Elsevier Editorial System(tm) for Food Research International

Manuscript Draft

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 ^{1}H NMR AND CHEMIOMETRICS AS USEFUL TOOL TO ASSESS THE GEOGRAPHICAL ORIGIN OF COCOA BEANS – COMPARISON WITH SOLUTION ^{1}H 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

HR MAS ^{1}H NMR AND CHEMIOMETRICS AS USEFUL TOOL TO ASSESS THE GEOGRAPHICAL ORIGIN OF COCOA BEANS – COMPARISON WITH SOLUTION ^{1}H NMR

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

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

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

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

*Corresponding author:

Augusta Caligiani

Dip. Scienze degli Alimenti

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

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

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 ^{1}H 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 ^{1}H NMR directly on cocoa powder after the method optimization. The same samples were also subjected to extraction and analysis with solution ^{1}H NMR.

HR MAS ^{1}H NMR, as ^{1}H 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 ^{1}H NMR. The data set obtained is therefore representative of all classes of cocoa compounds. Untargeted HR MAS ^{1}H NMR and ^{1}H 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.

Keywords

HRMAS ^{1}H NMR, solution ^{1}H NMR, cocoa beans, geographic origin, metabolic profile, chemometrics.

1. Introduction

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

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

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

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

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

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

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

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

2. Materials and Methods

2.1 Chemicals. Methanol**,** D_2O, CD_3OD and 3-(trimethylsilyl)-propionate-d4 (TSP, internal standard for NMR analysis) were purchased from Sigma-Aldrich (Milan, Italy).

2.2 Sampling. 61 fermented and dried cocoa beans samples of Forastero variety from 23 different geographical origins were considered. The sampling is representative of the average world production, (America Africa Asia), and, to the best of our knowledge, it represent the largest cocoa beans collection of different geographical origins considered in the literature. Sampled were kindly provided by Barry Callebaut Belgium, all fruits were harvested in 2012, from March to October, according to the cocoa season specific of the different countries. Table 1 reports the specific Countries of origins considered, the number of different lots for each origin and the months of harvesting. Samples represent cocoa beans traded from countries of origins 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. All the samples were considered well fermented.

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

2.3 Sample preparation for HR MAS ^{1}H NMR analysis. Cocoa powder were directly inserted in a 4 mm HRMAS rotor with a 50 µL disposable insert, together with D_2O containing TSP. The optimized amount of cocoa powder/D_2O was 4 mg in 20 µL D_2O.

2.4 HR MAS ^{1}H NMR acquisition and spectra processing. Samples were analysed with a Bruker Avance III 400 MHz NMR spectrometer operating at 9.4 T, using a 1H-13C HR-MAS probe with a BVT-3200 temperature controller and BCU cooling unit. The probe resonant circuits were tuned and matched with the sample in place. The ‘‘magic angle’’ (54.7) was adjusted using the 79Br signal from powdered KBr as standard. A spinning rate of 4000 Hz was chosen to move unwanted

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

2.5 Sample preparation for solution ^{1}H NMR analyses. 200 mg of fermented cocoa beans, finely ground, were extracted with 20 ml distilled water/methanol mixture (8:2 v/v), kept at the boiling point for 10 min under magnetic stirring. Extracts were cooled, filtered, taken to dryness, dissolved in 1 ml D_2O/CD_3OD (8:2 v/v) containing 0.1 % of 3-(trimethylsilyl)-propionate-*d4* (TSP), filtered again and transferred in a 5 mm NMR sample tube. TSP was used as internal standard for the quantitative analysis and for chemical shift referencing (δ=0 ppm).

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

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

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

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

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

3. Results and Discussion

3.1 Optimization of sample preparation for HR MAS ^{1}H NMR

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

rotor disposable insert, together with D_2O containing TSP. However, from the first experiment performed it appeared that the aspect of the HR MAS ^{1}H NMR spectra varied greatly based on sample preparation, and in particular completely different spectra were obtained utilizing different ratios of cocoa powder vs deuterated solvents. In figure 1 the comparison of HR MAS ^{1}H NMR spectra obtained with different ratio of cocoa powder: D_2O are reported and compared with high resolution liquid ^{1}H NMR spectrum. In the HR MAS spectrum (Fig. 1a) obtained utilizing cocoa powder added to the minimum D_2O (ratio about 1:1), only lipid signals are visible. Instead, spectrum obtained with a cocoa powder:D_2O ratio 1:5 (corresponding to optimized conditions, figure 1b) is very similar to high resolution ^{1}H NMR spectrum, and it contains also information on lipids. The main signals of the spectra were assigned as previously reported (Caligiani et al, 2010).

From this preliminary data, we started a more systematic study on the effect of solvent amount on the quantitative results that can be obtained from HR MAS ^{1}H NMR spectra. We selected some metabolites well observable and not overlapped in HR MAS spectra of cocoa and in the corresponding solution ^{1}H NMR spectrum, determining their absolute amounts by comparison with internal standard (TSP), as previously reported (Caligiani, Acquotti, Palla & Bocchi, 2007). HR MAS ^{1}H NMR spectra were registered on samples prepared with different ratio of cocoa powder:D_2O, namely 1:2, 1:5, 1:8, 1:15. Metabolites selected were aliphatic and aromatic amino acid (alanine and tyrosine), polyphenol (epicatechin) and methylxanthine (theobromine). Compared results are showed in the histograms of figure 2. For the most of the substances, the optimized amount of cocoa powder respect to D_2O, gives quantitative results close to solution NMR, but this is not true for all the substances considered. It is interesting to notice that substances as theobromine, and to a less extent epicatechin, are not visible in spectra obtained with low amounts of solvent, and they start to appear at higher dilutions. In general, the higher the dilution, the higher the spectral signal, probably because the dilution increases the mobility of the molecules linked to particular food structures. These data showed that absolute quantification in HR MAS is difficult to obtain respect to solution NMR, due to the compartmentation, binding to macromolecules and to

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

3.2. Description/classification of cocoa beans by chemometrics tools

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

3.2.1 HR MAS ^{1}H NMR fingerprint dataset

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

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

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

To check the predictive capability of the OPLS-DA model, balanced training and test sets were selected constituted by 14 samples. However, the predictive performance of the model was limited: 3/14 samples were incorrectly classified and 4/14 were classified in both classes with the same probability (Q2= 20.9%).

3.2.2. Solution 1H NMR fingerprint dataset

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

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

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

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

3.2.3. Solution ^{1}H NMR quantitative datasets

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

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

4. Conclusions

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

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

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

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

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

Legends to figures

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

Figure 2: Quantitative results (g/Kg) obtained for selected metabolites from HR MAS ^{1}H NMR spectra at different cocoa/D_2O ratios and compared with solution ^{1}H NMR

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

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

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

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

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). ^{1}H 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 ^{1}H 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

Choze, R., Alcantara, G.B., Alves Filho, E. de G. Silva, L. M. A., Faria, J. C., Liao, L. M. (2013). Distinction between a transgenic and a conventional common bean genotype by 1H HR-MAS NMR. *Food Chemistry*, 141(3), 2841-2847.

Cidell, J. L., Alberts, H. C. (2006). Constructing quality: the multinational histories of chocolate. *Geoforum*, 37, 999–1007.

Consonni, R., Cagliani, L. R. (2008). Geographical characterization of polyfloral and acacia honeys by nuclear magnetic resonance and chemometrics. *Journal of Agricultural and Food Chemistry*, 56, 6873–6880.)

Consonni, R., Cagliani, L. R., Stocchero, M., Porretta, S. (2009). Italian and Chinese tomato paste: Geographical discrimination by NMR and chemometrics. *Journal of Agricultural and Food Chemistry*, 57, 4506–4513.

Consonni, R., Cagliani, L.R. (2010). Nuclear Magnetic Resonance and Chemometrics to Assess Geographical Origin and Quality of Traditional Food Products. In Steve L. Taylor, editor: *Advances in Food and Nutrition Research*, Burlington: Academic Press, Vol. 59, pp. 87-165.

Consonni, R., Cagliani, L.R., Cogliati, R. (2012). NMR based geographical characterization of roasted coffee. *Talanta*, 88, 420-426.

Jinap, S., Dimick, P. S., Hollender, R. (1995). Flavour evaluation of chocolate formulated from cocoa beans from different countries. *Food Control*, 6, 105–110.

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

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

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

Figure1

Click here to download high resolution image

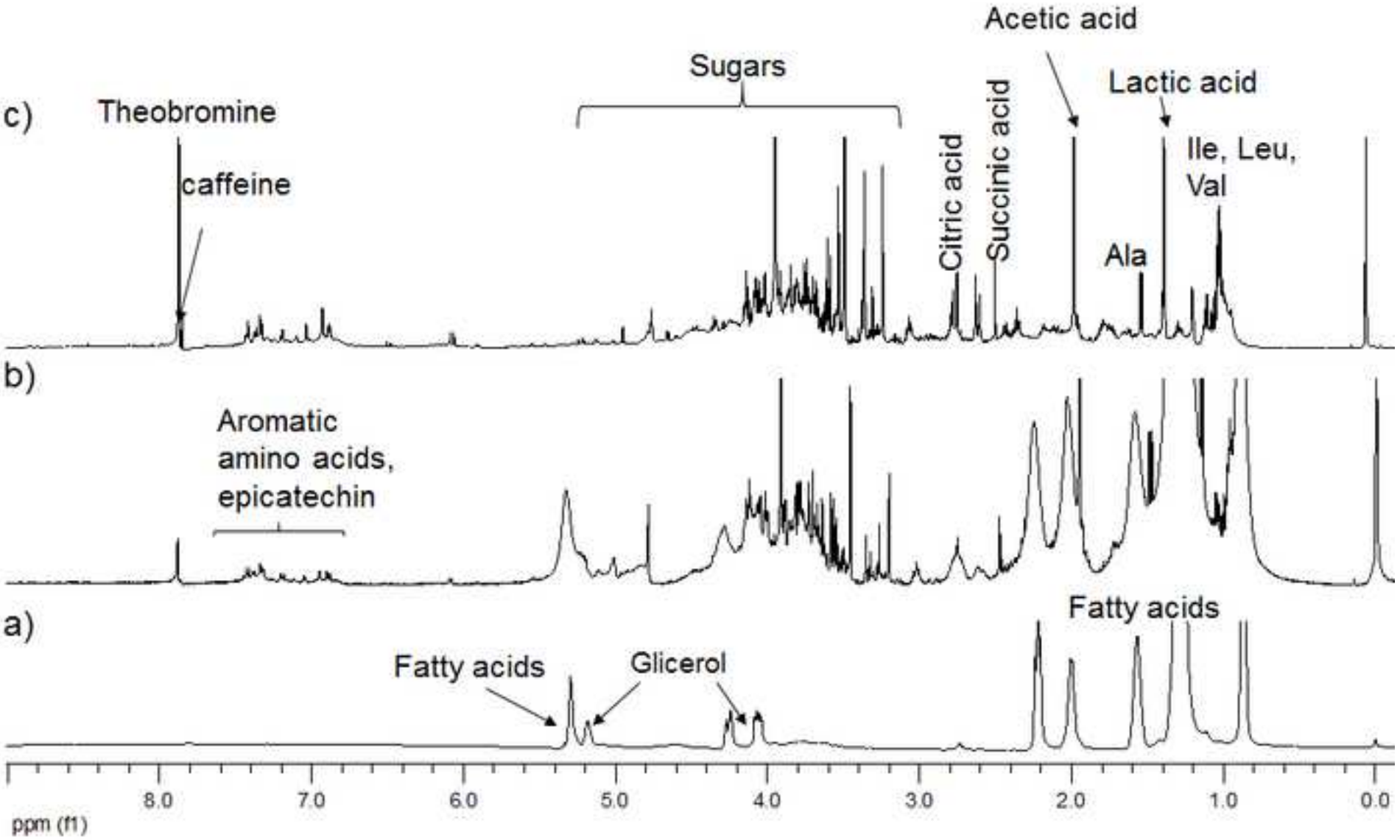


Figure2

Click here to download high resolution image

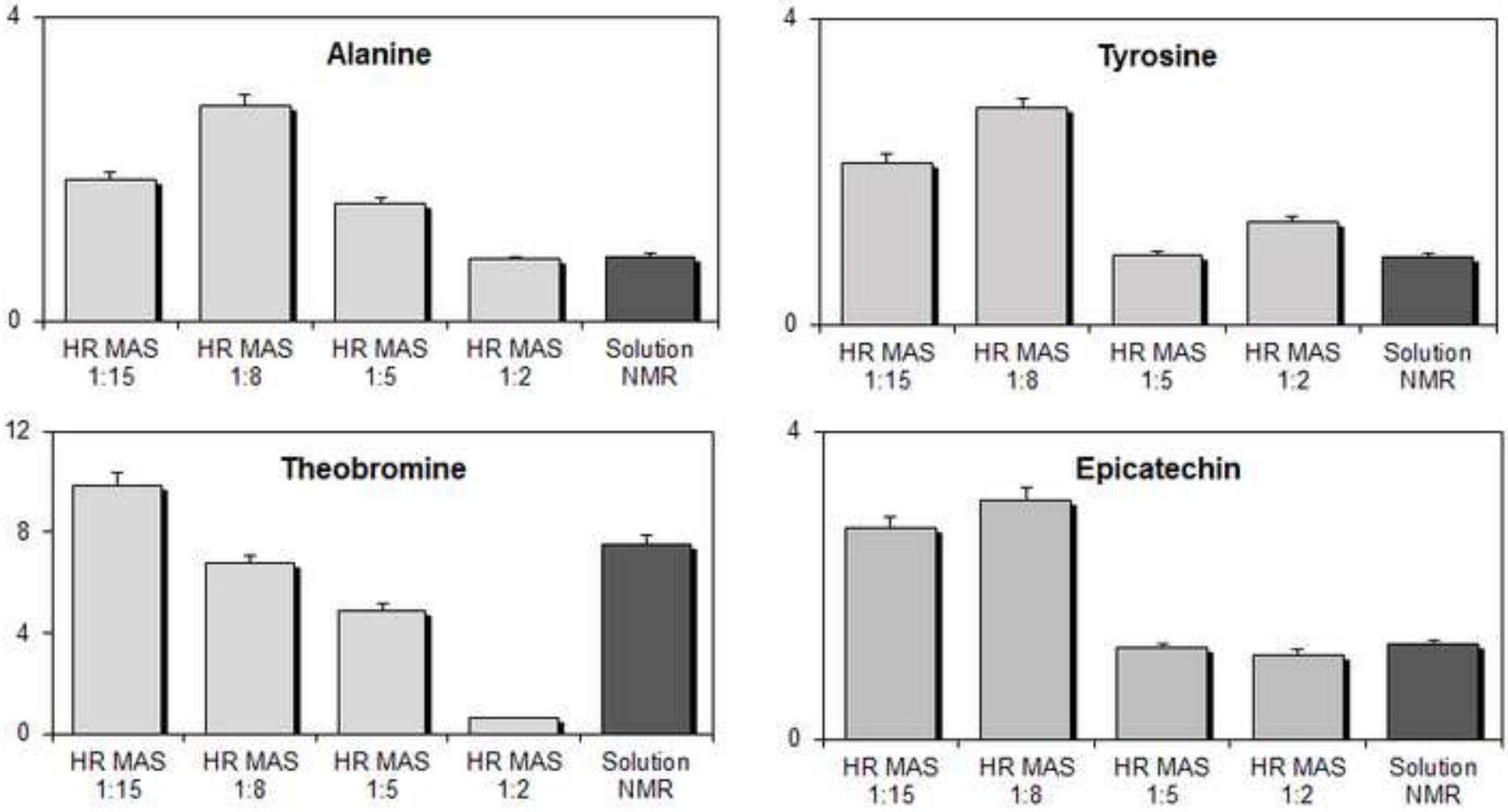


Figure3

Click here to download high resolution image

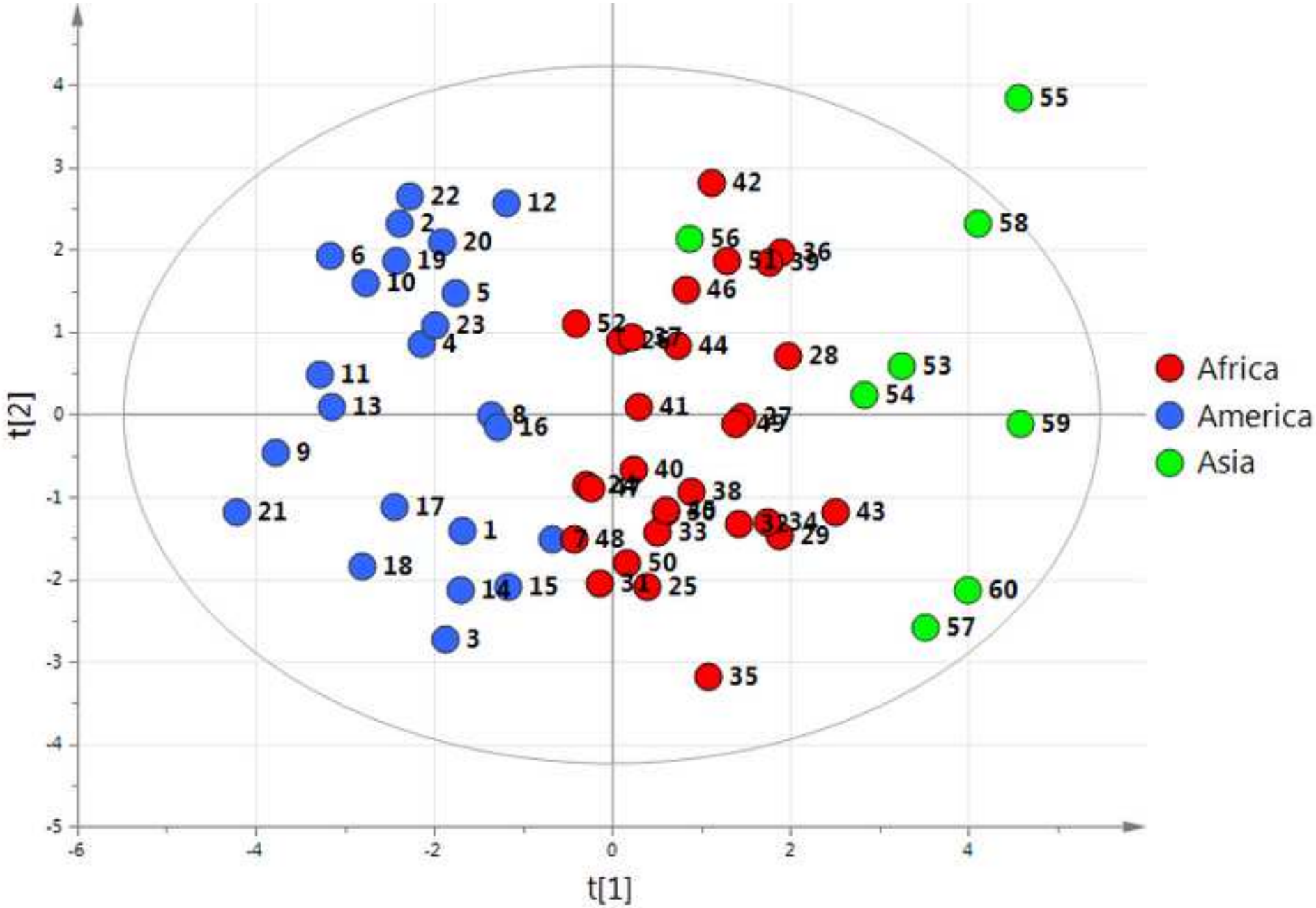


Figure4

Click here to download high resolution image

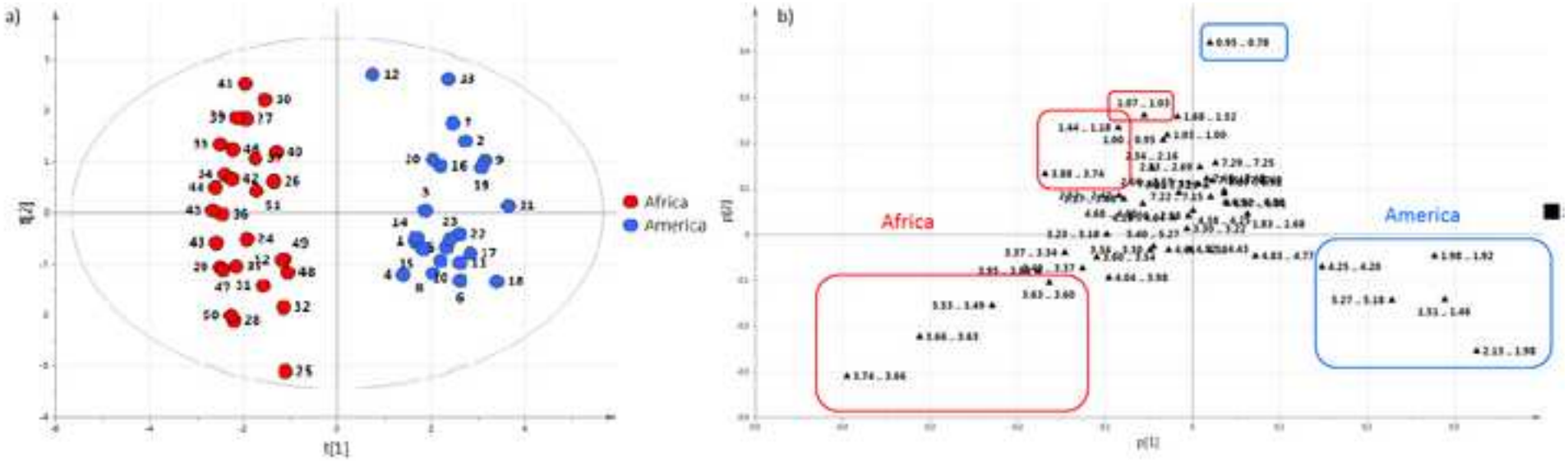


Figure5

Click here to download high resolution image

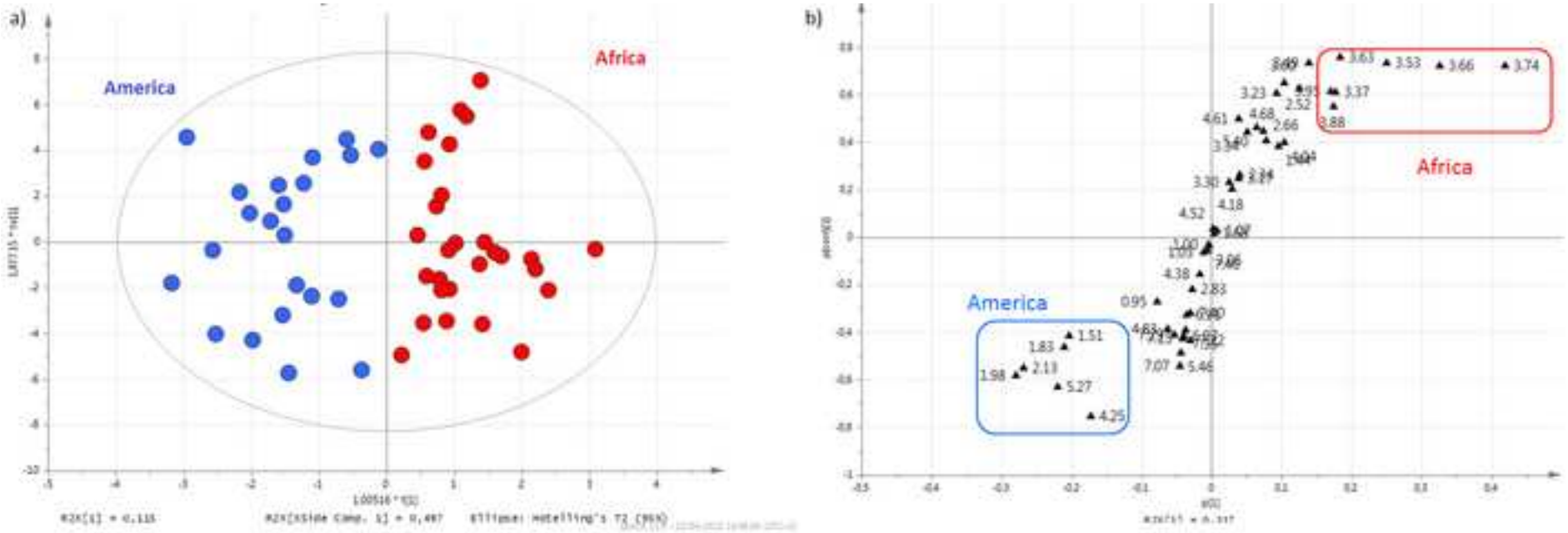
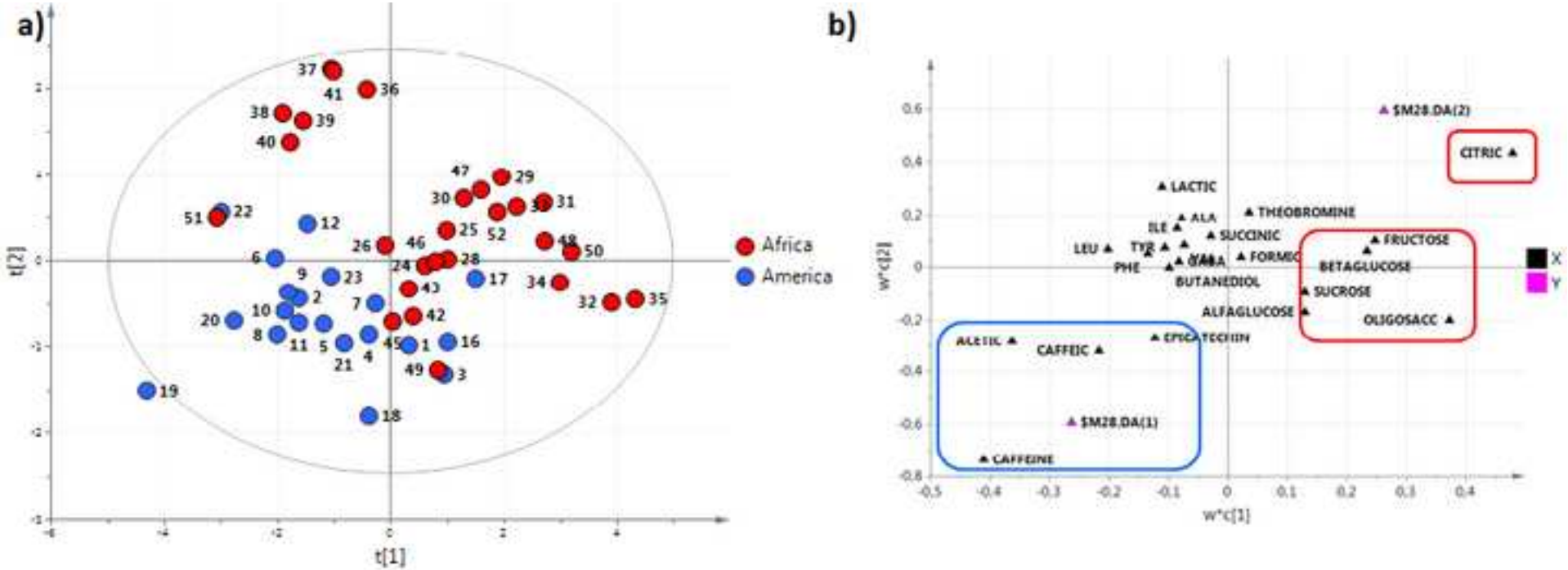


Figure6

Click here to download high resolution image



Table

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