and authentication of nuts by multivariate chemometric methods. *Sensors, 19*, 1388. https://doi.org/10.3390/s19061388

Campmajó, G., & Núñez, O. (2021). Authentication of nut and seed products. In G. Campmajó, & O. Núñez (Eds.), *Chromatographic and related separation techniques in food integrity and authenticity, B* pp. 129–155). https://doi.org/10.1142/9781786349972_0005

Campmajó, G., Saez-Vigo, R., Saurina, J., & Núñez, O. (2020). High-performance liquid chromatography with fluorescence detection fingerprinting combined with chemometrics for nut classification and the detection and quantitation of almond-based product adulterations. *Food Control, 114*, Article 107265. https://doi.org/10.1016/j.foodcont.2020.107265

Çavuş, F., Us, M. F., & Güzelsoy, N. A. (2018). Assessing pistachio nut (Pistacia vera L.) adulteration with green pea (Pisum sativum L.) by untargeted liquid chromatography-(quadrupole-time of flight)-mass spectrometry method and chemometrics. *Journal of Food and Feed Science - Technology, 19*, 34–41.

Chambers, M. C., MacLean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., et al. (2012). A cross-platform toolkit for mass spectrometry and proteomics. *Nature Biotechnology, 30*, 918–920. https://doi.org/10.1038/nbt.2377

Creydt, M., & Fischer, M. (2018). Omics approaches for food authentication. *Electrophoresis, 39*, 1569–1581. https://doi.org/10.1002/elps.201800004

Ding, Y., Jiang, G., Huang, L., Chen, C., Sun, J., & Zhu, C. (2020). DNA barcoding coupled with high-resolution melting analysis for nut species and walnut milk beverage authentication. *Journal of the Science of Food and Agriculture, 100*, 2372–2379. https://doi.org/10.1002/jsfa.10241

Erşan, S., Güçlü Üstündağ, Ö., Carle, R., & Schweiggert, R. M. (2016). Identification of phenolic compounds in red and green pistachio (Pistacia vera l.) hulls (exo- and mesocarp) by HPLC-DAD-ESI-(HR)-MSn. *Journal of Agricultural and Food Chemistry, 64*, 5334–5344. https://doi.org/10.1021/acs.jafc.6b01745

Esteki, M., Farajmand, B., Kolahderazi, Y., & Simal-Gandara, J. (2017). Chromatographic fingerprinting with multivariate data analysis for detection and quantification of

apricot kernel in almond powder. *Food Analytical Methods, 10*, 3312–3320. https://doi.org/10.1007/s12161-017-0903-5

Esteki, M., Vander Heyden, Y., Farajmand, B., & Kolahderazi, Y. (2017). Qualitative and quantitative analysis of peanut adulteration in almond powder samples using multi-elemental fingerprinting combined with multivariate data analysis methods. *Food Control, 82*, 31–41. https://doi.org/10.1016/j.foodcont.2017.06.014

Food Fraud Advisors. (2017). *Food fraud risk information*. Retrieved October 10, 2020, from https://trello.com/b/aoFO1UEf/food-fraud-risk-information.

Fu, Y., Zhao, C., Lu, X., & Xu, G. (2017). Nontargeted screening of chemical contaminants and illegal additives in food based on liquid chromatography–high resolution mass spectrometry. *TrAC, Trends in Analytical Chemistry, 96*, 89–98. https://doi.org/10.1016/j.trac.2017.07.014

Gil Solsona, R., Boix, C., Ibáñez, M., & Sancho, J. V. (2018). The classification of almonds (Prunus dulcis) by country and variety using UHPLC-HRMS-based untargeted metabolomics. *Food Additives & Contaminants Part A Chemistry, Analysis, Control, Exposure and Risk Assessment, 35*, 395–403. https://doi.org/10.1080/19440049.2017.1416679

Huang, Y. P., Robinson, R. C., Dias, F. F. G., de Moura Bell, J. M. L. N., & Barile, D. (2022). Solid-phase extraction approaches for improving oligosaccharide and small peptide identification with liquid chromatography-high-resolution mass spectrometry: A case study on proteolyzed almond extract. *Foods, 11*, 4. https://doi.org/10.3390/foods11030340

Lacalle-Bergeron, L., Izquierdo-Sandoval, D., Sancho, J. V., López, F. J., Hernández, F., & Portolés, T. (2021). Chromatography hyphenated to high resolution mass spectrometry in untargeted metabolomics for investigation of food (bio)markers. *TrAC, Trends in Analytical Chemistry, 135*, Article 116161. https://doi.org/10.1016/j.trac.2020.116161

Lee, J., Zhang, G., Wood, E., Rogel Castillo, C., & Mitchell, A. E. (2013). Quantification of amygdalin in nonbitter, semibitter, and bitter almonds (prunus dulcis) by UHPLC-(ESI)QqQ MS/MS. *Journal of Agricultural and Food Chemistry, 61*, 7754–7759. https://doi.org/10.1021/jf402295u

Lucci, P., Saurina, J., & Núñez, O. (2017). Trends in LC-MS and LC-HRMS analysis and characterization of polyphenols in food. *TrAC, Trends in Analytical Chemistry, 88*, 1–24. https://doi.org/10.1016/j.trac.2016.12.006

Luparelli, A., Losito, I., De Angelis, E., Pilolli, R., Lambertini, F., & Monaci, L. (2022). Tree nuts and peanuts as a source of beneficial compounds and a threat for allergic consumers: Overview on methods for their detection in complex food products. *Foods, 11*, 728. https://doi.org/10.3390/foods11050728

Ma, Y., Kosińska-Cagnazzo, A., Kerr, W. L., Amarowicz, R., Swanson, R. B., & Pegg, R. B. (2014). Separation and characterization of phenolic compounds from dry-blanched peanut skins by liquid chromatography-electrospray ionization mass spectrometry. *Journal of Chromatography A, 1356*, 64–81. https://doi.org/10.1016/j.chroma.2014.06.027

Massart, D. L., Vandeginste, B. G. M., Buydens, L. M. C., de Jong, S., Lewi, P. J., & Smeyers-Verbeke, J. (1997). *Handbook of chemometrics and qualimetrics* (1st ed.). Amsterdam, The Netherlands: Elsevier.

Medina, S., Perestrelo, R., Silva, P., Pereira, J. A. M., & Câmara, J. S. (2019). Current trends and recent advances on food authenticity technologies and chemometric approaches. *Trends in Food Science and Technology, 85*, 163–176. https://doi.org/10.1016/j.tifs.2019.01.017

Monaci, L., De Angelis, E., Bavaro, S. L., & Pilolli, R. (2015). High-resolution Orbitrap™-based mass spectrometry for rapid detection of peanuts in nuts. *Food Additives & Contaminants Part A Chemistry, Analysis, Control, Exposure and Risk Assessment, 32*, 1607–1616. https://doi.org/10.1080/19440049.2015.1070235

Pluskal, T., Castillo, S., Villar-Briones, A., & Orešič, M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics, 11*, 395. https://doi.org/10.1186/1471-2105-11-395

Proestos, C., & Pesic, M. (2022). Editorial: Polyphenols in food authenticity through the application of advanced omics technologies. *Frontiers in Nutrition, 9*, Article 998635. https://doi.org/10.3389/fnut.2022.998635

Regueiro, J., Sánchez-González, C., Vallverdú-Queralt, A., Simal-Gándara, J., Lamuela-Raventós, R., & Izquierdo-Pulido, M. (2014). Comprehensive identification of walnut polyphenols by liquid chromatography coupled to linear ion trap-Orbitrap mass spectrometry. *Food Chemistry, 152*, 340–348. https://doi.org/10.1016/j.foodchem.2013.11.158

Riedl, J., Esslinger, S., & Fauhl-Hassek, C. (2015). Review of validation and reporting of non-targeted fingerprinting approaches for food authentication. *Analytica Chimica Acta, 885*, 17–32. https://doi.org/10.1016/j.aca.2015.06.003

Rochat, B. (2016). From targeted quantification to untargeted metabolomics: Why LC-high-resolution-MS will become a key instrument in clinical labs. *TrAC, Trends in Analytical Chemistry, 84*, 151–164. https://doi.org/10.1016/j.trac.2016.02.009

Romani, A., Pinelli, P., Moschini, V., & Heimler, D. (2017). Seeds and oil polyphenol content of sunflower (Helianthus annuus l.) grown with different agricultural management. *Advances in Horticultural Science, 31*, 85–88. https://doi.org/10.13128/ahs-20608

Rothwell, J. A., Perez-Jimenez, J., Neveu, V., Medina-Remón, A., M'Hiri, N., García-Lobato, P., et al. (2013). *Phenol-explorer 3.0: A major update of the phenol-explorer database to incorporate data on the effects of food processing on polyphenol content*. Database. https://doi.org/10.1093/database/bat070, 2013, bat070.

Rovira, G., Miaw, C. S. W., Martins, M. L. C., Sena, M. M., de Souza, S. V. C., Ruisánchez, I., et al. (2022). In-depth chemometric strategy to detect up to four adulterants in cashew nuts by IR spectroscopic techniques. *Microchemical Journal, 181*, Article 107816. https://doi.org/10.1016/j.microc.2022.107816

Sang, S., Li, G., Tian, S., Lapsley, K., Stark, R. E., Pandey, R. K., et al. (2003). An unusual diterpene glycoside from the nuts of almond (Prunus amygdalus Batsch). *Tetrahedron Letters, 44*, 1199–1202. https://doi.org/10.1016/S0040-4039(02)02794-6

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

Shataer, D., Li, J., Duan, X. M., Liu, L., Xin, X. L., & Aisa, H. A. (2021). Chemical composition of the hazelnut kernel (corylus avellana L.) and its anti-inflammatory, antimicrobial, and antioxidant activities. *Journal of Agricultural and Food Chemistry, 69*, 4111–4119. https://doi.org/10.1021/acs.jafc.1c00297

Singldinger, B., Dunkel, A., Bahmann, D., Bahmann, C., Kadow, D., Bisping, B., et al. (2018). New taste-active 3-(O-β- d -Glucosyl)-2-oxoindole-3-acetic acids and diarylheptanoids in cimiciato-infected hazelnuts. *Journal of Agricultural and Food Chemistry, 66*, 4662–4673. https://doi.org/10.1021/acs.jafc.8b01216

Sud, M., Fahy, E., Cotter, D., Brown, A., Dennis, E. A., Glass, C. K., et al. (2006). LMSD: LIPID MAPS structure database. *Nucleic Acids Research, 35*, 527–532. https://doi.org/10.1093/nar/gkl838

Suman, M., Cavanna, D., Sammarco, G., Lambertini, F., & Loffi, C. (2021). Fighting food frauds exploiting chromatography-mass spectrometry technologies: Scenario comparison between solutions in scientific literature and real approaches in place in industrial facilities. *TrAC, Trends in Analytical Chemistry, 142*, Article 116305. https://doi.org/10.1016/j.trac.2021.116305

Taylan, O., Cebi, N., Yilmaz, M. T., Sagdic, O., Ozdemir, D., & Balubaid, M. (2021). Rapid detection of green-pea adulteration in pistachio nuts using Raman spectroscopy and chemometrics. *Journal of the Science of Food and Agriculture, 101*, 1699–1708. https://doi.org/10.1002/jsfa.10845

Trox, J., Vadivel, V., Vetter, W., Stuetz, W., Kammerer, D. R., Carle, R., et al. (2011). Catechin and epicatechin in testa and their association with bioactive compounds in kernels of cashew nut (Anacardium occidentale L.). *Food Chemistry, 128*, 1094–1099. https://doi.org/10.1016/j.foodchem.2011.04.018

Wishart, D. S., Feunang, Y. D., Marcu, A., Guo, A. C., Liang, K., Vázquez-Fresno, R., et al. (2018). HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Research, 46*, D608–D617. https://doi.org/10.1093/nar/gkx1089

Wolf, S., Schmidt, S., Müller-Hannemann, M., & Neumann, S. (2010). In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinformatics, 11*, 148. https://doi.org/10.1186/1471-2105-11-148

Zhong, P., Wei, X., Li, X., Wei, X., Wu, S., Huang, W., et al. (2022). Untargeted metabolomics by liquid chromatography-mass spectrometry for food authentication: A review. *Comprehensive Reviews in Food Science and Food Safety, 21*, 2455–2488. https://doi.org/10.1111/1541-4337.12938