Rapid identification of the volatile markers in lotus seed using headspace-gas chromatography ion-mobility spectrometry

Xiaoluan Tang[1,2], Min Ni[1], Xiao Guo[3], Jiayao Liu[3], Shanshuo Liu[3], Dan Huang[3,*]

[1]School of Pharmacy, Fujian Health College, Fuzhou, China; [2]College of Chemistry, Fuzhou University, Fuzhou, China; [3]State Key Laboratory of Chinese Medicine Powder and Medicine Innovation in Hunan (Incubation), Science and Technology Innovation Center, Hunan University of Chinese Medicine, Changsha, China

***Corresponding Author:** Dan Huang, State Key Laboratory of Chinese Medicine Powder and Medicine Innovation in Hunan (Incubation), Science and Technology Innovation Center, Hunan University of Chinese Medicine, Changsha 410208, China. Email: huangdan110@hnucm.edu.cn

Academic Editor: Mehran Moradi, PhD., Department of Food Hygiene and Quality Control, Faculty of Veterinary Medicine, Urmia University, Urmia, Iran

Received: 8 October 2024; Accepted: 23 December 2024; Published: 22 January 2025

ORIGINAL ARTICLE

Abstract

To rapidly identify the volatile markers in lotus seeds, this research compared the volatile organic compounds (VOCs) of five lotus seed samples using gas chromatography-ion mobility spectroscopy (GC-IMS) and chemometric analysis. The results revealed that 49 VOCs were identified from the lotus seed samples, including 16 aldehydes, 15 alcohols, 8 ketones, 3 esters, 3 acids, 3 terpenes, and 1 heterocyclic compound, respectively. Among these, 1-pentanol M, 2-methylbutan-1-ol M, 2-methylbutan-1-ol D, and 1-hexanol M were identified as the volatile markers. Based on the VOC analysis using GC-IMS, effective differentiation of lotus seeds was achieved through chemometric techniques, such as principal component analysis (PCA), cluster analysis (CA), and partial least squares discriminant analysis (PLS-DA). 2-furaldehyde D, hexanoic acid D, and 1-propanol M were found to be the most important components contributing to the differences among the five lotus seed samples. This research demonstrates that GC-IMS coupled with chemometric analysis provides valuable reference information for the identification and authenticity evaluation of lotus seeds, helping ensure their quality on the market and offering theoretical support for their identification and quality assessment.

Keywords: cluster analysis; gas chromatography ion-mobility spectrometry; lotus seed; partial least squares discriminant analysis; principal component analysis; volatile organic compounds

Introduction

Lotus seeds are widely used worldwide and hold significant medicinal and nutritional value. They have a history of over 2000 years as a functional food (Yu *et al*., 2022; Feng *et al*., 2016). Lotus seeds are the dried, ripe seeds of *Nelumbo nucifera* Gaertn. They are harvested in the autumn when ripe, separated from the pericarp, and dried (Chinese Pharmacopoeia Commission, 2020). Lotus seed-based foods include lotus seed porridge, lotus seed cake, lotus seed biscuits, lotus seed tea, lotus seed soup, and more. These seeds are typically used to tonify the spleen, relieve diarrhea, replenish the kidneys, arrest seminal emission, nourish the heart, and induce tranquility. They are also used to treat leukorrhea, palpitations, and insomnia. Lotus seeds contain a variety of phytochemicals, including alkaloids, flavonoids, polysaccharides, essential oils, glycosides, polyphenols, and triterpenes

(Chen *et al*., 2012; Huang *et al*., 2010; Zheng *et al*., 2010). Lotus seeds are abundant in China, with varieties such as Jianning lotus seeds from Jianning, Fujian province; Xiangtan lotus seeds from Xiangtan, Hunan province; and Xuanping lotus seeds from Xuanping, Zhejiang province, collectively known as the three major lotus seeds in China. Additionally, lotus seeds from Guangchang, Jiangxi province, and lotus seeds from Honghu, Hubei province, have been recognized as national geographical indication products in China.

Volatile organic compounds (VOCs) are key components (either harmful or beneficial) in food, and the types and concentrations of volatile components can vary significantly across different food types. VOCs are closely related to the quality, flavor, and aroma of food. They play a crucial role in identifying different varieties and origins of food, providing a theoretical basis for food quality control, origin traceability, and flavor analysis (Yin *et al*., 2024; Duan *et al*., 2023). Gas chromatography–mass spectrometry (GC-MS) has been used to analyze the volatile constituents of Nelumbinis stamen and Nelumbinis plumula in different medicinal parts of the lotus (Wang *et al*., 2020); however, there is currently no research analyzing the VOCs of lotus seeds for identification purposes.

Several fast methods are currently available for identifying plant-based foods, including gas chromatography-flame ionization detection (GC-FID) (Welke *et al*., 2022; Nedeltcheva-Antonova *et al*., 2022), gas chromatography with electron capture detection (GC-ECD) (da Silva *et al*., 2015; Pendem *et al*., 2010), gas chromatography–mass spectrometry (GC-MS), and near-infrared spectroscopy (NIRS). In addition to these techniques, gas chromatography–ion mobility spectroscopy (GC-IMS) may serve as a feasible alternative to traditional flavor analysis methods (Wang *et al*., 2019; Zhu, *et al*., 2023; Duan *et al*., 2023; Li *et al*., 2022; Tian *et al*., 2020; Valli *et al*., 2020).

GC-IMS can qualitatively analyze different plant-based foods and enable the rapid and accurate identification and separation of these foods through fingerprint recognition and isomer differentiation of VOCs. The robustness and simplicity of the instrument significantly enhance the application of GC-IMS in food certification, processing, storage monitoring, identification of illegal additives, and detection of harmful compounds.

GC-IMS is a powerful technique that combines gas chromatography and ion mobility spectroscopy for the separation and sensitive detection of VOCs. It is characterized by fast response speed, high sensitivity, ease of operation, and low cost. In the field of food analysis, it has been widely used for various purposes, including flavor and quality analysis, trace detection of toxic chemicals, and adulteration identification. The rapid, non-destructive, high-throughput detection and screening of volatile components play a crucial role in food flavor analysis (Valli *et al*., 2020; Zhang *et al*., 2020; Tian *et al*., 2020; Yan *et al*., 2024). For example, GC-IMS has been employed in flavor analysis to evaluate VOCs in apple cider (Wu *et al*., 2023), tea (Xu *et al*., 2023), and yellow croaker (Zhao *et al*., 2021).

PCA is an unsupervised machine learning and statistical technique used to identify patterns and relationships in large datasets. It is a widely used method for data and dimensionality reduction, which involves reducing the number of variables in the dataset while preserving as much original information as possible (Younes *et al*., 2023). CA is an unsupervised machine learning algorithm designed to identify subgroups in a dataset, characterized by discrete differences. Due to this unique feature, CA has gradually gained popularity over traditional statistical analysis (Dalmaijer *et al*., 2022). It has been widely used to analyze changes in food ingredients (Duan *et al*., 2023). PLS-DA is a supervised analysis method that can more effectively reveal differences and similarities between groups compared to unsupervised PCA analysis (Yin, *et al*., 2024).

In this study, GC-IMS technology was combined with principal component analysis (PCA), cluster analysis (CA), and partial least squares discriminant analysis (PLS-DA) to detect and analyze VOCs in five lotus seed varieties from China, enabling the rapid identification of volatile markers in lotus seeds. This approach provides valuable insights for the identification of lotus seeds, which helps ensure their quality in the market and offers theoretical support for the quality evaluation of lotus seeds.

Materials and Methods

Materials

The five types of lotus seeds (dried, pre-packaged) were purchased from Jingdong, Beijing, China, and then crushed into powder. The lotus seeds from Jianning, Fujian province of China (Jianlian) were designated as LZ-01; the lotus seeds from Xiangtan, Hunan province of China (Xianglian) were designated as LZ-02; the lotus seeds from Xuanping, Zhejiang province of China (Xuanlian) were designated as LZ-03; the lotus seeds from Guangchang, Jiangxi province of China (Guangchang white lotus seed) were designated as LZ-04; and the lotus seeds from Honghu, Hubei province of China (Honghu lotus seed) were designated as LZ-05. A voucher specimen (HNUCM2024-LZ001) was stored in the sample room of the Science and Technology Innovation Center at Hunan University of Chinese Medicine.

Analysis by GC–IMS

Sample Preparation

An amount of 0.5 g of the powder from each sample was placed in a 20 mL headspace vial.

Headspace conditions

The samples were incubated at 80 °C for 15 min. After incubation, 500 μL were injected into the headspace using non-shunt injection, and the vials were rotated at 500 revolutions per minute (rpm) for 20 mins. The temperature of the injection needle was set to 85°C.

GC conditions

The instrument used in this study was the FlavorSpec® Gas Phase Ion Mobility Spectrometer from GAS (Dortmund, Germany). An MXT-WAX GC Metal Capillary Column (15 m × 0.53 mm × 1.0 μm, Restek Inc, USA) was used for chromatography. The column temperature was set to 60°C. Initially, 2.00 mL/min of high-purity N2 was used as the carrier gas. Over the course of 8 min, the flow rate increased linearly to 10.00 mL/min, then to 100.00 mL/min over the next 10 min, and was held for an additional 10 min. The chromatography runtime was 30 min, and the injection temperature was set to 80°C.

IMS conditions

The analysis was conducted using tritium (^{3}H), a 53 mm drift tube, with an electric field intensity of 500 V/cm, a drift tube temperature of 45°C, and high-purity N2 (99.999%) at a flow rate of 150 mL/min, in positive ionization mode.

Statistical analysis

This instrument is paired with the analysis software Vocal, which displays spectra and data for both qualitative and quantitative analysis. Databases from NIST and IMS are included in the application software to facilitate the qualitative analysis of substances. A Porter plugin was used to analyze and compare the differences in spectral characteristics between samples, including three-dimensional spectra, two-dimensional top views, and difference spectra. VOCs were intuitively and quantitatively compared using the gallery plot plugin. Principal component analysis (PCA) was performed using OriginPro 2023b software, cluster analysis (CA) was conducted using TBtools, and partial least squares discriminant analysis (PLS-DA) was performed using SIMCA.

Research manuscripts reporting large datasets deposited in publicly available databases should specify the location where the data have been deposited and provide the relevant accession numbers. If the accession numbers have not yet been obtained at the time of submission, please indicate that they will be provided during the review. The accession numbers must be provided before publication.

Interventional studies involving animals or humans, as well as other studies requiring ethical approval, must specify the authority that provided the approval and include the corresponding ethical approval code.

Results and Discussion

GC-IMS Analysis of VOCs in five lotus seed samples

The three-dimensional spectrum of the VOCs is shown in Figure 1, with the samples labeled as LZ-01, LZ-02, LZ-03, LZ-04, and LZ-05. The x, y, and z axes in the figure represent drift time, gas chromatography retention time, and signal peak intensity, respectively. From the three-dimensional spectrum of the five lotus seed samples, it can be directly observed that there are distinct differences in the VOCs among the samples from different sources.

From Figure 1, it can be visually observed that there are certain differences in the VOCs among the samples from different sources. For easier comparison, the top view is shown in Figure 2. Each point on either side of the RIP peak represents a volatile organic compound. The color indicates the peak intensity of a substance, ranging from blue to red, with darker colors signifying higher peak intensity. The background of the figure is blue, and the red vertical line at 1.0 represents the RIP peak (reactive ion peak). The vertical axis represents the gas chromatography retention time (s), and the horizontal axis represents the relative drift time (normalized treatment).

To further visually compare the differences in their VOCs, the spectra of the LZ-01 sample were selected as a reference, and the spectra of the other samples were subtracted from the reference to generate a comparison chart of the differences between the samples, as shown in Figure 3. Based on Figure 2, if the VOC content in the target sample is the same as that in the LZ-01 sample, the point cancels out and is displayed as white. If the concentration of the substance in the target sample is higher than that in the LZ-01 sample, it is displayed as red; if it is lower, it is displayed as blue.

From Figure 1 to Figure 3, it is evident that there are certain differences in the VOCs of the five lotus seed samples. The specific differences in VOCs will be further analyzed based on fingerprint spectra and other data.

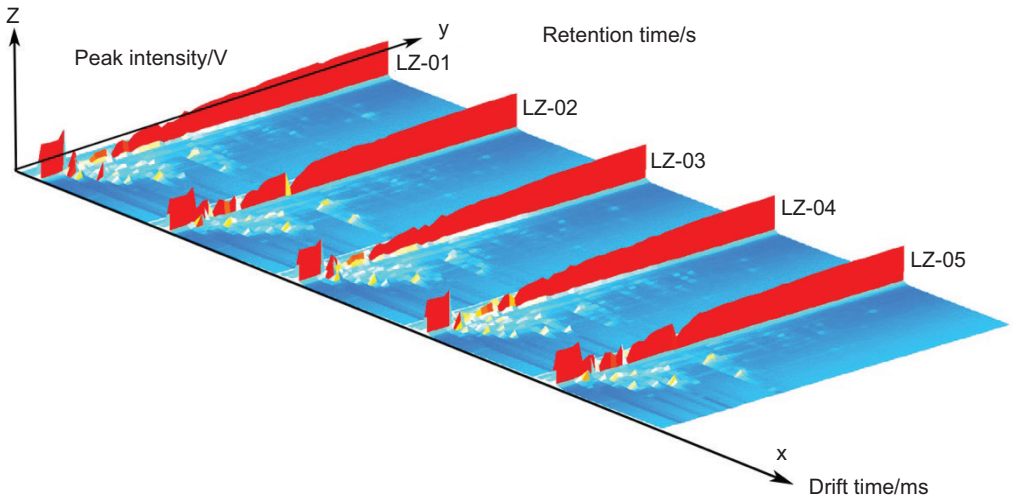

Figure 1. The 3D spectra of VOCs from five lotus seed samples.

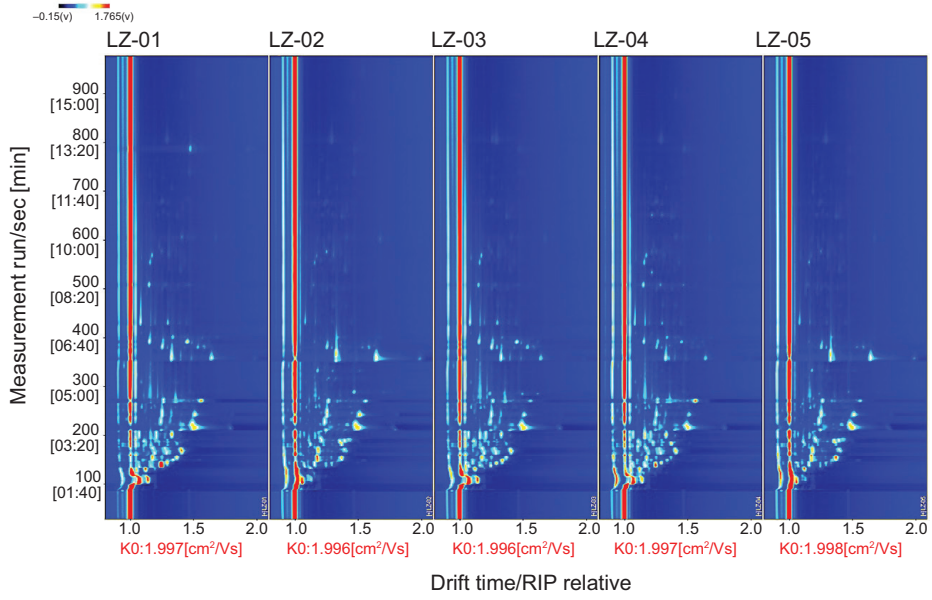

Figure 2. The 2D spectra of VOCs from five lotus seed samples. The graph is blue, with red vertical line representing the RIP (normalized reaction ion peak) at 1.0 on the horizontal axis. Retention time in gas chromatography is displayed on the vertical axis, and drift time is displayed on the horizontal axis (normalized, a.u.). Each point on either side of the RIP peak represents a volatile organic compound. The colors range from blue to red, with darker colors indicating higher peak intensities.

Qualitative analysis of VOCs in five lotus seed samples

In the GC-IMS two-dimensional spectrum, the difference in VOC content between the five groups of samples is reflected in the concentration of each volatile substance. Based on this, qualitative analysis of the VOCs was conducted by combining the NIST and IMS databases built into the software. This study detected a total of 49 VOCs, including 16 aldehydes, 15 alcohols, 8 ketones, 3 esters, 3 acids, 3 terpenes, and 1 pyridine. The qualitative analysis results of the VOCs are shown in Table 1.

Fingerprint analysis of VOCs in five lotus seed samples

The VOCs in the samples were further compared, and a fingerprint analysis was performed on all the volatile substances, as shown in Figure 4. The results of the comparison and analysis of the VOCs in samples LZ-01, LZ-02, LZ-03, LZ-04, and LZ-05 are also presented in Figure 4. As indicated in the purple box, nonanal, 2-ethyl-1-hexanol, 1-propanol, and benzeneacetaldehyde have a higher content in LZ-01. As shown in the red box, 1-hexanol, 1-pentanol, 2-methylbutan-1-ol, 3-methylbutan-1-ol,

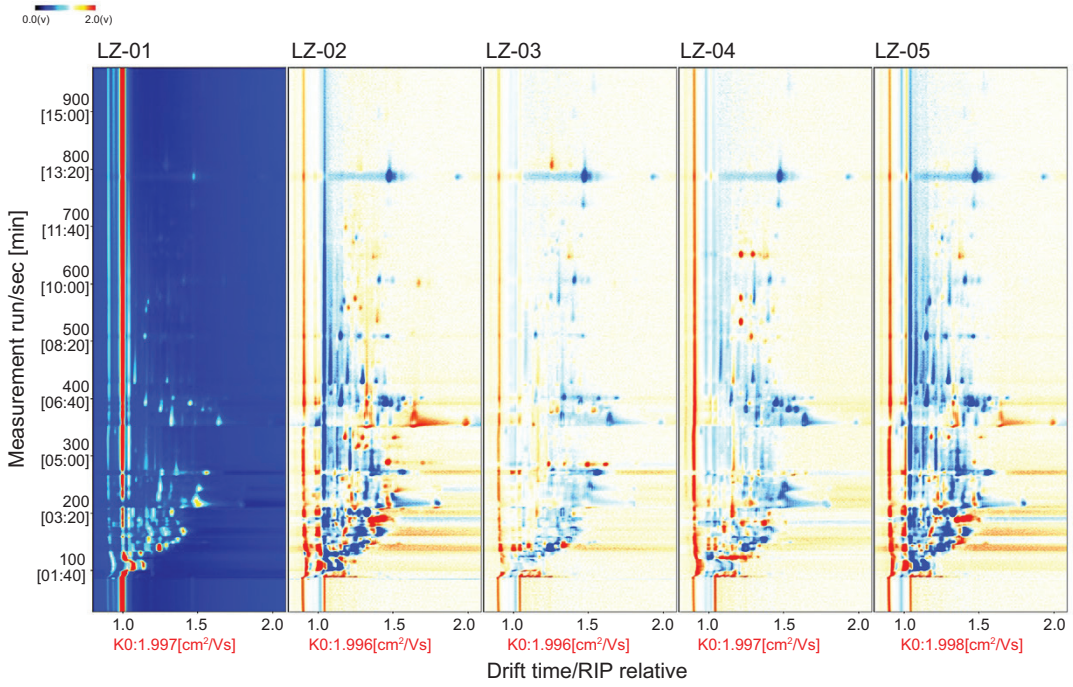

Figure 3. Differences in the 2D spectra of VOCs from five lotus seed samples.

pyridine, 2-hexenal, octanal, 2-furaldehyde, and 2-methylpropan-1-ol have a higher content in LZ-02 and LZ-05. As shown in the green box, 2-methyl-2-propanol, ethyl acetate, butyrolactone, 2-heptanone, cyclohexanone, and butanal have a higher content in LZ-01 and LZ-03. As indicated in the orange box, 1-octen-3-ol, benzaldehyde, 1-hexanal, butanoic acid, and 3-methyl-3-buten-1-ol have a higher content in LZ-01 and LZ-04. As shown in the yellow box, 2,3-butanedione, 1-hydroxy-2-propanone, 3-carene, and pinene have a higher content in LZ-04. Among these, 1-pentanol M, 2-methylbutan-1-ol M, 2-methylbutan-1-ol D, and 1-hexanol M are the volatile markers.

Chemometric analysis

Principal Component Analysis (PCA)

This study used OriginPro 2023b software to conduct a PCA on the VOCs in the five lotus seed samples. The results are shown in Figure 5, where different colors represent the different lotus seed samples. The PCA arranges the principal component scores from high to low based on the contribution rates and visualizes the scores of the first two principal components. According to Figure 5, PCA1 accounts for 59.3%, PCA2 accounts for 18.7%, and the cumulative contribution of the two principal components is 78%. The larger the distance between samples from different sources, the more significant the difference in VOCs among the samples. Conversely, the closer the distance between samples, the smaller the difference in VOCs. The distance between LZ-02 and LZ-05, as well as between LZ-01 and LZ-03, is relatively small, indicating that these samples have similar volatile organic compounds.

Cluster Analysis (CA)

To further analyze the differences in the VOCs among the five lotus seed samples, CA graphs were generated. Forty-nine VOCs from the lotus seed samples LZ-01, LZ-02, LZ-03, LZ-04, and LZ-05 were processed and imported into TBtools software for CA. The results are shown in Figure 6. It can be observed that LZ-05 and LZ-02 have the smallest difference in VOCs, while LZ-01, LZ-03, and LZ-04 have relatively small differences in their components, which is consistent with the results from the PCA plot. Among the components of LZ-04, 3-carene M, 3-carene D, pyridine, and 1-hydroxy-2-propanone have the highest content. In contrast, 2-ethyl-1-hexanol, 1-propanol D, nonanal D, and nonanal M are relatively abundant in the LZ-01 components. The results from the CA clearly reflect the differences in the content of each volatile organic compound among the different groups.

Partial Least Squares Discriminant Analysis (PLS-DA)

SIMCA software was used to perform PLS-DA on the various samples using a supervised pattern recognition method to observe the differences in the VOCs of the lotus seed samples LZ-01, LZ-02, LZ-03, LZ-04, and LZ-05. The results are shown in Figure 7(A). From the

Table 1. Results of the VOC analysis of five lotus seed samples.

No	Compounds	CAS	Molecular Formula	RI	Rt/s	Dt/ms
1	Nonanal M	C124196	$C_9H_{18}O$	1105.9	790.521	1.467
2	Nonanal D	C124196	$C_9H_{18}O$	1105.1	788.781	1.93816
3	1-Hexanol M	C111273	$C_6H_{14}O$	882.5	367.221	1.33177
4	1-Hexanol D	C111273	$C_6H_{14}O$	880.3	364.417	1.65191
5	1-Hexanol P	C111273	$C_6H_{14}O$	876.5	359.744	1.98831
6	Benzaldehyde D	C100527	C_7H_6O	967.3	512.602	1.46735
7	Benzaldehyde M	C100527	C_7H_6O	966.7	511.353	1.14862
8	Butyrolactone M	C96480	$C_4H_6O_2$	925.3	433.327	1.08407
9	Butyrolactone D	C96480	$C_4H_6O_2$	926	434.575	1.2979
10	Cyclohexanone M	C108941	$C_6H_{10}O$	903.5	397.122	1.16073
11	Cyclohexanone D	C108941	$C_6H_{10}O$	901.2	393.377	1.45525
12	Benzeneacetaldehyde	C122781	C_8H_8O	1054.1	680.516	1.25621
13	Hexanal D	C66251	$C_6H_{12}O$	792.9	270.599	1.5639
14	1-hexanal M	C66251	$C_6H_{12}O$	794.6	272.137	1.26461
15	Butanoic acid	C107926	$C_4H_8O_2$	804.4	281.367	1.16875
16	2-Methylbutan-1-ol D	C137326	$C_5H_{12}O$	735.6	218.623	1.50268
17	2-Methylbutan-1-ol M	C137326	$C_5H_{12}O$	747.9	228.978	1.23704
18	1-Pentanol D	C71410	$C_5H_{12}O$	766.2	245.35	1.51093
19	1-Pentanol M	C71410	$C_5H_{12}O$	770.2	249.092	1.25641
20	1-propanol D	C71238	C_3H_8O	617.3	142.437	1.24396
21	1-propanol M	C71238	C_3H_8O	622.9	145.244	1.10563
22	1-Propanol, 2-methyl- D	C78831	$C_4H_{10}O$	642.4	155.535	1.36845
23	1-Propanol, 2-methyl- M	C78831	$C_4H_{10}O$	644.1	156.471	1.17894
24	2-Pentanone D	C107879	$C_5H_{10}O$	694.8	187.345	1.3726
25	2-Pentanone M	C107879	$C_5H_{10}O$	692.8	185.941	1.11393
26	3-Methyl-3-buten-1-ol	C763326	$C_5H_{10}O$	715.8	202.782	1.16649
27	3-Methyl butanal M	C590863	$C_5H_{10}O$	673.1	173.311	1.18309
28	3-Methylbutanal D	C590863	$C_5H_{10}O$	671.6	172.376	1.40995
29	Ethyl Acetate	C141786	$C_4H_8O_2$	631.6	149.749	1.34133
30	Butanal	C123728	C_4H_8O	625.1	146.415	1.2973
31	2,3-Butanedione	C431038	$C_4H_6O_2$	615.1	141.34	1.1714
32	2-Hexenal D	C505577	$C_6H_{10}O$	855.2	334.622	1.51414
33	2-Hexenal M	C505577	$C_6H_{10}O$	855.9	335.331	1.18411
34	Hexanoic acid D	C142621	$C_6H_{12}O_2$	970	518.194	1.63352
35	Hexanoic acid M	C142621	$C_6H_{12}O_2$	965.5	508.98	1.2871
36	Pinene	C127913	$C_{10}H_{16}$	978.4	535.914	1.22039
37	2-furaldehyde M	C98011	$C_5H_4O_2$	833.3	310.524	1.0858
38	2-furaldehyde D	C98011	$C_5H_4O_2$	837.3	314.776	1.32455
39	3-Carene M	C13466789	$C_{10}H_{16}$	1040.5	654.279	1.22039
40	3-Carene D	C13466789	$C_{10}H_{16}$	1040.1	653.57	1.29295
41	Octanal	C124130	$C_8H_{16}O$	982.2	543.963	1.40269
42	6-Methyl-5-hepten-2-one	C110930	$C_8H_{14}O$	994.4	571.148	1.16736
43	1-Octen-3-ol	C3391864	$C_8H_{16}O$	986.1	552.701	1.16039
44	2-Ethyl-1-hexanol	C104767	$C_8H_{18}O$	1013.5	605.13	1.40618
45	2-Heptanone	C110430	$C_7H_{14}O$	897.5	387.648	1.25103
46	1-Hydroxy-2-propanone	C116096	$C_3H_6O_2$	691.8	185.219	1.22922
47	n-pentanal	C110623	$C_5H_{10}O$	702.5	192.84	1.42591
48	2-Methyl-2-propanol	C75650	$C_4H_{10}O$	541.2	109.009	1.14551
49	Pyridine	C110861	C_5H_5N	727.9	212.316	1.24637

The substance suffixes M, D, or P represent monomers, dimers, and polymers of the same substance, respectively. Dt represents drift time, Rt represents retention time, and RI represents retention index.

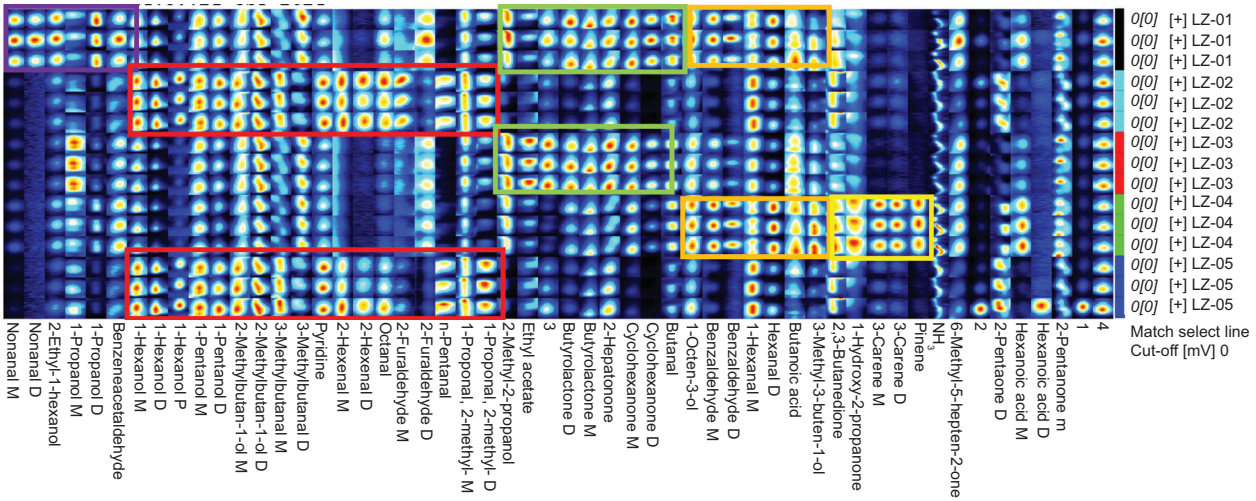

Figure 4. Fingerprints of VOCs from five lotus seed samples. Each row on the graph represents a sample with selected signal peaks. Each column represents the signal peak of the same VOC across samples. The graph provides complete information on the VOCs in each sample and highlights the differences between samples.

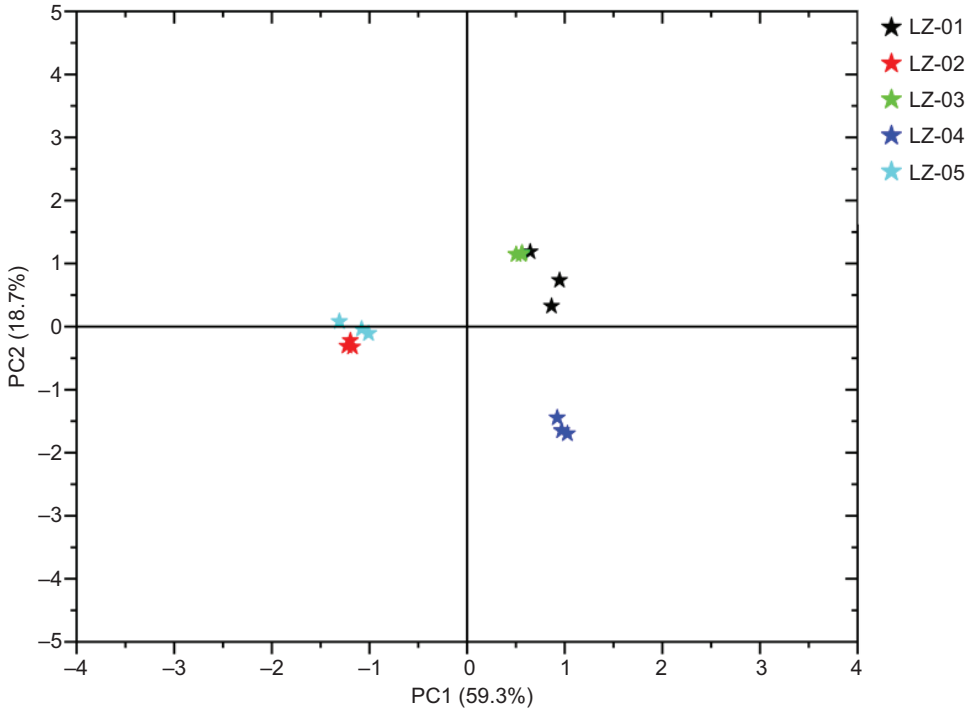

Figure 5. Results of the PCA of five lotus seed samples.

figure, it can be seen that the LZ-02 and LZ-05 samples are on the left side of the coordinate axis, with the distance between them being closer than to the other samples, indicating the smallest difference in VOCs. The LZ-01, LZ-03, and LZ-04 samples are all on the right side of the coordinate axis, but the LZ-01 and LZ-03 samples are closer to each other. LZ-04 does not overlap with the other samples and shows a clear distinction, indicating that the difference in VOCs between the LZ-01 and LZ-03 samples is small, while there is a significant difference between the LZ-04 samples and the other samples. This is consistent with the PCA results. Additionally, according to the processed data, R^2X = 0.963, R^2Y = 0.982, and Q^2 = 0.934. When R^2 and Q^2 are greater than 0.5, it indicates that the established model has relatively accurate generalization and predictive ability. A total of 200 permutation tests were conducted on the established PLS-DA model, with an R^2 intercept

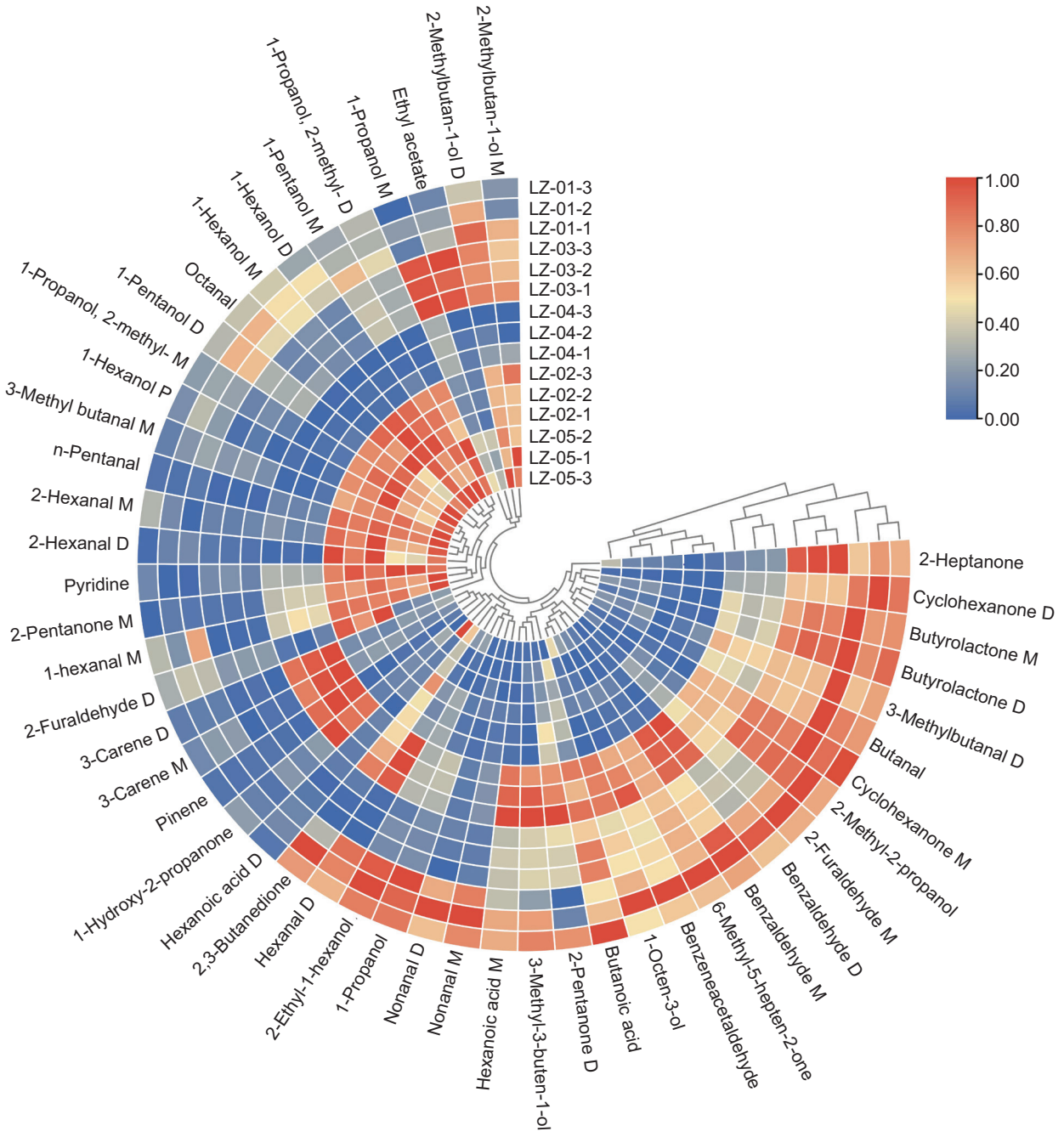

Figure 6. Results of the CA of five lotus seed samples.

value of 0.271 and a Q^2 intercept value of -0.602 (<0), indicating that the model results were not overfitting Figure 7(B). Furthermore, a variable projection importance map was constructed, as shown in Figure 7(C). The VIP value is a quantitative indicator of the influence of each volatile component on the lotus seed samples. The larger the VIP value, the more important the component. Variables with a VIP value greater than 1 are considered more important, while those with a VIP value less than 0.5 are considered unimportant. From Figure 7(C), it can be seen that the VIP values of 2-furaldehyde D, hexanoic acid D, 1-propanol M, nonanal D, ethyl acetate, nonanal M, 2-pentanone D, 1-propanol D, 2,3-butanedione, 3-carene M, and other substances are greater than 1, indicating that they are important components affecting the differences in the five lotus seed samples. The difference in VOCs between the LZ-02 and LZ-05 samples and the LZ-01 and LZ-03 samples is relatively small, and these results are consistent with the PCA findings.

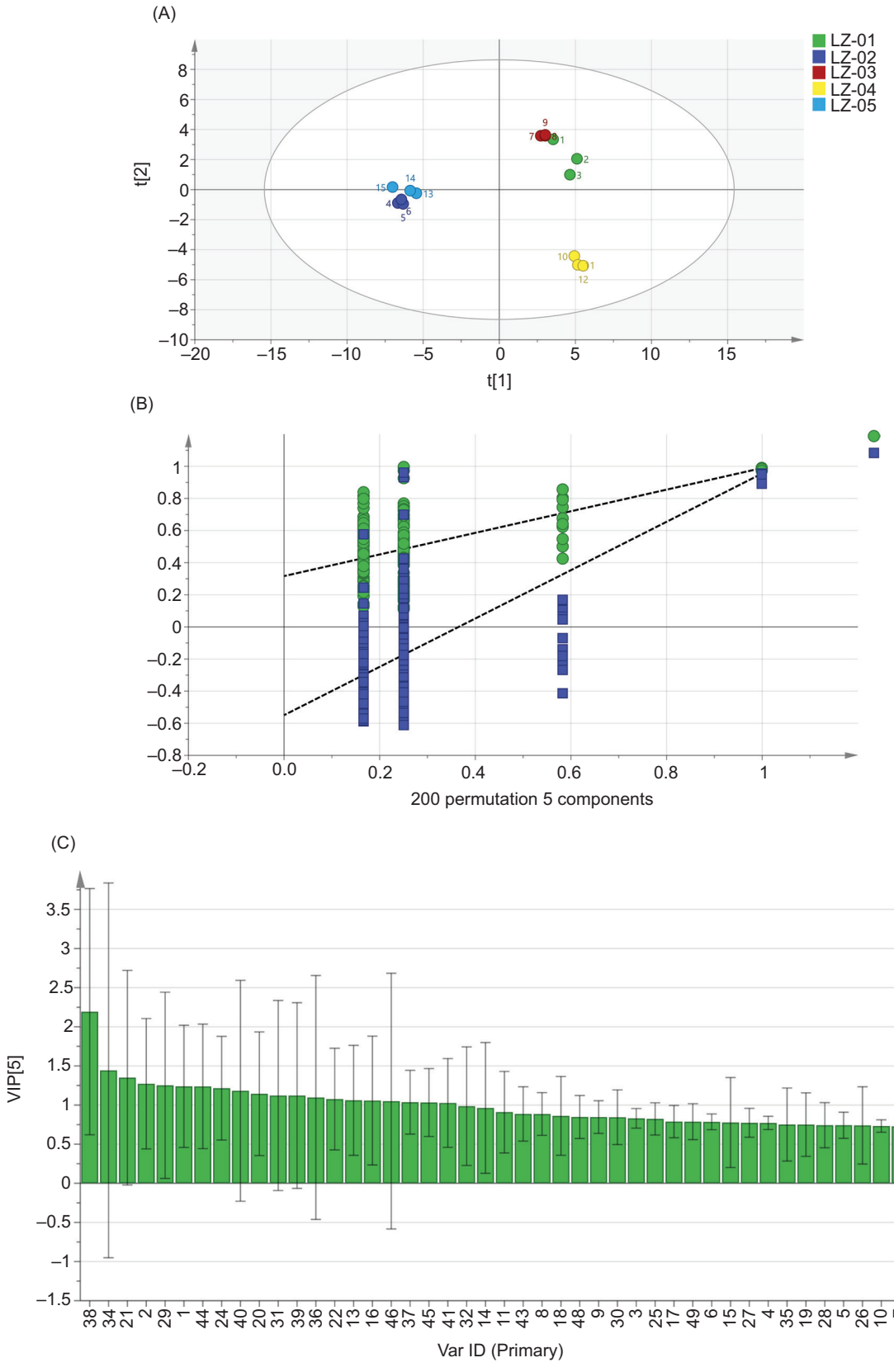

Figure 7. PLS-DA analysis of VOCs from five lotus seed samples: (A) score plots; (B) cross-validation plot with 200 permutation tests; (C) VIP diagram.

Modern research has shown that there are significant differences in the types and quantities of chemical components in food or traditional Chinese medicine from different geographical sources (Li *et al*., 2022; Hu *et al*., 2023; Zhang *et al*., 2023; Sammarco *et al*., 2023; Hai *et al*., 2023; Ji *et al*., 2023).

According to the research results, the differences in VOC content across the five groups of lotus seeds may be linked to their soil conditions, climate environment, and water quality. By analyzing the VOCs in lotus seeds using the GC-IMS method, lotus seeds can be identified.

Currently, there is limited research on the VOCs of lotus seeds in China. This study will contribute to the development of a range of lotus seed products rich in functional ingredients and promote the high-value utilization of lotus seeds.

Conclusions

Through the analysis of VOCs in five lotus seed varieties, it was found that there are distinct differences in the GC-IMS profiles among lotus seed varieties from different sources. Based on the feature components identified using the graph plugin software, a fingerprint map was created, which indicated that 1-Pentanol M, 2-Methylbutan-1-ol M, 2-Methylbutan-1-ol D, and 1-hexanal M are volatile markers. After performing PCA, CA, and PLS-DA, the lotus seed varieties from different sources were effectively and quickly distinguished. Additionally, VIP values were used in this experiment to assess the influence of each volatile component on the lotus seed samples, providing a basis for quality characteristic evaluation and germplasm identification. The combination of GC-IMS and chemometric analysis allows for efficient and rapid identification of lotus seeds. This research demonstrates that GC-IMS coupled with chemometric analysis provides valuable reference data for the identification and evaluation of lotus seeds from different sources, which aids in ensuring the quality of lotus seeds in the market and offers theoretical support for the identification and quality evaluation of lotus seeds from different sources.

Author Contributions

Conceptualization, X.T.; methodology, X.T., and M.N.; software, M.N., S.L.; validation, X.G.; formal analysis, X.G.; investigation, J.L.; resources, J.L.; data curation, S.L.; writing—original draft preparation, X.T., X.G., and J.L.; writing—review and editing, X.T. and D.H.; visualization, M.N.; supervision, D.H.; project administration, D.H.; funding acquisition, D.H. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Funding

This research was funded by the Hunan Provincial Natural Science Foundation of China (No. 2024JJ7358) and the Key Discipline Project on Chinese Pharmacology of Hunan University of Chinese Medicine [202302].

References

Chen, S., Fang, L., Xi, H., Guan, L., Fang, J., Liu, Y., Wu, B., & Li, S. (2012). Simultaneous qualitative assessment and quantitative analysis of flavonoids in various tissues of lotus (Nelumbo nucifera) using high performance liquid chromatography coupled with triple quad mass spectrometry. Analytica Chimica Acta, 724, 127–135. https://doi.org/10.1016/j.aca.2012.02.051

Chinese Pharmacopoeia Commission, 2020. *Pharmacopoeia of the People's Republic of China* (Vol. I). China Medical Science and Technology Press, Beijing, China.

da Silva, G.C., da Silva, A.A., da Silva, L.S., Godoy, R.L., Nogueira, L.C., Quitério, S.L., & Raices, R.S. (2015). Method development by GC-ECD and HS-SPME-GC-MS for beer volatile analysis. Food Chemistry, 167, 71–7. https://doi.org/10.1016/j.foodchem.2014.06.033

Dalmaijer, E., Nord, C., & Astle, D. (2022). Statistical power for cluster analysis. BMC Bioinformatics, 23, 205. https://doi.org/10.1186/s12859-022-04675-1

Duan, H., Zhou, Y., Wang, D., & Yan, W. (2023). Differences in Volatile Organic Compounds in Rhizoma gastrodiae (Tian Ma) of Different Origins Determined by HS-GC-IMS. Molecules, 28, 4883. https://doi.org/10.3390/molecules28134883

Feng, C.Y., Li, S.S., Yin, D.D., Zhang, H.J., Tian, D.K., Wu, Q, Wang, L.J., Su, S., & Wang, L.S. (2016). Rapid determination of flavonoids in plumules of sacred lotus cultivars and assessment of their antioxidant activities. Industrial Crops and Products, 87, 96–104. https://doi.org/10.1016/j.indcrop.2016.04.030

Hai, C., Yang, X., Fu, H., Chen, H., He, S., Kang, L., Hao, Q., Zhou, L., Zhan, Z., Liu, Z., Yang, J., & Guo, L. (2023). Determination of geographical origin for Atractylodes macrocephala Koidz by stable isotope and multielement analyses combined with chemometrics. Journal of Food Science, 88, 939–1953. https://doi.org/10.1111/1750-3841.16536

Hu Z., Luo Z., Wang Y., Zhou Q., Liu S., & Wang Q. (2023). Texture Feature Extraction from 1H NMR Spectra for the Geographical Origin Traceability of Chinese Yam. Foods, 12, 476. https://doi.org/10.3390/foods12132476

Huang, B., Ban, X., He, J., Tong, J., Tian, J., & Wang, Y. (2010). Comparative analysis of essential oil components and antioxidant activity of extracts of *Nelumbo nucifera* from various areas of China. Journal of Agricultural and Food Chemistry, 58(1), 441–448. https://doi.org/10.1021/jf902643e

Ji, Q., Li, C., Fu, X., Liao, J., Hong, X., Yu, X., Ye, Z., Zhang, M., & Qiu, Y. (2023). Protected Geographical Indication Discrimination of Zhejiang and Non-Zhejiang Ophiopogonis japonicus by Near-Infrared (NIR) Spectroscopy Combined with Chemometrics: The Influence of Different Stoichiometric and Spectrogram Pretreatment Methods. Molecules, 28, 2803. https://doi.org/10.3390/molecules28062803

Li, C., Wan, H., Wu, X., Yin, J., Zhu, L., Chen, H., Song, X., Han, L., Yang, W., Yu, H. & Li, Z. (2022). Discrimination and Characterization of the Volatile Organic Compounds in Schizonepetae Spica from Six Regions of China Using HS-GC-IMS and HS-SPME-GC-MS. Molecules, 27, 4393. https://doi.org/10.3390/molecules27144393

Nedeltcheva-Antonova, D., Gechovska, K., Bozhanov, S., & Antonov, L. (2022). Exploring the Chemical Composition of Bulgarian Lavender Absolute (Lavandula Angustifolia Mill.) by GC/MS and GC-FID. Plants (Basel), 11, 3150. https://doi.org/10.3390/plants11223150

Pendem, A., Pawar, V.M., & Jayaraman, S. (2010). Development of a gas chromatography method for the estimation of alkylamines in foods. Journal of Agricultural and Food Chemistry, 58, 8904–10. https://doi.org/110.1021/jf100507r

Sammarco, G., Bardin, D., Quaini, F., Dall'Asta, C., Christmann, J., Weller, P. & Suman, M. (2023). A geographical origin assessment of Italian hazelnuts: Gas chromatography-ion mobility spectrometry coupled with multivariate statistical analysis and data fusion approach. Food Research International, 171, 113085. https://doi.org/10.1016/j.foodres.2023.113085

Tian, X., Li, Z.J., Chao, Y.Z., Wu, Z.Q., Zhou, M.X., Xiao, S.T., Zeng, J. & Zhe, J. (2020). Evaluation by electronic tongue and headspace-GC-IMS analyses of the flavor compounds in dry-cured pork with different salt content. Food Research International. 137, 109456. https://doi.org/10.1016/j.foodres.2020.109456

Valli, E., Panni, F., Casadei, E., Barbieri, S., Cevoli, C., Bendini, A., Garcia-Gonzalez, D.L. & Gallina Toschi, T. (2020). An HS-GC-IMS method for the quality classification of virgin olive oils as screening support for the panel test. Foods, 9, 657. https://doi.org/10.3390/foods9050657

Wang, S., Chen, H., & Sun, B. (2019). Recent progress in food flavor analysis using gas chromatography-ion mobility spectrometry (GC-IMS). Food Chemistry, 315, 126158. https://doi.org/10.1016/j.foodchem.2019.126158

Wang, W., Jiang H., & Xuan, F. (2020). Analysis of volatile components in Nelumbinis stamen and Nelumbinis plumula by HS-SPME-GC-MS. Hubei Agricultural Sciences, 59, 137–140. https://doi.org/10.14088/j.cnki.issn0439-8114.2020.03.030

Welke, J. E., Nicolli, K. P., Hernandes, K. C., Biasoto, A. C. T., & Zini C. A. (2022). Adaptation of an olfactometric system in a GC-FID in combination with GCxGC/MS to evaluate odor-active compounds of wine. Food Chemistry, 370, 131004. https://doi.org/10.1016/j.foodchem.2021.131004

Wu, Y., Li, Z., Zou, S., Dong, L., Lin, X., Chen, Y., Zhang, S., Ji, C., & Liang, H. (2023). Chemical Composition and Flavor Characteristics of Cider Fermented with Saccharomyces cerevisiae and non-Saccharomyces cerevisiae. Foods, 12, 3565. https://doi.org/10.3390/foods12193565

Xu, J., Zhang, Y., Yan, F., Tang, Y., Yu, B., Chen, B., Lu, L., Yuan, L., Wu, Z., & Chen, H. (2023). Monitoring Changes in the Volatile Compounds of Tea Made from Summer Tea Leaves by GC-IMS and HS-SPME-GC-MS. Foods, 12, 146. https://doi.org/10.3390/foods12010146.

Yan, H., Huang, W., Zhao, X., Ge, C., Shen, Y., & Liu, L. (2024). Effect of fatty acid composition on the volatile compounds of pasteurized milk during low-temperature storage. Food Research International, 195, 114990. https://doi.org/10.1016/j.foodres.2024.114990

Yin, X., Xiao, W., Zhang, S., Yu, Z., Ai, W., Fu, S., Liu, J., & Huang, D. (2024). Impacts of Five Different Drying Methods on Volatile Organic Compounds in Mulberry Fruits. Foods, 13, 3514. https://doi.org/10.3390/foods13213514

Younes, K., Kharboutly, Y., Antar, M., Chaouk, H., Obeid, E., Mouhtady, O., Abu-Samha, M., Halwani, J., & Murshid, N. (2023). Application of Unsupervised Machine Learning for the Evaluation of Aerogels' Efficiency towards Ion Removal-A Principal Component Analysis (PCA) Approach. Gels, 9, 304. https://doi.org/10.3390/gels9040304

Yu, Y., Wei, X., Liu, Y., Dong, G., Hao, C., Zhang, J., Jiang, J., Cheng, J., Liu, A., & Chen S. (2022). Identification and quantification of oligomeric proanthocyanidins, alkaloids, and flavonoids in lotus seed: A potentially rich source of bioactive compounds. Food Chemistry, 379, 132124. https://doi.org/10.1016/j.foodchem.2022.132124

Zhang, Q., Ding, Y., Gu, S., Zhu, S., Zhou, X., & Ding, Y. (2020). Identification of changes in volatile compounds in dry-cured fish during storage using HS-GC-IMS. Food Research International, 137, 109339. https://doi.org/10.1016/j.foodres.2020.109339

Zhang, S., Fang, K., Ding, Z., Wu, J., Lin, J., Xu, D., Zhong, J., Xia, F., Feng, J., & Shen, G. (2023). Untargeted Metabolomics Analysis Revealed the Difference of Component and Geographical Indication Markers of Panax notoginseng in Different Production Areas. Foods, 12, 2377. https://doi.org/10.3390/foods12122377

Zhao, T., Benjakul, S., Sanmartin, C., Ying, X., Ma, L., Xiao, G., Yu, J., Liu, G., & Deng, S. (2021). Changes of Volatile Flavor Compounds in Large Yellow Croaker (Larimichthys crocea) during Storage, as Evaluated by Headspace Gas Chromatography-Ion Mobility Spectrometry and Principal Component Analysis. Foods, 10, 2917. https://doi.org/10.3390/foods10122917

Zheng, Z., Wang, M., Wang, D., Duan, W., Wang, X., & Zheng, C. (2010). Preparative separation of alkaloids from *Nelumbo nucifera* leaves by conventional and pH-zone-refining countercurrent chromatography. Journal of Chromatography B, 878, 1647-1651. https://doi.org/10.1016/j.jchromb.2010.04.020

Zhu, H., Zhu, D., & Sun, J. (2023). Application of GC-IMS coupled with chemometric analysis for the classification and authentication of geographical indication agricultural products and food. Frontiers in Nutrition, 10, 1247695. https://doi.org/10.3389/fnut.2023.1247695