Microchemical Journal 196 (2024) 109578

Contents lists available at ScienceDirect

Microchemical Journal

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

Investigating sensory-classified roasted arabica coffee with GC × GC-TOFMS and chemometrics to understand potato taste defect

Caitlin N. Cain [a], Meriem Gaida [b], Pierre-Hugues Stefanuto [b], Jean-François Focant [b], Robert E. Synovec [a], Susan C. Jackels [c], Kristen J. Skogerboe [c,*]

[a] Department of Chemistry, University of Washington, Box 351700, Seattle, WA 98195, USA
[b] Organic and Biological Analytical Chemistry Group, MolSys Research Unit, University of Liège, Allée du Six Août, 11, B6c, Liège 4000, Belgium
[c] Department of Chemistry, Seattle University, 901 12th Avenue, Seattle, WA 98122, USA

ARTICLE INFO

Keywords:
Coffee
Potato taste defect
Comprehensive two-dimensional gas chromatography-mass spectrometry
Chemometrics
Fisher ratio analysis
Principal components analysis
Partial least squares regression

ABSTRACT

The presence of flavor defects in coffee beans can negatively impact quality, the consumer experience, and commercial trade. Potato taste defect (PTD), a flavor defect specific to East African coffee, is often characterized by a musty, vegetable-like aroma. While previous work has correlated PTD with the presence of 2-isopropyl-3-methoxypyrazine (IPMP), additional changes in the volatile profile of these beans can further amplify the distinct odor of this defect. The aim of this work was to develop a volatile fingerprint of PTD in roasted arabica coffee using headspace solid-phase microextraction coupled to comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry (HS-SPME-GC × GC-TOFMS) and chemometrics. Examination of the HS-SPME-GC × GC-TOFMS data with tile-based Fisher ratio (F-ratio) analysis discovered 359 analytes that differentiated clean coffee samples from those impacted by severe PTD (p-value < 0.01). It was determined that 327 of the identified analytes were more prevalent in the clean coffee samples while 32 analytes, including IPMP, exhibited higher signals in the impacted coffee samples. Principal components analysis (PCA) of the F-ratio results demonstrated that the coffee samples clustered based on the presence of PTD. Partial least squares (PLS) regression modeling further demonstrated that the compounds discovered by F-ratio analysis were correlated with PTD by accurately predicting the concentration of IPMP in the samples. Investigation of the compounds highly weighted in both the PCA and PLS loadings suggest that the presence of microorganisms on coffee beans after antestia bug damage could be a potential pathway for PTD. This damage results in an overall decrease of analytes that are known to have positive sensory contributions to coffee aroma. Collectively, the volatile fingerprint shown herein illustrates that PTD alters the biochemical process in coffee beans.

1. Introduction

Since coffee quality is directly linked to taste and aroma, the presence of product defects due to agricultural practices and/or an adverse agricultural environment can cause off-flavors and devalue the crop [1–3]. Potato taste defect (PTD) is a sporadic defect that occurs in coffee beans grown in the African Great Lakes region, namely those cultivated in Burundi, Rwanda, and Uganda [4,5]. As the name implies, this defect is characterized by the distinct musty, potato aroma of affected coffee beans. The occurrence of this defect has been linked to the presence of *Antestiopsis orbitalis*, an antestia bug native to this region that feeds on the coffee plant [5–7]. Chemically, studies of both green and roasted coffee beans have discovered that the presence and strength of PTD is correlated to 2-isopropyl-3-methoxypyrazine (IPMP) [4,8–11]. However, the mechanism linking coffee plant damage from the antestia bug to the presence of IPMP in volatile headspace of affected beans remains unclear, especially since PTD can be present in coffee beans with and without visual insect damage [9]. It has been hypothesized that damage from the bug either provides favorable growth conditions for microorganisms that can produce IPMP [12–14] or initiates the conversion of hydroxypyrazines naturally produced by the plant into methoxypyrazines like IPMP via *O*-methyltransferase expression [15].

While IPMP has been chemically linked to PTD [4,8–11], it is important to note that the chemical composition of coffee is highly complex with numerous other volatiles contributing to aroma [16]. The presence of these other analytes could, in turn, either heighten or mask

* Corresponding author.
E-mail address: skogerbo@seattleu.edu (K.J. Skogerboe).

https://doi.org/10.1016/j.microc.2023.109578
Received 25 May 2023; Received in revised form 9 October 2023; Accepted 27 October 2023
Available online 30 October 2023

the odor of PTD. Volatile fingerprinting of coffee, especially of beans affected by PTD, has primarily been performed using one-dimensional (1D) gas chromatography-quadrupole mass spectrometry (GC–MS) [1–4,8–11]. For example, a recent study demonstrated that two key volatiles in roasted coffee aroma, 2-ethyl-3,5-dimethylpyrazine and 2-furfurylthiol, can mask the odor attributed to IPMP [11]. Other previous work applying headspace solid-phase microextraction GC–MS (HS-SPME-GC–MS) identified 22 analytes, including IPMP, whose signals differentiated samples that did not have a detectable off-odor ("clean" samples as scored by a sensory panel) from those affected by PTD [10]. Compounds with a higher abundance in the clean coffee samples generally had desirable aromas, whereas compounds linked to undesirable aromas had larger signals in samples affected by PTD [10]. The results of these studies highlight the potential for PTD to affect the concentration of other volatiles present in coffee.

However, 1D separation methods are inherently limited in their peak capacity, which constrains the number of analytes that can be resolved in a reasonable analysis time [17]. As a result, the discovery of volatiles impacted by the presence of PTD will also be limited due to the use of GC–MS. Fortunately, use of comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry (GC × GC-TOFMS) can readily increase the peak capacity of a 1D-GC separation. A GC × GC separation connects two complementary columns via a modulator, which continuously collects fractions of effluent from the first dimension (^{1}D) separation and reinjects those fractions on the second dimension (^{2}D) [18]. Due to this modulation process, the peak capacity of an ideal GC × GC separation is approximately 10-fold higher than its 1D-GC counterpart [19] and the sensitivity of the GC × GC separation is also enhanced [20]. The advantages of this separation platform has been illustrated in various food analysis studies [21,22], including those aimed at profiling coffee aroma [23–28].

While the increased resolving power provided by GC × GC-TOFMS is beneficial for chemical fingerprinting studies, manual identification, and signal integration of every peak present in a chromatogram can be burdensome due to the size and complexity of the data. Fortunately, the data set produced by GC × GC-TOFMS analysis enables the use of non-targeted chemometrics, which can identify statistically significant chemical signals and elucidate relationships among different samples in an automated fashion. For example, GC × GC-TOFMS and non-targeted chemometrics have been coupled to profile commercial espresso capsules [26] and differentiate decaffeinated coffee from its regular coffee counterpart [27]. Fisher ratio (F-ratio) analysis has been one of the prominent methods utilized in food analysis studies for the discover of class-distinguishing analytes [27,29–33]. Using *a priori* knowledge of the sample classes, the ratio of the between-class variance to the pooled within-class variance (a F-ratio) can be calculated for every data point [34], peak listed in a table [35], or a tile (binned data) in a chromatographic data set [36,37]. The output is a list of retention times for peaks ranked in descending order of their F-ratio, referred to as a "hit list". A peak with a large F-ratio is likely class-distinguishing since there is a large variance between classes relative to the within-class variance.

Implementation of F-ratio analysis should ensure that the discovery of chemically relevant differences is not hindered by retention time misalignment and spurious detector noise. The use of tiling (i.e., "smart binning") can prevent the discovery of these instrumental artifacts by dividing the chromatogram into small, rectangular tile sections which capture the entire peak signal along with any retention time shifting. Furthermore, tile-based non-targeted methods can also improve the discovery of low-level peaks due to enhancements in the relative signal-to-noise (*S/N*) while also automatically removing redundant hits [36,37]. Given these advantages, a tile-based algorithmic platform has been recently commercialized and adapted for various experimental designs [38–41]. Herein, the capabilities of tile-based F-ratio analysis are illustrated on a HS-SPME-GC × GC-TOFMS data set of PTD in roasted arabica coffee. Olfactory analysis categorized these samples into four classes based on odor severity: clean, mild PTD, medium PTD, and strong PTD [10]. Class-distinguishing analytes discovered by tile-based F-ratio analysis were quantified using a pure mass channel (*m/z*) identified for the analyte. With GC × GC, a nearly a 10-fold larger peak capacity is anticipated [19] and enhanced *S/N* due to the use of thermal modulation [20]; thus, giving rise to the hypothesis that a significantly larger number of compounds will be discovered relative to our previous 1D-GC–MS study [10]. The consequences of this improved discovery of relevant sensory compounds will be further discussed. Principal components analysis (PCA) and partial least squares (PLS) regression are also performed to illustrate how the analytes discovered by tile-based F-ratio analysis contribute to the biochemical understanding PTD in coffee beans from the African Great Lakes region.

2. Methods and materials

2.1. Acquisition and assessment of coffee samples

Arabica coffee samples from Burundi, Rwanda, and Uganda were sourced, roasted, and initially assessed by Counter Culture Coffee (Durham, NC, USA). The green coffee beans were sourced according to standard procedures, ensuring similarities in their screen size and moisture content and lack of primary and secondary defects [42]. Details of the roasting conditions and olfactory evaluation of the ground coffee samples were previously reported [10,42]. Briefly, the green coffee beans were roasted to a light-medium roast, resulting in a weight loss of 11–13 % and score of 77–80 on the Agtron color scale, and then ground. These coffee samples (56 total) were then classified by a sensory panel into one of four groups (clean, mild PTD, medium PTD, and strong PTD) based on a previously established protocol [42]. This protocol involved grinding 10 g portions of whole roasted coffee beans into a clean cup, carefully smelling the grounds, and categorizing them based on their odor intensity. Suspected PTD samples were then confirmed via cupping protocols published by the Specialty Coffee Association [43]. After olfactory analysis, a total of 14 clean samples, 11 mild PTD samples, 13 medium PTD samples, and 18 strong PTD samples were forwarded for analytical characterization (Table S1). The concentration of IPMP ranged from 0 ng/g – 3.1 ng/g for the clean samples, 1.6 ng/g – 72.4 ng/g for the mild PTD samples, 4.9 ng/g – 79.8 ng/g for the medium PTD samples, and 4.1 ng/g – 529.9 ng/g for the strong PTD samples. All samples were stored in airtight glass containers prior to sample preparation, extraction, and GC × GC-TOFMS analysis.

2.2. Sample preparation and extraction

All analytical standards were procured from Fisher Scientific or Sigma-Aldrich (USA). Preparation of the coffee samples for this study was similar to our previous methodology [10]. Ground roasted coffee (0.1 g) were added to a 20 mL SPME vial along with 1.5 g of sodium chloride, 4.5 mL of ethylenediaminetetraacetic acid (EDTA; 0.111 M), 480 μL of methanol, and 20 μL of the internal standard, deuterated-2-isobutyl-3-methoxypyrazine (d_3-IBMP; 5 ppm). A divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS; fiber thickness: 50/30 μm; Restek, Bellefonte, PA, USA) SPME fiber was utilized for the extraction of the headspace volatiles. This SPME fiber was selected due to its popularity and effectiveness in coffee studies [2,3,8,10,23,26,27,44], where it has been shown to extract a wide range of molecular weights and compound classes. Initial conditioning of a new DVB/CAR/PDMS fiber followed the manufacturer's guidelines, where it was held in a 270 °C GC inlet for 1 hr. Prior to each sample extraction, the coffee samples were incubated at 60 °C for 15 min while being agitated at 250 rpm with on/off times of 5 s and 2 s, respectively. The headspace of the coffee samples was extracted for 30 min at 60 °C. Slightly different from previous work [10], this extraction temperature and time was found to provide an optimal balance between amplifying the chromatographic signal of IPMP while detecting other coffee volatiles. The SPME fiber with the extracted volatiles was desorbed splitless

in the GC inlet for 5 min at 250 °C. Between sample extractions and chromatographic runs, the fiber was re-conditioned at 270 °C for 10 min. Note, chromatographic blanks were periodically collected to ensure the SPME fiber was properly cleaned in between sample extractions. These chromatographic blanks did not show the presence of peaks and/or contaminants that could affect the results of this study. The sample extraction method utilized the L-PAL3 autosampler (LECO, St. Joseph, MI, USA).

2.3. Chromatographic conditions

Separations of each coffee sample were collected in duplicate using the LECO Pegasus BT 4D GC × GC-TOFMS equipped with an Agilent 7890 GC (Agilent Technologies, Palo Alto, CA, USA) and a stock quad-jet thermal modulator. To prevent contamination and carryover from other experiments performed in the laboratory, a new set of conditioned GC columns was installed in the GC × GC-TOFMS, along with a new inlet liner. Splitless sample injections were separated on a polar Rtx-Wax ^{1}D column (30 m × 0.25 mm × 0.25 μm; Restek), and a non-polar Rxi-1MS ^{2}D column (1.7 m × 0.18 mm × 0.18 μm; Restek). A Rtx-Wax column was used in the ^{1}D to maintain a similar primary separation as in our previous 1D-GC–MS study, which also used this column type [10]. Furthermore, previous GC × GC studies on coffee volatiles have also used this GC × GC column configuration [23,27]. The ^{1}D column was held at 40 °C for 5.5 min before ramping to 240 °C at 5 °C/min, where it was held for 5 min. The same temperature program was used for the ^{2}D oven and modulator with an offset of +5 °C and +15 °C, respectively. The carrier gas, ultra-high purity helium (Grade 5, 99.999 %, Praxair, Seattle, WA, USA), operated at a constant flow rate of 1 mL/min. The ^{1}D effluent was reinjected on the ^{2}D column at a modulation period of 2 s. The ion source and transfer line temperatures were set to 225 °C and 285 °C, respectively. The TOFMS collected *m/z* 45–350 at 100 Hz with an electron ionization energy of 70 eV after a 10 s acquisition delay.

2.4. Data analysis

Following data acquisition, the chromatograms were imported into Matlab 2019b (Mathworks, Inc., Natick, MA, USA) for further analysis. The chromatograms were baseline corrected and normalized to the peak area of the internal standard, d_3-IBMP, at *m/z* 127. F-ratio analysis was performed by comparing the clean and strong PTD coffee samples. Note, since the number of samples in the clean and strong PTD class was unbalanced, 14 strong PTD coffee samples were arbitrarily selected for the F-ratio comparison. Hence, with 14 coffee samples per-class and two replicates collected per-sample, a total of 28 chromatograms per-class were compared. A tile size of 10 s × 200 ms (^{1}D × ^{2}D) and cluster window size of 6 s × 120 ms was selected based on the peak widths and degree of retention time shifting observed in the chromatograms. The verification that these tile bin parameters were appropriate for these samples is provided in the next section. Using the four-grid schemes, F-ratios were calculated on every tile per-*m/z* that had a *S/N* greater than 10. The resulting hits were then ranked according to their top F-ratio *m/z*. Any remaining redundant hits and artifacts from the SPME fiber or column bleed were removed from the final hit list. As a result of this methodology, the final hit list contained both class-distinguishing (i.e., true positives) and non-class-distinguishing (i.e., false positives) analytes.

To identify the class-distinguishing analytes in the hit list (i.e., the true positives), a *t*-test (assuming unequal variances) was calculated for every analyte discovered. A *p*-value was calculated using the signal encapsulated in the tile surrounding each hit at the top F-ratio *m/z*. Hits with a *p*-value < 0.01 from a *t*-test were considered as true positives, which were subjected to later identification and quantitation efforts, while hits with a *p*-value > 0.01 were labeled as false positives [33]. Next, for those analytes determined to be true positives, tentative compound identifications were determined by matching the acquired mass spectrum to the NIST 11 library (National Institute of Standards and Technology, Gaithersburg, MD, USA). A match value (MV) ≥ 800 was required for (tentative) identification [45]. Multivariate curve resolution-alternating least squares (MCR-ALS) was used to improve identification efforts for 283 analytes by resolving the hit mass spectrum from the background noise and interferences [46]. Hits that still could not be identified with a MV ≥ 800 after applying MCR-ALS are labeled as "unknown" and numbered. Lastly, for those true positive analytes, accurate concentration ratios between the clean and strong PTD samples, referred to as [Strong]/[Clean] herein, were determined using a pure *m/z* for each hit. Pure analyte *m/z* were discovered using a recent extension to the tile-based software known as the signal ratio (S-ratio) algorithm [47]. Two signal consistency metrics are implemented to discover pure *m/z* for the target analyte using two metrics, a lack-of-fit (*LOF*) and *p*-value from a *t*-test (assuming unequal variances). Ideally, a pure analyte *m/z* should have a significant difference in signals between two classes (i.e., a small *p*-value), but the peak shapes between classes should be similar (i.e., no significant *LOF* is identified). Therefore, the S-ratio calculated for a sufficiently pure analyte *m/z* accurately approximates the true concentration ratio [47]. Meanwhile, *m/z* from an unresolved analyte(s) with little chemical variation between classes will have a high *p*-value, low *LOF*, and S-ratio equal to 1. Using these metrics, the purest *m/z* for quantitation had the lowest *LOF* (maximum *LOF* tolerated = 20 %) and *p*-value (maximum *p*-value tolerated = 1×10^{-4}). Every class-distinguishing analyte had at least one *m/z* that met the *LOF* and *p*-value thresholds to be labeled as sufficiently pure for quantitation. For improved visualization of solely the analytes discovered by F-ratio analysis, "stitch" GC × GC chromatograms were constructed. The principles of this visualization technique have been described previously [48,49]. Briefly, this method extracts the chromatographic signal within a 10 s × 200 ms tile at a pure analyte *m/z* for every class-distinguishing analyte. A *S/N* filter of 10 was implemented to remove noise from the extracted tiles. The tiles were then inserted back into an empty "chromatogram" at their original retention time locations.

Both PCA and PLS models were developed using these "stitch" chromatograms with PLS Toolbox 8.9 (Eigenvector Research, Manson, WA, USA). The signals from each sample, quantified at the pure analyte *m/z*, were mean-centered and inputted into PCA to demonstrate the differences between PTD odor classes after F-ratio analysis. The hits discovered by F-ratio analysis were also utilized to build a predictive model of IPMP concentration via PLS regression. For PLS modeling, the data was divided into a calibration and validation data set using the Kennard-Stone algorithm. A calibration model using 42 coffee samples was built using the mean-centered signal data and auto-scaled values for IPMP concentration. Venetian-blinds cross-validation with 6-splits was performed to determine the appropriate number of latent variables (LVs) for the model and calculate a normalized root-mean-square error of cross-validation (NRMSECV). The validation data set was then input into the PLS model to determine the NRMSE of prediction (NRMSEP).

3. Results and discussion

The total ion current (TIC) chromatograms of a clean (A) and strong PTD (B) sample are shown in Fig. 1, illustrating the complexity of the volatile headspace and need for a GC × GC separation to discover analytes indicative of PTD. A peak detection algorithm identified approximately 500 peaks present in these TIC GC × GC chromatograms, excluding artifacts such as streaks. The total peak capacity for these GC × GC separations was 3750 based on an average ^{1}D and ^{2}D width-at-base (W_b) of 8 s and 200 ms, respectively. For reference, the 1D-GC TIC chromatograms for these coffee samples had a peak capacity of 390, based on an average W_b of 10 s, and only identified ~100 peaks present [10]. The ~5-fold increase in the number of peaks resolved can be attributed to the increased peak capacity and *S/N* provided by a GC × GC separation. Visual comparison of the chromatograms shown in Fig. 1 demonstrates that the strong PTD sample (B) has an overall reduction in

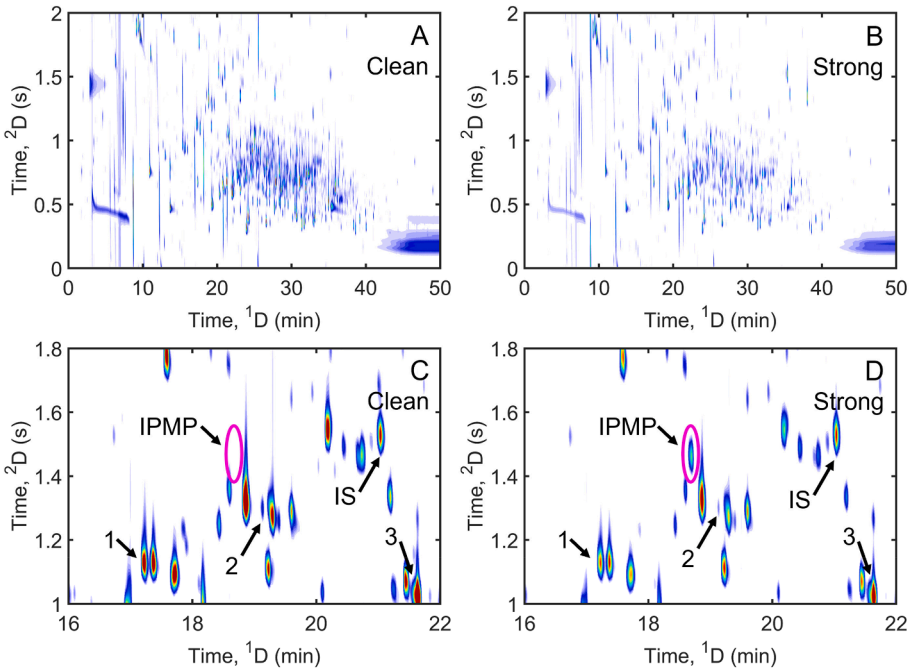

Fig. 1. Normalized TIC GC × GC chromatograms of coffee samples categorized as clean (A) or strong PTD (B) based on their odor. Both chromatograms are plotted on the same color scale. (C-D) A zoom-in on the chromatograms from 16 to 22 min on ^{1}D and 1–1.8 s on ^{2}D. IPMP and the internal standard (IS) are labeled along with three representative analytes: (1) 2-ethyl-6-methylpyrazine, (2) 2,3-diethylpyrazine, and (3) linalool. The pink oval outlines the peak location of IPMP for improved visualization of the absence of IPMP in the clean coffee sample. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

signal compared to the intensity of the peaks present in the clean sample (A). The scale-expanded chromatograms between 16 and 22 min on the ^{1}D and 1–1.8 s on the ^{2}D (Fig. 1C and D) further highlight the differences in these samples. Fig. 1C and D denotes the locations of three representative analytes known to contribute to the aroma of coffee: (1) 2-ethyl-6-methylpyrazine, (2) 2,3-diethylpyrazine, and (3) linalool. Note, the peaks corresponding to IPMP and the internal standard (IS), d_3-IBMP, are also labeled in Fig. 1C and D. Close examination of this region highlights not only the complexity of these chromatograms, but also noticeable concentration differences between the clean and strong PTD coffee samples. For example, alkylpyrazines like 2-ethyl-6-methylpyrazine (labeled as analyte 1) and 2,3-diethylpyrazine (labeled as analyte 2) are known to contribute a roasted, nutty aroma in the coffee headspace [44,50]. Linalool (labeled as analyte 3) has also been identified as one of the few alcohols pertinent to the odor of arabica coffee [3], providing a floral and fruity aroma in roasted coffee [50]. However, Fig. 1C and D shows that the signal for these analytes are ~1.5-fold to ~50-fold lower in the strong PTD sample compared to the clean coffee sample. Meanwhile, Fig. 1C and D demonstrates that the signal for IPMP (outlined by the pink oval) becomes present in the strong PTD coffee sample, which is consistent with previous studies [9,10]. The chromatographic complexity highlighted in Fig. 1 illustrates the utility in applying non-targeted chemometric methods to develop a "comprehensive" volatile fingerprint of PTD in coffee.

To ensure consistency between the work presented herein and our previous 1D-GC–MS study [10], the normalized signal for IPMP in the GC × GC-TOFMS chromatograms (Fig. 1) collected for every coffee sample was quantified using *m/z* 152, the molecular ion for IPMP. The signals for the two replicates were then averaged together and plotted against the IPMP concentration determined with 1D-GC–MS (Fig. 2A) [10]. The data points in Fig. 2A are color coded according to their sensory classification: clean (red), mild PTD (yellow), medium PTD

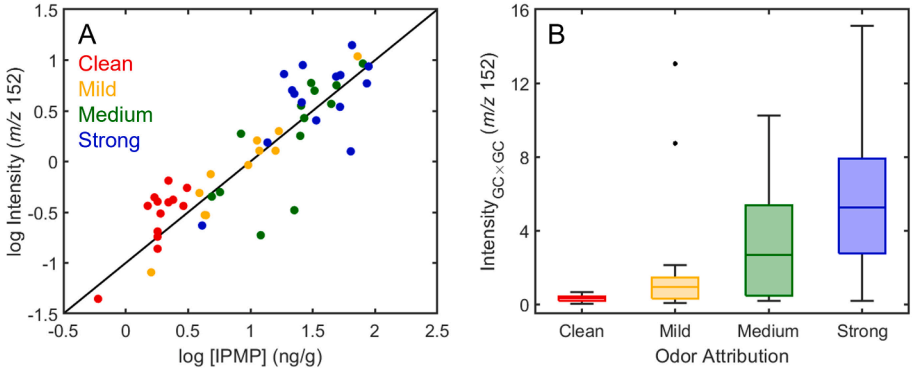

Fig. 2. (A) Relationship between the average peak intensity of IPMP measured herein using GC × GC-TOFMS and the IPMP concentration determined in a previous 1D-GC–MS study [10]. Samples are color coded based on their sensory classification: clean (red), mild PTD (yellow), medium PTD (green), and strong PTD (blue). (B) Box-and-whiskers plot relating the peak intensity of IPMP measured herein to the different PTD odor attributions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(green), and strong PTD (blue). Despite a few outlier samples, the GC × GC-TOFMS and 1D-GC–MS measurements for IPMP are in good agreement. Fig. 2B relates the GC × GC-TOFMS intensity of IPMP to the different sensory classifications of PTD. With a p-value < 0.01, a one-way analysis of variance (ANOVA) statistical test demonstrates that the intensity of IPMP is statistically significantly different between the four odor attributions. Again, these results are consistent with those previously obtained [10]. Based on the results in Fig. 2, a supervised chemometric comparison of the GC × GC-TOFMS chromatograms obtained from the clean and strong PTD coffee samples is the most promising approach for identifying analytes that contribute to the volatile fingerprint of PTD. Note, PCA was first applied to determine if the samples naturally cluster based on the presence or severity of PTD; however, no distinct clustering could be observed using the first few principal components (Fig. S2). Therefore, tile-based F-ratio analysis of the clean and strong PTD samples (the two extremes of the sensory panel classification) was selected to elucidate the PTD-related differences in the headspace of roasted coffee.

For tile-based F-ratio analysis, selection of the appropriate ^{1}D and ^{2}D tile dimensions is crucial to mitigate the discovery of false positives (i.e., non-class-distinguishing analytes) due to the variance produced from retention time shifting [51]. Ideally, the tile size should be large enough to encompass both the typical width of a peak and the degree of retention time shifting present between samples. Fig. 3A and B shows the summed ^{1}D and ^{2}D peak profiles of IPMP in the clean (red) and strong PTD (blue) classes. Minor run-to-run shifting of 1 modulation (2 s) is observed as irregular peak shape in the ^{1}D peak profiles while the ^{2}D profiles show no significant signs of shifting. Therefore, as illustrated by the vertical dashed lines in Fig. 3A and B, a tile size of 10 s × 200 ms (^{1}D × ^{2}D) was selected. Using this tile size, F-ratio analysis was performed by comparing the clean and strong PTD chromatograms. Fig. 3C shows the F-ratio distribution for the 495 hits initially discovered, with F-ratios ranging from 65.2 to 0.01. As indicated by the arrow, IPMP was discovered near the top of the hit list (hit #2) with an F-ratio of 54.3. Since an F-ratio was calculated for every tile possessing a summed signal greater than the S/N of 10 threshold, the initial hit list encompasses both class-distinguishing true positive and false positive hits. Typically, hits discovered near the top of the hit list (larger F-ratios) are more likely to be class-distinguishing, with a larger between-class variance relative to the within-class variance, while hits at the bottom of the list (smaller F-ratios) are more likely to be false positives.

Identifying, quantifying, and determining the statistical significance for all 495 hits discovered by F-ratio analysis using a manual top-down mining approach can be burdensome. To focus identification and quantitation efforts to solely class-distinguishing analytes, a p-value from a t-test (assuming unequal variances) was calculated using the signal encapsulated within a 10 s × 200 ms tile surrounding each hit at the top F-ratio m/z. Hits with a p-value < 0.01 (orange) are identified as class-distinguishing (i.e., true positive) while hits with a p-value > 0.01 (purple) are identified as a false positive (Fig. 4A). This p-value threshold, corresponding to the 99 % confidence level, was selected to minimize the erroneous inclusion of false positive hits during later identification and quantitation efforts [33]. In total, 359 out of 495 hits were determined to be class-distinguishing (p-value < 0.01) and became the focus for later data analysis efforts. Furthermore, the arrows on Fig. 4A indicate the location of the first false positive (hit #121; F-ratio = 21.1) and last true positive hit identified in the hit list (hit #491; F-ratio = 0.14). Hence, 239 true positives were interspersed with 136 false positives at F-ratios below 21.1, which could have been missed if the F-ratio hit list was cut-off at a certain number of hits or at a pre-determined F-ratio threshold.

The receiver operating characteristic (ROC) curve shown in Fig. 4B is generated for the hit list using the labels of true or false positive, defined in Fig. 4A. ROC curves can be beneficial in chemometrics for optimizing specific parameters within a method [51,52] or comparing the performance of different chemometric methods [39,40,53]. For this study, the ROC curve highlights the importance of using a p-value threshold to discover analytes related to the occurrence of PTD in roasted arabica coffee. To define, a ROC curve shows the relationship between the true positive rate versus the false positive rate for a given analytical method [54]. Moving down the hit list, the true positive rate was calculated as the cumulative sum of true positive hits divided by the total number of true positives (i.e., the 359 hits with a p-value < 0.01). The false positive rate was calculated in a similar fashion by keeping track of the running sum and total number of false positive hits discovered. As shown in Fig. 4B, the steps between the first false positive and last true positive hit (denoted with arrows) demonstrate how most of the class-distinguishing analytes before the last true positive hit were intermingled with a small number of false positives at low F-ratios. Additionally, the area under the ROC curve (AUC) defines the probability that the p-value threshold correctly distinguished between true or false positives, where an AUC of 1 represents the maximum classifying power [54]. The AUC for the ROC curve in Fig. 4B equals 0.93, which means that the p-value threshold of 0.01 could distinguish between significant chemical differences and background variation with high accuracy. Ultimately, the high AUC speaks to the outstanding performance of the tile-based F-ratio software [55].

The first 30 identifiable class-distinguishing (i.e., true positive) analytes discovered by F-ratio analysis are listed in Table 1, while Table S2 provides similar information for all 359 true positives identified in Fig. 4. A MV ≥ 800 between the hit and library mass spectrum was required for tentative analyte identification [45]. However, identification for some of these class-distinguishing analytes was challenging due to the presence of larger, overlapping interferent signals, so a

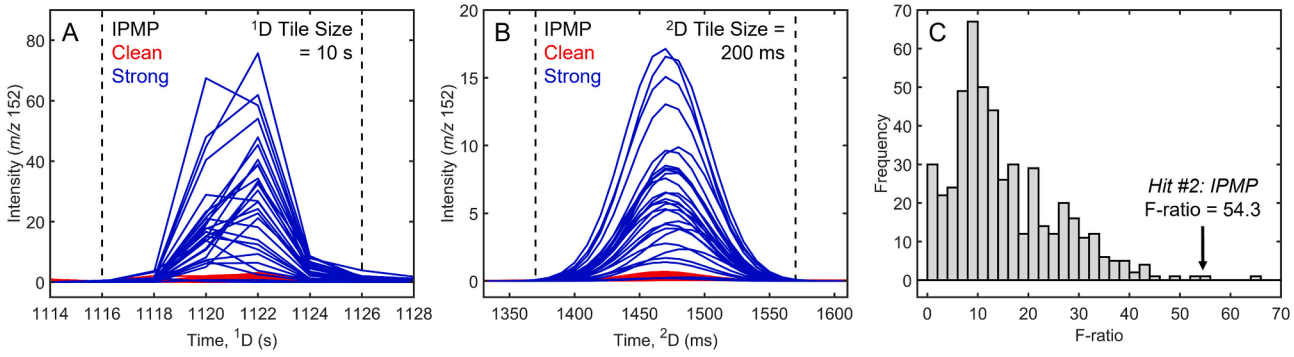

Fig. 3. (A) The ^{1}D peak profile of IPMP at m/z 152 in for the clean (red) and strong PTD (blue) coffee samples. The dashed lines represent the ^{1}D tile size of 10 s. (B) The ^{2}D peak profile of IPMP at m/z 152 with dashed lines representing the ^{2}D tile size of 200 ms. (C) The F-ratio distribution for all 495 hits discovered. The arrow indicates the hit number and F-ratio for IPMP. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

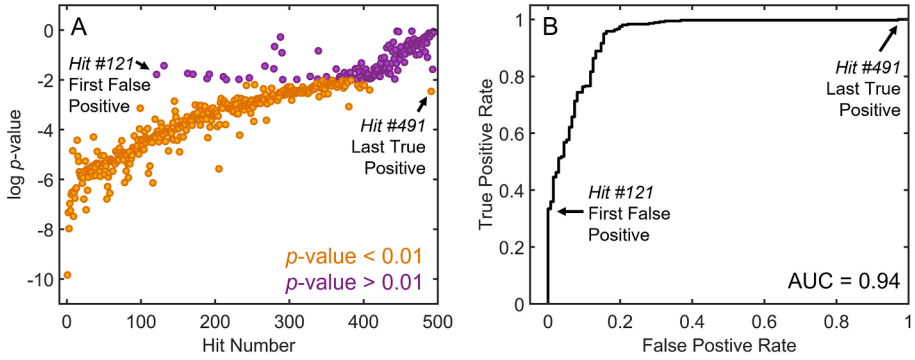

Fig. 4. Reduction of the F-ratio hit list by determining a *p*-value threshold. (A) The *p*-value calculated for each hit using their top F-ratio *m/z*. A total of 359 hits (orange) were determined to be true positives (i.e., class-distinguishing) since their *p*-value < 0.01, which was the *p*-value threshold. The remaining 136 hits (purple) were determined to be false positives (i.e., not class-distinguishing) since their *p*-value > 0.01. The black arrows denote the first false positive (hit #121) and last true positive (hit #491). (B) A receiver operating characteristic (ROC) curve prepared using the results shown in (A). The first false positive (hit #121) and last true positive (hit #491) are denoted again for reference. The area under the curve (AUC) is also provided. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

chemometric decomposition method known as MCR-ALS was applied in these situations. In total, 145 analytes could not be identified even after MCR-ALS decomposition (Table S2). This work discovered many analytes that were previously identified in the literature as a potential marker for PTD [10]. However, discrepancies between this work and previous literature [10,11] may exist due to differences in the roasting degree of the coffee beans or chromatographic methodology. Table 1 also highlights several analytes that have not been documented in coffee beans previously, denoted by a dagger (†). Note, these compound names are tentative since the identifications are based off the analyte mass spectrum alone; future studies with chemical standards can confirm their identification. However, many of these tentative compounds are structurally like other volatiles documented in coffee [2,3,56–60]. The pure analyte *m/z* used for quantifying the concentration ratio ([Strong]/[Clean]) is also provided along with its *p*-value and *LOF*. Twenty-three out of the 359 total analytes were only present in one class (marked with an asterisk; Table 1 & Table S2). The [Strong]/[Clean] for analytes that had a larger signal in the clean samples ranged from 0.02 to 0.82 while the [Strong]/[Clean] for analytes with a larger signal in the strong PTD samples ranged from 1.36 to 29.81 (Table S2). Since many volatiles in the coffee headspace are responsible for the final aroma of the beans, Table 1 and Table S2 reports the known organoleptic/odor properties [50] for 96 of the discovered analytes. Closer examination of these sensory descriptions highlights that analytes with desirable coffee aromas (e.g., roasted, nutty, cocoa, fruity) were found in decreased abundance in the strong PTD samples. Conversely, analytes discovered with larger signals in the strong PTD samples had less favorable aromas (e.g., vegetable-like, musty, aldehydic). These results further support the hypothesis that the diminished concentration of compounds with pleasant aromas further amplifies the odor attributed to PTD [10].

The locations of all 359 class-distinguishing analytes (Table 1 & Table S2) can be visualized as stitch GC × GC chromatograms, which are shown in Fig. 5A and B. These stitch chromatograms categorize the analytes as either having higher signal in the clean (A) or strong PTD (B) coffee samples. This grouping was based on their determined concentration ratio ([Strong]/[Clean]) using the S-ratio algorithm [47]. Based on Fig. 5A and B, 327 analytes discovered had statistically higher abundance in the clean samples ([Strong]/[Clean] < 0.82), whereas the remaining 32 analytes had higher signals in the strong PTD class ([Strong]/[Clean] > 1.36). These stitch chromatograms can also be visualized by projecting the determined concentration ratios onto windows surrounding every peak (Fig. 5C). As referenced earlier, the TIC chromatograms of a clean and strong PTD sample highlighted an overall decrease in signal for the strong PTD samples (Fig. 1). The results in Fig. 5 demonstrate that this overall lower signal in the TIC chromatogram is due to a decrease in analyte concentrations in the strong PTD samples. Additionally, visual comparison of the TIC (Fig. 1) and stitch (Fig. 5) chromatograms demonstrates that F-ratio analysis discovered many class-distinguishing analytes, which were not observed in Fig. 1. These analytes were not observed in the TIC chromatogram due to their signal being smaller than the noise when all the *m/z* were summed together [48]. This result illustrates that both F-ratio analysis is capable of discovering analytes near the limit of quantitation [61] and PTD is responsible for altering the volatile profile of coffee by affecting analytes at all concentration levels.

The resulting PCA scores (A) and loadings (B) plots shown in Fig. 6 was produced using the 359 analytes discovered by F-ratio analysisthat have a significant difference in concentration between the clean and strong PTD sample classes. Compared to the original PCA model (Fig. S2), which captured only 53.6 % of the variance, the PCA model shown in Fig. 6A and B now captures 91.79 % of the total variance within the data. The increased percent variance captured is due to the reduction of background noise and retention time misalignment in the data [62]. The scores plot in Fig. 6A also now illustrates that the coffee samples cluster based on the presence of PTD, where the clean samples (red) are separated from those affected by PTD at any odor attribution level (yellow, green, and blue). Note, there is no observable difference in the PCA model after excluding the signal of IPMP from the data set (Fig. S3), so the sample clustering observed shown in Fig. 6 was not solely due to IPMP. The differentiation between the clean samples and those affected by PTD primarily occurs along PC 1 (Fig. 6A). Inspection of the PCA loadings (Fig. 6B) emphasizes that this differentiation along PC 1 is driven by five class-distinguishing analytes: IPMP (hit #2), 3-(2-cyclopentenyl)-2-methyl-1,1-diphenyl-1-propene (hit #24), 2,4-di-*tert*-butylphenol (hit #34), 1,1,3-trimethyl-3-phenylindan (hit #37), and 2,4-diphenyl-4-methyl-2(*E*)-pentene (hit #40). All five of these highly loaded analytes were found to have large signal differences between the clean and strong PTD class, with [Strong]/[Clean] ranging from 4.5 to 29.8 (Table 1). Fig. 6B also demonstrates that most of the analytes discovered have more subtle differences between their classes since their loadings cluster around zero. However, these analytes with minor concentration differences are still pertinent since they contribute to the aroma profile of coffee.

The results presented in Fig. 6 and previous literature [4,8–11] highlight that IPMP and a small handful of analytes are responsible for differentiating clean and PTD-affected coffee beans. However, this current study aims to delve deeper and showcase how a complete volatile profile of these roasted coffee beans can contribute to a fuller understanding of PTD. For this goal, PLS regression was selected to relate the compounds discovered by F-ratio analysis to the concentration of IPMP.

Table 1
The first 30 identifiable hits (MV ≥ 800) discovered by F-ratio analysis. The hit list is ranked in descending order of their F-ratio hit number. Compounds not previously identified in coffee are denoted by a dagger (†). A concentration ratio for each analyte was calculated as [Strong]/[Clean] using a pure *m/z* based upon applying the S-ratio algorithm [47]. The metrics for determining *m/z* purity (*p*-value and *LOF*) are also reported. Analytes present in only one sample class are denoted by an asterisk (*). For these analytes, only a *p*-value is reported. Sensory descriptions are listed for known analytes [50].

Hit Number	Time, ^{1}D (min)	Time, ^{2}D (s)	Compound	MV	S-ratio *m/z*	[Strong]/[Clean]	*p*-value	*LOF* (%)	Sensory Description
1 †	26.57	0.77	7-Benzofuranamine, 2-methyl-	811	147	0.55	7.5E-11	15.0	
2	18.67	1.47	IPMP	849	152	20.4	4.9E-08	8.60	Earthy, Vegetable, Potato
4 †	38.30	1.16	1,3-Pentadiene, 1,1-diphenyl-, (*Z*)-	823	205	10.9	9.7E-08	15.7	
5 †	25.03	0.86	3(2H)-Benzofuranone, 7-methyl-	801	148	0.47	4.7E-08	10.3	
6 †	39.77	1.43	Benzene, 1,1′-(1,1,2,2-tetramethyl-1,2-ethanediyl)bis-	817	119	20.7	1.8E-07	5.60	
7 †,*	37.23	1.60	Benzene, 1,1′-(1,4-dimethyl-1-butene-1,4-diyl)bis-	809	221	Strong only	3.1E-07		
8 †	37.83	1.29	1,5,6,7-Tetramethyl-3-phenylbicyclo[3.2.0]hepta-2,6-diene	803	194	15.0	2.6E-05	19.7	
9	17.07	0.33	Pyridine, 3-ethyl-	813	136	0.31	5E-08	18.2	Caramellic, Roasted, Hazelnut
10	22.80	0.94	Benzofuran, 2-methyl-	826	103	0.42	5.3E-07	12.7	Burnt, Phenolic
12	17.67	0.27	2,4,6-Octatriene, 2,6-dimethyl-	855	79	0.46	7.8E-09	15.6	Sweet, Floral, Nutty
13 †	23.70	0.87	1H-Indole, 2,3-dihydro-	801	117	0.49	1.9E-07	15.6	
16	22.63	0.93	2-Propenal, 3-phenyl-	869	133	0.17	5.9E-08	7.60	Sweet, Spicy, Honey, Cinnamon
17	14.40	1.02	Cyclohexene, 1-methyl-4-(1-methylethylidene)-	902	103	0.81	3.1E-04	16.8	Fresh, Sweet, Pine, Citrus
21	21.07	1.20	3-Methyl-2,3-dihydro-benzofuran	803	105	0.39	2.5E-07	14.2	
24 †	35.77	1.37	1-Propene, 3-(2-cyclopentenyl)-2-methyl-1,1-diphenyl-	810	222	29.8	1.9E-06	10.8	
25	17.97	1.34	2-Methyl-3-isopropylpyrazine	820	108	0.43	7E-07	10.7	Coffee
28	24.70	0.86	Furan, 2-(2-furanylmethyl)-5-methyl-	910	74	0.46	2.9E-07	9.10	
31	13.50	0.79	1,3,7-Octatriene, 3,7-dimethyl-	924	93	0.32	2.6E-06	14.0	Fruity, Floral
32	27.60	0.85	Phenylethyl acetate	802	159	0.53	1.1E-06	10.0	Floral, Sweet, Honey, Fruity, Cocoa
33	17.43	1.51	Pyrazine, 2-ethyl-5-methyl-	877	93	0.14	1.1E-07	11.6	Coffee, Nutty, Roasted
34	37.23	0.66	2,4-Di-*tert*-butylphenol	908	191	4.50	2.5E-06	4.00	
36	21.40	1.52	*cis*-4-Decenal	809	98	0.41	3.5E-06	5.10	Citrus, Aldehydic, Cardamom
37 †	35.40	1.53	1,1,3-Trimethyl-3-phenylindan	896	236	19.6	6.8E-06	4.80	
38	21.83	1.03	5,6,7,8-Tetrahydroquinoxaline	805	119	0.50	6.7E-07	19.5	Nutty, Roasted, Cereal
39 †	24.97	1.09	Benzofuran, 4,7-dimethyl-	862	144	0.57	2.9E-08	5.60	
40 †	38.90	1.31	2,4-Diphenyl-4-methyl-2(*E*)-pentene	883	236	19.8	4.3E-06	4.90	
41	23.03	0.82	2-Furfurylfuran	901	100	0.47	6.8E-06	14.8	Rich, Roasted
42	17.83	1.24	3-Octen-2-one, (*E*)-	874	68	0.49	9.2E-07	18.4	
43 †,*	37.23	1.23	Benzene, (1,3-dimethyl-3-butenyl)-	807	105	Strong only	3.9E-06		
45 †,*	39.33	1.44	Benzene, [2-methyl-1-(1-methylethyl) propyl]-	801	91	Strong only	7.8E-07		

Fig. 7 shows the PLS prediction of IPMP concentration using the entire F-ratio hit list, excluding IPMP (hit #2) since it was being predicted. The regression plot in Fig. 7A highlights the relationship between the IPMP concentration measured previously [10] and the concentration predicted by the PLS model, where each sample is color coded according to its sensory information. While the data in Fig. 7A is color coded according to the odor attributed to PTD, it is important to note that the PLS regression does not take this sensory information into account when developing a model. Ideally, the measured and predicted concentrations should be equal (i.e., fall along the black 1:1 line). Fig. 7A shows that the PLS model developed using the F-ratio hit list can accurately predict IPMP concentration since the samples cluster closely around the 1:1 line and the model has low prediction errors (< 11 %). Using this PLS model, the linear regression vector (LRV) can be investigated to determine how each volatile analyte discovered by F-ratio analysis correlates with the concentration of IPMP. Note, discovering the relationship that each

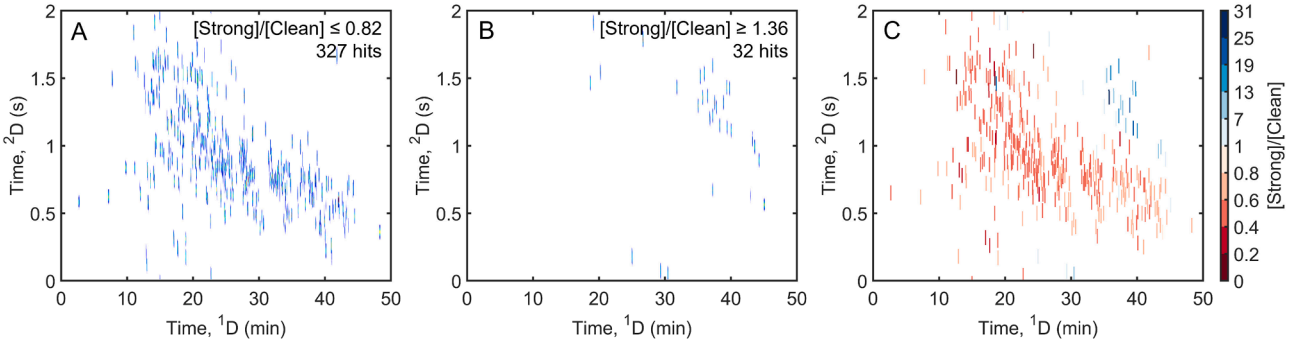

Fig. 5. Visualization of the 359 class-distinguishing hits discovered by F-ratio analysis. (A) Stitch GC × GC chromatogram of the 327 hits that were discovered to have a higher signal in clean coffee samples ([Strong]/[Clean] ≤ 0.82). For each hit, the sample with maximum signal at the S-ratio *m/z* [47] was extracted from the data and placed into the stitch chromatogram. (B) Stitch GC × GC chromatogram of the 32 hits that were discovered to have a higher signal in strong PTD coffee samples ([Strong]/[Clean] ≥ 1.36). (C) Projection of the calculated concentration ratios on the window surrounding each peak shown in (A-B).

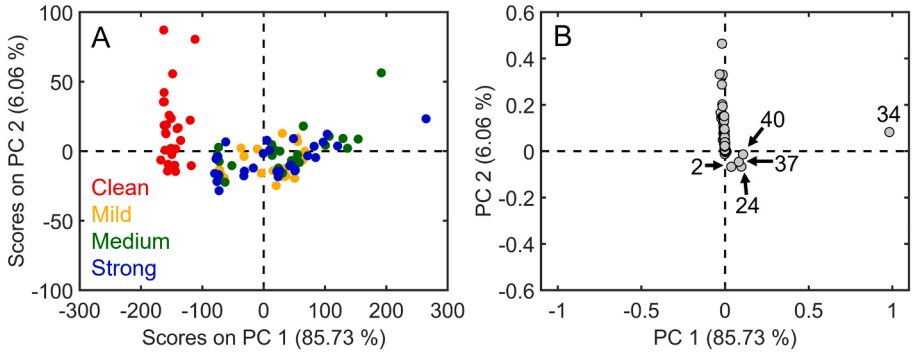

Fig. 6. Results from PCA using the normalized intensity measured at the S-ratio *m/z* [47] for the discovered hits. (A) Scores plot for the model built using the signal for all 359 statistically significant hits in the clean (red), mild PTD (yellow), medium PTD (green), and strong PTD (blue) samples. (B) Loadings plot for the model shown in (A), where each gray dot corresponds to one of the statistically significant hits discovered. Five highly loaded hits (hit #2, 24, 34, 37, and 40) on PC 1 are labeled. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

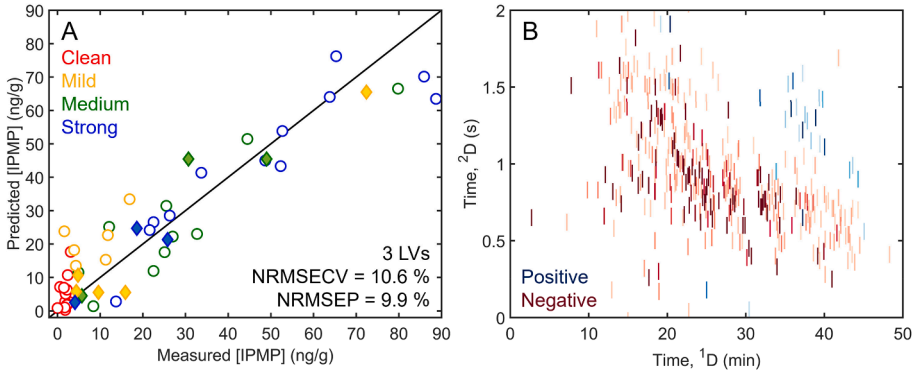

Fig. 7. PLS prediction of IPMP concentration using the normalized intensity measured at the S-ratio *m/z* [47] for all discovered hits except for IPMP (hit #2). (A) Regression plot for the PLS model. The black line symbolizes ideal agreement between the predicted and measured concentrations. Samples used to build the calibration model are shown as unfilled circles while samples used in the external validation set are shown as filled diamonds. Samples are color coded based on their sensory classification: clean (red), mild PTD (yellow), medium PTD (green), and strong PTD (blue). The number of LVs, NRMSECV, and NRMSEP for each PLS model is provided. (B) Projection of the linear regression vector value for each peak on its surrounding window. Positive loadings are highlighted in blue while negative loadings are highlighted in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

analyte has with IPMP is a direct benefit of coupling PLS modeling with tile-based F-ratio analysis. The LRV value for each analyte is provided in Table S2. Fig. 7B displays the LRV as a GC × GC chromatogram, where the value for each hit is projected onto a 2D window surrounding its retention time location. For the PLS model, a positive LRV (blue) indicates that the given compound has a direct relationship with IPMP concentration while a negative LRV (red) indicates an inverse relationship between the compound and IPMP. A visual comparison of the LRV in Fig. 7B and the projection of the concentration ratio for each analyte in Fig. 5C shows a striking similarity, where every analyte with a [Strong]/[Clean] > 1.36 has a positive LRV value and every analyte with a [Strong]/[Clean] < 0.82 has a negative LRV value in the PLS model. Tables 2 and 3 provide a list of the top 20 compounds identified in the LRV with a positive and negative loading, respectively. Despite

Table 2
The top 20 discovered hits with a positive loading in the LRV of the PLS model. The hit list is ranked in descending order of their F-ratio hit number. S-ratios for each analyte were calculated as [Strong]/[Clean] using a pure *m/z*. Tentative compound identifications were made if the mass spectrum match a library spectrum with a MV ≥ 800. Peaks that could not be identified are listed as an unknown (Unk) and numbered according to their order in the hit list (Table S2).

LRV Ranking	LRV	F-ratio Hit Number	Time, ^{1}D (min)	Time, ^{2}D (s)	Compound	MV	[Strong]/[Clean]
1	1.16E-03	34	37.23	0.66	2,4-Di-*tert*-butylphenol	908	4.50
2	8.52E-05	40	38.90	1.31	2,4-Diphenyl-4-methyl-2(*E*)-pentene	883	19.8
3	5.85E-05	116	20.37	1.92	Decanal	805	1.63
4	1.34E-05	24	35.77	1.37	1-Propene, 3-(2-cyclopentenyl)-2-methyl-1,1-diphenyl-	810	29.8
5	1.14E-05	43	37.23	1.23	Benzene, (1,3-dimethyl-3-butenyl)-	807	Strong only
6	8.89E-06	20	40.03	1.12	Unk7		Strong only
7	8.23E-06	149	20.23	1.55	Unk50		2.70
8	7.76E-06	11	31.80	1.44	Unk2		5.17
9	7.51E-06	155	36.00	1.55	Propane, 2-cyclohexyl-2-phenyl-	801	Strong only
10	7.06E-06	99	25.00	0.19	2-Undecanone, 6,10-dimethyl-	923	1.36
11	5.14E-06	282	44.30	0.91	Unk92		4.86
12	4.92E-06	6	39.77	1.43	Benzene, 1,1′-(1,1,2,2-tetramethyl-1,2-ethanediyl)bis-	817	20.7
13	4.80E-06	3	43.23	0.97	Unk1		2.81
14	4.22E-06	252	33.87	0.96	Acenaphthene	901	2.04
15	3.11E-06	325	35.03	1.30	1,1′-Biphenyl, 3,4-diethyl-	812	1.43
16	3.07E-06	37	35.40	1.53	1,1,3-Trimethyl-3-phenylindan	896	19.6
17	3.07E-06	8	37.83	1.29	1,5,6,7-Tetramethyl-3-phenylbicyclo[3.2.0]hepta-2,6-diene	803	15.1
18	2.93E-06	19	37.53	1.33	Unk6		Strong only
19	2.86E-06	4	38.30	1.16	1,3-Pentadiene, 1,1-diphenyl-, (*Z*)-	823	10.9
20	2.41E-06	364	45.07	0.57	1,4-Benzenediol, 2,6-bis(1,1-dimethylethyl)-	802	1.38

Table 3
The top 20 discovered hits with a negative loading in the LRV of the PLS model. The hit list is ranked in descending order of their F-ratio hit number. S-ratios for each analyte were calculated as [Strong]/[Clean] using a pure *m/z*. Tentative compound identifications were made if the mass spectrum match a library spectrum with a MV ≥ 800. Peaks that could not be identified are listed as an unknown (Unk) and numbered according to their order in the hit list (Table S2).

LRV Ranking	LRV	F-ratio Hit Number	Time, ^{1}D (min)	Time, ^{2}D (s)	Compound	MV	[Strong]/[Clean]
1	−1.29E-03	210	15.53	0.95	Pyrazine, 2,6-dimethyl-	915	0.56
2	−8.74E-04	153	28.37	0.80	2-Naphthalenol	886	0.60
3	−8.74E-04	306	30.70	0.41	3-Acetylpyrrole	869	0.60
4	−4.34E-04	261	32.40	0.78	Unk84		0.63
5	−4.00E-04	271	22.17	0.68	Furyl ethyl ketone	874	0.59
6	−2.02E-04	200	31.63	0.78	2,7-Naphthalenediol	859	0.55
7	−1.91E-04	120	24.07	0.97	Pyrazine, 2-methyl-5-(1-propenyl)-, (*Z*)-	875	0.50
8	−8.29E-05	148	19.37	1.28	Pyrazine, 2-methyl-6-propyl-	845	0.31
9	−7.20E-05	258	32.20	0.76	Unk83		0.57
10	−6.20E-05	5	25.03	0.86	3(2H)-Benzofuranone, 7-methyl-	801	0.47
11	−5.70E-05	81	24.23	0.63	2,2′-Bifuran	801	0.45
12	−5.55E-05	217	22.37	0.89	3-Acetyl-2,5-dimethyl furan	863	0.52
13	−5.48E-05	117	25.73	0.65	3-Methyl-2-thiophenecarboxaldehyde	901	0.52
14	−5.46E-05	51	28.57	0.75	4-Hydroxybenzo[b]thiophene	809	0.54
15	−5.35E-05	113	24.77	0.76	2-Acetyl-3-methylpyrazine	863	0.10
16	−4.10E-05	255	34.40	0.81	Unk81		0.48
17	−4.08E-05	1	26.57	0.77	7-Benzofuranamine, 2-methyl-	811	0.55
18	−4.06E-05	223	29.63	0.61	2-Thiophenecarboxylic acid, 4-nitrophenyl ester	805	0.54
19	−3.88E-05	245	33.80	0.77	Thiophene, 2-phenyl-	890	0.29
20	−3.74E-05	10	22.80	0.94	Benzofuran, 2-methyl-	826	0.42

the variation observed in each sensory class (Fig. 2), the PLS model accurately utilizes the entire volatile profile of PTD to predict IPMP concentration. Ultimately, the PLS modeling results underscore that the entire coffee headspace, not just IPMP and a select number of analytes, plays a significant role in PTD.

Box-and-whiskers plots relating the normalized intensity for eight exemplary analytes to their PTD odor attribution are shown in Fig. 8. Additional compounds are shown in Fig. S4. The top row (Fig. 8A – D) focuses on four analytes that were responsible for the sample clustering on the PCA scores plot (Fig. 6) and had positive loading in the PLS model (Fig. 7; Table 2): 3-(2-cyclopentenyl)-2-methyl-1,1-diphenyl-1-propene, 2,4-diphenyl-4-methyl-2(*E*)-pentene, 2,4-di-*tert*-butylphenol, and 1,1,3-trimethyl-3-phenylindan. Many of the volatiles highlighted in Fig. 8A – D and Table 2 are aromatic hydrocarbons and oxygenated compounds, with their signals elevated in the PTD-affected coffee samples. The presence of these analytes can potentially elucidate the biochemical mechanism linking antestia bug damage to PTD. For example, one potential pathway is that bug damage to the coffee plant creates favorable conditions for microorganisms, with research highlighting that bacteria and fungi found on PTD-affected coffee beans produced IPMP as a metabolite [12–14]. Previous research has shown that the presence of microorganisms on coffee beans can cause the concentration of various compound classes like hydrocarbons, phenols, ketones, and aldehydes in coffee to increase [3,63,64]. These types of compounds can form by microorganisms oxidizing the lipids naturally present in the coffee beans [3]. Lipid oxidation induced by microorganisms can potentially explain the increased signals observed in Fig. 8A and B and other analytes in Table 2 like 6,10-dimethyl-2-undecanone (hit #99) and decanal (hit #116). Phenols and phenylindanes like the analytes highlighted in Fig. 8C and D are formed during the roasting process via the degradation of chlorogenic acids [3,58]. Studies on defective Brazilian coffee beans found higher levels of chlorogenic acids in microbe affected green coffee [3,65]. Hence, the increased abundance of 2,4-di-*tert*-butylphenol and 1,1,3-trimethyl-3-phenylindan could be due to the PTD-affected coffee beans having a higher concentration of chlorogenic acids prior to roasting. More specifically, 2,4-di-*tert*-butylphenol has also been

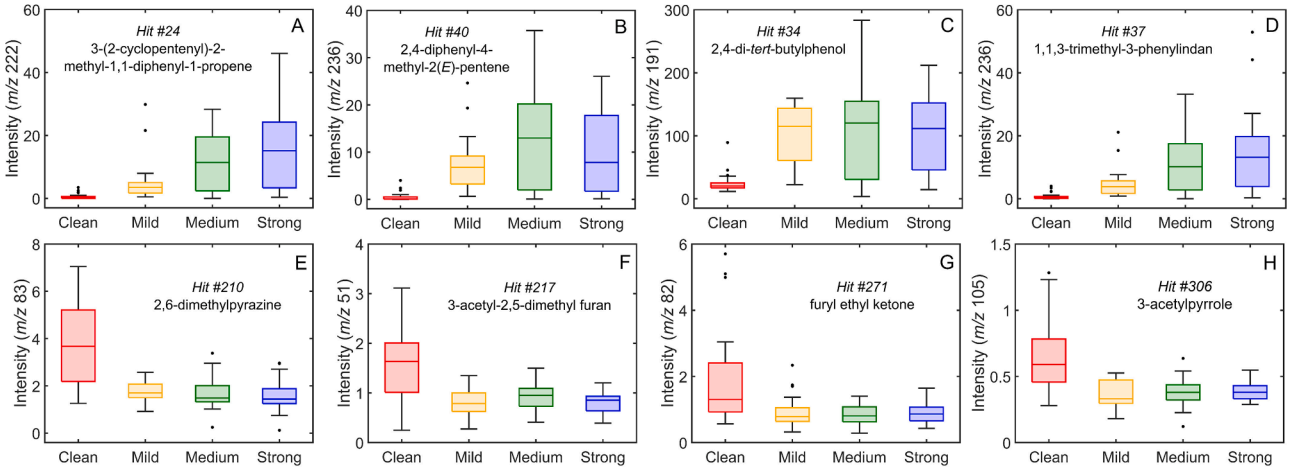

Fig. 8. Box-and-whiskers plots relating the intensity measured at the S-ratio *m/z* [47] to their PTD odor attribution for eight analytes that were highly loaded in the PCA and PLS models. The top row highlights analytes with signals larger in the PTD affected samples: (A) 3-(2-cyclopentenyl)-2-methyl-1,1-diphenyl-1-propene, (B) 2,4-diphenyl-4-methyl-2(E)-pentene, (C) 2,4-di-tert-butylphenol, and (D) 1,1,3-trimethyl-3-phenylindan. The bottom row highlights analytes with signals larger in the clean coffee samples: (E) 2,6-dimethylpyrazine, (F) 3-acetyl-2,5-dimethyl furan, (G) furyl ethyl ketone, and (H) 3-acetylpyrrole.

identified as a volatile indicative of bacterial growth on food products [66,67]. Thus, these results highlight that microbe damage could be a potential cause of PTD.

In contrast, Fig. 8E-H shows the diminished intensity of four analytes that had a negative loading in the PLS model (Fig. 7; Table 3): 2,6-dimethylpyrazine, 3-acetyl-2,5-dimethyl furan, furyl ethyl ketone, and 3-acetylpyrrole. Although these analytes were downregulated in samples affected by PTD, their signal remains invariant across the mild, medium, and strong PTD beans (Fig. 8D – H). Pyrazines and furans like those volatiles highlighted in Fig. 8E and F are the most important contributors to coffee aroma [3], providing cocoa, nutty, and roasted notes [50]. Ketones and pyrroles (examples shown in Fig. 8G and H), while not primary coffee odorants [3], can also provide sweet and fruity aromas [50]. These analytes primarily form during the roasting process, where major chemical reactions convert sugars, lipids, proteins, and chlorogenic acids into hundreds of volatile components [16]. For instance, a previous study found that sucrose and other carbohydrates were more concentrated in non-defective green coffee beans relative to those affected by microbial growth [3]. In turn, the non-defective beans had higher levels of furans and pyrroles after roasting [3]. The same study also indicated that roasted defective coffee beans could exhibit a higher concentration of alkylpyrazines due to an imbalance in amino acid and sugar content, thereby promoting the formation of pyrazines [3]. The work herein and previously reported [10] suggests that alkylpyrazines are less concentrated in PTD-impacted samples [10]. Given the complexity of the roasting process, future work is needed to elucidate how the chemical composition in both clean coffee beans and those affected by PTD impacts the volatile profile of the ultimate roasted product. Nonetheless, the decreased abundance of these analytes suggests that their absence contributes to PTD odor severity and may be further indication that biochemical changes are occurring in PTD-affected beans prior to roasting.

4. Conclusion

Volatile fingerprinting of PTD in roasted arabica coffee beans was performed using HS-SPME-GC × GC-TOFMS and non-targeted chemometrics. Tile-based F-ratio analysis discovered 359 analytes that changed with statistical significance, including IPMP, that differentiated clean and strong PTD coffee samples. These analytes were quantified using a pure *m/z*, providing a determination of the concentration ratio. Most of the analytes (327 out of 359 hits) had higher signals in the clean coffee samples with analyte concentrations ranging from present only in clean class to a [Strong]/[Clean] equal to 0.82. Meanwhile, only 32 analytes had larger abundances for the strong PTD class with their [Strong]/[Clean] ranging from 1.36 to only being present in these samples. Notably, examination of the known sensory properties for some volatiles revealed that analytes with desirable coffee aromas were found in decreased abundance in the strong PTD samples. PCA using the signals for these 359 hits illustrated sample clustering based on the presence of PTD while PLS modeling demonstrated that the compounds discovered by F-ratio analysis can accurately predict the concentration of IPMP. Compounds that are heavily weighted in both the PCA and PLS loadings implies that damage from microorganisms is one possible pathway for PTD. In turn, the damage from microorganisms can induce biochemical changes, which decrease the concentration of analytes with positive aroma descriptions. However, it is important to note that this research cannot exclude that these biochemical changes are also caused by the stress response pathway of *O*-methyltransferase expression in the coffee plant [15]. To further understand the differences in the volatile profile of samples affected by PTD, more work is required to deepen the link between antestia bug predation and the biochemical processes occurring inside green coffee beans and during roasting.

CRediT authorship contribution statement

Caitlin N. Cain: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Meriem Gaida:** Methodology, Software, Validation, Investigation, Visualization, Writing – review & editing. **Pierre-Hugues Stefanuto:** Supervision, Writing – review & editing. **Jean-François Focant:** Supervision, Writing – review & editing. **Robert E. Synovec:** Conceptualization, Investigation, Methodology, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. **Susan C. Jackels:** Conceptualization, Methodology, Resources. **Kristen J. Skogerboe:** Conceptualization, Investigation, Methodology, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

Acknowledgements

C. N. Cain acknowledges the U.S. National Science Foundation Graduate Research Fellowship (DGE-1762114). The authors also thank Counter Culture Coffee for performing the olfactory analysis and providing all of the sorted coffee samples.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.microc.2023.109578.

References

[1] P.D.C. Mancha Agresti, A.S. Franca, L.S. Oliveira, R. Augusti, Discrimination between defective and non-defective Brazilian coffee beans by their volatile profile, Food Chem. 106 (2008) 787–796, https://doi.org/10.1016/j.foodchem.2007.06.019.

[2] A.T. Toci, A. Farah, Volatile compounds as potential defective coffee beans' markers, Food Chem. 108 (2008) 1133–1141, https://doi.org/10.1016/j.foodchem.2007.11.064.

[3] A.T. Toci, A. Farah, Volatile fingerprint of Brazilian defective coffee seeds: Corroboration of potential marker compounds and identification of new low quality indicators, Food Chem. 153 (2014) 298–314, https://doi.org/10.1016/j.foodchem.2013.12.040.

[4] R. Becker, B. Dohla, S. Nitz, O.G. Vitzthum, Identification of the "Peasy" Off-Flavour Note in Central African Coffees, in: 12th Int. Sci. Colloq. Coffee, Montreaux, Switzerland, 29 June - 3 July 1987, Association for Science and Information on Coffee (ASIC), Paris, France, 1987: pp. 203–215.

[5] B. Bouyjou, B. Decazy, G. Fourny, Removing the "potato taste" from Burundian Arabica, Plant. Rech. Dev. 6 (1999) 107–115.

[6] A.G. Ahmed, L.K. Murungi, R. Babin, Developmental biology and demographic parameters of antestia bug Antestiopsis thunbergii (Hemiptera: Pentatomidae), on Coffea arabica (Rubiaceae) at different constant temperatures, Int. J. Trop. Insect Sci. 36 (2016) 119–127, https://doi.org/10.1017/S1742758416000072.

[7] J. Bigirimana, A. Gerard, D. Mota-Sanchez, L.J. Gut, Options for Managing Antestiopsis thunbergii (Hemiptera: Pentatomidae) and the Relationship of Bug Density to the Occurrence of Potato Taste Defect in Coffee, Florida Entomol. 101 (2018) 580–586, https://doi.org/10.1653/024.101.0418.

[8] S.C. Jackels, E.E. Marshall, A.G. Omaiye, R.L. Gianan, F.T. Lee, C.F. Jackels, GCMS investigation of volatile compounds in green coffee affected by potato taste defect and the antestia bug, J. Agric. Food Chem. 62 (2014) 10222–10229, https://doi.org/10.1021/jf5034416.

[9] D. Mutarutwa, L. Navarini, V. Lonzarich, P. Crisafulli, D. Compagnone, P. Pittia, Determination of 3-Alkyl-2-methoxypyrazines in Green Coffee: A Study to Unravel Their Role on Coffee Quality, J. Agric. Food Chem. 68 (2020) 4743–4751, https://doi.org/10.1021/acs.jafc.9b07476.

[10] C.N. Cain, N.J. Haughn, H.J. Purcell, L.C. Marney, R.E. Synovec, C.T. Thoumsin, S. C. Jackels, K.J. Skogerboe, Analytical Determination of the Severity of Potato Taste Defect in Roasted East African Arabica Coffee, J. Agric. Food Chem. (2021), https://doi.org/10.1021/acs.jafc.1c00605.

[11] J.B. Shingiro, P.K. Shee, R.M. Beaudry, D. Thiagarajan, L.D. Bourquin, K.D. Walker, Assessing Alkyl Methoxypyrazines as Predictors of the Potato-Taste Defect in Coffee, ACS Food Sci. Technol. 2 (2022) 1738–1745, https://doi.org/10.1021/acsfoodscitech.2c00233.

[12] D. Gueule, G. Fourny, E. Ageron, A. Le Flèche-Matéos, M. Vandenbogaert, P.A. D. Grimont, C. Cilas, Pantoea coffeiphila sp. nov., cause of the 'potato taste' of Arabica coffee from the African great lakes region, Int. J. Syst. Evol. Microbiol. 65 (2015) 23–29, https://doi.org/10.1099/ijs.0.063545-0.

[13] J.B. Ndayambaje, A. Nsabimana, S. Dushime, F. Ishimwe, H. Janvier, M.P. Ongol, Microbial identification of potato taste defect from coffee beans, Food Sci. Nutr. 7 (2019) 287–292, https://doi.org/10.1002/fsn3.887.

[14] A.R. Hale, P.M. Ruegger, P. Rolshausen, J. Borneman, J. in Yang, Fungi associated with the potato taste defect in coffee beans from Rwanda, Bot. Stud. 63 (2022), https://doi.org/10.1186/s40529-022-00346-9.

[15] K.E. Frato, Identification of Hydroxypyrazine O-Methyltransferase Genes in Coffea arabica: A Potential Source of Methoxypyrazines That Cause Potato Taste Defect, J. Agric. Food Chem. 67 (2019) 341–351, https://doi.org/10.1021/acs.jafc.8b04541.

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

[17] J.M. Davis, J.C. Giddings, Statistical theory of component overlap in multicomponent chromatograms, Anal. Chem. 55 (1983) 418–424, https://doi.org/10.1021/ac00254a003.

[18] Z. Liu, J.B. Phillips, Comprehensive two-dimensional gas chromatography using an on-column thermal modulator interface, J. Chromatogr. Sci. 29 (1991) 227–231, https://doi.org/10.1093/chromsci/29.6.227.

[19] M.S. Klee, J. Cochran, M. Merrick, L.M. Blumberg, Evaluation of conditions of comprehensive two-dimensional gas chromatography that yield a near-theoretical maximum in peak capacity gain, J. Chromatogr. A 1383 (2015) 151–159, https://doi.org/10.1016/j.chroma.2015.01.031.

[20] A.L. Lee, K.D. Bartle, A.C. Lewis, A model of peak amplitude enhancement in orthogonal two-dimensional gas chromatography, Anal. Chem. 73 (2001) 1330–1335, https://doi.org/10.1021/ac001120s.

[21] C. Cordero, J. Kiefl, P. Schieberle, S.E. Reichenbach, C. Bicchi, Comprehensive two-dimensional gas chromatography and food sensory properties: Potential and challenges, Anal. Bioanal. Chem. 407 (2015) 169–191, https://doi.org/10.1007/s00216-014-8248-z.

[22] F. Stilo, C. Bicchi, A. Robbat, S.E. Reichenbach, C. Cordero, Untargeted approaches in food-omics: The potential of comprehensive two-dimensional gas chromatography/mass spectrometry, TrAC - Trends Anal. Chem. 135 (2021), 116162, https://doi.org/10.1016/j.trac.2020.116162.

[23] D. Ryan, R. Shellie, P. Tranchida, A. Casilli, L. Mondello, P. Marriott, Analysis of roasted coffee bean volatiles by using comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry, J. Chromatogr. A 1054 (2004) 57–65, https://doi.org/10.1016/j.chroma.2004.08.057.

[24] S.T. Chin, G.T. Eyres, P.J. Marriott, Identification of potent odourants in wine and brewed coffee using gas chromatography-olfactometry and comprehensive two-dimensional gas chromatography, J. Chromatogr. A 1218 (2011) 7487–7498, https://doi.org/10.1016/j.chroma.2011.06.039.

[25] F.J.M. Novaes, A.I. da Silva Junior, C. Kulsing, Y. Nolvachai, H.R. Bizzo, F.R. de Aquino Neto, C.M. Rezende, P.J. Marriott, New approaches to monitor semi-volatile organic compounds released during coffee roasting using flow-through/active sampling and comprehensive two-dimensional gas chromatography, Food Res. Int. 119 (2019) 349–358, https://doi.org/10.1016/j.foodres.2019.02.009.

[26] G.R. Lopes, S. Petronilho, A.S. Ferreira, M. Pinto, C.P. Passos, E. Coelho, C. Rodrigues, C. Figueira, S.M. Rocha, M.A. Coimbra, Insights on Single-Dose Espresso Coffee Capsules' Volatile Profile: From Ground Powder Volatiles to Prediction of Espresso Brew Aroma Properties, Foods 10 (2021) 2508, https://doi.org/10.3390/foods10102508.

[27] Y. Zou, M. Gaida, F.A. Franchina, P.H. Stefanuto, J.F. Focant, Distinguishing between Decaffeinated and Regular Coffee by HS-SPME-GC×GC-TOFMS, Chemometrics, and Machine Learning, Molecules 27 (2022), https://doi.org/10.3390/molecules27061806.

[28] A. Pua, Y. Huang, R.M. Vivian Goh, K.-H. Ee, L. Li, M. Cornuz, B. Lassabliere, L. Jublot, S.Q. Liu, B. Yu, Multidimensional Gas Chromatography of Organosulfur Compounds in Coffee and Structure-Odor Analysis of 2-Methyltetrahydrothiophen-3-one, J. Agric. Food Chem. 71 (2023) 4337–4345, https://doi.org/10.1021/acs.jafc.2c08842.

[29] E.M. Humston, J.D. Knowles, A. McShea, R.E. Synovec, Quantitative assessment of moisture damage for cacao bean quality using two-dimensional gas chromatography combined with time-of-flight mass spectrometry and chemometrics, J. Chromatogr. A 1217 (2010) 1963–1970, https://doi.org/10.1016/j.chroma.2010.01.069.

[30] P.H. Stefanuto, K.A. Perrault, L.M. Dubois, B. L'Homme, C. Allen, C. Loughnane, N. Ochiai, J.F. Focant, Advanced method optimization for volatile aroma profiling of beer using two-dimensional gas chromatography time-of-flight mass spectrometry, J. Chromatogr. A 1507 (2017) 45–52, https://doi.org/10.1016/j.chroma.2017.05.064.

[31] M. Cialiè Rosso, E. Liberto, N. Spigolon, M. Fontana, M. Somenzi, C. Bicchi, C. Cordero, Evolution of potent odorants within the volatile metabolome of high-quality hazelnuts (Corylus avellana L.): evaluation by comprehensive two-dimensional gas chromatography coupled with mass spectrometry, Anal. Bioanal. Chem. 410 (2018) 3491–3506, https://doi.org/10.1007/s00216-017-0832-6.

[32] J. Crucello, L.F.O. Miron, V.H.C. Ferreira, H. Nan, M.O.M. Marques, P.S. Ritschel, M.C. Zanus, J.L. Anderson, R.J. Poppi, L.W. Hantao, Characterization of the aroma profile of novel Brazilian wines by solid-phase microextraction using polymeric ionic liquid sorbent coatings, Anal. Bioanal. Chem. 410 (2018) 4749–4762, https://doi.org/10.1007/s00216-018-1134-3.

[33] P.E. Sudol, M. Galletta, P.Q. Tranchida, M. Zoccali, L. Mondello, R.E. Synovec, Untargeted profiling and differentiation of geographical variants of wine samples using headspace solid-phase microextraction flow-modulated comprehensive two-dimensional gas chromatography with the support of tile-based Fisher ratio analysis, J. Chromatogr. A 1662 (2022), 462735, https://doi.org/10.1016/j.chroma.2021.462735.

[34] K.J. Johnson, R.E. Synovec, Pattern recognition of jet fuels: Comprehensive GC × GC with ANOVA-based feature selection and principal component analysis, Chemom. Intell. Lab. Syst. 60 (2002) 225–237, https://doi.org/10.1016/S0169-7439(01)00198-8.

[35] H.D. Bean, J.E. Hill, J.M.D. Dimandja, Improving the quality of biomarker candidates in untargeted metabolomics via peak table-based alignment of comprehensive two-dimensional gas chromatography-mass spectrometry data, J. Chromatogr. A 1394 (2015) 111–117, https://doi.org/10.1016/j.chroma.2015.03.001.

[36] L.C. Marney, W.C. Siegler, B.A. Parsons, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-based Fisher-ratio software for improved feature selection analysis of comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry data, Talanta 115 (2013) 887–895, https://doi.org/10.1016/j.talanta.2013.06.038.

[37] B.A. Parsons, L.C. Marney, W.C. Siegler, J.C. Hoggard, B.W. Wright, R.E. Synovec, Tile-Based Fisher Ratio Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry (GC × GC–TOFMS) Data Using a Null Distribution Approach, Anal. Chem. 87 (2015) 3812–3819, https://doi.org/10.1021/ac504472s.
[38] S.E. Prebihalo, G.S. Ochoa, K.L. Berrier, K.J. Skogerboe, K.L. Cameron, J.R. Trump, S.J. Svoboda, J.K. Wickiser, R.E. Synovec, Control-Normalized Fisher Ratio Analysis of Comprehensive Two-Dimensional Gas Chromatography Time-of-Flight Mass Spectrometry Data for Enhanced Biomarker Discovery in a Metabolomic Study of Orthopedic Knee-Ligament Injury, Anal. Chem. (2020), https://doi.org/10.1021/acs.analchem.0c03456.
[39] C.N. Cain, T.J. Trinklein, G.S. Ochoa, R.E. Synovec, Tile-Based Pairwise Analysis of GC × GC-TOFMS Data to Facilitate Analyte Discovery and Mass Spectrum Purification, Anal. Chem. 94 (2022) 5658–5666, https://doi.org/10.1021/acs.analchem.2c00223.
[40] S. Schöneich, G.S. Ochoa, C.M. Monzón, R.E. Synovec, Minimum variance optimized Fisher ratio analysis of comprehensive two-dimensional gas chromatography/mass spectrometry data: Study of the pacu fish metabolome, J. Chromatogr. A 1667 (2022), 462868, https://doi.org/10.1016/j.chroma.2022.462868.
[41] P.E. Sudol, G.S. Ochoa, C.N. Cain, R.E. Synovec, Tile-based variance rank initiated-unsupervised sample indexing for comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry, Anal. Chim. Acta 1209 (2022), 339847, https://doi.org/10.1016/j.aca.2022.339847.
[42] C. Thoumsin, Data Collection Methodology and Accurate Instance Rate Determination in Coffees With Potato Taste Defect (PTD), Count. Cult. Coffee. (2019) 1–9. https://counterculturecoffee.com/wp-content/uploads/2020/04/CCC-PTD-Paper-Final.pdf.
[43] Specialty Coffee Association of America, Cupping Specialty Coffee, (2015) 1–10. http://www.scaa.org/PDF/resources/ cupping-protocols.pdf (Accessed (accessed September 17, 2020).
[44] N. Caporaso, M.B. Whitworth, C. Cui, I.D. Fisk, Variability of single bean coffee volatile compounds of Arabica and robusta roasted coffees analysed by SPME-GC-MS, Food Res. Int. 108 (2018) 628–640, https://doi.org/10.1016/j.foodres.2018.03.077.
[45] S.E. Stein, D.R. Scott, Optimization and testing of mass spectral library search algorithms for compound identification, J. Am. Soc. Mass Spectrom. 5 (1994) 859–866, https://doi.org/10.1016/1044-0305(94)87009-8.
[46] S.C. Rutan, A. de Juan, R. Tauler, Introduction to Multivariate Curve Resolution, in: S.D. Brown, R. Tauler, B. Walczak (Eds.), Compr. Chemom., Vol. 2, Elsevier, 2009: pp. 249–259.
[47] G.S. Ochoa, S.E. Prebihalo, B.C. Reaser, L.C. Marney, R.E. Synovec, Statistical inference of mass channel purity from Fisher ratio analysis using comprehensive two-dimensional gas chromatography with time of flight mass spectrometry data, J. Chromatogr. A 1627 (2020), 461401, https://doi.org/10.1016/j.chroma.2020.461401.
[48] C.N. Cain, S. Schöneich, R.E. Synovec, Development of an Enhanced Total Ion Current Chromatogram Algorithm to Improve Untargeted Peak Detection, Anal. Chem. 92 (2020) 11365–11373, https://doi.org/10.1021/acs.analchem.0c02136.
[49] G.S. Ochoa, M.C. Billingsley, R.E. Synovec, Using solid-phase extraction to facilitate a focused tile-based Fisher ratio analysis of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry data: comparative analysis of aerospace fuel composition, Anal. Bioanal. Chem. (2022), https://doi.org/10.1007/s00216-022-04348-1.
[50] The Good Scents Company, The Good Scents Company Information System, (2018). http://www.thegoodscentscompany.com/.
[51] T.J. Trinklein, R.E. Synovec, Simulating comprehensive two-dimensional gas chromatography mass spectrometry data with realistic run-to-run shifting to evaluate the robustness of tile-based Fisher ratio analysis, J. Chromatogr. A 1677 (2022), 463321, https://doi.org/10.1016/j.chroma.2022.463321.
[52] B.C. Reaser, B.W. Wright, R.E. Synovec, Using Receiver Operating Characteristic Curves To Optimize Discovery-Based Software with Comprehensive Two-Dimensional Gas Chromatography with Time-of-Flight Mass Spectrometry, Anal. Chem. 89 (2017) 3606–3612, https://doi.org/10.1021/acs.analchem.6b04991.
[53] R. Rousseau, B. Govaerts, M. Verleysen, B. Boulanger, Comparison of some chemometric tools for metabonomics biomarker identification, Chemom. Intell. Lab. Syst. 91 (2008) 54–66, https://doi.org/10.1016/j.chemolab.2007.06.008.
[54] C.D. Brown, H.T. Davis, Receiver operating characteristics curves and related decision measures: A tutorial, Chemom. Intell. Lab. Syst. 80 (2006) 24–38, https://doi.org/10.1016/j.chemolab.2005.05.004.
[55] D.W. Hosmer, S. Lemeshow, R.X. Sturdivant, Assessing the Fit of the Model, in: Appl. Logist. Regres., 3rd ed., Wiley, Hoboken, NJ, 2013: pp. 153–226.
[56] D. Bressanello, E. Liberto, C. Cordero, B. Sgorbini, P. Rubiolo, G. Pellegrino, M. R. Ruosi, C. Bicchi, Chemometric Modeling of Coffee Sensory Notes through Their Chemical Signatures: Potential and Limits in Defining an Analytical Tool for Quality Control, J. Agric. Food Chem. 66 (2018) 7096–7109, https://doi.org/10.1021/acs.jafc.8b01340.
[57] L.F. Huang, M.J. Wu, K.J. Zhong, X.J. Sun, Y.Z. Liang, Y.H. Dai, K.L. Huang, F. Q. Guo, Fingerprint developing of coffee flavor by gas chromatography-mass spectrometry and combined chemometrics methods, Anal. Chim. Acta 588 (2007) 216–223, https://doi.org/10.1016/j.aca.2007.02.013.
[58] S. Blumberg, O. Frank, T. Hofmann, Quantitative studies on the influence of the bean roasting parameters and hot water percolation on the concentrations of bitter compounds in coffee brew, J. Agric. Food Chem. 58 (2010) 3720–3728, https://doi.org/10.1021/jf9044606.
[59] L. Poisson, I. Blank, A. Dunkel, T. Hofmann, The Chemistry of Roasting—Decoding Flavor Formation, in: Cr. Sci. Coffee, Elsevier, London, UK, 2017: pp. 273–309. https://doi.org/10.1016/B978-0-12-803520-7.00012-8.
[60] G. Strocchi, E. Bagnulo, M.R. Ruosi, G. Ravaioli, F. Trapani, C. Bicchi, G. Pellegrino, E. Liberto, Potential Aroma Chemical Fingerprint of Oxidised Coffee Note by HS-SPME-GC-MS and Machine Learning, Foods 11 (2022) 1–14, https://doi.org/10.3390/foods11244083.
[61] P.E. Sudol, G.S. Ochoa, R.E. Synovec, Investigation of the limit of discovery using tile-based Fisher ratio analysis with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry, J. Chromatogr. A 1644 (2021), https://doi.org/10.1016/j.chroma.2021.462092.
[62] V.E. de Almeida, D.D. de Sousa Fernandes, P.H.G.D. Diniz, A. de Araújo Gomes, G. Véras, R.K.H. Galvão, M.C.U. Araujo, Scores selection via Fisher's discriminant power in PCA-LDA to improve the classification of food data, Food Chem. 363 (2021), https://doi.org/10.1016/j.foodchem.2021.130296.
[63] A.C. de Oliveira Junqueira, G.V. de Melo Pereira, J.D. Coral Medina, M.C.R. Alvear, R. Rosero, D.P. de Carvalho Neto, H.G. Enríquez, C.R. Soccol, First description of bacterial and fungal communities in Colombian coffee beans fermentation analysed using Illumina-based amplicon sequencing, Sci. Rep. 9 (2019) 1–10, https://doi.org/10.1038/s41598-019-45002-8.
[64] X. Shen, B. Wang, C. Zi, L. Huang, Q. Wang, C. Zhou, W. Wen, K. Liu, W. Yuan, X. Li, Interaction and Metabolic Function of Microbiota during the Washed Processing of Coffea arabica, Molecules 28 (2023), https://doi.org/10.3390/molecules28166092.
[65] A. Farah, M.C. Monteiro, V. Calado, A.S. Franca, L.C. Trugo, Correlation between cup quality and chemical attributes of Brazilian coffee, Food Chem. 98 (2006) 373–380, https://doi.org/10.1016/j.foodchem.2005.07.032.
[66] F. Zhao, P. Wang, R.D. Lucardi, Z. Su, S. Li, Natural sources and bioactivities of 2,4-di-tert-butylphenol and its analogs, Toxins 12 (2020) 1–26, https://doi.org/10.3390/toxins12010035.
[67] S. Fang, S. Liu, J. Song, Q. Huang, Z. Xiang, Recognition of pathogens in food matrixes based on the untargeted in vivo microbial metabolite profiling via a novel SPME/GC × GC-QTOFMS approach, Food Res. Int. 142 (2021), https://doi.org/10.1016/j.foodres.2021.110213.