

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Chromatographic Fingerprinting Strategy to Delineate Chemical Patterns Correlated to Coffee Odor and Taste Attributes

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1799264> since 2025-01-22T14:01:36Z

Published version:

DOI:10.1021/acs.jafc.1c00509

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

Chromatographic Fingerprinting Strategy to Delineate Chemical Patterns Correlated to Coffee

Odor and Taste attributes

Bressanello D.^a, Marengo A.^a, Cordero C.^a, Strocchi G.^a, Rubiolo P.^a, Pellegrino G.^b, Ruosi M.R.^b,
Bicchi C.^a, Liberto E.^{a*}

^aDipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino, Via Pietro Giuria
9, Turin, Italy

^b Lavazza S.p.A., Strada Settimo 410, Turin, Italy

**Corresponding author:*

Tel: +39 011 6707134,

Fax: +39 011 6707687;

e-mail: erica.liberto@unito.it

1 **ABSTRACT**

2 Coffee cupping includes both aroma and taste, and its evaluation considers several different
3 attributes simultaneously to define flavor quality therefore requiring complementary data from
4 aroma and taste. This study investigates the potential and limits of a data-driven approach to
5 describe the sensory quality of coffee using complementary analytical techniques usually available
6 in routinely quality control laboratory. Coffee flavor chemical data from 155 samples were obtained
7 by analyzing volatile (HS-SPME-GC-MS), and non-volatile (HPLC-UV/DAD) fractions, as well as from
8 sensory data. Chemometric tools were used to explore the data sets, select relevant features,
9 predict sensory scores and investigate the networks between features. A comparison of the Q
10 model parameter and RMSEP highlights the variable influence that the non-volatile fraction has on
11 prediction, showing that it has a higher impact on describing *Acid*, *Bitter* and *Woody* notes than on
12 *Flowery* and *Fruity*. The data fusion emphasized the aroma contribution to driving sensory
13 perceptions, although the correlative networks highlighted from the volatile and non-volatile data
14 deserves a thorough investigation to verify the potential of odor-taste integration.

15

16

17 **Keywords:** Coffee, HPLC-UV, HS-SPME-GC-MS, chemometrics, sensory data

18 INTRODUCTION

19 The characteristics and sensory properties of coffee flavor are unique and of high appeal for
20 consumers.¹ The pleasure system includes different brain areas that are linked to emotional,
21 memory-related, motivational and linguistic aspects of food evaluation that are, in turn, mediated
22 by several sensory modalities and sub-modalities that contribute to flavor perception.¹⁻³

23 Flavor perception is therefore a complex sensation given by the interaction between aroma,
24 perceived through the sense of smell [orthonasally and retronasally (food aroma)] and taste,
25 perceived at the level of the oral cavity.^{3,4} However, the sense of smell often dominates flavor
26 perception, and aroma active compounds can modulate taste intensity.^{1,3,5} Cross-modal interactions
27 are therefore fundamental to delineating the hedonic profile of a food, and this has become a
28 strategic route for industry in designing new food products and healthier formulations (e.g. with
29 less salt or sugars).⁶⁻¹¹ On the other hand, taste-active compounds may influence the release of
30 volatiles and impact upon their actual concentration in the headspace.⁶

31 The evaluation of coffee flavor is a key step of the production chain; from raw-material selection to
32 the creation of new valuable blends. Coffee sensory quality is nowadays rated by cupping protocols
33 that intrinsically satisfy the multimodal-perception concept. The sensory panel therefore plays a
34 fundamental role in this respect.¹² The cupping protocol is however time-consuming, requires
35 properly trained and aligned professional panelists, and may suffer from subjectivity. The ever-
36 increasing consumption of coffee highlights the need for analytical techniques at supporting sensory
37 panel evaluation in routine quality controls (QC) or formulation and design of new blends. Suitable
38 instrumental methods should therefore: i) provide reliable information on coffee aroma and taste,
39 including their possible interactions; ii) support a consistent prediction of key-sensory attributes in
40 line with sensory panel outcomes; and iii) inform, at a molecular level, on the presence of key-
41 analytes patterns within a quality control laboratory routine.

42 Several analytical approaches and/or integrated strategies attempted, over the years, to combine
43 the chemical composition of a product with its flavor profile^{13–22}; however, to date, they have not
44 replaced a sensory panel evaluation, especially in regulatory contexts where sensory quality concur
45 to define labelling (e.g., extra-virgin olive oil^{22,23}) or a commercial value (e.g., coffee^{13,15,17,18,21,24–26}).

46 The limited application of instrumental tools in the industrial context might be explained taking in
47 consideration the breath and reliability – and thereby rate of success - of several approaches
48 modelled on set of samples with low/limited representativeness or trained by *a priori* defined lists
49 of targeted chemical variables^{15,19,24}. Issues related to their applicability might therefore be related
50 to the attempt of reducing the extremely complex phenomenon of sensory perception, triggered by
51 multiple ligands (volatiles and non-volatiles) and modulated by their cross-modal interactions, to a
52 few correlations (e.g., reductionist approach²⁷) or to the adoption of spectroscopic/spectrometric
53 technologies that have limited or not univocal “molecular resolution”. A further point is the natural
54 and reasonable skepticism of quality control decision makers to abandon, or replace, a
55 normalized/established protocol in favor of alternative procedures.

56 In this context, modern omics disciplines dealing with food (sensomics²⁸, flavoromics²⁹ and food
57 metabolomics³⁰) can be of help suggesting more systemic approaches to the chemical interpretation
58 of complex biological phenomena by untargeted investigations (e.g., integrationist approach²⁷).

59 Many research provides proof-of-evidence on the potential of applying omics workflows and
60 concepts to identify “features patterns” (i.e. patterns of potential informative components) with
61 high correlation to a biological output.³¹ Very recently, Nicolotti et al. ³² conceptually validated an
62 “artificial intelligence smelling machine”, an analytical workflow based on sensomics, that attempts
63 to simulate human olfaction by accurately define key-odorants patterns responsible of the aroma
64 of a food.

65 Moreover, machine learning applied to fingerprinting and/or profiling technologies highlighted
66 strong relationships and networking between aroma and flavor. In a meta-analysis study on various
67 food aroma and flavor, Dunkel *et al.*²⁸ evidenced several groups of odorants, validated by sensomics,
68 with specific associations to odor notes, and showed that the networking of odor notes might open
69 up possibilities in the exploration of these associations. Tromelin *et al.*³³ found potential similarities
70 and links between odorant and odor spaces using a multivariate-driven approach on a large odorant
71 database, while Wang *et al.*³⁴ confirmed that odorant and non-odorant compounds interact in the
72 expression of a perceived sensory attribute. Very recently, Guichard *et al.*³⁵ investigated odor-taste
73 networks in commercial multi-fruit juices using cheminformatics, and showed that network
74 visualization link between odor (green, grass, vegetal) and taste (bitterness) descriptors had strong
75 associations. Barba *et al.*³⁶ demonstrated that odorants enhancing targeted taste perception might
76 be exploited to modulate overall taste profile in foods and beverages.

77 In this complex and intriguing scenario, starting by preliminary results obtained by applying *omics*
78 principles to the modelling of specific coffee aroma notes¹⁵, this study is a step forward in evaluating
79 chromatographic fingerprints of volatile and non-volatile components as diagnostic signatures with
80 strong correlation with selected taste and aroma attributes, i.e. *bitterness, acidity, flowery, fruity,*
81 *woody, and spicy.*

82 Moreover, fingerprinting is combined to machine learning, by partial least squares (PLS) algorithms,
83 and extended to a comprehensive data matrix obtained by combining together peak features
84 information deriving from volatiles and non-volatiles. PLS drives features selection toward those
85 informative patterns capable of predicting sensory attributes and explaining correlations between
86 them. Analytical platforms for fingerprinting were selected in light of routine control laboratories
87 requirements for high batch-to-batch reproducibility, separation efficiency and confirmatory
88 potentials. Selected coffee powders were analyzed for their volatile fingerprints by headspace solid

89 phase micro extraction followed by gas chromatography mass spectrometric detection (HS-SPME-
90 GC-MS), and for their non-volatile fingerprints from the corresponding brews analyzed by liquid
91 chromatography with UV/DAD detection (LC-UV/DAD) to simulate the main phases of the cupping
92 protocol according to the Specialty Coffee Association (SCA)³⁷ that evaluates both smell and taste,
93 and here used to describe samples sensory notes.

94

95 **MATERIALS AND METHODS**

96 **Samples and chemicals**

97 Samples, consisting of roasted and ground coffees to suit a coffee-filter machine, were kindly
98 supplied over a period of 24 months by Lavazza Spa (Turin, Italy). The grinder was a Superjolly
99 grinder with a stepless micrometrical grinding adjustment by Mazzer (Venice, Italy), the particle
100 size average was: $425\ \mu\text{m} \pm 75\ \mu\text{m}$ and percentages dust (% of particles below $100\ \mu\text{m}$): $13\% \pm 3\%$.
101 Mono-origin samples from different countries were selected for their distinctive and peculiar
102 sensory notes, they accounted for a total of 155 samples belonging to *Coffea Arabica* L. (Arabica,
103 $n= 85$) and *Coffea canephora* Pierre ex- A. Froehner (Robusta, $n= 70$) species. Supplementary Table
104 1 (Table S1) reports details on all analyzed samples. The roasting was carried out on a laboratory
105 Probat BRZ2 drum roaster (Emmerich am Rhein, Germany) by applying optimized protocols. A 150
106 g of coffee beans were roasted at 200°C for 8-12 min until reaching a color of 55°Nh (Neuhaus
107 degrees) in line with the international standardization protocol for cupping (SCAA protocol).³⁷
108 Coffee color was carefully measured by ground-bean light reflectance, with a single-beam Neuhaus
109 Neotec Color Test II instrument (Genderkese, Germany) at a wavelength of 900 nm. Samples were
110 roasted no more than 24 hours prior to cupping and left for at least 8 hours to stabilize as indicated
111 by the protocol.³⁷

112 The coffee brew for cupping and analysis was prepared from 18 g of coffee powder and 300 mL of
113 water at 88-94°C with a commercially available coffee filter machine Xlong TSK-197A (Lavazza Spa,
114 Turin, Italy). Two milliliters of brew were then filtered using a 0.2 µm 13 mm nylon membrane
115 syringe filter (Agilent, Little Falls, DE, USA) and 20 µL were directly injected for LC-UV/DAD analysis.
116 LC-grade acetonitrile (LC-MS grade) and formic acid (>98% purity) were obtained from Merck while
117 de-ionized water (18.2 MΩ cm) was obtained from a Milli-Q purification system (Millipore, Bedford,
118 MA, USA).
119 Cryptochlorogenic acid; 3,5-dicaffeoylquinic acid and 4,5-dicaffeoylquinic acid were obtained from
120 Phytolab (Vestenbergsgreuth, Germany). Chlorogenic acid, neochlorogenic acid, 3,4-
121 dicaffeoylquinic acid, trigonelline and caffeine were obtained from Sigma Aldrich (Bellefonte, USA).
122 Normal alkanes, ISTD (*n*-C13), dibutyl phthalate and reference compounds for identity confirmation
123 of volatiles reported in Table 2S (a) were all from Merk (Milan, Italy).

124

125 **Sensory analysis of coffee samples**

126 Cup quality was assessed for several flavor attributes: *Acid*, *Bitter*, *Flowery*, *Fruity*, *Spicy*, and *Woody*
127 by an external trained panel of six assessors. The intensities of each attribute were evaluated
128 simultaneously on a scale from 0 to 10. The sensory data provided by the external panel had already
129 been verified at the origin by the ANOVA analysis with a *post-hoc* test. Average scores from the
130 panel were used for the investigated attributes.

131 Coffee sensory properties were evaluated both by sniffing the powder and the brew obtained using
132 the filter method, and by tasting aspiring the beverage into the mouth.³⁷ This multistep protocol
133 allows panelists to evaluate different attributes, with some being more closely linked to aroma
134 (sensory notes like *flowery*, *fruity*, *woody* and *spicy*) and others more closely to taste (*acidity* and
135 *bitterness*).

136 **Volatile fingerprints: sampling and analysis conditions**

137 The fingerprint corresponding to the volatile fraction of coffee, including also aroma active
138 compounds, was obtained on dry roasted and ground coffee powders by HS-SPME followed by GC-
139 MS analysis. SPME sampling was performed by a Combi-PAL AOC 5000 (Shimadzu, Milan, Italy) with
140 a Polydimethylsiloxane/Divinylbenzene (PDMS/DVB) fiber of d_f 65 μm and 1 cm length from Merck
141 (Bellefonte, PA, USA). Fiber selection was based on the results about profiles representativeness
142 obtained in a previous study³⁸ while its conditioning was performed as recommended by the
143 manufacturer.

144 Coffee samples (1.50 g of fine and homogeneous powder) were accurately weighed in headspace
145 vials (20 mL) and immediately sealed after the operation. Headspace sampling was performed for
146 40 minutes at 50°C vibrated at a constant speed. The internal standard was pre-loaded onto the
147 fiber by sampling 5 μL of a 1000 mg/L solution of *n*-C13 in dibutyl phthalate (DBP) placed in a 20 mL
148 headspace vial kept for 20 min at 50°C at a constant speed. After sampling, the analytes were
149 recovered via the thermal desorption of the fiber, for 5 min at 250°C, into the GC injector. All
150 samples were analyzed in triplicate.

151 Analysis was performed with a Shimadzu QP2010 GC-MS system equipped with Shimadzu GC-MS
152 Solution 2.51 software (Shimadzu, Milan, Italy). Chromatographic conditions: injector temperature:
153 250°C; injection mode: splitless; carrier gas: helium at a flow rate of 1 mL/min. Capillary column:
154 SGE SolGelwax (100% polyethylene glycol) 30 m x 0.25 mm d_c x 0.25 μm d_f (Trajan Scientific and
155 Medical, Melbourne, Australia). Temperature program, from 40°C (1 min) to 200°C at 3°C/min, then
156 to 250°C (5 min) at 10°C/min. MS conditions: ionization mode: EI (70 eV); temperatures: ion source
157 at 200°C; transfer line at 250°C. Scan range: 35-350 m/z ; scan speed 666 amu/sec.

158 Analytes identification was performed using linear retention indices (I^T_s) and EI-MS spectrum that
159 were either compared to those of authentic standards, to those collected in-house or in commercial
160 libraries (Wiley 7N and NIST 14 Mass Spectral Data).

161

162 **Non-volatile fingerprint: analysis conditions**

163 The non-volatile fraction was analyzed using a LC-UV/DAD system, Model 1200 Agilent, Little Falls,
164 DE, USA), equipped with a Spectra System UV Diode Array Detector 1100 series (Agilent, Little Falls,
165 DE, USA). Data acquisition and data handling were performed by Chemstation LC 3D software
166 Rev.3.03 01-SR1 (Agilent, Little Falls, USA). The LC column was a Platinum EPS C18 (250 × 4.6 mm,
167 80A, 4 μm) (Alltech, Deerfield, USA).

168 LC operative conditions: injection volume 20 μL; mobile phase: A: water/formic acid (999:1, v/v) B:
169 acetonitrile/formic acid (999:1, v/v); flow rate, 1.0 mL/min. The gradient program was as follows:
170 15% B for 7 min, 15-55% B in 20 min, 55-100% B in 25 min, 100% B for 2 min. Before re-injection,
171 the LC system was stabilized for at least 5 min. The UV/DAD acquired within the wavelength range
172 210-600 nm and at a 2.5 spectra/sec. Acquisition wavelengths were 276 and 325 nm.

173 Compounds identity confirmation and putative identifications were carried out on a LC-MS/MS
174 system consisting of a Shimadzu Nexera X2 unit equipped with a photodiode detector SPD-M20A
175 connected, in series, to a triple quadrupole Shimadzu LCMS-8040 MS system equipped with an
176 electrospray ionization (ESI) source (Shimadzu, Dusseldorf Germany).

177 The separation column was an Ascentis Express C18 (15 cm x 2.1 mm, 2.7μm) from Supelco
178 (Bellefonte, USA). Operative conditions: injection volume, 5 μL; mobile phases, A: water/formic acid
179 (999:1, v/v), B: acetonitrile/formic acid (999:1, v/v); flow rate, 0.4 mL/min. Mobile-phase program:
180 15% B for 7 min, 15-55% B in 3 min, 55-100% B in 1.5 min, 100% B for 1 min, total pre-running and
181 post-running time 23 min. UV/DAD detection for profiles monitoring was set within the wavelength

182 range 220-450 nm. MS operative conditions were as follows: heat-block temperature: 200°C;
183 desolvation line (DL) temperature: 250°C; nebulizer gas flow rate: 3 L/min; drying gas flow rate: 15
184 L/min. Mass spectra were acquired both in positive and negative full-scan modes over the 100-1000
185 m/z range at an event time of 0.5 sec. Product Ion Scan mode (collision energy: - 35.0 V for ESI+ and
186 35.0 V for ESI-, event time: 0.2 sec) was applied to compounds for which a correspondence between
187 the pseudomolecular ions $[M+H]^+$ in ESI+ and $[M-H]^-$ in ESI- had been confirmed.

188

189 **Data Processing**

190 Data sets (Sensory scores, GC-MS and LC-UV/DAD fingerprints) were explored by Principal
191 Component Analysis (PCA) followed by Multiple Factorial Analysis (MFA). This statistical tool enables
192 to investigate the relationships between chromatographic fingerprints from the two analytical
193 platforms and to compare them with the sensory data. Features selection for each data set, related
194 to the sensory note, was performed using VIP>1 (variable importance in projection) from Partial
195 Least Square-Discriminant Analysis (PLS-DA) on samples that were suitable to minimize and
196 maximize the sensory notes expression. This procedure was used to select peak features that were
197 either used afterward to model, by PLS regression, single perception modalities (*i.e.*, aroma or taste)
198 or flavor *in toto* by combining the two data set (*i.e.*, volatiles and non-volatiles fingerprints. Data
199 elaboration was performed using XLSTAT statistical and data analysis solution software (version
200 2020.1.3 - copyright Addinsoft 2020).

201 Raw analytical data representing the chromatographic fingerprints of volatiles and non-volatiles
202 were pre-processed, as shown in the work-flow of **Figure 1**, via the temporal alignment of
203 chromatograms and background noise subtraction. This pre-processing was made with Pirouette
204 software ver. 4.5 (Infometrix, Inc., Bothell, WA, USA). Raw signals (GC and LC) were converted into
205 an array $X (I \times J)$ where I corresponds to detector intensities and J to corresponding retention times.

206 Replicated analyses were treated independently. Therefore, each analytical platform provided a
207 tensor that included raw signals arrays for all the analyzed samples.^{39,40} The output was a table
208 listing, in the columns, all detected features together corresponding response and rows reporting
209 sample analytical replicates. A response threshold was set to filter out noise peaks, it corresponded
210 to 3 times the standard deviation of signal-to-noise ratio (S/N) values sampled at different time
211 points with both techniques. The S/N for GC-MS was set at 15 and for LC-UV/DAD at 8 and 10
212 respectively for 276 and 325 nm.

213 Data then underwent unsupervised multivariate analysis to highlight, if present, diagnostic patterns,
214 and were next treated with supervised methods (PLS-DA) to select features that were most related
215 to each sensory attribute. Selected features were then used to evaluate the ability to predict sensory
216 scores for the investigated attributes using PLS models.⁴¹ The instrumental data from both the
217 volatile and non-volatile fractions were first independently elaborated, and then combined/fused
218 in a single data matrix to investigate their contribution and/or ability in predicting selected sensory
219 attributes.

220 The models based on HS-SPME-GC-MS fingerprint, on HPLC-UV/DAD fingerprint and those obtained
221 by elaborating the fused data, were evaluated and compared on the basis of their model quality
222 index (Q^2), Coefficient of Determination (R^2) and the Root Mean Squared Error Cross-Validation
223 (RMSECV) and Prediction (RMSEP). The model quality index (Q^2) measures the global goodness-of-
224 fit and the predictive quality of the analytes used in the model (volatiles, non-volatiles and data
225 fusion). The maximum value of Q^2 is equivalent to the most stable model. The Coefficient of
226 Determination of the model (R^2) indicates the proportion of variability in the dependent variable
227 (sensory score) explained by the model, and ranges between 0 and 1; the closer R^2 is to 1, the better
228 the model. The main issue with R^2 is that it does not take into account the number of variables used
229 to fit the model. This limit has been overcome by the Adjusted R^2 . The number of variables used to

230 develop the model is important since the number of unnecessary variables penalizes the model;
231 unlike R^2 , Adjusted R^2 is sensitive to these penalties. Adjusted R^2 can be calculated using the
232 following equation:

233

$$234 \text{Adj}R^2 = 1 - (1 - R^2) \times \frac{n-1}{n-p}$$

235

236 where R^2 is the determination coefficient of the model, n and p are the numbers of observations
237 and variables used to fit the model, respectively. The differences between the predicted values and
238 those measured is given by the Root Mean Squared Error (RMSE), which determines the average of
239 the squares of the errors or deviations. The error calculated in the cross-validated data is known as
240 root mean squared error in Cross Validation (RMSECV), while the value calculated in the prediction
241 data is the root mean squared error in prediction (RMSEP).

242

$$243 \text{RMSEP} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_{i/i})^2}{n}}$$

244

245 y_i is the experimental response and $\hat{y}_{i/i}$ is the response predicted by the regression model, where
246 i/i indicates that the response is predicted by a model that was estimated when the i – th sample
247 was left out from the training set.⁴²

248

249 **RESULTS AND DISCUSSION**

250 This section mainly focuses on the chromatographic fingerprinting approach and on its correlation
251 to sensory data.

252

253 **Exploring data matrices**

254 Data exploration was, at first, performed using Principal Component Analysis (PCA) on volatiles and
255 non-volatile fingerprint features, treated independently. Each sample (observation) was described
256 by all of the features detected above the fixed S/N threshold (variables).

257 Results are illustrated in Figure 2A and 2B, which show the distribution of the samples over the two
258 PCs, that were able to cover more than 50% of the total variance of the data matrix [68% for volatiles
259 (Figure 2A) and 53% for non-volatiles (Figure2B)]. In both cases, good separation can be observed
260 between the Arabica (Blue) and Robusta (Green) samples, suggesting that both chemical fractions
261 provide a similar contribution to sample discrimination.³⁸

262 Multiple Factor Analysis (MFA) was then used to compare the three different data matrices:
263 chemical fingerprints (volatiles and non-volatiles) and the sensory data related to the seven sensory
264 notes considered (*i.e.*, *Acid*, *Bitter*, *Flowery*, *Fruity*, *Woody* and *Spicy*).

265 MFA proceeds in two steps: i) first it computes a PCA of each data table, and 'normalizes' each data
266 table by dividing all elements by the first singular value obtained from its PCA; ii) secondly, all
267 normalized data tables are aggregated into a single data table that is then analyzed via a (non-
268 normalized) PCA, which provides a set of factor scores for observations and loadings for the
269 variables.⁴³ The results, displayed by filtered group (by origins) for better visualization, are reported
270 in Figure 2C, which shows how volatiles (conventionally indicated in the figure as Ar) and non-
271 volatiles (conventionally indicated in the figure as Ta) for the different samples have quite similar
272 branches. RV correlation coefficients, shown in Table 1, indicate to what extent the distribution of
273 the tables/variables are related two-by-two, reflecting the amount of variance shared from the
274 tables. The more correlated the variables are, the higher the RV coefficient is (variation range 0-1).
275 The mutual correlation between volatiles and non-volatiles was 0.921, between sensory data and
276 non-volatiles it was 0.505, while for sensory data and volatiles it was 0.549 (Table 1).

277 These data suggest a possible relationship between chemical fingerprints, although the correlation
278 is not particularly high. Results confirm what already reported in the literature; i) the volatiles have
279 an important role in the definition of the coffee sensory profile; and ii) volatiles, including aroma
280 active compounds, better correlate with sensory features than non-volatiles, including taste active
281 compounds.^{3,5,44}

282 The correlation values (MFA values) reported in Table 1 indicate that the multicollinearity between
283 the information provided by the chemical fingerprints and the sensory analysis is weak, while
284 suggesting that they both contribute to the definition of the overall coffee flavor and that no aspect
285 can fully be ignored.

286 The next sections focus on the workflow that was adopted to develop a predictive model for the
287 *Bitter* note. In particular, the information provided by volatiles and non-volatiles fingerprints
288 independently or combined together will be explored. The *Bitter* note was also taken as a test bench
289 because of its relevance in the hedonic profile of the coffee brew. The adopted strategy was then
290 also applied to all other sensory notes considered.

291

292 ***Bitter* flavor evaluation: chemical components-selection strategy**

293 Bitter and acid notes are typically perceived as taste attributes. *Bitter* was here chosen as model
294 note to explore how the volatile, non-volatile fraction or combined data, might correctly describe
295 the bitter-score in prediction. The objective is to understand whether a traditional taste attribute
296 can be (more) correctly described by combining chemical data from both taste-active and odor-
297 active features.

298 The features that were highly correlated to the expression of high and low *Bitter* notes were first
299 selected by PLS-DA. This step was applied to volatile and non-volatile data sets separately, and later,
300 to the fused data matrix. Figure 3 shows the chromatographic fingerprints resulting from GC-MS

301 (volatiles Figure 3A) and LC-UV/DAD (non-volatiles Figure 3B). The grey lines indicate the retention
302 times of the most relevant features designated by PLS-DA.

303 The values for the Variable Impact on Projections (VIP) were used as a filter parameter, as VIP
304 coefficients reflect the relative importance of each X variable in the prediction model. A cut-off of 1
305 and a non-zero standard deviation (SD) were used to select features. Figures S1A and S1B in
306 supporting information show the results for volatiles, while Figures S1C and S1D indicate those for
307 non-volatiles, treated independently.

308 The prediction models were then developed by applying a PLS regression algorithm to the selected
309 *Bitter*-related features.

310 150 samples were used to build up the regression model, 20 of them were randomly employed as
311 a validation set, and 30 were excluded from the training set and adopted as an external test set. A
312 comparison of the PLS regression model parameters is reported in Table 2. The results unexpectedly
313 suggest that the data from the volatiles and from the fused data matrix (volatiles + non-volatiles)
314 show a similar behavior in the description of the *Bitter* note. Although the fingerprinting of non-
315 volatiles was obtained by applying selected wavelengths characteristic of bitter-related chemicals
316 (*i.e.* caffeine, trigonelline and chlorogenic acid derivatives), it provided less information than
317 volatiles alone (see both the Q^2 and coefficient of determination (R^2) values in Table 2). This is
318 probably due both to other inferences in the description of this note, and to the partial (not
319 comprehensive) fingerprinting of the non-volatile fraction by LC-UV.

320

321 *Bitter-related components in the volatile fraction: informative volatiles and aroma-active*
322 *compounds*

323 The volatile fraction provided information that is useful to characterize the bitter-note signature in
324 the analyzed samples. Further MS and retention data investigation into *Bitter*-related features led

325 to the identification of several aroma-active compounds, including: pyrazines (1-
326 methylethenylpyrazine (Ar6), 5-Methyl-6,7-dihydro-5H-cyclopentapyrazine (Ar24), 2-*n*-propyl
327 pyrazine (Ar28), 2,3-dimethylpyrazine (Ar32), 2-methyl-5H-6,7-dihydrocyclopentapyrazine (Ar25),
328 2,3-dimethylpyrazine (Ar32), 2-ethyl-3,5-dimethylpyrazine[§] (Ar39)); phenols (4-ethylguaiacol[§]
329 (Ar46), 4-vinylguaiacol[§] (Ar48), guaiacol[§] (Ar66)); 2-phenylethanol (Ar55); 1-H-pyrrole (Ar8); 2-
330 furanmethanethiol[§] (Ar17); furfuryl methyl sulfide (Ar63); and furfuryl pyrrole (Ar65).
331 Supplementary Table 2S lists the identified *Bitter*-related volatiles together with their odor quality,
332 experimental and tabulated I^T s, MS similarity match with the reference spectra from commercial
333 databases and/or pure standards. Several analytes (§) are coffee key-aroma compounds, as
334 indicated by Blank *et al.* ⁴⁵. Interestingly, these compounds are described as earthy, roasty, burnt
335 and phenolic, but none of them was directly related to bitterness. Nevertheless, Barba *et al.* ³⁶
336 suggested that 8% of the panelists associated furfural with the description *Bitter* taste, although it
337 is conventionally reported as *bready* and *caramellic*. Moreover, in a study on fruit juices, Guichard
338 *et al.* ³⁵ observed that an enhancing effect on *Bitter* perception was triggered by a correlative pattern
339 of non-*Bitter*-eliciting odorants, such as ethyl-2-methyl and 2-ethylbutanoate, γ -decalactone,
340 furfural, allo-ocimene, butyl-acetate, β -myrcene and pentanoic acid. These authors stated that
341 odorants that enhance a target-taste perception may be exploited to modulate the overall taste
342 perception in foods and beverages.

343

344 *Bitter-related chemicals in the non-volatile fraction: informative analytes and taste-active* 345 *compounds*

346 The non-volatile fingerprints were elaborated using the same strategy as for the volatiles (see
347 section above). Model performance is reported in Table 2, and confirmed the existing positive
348 correlation between the *Bitter* note and some chemical features detected by LC-UV/DAD. *Bitter-*

349 related chemicals were then identified (or putatively identified) using *post-hoc* LC-MS/MS
350 analysis.⁴⁶⁻⁴⁸

351 Peaks with a maximum of absorption at 325 nm are characterized by pseudomolecular ions at 337
352 m/z and 335 m/z, in the ESI⁺ and ESI⁻ acquisition modes, respectively, with diagnostic fragments at
353 163 m/z (ESI⁺) and 161 m/z (ESI⁻) that correspond to the caffeic acid moiety with a loss of a water
354 molecule, and that can be putatively identified as caffeoylquinic lactones. Similarly, peaks with
355 characteristic UV-absorption maxima at 323 nm and 310 nm can tentatively be attributed to
356 feruloylquinic acid isomers and coumaroylquinic acid, respectively. This is confirmed by the
357 presence of pseudomolecular ions at 369 m/z (ESI⁺) and 367 m/z (ESI⁻) for feruloylquinic acid, and
358 at 339 m/z (ESI⁺) and 337 m/z (ESI⁻) for coumaroylquinic acid and other diagnostic fragments, as
359 reported by Martini *et al.* 2017⁴⁹ (Figure 2S).

360 Table S3, in supporting information, lists the identified bitter-related non-volatiles together with
361 their retention times, λ -max, molecular weight, molecular ions and MS/MS data.

362 Some of the most relevant components are two feruloylquinic acid isomers (FQA – Ta6, Ta18), three
363 caffeoylquinic lactone isomers[§] (CQL – Ta11, Ta10, Ta13), one feruloylquinic acid isomer[§] (FQA–
364 Ta10), 3,4 and 4,5 dicaffeoylquinic acid (diCQA-Ta21, Ta25) and caffeine[§] (Ta14) (Figure 3B and Table
365 S3 in supporting information). Most of these (§) were already associated to the *Bitter* note by
366 Hofmann *et al.*^{47,50,51}, although, rather surprisingly, here caffeoylquinic acid isomers (CQAs-Ta3, 5,
367 7) and trigonelline (Ta2) were not strongly correlated to *Bitter*. Moreover, although to a different
368 extent, the heat map in Figure 4 shows that most of the identified taste active compounds were
369 positively correlated with volatiles related to *Bitter* (Ar6, 8, 17, 24-25, 32, 39, 44-48, 55, in light
370 green). In this figure, the Pearson's correlation matrix between volatiles (Ar) and non-volatiles (Ta)
371 at 5% of confidence level is visualized in a green-to-red color scale, with the colors ranging from

372 light green $p=+1$ to red $p=-1$. Most of the targeted compounds in Table S3 increase with the same
373 trend in the most bitter samples, albeit with different magnitudes.

374

375 *Bitter evaluation: performance evaluation of data-fusion strategy*

376 This paragraph investigates the possible gain in explanatory and predictive power when chemical
377 information from volatiles and non-volatiles are combined together in developing models. GC and
378 LC analyses are partly complementary, in terms of compounds analyzed, and the combination of
379 the data sets may be more informative by revealing, for instance, possible associations between
380 volatiles and non-volatiles (Figure 3).

381 As flavor perception derives from the interaction between aroma and taste, the combination of the
382 information provided by the two different fractions was expected to increase the performance of
383 the predictive model. The regression model to predict the scores of the *Bitter* sensory attribute was
384 built from the fused data matrix (GC and LC data), without a preliminary selection of variables from
385 the PLS-DA of the single approaches, and by re-submitting the fused data set to the work-flow
386 established for each single analytical technique. The performance of the fused data matrix
387 regression model was in line with those obtained using fingerprint data from volatile and non-
388 volatile models (Table 2). Data fusion did not improve the overall prediction quality of the model
389 (Q^2 in bold in Table 2), the error in the cross-validation set (RMSECV), nor the prediction of the
390 external test set (RMSEP). Results showed that the model had better prediction quality than the
391 non-volatile fraction alone, but worse than that of the volatile fraction alone. The non-volatile
392 fraction did not add information to better understand the perception while volatiles alone had
393 better performances even to model a taste perception. Although the coffee non-volatile fraction
394 analyzed is not fully representative, nevertheless the considered non-volatile markers are well-
395 established sensory quality marker in routine controls.^{46,47,50,51,52} The volatile fraction possibly have

396 an actual influence in driving the description of this sensory attribute.^{53,54} This possibility, together
397 with the correlation pathways between volatiles and non-volatiles, deserves much more in-depth
398 investigation.

399

400 **Investigation of the integrated approach into all sensory notes**

401 The data-elaboration workflow, validated on the *Bitter* note, was also used to investigate the other
402 flavor attributes, and to understand to which extent the two chemical fractions (single or combined)
403 play a role in sensory-quality description. Table 2 summarizes the performance of the models in
404 predicting sensory scores.

405 The Q^2 values clearly indicate that the non-volatile fraction had a differential impact on the
406 prediction models. Non-volatiles showed better performance for *Acid*, *Spicy* and *Woody* notes than
407 for *Flowery* and *Fruity* notes. This trend was also confirmed by the R^2 values, which were higher in
408 *Acid*, *Spicy* and *Woody* notes.

409 The behavior of RMSECV slightly differed; these values were in compliance with the previous
410 observations on *Bitter* only, for the *Acid* note, and less evident for the others. The non-volatile
411 fraction was expected to have a lower impact on *Fruity* and *Flowery* notes since these notes are
412 considered to be closer to aroma attributes, and the associations with components detected using
413 the adopted analytical method have not yet been found.⁵⁵

414 As a general consideration, most of the notes (with the exception of *Spicy*) showed better
415 performance when the predictive model was based on volatile features (Table 2), suggesting a
416 better agreement between HS-SPME-GC-MS data and sensory scores within the investigated sample
417 set.

418 The results of the fused data bring to three different scenarios:

- 419 - *Acid* and *Woody* notes: the models on fused data showed acceptable performance (Q^2
420 around 0.7, $R^2 > 0.8$ and a RMSECV lower than 1). Their overlap with the results of the PLS
421 models, which were developed using volatiles data alone, suggests that the non-volatiles
422 provides a negligible contribution into the flavor note description.
- 423 - *Flowery* and *Fruity* notes: the performance of the models based on fused data were worse
424 than that based solely on volatiles. As expected, here volatiles provide meaningful and
425 consistent information than the fused-data sets. Non-volatiles increase the noise and act as
426 confounding elements.
- 427 - *Spicy* note: the performance of the fused data model is slightly better than that obtained
428 from the single fractions. The most significant improvements were observed in the Q^2 and
429 RMSECV values. This result is supported by the fact that some key spicy volatile compounds
430 (mainly phenolics, such as guaiacoles) originated from the thermal degradation of
431 chlorogenic acids (i.e. those detected in HPLC analyses)⁵⁶ monitored on the LC-UV/DAD
432 fingerprints.
- 433 - The low model stability registered for *Flowery*, *Fruity* and *Spicy* notes might be linked to the
434 unbalanced distribution of the samples within the training set; the number of samples with
435 low scores for these notes was much higher than that for samples with high scores. This
436 unbalanced sample distribution makes the use of PLS algorithm challenging, and requires
437 suitable algorithms to better follow data evolution. The optimization of regression, via a non-
438 parametric algorithm on volatile data, might improve prediction ability of the models for
439 these sensory attributes.⁵⁷
- 440 These results show that the screening carried out with two different analytical platforms routinely
441 used in quality control laboratories have a complementary role but with different relevance in
442 describing coffee sensory quality. While MFA suggests the existence of a certain orthogonality

443 between volatile and non-volatile data, the regression models highlight the key role played by the
444 volatile fraction, and therefore of the aroma, in the sample sensory characterization. The
445 performance of the PLS models, built up with the fused fingerprints, is comparable to that obtained
446 from HS-SPME-GC-MS. These correlative results were already sensorially confirmed by some
447 authors, who have reported that flavor perception, in all its aspects, is mostly linked to aroma
448 composition and impact.^{1,3,5} These observations, together with the good results obtained in the
449 definition of *Acid* and *Bitter* notes (considered as “typical taste notes”) from the volatile data, make
450 it possible to hypothesize that the analysis of the volatile fraction may be sufficiently representative
451 to delineate coffee flavor, and provide reliable chemical fingerprints that can be associated to some
452 sensory notes, including those typical of taste. Moreover, the correlative results highlighted from
453 the volatile-non volatile fused data deserves a thorough investigation to verify the potential of odor-
454 taste integration.

455 The reported correlative patterns indicate that the integrated approach can successfully be used as
456 a complement to sensory analysis, in particular to design coffees with specific flavor profiles.

457 As a general consideration, the success in the development of these methods requires a high
458 consistency and alignment of the sensory panel in products evaluation, since subjectivity in data
459 collection can influence the development of the mathematical model for scores prediction.
460 However, the natural variability of the coffee matrix and its complexity makes difficult to achieve a
461 good representativeness for all commercial coffees treated at the industrial level.

462 To make more robust and reliable instrumental tools for sensory prediction, a huge amount of data,
463 both from sensory profiling and chemical fingerprinting, are necessary. Modern artificial intelligence
464 algorithms might be of help to simulate human skills but training data should match with the actual
465 complexity of the phenomenon of multimodal flavor perception.

466

467 **Supporting Information**

468 In supporting information are reported the PLS-DA results on volatiles and non-volatiles in Figure
469 S1, LC-DAD-MS signals in Figure S2. The list of the coffee origins and species investigated are
470 reported in Table S1. *Bitter*-related volatile and non-volatile compounds, identified or tentatively
471 identified with the workflow proposed, are displayed respectively in Table S2 and S3.

472 This material is available free of charge via the Internet at <http://pubs.acs.org>.

473

474 **Acknowledgements**

475 This study was carried out thanks to the financial support of Lavazza and was supported by Ricerca
476 Finanziata da Università - Fondo per la Ricerca Locale (Ex 60%) Anno2019.

477

478 **Compliance with ethical standards Notes**

479 Pellegrino Gloria and Ruosi M. Rosanna are presently employees of Lavazza S.p.a., Torino, Italy.

480

481 **REFERENCES**

- 482 (1) Prescott, J. ScienceDirect Multisensory Processes in Flavour Perception and Their Influence
483 on Food Choice. *Curr. Opin. Food Sci.* **2015**, *3*, 47–52.
484 <https://doi.org/10.1016/j.cofs.2015.02.007>.
- 485 (2) Prescott, J.; Johnstone, V.; Francis, J. Odor – Taste Interactions : Effects of Attentional
486 Strategies during Exposure. **2004**, *29* (4), 331–340. <https://doi.org/10.1093/chemse/bjh036>.
- 487 (3) Spence, C. The Multisensory Perception of Flavour. *Psychologist* **2010**, *23* (9), 720–723.
488 <https://doi.org/10.1016/j.concog.2007.06.005>.
- 489 (4) Thomas-Danguin, T.; Sinding, C.; Tournier, C.; Saint-Eve, A. Multimodal Interactions. In *Flavor:*
490 *From Food to Behaviors, Wellbeing and Health*; P. Etievant, P. E. Guichard, C. Salles, & A. E.
491 V., Ed.; Woodhead Publishing, 2016; pp 121–141.
- 492 (5) Sunarharum, W. B.; Williams, D. J.; Smyth, H. E. Complexity of Coffee Flavor: A Compositional
493 and Sensory Perspective. *Food Res. Int.* **2014**, *62*, 315–325.
494 <https://doi.org/10.1016/j.foodres.2014.02.030>.
- 495 (6) Alcaire, F.; Antúnez, L.; Vidal, L.; Giménez, A.; Ares, G. Aroma-Related Cross-Modal
496 Interactions for Sugar Reduction in Milk Desserts: Influence on Consumer Perception. *Food*
497 *Res. Int.* **2017**, *97*, 45–50. <https://doi.org/10.1016/j.foodres.2017.02.019>.
- 498 (7) Arla. Why sugar is on everyone’s lips | Arla [https://www.arla.com/food-for-](https://www.arla.com/food-for-thought/health/why-sugar-is-on-everyones-lips/)
499 [thought/health/why-sugar-is-on-everyones-lips/](https://www.arla.com/food-for-thought/health/why-sugar-is-on-everyones-lips/) (accessed Apr 23, 2020).
- 500 (8) Askew, K. Sugar reduction: Blend the trends with functional formulations
501 [https://www.foodnavigator.com/Article/2019/05/15/Sugar-reduction-Blend-the-trends-](https://www.foodnavigator.com/Article/2019/05/15/Sugar-reduction-Blend-the-trends-with-functional-formulations)
502 [with-functional-formulations](https://www.foodnavigator.com/Article/2019/05/15/Sugar-reduction-Blend-the-trends-with-functional-formulations) (accessed Apr 23, 2020).
- 503 (9) Khan, M. PepsiCo R&D, a Catalyst for Change in the Beverage Industry. *New Food* **2015**, *16*,
504 10–13.

- 505 (10) Thomas-Danguin, T.; Guichard, E.; Salles, C. Cross-Modal Interactions as a Strategy to
506 Enhance Salty Taste and to Maintain Liking of Low-Salt Food: A Review. *Food Funct.* **2019**, *10*,
507 5269. <https://doi.org/10.1039/c8fo02006j>.
- 508 (11) Wang, Q. J.; Mielby, L. A.; Junge, J. Y.; Bertelsen, A. S.; Kidmose, U.; Spence, C.; Byrne, D. V.
509 The Role of Intrinsic and Extrinsic Sensory Factors in Sweetness Perception of Food and
510 Beverages: A Review. *Foods* **2019**, *8* (6). <https://doi.org/10.3390/foods8060211>.
- 511 (12) Chambers, E.; Koppel, K. Associations of Volatile Compounds with Sensory Aroma and Flavor:
512 The Complex Nature of Flavor. *Molecules* **2013**, *18* (5), 4887–4905.
513 <https://doi.org/10.3390/molecules18054887>.
- 514 (13) Tolessa, K.; Rademaker, M.; De Baets, B.; Boeckx, P. Prediction of Specialty Coffee Cup Quality
515 Based on near Infrared Spectra of Green Coffee Beans. *Talanta* **2016**, *150*, 367–374.
516 <https://doi.org/10.1016/j.talanta.2015.12.039>.
- 517 (14) Craig, A. P.; Botelho, B. G.; Oliveira, L. S.; Franca, A. S. Mid Infrared Spectroscopy and
518 Chemometrics as Tools for the Classification of Roasted Coffees by Cup Quality. *Food Chem.*
519 **2018**, *245*, 1052–1061. <https://doi.org/10.1016/J.FOODCHEM.2017.11.066>.
- 520 (15) Bressanello, D.; Liberto, E.; Cordero, C.; Sgorbini, B.; Rubiolo, P.; Pellegrino, G.; Ruosi, M. R.;
521 Bicchi, C. Chemometric Modelling of Coffee Sensory Notes through Their Chemical
522 Signatures: Potential and Limits in Defining an Analytical Tool for Quality Control
523 Chemometric Modelling of Coffee Sensory Notes through Their Chemical Signatures:
524 Potential and Limit. **2018**. <https://doi.org/10.1021/acs.jafc.8b01340>.
- 525 (16) Liberto, E.; Bressanello, D.; Strocchi, G.; Cordero, C.; Ruosi, M. R.; Pellegrino, G.; Bicchi, C.;
526 Sgorbini, B. HS-SPME-MS-Enose Coupled with Chemometrics as an Analytical Decision Maker
527 to Predict in-Cup Coffee Sensory Quality in Routine Controls: Possibilities and Limits.
528 *Molecules* **2019**, *24* (24). <https://doi.org/10.3390/molecules24244515>.

- 529 (17) Baqueta, M. R.; Coqueiro, A.; Valderrama, P. Brazilian Coffee Blends: A Simple and Fast
530 Method by Near-Infrared Spectroscopy for the Determination of the Sensory Attributes
531 Elicited in Professional Coffee Cupping. *J. Food Sci.* **2019**, *84* (6), 1247–1255.
532 <https://doi.org/10.1111/1750-3841.14617>.
- 533 (18) Rocchetti, G.; Braceschi, G. P.; Odello, L.; Bertuzzi, T.; Trevisan, M.; Lucini, L. Identification of
534 Markers of Sensory Quality in Ground Coffee: An Untargeted Metabolomics Approach.
535 *Metabolomics* **2020**, *16* (12). <https://doi.org/10.1007/s11306-020-01751-6>.
- 536 (19) Domingues, L. O. C.; Garcia, A. O.; Ferreira, M. M. C.; Morgano, M. A. Sensory Quality
537 Prediction of Coffee Assessed by Physicochemical Parameters and Multivariate Model. *Coffee*
538 *Sci.* **2020**, *15* (1), 1–11. <https://doi.org/10.25186/cs.v15i.1654>.
- 539 (20) Oliveira, E. C. D. S.; Guarçoni, R. C.; de Castro, E. V. R.; de Castro, M. G.; Pereira, L. L. Chemical
540 and Sensory Perception of Robusta Coffees under Wet Processing. *Coffee Sci.* **2020**, *15* (1), 1–
541 8. <https://doi.org/10.25186/.v15i.1672>.
- 542 (21) Yu-Tang Chang, Meng-Chien Hsueh, Shu-Pin Hung, Juin-Ming Lu, Jia-Hung Peng, Shih-Fang
543 Chen. Prediction of Specialty Coffee Flavors Based on Near-infrared Spectra Using Machine-
544 and Deep-learning Methods. *J. Sci. food Agric.* **2021**.
545 <https://doi.org/https://doi.org/10.1002/jsfa.11116>.
- 546 (22) Quintanilla-Casas, B.; Bustamante, J.; Guardiola, F.; García-González, D. L.; Barbieri, S.;
547 Bendini, A.; Toschi, T. G.; Vichi, S.; Tres, A. Virgin Olive Oil Volatile Fingerprint and
548 Chemometrics: Towards an Instrumental Screening Tool to Grade the Sensory Quality. *LWT*
549 **2020**, *121*, 108936. <https://doi.org/10.1016/j.lwt.2019.108936>.
- 550 (23) Quintanilla-Casas, B.; Marin, M.; Guardiola, F.; García-González, D. L.; Barbieri, S.; Bendini, A.;
551 Toschi, T. G.; Vichi, S.; Tres, A. Supporting the Sensory Panel to Grade Virgin Olive Oils: An in-
552 House-Validated Screening Tool by Volatile Fingerprinting and Chemometrics. *Foods* **2020**, *9*

- 553 (10), 1–14. <https://doi.org/10.3390/foods9101509>.
- 554 (24) Lindinger, C.; Labbe, D.; Pollien, P.; Rytz, A.; Juillerat, M. A.; Yeretian, C.; Blank, I. When
555 Machine Tastes Coffee : Instrumental Espresso Coffee. *Anal. Chem.* **2008**, *80* (5), 1574–1581.
- 556 (25) Ribeiro, J. S.; Ferreira, M. M. C.; Salva, T. J. G. Chemometric Models for the Quantitative
557 Descriptive Sensory Analysis of Arabica Coffee Beverages Using near Infrared Spectroscopy.
558 *Talanta* **2011**, *83* (5), 1352–1358. <https://doi.org/10.1016/j.talanta.2010.11.001>.
- 559 (26) Ribeiro, J. S.; Augusto, F.; Salva, T. J. G.; Ferreira, M. M. C. Prediction Models for Arabica
560 Coffee Beverage Quality Based on Aroma Analyses and Chemometrics. *Talanta* **2012**, *101*,
561 253–260. <https://doi.org/10.1016/j.talanta.2012.09.022>.
- 562 (27) Peterson, R. T. Chemical Biology and the Limits of Reductionism. *Nat Chem Biol* **2008**, *4* (11),
563 635–638.
- 564 (28) Dunkel, A.; Steinhaus, M.; Kotthoff, M.; Nowak, B.; Krautwurst, D.; Schieberle, P.; Hofmann,
565 T. Nature s Chemical Signatures in Human Olfaction : A Foodborne Perspective for Future
566 Biotechnology Nature ’ s Chemical Signatures in Human Olfaction – a Foodborne Perspective
567 for Future Biotechnology. *Angew. Chem. Int. Ed.* **2014**, *53*, 2 – 22.
- 568 (29) Charve, J.; Chen, C.; Hegeman, A. D.; Reineccius, G. A. Evaluation of Instrumental Methods
569 for the Untargeted Analysis of Chemical Stimuli of Orange Juice Flavour. *Flavour Fragr. J.*
570 **2011**, *26* (6), 429–440. <https://doi.org/10.1002/ffj.2078>.
- 571 (30) Stilo, F.; Bicchi, C.; Robbat, A.; Reichenbach, S. E.; Cordero, C. Untargeted Approaches in Food-
572 Omics: The Potential of Comprehensive Two-Dimensional Gas Chromatography/Mass
573 Spectrometry. *TrAC Trends Anal. Chem.* **2020**, *135*, 116162.
574 <https://doi.org/10.1016/j.trac.2020.116162>.
- 575 (31) Cordero, C.; Kiefl, J.; Reichenbach, S. E.; Bicchi, C. Characterization of Odorant Patterns by
576 Comprehensive Two-Dimensional Gas Chromatography: A Challenge in Omic Studies. *TrAC -*

- 577 *Trends in Analytical Chemistry*. Elsevier B.V. **2019**, pp 364–378.
578 <https://doi.org/10.1016/j.trac.2018.06.005>.
- 579 (32) Nicolotti, L.; Mall, V.; Schieberle, P. Characterization of Key Aroma Compounds in a
580 Commercial Rum and an Australian Red Wine by Means of a New Sensomics-Based Expert
581 System (SEBES) - An Approach to Use Artificial Intelligence in Determining Food Odor Codes.
582 *J. Agric. Food Chem.* **2019**, *67* (14), 4011–4022. <https://doi.org/10.1021/acs.jafc.9b00708>.
- 583 (33) Tromelin, A.; Chabanet, C.; Audouze, K.; Koensgen, F.; Guichard, E. Multivariate Statistical
584 Analysis of a Large Odorants Database Aimed at Revealing Similarities and Links between
585 Odorants and Odors. *Flavour Fragr. J.* **2018**, *33* (1), 106–126.
586 <https://doi.org/10.1002/ffj.3430>.
- 587 (34) Wang, H.; Chambers, E.; Kan, J. Sensory Characteristics of Combinations of Phenolic
588 Compounds Potentially Associated with Smoked Aroma in Foods. *Molecules* **2018**, *23* (8).
589 <https://doi.org/10.3390/molecules23081867>.
- 590 (35) Guichard, E.; Barba, C.; Thomas-Danguin, T.; Tromelin, A. Multivariate Statistical Analysis and
591 Odor–Taste Network To Reveal Odor–Taste Associations. *J. Agric. Food Chem.* **2020**, *68*, 38,
592 10318–10328. <https://doi.org/10.1021/acs.jafc.9b05462>.
- 593 (36) Barba, C.; Beno, N.; Guichard, E.; Thomas-Danguin, T. Selecting Odorant Compounds to
594 Enhance Sweet Flavor Perception by Gas Chromatography/Olfactometry-Associated Taste
595 (GC/O-AT). *Food Chem.* **2018**, *257*, 172–181.
596 <https://doi.org/10.1016/J.FOODCHEM.2018.02.152>.
- 597 (37) SCAA. SCAA Protocols Cupping Specialty Coffee. *Spec. Coffee Assoc. Am.* **2015**, 1–10.
- 598 (38) Bressanello, D.; Liberto, E.; Cordero, C.; Rubiolo, P.; Pellegrino, G.; Ruosi, M. R.; Bicchi, C.
599 Coffee Aroma: Chemometric Comparison of the Chemical Information Provided by Three
600 Different Samplings Combined with GC–MS to Describe the Sensory Properties in Cup. *Food*

- 601 *Chemistry*. **2017**, pp 218–226. <https://doi.org/10.1016/j.foodchem.2016.07.088>.
- 602 (39) Valverde-som, L.; Cuadros-rodríguez, L.; Ruiz-sambl, C.; Gonz, A. *Analytica Chimica Acta*
603 *Chromatographic Fingerprinting : An Innovative Approach for Food ' Identification ' and Food*
604 *Authentication e A Tutorial*. **2016**, *909*. <https://doi.org/10.1016/j.aca.2015.12.042>.
- 605 (40) Stilo, F.; Bicchi, C.; Jimenez-Carvelo, A. M.; Cuadros-Rodriguez, L.; Reichenbach, S. E.; Cordero,
606 C. *Chromatographic Fingerprinting by Comprehensive Two-Dimensional Chromatography:*
607 *Fundamentals and Tools. TrAC - Trends Anal. Chem.* **2021**, *134*, 116133.
608 <https://doi.org/10.1016/j.trac.2020.116133>.
- 609 (41) Ribeiro, J. S.; Augusto, F.; Salva, T. J. G.; Thomaziello, R. A.; Ferreira, M. M. C. *Prediction of*
610 *Sensory Properties of Brazilian Arabica Roasted Coffees by Headspace Solid Phase*
611 *Microextraction-Gas Chromatography and Partial Least Squares. Anal. Chim. Acta* **2009**, *634*
612 *(2)*, 172–179. <https://doi.org/10.1016/j.aca.2008.12.028>.
- 613 (42) Brereton, R. G.; Jansen, J.; Lopes, J.; Marini, F.; Pomerantsev, A.; Rodionova, O.; Roger, J. M.;
614 Walczak, B.; Tauler, R. *Chemometrics in Analytical Chemistry—Part II: Modeling, Validation,*
615 *and Applications. Anal. Bioanal. Chem.* **2018**, *410* (26), 6691–6704.
616 <https://doi.org/10.1007/s00216-018-1283-4>.
- 617 (43) Abdi, H.; Williams, L. J.; Valentin, D. *Multiple Factor Analysis: Principal Component Analysis*
618 *for Multitable and Multiblock Data Sets. Wiley Interdiscip. Rev. Comput. Stat.* **2013**, *5* (2), 149–
619 179. <https://doi.org/10.1002/wics.1246>.
- 620 (44) Spinelli, S.; Masi, C.; Zoboli, G. P.; Prescott, J.; Monteleone, E. *Emotional Responses to*
621 *Branded and Unbranded Foods.* **2015**, *42*, 1–11.
- 622 (45) Blank, I.; Sen, A.; Grosch, W. *Potent Odorants of the Roasted Powder and Brew of Arabica*
623 *Coffee. Leb. Und-forsch.* **1992**, *195*, 239–245.
- 624 (46) Perrone, D.; Farah, A.; Donangelo, C. M.; Paulis, T. De; Martin, P. R. *Food Chemistry*

- 625 Comprehensive Analysis of Major and Minor Chlorogenic Acids and Lactones in Economically
626 Relevant Brazilian Coffee Cultivars. **2008**, *106*, 859–867.
627 <https://doi.org/10.1016/j.foodchem.2007.06.053>.
- 628 (47) Frank, O.; Zehentbauer, G. Bioresponse-Guided Decomposition of Roast Coffee Beverage and
629 Identification of Key Bitter Taste Compounds. *Eur Food Res Technol* **2006**, 492–508.
630 <https://doi.org/10.1007/s00217-005-0143-6>.
- 631 (48) Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W. M.;
632 Fiehn, O.; Goodacre, R.; Griffin, J. L.; et al. Proposed Minimum Reporting Standards for
633 Chemical Analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards
634 Initiative (MSI). *Metabolomics* **2007**, *3* (3), 211–221. [https://doi.org/10.1007/s11306-007-](https://doi.org/10.1007/s11306-007-0082-2)
635 [0082-2](https://doi.org/10.1007/s11306-007-0082-2).
- 636 (49) Martini, S.; Conte, A.; Tagliazucchi, D. Phenolic Compounds Profile and Antioxidant Properties
637 of Six Sweet Cherry (*Prunus Avium*) Cultivars. *Food Res. Int.* **2017**, *97*, 15–26.
638 <https://doi.org/10.1016/j.foodres.2017.03.030>.
- 639 (50) Frank, O.; Blumberg, S.; Krümpel, G.; Hofmann, T. Structure Determination of 3-O-Caffeoyl-
640 Epi- γ -Quinide, an Orphan Bitter Lactone in Roasted Coffee. *J. Agric. Food Chem.* **2008**, *56* (20),
641 9581–9585. <https://doi.org/10.1021/jf802210a>.
- 642 (51) Kreppenhofer, S.; Frank, O.; Hofmann, T. Identification of (Furan-2-Yl) Methylated Benzene
643 Diols and Triols as a Novel Class of Bitter Compounds in Roasted Coffee. *Food Chem.* **2011**,
644 *126* (2), 441–449. <https://doi.org/10.1016/j.foodchem.2010.11.008>.
- 645 (52) Farah, A. Food Chemistry Correlation between Cup Quality and Chemical Attributes of
646 Brazilian Coffee. **2006**, *98*, 373–380. <https://doi.org/10.1016/j.foodchem.2005.07.032>.
- 647 (53) Paravisini, L.; Soldavini, A.; Peterson, J.; Simons, C. T.; Peterson, D. G. Impact of Bitter Tastant
648 Sub-Qualities on Retronasal Coffee Aroma Perception. **2019**.

- 649 <https://doi.org/10.1371/journal.pone.0223280>.
- 650 (54) Sittipod, S.; Schwartz, E.; Paravisini, L.; Peterson, D. G. Identification of Flavor Modulating
651 Compounds That Positively Impact Coffee Quality. *Food Chem.* **2019**.
652 <https://doi.org/10.1016/j.foodchem.2019.125250>.
- 653 (55) De Oliveira Fassio L. , Ribeiro Malta M., C. G. R.; G.R., L.; P.M., D. L.; C.J., P. Sensory Description
654 of Cultivars (*Coffea Arabica* L.) Resistant to Rust and Its Correlation with Caffeine,
655 Trigonelline, and Chlorogenic Acid Compounds. *Beverages* **2016**, 2 (1), 1.
656 <https://doi.org/https://doi.org/10.3390/beverages2010001>.
- 657 (56) Flament, I. *Coffee Flavor Chemistry*; Wiley, Chichester UK, **2002**.
- 658 (57) Bressanello, D. PhD Thesis: Multidisciplinary Approaches In Foodomic Studies: Analytical and
659 Chemometric Relationships between Sensory Evaluation and Chemical Composition of
660 Different Coffees, Università degli studi di Torino, **2018**.
- 661

662 **Figures captions**

663 **Figure 1** Work-flows of chemical and sensory data treatment.

664

665 **Figure 2** The score plots obtained from the PCA of Volatiles (A) and Non-volatiles (B) data
666 respectively and (C) Coordinates of projection points from Multiple factor analysis (MFA). Legend
667 (C): B, C, J, U, In, Id, K, P indicate coffee origins; Se: sensory data; Ar: aroma data (volatile fraction
668 by HS-SPME-GC-MS); Ta: taste data (non-volatile fraction by LC-UV/DAD).

669

670 **Figure 3** HS-SPME-GC-MS of coffee powder (A) and an LC-UV/DAD coffee brew (the wavelengths
671 were set at 276 and 325 nm) (B), fingerprints of a coffee sample with high bitter notes. Grey bars
672 show features that were related to the bitter note and subjected to identification with I^T s, MS
673 commercial libraries and/or pure standards, or that were putatively-identified by LC-MS.

674

675 **Figure 4** Pearson's correlation matrix of volatile (Ar) and non-volatile (Ta) fractions represented by
676 a heat map, color scale from light green $\rho=+1$ to red $\rho=-1$, confidence level 5%.

677

Table 1 Multiple factor analysis results and RV correlation coefficients.

	Volatiles (Ar)	Non-volatiles (Ta)	Sensory (Se)	MFA
Volatiles (Ar)	1.000	0.921	0.549	0.951
Non-volatiles (Ta)	0.921	1.000	0.505	0.934
Sensory (Se)	0.549	0.505	1.000	0.748
MFA	0.951	0.934	0.748	1.000

Table 2 Comparison of the parameters of the PLS-regression models, in validation and prediction, that were obtained using aroma and taste, singularly, and data fusion (Volatiles+Non-Volatiles) for the six investigated notes. Models are built with specific selected features that were derived from PLS-DA analysis carried out on each sensory attribute following the strategy described in relative sections.

PLS model performance

Bitter	Volatiles	Non-volatiles	Fused data
<i>n°variables</i>	22	14	39
Q^2	0.742	0.666	0.692
R^2	0.892	0.810	0.888
<i>RMSECV</i>	0.579	0.659	0.575
<i>RMSEP</i>	1.073	0.929	1.120
Acid			
<i>n°variables</i>	22	10	26
Q^2	0.723	0.450	0.703
R^2	0.829	0.636	0.825
<i>RMSECV</i>	0.594	0.854	0.605
<i>RMSEP</i>	0.898	1.069	0.875
Flowery			
<i>n°variables</i>	20	14	27
Q^2	0.223	0.199	0.099
R^2	0.585	0.498	0.597
<i>RMSECV</i>	0.806	1.042	0.847
<i>RMSEP</i>	0.972	1.067	1.020
Fruity			
<i>n°variables</i>	19	16	39

Q^2	0.158	0.184	0.033
R^2	0.607	0.508	0.786
<i>RMSECV</i>	0.814	0.922	0.619
<i>RMSEP</i>	0.615	0.610	0.876
Spicy			
<i>n°variables</i>	22	16	32
Q^2	0.320	0.331	0.458
R^2	0.709	0.720	0.823
<i>RMSECV</i>	1.063	1.051	0.821
<i>RMSEP</i>	1.066	0.971	1.217
Woody			
<i>n°variables</i>	23	9	37
Q^2	0.708	0.472	0.706
R^2	0.879	0.714	0.885
<i>RMSECV</i>	0.798	1.228	0.782
<i>RMSEP</i>	0.920	0.948	1.129

Figure 1

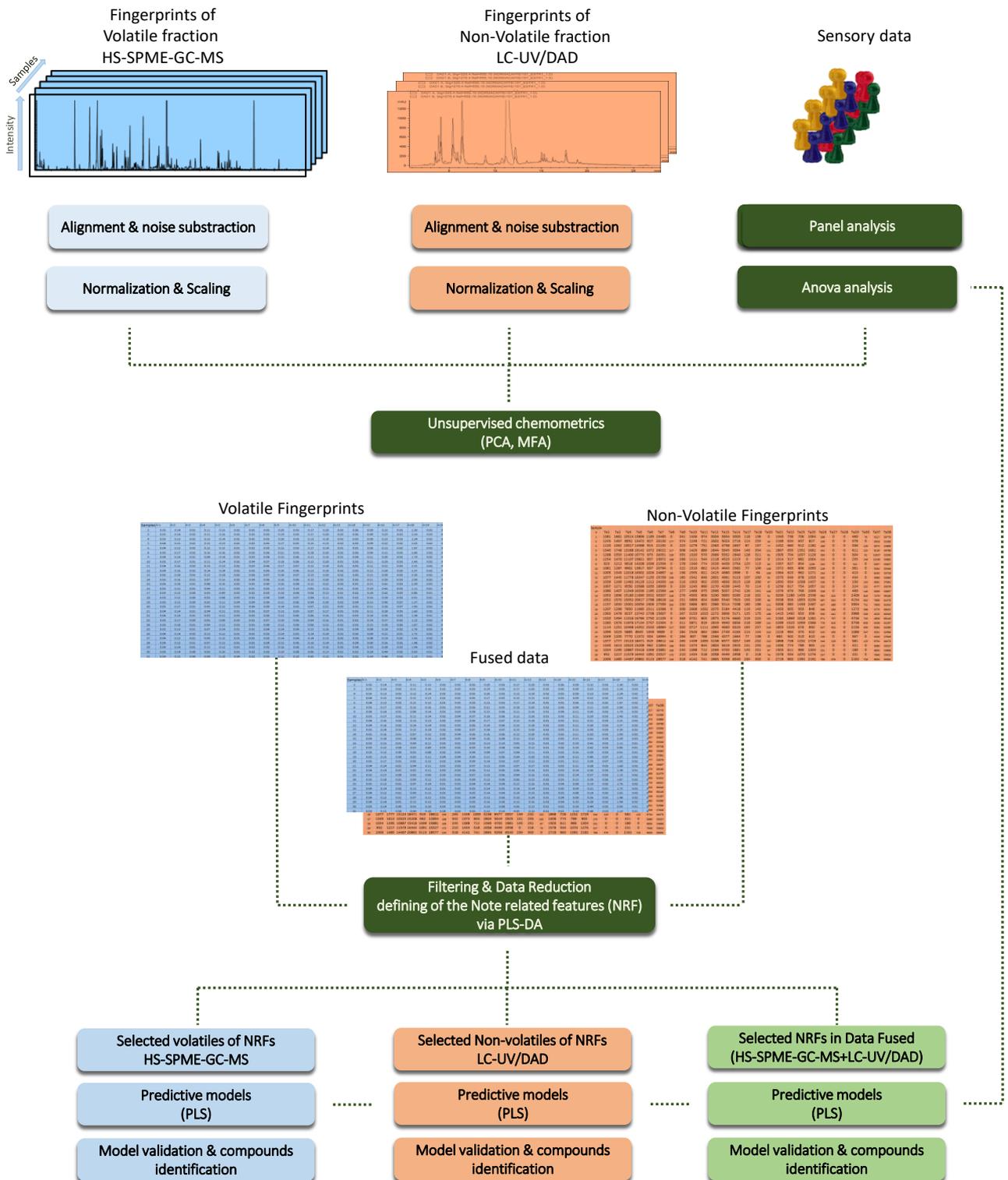


Figure 2

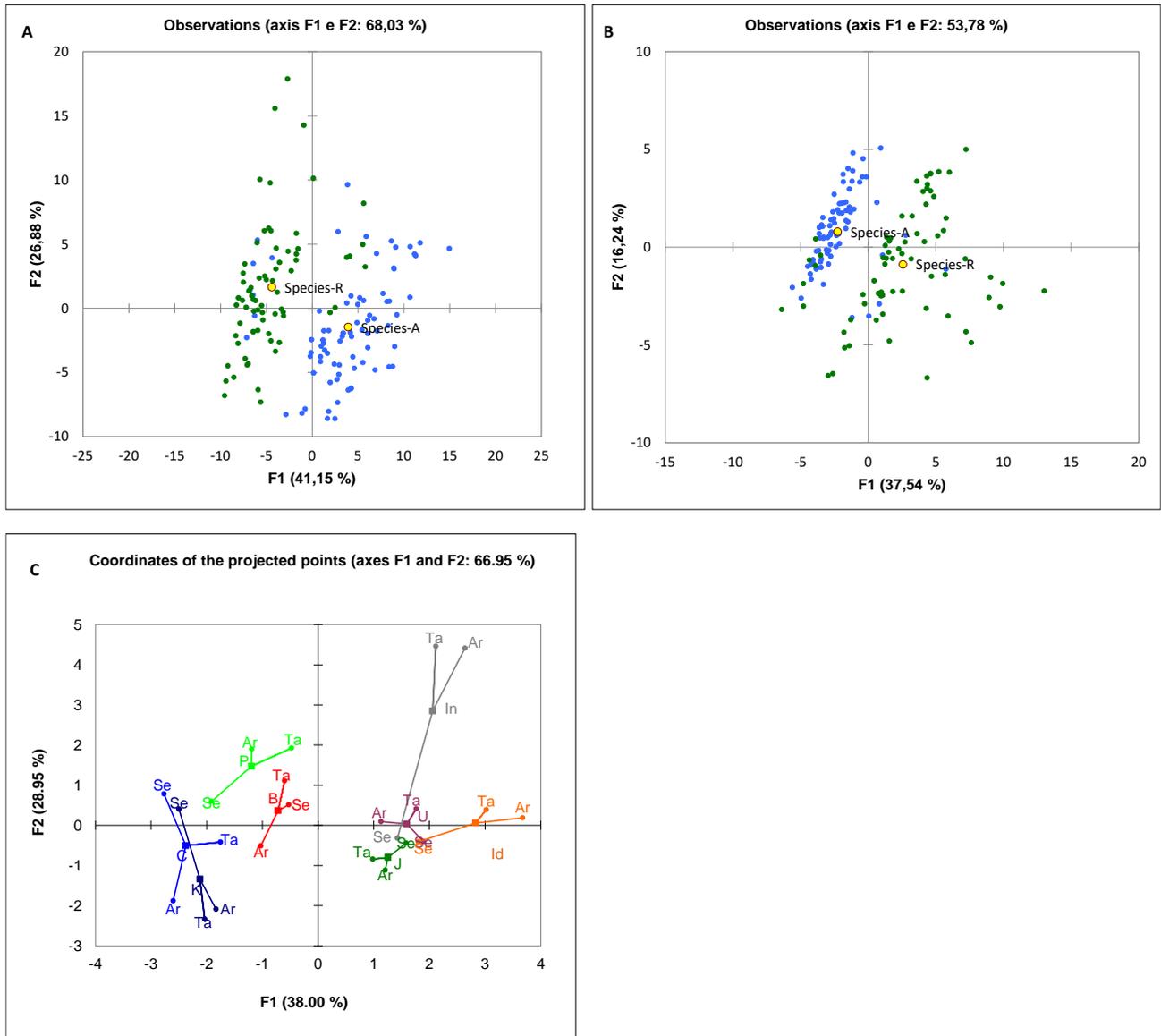


Figure 3

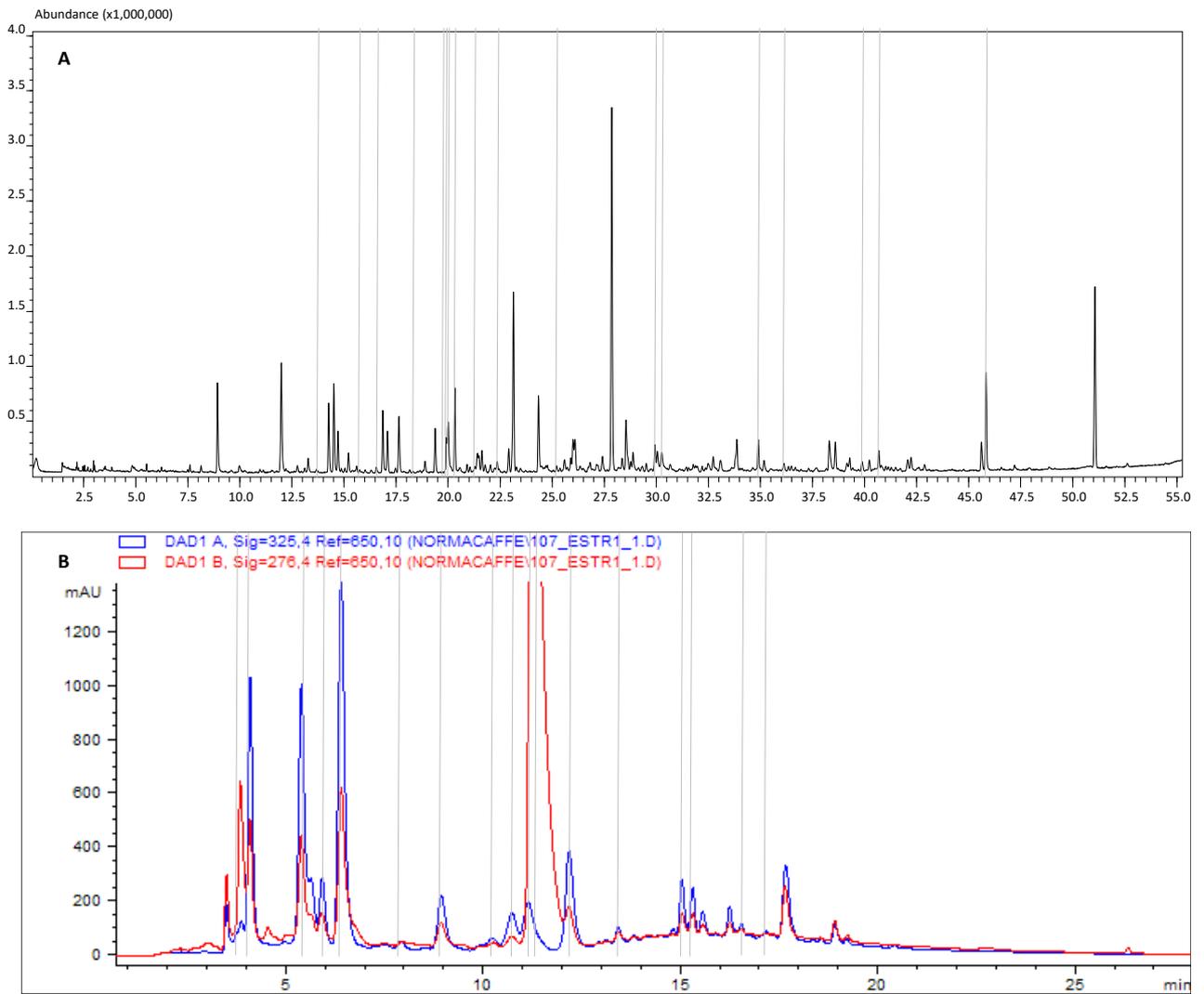
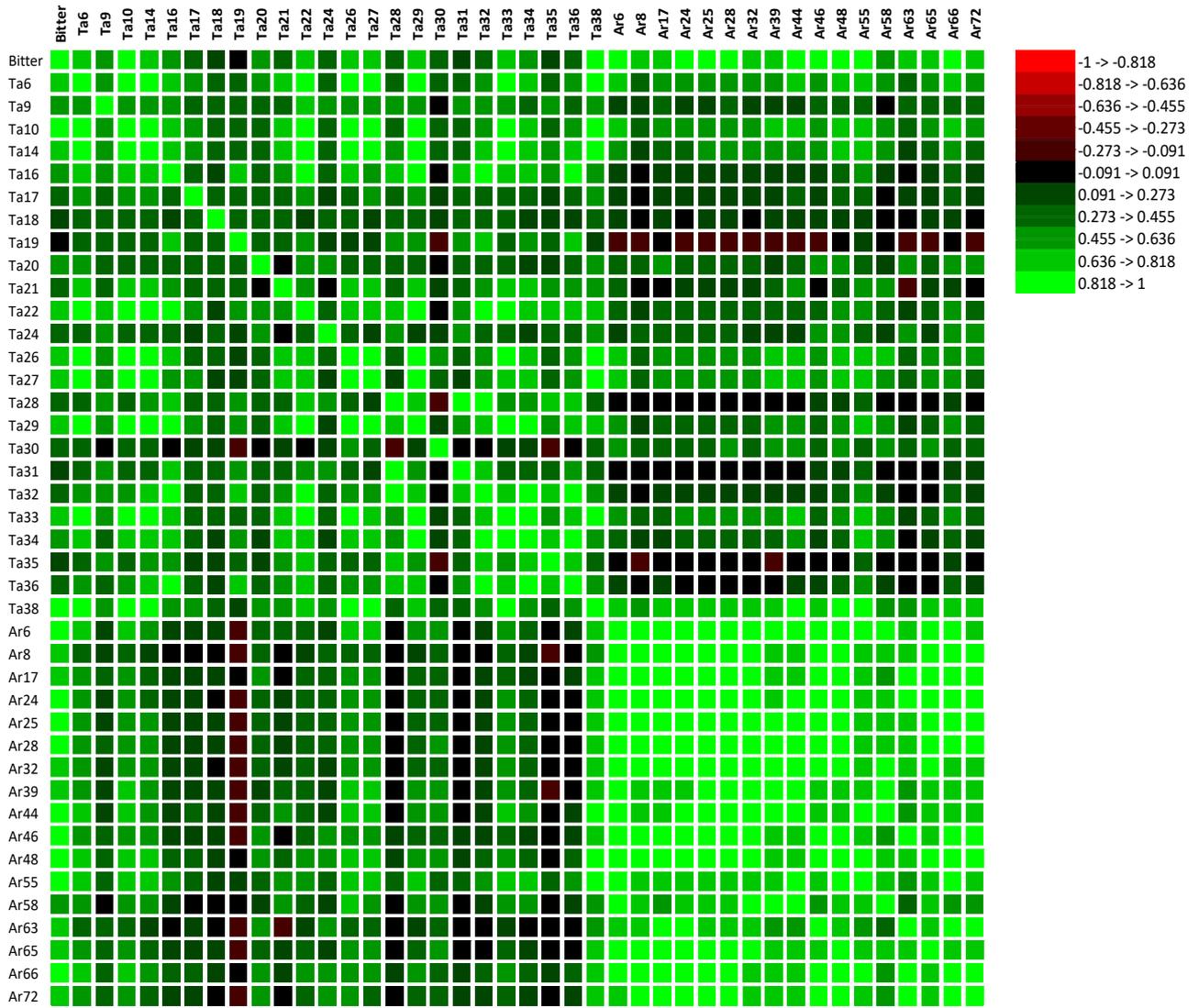


Figure 4



TOC graphic

