*Published in : Food Chemistry (2024), vol. 460, part 3, 140702*
*DOI: https://doi.org/10.1016/j.foodchem.2024.140702*
*Status : Postprint (Author's version)*

# OPTIMIZATION OF HEADSPACE HIGH-CAPACITY TOOL COUPLED TO TWO-DIMENSIONAL GAS CHROMATOGRAPHY-MASS SPECTROMETRY FOR MAPPING THE VOLATILE ORGANIC COMPOUNDS OF RAW PISTACHIOS. A PROOF-OF-CONCEPT ON THE CLASSIFICATION ABILITY BY GEOGRAPHIC ORIGIN

Andrea Schincaglia [a,b], Luisa Pasti [c], Alberto Cavazzini [a,e], Giorgia Purcaro [b], Marco Beccaria [a,d]

[a] *Department of Chemical Pharmaceutical, and Agricultural Sciences, Via Luigi Borsari 46, 44121, University of Ferrara, Ferrara, Italy*
[b] *Gembloux Agro-Bio Tech, Passage des Déportés 2, 5030, Gembloux, University of Liège, Belgium*
[c] *Department of Environmental and Prevention Sciences, Via L. Borsari 46, 44121, University of Ferrara, Ferrara, Italy*
[d] *Organic and Biological Analytical Chemistry Group, MolSys Research Unit, University of Liège, 4000 Liège, Belgium*
[e] *Council for Agricultural Research and Economics, CREA, via della Navicella 2/4, Rome, 00184, Italy*

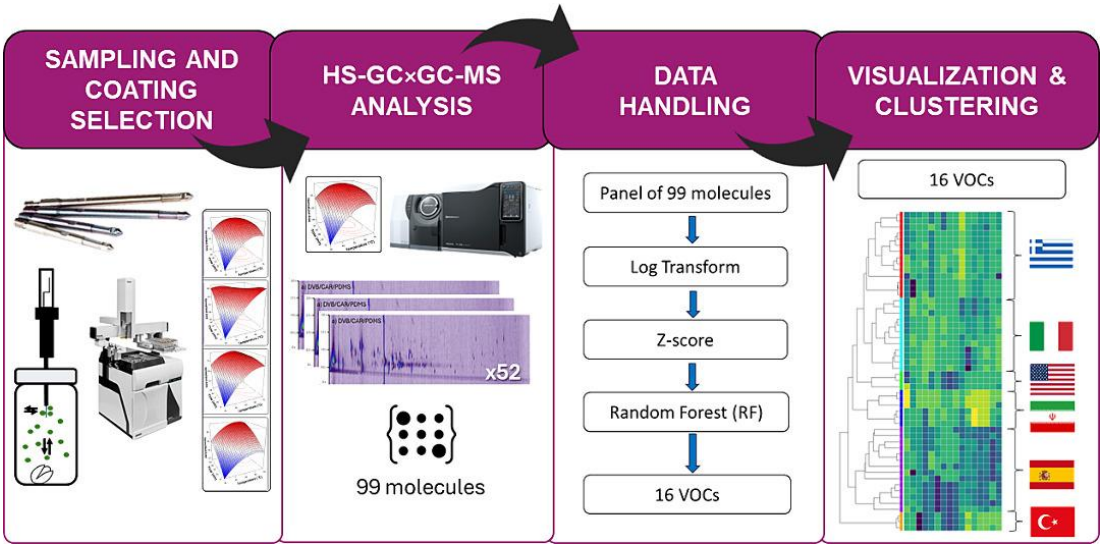
**KEYWORDS**: High-capacity tools ; GC × GC-MS ; VOCs ; Design of experiment ; Raw pistachios Geographic origin ; Classification

**ABSTRACT**

An optimized procedure for extracting and analyzing raw pistachio volatiles was developed through headspace sampling with high-capacity tools and subsequent analysis using comprehensive two-dimensional gas chromatography coupled with mass spectrometry. The examination of 18 pistachio samples belonging to different geographic areas led to the identification of a set of 99 volatile organic compounds (VOCs). Molecules were putatively identified using linear retention index, mass spectra similarity, and two-dimensional plot location. The impact of preprocessing and processing techniques on the aligned data matrix from a set of samples of different geographical origins, after removing contaminants, was evaluated. The combination of scaling with log-transformation, normalization with z-score, and data reduction with random forest machine learning algorithm generated a panel of 16 discriminatory VOC molecules. As a proof of concept, raw pistachios' VOC profile was employed for the first time to tentatively classify them based on their geographical origin.

*Published in : Food Chemistry (2024), vol. 460, part 3, 140702*
*DOI: https://doi.org/10.1016/j.foodchem.2024.140702*
*Status : Postprint (Author's version)*

**GRAPHICAL ABSTRACT**




# 1. Introduction

Pistachios have been consumed by humans for millennia, with remnants dating back to the 6th millennium B.C. found in Shortughai, Afghanistan (Fruits and Nuts, 2007). Through the centuries, pistachios' cultivation spread from this region to the rest of the globe, being imported for the first time in the United States in the 1880s (Mandalari et al., 2021). Currently, pistachios are cultivated across temperate and sub-tropical areas, including the Mediterranean, the Middle East, and California (Elloumi et al., 2024). Noteworthy, the last decade has seen a modest increase in worldwide production (+6% between 2014 and 2023), with an estimated worldwide production in 2023 of 776.869 metric tons (Pistachios | USDA Foreign Agricultural Service, 2024). The global spread and economic significance of pistachios underscore their value not only as a nutritious food source but also as a key ingredient in various products ranging from confectionery to pastry. As pistachios continue to be an integral part of diets and cuisines worldwide, their production and processing have evolved to meet growing demand, ensuring the availability of high-quality nuts for consumption and industrial use. The authenticity and quality of pistachios are of utmost importance, given their economic value and widespread consumption (Kashaninejad & Tabil, 2011). Over time, various research groups have explored diverse aspects, including physical properties (such as kernel dimensions and color components), biomolecular factors (such as DNA analysis), and chemical screenings (such as fatty acids, polyphenols, rare earth elements composition, and the ratio of chlorophyll *a* to chlorophyll *b*), to determine the geographical origin of pistachios (Boukid et al., 2019; Kalogiouri et al., 2021; Mannino et al., 2019; Wilson et al., 2018).

However, to our knowledge, there has been no investigation into the volatile organic compounds (VOCs) of raw pistachio nuts to assess their geographic origin. Indeed, VOCs can represent a key to understanding agricultural products' sensory profiles and authenticity (Lytou et al., 2019; Medina et al., 2019; Zakaria et al., 2018). In pistachios, these VOCs not only contribute to their distinctive aroma

*Published in : Food Chemistry (2024), vol. 460, part 3, 140702*
*DOI: https://doi.org/10.1016/j.foodchem.2024.140702*
*Status : Postprint (Author's version)*

and flavor but also reflect the nuts' environmental and cultivation conditions (Polari et al., 2019; Roxas et al., 2020). Despite previous attempts to analyze volatile in raw pistachios, the outcomes yielded only a limited number of VOCs, predominantly terpenes, alkanes, alcohols, ketones, aromatic amine, and aromatic hydrocarbons, using solid-phase microextraction (SPME) (Güler et al., 2022; Kendirci et al., 2011; Noguera-Artiaga et al., 2019; Penci et al., 2013; Şahan & Bozkurt, 2020; Valdés García et al., 2021). Indeed, the VOCs' profile of raw pistachios is inherently simpler than that observed in roasted pistachios (Hojjati et al., 2013). This simplicity in the VOCs' composition of raw pistachios poses challenges for distinguishing their geographic origins using VOCs' analysis alone.

This study aimed to optimize an analytical methodology based on high-capacity tool (HCT) headspace (HS) extraction of VOCs from raw pistachio samples, followed by two-dimensional gas chromatography (GC × GC) - mass spectrometry (MS). In addition, the VOCs profile obtained was used as a proof-of-concept to discriminate samples of raw pistachio nuts according to their geographical areas. A careful optimization was carried out on the HS analysis, considering both the ideal amount of sample in VOC extraction to avoid the HS saturation (using multiple headspace solid-phase microextraction, MHE-SPME) and the selectivity of different sorbent phases available (using HCTs) to ensure a representative and as comprehensive as possible analysis (Kolb & Ettre, 2006; Mascrez & Purcaro, 2020). To increase the extraction uptake of volatiles from the raw pistachios, an HCT characterized by a significantly higher sorbent volume compared to SPME was used (i.e., Hisorb®). This probe-like tool has been recently implemented in various analytical fields such as food, environmental, and clinical research (Eggermont, Spadafora, Aspromonte, Pellegrino, & Purcaro, 2023; Elia et al., 2023; Lenzi et al., 2023; van Vorstenbosch et al., 2024). The greater sorbent volume of HiSorb®, compared to that of traditional SPME fibers, enhances the yield of extracted analytes, extending the coverage of more polar compounds, making it suitable for extracting volatile and semi-volatile analytes in complex matrices.

Extraction with HCT (HiSorb®) was optimized for four sorbent phases available by means of a design of experiment (DoE) considering time and temperature of extraction as variables. The use of the combination of HCT, high-resolution chromatographic separation (GC × GC), and MS detection provided highly informative chromatographic fingerprints. After alignment, the resulting data matrix was analyzed through four data mining methods, using both preprocessing and processing techniques to select the most discriminatory molecules. Analytes were identified based on their linear retention index (LRI), spectral similarities to commercial libraries, and 2D-GC plot location. To the best of our knowledge, this is the first study where a panel of 99 volatile molecules were tentatively identified from the HS of raw pistachio samples. In addition, as a proof of concept, the most discriminatory molecules of VOCs' profile of raw pistachios were used to classify raw pistachios samples according to their geographic origin.

*Published in : Food Chemistry (2024), vol. 460, part 3, 140702*
*DOI: https://doi.org/10.1016/j.foodchem.2024.140702*
*Status : Postprint (Author's version)*

# 2. Materials and methods

## 2.1. CHEMICAL AND REAGENTS

Normal alkanes ($C_7$-$C_{30}$) mixture (Sigma-Aldrich, DE) was used for quality control of the instrument performance and to calculate the linear retention index (LRI) for peak identification. The alkane mixture was diluted with n-Hexane HPLC grade (Biosolve, FR). For the sample size evaluation, SPME fiber divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS) df 50/30 µm/ 2 cm length, kindly provided by MilliporeSigma (USA), was used. Further analyses were carried out with HiSorb® probes, standard length, coated with DVB/CAR/PDMS (H4-AXAAC), DVB/PDMS (H3-AXAAC), CAR/PDMS (H2-XXAAC) and PDMS (H1-AXAAC) kindly provided by Markes International Ltd. (Bridgend, UK).

## 2.2. PISTACHIOS SAMPLES

A total of 18 certified origin raw pistachios (unroasted, unsalted, and unshelled) belonging to different geographic areas, namely Greece ($n = 5$), Iran ($n = 2$), Italy (Bronte, Sicily) ($n = 4$), Spain ($n = 5$), Türkiye ($n = 1$), and USA (California) ($n = 1$), were kindly provided by Di Bartolo S.r. l. (Catania, IT). Due to the low numerosity of raw pistachio samples from California, Iran, and Türkiye ($n = 4$), these non-EU pistachios were grouped into one category named "Extra-EU". Pistachios were grounded (10 s) right before analysis with a coffee grinder until a homogenous powder was obtained and immediately subjected to analysis.

## 2.3. HEADSPACE EXTRACTION

Both SPME and HiSorb® extractions were performed using a Centri® sample extraction and enrichment platform (Markes International Ltd., Bridgend, UK) equipped with an electrically cooled focusing trap (U-T12ME-2S, Markes International, general-purpose in the C4-C32 volatile range).

### 2.3.1. MHE-SPME EXTRACTION

A DVB/CAR/PDMS SPME fiber was used. A blank analysis was performed at the start of the sampling batch at different points during sequences, ensuring the absence of carry-over. Various quantities of ground pistachios (1.5 g, 1 g, 0.5 g, 0.1 g) were weighed into 20 mL screw-top vials equipped with metallic caps featuring a central opening and polytetrafluoroethylene (PTFE)/silicone septa sourced from Restek (USA). SPME fibers were properly conditioned before use according to the manufacturer: 30 min at 270 °C. For each of the sizes, HS-SPME extraction was performed 3 times from the same vial. Sampling was performed for 50 min at 50 °C under agitation, consisting of cycles of 10s at 350 rpm followed by 2 s rest, using the heated mechanical agitating unit embedded in the Centri platform. After extraction, SPME fiber was fully desorbed at 250 °C for 5 min and the volatiles were refocused on the trap (U-T12ME-2S) set at 0 °C. After 1 min of helium purge at 50 mL/min, the trap was heated to 300 °C (hold 3 min), and the split flow was set at 10 mL/min. Analyses were run in

*Published in : Food Chemistry (2024), vol. 460, part 3, 140702*
*DOI: https://doi.org/10.1016/j.foodchem.2024.140702*
*Status : Postprint (Author's version)*

triplicate. The chromatographic areas of 3 compounds, namely 1-methylpyrrole, limonene, and nonanal, were used to evaluate the saturation of the HS.

#### 2.3.2. PROBE TOOL EXTRACTION AND OPTIMIZATION

HiSorb® probes (DVB/CAR/PDMS, DVB/PDMS, CAR/PDMS, and PDMS) were preconditioned according to the manufacturer: 10 min at 50 °C, then 90 min at 270 °C.

To assess the most suitable conditions of extraction, with each of the coatings available for HiSorb®, an experimental design was performed. Two variables ($k = 2$) inscribed rotatable ($\alpha = 1/\sqrt{k}$) central composite design (CCD) was used, considering as variables extraction time (5—50 min range) and extraction temperature (30—70 °C range). These variables were chosen according to other contributions on similar matrices (Georgiadou et al., 2015; Stilo et al., 2021). In total nine different points (1 central, 4 axials, and 4 factorials) were evaluated. The central point was run four times to evaluate the method's repeatability, also, experimental runs were randomized to minimize the effect of unexpected variability. Agitation consisted of cycles of 10s at 350 rpm followed by 2 s rest, using the heated mechanical agitating unit embedded in the Centri platform. After extraction, the probe was fully desorbed at 250 °C for 15 min. The volatiles were refocused on a trap (U-T12ME-2S) set at 0 °C. After 1 min of helium purge at 50 mL/min, the trap was heated to 300 °C (hold 5 min) and the split flow was set at 10 mL/min.

The total peak area obtained from each chromatographic run was used to evaluate the extraction efficiency throughout the response surface plot methodology (Marrubini et al., 2020). Among the integrated compounds, the ones recognized as siloxanes were not considered for the calculation of the total peak area since they are known to be artifacts coming from the instrumentation and not from the matrix of interest. After each extraction, the probe was thermally desorbed on the autosampler for 15 min at 250 °C.

### 2.4. GC×GC-qMS ANALYSIS

All the GC analyses were performed on a Shimadzu GCMS-TQ8050 NX (JP), consisting of a GC2030 coupled to a triple-quadrupole mass spectrometer detector (QQQ-MS) (Shimadzu, DE). The analytical GC columns were connected through an INSIGHT™ reverse fill-flush (RFF) flow modulator (SepSolve Analytical Ltd., UK). For the GC × GC separation, a conventional (non-polar/polar) set of columns was used. The first dimension ($^1$D) column was a BPX-5 ms 20 m × 0.18 mm i.d. × 0.18 µm film thickness; equivalent in polarity to 5% phenyl / 95% dimethyl polysiloxane (Trajan, AU). The second dimension ($^2$D) column was a SLB-50 ms 3 m × 0.25 mm i.d. × 0.25 film thickness equivalent in polarity to 50% phenyl / 50% dimethyl polysiloxane (Supelco, USA). The outlet of the 2D column was connected through a Y-union (Restek, USA) to a 1.1 m × 0.1 mm i.d. uncoated capillary to the MS and a 20 cm × 0.25 mm i.d. uncoated capillary to a VUV detector (VGA-101, VUV Analytics, USA), the data acquired by the VUV were not used in the present study.

The split flow ratio between the MS and the VUV was calculated to be 37% and 63%, respectively, at the beginning of the run. The custom temperature program for the GC oven was 30 °C, held for 3 min,

*Published in : Food Chemistry (2024), vol. 460, part 3, 140702*
*DOI: https://doi.org/10.1016/j.foodchem.2024.140702*
*Status : Postprint (Author's version)*

up to 220 °C, held for 3 min at 7 °C/min. The carrier gas used was helium and the auxiliary pressure-controlled module was set to have a flow corresponding to 0.5 mL/min in the $^{1}$D column and 19 mL/min in the $^{2}$D column. The installed bleed line was 1 m × 0.1 mm i.d. connected to an auxiliary pressure controller. The bleed line flow was regulated at 0.55 mL/min. The modulation time was set at 2.0 s, with 150 ms of reinjection time. The MS was operated in single-Q mode, using electron-impact ionization (EI) at 70 eV. The ion source and transfer line temperatures were 200 °C and 250 °C, respectively. The scan range was set to 35—350 $m/z$, with an acquisition frequency of 50 Hz.

Data was acquired using GCMS-solution version 4.53 software from Shimadzu. Compounds eluting before 4 min of retention time were excluded. Siloxanes were manually identified considering their specific parent ions (e.g. 73, 207, 281) and excluded. For peak identification, a signal-to-noise (S/N) cutoff was chosen at least of 20:1 (S/N). Compounds were determined by matching their spectra to the NIST17 and FFNSC 3.0 MS commercial libraries, using as identification criteria with a match score of ≥80%, an LRI window of ±20, and 2D-GC plot location (Beccaria et al., 2022).

## 2.5. DATA ELABORATION AND STATISTICAL ANALYSIS

Raw data files were converted to *.cdf* files and imported into ChromSpace version 2.1.7 software (SepSolve Analytical Ltd., UK). Chromatograms were aligned by a one-user-selected reference chromatogram in ChromSpace version 2.1.7 (SepSolve Analytical Ltd., UK). The alignment algorithm was used to overcome possible retention time drifts across the dataset. Utilizing the available spectral information, the algorithm aligned each chromatogram in the dataset with a designated reference chromatogram. No additional data pre-treatment was necessary for this alignment process (Mascrez et al., 2024).

The complete pistachio VOCs data matrix was handled and subjected to preprocessing techniques such as logarithmic transformation and normalization (z-score and quantile normalization). Random forest (RF) machine learning algorithm was used for classification purposes. Specifically, the RF model was configured with 2000 trees, while to determine the optimal number of variables tried at each split (*mtry*), a tuning process was employed using the *tuneRF* function. The best *mtry* value was systematically selected through this tuning procedure, which evaluated different *mtry* values based on a step factor of 1.5 and an improvement threshold of 0.01, with the results visualized and the minimum error rate determining the optimal *mtry*. Results were visualized using heatmaps (HMs) with hierarchical clustering analysis (HCA) with Euclidean distance and principal component analysis (PCA). Statistical elaborations were performed using Minitab software version 20.4 (https://www.minitab.com/en-us/). Excel® (Microsoft Office, version 16.75.2), and R v4.3.2 (R Foundation for Statistical Computing, Vienna, AT).

*Published in : Food Chemistry (2024), vol. 460, part 3, 140702*
*DOI: https://doi.org/10.1016/j.foodchem.2024.140702*
*Status : Postprint (Author's version)*

# 3. Results and discussion

## 3.1. OPTIMIZATION OF THE HEADSPACE EXTRACTION CONDITIONS

### 3.1.1. SAMPLE SIZE EVALUATION

In HS analysis, two equilibria involving three phases (i.e., sample, HS, and fiber) are crucial: one equilibrium between the sample and the headspace, and the other between the headspace and the probe (Paw-liszyn, 2000). The amount extracted by the probe (n) ideally reflects the initial sample concentration ($C_0$) under both equilibrium and nonequilibrium conditions. However, direct proportionality between chromatographic area and $C_0$ is assured only in unsaturated HS conditions (Kolb & Ettre, 2006). As in most untargeted studies, the area intensity for various compounds served as an indicator of absolute concentration, therefore, to ensure a truly representative analysis, the sample quantity to maintain HS linearity needs to be assessed (Mascrez & Purcaro, 2020). When such linearity is achieved, conducting multiple headspace extractions (MHE) results in a proportional decrease in recorded chromatographic area (Mascrez & Purcaro, 2020).

Although the work aimed to investigate the potential of Hisorb® for profiling pistachio's VOCs, the minimum sample quantity required to maintain unsaturated conditions (i.e., linear conditions) was assessed using MHE-SPME-GC × GC—MS. Indeed, unlike SPME fibers, which permit MHE due to partial resealing of PTFE/silicone septa upon fiber withdrawal, HiSorb®'s larger stainless-steel rod precludes such resealing, preventing it from being employed in MHE.

Four different sample sizes (sample Sicily 1, Italy), namely 1.5 g, 1 g, 0.5 g, and 0.1 g of freshly ground pistachios, were transferred to a 20 mL SPME vial and analyzed. Each size underwent HS-SPME extraction three times from the same vial, using the same extraction conditions: 50 min at 50 °C. The chromatographic areas of three major compounds, namely 1-methylpyrrole, limonene, and nonanal, were recorded to assess HS saturation. The three selected compounds cover different chemical classes and volatility ranges; moreover, being present at high concentrations, it can be reasonably hypothesized that when they are not saturating the HS, the entire volatile profile respects such a condition. These analytes were present in all of the raw pistachio samples analyzed.

As shown in Figure 1, after three HS extraction using 1.5 g, 1 g, or 0.5 g of sample, the peak area of the considered compounds remains almost constant. Instead, with 0.1 g of sample, a clear decrease in the chromatographic areas was observed, indicating unsaturated HS. Thus, further analyses were conducted using 0.1 g of freshly grounded sample.

### 3.1.2. HISORB OPTIMIZATION AND COATING CHOICE

While SPME stands as a well-recognized HC tool extensively utilized in VOCs' analysis in food samples, it may lack sensitivity (Aspromonte et al., 2023; Eggermont, Spadafora, Aspromonte, & Purcaro, 2023), particularly when dealing with samples possessing a lower abundance of VOCs such as raw pistachios (compared to their roasted counterparts) (Hojjati et al., 2013). Therefore, the use of a higher volume of sorbent was preferred for further analysis.

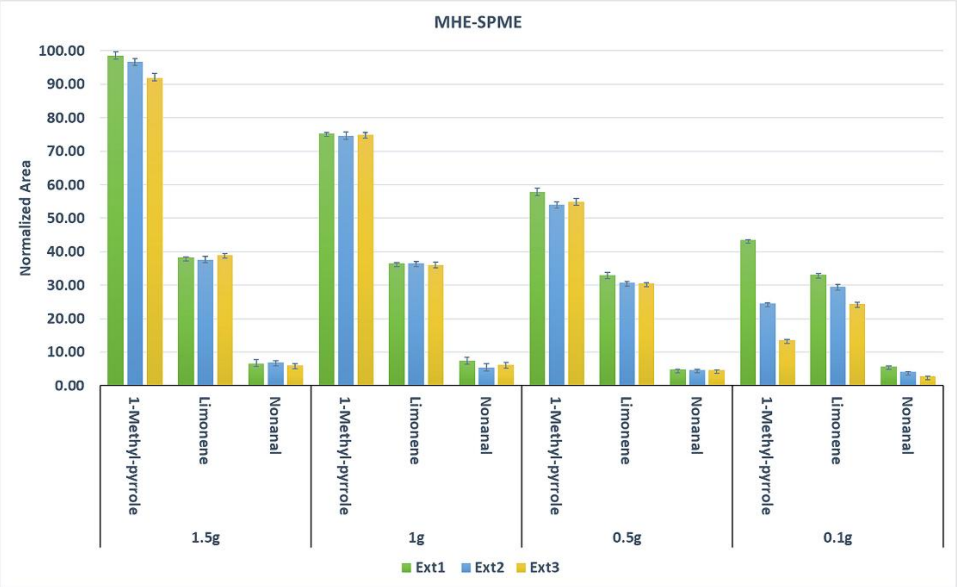
*Published in : Food Chemistry (2024), vol. 460, part 3, 140702*
*DOI: https://doi.org/10.1016/j.foodchem.2024.140702*
*Status : Postprint (Author's version)*

Following the determination that 0.1 g of freshly ground pistachios ensures unsaturated HS conditions, four HiSorb® coatings, namely DVB/CAR/PDMS, DVB/PDMS, CAR/PDMS, and PDMS, were compared to determine their extraction abilities across a wide range of compounds. To compare the optimal efficiency for each Hisorb® coating, the extraction conditions for each coating were separately optimized through a DoE, employing a CCD. The efficiency of extraction was based on the total peak area (excluding siloxanes), with optimal conditions summarized in Table 1.

Finally, the optimal conditions for each coating were used to analyze freshly ground pistachios (sample Sicily 1, Italy) three times to compare the volatile coverage of each coating. Evaluation of the compound range extracted with each coating revealed a wider variety of VOCs with the triphasic coating (DVB/CAR/PDMS) compared to alternative coatings. Notably, the DVB/CAR/PDMS coating exhibited efficacy in extracting VOCs spanning various chemical classes, covering both less polar analytes (e.g., alkanes) and more polar compounds like alcohols and aldehydes (Figure 2). For instance, compounds such as decanal, nonanal, and benzaldehyde—responsible for citrus-like and burnt sugar flavors (Acree & Heinrich, 2004)—were successfully extracted using the DVB/CAR/ PDMS coating. Although the CAR/PDMS coating demonstrated comparable performance in extracting diverse chemical classes, preference was given to the triphasic coating due to its ability to extract a higher number of VOCs overall, thereby providing a more comprehensive profile of the matrix. A comparison of the chromatograms obtained from the extraction of the VOCs using the four coatings available is reported in Figure 3.

The DVB/CAR/PDMS coating gave by far the richest chromatograms in terms of compound coverage and intensity. Therefore it was decided to continue with this probe for further evaluation on the value of the information that can be obtained with the volatile profile.

***Figure 1***. *Comparison of MHE (three extractions were performed) using four different sample weights (i.e., 1.5 g, 1.0 g, 0.5 g, and 0.1 g) evaluated on 3 compounds ($n = 3$), namely 1-methylpyrrole, limonene, and nonanal. Data normalized to the highest value. Extraction conditions: 50 °C for 50 min.*



*Published in : Food Chemistry (2024), vol. 460, part 3, 140702*
*DOI: https://doi.org/10.1016/j.foodchem.2024.140702*
*Status : Postprint (Author's version)*

***Figure 2**. A bar chart displaying the number of compounds, divided by chemical group, extracted by each coating type (sample: Sicily 1; $n = 3$ for each coating). At the top of each bar is reported the overall sum of VOCs.*

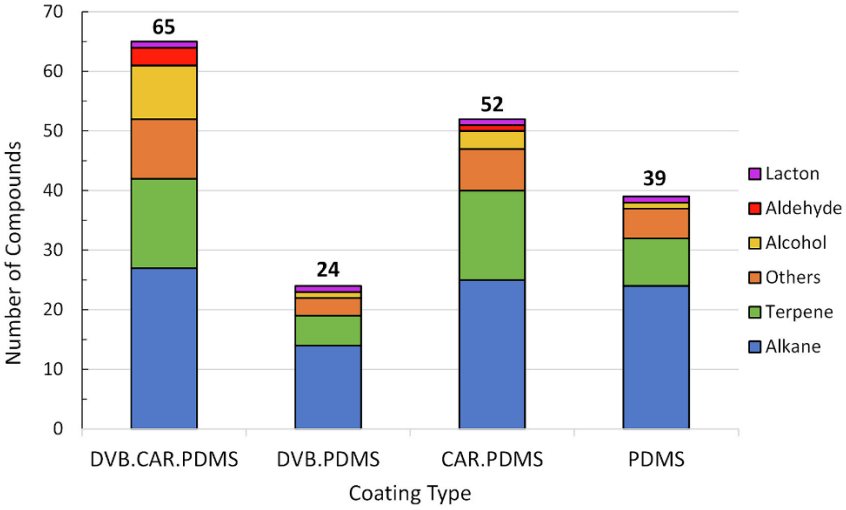
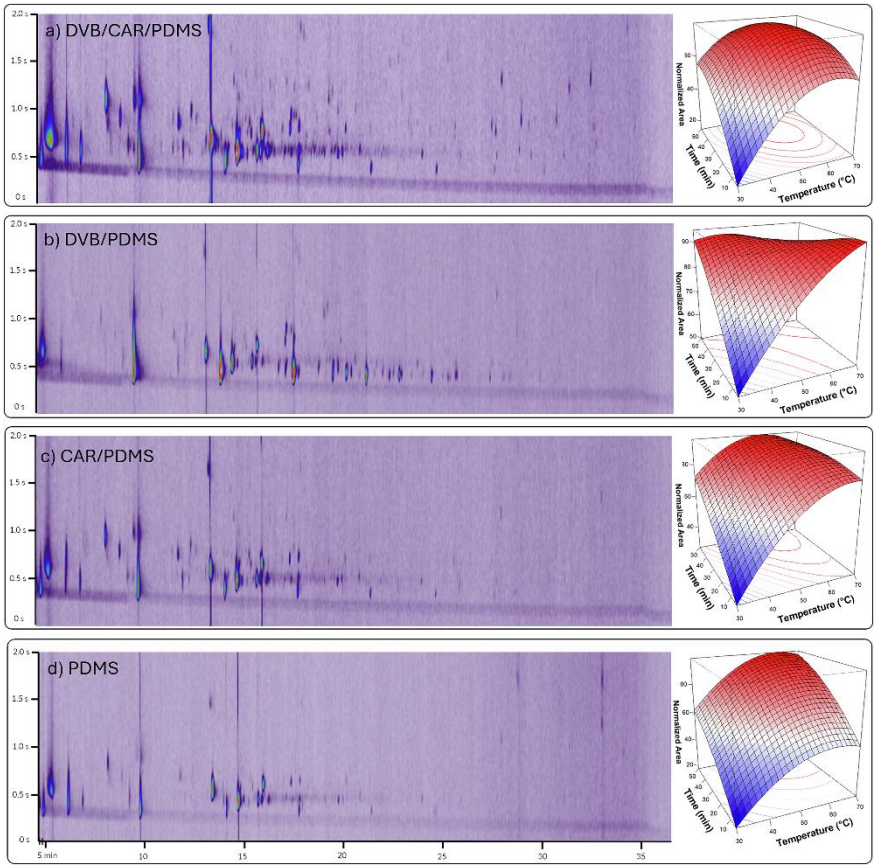



***Figure 3**. Comparison of the different chromatograms obtained via HS-HCT-GC × GC-MS using the optimal conditions of extraction for each available coating, alongside the corresponding response surfaces. a) DVB/CAR/PDMS; b) DVB/PDMS; c) CAR/PDMS; d) PDMS.*



*Published in : Food Chemistry (2024), vol. 460, part 3, 140702*
*DOI: https://doi.org/10.1016/j.foodchem.2024.140702*
*Status : Postprint (Author's version)*

***Table 1**. Optimized extraction conditions, given by the DoE, for each HiSorb® available coating.*

| Coating | Time (min) | Temperature (°C) |
|---|---|---|
| DVB/CAR/PDMS | 36 | 54 |
| DVB/PDMS | 7 | 70 |
| CAR/PDMS | 50 | 55 |
| PDMS | 50 | 62 |

## 3.2. CLASSIFICATION OF PISTACHIO SAMPLES BASED ON THEIR GEOGRAPHIC ORIGIN: A PROOF-OF-CONCEPT

VOCs are unique chemical signatures influenced by the environment where they originate. This characteristic makes them valuable tools for identifying the geographical source of the matrix of interest (De Flaviis et al., 2021; Gil-Solsona et al., 2016; Sater et al., 2020). However, VOCs are also sensitive to experimental conditions and slight changes in the analytical process can alter the detected VOC profile (Zhang & Li, 2010).

The optimized HiSorb® extraction using the DVB/CAR/PDMS coating was employed to create a data matrix from the analysis of 18 samples of raw pistachios belonging to different geographical origins, namely Greece ($n = 5$), Italy (Sicily) ($n = 4$), Spain ($n = 5$), and Extra-EU countries (US, Iran, and Türkiye) ($n = 4$). Each sample was analyzed in triplicate, except for two samples, one from Spain and the other from Greece, that were in duplicate, generating a matrix of 52 samples, including technical replicates. After alignment by one user-selected chromatogram (see section: 2.7 Data elaboration and statistical analysis), a panel of 113 compounds was identified. A first reduction was manually performed by removing artifacts (e.g. siloxanes), resulting in a panel of 99 molecules. Compound classes among the identified molecules mainly include alkanes (44%), followed by terpenes (19%) and alcohols (11%).

To improve the discrimination capability of the overall classification, the RF algorithm was applied to select and retain the most discriminatory features. The RF machine learning algorithm functions by creating numerous classification trees, employing randomly chosen subsets of both features and data points, with features ultimately selected based on their ability to effectively divide the data into classes at each split (Breiman, 2001). Mean decrease accuracy was used as the measure of variable importance for the RF, features were selected with the elbow method. This method involves plotting the features and selecting a cutoff capturing the elbow of the graph, only the features above the elbow were retained (Beccaria et al., 2018).

In applying data transformation, normalization, and feature reduction/selection, a similar workflow previously reported in VOC analysis was implemented (Purcaro et al., 2018). Four workflows were investigated, with and without the use of preprocessing techniques (log-transformation and normalization), in combination with RF as a data reduction step: I) the raw matrix directly underwent RF; II) log transformation followed by RF; III) log transformation followed by Z-score and RF; IV) log

*Published in : Food Chemistry (2024), vol. 460, part 3, 140702*
*DOI: https://doi.org/10.1016/j.foodchem.2024.140702*
*Status : Postprint (Author's version)*

transformation followed by quantile normalization and RF. The scheme of these data mining comparisons is reported in Figure 4.

Log transformation is effective for correcting skewed data since many statistical techniques require data to follow a normal distribution (Hammouri et al., 2020). Quantile normalization adjusts datasets so that their distribution matches a reference distribution, typically the averaged distribution across all samples under consideration. By enforcing a uniform statistical distribution, quantile normalization ensures that differences among samples are reflective of genuine variations in analyte concentrations rather than methodological discrepancies (Amaratunga & Cabrera, 2001; Bolstad et al., 2003; Hicks & Irizarry, 2015; Zhao et al., 2020). Z-score normalization, or standardization, rescales data to have a mean of zero and a standard deviation of one. Z- score can help enhance comparability, it facilitates the comparison of analytes concentrations across different samples, even when those compounds vary widely in their natural concentrations (Wang et al., 2010; Wei et al., 2012). These normalization techniques are pivotal in addressing the heterogeneity inherent in VOC datasets, facilitating the identification of significant patterns and/or outliers.

Following these transformations and normalization phases, data reduction was carried out to reduce noise and extract the most useful information.

Overall, out of the 4 workflows, 22 discriminatory features were selected. The tentative identification of the selected features is reported in Table 2. The identity of 21 out of 22 discriminatory compounds was confirmed by NIST library match ≥80 (out of 100) and LRI deviation within ±20. One out of 22 compounds with match scores below 80 was tentatively classified by chemical class, relying on its mass spectra resemblance to the highest-scoring compound under the 80 thresholds within an LRI deviation of ±10. Moreover, the position in the 2D plot was used as additional information for chemical class identification.

Results obtained for each workflow were visualized using PCA and HMs with HCA applying Euclidean distance to measure the similarity between data points (e.g., VOC measurements). In HCA points that are closer together in the space defined by their characteristics are considered more similar and are more likely to be grouped into the same cluster. The average linkage method chosen calculates the distance between two clusters as the average distance between all pairs of points in the two clusters (Cabezas et al., 2023). The linkage method in HCA determines how the distance between clusters is calculated, which in turn affects how clusters are formed at each step of the analysis. This approach is well-suited for identifying patterns and relationships in data where the physical distances (reflecting similarity in characteristics) and the average relationships between data points (reflecting group cohesion) are key to understanding the underlying structure of the data (Cabezas et al., 2023).

When the raw data matrix, after removing contaminants, is directly subjected to RF, a panel of 16 VOCs is generated (Table 2). However, both PCA and heatmap resulted in clusters that are not distinctly separated by geographical origin (Figs. S1 and S2). The lack of preprocessing may retain the influence of outliers or extreme values, which can hide the natural clustering in the data.

*Published in : Food Chemistry (2024), vol. 460, part 3, 140702*
*DOI: https://doi.org/10.1016/j.foodchem.2024.140702*
*Status : Postprint (Author's version)*

The sole log transformation before RF helps with skewness generating a panel of 16 VOCs, however, the absence of a normalization step may leave some features on scales that diminish RF's ability to select the most relevant features for geographical clustering (Figure S3 and S4).

The combination of log transformation, quantile normalization, and RF generated a panel of 13 VOCs, resulting in clusters not perfectly aligned with geographical origins in both the heatmap and the PCA (Figure S5 and S6), possibly because quantile normalization standardizes the distribution across all samples, potentially over-smoothing unique regional characteristics.

The combination of log transformation, Z-score normalization, and RF preprocessed the data in a way that enhanced the natural clustering based on geographical origin, Figure 5 and Figure 6. This combination likely improved the internal consistency within geographical clusters while maximizing the differences between them, thereby helping the RF algorithm in feature selection and the HCA in generating more distinct clusters.

Of the discriminatory molecules selected by the four combinations of preprocessing and processing techniques on the data matrix (Table 2), 12 out of 22 were terpenes, five of which (camphene, α-terpinene, limonene, terpinolene, and p-cymenene), together with decanal and 1-methylpyrrole, were in common among the different data mining approaches. These VOCs account for the citrus, pine, and camphor-like fragrances of raw pistachios (Acree & Heinrich, 2004).

Terpenes play crucial roles not just in imparting distinctive aromas to raw pistachios but also in the broader context of plant interactions and defense mechanisms, being part of a sophisticated chemical language that plants use for communication with the environment, other plants, and insects. They can signal to deter herbivores, attract pollinators, or even cue neighboring plants about potential threats, enhancing not only their survival and reproductive success but also influencing ecological community dynamics (Boncan et al., 2020; Bouwmeester et al., 2019; Ninkuu et al., 2021; Rosenkranz et al., 2021).

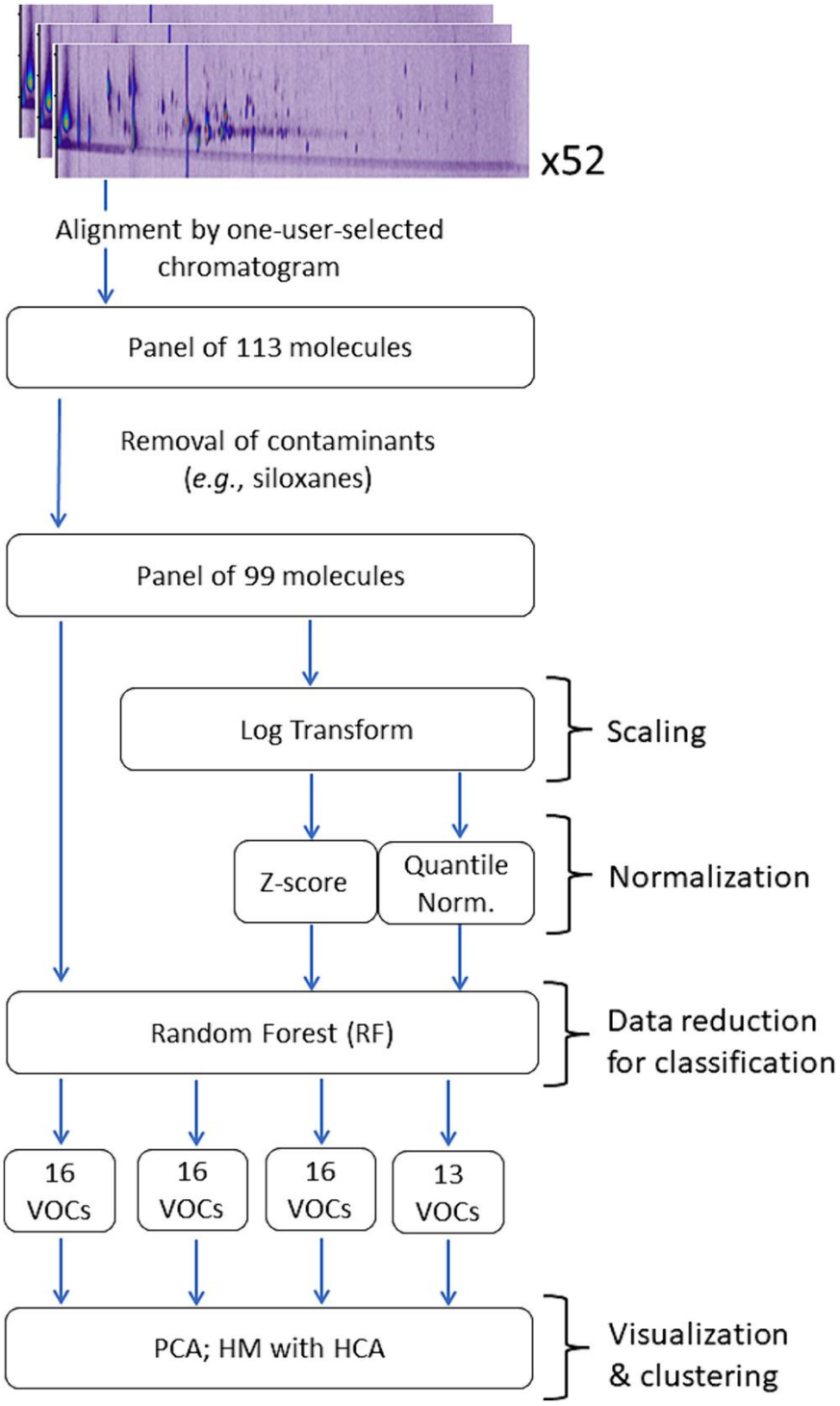
***Table 2**. Putative identification of the 22 discriminatory molecules selected with RF applying or not log transform and normalization (Z-Score or Quantile Normalization). Odor descriptors are also reported.*

| Compound Name | CAS | $LRI_{Exp}$ | $LIR_{Lib}$ | MS Similarity (%) | Random Forest | | | | Odor descriptor* |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | No Log | Log transform | | | |
| | | | | | No norm | No norm | Quantile Norm | Z-Score | |
| 1-Methylpyrrole | 96–54-8 | 683 | 702 | 84 | X | X | X | X | – |
| 1-butanol-3-methyl | 123–51-3 | 692 | 697 | 86 | | | X | | Whiskey, malt, burnt |
| Disulfide-dimethyl | 624–92-0 | 699 | 718 | 82 | | | X | | Onion, cabbage |
| 2–3-Butanediol | 513–85-9 | 802 | 802 | 85 | | | X | | Fruit, onion |
| Branched Alcohol | – | 811 | – | – | X | | X | | – |
| Butyrolactone | 96–48-0 | 936 | 922 | 95 | | | X | | Caramel, sweet |
| α-Pinene | 80–56-8 | 941 | 948 | 94 | X | X | | X | Pine, turpentine |
| β-Pinene | 80–56-8 | 941 | 948 | 94 | X | X | | X | Pine, resin, turpentine |
| Camphene | 79–92-5 | 960 | 953 | 93 | X | X | X | X | Camphor |
| β-Myrcene | 123–35-3 | 993 | 981 | 93 | X | X | | X | Balsamic, must, spice |
| Decane | 124–18-5 | 1002 | 1015 | 94 | | | X | | – |
| α-Terpinene | 99–86-5 | 1007 | 998 | 84 | X | X | X | X | Lemon |
| 3-Carene | 13,466–78-9 | 1018 | 1011 | 92 | X | X | | X | Lemon, resin |
| P-Cymene | 99–87-6 | 1037 | 1042 | 91 | X | X | | X | Solvent, gasoline, citrus |
| Limonene | 5989-27-5 | 1039 | 1028 | 93 | X | X | X | X | Lemon, orange |
| γ-Terpinene | 99–85-4 | 1070 | 1050 | 85 | X | X | | X | Gasoline, turpentine |
| Terpinolene | 586–62-9 | 1097 | 1088 | 94 | X | X | X | X | Pine, plastic |
| p-Cymenene | 1195-32-0 | 1106 | 1095 | 91 | X | X | X | X | Citrus, pine |
| Citrale | 5392-40-5 | 1199 | 1213 | 83 | X | X | | X | Lemon |
| Decanal | 112–31-2 | 1218 | 1204 | 87 | X | X | X | X | Orange peel, tallow |
| Isobornyl acetate | 92,618–89-8 | 1302 | 1302 | 84 | | X | | X | Piney, balsamic |
| Pentadecane | 629–62-9 | 1503 | 1500 | 80 | X | X | | X | Waxy |

* *Odor characteristics are taken from the following source (Acree & Heinrich, 2004).*

*Published in : Food Chemistry (2024), vol. 460, part 3, 140702*
*DOI: https://doi.org/10.1016/j.foodchem.2024.140702*
*Status : Postprint (Author's version)*

***Figure 4**. Schematic of the different statistical workflows applied.*



*Published in : Food Chemistry (2024), vol. 460, part 3, 140702*
*DOI: https://doi.org/10.1016/j.foodchem.2024.140702*
*Status : Postprint (Author's version)*

***Figure 5**. HM with HCA of the most discriminatory VOCs in raw Pistacia vera samples using log transformation, z-score normalization, and RF. The vertical axis represents individual samples, categorized by origin with corresponding color coding, while the horizontal axis lists the selected VOCs. The intensity of colors reflects the concentration or presence of each compound. Clusters are formed based on the similarity of VOC profiles, with the dendrogram above showing the hierarchical grouping.*

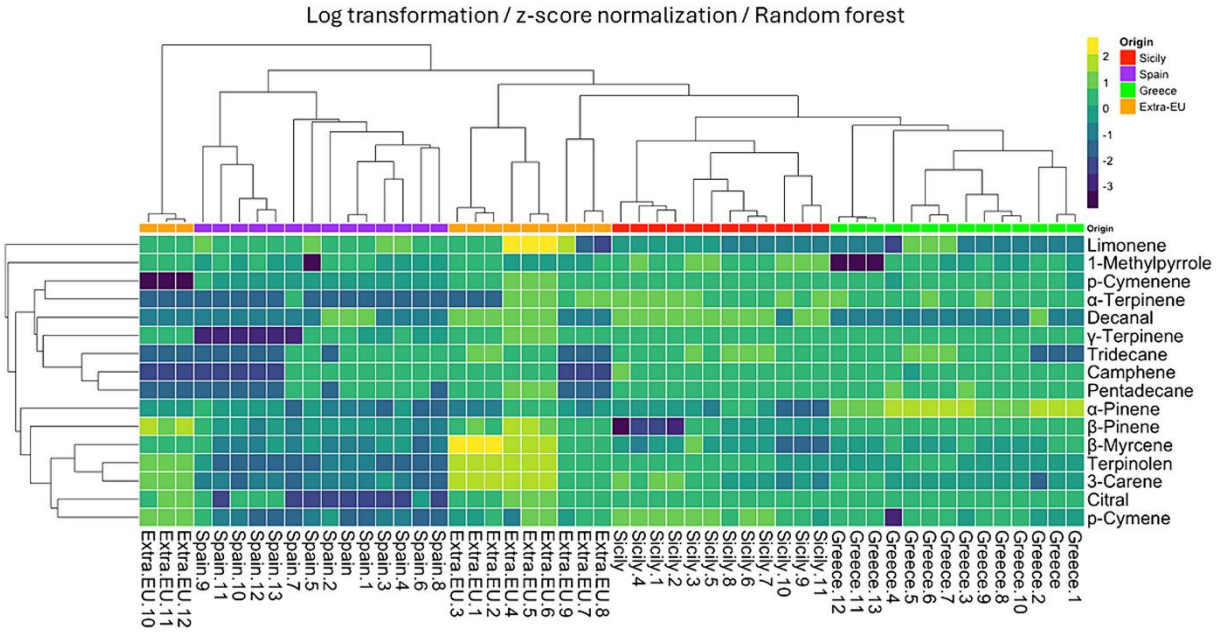
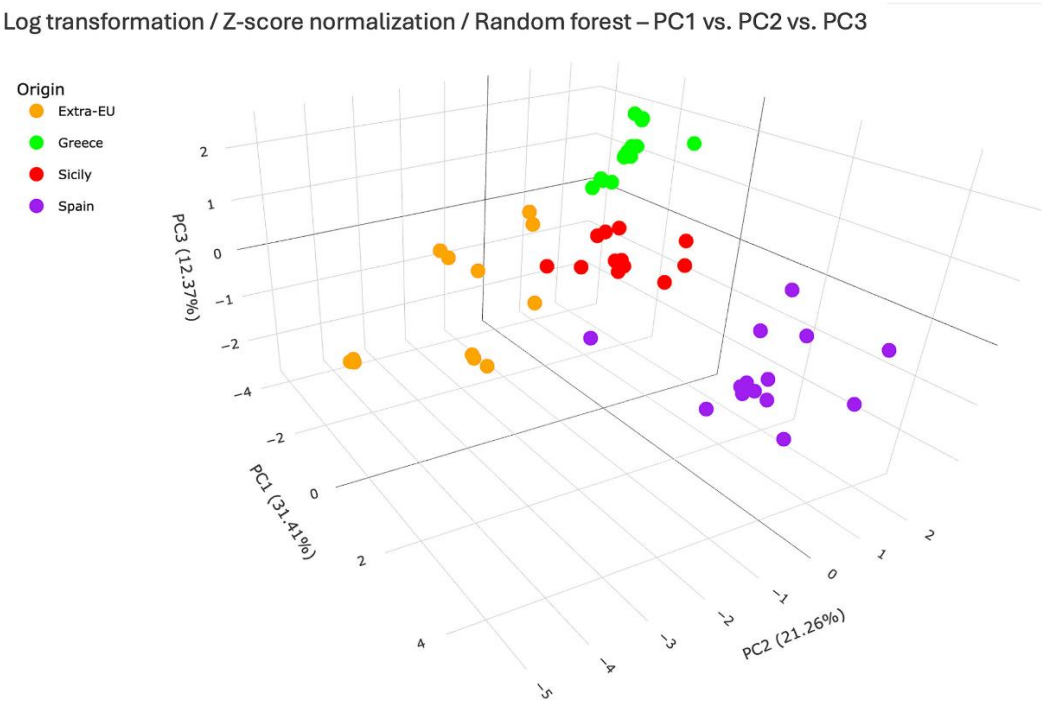



***Figure 6**. Score plot of the first three principal components of the PCA model, using the most discriminatory features selected via RF applied on log-transformed and z-score normalized data matrix.*



*Published in : Food Chemistry (2024), vol. 460, part 3, 140702*
*DOI: https://doi.org/10.1016/j.foodchem.2024.140702*
*Status : Postprint (Author's version)*

# 4. Conclusions

In this study, HiSorb® was used for the first time to profile VOCs in pistachios and correlate them with their geographical origin. The optimization process involved comparing the extraction efficiency of different HiSorb® coatings under optimal conditions for each. This allowed for a thorough evaluation and selection of the best coating. The chosen HiSorb® coating was then used to extract VOCs from 18 raw pistachio samples from Greece, Iran, Italy, Spain, Türkiye, and the United States. Analysis with an RFF-GC × GC-MS system identified 99 volatile molecules. Applying various data mining methods, the study evaluated the effects of normalization techniques (z-score and quantile normalization), data transformation (logarithmic scale) on feature selection algorithms (random forest) and followed discrimination ability based on the geographical origin. Different panels of molecules were generated using preprocessing and processing techniques. The combination of z-score normalization, log-transformation, and RF machine learning algorithm to the aligned data matrix gave the best results in classification, generating a pool of 16 molecules (mainly terpenes) and enhancing the classification of the samples inside their corresponding geographical origin.

Despite the sample numerosity being low to assert the identification of any markers of the geographical origin of raw pistachio, this proof-of-concept proved, for the first time, the feasibility of using the VOC profile of raw pistachio for this purpose. Although multivariate analysis showed promise in extracting information for classification purposes, we are aware that to confirm these results and minimize the risk of experimental errors in the identification of (bio)markers able to determine geographic origins, extensive studies involving a larger number of pistachio samples from different origins are needed, as well as a full statistical validation of the model with an external dataset.

## CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Andrea Schincaglia: Writing - review & editing, Writing - original draft, Software, Investigation, Formal analysis, Data curation. Luisa Pasti: Writing - review & editing, Writing - original draft, Visualization, Supervision. Alberto Cavazzini: Writing - review & editing, Writing - original draft, Visualization, Supervision. Giorgia Purcaro: Writing - review & editing, Writing - original draft, Visualization, Supervision, Resources, Methodology. Marco Beccaria: Writing - review & editing, Writing - original draft, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

## ACKNOWLEDGMENTS

GP, AS, and MB thank Supelco (MilliporeSigma), Markes, SepSolve, and Shimadzu for their continuous support. This article is based upon work from the Sample Preparation Study Group and Network, supported by the Division of Analytical Chemistry of the European Chemical Society.

## APPENDIX A. SUPPLEMENTARY DATA

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2024.140702

*Published in : Food Chemistry (2024), vol. 460, part 3, 140702*
*DOI: https://doi.org/10.1016/j.foodchem.2024.140702*
*Status : Postprint (Author's version)*

# References

Acree, T., & Heinrich, A. (2004). *Flavornet and human odor space*. https://www.flavornet.org/

Amaratunga, D., & Cabrera, J. (2001). Analysis of data from viral DNA microchips. *Journal of the American Statistical Association*, 96(456), 1161–1170. https://doi.org/10.1198/016214501753381814

Aspromonte, J., Mascrez, S., Eggermont, D., & Purcaro, G. (2023). Solid-phase microextraction coupled to comprehensive multidimensional gas chromatography for food analysis. *Analytical and Bioanalytical Chemistry*, 2023, 1–26. https://doi.org/10.1007/S00216-023-05048-0

Beccaria, M., Bobak, C., Maitshotlo, B., Mellors, T. R., Purcaro, G., Franchina, F. A., … Hill, J. E. (2018). Exhaled human breath analysis in active pulmonary tuberculosis diagnostics by comprehensive gas chromatography-mass spectrometry and chemometric techniques. *Journal of Breath Research*, 13(1), Article 016005. https://doi.org/10.1088/1752-7163/AAE80E

Beccaria, M., Zou, Y., Stefanuto, P. H., Siqueira, A. L. M., Maniquet, A., Piparo, M., … Focant, J. F. (2022). Deeper investigation of oxygen-containing compounds in oleaginous feedstock (animal fat) by preparative column chromatography and comprehensive two-dimensional gas chromatography coupled with high-resolution time-of-flight mass spectrometry. *Talanta*, 238, Article 123019. https://doi.org/10.1016/J.TALANTA.2021. 123019

Bolstad, B. M., Irizarry, R. A., Åstrand, M., & Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, 19(2), 185–193. https://doi.org/10.1093/BIOINFORMATICS/19.2.185

Boncan, D. A. T., Tsang, S. S. K., Li, C., Lee, I. H. T., Lam, H. M., Chan, T. F., & Hui, J. H. L. (2020). Terpenes and Terpenoids in plants: Interactions with environment and insects. *International Journal of Molecular Sciences*, 21(19), 1–19. https://doi.org/10.3390/IJMS21197382

Boukid, F., Abbattangelo, S., Carini, E., Marseglia, A., Caligiani, A., & Vittadini, E. (2019). Geographical origin discrimination of pistachio (Pistacia vera L.) through combined analysis of physical and chemical features. *European Food Research andTechnology*, 245(1), 143–150. https://doi.org/10.1007/S00217-018-3146-9/FIGURES/4

Bouwmeester, H., Schuurink, R. C., Bleeker, P. M., & Schiestl, F. (2019). The role of volatiles in plant communication. *The Plant Journal*, 100(5), 892. https://doi.org/10.1111/TPJ.14496

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324/ METRICS

Cabezas, L. M. C., Izbicki, R., & Stern, R. B. (2023). Hierarchical clustering: Visualization, feature importance and model selection. *Applied Soft Computing*, 141, Article 110303. https://doi.org/10.1016/J.ASOC.2023.110303

De Flaviis, R., Sacchetti, G., & Mastrocola, D. (2021). Wheat classification according to its origin by an implemented volatile organic compounds analysis. *Food Chemistry*, 341, Article 128217. https://doi.org/10.1016/J.FOODCHEM.2020.128217

Eggermont, D., Spadafora, N. D., Aspromonte, J., Pellegrino, R., & Purcaro, G. (2023). Exploring different high-capacity tools and extraction modes to characterize the aroma of brewed coffee. *Analytical and Bioanalytical Chemistry*, 415(18), 4501–4510. https://doi.org/10.1007/s00216-023-04654-2

Eggermont, D., Spadafora, N. D., Aspromonte, J., & Purcaro, G. (2023). Unraveling the impact of the capsule material on the aroma of brewed coffee by headspace analysis using a HiSorb probe followed by reverse fill/flush flow modulation GC×GC-MS. *Analytical and Bioanalytical Chemistry*, 415(13), 2511–2521. https://doi.org/10.1007/s00216-022-04457-x

Elia, S., Stylianou, M., & Agapiou, A. (2023). Combined EC/EO processes for treating goat cheese whey wastewater. *Sustainable Chemistry and Pharmacy*, 32, Article 100963. https://doi.org/10.1016/J.SCP. 2022.100963

Elloumi, O., Fki, M., Benmoussa, H., Feki, M., Ben Mimoun, M., & Ghrab, M. (2024). Agro-climatic requirements and reproductive potentials of pistachio pollinator specimens grown in warm area in North Africa. *Scientia Horticulturae*, 325, Article 112649. https://doi.org/10.1016/J.SCIENTA.2023.112649

*Published in : Food Chemistry (2024), vol. 460, part 3, 140702*
*DOI: https://doi.org/10.1016/j.foodchem.2024.140702*
*Status : Postprint (Author's version)*

Fruits and Nuts. (2007). In C. Kole (Ed.), *Fruits and nuts*. Berlin Heidelberg: Springer. https://doi.org/ 10.1007/978-3-540-34533-6

Georgiadou, M., Gardeli, C., Komaitis, M., Tsitsigiannis, D. I., Paplomatas, E. J., Sotirakoglou, K., & Yanniotis, S. (2015). Volatile profiles of healthy and aflatoxin contaminated pistachios. *Food Research International*, 74, 89–96. https://doi.org/10.1016/J.FOODRES.2015.03.021

Gil-Solsona, R., Raro, M., Sales, C., Lacalle, L., Díaz, R., Ibáñez, M., … Hernández, F. J. (2016). Metabolomic approach for extra virgin olive oil origin discrimination making use of ultra-high performance liquid chromatography – Quadrupole time-offlight mass spectrometry. *Food Control*, 70, 350–359. https://doi.org/10.1016/J.FOODCONT.2016.06.008

Güler, Z., Dursun, A., & Türkmen, D. (2022). Comparison of pistachio (Pistacia vera L.) kernel volatile compounds of "Siirt", "Uzun" and "Kirmizi" grown in Turkey: Effect of traditional and industrial in-shell roasting. *Journal of Food Process Engineering*, 45 (2), Article e13957. https://doi.org/10.1111/JFPE.13957

Hammouri, H. M., Sabo, R. T., Alsaadawi, R., & Kheirallah, K. A. (2020). Handling skewed data: A comparison of two popular methods. *Applied sciences*, 10(18), 6247. https://doi.org/10.3390/APP10186247

Hicks, S. C., & Irizarry, R. A. (2015). Quantro: A data-driven approach to guide the choice of an appropriate normalization method. *Genome Biology*, 16(1), 1–8. https://doi.org/10.1186/s13059-015-0679-0

Hojjati, M., Calín-Sánchez, Á., Razavi, S. H., & Carbonell-Barrachina, Á. A. (2013). Effect of roasting on colour and volatile composition of pistachios (Pistacia vera L.). *International Journal of Food Science & Technology*, 48(2), 437–443. https://doi.org/10.1111/J.1365-2621.2012.03206.X

Kalogiouri, N. P., Manousi, N., Klaoudatos, D., Spanos, T., Topi, V., & Zachariadis, G. A. (2021). Rare earths as authenticity markers for the discrimination of Greek and Turkish pistachios using elemental metabolomics and Chemometrics. *Foods*, 10(2), 349. https://doi.org/10.3390/FOODS10020349

Kashaninejad, M., & Tabil, L. G. (2011). Pistachio (Pistacia vera L.). *Postharvest Biology and Technology of Tropical and Subtropical Fruits*, 4, 218–247e. https://doi.org/10.1533/9780857092618.218

Kendirci, P., Tomris, Onoğur, A., Altu, T., & Gur, O. (2011). Investigation of volatile compounds and characterization of flavor profiles of fresh pistachio nuts (Pistacia vera L.). *International Journal of Food Properties*, 14(2), 319–330. https://doi.org/10.1080/10942910903177830

Kolb, B., & Ettre, L. S. (2006). Static headspace gas chromatography - theory and practice - Kobb, Ettre, 2nd ed. Lenzi, A., Biagini, D., Ghimenti, S., Vivaldi, F. M., Salvo, P., Di Francesco, F., & Lomonaco, T. (2023). HiSorb sorptive extraction for determining salivary short chain fatty acids and hydroxy acids in heart failure patients. *Journal of Chromatography B*, 1228, Article 123826. https://doi.org/10.1016/J.JCHROMB.2023.123826

Lytou, A. E., Panagou, E. Z., & Nychas, G. J. E. (2019). Volatilomics for food quality and authentication. *Current Opinion in Food Science*, 28, 88–95. https://doi.org/10.1016/J.COFS.2019.10.003

Mandalari, G., Barreca, D., Gervasi, T., Roussell, M. A., Klein, B., Feeney, M. J., & Carughi, A. (2021). Pistachio nuts (Pistacia vera L.): Production, nutrients, bioactives and novel health effects. *Plants*, 11(1), 18. https://doi.org/ 10.3390/PLANTS11010018

Mannino, G., Gentile, C., & Maffei, M. E. (2019). Chemical partitioning and DNA fingerprinting of some pistachio (Pistacia vera L.) varieties of different geographical origin. *Phytochemistry*, 160, 40–47. https://doi.org/10.1016/ J.PHYTOCHEM.2019.01.010

Marrubini, G., Dugheri, S., Cappelli, G., Arcangeli, G., Mucci, N., Appelblad, P., Melzi, C., & Speltini, A. (2020). Experimental designs for solid-phase microextraction method development in bioanalysis: A review. *Analytica Chimica Acta*, 1119, 77–100. https://doi.org/10.1016/J.ACA.2020.04.012

Mascrez, S., Aspromonte, J., Spadafora, N. D., & Purcaro, G. (2024). Vacuum-assisted and multi-cumulative trapping in headspace solid-phase microextraction combined with comprehensive multidimensional chromatography-mass

*Published in : Food Chemistry (2024), vol. 460, part 3, 140702*
*DOI: https://doi.org/10.1016/j.foodchem.2024.140702*
*Status : Postprint (Author's version)*

spectrometry for profiling virgin olive oil aroma. *Food Chemistry*, 442, Article 138409. https://doi.org/10.1016/J.FOODCHEM.2024.138409

Mascrez, S., & Purcaro, G. (2020). Enhancement of volatile profiling using multiplecumulative trapping solid-phase microextraction. Consideration on sample volume. *Analytica Chimica Acta*, 1122, 89–96. https://doi.org/10.1016/J.ACA.2020.05.007

Medina, S., Perestrelo, R., Santos, R., Pereira, R., & Cˆamara, J. S. (2019). Differential volatile organic compounds signatures of apple juices from Madeira Island according to variety and geographical origin. *Microchemical Journal*, 150, Article 104094. https://doi.org/10.1016/J.MICROC.2019.104094

Ninkuu, V., Zhang, L., Yan, J., Fu, Z., Yang, T., & Zeng, H. (2021). Biochemistry of terpenes and recent advances in plant protection. *International Journal of Molecular Sciences*, 22(11). https://doi.org/10.3390/IJMS22115710

Noguera-Artiaga, L., Salvador, M. D., Fregapane, G., Collado-González, J., Wojdyło, A., López-Lluch, D., & Carbonell-Barrachina, Á. A. (2019). Functional and sensory properties of pistachio nuts as affected by cultivar. *Journal of the Science of Food and Agriculture*, 99(15), 6696–6705. https://doi.org/10.1002/JSFA.9951

Pawliszyn, J. (2000). Theory of solid-phase microextraction. *Journal of Chromatographic Science*, 38(7), 270–278. https://doi.org/10.1093/CHROMSCI/38.7.270

Penci, M. C., Martinez, M. L., Fabani, M. P., Feresin, G. E., Tapia, A., Ighani, M., … Wunderlin, D. A. (2013). Matching changes in sensory evaluation with physical and chemical parameters: A case study: Argentinean pistachio nuts (Pistachia vera L. cv Kerman). *Food and Bioprocess Technology*, 6(12), 3305–3316. https://doi.org/10.1007/s11947-012-0993-4

Pistachios | USDA Foreign Agricultural Service. (2024). https://fas.usda.gov/data/prod uction/commodity/ 0577907.

Polari, J. J., Zhang, L., Ferguson, L., Maness, N. O., & Wang, S. C. (2019). Impact of microclimate on fatty acids and volatile terpenes in "Kerman" and "Golden Hills" pistachio (Pistacia vera) kernels. *Journal of Food Science*, 84(7), 1937–1942. https://doi.org/10.1111/1750-3841.14654

Purcaro, G., Stefanuto, P. H., Franchina, F. A., Beccaria, M., Wieland-Alter, W. F., Wright, P. F., & Hill, J. E. (2018). SPME-GC×GC-TOF MS fingerprint of virallyinfected cell culture: Sample preparation optimization and data processing evaluation. *Analytica Chimica Acta*, 1027, 158–167. https://doi.org/10.1016/J.ACA.2018.03.037

Rosenkranz, M., Chen, Y., Zhu, P., & Vlot, A. C. (2021). Volatile terpenes - mediators of plant-to-plant communication. *The Plant Journal : For Cell and Molecular Biology*, 108 (3), 617–631. https://doi.org/10.1111/ TPJ.15453

Roxas, A. A., Marino, G., Avellone, G., Caruso, T., & Marra, F. P. (2020). The effect of plant water status on the chemical composition of pistachio nuts (Pistacia vera L. cultivar Bianca). *Agriculture*, 10(5), 167. https://doi.org/10.3390/AGRICULTURE10050167

Şahan, A., & Bozkurt, H. (2020). Effects of harvesting time and irrigation on aroma active compounds and quality parameters of pistachio. *Scientia Horticulturae*, 261, Article 108905. https://doi.org/10.1016/ J.SCIENTA.2019.108905

Sater, H. M., Bizzio, L. N., Tieman, D. M., & Muñoz, P. D. (2020). A review of the fruit volatiles found in blueberry and other Vaccinium species. *Journal of Agricultural and Food Chemistry*, 68(21), 5777–5786. https://doi.org/10.1021/ACS.JAFC.0C01445/ASSET/IMAGES/LARGE/JF0C01445_0001.JPEG

Stilo, F., Liberto, E., Spigolon, N., Genova, G., Rosso, G., Fontana, M., … Cordero, C. (2021). An effective chromatographic fingerprinting workflow based on comprehensive two-dimensional gas chromatography – Mass spectrometry to establish volatiles patterns discriminative of spoiled hazelnuts (Corylus avellana L.). *Food Chemistry*, 340, Article 128135. https://doi.org/10.1016/J.FOODCHEM.2020.128135

Valdés García, A., Sánchez Romero, R., Juan Polo, A., Prats Moya, S., Pérez Maestre, S. E., & Beltrán Sanahuja, A. (2021). Volatile profile of nuts, key odorants and analytical methods for quantification. *Foods*, 10(7), 1611. https://doi.org/10.3390/FOODS10071611

*Published in : Food Chemistry (2024), vol. 460, part 3, 140702*
*DOI: https://doi.org/10.1016/j.foodchem.2024.140702*
*Status : Postprint (Author's version)*

van Vorstenbosch, R., Mommers, A., Pachen, D., van Schooten, F. J., & Smolinska, A. (2024). The optimization and comparison of two high-throughput faecal headspace sampling platforms: The microchamber/thermal extractor and hi-capacity sorptive extraction probes (HiSorb). *Journal of Breath Research*, 18(2), Article 026007. https://doi.org/10.1088/1752-7163/AD2002

Wang, B., Fang, A., Heim, J., Bogdanov, B., Pugh, S., Libardoni, M., & Zhang, X. (2010). DISCO: Distance and spectrum correlation optimization alignment for twodimensional gas chromatography time-of-flight mass spectrometry-based metabolomics. *Analytical Chemistry*, 82(12), 5069–5081. https://doi.org/10.1021/ AC100064B/ASSET/IMAGES/LARGE/AC-2010-00064B_0010.JPEG

Wei, X., Shi, X., Kim, S., Zhang, L., Patrick, J. S., Binkley, J., … Zhang, X. (2012). Data preprocessing method for liquid chromatography-mass spectrometry based metabolomics. *Analytical Chemistry*, 84(18), 7963–7971. https://doi.org/ 10.1021/AC3016856/SUPPL_FILE/AC3016856_SI_001.PDF

Wilson, J. S., Petino, G., & Knudsen, D. C. (2018). Geographic context of the green pistachio of Bronte, a protected designation of origin product. *Journal of Maps*, 14(2), 144–150. https://doi.org/10.1080/ 17445647.2018.1438318

Zakaria, S. R., Saim, N., Osman, R., Haiyee, Z. A., & Juahir, H. (2018). Combination compounds for the discrimination of authentic and unauthentic Harumanis mangoes. *Molecules*, 23(9), 2365. https://doi.org/ 10.3390/MOLECULES23092365

Zhang, Z., & Li, G. (2010). A review of advances and new developments in the analysis of biological volatile organic compounds. *Microchemical Journal*, 95(2), 127–139. https://doi.org/10.1016/J.MICROC.2009.12.017

Zhao, Y., Wong, L., Goh, W., & Bin, W. (2020). How to do quantile normalization correctly for gene expression data analyses. *Scientific Reports,* 10(1), 1–11. https://doi.org/10.1038/s41598-020-72664-6