





Article

# Untargeted/Targeted 2D Gas Chromatography/Mass Spectrometry Detection of the Total Volatile Tea Metabolome

Joshua Morimoto <sup>1</sup>, Marta Cialière Rosso <sup>2</sup>, Nicole Kfoury <sup>1</sup>, Carlo Bicchi <sup>2</sup>,  
Chiara Cordero <sup>2</sup> and Albert Robbat, Jr. <sup>1,\*</sup>

<sup>1</sup> Department of Chemistry, Tufts University, Medford, MA 02155, USA; joshua.morimoto@tufts.edu (J.M.); nicole.kfoury@tufts.edu (N.K.)

<sup>2</sup> Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino, 10125 Turin, Italy; marta.cialierosso@unito.it (M.C.R.); carlo.bicchi@unito.it (C.B.); chiara.cordero@unito.it (C.C.)

\* Correspondence: albert.robba@tufts.edu; Tel.: +1-617-627-3474

Received: 29 July 2019; Accepted: 15 October 2019; Published: 18 October 2019



**Abstract:** Identifying all analytes in a natural product is a daunting challenge, even if fractionated by volatility. In this study, comprehensive two-dimensional gas chromatography/mass spectrometry (GC×GC-MS) was used to investigate relative distribution of volatiles in green, pu-erh tea from leaves collected at two different elevations (1162 m and 1651 m). A total of 317 high and 280 low elevation compounds were detected, many of them known to have sensory and health beneficial properties. The samples were evaluated by two different software. The first, GC Image, used feature-based detection algorithms to identify spectral patterns and peak-regions, leading to tentative identification of 107 compounds. The software produced a composite map illustrating differences in the samples. The second, Ion Analytics, employed spectral deconvolution algorithms to detect target compounds, then subtracted their spectra from the total ion current chromatogram to reveal untargeted compounds. Compound identities were more easily assigned, since chromatogram complexities were reduced. Of the 317 compounds, for example, 34% were positively identified and 42% were tentatively identified, leaving 24% as unknowns. This study demonstrated the targeted/untargeted approach taken simplifies the analysis time for large data sets, leading to a better understanding of the chemistry behind biological phenomena.

**Keywords:** tea; metabolomics; GC/MS; software; database; MS subtraction; spectral deconvolution; 2DGC; volatilomics

## 1. Introduction

Obtaining the total metabolome of a natural or biological sample is a significant challenge. Even when analyzed by both gas and liquid chromatography (GC, LC), analyzing complex mixtures results in the detection of hundreds, if not thousands, of compounds. For example, we found that climate-induced tea plant (*Camellia sinensis* (L.) Kuntze) interactions cause significant differences in both the presence and concentration of secondary metabolites [1–3]. Abiotic pressures alone influence the relative distribution of more than 750 volatile compounds, with ~450 detected in any given sample. Differences in elevation, seasonal temperature, and rainfall from 2013 to 2016 caused two-thirds of the metabolites to either increase or decrease in concentration, one-third of them by more than 100%. We also measured nonvolatile concentration differences, including the catechins (flavan-3-ol derivatives with high antioxidant power). Only five days after the onset of the East Asian monsoon rains began, catechin concentrations decreased while total polyphenols and antioxidant potential increased. In these

studies, we consider the tea “environmental,” since we microwaved the leaves in the field to stop enzymatic oxidation; they were not processed to produce commercial tea.

We previously relied on automated sequential, 2D gas chromatography/mass spectrometry (GC-GC/MS) to build the environmental tea database and on GC/MS to quantify the analyte distribution in the samples. In this study, we used comprehensive 2D GC/MS (GC×GC/MS) to accomplish both tasks, building from the existing database. When comparing each technique, library-building by GC-GC/MS relied upon fifty 1-min sample transfers (heart-cuts) from the first (polar) to the second (nonpolar) column to obtain clean spectra. The total time to obtain 50 data files was 4.5 days. In contrast, GC×GC/MS, produced one data file in one hour. Although runtimes significantly differed, data analysis of the respective files for library-building purposes took about the same amount of time.

The terms targeted, untargeted, and feature detection are often applied to GC/MS and LC/MS data. In targeted analysis, a predetermined list of compounds is selected for the analysis, whereas untargeted involves the evaluation of all detectable compounds in the sample [4,5]. The following criteria affirm compound identity [6–8]. Positive identification requires confirmation by at least two independent measurements, such as sample and reference compound mass spectra and linear retention indices (LRI). Tentative identification is based on an acceptable match between sample and commercial library, database, or literature spectra. Because natural products and biological metabolomes contain hundreds to thousands of compounds [9], it is impractical to purchase reference standards to confirm all sample identities. Therefore, assignments are typically tentative for most compounds.

Given the numerous and diverse compounds in natural products and in metabolomic datasets, feature-based peak detection algorithms have become popular.  $m/z$  peak-retention time pairs are used to assess peak region commonalities and differences in large data sets to reduce data complexity. Because feature-based peak detection is overly sensitive to algorithmic parameters, coelution and instrument noise, it produces many more features than actual compounds [10–12], with less than 2% yielding identifiable analytes [13]. Although peak detection tools can differentiate samples [11,14,15], identifying why they differ is difficult, since features themselves provide no biological context [16,17]. Despite these drawbacks, feature-based detection software is popular among the metabolomics community [15].

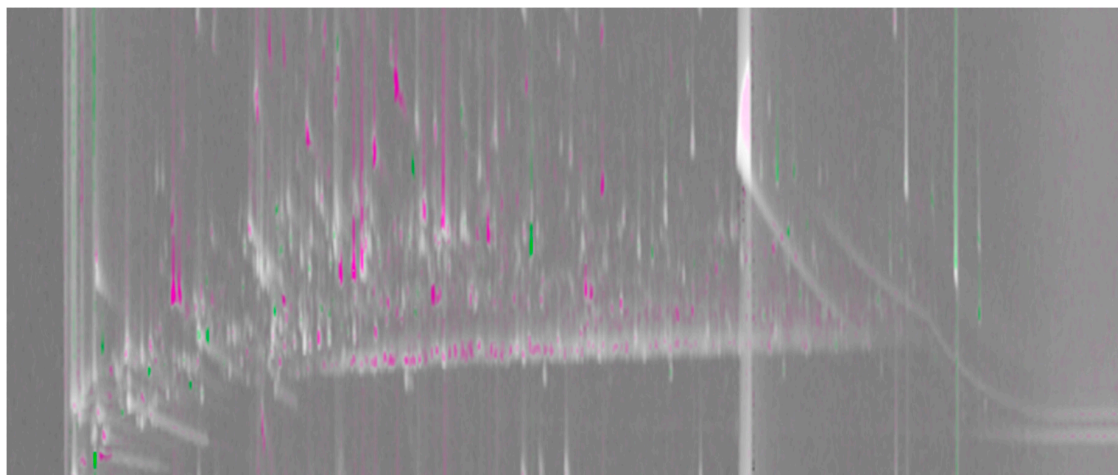
In this study, we used GC×GC/MS to reveal compositional differences in farmer-processed green, pu-erh teas (herein called pu-erh), collected from the same farm, plants, elevations, and time-period as environmental tea. 2D data were analyzed by GC Image and Ion Analytics. The first was adopted to detect 2D-peaks and to delineate peak-regions covering the full chromatographic space. Similarities and differences between high and low elevation samples were highlighted by both peak-region feature distribution and visual feature rendering [18]. Additionally, we identified compounds where possible. The second provided an untargeted/targeted workflow, based on spectral deconvolution and spectral subtraction of analytes found in the sample, to produce clean mass spectra and quantifiable constituent differences. New algorithms facilitated the annotation process.

## 2. Results

### 2.1. Molecular Feature Detection

Figure 1 shows a comparative visualization, based on a visual features approach [18], highlighting compositional similarities and differences in the high (reference sample) and low (analyzed sample) elevation pu-erh teas. This pair-wise comparison is done on a composite chromatogram obtained by summing each sample's 2D chromatograms ( $n = 2$ ) after transformation and re-alignment. In total, 1450 peak-regions were delineated; 107 with spectra clean enough to make tentative compound assignments, see Supplementary Table S1. Each pixel in Figure 1 corresponds to detector acquisition points (scans) that show response differences. The algorithm subtracts the pixel value registered in the reference image from the corresponding value in the analyzed image, with the difference divided by the larger of the pixel values. The brighter the pixel, the larger the difference is relative to the analyzed pixel value; the darker the pixel, the larger the difference is relative to the reference pixel

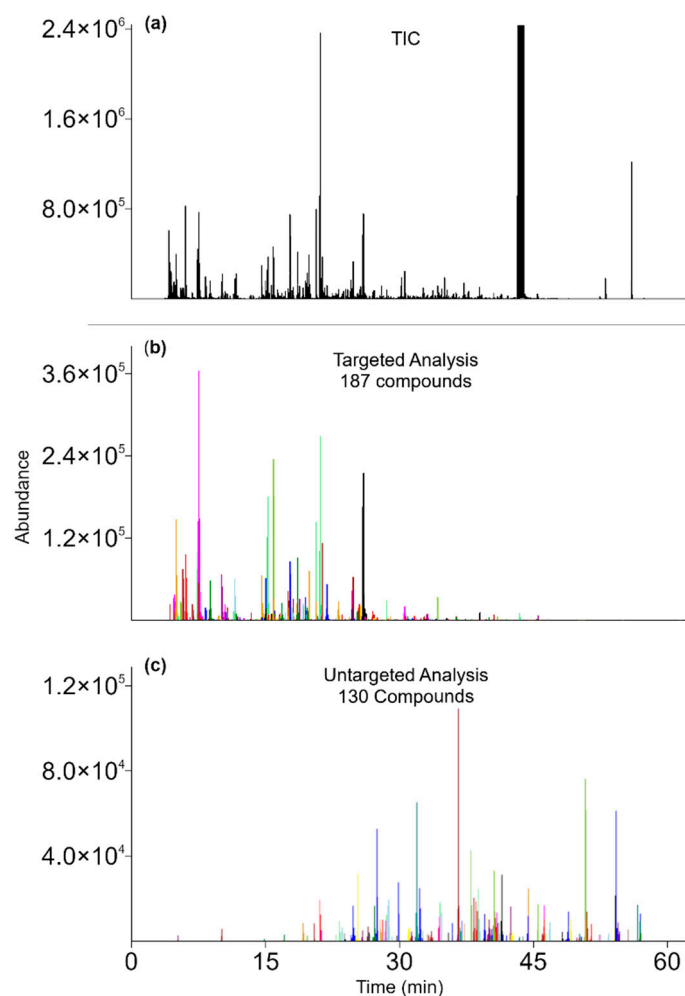
value. Medium gray indicates the difference is small relative to the value. The hue of a pixel indicates whether the analyzed image (green) or reference image (magenta) has the higher value. The saturation (color vs. grey) of a pixel indicates the magnitude of the difference ratio between the analyzed and reference images, with grey indicating equal pixel values and bold colors large differences.



**Figure 1.** A comparative visualization between composite GC×GC/MS chromatograms of the high (reference image) and low (analyzed image) elevation teas. The magenta and green peak regions show the areas in the chromatogram where relative distribution of common analytes is higher in the high and low elevation teas, respectively. Light grey peak regions correspond to analytes with similar percentage response in the two chromatograms.

## 2.2. Untargeted/Targeted Analysis

Shown in Figure 2 are the (a) total and (b and c) reconstructed ion current (TIC and RIC) chromatograms for high elevation tea. Using the environmental database as target compounds, spectral deconvolution produced chromatogram (b), which reveals the 187 compounds in pu-erh tea that survived farmer processing. After subtracting the target compound spectra from the TIC chromatogram, untargeted analysis revealed another 130 compounds in the sample (c). Table 1 lists the breakdown of the number of compounds by their identification levels, showing 107, 132, and 78 positively and tentatively identified compounds, and unknowns in the high elevation pu-erh tea, respectively. Although 25% of the volatile constituents are unknowns, should statistical analysis reveal their importance, we know where in the separation to collect the compound by GC-GC for further analysis. In addition to retention and spectral information, the database includes sample type, the sensory characteristics of the compound in the sample, and known health benefits.



**Figure 2.** The TIC (a) and RIC chromatograms of targeted (b, in environmental tea) and untargeted (c, due to farmer processing) compounds in high elevation pu-erh tea.

**Table 1.** Targeted and untargeted compounds found by Ion Analytics.

Identity Level	High Elevation		Low Elevation	
	Targeted	Untargeted	Targeted	Untargeted
Positive	92	15	82	13
Tentative	78	54	69	43
* Unknown	17	61	15	58
<b>Total</b>	<b>187</b>	<b>130</b>	<b>166</b>	<b>114</b>

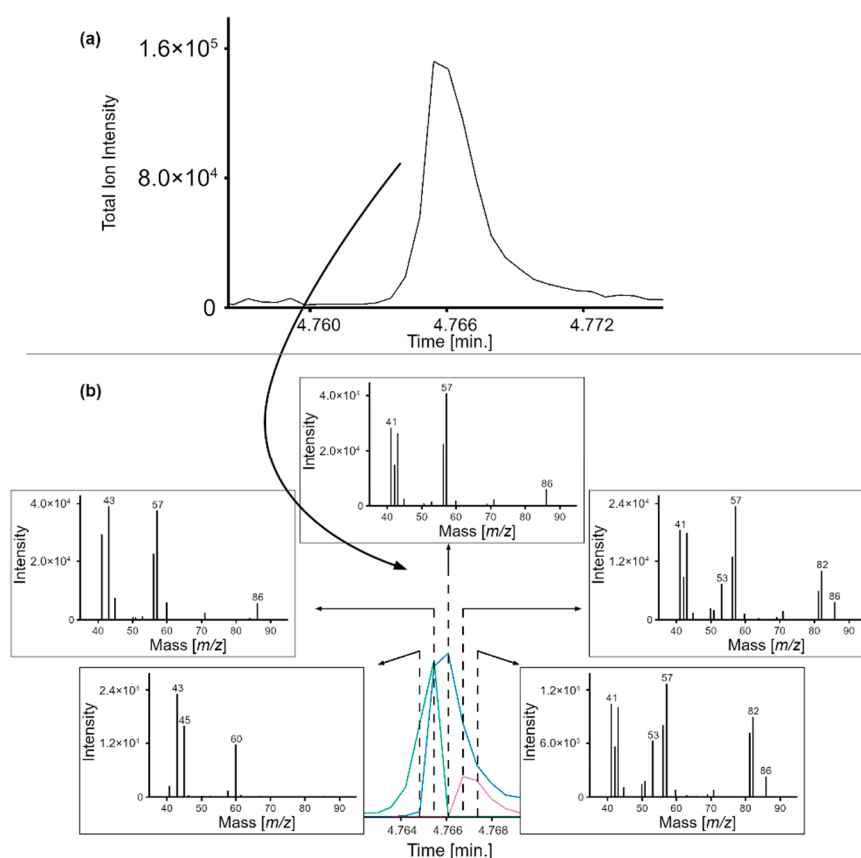
\* Unknowns are assigned a numerical identifier in the database.

Compared to the 450 compounds we typically detect by stir bar sorptive extraction [19], head space solid phase microextraction (HS-SPME) produced a lower extraction yield, as expected, with 317 and 280 compounds in the high and low elevation samples, see Table 1. The high elevation sample contained 42 unique compounds; whereas only five were found in the low elevation teas. Also listed in the table are the number of targeted and untargeted compounds, see Supplementary Table S2 for identities. Once the total, detectable profile was established, GC Image (by pattern recognition) confirmed 272 of the 322 (84%) assignments, see Supplementary Table S1.

### 3. Discussion

Supplementary Figure S1 illustrates the process we use to analyze complex natural products such as botanicals, essential oils, herbs, and spices caused by environmental and/or manufacturing perturbations. In this study, we were interested in learning how sensory-active and health beneficial compounds in pu-erh tea differ in plants grown at high (1651 m) and low (1162 m) elevations. Since the leaves are processed in the same manner, differences in chemistry are primarily due to differences in temperature ( $\Delta 3$  °C). Target compounds are leaf metabolites that survived farmer processing. Untargeted compounds are those produced by the fire-heated kettle process when making pu-erh tea. Spectral deconvolution of the target (environmental database) compounds, ~60% (187/317 and 166/280) of the constituents in each sample, followed by subtraction of their spectra simplified the analysis.

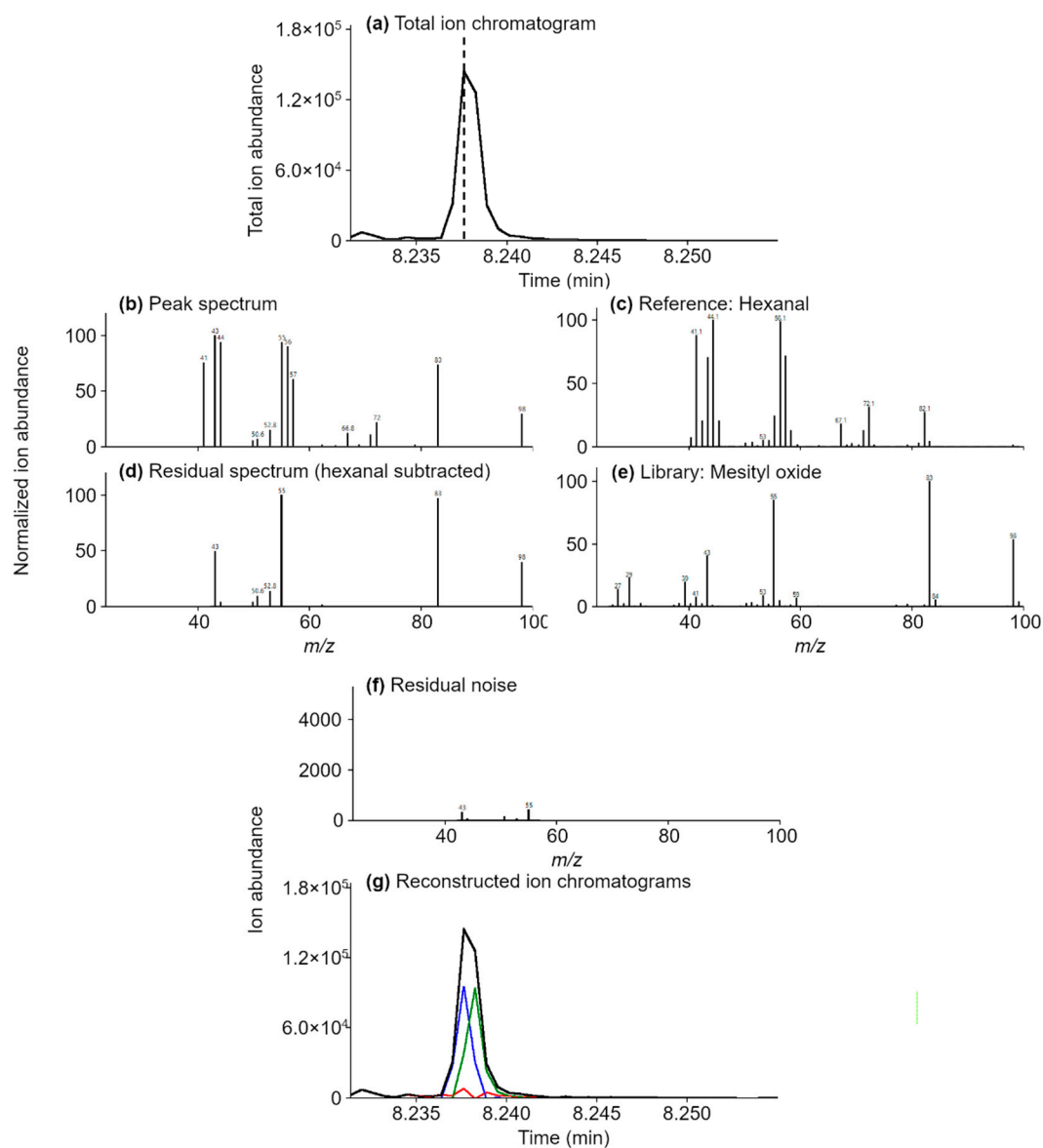
Figure 3 shows the TIC peak of three coeluting target compounds (acetic acid, 2-methylfuran, and hexane) in high elevation tea. Although the spectrum at each peak scan is different (Figure 3b), Ion Analytics correctly identified each analyte by deconvolving their spectral signature. Shown in the bottom panel are the RIC peaks for acetic acid (green), 2-methylfuran (pink), and hexane (blue), whose identities were confirmed using reference standards. The RIC peaks provide the means to measure the relative distribution of each analyte in the sample.



**Figure 3.** (a) Targeted analysis example of high elevation tea. (b) Spectral deconvolution of acetic acid (green), hexane (blue), and 2-methylfuran (pink) ions and relative abundances.

Figure 4 is an example of an untargeted analysis where the spectrum (c) for the target compound, hexanal, is subtracted from the TIC (a), see example spectra (b), exposing the spectrum of an unknown. Since spectrum (d) is constant at each peak scan, the MS subtraction algorithm reveals another compound. The retention time and mass spectrum match mesityl oxide (e). Subtracting the mesityl oxide reference spectrum from the remaining signal yields baseline noise (f and g), positively confirming

the peak assignments and that no other compounds elute in the corresponding peak-region. Also shown in (g) are the RIC peaks for both hexanal (blue) and mesityl oxide (green).



**Figure 4.** Untargeted analysis example, high elevation tea. Spectral deconvolution and MS subtraction of hexanal, the target compound, ions and relative abundances (c) from the TIC (a) peak spectra (b) yields the residual spectrum (d) for mesityl oxide (e). Subtraction of the ions and relative abundances of both compounds results in baseline noise (f,g). The RIC chromatograms for hexanal (blue) and mesityl oxide (green) are also in (g). Experimental spectra were acquired in the range of 40–280  $m/z$ , and therefore, ions below 40  $m/z$  are missing from spectra (b–d,f).

This approach reduces the complexity of each subsequent analysis by using the results of preceding samples as target compounds. By annotating the database, analytes can be tracked independent of sample type and from sample-to-sample, year-to-year. Both software provide data to analyze statistically. How one starts depends on the goal of the investigation. How one ends depends on whether speciation provides meaningful input into the system under investigation, see below.

## Effects of Processing and Elevation on Tea

The most flavorful teas in Yunnan, China, are those grown at the highest elevations [20,21]. Earlier, we showed that high compared to low elevation “environmental” teas contained more and higher concentrations of metabolites that exhibit sweet, floral notes [1]. We assumed, therefore, that the farmer-processed pu-erh tea would as well. The results of this study support this finding. The high elevation tea retained more of the sensory-pleasing environmental metabolites than the low elevation tea, which possessed more of the earthy, harsher tasting aromatics. The comparative chromatogram, depicted in Figure 1, shows more magenta peaks overall, indicating a decrease in volatiles in the low elevation pu-erh compared to the high. Further examination with Ion Analytics showed that of the 275 common analytes, 134 were higher and 19 were lower in abundance in the high elevation tea at the 95% confidence interval of the average relative percent difference, namely,  $23\% \pm 46\%$  ( $n = 2$  samples), with 119 and 60 higher than 100% and 200%, respectively. We found substantial concentration differences in the oxygenated monoterpenes. The high elevation tea contained more of the *trans*- (14000%) and *cis*- (3400%) furanoid linalool oxides, compounds characterized as sweet and floral, than the low elevation tea. Similarly, the *cis*- (700%) and *trans*- (750%) pyranoid linalool oxides, which exhibit citrusy, woody notes, were also higher in concentration.

Table 2 lists the 42 compounds unique to high elevation pu-erh and the five compounds unique to low elevation pu-erh, along with their sensory and/or health beneficial properties. Although high elevation spring teas are higher in quality, the occurrence of the summer monsoon rains offsets the elevational (temperature) effect, resulting in less desirable tea, as evidenced by farmers receiving 50% less for the summer compared to spring teas [22]. These findings illustrate why it is important to measure the total volatile profile, especially when evaluating the health benefits of green tea in clinical trials [23–26].

**Table 2.** Unique compounds in high and low elevation pu-erh tea and their sensory active and/or health beneficial properties.

High Elevation Compounds	Aroma *	Health Benefits
furfural	woody, almond, baked bread	—
18	—	—
(2E)-hexenal	green, banana, aldehydic	antimicrobial [27]
2-furanmethanol	sweet, caramel, burnt	—
(2E)-hexenol	leafy, fruity, unripe banana	—
2-heptanol	fruity, oily, fatty	—
2,5-dimethylpyrazine	cocoa, roasted nuts	—
2(5H)-furanone	buttery	—
heptanol	musty, leafy, herbal, peony	cardioprotective [28]
(3E)-hexenoic acid	fruity, honey, acidic	—
101	—	—
(3Z)-hexenyl acetate	green, banana, apple	—
heptanoic acid	rancid, sour, sweat	—
2-methoxyphenol	phenolic, smoke, spice	—
maltol	caramel, cotton candy, fruity	antianxiety [29], antioxidant [30]
114	—	—
511	—	—
(3Z)-hexenyl butyrate	green apple, fruity, wine	—
(2E)-hexenyl butyrate	green, apricot, ripe banana	—
hexyl butyrate	fruity, apple, waxy	—
512	—	—
514	—	—
nerol	neroli, citrus, magnolia	antibacterial [31], antifungal [32], antinociceptive/anti-inflammatory [33]
(3Z)-hexenyl valerate	apple, kiwi, unripe banana, tropical	—
(3Z)-hexenyl isovalerate	green apple, tropical, pineapple	—
phenylethyl acetate	rose, fruity	—
pentyl hexanoate	pineapple, apple, pear	—
1-nitro-2-phenyl ethane	floral, spice	cardioprotective [34]
$\gamma$ -nonalactone	coconut, creamy, waxy, buttery	—

Table 2. Cont.

High Elevation Compounds	Aroma *	Health Benefits
(3Z)-hexenyl hexenoate	waxy, pear, winey, grassy, pineapple	—
hexyl hexanoate	fresh cut grass, vegetable	—
(2E)-hexenyl caproate	cognac, herbal, waxy	—
(Z)-jasmone	floral, woody, herbal, spicy	antibacterial [35], anticancer [36]
(E,E)- $\alpha$ -farnesene	citrus, lavender, bergamot,	—
2,4-di-tert-butylphenol	phenolic	antioxidant [37]
$\delta$ -cadinene	thyme, woody	—
(Z)-calamenene	herb, spice	antimalarial [38], antitumor [39]
dodecanoic acid	fatty, coconut, bay oil	cardioprotective [40], antibacterial/anti-inflammatory [41]
caryophyllene oxide	woody, spicy	anticancer/analgesic/anti-inflammatory [42]
$\tau$ -muurolol	herbal, spicy, honey	antibacterial [43], antioxidant [44]
$\alpha$ -cadinol	herbal, woody	antibacterial/antioxidant [43], anticancer [45], anti-inflammatory [45]
bancroftinone	—	—
Low Elevation Compounds	Aroma *	Health Benefit
ethyl acetate	weedy, green	—
isoamyl alcohol	alcoholic, banana	antifungal [46]
(2E)-pental	pungent, green apple, orange, tomato	—
m-ethyltoluene	—	—
118	—	—

\* Aroma information was obtained from The Good Scents Company (<http://thegoodscentscompany.com/>).

## 4. Materials and Methods

### 4.1. Sample Collection

Tea samples (var. *assamica*) were collected from Nannuo Mountain, Menghai County in Yunnan Province, China, in 2014. Low elevation teas, grown at 1162 m, were collected from Xiang Yang Village from June 8–10. High elevation teas, grown at 1651 m, were collected from Ya Kou Old Village from June 10–12. Elevation differences correspond to a 3 °C cooler temperature at high elevation (<https://www.spc.noaa.gov/exper/soundings/help/lapse.html>). Farmers at the study site processed the samples as green, pu-erh tea [2,22].

### 4.2. Sample Preparation and HS-SPME

Added to a 20 mL glass vial, thermostatted at 50 °C, were 1.5 g of dried plant material and 2 mL of ultrapure water. Volatiles were sampled in the headspace of the vial by solid phase microextraction (SPME) for 50 min. The fibers were coated with 50/30  $\mu$ m thick divinylbenzene/carboxen on polydimethyl siloxane (PDMS), which was 2 cm long (Supelco, Bellefonte, PA, USA). Fibers were preconditioned according to the manufacturer before the addition of internal standards,  $\alpha$ - and  $\beta$ -thujone, which were sampled by exposing the SPME to 5  $\mu$ L of a stock solution at 1000 mg/L for 20 min at 50 °C. The internal standards were used to normalize the GC $\times$ GC/MS peak responses. Two replicates were prepared per sample.

### 4.3. GC $\times$ GC/MS Instrumentation

Tea samples were analyzed using a Gerstel (Mülheim an der Ruhr, Germany) MPS-2 multipurpose sampler and Agilent (Santa Clara, CA, USA) models 6890 GC and 5975C MS. The MS, operated in electron ionization mode at 70 eV, scanned from 40 to 280  $m/z$  at 30 Hz. The GC $\times$ GC was equipped with a two-stage KT 2004 loop thermal modulator (Zoex Corporation, Houston, TX, USA), which was cooled by liquid nitrogen. The hot jet pulse time was 250 ms. The modulation period was 4 s. A mass flow controller linearly reduced the cold-jet total flow from 40% to 8% by the end of the run. A deactivated fused silica capillary loop (1 m  $\times$  0.1 mm  $d_c$ ) transferred sample portions from the first to the second column. The first column was an SE52 column (95% PDMS, 5% phenyl, 30 m  $\times$  0.25 mm  $\times$  0.25  $\mu$ m), which was coupled to an OV1701 column (86% PDMS, 7% phenyl, 7% cyanopropyl, 1 m  $\times$  0.1 mm  $\times$



0.10  $\mu\text{m}$ ) as the second column. Both columns were purchased from Mega (Legnano, Milan, Italy). The temperature program was 50 °C (1 min) to 210 °C at 3 °C/min, then at 280 °C (10 min) at 10 °C/min. The SPME fiber was thermally desorbed into the split/splitless injector for 5 min using a split ratio of 1:5. Helium served as the carrier gas, operating at constant flow (1.3 mL/min) with an initial head pressure of 298 kPa.

#### 4.4. GC-GC/MS Instrumentation

Heart-cutting GC-GC/MS instrumentation and parameters for tea library-building have been described in detail in our previous works [1–3,19]. Briefly, PDMS-coated stir bars (Gerstel, Mülheim an der Ruhr, Germany) were used to extract volatiles from aqueous tea infusions. A thermal desorption unit (TDU, Gerstel) was used to provide splitless transfer of the sample into the programmable temperature vaporization inlet (CIS, Gerstel), held at –100 °C. The TDU was heated from 40 °C (0.70 min) to 275 °C (3 min) at 600 °C/min under 50 mL/min helium gas flow. After 0.1 min, the CIS was heated to 275 °C at 12 °C/min and held for 5 min. The first GC (Agilent 6890, Santa Clara, CA, USA) housed column 1 (C1, 30 m  $\times$  250  $\mu\text{m}$   $\times$  0.25  $\mu\text{m}$  Rtx-Wax, Restek, Bellefonte, PA, USA) and was equipped with a flame ionization detector. The temperature of C1 was programmed from 40 °C (1 min) to 240 °C at 5 °C/min. C1 was connected to the CIS with a TDU on one end and to a five-port crosspiece (Gerstel) on the other. The second oven (Agilent 6890) housed column 2 (C2, 30 m  $\times$  250  $\mu\text{m}$   $\times$  0.25  $\mu\text{m}$  Rxi-5MS, Restek), which was connected to the crosspiece through a cryogenic freeze trap (CTS1, Gerstel) on one end and to the MS (Agilent 5975) on the other. The oven temperature was held at 40 °C for 1 min, and then increased to 280 °C at a rate of 5 °C/min. Both columns operated at 1.2 mL/min constant helium flow. The MS operating conditions were: 70 eV electron ionization source, 230 °C ion source, 150 °C quadrupole, and 40 to 250  $m/z$  scan range. A multipurpose autosampler (MPS, Gerstel) was used for automated sample injection, and a multicolumn switching device (MCS, Gerstel) supplied countercurrent flow to the crosspiece. A heart-cut was made each minute for a total of 40 heart-cuts per sample.

#### 4.5. Data Analysis

GC Image v 2.8 (LLC, Lincoln, NE, USA) was used for untargeted/targeted fingerprinting based on peak-region features and for comparative visualization of composite chromatogram pairs for high and low elevation tea samples. Two replicate preparations of each sample were used in this approach. The software identified compounds when the LRI match was within  $\pm 10$  units of published values and the NIST spectral match for that compound was  $>950$ . After the analysis, positively matched peaks were used to re-align chromatograms [47,48] and combined into a single, composite chromatogram. 2D-peak detection was performed on the composite chromatogram with peak outlines recorded as peak-region objects in the chromatographic plane. Then, all metadata belonging to positively matched peaks and peak-region objects (chemical names, retention times,  $m/z$  fragmentation pattern, retention index, and additional information about matching results) were combined into a template of features.

We used the Ion Analytics software (Gerstel, Mülheim an der Ruhr, Germany and Andover, MA, USA) as follows. First, we selected the environmental tea database to target known compounds in unprocessed tea by spectral deconvolution of the chromatogram. For each target compound, the normalized ion intensity,  $I_i(t)$  (relative to the user-defined main ion, usually the base ion, specified by  $i = 1$ ), at scan ( $t$ ) is

$$I_i(t) = \frac{A_i(t)}{R_i A_1}, \quad (1)$$

where  $A_i(t)$  is the  $i$ -th qualifier ion intensity at scan ( $t$ ) and  $R_i$  is the expected relative ion abundance ratio obtained from reference standards or NIST, Wiley, Adams, etc. All qualifier ions are normalized to the main ion ( $I_1 = 1$ ). The spectral match,  $\Delta I$ , is calculated as:

$$\Delta I = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^N \text{Abs}(I_i - I_j)}{\sum_{i=1}^{N-1} I_i}, \quad (2)$$

where  $\Delta I$  is the average normalized ion intensity deviation of each of the  $N$  specified qualifier ions. As  $\Delta I$  approaches zero, the quality of the match increases. Between three to five ions, including the main ion, were used for each compound. Another constraint is that the scan-to-scan variance,  $\Delta E$ , must be  $\leq 7$ , and is calculated as:

$$\Delta E = \Delta I \cdot \log A_1, \quad (3)$$

Additive ion signal due to coelution is eliminated by comparing all ion ratios against each other. The relative error is computed at each scan. If the ion ratio exceeds the expected ratio, and if all other ions are in agreement, the residual signal is subtracted from the matrix-affected ion. An acceptable match fits the criterion:

$$\Delta I \leq K + \frac{\Delta_0}{A_1}, \quad (4)$$

where  $K$  is the adjustable percentage difference and  $\Delta_0$  is the cumulative error from background signal and/or instrument noise. Target compound identification occurs when  $\Delta E$  or  $\Delta I$  is  $\leq 7$  in at least three consecutive peak scans. Additionally, the qualifier ion ratio deviation must be  $\leq 20\%$  at each scan across the peak to ensure consistency across the peak. The  $Q$ -value must be  $\geq 90$ . The  $Q$ -value measures compound hit quality on a scale of integers 1 to 100, where a higher value indicates a higher quality match between sample and library spectra. The  $Q$ -value is determined as the maximum of either  $100 - D$  or 1, where  $D$  is calculated as:

$$D = \frac{100 \sum_{i=1}^N |r_i^e - r_i^0| \cdot (\log(100 \cdot r_i^e + 1))^2}{21.3 \sum_{i=0}^N r_i^e}, \quad (5)$$

where  $r_i^e$  is the expected qualifier ion ratio for the  $i$ -th qualifier ion,  $r_i^0$  is the observed qualifier ion ratio for the  $i$ -th qualifier ion, and  $N$  is the number of qualifier ions. The  $Q$ -ratio must also be within  $\leq 20\%$  of the relative abundance. The  $Q$ -ratio is the peak area ratio of the extracted  $i$ -th and main ions, calculated as:

$$D = \frac{100 \sum_{i=1}^N |r_i^e - r_i^0| \cdot (\log(100 \cdot r_i^e + 1))^2}{21.3 \sum_{i=0}^N r_i^e}, \quad (6)$$

where  $S_i$  is the peak area extracted for the  $i$ -th common ion over the hit and  $S_1$  is the peak area for the main ion over the hit. Since GC×GC produces multiple modulated peaks per compound, the deconvolution algorithm searches for up to five peaks per compound by default, which can be increased when high concentration analytes are found in the sample.

Then, the mass spectrum of each target compound was subtracted from corresponding peak scans to reveal untargeted compounds. Peak assignments were made by matching sample and library spectra and retention indices, which were then confirmed using available reference standards. If assignments were not possible, numerical identifiers were used to assign peak names. Finally, once all peak assignments were made, the mass spectrum for each compound was subtracted from the TIC chromatogram to reveal missed peaks. This was established when residual signals approximated the baseline (background noise) of the chromatogram.

## 5. Conclusions

Our overall objective is to demonstrate two complimentary data analysis approaches to differentiating complex samples using tea as the model system. First, we used molecular feature detection to show which regions of the chromatograms differed, then untargeted/targeted analysis to identify all compounds in the sample, both in regions of the chromatogram that differed significantly and where it did not. The latter showed the quantifiable differences at the molecular level in high and low elevation tea. From a sensory perspective, our findings point to compounds consistent with local perspectives of quality in high and low elevation green tea. In this paper, we used comprehensive GC×GC/MS, rather than GC/MS or GC-GC/MS, to produce data on the same time scale as GC/MS with the resolution of GC-GC/MS, without taking days to produce data on one sample. The technology and approach taken differentiates complex samples at the molecular level, critically important to the study of systems biology.

**Supplementary Materials:** The following are available online, Figure S1: Ion Analytics and GC Image workflow schematic for the analysis of targeted and untargeted compounds, Table S1: Features detected by GC Image and confirmatory analysis, Table S2: Compounds identified by Ion Analytics. The unprocessed GC×GC/MS data can be found at: <http://dx.doi.org/10.25833/jxyj-8w58>.

**Author Contributions:** Conceptualization, A.R., C.C. and C.B.; methodology, A.R., C.C., and C.B.; software, A.R. and N.K.; validation, J.M., M.C.R. and C.C.; formal analysis, J.M. and M.C.R.; investigation, M.C.R. and N.K.; resources, A.R., C.C., and C.B.; data curation, J.M. and M.C.R.; writing—original draft preparation, J.M. and A.R.; writing—review and editing, J.M., A.R., N.K., M.C.R., and C.C.; visualization, J.M., A.R., M.C.R., and C.C.; supervision, A.R., C.C., and C.B.; project administration, A.R.; funding acquisition, A.R.

**Funding:** This work did not receive any specific funding from public, commercial, or not-for-profit agencies.

**Acknowledgments:** We thank Stephen Reichenbach and Qingping Tao for providing GC Image support.

**Conflicts of Interest:** A.R. is the founder of Ion Analytics and developer of the software.

## References

1. Kfoury, N.; Morimoto, J.; Kern, A.; Scott, E.R.; Orians, C.M.; Ahmed, S.; Griffin, T.; Cash, S.B.; Stepp, J.R.; Xue, D.Y.; et al. Striking changes in tea metabolites due to elevational effects. *Food Chem.* **2018**, *264*, 334–341. [[CrossRef](#)] [[PubMed](#)]
2. Kowalsick, A.; Kfoury, N.; Robbat, A., Jr.; Ahmed, S.; Orians, C.; Griffin, T.; Cash, S.B.; Stepp, J.R. Metabolite profiling of *Camellia sinensis* by automated sequential, multidimensional gas chromatography/mass spectrometry reveals strong monsoon effects on tea constituents. *J. Chromatogr. A* **2014**, *1370*, 230–239. [[CrossRef](#)] [[PubMed](#)]
3. Robbat, A.; Kfoury, N.; Baydakov, E.; Gankin, Y. Optimizing targeted/untargeted metabolomics by automating gas chromatography/mass spectrometry workflows. *J. Chromatogr. A* **2017**, *1505*, 96–105. [[CrossRef](#)]
4. Cordero, C.; Liberto, E.; Bicchì, C.; Rubiolo, P.; Reichenbach, S.E.; Tian, X.; Tao, Q.P. Targeted and Non-Targeted Approaches for Complex Natural Sample Profiling by GC×GC-qMS. *J. Chromatogr. Sci.* **2010**, *48*, 251–261. [[CrossRef](#)]
5. Skogerson, K.; Wohlgemuth, G.; Barupal, D.K.; Fiehn, O. The volatile compound BinBase mass spectral database. *BMC Bioinf.* **2011**, *12*, 321. [[CrossRef](#)]
6. Sumner, L.W.; Amberg, A.; Barrett, D.; Beale, M.H.; Beger, R.; Daykin, C.A.; Fan, T.W.M.; Fiehn, O.; Goodacre, R.; Griffin, J.L.; et al. Proposed minimum reporting standards for chemical analysis. *Metabolomics* **2007**, *3*, 211–221. [[CrossRef](#)]
7. Creek, D.J.; Dunn, W.B.; Fiehn, O.; Griffin, J.L.; Hall, R.D.; Lei, Z.; Mistrík, R.; Neumann, S.; Schymanski, E.L.; Sumner, L.W.; et al. Metabolite identification: are you sure? And how do your peers gauge your confidence? *Metabolomics* **2014**, *10*, 350–353. [[CrossRef](#)]
8. Sumner, L.W.; Lei, Z.; Nikolau, B.J.; Saito, K.; Roessner, U.; Trengove, R. Proposed quantitative and alphanumeric metabolite identification metrics. *Metabolomics* **2014**, *10*, 1047–1049. [[CrossRef](#)]
9. Alseekh, S.; Fernie, A.R. Metabolomics 20 years on: What have we learned and what hurdles remain? *Plant J.* **2018**, *94*, 933–942. [[CrossRef](#)]

10. Mahieu, N.G.; Patti, G.J. Systems-Level Annotation of a Metabolomics Data Set Reduces 25,000 Features to Fewer than 1000 Unique Metabolites. *Anal. Chem.* **2017**, *89*, 10397–10406. [[CrossRef](#)]
11. Tautenhahn, R.; Bottcher, C.; Neumann, S. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinf.* **2008**, *9*, 504. [[CrossRef](#)] [[PubMed](#)]
12. Stolt, R.; Torgrip, R.J.O.; Lindberg, J.; Csenki, L.; Kolmert, J.; Schuppe-Koistinen, I.; Jacobsson, S.P. Second-order peak detection for multicomponent high-resolution LC/MS data. *Anal. Chem.* **2006**, *78*, 975–983. [[CrossRef](#)] [[PubMed](#)]
13. da Silva, R.R.; Dorrestein, P.C.; Quinn, R.A. Illuminating the dark matter in metabolomics. *P. Natl. Acad. Sci. USA* **2015**, *112*, 12549–12550. [[CrossRef](#)] [[PubMed](#)]
14. Smith, C.A.; Want, E.J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: Processing mass spectrometry data for metabolite profiling using Nonlinear peak alignment, matching, and identification. *Anal. Chem.* **2006**, *78*, 779–787. [[CrossRef](#)]
15. Weber, R.J.M.; Lawson, T.N.; Salek, R.M.; Ebbels, T.M.D.; Glen, R.C.; Goodacre, R.; Griffin, J.L.; Haug, K.; Koulman, A.; Moreno, P.; et al. Computational tools and workflows in metabolomics: An international survey highlights the opportunity for harmonisation through Galaxy. *Metabolomics* **2016**, *13*, 12. [[CrossRef](#)]
16. Papadimitropoulos, M.-E.P.; Vasilopoulou, C.G.; Maga-Nteve, C.; Klapa, M.I. Untargeted GC-MS Metabolomics. In *Metabolic Profiling: Methods and Protocols*; Theodoridis, G.A., Gika, H.G., Wilson, I.D., Eds.; Springer: New York, NY, USA, 2018; pp. 133–147.
17. Cho, K.; Mahieu, N.G.; Johnson, S.L.; Patti, G.J. After the feature presentation: technologies bridging untargeted metabolomics and biology. *Curr. Opin. Biotechnol.* **2014**, *28*, 143–148. [[CrossRef](#)]
18. Reichenbach, S.E.; Tian, X.; Cordero, C.; Tao, Q. Features for non-targeted cross-sample analysis with comprehensive two-dimensional chromatography. *J. Chromatogr. A* **2012**, *1226*, 140–148. [[CrossRef](#)]
19. Kfoury, N.; Baydakov, E.; Gankin, Y.; Robbat, A. Differentiation of key biomarkers in tea infusions using a target/nontarget gas chromatography/mass spectrometry workflow. *Food Res. Int.* **2018**, *113*, 414–423. [[CrossRef](#)]
20. Han, W.Y.; Huang, J.G.; Li, X.; Li, Z.X.; Ahammed, G.J.; Yan, P.; Stepp, J.R. Altitudinal effects on the quality of green tea in east China: a climate change perspective. *Eur. Food Res. Technol.* **2017**, *243*, 323–330. [[CrossRef](#)]
21. Chen, G.-H.; Yang, C.-Y.; Lee, S.-J.; Wu, C.-C.; Tzen, J.T.C. Catechin content and the degree of its galloylation in oolong tea are inversely correlated with cultivation altitude. *J. Food Drug Anal.* **2014**, *22*, 303–309. [[CrossRef](#)]
22. Ahmed, S.; Stepp, J.R.; Orians, C.; Griffin, T.; Matyas, C.; Robbat, A.; Cash, S.; Xue, D.Y.; Long, C.L.; Unachukwu, U.; et al. Effects of Extreme Climate Events on Tea (*Camellia sinensis*) Functional Quality Validate Indigenous Farmer Knowledge and Sensory Preferences in Tropical China. *PLoS ONE* **2014**, *9*, e109126. [[CrossRef](#)] [[PubMed](#)]
23. Chen, I.J.; Liu, C.-Y.; Chiu, J.-P.; Hsu, C.-H. Therapeutic effect of high-dose green tea extract on weight reduction: A randomized, double-blind, placebo-controlled clinical trial. *Clin. Nutr.* **2016**, *35*, 592–599. [[CrossRef](#)] [[PubMed](#)]
24. Tsao, A.S.; Liu, D.; Martin, J.; Tang, X.-M.; Lee, J.J.; El-Naggar, A.K.; Wistuba, I.; Culotta, K.S.; Mao, L.; Gillenwater, A.; et al. Phase II Randomized, Placebo-Controlled Trial of Green Tea Extract in Patients with High-Risk Oral Premalignant Lesions. *Cancer Prev. Res.* **2009**, *2*, 931–941. [[CrossRef](#)] [[PubMed](#)]
25. Gee, J.R.; Saltzstein, D.R.; Kim, K.; Kolesar, J.; Huang, W.; Havighurst, T.C.; Wollmer, B.W.; Stublaski, J.; Downs, T.; Mukhtar, H.; et al. A Phase II Randomized, Double-blind, Presurgical Trial of Polyphenon E in Bladder Cancer Patients to Evaluate Pharmacodynamics and Bladder Tissue Biomarkers. *Cancer Prev. Res.* **2017**, *10*, 298–307. [[CrossRef](#)]
26. Samavat, H.; Ursin, G.; Emory, T.H.; Lee, E.; Wang, R.; Torkelson, C.J.; Dostal, A.M.; Swenson, K.; Le, C.T.; Yang, C.S.; et al. A Randomized Controlled Trial of Green Tea Extract Supplementation and Mammographic Density in Postmenopausal Women at Increased Risk of Breast Cancer. *Cancer Prev. Res.* **2017**, *10*, 710–718. [[CrossRef](#)]
27. Battinelli, L.; Daniele, C.; Cristiani, M.; Bisignano, G.; Saija, A.; Mazzanti, G. In vitro antifungal and anti-elastase activity of some aliphatic aldehydes from *Olea europaea* L. fruit. *Phytomedicine* **2006**, *13*, 558–563. [[CrossRef](#)]
28. Tse, G.; Yeo, J.M.; Tse, V.; Kwan, J.; Sun, B. Gap junction inhibition by heptanol increases ventricular arrhythmogenicity by reducing conduction velocity without affecting repolarization properties or myocardial refractoriness in Langendorff-perfused mouse hearts. *Mol. Med. Rep.* **2016**, *14*, 4069–4074. [[CrossRef](#)]

29. Kumar, D.; Kumar, S. Isolation and Characterization of Bioactive Phenolic Compounds from Abies Pindrow Aerial Parts. *Pharm. Chem. J.* **2017**, *51*, 205–210. [[CrossRef](#)]
30. Shiratsuchi, H.; Chang, S.; Wei, A.; El-Ghorab, A.H.; Shibamoto, T. Biological activities of low-molecular weight compounds found in foods and plants. *J. Food Drug Anal.* **2012**, *20*, 359–365.
31. Iscan, G. Antibacterial and Anticandidal Activities of Common Essential Oil Constituents. *Rec. Nat. Prod.* **2017**, *11*, 374–388.
32. Miron, D.; Battisti, F.; Silva, F.K.; Lana, A.D.; Pippi, B.; Casanova, B.; Gnoatto, S.; Fuentefria, A.; Mayorga, P.; Schapoval, E.E.S. Antifungal activity and mechanism of action of monoterpenes against dermatophytes and yeasts. *Rev. Bras. Farm.* **2014**, *24*, 660–667. [[CrossRef](#)]
33. González-Ramírez, A.E.; González-Trujano, M.E.; Orozco-Suárez, S.A.; Alvarado-Vásquez, N.; López-Muñoz, F.J. Nerol alleviates pathologic markers in the oxazolone-induced colitis model. *Eur. J. Pharmacol.* **2016**, *776*, 81–89. [[CrossRef](#)] [[PubMed](#)]
34. de Lima, A.B.; Santana, M.B.; Cardoso, A.S.; da Silva, J.K.R.; Maia, J.G.S.; Carvalho, J.C.T.; Sousa, P.J.C. Antinociceptive activity of 1-nitro-2-phenylethane, the main component of Aniba canelilla essential oil. *Phytomedicine* **2009**, *16*, 555–559. [[CrossRef](#)] [[PubMed](#)]
35. Muroi, H.; Kubo, I. Combination Effects of Antibacterial Compounds in Green Tea Flavor Against *Streptococcus-mutans*. *J. Agric. Food Chem.* **1993**, *41*, 1102–1105. [[CrossRef](#)]
36. Tong, Q.S.; Jiang, G.S.; Zheng, L.D.; Tang, S.T.; Cai, J.B.; Liu, Y.; Zeng, F.Q.; Dong, J.H. Natural jasmonates of different structures suppress the growth of human neuroblastoma cell line SH-SY5Y and its mechanisms. *Acta Pharmacol. Sin.* **2008**, *29*, 861–869. [[CrossRef](#)]
37. Choi, S.J.; Kim, J.K.; Kim, H.K.; Harris, K.; Kim, C.-J.; Park, G.G.; Park, C.-S.; Shin, D.-H. 2,4-Di-tert-butylphenol from Sweet Potato Protects Against Oxidative Stress in PC12 Cells and in Mice. *J. Med. Food* **2013**, *16*, 977–983. [[CrossRef](#)]
38. Afoulous, S.; Ferhout, H.; Raelison, E.G.; Valentin, A.; Moukarzel, B.; Couderc, F.; Bouajila, J. Helichrysum gymnocephalum Essential Oil: Chemical Composition and Cytotoxic, Antimalarial and Antioxidant Activities, Attribution of the Activity Origin by Correlations. *Molecules* **2011**, *16*, 8273–8291. [[CrossRef](#)]
39. Takei, M.; Umeyama, A.; Arihara, S. T-cadinol and calamenene induce dendritic cells from human monocytes and drive Th1 polarization. *Eur. J. Pharmacol.* **2006**, *537*, 190–199. [[CrossRef](#)]
40. Alves Naiane Ferraz, B.; Queiroz Thyago, M.; Almeida Travassos, R.; Magnani, M.; Andrade Braga, V. Acute Treatment with Lauric Acid Reduces Blood Pressure and Oxidative Stress in Spontaneously Hypertensive Rats. *Basic Clin. Pharmacol.* **2016**, *120*, 348–353. [[CrossRef](#)]
41. Wang, J.; Lu, J.; Xie, X.; Xiong, J.; Huang, N.; Wei, H.; Jiang, S.; Peng, J. Blend of organic acids and medium chain fatty acids prevents the inflammatory response and intestinal barrier dysfunction in mice challenged with enterohemorrhagic Escherichia coli O157:H7. *Int. Immunopharmacol.* **2018**, *58*, 64–71. [[CrossRef](#)]
42. Nguyen, L.T.; Myslivečková, Z.; Szotáková, B.; Špičáková, A.; Lněničková, K.; Ambrož, M.; Kubíček, V.; Krasulová, K.; Anzenbacher, P.; Skálová, L. The inhibitory effects of  $\beta$ -caryophyllene,  $\beta$ -caryophyllene oxide and  $\alpha$ -humulene on the activities of the main drug-metabolizing enzymes in rat and human liver in vitro. *Chem. Biol. Interact.* **2017**, *278*, 123–128. [[CrossRef](#)] [[PubMed](#)]
43. Guerrini, A.; Sacchetti, G.; Grandini, A.; Spagnoletti, A.; Asanza, M.; Scalvenzi, L. Cytotoxic Effect and TLC Bioautography-Guided Approach to Detect Health Properties of Amazonian Hedyosmum sprucei Essential Oil. *Evid. Based Complement. Altern. Med.* **2016**, *8*. [[CrossRef](#)]
44. Rossi, D.; Guerrini, A.; Maietti, S.; Bruni, R.; Paganetto, G.; Poli, F.; Scalvenzi, L.; Radice, M.; Saro, K.; Sacchetti, G. Chemical fingerprinting and bioactivity of Amazonian Ecuador *Croton lechleri* Mull. Arg. (Euphorbiaceae) stem bark essential oil: A new functional food ingredient? *Food Chem.* **2011**, *126*, 837–848. [[CrossRef](#)]
45. Tung, Y.T.; Yen, P.L.; Lin, C.Y.; Chang, S.T. Anti-inflammatory activities of essential oils and their constituents from different provenances of indigenous cinnamon (*Cinnamomum osmophloeum*) leaves. *Pharm. Biol.* **2010**, *48*, 1130–1136. [[CrossRef](#)] [[PubMed](#)]
46. Pimenta, R.S.; Silva, J.F.M.D.; Buyer, J.S.; Jansiewicz, W.J. Endophytic Fungi from Plums (*Prunus domestica*) and Their Antifungal Activity against *Monilinia fructicola*. *J. Food Prot.* **2012**, *75*, 1883–1889. [[CrossRef](#)] [[PubMed](#)]

47. Reichenbach, S.E.; Tian, X.; Boateng, A.A.; Mullen, C.A.; Cordero, C.; Tao, Q.P. Reliable Peak Selection for Multisample Analysis with Comprehensive Two-Dimensional Chromatography. *Anal. Chem.* **2013**, *85*, 4974–4981. [[CrossRef](#)]
48. Rempe, D.W.; Reichenbach, S.E.; Tao, Q.P.; Cordero, C.; Rathbun, W.E.; Zini, C.A. Effectiveness of Global, Low-Degree Polynomial Transformations for GCxGC Data Alignment. *Anal. Chem.* **2016**, *88*, 10028–10035. [[CrossRef](#)]

**Sample Availability:** Samples of the compounds are not available from the authors.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).