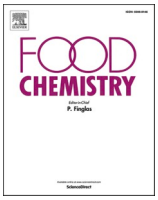
Food Chemistry 463 (2025) 141432

Contents lists available at ScienceDirect

Food Chemistry

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



The high-resolution molecular portrait of coffee: A gateway to insights into its roasting chemistry and comprehensive authenticity profiles

Stefan A. Pieczonka [a,b,*], Anna Dzemajili [a,c], Silke S. Heinzmann [b], Michael Rychlik [a], Philippe Schmitt-Kopplin [a,b,*]

[a] Analytical Food Chemistry, TUM School of Life Sciences, Technical University of Munich, Freising, Germany
[b] Analytical BioGeoChemistry, Helmholtz Association, Helmholtz Munich, Neuherberg, Germany
[c] Analytical Chemistry, Department of Applied Sciences and Mechatronics, Munich University of Applied Sciences, Munich, Germany

ARTICLE INFO

Keywords:
coffea arabica
coffea canephora
FT-ICR-MS
Food authenticity
Maillard reaction
Roasting
Metabolomics
Foodomics

ABSTRACT

The direct-infusion of 130 coffee samples into a Fourier-transform ion cyclotron mass spectrometer (FT-ICR-MS) provided an ultra-high resolution perspective on the molecular complexity of coffee: The exceptional resolving power and mass accuracy (± 0.2 ppm) facilitated the annotation of unambiguous molecular formulas to 11,500 mass signals.

Utilizing this molecular diversity, we extracted hundreds of compound signals linked to the roasting process through guided Orthogonal Partial Least Squares (OPLS) analysis. Visualizations such as van Krevelen diagrams and Kendrick mass defect analysis provided deeper insights into the intrinsic compositional nature of these compounds and the complex chemistry underlying coffee roasting.

Predictive OPLS-DA models established universal molecular profiles for rapid authentication of *Coffea arabica* versus *Coffea canephora* (Robusta) coffees. Compositional analysis revealed Robusta specific signals, indicative of tryptophan-conjugates of hydroxycinnamic acids. Complementary LC-ToF-MS^2 confirmed their compound class, building blocks and structures. Their water-soluble nature allows for application across raw and roasted beans, as well as in ready-made coffee products.

1. Introduction

Originating in Africa and having spread across the globe, coffee's rich history reflects its evolution from a traditional beverage to a staple of social and cultural life around the world (Aregay, 2009). Globally, over 10 billion kilograms of coffee are produced annually, with *Coffea arabica* (~ 60 %) and *Coffea canephora (~ 40 %,* commonly known as Robusta) (United States Department of Agriculture, 2023), collectively dominating the coffee growing sector, and giving rise to a wide variety of coffee products. Apart from the caffeine content, there is a significant difference in market price between these two types of coffee. Amid concerns about authenticity, the molecular composition of coffee has become the subject of numerous studies, as summarized by Perez, Dominguez-Lopez, Lopez-Yerena, and Vallverdu Queralt (2023): Certain diterpenes (Dias et al., 2010), diterpene-glucosides (Maier & Wewetzer, 1978), and tocopherols (Alves, Casal, Alves, & Oliveira, 2009) are characteristic of the more aromatic *Coffea arabica* bean, which achieves a market value up to 25 % higher. Cinnamoyl-amino acid conjugates (M. Clifford, 2004) are found in increased levels in the body and foam-producing Robusta beans. Regulatory authorities utilize liquid chromatography (HPLC) and nuclear resonance spectroscopy (^{1}H NMR) to quantify the lipophilic 16-O-Methylcafestol (16-OMC) as a marker substance for coffee brewed with Robusta beans that are considered to be of lower quality (German standard methods, 2011, 2023; Kurzrock & Speer, 2007; Monakhova et al., 2015; Schievano, Finotello, De Angelis, Mammi, & Navarini, 2014; Speer & Mischnick, 1989). Gunning et al. (2018) emphasize the need to establish concentration limits for the practical interpretation of results, as new methods and detection limits challenge the binary presence/absence determination of single marker compounds. Non-targeted spectroscopic and spectrometric methods oppose those individual indicator compounds with a molecular fingerprint, promising authenticity control based on multivariate statistical evaluation beyond single factors susceptible to falsification. While blend ratios can be accurately captured in individual studies (Garrett, Vaz, Hovell, Eberlin, & Rezende, 2012), the limitation of models to a few molecular descriptors leaves open the question of representativeness

* Corresponding authors at: Analytical Food Chemistry, TUM School of Life Sciences, Technical University of Munich, Freising, Germany.
E-mail addresses: stefan.pieczonka@tum.de (S.A. Pieczonka), schmitt-kopplin@tum.de (P. Schmitt-Kopplin).

https://doi.org/10.1016/j.foodchem.2024.141432
Received 21 June 2024; Received in revised form 30 August 2024; Accepted 23 September 2024
Available online 24 September 2024

and reproducibility in many cases.

In light of these considerations, our study aims to use a rapid high-throughput direct infusion method on a Fourier-transform ion cyclotron resonance mass spectrometer (FT-ICR-MS) to reveal an ultra-high resolution molecular profile of coffee The anticipated resolution of thousands of molecular compositions will establish a basis for distinguishing Arabica and Robusta coffee brews, concurrently shedding light on the molecular complexity inherent in coffee. A key determinant of coffee character and a major contributor to its complexity is the roasting process. Chemical changes during roasting include the thermally induced decomposition of hydroxycinnamic acid derivatives (Perez-Miguez, Sanchez-Lopez, Plaza, Castro-Puyana, & Marina, 2018) and lipid oxidation (Zhu et al., 2023). In addition, the Maillard reaction is instrumental in the formation of numerous volatile compounds that enrich coffee aroma (Sunarharum, Williams, & Smyth, 2014). The pathways leading to pertinent aroma compounds, as well as the foundational reaction scheme involving amino components and reducing sugars - from Schiff bases through Amadori intermediates and reactive α-dicarbonyls to Maillard reaction products like 5-hydroxymethylfurfural (HMF) or furfural, culminating in melanoidin condensation - have been thoroughly described (Murkovic & Bornik, 2007). Although many individual compounds have been identified and elucidated, the overall complexity of the Maillard reaction network during the coffee roasting process and the resulting diversity of compounds generated by roasting chemistry remains inadequately captured. As previously demonstrated in the context of the malting process in beer production (S. A. Pieczonka et al., 2021), ultra-high resolution mass spectrometry is proposed to provide a comprehensive insight into the molecular diversity resulting from the changes induced by Maillard and roasting chemistry during coffee roasting.

2. Materials & methods

2.1. Sample set and sample preparation

In the FT-ICR-MS study, 130 distinct coffee samples were analyzed, encompassing a wide variety of market-available coffees. The sample set included both whole bean and ground instant coffees derived from *Coffea arabica* and *Coffea canephora*. The samples were procured through both conventional and organic farming practices, covering a spectrum of green through lightly roasted, to deeply roasted coffee varieties. A comprehensive overview of all samples, representing the diversity of coffee in various combinations of meta information, is available in Supplementary Table S1. Information on the coffee samples was derived from the specifications provided by the manufacturers and distributors. Coffee variety information was obtained from the supplier and product labels, with any discrepancies resolved through direct confirmation with roasteries and their expert morphological assessment. The authenticity of two samples that remained in conflict with the statistical model (outlier in the training set; misclassified in the test set) was assessed using NMR-based analysis, adhering to the draft of a standardized DIN procedure. The assessment focused on the diterpene signal at 3.16 ppm, predominantly caused by 16-methylcafestol, on an 800 MHz Bruker AVANCE III spectrometer equipped with a with a quadruple inverse cryogenic probe at 300 K (German standard methods, 2023) (Supplementary Fig. S1). The roast level (420 nm) (S. A. Pieczonka et al., 2021; Yu, Zhao, Hu, Zeng, & Bai, 2012) and content of chlorogenic acid derivatives (325 nm) (Duarte, Pereira, & Farah, 2010) were semi-quantitatively estimated analyzing a 1 to 100 dilution of the prepared coffee by UV–Vis spectroscopy in Nunc UV-transparent 96-well microtiter plates performed on a Multiskan Sky UV–Vis reader (Thermo Fisher Scientific, Waltham, MO, USA) with temperature control (23 °C).

Coffee samples were homogenized using a CryoMill (Retsch, Haan, Germany), achieving a consistent average particle size of 60 μm (10 beans for 1 min at 30 Hz). Following the Specialty Coffee Association guidelines, coffee solution concentration was normalized to 55 mg/L in MilliQ-purified water from a Milli-Q Integral Water Purification System (18.2 MΩ, Billerica, MA, USA). Subsequently, 5 mL of this solution underwent mixing in a ThermoMixer C (Eppendorf, Hamburg, Germany) at 500 rpm for 8 min at 90 °C. Immediate cooling in ice ceased brewing, followed by centrifugation at 4000 rpm for 5 min to clarify the solution (Centrifuge 5425, Eppendorf, Hamburg, Germany). The clarified supernatant was stored at −20 °C, ensuring sample stability for subsequent analyses. The samples were prepared and measured in a series of batches over a period of three months.

2.2. FT-ICR-MS metabolome analysis

Dilution series, ionization polarities, and solid-phase extractions were tested on an aliquot sample in triplicates to mitigate ion suppression effects and achieve optimal ion yield (Supplementary Fig. S2). In the final protocol, the brewed coffee samples were diluted 1:1000 in methanol prior to injection into the solariX Ion Cyclotron Resonance FT-ICR-MS system (Bruker Daltonics GmbH, Bremen, Germany) equipped with a 12 T superconducting magnet (Magnex Scientific Inc., Yarton, GB) and a APOLO II ESI source (BrukerDaltonics GmbH, Bremen, Germany) operated in negative ionization mode (300 scans, m/z 120 to 1000). Parallel measurement in both polarity modes was omitted due to the limited amount of additional information, the distinct $[Na]^+$ and $[K]^+$ adduct formations in ESI^+ and the resulting different chemical spaces for annotation. Source and ICR cell parameters, data calibration, processing, and filtering followed established protocols from previous studies (S. A. Pieczonka et al., 2021; S.A. Pieczonka, Lucio, Rychlik, & Schmitt-Kopplin, 2020). The mass spectrometer demonstrated stable resolving power of 420,000 at m/z 400, allowing for the assignment of 11,500 monoisotopic signals to molecular formulas within the CHNO compositional space, an average mass error of ±0.2 ppm, and a signal-to-noise ratio exceeding five. Sulfur annotations and potential dihydrogen phosphate or chloride adducts were omitted due to annotation interference with isotopologues and multiply charged ions, as they accounted for less than 2 % of the formulas.

2.3. LC-ToF-MS structure verification

Prior to reversed-phase LC-ToF-MS analysis of eight representative coffee samples (three Robusta samples 2, 6, 16; the questionable samples 1 and 72; and three Arabica samples 35, 83,86 as negative control), we have separated interfering polar matrix components by solid-phase extraction (SPE) (Supplementary Table S2). The SPE treated, undiluted coffee samples were analyzed using a Sciex X500R QTOF system (AB Sciex, Darmstadt, Germany) in data dependent acquisition and negative ionization mode, coupled to a UPLC ExionLC (AB Sciex, Darmstadt, Germany) chromatography. The chromatographic and mass spectrometric parameters are summarized in Supplementary Table S2. Accurate molecular masses identified as significant in the FT-ICR-MS analysis (method section 2.4.1) were preferentially fragmented within the chromatographic run (fragmentation inclusion list). Resulting .wiff2 files were centered and converted to mzML format using the MSConvert software (ProteoWizard) (Chambers et al., 2012). Parameters for subsequent data processing with mzMine3 (Schmid et al., 2023) (Version 3.9.0) are detailed in Supplementary Table S3. The feature list was exported as an MGF file and analyzed within the Sirius 5.8.2 software environment (Dührkop et al., 2019). Tandem mass spectrometric data were utilized to determine molecular formulas via fragmentation trees (Bocker & Duhrkop, 2016), characterize compound classes using CANOPUS (Djoumbou Feunang et al., 2016), and generate structural suggestions via CSI:FingerID (Duhrkop, Shen, Meusel, Rousu, & Bocker, 2015). Databases consulted and parameters applied in this process are detailed in Supplementary Table S3. Identification confidence was classified as proposed by Schymanski et al. (2014), incorporating results from database searches, literature-known, and in silico calculated diagnostic ions and neutral losses (M. N. Clifford, Marks, Knight, &

Kuhnert, 2006; De Rosso, Colomban, Flamini, & Navarini, 2018; Jaiswal, Patras, Eravuchira, & Kuhnert, 2010; Wu et al., 2009).

2.4. Statistical data analysis and data visualization

2.4.1. PCA and OPLS analysis

Following zero-filling, z-score normalization, and centering (Pieczonka et al., 2023), we performed principal component analysis (PCA) to the entire dataset to evaluate its variance, identify any potential batch effects, and ensure the coherence between the test and training sets for subsequent supervised statistical analysis. We organized the coffee dataset into a training set (81 samples) and a test set (49 samples) for orthogonal partial least squares discriminant analysis (OPLS-DA) analysis. We performed supervised OPLS analysis of the training set using the absorbance of the samples at Maillard-wavelength 420 nm as the y-variable, utilizing the ropls package (R 4.1.2) within the RStudio environment (version 2023.12.0). For comparative purposes, we created an OPLS model based on absorbance values at 325 nm (chlorogenic acid derivatives). Furthermore, we distinguished the molecular data of *Coffea arabica*-labeled coffees (46 samples) from pure *Coffea canephora* coffees (17 samples) using OPLS-DA, excluding mixed coffee varieties from the analysis. Subsequently, the positions of the test set coffees were predicted within this model's score plot and the correctness of their classification was determined.

To mitigate the impact of significant outliers on the models, we employed Hotelling's T2 test at a 95 % data level. We assessed the model's accuracy and predictive capability through the R^2Y and Q^2 metrics (Golbraikh & Tropsha, 2002). We limited the number of principal components in OPLS-(DA) models to three in order to assure an accurate representation of the Q^2 parameter and the prediction capability. Given the distinction between regression (OPLS) and classification (OPLS-DA) modeling, Q^2 thresholds were set at >0.4 and > 0.7, respectively, while considering consistent R^2Y-values. A model differentiating organic against conventional farming did not meet these criteria. To further rule out the possible overfitting, the *p*-value from the Cross-Validation Analysis of Variance (CV-ANOVA) was calculated. Significant features were identified based on a threshold for the Variable Importance for Projection (VIP) value, which was set at 1.5 for OPLS and 3 for OPLS-DA. Significant annotations were assessed for plausibility confirming compositional patterns in van Krevelen diagrams and molecular networks.

2.4.2. Data visualization

The significant molecular formulas of the OPLS-(DA-)models were visualized in van Krevelen diagrams (hydrogen-to-carbon ratio against oxygen-to-carbon ratio) to identify compositional clusters and tentatively describe their respective (bio)chemical origin and compound class (Dou, Mei, Kettunen, Makinen, & Janis, 2022; He, Liu, Nam, Klasson, & Cheng, 2023; Roullier-Gall, Boutegrabet, Gougeon, & Schmitt-Kopplin, 2014; Schmitt-Kopplin et al., 2019). For molecular formulas significant in the OPLS-model (420 nm), their double-bond-equivalents (DBE) were calculated and plotted against the number of carbons. We applied a modified Kendrick mass defect analysis (Kim, Kramer, & Hatcher, 2003) to visualize the role of dehydration reaction cascades.

The mass difference network, derived from the annotation of mass signals (Tziotis, Hertkorn, & Schmitt-Kopplin, 2011), was employed to establish relationships among significant compounds. The mass differences used for this analysis are compiled in Supplementary Table S4. The resulting mass difference networks were visualized using the openly available Gephi Viz Platform (Bastian, Heymann, & Jacomy, 2009) through the application of the forced atlas algorithm.

3. Results and discussion

3.1. The compositional space of coffee

The selection of a diverse set of 130 samples (Supplementary Table 1) is a fundamental aspect of our study, facilitating the exploration of coffee's molecular diversity and complexity through ultra-high resolution analysis. Using FT-ICR mass spectrometry, we assigned monoisotopic molecular formulas to over 11,500 mass signals in the coffee sample set, thereby accessing a level of molecular complexity that exceeds comparable studies by two orders of magnitude (Garrett et al., 2012). As shown by Garrett et al. (2012), ion suppression effects, particularly due to metal ion adducts (K^+, Na^+), pose a significant challenge to the detection of minor components in direct infusion MS of coffee. Therefore, to mitigate these effects, we opted for a 1000-fold sample dilution in methanol and the use of negative ionization, which provided significant advantages (Supplementary Fig. 1). By optimizing the sample preparation, dilution, and polarity, we were able to achieve up to a fivefold increase in accurate mass signals. Despite the dominant ionization of caffeic and quinic acid derivatives in coffee ESI^- (Fig. 1A), our approach allowed for the annotation of up to 30 isobaric molecular compositions within a single nominal mass, resulting in an average of 4315 molecular formulae per sample (Fig. 1B).

Employing network calculations for annotation (Tziotis et al., 2011) revealed the molecular diversity of coffee within a single nominal mass, capturing a range of compounds from highly saturated ($C_{34}H_{66}O_4$) to very unsaturated ($C_{28}H_{26}O_{11}$) and oxygenated ($C_{21}H_{30}O_{16}$) structures, as well as those rich in nitrogen ($C_{26}H_{46}N_6O_6$) and multi-isotopologue peaks ($^{13}C_1C_{26}H_{54}N_9O_2$). With an average of 4315 annotated signals per sample and an average measurement error of <0.1 ppm, the complexity of the sample set culminates in a total of 11,500 monoisotopic compositions. Visualized in a van Krevelen diagram (plotting H/C versus O/C ratios), the predominant intensities of caffeic acid, quinic acid, and chlorogenic acid derivatives become apparent (Fig. 2). Beyond these dominant peaks in the coffee spectrum, a unique fingerprint emerges, revealing a rich compositional landscape that includes small organic acids, lipids, sugars, peptides, polyphenols, and the Maillard reaction products (MRPs) being characteristic of the roasting process. The chemistry of such MRPs has been the focus of extensive research through model system experiments (Hemmler et al., 2017, 2018; Kanzler, Wustrack, & Rohn, 2021) and studies targeting pathways leading to key aroma components (Buffo & Cardelli-Freire, 2004; Poisson, Blank, Dunkel, & Hofmann, 2017). The intricate complexity observed in real food matrices, as manifested in the diverse and multifaceted reaction products of the Maillard and roasting process, far exceeds what can be captured by simplified models or targeted approaches alone (S. A. Pieczonka et al., 2021).

3.2. The complex chemistry of roasting products

To explore the chemical intricacies of Maillard reaction products, we selected the 420 nm absorbance of coffee samples as the y-variable in our OPLS statistical analysis, designed to isolate these compounds (S. A. Pieczonka et al., 2021). This region in the UV/Vis spectrum is generally recognized as indicative of advanced reaction products (Yu et al., 2012) and, unlike the absorption of intermediates (294 nm), is not superimposed by bands of chlorogenic acids (325 nm) (M. N. Clifford, 1985) and caffeine (273 nm) (Ahmad Bhawani, Fong, & Mohamad Ibrahim, 2015) (Supplementary Fig. S3). Through PCA analysis, we excluded batch effects and verified the coherence of the dataset, specifically between the training and test sets for the OPLS-(DA) analyses (Supplementary Fig. S4). The goodness of fit (R^2Y 0.94) and prediction capability (Q2 0.51), together with the ANOVA-CV (*p*-value $<<$ 0.01), confirm a significant OPLS model without overfitting (Fig. 3A). The model is designed to extract significant signals that correlate with absorbance values and heat impact. An accurate prediction of

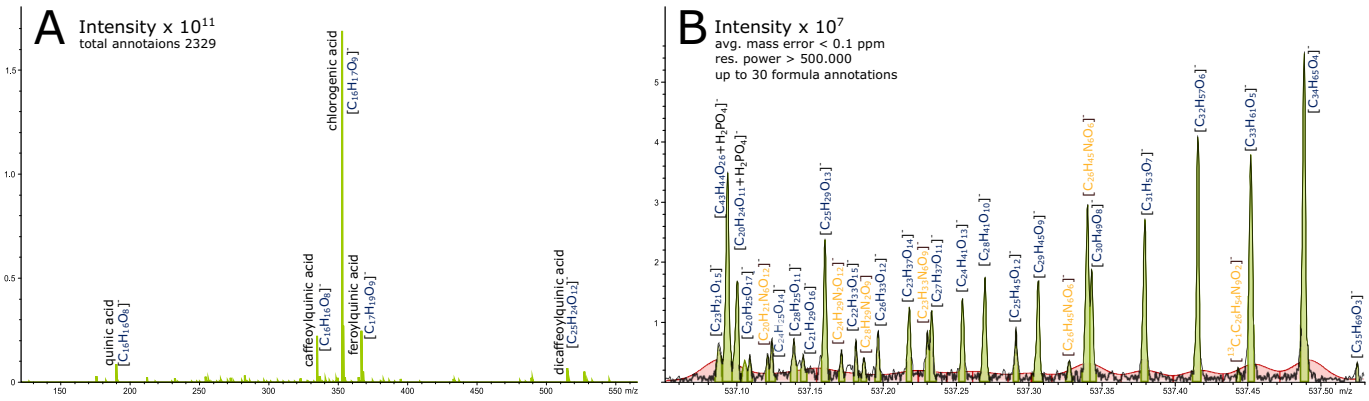


Fig. 1. Wide dynamic range of DI-FT-ICR mass spectra of coffee displayed as the whole *m/z* spectrum (A) and the segment of signals at the single nominal mass *m/z* 537 (B). The spectrum (sample 55) is characterized by base peaks of coffee and quinic acid derivatives, which cause substantial ion suppression. Sufficient dilution (1 to 1000) makes it is possible to annotate over 30 different molecular compositions within a single nominal mass. In B, a computational reduction of the mass resolution power to 40,000 is indicated in red in the background. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

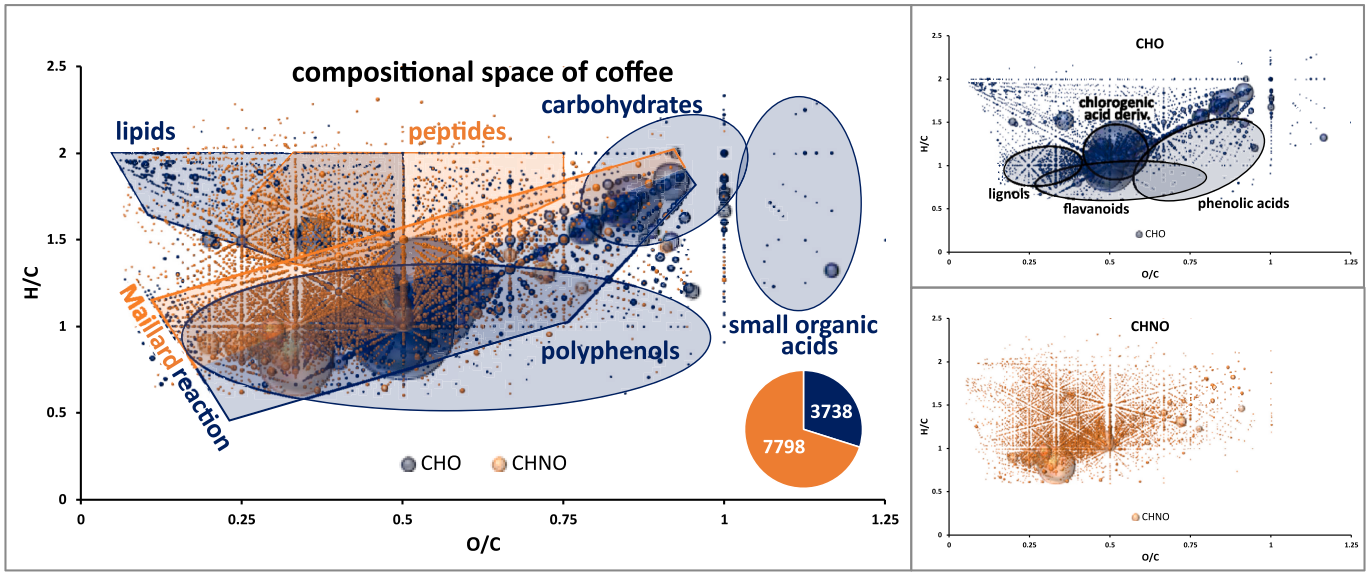


Fig. 2. Van Krevelen diagram of annotated CHO/CHNO-compositions in the coffee sample set. Tentative compound classes are highlights; molecular formulas containing only carbon, hydrogen, and oxygen are depicted in blue, while those containing nitrogen are shown in orange. Bubble sizes represent intensity values. The compositional spaces (CHO with a detailed structuring of phenolics; CHNO) are individually illustrated on the right. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

absorption values based on compositional data, in the context of a predictive regression model, is not anticipated.

The molecular formulas of nearly 1500 mass signals that correlate with higher absorbance values, thus indicating strong Maillard and roasting chemistry, show a characteristically diverse yet clearly structured arrangement in the van Krevelen representation (Fig. 3B-I). The formulas range from simple desoxyosones ($C_6H_{10}O_5$, hexose-derived desoxiosone) to well-known MRPs of amino acid or peptide conjugates ($C_{15}H_{24}N_2O_6$, pronyl-lysine). As the reaction progresses, dehydration cascades finally lead to extremely unsaturated and low-oxygenated advanced MRPs ($C_{16}H_{13}NO_3$). Sulfur-containing compounds, possibly originating from cysteine-including peptide products, were omitted due to their low abundance (<1 %) and annotation interferences. The list of all annotations is compiled in Supplementary Table S5 and contrasted with the van Krevelen representation of compounds predominantly found in lightly roasted coffees (Fig. 3B-II), notable for a cluster of disaccharide carbohydrates and chlorogenic acid derivatives. Both classes of compounds are known to react or decompose under thermal influence (Souto et al., 2010). Consequently, the OPLS model targeting the 325 nm absorbance band of chlorogenic acid shows almost identical Van Krevelen plots with reversed signs (Supplementary Fig. S5). Additional visualizations outline the intrinsic structure of the MRPs underlying the kinetic-chemical reaction cascade occurring in coffee: The number of identified compositions decreases linearly with the number of nitrogen atoms bound to the sugar backbone by conjugations (e.g., Amadori or Heyns products) (Fig. 3C-I). The distribution of oxygen atoms follows a Gaussian pattern (Fig. 3C-II). The number of double bond equivalents increases with molecular size, suggesting that many high molecular weight compounds are products of the advanced Maillard reaction phase (Fig. 3C-III). The Kendrick mass defect plot, adjusted for the mass of water (H_2O, 18.010565), reflects the dehydration cascades that occur particularly in the intermediate phase of the Maillard reaction (Fig. 3C-IV). This structured compositional space is not observed for the lightly roasted compounds (Supplementary Fig. S6).

Using these visualization techniques, we not only elucidate the intrinsic compositional nature of Maillard and caramelization chemistry

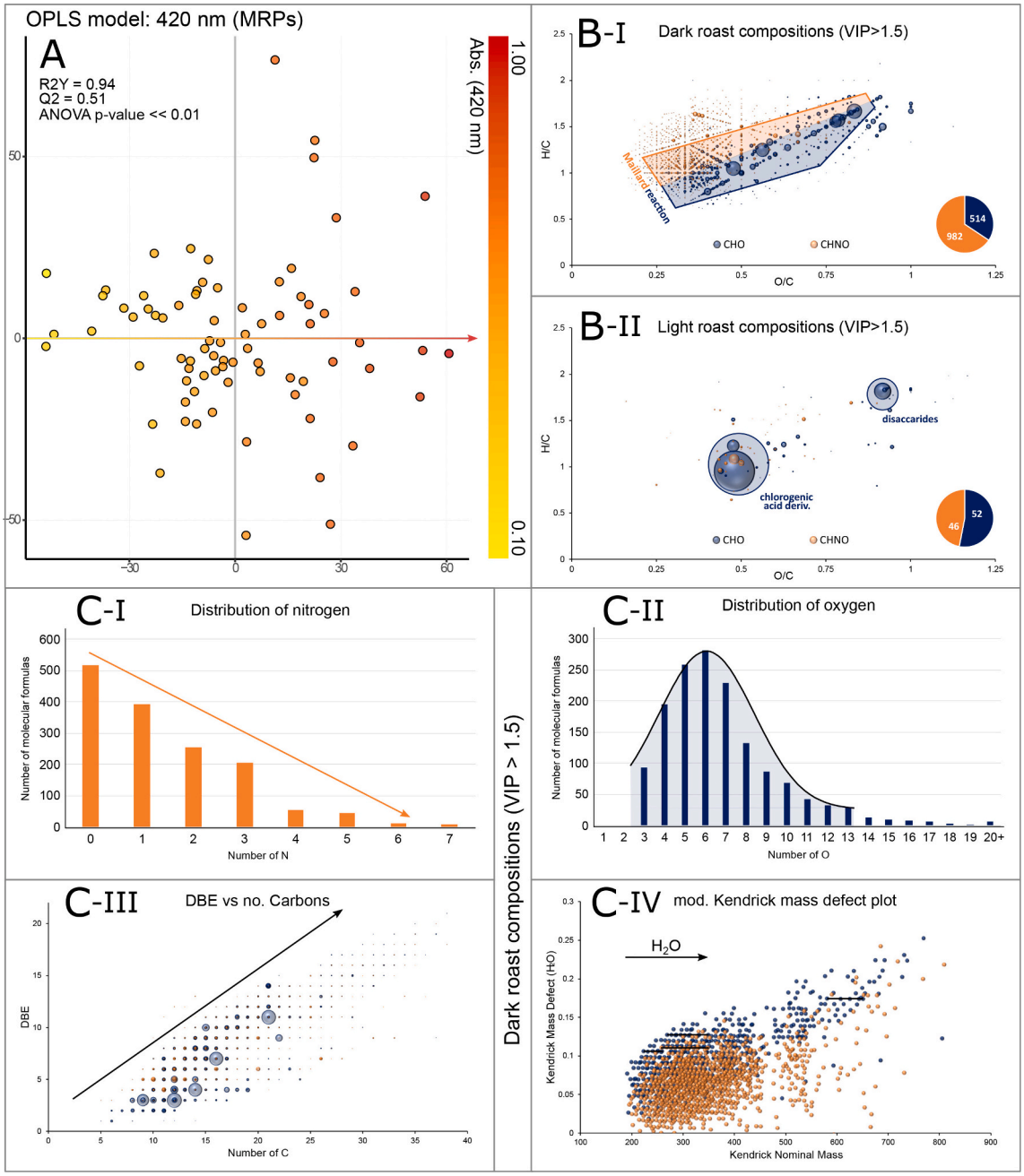


Fig. 3. OPLS model of the training set ($n = 81$) with the absorbance at 420 nm as y-variable (A) and characterization of the significant compositions for light and dark roasting (B—C). The score plot is shown in (A). Van Krevelen plots of features with significantly high and low loading values (VIP > 1.5) are shown in B—I and B-II, respectively. Nearly 1500 positively correlating compositions are found in the Maillard reaction chemical space, while negatively correlating compositions are distributed in regions associated with chlorogenic acid derivatives and carbohydrates. The molecular composition of mass signals correlating with darker roasts are further characterized by their linearly decreasing nitrogen distribution (C—I), the Gaussian distribution of oxygen (C-II), an increasing value of DBE with the carbon number (C-II), and the H_2O dehydration cascades in a modified Kendrick mass defect plot (C-IV).

in coffee, but also find that our results are consistent with studies of roasting-derived compounds in beer (S. A. Pieczonka et al., 2021). While the basic trends are identical, it is noteworthy that we identified only 339 (50 %) of the molecular compositions found as characteristic for dark beers to be equally significant for strong roasting in coffee. Despite the same reaction principles and mechanisms, differences in temperature control and the composition of sugars and amino components lead to a varied manifestation of complexity among hundreds to thousands of Maillard reaction products in these beverages. A clear distinction between the Maillard chemistry and other heat-induced reaction pathways, such as caramelization and pyrolysis, can be achieved in subsequent projects through isotopically labeled experiments. The potential impact of hundreds to thousands of compounds on the overall sensory impression of coffee through flavor contributions and interactions with aroma components (Gigl, Hofmann, & Frank, 2021; Hofmann, Czerny, Calligaris, & Schieberle, 2001) provides a further basis for complementary studies involving sensorics.

3.3. *Comprehensive molecular profiles of Arabica versus Robusta coffees*

Beyond the complexity of the chemical processes involved in roasting, the authenticity control of coffee is emerging as a critical concern.

While individual marker compounds indicative of the more prized Arabica or the less delicate Robusta variety are limited in their view and may become less definite due to climate-induced epigenetic factors, new breeds and hybrids, or improved detection limits (Gunning et al., 2018), the challenge for holistic molecular profiles or fingerprints lies in their universal applicability and reproducibility (Medina, Pereira, Silva, Perestrelo, & Camara, 2019; Medina, Perestrelo, Silva, Pereira, & Câmara, 2019). With our diverse collection coffee samples and comprehensive molecular data from FT-ICR-MS measurements (11,500 molecular compositions), we ensure the necessary representativeness of statistically relevant molecular profiles. By organizing the samples into different measurement batches, as well as a statistical training set and a test set, we further guarantee the reproducibility of fingerprints, which is crucial for authenticity verification.

To identify characteristic compounds hidden among the thousands of mass signals, we performed an OPLS-DA of the training set based on the single species labels of the commercially available coffees (Arabica $n = 46$; Robusta $n = 17$) (Fig. 4). Sample 1, officially labeled as Arabica, was removed from the training set as an outlier and subjected to further complementary NMR analysis. The statistical parameters confirm the high significance of the OPLS-DA model. Goodness of the fit R^2Y (0.98) and prediction capability Q^2 (0.91) values close to 1 indicate the high statistical power of the model but may also suggest potential overfitting (Golbraikh & Tropsha, 2002). However, the correct classification of 100 % of the test set samples (39 out of 39) and an additional ANOVA cross-validation ($p < <0.01$) mitigate this concern. One sample of the test set (sample 72), officially labeled as *Coffea arabica,* was classified by our analysis as *Coffea canephora* (Robusta). This apparent misclassification may imply that the sample in question was at least not entirely composed of the higher quality bean. The commonly used marker 16-

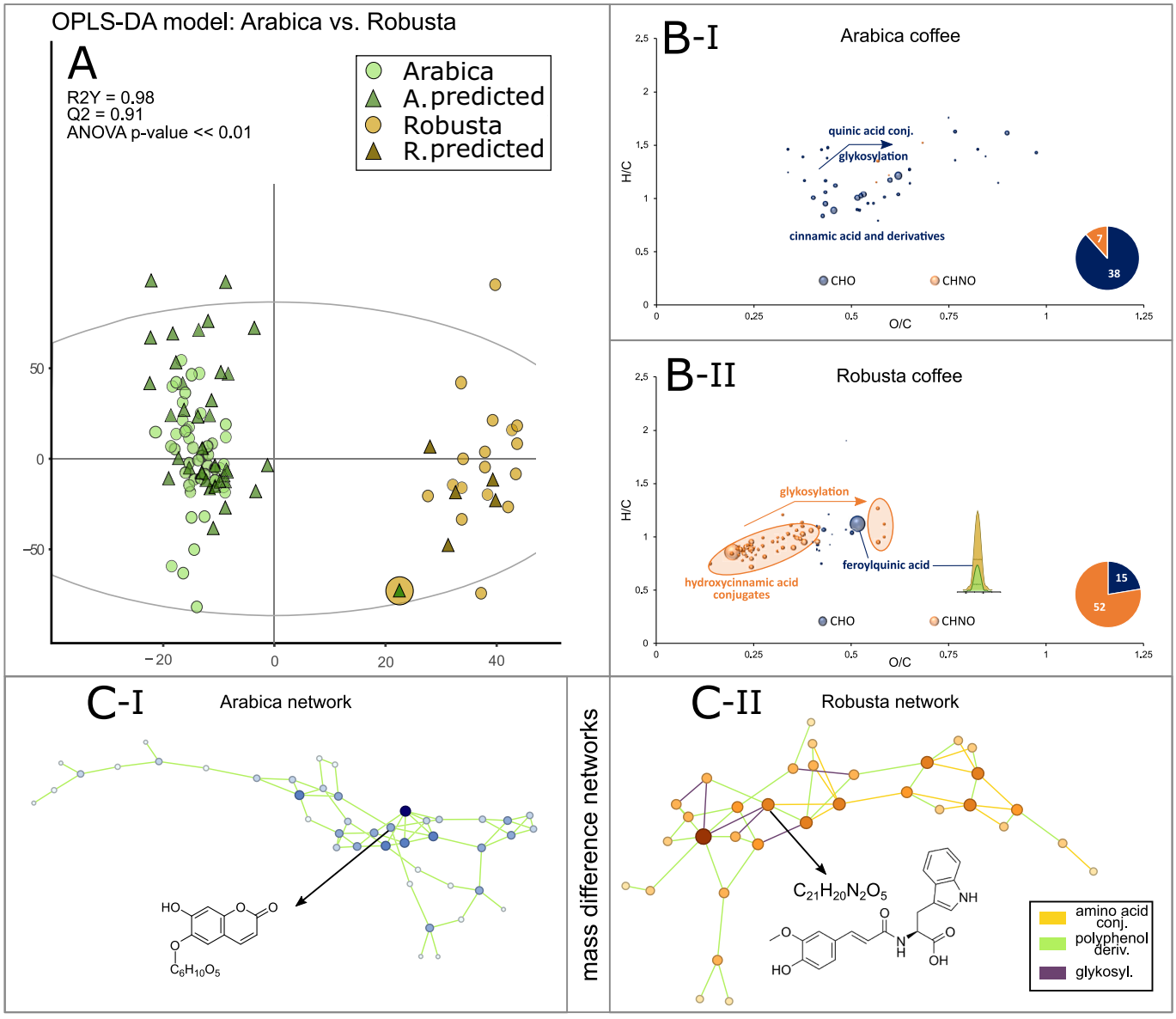


Fig. 4. OPLS-DA model of the single-variety training set ($n = 63$) and test set ($n = 39$) projections differentiating Arabica versus Robusta coffees (A) and characterization of the respective significant compositions in a van Krevelen representation (B) and mass difference networks (C). The score plot is shown in (A), including the training set and the projection of the position of the test set. The score plot prediction of sample 72, labeled as Arabica coffee but ultimately identified as containing Robusta markers, is highlighted. Van Krevelen diagrams of features with significantly high and low loading values (VIP > 3) are shown in B—I and B-II, respectively. Arabica compositions are in the range of hydroxycinnamic acid (HCA) derivatives, with more oxygenated compositions indicating glycosyl or quinic acid conjugation. The intensity value of feruloylquinic acid is shown as the square root. Robusta signals are highly unsaturated and indicate amino acid (tryptophan) conjugates of HCAs. The mass difference network plots (C) show the compositional similarity of the potential markers.

OMC (German standard methods, 2011, 2023) has limited ionizability in ESI^- and requires specific extraction due to its lipophilicity. Consequently, this compound is not included in our FT-MS data. Therefore, we employed a drafted standardized NMR method (German standard methods, 2023), identifying the diterpene signal at 3.16 ppm, recognized as characteristic. The signal predominantly originating from 16-OMC indicates a mixture or hybrid for Sample 72 and confirms the suspected mislabeling of outlier sample 1 (Supplementary Fig. S1).

Despite the questionable nature of two samples, the collective fingerprint used for the statistics highlights signals that significantly distinguish the Arabica and Robusta classes (VIP > 3). The van Krevelen diagram of the 45 highly significant Arabica compositions (Supplementary Table S6) not only suggests coumaroylquinic acid ($C_{16}H_{18}O_8$) (Baeza, Sarria, Bravo, & Mateos, 2016) as a known hydroxycinnamic acid conjugate and mozambioside ($C_{26}H_{36}O_{10}$) (Lang, Klade, Beusch, Dunkel, & Hofmann, 2015) as an Arabica-specific furokaurane glucoside but also reveals numerous other derivatives (Fig. 4B-I). The relatively high level of oxygenation (O/C > 0.5) suggests possible glycosidic or quinic acid-bound derivatives. Through in silico deglucosylation and quinic acid cleavage, we identified potential matches in the FooDB database (WishartLab, 2021) for esculetin glucoside $C_{15}H_{16}O_9$ (Tattini et al., 2014) and the lactone of feruloylcaffeoylquinic acid $C_{26}H_{24}O_{11}$ (Farah, de Paulis, Trugo, & Martin, 2005), but the majority of molecular compositions remain without known structural equivalents in the literature (Baeza et al., 2016; M. N. Clifford, 1985; Farah et al., 2005; Jiang, Chiaro, Maddali, Prabhu, & Peterson, 2009; Upadhyay & Mohan Rao, 2013). The previously unnoticed presence of these compounds may be due to the complex derivatization reactions of polyphenols, similar to the Maillard process, triggered by the heat of roasting (Bork, Proksch, Rohn, & Kanzler, 2024; Gigl, Frank, Irmer, & Hofmann, 2022; Jiang et al., 2009). The mass difference network for Arabica compounds (Fig. 4C-I), constructed on (bio)chemical derivatizations (Supplementary Table S4), reinforces the relatedness of the compounds. Among the dozens of characteristic molecular signals, some are found exclusively in all Arabica samples (Supplementary Fig. S7), underlying the unique chemical signature of Arabica coffee that we have traced.

In the context of coffee authenticity testing, the molecular fingerprint left by *Coffea canephora* (Robusta) metabolites is of considerably higher significant importance. The van Krevelen diagram shows chlorogenic acid metabolite derivatives such as feruloylquinic acid ($C_{15}H_{16}O_9$), which are recognized for their significance in coffee authenticity testing (Garrett et al., 2012). In particular, the van Krevelen visualization highlights highly unsaturated nitrogen-containing molecules (Fig. 4B-II), some of whose molecular compositions are partially known from the literature: Berti, Navarini, Colomban, and Forzato (2020) identified the p-coumaroyl-N-tryptophan conjugate ($C_{20}H_{18}N_2O_4$) as characteristic of Robusta coffees in a study focused on amino acid conjugates. Alonso-Salces, Serra, Reniero, and Heberger (2009) reported finding the ferulic acid equivalent ($C_{21}H_{20}N_2O_5$) exclusively in samples prepared with Robusta beans. They specifically observed a high abundance of these compounds in coffees from the Uganda region, aligning with the pronounced intensities found in our single-origin Uganda samples. In addition to these known tryptophan conjugates of hydroxycinnamic acid derivatives, the mass difference network (Fig. 4C-II), constructed with polyphenol modifications like methylation, methoxylation, hydroxylation, glycosylation, and amino acid conjugation (Supplementary Table S4), points to further derivatives characteristic for Robusta coffee. The significant compositions for both varieties are compiled in Supplementary Table S6 and made available to the research community for reuse and further investigation, including their potential relevance for bioactivity.

3.4. LC-ToF-MS verification of Robusta marker compounds

The cluster of compounds with similar compositions, identified as statistically significant signals in the FT-ICR-MS data, includes the p-coumaroyl- and feruloyl-N-tryptophan conjugates known in literature (Alonso-Salces et al., 2009; Berti et al., 2020). In order to validate our hypothesis on the distinct metabolism of tryptophan conjugates in Arabica and Robusta coffee varieties, potentially uncovering numerous previously unidentified derivatives, we employed tandem mass spectrometry for structural elucidation. Robusta samples with the highest OPLS-DA scores were analyzed by LC-ToF-MS, together with Arabica negative controls and the suspect samples 1 and 72 mentioned above.

By incorporating accurate masses of predicted candidates into the fragmentation inclusion list, MS^2 spectra were obtained for all 14 rediscovered compositions, even at intensities close to the detection limit. Molecular formulas annotated from the FT-ICR-MS data were confirmed using fragmentation tree calculations (Bocker & Duhrkop, 2016) in the SIRIUS platform (Dührkop et al., 2019). The resulting 25 compounds, including up to five isomers, all exhibited the characteristic signals of the fragments ($[C_{19}H_{17}N_2O_2]^-$, $[C_{12}H_9N_2O_3]^-$, $[C_{11}H_8NO_2]^-$, $[C_{10}H_9N_2]^-$, $[C_{10}H_8N]^-$, $[C_8H_6N]^-$, $[C_3H_2NO_3]^-$) and neutral losses ($-CO_2$, $-C_{11}H_{12}N_2O_2$) of tryptophan (Supplementary Fig. S8). The annotation as tryptophan conjugates was confirmed by CSI:FingerID (Duhrkop et al., 2015) and N-acyl-α-amino acid conjugate compound classes in CANOPUS (Djoumbou Feunang et al., 2016). The assumption of tryptophan conjugation with hydroxycinnamic acid derivatives was supported by the level 2 (Schymanski et al., 2014) identification of tryptophan conjugates of coumaric acid, ferulic acid, and chlorogenic acid isomers among the marker compounds (Table 1). The structural features of coumaric acid ($-C_9H_6O_2$, $[C_9H_5O_2]^-$, $[C_8H_7O]^-$, $[C_6H_5O]^-$), ferulic acid ($-C_{10}H_8O_3$, $[C_{10}H_7O_3]^-$, $[C_9H_9O_2]^-$, $[C_8H_7O_2]^-$, $[C_7H_7O_2]^-$), caffeic acid [$-C_9H_6O_3$, $[C_9H_7O_4]^-$, $[C_9H_5O_3]^-$, $[C_8H_7O_2]^-$], and quinic acid ($-C_7H_{10}O_5]^-$, $[C_7H_{11}O_6]^-$, $[C_7H_9O_5]^-$, $[C_7H_7O_4]^-$) were identified through the corresponding neutral losses and diagnostic ions (M. N. Clifford et al., 2006; De Rosso et al., 2018; Jaiswal et al., 2010; Wu et al., 2009) (Supplementary Fig. S8). Further derivatives were characterized at confidence level 3 using the same criteria (Table 1). Moieties not explained by diagnostic ions of tryptophan or hydroxycinnamic acid derivatives remain characterized on the basis of their molecular composition. They indicate heat-induced modifications of native hydroxycinnamic acids and reactions with Maillard intermediates, as demonstrated by Bork et al. in model studies (Bork, Proksch, Rohn and Kanzler, 2024; Bork, Stobernack, Rohn, & Kanzler, 2024; Bork, Proksch, Stobernack, Rohn, & Kanzler, 2024; Bork, Rohn, & Kanzler, 2022). Their accurate masses appear as neutral losses and/or fragment ions in the tandem mass spectra (Supplementary Table S7). All compounds were detected exclusively in the Robusta samples and the two suspect samples (1 and 72) erroneously labeled pure Arabica (Supplementary Fig. S9). The origin of minimal concentrations (1 % of the intensity of Robusta samples) of p-coumaroyl and feruloyl-N-tryptophan conjugates in a negative control remains inconclusive. It could be explained by technologically induced contamination. The compounds might also originate in very small quantities from the metabolism of the *Arabica* variety, as observed with 16-OMC (Gunning et al., 2018). Definitive conclusions could be drawn from genetic analyses that exclude the possibility of a crossed hybrid variety.

4. Conclusion

In our study, we demonstrate the ability of ultra-high resolution mass spectrometry (FT-ICR-MS) to rapidly map the complex molecular landscape of a coffee sample in a less than 10 min measurement, including thousands of metabolites and reaction products. A significant factor in the overall molecular complexity revealed is the roasting chemistry, which manifests itself in nearly 1500 signals across the coffee spectrum. The precise mass accuracy of FT-ICR-MS and the resulting annotation of molecular formulas enabled us to visualize the intrinsic compositional patterns of Maillard products, mapping the reaction network from desoxyosones and amino conjugates to highly unsaturated heteroaromatic compounds. Changes in atomic compositions due to dehydration

Table 1
Identification and characterization of Robusta specific Tryptophan (Trp) conjugates of hydroxycinnamic acids. The caffeoyl (C), feruloyl (F), coumaroyl (Co) and quinic acid (Q) moieties were identified by their specific neutral losses and diagnostic ions. Moieties that cannot be explained by either tryptophan or hydroxycinnamic acids remain characterized as molecular compositions by their fragment ions and/or neutral losses. Potential heat-induced modifications, as observed in model studies by Bork, L. et al., are referenced in a footnote.

Mass *m/z*	*Retention [min]*	Formula	unexpl. Moiety	Trp	C	F	Co	Q
733.2261	4.96	$C_{37}H_{38}N_2O_{14}$	$C_{10}H_{10}O_4$[a]	1	1	0	0	1
719.2112	4.69	$C_{36}H_{36}N_2O_{14}$	$C_9H_8O_4$[b]	1	1	0	0	1
611.2261	5.76	$C_{31}H_{36}N_2O_{11}$	$C_{11}H_{18}O_6$	1	1	0	0	0
539.1681	4.94	$C_{27}H_{28}N_2O_{10}$	–	1	1	0	0	1
539.1682	4.78	$C_{27}H_{28}N_2O_{10}$	–	1	1	0	0	1
515.1837	6.11	$C_{29}H_{28}N_2O_7$	$C_9H_{10}O_2$[c]	1	1	0	0	0
515.1832	5.82	$C_{29}H_{28}N_2O_7$	$C_9H_{10}O_2$[c]	1	1	0	0	0
501.1679	5.82	$C_{28}H_{26}N_2O_7$	$C_8H_8O_2$[d]	1	1	0	0	0
501.1680	5.74	$C_{28}H_{26}N_2O_7$	$C_8H_8O_2$[d]	1	1	0	0	0
501.1678	5.61	$C_{28}H_{26}N_2O_7$	$C_8H_8O_2$[d]	1	1	0	0	0
501.1680	5.56	$C_{28}H_{26}N_2O_7$	$C_8H_8O_2$[d]	1	1	0	0	0
501.1677	5.42	$C_{28}H_{26}N_2O_7$	$C_8H_8O_2$[d]	1	1	0	0	0
501.1676	5.15	$C_{28}H_{26}N_2O_7$	$C_8H_8O_2$[d]	1	1	0	0	0
485.1729	5.75	$C_{28}H_{26}N_2O_6$	$C_8H_8O_2$[d]	1	0	0	1	0
485.1729	6.05	$C_{28}H_{26}N_2O_6$	$C_8H_8O_2$[d]	1	0	0	1	0
475.1532	5.34	$C_{26}H_{24}N_2O_7$	$C_6H_6O_2$[e]	1	1	0	0	0
459.1574	6.19	$C_{26}H_{24}N_2O_6$	C_6H_6O	1	1	0	0	0
429.1468	6.51	$C_{25}H_{22}N_2O_5$	C_5H_4O[f]	1	0	0	1	0
429.1464	6.19	$C_{25}H_{22}N_2O_5$	C_5H_4O[f]	1	0	0	1	0
429.1473	6.05	$C_{25}H_{22}N_2O_5$	C_5H_4O[f]	1	0	0	1	0
403.1308	5.58	$C_{23}H_{20}N_2O_5$	$C_{12}H_8O_3$[g]	1	0	0	0	0
393.1097	5.35	$C_{21}H_{18}N_2O_6$	CO	1	1	0	0	0
379.1309	5.59	$C_{21}H_{20}N_2O_5$	–	1	0	1	0	0
379.1308	5.88	$C_{21}H_{20}N_2O_5$	–	1	0	1	0	0
349.1199	5.51	$C_{20}H_{18}N_2O_4$	–	1	0	0	1	0

[a] $C_{10}H_{10}O_4$ ferulic acid (addition)
[b] $C_9H_8O_4$ caffeic acid (addition)
[c] $C_9H_{10}O_2$ vinylguaiacol (addition)
[d] $C_8H_8O_2$ vinylcatechol (addition)
[e] $C_6H_6O_2$ furaneol (condensation) or methylfurfural (addition)
[f] C_5H_4O furfuryl alcohol (condensation)
[g] $C_{12}H_8O_3$ furan-dimer (condensation).

cascades, cleavage, and addition of fission products reveal that the roasting chemical processes follow a structured scheme that we characterized. The endpoints include not only small molecules but also larger, condensed molecules with up to 38 double-bond equivalents. While key aroma compounds in coffee have been largely deciphered (Sanz, Czerny, Cid, & Schieberle, 2002; Semmelroch & Grosch, 1996), we explore the complexity of the diverse coffee matrix, its numerous precursors, and intermediates. We raise awareness of the thousands of non- and semi-volatile compounds with a potential impact on the overall sensory impression of coffee. This impression is shaped not only by flavor compounds themselves but also by their possible interaction with the coffee matrix (Gigl et al., 2021; Hofmann et al., 2001). Despite the exceptional mass accuracy and resolution of FT-ICR-MS, it remains limited by ion suppression inherent to direct-infusion methods, which yields only semi-quantitative data. Additionally, FT-ICR-MS cannot provide structural information beyond basic atomic compositions. Therefore, further complementary research into the molecular complexity of coffee, which could potentially provide future process guidance through molecular profiles, may prove essential in improving coffee quality and overall sensory experience.

Besides the degradation of chlorogenic acids during roasting, cinnamic acids and their conjugates serve as crucial indicators of coffee varieties, underscoring the significant role of polyphenols in coffee composition. Based on the extensive data of 11,500 resolved compound signals, we extracted molecular signatures indicative of the Arabica and Robusta coffees varieties and highlighted their compositional relationship in a mass difference network. Both the comprehensive molecular data and the representative coffee sample set provided the basis for the reproducible statistical differentiation, across four measurement batches. Comprehensive fingerprints, obtained via direct-infusion mass spectrometry, are considered to be resilient against targeted fraud attempts. When built on a representative sample set and continuously expanded and adjusted (Medina, Pereira, et al., 2019), their reliability is expected to be robust to changes in consumer behaviors in the face of disruptions to vulnerable supply chains and changing climatic conditions. The discovery of two mislabeled coffee samples and traces of Robusta characteristic compounds in others underlines the critical importance of secure reference material in future studies, which may include testing of blends of reference standards for quantitative assessments.

Our Robusta molecular profile contained numerous highly significant molecular compositions, hypothesized to be tryptophan derivatives of hydroxycinnamic acid compounds previously found in coffee (Alonso-Salces et al., 2009; Baeza et al., 2016; Berti et al., 2020; Garrett et al., 2012; Lang et al., 2015; Tattini et al., 2014). The identification of numerous tryptophan conjugates as substances specific to Robusta coffee by LC-ToF tandem mass spectrometry substantiates the hypothesis of different tryptophan conjugation metabolisms of the coffee varieties, manifested in the coffee beverage. On the one hand, the presence and characterization of our proposed compounds by LC-ToF-MS offers the prospect of translating our findings into a targeted method. To address this, further studies must include a complete identification of compounds using NMR, ensure their availability through synthesis or extraction methods, and their quantitative screening in verified authentic samples. This approach will facilitate the establishment of a standardized method with proposed thresholds, ensuring authenticity both in raw and roasted coffee beans. Considering the water solubility of the target analytes compared to the lipophilic nature of 16-methylcafestol, the proposed marker compounds are appropriately suited for application in ready-made products. On the other hand, direct infusion into an ultra-high resolution mass spectrometer is fast, requires no sample preparation beyond dilution of the brewed coffee or coffee

product, and promises insights beyond the limited perspective of potentially falsifiable individual parameters. Standardized methods are also required, to be applied across hundreds of samples with verified authentic provenance. Once the statistical prediction models are established, they can be adapted to metabolic variations (e.g., vintage and climate) using control samples and be incorporated into databases for retrieval. However, initial characterization of the key compound classes remains a crucial step to ensure the analytical results are defensible and robust. The evolution of FT-ICR mass spectrometers towards more user-friendly instruments, the reproducibility and interoperability of data, the trend towards increasingly shorter retention times, and the ever-increasing mass resolving power of more affordable FT-Orbitrap instruments (Desligniere et al., 2024) pave a clear path for the integration of direct-infusion high-resolution mass spectrometry into food authenticity testing.

CRediT authorship contribution statement

Stefan A. Pieczonka: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Anna Dzemajili:** Writing – review & editing, Visualization, Validation, Investigation, Formal analysis, Data curation, Conceptualization. **Silke S. Heinzmann:** Conceptualization, Data curation, Investigation, Methodology, Visualization, Writing – review & editing. **Michael Rychlik:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Philippe Schmitt-Kopplin:** Writing – review & editing, Supervision, Software, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the corresponding author used DeepL (DeepL SE; April 2024) exclusively to improve readability and language of the manuscript. The author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Declaration of competing interest

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

Data availability

The FT-ICR-MS raw data will be made available, without undue reservation. The LC-ToF-MS data set and mgf file were uploaded to the MassIVE (UC San Diego) repository (doi:10.25345/C5JS9HK4K).

Acknowledgements

We gratefully acknowledge the support of the Research Center Weihenstephan for Brewing and Food Quality (TUM) for conducting the particle size measurements of the ground coffee beans and the helpful discussions with the Department of Food Chemistry and Analysis (TU Berlin) concerning heat-induced metabolite modifications.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2024.141432.

References

Ahmad Bhawani, S., Fong, S. S., & Mohamad Ibrahim, M. N. (2015). Spectrophotometric analysis of caffeine. *International Journal of Analytical Chemistry, 2015*, Article 170239. https://doi.org/10.1155/2015/170239

Alonso-Salces, R. M., Serra, F., Reniero, F., & Heberger, K. (2009). Botanical and geographical characterization of green coffee (Coffea arabica and Coffea canephora): Chemometric evaluation of phenolic and methylxanthine contents. *Journal of Agricultural and Food Chemistry, 57*(10), 4224–4235. https://doi.org/10.1021/jf8037117

Alves, R. C., Casal, S., Alves, M. R., & Oliveira, M. B. (2009). Discrimination between arabica and robusta coffee species on the basis of their tocopherol profiles. *Food Chemistry, 114*(1), 295–299. https://doi.org/10.1016/j.foodchem.2008.08.093

Aregay, M. W. (2009). The early history of Ethiopia's coffee trade and the rise of Shawa. *The Journal of African History, 29*(1), 19–25. https://doi.org/10.1017/s0021853700035969

Baeza, G., Sarria, B., Bravo, L., & Mateos, R. (2016). Exhaustive qualitative LC-DAD-MS (n) analysis of Arabica green coffee beans: Cinnamoyl-glycosides and Cinnamoylshikimic acids as new polyphenols in green coffee. *Journal of Agricultural and Food Chemistry, 64*(51), 9663–9674. https://doi.org/10.1021/acs.jafc.6b04022

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *Proceedings of the International AAAI Conference on Web and Social Media, 3*(1), 361–362. https://doi.org/10.1609/icwsm.v3i1.13937

Berti, F., Navarini, L., Colomban, S., & Forzato, C. (2020). Hydroxycinnamoyl amino acids conjugates: A chiral Pool to distinguish commercially exploited Coffea spp. *Molecules, 25*(7). https://doi.org/10.3390/molecules25071704

Bocker, S., & Duhrkop, K. (2016). *Fragmentation trees reloaded. J Cheminform, 8*, 5. https://doi.org/10.1186/s13321-016-0116-8

Bork, L. V., Proksch, N., Rohn, S., & Kanzler, C. (2024). Contribution of Hydroxycinnamic acids to color formation in nonenzymatic Browning reactions with key Maillard reaction intermediates. *Journal of Agricultural and Food Chemistry, 72*(3), 1708–1720. https://doi.org/10.1021/acs.jafc.3c07168

Bork, L. V., Proksch, N., Stobernack, T., Rohn, S., & Kanzler, C. (2024). Influence of Hydroxycinnamic Acids on the Maillard Reaction ofArabinose and Galactose beyond Carbonyl-Trapping. *Journal of Agricultural and Food Chemistry, 72*(28), 15933–15947. https://doi.org/10.1021/acs.jafc.4c02959

Bork, L. V., Stobernack, T., Rohn, S., & Kanzler, C. (2024). Browning reactions of hydroxycinnamic acids and heterocyclic Maillard reaction intermediates – Formation of phenol-containing colorants. *Food Chemistry, 449*(139189), 1–15, 10.1016/j.foodchem.2024.139189.

Bork, L. V., Rohn, S., & Kanzler, C. (2022). Characterization of Colorants Formed by Non-Enzymatic Browning Reactions of Hydroxycinnamic Acid Derivatives. *Molecules, 27*(21), 7564–7582. https://doi.org/10.1021/acs.jafc.4c02959, 1-19.

Buffo, R. A., & Cardelli-Freire, C. (2004). Coffee flavour: An overview. *Flavour and Fragrance Journal, 19*(2), 99–104. https://doi.org/10.1002/ffj.1325

Chambers, M. C., et al. (2012). *Nature Biotechnol., 30*, 918–920. https://doi.org/10.1038/nbt.2377

Clifford, M. (2004). The cinnamoyl-amino acid conjugates of green robusta coffee beans. *Food Chemistry, 87*(3), 457–463. https://doi.org/10.1016/j.foodchem.2003.12.020

Clifford, M. N. (1985). In R. J. Clarke, & R. Macrae (Eds.), *Chlorogenic acids*. Coffee: Springer.

Clifford, M. N., Marks, S., Knight, S., & Kuhnert, N. (2006). Characterization by LC-MS(n) of four new classes of p-coumaric acid-containing diacyl chlorogenic acids in green coffee beans. *Journal of Agricultural and Food Chemistry, 54*(12), 4095–4101. https://doi.org/10.1021/jf060536p

De Rosso, M., Colomban, S., Flamini, R., & Navarini, L. (2018). UHPLC-ESI-QqTOF-MS/MS characterization of minor chlorogenic acids in roasted Coffea arabica from different geographical origin. *Journal of Mass Spectrometry, 53*(9), 763–771. https://doi.org/10.1002/jms.4263

Desligniere, E., Yin, V. C., Ebberink, E., Rolland, A. D., Barendregt, A., Worner, T. P., & Heck, A. J. R. (2024). Ultralong transients enhance sensitivity and resolution in Orbitrap-based single-ion mass spectrometry. *Nature Methods*. https://doi.org/10.1038/s41592-024-02207-8

Dias, R. C., Campanha, F. G., Vieira, L. G., Ferreira, L. P., Pot, D., Marraccini, P., & De Toledo Benassi, M. (2010). Evaluation of kahweol and cafestol in coffee tissues and roasted coffee by a new high-performance liquid chromatography methodology. *Journal of Agricultural and Food Chemistry, 58*(1), 88–93. https://doi.org/10.1021/jf9027427

Djoumbou Feunang, Y., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G., & Wishart, D. S. (2016). ClassyFire: Automated chemical classification with a comprehensive, computable taxonomy. *Journal of Cheminformatics, 8*, 61. https://doi.org/10.1186/s13321-016-0174-y

Dou, Y., Mei, M., Kettunen, T., Makinen, M., & Janis, J. (2022). Chemical fingerprinting of phenolic compounds in Finnish berry wines using Fourier transform ion cyclotron resonance mass spectrometry. *Food Chemistry, 383*, Article 132303. https://doi.org/10.1016/j.foodchem.2022.132303

Duarte, G. S., Pereira, A. A., & Farah, A. (2010). Chlorogenic acids and other relevant compounds in Brazilian coffees processed by semi-dry and wet post-harvesting methods. *Food Chemistry, 118*(3), 851–855. https://doi.org/10.1016/j.foodchem.2009.05.042

Duhrkop, K., Shen, H., Meusel, M., Rousu, J., & Bocker, S. (2015). Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proceedings of the National Academy of Sciences of the United States of America, 112*(41), 12580–12585. https://doi.org/10.1073/pnas.1509788112

Dührkop, K., et al. (2019). *Nature Methods, 16*, 299–302. https://doi.org/10.1038/s41592-019-0344-8

Farah, A., de Paulis, T., Trugo, L. C., & Martin, P. R. (2005). Effect of roasting on the formation of chlorogenic acid lactones in coffee. *Journal of Agricultural and Food Chemistry, 53*(5), 1505–1513. https://doi.org/10.1021/jf048701t

Garrett, R., Vaz, B. G., Hovell, A. M., Eberlin, M. N., & Rezende, C. M. (2012). Arabica and robusta coffees: Identification of major polar compounds and quantification of blends by direct-infusion electrospray ionization-mass spectrometry. *Journal of Agricultural and Food Chemistry, 60*(17), 4253–4258. https://doi.org/10.1021/jf300388m

German standard methods. (2011). *DIN 10779:2011 DE Untersuchung von Kaffee und Kaffee-Erzeugnissen - Bestimmung des Gehaltes an 16-O-Methylcafestol in Röstkaffee - HPLC-Verfahren*.

German standard methods. (2023). *DIN EN 17992:2023–07 (draft) - Determination of the sum of 16-O-Methylcafestol, 16-O-Methylkahweol and their derivatives in roasted coffee by ^{1}H-qNM*.

Gigl, M., Frank, O., Irmer, L., & Hofmann, T. (2022). Identification and quantitation of reaction products from Chlorogenic acid, Caffeic acid, and their thermal degradation products with odor-active thiols in coffee beverages. *Journal of Agricultural and Food Chemistry, 70*(17), 5427–5437. https://doi.org/10.1021/acs.jafc.2c01378

Gigl, M., Hofmann, T., & Frank, O. (2021). NMR-based studies on odorant-Melanoidin interactions in coffee beverages. *Journal of Agricultural and Food Chemistry, 69*(50), 15334–15344. https://doi.org/10.1021/acs.jafc.1c06163

Golbraikh, A., & Tropsha, A. (2002). Beware of q2! *J. Mol. Graphics and Modelling, 20*, 269–276.

Gunning, Y., Defernez, M., Watson, A. D., Beadman, N., Colquhoun, I. J., Le Gall, G., & Kemsley, E. K. (2018). 16-O-methylcafestol is present in ground roast Arabica coffees: Implications for authenticity testing. *Food Chemistry, 248*, 52–60. https://doi.org/10.1016/j.foodchem.2017.12.034

He, Z., Liu, S., Nam, S., Klasson, K. T., & Cheng, H. N. (2023). Molecular level characterization of the effect of roasting on the extractable components of glandless cottonseed by Fourier transform ion cyclotron resonance mass spectrometry. *Food Chemistry, 403*, Article 134404. https://doi.org/10.1016/j.foodchem.2022.134404

Hemmler, D., Roullier-Gall, C., Marshall, J. W., Rychlik, M., Taylor, A. J., & Schmitt-Kopplin, P. (2017). Evolution of complex Maillard chemical reactions. *Resolved in Time. Sci Rep, 7*(1), 3227. https://doi.org/10.1038/s41598-017-03691-z

Hemmler, D., Roullier-Gall, C., Marshall, J. W., Rychlik, M., Taylor, A. J., & Schmitt-Kopplin, P. (2018). Insights into the chemistry of non-enzymatic Browning reactions in different ribose-amino acid model systems. *Scientific Reports, 8*(1), 16879. https://doi.org/10.1038/s41598-018-34335-5

Hofmann, T., Czerny, M., Calligaris, S., & Schieberle, P. (2001). Model studies on the influence of coffee melanoidins on flavor volatiles of coffee beverages. *Journal of Agricultural and Food Chemistry, 49*(5), 2382–2386. https://doi.org/10.1021/jf0012042

Jaiswal, R., Patras, M. A., Eravuchira, P. J., & Kuhnert, N. (2010). Profile and characterization of the chlorogenic acids in green Robusta coffee beans by LC-MS(n): Identification of seven new classes of compounds. *Journal of Agricultural and Food Chemistry, 58*(15), 8722–8737. https://doi.org/10.1021/jf1014457

Jiang, D., Chiaro, C., Maddali, P., Prabhu, K. S., & Peterson, D. G. (2009). Identification of hydroxycinnamic acid-maillard reaction products in low-moisture baking model systems. *Journal of Agricultural and Food Chemistry, 57*(21), 9932–9943. https://doi.org/10.1021/jf900932h

Kanzler, C., Wustrack, F., & Rohn, S. (2021). High-resolution mass spectrometry analysis of Melanoidins and their precursors formed in a model study of the Maillard reaction of methylglyoxal with l-alanine or l-lysine. *Journal of Agricultural and Food Chemistry, 69*(40), 11960–11970. https://doi.org/10.1021/acs.jafc.1c04594

Kim, S., Kramer, R. W., & Hatcher, P. G. (2003). Graphical method for analysis of ultrahigh-resolution broadband mass spectra of natural organic matter, the van Krevelen diagram. *Analytical Chemistry, 75*(20), 5336–5344. https://doi.org/10.1021/ac034415p

Kurzrock, T., & Speer, K. (2007). Diterpenes and Diterpene esters in coffee. *Food Review International, 17*(4).

Lang, R., Klade, S., Beusch, A., Dunkel, A., & Hofmann, T. (2015). Mozambioside is an Arabica-specific bitter-tasting Furokaurane glucoside in coffee beans. *Journal of Agricultural and Food Chemistry, 63*(48), 10492–10499. https://doi.org/10.1021/acs.jafc.5b04847

Maier, H. G., & Wewetzer, H. (1978). Bestimmung von Diterpen-Glykosiden im Bohnenkaffee. *Zeitschrift für Lebensmittel-Untersuchung und -Forschung, 167*(2), 105–107. https://doi.org/10.1007/bf01136137

Medina, S., Pereira, J. A., Silva, P., Perestrelo, R., & Camara, J. S. (2019). Food fingerprints - a valuable tool to monitor food authenticity and safety. *Food Chemistry, 278*, 144–162. https://doi.org/10.1016/j.foodchem.2018.11.046

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

Monakhova, Y. B., Ruge, W., Kuballa, T., Ilse, M., Winkelmann, O., Diehl, B., & Lachenmeier, D. W. (2015). Rapid approach to identify the presence of Arabica and Robusta species in coffee using 1H NMR spectroscopy. *Food Chemistry, 182*, 178–184. https://doi.org/10.1016/j.foodchem.2015.02.132

Murkovic, M., & Bornik, M. A. (2007). Formation of 5-hydroxymethyl-2-furfural (HMF) and 5-hydroxymethyl-2-furoic acid during roasting of coffee. *Molecular Nutrition & Food Research, 51*(4), 390–394. https://doi.org/10.1002/mnfr.200600251

Perez, M., Dominguez-Lopez, I., Lopez-Yerena, A., & Vallverdu Queralt, A. (2023). Current strategies to guarantee the authenticity of coffee. *Critical Reviews in Food Science and Nutrition, 63*(4), 539–554. https://doi.org/10.1080/10408398.2021.1951651

Perez-Miguez, R., Sanchez-Lopez, E., Plaza, M., Castro-Puyana, M., & Marina, M. L. (2018). A non-targeted metabolomic approach based on reversed-phase liquid chromatography-mass spectrometry to evaluate coffee roasting process. *Analytical and Bioanalytical Chemistry, 410*(30), 7859–7870. https://doi.org/10.1007/s00216-018-1405-z

Pieczonka, S. A., Hemmler, D., Moritz, F., Lucio, M., Zarnkow, M., Jacob, F., & Schmitt-Kopplin, P. (2021). Hidden in its color: A molecular-level analysis of the beer's Maillard reaction network. *Food Chemistry, 361*, Article 130112. https://doi.org/10.1016/j.foodchem.2021.130112

Pieczonka, S. A., Lucio, M., Rychlik, M., & Schmitt-Kopplin, P. (2020). Decomposing the molecular complexity of brewing. *npj Science of Food, 4*(11), 1–10. https://doi.org/10.1038/s41538-020-00070-3

Pieczonka, S. A., Zarnkow, M., Ampenberger, F., Gastl, M., Rychlik, M., & Schmitt-Kopplin, P. (2023). FT-ICR-MS reveals the molecular imprints of the brewing process. *Frontiers in Nutrition, 10*(1243503). https://doi.org/10.3389/fnut.2023.1243503

Poisson, L., Blank, I., Dunkel, A., & Hofmann, T. (2017). *The chemistry of roasting—Decoding flavor formation. In the craft and science of coffee (pp. 273-309)*.

Roullier-Gall, C., Boutegrabet, L., Gougeon, R. D., & Schmitt-Kopplin, P. (2014). A grape wine chemodiversity comparison of different appellations in burgundy: Vintage vs terroir effects. *Food Chemistry, 152*, 100–107.

Sanz, C., Czerny, M., Cid, C., & Schieberle, P. (2002). Comparison of potent odorants in a filtered coffee brew and in an instant coffee beverage by aroma extract dilution analysis (AEDA). *European Food Research and Technology, 214*(4), 299–302. https://doi.org/10.1007/s00217-001-0459-9

Schievano, E., Finotello, C., De Angelis, E., Mammi, S., & Navarini, L. (2014). Rapid authentication of coffee blends and quantification of 16-O-methylcafestol in roasted coffee beans by nuclear magnetic resonance. *Journal of Agricultural and Food Chemistry, 62*(51), 12309–12314. https://doi.org/10.1021/jf505013d

Schmid, R., Heuckeroth, S., Korf, A., Smirnov, A., Myers, O., Dyrlund, T. S., & Pluskal, T. (2023). Integrative analysis of multimodal mass spectrometry data in MZmine 3. *Nature Biotechnology, 41*(4), 447–449. https://doi.org/10.1038/s41587-023-01690-2

Schmitt-Kopplin, P., Hemmler, D., Moritz, F., Gougeon, R. D., Lucio, M., Meringer, M., & Hertkorn, N. (2019). Systems chemical analytics: Introduction to the challenges of chemical complexity analysis. *Faraday Discussions, 218*(0), 9–28. https://doi.org/10.1039/c9fd00078j

Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H. P., & Hollender, J. (2014). Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environmental Science & Technology, 48*(4), 2097–2098. https://doi.org/10.1021/es5002105

Semmelroch, P., & Grosch, W. (1996). Studies on character impact odorants of coffee brews. *Journal of Agricultural and Food Chemistry, 44*(2), 537–543. https://doi.org/10.1021/jf9505988

Souto, U. T. C. P., Pontes, M. J. C., Silva, E. C., Galvão, R. K. H., Araújo, M. C. U., Sanches, F. A. C., & Oliveira, M. S. R. (2010). UV–vis spectrometric classification of coffees by SPA–LDA. *Food Chemistry, 119*(1), 368–371. https://doi.org/10.1016/j.foodchem.2009.05.078

Speer, K., & Mischnick, P. (1989). 16-O-methylcafestol - a new diterpene in coffee - discovery and identification. *Zeitschrift Fuer Lebensmitteluntersuchung- und Forschung, 189*(3), 219–222.

Sunarharum, W. B., Williams, D. J., & Smyth, H. E. (2014). Complexity of coffee flavor: A compositional and sensory perspective. *Food Research International, 62*, 315–325. https://doi.org/10.1016/j.foodres.2014.02.030

Tattini, M., Di Ferdinando, M., Brunetti, C., Goti, A., Pollastri, S., Bellasio, C., & Agati, G. (2014). Esculetin and esculin (esculetin 6-O-glucoside) occur as inclusions and are differentially distributed in the vacuole of palisade cells in Fraxinus ornus leaves: A fluorescence microscopy analysis. *Journal of Photochemistry and Photobiology. B, 140*, 28–35. https://doi.org/10.1016/j.jphotobiol.2014.06.012

Tziotis, D., Hertkorn, N., & Schmitt-Kopplin, P. (2011). Kendrick-analogous network visualization of ion cyclotron resonance Fourier transform mass spectra: Improved options for the assignment of elemental compositions and the classification of organic molecular complexity. *European Journal of Mass Spectrometry, 17*, 415–421. https://doi.org/10.1255/ejms.1135

United States Department of Agriculture. (2023). *Coffee: World markets and trade* (Foreign Agricultural Service).

Upadhyay, R., & Mohan Rao, L. J. (2013). An outlook on chlorogenic acids-occurrence, chemistry, technology, and biological activities. *Critical Reviews in Food Science and Nutrition, 53*(9), 968–984. https://doi.org/10.1080/10408398.2011.576319

WishartLab. (2021). *FooDB*. Retrieved from: www.foodb.ca Accessed April 2024.

Wu, Z. J., Ma, X. L., Fang, D. M., Qi, H. Y., Ren, W. J., & Zhang, G. L. (2009). Analysis of caffeic acid derivatives from Osmanthus yunnanensis using electrospray ionization quadrupole time-of-flight mass spectrometry. *Eur J Mass Spectrom (Chichester), 15*(3), 415–429. https://doi.org/10.1255/ejms.992

Yu, X., Zhao, M., Hu, J., Zeng, S., & Bai, X. (2012). Correspondence analysis of antioxidant activity and UV–vis absorbance of Maillard reaction products as related to reactants. *LWT - Food Science and Technology, 46*(1), 1–9. https://doi.org/10.1016/j.lwt.2011.11.010

Zhu, J., Zhou, L., Zhao, M., Wei, F., Fu, H., & Marchioni, E. (2023). Revealing the dynamic changes of lipids in coffee beans during roasting based on UHPLC-QE-HR-AM/MS/MS. *Food Research International, 174*(Pt 1), Article 113507. https://doi.org/10.1016/j.foodres.2023.113507