Tutors

Dr. Javier Saurina Purroy
Departament de Química Analítica

Dra. Sònia Sentellas Minguillón
Departament de Química Analítica

Grau de Química

Treball Final de Grau

HPLC-UV and HPLC-FLD polyphenolic analysis for the discrimination and characterization of dark chocolates
Discriminació i caracterització de xocolates negres mitjançant l'anàlisi de polifenols per HPLC-UV i HPLC-FLD

Pablo Pardo Menacho
(June 2021)

Universitat de Barcelona

B:KC Barcelona Knowledge Campus
Campus d'Excel·lència Internacional

Quiero agradecer en primer lugar a mis tutores, el Dr. Javier Saurina Purroy y la Dra. Sònia Sentellas Minguillón, por toda su ayuda y dedicación durante estos meses.

Agradecer también la ayuda y el trato recibido por los compañeros del laboratorio de quimometría; en especial a Tamara Parada, ya que sus consejos han sido muy importantes durante la realización de este proyecto.

Por supuesto también quiero agradecer a mi familia y amigos por todo el apoyo brindado, no solo durante el TFG sino a lo largo de toda la carrera. Su cariño y amistad me motivan a seguir adelante.

REPORT

CONTENTS

1. SUMMARY

The popularity of chocolate has increased greatly in recent years not only because it is considered a delicatessen, but also because the high polyphenolic content of cocoa provides to chocolate some interesting health properties, mainly antioxidant and anti-carcinogenic activities. However, cocoa production has been involved in polemics, as the obtention of this praised food has been related to slavery, biodiversity loss, and deforestation. In order to ensure consumers that the products they buy are environmentally and socially responsible, the development of reliable chocolate authentication methods is required.

The aim of this work is to classify and characterize chocolate samples based on their geographical origin, cocoa variety and cocoa content by analysing their polyphenolic composition using high-performance liquid chromatography with UV (HPLC-UV) and fluorescence detection (HPLC-FLD). The chromatographic separation was performed on a C18 reversed phase column with an elution gradient using water 0.1 % in formic acid and acetonitrile as the mobile phases. The chromatographic data of 42 chocolate samples from 4 different bean varieties (Criollo, Forastero, Nacional and Trinitario) and 3 different geographical origins (Africa, America and Asia) were analysed using principal component analysis (PCA) and partial least squares – discriminant analysis (PLS-DA). The obtained score plots were used to achieve the classification and discrimination of the chocolates depending on their variety, origin and cocoa content.

Keywords: Cocoa, chocolate authentication, HPLC-UV, HPLC-FLD, PCA, PLS-DA, polyphenols.

2. RESUM

La popularitat de la xocolata ha crescut extremadament en els darrers anys no només pel fet de ser considerat com una "delicatessen", sinó també perquè el seu elevat contingut de polifenols dona a la xocolata unes propietats saludables molt interessants, principalment per la seva activitat antioxidant i antiinflamatòria. Però, la producció de cacau s'ha vist envoltada de certes polèmiques, ja que l'obtenció d'aquest valuós aliment s'ha relacionat en alguns casos amb esclavitud, pèrdua de biodiversitat i desforestació. Per tal d'assegurar als consumidors que els productes que compren son responsables amb el medi ambient i amb la societat, el desenvolupament d'un mètode fiable d'autentificació de la xocolata és necessari.

L'objectiu d'aquest treball és la classificació i caracterització d'unes mostres de xocolata en funció del seu origen geogràfic, de la seva varietat de cacau i del seu contingut de cacau analitzant la seva composició polifenòlica mitjançant un mètode de cromatografia de líquids d'alta eficàcia amb detecció UV (HPLC-UV) i amb detecció de fluorescència (HPLC-FLD). La separació cromatogràfica s'ha dut a terme mitjançant una columna de fase inversa C18 amb elució de gradient utilitzant aigua acidificada amb àcid fòrmic al 0.1 % i acetonitril com a fase mòbil. Les dades cromatogràfiques de 42 mostres de xocolata de 4 varietats diferents (Criollo, Forastero, Nacional i Trinitario) i de 3 orígens geogràfics diferents (Àfrica, Amèrica i Àsia) s'han analitzat mitjançant anàlisi de components principals (PCA) i regressió de mínims quadrats – anàlisi discriminant (PLS-DA). Els mapes de scores obtinguts han permès la classificació i discriminació de les diferents mostres en funció del seu origen, varietat i contingut de cacau.

Paraules clau: Cacau, autentificació de xocolata, HPLC-UV, HPLC-FLD, PCA, PLS-DA, polifenols.

3. INTRODUCTION

3.1. COCOA AND CHOCOLATE

Cocoa, the seed of the tree *Theobroma Cacao,* is the main ingredient of chocolate and its derivate products. The cacao tree naturally grows in the Amazonian and Orinoco basins, but humans cultivate it in all the tropical zones in the World; in fact, it is not America but rather Africa the world's largest producer of cocoa nowadays. The whole of Africa represents 70% of the total production of cocoa beans, with Ivory Coast alone sharing 40% of the global produce of this food. The next major producer is Asia representing 17% of the global production of cocoa. The remaining 13% corresponds to America's cocoa bean production (1).

Pre-Columbian civilizations already consumed beverages made from grinded cocoa beans long before the arrival of the Europeans, although only the most relevant people of their societies were allowed to consume these drinks. When Europeans arrived in America, they brought back cocoa with them to the old continent, where the popularity of this exotic food grew. To reduce the bitter taste of cocoa sugar was added to cocoa products, which ultimately led to the creation of chocolate (2).

The popularity of chocolate, and dark chocolate in particular, has grown significantly in recent years as many health benefits such as anti-inflammatory and anti-carcinogenic activity have been related to its consumption. Many of the attributed properties are thought to be related to the high polyphenolic content of chocolates and cocoa products (3).

3.1.1. Cocoa varieties

Usually *Theobroma Cacao* trees are divided into three major varieties, namely: Criollo, Forastero and Trinitario.

Criollo cocoa is considered a high-quality variety due to its pleasant texture, scents, and flavours. However, it is a more fragile variety that needs optimum environmental conditions to correctly develop.

Forastero beans belong to a much more resistant variety but do not have the taste and aroma of Criollo beans so it less appreciated by chocolate experts. Due to its resilience and the lower price of the beans, Forastero is the most commonly grown variety worldwide, corresponding to about 95% of global cocoa production(4). Forastero variety has many subvarieties; perhaps the

most famous is the one known as Nacional or Arriba Nacional, grown exclusively in Ecuador. Nacional is considered an independent variety in many cases because it has much better taste and aroma than the rest of Forastero cocoas (5).

From the hybridization of Criollo and Forastero humans developed a variety known as Trinitario, as it was first cultivated in Trinidad. This hybrid conserves the characteristic and desirable taste of the Criollo variety whilst being resistant and bulky as the Forastero, thus making it a very popular variety worldwide (1,6).

The variety of cocoa used by manufacturers to produce chocolate will depend greatly on the desired quality of the final product; usually Criollo and Nacional are the choice for high-quality chocolates, whereas Forastero and Trinitario are chosen when a cheaper product is desired. The most common varieties planted in different countries of the world are shown in figure 1.

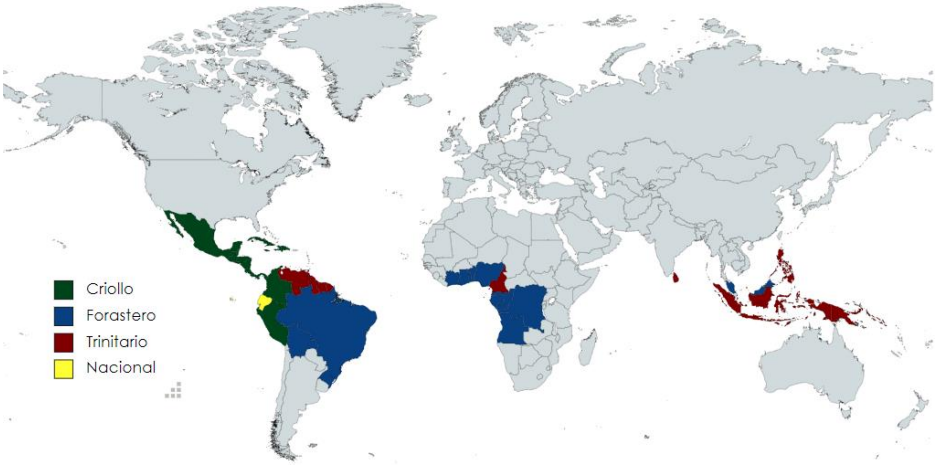


Figure 1. Most commonly grown variety in the largest cocoa producing countries.

3.1.2. Chocolate production

Chocolate production is a complex process that requires many different steps. These stages in the transformation from cocoa to chocolate can cause great changes in the composition of the chocolate (7), especially on its polyphenolic composition, as we will see further on in this section.

Once the cocoa pods have been harvested, the first two steps of the process, fermentation and drying, usually take place in the same region where the cocoa has been cultivated.

The aim of fermentation is to avoid the germination of cocoa seeds; although there are many different fermentation methods, they usually take less than a week and occur at a temperature of around 50 ºC. The process of fermentation is carried out mainly by yeasts present in cocoa pods and the main products generated are organic acids (like acetic acid and lactic acid), alcohols and esters. Some interesting compounds like caffeine and theobromine begin to appear during this stage also. During the fermentation many oxidative processes take place; polyphenol oxidase is one of the enzymes involved, which causes the oxidation and polymerization of polyphenols. This polymerization forms insoluble tannins, thus reducing drastically the content of polyphenols in cocoa, which in turn reduces the astringency and bitterness of the beans. After fermentation comes drying; this is done to avoid mold development during the transportation of cocoa beans from their origin to the chocolate producing countries (8). The oxidation and polymerization of polyphenols still continues to happen during drying although to a lesser extent. These first two stages cause the biggest reduction of polyphenol content in cocoa beans; in some cases, a decrease of up to 90% of the total content of (-)-epicatechin (the most abundant polyphenol in chocolate) has been observed (9).

Once the cocoa beans have been transported to the chocolate producing areas a roasting of the beans takes place. This process may last from 10 minutes to an hour and temperatures range from 110 to 140 ºC. The purpose of roasting the cocoa beans is to avoid bacterial infection and to toast the cocoa cotyledons, but many other chemical processes are taking place. For instance, bitterness is reduced due to the volatilization of acetic acid, cocoa gains its characteristic dark-brown colour as a consequence of starch hydrolysis (dextrinization), Maillard reactions and oxidation of polyphenols; lastly, some flavour and smell compounds of chocolate are formed as products of different Maillard reactions and Strecker degradations (10,11).

Prior to roasting the beans some chocolate producers choose to do a process known as alkalinization or dutching to further reduce the natural bitterness of cocoa, as well as to give the final product a much darker colour. Usually a sodium or potassium carbonate solution is used, which will neutralize the organic acids present in cocoa thus reducing the bitter taste and astringency of the final product (12). The alkaline pH also produces degradation of polyphenols, mainly flavonoids, which will further reduce the bitterness of the product and give it a brownish colour (9).

After the cocoa beans have been roasted the final stages of the process begin: grinding, pressing, conching and tempering. The cocoa beans are finely grinded and pressed so that the fat content of cocoa (cocoa butter) can be separated from what is known as cocoa liquor or cocoa mass. Conching is a long process that can take days in which the less volatile lactic and oxalic acids are eliminated, moisture is reduced and cocoa liquor is mixed with sugar, milk, cocoa butter or other additives to give the final product the desired viscosity. After conching is completed the chocolate is tempered to give it a more appealing appearance, and that is how the chocolate bar is finally obtained (7,11).

3.2. COCOA COMPOSITION

Although fats, proteins and carbohydrates are the major components of cocoa, it also contains considerable amounts of polyphenols and alkaloids which arises a lot of interest in chocolate and cocoa products from both a chemical and medical perspective.

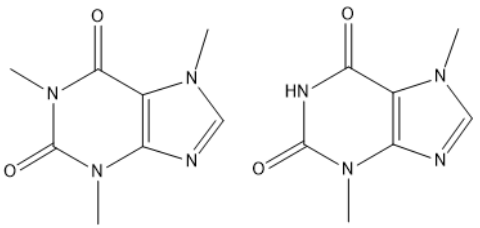
3.2.1. Macronutrients

Cocoa is well known for having a high fat content, easily corresponding to 50% or 60% of the total cocoa liquor. The main fatty acids in cocoa beans are oleic acid (C18:1) and stearic acid (C18:0), with each representing well over 30% of the total fat content. Palmitic acid is another unsaturated fatty acid that is fairly common in cocoa, in some cases representing 25% of the total fat content. The protein content in cocoa can be quite high also, representing 15% of cocoa beans and around 10% of the total dry weight of cocoa liquor. Starch and fiber contents can vary greatly from one case to another, although they often represent 12% and 6% of cocoa mass respectively (3).

3.2.2. Alkaloids

Alkaloids are some of the most common metabolites found in plants, they derive from amino acids and thus contain nitrogen atoms. Many alkaloids have psychoactive properties; some examples are caffeine, cocaine or nicotine.

In cocoa we mainly find methylxanthines, which are a specific type of purine alkaloids derived from cysteine and partly responsible for the bitter taste of cocoa. Theobromine is the most important xanthine and alkaloid in cocoa, as it represents 2% of the total cocoa mass. Actually, the name theobromine comes from Theobroma, that as we mentioned in section 3.1 is the genus to which the cocoa tree belongs. Caffeine, another methylxanthine present in cocoa, is the precursor of theobromine and its content in cocoa is around 0,5% (13,14). The structures of caffeine and theobromine are shown in figure 2.



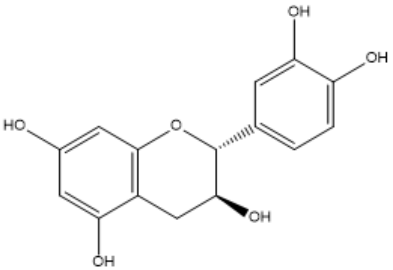
1 **2**

Figure 2. Structures of caffeine (1) and theobromine (2)

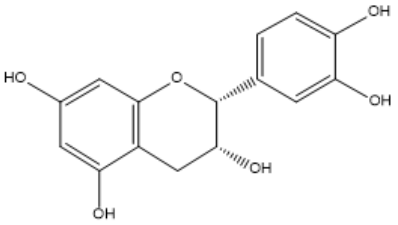
3.2.3. Polyphenols

Polyphenols are secondary metabolites of plants that form a large group of chemical substances that contain at least one phenol group on its structure. In recent years polyphenols have attracted a lot of interest as they are the most important group of natural antioxidants, present in many different foods like coffee, grapes, turmeric and, of course, chocolate. Apart from their great antioxidant activity, polyphenols are believed to have some impressive health benefits such as anti-carcinogenic and anti-inflammatory properties (3).

Polyphenols in cocoa are partly responsible for the characteristic taste, flavour and dark brown colour of cocoa products. The most important group of phenolic compounds present in cocoa are flavonoids; although small amounts of other polyphenols such as hydroxybenzoic acids and hydroxycinnamic acids have been found (1,15).



3



4

Figure 3. Structures of some common flavanols in cocoa: (+)-catechin (3) and (-)-epicatechin (4)

Flavonoids have a common structure formed by two phenol rings and a central heterocyclic ring with an oxygen atom; however, they can be classified into many subgroups as follows: anthocyanins, anthoxanthines, flavones, flavanols, isoflavonoids, etc. Flavanols (flavan-3-ols or catechins) are the most abundant kind in cocoa beans (figure 3); among which (-)-epicatechin is usually found in higher quantities, accounting in some cases more than a third of the total phenolic content of cocoa (16). Other flavanols found in smaller quantities include (+)-catechin, (+)-gallocatechin and (-)-epigallocathechin. When (-)-epicatechin and/or (+)-catechin are linked together form oligomers known as proanthocyanidins (PACs) or condensed tannins (figure 4), which are quite common in cocoa, as they can make up to 60% of the total phenolic content. The most common polymers are dimers (procyanidins A2 and B2) or trimers (procyanidin C1). Other groups of flavonoids present in cocoa, although in smaller amounts, are anthocyanidins and flavonols; which are commonly found as glycosides such as cyanidin-3-O-galactoside or quercetin-3-O-arabinoside (1,15,17,18).

Hydroxycinnamic acids and hydroxybenzoic acids are other families of polyphenols that can be found in cocoa beans in smaller concentrations (figure 4). Ferulic acid, coumaric acid, caffeic acid and chlorogenic acid are the main cinnamic acid derivates; whilst 3,4-hidroxibenzoic acid, gallic acid, vanillic acid and syringic acid are the most important hydroxybenzoic acids (1,18,19).

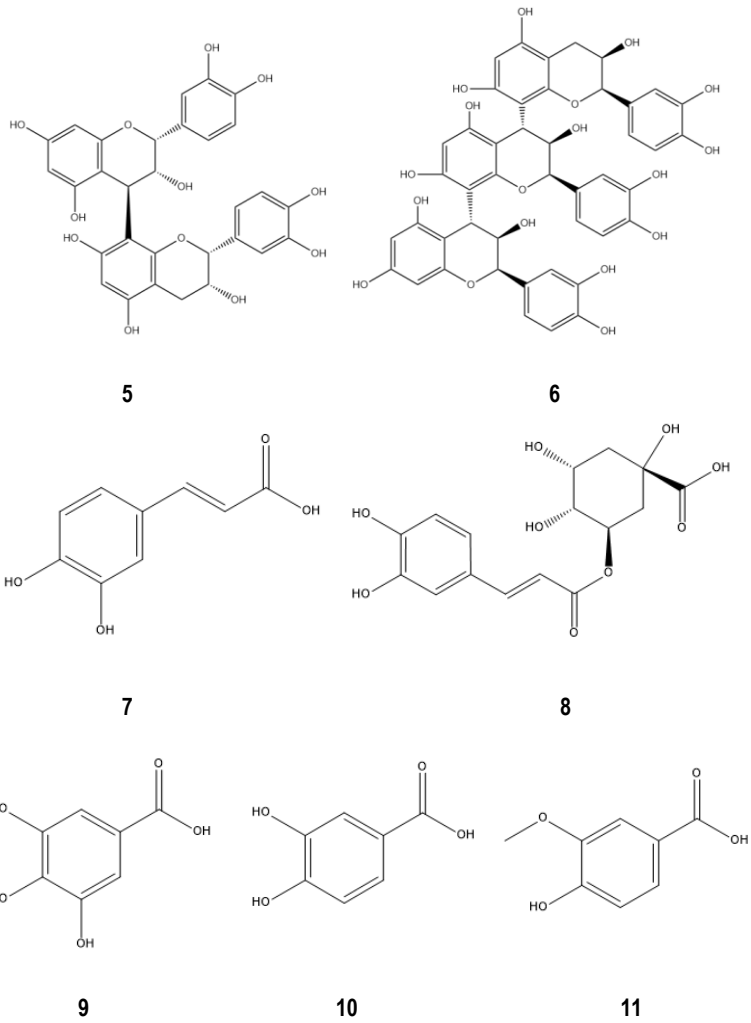


Figure 4. Procyanidins B2 (5) and C1 (6), caffeic acid (7), chlorogenic acid (8), gallic acid (9), 3,4-hydroxibenzoic acid (10) and vanillic acid (11).

3.3. CHOCOLATE AUTHENTICATION

As a consequence of the growing popularity of chocolate products in recent years the need to develop a reliable authentication method for chocolate origins and varieties is pressing. In fact, cocoa harvesting and chocolate production has been related to extensive forestry, carbon dioxide

emissions, biodiversity loss, and in some cases even slavery. Moreover, consumers often consider certain cocoa varieties or origins to be of higher quality, thus chocolate authentication is not only important for environmental and humanitarian purposes, but also to assure the quality of the products consumers acquire (20,21).

Many methods have been developed to discriminate between different cocoa varieties, origins and manufacturing process; some of which include: metabolic profiling by ^{1}H-NMR spectroscopy (4), multi-elemental fingerprinting by ICP-MS (22), fatty acid profiling by GC-MS (23) or polyphenolic profile by HPLC (24,25).

In our particular case, the methods developed by Cambrai et al. (24) and D'Souza (25) are the ones with the most interest. In the first case, the polyphenolic content of 47 different chocolates from 12 different geographical origins and 4 cocoa varieties was determined by an HPLC-DAD-MS method. Twenty-one different variables, corresponding to twenty different chromatogram peaks and the total polyphenolic content, were used to build a principal component analysis (PCA) model that allowed the researchers to correctly classify chocolates both by their origin and variety. In D'Souza's study an UHPLC coupled to time-of-flight MS analysis was used to obtain the polyphenolic fingerprinting of 86 samples of fermented and unfermented cocoa beans from six countries (Indonesia, Ivory Coast, Tanzania, Malaysia, Ecuador and Brazil). PCA was performed on 57 peaks present on the samples, and the built model was able to discriminate beans based on their fermentation; there was also good separation between some countries of origin, although not nearly as good as for the fermentation of the beans.

3.4. CHEMOMETRICS

3.4.1. Principal Component Analysis

PCA is an exploratory method widely used to detect trends and patterns of variation in complex sample pools. PCA is categorized as a projection method because it transports the original multidimensional data set (X-matrix, of values or responses of different variables) to a low-dimensional space constituted by optimized variables known as principal components (PCs).

PCs are orthogonal to each other and are formed by a linear combination of the multiple variables that described the original data. Every PC explains a certain amount of the variation of the samples; so, for example PC1, the first principal component, corresponds to the direction of maximum data variance. The following PCs: PC2, PC3 and so forth, explain less and less

variance. The first PCs are the most useful to data analysts, as they provide information about general trends in the data set, whereas higher PCs may describe individual sample differences or instrumental noise.

Once the number of PCs desired to build the PCA model is decided, the samples can be projected into the PC space thus obtaining the scores plot. The position of samples in the scores plot is used to identify patterns and to classify samples into groups. Objects that are close to each other in the PC space will have similar values of the original variables (concentrations, absorbances…) whilst samples that are further away in the scores plot will have greater differences between them. Similarly, the original variables can be projected into the PC space of loadings, allowing us to know how each principal component is affected by the variables. Since scores and loadings are represented in the same PC, by comparing the two graphs we can determine which variables are the most important to distinguish between the different sample groups identified on the scores plot (26,27).

3.4.1. Partial Least Square – Discriminant Analysis

Partial least squares – discriminant analysis (PLS-DA) is a supervised classification method derived from PLS. In PLS-DA, the X matrix corresponds to experimental data of the samples whilst the Y matrix contains qualitative information, for example origin or variety in the case of chocolate. PLS-DA will relate the experimental data with the class belonging to build a classification method that can be used later on to classify unknown or known samples (28,29).

PLS consists on performing simultaneously a PCA on an X and a Y matrix to obtain a combination of PCs that extract the maximum information about the correlation between X and Y. These new hybrid PCs are known as latent variables; and much like principal components, LVs progressively explain less variance.

4. OBJECTIVES

The goal of this project is to develop a method for the classification and discrimination of dark chocolates depending on their variety and origin based on their polyphenolic composition, which will be determined by HPLC-UV and HPLC-FLD. To do so, the steps that will be taken are:

1. To perform a defatting of the samples and subsequently, extraction of polyphenols with a previously optimized and validated method.
2. Analysis of the polyphenolic content from 42 chocolate samples from different geographical origins, varieties, and cocoa content using an HPLC-DAD-UV and HPLC-FLD method.
3. Identification of the main polyphenols and alkaloids present in chocolate samples.
4. To build PCA and PLS-DA models using 16 variables obtained from the chromatographic data that allows to classify and characterize chocolates regarding their geographical origin and cocoa variety.

5. EXPERIMENTAL SECTION

5.1. STANDARDS AND REAGENTS

Individual stock standards of 5000 mg L^{-1} were prepared in dimethyl sulfoxide (DMSO, VWR chemicals). These stock solutions were used to prepare working and calibration solutions of 20 mg/L. The phenolic compounds used for identification purposes were as follows:

- Catechol, Sigma Aldrich (St Louis, USA).
- 4-hydroxybenzoic acid, Sigma Aldrich (St Louis, USA).
- p-coumaric acid, Sigma Aldrich (St Louis, USA).
- Ellagic acid, Sigma Aldrich (St Louis, USA).
- Resveratrol, Sigma Aldrich (St Louis, USA).
- Gallic acid, Sigma Aldrich (St Louis, USA).
- Epigallocatechin, Carbosynth (Berkshire, UK).
- Chlorogenic acid, Sigma Aldrich (St Louis, USA).
- Vanillin, Fluka (Germany).
- Quercetin, Sigma Aldrich (St Louis, USA).
- 2,5-dihydroxybenzoic acid, Sigma Aldrich (St Louis, USA).
- (-)-epicatechin, Fluka (Germany)
- 4-Methyl catechol, Sigma Aldrich (St Louis, USA).
- Ferulic acid, Sigma Aldrich (St Louis, USA).
- 4-ethyl catechol, Sigma Aldrich (St Louis, USA).
- 3,4-dihydroxybenzoic acid, Sigma Aldrich (St Louis, USA).
- (+)-catechin, Fluka (Germany).
- Vanillic acid, Sigma Aldrich (St Louis, USA).
- Procyanidin A2, Phytolab (Vestenbergsgreuth, Germany).
- Caffeine, Sigma Aldrich (St Louis, USA).
- Procyanidin B2, Chengdu Biopurify Phytochemicals LTD (China).
- Caffeic acid, Sigma Aldrich (St Louis, USA).
- Procyanidin C1 Phytolab (Vestenbergsgreuth, Germany).
- Theobromine, TRC (Toronto, Canada)
- Ruthin, Sigma Aldrich (St Louis, USA).

In order to identify the different polyphenols five standard solutions were analysed. Each of the solutions contained a mixture of different compounds, 20 mg/L for all standards, of which the elution order had already been determined by the research group.

Table 1. Elution order and composition of the used standard solutions.

Elution Order	1st Standard (T1)	2nd Standard (T2)	3rd Standard (T3)
1	Catechol	Gallic acid	2,5-dihydroxibenzoic acid
2	4-hydroxibenzoic acid	Epigallocatechin	(-)-epicatechin
3	p-coumaric acid	Chlorogenic acid	Methyl catechol
4	Ellagic acid	Vanillin	Ferulic acid
5	Resveratrol	Quercetin	4-ethylcatechol
Elution Order	**4th Standard (T4)**	**5th Standard (T5)**	
1	3,4-dihydroxibenzoic acid	Procyanidin A2	
2	(+)-catechin	Caffeine	
3	Vanillic acid	Procyanidin B2	
4	Methyl catechol	Caffeic acid	
5	-	Procyanidin C1	

Cyclohexane ≥99.9% (Sigma-Aldrich, analytical reagent grade), methanol ≥99.9% (Panreac, UHPLC grade) and formic acid ≥99.9% (Sigma-Aldrich, ACS grade) were used for the defatting of chocolates and for the polyphenol extraction. Mili-Q grade water and acetonitrile (Panreac, HPLC quality) were used for the chromatographic separation. The Mili-Q water used for the HPLC analysis was obtained by purifying water through an Elix-3 purifying system (Millipore, Bedford, MA, USA) and a 0.22 µm nylon filter integrated in the Mili-Q system.

5.2. SAMPLES AND SAMPLE TREATMENT

5.2.1. Sample pool

A total of 42 different dark chocolates available in supermarkets and specialized shops were analysed in this work. The samples can be classified into different groups regarding their origin, variety and cocoa content (see appendix 1); however, there are 3 samples from which both bean origin and variety are unknown and 5 others for which only the variety is unknown. Moreover, some chocolates like Lindt or Suchard are made from a mixture of cocoa varieties and/or origins.

Samples can be classified in terms of origin in three groups that correspond to three continents: Africa, America and Asia; regarding cocoa variety, the analysed chocolates contain at least one of the following varieties: Criollo, Forastero, Nacional or Trinitario. The cocoa content varied from 60% to 100%, with the majority of chocolates (30 out of 42) being in the range of 70-85%.

5.2.2. Chocolate defatting

As we have seen in section 3.2.1, the fat content of cocoa is really high, in some cases contributing to more than 50% of the total chocolate mass. To avoid possible interferences, the first step required before analysing the samples was a qualitative defatting of chocolates.

The performed procedure was based on the method used elsewhere for the study of similar samples using liquid chromatography (30,31). The defatting process consisted of adding 5 mL of cyclohexane for every gram of sample; the mixture was stirred in a vortex until no sediment was observed on the bottom of the vial, left in an ultrasound bath for 15 minutes and finally, centrifuged for 5 minutes at 4500 rpm. The overfloating liquid (triglyceride solution in cyclohexane) was collected and kept in the fridge as it could be used for future studies. This whole process was carried out three times for every sample to ensure that most of the fat had been extracted. Finally, the defatted cocoa was left to dry.

5.2.3. Polyphenol extraction

Once the samples had been defatted, the polyphenols needed to be extracted in order to perform the HPLC analysis. The optimization of the polyphenol extractant and methodology is explained later on in this work.

The polyphenol extraction method was as follows: 4 mL of a 60:40 (v:v) methanol/water and 1% formic acid solution were added to 0.2 g of defatted chocolate and the mixture was stirred in a vortex, then sonicated for 15 minutes at room temperature and lastly centrifuged for 5 minutes at 4500 rpm. The overfloating liquid was collected with a syringe, filtered using a 0.22 μm nylon filter and kept in the freezer until analysis to avoid degradation.

5.3. INSTRUMENTATION AND METHODS

The chromatographic separation was carried out in an Agilent 1100 HPLC with a binary pump (G1312A), a vacuum degasser (G1322A), dual detection with a diode array detector (G1315B) in line with a fluorescence emission detector (G1321A) and an automatic sampler (G1367A). The software used was Agilent ChemStation.

The column used was a 150 x 4.6mm reversed phase Kinetex C18 of 100 Å pore size and 2.6 µm particle size from Phenomenex. The flow rate used was 0.7 mL/min and the two solvents used as mobile phase were a 0.1% formic acid in water (A) and acetonitrile (B). The elution gradient had already been optimized and is shown in table 2. HPLC-UV chromatographic data of the samples was acquired at three wavelengths: 280 nm, 325 nm and 370 nm, whereas the HPLC-FLD data was acquired using 280 nm as the excitation wavelength and 330 nm as the emission wavelength.

Table 2. Elution gradient used for the chromatographic separation.

Time (min)	%B
0	18
20	48
22	90
24	90
24.2	18
29	18

To be able to examine the repeatability of the method, a quality control sample (QC) made from 0,07 mL of every sample was analysed at the beginning and the end of the sequence of analysis as well as every 10 chocolate samples. In order to avoid the influence of possible instrumental drift during the analysis the samples were injected randomly throughout the chromatographic sequence. The standard solutions exposed in section 5.1. were also included in the analysis to tentatively identify some of the polyphenols present.

5.4. DATA ANALYSIS

PCA and PLS-DA models were calculated using SOLO, a standalone chemometrics software developed by Eigenvector Research.

The X-matrix used in both models was built combining data from the HPLC-UV and HPLC-FLD analysis. Out of the 16 variables used to form the matrix, 14 corresponded to the area of the most important peaks observed on the HPLC-FLD, HPLC-UV (λ=280nm) and HPLC-UV (λ=370 nm) chromatograms; the remaining two variables correspond to the total areas of the HPLC-UV (λ=280nm) and HPLC-FLD chromatograms. Since the data of the matrix came from different detection methods, autoscaling was applied as the data preprocessing method to equalize the contribution of major and minor components to the model. Moreover, the peaks selected to build the data matrix were all different, since using the data of the same compound from HPLC-UV and HPLC-FLD would be redundant. The Y-matrix, necessary only in PLS-DA, contained the class membership of each sample regarding their origin or variety.

Once the PCA and PLS-DA models had been calculated, the scores and loading plots were used to discriminate chocolates by variety, origin and cocoa content as well as to determine possible sample structures and patterns.

6. RESULTS AND DISCUSSION

6.1. CHOCOLATE DEFATTING

The defatting process carried out on the chocolate samples was necessary to avoid any difficulties that triglycerides could cause in the extraction of polar compounds. Although this process may not seem too necessary at first glance since the polyphenols in chocolates were firstly studied by UV-absorption and fluorescence emission; it was crucial for a further analysis by HPLC-MS were the presence of triglycerides would definitely cause interferences.

Once all the samples were defatted and dried, they were weighed. By comparing the initial weight of the sample and the weight after the defatting process, is possible to determine the percentage of fat and then to infer if the defatting had been quantitative or if a big percentage of triglycerides remained in the samples. In Appendix 2 the obtained and nominal fat percentages, the fat recovery, and the cocoa content of every sample is given. Overall, we can say that the defatting was done correctly, since the obtained percentages in this work are quite similar to the ones specified on the chocolate packages. Recovery of fat ranged from 82.3% to 114.3%. The lowest values correspond to chocolates with very high cocoa contents (90% or above).

In figure 5, the correlation between the obtained and the nominal fat content is shown. As can be seen in the graph, the trend line has a slope close to 1, which suggests that although there are some slight differences, the data obtained in this work and the data reported by manufacturers are very similar. This regression is statistically significant thus proving the good agreement among declared and experimental values.

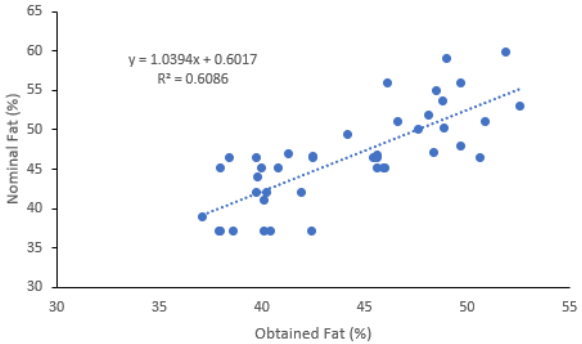


Figure 5. Representation of the obtained fat content versus the content specified by the manufacturers.

In order to ascertain the relevance of the defatting process on the recovery of the polyphenols present in the samples, 4 random chocolates were selected and, for each one, a non-defatted sample, a defatted sample and the lipidic fraction obtained during the defatting process were analysed by HPLC-UV at 280 nm. As an example, the obtained chromatograms for X6 chocolate are shown in figure 6. As can be seen very few peaks and at very low intensity appear for the lipidic fraction. The two main peaks observed in the chromatogram corresponded to theobromine (at 6.0 min) and caffeine (at 9.7 min).

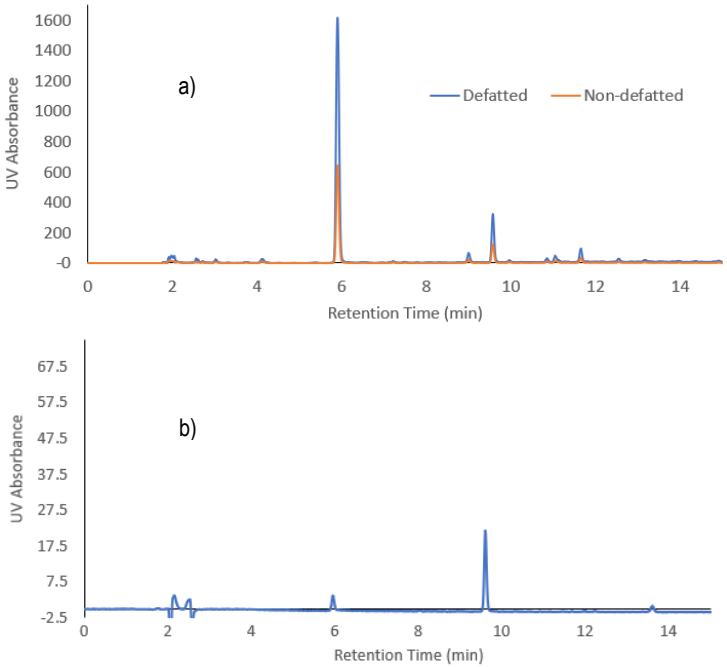


Figure 6. Chromatograms of X6 sample obtained by HPLC-UV at 280 nm: defatted and non-defatted (a) and lipidic fraction (b).

6.2. OPTIMIZATION OF THE POLYPHENOL EXTRACTION

Different variables, namely: extraction solvent, MeOH/H2O ratio, formic acid content, extraction temperature and extraction time were studied in order to optimize the extraction of polyphenols from defatted chocolates. For optimization purposes, to avoid high variability due to particularities of the different samples, all the studied variables were optimized using a single

sample. The total areas for the HPLC-UV chromatograms of the sample at 280 nm obtained for the different variables studied were analysed using ANOVA tests to determine the optimum extraction conditions. Each condition was assayed in triplicates.

The first step was to select the solvent that would be used for the extraction; and to do so, 10 different solvents were used to extract the polyphenols of the same chocolate sample and the solution obtained was then analysed by HPLC-UV at 280 nm. Figure 7 shows the obtained total areas for each of the solvents evaluated. The three best options were clearly acetone/water, ACN/water and MeOH/water as in the obtained chromatograms for these solvents appear more peaks and with more intensity than with any of the other extractants used. Although MeOH/H_2O did not provide the biggest areas at 280 nm, it was chosen as the solvent because it was the one that provided the best signals overall for the three wavelengths (280 nm, 325 nm and 370 nm) at which the analysis of all the samples would be carried out.

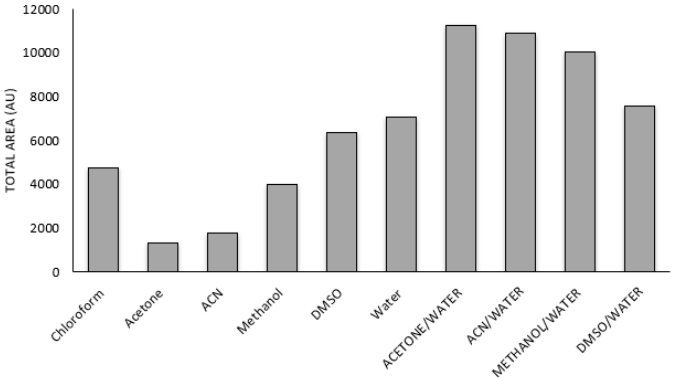


Figure 7. Total areas obtained in the chromatograms of every solvent used for the polyphenol extraction.

Once MeOH/water had been chosen as the solvent used for the polyphenol extraction, the next step was to determine the optimum percentage of methanol in water. The extraction was carried out using MeOH/H_2O solutions containing different amounts of methanol (20%, 40%, 60% and 80). The ANOVA test indicated that there were no statistically significant differences in the polyphenol extraction based on the percenatge of methanol used. However, when the areas of the minor compounds were considered there were significant differences for the different solvents used. However, the differences were due to the extraction with a 20% solution, which means that

with the 40%, 60% and 80% solutions the obtained results were very similar. The 60:40 MeOH/H_2O solution was arbitrary selected as the extractant because it was the middle ground and because we obtained the minimum standard deviation (figure 8).

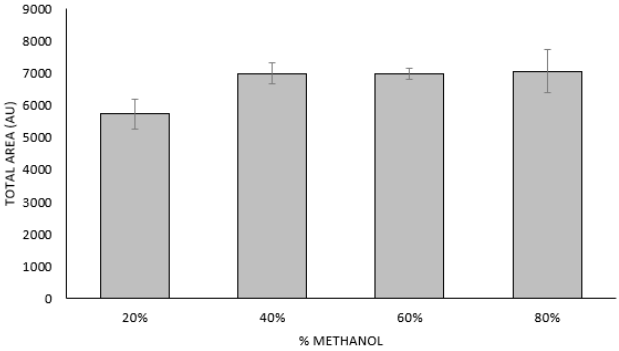


Figure 8. Total areas of the minoritarian compounds obtained for the different MeOH/H_2O solutions. Error bars indicate standard deviation of extraction replicates

Formic acid was added into the extraction solvent because it provides better stability to the recovered polyphenols in the extracts. Four 60:40 MeOH/H_2O solutions containing different amounts of formic acid (0%, 1%, 2% and 4%) were evaluated for the extraction. The total areas of the chromatograms obtained with the different solutions (figure 9) were analysed using ANOVA, and no statistically significant differences were found. However, the solution containing 1% formic acid gave the lowest standard deviation and the highest total area, so the solvent selected to extract the polyphenols from all the studied samples was decided to be a 60:40 MeOH/H_2O with 1% formic acid solution.

Once the composition of the extracting solution had been defined, the temperature and the time of sonication were evaluated. The extraction was carried out at room temperature and in a water bath at 50 ºC (15-minute sonication) to determine the effect of temperature; and sonicating the sample 10, 30, 60 and 120 minutes to determine the effect of time. The ANOVA results showed that neither the time nor the temperature cause statistically significant differences in the polyphenol extraction (figure 10). Since there were no significant differences for the temperature or time of extraction it was decided to keep working in the same conditions that we had been using so far, which were 15 minutes of sonication at room temperature.

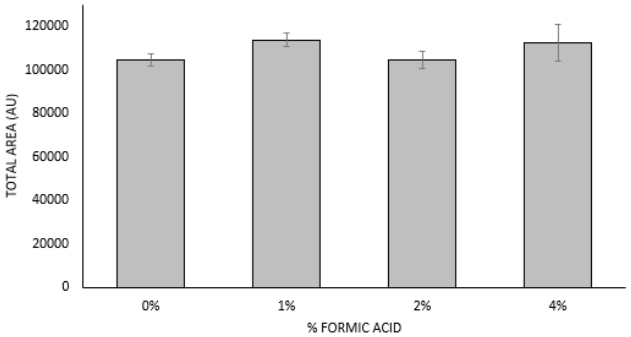


Figure 9. Total areas obtained for the solutions containing different proportions of formic acid. Error bars indicate standard deviation of extraction replicates.

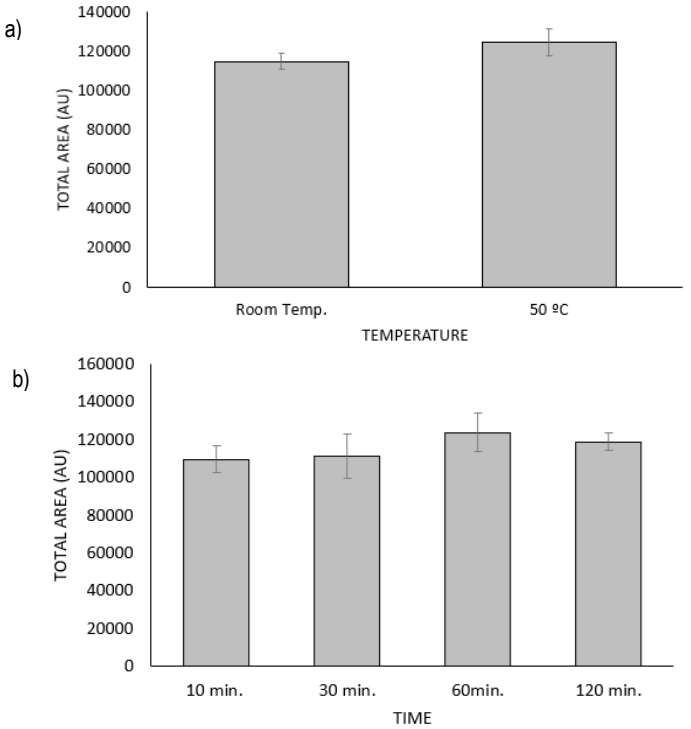


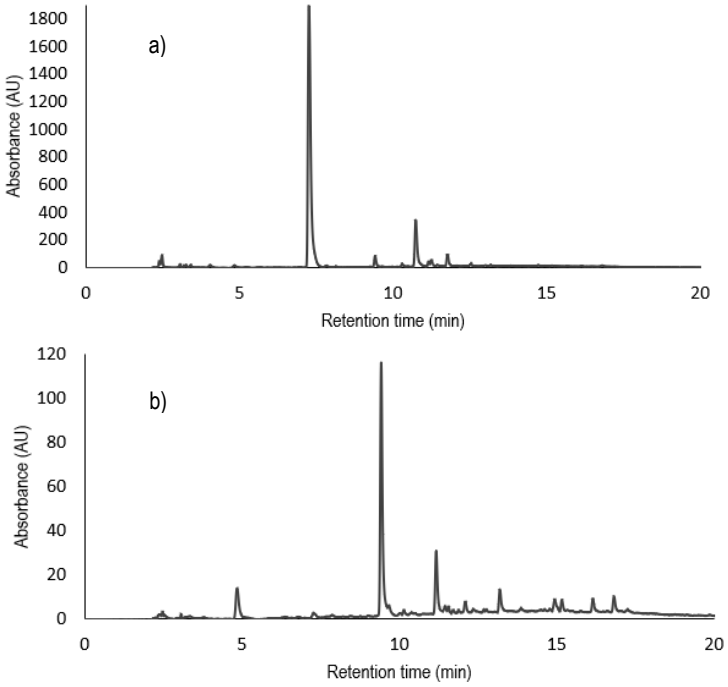
Figure 10. Total areas obtained for the extraction performed at different temperatures (a) and during different times (b). Error bars indicate standard deviation of extraction replicates.

The optimized extraction conditions consisted of a 15-minute sonication at room temperature using 4 mL 60:40 MeOH/H_2O 1% in formic acid as extractant for every 0.2 g of defatted chocolate.

6.3. HPLC-UV AND HPLC-FLD POLYPHENOL SEPARATION AND IDENTIFICATION

One of the main objectives of this project was to analyse the polyphenolic content and to identify the most common polyphenols in dark chocolate.

In order to separate the different polyphenols in chocolate a method based on HPLC-UV and HPLC-FLD was used. The selected wavelengths for the analysis were 280 nm, 325 nm and 370 nm for UV detection and 280 nm and 330 nm as exciting and emitting wavelengths respectively for the fluorescence detection. These wavelengths were chosen based on the absorbance spectra of the main classes of polyphenols found in cocoa which, as we discussed in section 3, are flavanols, flavonols, hydroxycinnamic acids and hydroxybenzoic acids. Flavanols have their maximum absorbance at 280 nm, hydroxycinnamic acids at 325 nm, hydroxybenzoic acids both at 280 and 325 nm and lastly, flavonols at 370 nm. For the fluorescence detection, the exciting wavelength chosen was 280 nm as it was the wavelength for which the number of obtained peaks and their intensities was the greatest. The obtained chromatographic profiles from a QC sample for both UV and fluorescence detection are shown in figure 11.



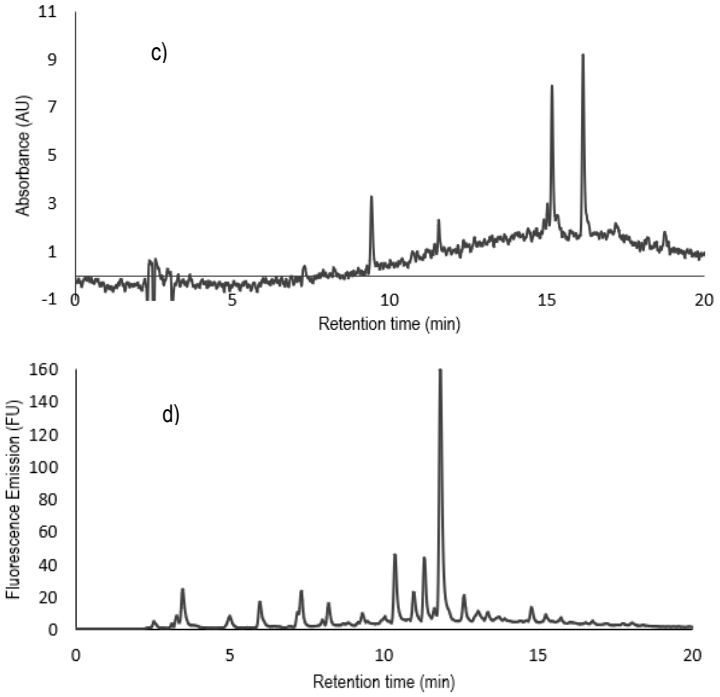


Figure 11. Acquired chromatograms by HPLC-UV at 280 nm (a), 325 nm (b), 370 nm (c) and by HPLC-FLD (d).

The main polyphenols present in the chocolate samples were preliminarily identified by LC-MS using full scan mode by the research group (data not shown). Since the chromatographic conditions for the method used in this project were identical to the ones used on the LC-MS analysis, the retention times of the identified compounds had to be similar in both cases. With that in my mind, and using standard solutions containing selected polyphenols, the identification of the main peaks that appeared on the chromatographs were performed. Moreover, the spectrum of a peak, was used as additional proof to confirm that a given peak was a certain compound. In figure 12 the identification of peaks from the HPLC-UV (280 nm) and HPLC-FLD chromatograms are shown. The correspondence between the assigned number to a peak and the compound is as follows: 1 theobromine, 2 caffeine, 3 3,4-dihydroxybenzoic acid, 4 (+)-catechin, 5 procyanidin B2, 6 (-)-epicatechin and 7 procyanidin C1.

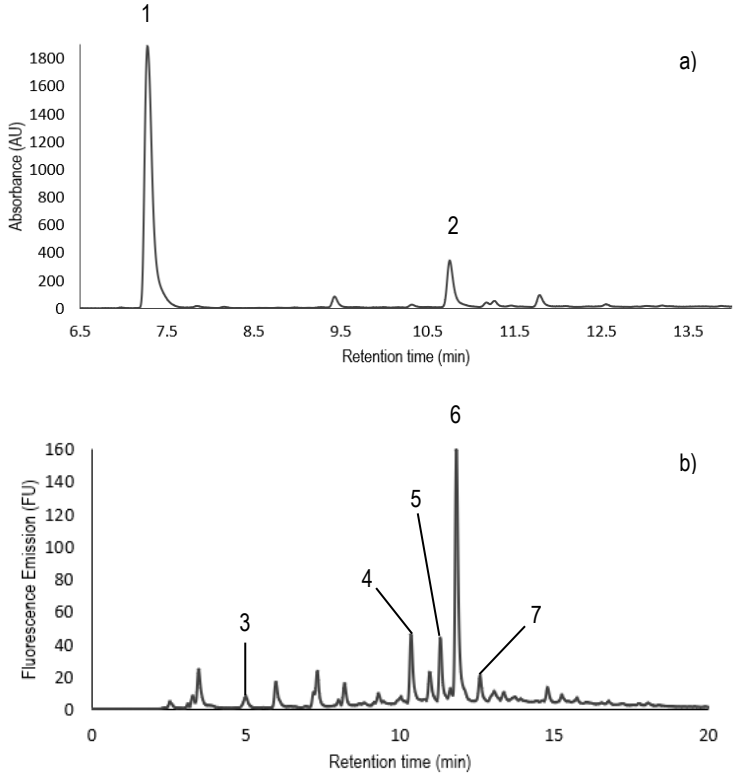


Figure 12. Chromatographic profiles of a QC sample. HPLC-UV at 280 nm (a) and HPLC-FLD (b).

6.4. PCA ANALYSIS

The X-matrix used to build the PCA models consisted of 141 samples (3 replicated extracts per sample and 15 QCs) and 16 variables obtained from the chromatographic profiles of the chocolates analysed by HPLC-UV and HPLC-FLD. In particular, the following features were taken as follows. The peak areas of all the identified compounds using UV-absorbance detection at 280 nm (Theobromine, caffeine, 3,4-dihydroxibenzoic acid, B2 and epicatechin) and fluorescence detection ((+)-catechin, (-)-epicatechin and C1) were included, as well as the areas of some unidentified compounds and the total areas of the HPLC-UV (280 nm) and HPLC-FLD chromatograms.

The built PCA model allows us to study the distribution of samples in the PC space, and patterns in sample distribution may be associated to certain characteristics of chocolates such as geographical origin, variety and cocoa content.

The PCA model showed that PC1 and PC2 accounted for 80% of the total variance of the samples; and as can be seen in figure 13, all the QC samples analysed form a compact group close to the center of the diagram which suggests that that the repeatability of the method is good and that no instrumental drift occurred throughout the analysis. Also, in figure 13 the samples are classified in groups based on their cocoa content, and a clear tendency is visible: samples with high cocoa content tend to appear on the right of the plot whereas samples with lower cocoa content appear on the opposite side. Thus, PC1 is responsible for the separation of chocolates by their cocoa content. Polyphenols present in chocolate come from cocoa, so it makes sense that the biggest differences in polyphenolic composition among the samples are caused by their cocoa content.

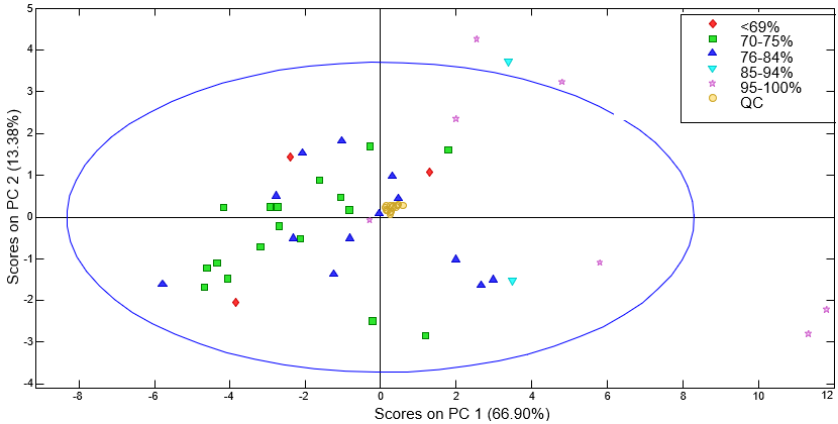


Figure 13. Scores plot on PC1 versus PC2. Samples classified by cocoa content.

When using PCA to classify the samples regarding their geographical origin (figure 14), some sample grouping is observed, with American samples in the center and African ones on the bottom-left area; however, there are some overlapping zones and many American chocolates are dispersed throughout the scores plot.

With cocoa varieties something similar occurs (figure 15), as chocolates of the same varieties are grouped together; especially Trinitario samples, which form a group close to the center of the scores plot; and Nacional samples, that can be found on the right side of the diagram separated

from all the other varieties. In this case though, overlapping between chocolates of different samples is greater, and it is difficult to really assign an area of the diagram to Criollo or Trinitario cocoas.

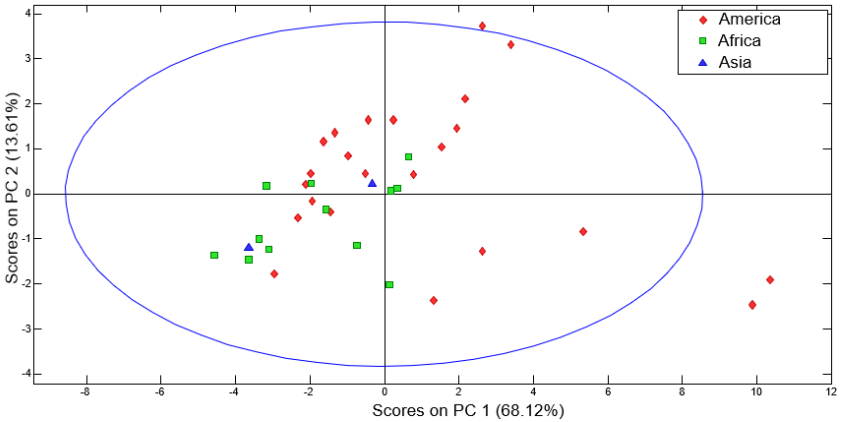


Figure 14. Scores plot on PC1 versus PC2. Samples classified by geographical origin.

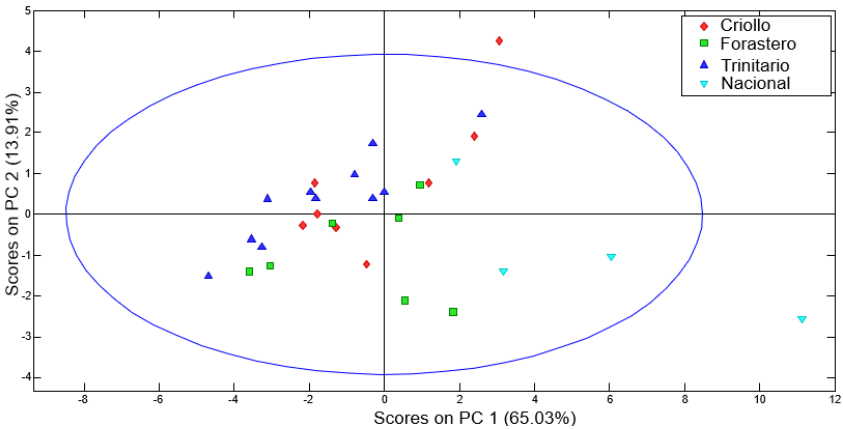


Figure 15. Scores plot on PC1 versus PC2. Samples classified by cocoa variety.

6.4.1. Sampaka chocolates

The polyphenolic composition of a chocolate is determined by many factors, which as we have already discussed include their geographical origin, cocoa content, manufacturing process and cocoa variety. The analysed sample pool contained chocolates that varied in all the factors mentioned above, and perhaps that multifactorial influence caused difficulties to discriminate

chocolates. 12 out of the 42 samples analysed belonged to a brand named Sampaka. These chocolates contained really similar cocoa content (72-84%) and since the manufacturer was the same, the processing of the cocoa beans to obtain chocolate was presumably the same.

So, a new PCA model was built using exclusively the data of these chocolates. PC1 was once again responsible for the separation based on the cocoa content; however, discrimination regarding geographical origin or variety was better achieved by representing the scores plot on PC2 and PC3. The fact that differences between varieties and origins are better explained by higher PCs (PC2 and PC3), which explain less variance than PC1, suggested that the overall polyphenolic content mainly influenced on PC1.

Figure 16 shows the different Sampaka samples classified by geographical origin, and one can easily see that two clear sample groups are formed. PC3 seems to be responsible for this separation, as American chocolates can be found on the upper side of the diagram whilst African samples are found on the bottom. The three blue triangles on the bottom of the plot, detached to the rest of American chocolates correspond to the triplicates of sample X25, which is the only American sample made from Forastero beans. By representing the loadings on PC2 and PC3 we can determine which polyphenols and xanthines are more abundant on the two groups of samples. It appears that African chocolates contain an overall higher polyphenolic content than American samples, especially regarding flavonoids ((-)-epicatechin and (+)-catechin) and the two unidentified flavonols that appear around 16 minutes in the 370 nm chromatograph (figure 11c). American chocolates though seem to contain a higher content of caffeine and theobromine.

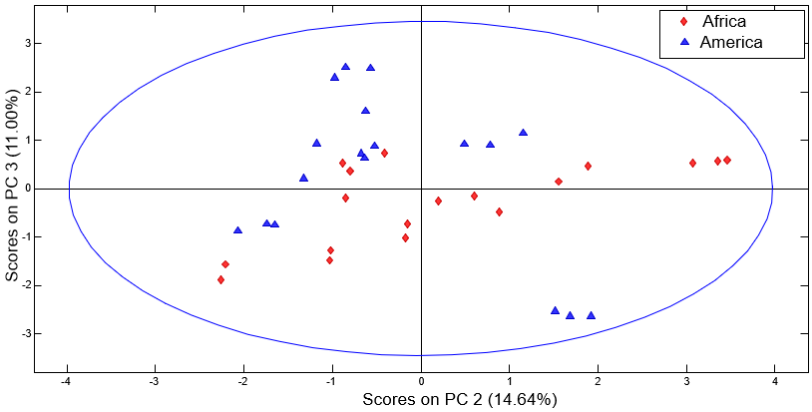


Figure 16. Scores plot on PC2 versus PC3. Samples classified by geographical origin.

Figure 17 on the other hand, shows the Sampaka samples classified by cocoa varieties. In this case the separation is also due to PC3, with Criollo and Forastero chocolates forming two separated groups. Trinitario beans are a hybrid of Criollo and Forastero beans, so they have a composition similar to both these varieties and perhaps that is the reason why samples from this variety appear scattered. In this case, the loading plot suggests that Criollo samples contain a higher content of caffeine and theobromine whereas Forastero samples contain more procyanidins C1 and B2, and flavanols. Trinitario samples appear too scattered to determine any distinct traits in their composition.

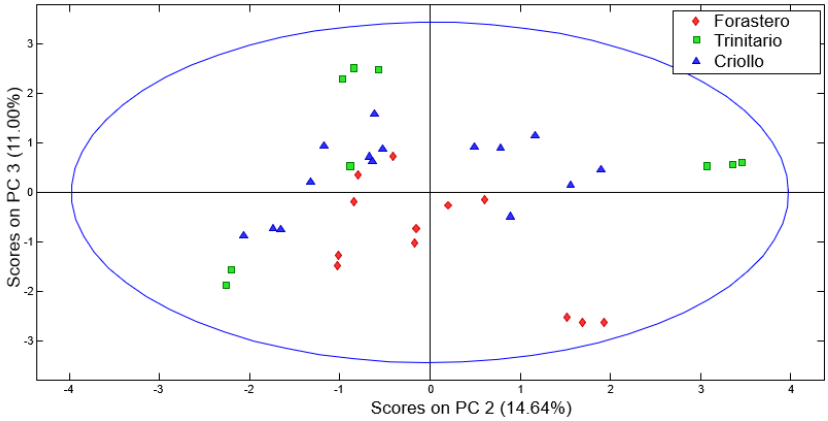


Figure 17. Scores plot on PC2 versus PC3. Samples classified by cocoa bean variety.

6.5. PLS-DA ANALYSIS

PLS-DA, unlike PCA, is a supervised classification method, which means that class belonging of a sample (Y-matrix) is considered in order to build the model. So, in general PLS-DA provides better separation of the samples based on a certain characteristic (e.g. geographical origin) than PCA does. The X-matrix used to build the PLS-DA model was the same used for the PCA analysis.

When using geographical origin belonging as the Y-matrix, the obtained PLS-DA model separates the samples much better than the PCA model as expected. In figure 18 we can see how LV1 is responsible for the separation of American (on the left) and African (on the right) samples. LV2 on the other hand is responsible for the separation of Asian samples from the rest

of the chocolates, as they appear on the bottom of the scores plot. Nevertheless, since there are only two samples from Asia, we cannot be sure that it is the real behavior these chocolates have. The information provided by the loading plot agrees to the conclusions drawn from the PCA analysis: African chocolates are richer in flavanols, flavonols, C1 and B2 than American ones. The data suggests that Asian samples have higher contents of caffeine and lesser flavonoids; but once again, not enough samples from this origin have been studied in order to consider this a general trend.

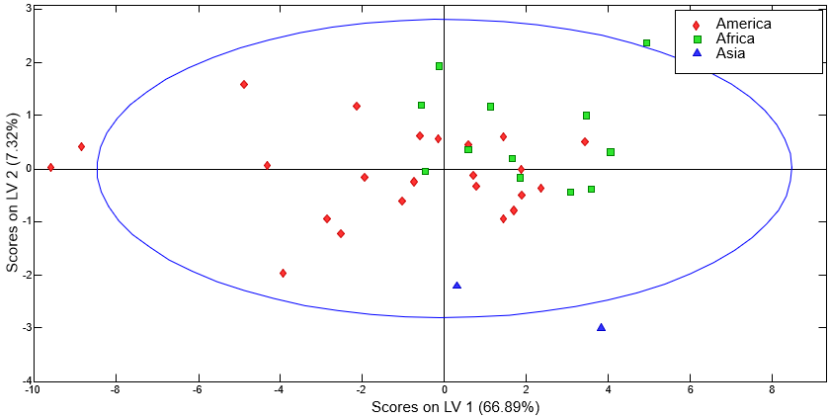


Figure 18. Scores plot on LV1 versus LV2. Samples classified by geographical origin.

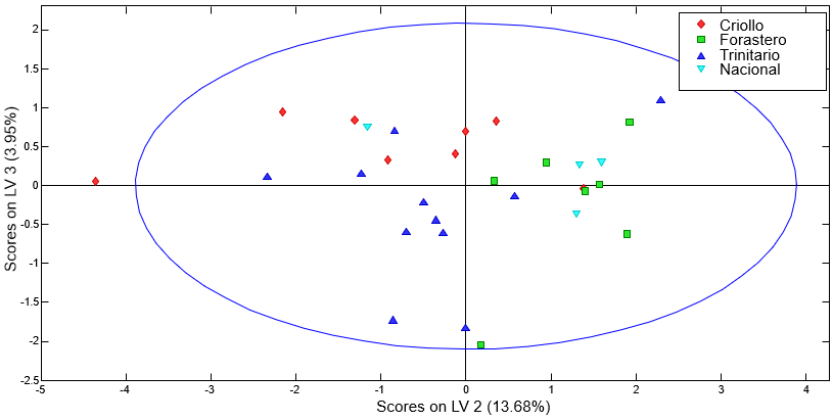


Figure 19. Scores plot on LV2 versus LV3. Samples classified by cocoa bean variety.

The obtained scores plot for a PLS-DA model built using cocoa varieties as classes (figure 19), also improves the separation of the samples when compared with PCA. In this case, LV2 and LV3 provide a better discrimination. As can be seen; Criollo chocolates appear on the upper-center part of the plot; Trinitario appears on the bottom, with some samples scattered; and Forastero samples appear on the right. Notice that chocolates made from Nacional beans, which are a subvariety of Forastero only found in Ecuador, are found within the area in which Forastero chocolates appear. In terms of their composition, the results drawn are in consonance with the PCA analysis. Forastero (Nacional included) samples are rich in flavanols, flavonols and procyanidins; in Criollo chocolates caffeine is abundant but not flavonols; and lastly, Trinitario is characterized for an abundance of B2 and small concentrations of (+)-catechin.

6.5.1. Sampaka chocolates

Much like we did in the PCA analysis, the Sampaka chocolates were analysed individually to determine if samples with similar cocoa contents and manufacturing processes would allow to discriminate better the chocolates by their geographical origin and variety.

Regarding geographical origin, using LV1 and LV2 the samples were completely separated with African samples situated on the bottom-left quarter of the diagram and American samples on top of the scores plot (figure 20). However, like with PCA, the three replicates of sample X25 are disintegrated from the rest of the American chocolates.

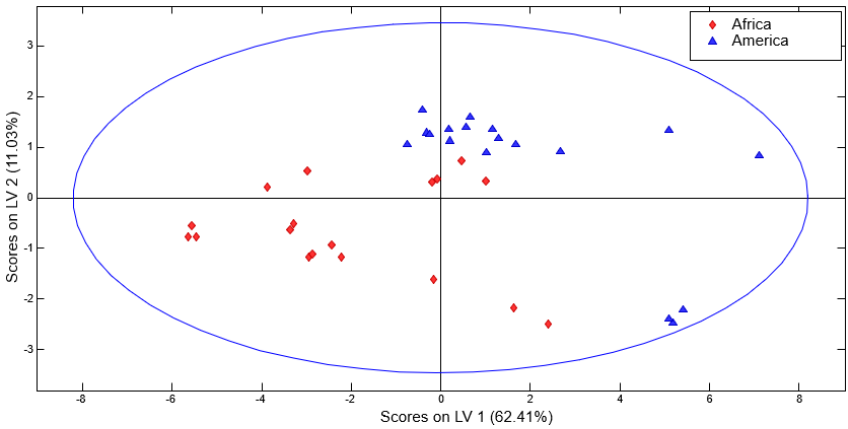


Figure 20. Scores plot on LV1 versus LV2. Samples classified by geographical origin.

LV2 and LV3 once again provided a better separation of the chocolates based on their variety. In figure 21 we can see how Criollo samples can be found on the top-right corner and Forastero ones on the left. In the area between Criollo and Forastero samples we can find the Trinitario chocolates. This is not unexpected since it as hybrid from Criollo and Forastero beans.

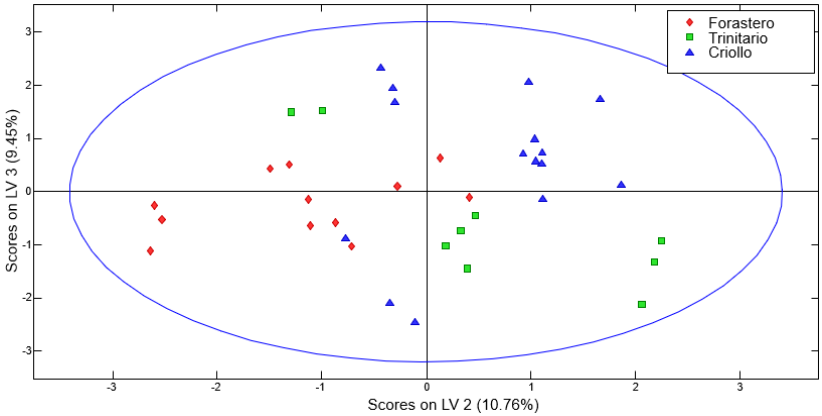


Figure 21. Scores plot on LV2 versus LV3. Samples classified by geographical origin.

The information provided by the loading plot remains consistent to the conclusions drawn from the PLS-DA analysis of the whole sample pool. That is, African chocolates generally have a higher polyphenolic content and lower xanthine abundance than American samples. Regarding cocoa varieties, Criollo is characterized by an overall lower content of polyphenols. Forastero is rich in (-)-epicatechin, (+)-catechin and procyanidins whereas Trinitario usually has a higher content of flavonols and a lower content of (+)-catechin than the other varieties.

7. CONCLUSIONS

In this project an HPLC-UV and HPLC-FLD method was developed in order to separate and identify the polyphenols from chocolate samples. The acquired chromatographic data was used to build PCA and PLS-DA models to classify the studied samples based on their geographical origin, bean variety and cocoa content.

The sample treatment applied consisted of defatting the chocolates and subsequently extracting the polyphenols from the defatted samples. The defatting process has proven to be effective as it almost completely eliminated the triglyceride content of the samples whilst extracting a negligible amount of polyphenols. The optimized polyphenol extraction method and gradient elution allowed us to obtain a good chromatographic separation which made possible the identification of the main polyphenols and alkaloids present in chocolates.

The PCA and PLS-DA analysis showed that the repeatability of the method is excellent since all the QC samples appeared cluttered on the center of the scores plot. Moreover, the chemometric analysis elucidates that polyphenols are chemical descriptors that can be used to classify and discriminate chocolates regarding some of their characteristics.

Scores plot on PC1 and PC2 (LV1 and LV2 for PLS-DA) separated the samples for their cocoa content and geographical origin; whereas scores plots on PC2 and PC3 (LV2 and LV3 for PLS-DA) separated the samples based on their bean variety. This suggests that the effect of cocoa content or origin on the polyphenolic profile of a chocolate is greater than the effect of cocoa variety.

The study of the loading plots has provided some insight on what polyphenols and alkaloids are most abundant on the different sample groups. The obtained results suggest that African samples are characterized by the abundance of flavanols, flavonols and procyanidins; American samples are rich in alkaloids but have lower polyphenolic concentrations. Regarding cocoa varieties, Forastero and Nacional often contain higher amounts of (-)-epicatechin, (+)-catechin and procyanidins B2 and C1. Criollo samples are not as rich in polyphenols, but caffeine levels are high. Lastly, Trinitario is characterized by the abundance of flavonols and a poor (+)-catechin content.

8. REFERENCES AND NOTES

1. Oracz J, Zyzelewicz D, Nebesny E. The Content of Polyphenolic Compounds in Cocoa Beans (Theobroma cacao L.), Depending on Variety, Growing Region, and Processing Operations: A Review. Crit Rev Food Sci Nutr. 2015;55(9):1176–92.
2. Ro S. Review Chocolate Divine food, fattening junk or nutritious supplementation? 1997;341–5.
3. Borchers AT, Keen CL, Hannum SM, Gershwin ME. Cocoa and chocolate: Composition, bioavailability, and health implications. J Med Food. 2000;3(2):77–105.
4. Caligiani A, Acquotti D, Cirlini M, Palla G. 1H NMR study of fermented cocoa (Theobroma cacao L.) beans. J Agric Food Chem. 2010;58(23):12105–11.
5. Rusconi M, Conti A. Theobroma cacao L., the Food of the Gods: A scientific approach beyond myths and claims. Pharmacol Res. 2010;61(1):5–13.
6. Elwers S, Zambrano A, Rohsius C, Lieberei R. Differences between the content of phenolic compounds in Criollo, Forastero and Trinitario cocoa seed (Theobroma cacao L.). Eur. Food. Res. Technol. 2009;229(6):937–48.
7. Perez M, Lopez-Yerena A, Vallverdú-Queralt A. Traceability, authenticity and sustainability of cocoa and chocolate products: a challenge for the chocolate industry. Crit. Rev. Food. Sci. Nutr. 2020;0(0):1–15.
8. Roelofsen PA. Fermentation, Drying, and Storage of Cacao Beans. Adv. Food. Res. 1958 Jan 1;8(C):225–96.
9. Cienfuegos-Jovellanos E, Del Mar Quiñones M, Muguerza B, Moulay L, Miguel M, Aleixandre A. Antihypertensive effect of a polyphenol-rich cocoa powder industrially processed to preserve the original flavonoids of the cocoa beans. J. Agric. Food. Chem. 2009;57(14):6156–62.
10. Tomas-Barberán FA, Cienfuegos-Jovellanos E, Marín A, Muguerza B, Gil-Izquierdo A, Cerdá B, et al. A new process to develop a cocoa powder with higher flavonoid monomer content and enhanced bioavailability in healthy humans. J. Agric. Food. Chem. 2007;55(10):3926–35.
11. Barišić V, Kopjar M, Jozinović A, Flanjak I, Ačkar D, Miličević B, et al. The chemistry behind chocolate production. Molecules. 2019;24(17).
12. Miller KB, Stuart DA, Smith NL, Lee CY, Mchale NL, Flanagan JA, et al. Antioxidant activity and polyphenol and procyanidin contents of selected commercially available cocoa-containing and chocolate products in the United States. J. Agric. Food. Chem. 2006;54(11):4062–8.
13. Serra Bonvehí J, Ventura Coll F. Evaluation of purine alkaloids and diketopiperazines contents in processed cocoa powder. Eur. Food. Res. Technol. 2000;210(3):189–95.
14. Meyer A, Ngiruwonsanga T, Henze G. Determination of adenine, caffeine, theophylline and theobromine by HPLC with amperometric detection. Fresenius J. Anal. Chem.

1996;356(3–4):284–7.

15. Niemenak N, Rohsius C, Elwers S, Omokolo Ndoumou D, Lieberei R. Comparative study of different cocoa (Theobroma cacao L.) clones in terms of their phenolics and anthocyanins contents. J. Food. Compos. Anal. 2006;19(6–7):612–9.
16. Gu L, House SE, Wu X, Ou B, Prior RL. Procyanidin and catechin contents and antioxidant capacity of cocoa and chocolate products. J. Agric. Food. Chem. 2006;54(11):4057–61.
17. Andres-Lacueva C, Monagas M, Khan N, Izquierdo-Pulido M, Urpi-Sarda M, Permanyer J, et al. Flavanol and flavonol contents of cocoa powder products: Influence of the manufacturing process. J. Agric. Food. Chem. 2008;56(9):3111–7.
18. Jalil AMM, Ismail A. Polyphenols in cocoa and cocoa products: Is there a link between antioxidant properties and health? Molecules. 2008;13(9):2190–219.
19. Jonfia-Essien WA, West G, Alderson PG, Tucker G. Phenolic content and antioxidant capacity of hybrid variety cocoa beans. Food Chem. 2008;108(3):1155–9.
20. Poore J, Nemecek T. Reducing food's environmental impacts through producers and consumers. Science (80-). 2018;360(6392):987–92.
21. Wessel M, Quist-Wessel PMF. Cocoa production in West Africa, a review and analysis of recent developments. NJAS - Wageningen J. Life. Sci. 2015;74–75:1–7.
22. Bertoldi D, Barbero A, Camin F, Caligiani A, Larcher R. Multielemental fingerprinting and geographic traceability of Theobroma cacao beans and cocoa products. Food Control. 2016;65:46–53.
23. Torres-Moreno M, Torrescasana E, Salas-Salvadó J, Blanch C. Nutritional composition and fatty acids profile in cocoa beans and chocolates with different geographical origin and processing conditions. Food Chem. 2015;166:125–32.
24. Cambrai A, Marchioni E, Julien-David D, Marcic C. Discrimination of Cocoa Bean Origin by Chocolate Polyphenol Chromatographic Analysis and Chemometrics. Food Anal. Methods. 2017;10(6):1991–2000.
25. D'Souza RN, Grimbs S, Behrends B, Bernaert H, Ullrich MS, Kuhnert N. Origin-based polyphenolic fingerprinting of Theobroma cacao in unfermented and fermented beans. Food Res. Int. 2017;99(March):550–9.
26. Peres-Neto PR, Jackson DA, Somers KM. How many principal components? stopping rules for determining the number of non-trivial axes revisited. Comput. Stat. Data Anal. 2005;49(4):974–97.
27. Trygg J, Holmes E, Lundstedt T. Chemometrics in metabonomics. J. Proteome Res. 2007;6(2):469–79.
28. Granato D, Putnik P, Kovačević DB, Santos JS, Calado V, Rocha RS, et al. Trends in Chemometrics: Food Authentication, Microbiology, and Effects of Processing. Compr. Rev. Food. Sci. Food Saf. 2018;17(3):663–77.

29. Wold S, Sjöström M, Eriksson L. PLS-regression: A basic tool of chemometrics. Chemom. Intell. Lab. Syst. 2001;58(2):109–30.

30. Fayeulle N, Meudec E, Boulet JC, Vallverdu-Queralt A, Hue C, Boulanger R, et al. Fast Discrimination of Chocolate Quality Based on Average-Mass-Spectra Fingerprints of Cocoa Polyphenols. J. Agric. Food Chem. 2019;67(9):2723–31.

31. Carmo Brito B de N, Campos Chisté R, da Silva Pena R, Abreu Gloria MB, Santos Lopes A. Bioactive amines and phenolic compounds in cocoa beans are affected by fermentation. Food Chem. 2017;228:484–90.

9. ACRONYMS

ACN: acetonitrile

ANOVA: analysis of variance

DAD: diode array detector

FLD: fluorescence detection

GC: gas chromatography

HPLC: high performance liquid chromatography

ICP: inductively coupled plasma

LV: latent variable

MeOH: methanol

MS: mass spectrometry

NMR: nuclear magnetic resonance

PAC: proanthocyanidin

PCA: principal component analysis

PLS-DA: partial least squares – discriminant analysis

QC: quality control

UV: ultraviolet

APPENDICES

APPENDIX 1: SAMPLE POOL

CODE	CHOCOLATE	ORIGIN	VARIETY	% COCOA
X1	Suchard BIO	America	Trinitario	70
X2	Lindt	Mixture	Mixture	90
X3	Torras	America	Mixture	100
X4	Simon Coll	Africa	Forastero	99
X5	Alternativa 3	America	Criollo	75
X6	Alternativa 3	America	Nacional	65
X7	Pacari	America	Nacional	100
X8	Ethiquetable	Mixture	-	98
X9	Ethiquetable	Africa	-	85
X10	Ethiquetable	America	-	72
X11	Solé	America	-	86
X12	Nestlé	**-**	-	85
X13	Lacasa	**-**	-	92
X14	Pancracio	**-**	-	80
X15	J.D.Gross	America	-	60
X16	J.D.Gross	America	Nacional	95
X17	J.D.Gross	America	Nacional	85
X18	MX BCN 01	Africa	Forastero	70
X19	MX BCN 02	Africa	Trinitario	70
X20	MX BCN 03	America	Mixture	68
X21	MX BCN 04	Africa	Trinitario	70
X22	MX BCN 05	America	Trinitario	70
X23	MX BCN 06	Asia	Trinitario	70
X24	Sampaka America 1	America	Criollo	73
X25	Sampaka America 2	America	Forastero	72
X26	Sampaka America 3	America	Criollo	73
X27	Sampaka America 4	America	Criollo	75
X28	Sampaka America 5	America	Trinitario	74
X29	Sampaka America 6	America	Criollo	84
X30	Sampaka Africa 1	Africa	Forastero	72
X31	Sampaka Africa 2	Africa	Forastero	79

X32	Sampaka Africa 3	Africa	Trinitario	79
X33	Sampaka Africa 4	Africa	Forastero	72
X34	Sampaka Africa 5	Africa	Trinitario	73
X35	Sampaka Africa 6	Africa	Criollo	79
X36	Blanxart	America	-	82
X37	Blanxart	America	Criollo	100
X38	Blanxart	Asia	Trinitario	71
X39	Blanxart	America	Criollo	77
X40	Blanxart	America	Trinitario	95
X41	Blanxart	Africa	Forastero	82
X42	Blanxart	America	Trinitario	85

APPENDIX 2: FAT CONTENT OF THE STUDIED SAMPLES

Sample	Obtained Fat Content (%)	Declared Fat Content (%)	Cocoa content (%)	Recovery
X1	39.7	42	70	94.5%
X2	48.5	55	90	88.2%
X3	46.1	56	100	82.3%
X4	49.7	56	99	88.8%
X5	39.8	44	75	90.5%
X6	37.1	39	65	95.1%
X7	49	59	100	83.1%
X8	48.8	53.7	98	90.9%
X9	46.6	51	85	91.4%
X10	40.2	42.1	72	95.5%
X11	41.9	42.1	86	99.5%
X12	42.5	46.6	85	91.2%
X13	47.6	50	92	95.2%
X14	49.7	48	80	103.5%
X15	40.1	41	60	97.8%
X16	50.9	51.1	95	99.6%
X17	45.6	46.8	85	97.4%
X18	37.9	37.1	70	102.2%
X19	38	37.1	70	102.4%
X20	38.6	37.1	68	104.0%
X21	40.1	37.1	70	108.1%
X22	42.4	37.1	70	114.3%
X23	40.4	37.1	70	108.9%
X24	38.4	46.4	73	82.8%
X25	39.7	46.4	72	85.6%
X26	42.5	46.4	73	91.6%
X27	45.4	46.4	75	97.8%
X28	45.6	46.4	74	98.3%

X29	50.6	46.4	84	109.1%
X30	40.8	45.1	72	90.5%
X31	45.9	45.1	79	101.8%
X32	45.6	45.1	79	101.1%
X33	40	45.1	72	88.7%
X34	38	45.1	73	84.3%
X35	46	45.1	79	102.0%
X36	48.9	50.3	82	97.2%
X37	52.6	53	100	99.2%
X38	41.3	46.9	71	88.1%
X39	44.2	49.4	77	89.5%
X40	51.9	59.9	95	86.6%
X41	48.1	51.8	82	92.9%
X42	48.4	47.2	85	102.5%