Food Research International 123 (2019) 684–696

Contents lists available at ScienceDirect

Food Research International

journal homepage: www.elsevier.com/locate/foodres

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

Letricia Barbosa-Pereira[a,*], Olga Rojo-Poveda[a,b], Ilario Ferrocino[a], Manuela Giordano[a], Giuseppe Zeppa[a]

[a] Department of Agriculture, Forestry and Food Sciences (DISAFA), University of Turin, Grugliasco, Italy
[b] RD3 Department-Unit of Pharmacognosy, Bioanalysis and Drug Discovery, Faculty of Pharmacy, Université Libre de Bruxelles, Brussels, Belgium

ARTICLE INFO

Keywords:
Cocoa bean shell
Cocoa by-product
HS-SMPE/GC-qMS
E-nose
Principal component analysis
Volatile fingerprint
Chemical markers
Cocoa flavour

ABSTRACT

The cocoa bean shell (CBS) is a main by-product of cocoa processing, with great potential to be used as an ingredient for functional foods because of its nutritional and flavour properties. This study aimed to characterise and classify CBSs obtained from cocoa beans of diverse cultivars and collected in different geographical origins through their volatile profile assessed using headspace solid-phase microextraction gas chromatography–mass spectrometry (HS-SPME/GC-qMS) and *E*-nose combined with principal component analysis (PCA). The study provides, for the first time in a representative set of samples, a comprehensive fingerprint and semi-quantitative data for > 100 volatile organic compounds (VOCs), such as aldehydes, ketones, pyrazines, alcohols, and acids. Through PCA, a clear separation of the *Criollo* cultivar from the other cultivars was achieved with both GC-qMS and *E*-nose analytical techniques because of 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 beans, were found suitable for the classification of CBSs according to their quality and origin. GC-qMS and *E*-nose appeared to be suitable analytical approaches to classify CBSs, with a high correlation between both analytical techniques. The volatile fingerprint and classification of CBSs could allow for 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.

1. Introduction

The cocoa bean (*Theobroma cacao* L.) is a ubiquitous edible product, consumed across the world, of great economic significance, and the key raw material in 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 2016–2017 season, and the major producers were 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 into 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 the *Forastero* cultivar, which has ordinary flavour properties and makes up 95% of the world's total cocoa production. Fine grade cocoa is exclusively produced from *Criollo*, *Trinitario*, and *Nacional* cultivars, the latter is grown in Ecuador and is characterized by its remarkable flavour properties, due to the fruity and floral aroma attributes (Aprotosoaie et al., 2016; Saltini, Akkerman, & Frosch, 2013). 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; Tran et al., 2015). Indeed, the quality and flavour of cocoa are not simply affected by genotype and geographical origin but also by other factors, such as growth conditions, post-harvest treatments, and industrial processing of beans (Kongor et al., 2016). In particular, fermentation and roasting are the key steps responsible for the characteristics and desirable organoleptic properties of cocoa, such as

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; DVB/CAR/PDMS, divinylbenzene/carboxen/polydimethylsiloxane; ISTD, internal standard; PCA, principal component analysis.

* Corresponding author.
E-mail address: letricia.barbosapereira@unito.it (L. Barbosa-Pereira).

https://doi.org/10.1016/j.foodres.2019.05.041
Received 22 February 2019; Received in revised form 8 May 2019; Accepted 29 May 2019
Available online 31 May 2019

Table 1
Origin of fermented and dried cocoa beans used to obtain the cocoa bean shells.

Sample code	Cultivar	Country	Region	Continent	Macroarea
BRA	*Trinitario*	Brazil	Cabruca	America	South America
CAM1	*Forastero*	Cameroon	N.D.	Africa	West Africa
CAM2	*Trinitario*	Cameroon	N.D.	Africa	West Africa
COL1	*Forastero*	Colombia	N.D.	America	South America
COL2	*Trinitario*	Colombia	Santander	America	South America
CON1	*Forastero*	Congo	N.D.	Africa	West Africa
CON2	*Forastero*	Congo	N.D.	Africa	West Africa
DOR1	*Trinitario*	Dominican Republic	N.D.	America	Central America
DOR2	*Forastero*	Dominican Republic	N.D.	America	Central America
DOR3	*Trinitario*	Dominican Republic	N.D.	America	Central America
DOR4	*Trinitario*	Dominican Republic	N.D.	America	Central America
ECU1	*Forastero*	Ecuador	Arriba	America	South America
ECU2	*Trinitario*	Ecuador	N.D.	America	South America
ECU3	*Forastero*	Ecuador	N.D.	America	South America
ECU4	*Nacional*	Ecuador	N.D.	America	South America
ECU5	*Nacional*	Ecuador	N.D.	America	South America
ECU6	*Forastero*	Ecuador	Palo Santo	America	South America
ECU7	*Criollo*	Ecuador	N.D.	America	South America
GHA	*Forastero*	Ghana	N.D.	Africa	West Africa
IVC	*Forastero*	Ivory Coast	N.D.	Africa	West Africa
JAM	*Trinitario*	Jamaica	N.D.	America	Central America
MAD	*Forastero*	Madagascar	N.D.	Africa	East Africa
MEX	*Trinitario*	Mexico	Chontalpa	America	Central America
PER1	*Forastero*	Peru	N.D.	America	South America
PER2	*Trinitario*	Peru	N.D.	America	South America
SAT1	*Forastero*	São Tomé	N.D.	Africa	West Africa
SAT2	*Forastero*	São Tomé	N.D.	Africa	West Africa
SAT3	*Forastero*	São Tomé	N.D.	Africa	West Africa
SLE	*Forastero*	Sierra Leone	N.D.	Africa	West Africa
TAN	*Forastero*	Tanzania	N.D.	Africa	East Africa
TOG1	*Forastero*	Togo	N.D.	Africa	West Africa
TOG2	*Forastero*	Togo	N.D.	Africa	West Africa
UGA1	*Forastero*	Uganda	N.D.	Africa	East Africa
UGA2	*Forastero*	Uganda	N.D.	Africa	East Africa
VEN1	*Trinitario*	Venezuela	Sur del Lago	America	South America
VEN2	*Trinitario*	Venezuela	Caucagua	America	South America
VEN3	*Trinitario*	Venezuela	Merida_1	America	South America
VEN4	*Trinitario*	Venezuela	Cuyagua	America	South America
VEN5	*Criollo*	Venezuela	Canoabo	America	South America
VEN6	*Trinitario*	Venezuela	Ocumare_1	America	South America
VEN7	*Criollo*	Venezuela	Merida_2	America	South America
VEN8	*Criollo*	Venezuela	Merida_3	America	South America
VEN9	*Criollo*	Venezuela	Carenero	America	South America
VEN10	*Criollo*	Venezuela	Ocumare_2	America	South America

N.D. Not determined.

aroma and flavour, that are important quality attributes for consumer 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 certify the quality and enable the discrimination of the cocoa products with label of origin to ensure food authentication, a new market trend of great interest for law enforcement, food producers, importers and exporters, and consumers (Danezis, Tsagkaris, Brusic, & Georgiou, 2016; Magagna et al., 2017).

The solid-phase microextraction (SPME) coupled with gas chromatography (GC) mass spectrometry (MS) methodology has been widely used to identify and quantify 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 (Caprioli et al., 2016; Magagna et al., 2017; Oliveira et al., 2016; Tran et al., 2015). However, a limited number of studies have explored the potential applicability of electronic nose technologies to assess 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. The cocoa bean shell (CBS) is one of the main by-products generated after the roasting and husking of cocoa beans (about 12% of the total weight), and consequently > 500 thousand tonnes are produced every year that represent a disposal problem for the cocoa sector. However, recent studies have established that the CBS might also be an economic source of fibre, minerals, polyphenols, and methylxanthines with potential health benefits and, therefore, with great potential to be used as an ingredient for functional foods, creating new food market perspectives (Barbosa-Pereira et al., 2017; Barbosa-Pereira, Guglielmetti, & Zeppa, 2018; Mandrile et al., 2019; Nsor-Atindana, Zhong, & Mothibe, 2012). Wang (2015) patented a process for chocolate flavour production, with a real chocolate aroma, from dried CBSs using an enzymatic technology. Nevertheless, despite this product exhibiting great potential as a food ingredient, to the best of our knowledge, no information is available in the literature describing the volatile composition of CBSs. As for cocoa beans, the study of the CBS volatile fingerprint is very important to define the quality and flavour of the product. Moreover, selecting the CBS with a specific flavour profile according to the food application could be an approach to valorise this by-product as a food ingredient. Hence, the aim of this study was to describe, for the first time, the

Table 2
Volatile compounds (VOCs) identified in cocoa bean shell (CBS) powders by HS-SPME/GC-qMS with their experimental retention index (RI_{exp}), odour description and concentration range (μg kg^{-1}).

Peak no.	VOC[a]	RI_{Exp}	Odour description[b]	Concentration range (μg kg^{-1})		
Aldehydes						
1	2-Methylpropanal	< 1000	Cocoa, chocolate, malty, roasted	55.87	–	470.62
2	3-Methylbutanal	< 1000	Cocoa, chocolate, malty, roasted	100.97	–	891.32
3	Hexanal	1085	Green, fermented	26.26	–	321.47
4	Heptanal	1193	Green, Oily, fatty	1.64	–	174.35
5	Octanal	1298	Orange peel, oily, fatty, soapy	6.86	–	160.77
6	2-Isopropyl-5-methyl-hex-2-enal (isomer 1)	1369	Cocoa	11.99	–	357.04
7	2-Isopropyl-5-methyl-hex-2-enal (isomer 2)	1369	Cocoa	4.40	–	65.16
8	Nonanal	1405	Tallowy, soapy-fruity, fatty, waxy, pungent	40.31	–	1420.63
9	(*E*)-2-Octenal	1440	Fatty, waxy	4.13	–	29.61
10	Decanal	1511	Sweet, orange, waxy	1.49	–	56.53
11	Benzaldehyde	1536	Bitter, almond-like, burnt sugar, grass, earthy	243.98	–	1318.30
12	(*E*)-2-Nonenal	1548	Green	2.77	–	21.83
13	Phenylacetaldehyde	1658	Flowery, honey, sweet, rose, green, berry, nutty	260.13	–	2331.19
14	2-Phenyl-2-butenal	1898	Flowery, sweet, cocoa, roasted, rum	5.46	–	83.31
15	5-Methyl-2-phenyl-2-hexenal	2078	Sweet, roasted cocoa	0.04	–	113.09
Ketones						
16	2,3-Butanedione	< 1000	Buttery, sweet	97.77	–	556.04
17	2-Heptanone	1189	Fruity, banana-like, green	29.31	–	362.87
18	6-Methyl-5-heptene-2-one	1348	Pungent, green	12.91	–	111.32
19	2-Nonanone	1399	Flowery, fatty	30.42	–	293.89
20	2-Decanone*	1505	Floral	8.22	–	109.41
21	3-Methyl-2-cyclohexen-1-one*	1605	Sweet, nutty	0.00	–	84.86
22	2-Undecanone*	1611	Waxy, fruity	3.07	–	67.13
23	Acetophenone	1665	Floral, sweet, almond, must, pungent	22.28	–	169.75
24	Phenylacetone*	1739	Almond	0.64	–	9.90
Sulfur compounds						
25	Dimethyl disulfide	1074	Rubbery, onion, meaty, sulphurous	5.52	–	291.23
26	Dimethyl trisulfide	1389	Onion, cabbage, sweaty, sulphurous	3.86	–	284.33
27	3-Phenylthiophene*	2126	–	0.12	–	10.96
28	2,3-Dihydrothiophene*	2333	–	0.35	–	11.82
Esters						
29	3-Methylbutyl acetate	1131	Banana, pear, sweet, fruity	6.57	–	449.32
30	Ethyl octanoate	1447	Fruity, floral	1.51	–	35.33
31	Methyl 2-phenylacetate	1774	Honey, sweet, jasmine	9.43	–	124.49
32	Ethylbenzene acetate	1800	Floral, honey, rose, fruity, sweet, green, cucumber	2.21	–	63.97
33	2-Phenylethyl acetate	1829	Honey, sweet, flowery	1.69	–	506.70
34	Phenylethyl butyrate*	1877	Musty, sweet, floral	0.00	–	7.15
35	2-Methylpropyl benzoate	2045	Fruity, balsamic	0.25	–	6.45
36	Methyl hexadecanoate	2210	Fruity	0.96	–	11.41
Hydrocarbons						
37	Dodecane	1203	–	7.29	–	1105.30
38	3-Phenylpentane*	1710	–	0.54	–	26.72
Furan derivates						
39	2-*n*-Pentylfuran	1241	Musty, green bean	21.69	–	409.93
40	Furfural	1478	Almond, caramel, sweet, woody, flowery, green	18.97	–	3911.55
41	Acetylfuran	1518	Sweet, balsamic, cocoa, green, grassy	4.71	–	191.66
Pyrazines						
42	2-Methylpyrazine	1273	Cocoa, chocolate, roasted nuts, green, hazelnut	11.49	–	200.16
43	2,6-Dimethylpyrazine	1334	Nutty, coffee, green, roasted, earthy	12.25	–	107.00
44	2-Ethylpyrazine	1341	Peanut-butter, musty, nutty, rum, earthy	7.77	–	57.30
45	2,3-Dimethylpyrazine	1352	Caramel, cocoa, sweet, baked, hazelnut, roasted, earthy	11.98	–	310.35
46	2-Ethyl-6-methylpyrazine	1393	Potato, earthy, halznut, cocoa, roasted, green	6.91	–	891.63
47	2,3,5-Trimethylpyrazine	1411	Cocoa, roasted nuts, peanut, earthy, green	41.66	–	891.63
48	2-Ethyl-3,5-dimethylpyrazine	1443	Earthy	1.66	–	106.10
49	2,5-Dimethyl-3-ethylpyrazine	1453	Potato, roast, earthy	3.84	–	309.79
50	2,3-Dimethyl-5-ethylpyrazine	1470	Cocoa, chocolate, praline	8.51	–	316.88
51	2,3,5,6-Tetramethylpyrazine	1482	Chocolate, cocoa, roasted, green, earthy	69.38	–	3298.06
52	2-Isobutyl-3-methylpyrazine	1497	Caramellic	0.68	–	32.66
53	2,3,5-Trimethyl-6-ethylpyrazine	1522	Candy, sweet, chocolate, cocoa, hazelnut, roasted	2.59	–	383.29
54	3,5-Dimethyl-2-isobutylpyrazine*	1543	Cocoa, hazelnut, musty, earthy, roasted, nutty	0.26	–	12.26
55	3-Methylbutylpyrazine*	1592	–	0.47	–	13.27
56	2-Isobutyl-3,5,6-trimethylpyrazine*	1602	–	0.40	–	43.65
57	2-Isoamyl-6-methylpyrazine*	1614	–	1.62	–	94.77
58	2,5-Dimethyl-3-isopentylpyrazine*	1669	–	0.65	–	321.09
59	2,3,5-Trimethyl-6-(2-methylbutyl)-pyrazine*	1682	–	0.00	–	24.12
60	2,5-Dimethyl-3-(1-methylpropyl)-pyrazine*	1710	–	0.00	–	74.33
61	2,3,5-Trimethyl-6-isopentylpyrazine*	1714	Green, floral	0.39	–	109.62
62	2-Acetyl-3,5-dimethylpyrazine*	1777	Nutty, roasted hazelnut	0.18	–	4.90
Alcohols						

(*continued on next page*)

Table 2 (*continued*)

Peak no.	VOC[a]	RI_{Exp}	Odour description[b]	Concentration range ($\mu g\,kg^{-1}$)		
63	2-Heptanol	1328	Citrus, earthy, oily	19.36	–	1655.07
64	1-Hexanol	1362	Fruity, green, herbaceous	2.58	–	98.75
65	2-Nonanol	1529	Waxy, green, creamy, citrus orange	6.79	–	137.23
66	1-Octanol	1568	Sharp, fatty, waxy, citrus, moss, nut, mushroom	11.90	–	141.35
67	1-Phenylethanol	1817	Honey, floral, rose, fragant	1.11	–	25.62
68	Benzyl Alcohol	1871	Sweet, fruity, floral	2.84	–	43.57
69	2-Phenylethanol	1892	Flowery, honey-like, rose, lilac, caramel, sweet	52.01	–	936.97
Pyrroles						
70	1-Pentyl-1H-pyrrole*	1381	Nutty	0.00	–	2.85
71	2-Acetylpyrrole	1968	Hazelnut, cocoa, chocolate, roasted, popcorn, musty, bread	14.98	–	223.47
72	1H-Pyrrole-2-carboxaldehyde	2025	Nutty, honey, candy	12.11	–	437.06
73	1-Methyl- 1H-pyrrole-2-carboxaldehyde*	2102	Nutty	1.43	–	12.31
Terpenes/Isoprenoids						
74	Linalool oxide (*cis*-Furanoid)	1451	Sweet, nutty, fruity, floral/flowery	3.12	–	256.77
75	Linalool	1559	Rose, flowery, floral, citrus	4.79	–	132.09
76	α-Terpineol	1705	Herbaceous, Fruity	1.19	–	21.15
77	ß-Cadinene*	1757	Green, woody	0.00	–	7.84
78	ß-Damascenone*	1833	Floral	0.00	–	2.55
79	α-Ionone*	1848	Floral	0.47	–	4.48
80	Geranylacetone*	1858	Rose, floral	7.28	–	88.69
81	ß-Ionone*	1928	Woody, sweat, fruity	1.63	–	10.92
82	*trans*-Nerolidol*	2056	Floral, green and citrus like	0.00	–	3.10
83	*cis*-Methyl dihydrojasmonate*	2305	Sweet, fruity, floral, citrus lemon	0.33	–	1.86
Acids						
84	Acetic acid	1463	Sour, vinegar-like, pungent	21.76	–	1612.24
85	2-Methylpropanoic acid	1580	Acidic, sour, cheese, dairy, buttery, rancid, hammy	2.07	–	454.27
86	3-Methylbutanoic acid	1674	Old cheese, acidic, rancid, hammy	36.27	–	2154.25
87	3-Methylpentanoic acid*	1803	Sour, cheesy, fresh, fruity	0.70	–	26.07
88	4-Methylpentanoic acid	1809	Cheesy	2.14	–	45.19
89	2-Butylbutanoic acid	1956	–	1.43	–	33.13
90	Heptanoic acid	1964	Rancid, sour, sweaty	5.21	–	91.06
91	Octanoic Acid	2059	Sweat, cheese, oily, fatty	9.05	–	403.22
92	Nonanoic acid	2160	Green, fat, sweat, waxy	0.23	–	130.83
93	2-Phenylacetic acid	2487	Sweet, floral, chocolate, honey, tobacco	0.49	–	43.55
Lactones						
94	(*S*)- Massoialactone*	2009	Coconut, nutty	0.49	–	258.64
95	(*R*)- Massoialactone*	2242	Coconut, nutty	0.93	–	364.54
96	4,5,7,7a-Tetrahydro-4,4,7a-trimethyl-2(6H) benzofuranone	2338	Nutty	0.78	–	6.19
Others						
97	Styrene*	1266	Balsamic, sweet, floral	5.70	–	42.09
98	2-Acetylpyridine	1616	Caramel-like, sweet, Popcorn	0.80	–	3.21
99	*o*-Guaiacol	1867	Woody, spicy, smoky, burnt, musty	1.87	–	37.08
100	*p*-Cresol	2096	Smoke	0.76	–	93.71
101	3-Methyl-1H-Indole	2477	Mothball, faecal	0.36	–	3.94

[a] Volatile organic compounds (VOC) identified by probability based matching of mass spectra available in commercial libraries (NIST, Wiley) and the retention index (RI). The key-aroma markers are shown highlighted in bold according to Afoakwa et al. (2008), Frauendorfer and Schieberle (2006), and Frauendorfer and Schieberle (2008). *Aromatic compounds described for the first time for cocoa and its related products.

[b] Odour descriptors according to literature (Afoakwa et al. (2009), Bonvehí (2005), Counet, Callemien, Ouwerx, and Collin (2002), Crafack et al. (2014), Frauendorfer and Schieberle (2008), Magagna et al. (2017), Magagna et al. (2017), Owusu et al. (2012), Rodriguez-Campos et al. (2012), Tran et al. (2015)) and from online databases: Flavornet (http://www.flavornet.org, accessed November 2018) and TGSC (The Good Scents Company) (http://www.thegoodscentscompany.com, accessed June 2018).

volatile fingerprint of the CBS by HS-SPME/GC-qMS and determine the volatile compounds responsible for differences among several CBSs yielded from cocoa beans collected in different geographical origins and cultivars to allow for the traceability of this material. Moreover, we also explored the applicability of *E*-nose as a rapid methodology for the classification of the CBS and evaluated the correlation between *E*-nose and GC-qMS data sets.

2. Materials and methods

2.1. Chemicals and standards

Methanol (≥99.9%), sodium chloride (≥99%), sodium hydroxide standard solution (0.1001 mol L^{-1}), and n-alkanes (n-C7–n-C30) mix standard (Supelco, Italy) for retention index determination were obtained from Sigma-Aldrich (Milano, Italy). Ultrapure water was prepared in a Milli-Q filter system (Millipore, Milan, Italy).

The internal standard (ISTD), 5-nonanol (≥95% GC), for analyte response normalisation was provided by Sigma-Aldrich (Milano, Italy). A standard stock solution of 5-nonanol was prepared in ultrapure water at a 10 mg L^{-1} concentration for the semi-quantification and stored in a sealed vial at −20 °C.

2.2. CBS samples

Fermented and dried cocoa beans (*Theobroma cacao* L.) from different cultivars and countries across the world, harvested during the seasons of 2014 and 2015, were purchased from several local cocoa companies. In total, 44 samples (2 batches each) from different geographical areas in 19 countries and four cultivars (*Criollo* (*n* = 6),

Trinitario ($n = 15$), *Nacional* ($n = 2$), and *Forastero* ($n = 21$)) were collected as described in Table 1. Specific information related to fermentation and drying conditions was not available, since the suppliers retained this confidential information. To obtain the CBSs, all samples were roasted individually using a standardized process performed in a laboratory at 130 °C (isothermal) for 20 min using a ventilated oven Memmert UFE 550 (ENCO, Spinea, Italy). Then, CBS samples were separated from the beans and ground into a powder with a 250 µm mesh size using an ultra-centrifugal mill Retsch ZM 200 (Retsch Gmbh, Haan, Germany). Samples were stored under a vacuum at −20 °C before sample preparation and headspace analysis. The humidity content of the CBS samples ranged between 5.46% and 9.22% as reported by Barbosa-Pereira el al., *submitted*.

2.3. HS-SPME/GC-qMS analysis

The VOCs from the CBS samples were identified and analysed using a headspace solid phase micro extraction (HS-SPME) coupled with gas chromatography/quadrupole mass spectrometry (GC-qMS).

2.3.1. HS-SPME/GC-qMS conditions

For the extraction of VOCs, 0.1 g of CBS powder was accurately weighed in a 20 mL headspace vial. Then, 2 mL of sodium chloride (40% *w/v*) and 10 µL of internal standard (IS) 5-nonanol (10 µg/mL) were added to the sample, and the vial was immediately hermetically capped with a PTFE-silicon septum. The extraction was performed in a COMBI PAL System Autosampler for SPME (CTC Analytics AG, Zwingen, Switzerland) equipped with an HS-SPME unit. The sample was equilibrated at 60 °C with stirring at 250 rpm for 10 min to reach equilibrium. Next, a well-conditioned SPME fibre coated with divinylbenzene/carboxen/ polydimethylsiloxane (DVB/CAR/PDMS) (d_f 50/30 µm, 1 cm) (Supelco, Bellefonte, PA, USA) was exposed to the headspace of the sample for another 30 min with continuous heating and agitation. After extraction, the fibre was inserted into the injection port of the GC system in splitless mode and desorbed at 260 °C for 2 min. Three identical samples were prepared for each analysis.

GC-qMS analyses were performed on a Shimadzu GC-2010 gas chromatograph equipped with a Shimadzu QP-2010 Plus quadrupole mass spectrometer (Shimadzu Corporation, Kyoto, Japan). Separation of VOCs was performed on a DB-WAXETR capillary column of 30 m length, 0.25 mm internal diameter, and 0.25 µm film thickness (J&W Scientific Inc., Folsom, CA, USA). The oven time-temperature programme was as follows: initial temperature 40 °C held for 5 min, from 40 °C 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 temperature of the ion source and quadrupole was 240 °C. The data were recorded in full-scan mode in the mass acquisition range of 30–450 *m/z* with 0.30 s scan time.

2.3.2. 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. The compounds in trace whose similarity was > 75% were considered (tentative identification) even if higher similarity was achieved for the identified compounds. Additionally, the confirmation of molecule identity was performed by comparing the gas chromatographic retention indices (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

E-nose analyses were performed using a portable electronic nose system PEN3 (Airsense Analytics GmbH., Germany). The system consists of a sampling unit and the gas detection system composed of 10 Metal Oxide Semiconductor (MOS) sensors, which are differentially sensitive to each characteristic volatile compound. The chemical sensors that composed the sensor array system are the following: S1, aromatic; S2, broad range; S3, aromatic; S4, hydrogen; S5, aromatic and aliphatics; S6, broad range and methane; S7, sulfur organic; S8, broad-range alcohol; S9, sulfur and chlorinate; and S10, methane and aliphatics (Benedetti, Buratti, Spinardi, Mannino, & Mignani, 2008).

For the analysis, 2 g of CBS powder was placed in a 20-mL glass vial and capped with a PTFE septum. Then, each vial was incubated at 30 °C for 30 min to reach the headspace equilibrium. The gas headspace was injected into the *E*-nose carried by air for 90 s at a constant flow rate of 400 mL min^{-1}, and during this time, the sensor signals were recorded at each second. After each analysis, the sensor system was purged with filtered air for 120 s, to allow reestablishment of the instrument baseline prior to the next sample injection. The sensor response, G/G0 (G and G0 stand for the conductance of the MOS connected with the sample and clean gas, respectively), is expressed as resistivity (Ohm) and changed accordingly to the composition of volatile compounds. Data were collected by the pattern recognition software (WinMuster, v.1.6., Airsense Analytics GmbH., Germany). Three replicates of each CBS sample were independently analysed, and the average of sensor responses (area under the curve) was used for the subsequent statistical analysis.

2.5. Chemometric analysis

A total of 44 CBS samples obtained from the continents Africa and America were used to perform chemometric analysis. From these 44 samples, two batches of cocoa beans provided from the same producer were available (see Table 1). All samples were analysed in triplicate with 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 cultivar principal component analysis (PCA) based on the normalized data (log10) was performed by using the *made4* package of R (https://www.r-project.org) and the function *dudi.pca*. Analysis of similarity based on the VOCs and *E*-nose table was applied with 999 permutations to detect significant differences as a function of the continent, macro-area, latitude, country of production, or cultivar, by using the anosim function in *vegan* package of R. Non-parametric Kruskal-Wallis as well as Wilcoxon tests were carried out in order to find VOCs differentially abundant between all the variables. Data were visualized as box plots, which represented the interquartile range between the first and the third quartiles, with the error bars showing the lowest and the highest values. 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 the CBS is very important to define the quality and flavour of the product to be used as a

food ingredient. The present study was divided into two main parts: the first one was dedicated to the analyses of all samples using GC-qMS and *E*-nose to define the volatile profile and fingerprint of CBS, followed by the classification of samples, using PCA analysis, and the identification of key compounds that differentiate the samples classes.

3.1. Volatile profile of the 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, ketones, sulfur compounds, esters, hydrocarbons, furans, pyrazines, alcohols, pyrroles, terpenes, isoprenoids and terpene alcohols, acids, lactones, and others were semi-quantified as $\mu g\ kg^{-1}$ of 5-nonanol equivalents. The average of the amounts of each VOC, the sum of each class of compound, and the total amount of VOCs presented in a single sample are shown in detail in Table 3 described by Barbosa-Pereira et al., *submitted*. The total amount of VOCs ranged between $4.92\ \mu g\ g^{-1}$ (VEN3) and $16.10\ \mu g\ g^{-1}$ (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 $142.5\ \mu g\ g^{-1}$). In general, the most representative classes of compounds in the 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 roasting has a great impact on cocoa aroma, and the alkyl pyrazines and Strecker aldehydes increased significantly at this stage in cocoa and consequently in the CBS, which is a main by-product produced during this stage. This distribution is slightly different from that found in literature for roasted cocoa beans, which presented acids and alcohols as the main compounds at high concentrations, or for roasted cocoa liquor, which displayed higher amounts of aldehydes, alcohols, and ketones (Caprioli et al., 2016; Crafack et al., 2014; Tran et al., 2015). However, the amounts of the several classes of compounds change with the cultivar and geographical origin of the cocoa bean (Bonvehí, 2005; Tran et al., 2015). In general, the CBSs from *Trinitario*, *Criollo*, and *Nacional* cocoa cultivars are those with the higher amounts of VOCs compared with those of the *Forastero* group. *Criollo* and *Nacional* cultivars display, on average, higher amounts of pyrazines, acids, alcohols, and ketones than the *Trinitario* and *Forastero* cultivars. This data is in accordance with that found in literature for cocoa beans (Qin et al., 2017).

Since no data are available in the literature, the results of the present work will be discussed by comparing with studies performed for roasted cocoa beans and cocoa products, such as dark chocolate and cocoa powder, described in numerous reports (Tran et al., 2015, Bonvehí, 2005; Menezes et al., 2016; Owusu, Petersen, & Heimdal, 2012; Afoakwa, Paterson, Fowler, & Ryan, 2009).

Aldehydes were the most representative aroma compounds in the CBS with total amounts up to $5123.89\ \mu g\ kg^{-1}$ quantified in samples yielded from cocoa beans from Dominican Republic (DOR1), with similar or higher amounts than that found in roasted cocoa beans ($1.22–3.84\ \mu g\ g^{-1}$) (Tran et al., 2015). Among aldehydes, 2-methylpropanal, 3-methylbutanal, nonanal, benzaldehyde, and phenylacetaldehyde were the most abundant in the CBS, as in cocoa beans (Bonvehí et al., 2005; Tran et al., 2015). The Strecker aldehydes, 2-methylpropanal, 3-methylbutanal, and phenylacetaldehyde, formed during fermentation and roasting processes, are described in the literature as flavour-active compounds and as key-aroma markers having a strong chocolate character with malty and buttery notes for the first two compounds and pleasant honey-like and nutty notes for phenylacetaldehyde (Afoakwa et al., 2009). Other aldehydes identified in CBSs, such as 2-phenyl-2-butenal, nonanal, 5-methyl-2-phenyl-2-hexenal, and 2-isopropyl-5-methyl-hex-2-enal (isomers 1 and 2), have also been described as contributors to the cocoa odour and quality of final products conferring cocoa and fruity notes (Menezes et al., 2016; Owusu et al., 2012; Bonvehí, 2005).

Pyrazines were one of the most representative groups of VOCs present in CBSs with concentrations up to $5285.68\ \mu g\ kg^{-1}$ for the Venezuela (VEN9) samples, as observed in several cocoa beans. In this study, 2,3,5,6-tetramethylpyrazine was the most abundant pyrazine in the CBS that represented > 50% of the total amount of pyrazines present in all CBS samples. 2,3,5,6-Tetramethylpyrazine is one of the main components of CBS aroma that exhibits chocolate flavour notes as described in the literature for dark chocolate (Afoakwa et al., 2009). Other pyrazines identified in the CBS were 2,3,5-trimethylpyrazine, 2,3-dimethyl-5-ethylpyrazine, 2,3-dimethylpyrazine, 2-ethyl-6-methylpyrazine, 2,3,5-trimethyl-6-ethylpyrazine, 2-methylpyrazine, and 2,6-dimethylpyrazine. All these compounds, derived from Maillard reactions, are characteristic and responsible for the cocoa aroma, providing the CBS samples with the essential notes of cocoa. 2,3,5-Trimethylpyrazine and 2-ethyl-3,5-dimethylpyrazine are recognized as key-aroma compounds for cocoa and cocoa products and therefore also for the CBS (Frauendorfer et al., 2006; Frauendorfer et al., 2008). A tentative identification was also performed for the pyrazine 2,5-dimethyl-3-isopentylpyrazine present in significant amounts, which is described for the first time in this work for cocoa products.

Another important group of VOCs consists of short- and branched-chain fatty acids such as acetic acid, 2-methylpropanoic acid, and 3-methylbutanoic acid, which are key-aroma compounds used as markers of cocoa and cocoa products. The total amount of acids found in CBS samples were lower than that described by Tran et al. (2015) for cocoa beans. Although acetic acid was found to be a major compound in the CBSs, the concentrations of this acid were lower than those described for cocoa beans. However, the concentrations found for 2-methylpropanoic acid and 3-methylbutanoic acid were similar to those described for cocoa products in the literature (Bonvehí, 2005; Tran et al., 2015). Although these acids are generally associated with unpleasant odour because of their rancid, sour-vinegar, and hammy notes in cocoa products, some acids present in the CBSs, such as octanoic acid and nonanoic acid, could present a pleasant odour with sweet notes.

Concerning alcohols, the total amount semi-quantified for these group of VOCs was lower than that described by Tran et al. (2015) for cocoa beans. The main alcohols found in the CBS were two key-aroma compounds of cocoa: 2-heptanol and 2-phenylethanol that were present at higher concentrations than those 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 present in *Criollo* and *Nacional* CBS samples from Ecuador (ECU4, ECU5, and ECU7) that contribute to the aroma with sweet, fruity, and flowery notes.

Esters were other key VOCs present in the CBS associated with the fruity, floral, and sweet notes attributed to cocoa aroma. The key-aroma marker, 2-phenylethyl acetate was the main ester present in the CBS followed by 3-methylbutyl acetate, methyl 2-phenylacetate, and ethyl benzeneacetate. Although the total amount in this group of VOCs was lower than that 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 concentrations to that found for cocoa beans (Tran et al., 2015).

Other VOCs derived from roasted cocoa identified in the CBS were the terpenes linalool and linalool oxide, both with characteristic key chocolate flavours (Afoakwa et al., 2009; Bonvehí, 2005). Furthermore, the pyrroles (1H-pyrrole-2-carboxaldehyde and 2-acetylpyrrole) were identified and quantified in the CBS in noteworthy amounts. An additional compound that contributes to the CBS flavour of cocoa is acetylfuran. Likewise, in the furan derivate group, furfural was identified and quantified at high concentrations in *Criollo* CBS samples from Venezuela (VEN10). Finally, dimethyl trisulfide, also described as a key-aroma compound for cocoa products, and dimethyl disulfide were identified and quantified. The highest amounts of both compounds

were detected in CBS samples from the Dominican Republic (DOR1) of the *Trinitario* cultivar.

The present study identified other VOCs, present in lower concentrations, which are described for the first time for cocoa-related products and may contribute to the total pleasant aroma of the CBS (see Table 2, compounds highlighted with asterisk symbol, *). Some of these molecules were ketones such as 2-decanone, 3-methyl-2-cyclohexen-1-one, 2-undecanone; pyrazines such as 2,3,5-trimethyl-6-isopentylpyrazine; several terpenes; and Massoialactone (S and R).

In this study, the roasting process was performed under standardized conditions to avoid the influence of this process on the volatile profile of CBS and to better evaluate the effect of the cultivar and/or the geographical origin. However, the CBS is a by-product produced at industrial scale by several companies with specific roasting treatments that could result in different volatile profiles for the same batch of cocoa beans. Therefore, further research needs to be done in order to access the volatile profile under different roasting conditions and evaluate the effect of this key-step on the final flavour of CBS.

3.2. Classification of CBSs based on VOCs determined by SPME-HS-GC-qMS

3.2.1. Classification of CBSs according to the cultivar and continent of origin – First approach

Fig. 1 shows the principal component analysis (PCA) based on volatile fingerprinting of the CBS that was used to find differences among types of cultivars (Fig. 1a) and the continent of provenience (Fig. 1b). The PCA clearly showed a separation ($p < 0.001$) of the *Criollo* CBS compared with the other cultivars that clustered together (Fig. 1a), and this was confirmed by the ANOSIM statistical test. By taking into account the continent of provenience, it was possible to observe a clear separation ($p < 0.007$) of American and African CBS samples (Fig. 1b). However, it can be noted that the marked separation of samples from *Criollo* cultivar makes difficult the differentiation of *Forastero* samples from the others "fine aroma" cultivars (*Trinitario* and *Nacional*). Furthermore, *Criollo* CBS is mainly from the American continent and may also interfere in the classification according to the geographical origin.

Going more deeply into the volatile composition, the level of diversity of the VOCs was clearly different based on the CBS cultivar. Several compounds (48 VOCs) were found to be significantly different according to the cultivar (FDR < 0.001 (10 VOCs), FDR < 0.01 (17 VOCs), and FDR < 0.05 (21 VOCs)), as shown in Table S1.1 (see supplementary material). Key aroma compounds, such as 2-methylpropanal, phenylacetaldehyde, 2,3,5-trimethylpyrazine, 2-ethyl-3,5-dimethylpyrazine, 2-heptanol, 2-phenylethanol, 2-methylpropanoic acid, and 3-methylbutanoic acid, were found to be significant for the CBS discrimination of the *Criollo* cultivar from the *Forastero* and *Trinitario* cultivars. Other compounds such as benzaldehyde, methyl-2-phenylacetate, 2,3-dimethylpyrazine, and 2,3,5,6-tetramethylpyrazine were present at high concentrations in *Criollo* CBSs and can also be putative markers of the *Criollo* cultivar. Tetramethylpyrazine was the most abundant pyrazine present in CBSs yielded from *Criollo* cocoa beans, as described by Tran et al. (2015) for cocoa beans of the same cultivar. The 3-methylbutanoic acid was found to be a potential marker for the *Forastero* CBS, while 2-phenylethanol and 2-heptanol were found to be potential markers for the *Trinitario* CBS, as already described for cocoa beans (Qin et al., 2017). The boxplot of three volatile compounds highly significant for the classification of CBS according to cultivar is shown in Fig. 1c. 2-methylpropanoic acid was found at high concentrations in *Criollo* CBSs and allowed for the discrimination of this

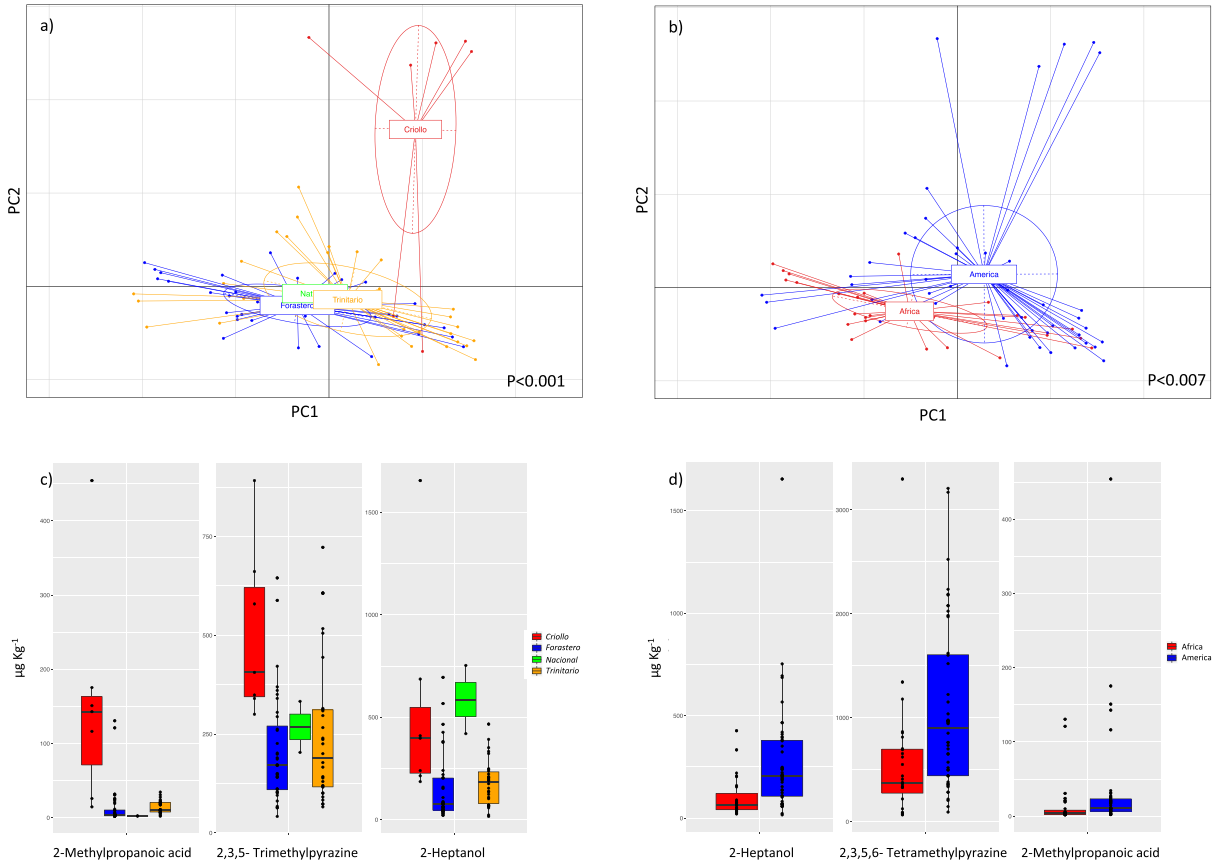

Fig. 1. PCA based on the VOCs (µg kg^{-1}) identified by HS-SPME/GC-qMS in all CBS samples as a function of: (a) cultivar, (b) continent of origin. The variance explained by the first component of PCA (PC1) was 28.54%, while the second component (PC2) explained 15.10%. Box plots showing abundance of key VOCs that can be used as possible markers for (c) cultivar and (d) geographical origin. For interpretation of the legends, see Table 1.

cultivar from *Forastero* and *Trinitario* cultivars (FDR < 0.001) and distinguished the CBS *Trinitario* from the *Forastero* and *Nacional* cultivars (FDR < 0.01 and FDR < 0.05, respectively). Also 2-heptanol allowed for the differentiation of *Trinitario* CBSs from the other cultivars (FDR < 0.05) and additionally distinguished *Forastero* CBSs from *Criollo* and *Nacional* (FDR < 0.01 and FDR < 0.05, respectively). Finally, 2,3,5-trimethylpyrazine was found to be highly significant, differentiating *Criollo* CBSs from *Forastero* and *Trinitario* cultivars (FDR < 0.01).

Taking into account the geographical origin, several compounds (47 VOCs) were found to be significant in the classification of CBSs according to the continent of origin (FDR < 0.001 (11 VOCs), FDR < 0.01 (19 VOCs), and FDR < 0.05 (17 VOCs)) as shown in Table S1.2 (see supplementary material). Considering the key aroma markers identified in cocoa samples (Frauendorfer et al., 2008), the boxplot showed that 2-heptanol (FDR < 0.001) and 2-methylpropanoic acid (FDR < 0.001) were those volatiles with the highest concentration in CBSs of *Criollo* cocoa from the American continent. Moreover, the most abundant pyrazine detected in CBS, 2,3,5,6-tetramethylpyrazine (FDR < 0.01), was also associated with American samples (Fig. 1d).

3.2.2. Classification of Forastero CBS samples according to their geographical origins

PCA analysis was performed to evaluate the sample separation according to the CBS origin of the *Forastero* cultivar among 14 countries from Africa and America (see Fig. 2). A clear separation ($p < 0.001$) was observed according to macroarea (Fig. 2a), latitude (Fig. 2b), and the country of origin (Fig. 2c). For the classification according to latitude, the countries of production were distributed into four main groups: L1 (5°S–5°N); L2 (5°S–20°S); L3 (5°N–10°N); and L4 (10°N–20°N). By taking into account the macroarea as a discriminant factor, it was possible to observe that West African and South American CBS samples cluster together and were well-separated ($p < .001$) from samples from East Africa and Central America (Fig. 2a). Moreover, the different volatile profile drove the impressive cluster separation ($p < 0.001$) according to the latitudes (Fig. 2b). In particular, the CBSs from West Africa and latitude L3 (5°N–10°N) were the samples with low amounts of total VOCs, mainly aldehydes, pyrazines, and sulfur compounds and high amounts of acids. Going more deeply into the classification of the CBS samples as a function of the geographical origin, it was possible to differentiate ($p < 0.001$) the CBS sample according to the country of origin (Fig. 2c). We observed that Congo clustered together with Uganda, and the two South American countries, Ecuador and Colombia, formed a central group in the centre of the PCA. The CBSs from cocoa beans produced in countries located at the latitude L2 (5°S–20°S), such Madagascar, Peru, and Tanzania, were those with high amounts of VOCs among *Forastero* cultivars (see Table 3 in Barbosa-Pereira et al., *submitted*). For these samples, cocoa key aroma compounds were found at high concentrations such as acids (acetic acid), aldehydes (benzaldehyde, 2-methylpropanal, and 3-methylbutanal), esters (3-methylbutyl acetate and 2-phenylethyl acetate), and pyrazines such as 2,3,5-trimethylpyrazine and 2,3,5,6-tetramethylpyrazine. It should be pointed out that CBSs from Madagascar were well-separated from Sao Tomé. Even though both countries are African, they belong to different macroareas and grow at different latitudes. However, they are both islands, with specific climate conditions that may affect the volatile profile of cocoa beans and their products (Afoakwa et al., 2008). CBS samples from these countries were characterized by low concentrations of alcohols and by the presence of 3-methylbutanoic acid, 3-

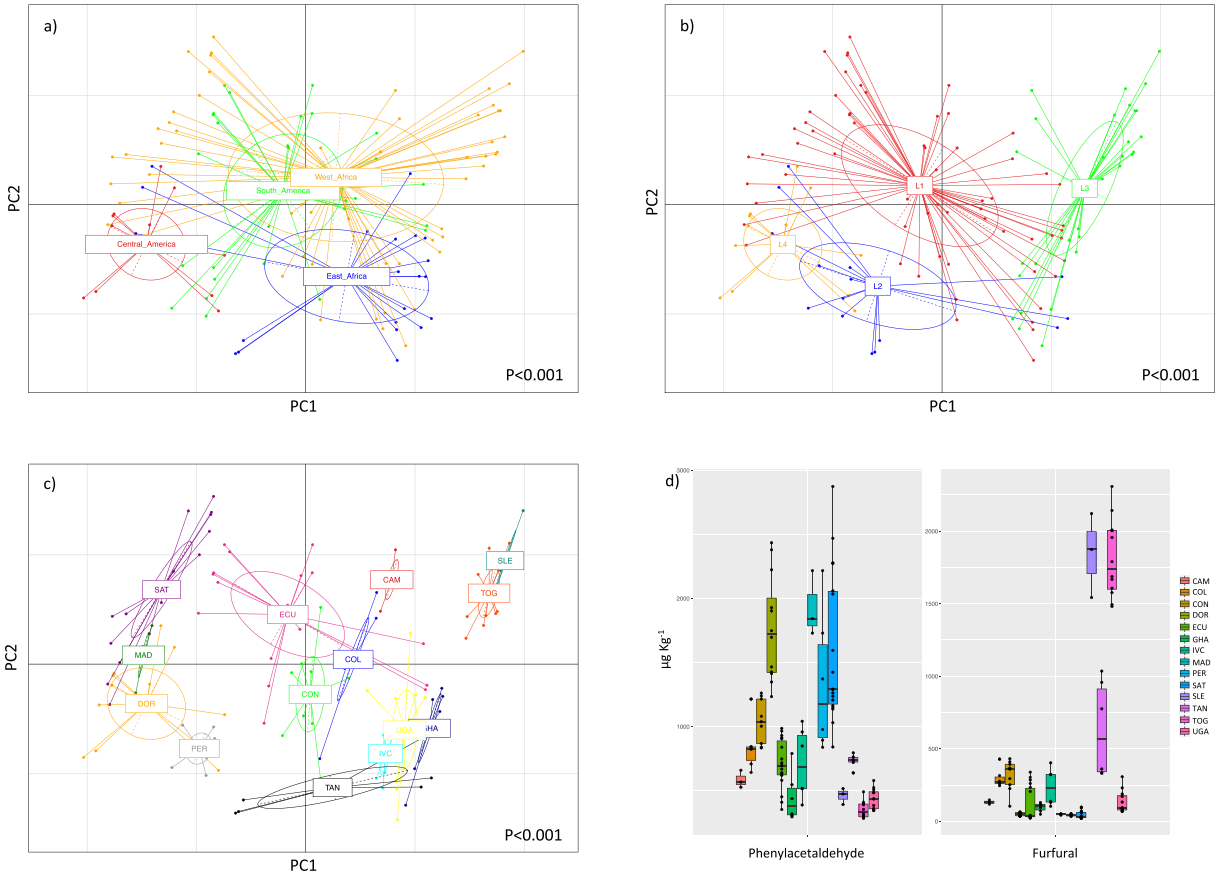

Fig. 2. PCA based on the VOCs (µg kg^{-1}) identified by HS-SPME/GC-qMS in CBS samples of the *Forastero* cultivar according to: (a) macroarea, (b) latitude, and (c) country of origin. The variance explained by the first component of PCA was 32.02%, while the second component explained 13.71%. (d) Boxplot showing abundance of VOCs that can be used as potential markers of origin: phenylacetaldehyde and furfural. For interpretation of the legends, see Table 1.

methylbutanal, phenylacetaldehyde, dimethyl trisulfide, trimethylpyrazine, and tetramethylpyrazine present at high concentrations conferring important flavour characteristics to the CBS that can valorise the product. CBSs from the Central American country the Dominican Republic (L4 10°N–20°N) were characterized by the presence of high amounts of aldehydes and pyrazines. Therefore, among the *Forastero* cultivar, CBS samples from Sao Tomé, Madagascar, the Dominican Republic, and Peru were those with high amounts of VOCs and could be distinguished from the rest of the samples (Fig. 2c).

By taking into account the key VOCs of *Forastero* that drove this separation (Table S2.1, see supplementary material), 2-methylpropanal, 3-methylbutanal, phenylacetaldehyde, dimethyl trisulfide, 2-phenethyl acetate, 2-heptanol, 2-phenylethanol, 2-methylpropanoic acid, 3-methylbutanoic acid, 2,3,5-trimethylpyrazine, and 2-ethyl-3,5-dimethylpyrazine, were found to be putative markers of *Forastero* CBS according to the country of origin. In detail, it was possible to identify two main components that drove the separation among the different origins: phenylacetaldehyde and furfural. Particularly, phenylacetaldehyde was most present in CBSs from the Dominican Republic, Madagascar, Peru, and Sao Tomé, while furfural was in CBSs from Sierra Leone, Togo, and Tanzania (Fig. 2d).

3.2.3. Classification of Trinitario CBS samples according to their geographical origins

Considering *Trinitario* CBS samples (see Fig. 3), we clearly observed a separation ($p < 0.001$) between samples from Central America and South America (Fig. 3a). Moreover, taking into account the latitude, we observed that samples from L1 (5°S–5°N) and L2 (5°S–20°S) clustered together and were well-separated ($p < 0.001$) from L3 (5°N–10°N) and L4 (10°N–20°N) (Fig. 3b). Taking into account the CBS origin, we observed that CBSs yielded from cocoa beans grown in Central America, the Dominican Republic, Jamaica, and Mexico, located at the latitude L4 (10°N–20°N), clustered together. For South America, CBS samples were divided into three latitudes, L1 (5°S–5°N) comprising Ecuador and Colombia, L2 (5°S–20°S) comprising Brazil and Peru that were not separated among them, and finally CBSs from Venezuela at the latitude L3 (5°N–10°N) that were separated from the other three groups (Fig. 3c). However, the high number of samples from Venezuela respect to the other countries of origin could also contribute to this separation, since the sample set is not completely balanced. CBSs from the Dominican Republic, Mexico, and Peru were those with high amounts of total VOCs among *Trinitario* CBSs, including pyrazines (e.g., 2,3,5-trimethylpyrazine, see Fig. 3d), ketones (2-nonanone), acids (acetic acid), and aldehydes (phenylacetaldehyde). Also, CBSs from Colombia and Jamaica displayed intermediate amounts of ketones (2-nonanone), terpenes, and aldehydes (3-methylbutanal) (See Fig. 3d and see Table 1 in Barbosa-Pereira et al., *submitted*). However, Jamaican CBSs were 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; high amounts of nonanal and heptanal; low amounts of pyrazines, esters, and acids; and the presence of high concentrations of furans (furfural and acetylfuran). CBSs from Ecuador were separated from Brazilian samples based on the high content of alcohols, mainly 2-heptanol and 2-phenylethanol (See Table 3 in Barbosa-Pereira et al., *submitted*).

As for *Forastero* CBSs, several VOCs were found as potential markers for the classification of the *Trinitario* CBSs. The volatile compounds that had significant differences (FDR < 0.05 or lower) among American countries are shown in Table S2.2 (see supplementary material). Fig. 3d

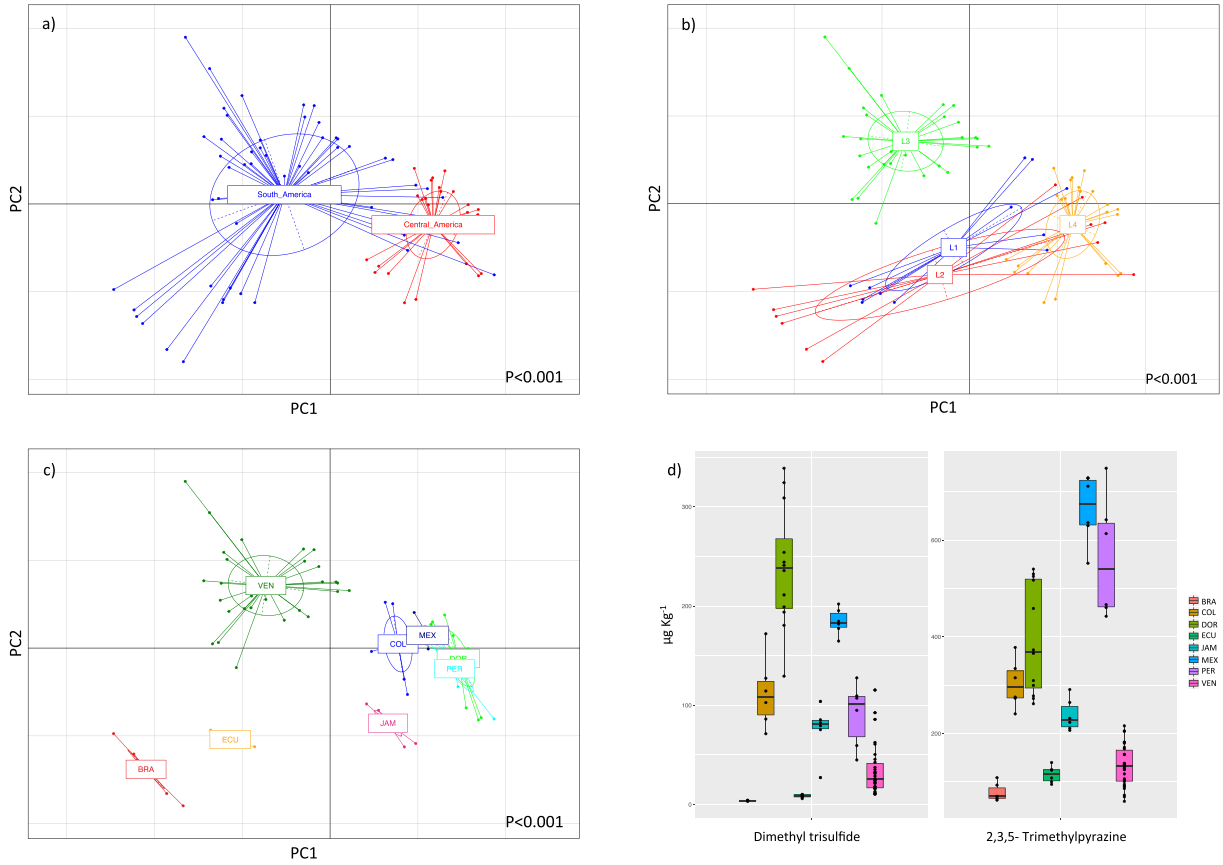

Fig. 3. PCA based on the VOCs (µg kg^{-1}) identified by HS-SPME/qGC-MS in CBS samples of *Trinitario* cultivar according to: (a) macroarea, (b) latitude, and (c) country of origin. The variance explained by the first and second principal component was 35.63% and 15.26%, respectively. (d) Boxplot showing abundance of VOCs that can be used as potential markers of origin: dimethyl trisulfide and 2,3,5-trimethylpyrazine. For interpretation of the legends, see Table 1.

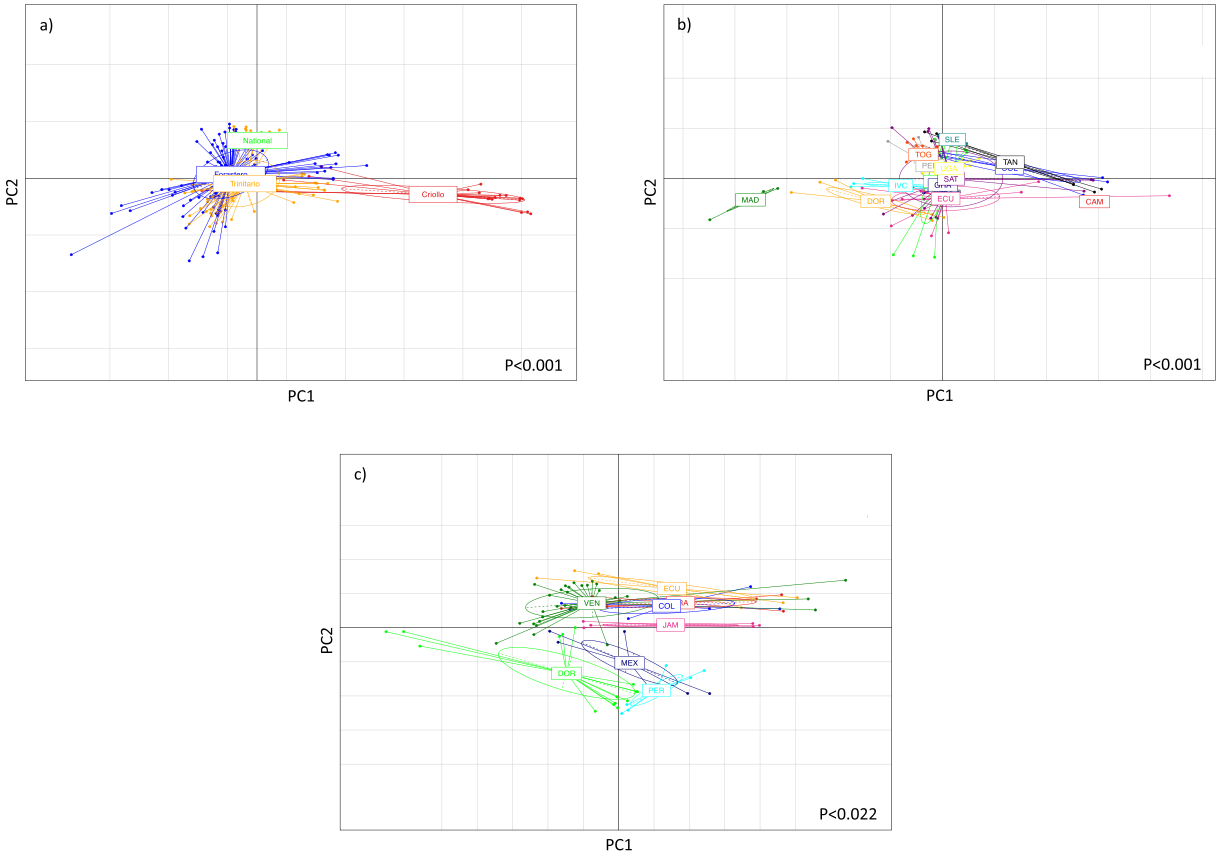

Fig. 4. PCA based on *E*-nose data set for: (a) all CBS samples as a function of cultivar (the variance explained by the first and second principal component was 78.94% and 11.06%, respectively); (b) CBS samples of *Forastero* according to country of origin (the variance explained by the first and second principal component was 74.89% and 13.51%, respectively); (c) CBS samples of *Trinitario* cultivar according to country of origin (the variance explained by the first and second principal component was 68.30% and 13.97%, respectively). For interpretation of the legends, see Table 1.

shows the boxplot of two key aroma compounds for cocoa, dimethyl trisulfide and 2,3,5-trimethylpyrazine, which contributed significantly (FDR < 0.05 or lower) to the classification of CBSs. Specifically, dimethyl trisulfide was most present in CBSs from the Dominican Republic, Colombia, and Mexico, while 2,3,5-trimethylpyrazine was in CBSs from Mexico and Peru.

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

For all CBS samples, the changes in the variation of signals were found to be similar (data not shown). The sensors, S2 (broad), S6 (broad-methane), S7 (sulfur organic), S8 (sensitive broad alcohol), and S9 (sensitive to aromatics and organic sulfides) were those that displayed high response intensity (See Table 4 in Barbosa-Pereira et al., *submitted*). The PCA analysis of E-nose data showed that the most significant classification of CBS was according to cultivar and country of origin (Fig. 4). A clear separation of CBS ($p < 0.001$) samples yielded from cocoa beans of *Criollo* cultivar from the other cultivars was observed (Fig. 4a). In this case, the potentials of several sensors that compose the *E*-nose were considered for the classification of CBSs. The sensors that displayed significant differences (FDR < 0.05 or lower) among the cultivars are shown in Table S3 (see supplementary material). All sensors indicated highly significant separation of the CBSs from the *Criollo* cultivar from *Trinitario* and *Forastero* cultivars. These results confirm the discrimination of CBS samples from the *Criollo* cultivar obtained with GC-qMS data. Considering all cultivars, the most representative sensors for the classification of CBSs were S5 (sensitive to aromatic and aliphatics), S6 (broad-methane), S7 (sensitive to terpenes and sulfur-containing organic compounds), and S10 (methane and aliphatics).

Taking into account the geographical origin, the two groups of CBS samples from cultivars *Forastero* and *Trinitario* are shown in Fig. 4b and Fig. 4c, respectively. For the *Forastero* cultivar, CBSs from Madagascar were well-separated ($p < 0.001$) from the rest of the samples. Likewise, samples from Sierra Leone and Togo were separated from the CBSs from the Dominican Republic and Ecuador as observed by the GC-qMS data. However, *E*-nose was not able to separate Peru or Sao Tomé, as observed from the GC-qMS data (see Fig. 2c). Considering the *Trinitario* cultivar samples, displayed in Fig. 4c, a significant separation ($p < 0.022$) of CBS samples was observed. However, the efficacy of separation was lower than that observed for the GC-qMS analysis. The results highlighted that E-nose can be used as a tool for rapid discrimination of CBS samples from different cultivars, mainly for *Criollo* cultivar. Nevertheless, this methodology presented limitations for the classification of CBS for a single country compared to GC-qMS.

3.4. GC-qMS vs. E-nose – Two case studies

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

In the case of CBSs from Venezuela (Fig. 5a and b), both techniques allowed for the separation of samples from the *Criollo* cultivar (VEN7,

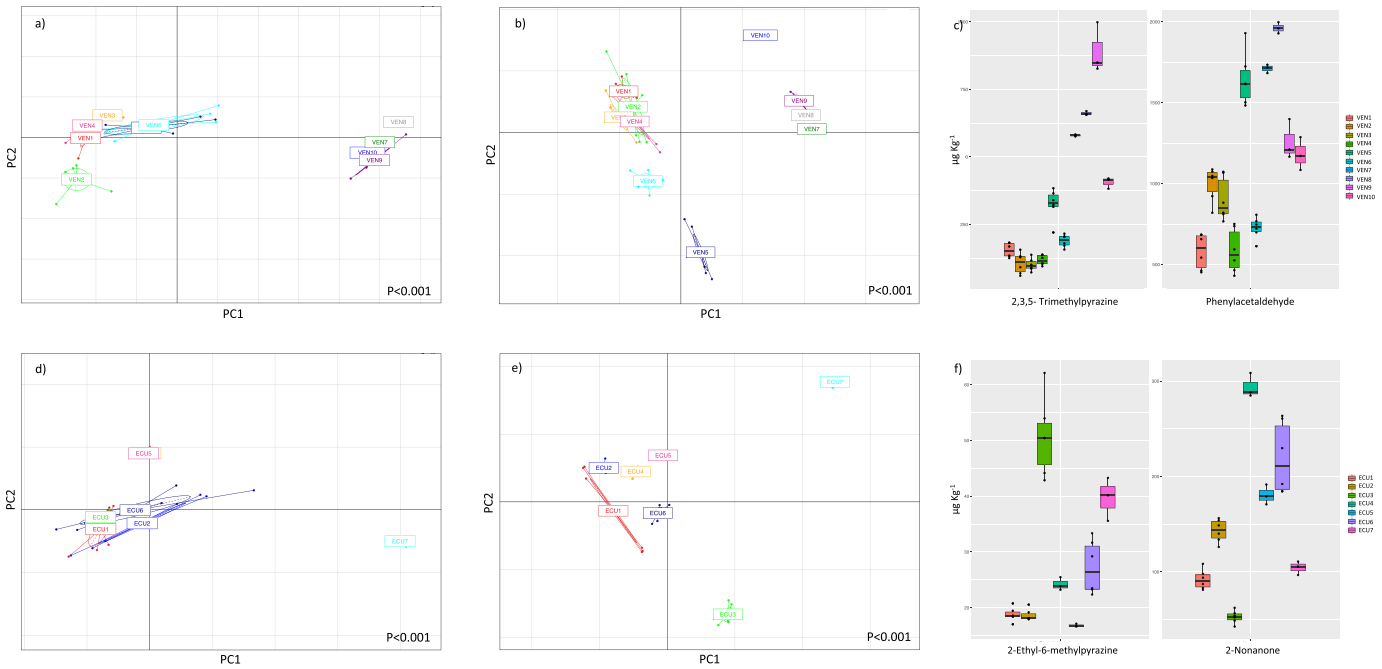

Fig. 5. PCA based on the VOCs (μg kg^{-1}) identified by HS-SPME/GC-qMS and E-nose data set for CBS samples from different regions of Venezuela: (a) PCA based on E-nose data set (the variance explained by the first and second principal component was 93.63% and 3.04%, respectively); (b) PCA based on HS-SPME/GC-qMS data set (the variance explained by the first and second principal component was 32.88% and 22.41%, respectively). (c) Boxplot showing abundance of VOCs that can be used as potential markers of origin: 2,3,5-trimethylpyrazine and phenylacetaldehyde.
PCA based on the VOCs (μg kg^{-1}) identified by HS-SPME/GC-qMS and E-nose data set for CBS samples from different regions of Ecuador: (d) PCA based on E-nose data set (the variance explained by the first and second principal component was 82.23% and 10.76%, respectively); (e) PCA based on HS-SPME/GC-qMS data set (the variance explained by the first and second principal component was 27.76% and 24.13%, respectively). (f) Boxplot showing abundance of VOCs that can be used as potential markers of origin; 2-ethyl-6-methylpyrazine and 2-nonanone.
For interpretation of the legends, see Table 1.

VEN8, VEN9, and VEN10) from other cultivars, except for samples from Canoabo (VEN5), which were not separated with *E*-nose (see Fig. 5a). However, E-nose was capable of differentiating CBSs from the Caucagua region (VEN2) of the *Trinitario* cultivar from the other regions of the same cultivar (VEN1, VEN3, VEN4, and VEN6), which was not accomplished with GC-qMS. Fig. 5b shows a clear separation of CBS samples of the *Criollo* cultivar from the other cultivars using GC-qMS data. Considering the GC-qMS data, CBS samples from the Ocumare region (VEN6 and VEN10) were separated according to cultivar. Moreover, CBSs from Canoabo (VEN5) were also separated from the samples clustered in the *Criollo* varietal group. As observed from the E-nose data, this technique allowed for the separation of a CBS from the rest of the samples of the *Trinitario* cultivar, but in this case, the CBS was from the Ocumare region (VEN6). Therefore, the use of GC-qMS coupled with the *E*-nose technique could be an interesting approach for the classification of Venezuelan CBSs. Volatile molecules such as 2,3,5-trimethylpyrazine and phenylacetaldehyde were identified as potential markers for the classification of CBSs 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 Venezuelan regions of production are shown in Table S4.1 (see supplementary material).

The PCA of CBSs from Ecuador using *E*-nose and GC-qMS data sets is shown in Fig. 5d and Fig. 5e, respectively. As shown in Fig. 5d, the *E*-nose technique allowed for the classification of CBS according to the cultivar. Samples from the *Criollo* cultivar (ECU7) and samples from the *Nacional* cultivar (ECU4 and ECU5) were clearly separated from the *Forastero* cultivar. However, this technique was not able to separate both *Nacional* cultivar CBSs (ECU4 and ECU5),as well the *Trinitario* cultivar CBS (ECU2) from *Forastero* cultivar samples (ECU1, ECU3, and ECU6), which were separated by GC-qMS. PCA based on GC-qMS data of CBSs from Ecuador showed a significant separation (FDR < 0.001) of the samples ECU7 (*Criollo* cultivar) and ECU3 (*Forastero* cultivar) among them and from the rest of the CBS samples. A clear separation between CBS samples from the *Forastero* cultivar (ECU1, ECU3, ECU6) and CBS yielded from cocoa beans of "fine aroma" (*Nacional*, *Trinitario*, and *Criollo* cultivars) was also observed. The GC-qMS technique allowed for a high separation of the CBSs compared to the *E*-nose that presented some limitations. Several VOCs were identified as potential markers for this classification (FDR < 0.05 or lower) of CBSs from Ecuador, such as 2-ethyl-6-methylpyrazine and 2-nonanone, as shown in Fig. 5f (see Table S4.2, supplementary material).

3.5. HS-SPME-GC-qMS vs E-nose – Correlation

According to the results described above, GC-qMS and E-nose could classify the CBSs at different levels. GC-MS methodology consents the identification and quantification of single molecules, it is more sensible and accurate and has demonstrated a remarkable discrimination potential with applicability for food authentication. However, the diffusion and practical application in the food industry of this technology is limited due to the high operational costs and personnel commitment required. Instead, E-nose is economical, non-destructive and easier to work with due to the operational simplicity and allows bulk sampling. This technology has already application on food industry for a rapid screening and identification of samples and verification of variety for traceability purposes. However, it presents some drawbacks including the low sensitivity and specificity and less accuracy in comparison with GC-MS. As described before in section 3.4 for the classification of Venezuelan CBS, the combination of both techniques improved the ability for sample discrimination and therefore could represent a good possibility for CBS classification.

VOCs found at higher concentrations than 100 μg kg^{-1} at least for one sample were selected for the correlation, since this is the limit of

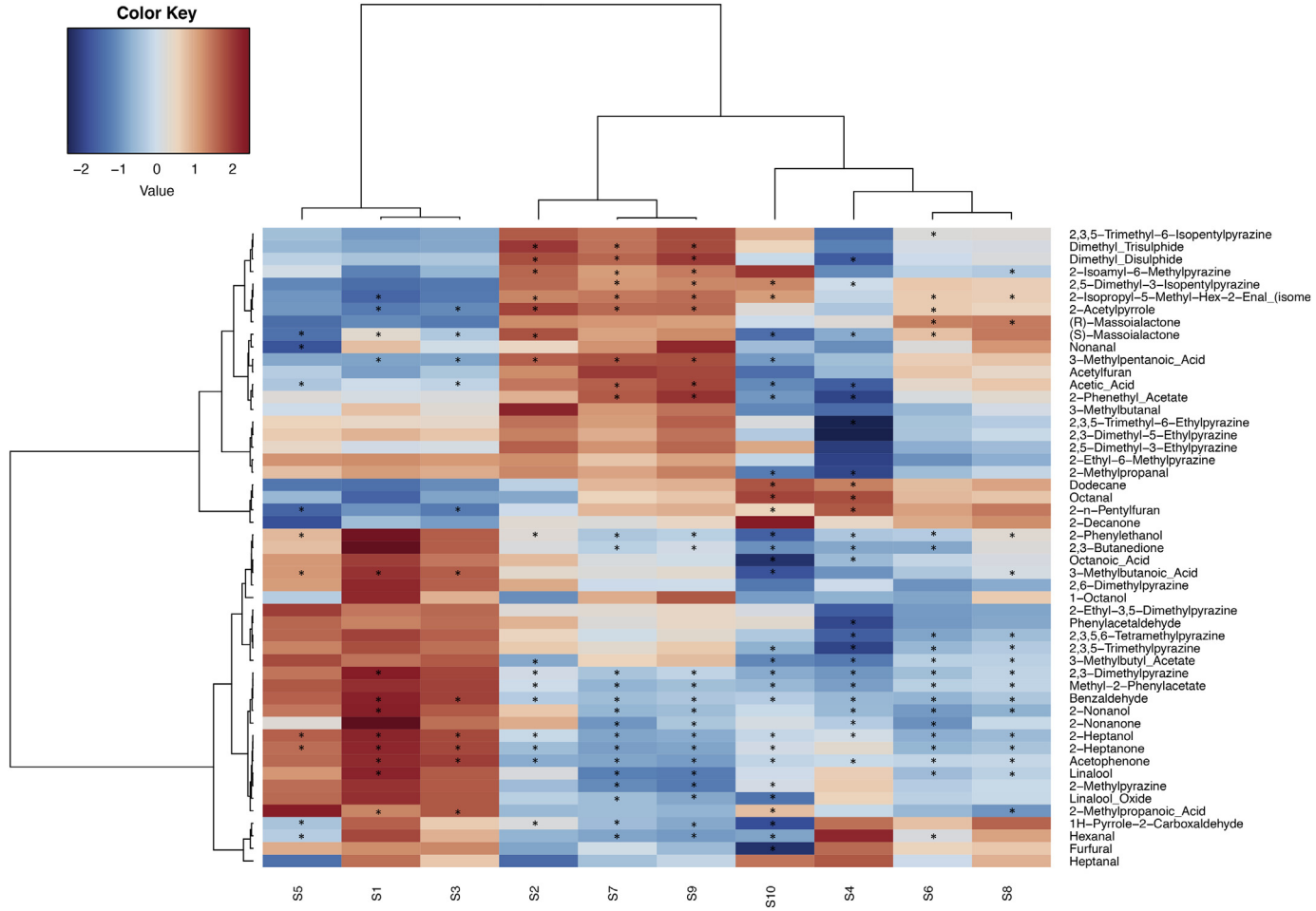

Fig. 6. Correlation between the abundance of VOCs ($\mu g\,kg^{-1}$) and E-nose sensors. Rows and columns are clustered by Ward linkage hierarchical clustering. The intensity of the colours represents the degree of correlation between the samples and VOCs as measured by the Spearman's correlations. Asterisks denote significant correlations after *P* value corrections (FDR $<$ 0.05).

detection of *E*-nose according to the manufacturer (Airsense Analytics GmbH., Germany). The correlation between VOCs and sensors is shown in Fig. 6.

The heatmap shows clearly three main clusters of sensors: S1, S3, and S5 (Cluster 1); S2, S7, and S9 (Cluster 2); and S10, S4, S6, and S8 (Cluster 3). For Cluster 1, S1, S3, and S5 were found to correlate positively with aromatic molecules (S1, S3, and S5), such benzaldeyde and 2-phenylethanol, aliphatic compounds (S5), such pyrazines (2,3-dmethylpyrazine), alcohols (2-heptanol and 2-nonanol), and ketones (2-heptanone and acetophenone), and acids (2-methylpropanoic acid and 3-methylbutanoic acid). These molecules were found at high concentrations for *Criollo* CBS and at lower concentration for samples from countries as Madagascar (*Forastero* classification). The response of the sensors of Cluster 1 showed the same behaviour with high response intensity for *Criollo* CBS and low intensity of response for Madagascar CBS (see Table 4 in Barbosa-Pereira et al., *submitted*). Cluster 2 results from the positive correlation of sensors S2, S7, and S9 with sulfur organic compounds representative of cocoa flavour: dimethyl trisulfide and dimethyl disulfide. These VOCs were found at high concentrations in samples from Mexico, Dominican Republic and Peru. The profile of these samples also showed high response intensity to Cluster 2 and confirm the classification provided by the *E*-nose (see Fig. 4c). Finally, Cluster 3 exhibited a high correlation between sensors related with long chain aliphatic compounds (mainly S4 and S10) such as dodecane (hydrocarbons) and octanal.

4. Conclusion

This study provides information, for the first time, on the volatile fingerprint of the CBS determined by HS-SPME/GC-qMS and identifies the molecules responsible for differences among a feasible number of samples yielded from cocoa beans collected from different geographical regions 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, valorises this by-product as a food ingredient. CBSs yielded form cocoa beans of the *Criollo* cultivar were those with high amounts of “fine aroma” molecules and, therefore, with more potential as a source of cocoa flavour.

GC-qMS-fingerprinting and *E*-nose data allowed for the identification and discrimination of fine flavour cocoa *Criollo*. Several markers, such as 2-methylpropanoic acid, 2,3-dimethylpyrazine, and 2-heptanol, were found to be mandatory for the classification of CBSs according to the cultivar. It was also possible to classify the CBS samples based on their different geographical origins using GC-qMS. Markers such as phenylacetaldehyde and furfural were associated with the CBSs of the *Forastero* cultivar from different countries from America and Africa. While for the *Trinitario* cultivar, dimethyl trisulfide and 3-methylbutanal, among others, were found to be markers capable of classifying CBSs from the American territory.

The results highlighted remarkable diversity in the volatile profile

of CBSs and confirmed the potential applicability of GC-qMS and E-nose for classification and future traceability of CBSs. The capability to identify common trends, leading variables and general indications through rapid and simple technology as E-nose is an encouraging result in this field. However, this technique easy to manipulate still requires the support of more accurate analytical techniques (GC-qMS) for comparison and calibration. Indeed, this study correlates, for the first time, the E-nose and GC-qMS data using a representative number of samples and contribute with new insights for this research field. Nevertheless, our findings are incomplete and further exhaustive investigation needs to be done considering a larger number of samples. The geographical traceability of CBS based on chemical analysis is complex and several intrinsic factors such as climatic conditions and processing conditions (e.g. roasting) should be considered in future studies.

Similar to cocoa beans, the CBS by-product might be considered for selective collection to yield a food ingredient with good aroma and specific flavour characteristics that could be recycled inside the cocoa industry, utilizing the concept of circular economy for a high add-value product.

Authors contributions

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

Funding

The present work was supported by COVALFOOD “Valorisation of high added-value compounds from cocoa industry by-products as food ingredients and additives”. This project has received funding from the European Union's Seventh Framework programme for research and innovation under the Marie Skłodowska-Curie grant agreement No 609402-2020 researchers: Train to Move (T2M).

Conflict of interests

The authors declare that they have no conflicts of interest.

Acknowledgments

L. Barbosa-Pereira gratefully acknowledges the European Union's Seventh Framework programme for the Marie Skłodowska-Curie grant. O. Rojo-Poveda is grateful to the University of Turin for her predoctoral fellowship. The authors are grateful to Silvio Bessone S.r.l., ICAM S.p.A., Ferrero International S.A., Guido Gobino S.r.l., Pastiglie Leone S.r.l., and Venchi S.p.A. for supplying the cocoa bean samples.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodres.2019.05.041.

References

Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2008). Flavor formation and character in cocoa and chocolate: A critical review. *Critical Reviews in Food Science and Nutrition, 48*(9), 840–857.

Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2009). Matrix effects on flavour volatiles release in dark chocolates varying in particle size distribution and fat content using GC–mass spectrometry and GC–olfactometry. *Food Chemistry, 113*(1), 208–215.

Aprotosoaie, A. C., Luca, S. V., & Miron, A. (2016). Flavor chemistry of cocoa and cocoa products—An overview. *Comprehensive Reviews in Food Science and Food Safety, 15*(1), 73–91.

Barbosa-Pereira, L., Dal Bello, B., Belviso, S., Giordano, M., Ferrocino, I., & Zeppa, G. (2017). Cocoa bean shell as functional ingredient to enhance the properties and nutritional quality of food products. *Poster session presentation at the meeting of the European Federation of Food Science and Technology (EFFoST)*Barcelona: Sitges November.

Barbosa-Pereira, L., Guglielmetti, A., & Zeppa, G. (2018). Pulsed electric field assisted extraction of bioactive compounds from cocoa bean shell and coffee Silverskin. *Food and Bioprocess Technology, 11*(4), 818–835.

Barbosa-Pereira, L., Rojo-Poveda, O., Ferrocino, I., Giordano, M. & Zeppa, G. Data on volatile compounds in cocoa bean shell analyzed by HS-SPME/GC-qMS. Data in Brief. (Submitted).

Benedetti, S., Buratti, S., Spinardi, A., Mannino, S., & Mignani, I. (2008). Electronic nose as a non-destructive tool to characterise peach cultivars and to monitor their ripening stage during shelf-life. *Postharvest Biology and Technology, 47*(2), 181–188.

Bonvehí, J. S. (2005). Investigation of aromatic compounds in roasted cocoa powder. *European Food Research and Technology, 221*(1–2), 19–29.

Caprioli, G., Fiorini, D., Maggi, F., Nicoletti, M., Ricciutelli, M., Toniolo, C., ... Sagratini, G. (2016). Nutritional composition, bioactive compounds and volatile profile of cocoa beans from different regions of Cameroon. *International Journal of Food Sciences and Nutrition, 67*(4), 422–430.

Counet, C., Callemien, D., Ouwerx, C., & Collin, S. (2002). Use of gas chromatography – olfactometry to identify key odorant compounds in dark chocolate. Comparison of samples before and after conching. *Journal of Agricultural and Food Chemistry, 50*(8), 2385–2391.

Crafack, M., Keul, H., Eskildsen, C. E., Petersen, M. A., Saerens, S., Blennow, A., ... Nielsen, D. S. (2014). Impact of starter cultures and fermentation techniques on the volatile aroma and sensory profile of chocolate. *Food Research International, 63*, 306–316.

Danezis, G. P., Tsagkaris, A. S., Brusic, V., & Georgiou, C. A. (2016). Food authentication: State of the art and prospects. *Current Opinion in Food Science, 10*, 22–31.

Frauendorfer, F., & Schieberle, P. (2006). Identification of the key aroma compounds in cocoa powder based on molecular sensory correlations. *Journal of Agricultural and Food Chemistry, 54*(15), 5521–5529.

Frauendorfer, F., & Schieberle, P. (2008). Changes in key aroma compounds of Criollo cocoa beans during roasting. *Journal of Agricultural and Food Chemistry, 56*(21), 10244–10251.

Gu, F., Tan, L., Wu, H., Fang, Y., Xu, F., Chu, Z., & Wang, Q. (2013). Comparison of cocoa beans from China, Indonesia and Papua New Guinea. *Foods, 2*(2), 183–197.

International Cocoa Organization. Statistics – Production. Quarterly bulletin of cocoa statistics. (2018). Vol. XLIV No. 3 https://www.icco.org/about-us/international-cocoa-agreements/cat_view/30-related-documents/46-statistics-production.html Accessed 31 November 2018.

Kongor, J. E., Hinneh, M., Van de Walle, D., Afoakwa, E. O., Boeckx, P., & Dewettinck, K. (2016). Factors influencing quality variation in cocoa (Theobroma cacao) bean flavour profile—A review. *Food Research International, 82*, 44–52.

Magagna, F., Guglielmetti, A., Liberto, E., Reichenbach, S. E., Allegrucci, E., Gobino, G., ... Cordero, C. (2017). Comprehensive chemical fingerprinting of high-quality cocoa at early stages of processing: Effectiveness of combined untargeted and targeted approaches for classification and discrimination. *Journal of Agricultural and Food Chemistry, 65*(30), 6329–6341.

Mandrile, L., Barbosa-Pereira, L., Sorensen, K. M., Giovannozzi, A. M., Zeppa, G., Engelsen, S. B., & Rossi, A. M. (2019). Authentication of cocoa bean shells by near-infrared and mid-infrared spectroscopy and inductive coupled plasma-optical emission spectroscopy. *Food Chemistry, 292*, 47–57.

Menezes, A. G. T., Batista, N. N., Ramos, C. L., e Silva, A. R. D. A., Efraim, P., Pinheiro, A. C. M., & Schwan, R. F. (2016). Investigation of chocolate produced from four different Brazilian varieties of cocoa (Theobroma cacao L.) inoculated with Saccharomyces cerevisiae. *Food Research International, 81*, 83–90.

Nsor-Atindana, J., Zhong, F., & Mothibe, K. J. (2012). In vitro hypoglycemic and cholesterol lowering effects of dietary fiber prepared from cocoa (Theobroma cacao L.) shells. *Food & Function, 3*(10), 1044–1050.

Oliveira, L. F., Braga, S. C., Augusto, F., Hashimoto, J. C., Efraim, P., & Poppi, R. J. (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., ... 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 amino acids, reducing sugar, solvent, reacting, cooling and filtering.* Patent CN104187539-A, China.