Food Chemistry 289 (2019) 591–602

Contents lists available at ScienceDirect

Food Chemistry

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

PTR-QiToF-MS and HSI for the characterization of fermented cocoa beans from different origins

Valentina Acierno[a,b,*], Giorgia Fasciani[a,c], Sajad Kiani[a,d], Augusta Caligiani[c], Saskia van Ruth[a,b]

[a] RIKILT Wageningen University and Research, P.O. Box 230, 6700 AE Wageningen, The Netherlands
[b] Food Quality and Design Group, Wageningen University and Research, P.O. Box 17, 6700 AA Wageningen, The Netherlands
[c] Department of Food Science, University of Parma, Parco Area delle Scienze 17/A, 43124 Parma, Italy
[d] Biosystems Engineering Department, Tarbiat Modares University, Tehran, Iran

ARTICLE INFO

Keywords:
Cocoa beans
Computer Vision System
Geographical origin
PTR-QiToF-MS

ABSTRACT

The wide range of geographical cocoa production areas and the increasing consumption trend towards single origin products induced the necessity to verify and certify cocoa beans origin for quality assurance purposes. In this study cocoa beans of various origins were examined by machine olfaction and machine vision techniques. Fifty-nine fermented and dried Forastero cocoa beans from 23 different geographical origins (Africa, Americas, Southeast Asia) were investigated using Proton Transfer Reaction-Quadrupole interface-Time of Flight-Mass Spectrometry and Hyperspectral Imaging to elucidate the geographical information in the beans. The volatile and spectral fingerprints showed the same tendency in clustering samples from Africa separate from those from the Americas. High variability was observed for the Southeast Asian samples, which is most likely related to differences in fermentation. Machine olfaction and machine vision characterization provided a similar degree of separation but are complementary rapid techniques, which may be further developed for use in practical settings.

1. Introduction

The demand in cocoa importing countries for certified cocoa has increased considerably over the last decades. The wide range of geographical production areas and the increasing consumption trend towards single origin food products induced the necessity to verify and certify bean origins during quality assurance (Yener et al., 2015). Moreover, certification is considered a tool to promote sustainability in the cocoa value chain and to improve the livelihoods of cocoa farmers (Icco, 2012 (web)). Certification and administrative controls are in place and useful but it were even better if these could be reinforced by analytical assessments that are able to authenticate the product's origin.

Cocoa characteristics have been studied with application of several analytical techniques in order to discriminate cocoa geographical origin, demonstrating the importance of the topic. Volatile compounds (Qin et al., 2017), amino acids (Marseglia, Palla, & Caligiani, 2014), peptides (Caligiani, Marseglia, Prandi, Palla, & Sforza, 2016), triglycerides (Hernandez, Castellote, & Permanyer, 1991), and volatile acids (Jinap & Dimick, 1990) have been analysed in cocoa beans from different origins.

Mass spectrometry (Qin et al., 2017), spectroscopic (Marseglia et al., 2016) and sensor techniques (Teye, Huang, Dai, & Chen, 2013) have been used to authenticate cocoa beans origin. Nevertheless, with the constant increase of authenticity challenges over the past decades (Lohumi, Lee, Lee, & Cho, 2015), the development of more new, fast and reliable analytical methods for geographical origin assessment is highly desirable. So far, looking for unique characteristics of cocoa beans originating from different continents, machine olfaction by of Proton Transfer Reaction -Mass Spectrometry (PTR-MS) for volatile compounds detection and machine vision by Hyperspectral Imaging (HSI) for spectral and spatial investigation has never been tested. The volatile compounds and spectral information have been widely used as sensitive and fast analytical techniques for authentication and quality analysis of food. PTR-Time of Flight-MS (PTR-ToF-MS) has been used to detect the geographical origin of several products such as coffee (Yener et al., 2014), saffron (Masi et al., 2016), olives (Masi, Romani, Pandolfi, Heimler, & Mancuso, 2015), and wines (Campbell-Sills et al., 2016). Moreover, this instrument has been used for chocolate analysis offering the possibility to compare cocoa and chocolate profiles. HSI has been successfully applied for the analysis of several food commodities (Lohumi et al., 2015) such as coffee beans (Cho, Bae, Cho, & Moon, 2017), tea (Zhao, Chen, Cai, & Ouyang, 2009) and turkey ham

* Corresponding author at: RIKILT Wageningen University and Research, P.O. Box 230, 6700 AE Wageningen, The Netherlands.
E-mail address: valentina.acierno@wur.nl (V. Acierno).

https://doi.org/10.1016/j.foodchem.2019.03.095
Received 3 September 2018; Received in revised form 13 March 2019; Accepted 19 March 2019
Available online 20 March 2019

(Elmasry, Iqbal, Sun, Allen, & Ward, 2011).

In the current study, we examine fermented and dried Forastero cocoa beans from 23 different geographical origins by machine olfaction and machine vision technology, i.e. using PTR-Quadrupole interface ToF-MS (PTR-QiToFMS) and HSI. The volatile and spectral profiles are used as cocoa fingerprint in order to elucidate the geographical information incorporated in the fermented and dried beans of various provenance.

2. Materials and methods

2.1. Materials

2.1.1. Samples

Fifty-nine fermented and dried cocoa beans samples from 23 different geographical origins were considered. The sampling is representative of the average world production, Africa (Af), Americas (Am), and Southeast Asia (As) and, to the best of our knowledge, it represents the largest cocoa beans collection of different geographical origins considered in the literature. Countries of origin and numbers of lots of cocoa beans samples collected are the following: Africa-Congo (1), Ghana (4), Ivory Coast (4), Madagascar (1), Nigeria (4), Sao Thome (6), Sierra Leone (2), Tanzania (5), Uganda (2); Americas- Brazil (1), Cuba (1), Dominican Republic (5), Ecuador (2), Grenada (1), Mexico (1), Perù (7), Trinidad (1), Venezuela (4); Southeast Asia- Flores Indonesia (1), Java (1), Malaysia (1), Sulawesi Indonesia (1), Papua New Guinea (4). The cocoa beans are belonging to the Forastero cocoa genetic group. Samples are characterised by cocoa beans traded from countries of origin to the transforming company, and in most cases, they are constituted of a blend made from cocoa collected in different farms, so they are most representative of the country instead of the specific farm.

2.2. Methods

2.2.1. Machine olfaction: PTR-QiToF-MS

2.2.1.1. Instrumental conditions. Cocoa beans samples were powdered using an electrical food grater. To prevent the grinder overheating, it was stopped every 15 s. The powder of each sample was sieved with a 630 microns (μm) mesh and kept at 4 °C prior to analysis. For the measurement, 0.1 g of ground cocoa beans was weighed into clean and odourless flasks of 250 ml. The closed flasks were placed in a water bath at 25 °C for 30 min to equilibrate the samples with their headspace.

The headspace measurements were performed using a PTR-QiToF-MS (Ionicon Analytik G.m.b.H., Innsbruck, Austria). Ionisation was carried out under drift tube conditions of 900 ± 1 V, drift pressure of 3.80 mbar, temperature 60 °C, and E/N value of 120 Td. Data acquisition was carried out at 1 spectrum per second. For each sample, a mass range between 0.00 and 512.15 was measured using a dwell time of 0.1 s mass^{-1}. Sampling was performed at a flow rate of 60 ± 1 sccm. For all samples, the headspace was measured for 1 min and the measurements were executed in three replicates. The first 10 s were subtracted as blank to the next 40 s of acquisition and the last 10 s were not included in the analysis. The average of the three replicates was taken into account for each sample.

2.2.1.2. Data treatment. Data were processed using PTRwid software (Holzinger, 2015). Peak detection, mass scale calibration were performed according to a procedure described in Holzinger (2015).The mass scale calibration relies on maximizing the matches with a library of ion masses (~2400 compounds); the correct set of parameters yield the most matches with the library. PTRwid has been also used for baseline signal, peak shape, and integration The experimental m/z values were reported up to three decimals. The VOC concentrations are presented in ppbv (part per billion by volume) according to the formula described by Lindinger, Hansel, Jordan, and Hansel (1998). A concentration threshold of 1pbbv was set for the data set, and the mass peaks with less concentration were filtered out. Peaks attributed to ^{13}C isotopologues and water clusters, were considered to be redundant and therefore discarded

2.2.1.3. Statistical analysis

2.2.1.3.1. Univariate analysis: non-parametric Kruskal Wallis test. The mass peaks were subjected to a rank-based nonparametric test, Kruskal Wallis test (IBM SPSS Statistics 23.0, IBM Corp., Armonk, NY, USA), to investigate statistical differences according to the geographical origins of the beans: Africa (Af), Americas (Am), Southeast Asia (As). The non-parametric test was followed by a rate of false discovery (RDF) correction. In order to find out where the differences between groups lie a pairwise Mann-Whitney test was applied. A significance level of $p \leq 0.05$ was used throughout the study.

2.2.1.3.2. Multivariate analysis: principal component analysis (PCA). PCA was used to graphically visualize the presence of any natural clustering in the data samples according to geographical origins of the beans (Af, Am, As). Multivariate analysis was conducted using Pirouette 4.0 Software (Infometrix, Seattle, WA, USA).

2.2.2. Machine vision: HSI

2.2.2.1. Instrumental conditions. A Specim IQ hyperspectral line scan camera (Specim, Spectral Imaging Ltd.) was applied for HSI. Imaging were performed with exposure of 30 ms, in the spectral range of 400–1000 nm, and with 2–3 nm spectral steps. A halogen-based lighting system was used to covers the full 400 to 1000 nm range. White references were utilized to calibrate captured images. The resulting image was a 3D data cube (X, Y, Z), two dimensions (X: 512 pixels and Y: 512 pixels) were utilized for spatial information and one for spectral bands (Z: 204 pixels). Each pixel represents a complete spectrum reflecting specific information against wavelengths. Before the measurements, homogeneity of the lighting system and sensitivity of the camera on the scanned view were evaluated and confirmed.

2.2.2.2. Data treatment. Reflectance spectra of each sample treatment were obtained from a region of interest (30 × 30 pixels) on the sample surface. In order to apply chemometric techniques, the hypercube (30 × 30 × 204) was unfolded into a two-dimensional matrix (900 × 200), and then averaged to a new matrix (20 × 200). After that, the smoothing operation was applied to the spectra matrix using Standard Normal Variant (SNV). Data treatment was performed using The Unscrambler, CAMO Software (Oslo, Norway). Afterwards, the 20 data points were averaged in order to have one spectral trend for each sample. The final data matrix was composed of 59 samples *200 wavelengths.

2.2.2.3. Statistical analysis

2.2.2.3.1. Multivariate analysis: principal component analysis (PCA). PCA was used to graphically visualize the presence of any natural clustering in the data samples according to geographical origins of the beans (Af, Am, As). Multivariate analysis was conducted using Pirouette 4.0 Software (Infometrix, Seattle, WA, USA).

2.2.3. Machine olfaction and machine vision comparison

Correlations between VOCs and wavelengths were evaluated by Pearson's correlation coefficients (r). T-test was used to determine significant correlations ($p < 0.05$) (Liu et al., 2018). All the statistical analyses of data were conducted using R 3.3.3 (R Foundation for Statistical Computing, Vienna, Austria).

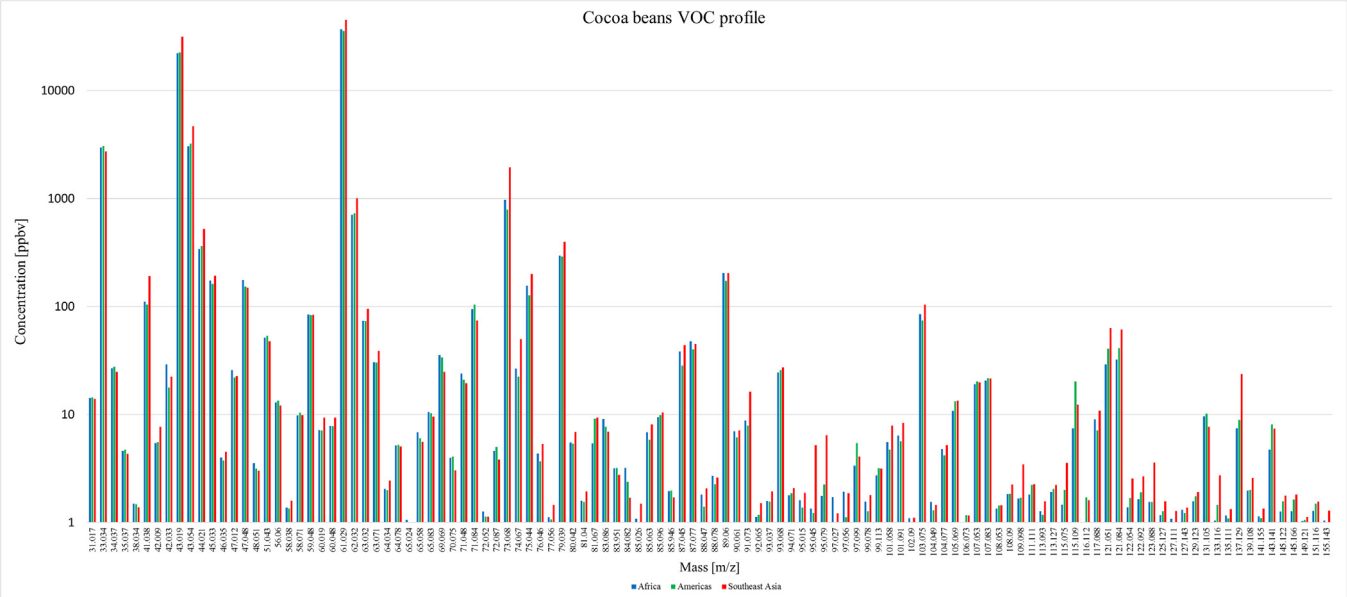

Fig. 1. Mean PTR-QiToF-MS mass spectra of the volatile compounds of African, American and Southeast Asian cocoa beans. Note, the chart has a logarithmic scale.

3. Results and discussion

3.1. Machine olfaction: PTR-QiToF-MS

3.1.1. Qualitative geographical comparison

3.1.1.1. Preliminary comparison of VOC mass spectra. The cocoa beans from all 23 provenances were analysed by PTR-QiToF-MS. The headspace, measured in the range 0.00–521.15 m/z, amounted to 415 mass peaks overall. The dataset was filtered based upon average concentration, the 114 most abundant peaks (higher than 1 ppb) were retained and a final profile including masses from m/z 31.017 to 155.143 was obtained. The profiles of the averaged PTR-QiToF-MS mass spectra of Af, Am and As cocoa beans are presented in Fig. 1.

The highest concentrations were observed for the ions m/z 33.033, 43.019, 43.054 and 61.028 in all samples. These masses were tentatively identified as methanol (33.033), fragment of diverse origin (43.019, 43.054), and acetic acid (61.028). Fig. 1 also shows m/z 73.068 at high concentration. This ion is tentatively identified as 2-methylpropanal, a flavour active compound with a strong chocolate character (Afoakwa, Paterson, Fowler, & Ryan, 2008). In Table 1, the averaged concentrations of the tentatively identified compounds are indicated for each continent. Acetic acid was also detected as the most abundant odorant in unroasted beans in another study (Frauendorfer & Schieberle, 2008). Moreover, masses 33.033, 43.019, and 61.028 were found to be predominant in previous studies focused on chocolates manufactured with cocoa beans from different origins analysed using PTR-QMS (Acierno, Yener, Alewijn, Biasioli, & Van Ruth, 2016) and PTR-QiToF-MS (Acierno, Liu, Alewijn, Stieger, & Van Ruth, 2019). Differently from chocolates, these three masses did not show significant differences between cocoa beans from the three different continents, however, the mass peaks averaged concentrations follow a similar trend both in cocoa beans and chocolates. The Southeast Asian samples showed the highest concentration of mass 43.019 and 61.028 both in beans and chocolates followed by the samples from the Americas. On the other hand, the Southeast Asian samples had the lowest concentration of mass 33.033 both in cocoa beans and chocolates. Taking into account that acid acetic contributes most to beans and chocolates acidity (Jinap & Dimick, 1990) the results of this study are in agreement with those of Counet, Ouwerx, Rosoux, and Collin (2004), that found dried and fermented beans from Southeast Asia and South Pacific to comprise higher acidity levels than those from South African origin due to the box of fermentation used, and possible application of rapid artificial drying. According to Fig. 1 and Table 1, the samples from Southeast Asia tended to be higher in VOCs concentrations compared to the other continents. Our results suggested that the differences in VOC profiles of fermented and dried beans from different origins are quantitative rather than qualitative (Counet et al., 2004; Frauendorfer & Schieberle, 2008).

3.1.1.2. Tentative identification of VOCs released during fermentation and drying. The accuracy of the PTR-QiToF-MS in measuring the exact mass allowed the assignment of sum formulas. The PTR-QiToF-MS results combined with literature data permitted a tentative identification of 64 masses/ions (Table 1). The highlighted compounds in Table 1 were tentatively identified, however, because of the low concentration (< 1 ppb) were not included in the analysis. The majority of the literature related to the volatile compounds of cocoa beans refers to the key aroma compounds that play an important role in the flavour of beans. The list of compounds developed by Huang and Barringer (2011) on cocoa and studies related to fermented and dry cocoa beans was taken into account. Moreover, several compounds were confirmed comparing these results with masses found in coffee studies applying the same instrument (Özdestan et al., 2013; Romano et al., 2014; Yener et al., 2014, 2015, 2016). Masses detected in chocolate (Acierno et al., 2016) were also considered.

The PTR-QiToF-MS measurements allowed detection of compounds of the main classes of cocoa odorant compounds, such as alcohols, acids, esters, aldehydes, pyridines, and pyrazines (Table 1). The cocoa beans analysed were fermented and dried. During these steps of production, the first aroma compounds are developed, which are predominantly aldehydes, alcohols, acids and acetates (Ziegleder, 2009). It is important, however, to underline that differences in the aroma profile of unroasted (fermented and dried) and roasted bean are caused by more quantitative more than qualitative changes (Frauendorfer & Schieberle, 2008).

In this study, within the alcohols group methanol (m/z 33.034), 2, 3-butanediol (m/z 91.073), and 2-phenylethanol (m/z 123.088) were tentatively identified. The latter was also identified as the most odour-active compounds in fermented and unroasted cocoa beans by Frauendorfer and Schieberle (2008). It has been reported in cocoa fermentation with *K. apiculata and S. cerevisiae* var. *chevalieri* yeasts and described as desirable compounds for high-quality cocoa products

Table 1
Average of concentrations (ppbv) of tentatively identified mass peaks in the headspace of cocoa beans of different origins detected by PTR-QiToF-MS. Masses mean showing geographical significant difference according to Kruskal Wallis tests ($p < 0.05$) are highlighted.

				Average VOC concentrations (ppbv)±SD		
m/z	Sum formula	Tentative identification	Chemical class	Africa	Americas	Southeast Asia
33.034	CH_5O^+	Methanol [3,4]	Alcohol	2962.9 ± 713.6	3050.5 ± 771.2	2724.5 ± 735.9
41.038	$C_3H_5^+$	Alkylic fragment [3,6]	Fragment	110.9 ± 51.9	104.5 ± 47.8	190.9 ± 277.3
43.019	$C_2H_3O^+$	Fragment (diverse origin) [4,7]	Fragment	22267.8 ± 16113.7	22510.5 ± 23284.8	31561.6 ± 28765.4
43.054	$C_3H_7^+$	Fragment (diverse origin) [4,7]	Fragment	3037.8 ± 2255.8	3232.7 ± 3501.7	4680.6 ± 4534.0
45.033	$C_2H_5O^+$	Acetaldehyde/ethanol [2,3,4]	Aldehyde/alcohol	173.7 ± 45.0	161.8 ± 55.6	193.2 ± 94.3
59.048	$C_3H_7O^+$	Acetone/propanal [3,5]	Ketone/aldehyde	84.1 ± 20.3	83.2 ± 14.8	83.9 ± 33.7
61.029	$C_2H_5O_2^+$	Acetic acid [3,4,5]	Acid/ester	37012.3 ± 23610.2	35550.2 ± 29906.5	44913.0 ± 35949.8
67.053	$C_5H_7^+$	Terpene fragment [6,7]	Fragment	0.7 ± 0.2	0.8 ± 0.2	0.8 ± 0.3
68.054	$C_4H_6N^+$	Pyrrole [4,5]	Pyrrole	0.3 ± 0.1	0.3 ± 0.1	0.4 ± 0.2
71.048	$C_4H_7O^+$	Dihydrofuran/2-methylpropenal/3-buten-2-one/alcohol fragment [3,4]	Furan/aldehyde/ketone/fragment	24.0 ± 15.1	21.0 ± 10.5	19.5 ± 11.4
71.084	$C_5H_1^{1+}$	Terpene fragment/pentene [3,6]	Fragment/alkene	94.9 ± 61.5	104.5 ± 69.6	73.9 ± 41.7
73.068	$C_4H_8O^+$	2-Methylpropanal (isobutanal)/Butanone[3,8]	Aldehydes/ketone	971.8 ± 499.1	784.9 ± 420.8	1933.2 ± 1470.0
75.043	$C_3H_7O_2^+$	Propanoic acid/methyl-acetate/acetol [3,4,7]	Acid/ester/ketone	156.2 ± 108.3	127.5 ± 96.2	201.3 ± 187.4
81.067	$C_6H_9^+$	Methylpentene/terpene fragment [4]	Alkene/fragment	5.4 ± 3.1[a]	9.1 ± 4.7[b]	9.4 ± 10.6[ab]
85.026	$C_4H_5O_2$+	2(5H)-furanone [4]	Ketone	1.1 ± 0.7	0.9 ± 0.8	1.5 ± 1.2
85.063	$C_5H_9O^+$	Methyl-butenal [3,5]	Aldehyde	6.8 ± 3.6	5.8 ± 3.8	8.1 ± 5.7
87.045	$C_4H_7O_2^+$	2,3-butandione (diacetyl)/butanedione/butyrolactone [3,5,6]	Ketone/ester	38.2 ± 23.8	28.3 ± 17.5	44.1 ± 23.7
87.077	$C_5H_{11}O^+$	3-methylbutanal/2-methylbutanal/methylbutanal[2,3,4,5]	Aldehyde/alcohol	47.9 ± 24.5	40.0 ± 23.1	45.0 ± 21.0
89.060	$C_4H_9O_2^+$	Methylpropanoate/2-methylpropanoic acid/hydroxybutanone/acetoin [2,3,4,5]	Ester/acid/ketone	203.1 ± 128.3	172.5 ± 119.2	204.0 ± 149.1
91.073	$C_4H_{11}O_2^+$	2,3- Butanediol [9]	Alcohol	8.8 ± 5.1	7.9 ± 5.4	16.2 ± 13.1
93.068	$C_7H_9^+$	Terpene fragment/Toluene [1,3,6]	Fragment	24.5 ± 6.2	25.8 ± 6.4	27.3 ± 8.0
95.045	$C_6H_7O^+$	Phenol/cyclohexadienone/1,4-cyclopentadiene-1-carbaldehyde	Alcohol/ketone/aldehyde	1.3 ± 0.6	1.2 ± 0.5	5.2 ± 8.3
96.081	$C_6H_{10}N^+$	Dimethyl-pyrrole/ethyl-pyrrole [4,5]	Pyrrole	0.3 ± 0.2	0.3 ± 0.1	0.7 ± 0.7
97.027	$C_5H_5O_2^+$	Furfural (furancarboxaldehyde) [3,5]	Aldehyde	1.7 ± 1.8	0.7 ± 0.5	1.2 ± 0.9
98.059	$C_5H_8ON^+$	Dimethyl-oxazole [5,8]	Pyrrole	0.2 ± 0.1	0.2 ± 0.1	0.2 ± 0.1
99.041	$C_5H_7O_2^+$	Furfuryl alcohol/a-angelica lactone [4,5]	Furan/ester	0.5 ± 0.3	0.4 ± 0.3	0.7 ± 0.4
99.078	$C_6H_{11}O^+$	Hexenal/methyl-pentenone [1,5]	Aldehyde/ketone	1.6 ± 0.8	1.3 ± 0.7	1.8 ± 1.0
101.058	$C_5H_9O_2^+$	Dihydro-2-methyl-3(2H)-furanone/2,3-pentanedione/pentanedione/methyl-tetrahydrofuranone [3,4,5]	Furan/ketone	5.5 ± 2.6	4.7 ± 3.1	7.9 ± 5.1
103.075	$C_5H_{11}O_2^+$	2,3-methylbutanoic acid/2-hydroxy-3-pentanone/3-methylbutanoic acid [3,4,5,6]	Acid/ketone	85.0 ± 58.7	74.4 ± 60.3	104.6 ± 85.7
104.049	$C_7H_6N^+$	Benzonitrile [3]	Nitrile	1.5 ± 0.6	1.3 ± 1.0	1.5 ± 0.9
105.069	$C_8H_9^+$	Styrene/phenyletanol fragment [1]	Alkene/fragment	10.8 ± 7.8	13.3 ± 11.8	13.5 ± 11.9
107.053	$C_7H_7O^+$	Benzaldehyde/benzyl alcohol [1,2]	Aldehyde/alcohol	19.0 ± 6.6	20.2 ± 8.8	19.9 ± 13.6
107.083	$C_8H_{11}^+$	Terpene fragment [1,7]	Fragment	20.6 ± 6.6	21.6 ± 9.1	21.5 ± 11.7

(*continued on next page*)

Table 1 (*continued*)

109.065	$C_7H_9O^+$	4-methylphenol/benzyl alcohol [3]	Alcohol	0.6 ± 0.5	0.8 ± 1.0	2.2 ± 2.4
110.060	$C_6H_8ON^+$	Formil-methylpyrrole/acetylpyrrole/2-acetylpyrrole [1,2,4]	Pyrrole	0.3 ± 0.3	0.3 ± 0.3	0.6 ± 0.5
111.046	$C_6H_7O_2^+$	2-acetylfuran/5-methylfurfural [4,5,6]	Furan	0.5 ± 0.4	0.3 ± 0.2	0.5 ± 0.3
111.078	$C_7H_{11}O^+$	2,3-dimethyl-2-cyclopente-1-one [4,5]	Ketone	0.8 ± 0.3	1.0 ± 0.5	1.2 ± 0.8
112.076	$C_6H_{10}ON^+$	Trimethyloxazole [4,5]	Furan	0.2 ± 0.2	0.2 ± 0.1	0.3 ± 0.3
113.053	$C_6H_9O_2^+$	Methylfurfuril alcohol/dimethylfuranone/methylciclopentanedione/cyclotene [4,5,7]	Furan/ketone	0.4 ± 0.2	0.3 ± 0.2	0.6 ± 0.3
113.093	$C_7H_{13}O^+$	Heptenone [1,2,4]	Ketone	1.3 ± 0.5	1.2 ± 0.6	1.6 ± 0.9
115.075	$C_6H_{11}O2^+$	4-methyltetrahydro-(2H)-pyran-2-one [4,5,7]	Ketone	1.5 ± 1.2	2.0 ± 2.1	3.6 ± 3.5
115.109	$C_7H_{15}O^+$	Heptanal/2-heptanone[1,2,3,5]	Aldehyde/ketone	7.4 ± 4.2[a]	20.2 ± 13.6[b]	12.4 ± 8.6[ab]
			l		b	
117.088	$C_6H_{13}O_2^+$	Hexanoic acid/C6 ester [1,3,5]	Acid/ester	9.0 ± 6.7	7.1 ± 4.7	10.9 ± 8.6
121.084	$C_7H_9N_2^+$	2-ethenyl-6-methylpyrazine/6,7-dihydro-(5H)-cyclopentapyrazine [1,3,5]	Pyrazine	32.2 ± 28.9	40.9 ± 64.6	61.2 ± 66.1
123.088	$C_8H_{11}O^+$	2-phenylethanol [3,9]	Alcohol	1.5 ± 1.3	1.5 ± 1.2	3.6 ± 3.5
125.058	$C_7H_9O_2^+$	Guaiacol/methyl-benzenediol/furyl acetone [1]	Ether	0.5 ± 0.3	0.5 ± 0.4	0.9 ± 1.0
125.093	$C_8H_{13}O^+$	Butylfuran/methylpropylfuran/octadienone/trimethylciclopentenone/alkylfurans/E-2-octenal [2,4,5]	Furan/ketone/aldehyde	0.6 ± 0.3	0.6 ± 0.3	1.0 ± 1.0
126.090	$C_7H_{12}ON^+$	Acethyl-dimethylpyrrole/alkyloxazole [1,5]	Pyrrole	0.1 ± 0.1	0.1 ± 0.1	0.2 ± 0.1
129.088	$C_7H_{13}O2^+$	Pentenylacetate/heptanedione [4,5,7]	Ester/ketone	0.6 ± 0.3	0.7 ± 0.4	1.1 ± 0.8
131.105	$C_7H_{15}O_2^+$	Ethyl valerate/heptanoic acid/C7 ester [1,3,5]	Ester/acid	9.6 ± 8.9	10.1 ± 11.4	7.6 ± 8.2
136.113	$C_9H_{14}N^+$	Butyl-pyridine/ethyl-propyl pyridine [1,5]	Pyridine	0.3 ± 0.2	0.2 ± 0.1	0.6 ± 0.8
137.129	$C_{10}H_{17}^+$	Various monoterpenes [1]	Terpene	7.4 ± 11.2	8.9 ± 5.1	23.7 ± 36.7
138.057	$C_7H_8O_2N^+$	Pyridinecarboxylic acid methyl ester [1,5]	Acid	0.1 ± 0.1	0.1 ± 0.1	[2,4,5] ± 0.3
139.108	$C_9H_{15}O^+$	E,E-2,4-nonadienal/E-2-nonenal [2,4,5]	Aldehyde	2.0 ± 0.5	2.0 ± 0.7	2.6 ± 1.5
143.141	$C_9H_{19}O^+$	Nonanal [3]	Aldehyde	4.7 ± 3.3[a]	8.1 ± 4.6[b]	7.4 ± 7.3[ab]
145.122	$C_8H_{17}O_2^+$	Ethyl-exanoic acid/C8 ester [1,5]	Acid/ester	1.3 ± 0.9	1.6 ± 1.5	1.8 ± 1.7
148.078	$C_9H_{10}ON^+$	1-furfurylpyrrole [4,5]	Pyrrole	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.0
153.126	$C_{10}H_{17}O^+$	E,E-2,4-decadienal [5,6,7]	Aldehyde	0.4 ± 0.2	0.4 ± 0.2	0.7 ± 0.5
155.143	$C_{10}H_{18}O^+$	Linalool [8]	Alcohol	1.0 ± 0.4	0.9 ± 0.5	1.3 ± 0.6
*159.137	$C_9H_{19}O_2^+$	Nonanoic acid/C9 ester [1,5]	Acid/ester	0.9 ± 0.5	0.9 ± 0.6	1.2 ± 0.9
*166.092	$C_9H_{12}O_2N^+$	Methyl-pyrrolyl-butanedione [1,5]	Pyrrole	0.1 ± 0.1	0.1 ± 0.1	0.1 ± 0.1
*167.114	$C_{10}H_{15}O_2^+$	Ethyl-dimethoxy-benzene [1,5]	Ether	0.2 ± 0.1	0.2 ± 0.1	0.2 ± 0.1
*169.191	$C_{13}H_{25}^+$	Dodecane [2]	Alkane	0.6 ± 0.2	0.6 ± 0.2	0.7 ± 0.2
*191.160	$C_{13}H_{19}O^+$	Beta-damascenone [5]	Ketone	0.1 ± 0.1	0.1 ± 0.1	0.1 ± 0.1

References: [1] Yener et al. (2016); [2] Huang and Barringer (2011); [3] Acierno et al. (2016); [4] Romano et al. (2014); [5] Yener et al. (2016); [6] Özdestan et al. (2013); [7] Yener et al. (2014), [8] Afoakwa et al. (2008); [9] Rodriguez-Campos et al. (2011).

*Compounds tentatively identified but with concentration lower than 1 ppb.

**Different superscripts indicate significant differences (Kruskal Wallis test and Mann-Whitney post hoc test, $p < 0.05$).

(Rodriguez-Campos, Escalona-Buendía, Orozco-Avila, Lugo-Cervantes, & Jaramillo-Flores, 2011). High alcohols content is desirable, however, their concentration decreases during drying and roasting because of the high temperature (Aprotosoaie, Luca, & Miron, 2016).

Among organic acids, acetic acid (m/z 61.029), formed during the fermentation of cocoa beans by acetic acid bacteria (Frauendorfer & Schieberle, 2008), methylpropanoic acid (m/z 89.060), and 2/3-methylbutanoic acid (m/z 103.75) were identified in this study. Kirchhoff, Biehl, and Crone (1989) proposed that 3-methylbutanoic acid is formed at the end of the fermentation from the amino acid leucine by aerobic putrefactive bacteria. Propanoic, hexanoic, and nonanoic acids found by Rodriguez-Campos et al. (2011), during the cocoa fermentation process were also detected in this study. Acids are typical of compounds of raw and fermented cocoa beans; their concentration decrease during drying and roasting.

Among the aldehydes, it was possible to identify acetaldehyde (m/z 45.033, from alanine), propanal (m/z 59.048), methyl-butenal (m/z 85.063), 2/3-methylbutanal (m/z 87.077, from isoleucine and leucine), benzaldehyde (m/z 107.053, from phenylalanine), and heptanal (m/z 115.109). Some of these compounds were identified as dominant odour-active volatiles in cocoa mass (Afoakwa et al., 2008). Aldehydes are formed by Strecker degradation of free amino acids during roasting. However, low concentrations of aldehydes may arise even during fermentation and drying (Aprotosoaie et al., 2016) as highlighted by the results in Table 1.

Within the ketones group in this study, acetone (59.048), 2,3-butandione (87.045), hydroxybutanone (89.060), 2,3-pentanedione (101.058) and 2-heptanone (115.109) were determined. 2,3-butandione has been found in fermented and dry beans (Rodriguez-Campos et al., 2011) and some of the aforementioned ketones were also found in dark chocolates by Afoakwa et al. (2008).

Pyrazines are usually products of roasting but several studies

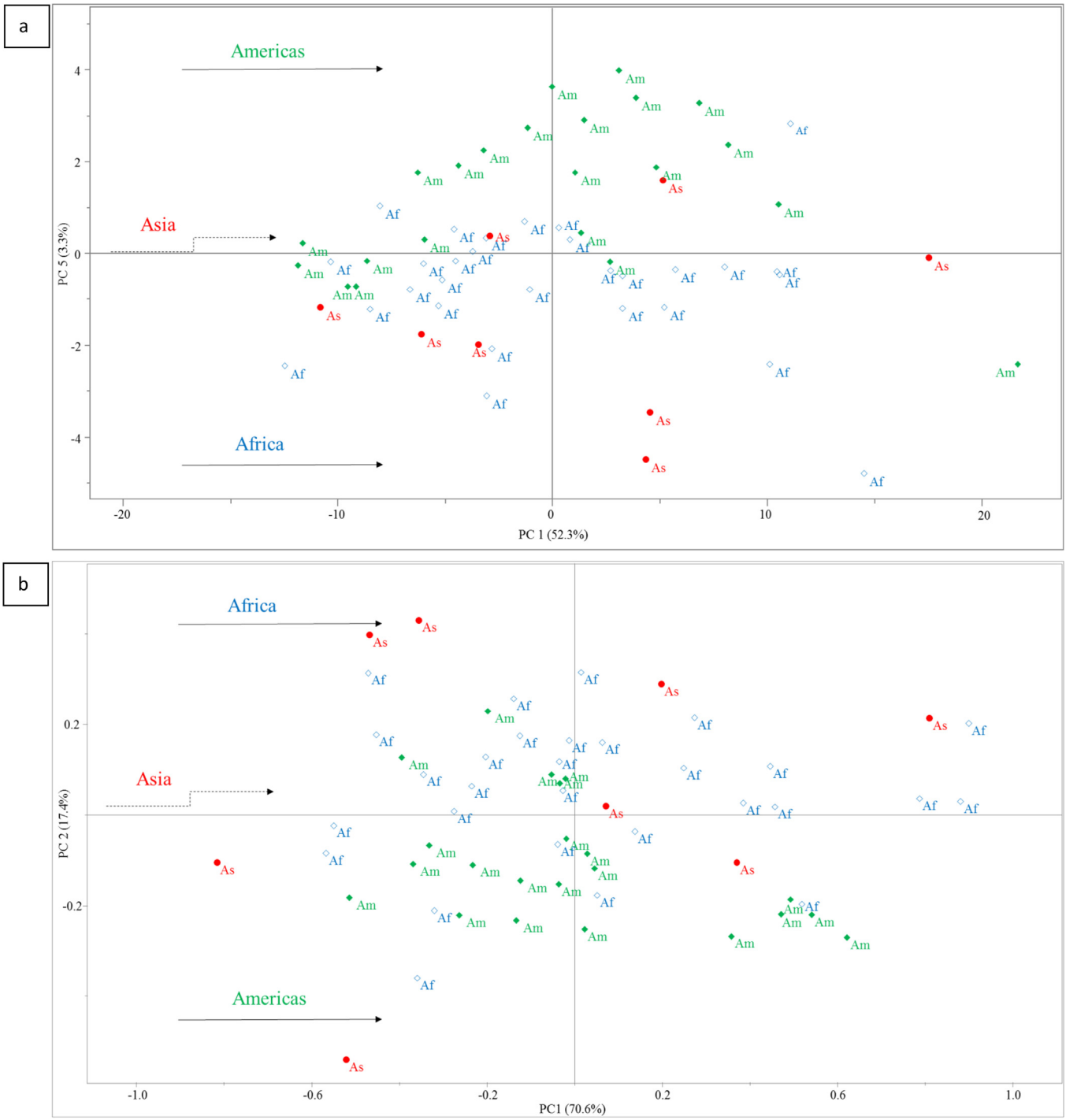

Fig. 2. a) PCA of normalized and auto-scaled PTR-QiToF-MS data of cocoa beans from single geographical origins: Africa (Af), America (Am) and Southeast Asia (As). The arrows indicate the trend of the samples distribution. The broken arrow indicate the tendency to overlap of Asian samples. b) PCA of mean centred and SNV visual camera system data of cocoa beans from single geographical origins: Africa (Af), America (Am) and Southeast Asia (As). The arrows indicate the trend of the samples distribution. The broken arrow indicate the tendency to overlap of Asian samples.

showed the presence in the unroasted beans, too (Frauendorfer & Schieberle, 2008). In this study, two pyrazines were identified: 2-ethenyl-6-methylpyrazine (m/z 121.084) and 6,7-dihydro-(5H)-cyclopentapyrazine (m/z 121.084). In Table 1, it is possible to highlight the presence of esters. This class of compounds, released during fermentation, is the second important class for cocoa and chocolate flavour after pyrazine and has been found both in unroasted and roasted cocoa beans (Aprotosoaie et al., 2016).

Moreover, it was possible to tentatively identify typical compounds of heated foods such as furan (Aprotosoaie et al., 2016; Ziegleder, 2009). Also, previous studies found these compounds as contributors to the aroma of unroasted cocoa beans (Frauendorfer & Schieberle, 2008).

3.1.2. Quantitative geographical comparison

3.1.2.1. Cocoa beans geographical origins: univariate comparison. The final data set (59 samples × 114 masses) was subjected to Kruskal-Wallis tests to detect significant differences in mass intensities with respect to the origins of the beans (Af, Am, As). The masses showing

statistical differences ($p < 0.05$) according to the rank-based test are listed in Table 1. Shapiro Wilk tests were conducted to examine normality within the groups. The data were not normally distributed (Shapiro-Wilk test < 0.05), because of this, significance of differences between the samples was evaluated by non-parametric tests Kruskall Wallis test and Mann-Whitney post-hoc.

None of the masses in Table 1 showed distinct concentrations allowing discrimination of the three continents simultaneously. What stands out in Table 1 is the possibility to statistically differentiate from Americas and Africa while the means of the Southeast Asian samples lie between the other two continents. The higher variation of the Southeast Asian samples (Table 1) compared to the other continents highlights the difficulties to statistically differentiate these cocoa beans. This is consistent with the results of previous work on the same sample set (Marseglia et al., 2016), highlighting that Asian samples have intermediate characteristics between African and South American cocoa beans; a strong variability was observed inside the Asian group. These data are also consistent with the fact that, in some regions of Asia such as Indonesia, cocoa fermentation is short or even omitted (Rohsius, Matissek, & Lieberei, 2005), inducing a large differentiation caused by fermentation level, overpassing that of geographic origin. For example, in a previous work on peptide pattern of cocoa beans, samples from Indonesia (Flores-Indonesia and Sulawesi-Indonesia), also analysed in the current study, showed the same peptide pattern as unfermented slaty beans, indicating that they were poorly fermented, while other Asian samples as those from Papua New Guinea had the highest amount of peptides, indicating an extensive fermentation (Caligiani et al., 2016).

Nevertheless, it is still possible to find information related to the continent but it is restricted to a limited number of masses (Table 1), similarly to the results of Hernandez and Rutledge (1994). Within the compounds in Table 1, the terpene fragment can be informative of the raw material as terpenes naturally occur in flowers and fruits. A study related to coffee beans confirmed that this compound could resist the roasting (Yener et al., 2016), indicating terpene as a probable geographical indicator for chocolate supply chain at least till the roasting step. Mass 115.109 tentatively identified as heptanal/2-heptanone was found to be also relevant in South American chocolate discrimination (Acierno et al., 2016), supporting the rank test results of the cocoa beans in this study and indicating this mass as a potential geographical indicator for products from Americas.

3.1.2.2. Cocoa beans geographical origins: multivariate comparison. For further evaluation of the data, a PCA was performed using the full volatile profiles obtained by PTR-QiToF-MS. The investigation of the total volatile profiles can help in extracting more information (Capuano & Van Ruth, 2012).

The data matrix (59 samples × 114 masses) was auto-scaled and normalized prior to PCA; in Fig. 2a the scores plot corresponding to PC1 versus PC5 is shown. In order to get geographical clusters, mass 43.019 and 61.028 were excluded. The distribution of the samples revealed a tendency to group cocoa beans according to their geographical origins (Fig. 2a). As shown in Fig. 2a, the variability within samples is mainly represented in PC1, while PC5 allows a geographical differentiation. A separation between samples from Africa and Americas is visible. On the other hand, a not clear trend is visible for the Southeast Asian cocoa beans: these do not form a separate cluster but are mostly grouped together with African cocoa beans (Africa - Southeast Asia cluster). A similar trend was also found by Marseglia et al. (2016), analysing the same sample set using HR MAS ^{1}H NMR and HR ^{1}H NMR. Mass 43.019 and 61.028 were bringing a lot of variability within the samples reducing the visualization of a geographical trend. In Fig. 3 it is possible to notice the differences in concentration of masses 43.019 and 61.028 between samples in general and also between samples from the same country. Wide variation in acetic acid values in samples from the same country was found also by Jinap and Dimick (1990). The main explanation for mass 61.028 variation is difference in fermentation conditions (Jinap & Dimick, 1990). High acidity can be partly due to long fermentation. Looking at the concentration of mass 61.028 for the samples from Papua New Guinea and Indonesia (Flores-Indonesia and Sulawesi-Indonesia) it is possible to support what stated in paragraph 3.1.2.1 about the Southeast Asian variability. In fact, the higher concentration of acid acetic for the sample from Papa New Guinea can be indication of well fermented beans while the low acid acetic concentration for the Indonesian samples can be informative of unfermented beans. Long fermentation steps determine higher release of flavour precursors (Afoakwa et al., 2008; Counet et al., 2004). In Fig. 3 the fermented and dried samples with higher concentrations of m/z 61.029 show a higher total volatile headspace concentration compared to the other cocoa beans highlighting consistency with the hypothesis about Papua New Guinea samples. Counet et al. (2004) reported too that, the samples from Papa New Guinea had been subjected to longer fermentation and presented higher concentration of total aroma. A trend between m/z 61.029 and the total ion count (not including m/z 43.019 and m/z 61.029) is evident for all the cocoa beans (Fig. 1, Supplementary material). This is indicative of a possible link between acid acetic -and fermentation-flavour formation already before roasting.

These results underline that at this stage of production (fermentation and drying) the impact of the geographical origin on the volatile profile is probably influenced by the variability related to differences in production within a country and between countries on the same continent. Excluding mass 43.019 and 61.028 the processing influence is partially decreased. However, the variability within Southeast Asian beans is still affecting the spread of the samples. As noted by Marseglia et al. (2016), analysing the same sample set using HR MAS 1H NMR and HR 1H NMR, the difficulties in differentiate the Southeast samples from the other origins are most likely related to the recentness of extensive cocoa cultivation, to the genetically similarity of this cocoa with cocoa hybrids of African countries and to the differences in fermentation that add further variability to the cocoa from these origin (Marseglia et al., 2016).

Investigating the highest loading values on PC1 and PC5 to understand which mass peaks contribute more to the PCs, it is possible to notice that the variability on PC1 is mainly related to furans, ketones, acids, and esters while alcohols and aldehydes contribute more to PC5. The high impact of furans on PC1 could indicate the effect of different sample-specific drying process not specified by the supplier.

Fig. 4 shows the 50 highest loading values related to the PC5. PC5 has been chosen as it is the one affected by the geographical information. Taking into account the sample distribution in Fig. 2a, the cluster of beans from Americas is linked with positive loadings on PC5 while the African-Southeast Asian one with negative loadings. The compounds showing statistical differences according to the origin in Table 1 are included in the most representative loadings values supporting an effective role in geographical differentiation. Specifically, heptanal/2-heptanone (m/z 115.109) and nonanal (m/z 143.141) can be confirmed as significant for South American samples. Several masses related to the beans from Americas have been found relevant also for the classification of chocolate produced with South American beans (Acierno et al., 2016). Within them is it possible to underline some compounds tentatively identified as benzonitrile (m/z 104.049), benzaldehyde (m/z 107.053), heptanal/2-heptanone (m/z 115.109). Heptanal/2-heptanone was the compound with a bigger role in the definition of the fifth PC in the positive range. Specifically, it was abundant in Peruvian samples. 2-heptanone has been found in beans fermented and dried at specific conditions (Rodriguez-campos, Escalona-buendía, Contreras-ramos, & Orozco-avila, 2012), for this reason m/z 115.109 can be indicative of specific farming practice of this area/continent and indirectly linked with the provenance of the beans. On the other hand, the terpene fragment (m/z 107.083) can be indicative of geographical origin (Yener et al., 2016). Concerning the negative loading

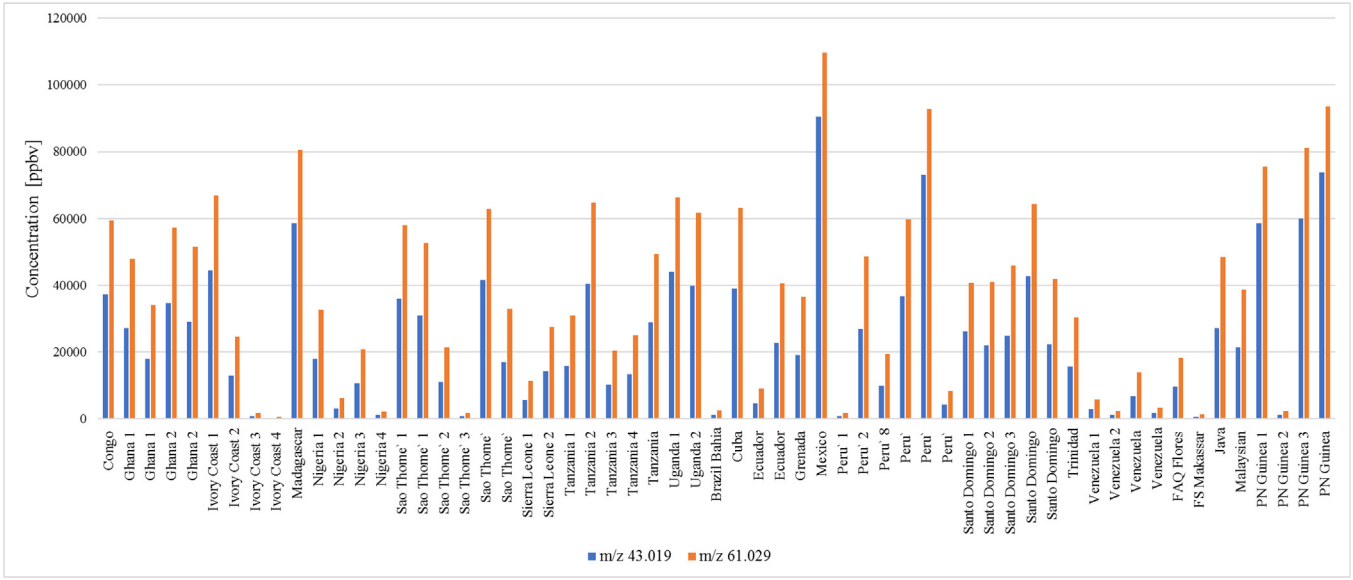

Fig. 3. Concentrations of mass 61.029 and 43.019 for each sample.

representative of the African – Southeast Asian samples, 2-methylpropanal (m/z 73.068) was the highest. Specifically, it was abundant in Madagascar and Papua New Guinea samples. Moreover, regarding the African cluster masses 41.038 and 45.033 were relevant for the classification of chocolate made from African beans (Acierno et al., 2016).

The majority of the compounds presenting loadings that are associated with African - Southeast Asian samples can be related to production steps. Several studies indicate compounds such as 2,3 butanedione (m/z 87.045), furfural (m/z 97.027) and hexanoic acid (m/z 117.088) as technological marker because they are linked with certain conditions of fermentation and drying (Magagna et al., 2017; Rodriguez-campos et al., 2012). These results are indicative of differences in farmer practices within Africa and South America and confirm similarity between African and Southeast Asian cocoa beans included in this set of sample. Specifically, 2(5)-furanone (m/z 85.026), furfural (m/z 97.027) and dihydro-2-methyl3(2H)-furanone (m/z 101.058)

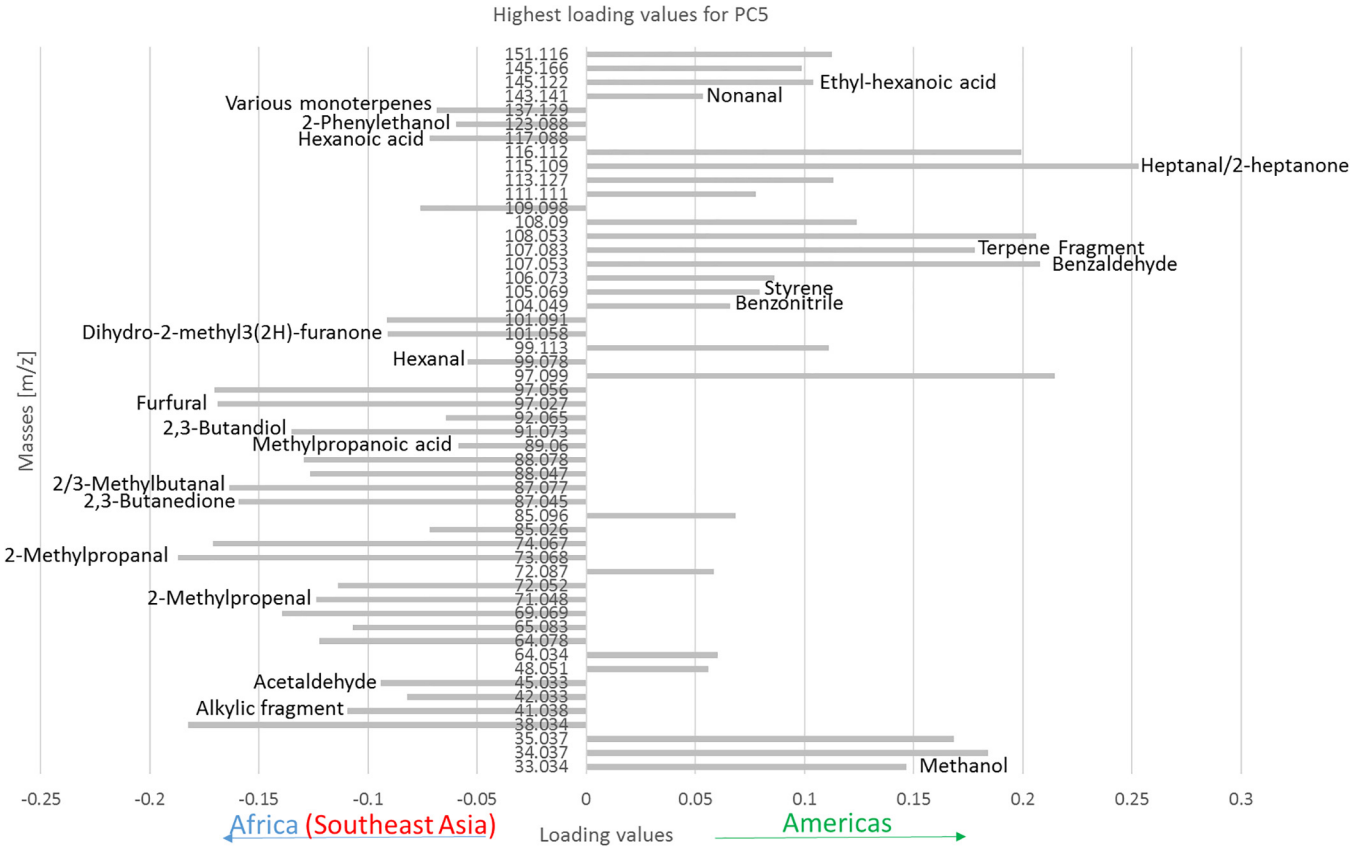

Fig. 4. Fifty highest positive and negative loading values associated with PC5. Negative values are related to African samples while positive values are related to the samples from the Americas.

could be indication of similarity in drying conditions. Within the other compounds with negarite loadings, 2/3 methylbutanal (m/z 87.077), 2 methylpropanoinc acid (m/z 89.060) and 2-phenylethanol (m/z 123.088) are reported as key aroma marker (Magagna et al., 2017).

3-Methylbutanal (m/z 87.077) and benzaldehyde (m/z 107.053) were found relevant by Hernandez and Rutledge (1994) to differentiate cocoa masses according to the geographical origin; which supports the influence of aldehydes on geographical distinction at PC5. Regarding 3-methylbutanal Uganda and Tanzania showed the higher concentrations. together with Madagascar ad Papua New Guinea as found in previous studies (Hernandez & Rutledge, 1994).

Therefore, the geographical information found by Hernandez and Rutledge (1994) could be further described as 3-methylbutanal relevant for the differentiation of African and Papua New Guinea samples from the rest and benzaldehyde (m/z 107.053) relevant for South American samples. Within the negative loading, 2-phenylethanol (m/z 123.088) can be found already in the pulp of the cocoa beans (Chetschik et al., 2018); it could represent a link with the raw material. Specifically, the Southeast Asian samples show the higher concentrations. The cocoa beans from Southeast Asia showed high concentration for 2,3-butanediol (m/z 91.073) as well; these masses (m/z 123.088 and 91,073) could be informative of this group even if it was difficult to differentiate these beans from the African ones. Within the African - Southeast Asian samples, is it important to highlight the monoterpenes (m/z 137.129) as possible geographical markers (Yener et al., 2016). Even if m/z 137.129 is relevant for the African - Southeast Asian cluster, it is mostly concentrated in the Asian beans from Java and Papa New Guinea. These results underline the possibility to characterise the Southeast Asian sample based on specific masses, however, more evidence are needed.

3.2. Machine vision: HSI

The cocoa beans samples were analysed by HSI and the spectra of each cocoa bean sample are shown in Fig. 5a. Samples are coloured and highlighted according to the continent (Af, As, Am). Fig. 5a shows that the spectra diverge after wavelength 625 nm. The patterns are fairly similar in the visible range, i.e. up to range 600–700 nm. This range reflects differences in colour, which obviously not differ with provenance. In the remaining Near Infrared range, patterns start to divert. In the 700–1000 nm range, the third overtone region, differences in composition are reflected, especially those presented as O–H and C–H bonds. The groups of South American and Southeast Asian samples showing higher reflectance compared to the African cocoa bean group in the higher end wavelength range.

Spectroscopic data may not be interpreted directly because of the effect of a large number of factors, such as a high number of variables, slight scattering, instrumental drift, baseline shift, and slope variation, caused by differences in particle size and physical properties of the samples. Therefore, the use of chemometrics is required to amplify the information of interest and lessen the undesirable information in the spectra (Lohumi et al., 2015). In this study, PCA was used to explore the visual camera system data (Fig. 3b). The matrix was mean centred and SNV (Standard Normal Variates) applied prior to PCA. In Fig. 5b, the cocoa beans profiles are shown after data treatment.

As presented in Fig. 2b, the first two principal components, which explain 88% of the total variability in the data, display a grouping of cocoa beans according to their geographic origin. PC2 is mainly affected by the geographical information while the variability within beans is mainly represented in PC1, probably affected by an unknown phenomenon such as lighting. A trend in the distribution of African and American cocoa beans is visible whereas a not so clear trend is shown by the cocoa beans from Southeast Asia.

The sample distribution of the spectroscopy data followed the same trend of the PTR-QiToF-MS indicating a possible link between the two measurements. Nevertheless, also in this case a similarity with the results showed by Marseglia et al. (2016), analysing the same sample set using HR MAS 1H NMR and HR 1H NMR, can be underlined. HR MAS ^{1}H NMR fingerprint of cocoa beans combined with chemometrics discriminates African and American cocoa samples, based mainly on the fatty acids, acetate and saccharidic components. Moreover, HR ^{1}H NMR data showed citrate and formiate as possible markers for African samples, while amino acids, caffeic acid, caffeine and epicatechin as possible markers for American cocoa samples. Taking into account the groups (CH_3-CH_2-CH) used by Caligiani, Acquotti, Cirlini, and Palla (2010) to identify these compounds it is possible to relate them to the spectral range in Fig. 5 and give a possible explanation to the similar trend in the sample distribution. Because of low evidences for the Southeast Asian samples the discussion is mainly focused on the African and American samples.

In Fig. 5b the wavelengths are highlighted according to their contribution on PC2. Taking into account the sample distribution in Fig. 2b, it is possible to state that the American cluster is linked with negative loading (highlighted area in Fig. 5b) on PC2 and the African group with positive loadings (not highlighted area in Fig. 5b). The wavelength range characteristic for the samples from the Americas include the wavelength range related to CH_3 groups (from ~700 nm to ~730 nm and from ~870 nm to ~910 nm); this group could be linked with particular organic compound groups, possibly fatty acids, etc. according to the NMR results. Amino acids may also play a role (RNH_2: from ~770 nm to ~830 nm) in the distinction of the beans from the Americas. Moreover, sugars such as glucose and fructose may be reflected in the spectral data and are relevant in discriminating the African cocoa beans according to the NMR studies.

3.3. Machine olfaction and machine vision comparison

The machine olfaction and machine vision results were correlated (Pearson correlation tests). Spectral range 625–1000 nm and the majority of the VOCs showed significant correlations (Fig. 2, Supplementary material), although this was even more pronounced for the 800–1000 nm range and lower molecular weight VOCs. The VOCs are unlikely to alter the HSI profiles by themselves, due to their very low concentrations in the beans. However, differences in the major components in the beans, which are reflected in the HIS spectra, will certainly affect formation of VOCs in the beans. For instance, the pronounced correlation of VOC with amine group wavelength range from ~770 nm to ~830 nm, can underline the significant role of amino acid in flavour development. Both VOCs and HSI spectra are closely related to local production circumstances and fermentation conditions. For instance 2-methylpropenal (m/z 71.048), 2,3-butanedione (m/z 87.045), acetic acid (m/z 61.029) and the furan class (m/z 71.048) and their correlating ranges may link specifically to the local/country fermentation/drying processes. Although the correlation between the machine olfaction and vision measurements is interesting, there is most likely not a direct relationship but only an indirect one. Further research is required to examine this correlation, which is not the aim of the current study. The correlation shows though, that there are multiple, linking characteristics that allow distinction between the cocoa beans produced in different geographical areas.

4. Conclusions

As a whole the results obtained in the present work point out the potential complementary use of machine olfaction by PTR-QiToF-MS and machine vision by HSI to have a fast screening of the cocoa beans geographical origin.

The PTR-QiToF-MS and HSI fingerprints showed the same tendency in clustering South American and African samples. Southeast Asian samples showed great variability with both approaches. Most likely this variation is due to differences in fermentation levels. They appear to influence the metabolic profiles of cocoa extensively. Machine olfaction and machine vision characterization provided a similar degree of

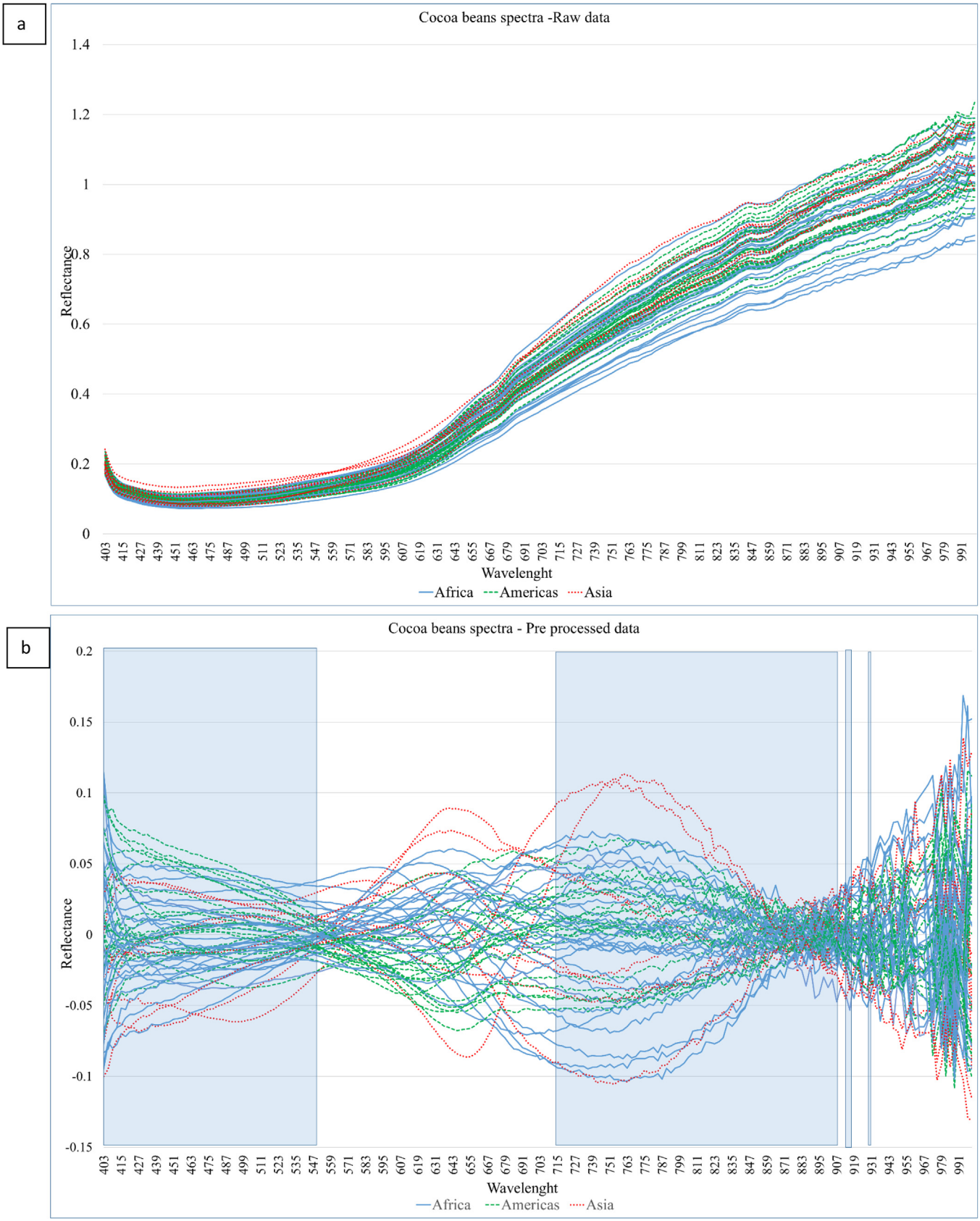

Fig. 5. a) HSI spectra of the cocoa beans. Samples are coloured and underlined according to the continent of provenance: Africa, Americas, Asia. b) Cocoa beans HSI spectra after mean centering and SNV pre-processing. The highlighted areas correspond to the wavelengths relevant for the PCA distribution of the samples from Americas while the non-highlighted areas are important with regard to the distribution of the African samples.

sample separation. Significant correlations highlighted the possibility to use these techniques as complementary rapid tools, which may be further developed for use in practical settings.

The cocoa beans volatile profiles measured using the PTR-QiToF-MS showed geographical similarities with the chocolate volatile profiles measured previously indicating the possibility to develop this technique further for assessments of the geographical origin along the whole cocoa-chocolate supply chain.

Declaration of interests

None.

Acknowledgments

This study has been funded by PIMMS (Proton Ionization Molecular Mass Spectrometry) ITN which is supported by the European Commission's 7th Framework Program under Grant Agreement Number 287382. The authors wish to thank dr. S. Yener for sharing knowledge about PTR-ToF-MS and offering assistance with PTR-ToF-MS data analysis, and N. Liu for offering support in the data analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2019.03.095.

References

Acierno, V., Liu, N., Alewijn, M., Stieger, M., & Van Ruth, S. M. (2019). Talanta which cocoa bean traits persist when eating chocolate ? Real-time nosespace analysis by PTR-QiToF-MS. *Talanta, 195*(July 2018), 676–682. https://doi.org/10.1016/j.talanta.2018.11.100.

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

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. https://doi.org/10.1080/10408390701719272.

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. https://doi.org/10.1111/1541-4337.12180.

Caligiani, A., Acquotti, D., Cirlini, M., & Palla, G. (2010). 1H NMR study of fermented cocoa (Theobroma cacao L.) beans. *Journal of Agricultural and Food Chemistry, 58*(23), 12105–12111. https://doi.org/10.1021/jf102985w.

Caligiani, A., Marseglia, A., Prandi, B., Palla, G., & Sforza, S. (2016). Influence of fermentation level and geographical origin on cocoa bean oligopeptide pattern. *Food Chemistry, 211*, 431–439. https://doi.org/10.1016/j.foodchem.2016.05.072.

Campbell-Sills, H., Capozzi, V., Romano, A., Cappellin, L., Spano, G., Breniaux, M., ... Biasioli, F. (2016). Advances in wine analysis by PTR-ToF-MS: Optimization of the method and discrimination of wines from different geographical origins and fermented with different malolactic starters. *International Journal of Mass Spectrometry, 397–398*, 42–51. https://doi.org/10.1016/j.ijms.2016.02.001.

Capuano, E., & Van Ruth, S. M. (2012). QA : Fraud control for foods and other biomaterials by product fingerprinting. In I. Akyar (Ed.). *Latest research into Qualitycontrol* (pp. 111–143). Rijeka, Croatia: Intech.

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

Cho, J. S., Bae, H. J., Cho, B. K., & Moon, K. D. (2017). Qualitative properties of roasting defect beans and development of its classification methods by hyperspectral imaging technology. *Food Chemistry, 220*, 505–509. https://doi.org/10.1016/j.foodchem.2016.09.189.

Counet, C., Ouwerx, C., Rosoux, D., & Collin, S. (2004). Relationship between procyanidin and flavor contents of cocoa liquors from different origins. *Journal of Agricultural and Food Chemistry, 52*(20), 6243–6249. https://doi.org/10.1021/jf040105b.

Elmasry, G., Iqbal, A., Sun, D. W., Allen, P., & Ward, P. (2011). Quality classification of cooked, sliced turkey hams using NIR hyperspectral imaging system. *Journal of Food Engineering, 103*(3), 333–344. https://doi.org/10.1016/j.jfoodeng.2010.10.031.

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. https://doi.org/10.1021/jf802098f.

Hernandez, B., Castellote, a. I., & Permanyer, J. J. (1991). Triglyceride analysis of cocoa beans from different geographical origins. *Food Chemistry, 41*(3), 269–276. https://doi.org/10.1016/0308-8146(91)90052-P.

Hernandez, V., & Rutledge, D. N. (1994). Multivariate statistical analysis of gas chromatograms to differentiate cocoa masses by geographical origin and roasting conditions. *Analyst, 119*(June), 1171–1176.

Holzinger, R. (2015). PTRwid: A new widget tool for processing PTR-TOF-MS data. *Atmospheric Measurement Techniques, 8*(9), 3903–3922. https://doi.org/10.5194/amt-8-3903-2015.

Huang, Y., & Barringer, S. A. (2011). Monitoring of cocoa volatiles produced during roasting by selected ion flow tube-mass spectrometry (SIFT-MS). *Journal of Food Science, 76*, C279–C286. https://doi.org/10.1111/j.1750-3841.2010.01984.x.

Icco (International Cocoa Organization). Study on the costs, advantages and disadvantages of cocoa certification. (2012). URL: file:///C:/Users/acier001/Downloads/121105_Study%20on%20the%20costs%20advantages%20and%20disadvantages%20of%20cocoa%20certification_FINAL_Erratum.pdf Accessed 02.09.2018.

Jinap, S., & Dimick, P. S. (1990). Acidic characteristics of fermented and dried cocoa beans from different countries of origin. *Journal of Food Science, 55*(2), 547–550. https://doi.org/10.1111/j.1365-2621.1990.tb06806.x.

Kirchhoff, P.-M., Biehl, B., & Crone, G. (1989). Peculiarity of the accumulation of free amino acids during cocoa fermentation. *Food Chemistry, 31*(4), 295–311. https://doi.org/10.1016/0308-8146(89)90071-X.

Lindinger, W., Hansel, A., Jordan, A., & Hansel, A. (1998). Proton-transfer-reaction mass spectrometry (PTR – MS): On-line monitoring of volatile organic compounds at pptv levels. *International Journal of Mass Spectrometryand Ion Processes, 173*, 191–241.

Liu, N., Parra, H. A., Pustjens, A., Hettinga, K., Mongondry, P., & van Ruth, S. M. (2018). Evaluation of portable near-infrared spectroscopy for organic milk authentication. *Talanta, 184*(December 2017), 128–135. https://doi.org/10.1016/j.talanta.2018.02.097.

Lohumi, S., Lee, S., Lee, H., & Cho, B. K. (2015). A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. *Trends in Food Science and Technology, 46*(1), 85–98. https://doi.org/10.1016/j.tifs.2015.08.003.

Magagna, F., Guglielmetti, A., Liberto, E., Reichenbach, S. E., Allegrucci, E., Gobino, G., ... Cordero, C. (2017). Comprehensive chemical fingerprinting of high-quality cocoa at early stages of processing: Effectiveness of combined untargeted and targeted approaches for classification and discrimination. *Journal of Agricultural and Food Chemistry, 65*, 6329–6341. https://doi.org/10.1021/acs.jafc.7b02167.

Marseglia, A., Acquotti, D., Consonni, R., Cagliani, L. R., Palla, G., & Caligiani, A. (2016). HR MAS1H NMR and chemometrics as useful tool to assess the geographical origin of cocoa beans - Comparison with HR1H NMR. *Food Research International, 85*, 273–281. https://doi.org/10.1016/j.foodres.2016.05.001.

Marseglia, A., Palla, G., & Caligiani, A. (2014). Presence and variation of γ-aminobutyric acid and other free amino acids in cocoa beans from different geographical origins. *Food Research International*. https://doi.org/10.1016/j.foodres.2014.05.026.

Masi, E., Romani, A., Pandolfi, C., Heimler, D., & Mancuso, S. (2015). PTR-TOF-MS analysis of volatile compounds in olive fruits. *Journal of the Science of Food and Agriculture, 95*(7), 1428–1434. https://doi.org/10.1002/jsfa.6837.

Masi, E., Taiti, C., Heimler, D., Vignolini, P., Romani, A., & Mancuso, S. (2016). PTR-TOF-MS and HPLC analysis in the characterization of saffron (Crocus sativus L.) from Italy and Iran. *Food Chemistry, 192*, 75–81. https://doi.org/10.1016/j.foodchem.2015.06.090.

Özdestan, Ö., van Ruth, S. M., Alewijn, M., Koot, A., Romano, A., Cappellin, L., & Biasioli, F. (2013). Differentiation of specialty coffees by proton transfer reaction-mass spectrometry. *Food Research International, 53*(1), 433–439. https://doi.org/10.1016/j.foodres.2013.05.013.

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. https://doi.org/10.1080/10942912.2016.1236270.

Rodriguez-campos, J., Escalona-buendía, H. B., Contreras-ramos, S. M., & Orozco-avila, I. (2012). Effect of fermentation time and drying temperature on volatile compounds in cocoa. *Food Chemistry, 132*(1), 277–288. https://doi.org/10.1016/j.foodchem.2011.10.078.

Rodriguez-Campos, J., Escalona-Buendía, H. B., Orozco-Avila, I., Lugo-Cervantes, E., & Jaramillo-Flores, M. E. (2011). Dynamics of volatile and non-volatile compounds in cocoa (Theobroma cacao L.) during fermentation and drying processes using principal components analysis. *Food Research International, 44*(1), 250–258. https://doi.org/10.1016/j.foodres.2010.10.028.

Rohsius, C., Matissek, R., & Lieberei, R. (2005). Free amino acid amounts in raw cocoas from different origins. *European Food Research and Technology, 222*(3–4), 432–438. https://doi.org/10.1007/s00217-005-0130-y.

Romano, A., Cappellin, L., Ting, V., Aprea, E., Navarini, L., Gasperi, F., & Biasioli, F. (2014). Nosespace analysis by PTR-ToF-MS for the characterization of food and tasters: The case study of coffee. *International Journal of Mass Spectrometry, 365–366*, 20–27. https://doi.org/10.1016/j.ijms.2013.12.001.

Teye, E., Huang, X., Dai, H., & Chen, Q. (2013). Rapid differentiation of Ghana cocoa beans by FT-NIR spectroscopy coupled with multivariate classification. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy, 114*, 183–189. https://doi.org/10.1016/j.saa.2013.05.063.

Yener, S., Navarini, L., Lonzarich, V., Cappellin, L., Märk, T. D., Bonn, G. K., & Biasioli, F. (2016). Monitoring single coffee bean roasting by direct volatile compound analysis with proton transfer reaction time-of-flight mass spectrometry. *Journal of Mass Spectrometry, July*, 690–697. https://doi.org/10.1002/jms.3825.

Yener, S., Romano, A., Cappellin, L., Granitto, P. M., Aprea, E., Navarini, L., ... Biasioli, F. (2015). Tracing coffee origin by direct injection headspace analysis with PTR/SRI-MS. *Food Research International, 69*, 235–243. https://doi.org/10.1016/j.foodres.2014.12.046.

Yener, S., Romano, A., Cappellin, L., Märk, T. D., Sánchez Del Pulgar, J., Gasperi, F., ... Biasioli, F. (2014). PTR-ToF-MS characterisation of roasted coffees (C. arabica) from different geographic origins. *Journal of Mass Spectrometry: JMS, 49*(9), 929–935. https://doi.org/10.1002/jms.3455.

Zhao, J., Chen, Q., Cai, J., & Ouyang, Q. (2009). Automated tea quality classification by hyperspectral imaging. *Applied Optics, 48*(19), 3557–3564. https://doi.org/10.1364/AO.48.003557.

Ziegleder, G. (2009). Flavour development in cocoa and chocolate. In S. T. Beckett, M. S. Fowler, & G. R. Ziegler (Eds.). *Beckett's Industrial Chocolate Manufacture and Use* (pp. 169–191). Chichester, West Sussex, UK: John Wiley & Sons Inc.