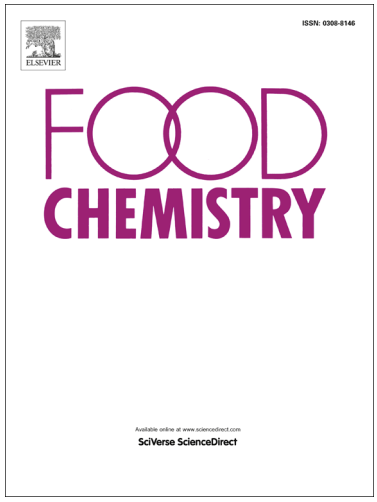


Short communication

A semi-automatic approach to the characterisation of dark chocolate by Nuclear Magnetic Resonance and multivariate analysis

Adam Le Gresley, Jean-Marie R. Peron

PII: S0308-8146(18)31661-3
DOI: https://doi.org/10.1016/j.foodchem.2018.09.089
Reference: FOCH 23576

To appear in: *Food Chemistry*

Received Date: 2 March 2018
Revised Date: 6 September 2018
Accepted Date: 14 September 2018

Please cite this article as: Gresley, A.L., Peron, J.R., A semi-automatic approach to the characterisation of dark chocolate by Nuclear Magnetic Resonance and multivariate analysis, *Food Chemistry* (2018), doi: https://doi.org/10.1016/j.foodchem.2018.09.089

A semi-automatic approach to the characterisation of dark chocolate by Nuclear Magnetic Resonance and multivariate analysis.

Adam LE GRESLEY[a*] and Jean-Marie R.PERON[a]

[a]Department of Chemistry and Pharmaceutical Sciences, SEC Faculty, Kingston University, Kingston-upon-Thames, Surrey, KT1 2EE, UK.

[*]Corresponding author:

Dr Adam Le Gresley, Tel + 44 (0)20 84177432 Email: a.legresley@kingston.ac.uk

Abstract

Tracing the geographical origin of chocolate is of increasing importance owing to the market growth of cocoa products of high quality and especially where value is derived from those products being of single origin. The NMR analysis of methanolic/aqueous extracts of dark chocolate samples from Peru, Venezuela and Madagascar is reported and 42 different chemical constituents are identified, quantified and analysed using multivariate techniques. This paper describes a simple non-destructive protocol, which look at the chemical profile for chocolate samples from these three geographical locations and demonstrates potential for assessing the provenance of chocolate products, which has implications in food quality, safety and authenticity.

Introduction

Determining the geographical origin of foods is a challenge with implications for food authenticity, especially when the existence of high-value products belonging to

specific geographical areas are granted Protected Geographical Indication (PGI) or Protected Denomination of Origin (PDO) marks. Chocolate, in particular, premium chocolate represents a rapidly growing industry of significant financial importance to many parts of the world. Often there are unique characteristics to this chocolate, linked specifically to the area of production.

This is of particular relevance when considering, chocolates made from fine flavour cocoa beans, such as those from Ghana, Ecuador or Venezuela, or organic and fair-trade chocolates (Afoakwa, 2010).

Despite the importance of the chocolate market there are comparatively few studies, which consider the chemical constituents of the finished chocolate product, that is the product of cocoa beans after fermentation, drying process, roasting, winnowing, alkalisation, conching and finally tempering. The majority focus on analytical methods for determining the geographical origin of unfermented cocoa beans and certain fermentation products. Investigations looking at fatty acid profiles (Torres-Moreno, Torrescasana, Salas-Salvadó, & Blanch, 2015) volatile compounds (Cambrai, Marcic, Morville, Sae-Houer, Bindler, & Marchioni, 2010) Solid State Magic Angle Spinning (MAS) NMR (Marseglia,Acquotti, Consonni, Cagliani, Palla, Caligliani, 2016) and Fourier Transform Near Infrared (FT-NIR) (Teye, Huang, Dai & Chen, 2013) are amongst those reported. Of particular interest is the recent use of Electrospray Mass Spectrometry (ESI-MS) to characterise chocolates using differences in polyphenols and other unassigned mass values via multivariate analysis (Acierno, Alewijn, Zomera & van Ruth 2018). Drawbacks to this process include the several extraction, agitation and centrifugation steps required before injection; it is also destructive. When considering the validity of using NMR directly to characterise chocolate it is important to consider the production process and how the

geographical origin of the cocoa bean and its specific local environmental conditions can still be evidenced after the many steps involved in its conversion to chocolate. The primary transformation is the fermentation and drying of cocoa beans, in their country of origin, during which time many of the polyphenolic constituents are lost. These fermented and dried beans then undergo secondary processing comprising roasting, milling, pressing, conching and tempering, to produce chocolate and cocoa powder. The secondary process often occurs outside the country of origin. By the time the chocolate is generated the composition has changed markedly, however, the taste and flavour of chocolate and cocoa-based products, strongly depends both on the cocoa variety and on the geographical origin(Acierno, Yener, Alewijn, Biasioli & van Ruth, 2016).

Analysis of finished chocolate by MS (Cambrai, Marchioni, Julien-David & Marcic, 2017) indicate that certain polyphenols and derivatives can be identified, but these concentrations can be low enough to prevent analysis with high resolution NMR, necessitating analysis by more sensitive techniques, such as HPLC-MS analysis (Kuhnert, Milev, Patras & Vrancken G, 2014). There is an implication that the characteristics of the resulting chocolate are influenced by its pattern of polyphenolic constituents. It can be hypothesised that those components (and their relative ratios) that remain in higher concentration in the finished dark chocolate can be evaluated by multivariate analysis and afford an equally robust set of discriminatory variables for distinguishing the geographical origin of the original cocoa bean. The strategy employed was to reduce the operator input as much as possible and automate the NMR analysis. This was achieved by running ^{1}H NMR experiments and automatically identifying and quantifying components using Chenomx software. The quantitative

data could then be analysed using PCA/PLS-DA to observe any combination of component variation that could be explained by geographical origin.

Materials and methods

Finished dark chocolate samples (x5 from each country) were obtained from Willies Cacao Ltd and were verified as being from Peru, Madagascar and Venezuela. All experiments were carried out on a 600MHz Bruker Avance III with TXI probe. For each experiment 25mg of a dark chocolate (Peru = P, Madagascar = M, Venezuela = V) was suspended in 600uL 1mM sodium 3-(trimethylsilyl)propionate-2,2,3,3-d_4(TSP) in 50/50 v/v D_2O/CD_3OD and sonicated for 10 minutes. It was then transferred in its entirety to a 5mm Norrell HT – 600MHz – 7Inch NMR tube. For each of the 5 samples (same production lot) from the 3 locations 3x replicates of this preparation were carried out to observe any intrasample variation caused by operator bias. ^{1}H spectra were obtained using a noesygppr1d pulse sequence and 65,536 complex data points over a sweep width of 20.57 ppm using a pre-saturation of the water signal at 4.7ppm and one spoil gradient. NS = 16, D1=8s. All samples were buffered (PBS) to a pH of 7.4.

Spectral binning in NMR experiments for multivariate analysis is a common technique, but essentially reduces the dimensionality down to the point that identifying a set of chemicals whose variation in concentration is important in explaining the difference between samples becomes difficult if not impossible. At this point the use of multivariate analysis only serves to describe rather than explain the data and as a result is rather limited. This is especially true if we would seek to understand and influence the principle components through causality/correlation with

a specific chemical entity e.g. through use of a specific fertiliser during cocoa bean growth.

As a result it was crucial to be able to identify the specific chemicals in the sample, when suggesting their contribution to that variation in the loadings plot. Chenomx Profiler (Chenomx NMR Suite, Chenomx, Alberta, Canada) was used with the TSP as a lineshape/quantitation standard. An automatic search through the Chenomx metabolomics database yielded 50-69 hits across all samples. It is worthy of note that Chenomx was unable to assign all signals and for the purposes of the multivariate analysis these were ignored. Chenomx cannot be used to provide the signal intensity of unassigned NMR signals. One of the reasons for this was to avoid operator bias when trying to integrate overlapping signals manually. The other was to see if a fast, semi-automated approach to this multivariate analysis was possible without recourse to manual component identification. Multivariate analysis was undertaken using the Unscrambler X (v10.1 Camo Software) with concentration data for 42 chemical components, common to all samples, being used for the multivariate analysis. This was done to avoid components specific to one or more of the chocolate samples exerting a strong influence, which would override the more subtle variations in the PCA, which it is believed are more likely to code for the geographic origin.

The Principle Component Analysis (PCA) was applied in the absence of scaling with confidence level set at 95% and minimum explained variance set at 95%.

Results and Discussion

Signals for the ^{1}H NMR were fit against a database of 332 compounds. An example of a fitted spectrum is shown in Figure. 1. Chenomx software was unable to fit all signals and for these reasons given, they did not factor in the multivariate analysis.

FIGURE 1a

FIGURE 1b

The 42 common components from all three types of chocolate are indicated in Table 1. The averages from the 5 different samples (each prepared 3 times) are also shown. The difference between the n=3 preparations was not statistically significant, indicating that the processing of the 5 samples from each geographical region was consistent.

		Concentration in NMR sample (mM)		
Compound Number (for Loadings plots)	Name of Compound	V	P	M
C1	2-Hydroxyglutarate	0.091	0.308	0.367
C2	2-Hydroxyisobutyrate	0.026	0.104	0.098
C3	2-Hydroxyvalerate	0.156	0.078	0.041
C4	3,5-Dibromotyrosine	0.056	0.066	0.056
C5	3-Chlorotyrosine	0.029	0.087	0.076
C6	3-Hydroxy-3-methylglutarate	0.078	0.044	0.08
C7	3-Phenyllactate	0.151	0.152	0.149
C8	Alanine	0.317	0.408	0.483
C9	Anserine	0.107	0.03	0.029
C10	Arabinose	0.615	1.066	0.611
C11	Betaine	0.012	0.022	0.02
C12	Biotin	0.121	0.193	0.166
C13	Theobromine	0.304	0.341	0.262
C14	Formate	0.112	0.07	0.074

C15	Fructose	0.507	0.413	0.183
C16	Fumarate	0.01	0.026	0.014
C17	Gluconate	0.376	0.747	0.373
C18	Glucose-6-phosphate	0.496	0.155	0.176
C19	Glycine	0.189	0.206	0.092
C20	Glycolate	2.688	3.022	2.467
C21	Glycylproline	0.69	0.965	0.547
C22	Guanidoacetate	0.226	0.195	0.267
C23	Homovanillate	0.079	0.054	0.036
C24	Imidazole	0.082	0.02	0.059
C25	Isocitrate	0.671	0.555	0.689
C26	Isoleucine	0.119	0.151	0.151
C27	Lactate	1.272	1.782	1.884
C28	Lactose	0.724	0.799	0.453
C29	Lactulose	0.502	0.411	0.366
C30	Mannose	0.323	0.117	0.269
C31	N-Acetylserotonin	0.051	0.038	0.033
C32	O-Phosphocholine	0.034	0.028	0.166
C33	O-Phosphoethanolamine	0.297	0.275	0.262
C34	Phenylalanine	0.212	0.291	0.268
C35	Ribose	0.311	0.251	0.151
C36	Sarcosine	0.115	0.072	0.044
C37	Threonate	0.168	0.332	0.23
C38	Trimethylamine N-oxide	0.01	0.013	0.011
C39	Valine	0.169	0.222	0.233
C40	Xylose	0.097	0.043	0.04
C41	π-Methylhistidine	0.072	0.024	0.029
C42	τ-Methylhistidine	0.019	0.023	0.009

Table 1: Identified metabolites common to all three geographical sources of chocolate V - Venezuelan, M – Madagascan, P - Peruvian. The mean average concentrations (3x5 for each origin) are given as mM in the 600uL NMR solvent with reference to 1mM TSP. All standard deviations were less than 5%.

The PCA scores plot (Figure 2) indicates clustering of samples based on geographical location. PC1 encodes for the difference in variables and indicates a strong distinction between the two South American types (Venezuelan and Peruvian). PC2 shows the difference between the Madagascan chocolate (Africa) and the two South American types. There exists a strong negative correlation (Figure 3) in PC1 between compounds glycolate and mannose, which largely describes the difference between the Venezuelan and Peruvian chocolate. The largest negatively

correlated variables between African and South American chocolate samples being alanine and 2-hydroxyvalerate.

FIGURE 2

FIGURE 3

The positively correlated components across both PC1 and PC2 are guanidinoacetate and O-phosphocholine. In all cases, the observation of multiple components being indicative of a specific chocolate type would potentially indicate the impact of a wide variety of factors, not least the origins of the cocoa beans themselves in creating a characteristic pattern of components, which persist after the several stages required to go from cocoa bean to processed chocolate.C4 and C5,for example, are unlikely in cocoa, they might be added e.g. to jutebags in transport. Acetic acid is a product of fermentation and its content reflects fermentation practice and conching practice where prolonged conching times aim at reducing chocolate acidity. Phosphoethanolamines are typical yeast and bacterial metabolites included in cocoa within fermentation. Fructose levels are largely derived from sucrose, which is added to all chocolates. So it is a marker of recipe and sugar addition rather than origin. Experimentally, it was also observed that the components caffeine and acetate had a high leverage and influence on the data, which whilst supportive of the observed discrimination in the PCA, were considered outliers in the cross validation and were hence removed. As a result of the initial PCA, a PLS-DA analysis was carried out based on the three different cocoa bean origins being specified as the response variable based on the predictor variables of observed components (C1-C42) in the matrix C4 and C5 were removed as they were deemed a post processing component and little impact was seen after their removal.

FIGURE 4

FIGURE 5.

The clustering observed in the scores plot of the PLS-DA (Figure 4) is statistically significant based on the REMSEC (Root Mean Square Error of Calibration) and R^2X/R^2Y values. As a result the X-loading weights for the individual variables were evaluated. The components homovanillate, mannose, glycine, lactose, gluconate, ribose and fructose show the best link for the PLS-DA model and it is worthy of note that the majority of the components are saccharides. The homovanillate, being a polyphenolic metabolite and the amino acid glycine are the only two components that do not fall into this category. The discussion of the biochemical reasons for this observation in the context of complex chemistry of cocoa bean processing fermentation/processing is beyond the scope of this conceptual communication on how the geographical origins of the finished chocolate can be determined.

Conclusion

In this communication we have demonstrated the potential for a semi-automated approach to finished chocolate analysis using NMR fitting software and a relatively straightforward multivariate treatment. This approach, may better enable the high throughput screening of dark chocolate samples from multiple sources to identify not just the potential geographical origin of the cocoa bean used to make it, but also to identify the types of process used in its conversation. The authors are acutely aware that differentiating different origins in finished chocolate is very ambitious. Variations in composition might result from different hybrids, plants cultivated under different agricultural conditions. The types of fermentation (type of microorganisms, starter

cultures, spontaneous); box fermentation or heap fermentation; fermentation length, temperature and pH profiles; the drying process (sun dried or mechanically or fire dried); transport and storage, roasting (temperature, time); nib or bean roasting; winnowing (before or not); alkalisation (base dependent); conching (additive dependent, e.g. amount and type of sugar/lecithin etc) and finally tempering.

The NMR data, assigned *via* automation with Chenomx and assessed *via* robust multivariate statistical analysis, indicates that additional work should be carried out with a wider geographical range of chocolate samples and further environmental factors such as where the fermented/dried cocoa beans were processed/refined should be taken into account to support/disprove the observations contained herein

The analysis we report in this communication does not require protracted and technically challenging separation of either polyphenolic or fatty acid components prior to analysis and as result considers the entirety of the chocolate chemistry *in situ* without the need for pre-treatment. Using this approach we retain some of the dimensionality afforded by NMR, which is lost when spectral bucketing is undertaken for PCA analysis of complex mixtures.

In terms of food authenticity, especially where high value products of significant economic importance to developing nations is concerned, finding a non-destructive, robust and reliable method to confirm the geographical origin of finished chocolate is crucial in order to ensure that PGI and PDO mark protection within the global chocolate industry is not open to abuse. We will be building on this initial study to assess the scope and limitations of this approach.

Acknowledgements

The authors are grateful to Kingston University for their support for this project.

Conflict of Interest

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.The authors declare no conflict of interest.

References

Afoakwa, E. (2010). Chocolate production and consumption patterns. In E. Afoackwa (Ed.), *Chocolate science and technology* (pp. 1–11). Oxford: Wiley-Blackwell.

Acierno, V., Alewijn, M., Zomera, P., van Ruth, S.M. (2018). Making cocoa origin traceable: Fingerprints of chocolates using Flow Infusion - Electro Spray Ionization - Mass Spectrometry. *Food Control*, 85, 245-252.

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.

Cambrai, A., Marchioni, E., Julien-David, D., Marcic, C. (2017). Discrimination of Cocoa Bean Origin by Chocolate Polyphenol Chromatographic Analysis and Chemometrics. *Food Analytical Methods*, 10, 6, 1991-2000.

Cambrai, A., Marcic, C., Morville, S., Sae Houer, P., Bindler, F., & Marchioni, E. (2010). Differentiation of chocolates according to the cocoa's geographical origin using chemometrics. *Journal of Agricultural and Food Chemistry*, 58, 1478–1483.

Kuhnert, N., Milev, B., Patras, M. A., Vrancken G. (2014) Fourier transform ion cyclotrone resonance mass spectrometrical analysis of raw fermented cocoa beans from Cameroon and Ivory Coast origin, *Food Research International,*64*,* 958-962.

Marseglia, A., Acquotti, D., Consonni, R., Cagliani L.R., Palla, G., Caligliani, A. (2016). HR MAS 1H NMR and chemometrics as useful tool to assess the geographical origin of cocoa beans – Comparison with HR 1H NMR. *Food Research International*, 85, 273–281.

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.

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.

Figure Captions

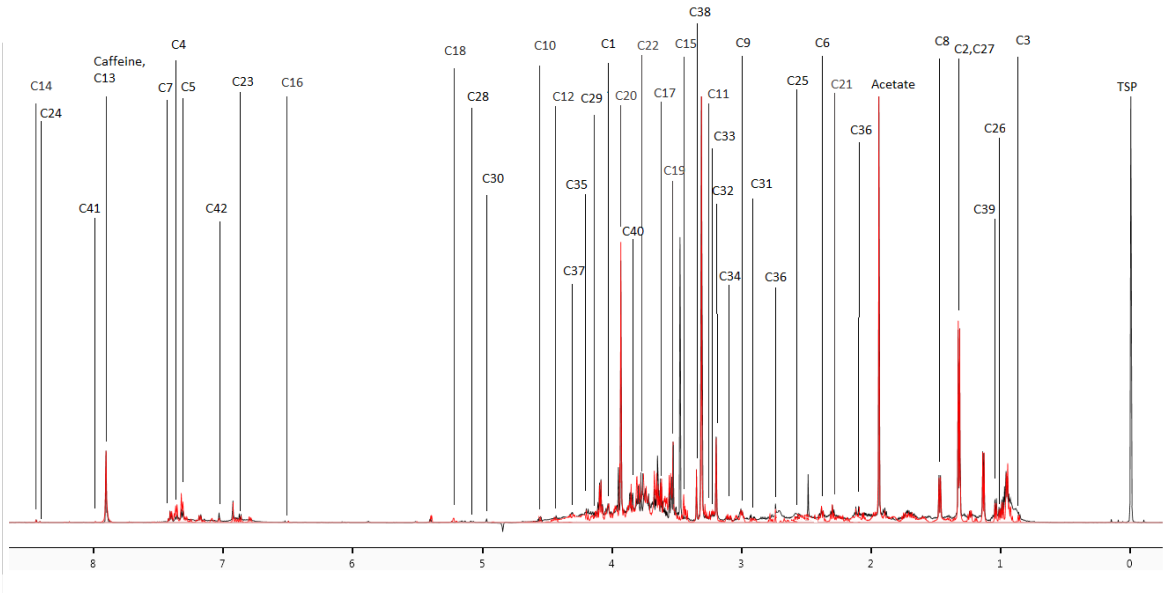
Figure 1. a) Fitted ^{1}H NMR spectrum for Madagascan chocolate (Sample 1, Replicate 1). X- Axis units: ppm. Sample NMR is in Black, Scaled fitted spectrum is in red. A single signal for each of the 42 fitted, common components is shown. b). Expanded ^{1}H NMR region for caffeine and theobromine for representative sample of Madagascan chocolate, showing reference signals (red) compared with actual signals observed (black).

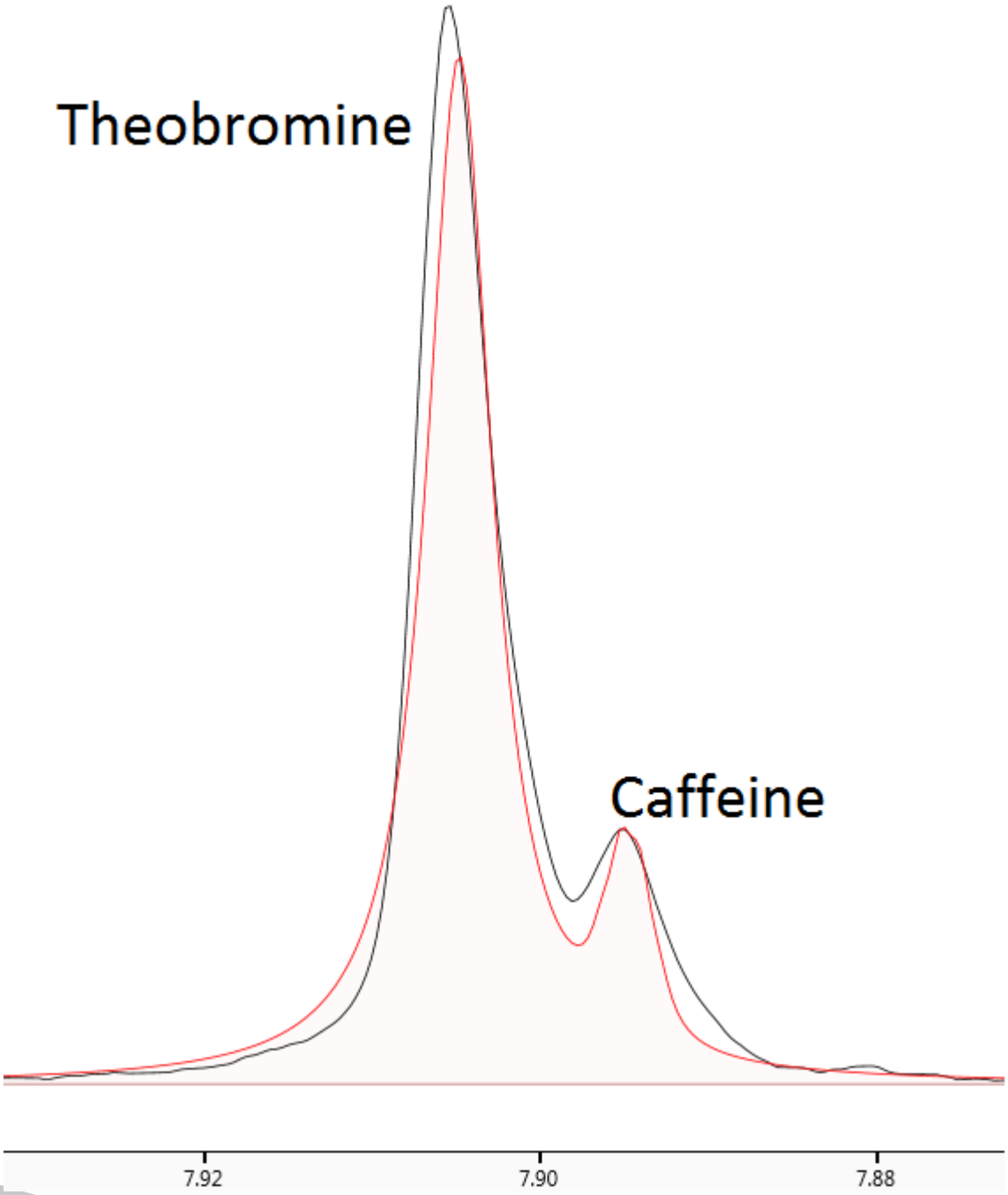
Figure 2. PCA Scores Plot for 5 x Chocolate Samples from Madagascar■, Venezuala▲and Peru ●, cross validated and with each component carrying a weighting of 1.00.

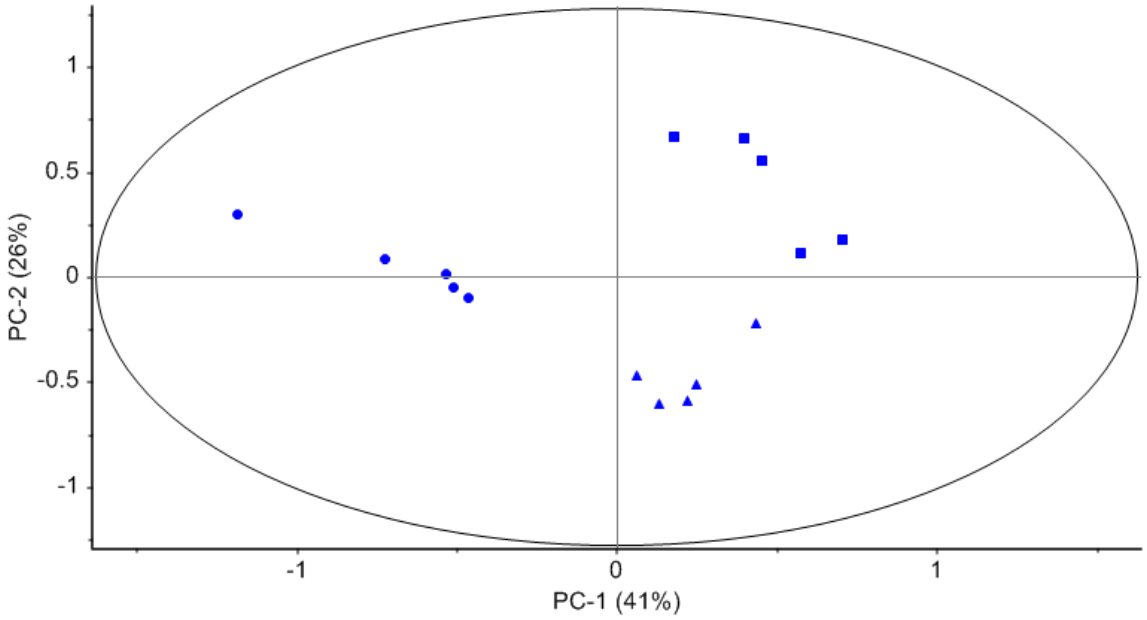
Figure 3. PCA Correlation loadings plot (X) for components C1-C41 with outer ellipse coding for 100% of observed variation between the samples.

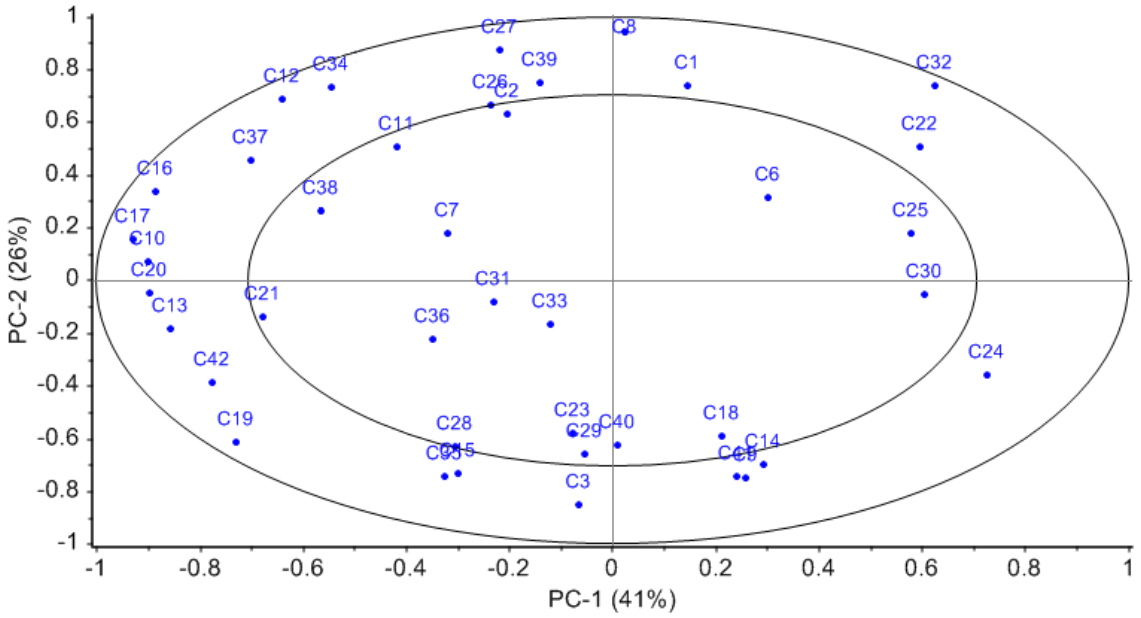
Figure 4. PLS-DA scores plot for 5 x Chocolate Samples from Madagascar■, Venezuala▲and Peru ●. R^2X = 0.998 R^2Y =0.910 and REMSEC 0.0269.

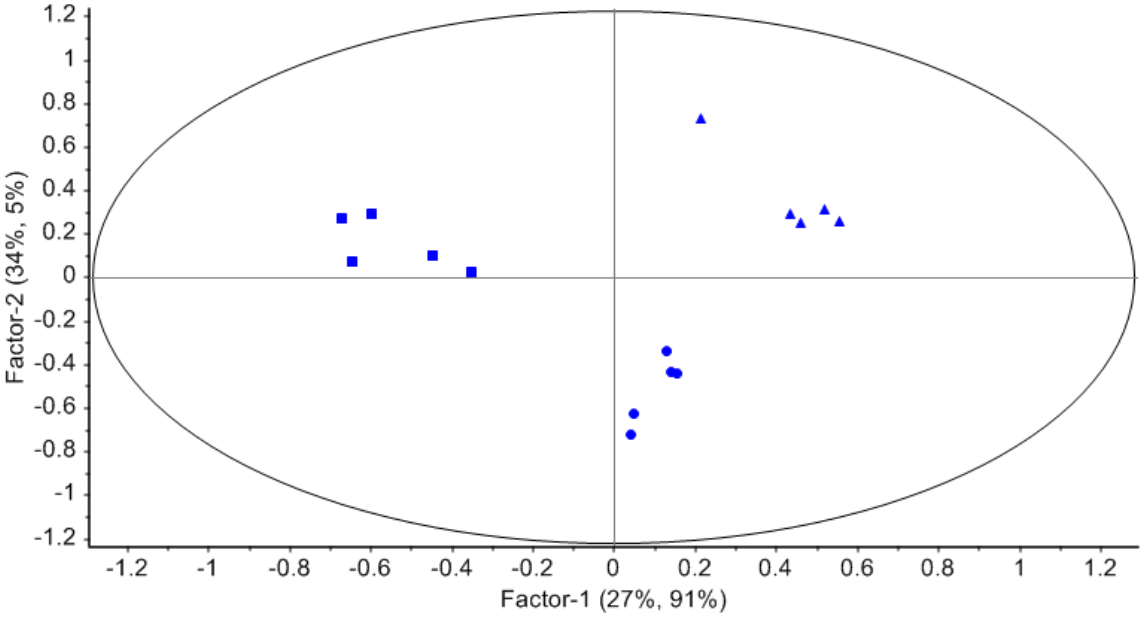
Figure 5. PLS-DA X-loadings plot for all 42 components. Those that have a large positive value show a positive link with the PLS-DA model. Those with a negligible value contribute little to the validity of the model.

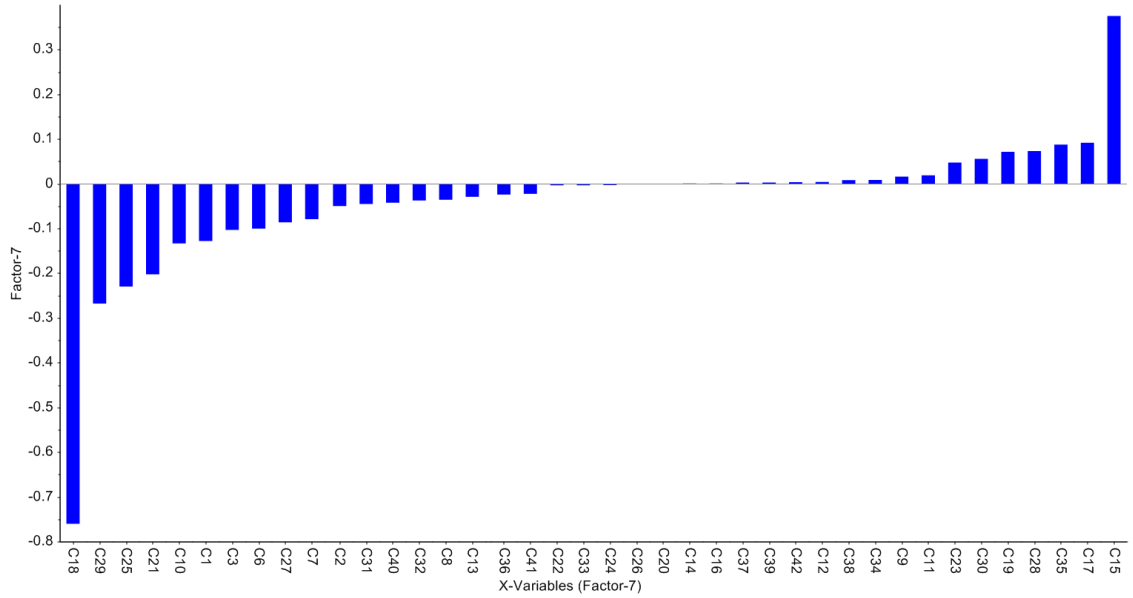












Highlights

High-resolution ^{1}H NMR analysis on finished dark chocolate.

Chocolate from South America and Africa was analysed non-destructively.

Multivariate analysis identify component clusters, which may enable distinguishing of finished chocolate based on geographical origin.