



UNIVERSITÀ DI PARMA

ARCHIVIO DELLA RICERCA

University of Parma Research Repository

HR MAS 1H NMR and chemometrics as useful tool to assess the geographical origin of cocoa beans - Comparison with HR 1H NMR

This is a pre print version of the following article:

Original

HR MAS 1H NMR and chemometrics as useful tool to assess the geographical origin of cocoa beans - Comparison with HR 1H NMR / Marseglia, A., Acquotti, D., Consonni, R., Cagliani, L.R., Palla, G.G., Caligiani, A.. - In: FOOD RESEARCH INTERNATIONAL. - ISSN 0963-9969. - 85:(2016), pp. 273-281. [10.1016/j.foodres.2016.05.001]

Availability:

This version is available at: 11381/2809700 since: 2021-11-26T15:17:23Z

Publisher:

Elsevier Ltd

Published

DOI:10.1016/j.foodres.2016.05.001

Terms of use:

Anyone can freely access the full text of works made available as "Open Access". Works made available

Publisher copyright

note finali coverpage

(Article begins on next page)

Manuscript Number:

Title: HR MAS 1H NMR AND CHEMIOMETRICS AS USEFUL TOOL TO ASSESS THE GEOGRAPHICAL ORIGIN OF COCOA BEANS - COMPARISON WITH SOLUTION 1H NMR

Article Type: Research Paper

Keywords: HRMAS 1H NMR; solution 1H NMR; cocoa beans; geographic origin; metabolic profile; chemometrics.

Corresponding Author: Dr. Augusta Caligiani, Ph.D.

Corresponding Author's Institution: University of Parma

First Author: Angela Marseglia

Order of Authors: Angela Marseglia; Domenico Acquotti; Roberto Consonni; Laura Cagliani; Gerardo Palla; Augusta Caligiani, Ph.D.

Abstract: Chocolate and cocoa-based products are among the goods with higher added value. A current trend of the cocoa market is to offer to the consumers' high quality cocoa products, namely mono-origin cocoa. However, a reliable analytical method able to trace the geographical origin of cocoa is lacking. In this work we tested the capability of HR MAS 1H NMR combined with chemometrics to assess the geographical origins of 61 fermented and dried cocoa beans of 23 different cocoa producing Countries from the three major crop-growing areas (Africa, Central/South America, Asia/Oceania). Metabolic profile was determined by HR MAS 1H NMR directly on cocoa powder after the method optimization. The same samples were also subjected to extraction and analysis with solution 1H NMR. HR MAS 1H NMR, as 1H NMR analysis, allowed the simultaneous detection of amino acids, polyalcohols, organic acids, sugars, methylxanthines, catechins, together with lipids, which are not present in the aqueous extract utilized for 1H NMR. The data set obtained is therefore representative of all classes of cocoa compounds. Untargeted HR MAS 1H NMR and 1H NMR datasets were utilized as fingerprint of the samples and elaborated with multivariate statistical methods (PCA). A targeted quantitative approach of selected metabolite was possible only with solution 1H NMR data, because HR MAS 1H NMR does not give reliable quantitative results. All the approaches adopted showed a discrimination of the cocoa origins, even if with different Q2. HR MAS gives the advantages to obtain a very rapid picture of the samples, comprising both lipophilic and hydrophilic components, avoiding any sample manipulation.

Dr. Augusta Caligiani
Department of Food Science
Università degli Studi di Parma
Parco Area delle Scienze 17A
43124-Parma, Italy
tel: +39 0521 905407;
fax: +39 0521 905472;
E-mail: augusta.caligiani@unipr.it

January, 13 2016

Dear Editor of Food Research International,

we would submit the manuscript entitled “**HR MAS ¹H NMR AND CHEMIOMETRICS AS USEFUL TOOL TO ASSESS THE GEOGRAPHICAL ORIGIN OF COCOA BEANS – COMPARISON WITH SOLUTION ¹H NMR** ” for the eventual publication in Food Research International as a research article.

Autors: A. Marseglia, D. Acquotti, R. Consonni, L. Cagliani, G. Palla, A. Caligiani.

We would greatly appreciate Your suggestions about this work.

Thank You in advance,

Yours sincerely

Augusta Caligiani

*Highlights (for review)

HR MAS ^1H NMR was applied for the first time in cocoa characterization

Comparison of HR MAS data with solution NMR data were provided

The combination of NMR data with chemometrics allow to classify cocoa based on origin

HR MAS NMR permit to simultaneously obtain cocoa profile of lipophilic and hydrophilic compounds

1 **HR MAS ¹H NMR AND CHEMIOMETRICS AS USEFUL TOOL TO ASSESS THE**
2 **GEOGRAPHICAL ORIGIN OF COCOA BEANS – COMPARISON WITH SOLUTION ¹H**
3 **NMR**

4 A. Marseglia ‡, D. Acquotti †, R. Consonni\$, L. Cagliani\$, G. Palla ‡, A. Caligiani ‡*

5

6 ‡ Department of Food Science, University of Parma, Parco Area delle Scienze 17/A, 43124, Parma,
7 Italy

8 † Centro Interdipartimentale Misure, University of Parma, Parco Area delle Scienze 23/A, 43124,
9 Parma, Italy

10 \$ Istituto per lo Studio delle Macromolecole, Lab. NMR, CNR, v. Bassini 15, 20133 Milan, Italy.

11 *Corresponding author:

12 Augusta Caligiani

13 Dip. Scienze degli Alimenti

14 Università degli Studi di Parma

15 Parco Area delle Scienze 17A

16 43124-Parma, Italy

17 tel: +39 0521 905407; fax: +39 0521 905472; E-mail: augusta.caligiani@unipr.it

18 **Short Title:** HR MAS ¹H NMR for cocoa bean characterization

19

20

21

22

23

24

25

26

27 **ABSTRACT**

28 Chocolate and cocoa-based products are among the goods with higher added value. A current trend
29 of the cocoa market is to offer to the consumers' high quality cocoa products, namely mono-origin
30 cocoa. However, a reliable analytical method able to trace the geographical origin of cocoa is
31 lacking. In this work we tested the capability of HR MAS ¹H NMR combined with chemometrics to
32 assess the geographical origins of 61 fermented and dried cocoa beans of 23 different cocoa
33 producing Countries from the three major crop-growing areas (Africa, Central/South America,
34 Asia/Oceania). Metabolic profile was determined by HR MAS ¹H NMR directly on cocoa powder
35 after the method optimization. The same samples were also subjected to extraction and analysis
36 with solution ¹H NMR.

37 HR MAS ¹H NMR, as ¹H NMR analysis, allowed the simultaneous detection of amino acids,
38 polyalcohols, organic acids, sugars, methylxanthines, catechins, together with lipids, which are not
39 present in the aqueous extract utilized for ¹H NMR. The data set obtained is therefore representative
40 of all classes of cocoa compounds. Untargeted HR MAS ¹H NMR and ¹H NMR datasets were
41 utilized as fingerprint of the samples and elaborated with multivariate statistical methods (PCA). A
42 targeted quantitative approach of selected metabolite was possible only with solution ¹H NMR data,
43 because HR MAS ¹H NMR does not give reliable quantitative results. All the approaches adopted
44 showed a discrimination of the cocoa origins, even if with different Q2. HR MAS gives the
45 advantages to obtain a very rapid picture of the samples, comprising both lipophilic and hydrophilic
46 components, avoiding any sample manipulation.

47

48

49 **Keywords**

50 HRMAS ¹H NMR, solution ¹H NMR, cocoa beans, geographic origin, metabolic profile,
51 chemometrics.

52

53

54 **1. Introduction**

55 Tracing the geographical origin of foods is a current challenging topic with different implications in
56 food quality, safety and authenticity. Products belonging to specific areas, awarded with PDO
57 (Protected Denomination of Origin) or PGI (Protected Geographical Indication) marks have higher
58 economical values justified by unique quality characteristics strictly linked to the area of
59 production.

60 The increasing consumer awareness of food quality and safety required the development of new
61 analytical techniques for authentication of highly quoted food commodities. In particular, there is an
62 increasing interest of consumers for high quality and mono-origin food products. However, the
63 detection of food geographical origin is not straightforward and generally requires combined
64 analytical approaches. The most useful analytical techniques for tracing the geographical origins of
65 foods are those based on natural abundance isotope variation and elemental concentrations (Kelly,
66 Heaton & Hoogewerff, 2005), but also metabolomics approached based on NMR have gained an
67 important part of the scientific interest about this topic (Consonni & Cagliani, 2010). NMR has
68 been successfully applied in the characterization of the origin of complex food matrices like tomato
69 sauce (Consonni, Cagliani, Stocchero & Porretta, 2009), honey (Consonni, & Cagliani, 2008) and
70 coffee (Consonni, Cagliani & Cogliati, 2012). Almost all the NMR studies reported in literature
71 concerning the assessment of the geographical origins of food, applied solution NMR on food
72 extracts or directly on liquid foods, and rare are the studies utilizing probes that permit the analysis
73 directly on solid or semisolid food samples as HR MAS NMR probe. HR MAS ^1H NMR was
74 recently used to differentiate conventional and specific transgenic common beans (Choze,
75 Alcantara, Alves Filho, Silva, Faria & Liao, 2013), apples grown in organic or conventional
76 cultivation (Vermathen, Marzorati, Baumgartner, Good & Vermathen, 2013), beef of different
77 geographical origins (Shintu, Caldarelli & Franke, 2007) and red chicories (Ritota, Casciani &
78 Valentini, 2013). To the best of our knowledge, HR MAS NMR has never been applied to the study

79 of cocoa beans metabolomics. We have previously applied solution ¹H NMR for the complete
80 characterization of cocoa beans metabolome (Caligiani, Acquotti, Cirlini & Palla, 2010) and for the
81 study of the influence of fermentation level, variety and origin on cocoa composition (Caligiani,
82 Palla, Acquotti, Marseglia, Palla, 2014). In the present work we test the capability of HR MAS ¹H
83 NMR to give information about cocoa beans origin, and, possibly, to classify samples based on
84 geographical areas. Regarding cocoa origin, scarce studies are present in literature about the
85 analytical methods for determining the geographical origin of cocoa products. Some attempts were
86 performed utilizing fatty acids profile (Torres-Moreno, Torrescasana, Salas-Salvadó & Blanch,
87 2015), volatile compounds (Cambrai, Marcic, Morville, Sae Houer, Bindler & Marchioni, 2010),
88 sensorial analysis (Torres-Moreno, Tarrega, Costell & Blanch, 2012), FT-NIR (Teye, Huang, Dai &
89 Chen, 2013).

90 Assessing the provenience of cocoa has becoming more and more important in the last years due to
91 the increasing market of cocoa products of high quality and in particular of mono-origin products.
92 Premium chocolate represents a fast-growing and dynamic market in many parts of the world, with
93 global sales having risen by over 18% within the last years, as consumers are now becoming more
94 knowledgeable about specific cocoa varieties and their origins. Consumers have become especially
95 interested in premium chocolates, chocolates made from single-origin cocoa beans, such as those
96 from Ghana, Ecuador or Venezuela, or organic and fair-trade chocolates (Afoakwa, 2010). The
97 effect of cocoa origin on cocoa and chocolate quality is not easy to be determined, because cocoa is
98 produced through a multistep process involving primary and secondary transformations. The
99 primary transformation is made in the producing countries, which are all developing countries
100 distributed around the Equatorial line, and involves cocoa beans fermentation and drying: local
101 variations in cocoa plant materials, fermentation procedures (generally carried out according to
102 traditional processes) and drying processes lead to a traded good typical of the country of origin
103 (Ardhana & Fleet, 2003). The fermented and dried beans are collected by local buyers from
104 different farmers, often blended and exported up to the cocoa Industries, most of them in the

105 industrialized Countries, which perform the secondary processing comprising roasting, milling,
106 pressing, conching and tempering, to ultimately produce chocolate and cocoa powder. Cocoa
107 beans composition, which is one of the most important factors influencing the taste and flavor of
108 chocolate and cocoa-based products, strongly depends both on the cocoa variety and on the
109 geographical origin. Several papers corroborate that the geographical origin can influence the cocoa
110 and cocoa by-product characteristics (Cambrai et al. 2010). In fact the final composition of raw
111 cocoa beans is linked both to human factors (local farmers procedures) farmer-dependent and
112 physical factors (climatic and soil factors) country dependent.

113 Consumers have no way of tracing the origin of the cocoa used to produce their chocolate to a
114 particular country, much less a particular site of agricultural production. To determine the quality of
115 chocolate and the veracity of labeling, consumers need to be informed of the cocoa production site's
116 country, even more with the rising market of "healthy chocolate" (Cidell & Alberts, 2006).

117 Moreover, from the point of view of the Companies performing the secondary transformation of
118 cocoa is very important to know the geographical origin and the relative cocoa beans characteristic
119 in order to select the appropriate roasting and conching conditions to produce selected chocolates
120 from single origins with a guarantee of maximum acceptability (Torres Moreno et al., 2012). In
121 fact, cocoas of different geographical origins have different organoleptic characteristics and
122 influence chocolate's quality (Kattenberg & Kemmink, 1993; Jinap, Dimick & Hollender, 1995). In
123 previous work (Caligiani et al., 2014) solution ^1H NMR spectra of hydro-alcoholic extracts of a series
124 of cocoa beans with different fermentation levels, varieties and geographical origins were
125 registered, obtaining a grouping of samples based on the fermentation level. In this work, we extend
126 this NMR approach to a large number of traded cocoa beans samples, representative of the entire
127 world cocoa production. NMR spectra were registered both with solution ^1H NMR on cocoa
128 hydroalcoholic extracts and with HR MAS ^1H NMR directly on cocoa powder, and the results were
129 compared. A chemometric approach was applied to obtain information about possible grouping of

130 samples based on the geographical origin and on the metabolites characterizing each group of cocoa
131 beans.

132 **2. Materials and Methods**

133 **2.1 Chemicals.** Methanol, D₂O, CD₃OD and 3-(trimethylsilyl)-propionate-d₄ (TSP, internal
134 standard for NMR analysis) were purchased from Sigma-Aldrich (Milan, Italy).

135 **2.2 Sampling.** 61 fermented and dried cocoa beans samples of Forastero variety from 23 different
136 geographical origins were considered. The sampling is representative of the average world
137 production, (America Africa Asia), and, to the best of our knowledge, it represent the largest cocoa
138 beans collection of different geographical origins considered in the literature. Sampled were kindly
139 provided by Barry Callebaut Belgium, all fruits were harvested in 2012, from March to October,
140 according to the cocoa season specific of the different countries. Table 1 reports the specific
141 Countries of origins considered, the number of different lots for each origin and the months of
142 harvesting. Samples represent cocoa beans traded from countries of origins to the transforming
143 Company, and in most cases, they are constituted of a blend made from cocoa collected in different
144 farms; so they are most representative of the country, instead of the specific farm. All the samples
145 were considered well fermented.

146 Samples were stored frozen and grinded under liquid nitrogen to obtain a fine cocoa powder before
147 analysis.

148 **2.3 Sample preparation for HR MAS ¹H NMR analysis.** Cocoa powder were directly inserted in
149 a 4 mm HRMAS rotor with a 50 μL disposable insert, together with D₂O containing TSP. The
150 optimized amount of cocoa powder/D₂O was 4 mg in 20 μL D₂O.

151 **2.4 HR MAS ¹H NMR acquisition and spectra processing.** Samples were analysed with a Bruker
152 Avance III 400 MHz NMR spectrometer operating at 9.4 T, using a 1H-13C HR-MAS probe with a
153 BVT-3200 temperature controller and BCU cooling unit. The probe resonant circuits were tuned
154 and matched with the sample in place. The “magic angle” (54.7) was adjusted using the 79Br
155 signal from powdered KBr as standard. A spinning rate of 4000 Hz was chosen to move unwanted

156 spinning sidebands outside the spectral range of interest. The sample temperature was set to 308 K.
157 The sample was locked on the deuterium signal from D₂O, and the magnetic field homogeneity was
158 optimized. The receiver gain was checked prior to each acquisition to avoid overloading the
159 spectrometer digitizer with a too intense water signal. Spectra were acquire with 512 scans, a time
160 domain of 16K, a spectral width of 4000 Hz and the relaxation delay was 1s. FIDs were Fourier
161 transformed with FT size of 32K and 0.2 Hz line-broadening factor, phased and baseline corrected
162 with Topspin 2.1 software.

163 **2.5 Sample preparation for solution ¹H NMR analyses.** 200 mg of fermented cocoa beans, finely
164 ground, were extracted with 20 ml distilled water/methanol mixture (8:2 v/v), kept at the boiling
165 point for 10 min under magnetic stirring. Extracts were cooled, filtered, taken to dryness, dissolved
166 in 1 ml D₂O/CD₃OD (8:2 v/v) containing 0.1 % of 3-(trimethylsilyl)-propionate-*d*₄ (TSP), filtered
167 again and transferred in a 5 mm NMR sample tube. TSP was used as internal standard for the
168 quantitative analysis and for chemical shift referencing ($\delta=0$ ppm).

169 **2.6 ¹H NMR acquisition.** ¹H NMR spectra were acquired on a VARIAN-INOVA 600 MHz
170 spectrometer, equipped with a triple resonance inverse probe (HCN), operating at 599.729 MHz for
171 proton. The experiments were carried out with water suppression by low power selective water
172 signal presaturation of 1.5 s. Spectra were acquired at 308 K, with 32K complex points, using a 45°
173 pulse length. 128 scans were acquired with a spectral width of 9611.9 Hz, an acquisition time of 1.3
174 s and a relaxation delay of 3s. Spectra were Fourier transformed with FT size of 64K and 0.2 Hz
175 line-broadening factor, phased and baseline corrected.

176 **2.7 NMR spectra processing.** To analyse the profiles by pattern recognition techniques, NMR
177 spectra were subsequently transferred to MestreNova software and referenced to TSP. Integration
178 pattern was defined by a supervised bucketing integration, choosing buckets manually on all the
179 considered spectra in the overlapped form. Buckets were chosen as large as to compensate the little
180 chemical shifts fluctuation in each single spectrum. In the case of HR MAS 1H NMR, a total of 51
181 signals were chosen in the 1H NMR spectra of cocoa beans, considering both identified and

182 unknown signals (fingerprint dataset). In the case of solution ^1H NMR, it was possible to
183 individuate a total of 91 signals or groups of signals that were considered in the integration pattern,
184 considering also in this case both previously identified and unknown signals (fingerprint dataset).
185 The defined patterns were used for the automatic integration of all the spectra and the integrals were
186 normalized to the total area. For solution ^1H NMR, a second pattern (quantitative dataset) was
187 defined by manually integrating only the identified signals, corresponding to 20 substances, as
188 previously reported (Caligiani et al., 2010); the absolute amounts of the selected substances were
189 calculated utilizing TSP as internal standard.

190 **2.8 Statistical methods.** The NMR integrals datasheets were imported into SIMCA-P+13
191 (Umetrics, Umeå, Sweden) for statistical data analysis. Unsupervised principal component analysis
192 (PCA) was initially used to explore the dataset. Projections to Latent Structures-Discriminant
193 Analysis (PLS-DA) and Orthogonal Projections to Latent Structures-Discriminant Analysis (OPLS-
194 DA) were used as supervised approach to construct a model able to classify cocoa samples
195 according to the geographic origin. Data pretreatment was performed with “Pareto” scaling. T2 and
196 the distance to the model (DModX) tests were applied to check for outliers and evaluate whether the
197 test set samples fall within the model applicability domain. The model was validated with the
198 permutation test on the Y block to safely overcome randomness or overfitting to the model.

199 In some cases spectral filtering has been applied to dataset to remove information not correlated
200 with the response, usually denoted as “structured noise”. In particular, Orthogonal Signal
201 Correction filters (OSC) has been applied to lead more simple models with easier interpretation.

202 **3. Results and Discussion**

203 **3.1 Optimization of sample preparation for HR MAS ^1H NMR**

204 One of the most important advantages in using HR MAS is the possibility to analyse food samples
205 exactly as they are, without sample manipulation, and the only requested operation is the addition to
206 the food sample of a minimal amount of deuterated solvent for the locking of the NMR frequency.
207 In fact, the sample preparation consists simply in inserting the cocoa powder directly in HR MAS

208 rotor disposable insert, together with D₂O containing TSP. However, from the first experiment
209 performed it appeared that the aspect of the HR MAS ¹H NMR spectra varied greatly based on
210 sample preparation, and in particular completely different spectra were obtained utilizing different
211 ratios of cocoa powder vs deuterated solvents. In figure 1 the comparison of HR MAS ¹H NMR
212 spectra obtained with different ratio of cocoa powder: D₂O are reported and compared with high
213 resolution liquid ¹H NMR spectrum. In the HR MAS spectrum (Fig. 1a) obtained utilizing cocoa
214 powder added to the minimum D₂O (ratio about 1:1), only lipid signals are visible. Instead,
215 spectrum obtained with a cocoa powder:D₂O ratio 1:5 (corresponding to optimized conditions,
216 figure 1b) is very similar to high resolution ¹H NMR spectrum, and it contains also information on
217 lipids. The main signals of the spectra were assigned as previously reported (Caligiani et al, 2010).
218 From this preliminary data, we started a more systematic study on the effect of solvent amount on
219 the quantitative results that can be obtained from HR MAS ¹H NMR spectra. We selected some
220 metabolites well observable and not overlapped in HR MAS spectra of cocoa and in the
221 corresponding solution ¹H NMR spectrum, determining their absolute amounts by comparison with
222 internal standard (TSP), as previously reported (Caligiani, Acquotti, Palla & Bocchi, 2007). HR
223 MAS ¹H NMR spectra were registered on samples prepared with different ratio of cocoa
224 powder:D₂O, namely 1:2, 1:5, 1:8, 1:15. Metabolites selected were aliphatic and aromatic amino
225 acid (alanine and tyrosine), polyphenol (epicatechin) and methylxanthine (theobromine). Compared
226 results are showed in the histograms of figure 2. For the most of the substances, the optimized
227 amount of cocoa powder respect to D₂O, gives quantitative results close to solution NMR, but this
228 is not true for all the substances considered. It is interesting to notice that substances as
229 theobromine, and to a less extent epicatechin, are not visible in spectra obtained with low amounts
230 of solvent, and they start to appear at higher dilutions. In general, the higher the dilution, the higher
231 the spectral signal, probably because the dilution increases the mobility of the molecules linked to
232 particular food structures. These data showed that absolute quantification in HR MAS is difficult to
233 obtain respect to solution NMR, due to the compartmentation, binding to macromolecules and to

234 repartition in different phases (aqueous/lipidic). So in the case of HR MAS ^1H NMR, we excluded
235 the possibility to obtain quantitative results as dataset and we chose the ratio cocoa powder 1:5 as a
236 compromise that permits to detect as many metabolites as possible compared with ^1H NMR spectra
237 in solution avoiding an excessive dilution of the sample. As a general remark, it is important to
238 highlight that to obtain comparable spectra it is mandatory that the amounts of cocoa powder and
239 D_2O in the rotor are the same for all the samples considered, because only in this case it is possible
240 to use the spectra as fingerprint. In fact the metabolite profile that can be obtained in HR MAS ^1H
241 NMR spectra strongly depends on the cocoa amount in the rotor respect to D_2O (ratio cocoa: D_2O)
242 and the optimization of the method is a prerequisite to obtain reliable results.

243 244 **3.2. Description/classification of cocoa beans by chemometrics tools**

245 In order to obtain a quali-quantitative description of the samples and to identify possible existing
246 relations or differences among cocoa beans of different origins based on their metabolites content,
247 PCA and OPLS-DA were performed on the different NMR dataset obtained.

248 **3.2.1 HR MAS ^1H NMR fingerprint dataset**

249 HR MAS fingerprint dataset is constituted by a total of 51 signals chosen in the HR MAS ^1H NMR
250 spectra of the 61 cocoa beans samples, considering both identified and unknown signals. PCA
251 conducted on raw data did not give information because no specific clustering was observed, so a
252 spectral filtering has been applied to dataset to remove information not correlated with the response,
253 usually denoted as “structured noise”. In particular, three Orthogonal Signal Correction filters
254 (OSC) has been applied to lead more simple models with easier interpretation. From the score plot
255 reported in figure 3a, clustering tendency for the three different continent of origin was evident after
256 scoring the first and the second PCs of the model. However, the separation is only on PC1, so it is
257 difficult to find specific markers of the origins, in particular for Asian samples. Moreover, due to
258 the low number of samples from Asia/Oceania respect to the other two Continents, we decided to
259 perform a PCA excluding these sample (Figure 4).

260 Considering only two origins, the separation is much better. In the loadings plot (figure 4b)
261 American cocoa beans were characterized by fatty acids (buckets 0.78-0.95 ppm and 1.46-1.51
262 ppm), acetic acid (bucket 1.98-2.13 ppm), oligosaccharides (bucket at 4.20-4.25 ppm) and
263 monosaccharides (bucket at 5.17-5.27 ppm). African samples were characterized by branched chain
264 amino acids (ile, leu, val, bucket at 1.03-1.07 ppm) and sugars (series of buckets between 3.49 and
265 3.88 ppm). The buckets were identified according to the identification of cocoa signals previously
266 reported for solution ^1H NMR (Caligiani et al., 2010).

267 A two-class OPLS-DA model was performed to better highlight the variables responsible of the
268 clustering. This model results in one predictive and two orthogonal components and the
269 corresponding S-plot (not shown) confirms what observed by PCA with OSC filters: African
270 samples are characterized mainly by sugars, while American samples by fatty acids, acetic acid,
271 oligosaccharides and monosaccharides. This means also that the OSC filters applied to PCA
272 effectively eliminated no-informative variables.

273 To check the predictive capability of the OPLS-DA model, balanced training and test sets were
274 selected constituted by 14 samples. However, the predictive performance of the model was limited:
275 3/14 samples were incorrectly classified and 4/14 were classified in both classes with the same
276 probability ($Q^2 = 20.9\%$).

277 **3.2.2. Solution ^1H NMR fingerprint dataset**

278 In order to compare the results of HR MAS with conventional ^1H NMR experiments, the same
279 approach was applied to the dataset obtained by integration of solution ^1H NMR signals. In this
280 case, the fingerprint dataset submitted to statistical analysis was constituted by 91 spectral integrals
281 for each of the 61 cocoa sample; as for HR MAS, a group separation was obtained only excluding
282 samples from Asia/Oceania. The better model was constituted by OPLS-DA (Figure 5).

283 The group separation (figure 5a) is better respect to the corresponding OPLS-DA performed on HR
284 MAS dataset and also the predictive performance of the model was slightly better: 3/13 samples

285 were incorrectly classified and 1/13 was classified in both classes with the same probability (Q2=
286 22.1%).

287 From the S-plot (figure 5b) it arise that African samples are characterized by a bucket at 2.64 ppm
288 (citric acid) and a series of buckets all concentrated in the spectral zone 3.6-4.2 ppm, corresponding
289 to sugar zone, in accordance with HR MAS results. American samples are characterized by bucket
290 at 1.98 ppm (acetic acid, important also for the HR MAS model), 0.99 ppm (valine), 1.85 ppm
291 (amino acids). Group of American samples in the HR MAS model was characterized also by lipids,
292 which in the aqueous extract utilized for solution NMR are obviously not present.

293 **3.2.3. Solution ¹H NMR quantitative datasets**

294 For solution ¹H NMR it was also possible to obtain a quantitative dataset manually integrating
295 twenty signals of the substances identified previously (Caligiani et al., 2010). The PCA performed
296 utilizing as variables only the quantitative data (not shown) does not permit a clear differentiation of
297 groups. PLS-DA with Pareto scaling, considering only two origins (figure 6a), showed a grouping
298 of the samples according to the two macro areas, with a sub-group of African samples. Observing
299 the loading plot (figure 6b) it arise that main group of African samples are characterized by citric
300 acid and sugars, in accordance with both HR MAS and solution fingerprint datasets. The sub group
301 of African samples is instead characterized by amino acids and lactic acid. American samples are
302 characterized by acetic acids and secondary metabolites as polyphenols and caffeine.

303 The sub group of African samples corresponded to the six samples coming from Sao Thome,
304 indicating specific Country-dependent characteristics for this origin. The variables associated with
305 the Sao Thome group are amino acids and lactic acid, which are both correlated with a good
306 fermentation process. Sao Thome samples were found rich of amino acids also in a previous work
307 (Marseglia, Palla, Caligiani, 2014).

308 **4. Conclusions**

309
310 Overall, the results confirm that NMR is a powerful tool to obtain quali-quantitative information on
311 metabolite composition of cocoa beans. The combination of the HR MAS ¹H NMR or solution ¹H

312 NMR datasets obtained from the whole spectra with chemometric techniques such as Principal
313 Component Analysis and OPLS-DA has emerged as an useful and rapid method to visualize the
314 characteristics of the beans in term of geographical origin. Our results showed that HR MAS ^1H
315 NMR fingerprint of cocoa beans combined with chemometrics is able to discriminate cocoa
316 samples from African regions and from America regions, based mainly on the lipidic and
317 saccharidic components. Solution NMR gives the advantage to obtain also quantitative results,
318 much more difficult with HR MAS.

319 It has to be considered that obtaining a discrimination of cocoa beans according to the geographical
320 origin is a challenging issue, because many factors concur to the metabolite profile, as variety,
321 climate, cultivation conditions, fermentation. Because fermentation level influences the metabolite
322 profile of cocoa to a higher level respect to the other factors linked to the geographic area, the
323 definition of cocoa geographic origin will remain very difficult until controlled fermentations will
324 not be introduced in the producing countries. However, also in this difficult contest, ^1H NMR, both
325 solution and HRMAS, have been proved to represent powerful tools to obtain rapid and objective
326 information on cocoa metabolite profile.

327 HR MAS ^1H NMR, in particular resulted a very effective in obtaining a cocoa metabolite profile
328 comprising both lipophilic and hydrophilic components, avoiding any sample manipulation and it is
329 maybe the only analytical technique that permits to obtain simultaneously structural information on
330 these two components of food.

331 Even if preliminary, our approach resulted well effective in geographical determination in terms of
332 Continents discrimination. Our future efforts will be routed towards Country discrimination by
333 using a wider collection of certified origin samples from those Countries that are main interested in
334 the production of high quality or mono-origin coca.

335

336

337

338

339

340 **Legends to figures**

341 Figure 1: 400 MHz HR MAS ^1H NMR spectrum of a cocoa bean sample (origin Santo Domingo)
342 obtained with a) cocoa powder:D₂O ratio 1:1; b) cocoa powder:D₂O ratio 1:5; c) 600 MHz ^1H
343 NMR spectrum of an aqueous extract of the same cocoa sample
344

345 Figure 2: Quantitative results (g/Kg) obtained for selected metabolites from HR MAS ^1H NMR
346 spectra at different cocoa/D₂O ratios and compared with solution ^1H NMR
347

348 Figure 3: Score plot of PCA performed on the HR MAS ^1H NMR fingerprint dataset by considering
349 all cocoa beans samples (3 OSC filters, R₂= 42.95%, Noise=57.05%, Pc₁=45.9%, Pc₂=27.4%,
350 R₂X=98.1%, Q₂=75.3%).
351

352 Figure 4: Score (a) and loading plots (b) of PCA performed on the HR MAS ^1H NMR fingerprint
353 dataset by considering only cocoa beans samples from African and American Countries (Pareto
354 scaling, 3 OSC filters, R₂= 31.39%, noise=68.61%, Pc₁=40.7%, Pc₂=14.8%, R₂X=55.5%,
355 Q₂=41.5%).
356

357 Figure 5: Score plot (a) and S-plot (b) of OPLS-DA, performed on the solution ^1H NMR fingerprint
358 dataset by considering cocoa beans samples from African and American Countries. Pareto scaling,
359 R₂X=82.5%, R₂Y=77.9%, Q₂=50.7%
360

361 Figure 6: Score plot (A) and (B) loading plots of PLS-DA, performed on the solution ^1H NMR
362 quantitative dataset by considering cocoa beans samples from Africa and America. 2 pc,
363 Pc₁=47.2%, Pc₂=9.58%, R₂X=56.8%, R₂Y=58.7%, Q₂=47.2%, Permutation X 200, Q₂ neg,
364 R₂=0.13/0.126
365

366

367

368

369

370

371

372

373

374

375 **References**

376 Afoakwa, E., (2010). Chocolate production and consumption patterns, in *Chocolate Science and*
377 *Technology*, ed. by Afoackwa E. Wiley- Blackwell, Oxford, pp. 1–11.

378

379 Ardhana, M.M. & Fleet, G.H. (2003). The microbial ecology of cocoa bean fermentations in
380 Indonesia. *International Journal of Food Microbiology*, 86, 87-99.

381

382 Caligiani, A., Acquotti, D., Palla, G. & Bocchi, V. (2007). Identification and quantification of the
383 main organic components of vinegars by high resolution ^1H NMR spectroscopy. *Analytica Chimica*
384 *Acta*, 585, 110-119.

385

386 Caligiani, A., Acquotti, D., Cirlini, M. & Palla, G. (2010). ^1H NMR Study of Fermented Cocoa
387 (*Theobroma Cacao L.*) Beans. *Journal of Agricultural and Food Chemistry*, 58,12105-12111.

388

389 Caligiani, A., Palla, L., Acquotti, D., Marseglia, A., Palla, G. (2014). Application of ^1H NMR for
390 the Characterization of Cocoa Beans of Different Geographical Origins and Fermentation Levels.
391 *Food Chemistry*, 157, 94-99.

392

393 Cambrai, A., Marcic, C., Morville, S., Sae Houer, P., Bindler, F., Marchioni, E. (2010).
394 Differentiation of Chocolates According to the Cocoa's Geographical Origin Using Chemometrics.
395 *J. Agric. Food Chem.*, 58, 1478–1483.

396

397 Choze, R., Alcantara, G.B., Alves Filho, E. de G. Silva, L. M. A., Faria, J. C., Liao, L. M. (2013).
398 Distinction between a transgenic and a conventional common bean genotype by ¹H HR-MAS
399 NMR. *Food Chemistry*, 141(3), 2841-2847.
400
401 Cidell, J. L., Alberts, H. C. (2006). Constructing quality: the multinational histories of chocolate.
402 *Geoforum*, 37, 999–1007.
403
404 Consonni, R., Cagliani, L. R. (2008). Geographical characterization of polyfloral and acacia honeys
405 by nuclear magnetic resonance and chemometrics. *Journal of Agricultural and Food Chemistry*, 56,
406 6873–6880.)
407
408 Consonni, R., Cagliani, L. R., Stocchero, M., Porretta, S. (2009). Italian and Chinese tomato paste:
409 Geographical discrimination by NMR and chemometrics. *Journal of Agricultural and Food*
410 *Chemistry*, 57, 4506–4513.
411
412 Consonni, R., Cagliani, L.R. (2010). Nuclear Magnetic Resonance and Chemometrics to Assess
413 Geographical Origin and Quality of Traditional Food Products. In Steve L. Taylor, editor: *Advances*
414 *in Food and Nutrition Research*, Burlington: Academic Press, Vol. 59, pp. 87-165.
415
416 Consonni, R., Cagliani, L.R., Cogliati, R. (2012). NMR based geographical characterization of
417 roasted coffee. *Talanta*, 88, 420-426.
418
419 Jinap, S., Dimick, P. S., Hollender, R. (1995). Flavour evaluation of chocolate formulated from
420 cocoa beans from different countries. *Food Control*, 6, 105–110.
421

422 Kattenberg, H.R., Kemmink, A. (1993). The flavor of cocoa in relation to the origin and processing
423 of the cocoa beans. *Development in Food Science*, 32, 1-22.

424

425 Kelly, S., Heaton, K., Hoogewerff, J. (2005). Tracing the geographical origin of food: The
426 application of multi-element and multi-isotope analysis. *Trends in Food Science & Technology*, 16
427 (12), 555-556.

428

429 Marseglia, A., Palla, G., Caligiani, A. (2014). Presence and variation of γ -aminobutyric acid and
430 other free amino acids in cocoa beans from different geographical origins. *Food Research*
431 *International*, 63(C), 360-366.

432

433 Ritota, M., Casciani L., Valentini M. (2013). PGI chicory (*Cichorium intybus L.*) traceability by
434 means of HRMAS-NMR spectroscopy: a preliminary study. *Journal of the science of food and*
435 *agriculture*, 93(7), 1665-1672.

436

437 Shintu, L., Caldarelli, S., Franke, B. M. (2007). Pre-selection of potential molecular markers for the
438 geographic origin of dried beef by HR-MAS NMR spectroscopy. *Meat Science*, 76(4), 700-707.

439

440 Teye, E., Huang, X., Dai, H., Chen, Q. (2013). Rapid differentiation of Ghana cocoa beans by FT-
441 NIR spectroscopy coupled with multivariate classification. *Spectrochimica Acta Part A: Molecular*
442 *and Biomolecular Spectroscopy*, 114, 183–189

443

444 Torres-Moreno, M., Tarrega, A., Costell, E., Blanch, C. (2012). Dark chocolate acceptability:
445 influence of cocoa origin and processing conditions. *J Sci Food Agric*, 92, 404–411.

446

447 Torres-Moreno, M., Torrecasana, E., Salas-Salvadó, J., Blanch, C. (2015). Nutritional composition
448 and fatty acids profile in cocoa beans and chocolates with different geographical origin and
449 processing conditions. *Food Chemistry*, 166, 125–132.

450 Vermathen, M., Marzorati, M., Baumgartner, D., Good, C., Vermathen, P. (2013). Impact of
451 different cultivation methods on the metabolic profile of apples studied by ¹H-MAS NMR
452 spectroscopy From Special Publication - Royal Society of Chemistry, *Magnetic Resonance in Food*

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

Figure1

[Click here to download high resolution image](#)

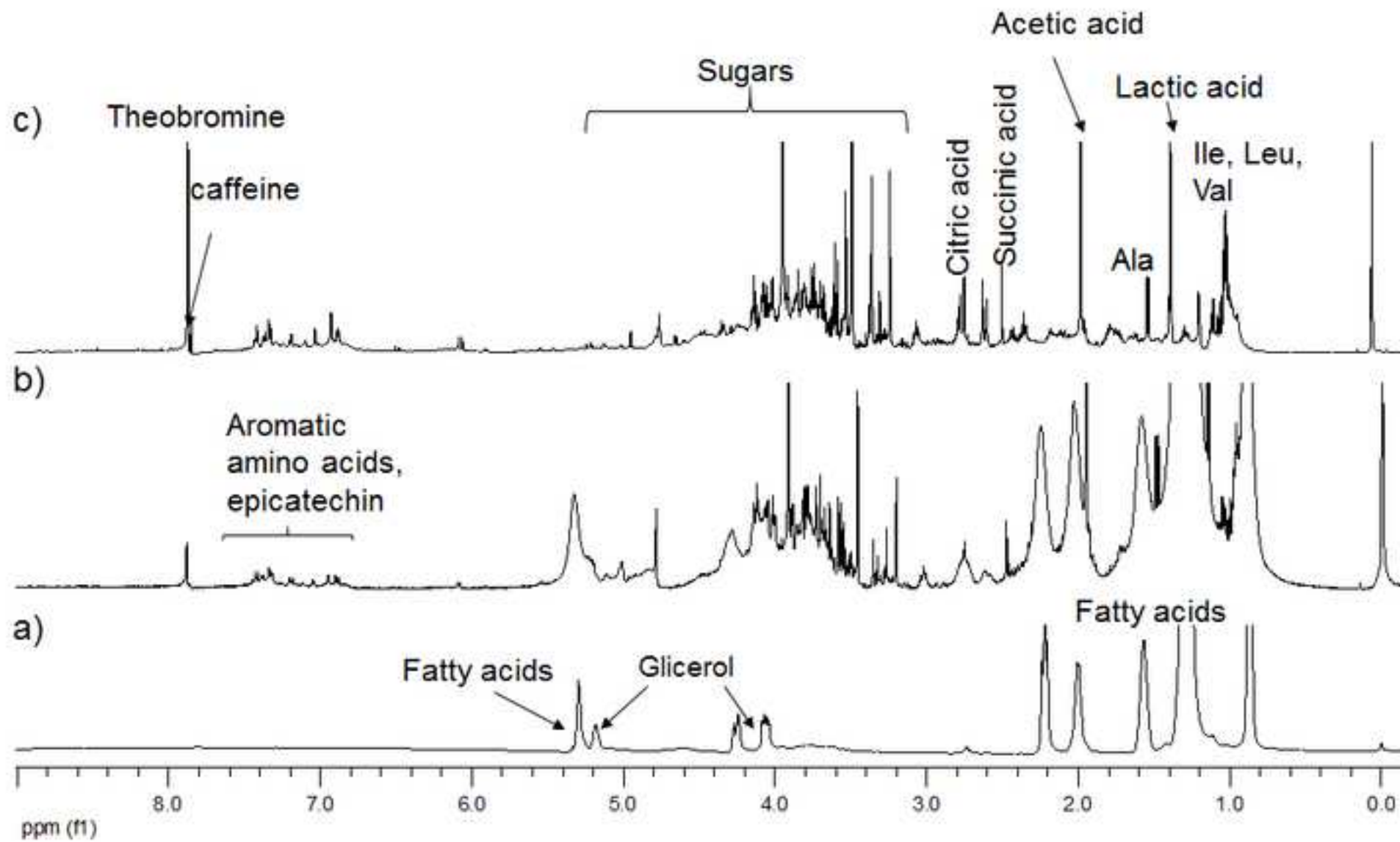


Figure2

[Click here to download high resolution image](#)

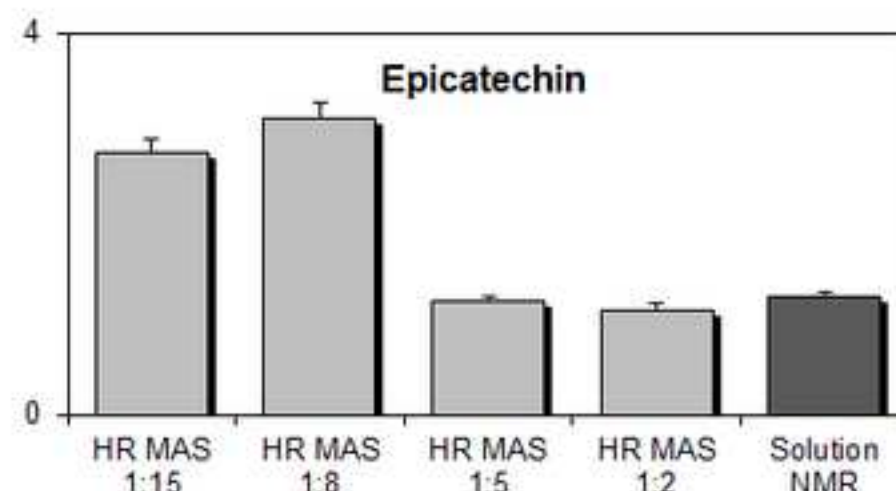
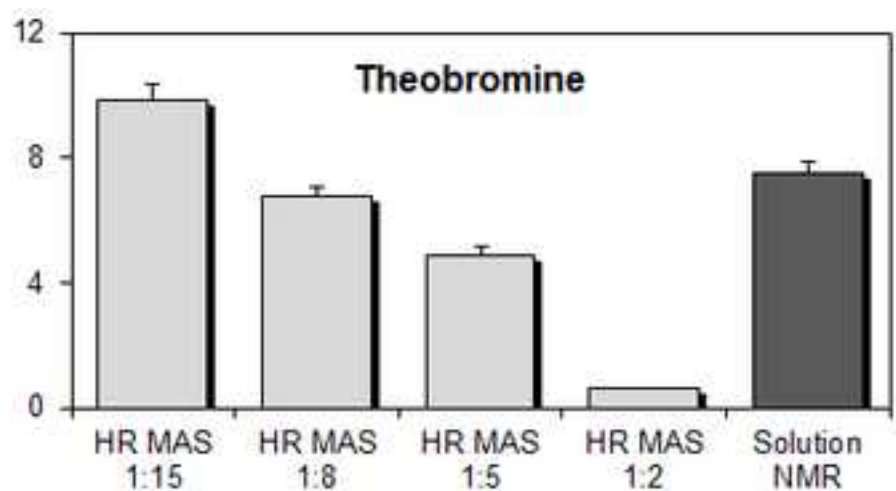
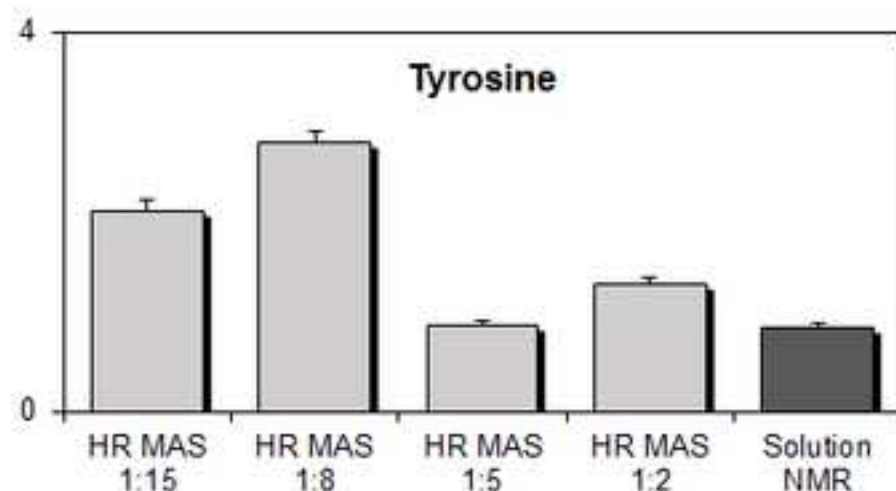
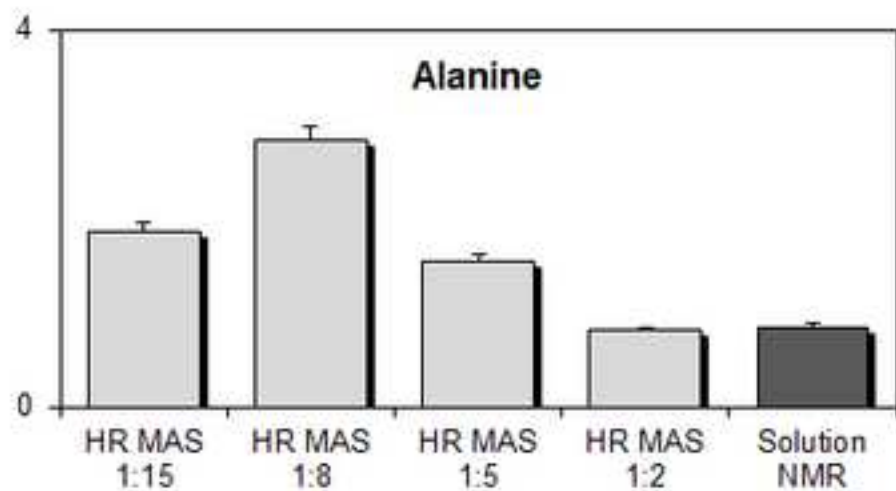


Figure3

[Click here to download high resolution image](#)

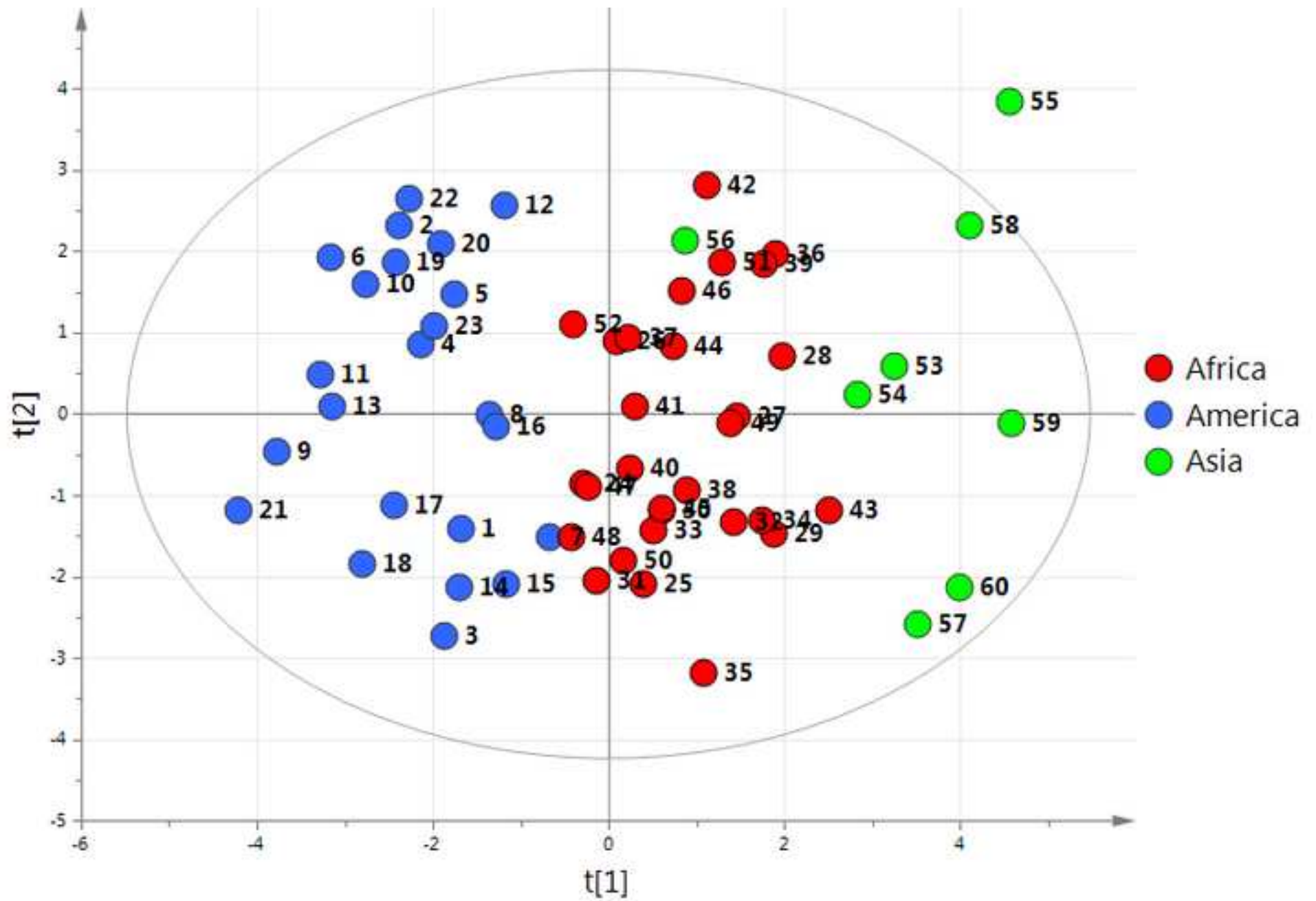


Figure4

[Click here to download high resolution image](#)

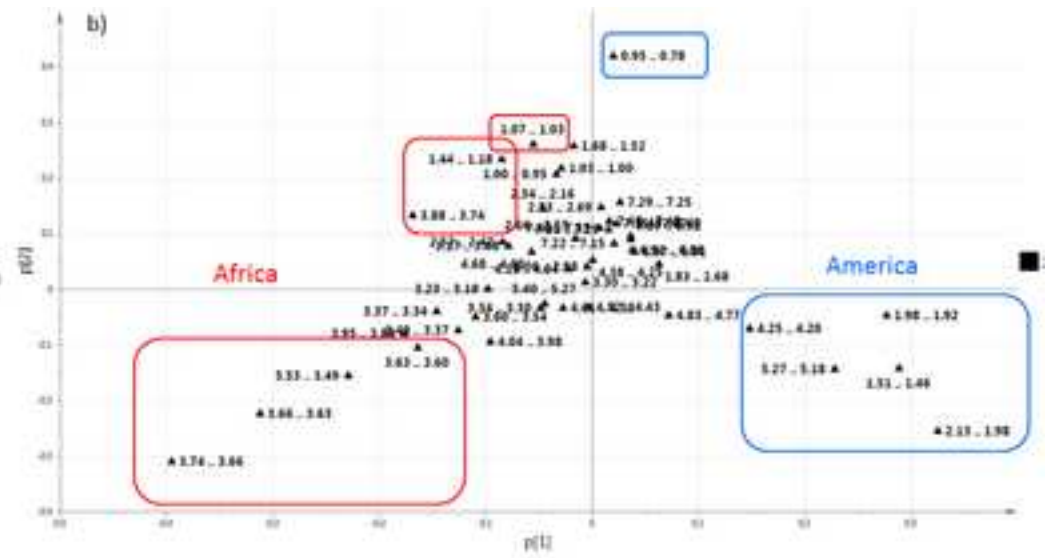
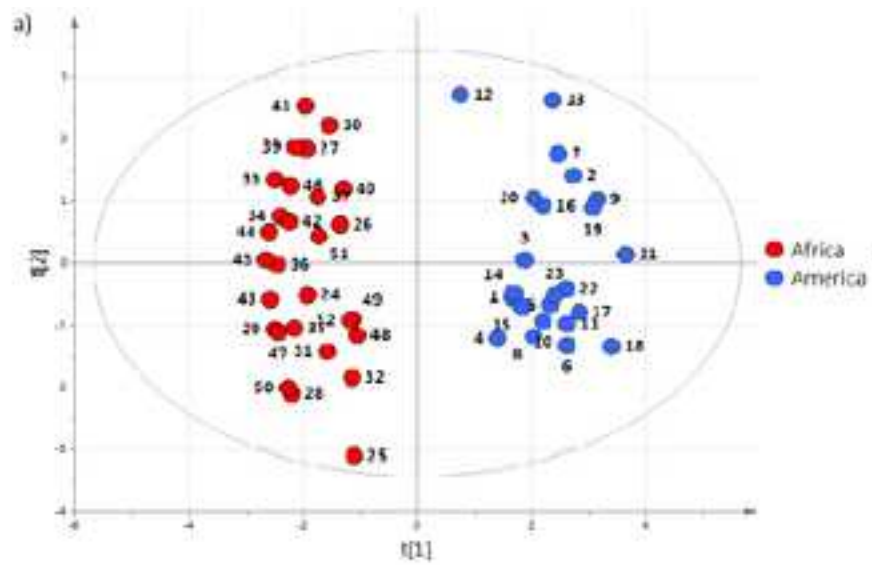


Figure5
[Click here to download high resolution image](#)

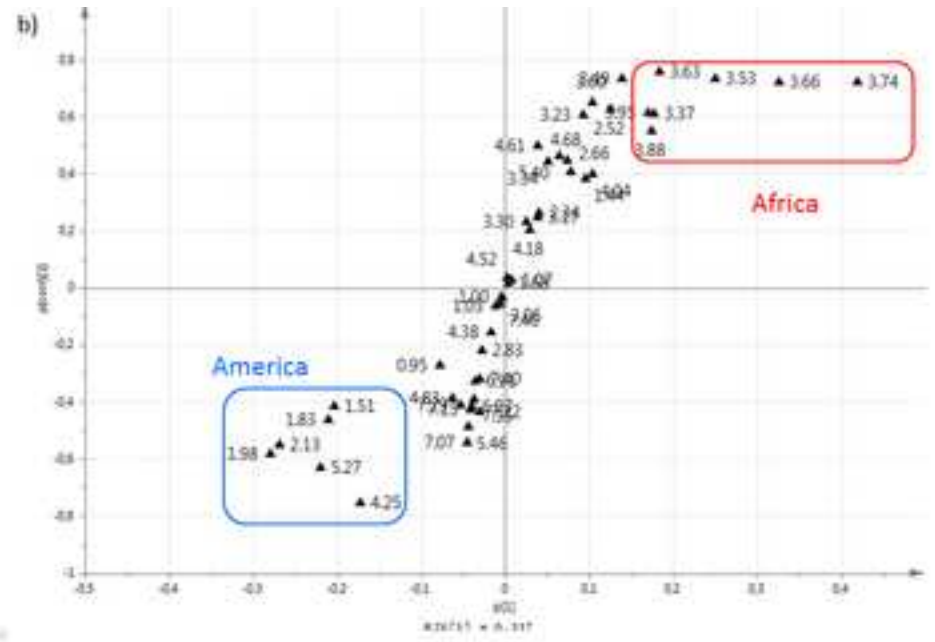
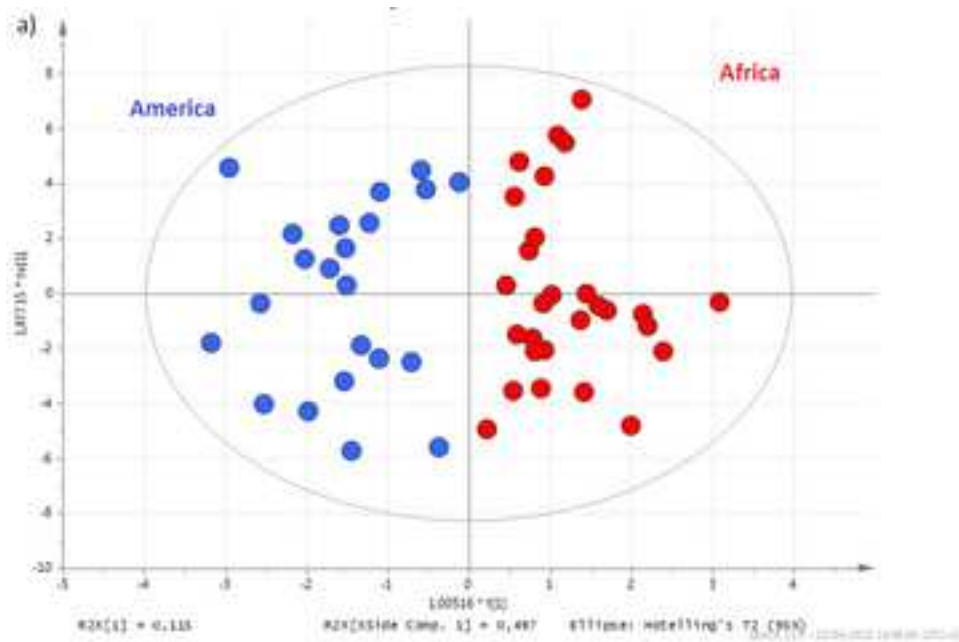


Figure6

[Click here to download high resolution image](#)

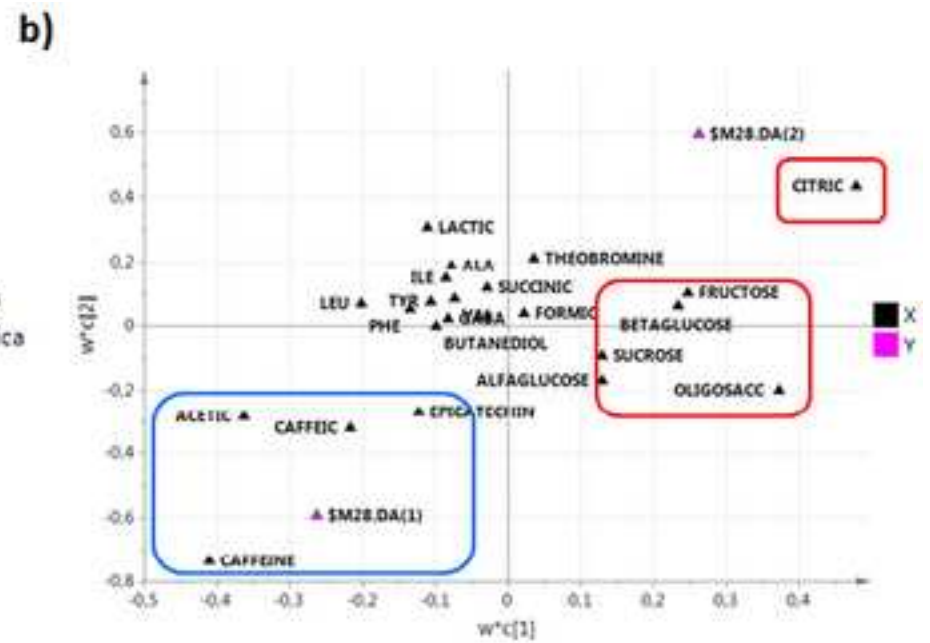
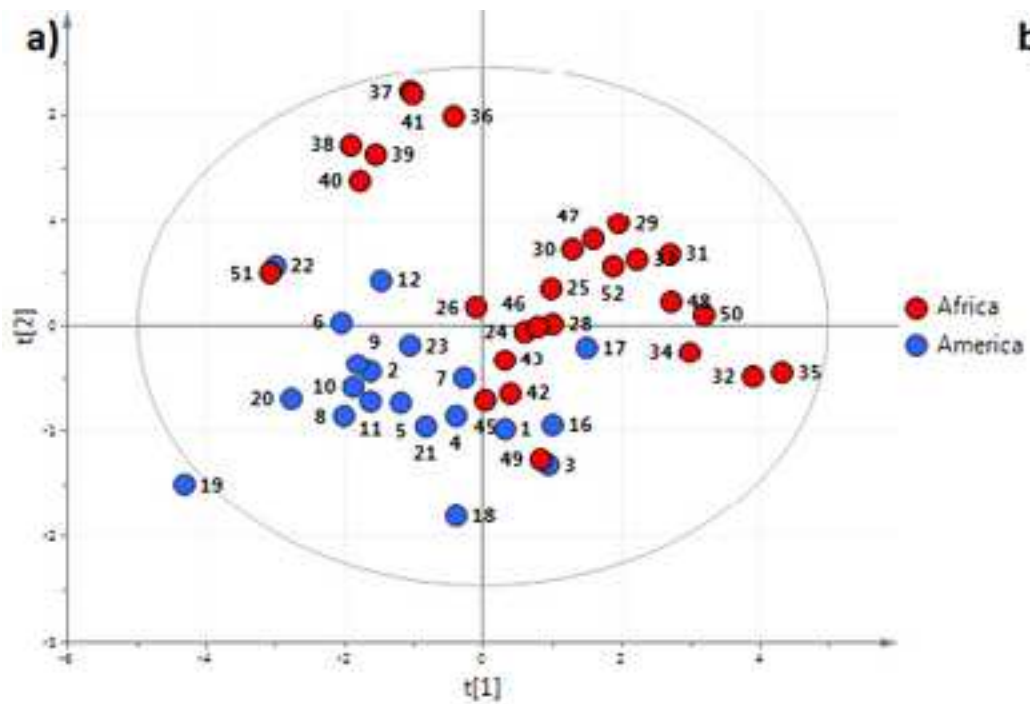


Table 1: Fermented and dried cocoa beans collection. All samples were harvested in 2012

Country	Month of harvesting	Number of samples
Brazil (Bahia)	October	1
Cuba	March	1
Congo	September	1
Ecuador	July	1
	September	1
Ghana	June	2
	July	2
Grenada	March	1
Indonesia-Flores	October	1
Indonesia-Sulawesi	October	1
Ivory Coast	June	4
Java	March	1
Madagascar	March	1
Malaysia	September	1
Mexico	September	1
Mexico (var. Criollo)	June	1
Nigeria	September	4
Papua New Guinea	June	3
	July	1
Perù	March	1
	June	1
	July	4
	September	2
Santo Domingo	June	3
	July	2
Sao Thome	March	1
	June	1
	September	4
Sierra Leone	September	2
Tanzania	July	1
	September	4
Trinidad	March	1
Uganda	September	2
Venezuela	March	1
	September	1
	October	1