

Article

Benchtop Volatilomics and Machine Learning for the Discrimination of Coffee Species

Catherine Kiefer ^{1,2}, Steffen Schwarz ³ , Nima Naderi ⁴, Hadi Parastar ^{1,4}, Sascha Rohn ²  and Philipp Weller ^{1,*} 

¹ Institute for Instrumental Analytics and Bioanalytics, Technical University of Applied Sciences Mannheim, Paul-Wittsack-Str. 10, 68163 Mannheim, Germany; catherine@kiefersonline.de (C.K.); h.parastar@sharif.edu (H.P.)

² Department of Food Chemistry and Analysis, Institute of Food Technology and Food Chemistry, Technische Universität Berlin, Kaiserin-Augusta-Allee 14, 10553 Berlin, Germany; rohn@tu-berlin.de

³ Coffee Consulate, Hans-Thoma-Strasse 20, 68163 Mannheim, Germany; schwarz@coffee-consulate.com

⁴ Department of Chemistry, Sharif University of Technology, Tehran P.O. Box 11155-9516, Iran; naderinima73@gmail.com

* Correspondence: p.weller@hs-mannheim.de; Tel.: +49-(0)621-292-6484

Abstract

The main characteristics of the large number of coffee species are differences in aroma and caffeine content. Labeled blends of *Coffea arabica* (*C. arabica*) and *Coffea canephora* (*C. canephora*) are common to broaden the flavor profile or enhance the stimulating effect of the beverage. New emerging species such as *Coffea liberica* (*C. liberica*) further increase the variability in blends. However, significant price differences between coffee species increase the risk of unlabeled blends and thus influence food quality and safety for consumers. In this study, a prototypic hyphenation of trapped headspace-gas chromatography-ion mobility spectrometry-quadrupole mass spectrometry (THS-GC-IMS-QMS) was used for the detection of characteristic compounds of *C. arabica*, *C. canephora*, and *C. liberica* in green and roasted coffee samples. For the discrimination of coffee species with IMS data, multivariate resolution with multivariate curve resolution–alternating least squares (MCR-ALS) prior to partial least squares–discriminant analysis (PLS-DA) was evaluated. With this approach, the classification accuracy, as well as sensitivity and specificity, of the PLS-DA model was significantly improved from an overall accuracy of 87% without prior feature selection to 92%. As MCR-ALS preserves the physical and chemical properties of the original data, characteristic features were determined for subsequent substance identification. The simultaneously generated QMS data allowed for partial annotation of the characteristic volatile organic compounds (VOC) of roasted coffee.

Keywords: THS-GC-IMS-QMS; ion mobility spectrometry; volatile organic compounds; authentication; coffee species; MCR-ALS; multivariate resolution; chemometrics



Received: 25 November 2025

Revised: 30 December 2025

Accepted: 22 January 2026

Published: 2 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

With substantially increasing prices of green coffee on the world market, simple and practicable strategies for product authentication are crucial aspects in assuring food quality as well as food safety. This is particularly important for “single-variety” products from *Coffea arabica* (*C. arabica*), as this still is the most popular coffee species. However, it is often blended with *Coffea canephora* (*C. canephora*) or other coffee species with lower prices. However, it is important to underline that it is not only the species that determines quality but also the origin and processing. Consequently, *Canephora*-based coffees are not per se of

lower value. A relatively new phenomenon is the growing interest in a third coffee species in the specialty coffee market: *Coffea liberica* (*C. liberica*) [1]. While, for decades, *Liberica* coffees were not in demand and the trees were historically primarily used as a trap crop for pest management, climate change has changed this view: due to its good resistance towards emerging climatic conditions, paired with its unique taste and limited availability, market prices for *C. liberica* have overtaken the other two coffee species. Now, in turn, *C. liberica* is a novel target for adulteration and mislabeling through blending with less expensive coffee species such as *C. arabica* and *C. canephora*. Overall, this underlines the clear need for routine-suitable methodologies for product authentication and quality analysis.

Differentiation of whole beans is typically described as straightforward, as the coffee beans of the different species have clearly defined morphological attributes. However, low-abundant adulterations are challenging to detect. An additional twist complicating authentication is the fact that morphologies differ within coffee species, such as the *C. arabica* varieties Catuai and Maragogype differing vastly in size. This underlines the need for authentication strategies on a deeper layer.

To date, different methods are described to detect adulteration of coffee. These include nuclear magnetic resonance (NMR) [2,3], high-performance liquid chromatography (HPLC) [4,5], near infrared (NIR) [5–7], and Raman spectroscopic methods [8,9], as well as solid phase micro-extraction-gas chromatography-mass spectrometry (SPME-GC-MS) [10–12]. Although NMR and SPME-GC-MS generate in-depth data and have the advantage of compound identification, these suffer from several limitations. NMR techniques require advanced sample preparations with a high rate of solvent consumption, while SPME-GC-MS methods are often time-consuming due to long extraction and desorption times. Similar to NMR, sample preparation for HPLC methods requires complex extraction methods with high solvent consumption before and during analysis while lacking the power of substance identification. Furthermore, the operation of these three analytical methods is confined to a highly specified laboratory environment and the availability of expensive gases with an increasing limitation of supplies such as helium. Fast spectroscopic methods like NIR and Raman are good screening tools with low infrastructural need, especially with the emerging trend of specified hand-held devices [13]. However, the chemical information generated is limited and the identification of marker substances for validation is not possible. This leads to a rising need for a fast analytical method, operated at low infrastructure, that generates a sample fingerprint with identified marker compounds for validation.

The combination of the separation of gas chromatography paired with the fast and highly sensitive ion mobility spectrometry generates comprehensive volatile organic compound (VOC) 2D-fingerprints based on retention time and drift time of complex samples. GC-IMS is an emerging analytical method for the authentication of foods [14,15], beverages [16,17], or essential oils [18,19]. One major advantage of GC-IMS is the simplicity and robustness of the platform, which allows for use at the point-of-need (PoN)—a crucial aspect for fast authentication of suspect samples. These systems are operated at ambient pressure based on easily available and cheap nitrogen as carrier gas and feature a low power consumption pose, which drastically reduces the demand on laboratory infrastructure. As GC-IMS is typically based on soft ionization sources (e.g., ^3H , UV, or corona discharge (CD)), sensitivity for polar to medium polar species is excellent, which typically reduces the need for sample preparation, which again is an important factor for fast sample analysis at the PoN. This is also reflected by the substantially better scores in the context of green analytical chemistry (GAC) in comparison with, e.g., GC-MS [20], as neither helium nor enrichment steps or solvents are required. The systems are typically benchtop-based. However, the latest developments in this field have continuously miniaturized IMS systems towards hand-held devices (e.g., [21]).

While there are a number of GC-IMS-related publications in the field of coffee aroma [22], geographic origin [23,24], and processing [25,26], the literature on species differentiation is surprisingly scarce. Konieczka et al. reported the application of GC-IMS for coffee species authentication [27]. However, this approach was based on the sum spectra of the IMS data, disregarding the separation power of the GC. This second-dimension separation is the “power-up” for IMS systems and increases both selectivity and (useable) sensitivity substantially, opening up the path for omics-based approaches with chemometric data analysis.

Chemometric methods enhance the extraction of valuable information from the generated volatilomic fingerprints. Diverse fields of application in combination with different data analysis strategies have already been reported. Among the chemometric methods described are principal component analysis (PCA) [28,29], partial least squares-discriminant analysis (PLS-DA) [30,31], and different options of artificial neural networks [32,33]. These data analysis strategies are mainly maintained for the detection of geographical and botanical origin, as well as the detection of adulterants in the agricultural and food sector. Recently, the use of non-targeted volatilomics in combination with multivariate curve resolution-alternating least squares (MCR-ALS) was described for the authentication of saffron [34]. This advanced chemometric tool enables the decomposition of highly complex spectrometric and chromatographic data matrices. The generated pure component and corresponding concentration profiles lead to the resolution of overlapping peaks and background contributions [35]. In this context, MCR is used to generate extracted features that are utilized for sample discrimination with PCA and PLS-DA for authentication. Another advantage of MCR is the power to resolve the pure component spectra, which facilitates the potential to identify the underlying metabolites.

However, the identification of substances with GC-IMS data is challenging as commercial databases are not yet available and identification is commonly carried out via the analysis of reference substances. This method is costly and time-consuming, and prototype installations of IMS to GC-QMS systems for simultaneous detection of mass spectra as well as IMS spectra have already been reported for different applications [14,36,37]. The profiling of the original dataspace coupled with the simultaneous mass spectrometry leads to a more accurate detection of influencing features and thus to a more accurate identification with the corresponding QMS data. Consequently, the application of GC-IMS in combination with advanced chemometric tools could offer a powerful, point-of-need-suitable approach for the differentiation of *Coffea* species.

The aim of this study was the development of fast, cost-efficient, and potentially point-of-need screening of coffee species with a minimal need for sample preparation. The focus was set on the identification of relevant signals in the IMS data by deconvolution with MCR-ALS. Parallel simultaneously generated MS data were used for a tentative identification of the assigned signals with commercially available databases. To the best of our knowledge, similar results for the differentiation of coffee species have not been published yet.

2. Materials and Methods

2.1. Reagents and Samples

Green and roasted coffee samples were provided by Coffee Consulate (Mannheim, Germany). The samples consisted of 30 green coffee samples and 30 roasted coffee samples including 8 *C. arabica* samples, 16 *C. canephora* samples, and 6 *C. liberica* samples, respectively. Apart from four *C. canephora* samples, namely samples no. 21–24, all samples were produced as “specialty coffees”, i.e., no defective beans were used. The samples were stored in sealed

bags protected from light at room temperature until analysis. The used sample sets with species, variety, geographical origins, and post-harvest processing is described in Table 1.

Table 1. Sample list of green and roasted coffees, n. d.: not determined, x: present in sample set.

Sample No.	Species	Variety	Geographical Origin	Post-Harvest Processing	Green Sample	Roasted Sample
1	<i>C. arabica</i>	Bourbon amarello	Brazil	Pulped natural	x	x
2	<i>C. arabica</i>	Catucai 785	Brazil	Pulped natural	x	x
3	<i>C. arabica</i>	S795	India	Fully washed	x	x
4	<i>C. arabica</i>	S795	India	Fully washed	x	x
5	<i>C. arabica</i>	Catuai	Mexico	Natural	x	x
6	<i>C. arabica</i>	Obata	Mexico	Fully washed	x	x
7	<i>C. arabica</i>	Marsellesa	El Salvador	Natural	x	x
8	<i>C. arabica</i>	Bourbon tekesic	El Salvador	Fully washed	x	x
9	<i>C. canephora</i>	SLN274/Old paradenia	India	Natural	x	x
10	<i>C. canephora</i>	SLN274/Old paradenia	India	Natural	x	x
11	<i>C. canephora</i>	SLN274/Old paradenia	India	Pulped natural	x	x
12	<i>C. canephora</i>	SLN274/Old paradenia	India	Pulped natural	x	x
13	<i>C. canephora</i>	SLN274/Old paradenia	India	Pulped natural	x	x
14	<i>C. canephora</i>	SLN274/Old paradenia	India	Fully washed	x	x
15	<i>C. canephora</i>	SLN274/Old paradenia	India	Fully washed	x	x
16	<i>C. canephora</i>	SLN274/Old paradenia	India	Fully washed	x	x
17	<i>C. canephora</i>	SLN274/Old paradenia	India	Fully washed	x	x
18	<i>C. canephora</i>	SLN274/Old paradenia	India	Fully washed	x	x
19	<i>C. canephora</i>	C × R	India	Fully washed	x	x
20	<i>C. canephora</i>	Conillon vermelho	Brazil	Natural	x	x
21	<i>C. canephora</i>	n. d.	Uganda	n. d.	x	x
22	<i>C. canephora</i>	n. d.	Uganda	n. d.	x	x
23	<i>C. canephora</i>	n. d.	Vietnam	n. d.	x	x
24	<i>C. canephora</i>	n. d.	Bali	n. d.	x	x
25	<i>C. liberica</i>	Liberica	India	Natural	x	x
26	<i>C. liberica</i>	Liberica	India	Natural	x	x
27	<i>C. liberica</i>	Liberica	India	Natural	x	x
28	<i>C. liberica</i>	Liberica	India	Natural	x	x
29	<i>C. liberica</i>	Liberica	India	Natural		x
30	<i>C. liberica</i>	Liberica	India	Natural		x
31	<i>C. liberica</i>	Liberica	Malaysia	Pulped natural	x	
32	<i>C. liberica</i>	Liberica	Malaysia	Pulped natural	x	

For the analysis, 5 g of coffee beans per sample were shock-frozen with liquid nitrogen and ground for 45 s at level 8.5 using a kitchen grade knife mill (Thermomix TM6, Vorwerk Deutschland Stiftung & Co., KG, Wuppertal, Germany). Subsequently, 1.4 g per sample of the green coffee grounds and 0.4 g per sample of the roasted coffee grounds were transferred into 20 mL headspace vials and closed tightly with a screwcap with butyl/PTFE septa.

2.2. Instrumentation

All measurements were performed on a prototypic THS-GC-IMS-QMS dual detection system, consisting of a Shimadzu HS 20 headspace sampler (Shimadzu Corporation, Kyoto, Japan), a Shimadzu Nexus GC-2030 (Shimadzu Deutschland GmbH, Duisburg, Germany) coupled to a Shimadzu QP-2020 NX MSD (electron impact (EI) mode) (Shimadzu Deutschland GmbH, Duisburg, Germany) and a FOCUS-ion mobility spectrometer module (Gesellschaft für Analytische Sensorsysteme mbH, Dortmund, Germany). The optimal instrument parameters are summarized in Table 2. Details on the hardware setup can be found here [14].

Table 2. Trapped headspace incubation settings and GC parameter.

Trapped Headspace Conditions and GC Settings	
Incubation time	15 min
Incubation temperature	80 °C
Shaker level	2
Sample loop	1 mL
Trap temperature	−10 °C
Trap cycles	2
Trap equilibration temperature	25 °C
Trap desorption temperature	280 °C
Split ratio	1:20
Inlet pressure	180 kPa
GC column	VF-23 ms (30 m × 0.25 mm × 0.25 µm)
Oven program	40 °C → 200 °C with 10 °C/min

Trapped headspace measurements were carried out with an incubation at 80 °C for 15 min and shaking level 1. A headspace volume of twice 1 mL was transferred onto a Tenax TA (Shimadzu Corporation, Kyoto, Japan) tube and trapped at −10 °C. The Tenax TA tube was equilibrated at 25 °C and desorbed onto the GC column at 280 °C with a split ratio of 1:20. Chromatographic separation was performed on a VF-23 ms capillary column (operating temperatures: 40–260 °C/260 °C; SN: NL10772427) with a 30 m × 0.25 mm × 0.25 µm film thickness (Agilent Technologies Inc., Santa Clara, CA, USA). The carrier gas was helium with a constant pressure of 180 kPa and a splitter advanced pressure controller (APC) pressure of 38 kPa. The GC oven program was as follows: 40 °C, held for one minute, 40 °C to 200 °C at 10 °C/min, held for 3 min, resulting in a run time of 20 min. At the end of the analytical column, the column gas flow was split by a SilFlow GC 4-port splitter plate (Trajan Scientific and Medical, Ringwood, Australia) into two retention gaps of 0.7 m length to the IMS and 1.6 m length to the QMS, with 0.15 mm inner diameter, respectively. Transfer lines were operated at 220 °C to both QMS (Shimadzu Corporation, Kyoto, Japan) and IMS (Hillesheim GmbH, Waghäusel, Germany). The ion source temperature of the QMS was set to 220 °C, the electron ionization energy was 70 eV, and the scan range was m/z 35 to m/z 500 with a duty cycle of 300 ms.

The OEM-Focus-IMS[®] cell consisted of a ³H-radioactive ionization source (100 MBq β -emission). It was operated at 100 °C in positive-ion mode at a constant voltage of 2.5 kV. The drift tube had a diameter of 15.2 mm and a length of 98 mm. The injection voltage was set to 2500 V, and the blocking voltage was set to 70 V. The drift gas was nitrogen of 99.9999% purity, controlled with a mass flow controller (Vögtlin Instruments GmbH, Aesch, Switzerland) to 150 mL/min. The injection pulse width was set to 100 µs and the sampling frequency was 228 kHz. To reduce data size, each spectrum was averaged over six scans with a repetition rate of 21 ms.

2.3. Data Processing and Evaluation

For data preprocessing, exploratory analysis and visualization of the IMS spectra, gc-ims-tools version 0.1.7 and Python version 3.11.4 were used [38]. Preprocessing is an important step in multivariate data analysis to remove analysis-related effects, such as signal shifts. To reduce the size of the data for more efficient processing, the first step in the preprocessing was a level 3 wavelet compression to lower the number of variables without losing important information. The data were aligned alongside the drift time to correct for pressure-dependent shifts, and the drift time was normalized to the reactant ion peak (RIP). Further, the drift times were aligned by dynamic time warping (DTW). The spectra were cropped to the relevant areas corresponding to 1.05–2.1 on the RIP-relative drift time axis (7 ms to 15 ms drift time) and 50–900 s on the retention time axis. Afterwards, the data were baseline corrected by asymmetric least squares (AsLS) (weighting of 0.001, smoothing at 10^7). Finally, data were Pareto-scaled and mean-centered. The preprocessed data were analyzed in an exploratory data analysis by PCA and subsequently checked for outliers via Hotelling T^2 -Q residuals plot. In a second step, supervised analysis by PLS-DA was performed.

For MCR-ALS, MATLAB version R2019b (MathWorks, Natick, MA, USA) in combination with the MCR-ALS 2.0 toolbox [39] was used. The parameters used are described here [34]. The GC-IMS data were RIP corrected and finally augmented column-wise with the retention times as rows and the drift times as columns. The augmented data were subjected to multivariate resolution by MCR-ALS. The number of components was determined by singular value decomposition (SVD) and the resulting scree plot as 50 for roast coffee and 55 for green coffee samples, respectively. Additionally, simple-to-use interactive self-modeling mixture analysis (SIMPLISMA) was used to calculate the initial estimate of IMS profiles as a starting point for ALS optimization. The constraints were spectral normalization, non-negativity in both RT and DT mode, and unimodality in RT mode. The PCA-based lack of fit (LOF) metric was used for MCR-ALS model evaluation. The threshold value as stop value was set to 0.1 and the iteration number was set to 1000. Convergence was achieved after 36 iterations with 2.40% LOF. The resolved elution profiles were subsequently used for discriminant analysis by PCA and PLS-DA, as well as compound identification in combination with the MS data.

The GC-MS data were analyzed with mzmne (mzio GmbH, Bremen, Germany). The data were analyzed as a batch. The steps of the batch queue were as follows: mass detection with a noise level set to 5.0×10^1 ; chromatogram builder; smoothing with the Loess smoothing algorithm; local minimum feature resolver; GC-EI spectral deconvolution with rt grouping and shape correlation algorithm; join aligner; spectral/molecular networking.

The assigned compounds were preliminarily annotated with NIST/EPA/NIH Mass Spectral Library 23 from the NIST (Gaithersburg, MD, USA) of the U.S. Department of Commerce. The TIC chromatograms were integrated and the retention time at peak maxima used for retention time comparison. For exploratory data analysis, the aligned feature list was exported and assessed with PCA and fold-change analysis in volcano plots with MetaboAnalyst 6.0 (NSERC, Ottawa, ON, Canada). To assess the need for data normalization prior to exploratory data analysis, the feature density and normalized intensity were evaluated. Based on the results, green coffee data were normalized by sum, $\log_{10}$ -transformed, and auto-scaled, while roast coffee data were only auto-scaled prior to PCA analysis and fold-change analysis. For the generation of volcano plots, data were used unpaired and p was set to 0.05.

3. Results

3.1. Exploratory Data Analysis

The GC-IMS data were preprocessed with the gc-ims-tools python package for drift time and retention time alignment, normalization and ROI definition. The preprocessed data matrix was subjected to PCA for exploratory analysis. Figure 1A shows the PCA scores plot of the first two principal components of green coffee.

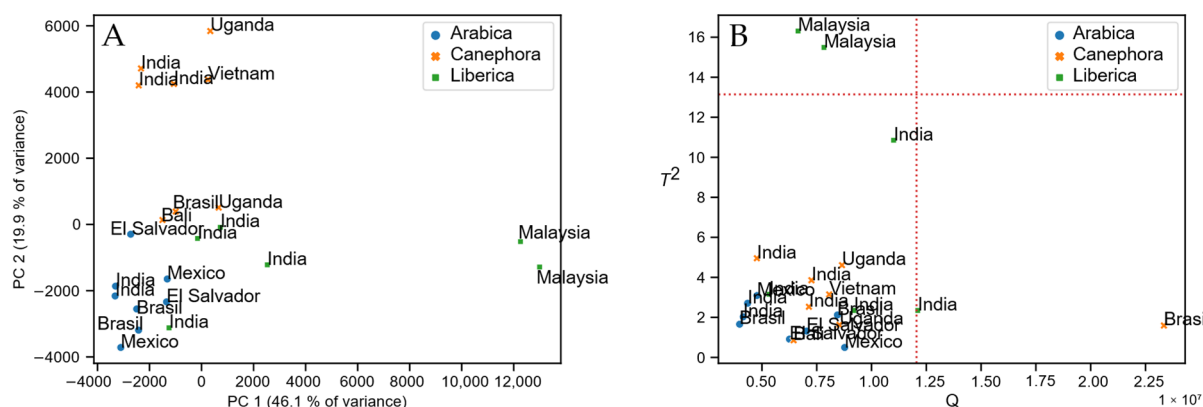


Figure 1. PCA scatter plot of PC1 and PC2 of green coffee samples generated with gc-ims-tools (preprocessed data after Pareto-scaling and mean-centering) (A). Corresponding Hotelling T^2 -Q residuals-plot with T^2 and Q thresholds (red dotted line) for outlier evaluation (B).

It can be seen that PC2 separates *C. canephora* from *C. arabica* and *C. liberica*. The *C. liberica* samples show a slight shift into the positive direction on the PC1; however, most of the *liberica* samples cluster together with *C. arabica*. Two of the *C. liberica* samples were located far in the positive range on PC1, which indicates a potential residual outlier. The Hotelling T^2 -Q residuals-plot in Figure 1B showed that the two *C. liberica* samples featured high T^2 values, i.e., deviations within the model, and were therefore not considered as residual outliers. The origin of these samples is Malaysia, and the samples were processed semi-dry with intact parchment. As the parchment was not removed from the coffee beans prior to shipping, the moisture content could not be measured before and after shipment. The parchment was removed shortly before the sample preparation and the olfactory properties of the dried coffee cherries were perceptibly more intense; furthermore, the coffee beans showed a different, more brown color compared to the other coffee samples. MS data analysis revealed a higher acetic acid content, but the overall compound profile was similar to the other samples.

Figure 2 shows the results of the PCA as well as the outlier test for the roasted coffee samples. PC2 and PC3 tentatively differentiate between the coffee species. The scores plot (Figure 2A) visualizes the tentative separation of *C. canephora* from the rest on PC2, while PC3 separates *C. arabica* from *C. liberica*.

The Hotelling T^2 -Q residuals-plot indicated one sample from Bali as a potential residual outlier. To determine whether the sample should be removed from the dataset, the volatile composition of the sample was analyzed using the corresponding QMS data. The mass spectrometric analysis showed a substantially higher amount of acetic acid in comparison to the other samples. This is most likely caused by a high ratio of defective coffee beans in the raw coffee, which are known to form an increased level of acetic acid during the roasting process [40]. However, as with the previous samples from Malaysia, after removal of this specific feature in the data analysis, the VOC profile still featured the typical properties of the other *canephora* samples and as such, the sample was kept in the dataset. Also, as a T^2 -Q plot from a PCA after MCR-ALS determined no residual outliers (Figure S1B), the

sample remained in the dataset. Similar to the results described by Konieczka et al. [27], a tentative separation between *C. arabica* and *C. canephora* was obtained (Figure 2A).

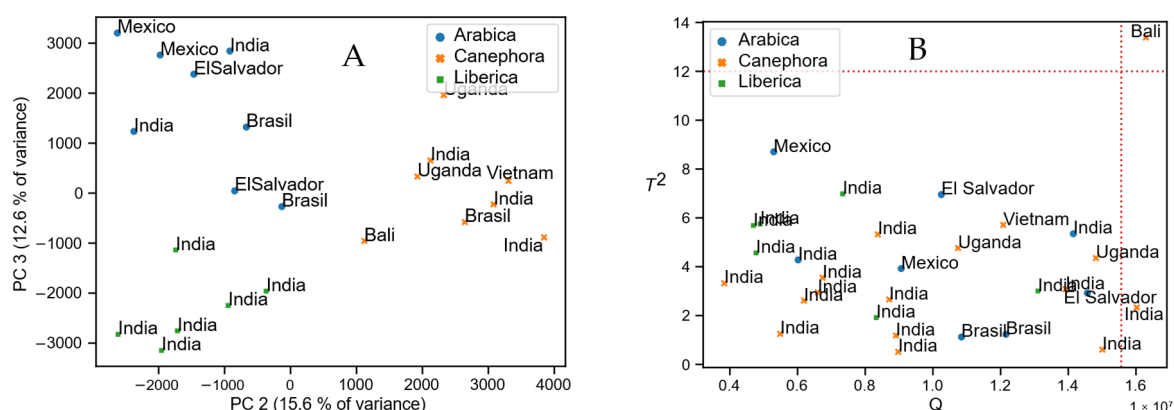


Figure 2. PCA scatter plot of PC2 and PC3 of GC-IMS data (preprocessed, Pareto-scaled and mean-centered) of roast coffee samples generated with gc-ims-tools (A). Corresponding T^2 -Q-plot with T^2 and Q thresholds (red dotted line) for residual outlier evaluation with one sample from Bali as potential residual outlier (B).

3.2. Classification of Coffee Species

PLS-DA was used for the classification of the coffee species. Based on the mean squared error, the optimal number of latent variables was seven. Model validation was carried out with 5-fold cross validation (CV) with a classification accuracy of 87%. However, sensitivity and specificity of the model indicate that the model stability was not optimal for the classification of the green coffee samples (Table 3).

Table 3. Figures of merit of PLS-DA classification of green coffee without prior feature extraction.

	Green Coffee		
	<i>Coffea arabica</i>	<i>Coffea canephora</i>	<i>Coffea liberica</i>
Accuracy	1.0 ± 0.00	0.80 ± 0.20	0.80 ± 0.40
Sensitivity	1.0 ± 0.00	0.67 ± 0.27	0.70 ± 0.40
Specificity	1.0 ± 0.00	0.95 ± 0.10	0.92 ± 0.16

One of the main reasons for inferior PLS model stability is a high number of variables used, which are often without true meaning [41]. As GC-IMS detects an enormous richness of signals due to the soft ionization, this could also be a potential aspect in this model. Therefore, a feature extraction strategy based on MCR-ALS was evaluated. Besides the feature selection itself, MCR-ALS also elevates the stability of subsequent classification models by correcting for baseline drifts, retention time shifts, and coelution [35]. Higher signal-to-noise ratios (SNR), as well as the deconvolution of overlapping or coeluting peaks, generate more chemically meaningful features [42]. MCR-ALS-based dimensionality reduction preserves both the concentration and spectral profiles. Therefore, this approach is an optimal pairing with following classification models, by which the relevant metabolite signals for classification can easily be determined from generated peak lists.

MCR-ALS resolved 55 components for the 30 coffee samples (Figure S1). A matrix of the corresponding peak areas was generated, which was then used for discriminant analysis of coffee species with PLS-DA. The data were plotted in a Hotelling T^2 -Q residuals-plot for the determination of residual outliers (Figure S2). The results showed no residual outliers in the green and roast dataset. Consequently, the whole datasets of green and roast coffee were used for PLS-DA. The optimal number of latent variables was determined as five

and validation was performed by a 5-fold CV. Table 4 shows the results of the PLS-DA after MCR-ALS-based deconvolution and underlines the effectiveness of intelligent feature extraction and selection.

Table 4. Figures of merit of PLS-DA classification with feature extraction by MCR-ALS of green and roasted coffee.

	Green Coffee			Roast Coffee		
	<i>Coffea arabica</i>	<i>Coffea canephora</i>	<i>Coffea liberica</i>	<i>Coffea arabica</i>	<i>Coffea canephora</i>	<i>Coffea liberica</i>
Accuracy	1.00 ± 0.00	1.00 ± 0.00	0.75 ± 0.12	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
Sensitivity	1.00 ± 0.00	0.94 ± 0.16	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
Specificity	1.00 ± 0.00	1.00 ± 0.00	0.96 ± 0.08	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00

The accuracy for the classification of *C. canephora* improved significantly, while the accuracy for *C. liberica* was slightly lower compared to results without prior deconvolution. This is most likely due to the imbalanced sample set with lower sample numbers for *C. liberica*. However, the prior inferior model performance for *C. canephora* and *C. liberica* improved substantially in terms of sensitivity and specificity and indicated a higher prediction power as well as a more accurate classification of *C. canephora* and *C. liberica* samples. The influence of the extracted features on the species classification was obtained from the PLS-DA coefficients for each species. (Figure 3).

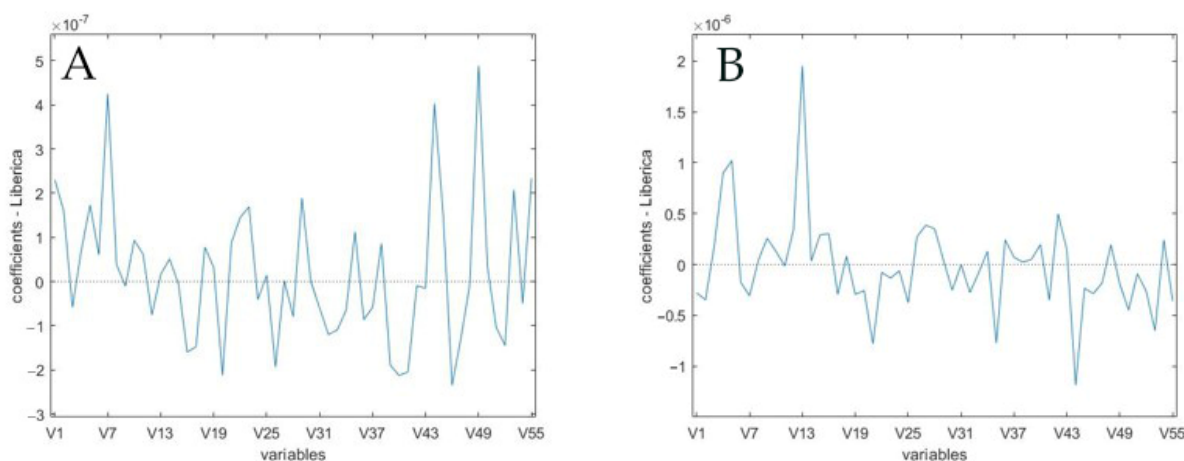


Figure 3. PLS-DA coefficients of green (A) and roasted (B) *C. liberica* GC-IMS data after preprocessing, and MCR-ALS for feature extraction.

As MCR-ALS resolution preserves the chemical and physical information of the original data, it was possible to extract the retention time for subsequent substance annotation by MS data. The features considered relevant for classification were selected to include only the features with the highest contributions to the model with thresholds set to 5×10^{-7} for green and 0.5×10^{-6} for roast coffee.

These results highlight the high robustness of the PLS-DA model when combined with feature extraction by MCR-ALS. Although models like Random Forest (RF) or artificial neural networks are highly powerful for classification problems, they often lack the power to determine relevant features for biomarker research, which was crucial for this study [43,44]. For this data evaluation, the coefficients calculated in the PLS-DA model were used to further analyze features relevant in the classification. The use of PLS-DA also enhances the explainability of the results compared to RF, ANNs, or other non-linear algorithms such as support vector machines (SVM), which in the context of the rising trend of explainable AI is gaining more importance [45]. Although PLS-DA models struggle with

very large datasets, the low number of samples per class for coffee species classification further promoted the use of less complex models such as PLS-DA.

3.3. Substance Annotation by QMS Data

For substance annotation, the extracted retention times of the compounds with a high influence on the classification were correlated with the simultaneously measured QMS data. Data analysis was carried out with mzmine 4.1.6 and the assigned features and annotated compounds were compared to the feature lists generated by MCR-ALS. The QMS data were deconvoluted, aligned, and subsequently introduced to the spectral library search with public databases such as MoNA and the GC-MS public KovatsRI database, as well as the commercial NIST23 GC-MS database. The resulting aligned feature list consisted of all the features detected in the data, as well as the spectral library matches including the corresponding match quality. Spectral library matches with a match quality of less than 70% were not considered for the feature lists. The compounds characteristic for the coffee species were evaluated by exploratory analysis in PCA in combination with the PCA loadings, as well as a two-tailed *t*-test in a volcano plot of the negative $\log_{10}$ *p*-value and $\log_2$ fold-change. Signals were considered significantly changed with *p*-values smaller or equal to 0.05 and a fold change of higher or equal to two. As the annotation was solely carried out by reference library matches, all annotations were considered as putative.

For green coffee samples, a total of 108 features were detected of which 39 were annotated by the spectral databases. Among the annotated features the compound classes ranged from aldehydes, ketones, esters, and alcohols to pyrazines, and acids (Table 5).

Table 5. Annotated compounds of green coffee QMS data.

No.	Retention Time [min]	Annotation
1	1.80	2-Methylheptane
2	1.92	Octane
3	2.54	Ethyl acetate
4	2.91	3-Methylbutanal
5	3.18	2-Butanol
6	3.41	Pentanal
7	3.60	2-Methylpropan-1-ol
8	4.09	3-Hydroxy-2-butanone (Acetoin)
9	4.11	Ethyl isovalerate
10	4.19	Ethyl 2-methylbutyrate
11	4.56	Hexanal
12	5.43	2-Pentylfuran
13	5.45	1-Pentanol
14	5.85	Heptanal
15	6.02	2-Heptanone
16	6.09	Hexanoic acid ethyl ester
17	6.34	Acetic acid
18	6.73	4-Penten-2-one
19	7.81	2,3-Butanediol
20	7.93	Trimethylpyrazine
21	8.44	Ethyl-3-hydroxy-methylbutanoate
22	8.45	Octanal
23	8.45	Nonanal
24	8.61	Tetramethylpyrazine
25	8.69	Furfural
26	8.95	3-Octen-2-one

Table 5. Cont.

No.	Retention Time [min]	Annotation
27	9.00	Propionic acid
28	9.14	Isovaleric acid
29	9.69	Benzaldehyde
30	10.53	5-Methylfurancarboxaldehyde
31	10.54	1,3,5-Trimethyl-1H-pyrazole
32	11.02	1-Methyl-1H-pyrrole-2-carboxaldehyde
33	11.73	2-Butyl-2,7-octadien-2-ol
34	12.09	Methyl salicylate
35	12.14	1,2-Dimethoxybenzene
36	12.62	Ethyl salicylate
37	12.90	Phenylethylacetate
38	12.94	Benzyl alcohol
39	12.94	3-Methyl-phenol

In the exploratory data analysis, the explained variance in the green coffee MS data was heavily influenced by the extreme samples of the Malaysian *C. liberica*. Due to a significantly higher content of acetic acid in these two samples, PC1 separated the two Malaysian coffee samples from the rest. Although there is a trend for the location of *C. canephora* samples in the positive region of the PC2 and *C. liberica* samples in the negative region of PC1, a clear separation between the three species was not achieved, as visualized in Figure 4A.

Excluding the Malaysian samples from the dataset did not influence the separation of the coffee species in the PCA. Therefore, it was not possible to define characteristic compounds from the biplot (Figure S3). However, the volcano plot (Figures 4B, S4 and S5) showed significant changes in the composition for each coffee species, where annotation was partially feasible. The results are visualized in Table 6.

Table 6. Up-regulated compounds for green coffee.

Species	Retention Time [min]	Annotation
<i>C. arabica</i>	1.80	2-Methylheptane
	2.54	Ethyl acetate
	5.43	2-Pentylfuran
	6.73	4-Penten-2-one
<i>C. canephora</i>	1.80	2-Methylheptane
	3.41	Pentanal
	4.56	Hexanal
	5.43	2-Pentylfuran
	5.45	1-Pentanol
	5.85	Heptanal
	6.02	2-Heptanone
	6.73	4-Penten-2-one
	9.69	Benzaldehyde
	11.73	2-Butyl-2,7-octadien-2-ol
<i>C. liberica</i>	2.54	Ethyl acetate
	6.34	Acetic acid
	7.81	2,3-Butandiol
	9.14	Isovaleric acid
	12.09	Methyl salicylate
	12.62	Ethyl salicylate

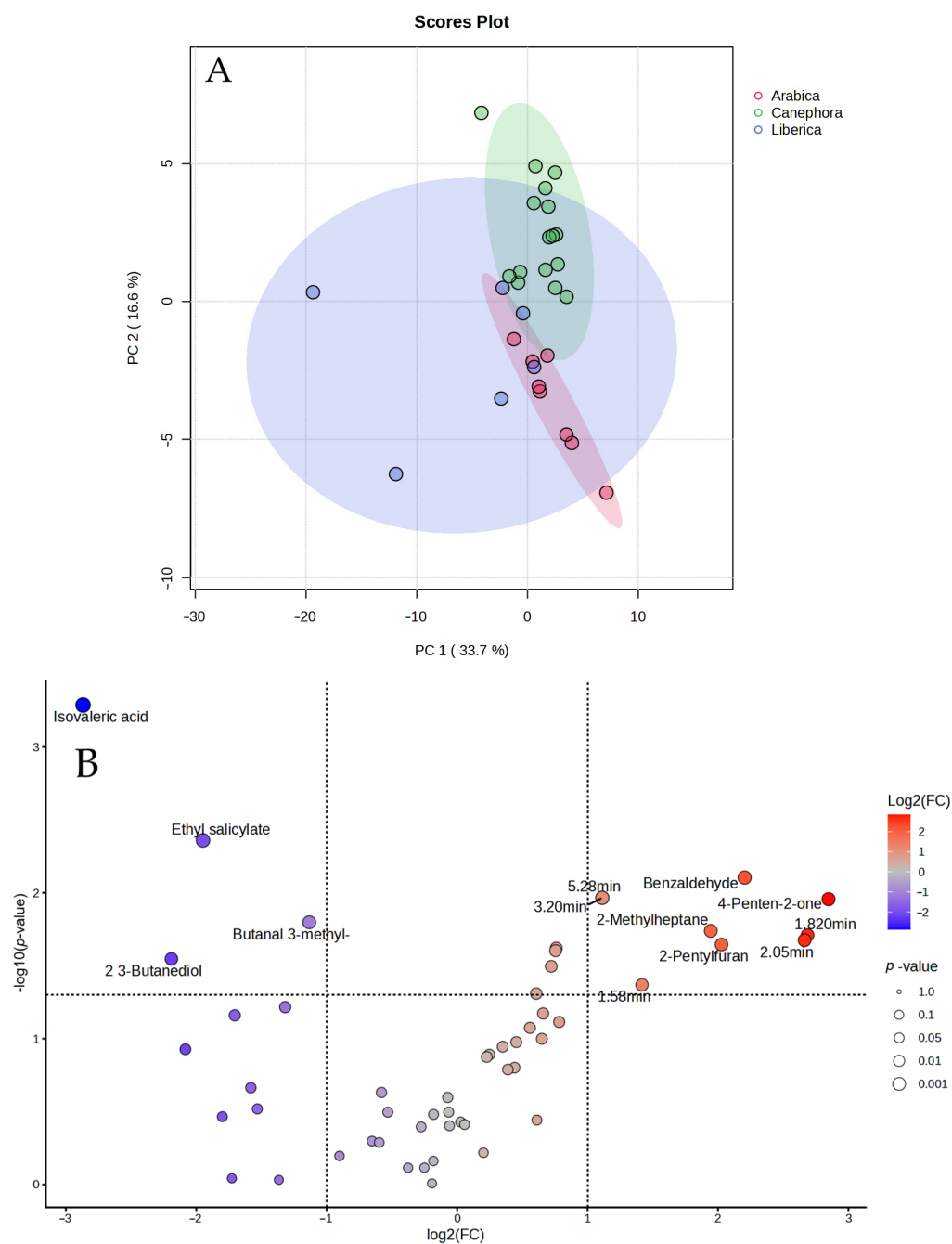


Figure 4. PCA of QMS data (log₁₀-transformed, auto-scaled, peak intensities) from green coffee samples generated in MetaboAnalyst (A). Volcano plot of *C. liberica* and *C. arabica* green coffee samples with thresholds marked by the dotted lines generated in MetaboAnalyst (log₁₀-transformed, auto-scaled, peak intensities, *p*-value: 0.05) (B).

For roasted coffee, 175 features were detected of which 41 were preliminarily annotated. Compounds characteristic of roasted coffee, such as furans, phenols, and sulfide-containing VOCs were found among the annotated features (Table 7).

Table 7. Annotated compounds of roast coffee QMS data.

No.	Retention Time [min]	Annotation
1	2.09	3-Methylfuran
2	3.51	2-Propionylthiazole
3	3.86	Dimethyl disulfide
4	5.24	Pyridine
5	6.13	Methylpyrazine
6	6.22	Acetic acid
7	6.53	1-Hydroxy-2-propanone *
8	6.77	1-Hydro-2-methyl-3(2H)-furanone
9	7.30	2,3-Dimethylpyrazine
10	7.70	2-Ethyl-6-methylpyrazine
11	7.86	1-Hydroxy-2-propanone *
12	7.93	Trimethylpyrazine
13	8.23	2-Cyclopenten-1-one
14	8.40	1,3-Di-tertbutylbenzene
15	8.69	Furfural
16	9.15	Isovaleric acid
17	9.21	Furfuryl acetate
18	9.24	β -Ocimene
19	9.45	1-(Acetyloxy)-2-propanone
20	9.70	Benzaldehyde
21	10.14	Furfuryl alcohol
22	10.36	2,3-Butanedione
23	10.44	1-Acetyloxy-2-butanone
24	10.54	5-Methylfurfural
25	10.78	2,5-Hexadione
26	11.27	1-Methyl-1H-pyrrole-2-carboxaldehyde
27	11.29	1-(1-Methyl-1H-pyrrol-2-yl)-ethanone
28	11.43	2-Acetyl-3-Methylpyrazine
29	11.74	2-Formylthiophene
30	11.76	2-Butyl-2-octenal
31	11.85	4-(2-Propenyl)-phenol
32	12.11	Methyl-salicylate
33	12.39	3-Methyl-1,2-cyclopentanedione-
34	12.58	2-Hydroxy-3-methyl-2-cyclopenten-1-one
35	12.62	Nona-3,5-dien-2-one
36	12.65	2-Hydroxybenzoic acid ethyl ester
37	13.25	2,6-Dimethylbenzaldehyde
38	14.02	1-Pentyl-1H-pyrrole
39	14.73	2-Amino-4-quinolinol
40	16.00	2,4-Di-tertbutylphenol
41	16.23	4-Vinylguaiaicol

* the two chromatographic signals are attributed to 1-hydroxy-2-propanone (RT 6.55 and 7.86 min), both exhibiting identical EI mass spectra. The dual 1-hydroxy-2-propanone signals are attributed to strong stationary-phase interactions and/or tautomeric effects on the polar column.

Similar to the results for green coffee, no distinct clusters for the coffee species could be determined. Although *C. arabica* shows a trend to the positive region of PC 3 and *C. liberica* to the negative region of PC 3, a clear separation was not achieved (Figure 5A).

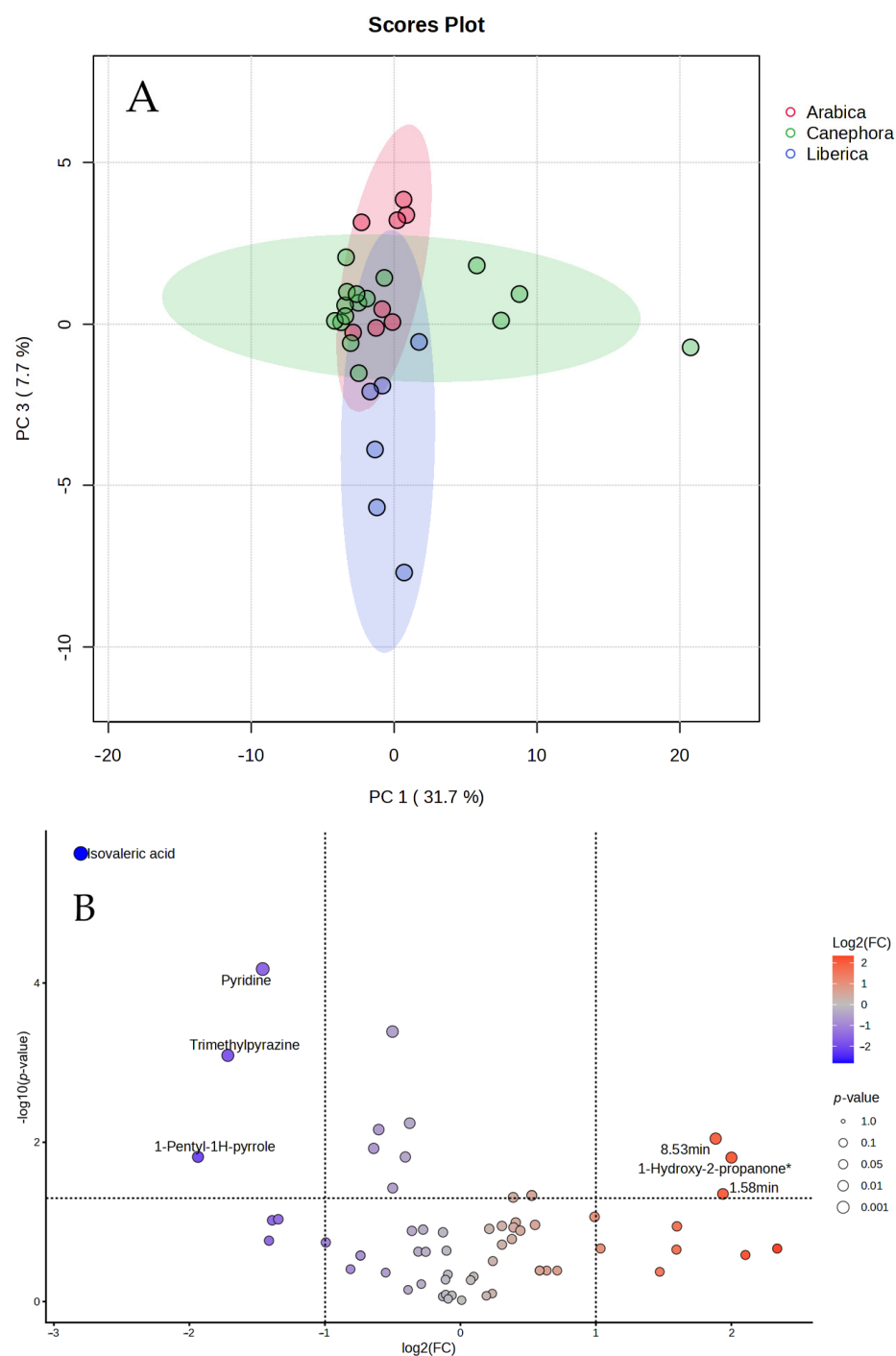


Figure 5. PCA of QMS data from roast coffee samples (auto-scaled, peak intensities) generated in MetaboAnalyst (A). Volcano plot of *C. liberica* and *C. arabica* roast coffee samples with thresholds marked by the dotted lines generated in MetaboAnalyst (auto-scaled, peak intensities, p -value: 0.05) (B). * two chromatographic signals are attributed to 1-hydroxy-2-propanone (RT 6.55 and 7.86 min), both exhibiting identical EI mass spectra.

Due to the close proximity of the clusters, the PCA biplot (Figure S6) could not be used to determine characteristic compounds. The volcano plot showed significant changes in the VOC composition for each coffee species in the roast samples, of which most were annotated (Figures 5B, S7 and S8). Compounds with significantly changed abundance for each coffee species are listed in Table 8.

Table 8. Up-regulated compounds for roast coffee species.

Species	Retention Time [min]	Annotation
<i>C. arabica</i>	3.51	2-Propionylthiazole
	6.22	Acetic acid
	6.53	1-Hydroxy-2-propanone *
	6.77	1-Hydro-2-methyl-3(2H)-furanone
	7.86	1-Hydroxy-2-propanone *
	8.69	Furfural
<i>C. canephora</i>	6.13	Methylpyrazine
	7.93	Trimethylpyrazine
<i>C. liberica</i>	3.51	2-Propionylthiazole
	6.22	Acetic acid
	5.24	Pyridine
	7.93	Trimethylpyrazine
	9.15	Isovaleric acid
	9.45	1-(Acetyloxy)-2-propanone
	9.70	2-Furanmethanol
	10.36	2,3-Butanedione
	10.78	2,5-Hexanedione
	11.27	1-Methyl-1H-pyrrole-2-carboxaldehyde
	14.02	1-Pentyl-1H-pyrrole

* the two chromatographic signals are attributed to 1-hydroxy-2-propanone (RT 6.55 and 7.86 min), both exhibiting identical EI mass spectra. The dual 1-hydroxy-2-propanone signals are attributed to strong stationary-phase interactions and/or tautomeric effects on the polar column.

3.4. Correlation of MS Data with MCR-ALS Feature Tables

To correlate the characteristic compounds annotated by MS data to the MCR feature tables of the IMS data, retention times were compared. Minor shifts in retention time in the prototypic setup were already discussed in previously published results [14] and therefore considered in the correlation between the datasets. For green coffee, a possible correlation of isovaleric acid (9.15 min, MS) was found in the IMS data at 9.09 min. However, as the shifts in retention time between MS and IMS data were positive for all other correlated compounds, propionic acid (9.00 min, MS) was considered as putative annotation for the IMS signal at 9.09 min. A further indication is the absence of the signal at 9.00 min (MS) and 9.09 min (IMS) in the roasted coffees. From a chemical perspective, the presence of propionic acid is in line with published results [46], and the reduction in the propionic acid content during the roasting process was already reported in [47]. Furthermore, ethyl acetate (2.54 min) could be correlated to the IMS feature at 2.60 min. Ethyl acetate, while commonly used for decaffeination of coffee, was already reported to be present in green coffee in literature [48]. However, for a more reliable annotation, further experiments will be required, including spiking experiments with reference materials and further refinement of alignment strategies for IMS and MS traces.

The substantial differences in the VOC detected by the two detectors can be attributed to the higher sensitivity of the IMS detector for polar and medium polar compounds compared to the QMS detector in full scan, as was already reported [36]. Vice versa, the ³H-based IMS system features limited sensitivity for non-polar substances; these are typically not detectable if present only at trace levels, which seems to be the case for green coffee. This underlines the challenge of correlating IMS and EI-MS for compounds that are not optimally detectable by either of the detectors.

In roasted coffee, five characteristic compounds were correlated with the IMS feature table by retention time, of which four were annotated (Table 9).

Table 9. Retention times of characteristic compounds for IMS after MCR-ALS and up-regulated QMS data.

Retention Time IMS [min]	Retention Time QMS [min]	Annotation
8.72	8.69	Furfural
6.57	6.53	1-Hydroxy-2-propanone *
6.16	6.22	Acetic acid
7.79	7.86	1-Hydroxy-2-propanone *
7.97	7.93	Trimethylpyrazine
10.80	10.78	2,5-Hexanedione
14.05	14.02	1-Pentyl-1H-pyrrole

* see Table 8.

The MCR-ALS-extracted features included acetic acid, 1-hydroxy-2-propanone, trimethylpyrazine, 1-pentyl-1H-pyrrole, and furfural. All of these compounds are known products of the Maillard reaction [49–51].

As the described compounds were significantly up-regulated, it was considered that the flavor and aromatic profiles could be influenced by these compounds. Therefore, for a further comparison of the annotated compounds, the characteristic flavor profiles reported in literature for the three coffee species were compared to the odor and flavor of the annotated compounds. *C. arabica* is commonly described as almond-like and caramelly tasting, while *C. canephora* has more cereal- and spice-like attributes [52]. *C. liberica* is often described as having an excessively fermented flavor [53]. While furfural and 1-hydroxy-2-propanone would be in line with the *C. arabica* description and isovaleric acid with the characterization of *C. liberica*, spiking experiments as well as quantitative measurements of these compounds are necessary to confirm the annotations, as well as a potential influence on the aromatic profiles.

Furthermore, this confirmation should also involve HRMS to reach at least a Schymanski [54] level of 2, as for the time being, there are no MS2 data available, which, however, are a prerequisite for a level 1 identification.

4. Conclusions

This study evaluated the potential of THS-GC-IMS paired with modern machine learning for the differentiation of coffee species and annotation of characteristic substances with the simultaneously measured QMS data. The classification without prior feature extraction was compared to MCR-ALS-resolved data with PLS-DA as the classification model. The feature extraction with MCR-ALS significantly enhanced classification accuracy and sensitivity for green coffee samples, which commonly have a less complex and highly similar volatile fingerprint throughout different species when compared to roast coffee. Furthermore, MCR-ALS preserves the chemical and physical information of the original data. This allows for precise determination of retention time and drift time of the characteristic signals in the IMS-spectra.

In order to annotate the IMS signals, the QMS data were first analyzed separately, and subsequently the retention times of the characteristic compounds were compared to the retention times of the characteristic IMS signals. Comparing the characteristic signals for each coffee species in IMS and QMS data, it was possible to tentatively annotate one characteristic compound for green coffee; however, for roast coffee samples, four common characteristic compounds were tentatively annotated in the QMS and IMS data. This study also made clear that there are still challenges to overcome, such as more precise alignment of the resulting data from the two very different detection systems, particularly for such highly complex samples. In a routine environment, the MS section of the system described here might not even be required, provided that the relevant features are identified

or verified and databases are established. Commercially available GC-IMS applications could then be sufficient for such tasks. In conclusion, this study demonstrates the power of THS-GC-IMS in combination with MCR-ALS and PLS-DA for a fast authentication of coffee species, creating the potential for a fast point-of-care technique for the detection of fraudulent coffee blends.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/chemosensors14020034/s1>, Figure S1: MCR-ALS resolved profiles for the retention time of green (A) and roast (C) coffee, as well as the resolved profiles of green (B) and roast (D) coffee samples. Figure S2: Hotelling T²-Q residuals-plot of IMS-data for green (A) and roast (B) coffee after MCR-ALS resolution for outlier evaluation with T2 and Q thresholds (red dotted line). Figure S3: Biplot corresponding to the PCA of green coffee samples of QMS-data generated with MetaboAnalyst (log₁₀-transformed, auto-scaled, peak intensities). Figure S4: Volcano plot of *C. canephora* and *C. liberica* green coffee samples with thresholds marked by the dotted lines generated in MetaboAnalyst (log₁₀-transformed, auto-scaled, peak intensities, *p*-value: 0.05). Figure S5: Volcano plot of *C. arabica* and *C. canephora* green coffee samples with thresholds marked by the dotted lines generated in MetaboAnalyst (log₁₀-transformed, auto-scaled, peak intensities, *p*-value: 0.05). Figure S6: Biplot corresponding to the PCA of roast coffee samples of QMS-data generated with MetaboAnalyst (auto-scaled, peak intensities). Figure S7: Volcano plot of *C. canephora* and *C. liberica* roasted coffee samples with thresholds marked by the dotted lines generated in MetaboAnalyst (auto-scaled, peak intensities, *p*-value: 0.05). Figure S8: Volcano plot of *C. canephora* and *C. arabica* roasted coffee samples with thresholds marked by the dotted lines generated in MetaboAnalyst (auto-scaled, peak intensities, *p*-value: 0.05).

Author Contributions: Conceptualization, P.W.; sample provision, S.S.; methodology, P.W., C.K.; software, C.K., P.W. and H.P.; data curation, C.K. and N.N.; study design of chemometric approaches, C.K., H.P., N.N. and P.W.; writing—original draft preparation, C.K. and P.W.; writing—review and editing, P.W., S.S., S.R. and H.P.; supervision, P.W., S.R.; funding acquisition, P.W.; final approval, C.K., S.R., S.S., H.P., N.N. and P.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Federal Ministry of Education and Research (BMBF), Berlin, Germany, grant number 13FH138KX0 (FH Kooperativ “Deep Authent”).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: Author Steffen Schwarz was employed by the Coffee Consulate (Germany). The remaining authors declare that this research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

References

1. Davis, A.P.; Kiwuka, C.; Faruk, A.; Walubiri, M.J.; Kalema, J. The Re-Emergence of Liberica Coffee as a Major Crop Plant. *Nat. Plants* **2022**, *8*, 1322–1328. [[CrossRef](#)]
2. Monakhova, Y.B.; Ruge, W.; Kuballa, T.; Ilse, M.; Winkelmann, O.; Diehl, B.; Thomas, F.; Lachenmeier, D.W. Rapid Approach to Identify the Presence of Arabica and Robusta Species in Coffee Using 1H NMR Spectroscopy. *Food Chem.* **2015**, *182*, 178–184. [[CrossRef](#)] [[PubMed](#)]
3. Schievano, E.; Finotello, C.; De Angelis, E.; Mammi, S.; Navarini, L. Rapid Authentication of Coffee Blends and Quantification of 16-O-Methylcafestol in Roasted Coffee Beans by Nuclear Magnetic Resonance. *J. Agric. Food Chem.* **2014**, *62*, 12309–12314. [[CrossRef](#)] [[PubMed](#)]
4. DIN EN 18003:2025-01; Lebensmittelauthentizität-Bestimmung des Gehaltes an 16-O-Methylcafestol in Roh-Und Röstkaffee-HPLC-Verfahren; Deutsche Fassung EN_18003:2024. DIN Media GmbH: Berlin, Germany, 2025. [[CrossRef](#)]

5. De Luca, S.; De Filippis, M.; Bucci, R.; Magrì, A.D.; Magrì, A.L.; Marini, F. Characterization of the Effects of Different Roasting Conditions on Coffee Samples of Different Geographical Origins by HPLC-DAD, NIR and Chemometrics. *Microchem. J.* **2016**, *129*, 348–361. [\[CrossRef\]](#)
6. Bertone, E.; Venturello, A.; Giraudo, A.; Pellegrino, G.; Geobaldo, F. Simultaneous Determination by NIR Spectroscopy of the Roasting Degree and Arabica/Robusta Ratio in Roasted and Ground Coffee. *Food Control* **2016**, *59*, 683–689. [\[CrossRef\]](#)
7. Mutz, Y.S.; do Rosario, D.; Galvan, D.; Schwan, R.F.; Bernardes, P.C.; Conte-Junior, C.A. Feasibility of NIR Spectroscopy Coupled with Chemometrics for Classification of Brazilian Specialty Coffee. *Food Control* **2023**, *149*, 109696. [\[CrossRef\]](#)
8. Wermelinger, T.; D'Ambrosio, L.; Klopprogge, B.; Yeretizian, C. Quantification of the Robusta Fraction in a Coffee Blend via Raman Spectroscopy: Proof of Principle. *J. Agric. Food Chem.* **2011**, *59*, 9074–9079. [\[CrossRef\]](#)
9. El-Abassy, R.M.; Donfack, P.; Materny, A. Discrimination between Arabica and Robusta Green Coffee Using Visible Micro Raman Spectroscopy and Chemometric Analysis. *Food Chem.* **2011**, *126*, 1443–1448. [\[CrossRef\]](#)
10. Caporaso, N.; Whitworth, M.B.; Cui, C.; Fisk, I.D. Variability of Single Bean Coffee Volatile Compounds of Arabica and Robusta Roasted Coffees Analysed by SPME-GC-MS. *Food Res. Int.* **2018**, *108*, 628–640. [\[CrossRef\]](#)
11. Vezzulli, F.; Lambri, M.; Bertuzzi, T. Volatile Compounds in Green and Roasted Arabica Specialty Coffee: Discrimination of Origins, Post-Harvesting Processes, and Roasting Level. *Foods* **2023**, *12*, 489. [\[CrossRef\]](#)
12. Zakidou, P.; Plati, F.; Matsakidou, A.; Varka, E.-M.; Blekas, G.; Paraskevopoulou, A. Single Origin Coffee Aroma: From Optimized Flavor Protocols and Coffee Customization to Instrumental Volatile Characterization and Chemometrics. *Molecules* **2021**, *26*, 4609. [\[CrossRef\]](#)
13. Nuguri, S.M.; Gonzalez, C.M.; Hyseni, B.; Aykas, D.P.; Barineau, M.; Rodriguez-Saona, L. Application of Handheld Near-Infrared Technology for in-Field Analysis of Non-Volatile Traits in Fresh Market Tomatoes. *Appl. Food Res.* **2025**, *5*, 101186. [\[CrossRef\]](#)
14. Bodenbender, L.; Rohn, S.; Sauer, S.; Jungen, M.; Weller, P. Chiral Trapped-Headspace GC-QMS-IMS: Boosting Untargeted Benchtop Volatilomics to the Next Level. *Chemosensors* **2024**, *12*, 165. [\[CrossRef\]](#)
15. Yao, W.; Cai, Y.; Liu, D.; Zhao, Z.; Zhang, Z.; Ma, S.; Zhang, M.; Zhang, H. Comparative Analysis of Characteristic Volatile Compounds in Chinese Traditional Smoked Chicken (Specialty Poultry Products) from Different Regions by Headspace–Gas Chromatography–ion Mobility Spectrometry. *Poult. Sci.* **2020**, *99*, 7192–7201. [\[CrossRef\]](#)
16. Capitain, C.C.; Nejati, F.; Zischka, M.; Berzak, M.; Junne, S.; Neubauer, P.; Weller, P. Volatilomics-Based Microbiome Evaluation of Fermented Dairy by Prototypic Headspace-Gas Chromatography–High-Temperature Ion Mobility Spectrometry (HS-GC-HTIMS) and Non-Negative Matrix Factorization (NNMF). *Metabolites* **2022**, *12*, 299. [\[CrossRef\]](#)
17. Papp, Z.; Nemeth, L.G.; Nzetchouang Siyapndjeu, S.; Bufa, A.; Marosvölgyi, T.; Gyöngyi, Z. Classification of Plant-Based Drinks Based on Volatile Compounds. *Foods* **2024**, *13*, 4086. [\[CrossRef\]](#)
18. Rodríguez-Maecker, R.; Vyhmeister, E.; Meisen, S.; Rosales Martinez, A.; Kuklya, A.; Telgheder, U. Identification of Terpenes and Essential Oils by Means of Static Headspace Gas Chromatography-Ion Mobility Spectrometry. *Anal. Bioanal. Chem.* **2017**, *409*, 6595–6603. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Lin, Y.; Yu, G.; Zhang, S.; Zhu, G.; Yi, F. Comparative Analysis of the Differences in Volatile Organic Components of Three Lavender Essential Oils in Ili Region Using Sensory Evaluation, GC-IMS and GC-MS Techniques. *J. Chromatogr. A* **2024**, *1731*, 465197. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Parastar, H.; Weller, P. Towards Greener Volatilomics: Is GC-IMS the New Swiss Army Knife of Gas Phase Analysis? *TrAC Trends Anal. Chem.* **2024**, *170*, 117438. [\[CrossRef\]](#)
21. Ahrens, A.; Zimmermann, S. Towards a Hand-Held, Fast, and Sensitive Gas Chromatograph-Ion Mobility Spectrometer for Detecting Volatile Compounds. *Anal. Bioanal. Chem.* **2021**, *413*, 1009–1016. [\[CrossRef\]](#)
22. Chen, Y.; Chen, H.; Cui, D.; Fang, X.; Gao, J.; Liu, Y. Fast and Non-Destructive Profiling of Commercial Coffee Aroma under Three Conditions (Beans, Powder, and Brews) Using GC-IMS. *Molecules* **2022**, *27*, 6262. [\[CrossRef\]](#)
23. Shi, X.; Li, Y.; Huang, D.; Chen, S.; Zhu, S. Characterization and Discrimination of Volatile Compounds in Roasted Arabica Coffee Beans from Different Origins by Combining GC-TOFMS, GC-IMS, and GC-E-Nose. *Food Chem.* **2025**, *481*, 144079. [\[CrossRef\]](#)
24. Bordiga, M.; Disca, V.; Manfredi, M.; Barberis, E.; Carrà, F.; Navarini, L.; Lonzarich, V.; Arlorio, M. Fingerprinting of Green Arabica Coffee Volatile Organic Compounds (VOCs): HS-GC-IMS Versus GC × GC-MS. *Int. J. Food Sci.* **2025**, *2025*, 1302823. [\[CrossRef\]](#)
25. Zhao, L.; Wang, Y.; Wang, D.; He, Z.; Gong, J.; Tan, C. Effects of Different Probiotics on the Volatile Components of Fermented Coffee Were Analyzed Based on Headspace-Gas Chromatography-Ion Mobility Spectrometry. *Foods* **2023**, *12*, 1015. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Zhai, H.; Dong, W.; Tang, Y.; Hu, R.; Yu, X.; Chen, X. Characterization of the Volatile Flavour Compounds in Yunnan Arabica Coffee Prepared by Different Primary Processing Methods Using HS-SPME/GC-MS and HS-GC-IMS. *LWT* **2024**, *192*, 115717. [\[CrossRef\]](#)
27. Piotr Konieczka, P.; Aliaño-González, M.J.; Ferreira-González, M.; Barbero, G.F.; Palma, M. Characterization of Arabica and Robusta Coffees by Ion Mobility Sum Spectrum. *Sensors* **2020**, *20*, 3123. [\[CrossRef\]](#)

28. Del Mar Contreras, M.; Arroyo-Manzanares, N.; Arce, C.; Arce, L. HS-GC-IMS and Chemometric Data Treatment for Food Authenticity Assessment: Olive Oil Mapping and Classification through Two Different Devices as an Example. *Food Control* **2019**, *98*, 82–93. [\[CrossRef\]](#)
29. Yang, L.; Liu, J.; Wang, X.; Wang, R.; Ren, F.; Zhang, Q.; Shan, Y.; Ding, S. Characterization of Volatile Component Changes in Jujube Fruits during Cold Storage by Using Headspace-Gas Chromatography-Ion Mobility Spectrometry. *Molecules* **2019**, *24*, 3904. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Martín-Gómez, A.; Rodríguez-Hernández, P.; Cardador, M.J.; Vega-Márquez, B.; Rodríguez-Estévez, V.; Arce, L. Guidelines to Build PLS-DA Chemometric Classification Models Using a GC-IMS Method: Dry-Cured Ham as a Case of Study. *Talanta Open* **2023**, *7*, 100175. [\[CrossRef\]](#)
31. Xu, N.; Lai, Y.; Shao, X.; Zeng, X.; Wang, P.; Han, M.; Xu, X. Different Analysis of Flavors among Soft-Boiled Chicken: Based on GC-IMS and PLS-DA. *Food Biosci.* **2023**, *56*, 103243. [\[CrossRef\]](#)
32. Zhu, W.; Benkwitz, F.; Sarmadi, B.; Kilmartin, P.A. Validation Study on the Simultaneous Quantitation of Multiple Wine Aroma Compounds with Static Headspace-Gas Chromatography-Ion Mobility Spectrometry. *J. Agric. Food Chem.* **2021**, *69*, 15020–15035. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Liu, S.; Dong, H.; Pan, S.; Kang, C.; Wang, X. Characterization of Gastrodia Rhizoma from Different Geographical Origins by HS-GC-IMS and Authenticity Identification Combined with Deep Learning. *J. Chromatogr. A* **2026**, *1765*, 466482. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Parastar, H.; Yazdanpanah, H.; Weller, P. Non-Targeted Volatilomics for the Authentication of Saffron by Gas Chromatography-Ion Mobility Spectrometry and Multivariate Curve Resolution. *Food Chem.* **2024**, *465*, 142074. [\[CrossRef\]](#)
35. Parastar, H.; Tauler, R. Multivariate Curve Resolution of Hyphenated and Multidimensional Chromatographic Measurements: A New Insight to Address Current Chromatographic Challenges. *Anal. Chem.* **2014**, *86*, 286–297. [\[CrossRef\]](#)
36. Brendel, R.; Schwolow, S.; Rohn, S.; Weller, P. Gas-Phase Volatilomic Approaches for Quality Control of Brewing Hops Based on Simultaneous GC-MS-IMS and Machine Learning. *Anal. Bioanal. Chem.* **2020**, *412*, 7085–7097. [\[CrossRef\]](#)
37. Schanzmann, H.; Ruzsanyi, V.; Ahmad-Nejad, P.; Telgheder, U.; Sielemann, S. A Novel Coupling Technique Based on Thermal Desorption Gas Chromatography with Mass Spectrometry and Ion Mobility Spectrometry for Breath Analysis. *J. Breath Res.* **2023**, *18*, 016009. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Christmann, J.; Rohn, S.; Weller, P. Gc-Ims-Tools—A New Python Package for Chemometric Analysis of GC-IMS Data. *Food Chem.* **2022**, *394*, 133476. [\[CrossRef\]](#)
39. Jaumot, J.; de Juan, A.; Tauler, R. MCR-ALS GUI 2.0: New Features and Applications. *Chemom. Intell. Lab. Syst.* **2015**, *140*, 1–12. [\[CrossRef\]](#)
40. Toci, A.T.; Farah, A. Volatile Fingerprint of Brazilian Defective Coffee Seeds: Corroboration of Potential Marker Compounds and Identification of New Low Quality Indicators. *Food Chem.* **2014**, *153*, 298–314. [\[CrossRef\]](#)
41. Christmann, J.; Rohn, S.; Weller, P. Finding Features—Variable Extraction Strategies for Dimensionality Reduction and Marker Compounds Identification in GC-IMS Data. *Food Res. Int.* **2022**, *161*, 111779. [\[CrossRef\]](#)
42. Parastar, H.; Weller, P. Feature Selection and Extraction Strategies for Non-Targeted Analysis Using GC-MS and GC-IMS: A Tutorial. *Anal. Chim. Acta* **2025**, *1343*, 343635. [\[CrossRef\]](#)
43. Bayat-Afshary, F.; Naderi Tehrani, N.; Bodenbender, L.; Weller, P.; Parastar, H. Benchtop Volatilomics and Advanced Convolutional Neural Network Workflows for Accurate and Explainable Food Authentication. *Food Chem.* **2026**, *500*, 147511. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Rogers, J.; Gunn, S. Identifying Feature Relevance Using a Random Forest. In *Subspace, Latent Structure and Feature Selection*; Saunders, C., Grobelnik, M., Gunn, S., Shawe-Taylor, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 173–184.
45. Ali, S.; Abuhmed, T.; El-Sappagh, S.; Muhammad, K.; Alonso-Moral, J.M.; Confalonieri, R.; Guidotti, R.; Del Ser, J.; Díaz-Rodríguez, N.; Herrera, F. Explainable Artificial Intelligence (XAI): What We Know and What Is Left to Attain Trustworthy Artificial Intelligence. *Inf. Fusion* **2023**, *99*, 101805. [\[CrossRef\]](#)
46. Haile, M.; Kang, W.H. The Role of Microbes in Coffee Fermentation and Their Impact on Coffee Quality. *J. Food Qual.* **2019**, *2019*, 4836709. [\[CrossRef\]](#)
47. Gloess, A.N.; Vietri, A.; Wieland, F.; Smrke, S.; Schönbächler, B.; López, J.A.S.; Petrozzi, S.; Bongers, S.; Koziorowski, T.; Yeretzyan, C. Evidence of Different Flavour Formation Dynamics by Roasting Coffee from Different Origins: On-Line Analysis with PTR-ToF-MS. *Int. J. Mass Spectrom.* **2014**, *365–366*, 324–337. [\[CrossRef\]](#)
48. Lazaridis, D.G.; Kokkosi, E.K.; Mylonaki, E.N.; Karabagias, V.K.; Andritsos, N.D.; Karabagias, I.K. Rapid Classification of Unroasted Green Coffee Beans and Spices Based on the Tentative Determination of Volatile Compounds by Solid-Phase Dynamic Extraction (SPDE) and Gas Chromatography–Mass Spectrometry (GC–MS) with Supervised Learning. *Separations* **2024**, *11*, 351. [\[CrossRef\]](#)

49. Nursten, H.E. The Mechanism of Formation of 3-Methylcyclopent-2-En-2-Olone. In *The Maillard Reaction in Foods and Medicine*; Woodhead Publishing Series in Food Science, Technology and Nutrition; O'Brien, J., Nursten, H.E., Crabbe, M.J.C., Ames, J.M., Eds.; Woodhead Publishing: Cambridge, UK, 2005; pp. 65–68, ISBN 978-1-85573-791-4.
50. Davidek, T.; Gouézec, E.; Devaud, S.; Blank, I. Origin and Yields of Acetic Acid in Pentose-Based Maillard Reaction Systems. *Ann. N. Y. Acad. Sci.* **2008**, *1126*, 241–243. [[CrossRef](#)]
51. Filipowska, W.; Jaskula-Goiris, B.; Ditrych, M.; Bustillo Trueba, P.; De Rouck, G.; Aerts, G.; Powell, C.; Cook, D.; Cooman, L. On the Contribution of Malt Quality and the Malting Process to the Formation of Beer Staling Aldehydes: A Review. *J. Inst. Brew.* **2021**, *127*, 107–126. [[CrossRef](#)]
52. Sunarharum, W.B.; Williams, D.J.; Smyth, H.E. Complexity of Coffee Flavor: A Compositional and Sensory Perspective. *Food Res. Int.* **2014**, *62*, 315–325. [[CrossRef](#)]
53. Lee, K.W.T. Liberica Coffee Development and Refinement Project in Sarawak Malaysia. *Proceedings* **2023**, *89*, 15. [[CrossRef](#)]
54. Schymanski, E.L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H.P.; Hollender, J. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* **2014**, *48*, 2097–2098. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.