The final publication is available at

https://doi.org/10.1016/j.foodchem.2022.133074

Additional Information

1 **Automatic and non-targeted analysis of the volatile profile of natural and alkalized**

2 **cocoa powders using SBSE-GC-MS and chemometrics**

3

4 Maribel Alexandra Quelal-Vásconez[a,c,f]*, Riccardo Macchioni, Greta Livi[e], Édgar

5 Pérez-Esteve[a], María Jesús Lerma-García[d], Pau Talens[a], José Manuel Barat[a], Mikael

6 Agerlin Petersen[b], Rasmus Bro[b]

7 *[a] Departamento de Tecnología de Alimentos. Universitat Politècnica de València.*

8 *Camino de Vera, s/n 46022, Valencia, Spain*

9 *[b] Department of Food Science, University of Copenhagen, DK-1958 Frederiksberg,*

10 *Denmark*

11 *[c] Data Innovation, Quito, Ecuador*

12 *[d] Department of Analytical Chemistry, University of Valencia, C/Dr. Moliner, 50, 46100*

13 *Burjassot, Valencia, Spain*

14 *[e] Department of Management, Sapienza University of Rome, Via del Castro Laurenziano*

15 *9, 00161, Rome, Italy*

16 *[f] Agroindustrial and Food Engineering, Universidad de las Americas, Quito, Ecuador*

17

18 * Corresponding author:

19

20 **Abstract**

21   A total of 56 key volatile compounds present in natural and alkalized cocoa powders

22 have been rapidly evaluated using a non-target approach using stir bar sorptive extraction

23 gas chromatography mass spectrometry (SBSE-GC-MS) coupled to Parallel Factor

24 Analysis 2 (PARAFAC2) automated in PARADISe. Principal component analysis (PCA)

25 explained 80% of the variability of the concentration, in four PCs, which revealed specific

groups of volatile characteristics. Partial least squares discriminant analysis (PLS-DA)

helped to identify volatile compounds that were correlated to the different degrees of

alkalization. Dynamics between compounds such as the acetophenone increasing and

toluene and furfural decreasing in medium and strongly alkalized cocoas allowed its

differentiation from natural cocoa samples. Thus, the proposed comprehensive analysis

is a useful tool for understanding volatiles, e.g., for the quality control of cocoa powders

with significant time and costs savings.

Keywords: Alkalized cocoa powder, chemometrics, volatile compounds, SBSE-GC-MS,

PARAFAC2, PLS-DA

1. Introduction

Cocoa powder is an appreciated and widely consumed product around the world due to its capacity to provide color and flavor to a wide range of food products such as beverages, confectionery, bakery products or pastries, with high consumer cocoa powder demands (Quelal et al, 2018).

It is known that cocoa flavor depends on the characteristics of the raw material (variety, origin, harvest conditions…) and the conditions of the primary (seeds fermentation) and secondary transformation (drying, roasting and alkalization). It consists of treating the cocoa with alkali at high temperature and pressure with the aim of darkening the color, improving solubility and reducing acidity and astringency of the so-called natural cocoa (Valverde, Pérez-Esteve, & Barat, 2020). Such is the importance of the degree of alkalization in cocoa powder properties, that cocoa powders are classified in different categories (from natural to strongly alkalized powders). In order to standardize this classification, Miller et al. proposed in 2008 a categorization based on the pH of the sample. However, since pH is modified by the type and amount of alkali, and not by other processing variables, such as temperature, pressure, oxygen presence… this accepted parameter may result in imprecise categorization, for instance, for the prediction of the flavor of a cocoa sample.

Characterization of the volatile profile of cocoa powder is generally done by gas chromatography (GC) coupled to mass spectrometry (MS) (Quelal et al., 2020) after recovering the compounds by solvent extraction (Bonvehí, 2005), thin-layer high-vacuum distillation (Krings, Zelena, Wu, & Berger, 2006), or solid phase micro extraction (Ducki, Miralles-Garcia, Zumbé, Tornero, & Storey, 2008). Once chromatograms are obtained, data must be carefully analyzed to identify and quantify the compounds. This data analysis has been traditionally done by using a suspect screening or target analysis

approach, for which references are used to determine the identity and concentration.

However, the volatile aroma fraction of complex matrices, such as cocoa, usually contains

unknown organic compounds. Hence, using this approach to analyze the high amount of

compounds found in a cocoa sample is unfeasible as a routine method, as more cost-

effective time-saving methods and procedures are required to monitor the optimal quality

of the volatile profile changes of cocoa powders.

In this way, we present herein a double approach to reach this goal. On the one hand,

using a fast and more environmentally friendly technique that does not require the use of

organic solvents to collect the analytes from the sample is proposed. This technique,

called stir bar sorptive extraction (SBSE) consists of using a polymer coating on a

magnetic stirring rod that extracts and enriches organic compounds from aqueous

matrices (Frank & Pat., 2007). This modern approach has been successfully applied for

the determination of the volatile profile of liquid foods such as high-quality wine vinegar

(Marrufo-Curtido et al., 2012), and combined with chemometrics for the differentiation

of variety and processing conditions in peach juices (Marsol-Vall et al., 2018), among

others, but as far as we are concerned, never for the determination of the volatile fraction

of cocoa powders. On the other hand, regarding data treatment, we propose the use of

non-targeted analysis. This novel approach has already been used for volatile compounds

(Dubrow et al., 2022), but also for contaminants or xenobiotics (Plassmann et al., 2016),

food metabolomics (Wei, Furihata, Miyakawa, & Tanokura, 2014) or foodomics (Díaz,

Pozo, Sancho, & Hernández, 2014), prior information for the compounds in the samples

is not assumed or required (Schymanski et al., 2014).

Among different options, we propose the use of the so-called PARAllel FACtor

analysis2 (PARAFAC2) model (Amigo et al., 2010) due to its capability to deconvolute

co-eluted, retention time shifted and low signal-to-noise (S/N) ratio chromatographic

90 peaks for all investigated samples simultaneously (Johnsen, Skou, Khakimov, & Bro,

91 2017). However, and since the use of this method requires the handling of complicated

92 mathematical computer coding tools, making its use difficult for non-expert users, the

93 employment of an integrated tool called PARAFAC2 based Deconvolution and

94 Identification System (PARADISe) (Petersen & Bro, 2018), is proposed.

95 PARADISe has emerged as a novel and unique application for GC-MS data

96 processing. Unlike previous tools like XCMS OR MzMine, among others, PARADISe

97 performs automatic tentative peak identification using deconvoluted mass spectra with

98 the National Institute of Standards and Technology (NIST) library (Lacalle-Bergeron et

99 al., 2020). As a result, the data matrix is reduced, as is the time spent on statistical analysis

100 and peaks identification. This software has been successfully applied for the data

101 processing and determination of volatile compounds used in the classification of smoked

102 seafood (Lacalle-Bergeron et al., 2020).

103 Bearing this in mind, the objective of this work is to study the modifications of the

104 most important cocoa volatiles during the alkalization step. For this purpose, cocoa

105 powder samples of different degree of alkalization (from natural to high alkalization

106 levels) were evaluated by direct extraction of the analytes with no lengthy intermediate

107 steps using SBSE-GC-MS and the automatic peak identification software PARADISe.

108

109 **2. Materials and Methods**

110 *2.1 Raw materials*

111 A total of 30 cocoa powder samples were evaluated in this study. These samples,

112 which covered different geographical origins (Bolivia, Perú, Ecuador, Ivory Coast,

113 Cameroon and Indonesia), were bought in physical or virtual stores or kindly donated by

114   OLAM Food Ingredients, Spain (Cheste, Valencia). Samples were placed inside a glass

115   container, and stored in a dry, dark atmosphere until used. According to product labels

116   and technical datasheets, of the 30 samples 7 were natural cocoas (NC), 7 light-alkalized

117   (LAC), 8 medium-alkalized (MAC) and 6 strong-alkalized (SAC). This information was

118   not declared (UK) in two of the samples.

119   *2.2 Extraction and GC-MS analysis*

120   Volatile compounds were extracted in 25 mL vials by weighing 2 g of cocoa powder

121   and mixing them with 10 mL of distilled water. Volatile compounds were collected on

122   'Twister' SBSE, covered with a polydimethylsiloxane (PDMS) phase (Gerstel, Mullheim

123   an der Ruhr, Germany), by stirring for 60 min at 1000 rpm at 20ºC. The stir bar was

124   removed from the vial, rinsed with water to eliminate cocoa residue, dried with lint-free

125   tissue and placed inside a stainless-steel desorption tube.

126   Trapped volatiles were desorbed with an automatic thermal desorption unit (Turbo

127   Matrix 350, Perkin Elmer, Shelton, USA). Primary desorption was carried out by heating

128   the tube to 250 °C with a flow (50 mL min^{-1}) of carrier gas (H_2) for 15.0 min. Stripped

129   volatiles were trapped in a Tenax TA cold trap (30 mg, held at 5°C), which was

130   subsequently heated to 300 °C for 4 min (secondary desorption, outlet split 1:10). This

131   allowed the rapid transfer of volatiles to the GC-MS instrument (7890A GC-system

132   interfaced with a 5975C VL MSD with Triple-Axis detector from Agilent Technologies,

133   Palo Alto, California, USA) by a heated (225 °C) transfer line.

134   Volatile separation was carried out using a ZB-wax capillary column (30 m long x

135   0.25 mm internal diameter, 0.50 µm film thickness). The column pressure was held

136   constant at 2.3 psi which resulted in an initial flow rate of 1.4 mL min^{-1} with H_2 as carrier

137   gas. The column temperature program was: 10 min at 30°C, from 30°C to 240°C at 8°C

138   min^{-1}, and finally 5 min at 240°C. The mass spectrometer was operated in the electron

139   ionization mode at 70 eV. Mass-to-charge ratios between 15 and 300 atomic mass units

140   (amu) were scanned. The 30 samples were analyzed by duplicate, obtaining 60

141   chromatograms.

142

143   ***2.3 Chemometric analysis***

144

145   *2.3.1 Calibration and validation of the PARAFAC2 models*

146

147   The GC-MS chromatograms of volatile compounds were arrayed into a dataset with a

148   three-way structure, X (I×J×K), where the first mode represents elution times (I scans),

149   the second the spectral domain (J *m/z* fragments) and the third samples (K). The peaks of

150   chromatograms were selected as intervals to fit the PARAFAC2 models with to non-

151   negativity   constraints,   using   the   software   PARADISe

152   (www.models.life.ku.dk\PARADISe, Version 3.90, accessed July. 2021, Copenhagen

153   University) (Amigo, Skov, Bro, Coello, & Maspoch, 2008; Johnsen et al., 2017). Models

154   were evaluated by the maximum fit and core consistency (range 0-100) to select the

155   proper number of PARAFAC2 components for each interval. For each chemical

156   compound, the mass spectrum is estimated as part of the PARAFAC2 model, which is

157   used for identification purposes using 'NIST MS Search 2.0' software (NIST/EPA/NIH

158   Mass Spectral Library, NIST Scientific and Technical Databases, Gaithersburg, MD

159   20899-8380) that contains a library containing 190,825 spectra of 163,198 compounds.

160   The identification for each mass spectrum obtained by PARAFAC2 was performed by

161   checking the similarity with the spectrum in the database and the Kovats retention index

162   reported in the literature. PARADISe output is a semiquantitative report based on the total

163   ion current (TIC) without any external calibration. An important feature of these relative

164   concentrations is that it needs only to be scaled to obtain the real concentrations.

165   However, these absolute TIC values can be evaluated in relative terms to provide an

166   averaged qualitative comparison between the analytes (Amigo et al., 2010).

167

168   *2.3.2 PCA analysis*

169

170   After obtaining the relative concentrations of volatile compounds, a dataset with

171   structure Z (KxL) (where K are the 60 chromatograms and L the number of identified

172   compounds) was analyzed by a principal component analysis (PCA) to seek the effects

173   of the different degrees of alkalization on the aroma profile of cocoa powders. The PCA

174   is a variable reduction technique that condenses the information of the chemical variables

175   (usually highly correlated) into a few principal components (PCs), that represent the

176   original variables in a few uncorrelated components. This is done by decomposing the

177   data matrix Z (KxL) into two submatrices described by equation 1:

178   $$Z = TP^T + E \qquad\qquad (1)$$

179   where T (KxF) and P^T (FxL) are the so-called scores and loadings matrices, depending

180   on the number of selected PCs (F). The scores matrix compiles all the useful information

181   about samples, whereas loading matrix collect the information of variables (volatile

182   compounds).

183   In order to interpret the PCA model based on the volatile compounds, plots were

184   colored and labeled according to the different categories: four alkalization levels (NC,

185   LAC, MAC, SAC) and non-declared alkalization level (UK). This analysis was done

186   using PLS_Toolbox (version 8.6.2, Eigenvector Research, Inc., Manson, WA, USA

187   98831; software available at http://www.eigenvector.com).

2.3.3 PLS-DA analysis

189

190   Partial least squares discriminant analysis (PLS-DA) is a classification method (Indahl

191   et al., 2007) based on the partial least squares PLS approach. PLS algorithm was applied

192   to the database in which the independent variables were the identified volatile

193   compounds, and the dependent Y vector was the class labels according to the degrees of

194   alkalization NC, LAC, MAC and SAC. In this case of more than 2 classes, dummy

195   variables were defined and a PLS2 algorithm was used. The analysis was performed to

196   obtain a descriptive model, which allows the description of patterns found by the

197   relationship between the volatile compounds and the degrees of alkalization. The

198   evaluation of the models was done by observing the root mean error of prediction and

199   cross validation (RME_P, RME_{CV}), and the sensitivity and specificity of the calibration

200   and cross-validation sets. A variable selection was performed using the Variable

201   Importance in Projection (VIP) (Botelho, Reis, Oliveira, & Sena, 2015). All the analyses

202   were done with the PLS_Toolbox.

203

204   **3. Results and Discussion**

205

206   *3.1 Automatic compounds determination and identification by PARADISe*

207

208   After dividing the chromatograms into 215 intervals, PARAFAC2 models were built

209   and a total of 56 analytes were identified (see Table 1 of APPENDIX 1), assessing their

210   belonging to the mass spectral profiles by using the NIST database mentioned in Section

211   2.3.1. It should be mentioned that resolved chromatographic profiles, baseline effects,

212   some interfering effects, as well as low signal-to-noise peaks, were also modeled. These

213 results are important to ensure the further qualitative and quantitative analysis of the

214 obtained peaks, which rendered the task of finding selective *m/z* ions unnecessary.

215 According to Table 1, most of the identified compounds are pyrazines, which is in

216 agreement with several studies performed on cocoa products (Deuscher et al., 2020).

217 Figure 1a depicts one of the selected intervals considered in this study, after modeling

218 this interval, three components could be identified: two volatile compounds (2,3-

219 dimethyl-5-ethylpyrazine and furfural) and baseline influence (Figure 1 b and c).

220   -Figure 1-

221 The number of compounds here identified is similar to those previously reported in

222 studies of cocoa powders at alkalization stage (Mohamadi Alasti et al., 2019) and in non-

223 alkalized, medium and high alkalized cocoas (Sioriki et al., 2022).

224

225 *3.2 Characterization by PCA*

226

227 The PCA of the relative concentrations explained 80% of total variability, being most

228 of the variance of the samples explained in the first four latent variables. The first

229 component explained 56%, the second 13.9%, the third 5% and the fourth 4.6% of

230 variability. For the different degrees of alkalization, clusters showed a different aroma

231 profile.

232 Sample groupings (Figure 2) can be seen across PC1. In this, NCs were on the positive

233 side of the PC1, while the alkalized ones were on the negative side. According to PC2,

234 LACs were on the positive side, while SACs were on the negative one. This negative

235 relationship between LACs and SACs suggested the presence or absence of volatile

236 compounds with direct relationship to the degree of alkalization. The positions of the

237   cocoas with non-declared degree of alkalization (UK) suggested that should be NCs or

238   LACs with similar concentrations of volatile compounds.

239   -   Figure 2 -

240   The dispersion of NCs suggested that there is variability associated to their volatile

241   profiles. This variability was also observed in other studies that have been carried out in

242   relation to the cocoa aroma profile, which indicated that wide variations in volatile

243   profiles and other molecules are linked to environmental and genetics (origin, variety)

244   (Afoakwa, Paterson, Fowler, & Ryan, 2008), postharvest (fermentation, drying) (Liu et

245   al., 2017) and processing factors (roasting, alkalization) (Frauendorfer, Schieberle, &

246   Chieberle, 2008; Kongor et al., 2016, Li et al., 2012). The variation in alkalized sample

247   groups tends to be narrow, which could be attributed to the quality standardization of

248   products, which is one of the benefits of alkalization: The alkalization process creates a

249   better-standardized product in terms of color, texture, and solubility (Pérez, Lerma,

250   Fuentes, Palomares, & Barat, 2016).

251   According to Figure 2, some MAC samples may have similar volatile profiles to

252   LACs: these similarities could be attributed to the process stage in which alkalization was

253   performed (Miller et al., 2008). During the alkalization process, volatile compounds can

254   be better conserved depending on the process method, conditions (temperature and alkali

255   levels), products (alkali types) and the presentation of the cocoa product to be alkalized:

256   beans, cake or powder.

257

258   *3.3 Description of cocoa volatiles according to PCA*

259

260   According to Figure 2, LAC samples, located on the positive side of PC2, were

261   associated   with   the   following   volatiles: α-ethylidenbenzeneacetaldehyde,   n-

262   hexadecenoic acid, 5 hydroxy-2-decenoic acid δ lactone, benzeneacetaldehyde, and 5-

263   methyl-2-phenyl-2-hexenal, which had high positive loadings and ethyl hexadecanoate,

264   heneicosane, 2-ethyl-3-methylpyrazine, 2,3,5-trimethyl-6-propylpyrazine, and

265   benzonitrile which had high negative loading values, among others. The location in

266   Figure 2 of n-hexadecanoic acid and the ethyl hexadecanoate which is a long-chain fatty

267   acid ethyl ester, suggested a negative relationship between saturated fatty acids and fatty

268   acid ethyl esters. This last compound resulted from the condensation of the carboxy group

269   of hexadecanoic acid with the hydroxy group of ethanol. This also contributed to

270   distinguishing between LACs and SACs.

271   Some studies mention that 70% of acids are eliminated by the roasting process. Thus,

272   the fact that acids were found in LACs means that they can still be present after the

273   alkalization process. However, with medium and strong degrees of alkalization, the

274   importance of these compounds is considerably reduced (Aprotosoaie et al., 2016).

275   On the other hand, MACs are associated to the presence of acetophenone,

276   benzaldehyde, 2-methylundecane, and 2,6,10,15-tetrmethyl-heptadecane, which

277   possessed negative loadings at PC1. Regarding SACs, they are located at the negative

278   side of both PC1 and PC2, which could be related to the following compounds with

279   negative loadings: benzaldehyde, acetophenone, 2-methylundecane, and 2,6,10,15-

280   tetrmethyl-heptadecane (Sioriki et al., 2022).

281   Among them, the ketone acetophenone has been described as a compound that

282   contributes with sweet floral (Aprotosoaie, Luca, & Miron, 2016) or almond notes, while

283   2-ethyl-3-methylpyrazine is responsible of nutty and raw potato notes, being 5-methyl-2-

284   phenyl-2-hexenal related to bitter taste in cocoas (Bonvehí, 2005 and Moreira et al, 2018).

285   These previous studies aimed to identify, on the one hand, the volatile compounds in

286   roasted cocoa powders (Bonvehí, 2005) and, on the other hand, the volatile compounds

287 and protein profiles of fermented cocoa beans and chocolates (Moreira, Vilela, Santos,

288 Lima, & Schwan, 2018), studies that show the importance of generating knowledge of

289 the volatile profile of cocoa during its processing.

290 Figure 3 represents the distribution of samples in the biplot obtained with PC1-PC3

291 (Figure 3a) and PC1-PC4 (Figure 3b).

292 - Figure 3 -

293 As it can be observed in Figure 3a, the dispersion of the loadings showed a reduction

294 in the volatile compounds as the degree of alkalization increased. NCs and LACs had a

295 high content of volatile compounds, being the ones with higher loadings values 2

296 heptanol, phenylethyl alcohol, ethyl acetate and benzeneacetaldehyde (positive values of

297 both PC1 and PC3), and benzaldehyde and 2,6,10,15-tetramethyl- heptadecane (positive

298 values of PC3 and negative of PC1). On the other hand, acetophenone had negative values

299 of both PC1 and PC3 and nonanal and α-ethyliden benzeneacetaldehyde had negative

300 loading values only at PC3 (Figure 3a). Some of the aforementioned compounds have

301 been identified in studies into the aroma of criollo cocoa beans (Álvarez, 2017; Bonvehí,

302 2005).

303 It should be noted that the non-targeted nature of this analysis enabled finding volatile

304 compounds that may be of interest to state the functional properties of cocoa powder,

305 although an in-depth study would be needed to quantify and confirm whether the presence

306 of functional compounds may be significant. This is the case of 2-phenylethyl acetate,

307 which has been identified as a potential biomarker of cocoa with functional properties

308 (Mota-Gutierrez, Barbosa-Pereira, Ferrocino, & Cocolin, 2019).

309 Thereby from the literature, and in relation to the correspondence of flavor to volatile

310 cocoa compounds, it is known that 2-heptanol confers a fruity, herbaceous, flowery and

311 spicy flavor.

312   It is interesting to note that some compounds associated with NCs were also associated

313   with LACs but to a lesser extent. As previously indicated, acetophenone is one of the

314   compounds found with positive loadings when MACs samples were discussed. As

315   previously mentioned, this compound was described among the most important volatile

316   compounds for the floral aroma of cocoa (Mohamadi Alasti et al., 2019). Moreover, it

317   has been found that acetophenone has the same precursor as 2-phenylethanol, although

318   each one has a specificity and in the acetophenone case the biosynthesis of L-

319   phenylalanine belongs to a β-oxidation pathway (Colonges, et al., 2021).

320   On the other hand, a negative relationship between benzaldehyde and nonanal is

321   noticeable (see Figure 3a), which corresponded to 5.49% of the total variability explained

322   by PC3. Benzaldehyde, which was on the positive side of PC3, was high in some NCs,

323   while nonanal (with a negative loading value) was related to some LACs and MACs. This

324   last relationship agrees with the findings of Cremer and Eichner (2000), who mentioned

325   that nonanal, among other compounds, appears due to cocoa alkalization. It is well

326   established that linear aldehydes such as nonanal stem from lipid oxidation.

327   PC4 explained 4.66% of data variability (Figure 3b), and the positive side was related

328   to 2-methylundecane, 2-6-10-15-tetramethyl-heptadecane, and tetradecane, while the

329   negative side was associated with benzaldehyde, among others. These positions showed

330   a negative relationship between these compounds, which increased the variability among

331   LACs. It was interesting to note that some LACs and MACs had high levels of

332   benzaldehyde, one of the compounds considered by Mota-Gutierrez et al. (2019) to be a

333   biomarker of the presence of compounds with functional properties. This compound was

334   persistently found after the roasting process. In Figure 3b, loadings of PC4 suggested that

335   this biomarker (benzaldehyde) remains in certain samples after a light and medium

336   alkalization process. This knowledge can help to maintain the process parameters related

337   to raw material, alkali type and the alkali application stage (beans, cake, powder) to obtain

338   a cocoa-alkalized powder with functional properties.

339   The identified compounds coincide with those previously identified by studies that

340   have been carried out to characterize varieties, origins in cocoa beans and cocoa powders

341   (Bonvehí, 2005; Mohamadi Alasti et al., 2019). Volatile compounds are frequently

342   classified as families of pyrazines, alcohols, acids, aldehydes, ketones and esters

343   (Moreira, Vilela, Santos, Lima, & Schwan, 2018). These compounds are usually

344   responsible of pleasant aromas such as fruity notes or sweet and caramel odor perceptions

345   (Aculey et al., 2010).

346

347   *3.4 Classification patterns*

348

349   In order to study the possibility of classifying the different cocoas according to the 4

350   degrees of alkalization considered, a PLS-DA analysis was performed. When the model

351   was constructed, three latent variables were able to discriminate the samples according to

352   their degree of alkalization with RMS_C values of 0.216 (NCs), 0.25 (LACs), 0.25

353   (MACs), 0.213 (SACs) and RMS_{CV} values of 0.239 (NCs), 0.28 (LACs), 0.28 (MACs),

354   0.231 (SACs). The discriminant plot obtained using the two first LVs is shown in Figure

355   4a.

356   -   Figure 4 –

357   According to this figure, LV1 (57.6%) separated NC and LAC samples from MAC

358   and SAC ones. This LV was mainly related to the presence of several pyrazines in the

359   NCs and their decreased quantity in the SACs. Thus, in Figure 4a on the positive side of

360   LV1 was acetophenone and on the negative side, the compounds listed in the figure as 2,

361   3, 4, 5, 6, 7, 8, 9, 11, 10, 12, 13, 14, 15, and furfural. On the other hand, the LV2 (15.82%)

362 was associated with the discrimination between NC and SAC samples, and LAC and

363 MAC ones. The volatile compounds responsible for this separation were: heneicosane,

364 ethyl hexadecanoate and toluene that were on the positive side of the LV2, n-

365 hexadecanoic acid, 5-hydroxy-2-decenoic acid δ-lactone, α-

366 ethylidenbenzeneacetaldehyde and acetophenone, with negative loadings. It is

367 noteworthy to highlight the direct negative relationship between, toluene, furfural and

368 acetophenone. The position at the score plot suggested that in samples with a high degree

369 of alkalization (i.e. MAC), the acetophenone levels increased while the toluene and

370 furfural decreased (Fig. 4a). According to the relative concentrations found,

371 acetophenone content was 106% and 130% higher in SACs and MACs, respectively, than

372 in NCs. On the other hand, toluene was 58% and 79% lower in SACs and MACs,

373 respectively, while furfural was 85% and 82% lower in SACs and MACs, respectively,

374 when compared to NCs.

375 Figure 4b shows the discriminant plot obtained using LV1 and LV3. The volatile

376 compounds associated with LV3 on the positive side were n-hexadecanoic acid (fatty

377 acid), dodecane (alkane) and ethyl hexadecanoate (fatty acid ethyl ester) in which a direct

378 positive relationship can be elucidated, this in part of the NC and LACs cocoa samples.

379 In this figure, a direct negative relationship between acetophenone (ketone) (negative side

380 of LV3), toluene (alkane) and tetramethylpyrazine (pyrazine) (positive side of LV3) was

381 observed, which could be responsible of discriminating between some LAC and MAC

382 samples. Focusing on tetramethylpyrazine, in the change from LAC to MAC its level was

383 lower while the acetophenone level was higher. Based on their relative concentrations,

384 acetophenone was a 10% higher in MACs with respect to its content in LACs. Regarding

385 toluene, it was 34% lower in MACs with respect to LACs, while tetramethylpyrazine was

386 38% lower in MACs with respect to its content in LAC. Thus, all these found

387 relationships are important aiming to standardize the quality of cocoa products (Valverde

388 et al., 2020; Calvo et al., 2021). It is noteworthy that acetophenone is a ketone produced

389 during fermentation, and it has been identified that while longer the fermentation, the

390 higher the content, it can be seen that this compound remains after the alkalization of

391 cocoa, this has been also found by Sioriki et al. (2022).

392   Finally, the model was able to classify the cross-validation set with a sensitivity of

393 100% for NC, 85.7 % for LAC, 87.5% f for MAC, and 100% for SAC, while the

394 specificity was 97.6% for NC, 93.5% for LAC, 93.2% for MAC and 97.9% for SAC,

395 showing that it is possible to easily discriminate between the samples that are not

396 alkalized (NC) with the strongly alkalized (SAC) ones, being not as absolutely classified

397 the LAC and MAC samples.

398

399 **4. Conclusions**

400   The SBSE-GC-MS coupled to PARAFAC2 automated in PARADISe allowed the

401 determination and accurate identification of 56 volatile compounds in 30 samples of

402 cocoa with different degrees of alkalization in a simultaneous way, which is timely

403 manner in comparison to the traditional methods for the monitoring of the volatile profile

404 of cocoa products.

405   Wide variability in relation to the volatile compounds in NC was evident, along with

406 a standardized volatile profile that was less variable after a marked alkalization process.

407   Key volatile compounds were associated with the level of alkalization of the cocoa

408 powder, most of them pyrazines, but also some such as nonanal, directly associated with

409 alkalization. The unsupervised method PCA explained 80% of the variability of the data

410 by four PCs, which allowed groupings of samples and variables associated with NC,

411 LAC, MAC and SAC samples. The supervised method PLS-DA with variable selection

412 enables better discrimination of the NC and SAC samples. Dynamics between volatile

413 compounds such as the increase of acetophenone and the decrease of toluene and furfural

414 allow the differentiation between NC-LAC and MAC-SAC samples.

415   This approach offers a profound knowledge of cocoa powder and can be an important

416 aid to produce high-quality and innovative cocoa products with the possibility of

417 optimizing cost savings, reducing time and improving processing practices. Facts that are

418 also extremely helpful in the specialty industry.

419

428

429

430 **References**

431 Aculey, P. C., Snitkjaer, P., Owusu, M., Bassompiere, M., Takrama, J., Nørgaard, L.,

432   Petersen M.A., Nielsen, D. S. (2010). Ghanaian cocoa bean fermentation

433   characterized by spectroscopic and chromatographic methods and chemometrics.

434   *Journal of Food Science, 75*(6), 300–307.

435 Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2008). Flavor Formation and

436   Character in Cocoa and Chocolate: A Critical Review. *Critical Reviews in Food*

437   *Science and Nutrition, 48*(9), 840–857. https://doi.org/10.1080/10408390701719272

438 Álvarez, C. (2017). Identification of the Volatile Compounds in the roasting Venezuela

439   Criollo cocoa beans by Gas Chromatography-Spectrometry Mass. *Journal of*

440   *Nutritional Health & Food Engineering, 5*(4), 659–666.

441   https://doi.org/10.15406/jnhfe.2016.05.00178

442 Amigo, J. M., Skov, T., Bro, R., Coello, J., & Maspoch, S. (2008). Solving GC-MS

443   problems with PARAFAC2. *TrAC - Trends in Analytical Chemistry*, *27*(8), 714–725.

444   https://doi.org/10.1016/j.trac.2008.05.011

445 Amigo, J. M., Popielarz, M. J., Callejón, R. M., Morales, M. L., Troncoso, A. M.,

446   Petersen, M. A., & Toldam-Andersen, T. B. (2010). Comprehensive analysis of

447   chromatographic data by using PARAFAC2 and principal components analysis.

448   *Journal of Chromatography A, 1217*(26), 4422–4429.

449 Aprotosoaie, A. C., Luca, S. V., & Miron, A. (2016). Flavor Chemistry of Cocoa and

450   Cocoa Products-An Overview. *Comprehensive Reviews in Food Science and Food*

451   *Safety*, *15*(1), 73–91.

452 Bonvehí, J. S. (2005). Investigation of aromatic compounds in roasted cocoa powder.

453   *European Food Research and Technology, 221*(1–2), 19–29.

454   https://doi.org/10.1007/s00217-005-1147-y

455 Botelho, B. G., Reis, N., Oliveira, L. S., & Sena, M. M. (2015). Development and
456   analytical validation of a screening method for simultaneous detection of five
457   adulterants in raw milk using mid-infrared spectroscopy and PLS-DA. *Food*
458   *Chemistry, 181*, 31–37. https://doi.org/10.1016/j.foodchem.2015.02.077

459 Calvo, A. M., Botina, B. L., García, M. C., Cardona, W. A., Montenegro, A. C., & Criollo,
460   J. (2021). Dynamics of cocoa fermentation and its effect on quality. *Scientific*
461   *Reports, 11*(1), 1–15. https://doi.org/10.1038/s41598-021-95703-2

462 Díaz, R., Pozo, O. J., Sancho, J. V., & Hernández, F. (2014). Metabolomic approaches
463   for orange origin discrimination by ultra-high performance liquid chromatography
464   coupled to quadrupole time-of-flight mass spectrometry. *Food Chemistry, 157*, 84–
465   93. https://doi.org/10.1016/j.foodchem.2014.02.009

466 Colonges, K., Jimenez, J. C., Saltos, A., Seguine, E., Loor Solorzano, R. G., Fouet, O.,
467   Argout, X., Assemat, S., Davrieux, F., Cros, E., Boulanger, R., & Lanaud, C. (2021).
468   Two Main Biosynthesis Pathways Involved in the Synthesis of the Floral Aroma of
469   the Nacional Cocoa Variety. *Frontiers in Plant Science, 12*(September), 1–24.
470   https://doi.org/10.3389/fpls.2021.681979

471 Cremer, D. R., & Eichner, K. (2000). The reaction kinetics for the formation of Strecker
472   aldehydes in low moisture model systems and in plant powders. *Food Chemistry*,
473   71(1), 37–43.

474 Deuscher, Z., Gourrat, K., Repoux, M., Boulanger, R., Laboure, H., & Le Quere, J.-L.
475   (2020). Key Aroma Compounds of Dark Chocolates Differing Organoleptic
476   Properties: A GC-O Comparative Study. Molecules, 25(1809), 32.

477 Dubrow, G. A., Forero, D. P., & Peterson, D. G. (2022). Identification of volatile
478   compounds correlated with consumer acceptability of strawberry preserves:
479   Untargeted   GC–MS   analysis.   *Food   Chemistry,   378*,   132043.

480   https://doi.org/10.1016/j.foodchem.2022.132043

481   Ducki, S., Miralles-Garcia, J., Zumbé, A., Tornero, A., & Storey, D. M. (2008).

482   Evaluation of solid-phase micro-extraction coupled to gas chromatography-mass

483   spectrometry for the headspace analysis of volatile compounds in cocoa products.

484   *Talanta, 74*(5), 1166–1174. https://doi.org/10.1016/j.talanta.2007.08.034

485   Frank, D., & Pat, S. (2007). Stir bar sorptive extraction for trace analysis. *Journal of*

486   *Chromatography   A,   1152*(1–2),   54–69.

487   https://doi.org/10.1016/j.chroma.2007.01.032

488   Frauendorfer, F., Schieberle, P., & Chieberle, P. E. S. (2008). Changes in Key Aroma

489   Compounds of Criollo Cocoa Beans During Roasting, *Journal of Agricultural and*

490   *Food Chemistry, 56,* 10244–*10251.* https://doi.org/10.1021/jf802098f

491   Indahl, U. G., Martens, H., & Næs, T. (2007). From dummy regression to prior

492   probabilities in PLS-DA. Journal of Chemometrics, 21(August), 529–536.

493   https://doi.org/10.1002/cem

494   Johnsen, L. G., Skou, P. B., Khakimov, B., & Bro, R. (2017). Gas chromatography – mass

495   spectrometry data processing made easy. *Journal of Chromatography A, 1503,* 57–

496   64. https://doi.org/10.1016/j.chroma.2017.04.052

497   Kongor, J. E., Hinneh, M., de Walle, D. Van, Afoakwa, E. O., Boeckx, P., & Dewettinck,

498   K. (2016). Factors influencing quality variation in cocoa (Theobroma cacao) bean

499   flavour   profile   -   A   review.   *Food   Research   International,   82,   44–52.*

500   https://doi.org/10.1016/j.foodres.2016.01.012

501   Krings, U., Zelena, K., Wu, S., & Berger, R. G. (2006). Thin-layer high-vacuum

502   distillation to isolate volatile flavour compounds of cocoa powder. *European Food*

503   *Research   and   Technology,   223*(5),   675–681.   https://doi.org/10.1007/s00217-006-

504   0252-x

505 Lacalle-Bergeron, L., Portolés, T., Sales, C., Carmen Corell, M., Domínguez, F., Beltrán,

506   J., Vicente Sancho, J., & Hernández, F. (2020). Gas chromatography-mass

507   spectrometry based untargeted volatolomics for smoked seafood classification. *Food*

508   *Research*   *International*,   *137*(September).

509   https://doi.org/10.1016/j.foodres.2020.109698

510 Li, Y., Feng, Y., Zhu, S., Luo, C., Ma, J., & Zhong, F. (2012). The effect of alkalization

511   on the bioactive and flavor related components in commercial cocoa powder. *Journal*

512   *of Food Composition and Analysis, 25*(1), 17–23.

513   https://doi.org/10.1016/j.jfca.2011.04.010

514 Liu, M., Liu, J., He, C., Song, H., Liu, Y., Zhang, Y., Wang, Y., Guo J., Yang, H., Su, X.

515   (2017). Characterization and comparison of key aroma-active compounds of cocoa

516   liquors from five different areas. *International Journal of Food Properties, 20*(10),

517   2396–2408. https://doi.org/10.1080/10942912.2016.1238929.

518 Marsol-Vall, A., Balcells, M., Eras, J., & Canela-Garayoa, R. (2018). Development of a

519   SBSE-TD method coupled to GC–MS and chemometrics for the differentiation of

520   variety and processing conditions in peach juices. Food Chemistry, 239, 119–125.

521   https://doi.org/10.1016/j.foodchem.2017.06.070

522 Marrufo-Curtido, A., Cejudo-Bastante, M. J., Durán-Guerrero, E., Castro-Mejías, R.,

523   Natera-Marín, R., Chinnici, F., & García-Barroso, C. (2012). Characterization and

524   differentiation of high quality vinegars by stir bar sorptive extraction coupled to gas

525   chromatography-mass spectrometry (SBSE-GC-MS). *LWT - Food Science and*

526   *Technology*, *47*(2), 332–341. https://doi.org/10.1016/j.lwt.2012.01.028

527 Miller, K. B., Hurst, W. J., Payne, M. J., Stuart, D. A., Apgar, J., Sweigart, D. S., & Ou,

528   B. (2008). Impact of alkalization on the antioxidant and flavanol content of

529   commercial cocoa powders. *Journal of Agricultural and Food Chemistry, 56*(18),

530   8527–8533. https://doi.org/10.1021/jf801670p

531   Mohamadi Alasti, F., Asefi, N., Maleki, R., & SeiiedlouHeris, S. S. (2019). Investigating

532   the flavor compounds in the cocoa powder production process. *Food Science and*

533   *Nutrition*, 7(12), 3892–3901. https://doi.org/10.1002/fsn3.1244

534   Moreira, I. M. da V., Vilela, L. de F., Santos, C., Lima, N., & Schwan, R. F. (2018).

535   Volatile compounds and protein profiles analyses of fermented cocoa beans and

536   chocolates from different hybrids cultivated in Brazil. *Food Research International,*

537   *109*(September 2017), 196–203. https://doi.org/10.1016/j.foodres.2018.04.012

538   Mota-Gutierrez, J., Barbosa-Pereira, L., Ferrocino, I., & Cocolin, L. (2019). Traceability

539   of functional volatile compounds generated on inoculated cocoa fermentation and its

540   potential health benefits. Nutrients, 11(4).

541   Pérez, É., Lerma, M. J., Fuentes, A., Palomares, C., & Barat, J. M. (2016). Control of

542   undeclared flavoring of cocoa powders by the determination of vanillin and ethyl

543   vanillin by HPLC. *Food Control, 67*, 171–176.

544   Petersen, M. A., & Bro, R. (2018). PARADISe - A ground-breaking tool to treat complex

545   GC-MS datasets, 421–426. https://doi.org/10.3217/978-3-85125-593-5-89

546   Plassmann, M. M., Tengstrand, E., Åberg, K. M., & Benskin, J. P. (2016). Non-target

547   time trend screening: a data reduction strategy for detecting emerging contaminants

548   in biological samples. *Analytical and Bioanalytical Chemistry, 408*(16), 4203–4208.

549   https://doi.org/10.1007/s00216-016-9563-3

550   Quelal-Vásconez, M. A., Pérez-Esteve, É., Arnau-Bonachera, A., Barat, J. M., & Talens,

551   P. (2018). Rapid fraud detection of cocoa powder with carob flour using near infrared

552   spectroscopy. *Food Control, 92*, 183–189.

553   Quelal-Vásconez, M. A., Lerma-García, M. J., Pérez-Esteve, É., Talens, P., & Barat, J.

554   M. (2020). Roadmap of cocoa quality and authenticity control in the industry: A

555   review of conventional and alternative methods. *Comprehensive Reviews in Food*

556   *Science and Food Safety*, 19(2), 448–478.

557   Schymanski, E. L., Singer, H. P., Longrée, P., Loos, M., Ruff, M., Stravs, M. A., Ripollés

558   Vidal, C., & Hollender, J. (2014). Strategies to characterize polar organic

559   contamination in wastewater: Exploring the capability of high resolution mass

560   spectrometry. *Environmental Science and Technology, 48*(3), 1811–1818.

561   https://doi.org/10.1021/es4044374

562   Sioriki, E., Tuenter, E., de Walle, D. Van, Lemarcq, V., Cazin, C. S. J., Nolan, S. P.,

563   Pieters, L., & Dewettinck, K. (2022). The effect of cocoa alkalization on the non-

564   volatile and volatile mood-enhancing compounds. *Food Chemistry, 381*(January),

565   132082. https://doi.org/10.1016/j.foodchem.2022.132082

566   Telgheder, U., Bader, N., & Alshelmani, N. (2018). Stir bar sorptive extraction as a

567   sample preparation technique for chromatographic analysis: An overview. Asian

568   *Journal of Nanoscience and Materials, 1*(March), 54–60.

569   Valverde García, D., Pérez Esteve, É., & Barat Baviera, J. M. (2020). Changes in cocoa

570   properties induced by the alkalization process: A review. *Comprehensive Reviews in*

571   *Food Science and Food Safety, 19*(4), 2200–2221.

572   Wei, F., Furihata, K., Miyakawa, T., & Tanokura, M. (2014). A pilot study of NMR-

573   based sensory prediction of roasted coffee bean extracts. *Food Chemistry, 152*, 363–

574   369. https://doi.org/10.1016/j.foodchem.2013.11.161

575

576   **Figure captions**

577   **Figure 1:** Example of the PARADISe results: a) Raw data of interval, b) estimated elution

578   profiles of compounds (green and light green) and baseline (blue), and c) estimated mass

579   spectra evaluation of 2,3-dimethyl-5-ethylpyrazine (upper) and furfural (lower).

580

581   **Figure 2:** Biplot of the PC1 and PC2, colored according to the degree of alkalization:

582   NC, LAC, MAC, SAC and UK.

583

584   **Figure 3:** a) Distribution of NCs and SACs in the score plot of PC1-PC3 and b)

585   distribution of LACs in the score plot of PC1-PC4.

586

587   **Figure 4:** a) PLS-DA biplot obtained using LV1 and LV2 and b) LV1-LV3,

588   discrimination of cocoas with different degree of alkalization.
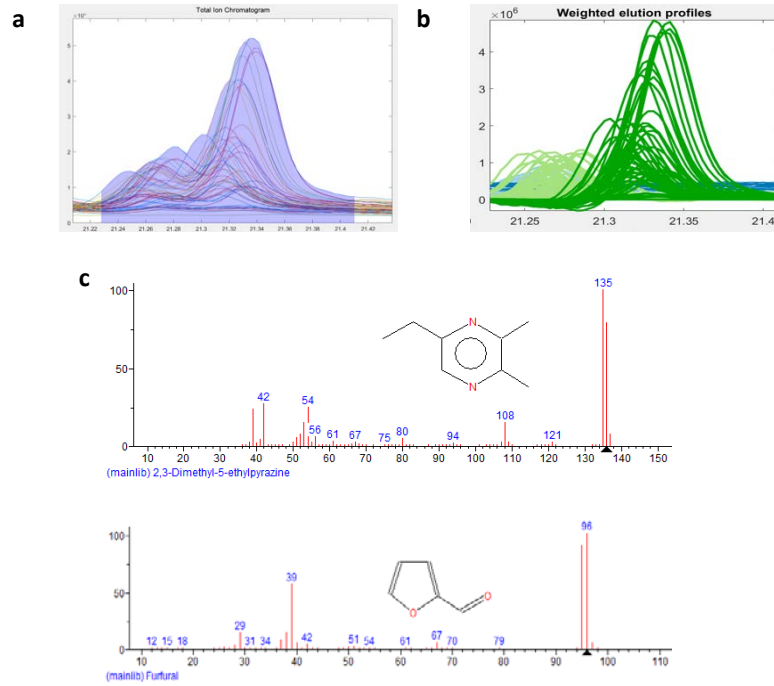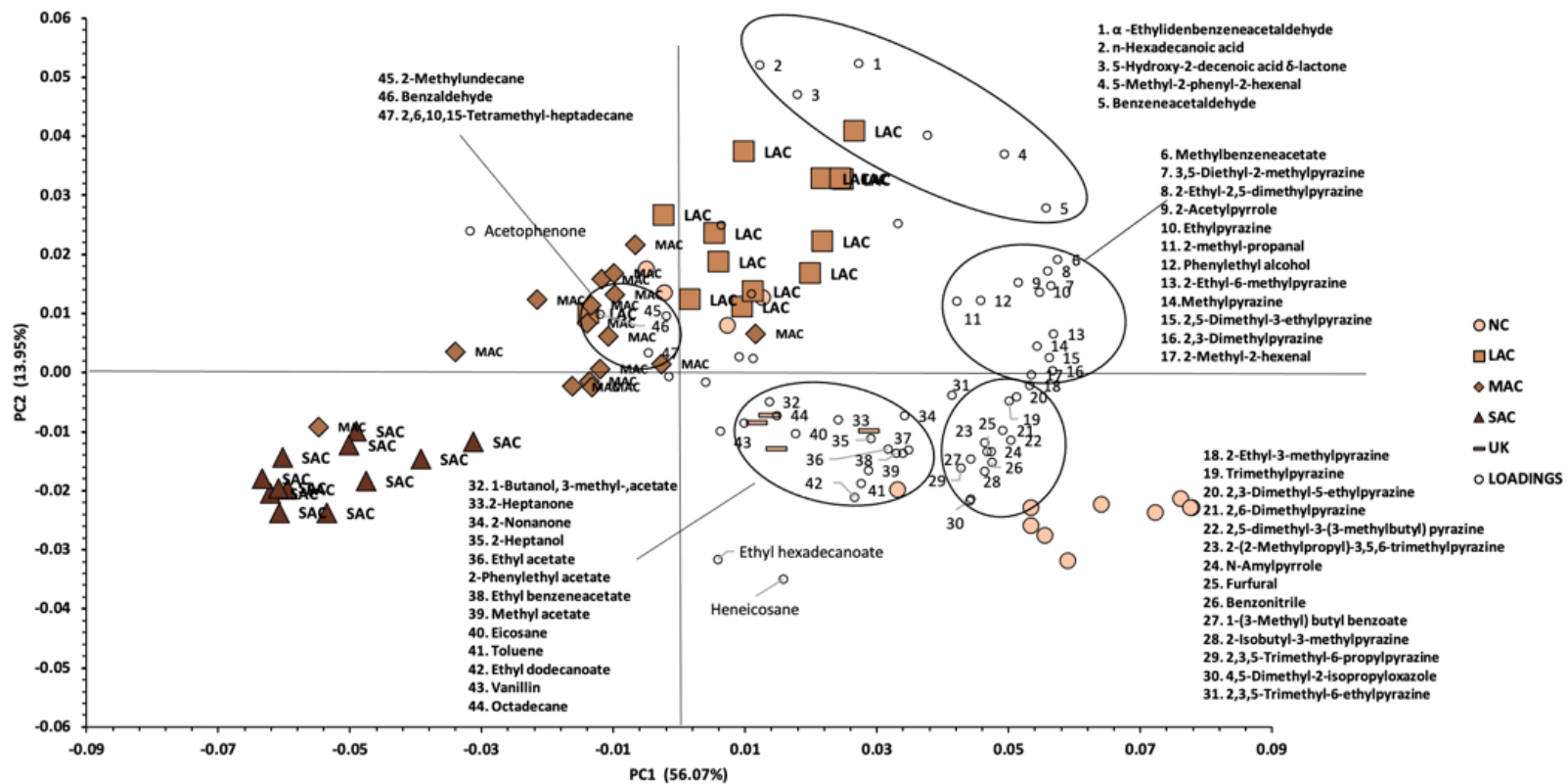
Figure 1: M.A. Quelal-Vásconez et al. Color

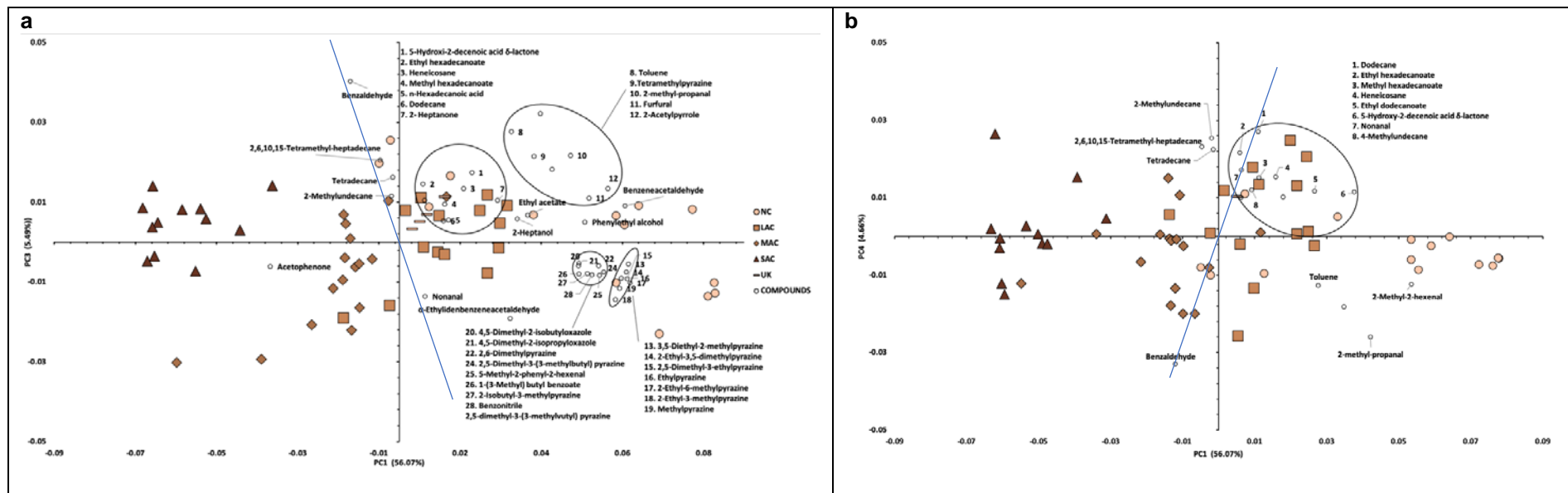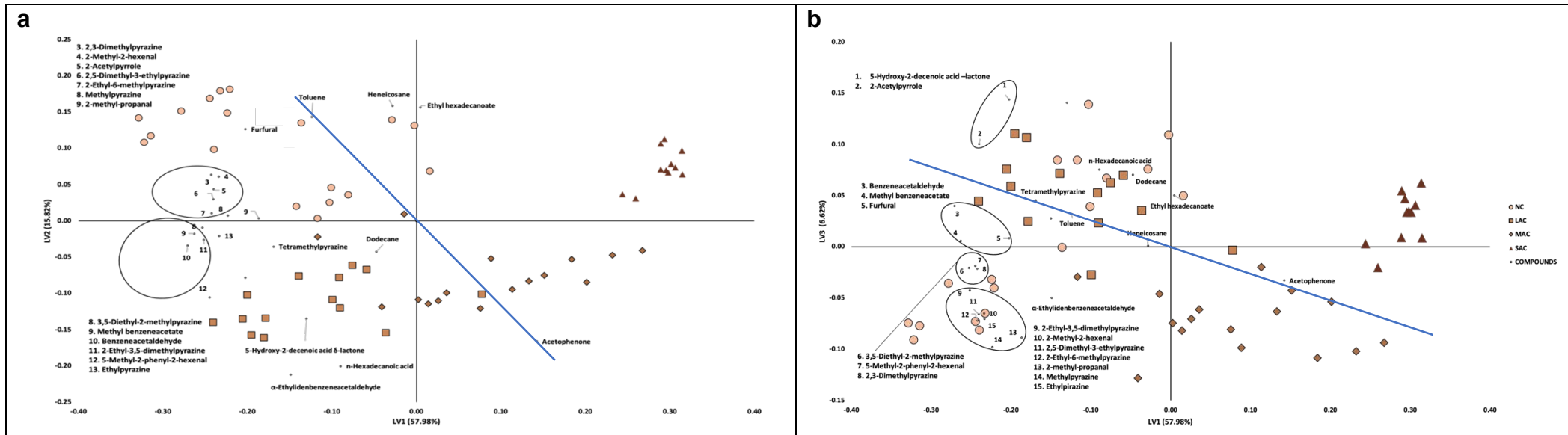**Figure 2:** M.A. Quelal-Vásconez et al. Color

Figure 3: M.A. Quelal-Vásconez et al. Color

Figure 4: M.A. Quelal-Vásconez et al. Color