Assessment of volatile fingerprint by HS-SPME/GC-qMS and E-nose for the classification of cocoa bean shells using chemometrics

Assessment of volatile fingerprint by HS-SPME/GC-qMS and E-nose for the classification of cocoa bean shells using chemometrics

Letricia Barbosa-Pereira*, Olga Rojo-Poveda, Ilario Ferrocino, Manuela Giordano, Giuseppe Zeppa

Department of Agriculture, Forest and Food Sciences (DISAFA), University of Turin, Grugliasco, Italy

Abstract

Cocoa bean shell (CBS) is a main by-product of cocoa processing, with great potential to be used as ingredient for functional foods due to its nutritional and flavour properties. This study aimed to characterize and classify CBSs yielded from cocoa beans of diverse cultivars and collected in different geographical origins through their volatile profile assessed by headspace solid-phase microextraction gas chromatography-mass spectrometry (HS-SPME/GCqMS) and E-nose combined with Principal Component Analysis (PCA). The study provides, for the first time, in a large set of samples a comprehensive fingerprint and semi-quantitative data for more than 100 volatile organic compounds (VOCs) such as aldehydes, ketones, pyrazines, alcohols, acids and others. Through PCA, a clear separation of *Criollo* cultivar from the others cultivars were achieved with both GC-qMS and E-nose analytical techniques due to the high content of key-aroma VOCs. Several biomarkers identified by GC-qMS, such as 2-hepanol, 2-methylpropanoic acid and 2,3,5-trimethylpyrazine, recognized as key-aroma compounds for cocoa, were found suitable for the classification of CBS according to the quality and origin of CBS. GC-qMS and E-nose appeared to be suitable analytical methodologies to classify CBS with high correlation between both analytical techniques. The volatile profile and classification of CBS will allow the selection of samples with a specific flavour profile according to the food application and therefore constitute an interesting approach to valorise this by-product as a food ingredient.

Keywords (6-12)

Cocoa bean shell, Cocoa by-product, HS-SMPE-GC-qMS, E-nose, Principal component analysis, volatile fingerprint, chemical markers, cocoa flavour/aroma

Highlights (85 characters, including spaces)

- The volatile fingerprint of CBS was established by HS-SPME/GC-qMS
- CBSs were classified by GC-qMS/PCA according to geographical origin and cultivar
- Several cocoa key-flavour markers present in CBSs contribute to the classification
- GC-qMS and E-nose allowed the discrimination of *Criollo* CBSs
- High correlations were found between E-nose and GC-qMS data sets

Abbreviations

CBS, cocoa bean shell; HS-SPME/GC-qMS, headspace solid-phase micro-extraction coupled with gas chromatography–quadrupole mass spectrometry; VOC, volatile organic compound; E-nose, electronic nose; ISTD, internal standard; PCA, principal component analysis.

1. Introduction

Cocoa bean (*Theobroma cacao* L.) is a ubiquitous edible product consumed across the world, of great economic significance, and the key raw material for chocolate manufacturing (Aprotosoaie, Luca, & Miron, 2016). According to the International Cocoa Organization, the world production of cocoa beans reached 4.7 million tonnes in the season 2016/2017 and the major producers are the West African countries, Ivory Coast and Ghana, and countries located in other tropical areas, like Central and South America (Brazil and Ecuador) or Southern Asia (ICCO, 2018).

The world cocoa market typically separates cocoa beans in two main categories according to their flavour, namely bulk or basic cocoa and fine or flavour cocoa (Afoakwa, Paterson, Fowler, & Ryan, 2008). Bulk cocoa is mainly produced from *Forastero* cultivar, with ordinary flavour properties and make up 95% of the world's total cocoa production. While, fine grade cocoa is exclusively produced from *Criollo*, *Trinitario* and *Nacional* cultivars, the last one grown in Ecuador, and is characterized for their remarkable flavour properties due to the fruity and floral aroma attributes (Saltini, Akkerman, & Frosch, 2013; Aprotosoaie, Luca, & Miron, 2016). Even though almost all the cocoa cultivated worldwide is *Forastero*, differences can be found in the flavour profiles of cocoa-derived products produced with cocoa beans from different geographical origins (Magagna et al., 2017; Oliveira et al., 2016, Than et al., 2015). Indeed, the quality and flavour of cocoa are not simply affected by genotype and geographical origin, but also of other factors, such as growth conditions, post-harvest treatments and industrial processing of beans (Kongor et al., 2016). In particular, fermentation and roasting are key steps responsible for the characteristics and desirable organoleptic properties of cocoa, such as aroma and flavour that are important quality attributes for consumer's acceptability (Afoakwa et al., 2008; Saltini et al., 2013).

To date, several hundreds of volatile organic compounds (VOCs) have been reported to characterise the cocoa aroma, mainly represented by pyrazines, aldehydes, ketones, alcohols, esters, furans, acids, pyrroles, phenols and terpenes (Afoakwa et al. 2008). Some of these molecules might be used as key indicators to certificate the quality and consent the discrimination of the cocoa products with label of origin, to ensure the food authentication, a new market trend of great interest for law enforcement, food producers, importers and exporters, and the consumers (Magagna et al., 2017; Danezis, Tsagkaris, Brusic, & Georgiou, 2016).

The solid-phase microextraction (SPME) coupled to Gas Chromatography (GC) Mass Spectrometry (MS) methodology has been widely used to identify and quantify the VOCs, and more recently in combination with multivariate analysis to consent the classification and discrimination of cocoa and cocoa-related products for the traceability of such products (Oliveira et al., 2016; Tran et al., 2015, Caprioli et al., 2016; Magagna et al., 2017). However, a reduce number of studies explored the potential applicability of electronic nose technologies to assess the cocoa quality and origin (Gu et al., 2013; Olunloyo, Ibidapo, & Dinrifo, 2012). To the best of our knowledge no studies are available in literature that explore and compare the potential applicability of both techniques for the classification of cocoa and related products.

Similar to other agro-food sectors, the cocoa industry also produces large amounts of by-products during manufacturing. Cocoa bean shell (CBS) is one of the main by-products generated after the roasting and husking processes of cocoa beans (about 12% of the total weight) and consequently more than 500 thousand of tonnes are produced every year that represents a disposal problem for the cocoa sector. However, recent studies have established that CBS might also be an economic source of fiber, minerals, polyphenols and methylxanthines with potential health benefits and therefore with great

99 potential to be used as an ingredient for functional foods, creating new food market
100 perspectives (Nsor-Atindana, Zhong, & Mothibe, 2012; Mandrile et al., 2019; Barbosa-
101 Pereira et al., 2017; Barbosa-Pereira, Guglielmetti, & Zeppa, 2018). Besides, Wang et
102 al. (2015) patented a process for chocolate flavour production, with a real chocolate
103 aroma, from dried CBSs using an enzymatic technology. Nevertheless, despite this
104 product exhibit great potential as a food ingredient, to the best of our knowledge, no
105 information is available in the literature describing the volatile composition of CBS. As
106 for cocoa beans, the study of CBS volatile fingerprint is very important to define the
107 quality and the flavour of the product. Moreover, selecting the CBS with a specific
108 flavour profile according to the food application could be also an interesting approach to
109 valorise this by-product as a food ingredient. Then, the aim of this study was to describe
110 for the first time the volatile fingerprint of CBS by HS-SPME/GC-qMS and determine
111 the volatile compounds responsible for differences among several CBSs yielded from
112 cocoa beans collected in different geographical origins and cultivars to allow the
113 traceability of this material. Moreover, we also explored the applicability of E-nose as a
114 rapid methodology for the classification of CBS and evaluate the correlation between E-
115 nose and GC-qMS data sets.
116
117 **2. Materials and Methods**
118 *2.1. Chemicals and Standards*
119 Methanol (≥ 99.9%), sodium chloride (≥ 99%), sodium hydroxide standard solution
120 (0.1001mol L^{-1}) and n-alkanes (n-C7–n-C30) mix standard (Supelco, Italy) for retention
121 index determination were obtained from Sigma-Aldrich (Milano, Italy). Ultrapure water
122 was prepared in a Milli-Q filter system (Millipore, Milan, Italy).

123 The Internal standard (ISTD), 5-nonanol (≥ 95% GC), for analyte response
124 normalisation was provided by Sigma-Aldrich (Milano, Italy). A standard stock solution
125 of 5-nonanol was prepared in ultrapure water at 50 mg L^{-1} concentration for the semi-
126 quantification, and stored in a sealed vial at -20 ºC.
127
128 *2.2. CBS Samples*
129 Fermented and dried cocoa beans (*Theobroma cacao* L.) from different cultivars and
130 countries across the world, harvested during the seasons of 2014 and 2015, were
131 purchased in several local cocoa companies. In total, 44 samples (2 batches each) from
132 different geographical areas in 19 countries and four cultivars (*Criollo* (n=6), *Trinitario*
133 (n=15), *Nacional* (n= 2) and *Forastero* (n=21) were collected as described in **Table 1**.
134 Specific information related with fermentation and drying conditions are not available,
135 since its suppliers retained this information confidential. To obtain CBS, object of this
136 study, all samples were roasted individually using a standardized process performed in
137 laboratory at 130 ºC (isothermal) for 20 min using a ventilated oven Memmert UFE 550
138 (ENCO, Spinea, Italy). Then, CBS samples were separated from the beans and ground
139 in a powder with 250 μm mesh size using an ultra-centrifugal mill Retsch ZM 200
140 (Retsch Gmbh, Haan, Germany). Samples were stored under vacuum at -20 °C before
141 sample preparation and headspace analysis. The humidity content of the CBS samples
142 determined using a Gibertini Eurotherm electronic moisture balance (Gibertini
143 Elettronica, Novate Milanese MI, Italy) ranged between 5.46 and 9.22 %.
144
145 *2.3. HS-SPME/GC-qMS analysis*

146 The VOCs from the CBS samples were identified and analysed using a headspace
147 solid phase micro extraction (HS-SPME) coupled with gas chromatography/quadrupole
148 mass spectrometry (GC/qMS).
149
150 *2.3.1 HS-SPME conditions*
151 For the extraction of VOCs, 2.0 g of CBS powder were accurately weighed in a 20
152 mL headspace vial for sealed. Then, 2 mL of sodium chloride (40% w/v) and 10 µL of
153 internal standard (IS) 5-nonanol (50 µg/mL) were added to the sample, and the vial was
154 immediately hermetically capped with a PTFE-silicon septum. The extraction was
155 performed in an Autosampler for SPME COMBI PAL (PAL System, Switzerland)
156 equipped with a HS-SPME unit. The sample was equilibrated at 60 ºC with stirring at
157 250 rpm for 10 min to reach equilibrium. Next, a well-conditioned SPME fiber coated
158 with divinylbenzene/ carboxen/ polydimethylsiloxane (DVB/CAR/PDMS) (d_f 50/30
159 µm, 1 cm) (Supelco, Bellefonte, PA, USA) was exposed to the headspace of the sample
160 for another 30 min with continuous heating and agitation. After extraction, the fiber was
161 inserted into the injection port of the GC system in splitless mode and desorbed at 260
162 °C for 2 min. Three identical samples were prepared for each analysis.
163
164 *2.3.2. GC–qMS Instrument and Analytical Conditions*
165 GC/qMS analyses were performed on a Shimadzu GC-2010 gas chromatograph
166 equipped with a Shimadzu QP-2010 Plus quadrupole mass spectrometer (Shimadzu
167 Corporation, Kyoto, Japan). Separation of VOCs was performed on a DB-WAXETR
168 capillary column of 30 m length, 0.25 mm internal diameter and 0.25 mm film thickness
169 (J&W Scientific Inc., Folsom, CA, USA). The oven time–temperature programme was
170 as follows: initial temperature 40 °C held for 5 min, from 40ºC to 180°C at the rate of 5

°C min^{-1} and then to 240°C at the rate of 10°C min^{-1}, which was held for 5 min. The carrier gas was helium at a constant flow of 1 ml min^{-1} with the splitless GC inlet mode. GC inlet and transfer lines were set at 260 and 240 °C, respectively. The MS fragmentation was performed by electron impact ionization mode (70 eV) and the temperatures of ion source and quadrupole were 240°C, respectively. The data were recorded in full-scan mode in the mass acquisition range of 30–450 *m/z* and 0.30 s scan time.

2.3.3. Qualitative and quantitative analysis

The identification of the volatile organic compounds, focused on 101 molecules described in **Table 2**, was performed by comparing the EI-MS fragmentation pattern of each compound with those available on the National Institute of Standards and Technology (NIST05) mass-spectral library and on our home-based library. Only compounds whose similarity is more than 75% were considered. Additionally, the confirmation of molecules identity was done by comparing the gas chromatographic retention indexes (RI) of volatile compounds, determined after injection of a series of *n*-alkane homologues (C7–C30) under the same GC–qMS analytical conditions described above, with literature data. The semi-quantitative concentrations of the VOCs identified were calculated as the area of the volatile marker component divided by the response factor of the ISTD 5-nonanol and expressed as micrograms of 5-nonanol equivalents per kg of sample (µg 5-nonanol Eq. kg^{-1} of CBS). Data were acquired and analysed by using GC/qMS Solution Workstation software (version 4.3) (GC-qMS Solution, Shimadzu Corporation, Kyoto, Japan).

2.4. E-nose analysis

196 E-nose analyses were performed using a portable electronic nose system PEN3
197 (Airsense Analytics GmbH., Germany). The system consists of a sampling unit and the
198 gas detection system composed of 10 Metal Oxide Semiconductor (MOS) sensors,
199 which are differentially sensitive to each characteristic volatile compound. The
200 chemical sensors that composed the sensor array system are the following: S1, aromatic;
201 S2, broad range; S3, aromatic; S4, hydrogen; S5, aromatic and aliphatics; S6, broad
202 range and methane; S7, sulphur organic; S8, broad range alcohol; S9, sulphur and
203 chlorinate; and S10, methane and aliphatics (Benedetti, Buratti, Spinardi, Mannino, &
204 Mignani, 2008).
205 For the analysis, 2 g of CBS powder was placed in a 20 mL glass vial and capped
206 with a PTFE septum. Then, each vial was incubated at 30ºC for 30 min to reach the
207 headspace equilibrium. The gas headspace was injected into the E-nose carried by air
208 for 90 s at a constant flow rate of 400 mL min^{-1} and during this time the sensor signals
209 were recorded each second. After each analysis, the sensor system was purged with
210 filtered air for 120 s, to allow reestablishment of the instrument baseline prior to the
211 next sample injection. The sensor response, G/G0 (G and G0 stand for the conductance
212 of the MOS connected with the sample and clean gas, respectively), is expressed as
213 resistivity (Ohm) and changed accordingly to the composition of volatile compounds.
214 Data were collected by the pattern recognition software (WinMuster, v.1.6., Airsense
215 Analytics GmbH., Germany). Three replicates of each CBS sample were independently
216 analysed and the average of sensor responses (area under the curve) was used for the
217 subsequent statistical analysis.
218
219 *2.5. Chemometric analysis / Statistical analysis*

A total of 44 CBS samples prevenient from the continents Africa and America were used to perform chemometric analysis. From these 44 samples, two badges of cocoa beans provided from the same producer were available (see **Table 1**). All samples were analysed in triplicate in a final number of 264 analyses for each methodology used (GC-qMS (101 VOCs each) and E-nose (10 sensors each). To discriminate the CBS samples as a function of geographical origin of production or variety principal-component analysis (PCA) based on the normalized data (log10) were build by using the *made4* package of R (https://www.r-project.org) and the function *dudi.pca.* Analysis of similarity based on VOCs and E-nose table was applied with 999 permutations to detect significant differences as a function of the continent, macroaera, latitude, country of production or variety, by using the anosim function in *vegan* package or R. Non-parametric Kruskal-Wallis as well as Wilcoxon tests were carried out in order to find VOCs differentially abundant between all the variable. Data were visualized as box plots represented the interquartile range between the first and the third quartile, with the error bars showing the lowest and the highest value. Pairwise Spearman's non-parametric correlations (*corr.test* function in *psych* package of R) were used to study the relationships between VOCs and sensors. The correlation plots were visualized in R using the *made4* package of R. P-values were adjusted for multiple testing and a false discovery rate (FDR) < 0.05 or lower was considered as statistically significant.

3. Results and Discussion

The study of volatile constituents (VOCs) of CBS is very important to define the quality and the flavour of the product to be used as food ingredient. The present study was divided on two main parts: the first one dedicated to the analyses of all samples using GC-qMS and E-nose to define the volatile profile and fingerprint of CBS; and then the classification of samples, using PCA analysis, and the identification of key compounds that differentiate the samples classes.

3.1. Volatile profile of CBS characterized by HS-SPME/GC-qMS

The volatile components of CBS samples, extracted and identified by HS-SPME/GC-qMS are described in **Table 2**. Each compound (VOC) is characterized by its retention index (RI), odour description as reported in literature and the different semi-quantitative concentration ranges determined in the group of samples analysed. A total of 101 compounds, comprising aldehydes (n=15), ketones (n=9), sulphur compounds (n=4), esters (n=8), hydrocarbons (n=2), furans (n=3), pyrazines (n=21), alcohols (n=7), pyrroles (n=4), terpenes, isoprenoids and terpene alcohols (n=10), acids (n=10), lactones (n=3) and others (n=5) were semi-quantified as μg kg^{-1} of 5-nonanol equivalents. The average of the amounts of each VOC, the sum of each class of compounds and the total amount of VOCs presented in the single sample is shown in detail in **Table S1** (see supplementary material). The total amount of VOCs ranged between 4.92 μg g^{-1} (VEN3) and 16.10 μg g^{-1} (VEN10), both from Venezuela, and these concentrations represent 10-20% of that total amount described by Tran et al., 2015 for roasted cocoa beans (20.6 to 142.5 μg g^{-1}). In general, the most representative classes of compounds in CBS were aldehydes (35.8%), pyrazines (18.7%), acids (11.0%), alcohols (7.9%), ketones (7.7%) and furan derivates (6.4%). The process of

265 roasting has a great impact on cocoa aroma and the alkyl pyrazines and Streker
266 aldehydes increased significantly with this stage in cocoa and consequently in CBS,
267 which is a main by-product produced during this stage. This distribution is slightly
268 different from that found in literature for roasted cocoa beans, that presented acids and
269 alcohols as the main compounds at high concentration, or for roasted cocoa liquor,
270 which displayed higher amounts of aldehydes, alcohols and ketones (Caprioli et al.,
271 2016; Tran et al., 2015; Crafack et al., 2014). However, the amounts of the several
272 classes of compounds change with the cultivars and geographical origin of the cocoa
273 beans (Bonvehí et al., 2005; Tram et al., 2015). In general, the CBS from *Trinitario*,
274 *Criollo* and *Nacional* cocoa cultivars are those with highest amounts of VOCs than
275 *Forastero* group. *Criollo* and *Nacional* cultivars display, in average, high amounts of
276 pyrazines, acids, alcohols and ketones than *Trinitario* and *Forastero* cultivars. This data
277 is in accordance with those find in literature for cocoa beans (Quin et al., 2017).
278 Since, no data are available in literature, the results of the present work will be
279 discussed by comparison with studies performed for roasted cocoa beans and cocoa
280 products, such as dark chocolate and cocoa powder described in several works found in
281 literature (Tram et al., 2015, Bonvehí et al., 2005; Menezes et al., 2016; Owusu,
282 Petersen, & Heimdal, 2012; Afoakwa et al., 2009).
283 Aldehydes were the most representative aroma compounds in CBS with total
284 amounts ranging from 1444.83 μg kg^{-1} to 5122.55 μg kg^{-1} quantified in samples yielded
285 from cocoa beans from Togo (TOG2) and Sao Tomé (SAT3) respectively, similar or
286 higher amounts than that found in roasted cocoa beans (1.22-3.84 μg g^{-1}) (Tran et al.,
287 2015). Among aldehydes, 2-methylpropanal, 3-methylbutanal, nonanal, benzaldehyde
288 and phenylacetaldehyde were the most abundant in CBS as in cocoa beans (Bonvehí et
289 al., 2005; Tram et al., 2015). The Strecker aldehydes 2-methylpropanal, 3-

290 methylbutanal and phenylacetaldehyde, formed during fermentation and roasting
291 processes, are described in literature as flavour-active compounds and as key-aroma
292 markers having a strong chocolate character with malty and buttery notes for the first
293 two compounds and pleasant honey-like and nutty notes for phenylacetaldehyde
294 (Afoakwa et al., 2009). Other aldehydes identified in CBS, such as 2-phenyl-2-butenal,
295 nonanal, 5-methyl-2-phenyl-2-hexenal and 2-isopropyl-5-methyl-hex-2-enal (isomers 1
296 and 2) has been also described as contributors for the cocoa odour and quality of final
297 products conferring cocoa and fruity notes (Menezes et al., 2016; Owusu, Petersen, &
298 Heimdal, 2012; Bonvehí et al., 2005).

299 Pyrazines were one of the most representative groups of VOCs present in CBS with
300 concentrations ranging from 199.34 μg kg^{-1} for the Sierra Leone (SLE) samples to
301 5285.68 μg kg^{-1} for the Venezuela (VEN9) samples as observed in several cocoa beans.
302 In this study, 2,3,5,6-tetramethylpyrazine was the most abundant pyrazine in CBS, up to
303 3298.06 μg kg^{-1} in samples from Madagascar (MAD) that represented more than 50%
304 of total amount of pyrazines present in all CBS samples. 2,3,5,6-Tetramethylpyrazine is
305 one of the main components of CBS aroma that exhibited nutty and roasted and
306 chocolate flavour notes as described in literature for dark chocolate (Afoakwa et al.,
307 2009). Other pyrazines identified in CBS were 2,3,5-trimethylpyrazine, 2,3-dimethyl-5-
308 ethylpyrazine, 2,3-dimethylpyrazine, 2-ethyl-6-methylpyrazine, 2,3,5-trimethyl-6-
309 ethylpyrazine, 2-methylpyrazine, 2,6-dimethylpyrazine. All these compounds, derived
310 from Maillard reactions and they are characteristic and responsible for the cocoa aroma,
311 providing to CBS samples essential notes of cocoa, roasted, caramel, baked, nutty and
312 earthy. 2,3,5-Trimethylpyrazine and 2-ethyl-3,5-dimethylpyrazine are recognised as
313 key-aroma compounds for cocoa and cocoa products and therefore also for CBS
314 (Frauendorfer et al., 2006; Frauendorfer et al., 2008). A tentative identification was also

315 performed for the pyrazine 2,5-dimethyl-3-isopentylpyrazine present in important
316 amounts, up to 321.09 μg kg^{-1} for CBS sample from Dominican Republic (DOR1),
317 which is described for the first time in this work for cocoa products.

318 Another important group of VOCs consists of short and branched chain fatty acids
319 such as acetic acid, 2-methylpropanoic acid and 3-methylbutanoic acid, which are key-
320 aroma compounds used as markers of cocoa and cocoa products. The total amount of
321 acids ranging from 100.43 μg kg^{-1} to 4450.96 μg kg^{-1} were found in samples yielded
322 from Cameroon (CAM2) and Venezuela (VEN10), respectively. These amounts were
323 lower than that described by Tran et al. (2015) for cocoa beans ranging from 8.27 to
324 95.47 μg g^{-1}. Despite acetic acid was found a major compound in CBS, the
325 concentrations of this acid (21.76 – 1612.24 μg kg^{-1}) were lower than those described
326 for cocoa beans (6.66 – 95.17 μg g^{-1}). However, the concentrations found for 2-
327 methylpropanoic acid (2.07 – 454.27 μg kg^{-1}) and 3-methylbutanoic acid (36.27 –
328 2154.25 μg kg^{-1}) were similar to those described for cocoa products in literature (Tran
329 et al., 2015; Bonvehí et al., 2005). Although these acids are generally related with
330 unpleasant odour because of their rancid, sour-vinegar and hammy notes in cocoa
331 products, some acids present in CBS such as octanoic acid and nonanoic acid could
332 show a pleasant odour with sweet notes.

333 Concerning alcohols, the total amount semi-quantified for these group of VOCs was
334 in the range of 135.83 for *Forastero* CBS from Sao Tomé (SAT1) and 2310.18 μg kg^{-1}
335 for *Criollo* CBS from Ecuador (ECU7), lower than that described by Tran et al., 2015
336 for cocoa beans (5.83 – 27.07 μg g^{-1}). The main alcohols found in CBS were two key-
337 aroma compounds of cocoa: 2-heptanol (19.36 – 1655.07 μg kg^{-1}) with citrus notes; and
338 2-phenylethanol (52.01 – 936.97 μg kg^{-1}) that confer flowery, honey caramel and sweet

notes and was present at higher concentrations than that found in cocoa beans and cocoa powder (Tran et al., 2015; Bonvehí et al., 2005).

2,3-Butanedione, 2-heptanone and 2-nonanone were the main ketones up to 556, 362.87 and 293.89 μg kg^{-1}, respectively, present in *Criollo* and *Nacional* CBS samples from Ecuador (ECU4, ECU5 and ECU7) that contribute to the aroma with sweet, buttery, fruity and flowery notes.

Esters were other key VOCs present in CBS associated with fruity, floral and sweet notes attributed to cocoa aroma. The key-aroma marker, 2-phenylethyl acetate, characteristic by their honey, sweet and fruity notes, was the main ester present in the CBS followed by 3-methylbutyl acetate, methyl 2-phenylacetate and ethyl benzeneacetate. Despite the total amount of this group of VOCs was lower (73.35 – 1036.65 μg kg^{-1}) than those found in cocoa products, the main esters identified in CBS samples from specific origins (e.g. Peru, Tanzania, Togo and Venezuela), were found in similar concentration than that found for cocoa beans (Tran et al., 2015).

Another VOCs derived from roasted cocoa identified in CBS, were the terpenes linalool and linalool oxide, both characteristic key chocolate flavours, characterized by sweet, nutty, fruity, floral/flowery notes (Afoakwa et al., 2009; Bonvehí et al., 2005). Furthermore, the pyrroles; 1H-pyrrole-2-carboxaldehyde, characterized by nutty, honey and candy notes; and 2-acetylpyrrole with notes of hazelnut, cocoa, chocolate, were identified and quantified in CBS in important amounts (up to 223.47 μg kg^{-1} for CBS from Brazil (BRA) and up to 437.06 μg kg^{-1} for CBS from Venezuela (VEN2), respectively). Further compound with odour description of cocoa, chocolate and roasted cocoa that contribute for the CBS flavour of cocoa was acetylfuran. Likewise in the furan derivate group, furfural is correlated with almond, caramel, sweet, woody and flowery notes of cocoa, was identified and quantified at high concentrations in *Criollo*

364 CBS samples from Venezuela (VEN10) (3911.55 μg kg^{-1}). Finally, dimethyl
365 trisulphide, also described as key-aroma compound for cocoa products, and dimethyl
366 disulphide were identified and quantified in a range between 3.86 – 284.33 μg kg^{-1} and
367 5.52 – 291.23 μg kg^{-1}, respectively. The highest amounts of both compounds were
368 detected in CBS samples from Dominican Republic (DOR1) of *Trinitario* cultivar.
369 The present study identified other VOCs, present in lower concentrations, which are
370 described for the first time for cocoa related products and may contribute for the total
371 pleasant aroma of CBS (see **Table 2**, compounds highlighted with asterisk symbol, *).
372 Some of these molecules were ketones such as 2-decanone, 3-methyl-2-cyclohexen-1-
373 one, 2-undecanone, characterized by nutty, floral and fruity notes; pyrazines such as
374 2,3,5-trimethyl-6-isopentylpyrazine with floral notes; several terpenes that contribute
375 with sweet, floral, fruity and citrus notes to the CBS aroma; and Massoialactone (S and
376 R) that may confer coconut and nutty notes.
377
378 *3.2. Classification of CBS based on VOCs determined by SPME-HS-GC-qMS*
379 *3.2.1. Classification of CBS according to the cultivar and continent of origin – first*
380 *approach*
381 Figure 1 shows the Principal Component Analysis (PCA) based on volatile
382 fingerprinting of CBS that was used to find difference among type of cultivars (Fig. 1a)
383 and continent of provenience (Fig. 1b). PCA clearly showed a separation ($p<0.001$) of
384 *Criollo* CBS if compared with the other cultivars that clustered together (Fig. 1a)
385 confirmed by ANOSIM statistical test. By taking into the account the continent of
386 provenience it was possible to observe a clear separation ($p<0.007$) of American and
387 African CBS samples (Fig. 1b).

388 Going more deeply in the volatile composition, the level of diversity of the VOCs
389 was clearly different based on the CBS cultivar. Several compounds (48 VOCs) were
390 found significant different according to the cultivar (FDR<0.001 (10 VOCs), FDR<0.01
391 (17 VOCs) and FDR<0.05 (21 VOCs)) as shown in **Table S2.1** (see supplementary
392 material). Key aroma compounds such 2-methylpropanal, phenylacetaldehyde, 2,3,5-
393 trimethylpyrazine, 2-ethyl-3,5-dimethylpyrazine, 2-heptanol, 2-phenylethanol, 2-
394 methylpropanoic acid and 3-methylbutanoic acid were found significant for the CBS
395 discrimination of *Criollo* cultivar from the *Forastero* and *Trinitario* cultivars. Other
396 compounds such as, benzaldehyde, methyl-2-phenylacetate, 2,3-dimethylpyrazine and
397 2,3,5,6-tetramethylpyrazine present at high concentrations in *Criollo* CBSs can be also
398 putative markers of *Criollo* cultivar. Tetramethylpyrazine was the most abundant
399 pyrazine present in CBS yielded from *Criollo* cocoa beans as described by Tran et al.
400 (2015) for cocoa beans of the same cultivar. The 3-methylbutanoic acid was found as a
401 potential marker for the *Forastero* CBS and 2-phenylethanol and 2-heptanol for
402 *Trinitario* CBS, as already described for cocoa beans (Quin et al., 2017). The boxplot of
403 three volatile compounds highly significant for the classification of CBS according to
404 cultivar is shown in **Fig. 1c** 2-methylpropanoic acid was found at high concentrations in
405 *Criollo* CBS allowed to discriminate this cultivar from *Forastero* and *Trinitario*
406 cultivars (FDR<0.001) and distinguished CBS *Trinitario* from the *Forastero* and
407 *Nacional* cultivars (FDR<0.01 and FDR<0.05, respectively). Also 2-heptanol allowed
408 to differentiate *Trinitario* CBS from the other cultivars (FDR<0.05) and additionally
409 distinguished *Forastero* CBS from *Criollo* and *Nacional* (FDR<0.01 and FDR<0.05,
410 respectively). Finally, 2,3,5-trimethylpyrazine was found highly significant to
411 differentiate *Criollo* CBS from *Forastero* and *Trinitario* cultivars (FDR<0.01).

412 Taking into the account the geographical origin, several compounds (47 VOCs) were
413 found significant to the classification of CBS according to the continent of origin
414 (FDR<0.001 (11 VOCs), FDR<0.01 (19 VOCs) and FDR<0.05 (17 VOCs)) as shown in
415 **Table S2.2** (see supplementary material). Considering the key aroma markers identified
416 in cocoa samples (Frauendorfer et al., 2008) the boxplot showed that 2-heptanol
417 (FDR<0.001) and 2-methylpropanoic acid (FDR <0.001) were found those volatiles
418 with the highest concentration in CBS of *Criollo* cocoa from American continent.
419 Moreover the most abundant pyrazine detected in CBS, 2,3,5,6-tetrametylpyrazine
420 (FDR<0.01) was also associated with American samples (Fig. 1d).
421
422 *3.2.2. Classification of Forastero CBS samples according to their geographical origins*
423 PCA analysis was performed to evaluate the sample separation according to the
424 CBSs origin of *Forastero* cultivar among 14 countries from Africa and America (see
425 **Figure 2**). A clear separation (p<0.001) was observed according to macroarea (Fig. 2a),
426 latitude (Fig. 2b) and among the country of origin (Fig. 2c). For the classification
427 according to latitude, the countries of production were distributed in four main groups:
428 L1 (5ºS–5ºN); L2 (5ºN–20ºS); L3 (5ºN–10ºN) and L4 (10ºN–20ºN). By taking into the
429 account the macroarea as a discriminant factor it was possible to observe that west
430 Africa and south Africa CBS samples cluster together and were well separated
431 (p<0.001) from east Africa and Central America (Fig. 2a). Moreover the different
432 volatile profile drove the impressive cluster separation (p<0.001) according to the
433 latitudes (Fig. 2b). In particular CBS from West Africa and latitude L3 (5ºN–10ºN)
434 were those samples with low amounts of total VOC's, mainly aldehydes, pyrazines,
435 sulphur compounds and high amounts of acids. Going more deeply in the classification
436 of the CBS samples as a function of the geographical origins it was possible to

437 differentiate ($p<0.001$) the CBS sample according to the country of origin (Fig. 2c). We
438 observed that Congo clustered together with Uganda and the two South America
439 countries, Ecuador and Colombia, formed a central group in the centre of the PCA.
440 Considering the CBS from cocoa beans produced in countries located at the latitude L2
441 (5ºN–20ºS) such Madagascar, Peru and Tanzania were those with high amounts of
442 VOCs among *Forastero* cultivars (see **Table S1**). For these samples, cocoa key aroma
443 compounds were found at high concentrations such as acids (acetic acid), aldehydes
444 (benzaldehyde, 2-methylpropanal and 3-methylbutanal), esters (3-methylbutyl acetate
445 and 2-phenylethyl acetate) and pyrazines such as 2,3,5-trimethylpyrazine and 2,3,5,6-
446 tetramethylpyrazine. It could be pointed out that CBSs from Madagascar were well
447 separated from Sao Tomé. Even though both countries are African, they belong to
448 different macroarea and growth at different latitude. However they are both islands,
449 with specific climate conditions that may affect the volatile profile of cocoa beans and
450 their products (Afoakwa et al. (2008). CBS samples from these countries were
451 characterized by low concentrations of alcohols and by the presence of 3-
452 methylbutanoic acid, 3-methylbutanal, phenylacetaldehyde, dimethyl trisulphide,
453 trimethylpyrazine and tetramethylpyrazine present at high concentrations conferring
454 important flavour characteristics to CBS that can valorise the product. CBS from the
455 Central America country Dominican Republic (L4 10ºN–20ºN) was characterized by
456 the presence of high amounts of aldehydes and pyrazines. Therefore, among *Forastero*
457 cultivar, CBS samples from Sao Tomé, Madagascar, Dominican Republic and Peru
458 were those with high amounts of VOCs and can be distinguished from the rest of
459 samples (Fig. 2c).

460 By taking into the account the key VOCs of *Forastero* that drove this separation
461 (**Table S3.1,** see supplementary material) 2-methylpropanal, 3-methylbutana,

462 phenylacetaldehyde, dimethyl trisulphide, 2-phenethyl acetate, 2-heptanol, 2-
463 phenylethanol, 2-methylpropanoic acid, 3-methylbutanoic acid, 2,3,5-trimethylpyrazine
464 and 2-ethyl-3,5-dimethylpyrazine were found as a putative markers of *Forastero* CBS
465 according to the country of origin. In details it was possible to identify two main
466 components that drove the separation among the different origins: phenylacetaldehyde
467 and furfural. In particular phenylacetaldehyde was most present in CBS from
468 Dominican Republic, Madagascar, Peru and Sao Tomé, while furfural in CBS from
469 Sierra Leone, Togo and Tanzania (Fig. 2d).
470
471 *3.2.3. Classification of Trinitario CBS samples according to their geographical origins*
472 Considering *Trinitario* CBS samples (see **Figure 3**) we can clearly observed a
473 separation ($p<0.001$) between samples from Central America and South Africa (Fig.
474 3a). Moreover taking into the account the latitude we observed that samples from L1
475 (5ºS–5ºN) and L2 (5ºN–20ºS) cluster together and were well separated ($p<0.001$) from
476 L3 (5ºN–10ºN) and L4 (10ºN–20ºN) (Fig. 3b). Going more deeply and taking into the
477 account the CBS origins we observed that CBS yielded from cocoa beans grown in
478 Central America, the following countries Dominican Republic, Jamaica and Mexico,
479 located at the latitude L4 (10ºN–20ºN), clustered together. While from South America,
480 CBS samples were divided in three latitudes, L1 (5ºS–5ºN) comprising Ecuador and
481 Colombia and L2 (5ºN–20ºS) comprising Brazil and Peru that were not separated
482 among them, and finally CBS from Venezuela at the latitude L3 (5ºN–10ºN) that were
483 separated from the other three groups (Fig. 3c). CBS from Dominican Republic, Mexico
484 and Peru were those with high amounts of total VOCs among *Trinitario* CBS, including
485 pyrazines (e.g. 2,3,5-trimethylpyrazine, see Fig. 3d.), ketones (2-nonanone) acids (acetic
486 acid) and aldehydes (phenylacetaldehyde). Also CBS from Colombia and Jamaica

displayed intermediate amount of ketones (2-nonanone) terpenes, and aldehydes (3-methylbutanal) (See Fig. 3d and **Table S1**). However, Jamaica CBS was also characterized by low amounts of esters, furan derivates and acids. CBSs yielded from cocoa beans grown in Brazil and Ecuador were characterized by low amounts of 2-methylpropanal, 2-methylbutanal and phenylacetaldehyde, and high amounts of nonanal and heptanal); the low amounts of pyrazines, esters and acids and the presence of high concentrations of furans (furfural and acetylfuran). CBS from Ecuador was separated from Brazil samples due to the high content of alcohols, mainly 2-heptanols and 2-phenylethanol (See **Table S1**).

As for *Forastero* CBS, also in this case several VOCs were found as potential markers for the classification of CBS. The volatile compounds that had significant differences (FDR<0.05 or lower) among American countries are shown in **Table S.3.2** (see supplementary material). Figure 3d, shows the boxplot of two key aroma compounds for cocoa, dimethyl trisulphide and 2,3,5-trimethylpyrazine, which contributed significantly (FDR<0.05 or lower) for the classification of CBS. In particular dimethyl trisulphide was most present in CBS from Dominican Republic, Colombia and Mexico, while 2,3,5-trimethylpyrazine in CBS from Mexico and Peru.

3.3. Volatile profile of CBS characterized by E-nose and classification based on E-nose data set

For all CBS samples, the changes in the variation of signals were found similar (data not shown). The sensors, S2 (broad), S6 (broad-methane), S7 (sulphur organic), S8 (sensitive broad alcohol) and S9 (sensitive to aromatics and organic sulphides) were those that display high response intensity. The PCA analysis of E-nose data showed that the most significant classification of CBS was according to cultivars and country of

512 origin, **Figure 4**. A clear separation of CBS ($p<0.001$) samples yielded from cocoa
513 beans of *Criollo* cultivar from the others cultivars was observed (Fig. 4a). In this case,
514 the potential of the several sensors that compose the E-nose were considered for the
515 classification of CBS. The sensors that displayed significant differences (FDR <0.05 or
516 lower) among the cultivars are shown in **Table S.4** (see supplementary material). All
517 sensors resulted highly significant for the separation of CBS from *Criollo* cultivar
518 respect to *Trinitario* and *Forastero* cultivars. These results confirm the discrimination
519 of CBS samples from *Criollo* cultivar obtained with GC-qMS data. Considering all
520 cultivars, the most representative sensors for the classification of CBS were S5
521 (sensitive to aromatic and aliphatics), S6 (broad-methane), S7 (sensitive to terpenes and
522 sulphur containing organic compounds), and S10 (methane and aliphatics).
523 Taking into the account the geographical origin, the two groups of CBS samples
524 from cultivars *Forastero* and *Trinitario* are shown in Fig. 4b and Fig. 4c, respectively.
525 For *Forastero* cultivar, CBS from Madagascar was well separated ($p<0.001$) from the
526 rest of samples. Likewise samples from Sierra Leone and Togo were separated from
527 CBS from Dominican Republic and Ecuador as observed by GC-qMS data. However,
528 E-nose was not able to separate Peru or Sao Tomé as we observed by GC-qMS data (see
529 Fig. 2c). Considering *Trinitario* cultivar samples, displayed in Fig. 4c, a significant
530 separation ($p<0.022$) of CBS samples was observed. Despite, the efficacy of separation
531 was slighter than that observed for GC-qMS analysis, the results highlighted that E-nose
532 can be used as a tool for rapid discrimination of CBS samples from different cultivars
533 and origins.
534
535 *3.4. GC-qMS vs E-nose – two case studies*

In this section CBS from two representative countries of cocoa production with the most representative number of samples, Venezuela (n=10) and Ecuador (n=7), were taking into the account to verify if both GC-qMS and E-nose were able to classify CBS among the same country of origin. A significant separation ($p<0.001$) of CBS from Venezuela and Ecuador was observed using both analytical techniques as shown in **Figure 5**.

In the case of CBS from Venezuela (Fig. 5a and 5b), both techniques allowed to separate samples from *Criollo* cultivar (VEN7, VEN8, VEN9 and VEN10) from other cultivars except for samples from Canoabo (VEN5), which was not separated with E-nose (see Fig. 5a). However, E-nose was capable to differentiate CBS from Caucagua region (VEN2) of *Trinitario* cultivar from the other regions of the same cultivar (VEN1, VEN3, VEN4 and VEN6) that was not accomplished with GC-qMS. Figure 5b shows a clear separation of CBS samples of *Criollo* cultivar from the others cultivars using GC-qMS data. Considering GC-qMS data, CBS samples from Ocumare region (VEN6 and VEN10) were separated according to cultivar. Moreover, CBS from Canoabo (VEN5) was also separated from the samples clustered in the *Criollo* varietal group. As observed for E-nose data, this technique allowed the separation of a CBS from the rest of samples of the *Trinitario* varietal, but in this case was CBS from Ocumare region (VEN6). Therefore, the use of GC-qMS coupled to E-nose technique could be and interesting approach for the classification of Venezuela CBS. Volatile molecules such as 2,3,5-trimethylpyrazine and phenylacetaldehyde were identified as potential markers for the classification of CBS from Venezuela as shown in the boxplot represented in Fig. 5c. These and other volatile compounds that had significant differences (FDR<0.05 or lower) among Venezuela regions of production are shown in **Table S.5.1** (see supplementary material).

561 The PCA of CBS from Ecuador using E-nose and GC-qMS data sets are shown in
562 Fig. 5d and Fig. 5e, respectively. As can be observed in Fig. 5d, E-nose technique
563 allowed the classification of CBS according to the cultivar. Samples from *Criollo*
564 cultivar (ECU7) and samples from *Nacional* cultivar (ECU4 and ECU5) were clearly
565 separated from *Forastero* cultivar. However, this technique were not able to separate
566 both *Nacional* cultivar CBSs (ECU4 and ECU5), as well the *Trinitario* cultivar CBS
567 (ECU2) from *Forastero* cultivar samples (ECU1, ECU3 and ECU6), which were
568 separated as for as GC-qMS. PCA based on GC-qMS data of CBS from Ecuador
569 showed a significant separation (FDR<0.001) of the samples ECU7 (*Criollo* varietal)
570 and ECU3 (*Forastero* varietal) among them and from the rest of the CBS samples. It
571 can be also observed a clear separation between CBS samples from *Forastero* cultivar
572 (ECU1, ECU3, ECU6) and CBS yielded from cocoa beans from "fine aroma"
573 (*Nacional*, *Trinitario* and *Criollo* cultivars). This technique allowed a high separation of
574 the CBS respect to E-nose that presented some limitations. Several VOCs were
575 identified as potential markers for this classification (FDR<0.05 or lower) of CBS from
576 Ecuador, such as 2-ethyl-6-methylpyrazine, 2-nonanone as shown in Fig. 5f (see **Table**
577 **S.5.2**, supplementary material).
578
579 *3.5. HS-SPME-GC-qMS vs E-nose – correlation*
580 According to the results described above, GC-qMS and E-nose were capable of
581 classify the CBS at different levels. VOC's found at higher concentrations than 100 μg
582 kg^{-1} at least for one sample were selected for the correlation since this is the limit of
583 detection of E-nose according the manufacturer (Airsense Analytics GmbH., Germany).
584 The correlation between VOCs and sensors are shown in **Figure 6**.

585 The heatmap shows clearly three main clusters of sensors: S1, S3 and S5 (Cluster 1);
586 S2, S7 and S9 (Cluster 2); and S10, S4, S6 and S8 (Cluster 3). For Cluster 1, S1, S3 and
587 S5 were found correlated positively with aromatic molecules (S1, S3 and S5) and
588 aliphatic (S5), such pyrazines (2,3-dmethylpyrazine), alcohols (2-heptanol and 2-
589 nonanol) and ketones (2-heptanone and acetophenone) and acids (2-methylpropanoic
590 acid and 3-methylbutanoic acid). Cluster 2 results from the positive correlation of
591 sensors S2, S7 and S9 with sulphur organic compounds representative of cocoa flavour:
592 dimethyl trisulphide and dimethyl disulphide. Finally, Cluster 3 exhibited a high
593 correlation between sensors related with long chain aliphatics compounds (mainly S4
594 and S10) such as dodecane (hydrocarbons) and octanal.
595

4. Conclusion

This study provides information for the first time on the volatile fingerprint of CBS performed by HS-SPME/GC-qMS and identifies the molecules responsible for differences among an elevated number of samples yielded from cocoa beans collected in different geographical origins and cultivars.

The presence of high amounts of cocoa key-aroma markers in CBS samples, such as 2-methylpropanal, 3-methylbutanal, phenylacetaldehyde, dimethyl trisulfide, 2-phenylethyl acetate, 2,3,5-trimethylpyrazine, 2-ethyl-3,5-dimethylpyrazine, 2-heptanol, 2-phenylethanol, 2-methylpropanoic acid and 3-methylbutanoic acid valorise this by-product as food ingredient. CBSs yielded form cocoa beans of *Criollo* cultivar were those with high amounts of "fine aroma" molecules and therefore those with more interest as source of cocoa flavour.

GC-qMS-fingerprinting and E-nose data proved to be capable of identifying the fine flavour cocoa *Criollo* and classify the CBS according to their cultivar. Several markers, such as 2-methylpropanoic acid, 2,3-dimethylpyrazine and 2-heptanol were found mandatory for the classification of CBS according to the cultivar. It was also possible to classify the CBS samples on the basis of their different geographical origins by using GC–qMS and electronic nose. Markers such as phenylacetaldehyde and furfural were related with the CBS of *Forastero* cultivar from different countries from America and Africa. While for *Trinitario* cultivar, dimethyl trisulphide and 3-methylbutanal, among others, were found those markers capable of classify CBS from American territory.

The results highlighted remarkable diversity in the volatile profile of CBS and confirm the applicability of GC–qMS and E-nose for classification and future traceability of CBS. Moreover, this study correlates for the first time the E-nose and GC-qMS data using a high number of samples and with high significance.

621 Similar to cocoa beans, also for the CBS by-product might be considered the
622 selective collection of this material, with the concept of single origin, to yield a food
623 ingredient with better aroma and specific flavour characteristics that could be recycled
624 inside the cocoa industry in a concept of circular economy as high add-value product.

625 **Authors Contributions**

626 Conceptualization, L.BP, G.Z; Validation, M.G, G.Z; Investigation, O.RP, L.BP, I.F;
627 Writing-original Draft Preparation, L.BP; Review and Editing, O.R.P, I.F, M.G, G.Z;
628 Supervision, M.G, G.Z; Project Administration, L.BP, G.Z.

629 **Funding**

630 The present work has been supported by COVALFOOD "Valorisation of high
631 added-value compounds from cocoa industry by-products as food ingredients and
632 additives". This project has received funding from the European Union's Seventh
633 Framework programme for research and innovation under the Marie Skłodowska-Curie
634 grant agreement No 609402 - 2020 researchers: Train to Move (T2M). L. Barbosa-
635 Pereira gratefully acknowledges the European Union's Seventh Framework programme
636 for the Marie Skłodowska-Curie grant. O. Rojo-Poveda is grateful to the University of
637 Turin for her predoctoral fellowship. We are grateful to Silvio Bessone S.r.l., ICAM
638 S.p.A., Ferrero International S.A., Guido Gobino S.r.l., Pastiglie Leone S.r.l. and
639 Venchi S.p.A. for supplying the samples of cocoa beans.

640 **Conflict of Interest**

641 The authors declare that they have no conflicts of interest.

642 **References**

643 Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2008). Flavor formation and
644 character in cocoa and chocolate: a critical review. *Critical Reviews in Food*
645 *Science and Nutrition*, 48(9), 840-857.
646 Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2009). Matrix effects on flavour
647 volatiles release in dark chocolates varying in particle size distribution and fat
648 content using GC–mass spectrometry and GC–olfactometry. *Food Chemistry*,
649 113(1), 208-215.
650 Aprotosoaie, A. C., Luca, S. V., & Miron, A. (2016). Flavor chemistry of cocoa and
651 cocoa products—an overview. *Comprehensive Reviews in Food Science and Food*
652 *Safety*, 15(1), 73-91.
653 Barbosa-Pereira, L., Dal Bello, B., Belviso, S., Giordano, M., Ferrocino, I., Zeppa, G.,
654 (2017, November). Cocoa bean shell as functional ingredient to enhance the
655 properties and nutritional quality of food products. Poster session presentation at
656 the meeting of the European Federation of Food Science and Technology
657 (EFFoST), Sitges, Barcelona.
658 Barbosa-Pereira, L., Guglielmetti, A., & Zeppa, G. (2018). Pulsed Electric Field
659 Assisted Extraction of Bioactive Compounds from Cocoa Bean Shell and Coffee
660 Silverskin. *Food and Bioprocess Technology*, 11(4), 818-835.
661 Benedetti, S., Buratti, S., Spinardi, A., Mannino, S., & Mignani, I. (2008). Electronic
662 nose as a non-destructive tool to characterise peach cultivars and to monitor their
663 ripening stage during shelf-life. *Postharvest Biology and Technology*, 47(2), 181-
664 188.
665 Bonvehí, J. S. (2005). Investigation of aromatic compounds in roasted cocoa powder.
666 *European Food Research and Technology*, 221(1-2), 19-29.

667 Caprioli, G., Fiorini, D., Maggi, F., Nicoletti, M., Ricciutelli, M., Toniolo, C., Prosper,
668 B., Vittori, S., & Sagratini, G. (2016). Nutritional composition, bioactive
669 compounds and volatile profile of cocoa beans from different regions of
670 Cameroon. *International Journal of Food Sciences and Nutrition*, 67(4), 422-430.
671 Counet, C., Callemien, D., Ouwerx, C., & Collin, S. (2002). Use of gas
672 chromatography− olfactometry to identify key odorant compounds in dark
673 chocolate. Comparison of samples before and after conching. *Journal of*
674 *Agricultural and Food Chemistry*, 50(8), 2385-2391.
675 Crafack, M., Keul, H., Eskildsen, C. E., Petersen, M. A., Saerens, S., Blennow, A.,
676 Skovmand-Larsen, M., Swiegers, J., Petersen, G. B., Heimdal, H., & Nielsen, D.
677 S. (2014). Impact of starter cultures and fermentation techniques on the volatile
678 aroma and sensory profile of chocolate. *Food Research Internation*al, 63, 306-
679 316.
680 Danezis, G. P., Tsagkaris, A. S., Brusic, V., & Georgiou, C. A. (2016). Food
681 authentication: state of the art and prospects. *Current Opinion in Food Science*,
682 10, 22-31.
683 Frauendorfer, F., & Schieberle, P. (2006). Identification of the key aroma compounds in
684 cocoa powder based on molecular sensory correlations. *Journal of agricultural*
685 *and food chemistry*, 54(15), 5521-5529.
686 Frauendorfer, F., & Schieberle, P. (2008). Changes in key aroma compounds of Criollo
687 cocoa beans during roasting. *Journal of Agricultural and Food Chemistry*, 56(21),
688 10244-10251.
689 Gu, F., Tan, L., Wu, H., Fang, Y., Xu, F., Chu, Z., & Wang, Q. (2013). Comparison of
690 cocoa beans from china, indonesia and papua new guinea. *Foods*, 2(2), 183-197.

691 International Cocoa Organization. Statistics – Production. Quarterly Bulletin of Cocoa
692 Statistics, Vol. XLIV No. 3. (2018). https://www.icco.org/about-us/international-
693 cocoa-agreements/cat_view/30-related-documents/46-statistics-production.html.
694 Accessed 31 November 2018.
695 Kongor, J. E., Hinneh, M., Van de Walle, D., Afoakwa, E. O., Boeckx, P., &
696 Dewettinck, K. (2016). Factors influencing quality variation in cocoa (Theobroma
697 cacao) bean flavour profile—A review. *Food Research International*, 82, 44-52.
698 Magagna, F., Guglielmetti, A., Liberto, E., Reichenbach, S. E., Allegrucci, E., Gobino,
699 G., Bichi, C. & Cordero, C. (2017). Comprehensive Chemical Fingerprinting of
700 High-Quality Cocoa at Early Stages of Processing: Effectiveness of Combined
701 Untargeted and Targeted Approaches for Classification and Discrimination.
702 *Journal of Agricultural and Food Chemistry*, 65(30), 6329-6341.
703 Mandrile, L., Barbosa-Pereira, L., Sorensen, K. M., Giovannozzi, A. M., Zeppa, G.,
704 Engelsen, S. B. & Rossi, A. M. (2019). Authentication of cocoa bean shells by
705 near-infrared and mid-infrared spectroscopy and inductive coupled plasma-optical
706 emission spectroscopy. *Food Chemistry*, *In press*.
707 Nazaruddin, R., Seng, L. K., Hassan, O., & Said, M. (2006). Effect of pulp
708 preconditioning on the content of polyphenols in cocoa beans (Theobroma cacao)
709 during fermentation. *Industrial Crops and Products*, 24(1), 87-94.
710 Nsor-Atindana, J., Zhong, F., & Mothibe, K. J. (2012). In vitro hypoglycemic and
711 cholesterol lowering effects of dietary fiber prepared from cocoa (Theobroma
712 cacao L.) shells. *Food & Function*, 3(10), 1044-1050.
713 Oliveira, L. F., Braga, S. C., Augusto, F., Hashimoto, J. C., Efraim, P., & Poppi, R. J.
714 (2016). Differentiation of cocoa nibs from distinct origins using comprehensive

two-dimensional gas chromatography and multivariate analysis. *Food Research International*, 90, 133-138.

Olunloyo, V. O., Ibidapo, T. A., & Dinrifo, R. R. (2012). Neural network-based electronic nose for cocoa beans quality assessment. *Agricultural Engineering International: CIGR Journal*, 13(4).

Owusu, M., Petersen, M. A., & Heimdal, H. (2012). Effect of fermentation method, roasting and conching conditions on the aroma volatiles of dark chocolate. *Journal of Food Processing and Preservation*, 36(5), 446-456.

Qin, X. W., Lai, J. X., Tan, L. H., Hao, C. Y., Li, F. P., He, S. Z., & Song, Y. H. (2017). Characterization of volatile compounds in Criollo, Forastero, and Trinitario cocoa seeds (Theobroma cacao L.) in China. *International Journal of Food Properties*, 20(10), 2261-2275.

Rodriguez-Campos, J., Escalona-Buendía, H. B., Contreras-Ramos, S. M., Orozco-Avila, I., Jaramillo-Flores, E., & Lugo-Cervantes, E. (2012). Effect of fermentation time and drying temperature on volatile compounds in cocoa. Food Chemistry, 132(1), 277-288.

Saltini, R., Akkerman, R., & Frosch, S. (2013). Optimizing chocolate production through traceability: A review of the influence of farming practices on cocoa bean quality. *Food Control*, 29(1), 167-187.

Tran, P. D., Van de Walle, D., De Clercq, N., De Winne, A., Kadow, D., Lieberei, R., Messens, K., Tran, D. N., Dewettinck, K., & Van Durme, J. (2015). Assessing cocoa aroma quality by multiple analytical approaches. *Food Research International*, 77, 657-669.

Wang, J. (2015). *Use of cocoa shells for preparing chocolate flavor by preparing enzymatic hydrolyzate of cocoa shells, adding cocoa shells enzyme solution with*

740 *amino acids, reducing sugar, solvent, reacting, cooling and filtering*. Patent
741 CN104187539-A, China.
742

743 **Figure captions**

744 **Figure 1.** PCA based on the VOC's (μg kg^{-1}) identified by HS-SPME/GC-qMS in all
745 CBS samples as function of: (a) cultivar, (b) continent of origin. The variance explained
746 by the first component of PCA (PC1) was 28.54 %, while the second component (PC2)
747 explained 15.10 %. Box plots showing abundance of key VOC's that can be used as
748 possible markers for (c) cultivar and (d) geographical origin. For interpretation of the
749 legends, see **Table 1**.
750
751 **Figure 2.** PCA based on the VOC's (μg kg^{-1}) identified by HS-SPME/GC-qMS in CBS
752 samples of *Forastero* cultivar according to: (a) macroarea, (b) latitude, and (c) country
753 of origin. The variance explained by the first component of PCA was 32.02 %, while
754 the second component explained 13.71 %. (d) Boxplot showing abundance of VOC's
755 that can be used as potential markers of origin: phenylacetaldehyde and furfural. For
756 interpretation of the legends, see **Table 1**.
757
758 **Figure 3.** PCA based on the VOC's (μg kg^{-1}) identified by HS-SPME/qGC-MS in CBS
759 samples of *Trinitario* cultivar according to: (a) macroarea, (b) latitude and (c) country
760 of origin. The variance explained by the first and second principal component was
761 35.63% and 15.26%, respectively. (d) Boxplot showing abundance of VOC's that can
762 be used as potential markers of origin: dimethyl trisulphide and 2,3,5-trimethylpyrazine.
763 For interpretation of the legends, see **Table 1**.
764
765 **Figure 4.** PCA based on E-nose data set for: (a) all CBS samples as function of cultivar
766 (the variance explained by the first and second principal component was 78.94% and
767 11.06% respectively); (b) CBS samples of *Forastero* according to country of origin (the

768 variance explained by the first and second principal component was 74.89% and
769 13.51%, respectively); (c) CBS samples of *Trinitario* cultivar according to country of
770 origin (the variance explained by the first and second principal component was 68.30%
771 and 13.97%, respectively). For interpretation of the legends, see **Table 1**.
772
773 **Figure 5.** PCA based on the VOC's (μg kg^{-1}) identified by HS-SPME/GC-qMS and E-
774 nose data set for CBS samples from different regions of Venezuela: (a) PCA based on
775 E-nose data set (the variance explained by the first and second principal component was
776 93.63% and 3.04%, respectively); (b) PCA based on HS-SPME/GC-qMS data set (the
777 variance explained by the first and second principal component was 32.88% and
778 22.41%, respectively). (c) Boxplot showing abundance of VOC's that can be used as
779 potential markers of origin: 2,3,5-trimethylpyrazine and phenylacetaldehyde.
780 PCA based on the VOC's (μg kg^{-1}) identified by HS-SPME/GC-qMS and E-nose data
781 set for CBS samples from different regions of Ecuador: (d) PCA based on E-nose data
782 set (the variance explained by the first and second principal component was 82.23% and
783 10.76%, respectively); (e) PCA based on HS-SPME/GC-qMS data set (the variance
784 explained by the first and second principal component was 27.76% and 24.13%,
785 respectively). (f) Boxplot showing abundance of VOC's that can be used as potential
786 markers of origin; 2-ethyl-6-methylpyrazine and 2-nonanone.
787 For interpretation of the legends, see **Table 1**.
788
789 **Figure 6.** Correlation between the abundance of VOCs (μg kg^{-1}) and E-nose sensors.
790 Rows and columns are clustered by Ward linkage hierarchical clustering. The intensity
791 of the colours represents the degree of correlation between the samples and VOCs as

792 measured by the Spearman's correlations. Asterisks denote significant correlations after
793 P value corrections (FDR < 0.05).