IRIS AperTO

UNIVERSITÀ DEGLI STUDI DI TORINO

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

**Cocoa quality: Chemical relationship of cocoa beans and liquors in origin identitation**

**This is a pre print version of the following article:**

*Original Citation:*

*Availability:*

This version is available http://hdl.handle.net/2318/1918351 since 2024-03-26T17:14:15Z

*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)

20 November 2024

Food Research International

# Cocoa quality: chemical relationship of cocoa beans and liquors in origin identitation

--Manuscript Draft--

| | |
|---|---|
| **Manuscript Number:** | |
| **Article Type:** | Research Paper |
| **Keywords:** | Quality, Cocoa beans and liquors, origin identitation, machine learning, fingerprinting, profiling |
| **Corresponding Author:** | erica liberto, Ph.D.<br>University of Turin Department of Pharmaceutical Science and Technology<br>torino, ITALY |
| **First Author:** | Eloisa Bagnulo |
| **Order of Authors:** | Eloisa Bagnulo |
| | Camilla Scavarda |
| | Cristian Bortolini |
| | Chiara Cordero |
| | carlo bicchi |
| | erica liberto, Ph.D. |
| **Abstract:** | In this study, HS-SPME-GC-MS was applied in combination with machine learning tools to the identitation of a set of cocoa samples of different origins. Untargeted fingerprinting and profiling approaches were tested for their informative, discriminative and classification ability provided by the volatilome of the raw beans and liquors inbound at the factory in search of robust tools exploitable for long-time studies. The ability to distinguish the country of origin on both beans and liquors is not so obvious due to processing steps accompanying the transformation of the beans, but this capacity is of particular interest to the chocolate industry as both beans and liquors can indifferently enter at different steps of the chocolate processing. Both fingerprinting (untargeted) and profiling (targeted) strategies enable to decipher of the information contained in the complex dataset and the cross-validation of the results, affording to discriminate between the origins with effective classification models. |
| **Suggested Reviewers:** | Luis Cuadros-Rodríguez, Prof<br>University of Granada<br>lcuadros@ugr.es<br>Prof. Cuadros is expert in analytical chemistry anc chemometrics |
| | Irene Chetschik, Prof.sa<br>University of Zurich<br>irene.chetschik@zhaw.ch<br>Prof. Chetschik has expertise in chocolate flavor |
| | Angela Marseglia<br>University of Parma<br>angela.marseglia@unipr.it<br>Prof. Marseglia has analytical expertise |

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

UNIVERSITÀ DEGLI STUDI DI TORINO
Dipartimento di Scienza e Tecnologia del Farmaco
Via Pietro Giuria 9 - 10125 – Torino Italy
Phone: 0039 011 6707172
Fax: 0039 011 6707162
Web site: www.http://www.dstf.unito.it/

Editorial Office
Food Research International

Dear Editor,
I'm sending you the manuscript entitled

"***Cocoa quality: chemical relationship of cocoa beans and liquors in origin identitation"***
*by Eloisa Bagnulo, Camilla Scavarda, Cristian Bortolini, Chiara Cordero, Carlo Bicchi, Erica Liberto**

In this study, untargeted fingerprinting and profiling approaches were tested for their informative, discriminative and classification ability provided by the volatilome of the raw cocoa beans and liquors inbound at the factory in search of robust and objective tools exploitable for long-time studies. HS-SPME-GC-MS combined with machine learning tools has been applied to the identitation of a set of cocoa samples of different origins. Untargeted fingerprinting capability and, at the same time, the potentiality of profiling in identitation of origin were tested and cross-validated as well as the chemical information they can provide from the volatilome of both beans and liquors. This information is important on an industrial level for the development of an artificial smelling machine to check the quality of incoming beans and liquors and to fulfil the flavour reference driver in product design.

The ability to distinguish the country of origin on both beans and liquors is not so obvious due to processing steps accompanying the transformation of the beans, but this capacity is of particular interest to the chocolate industry as both beans and liquors can indifferently enter at different steps of the chocolate processing. Both fingerprinting (untargeted) and profiling (targeted) strategies enable deciphering the information in the complex dataset and the cross-validation of the results, affording to discriminate between the origins with effective classification models.

The paper is unpublished and is not under consideration for publication elsewhere; please do not hesitate to contact us if further questions arise.

Looking forward to hearing from you
Sincerely yours
Erica Liberto

Address for correspondence: Prof.sa, PhD Erica Liberto,
Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino, Via Pietro Giuria 9, I-10125 Torino, Italy; Tel. +39 011 670 7134; e-mail: erica.liberto@unito.it

**Highlights**

- Cocoa volatilome of beans and liquors is a tool for origin identitation and authentication
- Identitation at the molecular level is an objective way to qualify flavour in long time studies
- Machine learning affords to define origin classification model for the chemical-sensory identitation
- Fingerprinting and profiling approaches provided comparable classification performances
- Targeted approach is mandatory when a quality certification of origin is requested

Manuscript Click here to view linked References

# Cocoa quality: chemical relationship of cocoa beans and liquors in origin identitation

Eloisa Bagnulo[1], Camilla Scavarda[3], Cristian Bortolini[2], Chiara Cordero[1], Carlo Bicchi[1], Erica Liberto[1]*

[1] Dipartimento di Scienza e Tecnologia del Farmaco, Università degli Studi di Torino, Turin, Italy

[2] Soremartec Italia S.r.l. (Ferrero Group), P.le P. Ferrero 1, 12051 Alba (CN), Italy

[3]A. Costantino & C. S.P.A, Via Francesco Romana, Favria, TO, Italy

**ABSTRACT**

In this study, HS-SPME-GC-MS was applied in combination with machine learning tools to the identitation of a set of cocoa samples of different origins. Untargeted fingerprinting and profiling approaches were tested for their informative, discriminative and classification ability provided by the volatilome of the raw beans and liquors inbound at the factory in search of robust tools exploitable for long-time studies. The ability to distinguish the country of origin on both beans and liquors is not so obvious due to processing steps accompanying the transformation of the beans, but this capacity is of particular interest to the chocolate industry as both beans and liquors can indifferently enter at different steps of the chocolate processing. Both fingerprinting (untargeted) and profiling (targeted) strategies enable to decipher of the information contained in the complex dataset and the cross-validation of the results, affording to discriminate between the origins with effective classification models.

Keywords: Quality, Cocoa beans and liquors, origin identitation, machine learning, fingerprinting, profiling

## 1. INTRODUCTION

Cacao (*Theobroma cacao* L.) is a perennial tropical crop economically important for the countries growing it. It is also a raw material of great relevance for different fields including confectionary and functional food and beverages (cocoa and chocolate derivatives) that cover more than 60% of the market (Research and Markets, 2021). The flavour is one of the main features linked to chocolate product quality besides brand and cost (CABISCO/ECA/FCC, 2015). Cocoa beans and the post-harvest treatments are uniquely responsible for the flavour and nutritional components of cocoa derivatives, which implies that the cocoa chain must be supplied with products with consistent standards of quality, i.e. uniform raw materials with specific quality parameters from the country of origin. This is a serious issue for chocolate manufacturers since they need bulks of raw or semi-finished cocoa products of consistent quality to respond to the demand.

However, standardisation of quality over time is difficult to achieve as cocoa beans are mostly produced in non-EU countries by a large number of independent farmers, resulting in remarkable fragmentation and heterogeneity of batches where socio-political instability and the impact of climate change negatively affect cocoa production and farm survival (Boeckx, Bauters, & Dewettinck, 2020; Danezis, Tsagkaris, Brusic, & Georgiou, 2016; Lahive, Hadley, & Daymond, 2019; Medina, Perestrelo, Silva, Pereira, & Câmara, 2019; Somarriba et al., 2021). Objective and robust tools to trace the authenticity and the quality stability of cocoa products are therefore necessary to support the continuity year-to-year in an ever-increasing global demand at the industrial level.

Different instrumental analytical methods have so far been applied to sustain cocoa-origin authentication based on bioactive components or flavour-related compounds (Febrianto & Zhu, 2022; Kumar et al., 2022; Magagna et al., 2017; Marseglia, Musci, Rinaldi, Palla, & Caligiani, 2020). Spectroscopic and spectrometric-based methods have recently been reported for cocoa, and more in general, for food and fingerprinting analysis (Gutiérrez, 2017; Buertt, Harris, & Klevorn, 2016; Medina, Perestrelo, et al., 2019a; Scavarda et al., 2021; Acierno, Alewijn, Zomer, & van Ruth, 2017). In food authentication fingerprinting and profiling strategies are becoming widely used and accepted to monitor food integrity

54 (Ballin & Laursen, 2019; Luis Cuadros-Rodríguez, Ortega-Gavilán, Martín-Torres, Arroyo-Cerezo, &
55 Jiménez-Carvelo, 2021; Medina, Pereira, Silva, Perestrelo, & Câmara, 2019; Suman, Cavanna, Sammarco,
56 Lambertini, & Loffi, 2021). Indeed, food authentication is often based on determining the degree of
57 similarity of the fingerprints of some diagnostic chemical characteristics of an unknown sample compared
58 to a representative reference sample (Cuadros-Rodríguez,Ruiz-Samblás,Valverde-Som,Pérez-
59 Castaño,&González Casado, 2016). Consequently, fingerprinting and profiling approaches require a
60 suitable number of pure and authentic samples to establish a representative database of the "genuine"
61 food population (Danezis et al., 2016). This step is known as food 'Identitation' and, if correctly carried
62 out, affords a reliable food authentication (L. Cuadros-Rodríguez et al., 2016). Furthermore, flavour quality
63 identitation requires analytical methods able to provide diagnostic detailed profiles correlated with the
64 sensory features that can be monitored and quantified for an objective assessment in quality control (QC)
65 (Bressanello et al., 2021; Bressanello et al., 2018; Magagna et al., 2017; Perotti et al., 2020; Stilo et al.,
66 2021). Chromatographic hyphenated platforms combined with artificial intelligence are the tools of choice
67 to obtain significant information encrypted in complex data sets to delineate significant trends and/or
68 data structures. In addition, they are of crucial importance in reliably assessing associations and/or
69 correlations between the chemical composition of foodstuffs and food extracts to determine compliance
70 with quality and legal standards for the authentication of samples (Bressanello et al., 2018; Bressanello et
71 al., 2017; Rodionova & Pomerantsev, 2020).

72 In this study, HS-SPME-GC-MS combined with machine learning tools has been applied to the
73 identitation of a set of cocoa samples of different origins. Untargeted fingerprinting capability and at the
74 same time the potentiality of profiling in identitation of origin were tested as well as the chemical
75 information they can provide from the volatilome of cocoa beans and liquors. This information is
76 important on an industrial level for the development of an artificial smelling machine to check the quality
77 of incoming beans and liquors and to fulfil the flavour reference driver in product design. The study is the
78 first step of a wider project that aims 1) to investigate the flavour profiles of different origins and define
79 their chemical-sensory identity card; 2) to guarantee consistent supplies from different origins over time
80 independently of climate change and geo-political conflicts; and 3) to ensure final standard quality of the

81 cocoa products from a sensorial and qualitative point of view by blending different origins through
82 objective assessments using chemical analysis and machine learning when supply difficulties may occur.

83

84 ## 2. EXPERIMENTAL

85 *2.1 Cocoa samples and reference compounds*

86 A total of 160 samples of unroasted beans (n=85) and liquors (n=75) (*Theobroma cacao* L.) of
87 cocoa were analysed. Samples were provided by Soremartec Italia s.r.l. (Alba, Italy) and were from 4
88 origins: Colombia (COL), Cameron (CAM) and West Africa (WA, a blend of Ghana, Ivory Coast and
89 Nigeria) all from *Forastero* variety, Ecuador (ECU) from CCN51 Clone, harvested in 2015 and 2016 by
90 different local regional farmers. Cocoa was of commercial grade (beans size "standard" based on
91 counting test under guidelines of the Federation of cocoa commerce) (FCC, 2015). Samples were
92 ground in liquid nitrogen to obtain a homogeneous powder and then stored at -80°C until analysis.

93 Pure reference standards for identity confirmation (key-aroma compounds (§) and informative
94 volatiles) as reported in Table A1, normal alkanes (*n*-alkanes *n*-C9 to *n*-C25) for Linear Retention Index
95 ($I^T_S$) determination and α-thujone as internal standard (ISTD) were from Merk (Milan, Italy). An α-
96 thujone solution (ISTD) at a concentration of 1000 mg/L was prepared in degassed sunflower seed oil
97 and stored in a sealed vial at -18°C.

98

99 *2.2 Automated Head Space Solid Phase Micro Extraction: sampling device and analysis conditions*

100 Automated Headspace Solid Phase Microextraction (auto-HS-SPME) was performed using a
101 Combi-PAL AOC 5000 (Shimadzu, Milan, Italy) online integrated with a Shimadzu QP2010 GC–MS
102 system provided with Shimadzu GC–MS Solution 2.51 software (Shimadzu, Milan, Italy). SPME fibre:
103 Divinylbenzene/Carboxen/Polydimethylsiloxane (DVB/CAR/PDMS) df 50/30 µm - 2 cm length from
104 Millipore (Bellefonte, PA, USA). Fibres were conditioned before use as recommended by the

105 manufacturer. The standard-in-fibre procedure was adopted to pre-load the ISTD (α-thujone) onto
106 the fibre before sampling (Wang, O'Reilly, Chen, & Pawliszyn, 2005). 5.0 µL of ISTD solution were
107 placed in a 20 mL glass vial and submitted to HS-SPME at 50°C for 20 min, stirring speed 350 rpm.
108 Cocoa powder (1.00 g) was weighed in the headspace glass vials (20 mL) and submitted to automated
109 HS-SPME sampling. After ISTD loading on the fibre, the SPME device was exposed to the headspace
110 of cocoa for 40 min at 50° at a shaking speed of 350 rpm.

111 *GC-MS analysis- Chromatographic conditions*: analyses were run on a Shimadzu QP2010 GC–MS
112 system, controlled by Shimadzu GC–MS Solution 2.5SU1 software (Shimadzu, Milan, Italy) Injector
113 temperature: 240°C, injection mode: splitless; carrier gas: helium, flow rate: 1 mL/min. Sampled
114 analytes were recovered by thermal desorption into the split/splitless (S/SL) injection port of the GC
115 system at 240°C for 5 min. GC column: SolGelwax (100% polyethene glycol) 30 m x 0.25 mm $d_c$ x 0.25
116 µm $d_f$ Trajan Analytical Science (Ringwood, Australia). Temperature program, from 40°C (2 min) to
117 200°C at 3,5°C/min, then to 240°C (5 min) at 10°C/min. MSD conditions: ionization mode: EI (70 eV);
118 temperatures: ion source: 200°C; quadrupole: 150°C; transfer line: 260°C; scan range: 35-350 amu.
119 Each sample was analyzed in triplicate.

120 *2.6 Analytes identification and data analysis*

121 Untargeted fingerprinting data elaboration, Principal Component Analysis (PCA), Partial Least
122 Square Discriminant Analysis (PLS-DA) and regression analysis, was carried out with Pirouette®
123 (Comprehensive Chemometrics Modelling Software, version 4.5-2014) (Infometrix, Inc. Bothell, WA).
124 Heat map and hierarchical clustering were obtained with Morpheus
125 (https://software.broadinstitute.org/morpheus/). Targeted analysis was on 55 compounds identified
126 by matching their EI-MS fragmentation patterns (NIST MS Search algorithm, version 2.0, National
127 Institute of Standards and Technology, Gaithersburg, MD, USA, with Direct Matching Factor > 900)
128 with those stored in commercial (NIST2014 and Wiley 7n) and in-house databases, and/or with
129 reference standards available in the laboratory. Linear retention indices ($I^T_S$) were taken as a further
130 parameter to support identification and experimental values were compared to tabulated values

131 **(Table A1 in Appendix A)**. Profiling data elaboration and the Kruskall-Wallis test were performed with
132 XLSTAT version 2021.4.1 statistical and data analysis solution. (Addinsoft (2022), New York, USA.
133 https://www.xlstat.com/en)

134

135 ## 3. RESULTS AND DISCUSSION

136 *3.1 Untargeted fingerprinting approach and origin identity*

137 Untargeted fingerprinting approaches are time-saving and can be exploited in screening for
138 authentication and very recently in detecting food fraud thanks to platforms with highly informative
139 power (Greño, Plaza, Luisa Marina, & Castro Puyana, 2023; Suman et al., 2021). The increasing
140 difficulties in supplying high-quality raw materials due to the climate and social instability crisis have
141 prompted the large-scale industry to look for new sources or new strategies to maintain or improve
142 the standard quality of raw and finished products. Cocoa volatilome is the final expression of the
143 different "chemical and biochemical activities" occurring in the beans as a result of the varieties,
144 agronomic practices, post-harvest treatments and industrial processing (Lytou, Panagou, & Nychas,
145 2019). With these perspectives, untargeted volatilomics can be applied to define the fingerprint of
146 quality standard and/or to benchmark with a quality reference e.g. linked to the origin. In particular,
147 the origin benchmark based on the similarity of the volatilome can be a useful strategy to deepen the
148 knowledge of the origin, to evaluate the yields with a view on the compatibility with large chocolate
149 productions, to detect the defects and evaluate roasting profiles or to define blending strategies to
150 standardize the reference quality. To define an origin identity based on the cocoa volatilome it is
151 necessary to have representative reference samples of each investigated origin. The reliability of the
152 identitation process depends on the parameters influencing the origin characteristics included within
153 the reference samples (i.e. seasonality, regional variations, post-harvest treatments, farm of
154 production, etc.).

155 Exploratory data analysis on untargeted chromatographic fingerprints shows a better origin
156 description with beans than with liquors (**Figure 1 a and b)** with respectively 74.3 and 59.7 % of
157 explained variance, but in any case, with very good results. This agreement was not obvious since
158 liquors undergo further processing steps that tend to homogenize origin information. Supervised
159 approach by PLS-DA for beans based on a training set of 66 samples internally cross-validated (CV 5)
160 and an external test set of 18 samples shows excellent results with 100% of correct classification
161 within origins. An OPLS-DA model cross-validated (CV 4) based on a liquor training set of 58 samples
162 presents a total ability in origin classification of 88.24% when applied to the external test set (17
163 samples), with high specificity for all origins but a lower sensitivity (71%) for Colombia (COL) compared
164 to other origins (**Figure 1c and d**).

165 This approach is very similar to others applied for origin discrimination but in addition shows that
166 beans and liquors provide similar chemical information (Acierno, Yener, Alewijn, Biasioli, & Van Ruth,
167 2016; Liu et al., 2017; Medina, Perestrelo, et al., 2019; Torres-Moreno, Tarrega, & Blanch, 2014;
168 Acierno et al., 2017). The ability to distinguish the country of origin of both beans and liquors is of
169 particular interest to the chocolate industry, as either beans or liquors can indifferently be processed
170 by the chocolate factory depending on the country of origin. Despite these excellent results in origin
171 identitation, the untargeted strategies do not provide detailed information about the aroma chemical
172 profile that, on the contrary, enables an in-depth knowledge of the aroma components that
173 characterise the standards quality for cocoa flavour.

174

175 *3.2 Profiling and origin identity: chemical information provided from beans and liquors*

176 The flavour is an essential criterion of quality for manufacturers of cocoa products
177 (CABISCO/ECA/FCC, 2015). It represents an overall sensation that emerges from the interaction of
178 taste, odour and textural feeling and it results from three classes of compounds, i.e. those responsible
179 for taste (generally non-volatile compounds), for odours (aroma volatile substances) or both of them
180 (Belitz, Grosch, & Schieberle, 2009). Cocoa quality and economic value are strictly related to unique

181 and complex flavours. Generally, the sensory profile (aroma, taste, mouth feel, and texture) is a key
182 factor in obtaining premium-quality products meeting consumer preferences. Several chemical
183 compounds are involved in the flavour composition (aldehydes, ketones, esters, alcohols, pyrazines,
184 quinoxalines, furans, pyrones, lactones, pyrroles, and diketopiperazines) and their concentration
185 strictly depends on the components formed during processing steps of the cocoa supply chain
186 (fermentation, drying, roasting etc.) (Aprotosoaie, Vlad Luca, & Miron, 2016). Therefore, the content
187 of these components is related to the processing methods applied both in the countries of origin and
188 by the chocolate manufacturer. In this respect, flavour composition can be evaluated to find a
189 characteristic chemical profile enabling to recognition cocoa beans and liquors coming from a specific
190 country to authenticate incoming raw materials from a sensory point of view. Furthermore, a more
191 detailed analysis of the volatile components is required when the aim is to find similarities in aroma
192 profiles to create similar blends and to know the chemical components responsible for different
193 sensory properties. Cocoa aroma is described by "generalist" odorants that are in common with other
194 foods with which they share the same processing such as fermentation and thermal treatment.
195 However, "individualist" odorants can be derived from particular processing methods and/or origins
196 (Dunkel et al., 2014). Moreover, several non-odorant compounds can significantly affect the sensory
197 profiles through the so-called odour synaesthesia (Chambers & Koppel, 2013; Prescott, 2015).
198 Therefore, gas chromatographic analysis of cocoa volatilome benefits from detailed profiling, which
199 can then be used for correlations with sensory data and characterisation of the chemical odour code.
200 From these points of view, it became interesting to evaluate the chemical information provided by
201 beans and liquors, because, in general, the panel(s) taste the liquors to evaluate the cocoa flavour.
202 Despite the further processing that the liquors undergo compared to the beans, the analytical
203 patterns of volatiles are similar when describing the origin of samples, albeit to different quantitative
204 extents, as shown in the heat map in **Figure 2**. Figure 2 shows the chromatographic profiles in plan
205 view, where L and B indicate liquors and beans, and the volatile distribution depending on the origins.
206 The clustering of the origins is based on the agglomerative average linkage method and the one-minus
207 Pearson correlation distance on the zeta score-adjusted data matrix. The colour scale ranges from red

208 (volatile substances in larger quantities) to blue (volatiles in smaller quantities). The representative
209 abundance of the volatilome of both beans and liquors for the different origins under study is
210 reported on the tree chart in Figure A1 in **Appendix A**.

211 Normalized responses of the main volatiles in beans and liquors were transformed by logarithmic
212 function and pre-processed by Pareto scaling before PCA analysis. Explorative data elaboration by PCA
213 of the target volatiles displays similar power in the origin discrimination both for beans and liquors on
214 the first 3 principal components (PCs) confirming an existing correlation between the two patterns of
215 volatiles **Figure 3 a and b**.

216 Beans from different countries of origin have different ratios between volatiles, which mainly
217 reflect the origin and primary processing in the country of origin (mainly fermentation and drying), as
218 we can see from the boxplots of the different chemical classes in **Figure 4**. The content of volatiles of
219 cocoa beans from Colombia was somewhat lower than in the other countries of origin. Factors such
220 as the system of storage of the pods and its duration influence the pH, titratable acidity and
221 temperature reached during fermentation affecting enzyme activities and flavour development
222 (Afoakwa, Paterson, Fowler, & Ryan, 2008; Aprotosoaie et al., 2016). Further steps in the processing
223 of the beans (e.g. roasting, alkalisation, batch blending, milling) affect the overall profile, especially
224 the abundance and ratios of volatiles. The relative content of volatiles was highest in cocoa liquors
225 from Ecuador as also reported by Lie et al. (Liu et al., 2017). The processing of the beans leads to a
226 quali-quantitative "flattening" of some volatiles between the origins, e.g. aromatics and aldehydes,
227 while it increases the differences in esters, alcohols, terpenes, ketones and acids although with a
228 different trend depending on the origins. For example, acids decrease more in COL and WA, but less
229 in ECU and CAM liquors, and they change their distribution with the origins. Acids are generated in
230 the fermentation process, in particular, acetic acid whose content is in strict relationship with the
231 fermentation conditions (Schwan & Wheals, 2004). Esters are an important class of volatiles formed
232 from amino acids in unroasted beans that impart fruity and floral aromas whose content is negatively

233 influenced by roasting. Liquors from Ecuador contain higher amounts of esters compared to other
234 origins (**Figure 5**).

235 However, supervised methods are needed to classify and predict origin based on chemical
236 characteristics (Casale M. et al., 2020). The main objective of supervised machine learning is to model
237 data to obtain qualitative prediction models, and thanks to dedicated algorithms to define target
238 compounds that are characteristic of the origin. The Variable Importance in Projection (VIP), for
239 example, is a well-known method to find significant variables in complex data sets. It enables to select
240 variables from the PLS model by calculating the VIP scores for each chemical variable and removing
241 those with a VIP score below a predefined threshold (default=1) (Medina, Perestrelo, et al., 2019).

242 Orthogonal signal correction Partial Least Square Discriminant Analysis (OPLS-DA) is a supervised
243 approach giving excellent results in the origin classification and identification of components
244 potentially associated with both beans and liquors with a classification rate of over 92% (i.e. 92.86%
245 for beans and 92.31% for liquors) **Figures 3 c and 3 d**. The 20% of samples were selected by the
246 Kennard-Stone algorithm within each sample set (i.e. beans and liquors) to create the external test
247 sets. OPLS-DA was built on the training data sets (65 samples for beans and 60 samples for liquors),
248 log10 transformed, autoscaled and cross-validated (10 CV). The models were then applied to the
249 external test sets (i.e. 14 beans and 13 liquors) (**Figures 3 c and d**). The results of origin classification
250 were similar to the untargeted fingerprinting approach but with a better classification rate in
251 particular for liquors.

252 **Figure 6 a and b** reports the VIP graphics from the OPLS-DA both from beans and liquors profiling, the
253 scale is from 0 and 1.8, but only those higher than 1 were selected as important for origin
254 classification. Several VIP (VIP>1) compounds are in common between beans and liquors:
255 Acetophenone, Benzaldehyde, 3-Methylbutanoic acid, 2,3-Butanediol, Tetramethylpyrazine, 2-
256 Heptanone, Styrene, 4-Heptanol **Figure 6 a and b**. These volatiles are important variables in the origin
257 classification and share similar relative abundance in beans and liquors of the same sample (**Figure**
258 **A2 in Appendix A**). Benzaldehyde, Tetramethylpyrazine and 3-Methylbutanoic acid have been also

evidenced as aroma-active compounds in liquors for WA and Cameroon besides other origins from Liu et al., while Isobutyl acetate, Benzaldehyde, Trimetylpyrazine, Butanal, 2 and 3-methyl-Butanal, β-Myrcene and 2-heptanone were indicated by Marseglia et al. as characteristic compounds both of unroasted and roasted beans although at different extents, higher in roasted beans (Liu et al., 2017; Marseglia et al., 2020). Tetramethylpyrazine whose odour quality is described as *cocoa*, *chocolate - coffee*, comes from Maillard reactions that can also occur in the beans during fermentation and drying due to the temperatures reached in this stage. However, roasting is the main factor contributing to its formation in liquors. WA cocoas result richer in Tetramethylpyrazine compared to other origins both in beans and in liquors and generally higher in liquors (**Figure 6c**). WA cocoa beans, and more in general from African regions, result richer in Acetophenone responsible for the *floral* and *sweet* notes among the ketones, and in 3-Methylbutanoic acid (*sweaty, rancid*), Trimethylpyrazine (*cocoa, roasted nutty*), Benzaldehyde (*bitter almond*), 4-Heptanol (*alcoholic*) (**Figure 6c**), confirming the observation of Marseglia et al. although their relative distribution, compared to other origins, are not respected for all compounds in liquors (Marseglia et al., 2020; Tuenter et al., 2020).

For example, some 2-Heptanol (*citrus*), 2-Heptanone (*sweet, fruity*) and 2-Heptanol acetate (*fruity*) tend to be characteristics of liquors from South American countries (**Figure 6c**), which agrees with Marseglia et al. but contrast with the findings of Tuenter et al. who compared West Africa bulk with Ecuadorian fine-flavour cocoas (Marseglia et al., 2020; Tuenter et al., 2020).

Besides these volatiles, a group of monoterpenoids contribute significantly to the definition of the origin of liquors, the main ones of them being α-pinene (*woody, resinous-piney*), δ-3-carene (*sweet citrus*), *trans*-β-ocimene (*floral*), β-myrcene (*balsamic, must, spicy, sweet*), limonene (*citrus-like*) and linalool (a key-odorant of cocoa conferring a *floral, leafy, and tea-like* note) **Figure 6 and b**. Although these volatiles are associated with odour description very far from the cocoa flavour and have a high odour threshold, they do affect its overall aroma (Liu et al., 2017; Ullrich et al., 2022). In addition, some of the above monoterpenoids are chiral compounds and their enantiomers can have different odour thresholds and/or sensory descriptions, and can differently influence the cocoa aroma, thus

285 making of particular interest the determination of their enantiomeric distribution in cocoa products
286 (Cagliero, Sgorbini, Cordero, Liberto, Rubiolo, Bicchi, 2017). These components derive from the
287 monoterpenoid biosynthetic pathway and, during fermentation, they are present in higher amounts
288 in beans than in pulp (Chetschik et al., 2018). Monoterpenoids have not yet been studied in depth in
289 cocoa, but they can play a marked role in defining/monitoring i) the origin through the liquor
290 volatilome, ii) the different fermentation processes, and iii) more in general, the post-harvest
291 processing in the countries of origin (Aprotosoaie et al., 2016; Afoakwa et al., 2008). For instance,
292 Calva-Estrada et al. have recently shown that limonene has an important role in the composition of
293 the dark-chocolates aroma profile from Latin American cocoas (Calva-Estrada, Utrilla-Vázquez,
294 Vallejo-Cardona, Roblero-Pérez, & Lugo-Cervantes, 2020).

295

296

297 **4. CONCLUSIONS**

298 The volatilome of cocoa beans and liquors is a very rich source of information to define the origin
299 of incoming raw material. Both fingerprinting (untargeted) and profiling (targeted) strategies allow
300 to decipher the information contained in the complex dataset resulting from HS-SPME-GC-MS
301 analysis, and the cross-validation of the results enables to discriminate between the origins of both
302 beans and liquors studied with suitable classification models. Fingerprinting and profiling approaches
303 provided comparable classification performance. The untargeted approach can be used for rapid
304 screening for quality control and for a preliminary evaluation of incoming lots, but it does not provide
305 chemical information. On the other hand, profiling is more time-consuming but provides detailed
306 chemical information on the samples, enabling the definition of the chemical marker(s) of origin and
307 is very useful, in particular, if carried out in connection with the sensory profile. Targeted methods
308 have allowed the identification of several origin-related key odorants in both bean and liquor
309 samples. Furthermore, the targeted approach enables the establishment of correlations with the
310 chemical pathways involved with the volatile formation. This information can be linked to the soil

311 and climate characteristics, and in-situ processing of the world region under consideration before
312 the next steps of the chocolate production chain. However, volatile fingerprinting to be reliable and
313 to be used as a robust authentication tool for origin requires a large number of representative
314 samples, including the variability of cocoa products, to build up a representative cocoa chemical
315 identity card correlated to the sensory profile. The information provided by the untargeted approach
316 can thereby be sufficient for the basic authentication of cocoa volatiles for routine control quality to
317 standardize the quality year-to-year. On the other hand, the targeted approach is mandatory when
318 a certification of origin is requested, as it requires an identity card detailing the chemical and sensory
319 characteristics of the origin itself to fulfil the flavour reference driver in product design.

320 **Acknowledgements**

321 The authors acknowledge the financial support of PhD program of the University of Turin and
322 Soremartec Italia Srl. The attribution is based on the PNRR, Mission 4, component 2 "From Research
323 to Business" - Investment 3.3 "Introduction of innovative doctorates that respond to the innovation
324 needs of companies and promote the recruitment of researchers from companies" DM 352/2022.

325

326 **Appendix A. Supplementary data**

327 **Table A1** Targeted volatiles identified with their experimental and literature retention indices ITs, Target
328 ion (Ti) and qualifier ions (Qis) and their mass spectral similarity index (SI). *Volatiles confirmed by the
329 reference standard (A), Relative retention index (RI), and Mass spectrum similarity (MS). [§] Key aroma
330 compounds.

331 **Figure A1** Tree chart of the comparison of the cocoa volatilome of beans and liquors

332 **Figure A2** Box plots of the compounds from OPLS-DA (with VIP>1) in describing the origins common to
333 both beans and liquors.

334

335 **Captions to figures**

336 **Figure 1** Untargeted data mining results: a) and b) are PCA scores plots of beans and liquors respectively,
337 and c) and d) are the results from the OPLS-DA Class Predicted plot for beans and liquors. Pre-
338 process: autoscale. Class 1: CAM, Class 2: COL, Class 3: ECU, Class 4: WA

339 **Figure 2** Heat-map of the target volatiles in describing the different origins; clustered data by Z score and
340 Pearson correlation. Colour rendering from low (blue) to high amount (red). L: target liquor
341 volatiles, B: target bean volatiles.

342 **Figure 3** Targeted metabolites profiles from PCA a) and b) scores plots of beans and liquors respectively,
343 c) and d) are the results from the OPLS-DA Class Predicted plot for beans and liquors. Pre-process:
344 autoscale. Class 1: COL, Class 2: ECU, Class 3: CAM, Class 4: WA

345 **Figure 4** Box-plots of the comparison of the composition of the beans by representative chemical classes.

346 **Figure 5** Box-plots of the comparison of the composition of liquors according to representative chemical
347 classes.

348 **Figure 6** VIP (higher than 1) from the OPLS-DA from the OPLS-DA on targeted profiling for beans and
349 liquors respectively in a), in b) Venn diagram of volatiles describing the origins with VIP>1 from
350 the OPLS-DA of beans and liquors, c) heat map of the VIP distribution between the origins.

351

352

353 **References**


354 Acierno, V., Alewijn, M., Zomer, P., & van Ruth, S. M. (2018). Making cocoa origin traceable: Fingerprints
355 of chocolates using Flow Infusion - Electro Spray Ionization - Mass Spectrometry. *Food Control*, *85*,
356 245–252. https://doi.org/10.1016/j.foodcont.2017.10.002

357 Acierno, V., Yener, S., Alewijn, M., Biasioli, F., & van Ruth, S. (2016). Factors contributing to the variation
358 in the volatile composition of chocolate: Botanical and geographical origins of the cocoa beans,
359 and brand-related formulation and processing. *Food Research International*, *84*, 86–95.
360 https://doi.org/10.1016/j.foodres.2016.03.022

361 Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2008). Flavour Formation and Character in Cocoa
362 and Chocolate: A Critical Review. *Critical Reviews in Food Science and Nutrition*, *48*(9), 840–857.
363 https://doi.org/10.1080/10408390701719272

364 Aprotosoaie, A. C., Vlad Luca, S., & Miron, A. (2016). Flavor Chemistry of Cocoa and Cocoa Products-An
365 Overview. *Comprehensive Reviews in Food Science and Food Safety*, *15*(1), 73–91.
366 https://doi.org/10.1111/1541-4337.12180

367 Ballin, N. Z., & Laursen, K. H. (2019). To target or not to target? Definitions and nomenclature for
368 targeted versus non-targeted analytical food authentication. *Trends in Food Science & Technology*,
369 *86*, 537–543. https://doi.org/10.1016/j.tifs.2018.09.025

370 Belitz, H. D., Grosch, W., & Schieberle, P. (2009). *Food Chemistry*. (Springer, Ed.).

371 Boeckx, P., Bauters, M., & Dewettinck, K. (2020). Poverty and climate change challenges for sustainable
372 intensification of cocoa systems. *Current Opinion in Environmental Sustainability*, *47*, 106–111.
373 https://doi.org/10.1016/j.cosust.2020.10.012

374 Bressanello, D., Marengo, A., Cordero, C., Strocchi, G., Rubiolo, P., Pellegrino, G., … Liberto, E. (2021).
375 Chromatographic Fingerprinting Strategy to Delineate Chemical Patterns Correlated to Coffee Odor
376 and Taste Attributes. *Journal of Agricultural and Food Chemistry*, *69*(15), 4550–4560.
377 https://doi.org/10.1021/acs.jafc.1c00509

Bressanello, D, Liberto, E., Cordero, C., Sgorbini, B., Rubiolo, P., Pellegrino, G., … Bicchi, C. (2018). Chemometric Modeling of Coffee Sensory Notes through Their Chemical Signatures: Potential and Limits in Defining an Analytical Tool for Quality Control. *Journal of Agricultural and Food Chemistry*, *66*(27), 7096–7109. https://doi.org/10.1021/acs.jafc.8b01340

Bressanello, D., Liberto, E., Cordero, C., Rubiolo, P., Pellegrino, G., Ruosi, M. R., & Bicchi, C. (2017). Coffee aroma: Chemometric comparison of the chemical information provided by three different samplings combined with GC–MS to describe the sensory properties in cup. *Food Chemistry*, *214*, 218–226. https://doi.org/10.1016/j.foodchem.2016.07.088

Casale M., Malegori, C., Oliveri, P., Liberto, E., Rubiolo, P., Bicchi, C., & Cordero, C. (2020). Chapter 10. Chemometrics: Basic Principles and Applications. In P. Tranquida (Ed.), *Advanced Gas Chromatography in Food Analysis* (Royal Soci, pp. 403–451). Croydon, UK.

Cagliero, C., Sgorbini,B., Cordero,C., Liberto,E., Rubiolo,P., Bicchi,C. (2017) Enantioselective Gas Chromatography with Cyclodextrinin Odorant Analysis. Chapter 19 in Andrea Buettner (Ed). *Handbook of odor*. Springer, Switezerland.

Cavanna, D., Righetti, L., Elliott, C., & Suman, M. (2018). The scientific challenges in moving from targeted to non-targeted mass spectrometric methods for food fraud analysis: A proposed validation workflow to bring about a harmonized approach. *Trends in Food Science & Technology*, *80*, 223–241. https://doi.org/10.1016/J.TIFS.2018.08.007

Chambers, E., & Koppel, K. (2013). Associations of Volatile Compounds with Sensory Aroma and Flavor: The Complex Nature of Flavor. *Molecules*, *18*(5), 4887–4905. https://doi.org/10.3390/molecules18054887

Chetschik, I., Kneubühl, M., Chatelain, K., Schlüter, A., Bernath, K., & Hühn, T. (2018). Investigations on the Aroma of Cocoa Pulp ( Theobroma cacao L.) and Its Influence on the Odor of Fermented Cocoa Beans. *Journal of Agricultural and Food Chemistry*, *66*(10), 2467–2472. https://doi.org/10.1021/acs.jafc.6b05008

Cuadros-Rodríguez, L., Ruiz-Samblás, C., Valverde-Som, L., Pérez-Castaño, E., & Casado, A. G. (2016). Chromatographic fingerprinting: An innovative approach for food 'identitation' and food authentication - A tutorial. *Analytica Chimica Acta*, *909*, 9–23. https://doi.org/10.1016/j.aca.2015.12.042

Cuadros-Rodríguez, Luis, Ortega-Gavilán, F., Martín-Torres, S., Arroyo-Cerezo, A., & Jiménez-Carvelo, A. M. (2021). Chromatographic Fingerprinting and Food Identity/Quality: Potentials and Challenges. *Journal of Agricultural and Food Chemistry*, *69*(48), 14428–14434. https://doi.org/10.1021/acs.jafc.1c05584

Danezis, G. P., Tsagkaris, A. S., Brusic, V., & Georgiou, C. A. (2016). Food authentication: state of the art and prospects. *Current Opinion in Food Science*, *10*, 22–31. https://doi.org/10.1016/j.cofs.2016.07.003

Dunkel, A., Steinhaus, M., Kotthoff, M., Nowak, B., Krautwurst, D., Schieberle, P., & Hofmann, T. (2014). Nature's chemical signatures in human olfaction: A foodborne perspective for future biotechnology. *Angewandte Chemie - International Edition*. https://doi.org/10.1002/anie.201309508

FCC. (2018). Services - Rules | The Federation of Cocoa Commerce. Retrieved 2 February 2020, from https://www.cocoafederation.com/services/guidelines

Febrianto, N. A., & Zhu, F. (2022). Composition of methylxanthines, polyphenols, key odorant volatiles and minerals in 22 cocoa beans obtained from different geographic origins. *LWT*, *153*, 112395. https://doi.org/10.1016/j.lwt.2021.112395

Greño, M., Plaza, M., Luisa Marina, M., & Castro Puyana, M. (2023). Untargeted HPLC-MS-based metabolomics approach to reveal cocoa powder adulterations. *Food Chemistry*, *402*, 134209. https://doi.org/10.1016/j.foodchem.2022.134209

Gutiérrez, T. J. (2017). State-of-the-Art Chocolate Manufacture: A Review. *Comprehensive Reviews in Food Science and Food Safety*, *16*(6), 1313–1344. https://doi.org/10.1111/1541-4337.12301

Johanningsmeier, S. D., Harris, G. K., & Klevorn, C. M. (2016). Metabolomic Technologies for Improving the Quality of Food: Practice and Promise. *Http://Dx.Doi.Org/10.1146/Annurev-Food-022814-015721*, *7*, 413–438. https://doi.org/10.1146/ANNUREV-FOOD-022814-015721

Kumar, S., D'Souza, R. N., Corno, M., Ullrich, M. S., Kuhnert, N., & Hütt, M.-T. (2022). Cocoa bean fingerprinting via correlation networks. *Npj Science of Food 2022 6:1*, *6*(1), 1–9. https://doi.org/10.1038/s41538-021-00120-4

Lahive, F., Hadley, P., & Daymond, A. J. (2019). The physiological responses of cacao to the environment and the implications for climate change resilience. A review. *Agronomy for Sustainable Development*, *39*(1), 5. https://doi.org/10.1007/s13593-018-0552-0

Liu, M., Liu, J., He, C., Song, H., Liu, Y., Zhang, Y., … Su, X. (2017). Characterization and comparison of key aroma-active compounds of cocoa liquors from five different areas. *International Journal of Food Properties*, *20*(10), 2396–2408. https://doi.org/10.1080/10942912.2016.1238929

Lytou, A. E., Panagou, E. Z., & Nychas, G.-J. E. (2019). Volatilomics for food quality and authentication. *Current Opinion in Food Science*, *28*, 88–95. https://doi.org/10.1016/j.cofs.2019.10.003

Magagna, F., Guglielmetti, A., Liberto, E., Reichenbach, S. E., Allegrucci, E., Gobino, G., … Cordero, C. (2017). Comprehensive Chemical Fingerprinting of High-Quality Cocoa at Early Stages of Processing: Effectiveness of Combined Untargeted and Targeted Approaches for Classification and Discrimination. *Journal of Agricultural and Food Chemistry*, *65*(30), 6329–6341. https://doi.org/10.1021/acs.jafc.7b02167

Marseglia, A., Musci, M., Rinaldi, M., Palla, G., & Caligiani, A. (2020). Volatile fingerprint of unroasted and roasted cocoa beans (Theobroma cacao L.) from different geographical origins. *Food Research International*, *132*, 109101. https://doi.org/10.1016/j.foodres.2020.109101

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

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

Perotti, P., Cordero, C., Bortolini, C., Rubiolo, P., Bicchi, C., & Liberto, E. (2020). Cocoa smoky off-flavor: Chemical characterization and objective evaluation for quality control. *Food Chemistry*, *309*, 125561. https://doi.org/10.1016/j.foodchem.2019.125561

Prescott, J. (2015). Multisensory processes in flavour perception and their influence on food choice. *Current Opinion in Food Science*, *3*, 47–52. https://doi.org/10.1016/j.cofs.2015.02.007

Research and Markets. (2021). *Cocoa Beans Market Size, Market Share, Application Analysis, Regional Outlook, Growth Trends, Key Players, Competitive Strategies and Forecasts, 2021 to 2029*. Retrieved from https://www.researchandmarkets.com/reports/5615897/cocoa-beans-market-size-market-share?gclid=EAIaIQobChMIr7Puqa2H-gIVNxkGAB32MgszEAAYAiAAEgI-n_D_BwE

Rodionova, O. Y., & Pomerantsev, A. L. (2020). Chemometric tools for food fraud detection: The role of target class in non-targeted analysis. *Food Chemistry*, *317*, 126448. https://doi.org/10.1016/j.foodchem.2020.126448

Scavarda, C., Cordero, C., Strocchi, G., Bortolini, C., Bicchi, C., & Liberto, E. (2021). Cocoa smoky off-flavour: A MS-based analytical decision maker for routine controls. *Food Chemistry*, *336*, 127691. https://doi.org/10.1016/j.foodchem.2020.127691

Schwan, R. F., & Wheals, A. E. (2004). The Microbiology of Cocoa Fermentation and its Role in Chocolate Quality. *Critical Reviews in Food Science and Nutrition*, *44*(4), 205–221. https://doi.org/10.1080/10408690490464104

Somarriba, E., Peguero, F., Cerda, R., Orozco-Aguilar, L., López-Sampson, A., Leandro-Muñoz, M. E., … Sinclair, F. L. (2021). Rehabilitation and renovation of cocoa (Theobroma cacao L.) agroforestry systems. A review. *Agronomy for Sustainable Development*, *41*(5), 64. https://doi.org/10.1007/s13593-021-00717-9

Stilo, F., Liberto, E., Spigolon, N., Genova, G., Rosso, G., Fontana, M., … Cordero, C. (2021). An effective chromatographic fingerprinting workflow based on comprehensive two-dimensional gas chromatography – Mass spectrometry to establish volatiles patterns discriminative of spoiled hazelnuts (Corylus avellana L.). *Food Chemistry*, *340*, 128135.

Suman, M., Cavanna, D., Sammarco, G., Lambertini, F., & Loffi, C. (2021). Fighting food frauds exploiting chromatography-mass spectrometry technologies: Scenario comparison between solutions in scientific literature and real approaches in place in industrial facilities. *TrAC Trends in Analytical Chemistry*, *142*, 116305. https://doi.org/10.1016/j.trac.2021.116305

Torres-Moreno, M., Tarrega, A., & Blanch, C. (2014). Characterization of Volatile Compounds in Dark Chocolates by HS-SPME and GC-MS. In *Flavour Science* (pp. 283–287). Elsevier. https://doi.org/10.1016/B978-0-12-398549-1.00054-4

Tuenter, E., Delbaere, C., De Winne, A., Bijttebier, S., Custers, D., Foubert, K., … Pieters, L. (2020). Non-volatile and volatile composition of West African bulk and Ecuadorian fine-flavor cocoa liquor and chocolate. *Food Research International*, *130*, 108943. https://doi.org/10.1016/j.foodres.2019.108943

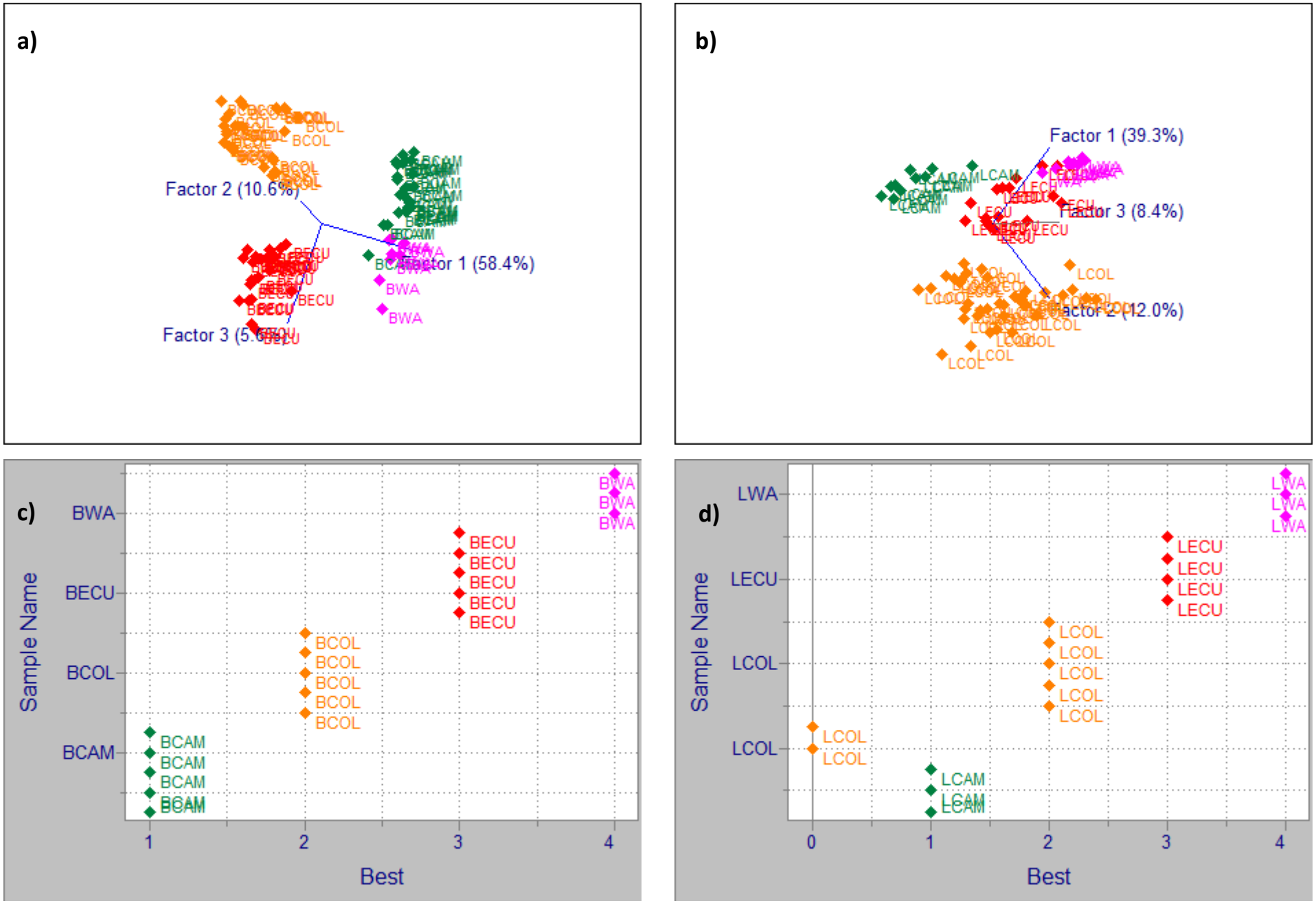
Ullrich, L., Casty, B., André, A., Hü, T., Steinhaus, M., & Chetschik, I. (2022). Decoding the Fine Flavor Properties of Dark Chocolates. *J. Agric. Food Chem*,70, 42, 13730–13740. https://doi.org/10.1021/acs.jafc.2c04166

Wang, Y., O'Reilly, J., Chen, Y., & Pawliszyn, J. (2005). Equilibrium in-fibre standardisation technique for solid-phase microextraction. *Journal of Chromatography A*, *1072*(1), 13–17. https://doi.org/10.1016/j.chroma.2004.12.084

Figure

Click here to access/download;Figure;Figures.pptx

# Figure 1



# Figure 2

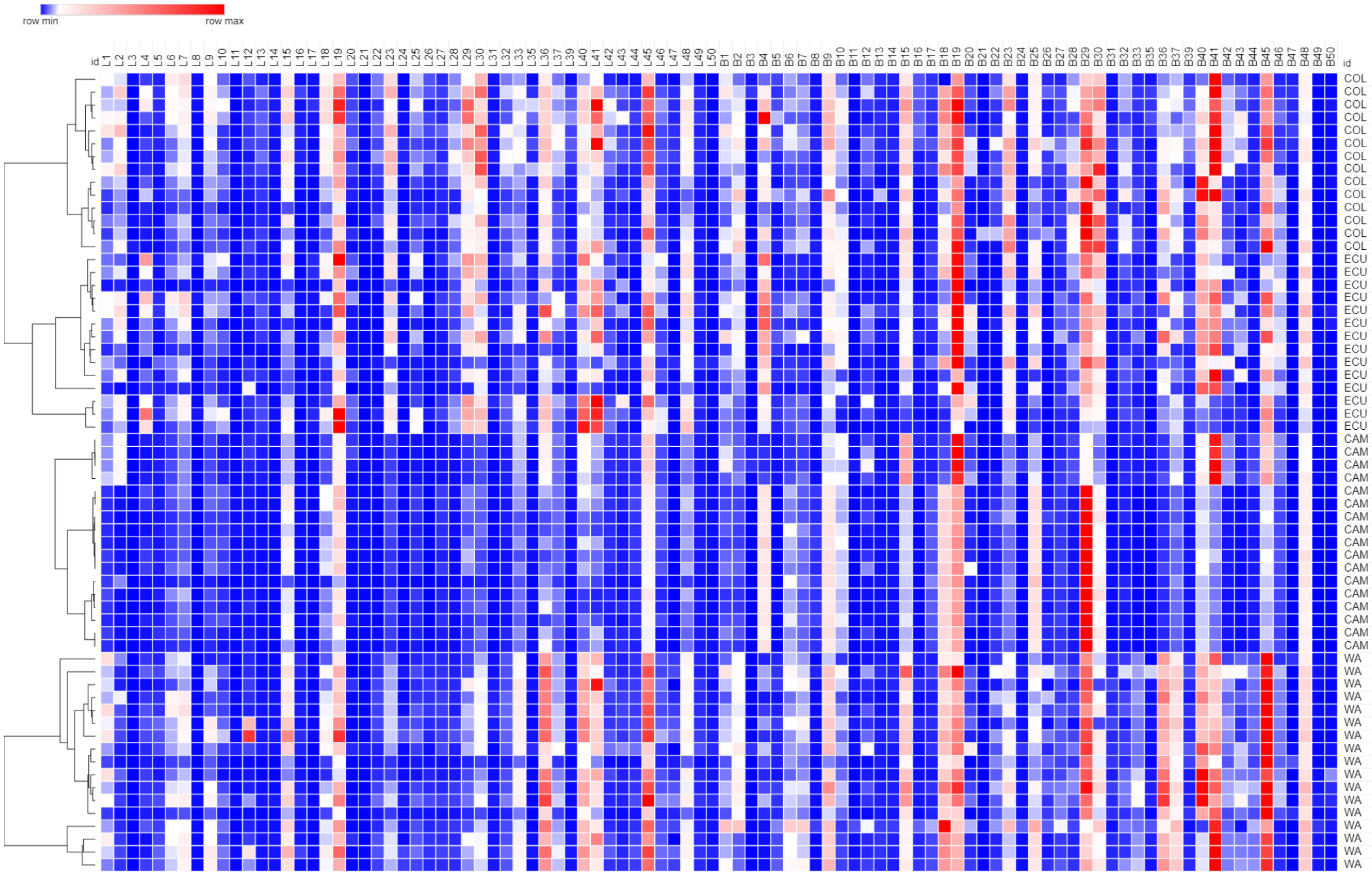
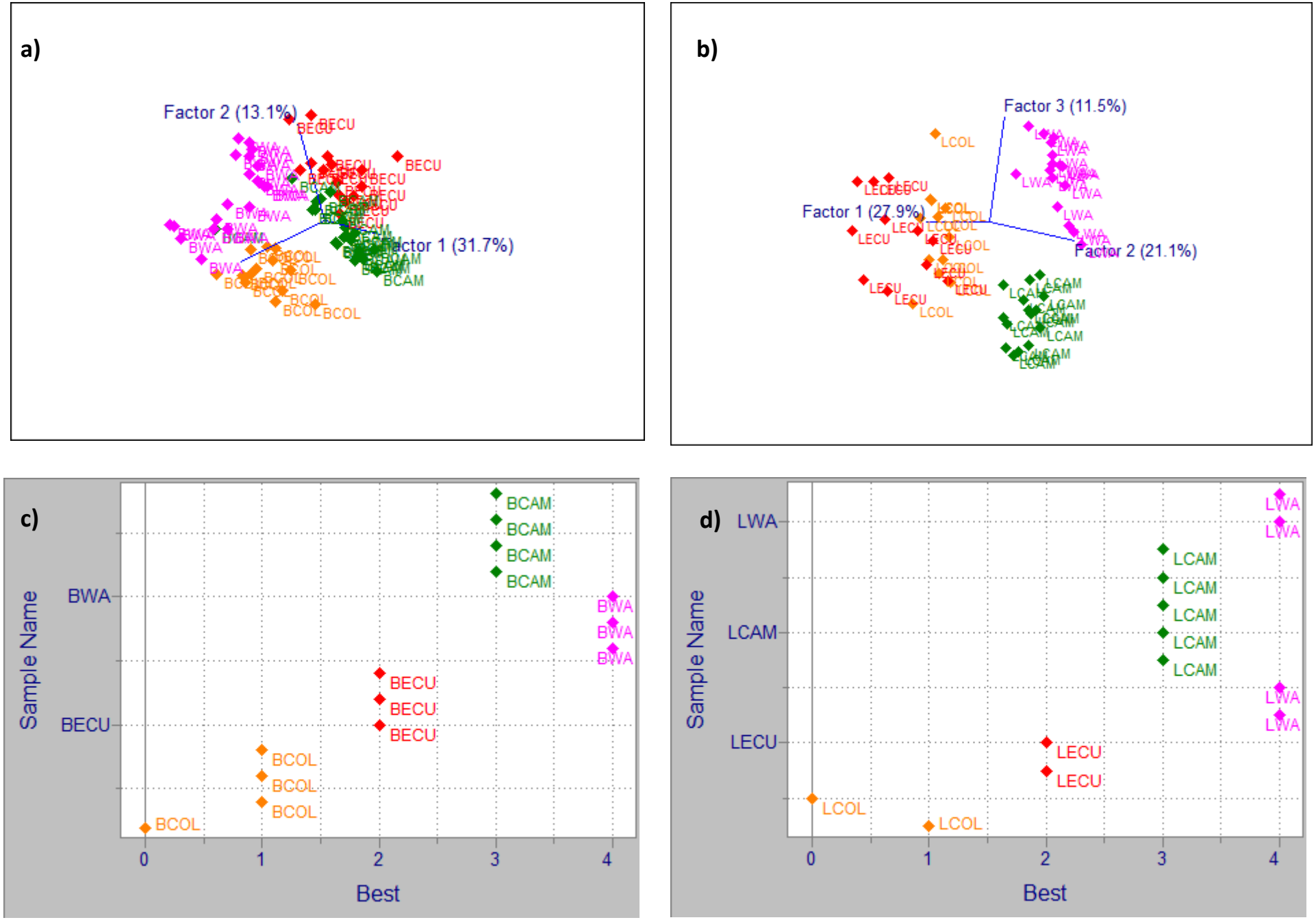
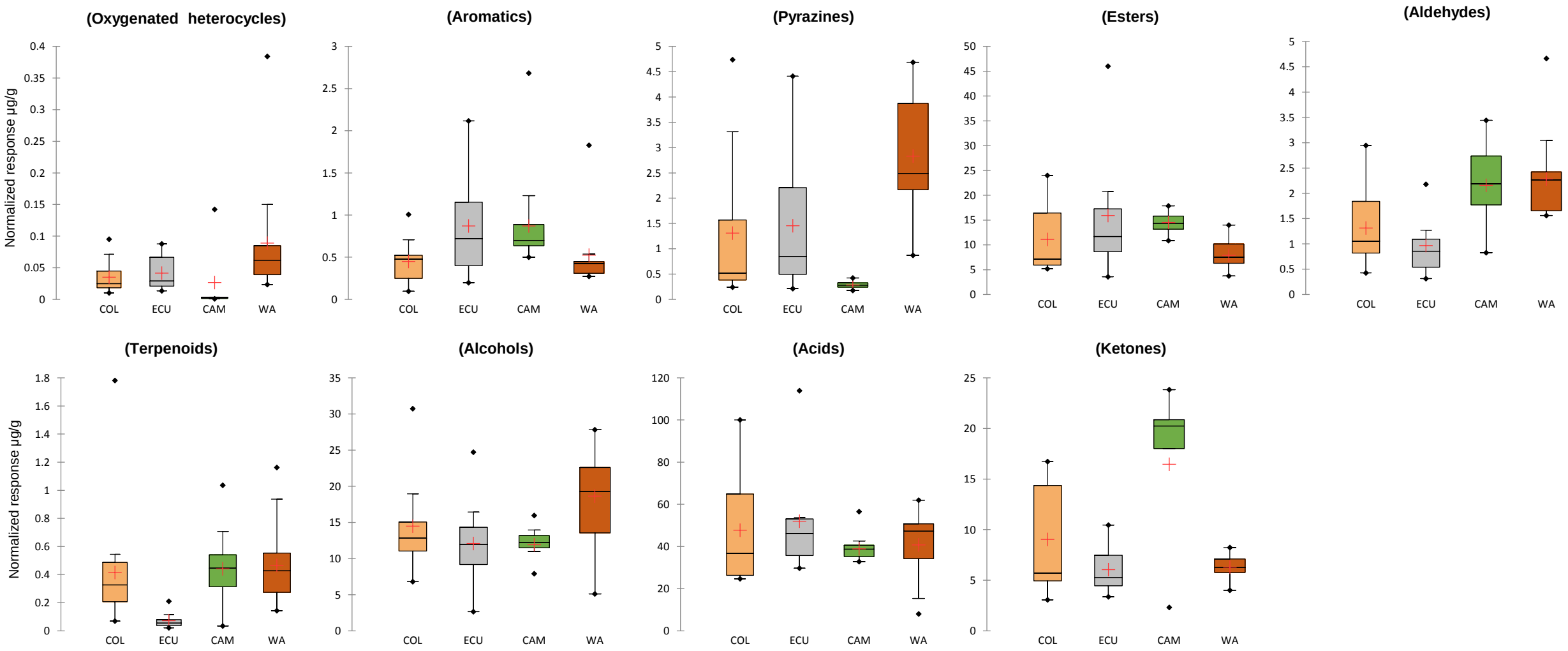

row min
row max
id
L1 L2 L3 L4 L5 L6 L7 L8 L9 L10 L11 L12 L13 L14 L15 L16 L17 L18 L19 L20 L21 L22 L23 L24 L25 L26 L27 L28 L29 L30 L31 L32 L33 L35 L36 L37 L39 L40 L41 L42 L43 L44 L45 L46 L47 L48 L49 L50 B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 B11 B12 B13 B14 B15 B16 B17 B18 B19 B20 B21 B22 B23 B24 B25 B26 B27 B28 B29 B30 B31 B32 B33 B35 B36 B37 B39 B40 B41 B42 B43 B44 B45 B46 B47 B48 B49 B50
id
COL
ECU
CAM
WA

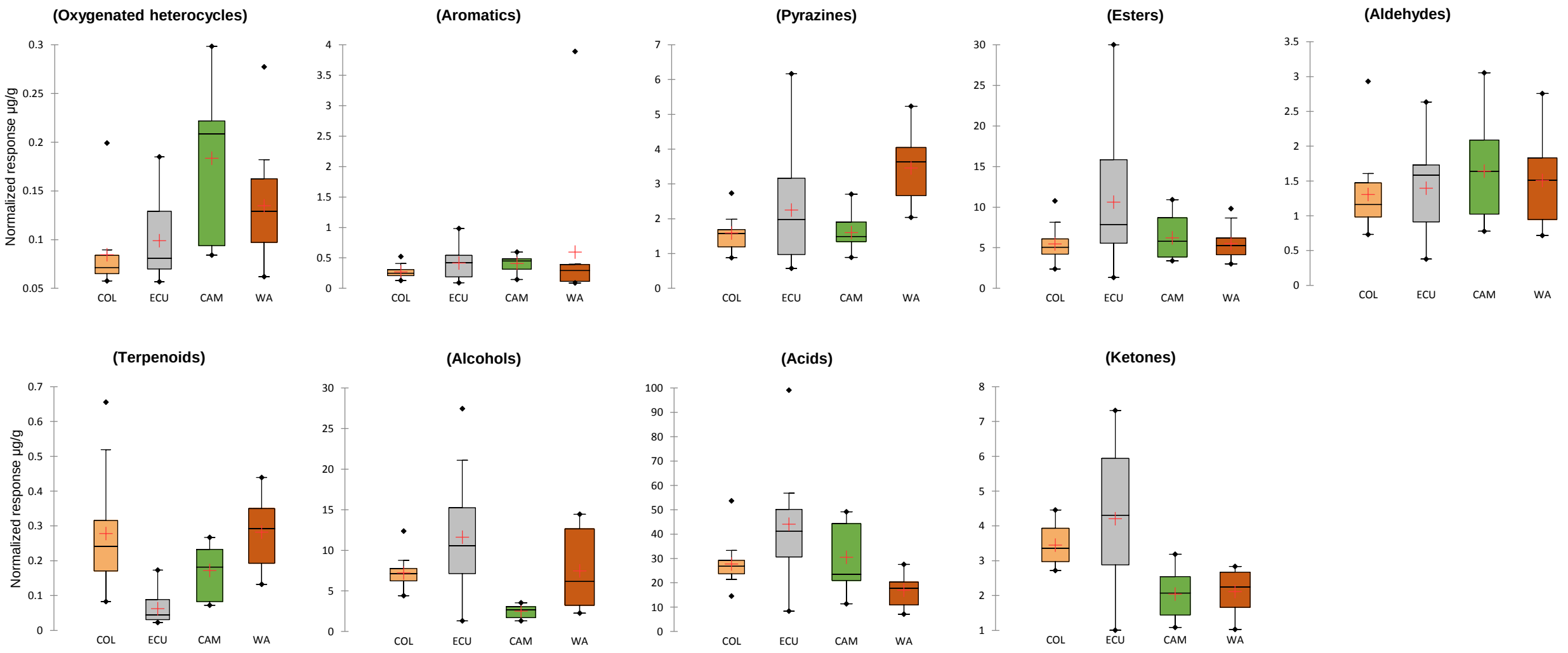
**Figure 3**


a)
Factor 2 (13.1%)
Factor 1 (31.7%)
BECU
BWA
BCOL
BCAM
b)
Factor 3 (11.5%)
Factor 1 (27.9%)
Factor 2 (21.1%)
LCOL
LECU
LWA
LCAM
c)
Sample Name
BWA
BECU
BCAM
BCOL
0
1
2
3
4
Best
d)
Sample Name
LWA
LCAM
LECU
LCOL
0
1
2
3
4
Best

# Figure 4

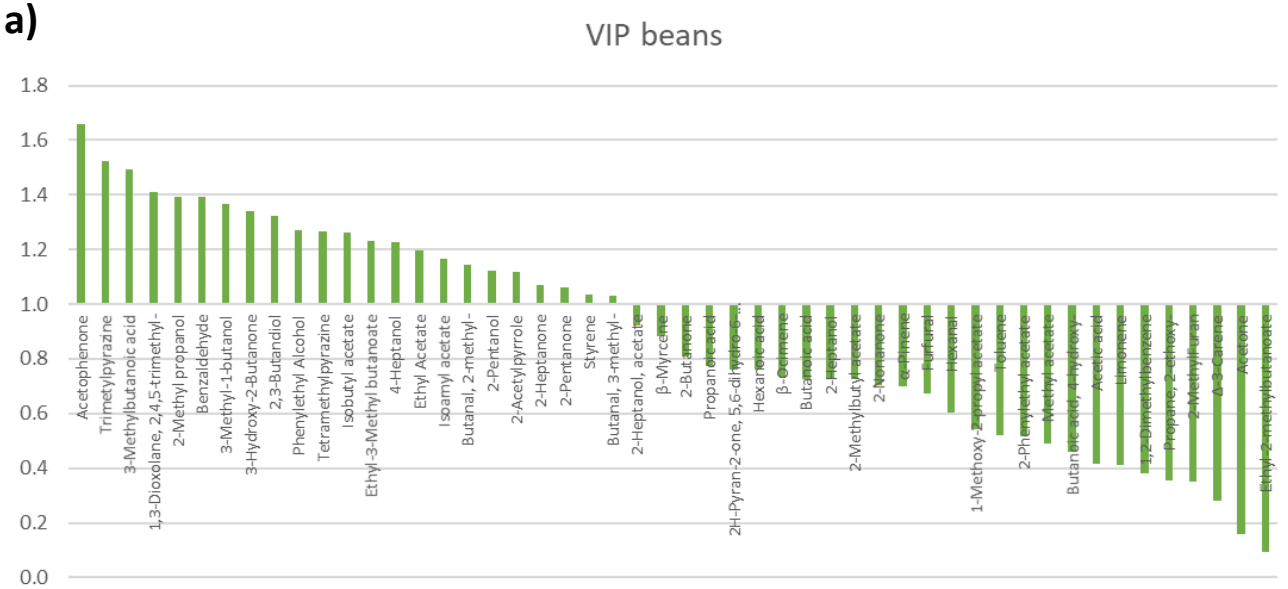

(Oxygenated heterocycles)
(Aromatics)
(Pyrazines)
(Esters)
(Aldehydes)
(Terpenoids)
(Alcohols)
(Acids)
(Ketones)
Normalized response µg/g
COL
ECU
CAM
WA

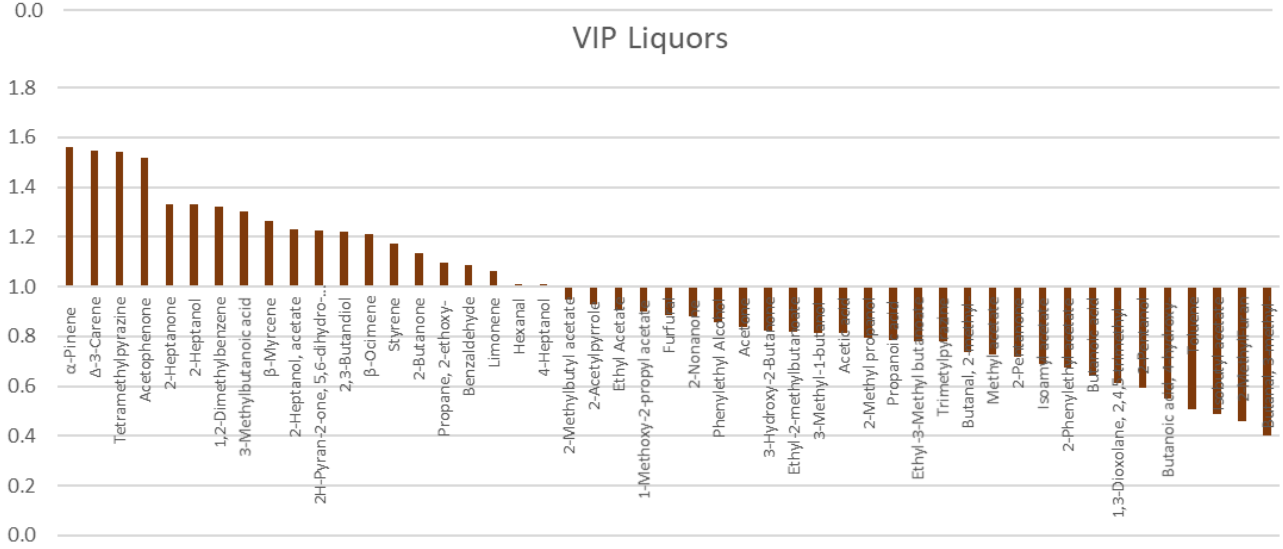
# Figure 5


(Oxygenated heterocycles)
(Aromatics)
(Pyrazines)
(Esters)
(Aldehydes)
(Terpenoids)
(Alcohols)
(Acids)
(Ketones)
Normalized response µg/g
COL
ECU
CAM
WA

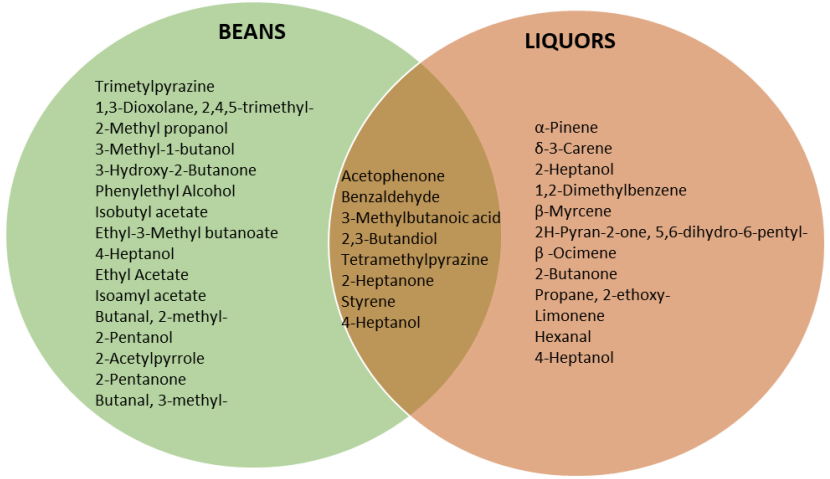
# Figure 6

a)

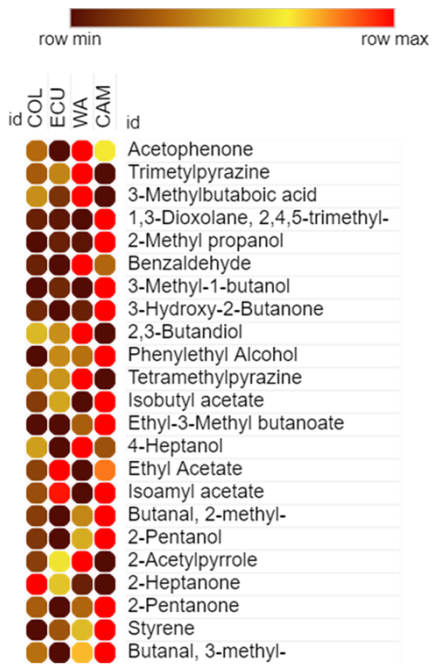

VIP beans
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Acetophenone
Trimetylpyrazine
3-Methylbutanoic acid
1,3-Dioxolane, 2,4,5-trimethyl-
2-Methyl propanol
Benzaldehyde
3-Methyl-1-butanol
3-Hydroxy-2-Butanone
2,3-Butandiol
Phenylethyl Alcohol
Tetramethylpyrazine
Isobutyl acetate
Ethyl-3-Methyl butanoate
4-Heptanol
Ethyl Acetate
Isoamyl acetate
Butanal, 2-methyl-
2-Pentanol
2-Acetylpyrrole
2-Heptanone
2-Pentanone
Styrene
Butanal, 3-methyl-
2-Heptanol, acetate
β-Myrcene
2-Butanone
Propane, 2-ethoxy-
2H-Pyran-2-one, 5,6-dihydro-6-...
Hexanoic acid
β-Ocimene
Butanoic acid
2-Heptanol
2-Methylbutyl acetate
2-Nonanone
α-Pinene
Furfural
Hexanal
1-Methoxy-2-propyl acetate
Toluene
2-Phenylethyl acetate
Methyl acetate
Butanoic acid, 4-hydroxy-
Acetic acid
Limonene
1,2-Dimethylbenzene
Propane, 2-ethoxy-
2-Methylfuran
δ-3-Carene
Acetone
Ethyl 2-methylbutanoate

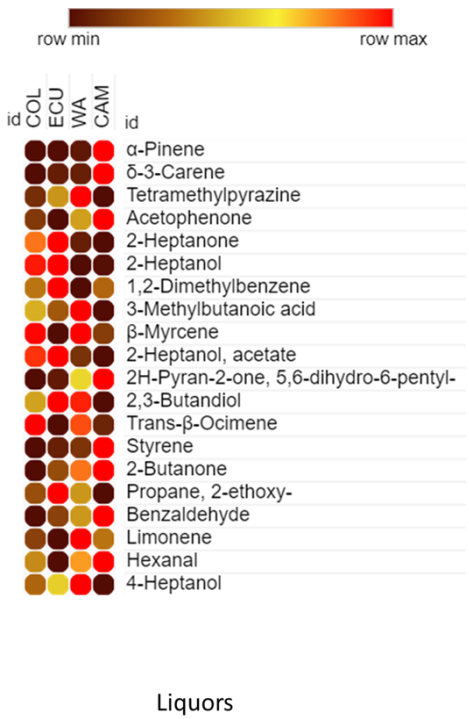

VIP Liquors
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
α-Pinene
Δ-3-Carene
Tetramethylpyrazine
Acetophenone
2-Heptanone
2-Heptanol
1,2-Dimethylbenzene
3-Methylbutanoic acid
β-Myrcene
2-Heptanol, acetate
2H-Pyran-2-one, 5,6-dihydro-...
2,3-Butandiol
β-Ocimene
Styrene
2-Butanone
Propane, 2-ethoxy-
Benzaldehyde
Limonene
Hexanal
4-Heptanol
2-Methylbutyl acetate
2-Acetylpyrrole
Ethyl Acetate
1-Methoxy-2-proppyl acetate
Furfural
2-Nonanone
Phenylethyl Alcohol
Acetone
3-Hydroxy-2-Butanone
Ethyl-2-methylbutanoate
3-Methyl-1-butanol
Acetic acid
2-Methyl propanol
Propanoic acid
Ethyl-3-Methyl butanoate
Trimetylpyrazine
Butanal, 2-methyl-
Methyl acetate
2-Pentanone
Isoamyl acetate
2-Phenylethyl acetate
Butanoic acid
1,3-Dioxolane, 2,4,5-trimethyl-
2-Butanone
Butanoic acid, 4-hydroxy-
Toluene
Isobutyl acetate
Furfural
Ethyl 2-methylbutanoate

b)


BEANS
Trimetylpyrazine
1,3-Dioxolane, 2,4,5-trimethyl-
2-Methyl propanol
3-Methyl-1-butanol
3-Hydroxy-2-Butanone
Phenylethyl Alcohol
Isobutyl acetate
Ethyl-3-Methyl butanoate
4-Heptanol
Ethyl Acetate
Isoamyl acetate
Butanal, 2-methyl-
2-Pentanol
2-Acetylpyrrole
2-Pentanone
Butanal, 3-methyl-
LIQUORS
α-Pinene
δ-3-Carene
2-Heptanol
1,2-Dimethylbenzene
β-Myrcene
2H-Pyran-2-one, 5,6-dihydro-6-pentyl-
β -Ocimene
2-Butanone
Propane, 2-ethoxy-
Limonene
Hexanal
4-Heptanol
Acetophenone
Benzaldehyde
3-Methylbutanoic acid
2,3-Butandiol
Tetramethylpyrazine
2-Heptanone
Styrene
4-Heptanol

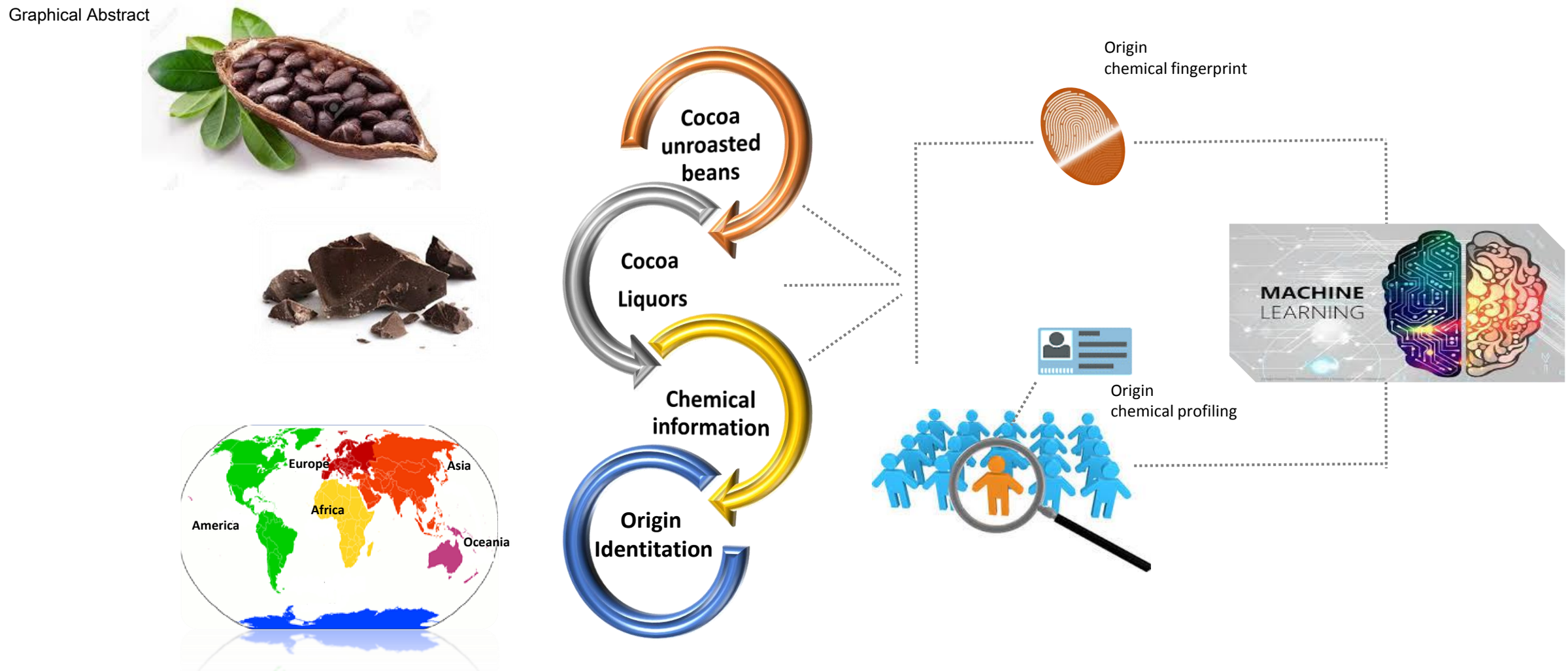
c)


row min
row max
id
COL
ECU
WA
CAM
id
Acetophenone
Trimetylpyrazine
3-Methylbutaboic acid
1,3-Dioxolane, 2,4,5-trimethyl-
2-Methyl propanol
Benzaldehyde
3-Methyl-1-butanol
3-Hydroxy-2-Butanone
2,3-Butandiol
Phenylethyl Alcohol
Tetramethylpyrazine
Isobutyl acetate
Ethyl-3-Methyl butanoate
4-Heptanol
Ethyl Acetate
Isoamyl acetate
Butanal, 2-methyl-
2-Pentanol
2-Acetylpyrrole
2-Heptanone
2-Pentanone
Styrene
Butanal, 3-methyl-
Beans


row min
row max
id
COL
ECU
WA
CAM
id
α-Pinene
δ-3-Carene
Tetramethylpyrazine
Acetophenone
2-Heptanone
2-Heptanol
1,2-Dimethylbenzene
3-Methylbutanoic acid
β-Myrcene
2-Heptanol, acetate
2H-Pyran-2-one, 5,6-dihydro-6-pentyl-
2,3-Butandiol
Trans-β-Ocimene
Styrene
2-Butanone
Propane, 2-ethoxy-
Benzaldehyde
Limonene
Hexanal
4-Heptanol
Liquors

Click here to access/download
**Supplementary material for online publication only**
Appendix A supplementary material.pdf

Graphical Abstract



**Declaration of interests**

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

☒ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

| |
|---|
| Dr. Cristian Bortolini is presently employees of Soremartec Italia s.r.l Alba, Italy. Academic authors declare no conflict of interest. |