FOLLOWING COCOA BEANS TO CHOCOLATE:

THE SEARCH FOR INTRINSIC CHARACTERISTICS

VALENTINA ACIERNO

Propositions

1. Brands often use the cocoa geographical origin in their marketing but it is mainly the cocoa botanical origin delineating the final product characteristics.
   (this thesis)

2. Cocoa's volatile profile has the potential to trace raw material origin and track it back in the finished product.
   (this thesis)

3. PTR-MS could trace our food from the fridge all the way back to the prehistoric times highlighting the large applicability and the multi-disciplinarily features of this technique.

4. Next to never underestimating the importance of the assumptions when choosing a statistical model (Vardeman & Morris, 2003), never underestimate the ethics behind the statement "all the assumptions were met".
   (Vardeman, S. B., & Morris, M. D. (2003). *The American statistician*, 57, 21-26).

5. The Elephant's journey from Portugal to Austria (Saramago, 2008) teaches us that when crossing borders and living abroad, one overcomes challenges with a vitality that one probably did not experience in his/her own country.
   (Saramago, J., (2008). *The elephant's journey*, Caminho (Ed.)).

6. Most of the time we tend to forget about the person behind the scientist.

Propositions belonging to the thesis entitled
"Following cocoa beans to chocolate: The search for intrinsic characteristics"

Valentina Acierno
Wageningen, 15 May 2020

Following cocoa beans to chocolate: The search for intrinsic characteristics

Valentina Acierno

Thesis committee

Promotor
Prof. Dr Saskia M. van Ruth
Professor of Food Authenticity and Integrity
Wageningen University & Research

Co-promotor
Dr Martin Alewijn
Wageningen Food Safety Research
Wageningen University & Research

Other members
Prof. Dr J.G.M. (Hans-Gerd) Janssen, Wageningen University & Research
Dr F.J.M. Harren, Radboud University, Nijmegen
Dr I.S. Sousa Gomes Mafra, Universidade do Porto, Portugal
Dr J.-L. Le Quéré, Centre des Sciences du Goût et de l'Alimentation, Dijon, France

Following cocoa beans to chocolate: The search for intrinsic characteristics

Valentina Acierno

Table of contents

Chapter 1
Introduction

1. Introduction

1.1 The historical cocoa-chocolate value chain

Cocoa was an important commodity for the ancient Olmec, Mayan, Toltec and Aztec civilizations. It was used as ingredient for a drink and as offering to the gods during religious rituals (Poelmans & Swinnen, 2016). The cocoa drink was called "chocol'ha" and not everyone was allowed to consume it (Tonantzin Chocolate, 2015). Maya considered the cocoa tree as one of the four cosmic trees located in the different directions of the universe. The painting in Figure 1.1 underlines the importance of cacao for this civilization. The painting is part of the universe representation as they imagined it. The cocoa tree is located in the southern part of the universe and had a crucial association with the sacred plant of Mesoamerica, the corn in the painting represented by the god of corn positioned at the right-hand side of the tree (Montserrat Camacho Ángeles, 2012). The other god represented is the god of death. In fact, cocoa was also associated with the underworld, the ancestors and their rebirth. Furthermore, cocoa was a commodity for exchange. The trading value given to cocoa beans by these civilizations is emphasised by archaeological findings of fake cocoa beans made with clay (Poelmans & Swinnen, 2016).

Figure 1.1. Maya cocoa tree representation (Autor unknown. The Codex Fejervary-Mayer, 15th century. From Wikimedia Commons, n.d.).

It was only when the Spanish arrived in those regions in the fifteenth and sixteenth centuries that the rest of the world learned about cocoa (Poelmans & Swinnen, 2016). Since then, cocoa lost its spiritual value but not its importance as a commodity. Nowadays, cocoa is a cash crop for developing countries and one of the main import commodities for processing and consuming countries. It is quoted on the exchange and it is produced in three different continents. In contrast to the ancient civilisations, cocoa now has a global supply chain and is completely transformed to produce a mouthful chocolate bar before reaching the consumers. Chocolate is no longer consumed by a selected group of people but it is one of the most popular foods all over the world. This made chocolate manufacturing a serious business.

1.2 Cocoa production, supply chain and concerns

Cocoa is primarily grown in West Africa, Central and South America and Southeast Asia. Africa is the largest producer covering ~75% of the production. Ivory Coast is the largest producing country by volume, covering ~43 % of the global supply chain followed by Ghana (Fountain & Huetz-Adams, 2018). Figure 1.2 summarises the global cocoa production and consumption during the year 2017/2018. Most of the cocoa beans are produced by small-sized farmers. According to the World Cocoa Foundation (WCF) (World Cocoa Foundation Report, 2014), 5 to 6 million cocoa farmers are contributing to the annual cocoa production. The worldwide production is about 4.2 million tons with a value of $11.8 billion and a growing rate of 3% per year from the past decade.

The chocolate market is estimated to grow at a compound annual growth rate (CAGR) of 2.3% from 2014 to 2019 and the demand for cocoa is predicted to have increased by 30% in 2020 compared to the last years (World Cocoa Foundation Report, 2014).

Production / Consumption

Cocoa Production in 1,000 tonnes 2017/18
Source: ICCO 2018, Table 2, 40
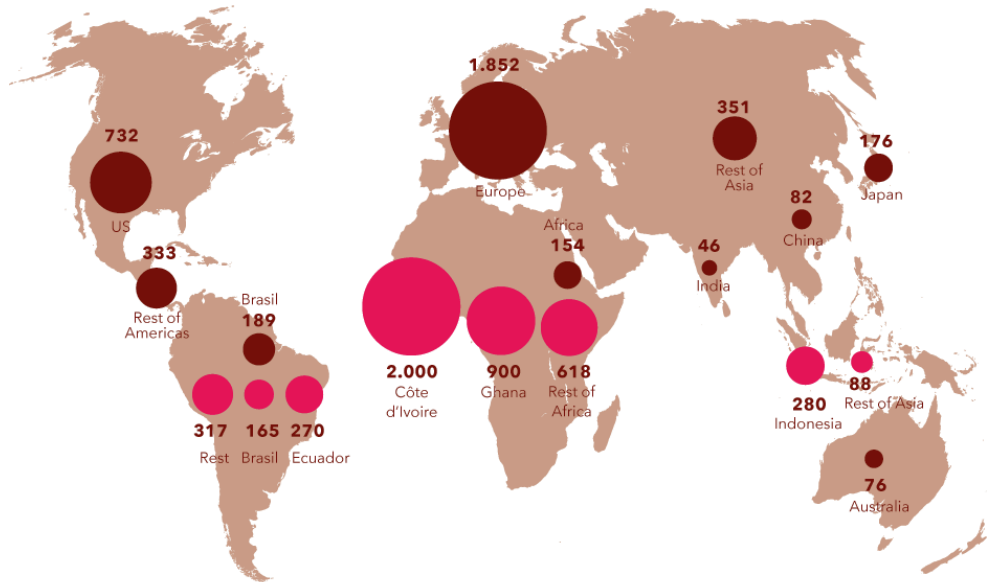
Figure 1.2. Cocoa production (pink circles) and consumption (brown circles) (Reproduced with permission: Fountain & Huetz-Adams, 2018).

One of the main concerns in cocoa production is the volatility in cocoa bean prices. This is due to the unstable cocoa production, vulnerability to weather conditions, pests and diseases, and to the speculation and political instability in the cocoa producing countries (Shavez, Ahmad, Jan, & Bashir, 2017). These issues have an effect not only on the quality and price of the cocoa beans but also on the income of the farmers. Cocoa is usually their main source of income and it contributes to 70–100% of annual household incomes (Kongor et al., 2016). Nevertheless, the families struggle to make a living from it because of large price asymmetries in the chain. For instance, farmers in West Africa are receiving only 3.5-6.4% of the final price of a chocolate bar (Shavez et al., 2017).

Moreover, child labour is not uncommon. For these reasons, initiatives such as sustainability projects and fair-trade chocolate production were set up. This all in view of corporate social responsibility in order to redress the cocoa production taking into account social, economic and environmental issues (Afoakwa, 2010a).

As mentioned before, the cocoa production is mainly localized in developing countries. On the other hand, as can be seen in Figure 1.2, the consumption mostly takes place in the developed countries. The consuming countries import the cocoa to grind it and transform it into intermediate and finished products. Europe is the largest cocoa processor, followed by the USA (Figure 1.2). A growing demand is also coming from emerging markets, such as Russia, China, India, and Africa (Poelmans & Swinnen, 2016). The global journey of cocoa includes numerous parties starting with the farmers followed by buyers, shipping organisations, processors, chocolatiers and distributors. After the farmers grow, harvest, ferment, dry and pack the cocoa beans, the batches from several growers are collected and mixed by local buyers or stations, traders and exporters until they reach the chocolate manufacturer. This global cocoa–chocolate chain going from 'South-to-North' impacts the management of the trade. Traceability is one of the main issues mainly because legislation applied in the EU countries are not applied in the producing countries (Saltini, Akkerman, & Frosch, 2013). This is the case for the European law regarding traceability (European Commission, 2002). While the actors located in Europe must be able to trace back a particular product up to their direct suppliers and customers, there is no legislation for the non-European actors involved to have traceability systems. This is illustrated by the study of Saltini and Akkerman (2012), in which the researchers were able to trace the cocoa up to the local exporter, but it was not possible to collect any further information, indicating a poor data exchange between farmers and chocolate manufacturers. Combining this lack of data exchange with the complexity of the supply chain makes the cocoa-chocolate chain more sensitive to opportunist behaviour of particular actors for individual economic benefit.

1.3 Chocolate manufacture: from cocoa beans to chocolate

The conversion from cocoa beans to chocolate involves technological changes that further intricate the cocoa-chocolate supply chain. After harvesting, the cocoa beans go through chemical and physical changes needed to enhance the seeds' palatability and chocolate flavour (Aculey et al., 2010; McShea et al., 2008).

The processing of the cocoa beans can be sub-divided in two phases: primary and secondary processing.

The **primary processing** phase includes the post-harvest procedures starting from the collection of the pods till the storage of the dried cocoa beans. The main steps of this phase are: *gathering of the pods, pod storage for pulp pre-conditioning, opening or breaking the pod, fermentation of the beans, drying, sorting, packaging and storage* (Afoakwa, 2014b).

The pulp pre-conditioning decreases the acidity of the dried fermented cocoa beans and can be conducted before fermentation. There are three pre-conditioning methods: pod storage, depulping (mechanical or enzymatic depulping) and spreading of the beans (Afoakwa, 2014c). With pod storage, the pulp is pre-conditioning inside the pods. On the contrary, with depulping and bean spreading the pre-conditioning of the pulp takes place outside of the pods.

Once the beans have been removed from the pods, the beans and adhering pulp are placed in heaps, boxes, or baskets and fermented for 3–7 days (Aprotosoaie, Luca, & Miron, 2016). Fermentation favours the removal of the residual pulp and the subsequent drying. Before breaking or opening of the pods, the beans are microbiologically sterile. Once the pods are opened, the pulp and the beans are exposed to air and therefore naturally occurring microorganisms start to degrade the pulp which is rich in sugars. The fermentation starts with an anaerobic phase called the hydrolytic phase, in which the anaerobic yeasts flourish initiating an alcoholic fermentation. The yeasts release enzymes that degrade the pectin constituents of the pulp cells causing the release of the cellular fluid. The breakdown of the pulp cells facilitates the entrance of air through the pulp mass, starting the second phase of fermentation. The second phase, known as the oxidative condensation phase, takes place under aerobic conditions and is characterised by the increase of the lactic acid bacteria activity and the inhibition of the yeast activity. Moreover, during this phase the microbial activity produces ethanol, acetic and lactic acids. The increased pulp aeration and temperature (45 °C), due to the microbial activity, favour the growth of acetic acid bacteria responsible for the oxidation of ethanol to acetic acid. The reduced activity of lactic acid bacteria and increased activity of acetic acid bacteria cause the so-called death of the beans that leads to the easier breakdown of cellular components. During fermentation the formation of flavour precursors, the reduction in bitterness and astringency, and the development of colour takes place (Afoakwa, Paterson, Fowler, & Ryan, 2008). Free amino acids, peptides and reducing sugars are the flavour precursors developed during fermentation. Reducing sugars are generated from the enzymatic degradation of sucrose and peptides and free amino acids are produced from the enzymatic degradation of cocoa proteins (Frauendorfer & Schieberle, 2008).

Cocoa bean fermentation is still a spontaneous process and is influenced by many factors, which in turn impact cocoa quality. Factors influencing fermentation include the method, duration and speed of fermentation, pod storage, cocoa bean variety and local process conditions (Afoakwa, 2012; Afoakwa, Quao, Takrama, & Budu, 2013).

After fermentation, the beans are dried. During the drying of the beans, the moisture content is reduced from about 60% to between 6 and 8% to prevent mould growth. This phase has an important role in reducing bitterness, astringency and acidity (Afoakwa et al., 2008).

Chemical changes occurred during fermentation also continue during drying with the development of a characteristic brown colour and Amadori compounds, the first intermediates of Maillard reaction. Specifically, they are the first intermediates of the reaction between free amino acids and glucose, and during subsequent roasting they will decompose into numerous volatile components (Ziegleder, 2009). Nevertheless, some of the chocolate aroma compounds such as aldehydes, alcohols, acids and acetates are already developed during fermentation and drying (Ziegleder, 2009). Usually, cocoa beans are dried under the sun on a concrete floor or on a raised platform. However, when the weather is not favourable, artificial drying is employed. For instance, in regions such as Brazil, Cameroon, Costa Rica, Malaysia and Panama the harvesting occurs in a period of rainfall so artificial drying is necessary (Afoakwa, 2014a).

After the first steps of production, the fermented and dried cocoa beans are packed into bags and sold to intermediaries and/or delivered to the grinder's plant. The processing continues with the **secondary processing** phase during which the fermented and dried cocoa beans are transformed into the finished product. This phase includes: *winnowing, roasting, alkalization, grinding, conching, tempering, moulding and packing.*

Before roasting the beans are cleaned and winnowing is performed to separate the cocoa nibs or cotyledons from their shells (Aprotosoaie et al., 2016); this procedure can take place either before or after roasting. Roasting and conching play a significant role, specifically for flavour development. During roasting, the flavour precursors (free amino acids, peptides and reducing sugars) produced during fermentation and drying go through the Maillard reaction and Strecker degradation to produce the typical chocolate flavour compounds such as pyrazines, alcohols, esters, aldehydes, ketones, furans, thiazoles, pyrones, acids, imines, amines, oxazoles, pyrroles and ethers (Kongor et al., 2016).

In the first step of Maillard reaction, the bond between the free amino groups of amino acids with the carbonyl groups of glucose leads to the formation of Schiff bases. The Schiff bases (glucosyl amines and fructosyl amines) tautomerizate to 1,2 enaminols to be further rearranged to Amadori compounds (1-amino-1-deoxy-2- ketoses) (Aprotosoaie et al., 2016). At this point the pH condition is important as it influences the formation of the next intermediates. In acidic conditions, the amino compounds formed are decomposed to 3-deoxyhexuloses. The latter loses water to form hydroxymethylfurfurals and other furfural products. In basic or neutral pH, the intermediates formed are 2,3-enediol and dehydroreductone which are responsible of the formation of maltol, isomaltol and α-dicarbonyl compounds (Afoakwa et al., 2008). α-dicarbonyl compounds can undergo further steps of dehydration, fragmentation and transamination, and form smaller aldehydes and ketones fundamental in cocoa flavour development. Another important path is the Strecker degradation that results in the formation of amino acid specific aldehydes, pyrazines and other heterocyclic compounds (Afoakwa et al., 2008). Complex additional reactions involve a second amino acid in a Strecker degradation and generate pyrroles and pyridines (Aprotosoaie et al., 2016). There are two ways of roasting: whole bean and nib roasting. The optimal roasting conditions depends on the raw material, the variety of cocoa and/or the final product desired flavour. An optional step is the alkalization. Nibs are alkalized usually with potassium or sodium carbonate to enhance flavour and colour (Afoakwa, 2014a). After roasting, the nibs are ground to produce cocoa liquor or cocoa mass. The temperature and degree of grinding depends on the type of nib and the final product. In certain cases, roasting can be performed after cocoa liquor production. When this occurs, a thermal pre-treatment is applied, then the shells are removed and the nibs are transformed into cocoa liquor before roasting (Afoakwa et al., 2008; Aprotosoaie et al., 2016).

Grinding has the purpose to make the cocoa particles small enough and to extract as much fat as possible from the cotyledon cells. Reducing the cocoa beans to a fine powder involves three types of mills: roll mills that break cocoa beans into pieces or nibs, pin (impact) mills that transform the nibs into cocoa liquor and ball mills that reduce the particle size of the liquor leaving it soft and silky. At this point, the cocoa liquor can be mixed with the other ingredients to produce chocolate or go through a pressing process that will divide the liquor into cocoa butter (fat extract) and cocoa cakes or cocoa solids (non-fat cocoa material). The cocoa cake or cocoa solid can be sold as such or pulverized into cocoa powder. On the other hand, the cocoa butter can be mixed with cocoa liquor, sugar and milk in case of milk chocolate production. This mixture, finely ground to a particle size smaller than 30 μm, is then placed into conches to smooth it and to guarantee that all the solid particles are coated with fat. During conching, the final mixture is stirred at a temperature of 30°C for several hours. In this phase, the moisture content and the concentration of volatile organic acids generated during fermentation (*i.e.* acetic acid, propionic acid, isobutyric acid) are reduced, therefore the final flavour is formed (Afoakwa, 2014a; Aprotosoaie et al., 2016). The final steps of chocolate production involve tempering. Cocoa butter is known to show a polymorphic transformation of fat crystals by crystallizing in six different forms (Afoakwa, 2014d). Tempering ensures that cocoa butter crystallizes into the stable form to guarantee good heat stability, good snap, glossy surface, and smooth and fast melting (Windhab, 2009).

There are two ways to finalise chocolate production. The first is to form the final, desired shape using a mould into which tempered chocolate is poured. The second way is to use a pre-formed centre over which chocolate is poured to produce a coating; this method is called enrobing. In both cases, subsequent cooling is required to prepare the chocolate for packaging (Gray, 2009).

1.4 Cocoa products: culinary aspects and trends

Cocoa is mainly used to produce chocolate confections, chocolate-coated products (biscuits, ice creams), or food products containing cocoa powder, such as beverages, cakes and snacks (Afoakwa, 2014a). Even though the chocolate sector faced economic difficulties in the past years, chocolate is considered an affordable luxury (Afoakwa, 2014a). The highest chocolate consumption takes place in Western Europe and North America: the areas that dominate the chocolate market (Figure 1.2).

In recent years, consumers became more sophisticated and are more aware of different cocoa varieties and different cocoa origins, moreover, the market of speciality chocolates is growing. The term speciality refers to products made with specific cocoa types or grades that require premium prices compared to the bulk cocoa. These products have their own supply chains to guarantee the provenance and quality of each lot before delivery (Fowler, 2009). The cocoas included in this class are cocoa from specific geographical or botanical origins, from environment-friendly farming practices such rainforest protection and certified organic, or cocoa purchased under schemes that help the farmers such as fair-trade. Specialty chocolate represents a fast-growing and dynamic market in many parts of the world, with global sales have risen by 18% within 2007 (Afoakwa, 2010a). For instance, due to the increasing social, environmental and economic concerns, the demand for organic and fair-trade chocolate has increased. Vlaeminck, Vandoren and Vranken (2016) showed that Belgium consumers are willing to pay a price premium of €0.84/100g for fair-trade chocolate. This happens especially when it is guaranteed that cocoa is produced under good labour conditions and that these conditions are frequently verified. These trends enlarge the concept of quality, including food safety, food origin, sustainability and environmental impact. The aforementioned aspects combined with the fact that the growing market of cocoa specialty did not match with a growth in production, challenges the complex cocoa-chocolate supply chain considerably (Fowler, 2009).

1.5 Traceability of cocoa

Due to the extensive geographical areas of cocoa production and the increasing consumption of specialty chocolates described in the previous section, the need to verify and certify cocoa bean origins has increased. In this regard, traceability can help ensuring food quality and authenticity. When bulk products such as cocoa are involved in the production, batch dispersion may happen already in the first steps of the supply chain. The cocoa beans produced by small-sized farmers are combined in larger and larger batches causing possible trading of unlabelled and/or mislabelled cocoa going unrecognised (Saltini et al., 2013). In the cocoa-chocolate sector, the length of the supply chain from South to North, the number of actors involved in the trade, the miscommunication between farmers and producers and the drastic transformation from beans to chocolate mentioned in the previous paragraphs increase the complexity of cocoa traceability.

Traceability is important not only in ensuring transparency along the supply chain but could also contribute to the sustainability of the production (Wognum, Bremmers, Trienekens, Van Der Vorst, & Bloemhof, 2011). Over the last decade, issues with corporate social responsibility in the cocoa sector emerged and focused on economic, social and environmental aspects. As a consequence, the demand in cocoa-importing countries for certified cocoa has increased and is expected to show continuous growth over the next years. For instance, the Dutch government requires 100% certified sustainable cocoa by 2025 (Government of the Netherlands, 2014; CBI, 2016). Promoting sustainability in the cocoa value chain will improve the livelihoods of cocoa farmers (Cocoa Barometer, 2012). For these reasons, an improvement in chocolate traceability is required.

Legislation frameworks are in place to ensure authenticity in terms of labelling compliance with the directive 2000/36/CE and the EC General Food Law Regulation 178/2002. For instance, the EC General Food Law Regulation 178/2002 (European Commission, 2002), applied since 2005, clearly states the establishment of a traceability system for all food products and food ingredients.

However, the General Food Law and the other legislations do not state specific methodologies that can be applied in the food industry. It would be beneficial for all parties involved if the legislation could be reinforced by reliable and automated technologies and/or analytical assessments that can authenticate the product's origin.

1.6 Extrinsic markers: current track and trace systems to assure quality and authenticity

Nowadays, several technologies based on (semi) automated data collection are available for tracing products and guarantee quality and authenticity using extrinsic product markers. Within these technologies, there are barcode, data matrix, Quick Response (QR) code and Radio Frequency Identification (RFID) devices. Barcode systems are the most used tracing technology as they offer high levels of automation and are not expensive but require manual intervention in entering the data in the database and/or sometimes in attaching the barcode on the packages, leaving room for human errors and manipulation. RFID allows tracking materials in real-time with high precision and no manual intervention. RFID devices have been successfully applied to different food products (Costa et al., 2013) and even to continuous granular flow (Kvarnström, 2011; Liang et al., 2013). These devices continue to build on previously developed offline solutions based on the introduction of specific tracers into grains using chemical compounds or radioactive tracers (Comba, Belforte, Dabbene, & Gay, 2013). However, the cost of tags is prohibitive in many applications (Bateman, 2015).

Blockchain technology is a new way to solve traceability issues and ensure transparency. This technology emerged in 2008 as key component of the bitcoin cryptocurrency, however, in the last years, the possibility to apply this technology in the food supply chain has been evaluated. This type of technology could be used to collect and store data such as farm origin, batch numbers, factory and processing steps, expiring dates, storage temperatures and shipping. Additionally, details from the chemical analysis could be stored. All parties involved in the network have to agree on the information that has to be acquired in each transaction. Once an agreement has been reached, permanent records cannot be changed. All information collected can provide critical data that potentially may reveal food safety issues concerned to the product, help retailers to manage the shelf life of products and allow consumers to access the story of foods on their label from any device. This technology can represent a valid solution to increase trust and transparency and further support the authentication of the product. However, blockchain technology is still at an early stage of implementation (Galvez, Mejuto, & Simal-gandara, 2018). The applicability of blockchain to cocoa-chocolate value chain has been tested by some producers (Tony's Chocoloney, 2018; Palamara, 2018), however, it has been estimated that 5–10 years of implementation are required.

1.7 Intrinsic markers

1.7.1 Current methodology to assure authenticity

An analytical authentication is a procedure that can detect products' origin information such as their botanical and geographical origin. This kind of verification detects the intrinsic product markers and could support both the legislatory framework and the track and trace systems. Analytical methods such as Mass Spectrometry (MS), spectroscopic and chromatographic techniques have been used for the assessment of food geographical and botanical origins (Luykx & van Ruth, 2008). Furthermore, deoxyribonucleic acid (DNA) analysis has been more and more applied in food authentication. Polymerase Chain Reaction (PCR) is a technique that is able to amplify low amount of DNA using known nucleic acid probes. Once amplified, the DNA fragments can be sequenced and identified. This technique has been used for olive oil cultivar identification (Woolfe & Primrose, 2004; Testolin & Lain, 2005) and species-specific methods for the detection of beef, chicken, lamb, pork and turkey (Carcea et al., 2009).

Within the group of MS techniques, Isotope Ratio-MS (IR-MS) for the measurements of stable isotope abundance and Inductively Coupled Plasma-MS (ICP-MS) for the detection of macro, micro- or trace-elements have been used to develop methods for the geographical assessment of several foodstuffs. These are within the most used analytical techniques for tracing foods (Kelly, Heaton, & Hoogewerff, 2005). What makes the isotope abundance a good geographical indication is that the isotope fraction is related to physico-chemical phenomena happening in nature. For example, the isotopic fractionation can occur with absorption and desorption, crystallisation and melting, evaporation and condensation. The natural variation that occurs in the isotopic ratio of hydrogen, carbon, nitrogen and oxygen and the heavy element strontium can be used for the geographical detection of food origin (Kelly et al., 2005). Moreover, the isotopic composition of the constituents of agricultural products (proteins, carbohydrates, fats, minerals) is depending on several factors both indicative of the geographical origin and some production steps. These factors include seasonal and geological variation such as altitude, soil composition etc. (coffee: Anderson & Smith, 2002; wine: Dutra et al., 2013; beef: Heaton, Kelly, Hoogewerff, & Woolfe, 2008), the use of fertilisers and the feed included in the diet of farm animals (Baroni et al., 2011),

Regarding ICP-MS, the macro, micro- or trace-elements are indicative of provenance, as they reflect trace mineral composition of the soil where the plant grows. Trace element availability in plants or plant products depends on several factors such as soil pH, humidity, porosity, clay and humic complex (Kelly et al., 2005). Consequently, the elemental composition of the different soils may provide unique markers in agricultural products that characterise geographical origin. ICP-MS has been successfully used to detect geographical origins of wines (Coetzee et al., 2005), tea and nuts (Gómez-Ariza, Arias-Borrego, & García-Barrera, 2006; Moreda-Piñeiro, Fisher, & Hill, 2003).

Other MS techniques have been used for geographical determination, such as Gas Chromatography-MS (GC-MS) and Proton Transfer Reaction–MS (PTR-MS). These techniques are used for the analysis of volatile organic compounds (VOCs). VOCs play an important role in food science and technology. They are generated along the entire production chain from the crop till consumption. For this reason, VOCs are useful to characterise the food product and the agro-industrial process (Biasoli, Yeretzian, Gasperi, & Märk, 2011). GC-MS has been used for detecting the origin of several dairy products such as cheese (i.e. mozzarella) and honey (Luykx & van Ruth, 2008). PTR-MS has been used to detect the geographical origin of different food products such as coffee (Yener et al., 2014), saffron (Masi et al., 2016), olives (Araghipour et al., 2008; Masi, Romani, Pandolfi, Heimler, & Mancuso, 2015), honey (Kuś & van Ruth, 2015), cheese (Galle et al., 2011), wines (Campbell-Sills et al., 2016) and also flower bulbs (van Ruth & de Visser, 2015). The possibility to trace the production of organic milk (Liu, Koot, Hettinga, Jong, & van Ruth, 2018) and characterise nutmeg (van Ruth et al., 2019) and other spices such as black/white pepper, chili paprika, cinnamon and saffron (Silvis, Luning, Klose, Jansen, & van Ruth, 2019) has also been tested.

Spectroscopic techniques offer information about the molecular structure of the samples. Nuclear Magnetic Resonance (NMR) (Consonni & Cagliani, 2010), fluorescence spectroscopy (Karoui, Bosset, Mazerolles, Kulmyrzaev, & Dufour, 2005) and Infrared (IR) spectroscopy (Lohumi, Lee, Lee, & Cho, 2015), have been widely used for food origin characterisation (Luykx & van Ruth, 2008). For instance, NMR has often been used for the geographical discrimination of wine. A study about Chinese wines highlighted the possibility to differentiate dry red wine from three different regions (Du, Bai, Zhang, & Liu, 2007). The separation of the wine samples was due to variation in fermentation and environmental conditions. These differences were linked with specific components detected by NMR such as ethyl esters, 2,3-butanediol, and 3-hydroxyl-2-butanone, organic acid, proline and glycerol.

Chromatographic techniques have been studied to verify the origin of foods by detecting (or: analysing/profiling) intrinsic markers such as aroma, sugar, phenolic and flavour compound profiling using high-performance liquid chromatography (HPLC) and gas chromatography (GC) (Carcea et al., 2009). For instance, European wines from different geographical origin have been classified according to the phenolic profiles using with HPLC in combination with an ultraviolet-visible (UV–VIS) and/or fluorescence detector (Rodriquez-Delgado, Gonzalez-Hernandez, Conde-Gonzales, & Perez-Trujillo, 2002). The possibility to differentiate the wine using the phenolic profile is mainly related to variation of the phenolic composition related to the location, climate, soil characteristics, cultivation system, production process and ageing.

1.7.1.1 Analytical fingerprinting

Geographical and botanical characteristics might be difficult to be measured with targeted methods. Targeted techniques in general, are focused on the determination of specific markers which typically cannot fully characterise a product according to the origin, therefore reducing its applicability in food authentication. A more comprehensive insight into the sample composition is possible by obtaining for instance a complete spectrum or an image of the test material (Capuano & van Ruth, 2012). This untargeted way of analysis is known as fingerprinting. All information included in the fingerprint (or a selected part of it) can be used to assess and guarantee the characteristics of the test sample.

Fingerprints can be obtained from several analytical techniques such as chromatography, spectroscopy and mass spectrometry. The profile generated by one of these techniques is then statistically analysed to detect markers that can be linked to product authentication such as country of origin (Luykx & van Ruth, 2008), organic (Tres & van Ruth, 2011), kosher, halal (Rohman & Che Man, 2012) or "free-range" production systems (Alonso et al., 2008).

1.7.2 Cocoa and chocolate intrinsic markers for botanical and geographical characteristics

Cocoa and chocolate intrinsic markers have been studied with the application of several analytical techniques to discriminate cocoa botanical and geographical origins. Mass spectrometry (Qin et al., 2017) and spectroscopy (Marseglia et al., 2016; Teye, Huang, Dai, & Chen, 2013) have been used to detect cocoa beans origin.

Cocoa beans from different botanical origin (botanical varieties) have significant differences in terms of acidity, flavour and chemical composition, flavour precursors and polyphenols, which are reflected in the characteristics of the final product. Three varieties of cocoa are commonly recognised: Criollo (fine grade), Forastero (bulk grade) and Trinitario (fine grade). The latter is a hybrid of the two former varieties. There is also a niche variety, genetically similar to Criollo, grown specifically in Ecuador, called Nacional (Afoakwa et al., 2008; Counet, Ouwerx, Rosoux, & Collin, 2004; Motamayor et al., 2008; Saltini et al., 2013). The varieties differ in pod characteristics, yields and resistance to pests and diseases (Afoakwa, 2010b; Afoakwa et al., 2008; Saltini et al., 2013). The botanical origin of the cocoa bean will influence the cocoa bean composition such as concentrations and the type of proteins, carbohydrates and polyphenols (Afoakwa et al., 2008). This determines the concentrations and type of flavour precursors formed during fermentation and drying, which in turn will influence both the type and intensity of flavour of the beans (Afoakwa et al., 2008; Counet et al., 2004). Each cocoa type has a unique potential flavour character (Afoakwa et al., 2008) that has been confirmed by several studies focused on the measurements of the volatile profiles (Cevallos-cevallos, Gysel, Maridueña-Zavala, & Molina-Miranda, 2018; Frauendorfer & Schieberle, 2008; Qin et al., 2017).

Criollo cocoa contains relatively large amounts of pyrazines (Reineccius, Keeney, & Weissberger, 1972) and has a low pH which easily affects the flavour profile (Saltini et al., 2013). Cocoa beans main polyphenol class are the procyanidins, which contribute to the astringency and bitterness taste of the chocolate. The Criollo variety contains the highest levels of these compounds, even when they originate from different countries (Counet et al., 2004). The chocolate produced with the Criollo variety has high levels of aroma compounds, supported by a high concentration of flavour precursors in cocoa beans such as reducing sugars before the roasting process.

Forastero cocoa have lower concentrations of aromatic compounds than Criollo cocoa beans and the flavour precursors before roasting are also expected to be slightly lower. After fermentation and drying, the Forastero beans have a higher pH when compared with Criollo beans (Saltini et al., 2013). Because of these chemical differences, chocolate produced from Forastero beans is less acid, less astringent and less bitter than chocolate produced from either Criollo or Trinitario beans (Kongor et al., 2016; Sukha, Butler, Umaharan & Boult, 2007). Chocolate produced with Forastero beans has low procyanidins levels and an amount of pyrazines similar to chocolate produced with Criollo beans (Counet et al., 2004).

Trinitario beans are known to have strong basic chocolate characters and wine notes that are not found in other varieties (Afoakwa et al., 2008). The chocolate produced with Trinitario cocoa beans showed similar levels of procyanidins to products manufactured with Forastero cocoa beans (Counet et al., 2004). However, only small differences were found in sensory studies when comparing cocoa liquors made with Trinitario cocoa beans and liquors made with bulk cocoa beans (Sukha et al., 2007).

Cocoa batches from different geographical origins might also show significant differences in several attributes such as the contents of amino acids (Marseglia, Palla & Caligiani, 2014; Rohsius, Matissek, & Lieberei, 2005), peptides (Caligiani, Marseglia, Prandi, Palla & Sforza, 2016), triglycerides (Hernandez, Castellote, & Permanyer, 1991), polyphenols (Arlorio et al., 2008; Caligiani, Cirlini, Palla, Ravaglia, & Arlorio, 2007) and volatile acids (Jinap & Dimick, 1990). For instance, cocoa beans from Congo, Mexico, Perù, Santo Domingo and Sao Thome contain high levels of total free amino acids compared to the ones from Indonesia, Ivory Coast, Nigeria, Sulawesi, Tanzania and Venezuela, while dried fermented cocoa beans from Southeast Asia and the South Pacific are characterised by a higher acidity (higher levels of lactic and acetic acids) than those of South African origin (Holm, Astong, & Douglas, 1993). These differences are mainly origin-related but several studies showed that also the fermentation and drying procedures locally applied can influence the final concentrations of these compounds (Caligiani et al., 2016). As in the case of the cocoa variety, all these differences in chemical composition are reflected in the flavour profile of the cocoa from different geographical areas as showed by Jinap, Dimick and Hollender (1995). Through the investigation of the flavour profile, it was possible to also detect the geographical origin of semi-finished product such as cocoa liquor (Counet et al., 2004) and finished products like chocolate (Cambrai et al., 2010; Jinap et al., 1995).

The transformation of raw material constituents along the production is influenced by the production steps making difficult to consistently link the origin of the raw material with the finished product. Several studies followed this transformation along the primary processing; the change in the amino acids, sugars, polyphenols, pyrazines, flavour profile and microflora according to the fermentation and drying conditions (Saltini et al., 2013). These components have been studied to optimise the production and guarantee a good quality of the raw beans. Specifically, a lot of attention has focused on the flavour profile being an essential attribute of cocoa quality and one of the most significant consumer parameters. Rodriguez-Campos, Escalona-Buendía, Contreras-Ramos, & Orozco-Avila, (2012) followed the effect of the fermentation and drying on volatile composition underlining the differences in volatile formation according to the fermentation temperature and the drying methods applied, such as sun drying or the application of convection ovens. The study highlighted that six days of fermentation were optimal to produce desirable cocoa flavour compounds such as ethyl phenyl acetate, 2-penthyl acetate, acetoin, 2-phenyl-2-butenal, 2,3-butanedione, ethyl 3-methylbutanoate, ethyl 2-methylbutanoate and the reduction of the off-flavours such as isovaleric, isobutyric and propionic acid and 2-methoxy phenol. Moreover, it was shown that drying the cocoa beans at 70°C after six days of fermentation were the optimal conditions for cocoa beans fermentation and drying. The volatile profile generated at these conditions was similar to the one obtained by sun drying with the production of desirable compounds such as tetramethylpyrazine.

The evolution of the volatiles (Frauendorfer & Schieberle, 2008), amino acids, sugars and phenols (Brito et al., 2001) during the secondary production steps, such roasting and conching (Counet et al., 2004), has been extensively investigated as well. Quantitative rather than qualitative differences were found by Frauendorfer & Schieberle, (2008) when analysing the volatile profile of unroasted and roasted cocoa beans. For instance, compounds detected in roasted beans such as 3-methylbutanoic acid, ethyl 2-methylbutanoate, and 2-phenylethanol, were already present before the heat treatment.

However, while the profile of the unroasted beans was dominated by 3-methylbutanoic acid and acetic acid the profile of the roasted beans showed a strong increase of Strecker aldehydes such as 3-methylbutanal and phenylacetaldehyde, and 4-hydroxy-2,5-dimethyl-3(2H)- furanone, linking these compounds with the aroma changes after roasting. Mainly quantitative differences were also found when comparing samples before and after conching. The concentrations of 2-phenyl-5-methyl- 2-hexenal, furaneol and branched pyrazines significantly increased during conching while most of the Strecker aldehydes were lost by evaporation.

Due to the alterations made to the product along the supply chain, it is necessary to investigate the link between the cocoa beans and the chocolate intrinsic markers. This has been done in a few studies in order to assess the quality in terms of flavour development (Ascrizzi, Flamini, Tessieri & Pistelli, 2017; Braga et al., 2018) and to predict the quality of the product by comparing cocoa beans and corresponding cocoa mass (Hernandez & Rutledge, 1994).

1.7.2.1 Cocoa and chocolate intrinsic markers: a preliminary study

To further assess the link between the raw material and the finished product, several studies focused on the possibility to perceive origin differences during chocolate consumption (Sukha et al., 2007). At the beginning of this thesis project, in a preliminary experiment, a consumer panel of Wageningen University visitors assessed chocolates of two different brands (Brand A and B) using cocoa beans of different origin. The chocolates were examined in a triangle test setting (Figure 1.3). According to the results of the binominal test ($p \leq 0.05$), the participants (360 in total, 60 for each comparison) could better differentiate chocolate of different origins of Brand B (Figure 1.3). Brand B is more oriented to a niche market while Brand A is more oriented to large-volume sales. Within the chocolates belonging to Brand B, the participants could differentiate chocolate produced with beans from a single origin compared to chocolate made from mixed beans, i.e. beans from multiple origins (MIX-GHANA-GHANA: p-values ≤ 0.05). Regarding the single-origin chocolates, the participants could distinguish Ghana bean chocolate from Peruvian bean chocolate (GHANA-PERÙ-PERÙ). However, it is difficult to determine if the botanical or geographical origin of the cocoa beans used for the chocolate production made the participants perceive the difference as Ghana chocolates were produced with Forastero beans while Perù chocolates were produced with Trinitario. The possibility for consumers to perceive differences related to cocoa origin in terms of botanical origin of the cocoa beans used to manufacture chocolate has been also considered in the study of Sukha et al. (2007) in which an organoleptic assessment protocol has been optimised to describe and quantify different flavour attributes of cocoa liquors made from Ghana and Trinitario beans. These studies highlighted the contribution of botanical and geographical origin characteristics and the impact of processing conditions (fermentation and drying) in the different locations on the final product flavour. Consequently, the concept of 'terroir' could also apply to chocolate. This concept, already well established for wines, can also provide a scientific basis for cacao quality certification programs.

LET'S EAT CHOCOLATE!

CAN CONSUMERS PERCEIVE THE ORIGIN DIFFERENCE?
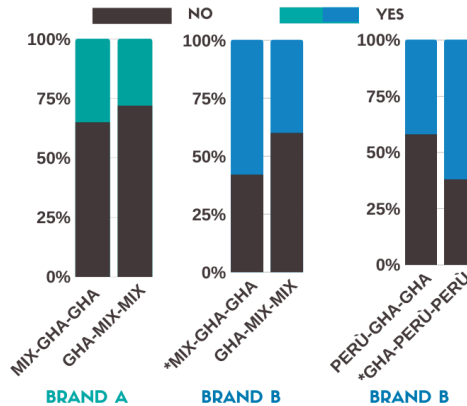
■ BRAND A   ■ BRAND B

TRIANGLE COMPARISONS:

MIX VS SINGLE-ORIGIN	MIX-GHANA-GHANA (30*2)
	GHANA-MIX-MIX (30*2)
MIX VS SINGLE-ORIGIN	MIX-GHANA-GHANA (30*2)
	GHANA-MIX-MIX (30*2)
SINGLE-ORIGIN	PERÙ-GHANA-GHANA (30*2)
	GHANA-PERÙ-PERÙ (30*2)

CAN YOU PERCEIVE THE DIFFERENT ONE?

■ NO   ■ YES

BRAND A — MIX-GHA-GHA, GHA-MIX-MIX
BRAND B — *MIX-GHA-GHA, GHA-MIX-MIX
BRAND B — PERÙ-GHA-GHA, *GHA-PERÙ-PERÙ

*p ≤ 0.05. Significance established on the minimum number of correct judgements for the triangle test
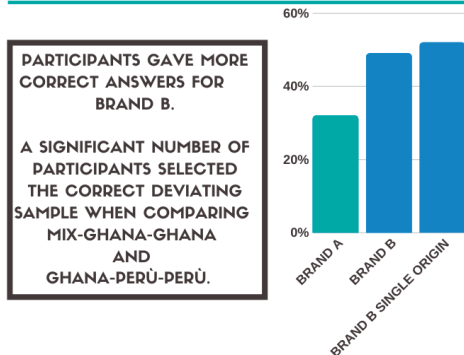
PARTICIPANTS GAVE MORE CORRECT ANSWERS FOR BRAND B.

A SIGNIFICANT NUMBER OF PARTICIPANTS SELECTED THE CORRECT DEVIATING SAMPLE WHEN COMPARING MIX-GHANA-GHANA AND GHANA-PERÙ-PERÙ.

BRAND A / BRAND B / BRAND B SINGLE ORIGIN

Figure 1.3. Consumer panel test at Wageningen University. All the chocolates investigated had the same cocoa percentage (75%). Chocolates from Brand B are from different botanical origins: MIX: *Criollo-Forastero*; GHANA: *Forastero*; PERÙ: *Trinitario*
*p-value ≤ 0.05. Significance established on the minimum number of correct judgements for the triangle test. 30*2: the test was performed in two rounds with 30 participants each.

1.8 Knowledge gap and study justification

As highlighted in the previous section, the majority of the studies available aim to characterise cocoa origin to improve the aroma quality of cocoa and chocolate. Few studies are focused on the investigation of techniques able to support cocoa/chocolate traceability. NIR spectroscopy was tested for geographical discrimination of cocoa beans (Teye et al., 2013) and a method to trace back the final product was investigated comparing the multi-elemental composition of cocoa beans and chocolate by ICP-MS (Bertoldi, Barbero, Camin, Caligiani, & Larcher, 2016). However, there are no studies following the botanical and geographical characteristics from cocoa beans to chocolate to assess the authenticity of the product along the supply chain.

Considering the consumer study described in section 1.7.2.1, it appears feasible for consumers to distinguish some terroir chocolates. Because of the relevance of volatile and non-volatile compounds on the final product as described in this section, the consumer panel results were the reason for selecting volatile and non-volatile compound analysis techniques for examining the cocoa origin differences and following them all the way from the consumer to the chocolate and to the raw material. If we wish to be able to discern the origin of beans in the final product, we must ensure that it is still present in the final product. Therefore, in this project, we started at the side of the final product and worked backwards towards the beans.

1.9 Objectives and general approach

The main objective of this thesis is to discover the intrinsic characteristics of cocoa beans that reflect their botanical and geographical origins and that are carried along the cocoa–chocolate supply chain. The extraction of this information along the supply chain is explored to provide a scientific link between raw materials and the finished product.

The detailed objectives are:

- To explore which intrinsic, compositional characteristics of the **botanical origin** of the cocoa beans are present after cocoa bean fermentation, drying, chocolate manufacturing and during consumption.
- To explore which intrinsic, compositional characteristics of the **geographical origin** of the cocoa beans are present after cocoa bean fermentation, drying, chocolate manufacturing and during consumption.
- To investigate which intrinsic compositional characteristics are most useful to identify botanical and/or geographical origin along the cocoa-chocolate supply chain.

Rapid and high throughput analytical techniques have been used to chemically describe the intrinsic compositional characteristics at various stages of cocoa-chocolate production. The reflectance of the cocoa beans origin along the supply chain is investigated starting from the chocolate consumption in *Chapter 2*. Moving back along the supply chain, the possibility to extract geographical and botanical characteristics is further tested in chocolates available in retail (*Chapters 3 and 4*). *Chapter 5* provides a direct comparison between chocolate and corresponding beans used in the production to determine the native cocoa beans compositional markers and the ones formed during processing from bean components and highlights the components added or modified by other ingredients of the chocolate. Finally, the raw material investigation in *Chapter 6* completes the assessment of botanical and geographical cocoa beans characteristics along the supply chain.

The various components of this study are presented in Figure 1.4.

COCOA REFLECTANCE IN CHOCOLATE

From consumer to farmer

2 CHOCOLATE IDENTITY: IN VIVO RELEASE OF VOCS DURING COMSUPTION OF DARK CHOCOLATES

- Proton Transfer Reaction-Quadrupole Interface TIme of Flight-Mass Spectrometry (PTR-Qi ToF-MS)

3 CHOCOLATE IDENTITY: VOCS COMPOSITION OF DARK CHOCOLATES

- High Sensitivity PTR-MS (HS-PTR-MS)

4 CHOCOLATE IDENTITY: NON-VOLATILE PROFILE OF DARK CHOCOLATES

- Flow Infusion- Electrospray Ionization-MS (FI-ESI-MS)

5 ALTERATIONS FROM COCOA BEANS TO CHOCOLATE

- HS-PTR-MS
- Inductively Coupled Plasma-MS (ICP-MS)
- isotope Ratio-MS (IR-MS)

6 COCOA BEANS IDENTITY

- PTR-Qi ToF- MS
- Hyperspectral camera (HSI)

7 GENERAL DISCUSSION

Figure 1.4. Structure of the thesis chapters.

21

1.10 Outline of the thesis

The thesis is outlined as follows. The current chapter (***Chapter 1***) provides a general introduction on the topics covered in the thesis to guide the reader through the different chapters.

In ***Chapter 2*** the reflectance of cocoa beans origin in the volatile organic compound (VOC) profiles released during consumption of dark chocolates is examined. Proton Transfer Reaction-Quadrupole interface Time of Flight-Mass Spectrometry (PTR-Qi ToF-MS) is applied to analyse nosespace (NS) VOC profiles of 10 subjects while they were eating 10 different chocolates manufactured with beans of different botanical origins (Criollo-Forastero-Trinitario) and geographical origins (Africa-South America-Asia). This chapter aims to investigate whether traits of cocoa botanical and geographical origins persist in the product and during chocolate consumption. By analysing the concentrations of VOCs over time in the nose of subjects while they are eating, new insights about aroma release according to chocolate origin are gained. Chapter 2 underlines how the intrinsic characteristics of chocolate and the technological process complicate the assessment of the typical features of this product.

Chapter 3 studies and maps the variation in VOCs composition of ninety dark chocolates available on the Dutch market. The analysis includes chocolates declared as single-origin chocolate (Africa-South America-Asia) and chocolate declared as mixed-origin chocolates. The botanical origin (Criollo-Forastero-Trinitario) of single-origin bars is also considered. The VOC profiles generated by High Sensitivity-PTR-MS (HS-PTR-MS) are used as fingerprints (136 masses per sample) and are investigated using chemometric tools to elucidate information on the geographical and botanical origin of cocoa beans. The impact of brand-related processing on the chocolate VOC composition is also examined.

Chapter 4 evaluates the non-volatile profiles of 57 single-origin dark chocolates available on the Dutch market. The applicability of Flow Infusion-Electrospray Ionization-MS (FI-ESI-MS) to assess the botanical and geographical origin intrinsic traits in finished chocolate is tested. The analysis included chocolates that according to the label were manufactured with beans from different botanical origins (Criollo-Forastero-Trinitario) and geographical origins (Africa-South America-Asia). The non-volatile profiles generated by FI-ESI-MS are used as fingerprints. In this study, the importance of diminishing the influence of brand-related factors to improve geographical separation is addressed as well.

Chapter 5 aims to link chocolate with the corresponding beans used for the production by MS techniques. Considering the results of Chapter 3 and 4, we decided to focus on a single brand. Volatile, elemental and stable isotope signatures of cocoa beans of 10 different origins and 11 corresponding chocolates are determined by HS-PTR-MS, ICP-MS and IR-MS, respectively. This was done to explore the magnitude of similarity and dissimilarity between the raw material and the finished product. The possibility to use the fingerprints generated by these MS techniques for different product properties is addressed. Moreover, it is defined which technique among HS-PTR-MS, ICP-MS and IR-MS correlates raw material best to its finished product.

Moving back along the supply chain, ***Chapter 6*** focuses on the cocoa beans. Fifty-nine fermented and dried Forastero cocoa beans from 23 different geographical origins (Africa-America-Southeast Asia) are investigated using PTR-QiToF-MS and Hyperspectral Imaging (HIS) to elucidate the geographical information in the beans. Beans from the same variety and same brand are analysed to reduce the variability within this natural product. The possibility to use volatile and spectral fingerprints to detect the geographical information is assessed. Moreover, the possibility to use the farming practises (fermentation and drying) as provenance trait is questioned in this chapter.

In Chapter 7, the ***General discussion***, the findings of the thesis are integrated. The impact of the research, together with limitations and recommendations for further research are presented.

References

Aculey, P. C., Snitkjaer, P., Owusu, M., Bassompiere, M., Takrama, J., Nørgaard, L., … Nielsen, D. S. (2010). Ghanaian cocoa bean fermentation characterized by spectroscopic and chromatographic methods and chemometrics. *Journal of Food Science*, *75*(6), S300-7. https://doi.org/10.1111/j.1750-3841.2010.01710.x

Afoakwa, E. O. (2012). Chocolate and cocoa, flavor and quality. *Kirk-Othmer Encyclopedia of Chemical Technology,* pp. 1–19.

Afoakwa, E. O. (2010a). Chocolate Production and Consumption Patterns. *Chocolate Science and Technology*, pp. 1–11. https://doi.org/10.1002/9781444319880.ch1

Afoakwa, E. O. (2010b). Cocoa Cultivation, Bean Composition and Chocolate Flavour Precursor Formation and Character. *Chocolate Science and Technology*, pp. 12–34. https://doi.org/10.1002/9781444319880.ch2

Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2008). Flavor formation and character in cocoa and chocolate: a critical review. *Critical Reviews in Food Science and Nutrition*, *48*(9), 840–857. https://doi.org/10.1080/10408390701719272

Afoakwa, E. O., Quao, J., Takrama, J., & Budu, A. S. (2013). Chemical composition and physical quality characteristics of Ghanaian cocoa beans as affected by pulp pre-conditioning and fermentation. *Journal of food science and technology, 50*(6), 1097–1105. https://doi.org/10.1007/s13197-011-0446-5

Afoakwa, E. (2014a). Drying Techniques, Storage Practices and Trading Systems. In E. Afoakwa (Ed.), *Cocoa production and processing technology* (pp. 139–146). Boca Raton, FL, USA: Taylor & Francis Group

Afoakwa, E. (2014b). Traditional and modern cocoa cultivation practices. In E. Afoakwa (Ed.), *Cocoa production and processing technology* (pp. 79-90). Boca Raton, FL, USA: Taylor & Francis Group

Afoakwa, E. (2014c). Post-harvest treatments and technologies of cocoa. In E. Afoakwa (Ed.), *Cocoa production and processing technology* (pp. 109-119). Boca Raton, FL, USA: Taylor & Francis Group

Afoakwa, E. (2014d). Chocolate manufacturing and processing technology. In E. Afoakwa (Ed.), *Cocoa production and processing technology* (pp. 183-196). Boca Raton, FL, USA: Taylor & Francis Group

Alonso, R., Rodríguez-Estévez, V., Domínguez-Vidal, A., Ayora-Cañada, M. J., Arce, L., & Valcárcel, M. (2008). Ion mobility spectrometry of volatile compounds from Iberian pig fat for fast feeding regime authentication. *Talanta*, *76*(3), 591–596. https://doi.org/10.1016/j.talanta.2008.03.052

Anderson, K. A., & Smith, B. W. (2002). Chemical profiling to differentiate geographic growing origins of coffee. *Journal of Agricultural and Food Chemistry*, *50*(7), 2068–2075. https://doi.org/10.1021/jf011056v

Aprotosoaie, A. C., Luca, S. V., & Miron, A. (2016). Flavor Chemistry of Cocoa and Cocoa Products-An Overview. *Comprehensive Reviews in Food Science and Food Safety*, *15*(1), 73–91. https://doi.org/10.1111/1541-4337.12180

Araghipour, N., Colineau, J., Koot, A., Akkermans, W., Rojas, J. M. M., Beauchamp, J., … van Ruth, S. M. (2008). Geographical origin classification of olive oils by PTR-MS. *Food Chemistry*, *108*(1), 374–383. https://doi.org/10.1016/j.foodchem.2007.10.056

Arlorio, M., Locatelli, M., Travaglia, F., Coïsson, J.-D., Grosso, E. Del, Minassi, A., … Martelli, A. (2008). Roasting impact on the contents of clovamide (N-caffeoyl-L-DOPA) and the antioxidant activity of cocoa beans (Theobroma cacao L.). *Food Chemistry*, *106*(3), 967–975. https://doi.org/10.1016/j.foodchem.2007.07.009

Ascrizzi, R., Flamini, G., Tessieri, C., Pistelli, L. (2017). From the raw seed to chocolate: Volatile profile of Blanco de Criollo in different phases of the processing chain. *Microchemical Journal, 133*, 474-479.

Baroni, M. V., Podio, N. S., Badini, R. G., Inga, M., Ostera, H. A., Cagnoni, M., … Wunderlin, D. A. (2011). How much do soil and water contribute to the composition of meat? A case study: Meat from three areas of Argentina. *Journal of Agricultural and Food Chemistry*, *59*(20), 11117–11128. https://doi.org/10.1021/jf2023929

Bateman, A.H (2015). Tracking the value of traceability. *Supply Chain Management Review, 9,* pp. 8-10. https://ctl.mit.edu/sites/ctl.mit.edu/files/SCMR1511_InnovStrategies.pdf

Bertoldi, D., Barbero, A., Camin, F., Caligiani, A., & Larcher, R. (2016). Multielemental fingerprinting and geographic traceability of Theobroma cacao beans and cocoa products. *Food Control*, *65*, 46–53. https://doi.org/10.1016/j.foodcont.2016.01.013

Braga, C.G.N., Luciana F. Oliveira, Juliana C. Hashimoto, Mariana R. Gama, Priscilla Efraim, Ronei J. Poppi, F. A. (2018). Study of volatile profile in cocoa nibs, cocoa liquor and chocolate on production process using GC × GC-QMS. *Microchemical Journal, 141*, pp. 353-361.

Brito, E. S. De, Garcı, Pezoa, N. H., Gallao, M. I., Cortelazzo, A. L., Fevereiro, P. S., & Braga, M. R. (2001). Structural and chemical changes in cocoa (Theobroma cacao L) during fermentation, drying and roasting. *Journal of the Science of Food and Agriculture, 81*(2), 281–288.

Caligiani, A., Cirlini, M., Palla, G., Ravaglia, R., & Arlorio, M. (2007). GC-MS Detection of Chiral Markers in Cocoa Beans of Different Quality and Geographic Origin. *Chirality: The Pharmacological, Biological, and Chemical Consequences of Molecular Asymmetry, 19*(4), 329-334. https://doi.org/10.1002/chir

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

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*(3), 1478–1483. https://doi.org/10.1021/jf903471e

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

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

Carcea, M., Brereton, P., Hsu, R., Kelly, S., Marmiroli, N., Melini, F., … Wenping, D. (2009). Food authenticity assessment: Ensuring compliance with food legislation and traceability requirements. *Quality Assurance and Safety of Crops and Foods*, *1*(2), 93–100. https://doi.org/10.1111/j.1757-837X.2009.00011.x

CBI. (2016). CBI product Factsheet: Cocoa in The Netherlands. Retrieved from https://www.cbi.eu/sites/default/files/market_information/researches/product-factsheet-netherlands-cocoa-2016.pdf/

Cevallos-cevallos, J. M., Gysel, L., Maridueña-zavala, M. G., & Molina-miranda, M. J. (2018). Time-Related Changes in Volatile Compounds during Fermentation of Bulk and Fine-Flavor Cocoa (Theobroma cacao) Beans. *Journal of food quality, 2018*.

Cocoa Barometer (2012). The Cocoa Barometer 2012. Retrieved from http://www.cocoabarometer.org/Cocoa_Barometer/Download_files/Cocoa%20Barometer%20Full%20 2012.pdf

Coetzee, P. P., Steffens, F. E., Eiselen, R. J., Augustyn, O. P., Balcaen, L., & Vanhaecke, F. (2005). Multi-element analysis of South African wines by ICP− MS and their classification according to geographical origin. *Journal of Agricultural and Food Chemistry, 53*(13), 5060-5066.

Comba, L., Belforte, G., Dabbene, F., & Gay, P. (2013). Methods for traceability in food production processes involving bulk products. *Biosystems Engineering*, *116*(1), 51–63. https://doi.org/10.1016/j.biosystemseng.2013.06.006

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 (Ed.), *Advances in food and nutrition research*, *59*, pp. 87–165. Burlington: Academic Press

Costa, C., Antonucci, F., Pallottino, F., Aguzzi, J., Sarriá, D., & Menesatti, P. (2013). A Review on Agri-food Supply Chain Traceability by Means of RFID Technology. *Food and Bioprocess Technology*, *6*(2), 353–366. https://doi.org/10.1007/s11947-012-0958-7

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

Du, Y. Y., Bai, G. Y., Zhang, X., and Liu, M. L. (2007). Classification of wines based oncombination of1H NMR spectroscopy and principal component analysis. *Chinese Journal of Chemistry, 25*(7), 930–93

Dutra, S. V., Adami, L., Marcon, A. R., Carnieli, G. J., Roani, C. A., Spinelli, F. R., … Vanderlinde, R. (2013). Characterization of wines according the geographical origin by analysis of isotopes and minerals and the influence of harvest on the isotope values. *Food Chemistry*, *141*(3), 2148–2153. https://doi.org/10.1016/j.foodchem.2013.04.106

Fountain, A., & Huetz-Adams, F. (2018). Cocoa barometer 2018. Retrieved from https://www.suedwindinstitut.de/files/Suedwind/Publikationen/2018/201809%20Cocoa%20Barometer %202018%20English.pdf

Fowler, (2009). Cocoa beans: from tree to factory. In Beckett, S.T. (Ed.), *Industrial Chocolate Manufacture and Use:* Fourth Edition (pp. 10-47). York, UK: Blackwell Publishing Ltd

Frauendorfer, F., & Schieberle, P. (2008). Changes in key aroma compounds of Criollo cocoa beans during roasting. *Journal of Agricultural and Food Chemistry*, *56*(21), 10244–10251. https://doi.org/10.1021/jf802098f

Galle, S. A., Koot, A., Soukoulis, C., Cappellin, L., Biasioli, F., Alewijn, M., & van Ruth, S. M. (2011). Typicality and geographical origin markers of protected origin cheese from the Netherlands revealed by PTR-MS. *Journal of Agricultural and Food Chemistry*, *59*(6), 2554–2563. https://doi.org/10.1021/jf104170r

Galvez, J. F., Mejuto, J. C., & Simal-gandara, J. (2018). Trends in Analytical Chemistry Future challenges on the use of blockchain for food traceability analysis. *Trends in Analytical Chemistry*, *107*, 222–232. https://doi.org/10.1016/j.trac.2018.08.011

Gómez-Ariza, J. L., Arias-Borrego, A., & García-Barrera, T. (2006). Multielemental fractionation in pine nuts (Pinus pinea) from different geographic origins by size-exclusion chromatography with UV and inductively coupled plasma mass spectrometry detection. *Journal of Chromatography A*, *1121*(2), 191–199. https://doi.org/10.1016/j.chroma.2006.04.025

Government of the Netherlands (2014). Sustainable chocolate depends on fair price for farmers. Retrieved from https://www.government.nl/latest/news/2014/06/10/sustainable-chocolate-depends-on-fair-price-for-farmers/

Gray, M.P. (2009). Moulding, enrobing and cooling chocolate products. In Beckett, S.T. (Ed.), *Industrial Chocolate Manufacture and Use:* Fourth Edition (pp. 320-357). York, UK: Blackwell Publishing Ltd

Heaton, K., Kelly, S. D., Hoogewerff, J., & Woolfe, M. (2008). Verifying the geographical origin of beef: The application of multi-element isotope and trace element analysis. *Food Chemistry*, *107*(1), 506–515. https://doi.org/10.1016/j.foodchem.2007.08.010

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

Hernandez, V., & Rutledge, D. N. (1994). Multivariate Statistical Analysis of Gas Chromatograms to Differentiate Cocoa Masses by Geographical Origin and Roasting Conditions. *Analyst, 119*(6), 1171-1176.

Holm, C. S., Astong, J. W., & Douglas, K. (1993). The Effects of the Organic Acids in Cocoa on the Flavour of Chocolate. *Journal of the Science of Food and Agriculture, 61*(1), 65-71.

Jinap, S., Dimick, P. S., & Hollender, R. (1995). Flavour evaluation of chocolate formulated from cocoa beans from different countries. *Food Control*, *6*(2), 105–110. https://doi.org/10.1016/0956-7135(95)98914-M

Jinap, S. & Dimick, P. S. (1990). Acidic Characteristics of Fermented and Dried Cocoa Beans from Different Countries of Origin. *Journal of Food Science*, *55*(2), 547–550. https://doi.org/10.1111/j.1365-2621.1990.tb06806.x

Karoui, R., Bosset, J. O., Mazerolles, G., Kulmyrzaev, A., & Dufour, É. (2005). Monitoring the geographic origin of both experimental French Jura hard cheeses and Swiss Gruyère and L'Etivaz PDO cheeses using mid-infrared and fluorescence spectroscopies: A preliminary investigation. *International Dairy Journal*, *15*(3), 275–286. https://doi.org/10.1016/j.idairyj.2004.07.006

Kelly, S., Heaton, K., & Hoogewerff, J. (2005). Tracing the geographical origin of food: The application of multi-element and multi-isotope analysis. *Trends in Food Science and Technology*, *16*(12), 555–567. https://doi.org/10.1016/j.tifs.2005.08.008

Kongor, J. E., Hinneh, M., de Walle, D. Van, Afoakwa, E. O., Boeckx, P., & Dewettinck, K. (2016). Factors influencing quality variation in cocoa (Theobroma cacao) bean flavour profile - A review. *Food Research International*, *82*(January), 44–52. https://doi.org/10.1016/j.foodres.2016.01.012

Kuś, P. M., & van Ruth, S. (2015). Discrimination of Polish unifloral honeys using overall PTR-MS and HPLC fingerprints combined with chemometrics. *LWT - Food Science and Technology*, *62*(1), 69–75. https://doi.org/10.1016/j.lwt.2014.12.060

Kvarnström, B. B. & K. V. (2011). RFID to Improve Traceability in Continuous Granular Flows — An Experimental Case Study. *Quality Engineering, 23*(4), 343-357. https://doi.org/10.1080/08982112.2011.602278

Liang, K., Thomasson, J. A., Shen, M. X., Armstrong, P. R., Ge, Y., Lee, K. M., & Herrman, T. J. (2013). Ruggedness of 2D code printed on grain tracers for implementing a prospective grain traceability system to the bulk grain delivery system. *Food Control, 33*(2), 359-365. https://doi.org/10.1016/j.foodcont.2013.03.029

Liu, N., Koot, A., Hettinga, K., Jong, J. De, & van Ruth, S. M. (2018). Portraying and tracing the impact of different production systems on the volatile organic compound composition of milk by PTR- (Quad) MS and PTR- (ToF) MS. *Food Chemistry*, *239*, 201–207. https://doi.org/10.1016/j.foodchem.2017.06.099

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

Luykx, D. M. a. M., & van Ruth, S. M. (2008). An overview of analytical methods for determining the geographical origin of food products. *Food Chemistry*, *107*(2), 897–911. https://doi.org/10.1016/j.foodchem.2007.09.038

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

Marseglia, A., Palla, G., & Caligiani, A. (2014). Presence and variation of γ-aminobutyric acid and other free amino acids in cocoa beans from different geographical origins. *Food research international, 63*, 360-366. https://doi.org/10.1016/j.foodres.2014.05.026

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

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

McShea, A., Ramiro-Puig, E., Munro, S. B., Casadesus, G., Castell, M., & Smith, M. a. (2008). Clinical benefit and preservation of flavonols in dark chocolate manufacturing. *Nutrition Reviews*, *66*(11), 630–641. https://doi.org/10.1111/j.1753-4887.2008.00114.x

Moreda-Piñeiro, A., Fisher, A., & Hill, S. J. (2003). The classification of tea according to region of origin using pattern recognition techniques and trace metal data. *Journal of Food Composition and Analysis*, *16*(2), 195–211. https://doi.org/10.1016/S0889-1575(02)00163-1

Montserrat Camacho Ángeles, M. (2012). La imagen bajo la perspectiva de la cosmovisión: cuatro cosmogramas precolombinos mesoamericanos. Retrieved from https://ddd.uab.cat/pub/tesis/2011/hdl_10803_97367/mmca1de1.pdf

Motamayor, J. C., Lachenaud, P., da Silva E Mota, J. W., Loor, R., Kuhn, D. N., Brown, J. S., & Schnell, R. J. (2008). Geographic and genetic population differentiation of the Amazonian chocolate tree (Theobroma cacao L). *PloS One*, *3*(10), 3311. https://doi.org/10.1371/journal.pone.0003311

Palamara P. (2018). Tracing and tracking with the blockchain. Retrieved from http://hdl.handle.net/10589/139387

Poelmans, E., & Swinnen J. (2016). A brief economic history of chocolate. In M.P. Squicciarini & J. Swinnen (Eds). *The economics of chocolate* (pp. 11-88). Oxford, UK: Oxford University press.

Qin, X. W., Lai, J. X., Tan, L. H., Hao, C. Y., Li, F. P., He, S. Z., & Song, Y. H. (2017). Characterization of volatile compounds in Criollo, Forastero, and Trinitario cocoa seeds (Theobroma cacao L.) in China. *International Journal of Food Properties*, *20*(10), 2261–2275. https://doi.org/10.1080/10942912.2016.1236270

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

Rohman, A. & Che Man, Y. B. (2012). Analysis of Pig Derivatives for Halal Authentication Studies. *Food Reviews International*, *28*(1), 97–112. https://doi.org/10.1080/87559129.2011.595862

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

Saltini, R., Akkerman, R., & Frosch, S. (2013). Optimizing chocolate production through traceability: A review of the influence of farming practices on cocoa bean quality. *Food Control*, *29*(1), 167–187. https://doi.org/10.1016/j.foodcont.2012.05.054

Saltini, R., & Akkerman, R. (2012). Testing improvements in the chocolate traceability system: Impact on product recalls and production efficiency. *Food Control*, *23*(1), 221–226. https://doi.org/10.1016/j.foodcont.2011.07.015

Shavez, M., Ahmad, S., Jan, K., & Bashir, K. (2017). Status , supply chain and processing of cocoa - A review. *Trends in Food Science & Technology*, *66*, 108–116. https://doi.org/10.1016/j.tifs.2017.06.007

Silvis, I. C. J., Luning, P. A., Klose, N., Jansen, M., & van Ruth, S. M. (2019). Similarities and differences of the volatile profiles of six spices explored by Proton Transfer Reaction Mass Spectrometry. *Food Chemistry*, *271*(July 2018), 318–327. https://doi.org/10.1016/j.foodchem.2018.07.021

Sukha, D. a., Butler, D. R., Umaharan, P., & Boult, E. (2007). The use of an optimised organoleptic assessment protocol to describe and quantify different flavour attributes of cocoa liquors made from Ghana and Trinitario beans. *European Food Research and Technology*, *226*(3), 405–413. https://doi.org/10.1007/s00217-006-0551-2

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

Tres, A., & van Ruth, S. M. (2011). Verification of organic feed identity by fatty acid fingerprinting. *Journal of Agricultural and Food Chemistry*, *59*(16), 8816–8821. https://doi.org/10.1021/jf2016682

Tony's Chocolonely (2018). Where blockchain and slave free chocolate come together. Retrieved from https://medium.com/@tonyschocolonely/where-blockchain-and-slave-free-chocolate-come-together-d637b14b1ef0

van Ruth, S. M., & de Visser, R. (2015). Provenancing Flower Bulbs by Analytical Fingerprinting: Convallaria Majalis. *Agriculture*, *5*(1), 17–29. https://doi.org/10.3390/agriculture5010017

van Ruth, S. M., Silvis, I. C. J., Alewijn, M., Liu, N., Jansen, M., & Luning, P. A. (2019). No more nutmegging with nutmeg: Analytical fingerprints for distinction of quality from low-grade nutmeg products. *Food Control*, *98*(October 2018), 439–448. https://doi.org/10.1016/j.foodcont.2018.12.005

Tonantzin Chocolate (2015). Cacao Is Sacred. Retrieved from https://tonantzinchocolate.com/cacao-is-sacred/

Vlaeminck, P., Vandoren, J., & Vranken, L. (2016). Consumers' willingness to pay for fair trade chocolate. In M.P. Squicciarini & J. Swinnen (Eds). *The economics of chocolate* (pp. 180-191). Oxford, UK: Oxford University press.

Wikimedia Commons (n.d.). Autor unknown. The Codex Fejervary-Mayer, 15th century. Retrieved from https://commons.wikimedia.org/wiki/File:Xiuhtecuhtli_1.jpg

Windhab, E.J. (2009). Tempering. In Beckett, S.T. (Ed.), *Industrial Chocolate Manufacture and Use: Fourth Edition* (pp 226-319). York, UK: Blackwell Publishing Ltd

Wognum, P. M., Bremmers, H., Trienekens, J. H., Van Der Vorst, J. G. A. J., & Bloemhof, J. M. (2011). Systems for sustainability and transparency of food supply chains - Current status and challenges. *Advanced Engineering Informatics*, *25*(1), 65–76. https://doi.org/10.1016/j.aei.2010.06.001

Woolfe, M., & Primrose, S. (2004). Food forensics: using DNA technology to combat misdescription and fraud. *TRENDS in Biotechnology, 22*(5), 222-226. https://doi.org/10.1016/j.tibtech.2004.03.010

World Cocoa Foundation Report: 13. (2014). Retrieved from http://worldcocoafoundation.org/about-cocoa/cocoa-market-statistics/

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

Ziegleder, G. (2009). Flavour development in cocoa and chocolate. *Industrial Chocolate Manufacture and Use: Fourth Edition.Stephen T. Beckett*, pp. 169–191.

Chapter 2

Which cocoa bean traits persist when eating chocolate? Real-time nosespace analysis by PTR-QiToF-MS

Abstract

More consumers have become aware of the existence of different cocoa genotypes and their origins, which resulted in a growing market of premium chocolates with single-origin beans. The question is whether traits of cocoa botanical and geographical origins persist in the end product, especially when it is consumed. By analysing the concentrations of volatile organic compounds (VOCs) in the nose of subjects over time while they are eating, new insights about aroma release can be gained. In the current study, in vivo release of VOCs during consumption of dark chocolates with different botanical and geographical origins was examined. Proton Transfer Reaction-Quadrupole interface Time of Flight-Mass Spectrometry (PTR-Qi ToF-MS) was applied to analyse nosespace VOC profiles of 10 subjects while they were eating 10 different chocolates manufactured with beans of different botanical origins (Criollo-Forastero-Trinitario) and geographical origins (Africa-South America-Asia). The headspace of the chocolates was also analysed for comparison. Cocoa botanical information appeared to affect the nosespace profiles more than geographical information. The subjects varied considerably in their VOC release, and inter-individual differences were larger than cocoa beans differences. Nevertheless, the botanical origin was consistently reflected in the nosespace profile during eating. It was possible to distinguish Criollo chocolates from the nosespace profiles despite inter-individual differences.

Keywords Chocolate, Cocoa beans origin, Nosespace, PTR-QiToF-MS

2.1 Introduction

To be called "chocolate", a product must contain cocoa (Beckett, 2008). This statement is relevant both for the quality and sensory perception of chocolate. The complex flavour of cocoa beans depends on genotype/origin/processing and is generated by several processes, such as fermentation, drying, and roasting. After all processing steps, the beans are mixed with other ingredients to produce chocolate with its peculiar flavour. This raises the question about which of the flavour characteristics of the beans persist in the chocolate. Answering this question is important to assure chocolate quality and consumer satisfaction.

More consumers become aware of the existence of different cocoa genotypes and their origins. This is linked to the growing market of premium chocolates and the trend of many companies to launch new chocolates with single-origin bean, organic and fairtrade chocolate (Afoakwa, 2010). For this reason, the analytical verification of which raw material characteristics such as botanical and geographical origin persist in the final product is significant in ensuring a high-quality product.

Several studies focused on the volatile profiles of different cocoa genotypes (Frauendorfer & Schieberle, 2008) and different origins (Afoakwa, Paterson, Fowler, & Ryan, 2008). The volatile profiles of cocoa products such as cocoa liquor (Counet, Ouwerx, Rosoux, & Collin, 2004) and chocolate (Jinap, Dimick & Hollender, 1995; Cambrai et al., 2010) were also evaluated according to the provenances of the raw materials. However, there are no studies focused on the persistence of cocoa beans botanical and/ or geographical origin flavour in the aroma released when eating chocolates; only a study optimised an organoleptic assessment protocol to describe and quantify different flavour attributes of cocoa liquors made from Ghana and Trinitario beans (Sukha, Butler, Umaharan, & Boult, 2007). For this analysis, a panel test was involved. An alternative approach is to use analytical techniques based on real-time measurements, which can be applied for the determination of volatile organic compounds (VOCs) released during food consumption. By analysing the concentration of VOCs over time, new insights about aroma release can be gained. Proton Transfer Reaction Mass Spectrometry (PTR-MS) has been used to quantify in-nose and in-mouth volatile flavour release from different foods. Several food products were tested such as custard (Aprea, Biasioli, Gasperi, Märk, & van Ruth, 2006), apple (Ting et al., 2012) and cereal bars (Heenan et al., 2012). Moreover, milk and aqueous sugar solution (Szymańska et al., 2015), flavoured vodka (Déléris et al., 2011), lipid solutions (Frank, Appelqvist, Piyasiri, Wooster, & Delahunty, 2011), coffee (Charles et al., 2015; Romano et al., 2014), and aromatized air (Sánchez-López, Ziere, Martins, Zimmermann, & Yeretzian, 2016) were investigated. There is no literature about in-nose and/or in-mouth volatile flavour released from chocolate.

This study aimed to assess the influence of the botanical and/or geographical origin of cocoa beans on the volatiles compounds released in-nose when eating dark chocolates. Ten identical chocolates manufactured with beans from different botanical and geographical origins were evaluated and the nosespace (NS) profiles - when eating those chocolates - were measured using PTR-Quadrupole interface Time of Flight (Qi ToF)-MS.

2.2 Materials and subjects

2.2.1 Samples

Ten dark chocolates (75% cocoa) of a single manufacturer were analysed. The chocolate samples were manufactured with beans belonging to three different varieties (Criollo, Forastero, Trinitario) and from ten different single origins (Ghana, Madagascar, Sao Tome, Tanzania, Cuba, Brazil, Trinidad, Venezuela, Papua New Guinea and Java). Regarding the botanical origin the samples were grouped in Criollo (Madagascar, Java), Forastero (Brazil, Ghana, Sao Tome, Tanzania) and Trinitario chocolates (Cuba, Papua New Guinea, Trinidad, Venezuela).

While for the geographical origin the chocolates were split into three clusters according to the continent of provenance Africa (Ghana, Madagascar, Sao Tome, Tanzania), South America (Cuba, Brazil, Trinidad, Venezuela) and Asia (Papua New Guinea, Java). The set of samples was used to carry out both headspace and nosespace measurements.

2.2.2 Subjects

The ten samples were presented to a panel of ten subjects (five females and five males, age 25–55 yrs.). Subjects were volunteers recruited from Wageningen University and Research and had no self-reported sensory issues. They were asked not to eat, drink or use persistent products at least 2 h before the session. Subjects attended six sessions on six different days. Each session lasted for about 25 min: 90 s of measurement followed by 5 min break before the next measurement. For each session, a maximum of four chocolates was tested.

2.3 Methods

2.3.1 PTR-QiTOF-MS Instrumental conditions and raw data processing

All measurements were performed using a PTR-QiToF-MS (Ionicon Analytik G.m.b.H., Innsbruck, Austria) to investigate the VOCs. Ionisation was carried out under drift tube voltage 999 ± 1 V, drift pressure of 3.80 mbar, temperature 60 °C, and E/N value of 133 Td. Data acquisition was carried out at 1 spectrum per second. For each sample, a mass range between 0.00 and 512.15 Th was measured using a dwell time of 0.1 s mass^{-1}. Sampling was performed at a flow rate of 60 ± 1 sccm for headspace and nosespace analysis.

2.3.1.1 Data processing

Mass scale calibration and peak extraction were performed using PTRwid (Holzinger, 2015).

2.3.2 Headspace analysis

Chocolate samples were powdered using an electrical grater and were kept at 4 °C before analysis. For the measurements, 3 g of ground chocolate were weighed into clean and odourless flasks of 250 ml. The closed flasks were placed in a water bath at 30 °C for 30 min to equilibrate samples with their headspace. For all samples, four repeated measurements were carried out. Data acquisition was performed over 60 spectra resulting in an analysis time of 60 s per sample. A blank (empty flask) was analysed before each sample. So eventually 10 samples x 4 replicates x 60 spectra = 2400 spectra were acquired. The first 30 scans per sample were averaged because the intensity of the signal was constant. The values obtained for the blank were subtracted from each sample.

2.3.2.1 Data treatment and analysis

A concentration threshold of 1 pbbv was set for the data set, and the mass peaks with lower concentrations were filtered out. Peaks attributed to 13C isotopologues and water clusters were considered to be redundant and therefore discarded.

The final data set was analysed using two-way ANOVA to understand whether there was an interaction between botanical and geographical origin of chocolate on the VOCs concentration. R 3.3.3 (the R Foundation for Statistical Computing, Vienna, Austria) was used to perform the analysis.

2.3.3 Nosespace analysis

The ten chocolates were presented in independent duplicates to the ten subjects. The pieces of chocolate tested had the same dimensions (2 cm × 2 cm) and weight (3 g). The order of chocolates and subjects was fully randomized. A maximum of four chocolates per day was given to the subjects. Sampling was carried out by applying an ergonomic nosepiece in glass to the nose of the subjects. The glass nosepiece was connected to the PTR-QiToF-MS by a PEEK tube (inner diameter = 0.5 mm); the nosepiece was changed for every test.

After positioning the nosepiece in the nostrils, subjects were asked to breathe normally through the nose (mouth closed) for 10 s. Then subjects received the sample. As soon as subjects put the sample in the mouth, they were allowed to masticate and swallow freely. The mastication period and the first swallowing were recorded for each session. During all evaluations, subjects had to keep their mouth closed. In total, the NS measurement lasted 90 s. Sampling was carried out in duplicate. So, eventually 10 chocolates x 10 subjects x 2 replicate measurements x 90 spectra = 18,000 spectra were acquired.

2.3.4 Data treatment

For each compound 200 release curves, corresponding to the NS sessions were acquired. Within each release curve, three-time windows were selected: breathing session (1–10 s), mastication session (11 s to first swallowing point) and post-swallowing session (first swallowing point plus 20 s). For each window, the part of the curve selected was superposed and averaged to generate an average release curve. Differences between mastication and swallowing were expected as has been shown in a previous study by Déléris et al. (2011). For this reason, they were analysed as a separate set.

According to Romano et al. (2014), in order to select mastication and post swallowing relevant peaks, the breathing part was compared with the mastication and post-swallowing part by means of a Mann–Whitney test and p-values were corrected using the Rate of False Discovery (RDF) (Romano et al., 2014; Benjamini & Hochberg, 1995). Mass peaks and corresponding release curves associated with a corrected p-value lower than 0.05 were retained for further analysis. After this selection step, peaks attributed to 13C isotopologues and water clusters were considered to be redundant and therefore discarded. From each selected peak, the baseline (obtained by averaging the first 10 cycles) was subtracted. A non-parametric two-way Friedman test comparing each compound released curve was used to investigate the interaction of botanical and geographical origin on the NS mastication and post-swallowing profiles.

For each selected peak five descriptors for mastication and post-swallowing were calculated: time-independent parameters, i.e. the maximum intensity (max), the area under the curve (area), and the average of the release curve (avg), as well as time-dependent parameters such as the time to reach the maximum (tmax), and the slope of the first descending section of the curve (slope) (Normand, Avison, & Parker, 2004; Frasnelli, van Ruth, Kriukova, & Hummel, 2005).

Non-parametric Kruskall Wallis tests were applied to the time-independent (max, area, avg) and time-dependent parameters (tmax, slope), to investigate statistical differences between the NS profile released during mastication and post-swallowing of the dark chocolates. P-values were corrected using the Rate of False Discovery (RDF). For analysing the specific sample pairs for stochastic dominance in post hoc testing, pairwise Mann-Whitney tests were performed.

Non-parametric tests were preferred for the NS session analysis because when using human data there is a higher probability of not satisfying the ANOVA assumptions. Shapiro Wilk tests were conducted to examine normality within the groups. Data analysis and statistical analysis were performed using IBM SPSS Statistics 23.0 (IBM Corp., Armonk, NY, USA) for one-way non-parametric tests, R 3.3.3 (the R Foundation for Statistical Computing, Vienna, Austria) for mastication and post-swallowing descriptors extraction and non-parametric tests- Friedman Two-Way ANOVA, and Pirouette 4.0 Software (Infometrix, Seattle, WA, USA) for multivariate analysis (Principal Component Analysis, PCA).

2.4 Results and discussion

2.4.1 Headspace analysis

A total of 695 mass peaks were measured in the range 19.081 –505.086 m/z. After setting the 1 ppb threshold and having discarded 20 peaks related to 13C isotopologues and water clusters, 205 peaks were retained.

The sum peaks concentration measured in the chocolate headspace was higher in Forastero samples (104.358 ± 32.695 ppbv) immediately followed by the Trinitario chocolates (102.391 ± 48.571 ppbv). Criollo showed the lowest headspace concentration (97.355 ± 67.651 ppbv). High concentrations for Forastero samples were found also in previous studies (Counet et al., 2004). Regarding the geographical origin, the South American (112.068 ± 45.860 ppbv) and African (111.490 ± 45.860 ppbv) chocolates scored higher values compared to the Asian samples (63.737 ± 20.109 ppbv). General prime ions with m/z 33.033, 43.017, and 61.028 tentatively identified as methanol, a fragment of diverse origin, and acetic acid, respectively, were observed in all samples (Table S2.1, Supplementary materials). These masses were considered predominant also in a previous study focused on the volatile compositions of chocolates using PTR-MS (Acierno, Yener, Alewijn, Biasioli, & van Ruth, 2016). Mass 61.028 did not show significant differences related to botanical or geographical origin. However, its concentration was highest in Criollo samples, which underpins the acidic tones of Criollo beans (Acierno et al., 2016; Saltini, Akkerman, & Frosch, 2013; Holm, Astong, & Douglas, 1993).

From our previous studies (Acierno et al., 2016), an interaction between geographical origin, botanical origin and brand are expected. To investigate the interaction between botanical and geographical origin, the dataset was submitted to two-way ANOVA tests (p < 0.05).

A subset of 56 significantly different peaks for geographical origin (Africa, Asia, South America) and 20 for botanical origin was obtained. Eleven masses were affected by an interaction of both parameters. Thus, a higher number of mass peaks were significant in differentiating the samples according to the geographical origin compared to the botanical one. A sum formula was applied to 48 of those masses, and 30 were tentatively identified based on literature knowledge of their occurrence in chocolate (Afoakwa et al., 2008; Acierno et al., 2016) (Table S2.1, Supplementary materials).

2.4.2 Nosespace analysis

2.4.2.1 PTR-ToF-MS spectra
The nosespace measurement generated 695 mass peaks over a total of 18,000 mass spectra, distributed over 200 sessions overall (90 spectra per session). Figure 2.1 illustrates a typical NS release curve; the curve of m/z 87.080, methylbutanal, an important contributor to the chocolate aroma (Frauendorfer & Schieberle, 2008). This plot supports the decision of analysing mastication and post-swallowing separately. Differences exist between intensities according to the mastication and post-swallowing phases. This trend can be supported by previous studies that demonstrated that the flavour release changes according to different phenomena occurring in the mouth such as salivation, mastication, swallowing (Buettner & Beauchamp, 2010).

2.4.2.2 Impact of cocoa bean traits in general
The NS profiles resulted in 311 mass peaks for the mastication and 441 for the post-swallowing part. The data were not normally distributed (Shapiro-Wilk test < 0.05) and the two-way non-parametric Friedman-tests were applied. The results showed that both in mastication and post-swallowing a higher number of masses showed significant differences for botanical origin compared to the geographical origin (Table S2.2 and S2.3, Supplementary materials). Taking into account this information, attention was focused on the botanical information of cocoa beans in the NS profile.

2.4.3 Influence of cocoa beans' botanical origin on NS profiles

2.4.3.1 Univariate characterisation
The botanical origin of the beans in the chocolates affected the VOCs concentration of NS profiles during mastication and post-swallowing for the time-independent parameters (area, avg, max) (Kruskall Wallis test and Mann-Whitney post-hoc test). The data were not normally distributed (Shapiro-Wilk test < 0.05) and significance of differences between the samples was evaluated by non-parametric tests (Kruskall Wallis test and Mann-Whitney post-hoc test). Significant differences (p < 0.05) were found for 30 mass peaks for the mastication phase and 39 mass peaks for the post-swallowing phase.

With help of data reported in literature (Frauendorfer & Schieberle, 2008; Afoakwa et al., 2008; Acierno et al., 2016; Diab, Hertz-Schünemann, Streibel, & Zimmermann, 2014; Yener et al., 2014; Nolvachai et al., 2015; Yener et al., 2016; Özdestan et al., 2013) a tentative identification was proposed for all the significant mass peaks (46 tentative identifications). These masses can be linked to chemical classes known to be responsible for the typical chocolate aroma, including Maillard reaction products such as aldehydes and pyrazine and other compounds as pyrroles, furans, pyridines, oxazoles (Afoakwa et al., 2008; Romano et al., 2014).

Table S2.4 (Supplementary materials) reports all mass peaks and parameters showing significant differences according to the botanical origins of the cocoa beans used for the manufacture of the chocolates.

Differences were found for parameters related to the area under the curve (area, occurring 24 times for mastication and 34 for post-swallowing), maximum intensity (max, 5 times for mastication and 19 for post-swallowing), mean intensity (avg, 8 times for mastication and 12 for post-swallowing). The influence of the time-independent parameters has been shown also in a previous study about coffee NS measurements (Charles et al., 2015). On the other hand, time-related parameters (i.e. slope, tmax) did not show significant differences; also in coffee NS were scarcely represented (Charles et al., 2015).

As expected from Figure 2.1, the aforementioned parameters were always higher in the post-swallowing session; the majority of the significant differences were found during the post-swallowing measurement as well (Table S2.4, Supplementary materials). An increase of the volatile released during the post-swallowing phase, when eating a fat matrix, was also found in another study measuring the flavour released from a lipid emulsion (Frank et al., 2011).

2.4.3.1.1 Mastication phase

During mastication, the main NS profile differences are observed for the chocolates made from Criollo beans ('Criollo chocolates') versus the rest of the chocolates, and for chocolates produced from Forastero beans ('Forastero chocolates') versus those with Trinitario beans ('Trinitario chocolates'). Only m/z 93.067, tentatively identified as toluene/ terpene fragment was able to differentiate the three botanical origins at the same time (area). About two-third of the compounds (Table S2.4, Supplementary materials) exhibit statistically different concentrations for 'Forastero chocolates' compared to 'Trinitario chocolates', ~20% show significant differences between 'Criollo chocolates' and the rest of the samples, and ca. 10% distinguish 'Trinitario chocolates' from the other samples.

Lower mass (lighter) compounds tend to characterise the differences between 'Criollo chocolates' from the others whereas those distinguishing 'Forastero and Trinitario chocolates' tend to be heavier. The group of compounds showing significant differences for the NS of 'Criollo chocolates' during mastication comprises mainly esters and acids. This was expected as according to the literature Criollo beans are characterised by acidic tones (Saltini et al., 2013). On the other hand, within the group of compounds relevant for the NS of 'Forastero chocolates', there are aldehydes, alcohols, furans, ketones, and pyrroles. The higher concentrations of these aroma compounds in the NS profiles of 'Forastero chocolates' can be attributed to different conditions of these beans fermentation compared to the Criollo beans. Usually, Forastero beans are more fermented; this maximizes the release of flavour precursors which subsequently increases the aroma compounds concentrations in the final product (Afoakwa, 2008; Counet et al., 2004; Holm et al., 1993) and consequently in the NS profile.

The mean values for max, avg and area (Table S2.4, Supplementary materials) indicate that Forastero chocolate showed generally higher concentrations compared to the other samples. The high concentrations of VOCs of 'Forastero chocolates' in the NS are in alignment with the high VOC concentrations in the headspace of the same chocolates.
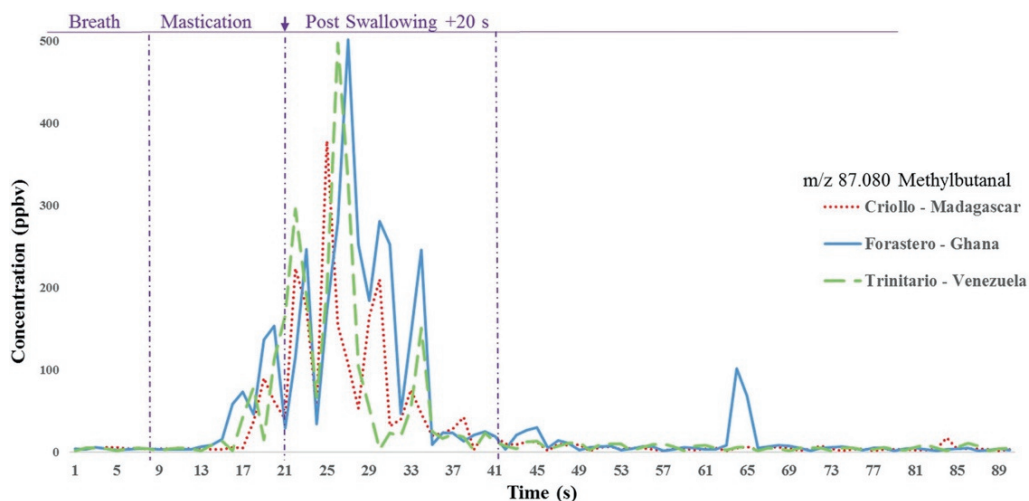
Figure 2.1. Release curve of mass 87.080, related to a single person when eating 3 different chocolates (swallowing points were aligned at 24 s for curves visualisation).

2.4.3.1.2 Post-swallowing phase

During the post-swallowing session, the situation changes. More compounds show significant botanical differences and the number of compounds exhibiting significant differences between 'Criollo chocolates' and the others increases. About 44% of the compounds show significant differences between 'Criollo chocolates' and the other chocolates. The number of mass peaks significantly differentiating between 'Forastero and Trinitario chocolates' is decreased during the post-swallowing phase to approximately 15%. Some of the peaks which distinguished 'Forastero chocolates' during the mastication phase now become in the post-swallowing phase significant in differentiating 'Trinitario chocolates' from the other samples (~25%) or 'Criollo chocolates' from the rest. During swallowing also the number of mass peaks differentiating 'Criollo chocolates' from 'Trinitario chocolate' increase (~15%).

This change in the relevance of compounds can be related to the different way in which compounds are retained in the chocolates. For instance, a different structure of the chocolate matrix, such as cocoa butter hardness and lipids composition, depending on the cocoa beans provenance, is expected (Chaiseri & Dimick, 1989); this could influence the release of the compounds.

According to Mann Whitney results, m/z 112.077 (Trimethyl-ox- azole/2-acetyl-1-pyrroline), 123.092 (Trimethylpyrazine/2-ethyl-3-methylpyrazine) and 131.071 (Acetyloxy-butanone/ethanrdioldiacetate/oxopropoxy-propanone/ethyl-oxobutanoate) differentiated chocolate from the three different botanical origins simultaneously. Mass 112.077 and 123.092 are related to oxazoles and pyrazine. These compounds are generated during the cocoa beans roasting through Maillard browning (Jinap, Jamilah, & Nazamid, 2004) and they can be an indication of different roasting effects as a result of the cocoa beans botanical origins.

During post-swallowing, the majority of the compounds show high concentrations for the 'Criollo chocolates'. Some of them do not so in the mastication phase (Table S2.4, Supplementary materials). They comprise ester/acids, oxazoles and pyrazines. The latter is known to be a more persistent group of compounds (Romano et al., 2014) and most probably for this reason more relevant during the post-swallowing phase.

Some of the other compounds showing significant differences due to origin during the mastication phase are indeed not showing these differences post-swallowing. They include for example alcohols, carbonyls, aldehydes, pyridines and acids. For instance, it is expected that aldehydes are released faster as they interact less with the mucosa (Charles et al., 2015); this could affect their influence on the NS during swallowing.

These results emphasise that difference in the botanical origin of cocoa beans used to produce the chocolate influences the NS profile released by different subjects when eating chocolates. Especially chocolates manufactured from Criollo cocoa beans stand out in the post-swallowing phase.

2.4.3.2 Multivariate characterisation

The area values of the peaks measured during the post-swallowing phase of the NS analyses were subjected to PCA; the results of the mean-centred and normalized data are displayed in Figure 2.2. The PCA highlights a cluster of NS area data of 'Criollo chocolates'. Areas related to 'Forastero and Trinitario chocolates' are not well distinguished, supporting the univariate trend discussed in the previous paragraphs. PCAs based on area, average and maximum values, both for mastication and post swallowing phases, clearly highlighted differences in VOCs of NS profiles released when eating chocolate produced with Criollo (results not shown). A geographical trend in the distribution of chocolates is visible when analysing the cluster related to the botanical origin separately. Specifically, for Criollo and Trinitario chocolate, it is possible to notice the NS of chocolate made with beans from the same single-origin are close together (Figure S2.1 and S2.2, Supplementary materials).

2.4.4 Influence of subject on NS profiles

Evaluation of the plot of the PCA of the area values (Fig. 2.2) reveals subject clusters in the part of the plot corresponding to the mixed area of 'Forastero and Trinitario chocolates'. This indicates that for these samples consistent inter-individual patterns exist. The consistent patterns for subjects are probably related to other inter-individual parameters such as genetic factors, and physiological factors, or naso- or opharyngeal volumes, saliva composition, mucus composition, and breath flow (Charles et al., 2015; Sánchez-López et al., 2016).

To better understand the influence of each subject, Figure 2.3 shows the release profiles of each participant, expressed as mean integrated release cumulated for all masses for the mastication and post-swallowing phases (area) for each cocoa botanical origin. Looking at the barplot (Fig. 2.3), it is possible to notice subjects characterised by low (S_1-S_5) and high (S_6-S_10) NS post-swallowing release.
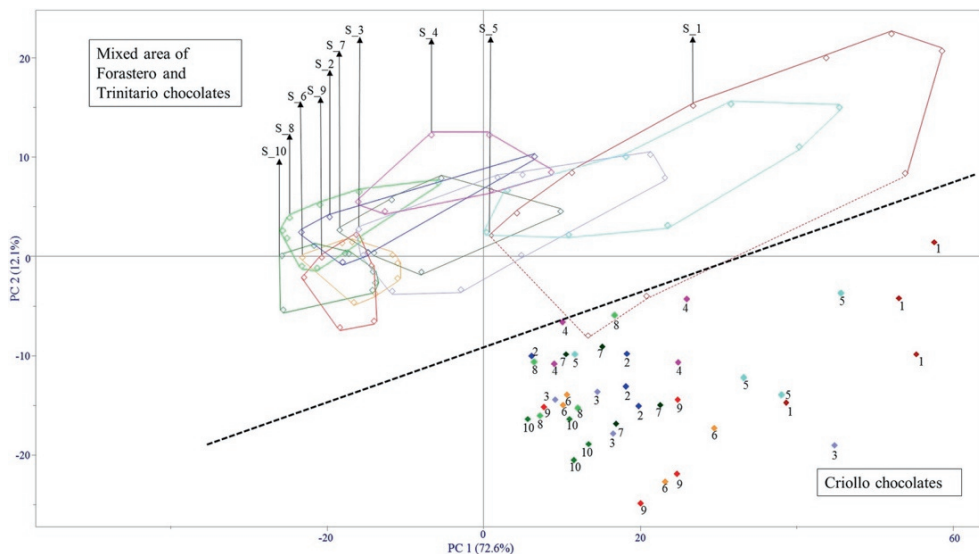
Figure 2.2. PCA scores plot of the mean-centred and normalized NS (area) data of the post-swallowing phase collected from 10 subjects when eating 10 different chocolates (replicates are shown). Plot showing Criollo chocolate clusters in the lower part and clusters of subjects in the upper part. The lines connect the external points of the clusters related to each subject from S_1 to S_10. S: Subject. The chocolates highlighted are produced from Criollo beans while the numbers from 1 to 10 indicate the different subjects. Empty symbols are related to Forastero and Trinitario chocolates.

The PCA in Figure 2.2 highlights that the distribution of the subjects is most probably concentration-related. Subjects characterised by higher cumulative area (S_6, _8, _9, 10) overlapped in the left-hand side of the plot. Subject 7, showing a high cumulative area, occupies a larger area in the plot compared to the other high releasers. Controversial is the trend for the low releasers; Subject 2 overlaps with the high releaser occupying a relatively small area compared to Subject 3 and 4. Subject 1, and_5 follow a different distribution compared to the rest. Apparently, the people characterised by a low cumulative area show more variation and a larger spreading of the chocolates compared to the subjects with a higher cumulative area (Fig. 2.2).

2.4.5 Interactions between traits of cocoa beans and subjects

Cumulative release shows a factor 3 difference between S1 and S10 (Fig. 2.3), but there appears also a sample effect. Larger differences are observed between 'Criollo chocolates' and others after swallowing; this trend is valid for almost each subject. The magnitude of difference between Criollo chocolate and the others is approximately a factor 2.

This confirms that the subject effect is probably larger than the bean influence on NS profile release when eating chocolate, emphasizing the distribution of subjects over the first dimension, and the separation of 'Criollo chocolates' and others in the second dimension in the PCA (Fig. 2.2). Moreover, analysing PCA related to each subject, it is possible to notice that for certain subjects there is a good separation within the three chocolate samples. This is more evident for subject 1 and 5 (Figures S2.3 and S2.4, Supplementary materials). Probably for this reason, they show a different distribution compared to the rest of the participants (Fig. 2.2). Despite the influence of the subjects, the chocolate-cluster in Figure 2.2 and the plot in Figure 2.3 can still be distinguished. In fact, it is important to highlight the consistent differences between NS profile when eating chocolate made with Criollo bean and the rest.
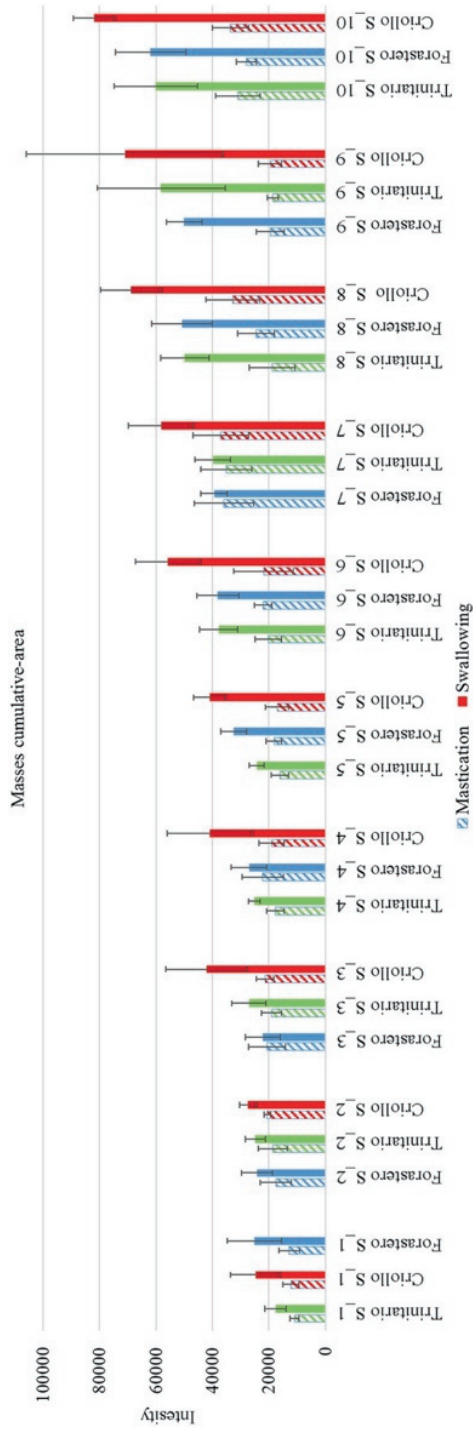
Figure 2.3. Plot showing the cumulative release profiles of each subject (S), expressed as an average of the cumulative areas of NS chocolates belonging to the same botanical cluster (Criollo, Forastero, Trinitario). The cumulative area includes all the masses. Mastication and post-swallowing comparison are shown. The chocolate areas are ordered according to the highest intensity per subject.

41
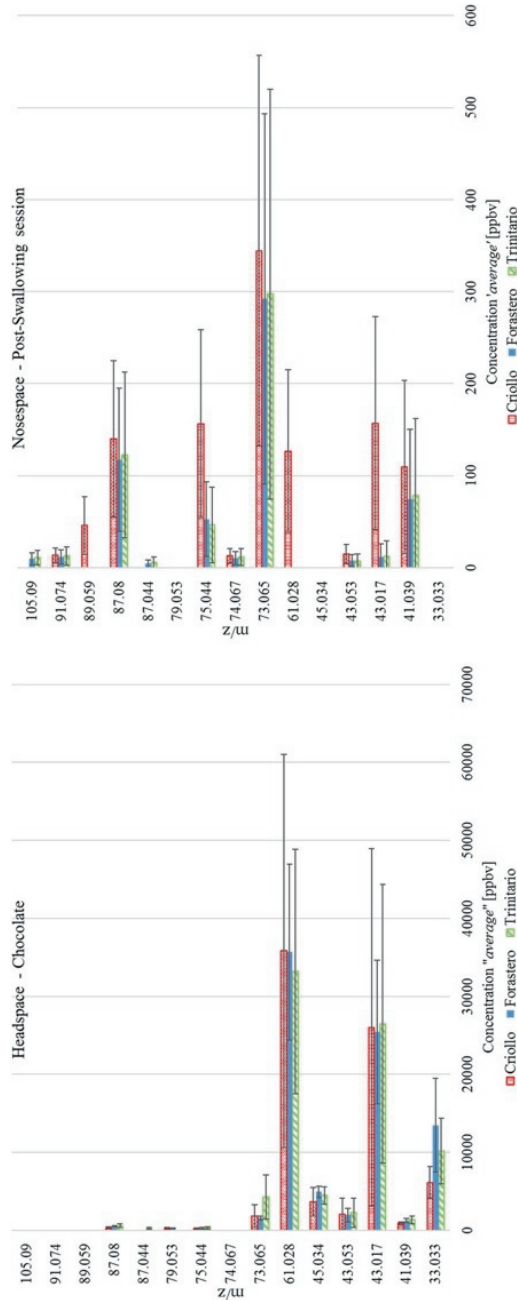
Figure 2.4. Botanical origin comparison based on the ten highest compounds of the average concentration of all headspace and NS measurements. Left figure: Headspace 10 highest compounds for Criollo, Foratsero, Trinitario; Right figure: NS 10 highest compounds for Criollo, Forastero, Trinitario.

2.4.6 Linking headspace and nosespace analysis

In the headspace of the chocolates, the geographical information was more dominant than the botanical traits (Section 2.4.1), whereas the order was reversed for the NS analyses (Section 2.4.2.2). The headspace concentrations of the VOCs showed highest total intensity for 'Forastero chocolates' (Section 2.4.1), even though some of the major compounds were present in higher concentrations in the headspace of the 'Criollo chocolates' (Fig. 2.4). The latter is reversed in the NS measurement (Fig. 2.4). This high release is in accordance with a previous study in which Criollo chocolate measured with PTR-MS showed higher concentration compared to the other botanical origins (Acierno et al., 2016). Dissimilarities in flavour delivery will depend on the affinity of the compounds for the food matrix. As mentioned before, considering the structure of the chocolate matrix according to the cocoa beans used, differences between the samples are expected (Chaiseri & Dimick, 1989); this explains the variation in the NS concentration. In a multiphase system, such as chocolate, aroma compounds are partitioned in the fat phase that can influence the release of the volatile compounds according to their lipophilicity. For these reasons, differences in the cocoa butter hardness and lipids composition can cause differences in the release of VOCs, especially when the matrix is changing due to salivation and mastication. Looking at Figure 2.4, comparing the 10 highest compounds in the headspace and NS of each botanical origin, it is possible to notice that for chocolate the VOCs release is almost homogeneous between the different botanical samples while in the NS the differences increase.

Figure 2.4 underlines the VOCs profile differences between headspace and NS. Whereas for the chocolates' headspace mass 61.028 is the one highest in concentration, in the NS mass 73.065 (methylpropanal) is predominant for all the chocolates together with mass 87.080 (methlbutanal), although overall concentrations are much lower due to dilution by the breath airflow. These compounds are typical contributors to the chocolate flavour. Masses 74.067, 91.074, 105.09 (methional) were not included in the chocolates' predominant compounds. Within them, mass 91.074 was found in a previous work of the authors of this study as an important compound for the discrimination of Trinitario chocolate (Acierno et al., 2016). Moreover, in the aforementioned study mass 45.053 (acetaldehyde) and 105.09 (methional) were relevant in the discrimination of the chocolates belonging to the brand analysed in the NS measurements.

Evaluating the trend of the compounds, in the headspace is clear the prevalence of few compounds on the other hand in the NS measurements even though the compounds are more diluted, the predominance of certain compounds is less evident and heavier compounds are becoming more dominant during eating. They are probably more hydrophobic and are being pushed out from the matrix that is becoming more and more hydrophilic with continued salivation. When air/product distribution coefficients change as the 'product' phase becomes very hydrophilic with the saliva, they tend to accumulate and increase their concentrations in air. Mastication and swallowing cause significant changes to the food in terms of surface area, hydration and time in the mouth and these factors, in turn, affect the mass transfer phase in the mouth (van Ruth & Roozen, 2000).

2.5 Conclusions

PTR-QiToF-MS was shown to allow rapid real-time screening of the cocoa beans characteristics that persist in the final product before and during consumption. Distinct profiles of the VOCs released during the eating of chocolates of different origins (botanical and geographical) during mastication and post-swallowing were defined and show consistent differences due to inter-individual subject and beans' botanical origin factors as well as product/subject interactions. The inter-individual differences were larger than cocoa beans differences, and cocoa botanical traits were more pronounced in the profiles than geographical traits. On the other hand, in the chocolate headspace profiles, the geographical information was more dominant than the botanical traits; this dissimilarity is mainly due to the affinity of the compounds for the food matrix and the air/product distribution changes caused during the mastication and swallowing.

This study underlines that the botanical origin is consistently reflected in the final chocolates and even in the breath of subjects during eating. Criollo chocolates were clearly distinguished from the other samples. Geographical differences were detectable within the three botanical origin groups. Further studies could focus on differences in VOC release due to cocoa bean origin and sensory properties to understand how sensory perception is eventually affected.

Acknowledgements

References

Acierno, V., Yener, S., Alewijn, M., Biasioli, F., van Ruth, S. M. (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,* https://doi.org/10.1016/j.foodres.2016.03.022.

Afoakwa, E.O. (2010). Chocolate production and consumption patterns, *Chocolate Science and Technology,* 1–11, https://doi.org/10.1002/9781444319880.ch1.

Afoakwa, E.O., Paterson, A., Fowler, M., Ryan, A. (2008). Flavor formation and character in cocoa and chocolate: a critical review, *Critical reviews in food science and nutrition*, *48*, 840–857, https://doi.org/10.1080/10408390701719272.

Aprea, E., Biasioli, F., Gasperi, F., Märk, T.D., van Ruth, S. M. (2006). In vivo monitoring of strawberry flavour release from model custards: effect of texture and oral processing, *Flavour and Fragrance Journal, 21,* 53–58, https://doi.org/10.1002/ffj.1702.

Beckett, S.T., Chocolate ingridiente, in: S.T. Beckett (Ed.) (2008). The Science of Chocolate, The Royal Society of Chemistry, Cambridge UK, pp. 11–38, https://doi.org/ 10.1089/jam.2007.0571.

Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal statistical society: series B (Methodological), 57*(1)*,* 289–300, https://doi.org/10.2307/2346101.

Buettner, A., Beauchamp, J. (2010). Chemical input - sensory output: diverse modes of physiology-flavour interaction, *Food Quality and Preference. 21,* 915–924, https://doi. org/10.1016/j.foodqual.2010.01.008.

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*(3)*,* 1478–1483, https://doi.org/10. 1021/jf903471e.

Chaiseri, S., Dimick, P.S. (1989). Lipid and hardness characteristics of cocoa butters from different geographic regions. *Journal of the American Oil Chemists' Society, 66*(12), 1771–1776.

Charles, M., Romano, A., Yener, S., Barnabà, M., Navarini, L., Märk, T.D., Biasoli, F., Gasperi, F. (2015). Understanding flavour perception of espresso coffee by the combination of a dynamic sensory method and in-vivo nosespace analysis, *Food Research International, 69,* 9–20, https://doi.org/10.1016/j.foodres.2014.11.036.

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

Déléris, I., Saint-Eve, A., Guo, Y., Lieben, P., Cypriani, M.L., Jacquet, N., Brunerie, P., Souchon, I. (2011). Impact of swallowing on the dynamics of aroma release and perception during the consumption of alcoholic beverages, *Chemical Senses 36*(8)*,* 701–713, https://doi.org/10.1093/chemse/bjr038.

Diab, J., Hertz–Schünemann, R., Streibel, T., Zimmermann, R. (2014). Online measurement of volatile organic compounds released during roasting of cocoa beans, *Food Research International, 63,* 344–352, https://doi.org/10.1016/j.foodres.2014.04.047.

Frank, D., Appelqvist, I., Piyasiri, U., Wooster, T.J., Delahunty, C. (2011). Proton transfer reaction mass spectrometry and time intensity perceptual measurement of flavour release from lipid emulsions using trained human subjects, *Journal of agricultural and food chemistry, 59*(9)*,* 4891–4903, https://doi.org/10.1021/jf104376b.

Frasnelli, J., van Ruth, S. M., Kriukova, I., Hummel, T. (2005). Intranasal concentrations of orally administered flavors, *Chemical Senses 30,* 575–582, https://doi.org/10.1093/chemse/bji051.

Frauendorfer, F., Schieberle, P. (2008). Changes in key aroma compounds of Criollo cocoa beans during roasting, *Journal of agricultural and food chemistry, 56,* 10244–10251, https://doi.org/10.1021/jf802098f.

Heenan, S., Soukoulis, C., Silcock, P., Fabris, A., Aprea, E., Cappellin, L., Märk, T.D., Gasperi, F., Biasioli, F. (2012). PTR-TOF-MS monitoring of in vitro and in vivo flavour release in cereal bars with varying sugar composition, *Food Chemistry, 131,* 477–484, https://doi.org/10.1016/j.foodchem.2011.09.010.

Holm, C.S., Astong, J.W., Douglas, K. (1993). The effects of the organic acids in cocoa on the flavour of chocolate *, *Journal of the Science of Food and Agriculture, 61*(1), 65–71.

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

Jinap, S., Jamilah, B., Nazamid, S. (2004). Sensory properties of cocoa liquor as affected by polyphenol concentration and duration of roasting. *Food Quality and Preference, 15*, 403–409, https://doi.org/10.1016/S0950-3293(03)00097-1.

Jinap, S., Dimick, P.S., Hollender, R. (1995). Flavour evaluation of chocolate formulated from cocoa beans from different countries, *Food Control 6,* 105–110, https:// doi.org/10.1016/0956-7135(95)98914-M.

Nolvachai, Y., Kulsing, C., Boysen, R.I., Matyska, M.T., Pesek, J.J., Marriott, P.J., Hearn, M.T.W. (2015). Comparison of the performance of different silica hydride particles for the solid-phase extraction of non-volatile analytes from dark chocolate with analysis by gas chromatography-quadrupole mass spectrometry, *Food Chemistry, 174*, 434–439, https://doi.org/10.1016/j.foodchem.2014.10.083.

Normand, V., Avison, S., Parker, A. (2004). Modeling the kinetics of flavour release during drinking, *Chemical Senses, 29*, 235–245, https://doi.org/10.1093/chemse/ bjh028.

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

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

Saltini, R., Akkerman, R., Frosch, S. (2013). Optimizing chocolate production through traceability: a review of the influence of farming practices on cocoa bean quality, *Food Control. 29,* 167–187, https://doi.org/10.1016/j.foodcont.2012.05.054.

Sánchez-López, J.A., Ziere, A., Martins, S.I.F.S., Zimmermann, R., Yeretzian, C. (2016). Persistence of aroma volatiles in the oral and nasal cavities: real-time monitoring of decay rate in air exhaled through the nose and mouth, *Journal of breath research, 10,* https://doi.org/10.1088/1752-7155/10/3/036005.

Sukha, D. a., Butler, D.R., Umaharan, P., Boult, E. (2007). The use of an optimised organoleptic assessment protocol to describe and quantify different flavour attributes of cocoa liquors made from Ghana and Trinitario beans, *European Food Research and Technology, 226*(3), 405–413, https://doi.org/10.1007/s00217-006-0551-2.

Szymańska, E., Brown, P.A., Ziere, A., Martins, S., Batenburg, M., Harren, F.J.M, Buydens, L.M.C. (2015). Comprehensive data scientific procedure for enhanced analysis and interpretation of real-time breath measurements in In vivo aroma-release studies, *Analytical Chemistry, 87*, 10338–10345, https://doi.org/10.1021/acs.analchem.5b02420.

Ting, V.J.L., Soukoulis, C., Silcock, P., Cappellin, L., Romano, A., Aprea, E., Bremer, P.J., Märk, T.D., Gasperi, F., Biasioli, F. (2012). In vitro and in vivo flavor release from intact and fresh-cut apple in relation with genetic, textural, and physicochemical parameters, *Journal of Food Science, 77*, https://doi.org/10.1111/j.1750-3841.2012.02947.x.

van Ruth, S.M., Roozen, J.P. (2000). Influence of mastication and saliva on aroma release in a model mouth system, *Food Chemistry, 71*, 339–345, https://doi.org/10.1016/ S0308-8146(00)00186-2.

Yener, S., Navarini, L., Lonzarich, V., Cappellin, L., Märk, T.D., Bonn, G.K., Biasioli, F. (2016). Monitoring single coffee bean roasting by direct volatile compound analysis with proton transfer reaction time-of-flight mass spectrometry, *Journal of Mass Spectrometry, 51*(9), 690–697, https://doi.org/10.1002/jms.3825.

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

Supplementary materials

Table S2.1. Masses showing significant average concentration in the headspace of chocolate according to Geographical origin (Africa, Asia, South America), Botanical origin (Criollo, Forastero, Trinitario), interaction Geographical * Botanical origin (two-way ANOVA, $p < 0.05$).

m/z	Sum formula	Tentative identification	Botanical origin [b]	Geographical origin	Geographical * Botanical origin
45.034	$C_2H_4OH^+$	Acetaldehyde		x	
55.053	$C_4H_6H^+$	N.I. [a]		x	
57.033	$C_3H_4OH^+$	Acrolein		x	
73.065	$C_4H_8OH^+$	2-methylpropanal		x	
77.059	$C_3H_8O_2H^+$	N.I.	x		x
83.049	$C_5H_6OH^+$	Methyl-furan	x		
83.085	$C_6H_{10}H^+$	N.I.	x		
84.083	$C_5H_9NH^+$	Pyridine	x	x	
91.074	$C_4H_{10}O_2H^+$	N.I.			
93.067	$C_7H_8H^+$	Toluene/ terpene fragment	x	x	
95.050	$C_6H_6OH^+$	Phenol		x	
97.064	$C_6H_8OH^+$	Dimethyl-furan		x	
97.100	$C_7H_{12}H^+$	N.I.	x		
98.061	$C_5H_7ONH^+$	Dimethyl-oxazole		x	
99.044	$C_5H_6O_2H^+$	Furfuryl alcohol/a-angelica lactone		x	
99.080	$C_6H_{10}OH^+$	Hexenal/methyl-pentenone/2,4-hexadien-1-ol	x		
103.076	$C_5H_{10}O_2H^+$	Hydroxy-pentanone/methyl-butanoic acid		x	
104.050	$C_7H_5NH^+$	Bezonitrile	x		
105.070	$C_8H_8H^+$	Styrene/phenylethanol fragment	x		x
105.090	$C_5H_{12}O_2H^+$	N.I.	x		x
107.083	$C_3H_{10}O_2N_2H^+$	N.I.	x		x
109.077	$C_6H_8N_2H^+$	2-5 Dimetilpyrazine/ethylpyrazine		x	
113.059	$C_6H_8O_2H^+$	Methyl-furfuryl-alcohol/dimethyl-furanone/methyl-cyclopentanedione/cyclotene		x	
113.095	$C_7H_{12}OH^+$	4-methylcyclohexanone		x	

m/z	Sum formula	Tentative identification	Botanical origin [b]	Geographical origin	Geographical * Botanical origin
115.111	$C_7H_{14}OH^+$	Heptanal/heptanone		x	
116.111	$C_6H_{13}ONH^+$	N.I.	x	x	
117.091	$C_6H_2O_2H^+$	Hexanoic acid		x	
119.085	$C_9H_{10}H^+$	N.I.			x
120.100	$C_5H_{13}O_2NH^+$	N.I.			x
121.096	$C_4H_{12}O_2N_2H^+$	N.I.			
125.095	$C_8H_{12}OH^+$	3,4,4-trimethyl-2-cyclopenten-1-one	x	x	
125.132	$C_9H_{16}H^+$	N.I.	x		
127.075	$C_7H_{10}O_2H^+$	3-Ethyl-1-2-cyclopentanedione		x	
129.090	$C_7H_{12}O_2H^+$	Furaneol		x	
131.105	$C_7H_{14}O_2H^+$	Heptanoic acid/Ethyl 2-methylbutanoate		x	
135.109	$C_5H_{14}O_2N_2H^+$	N.I.		x	
137.132	$C_{10}H_{16}H^+$	N.I.		x	
139.075	$C_8H_{10}O_2H^+$	4-Ethyl-1/2-benzenediol		x	
139.111	$C_9H_{14}OH^+$	E/E-2,4-nonadienal		x	
139.146	$C_{10}H_{18}H^+$	N.I.		x	
141.091	$C_8H_{12}O_2H^+$	Propyl cyclopentendione/ethyl methyl cyclopentenedione		x	
143.143	$C_9H_{18}OH^+$	Nonanal		x	
151.146	$C_{11}H_{18}H^+$	N.I.		x	
153.127	$C_{10}H_{16}OH^+$	E/E-2,4-decadienal		x	
159.140	$C_9H_{18}O_2H^+$	Nonanoic acid/C9 ester	x	x	
165.092	$C_{10}H_{12}O_2H^+$	2-phenylethylacetate		x	
169.123	$C_{10}H_{16}O_2H^+$	N.I.		x	
187.142	$C_9H_{18}O_2N_2H^+$	N.I.		x	

[a] N.I.: Not identified
[b] The symbol x indicates if the mass peak is significant ($p < 0.05$) for Geographical origin, Botanical origin, interaction Geographical * Botanical origin.

Table S2.2. Masses showing significant average concentration in the chocolate NS during mastication according to Geographical origin (Africa, Asia, South America), Botanical, origin (Criollo, Forastero, Trinitario), interaction Geographical * Botanical (non-parametric two-way Friedman-tests, p < 0.05).

m/z	Sum formula	Tentative identification	Botanical origin [b]	Geographical origin	Botanical*Geographical origin
43.053	$C_3H_6H^+$	Fragment (diverse origin)	x		
57.033	$C_3H_4OH^+$	Acrolein	x		
63.043	$C_2H_6O_2H^+$	N.I. [a]	x		
75.044	$C_3H_6O_2H^+$	Propanoinc acid	x		
83.085	$C_6H_{10}H^+$	N.I.		x	
87.044	$C_4H_6O_2H^+$	2-3 Butanedione	x		
89.059	$C_4H_8O_2H^+$	Methyl-propanoate/hydroxy-butanone	x		
93.067	$C_2H_8O_2N_2H^+$	Toluene	x		
98.061	$C_5H_7ONH^+$	Dimethyl-oxazole	x		
100.076	$C_5H_9ONH^+$	Ethylpyrrolidine/dimethylpyrrolidine	x		
102.091	$C_5H_{11}ONH^+$	N.I.	x		
103.040	$C_4H_6O_3H^+$	N.I.	x		x
103.076	$C_5H_{10}O_2H^+$	Hydroxy-pentanone/methyl-butanoic acid	x		
107.050	$C_7H_6OH^+$	Benzaldehyde	x		
107.083	$C_3H_{10}O_2N_2H^+$	N.I.	x		
114.093	$C_6H_{11}ONH^+$	N.I.		x	
115.111	$C_7H_{14}OH^+$	Heptanal/heptanone	x	x	
116.111	$C_6H_{13}ONH^+$	N.I.	x	x	
120.100	$C_5H_{13}O_2NH^+$	N.I.		x	x
123.045	$C_7H_6O_2H^+$	N.I.	x		
145.055	$C_5H_8O_3N_2H^+$	N.I.			x
157.159	$C_{10}H_{20}OH^+$	N.I.			x
164.072	$C_9H_9O_2NH^+$	N.I.		x	
167.057	$C_5H_{10}O_6H^+$	N.I.	x		
175.165	$C_{10}H_{22}O_2H^+$	N.I.	x		
190.174	$C_{10}H_{23}O_2NH^+$	N.I.	x		x

Table S2.2. Continued.

m/z	Sum formula	Tentative identification	Botanical origin [b]	Geographical origin	Botanical*Geographical origin
197.152	$C_{12}H_{20}O_2H^+$	N.I.	x		
199.203	$C_{13}H_{26}OH^+$	N.I.		x	
239.101	$C_{11}H_{14}O_4N_2H^+$	N.I.		x	
447.103	$C_{14}H_{22}O_{16}H^+$	N.I.			x

[a] N.I.: Not idenfied [b] The symbol x indicates if the mass peak is significant ($p < 0.05$) for Geographical origin, Botanical origin, interaction Geographical * Botanical origin.

51

Table S2.3. Masses showing significant average concentration in the chocolate NS during post-swallowing according to Geographical origin (Africa, Asia, South America), Botanical, origin (Criollo, Forastero, Trinitario), interaction Geographical * Botanical (non-parametric two-way Friedman-tests, $p < 0.05$).

m/z	Sum formula	Tentative identification	Botanical origin [b]	Geographical origin	Botanical*Geographical origin
43.017	$C_2H_2OH^+$	Fragment (diverse origin)	x		
43.053	$C_3H_6H^+$	Fragment (diverse origin)	x		
47.048	$C_2H_6OH^+$	N.I.	x		
57.069	$C_4H_8H^+$	Butene, propenal	x	x	
61.028	$C_2H_4O_2H^+$	Acetic acid	x	x	
75.044	$C_3H_6O_2H^+$	Propanoinc acid	x		
79.041	$C_2H_6O_3H^+$	N.I.	x	x	
79.053	$C_6H_6H^+$	Benzene	x		
83.049	$C_5H_6OH^+$	Methyl-furan	x		
83.085	$C_6H_{10}H^+$	N.I.	x	x	
84.083	$C_5H_9NH^+$	Pyridine	x		
87.044	$C_4H_6O_2H^+$	2-3 Butanedione	x		
89.059	$C_4H_8O_2H^+$	Methyl-propanoate/hydroxy-butanone	x	x	
90.019	$C_2H_3O_3NH^+$	N.I.	x		
93.067	$C_2H_8O_2N_2H^+$	Toluene	x		
97.064	$C_6H_8OH^+$	Dimethyl-furan	x		
97.100	$C_7H_{12}H^+$	N.I.	x	x	
98.061	$C_5H_7ONH^+$	Dimethyl-oxazole	x		
99.044	$C_5H_6O_2H^+$	Furfuryl alcohol/a-angelica lactone	x		
100.076	$C_5H_9ONH^+$	Ethylpyrrolidine/dimethylpyrrolidine	x		
101.06	$C_5H_8O_2H^+$	Dihydro-2-2methyl-furanone/Pentanedione/ methyl-tetrahydrofuranone	x		
102.091	$C_5H_{11}ONH^+$	N.I.	x		
103.040	$C_4H_6O_3H^+$	N.I.	x		
104.050	$C_7H_5NH^+$	Bezonitrile	x		
107.050	$C_7H_6OH^+$	Benzaldehyde	x		x

Table S2.3. Continued.

m/z	Sum formula	Tentative identification	Botanical origin [b]	Geographical origin	Botanical*Geographical origin
107.083	$C_3H_{10}O_2N_2H^+$	N.I.	x		
109.077	$C_6H_8N_2H^+$	2-5 Dimetilpyrazine/ethylpyrazine	x		
111.046	$C_6H_6O_2H^+$	Acetylfuran/methyl-furfural	x		
111.079	$C_7H_{10}OH^+$	2/3-Dimethyl-2-cyclopenten-1-one	x		
113.095	$C_7H_{12}OH^+$	Heptenone	x		x
114.093	$C_6H_{11}ONH^+$	N.I.			x
115.075	$C_6H_{10}O_2H^+$	4-Methyltetrahydro-2H-pyran-2-one	x		
115.111	$C_7H_{14}OH^+$	Heptanal/heptanone	x	x	
116.074	$C_5H_9O_2NH^+$	Proline	x		
116.111	$C_6H_{13}ONH^+$	N.I.	x	x	
120.046	$C_7H_5ONH^+$	N.I.	x		
120.100	$C_5H_{13}O_2NH^+$	N.I.	x		
123.092	$C_7H_{10}N_2H^+$	Trimethylpyrazine/2-ethyl-3-methylpyrazine	x		
125.095	$C_8H_{12}OH^+$	Butylfuran/methyl-propylfuran/3,4,4-trimethyl-2-cyclopenten-1-one		x	
125.132	$C_9H_{16}H^+$	N.I.		x	
129.090	$C_7H_{12}O_2H^+$	Pentenyl-acetate/heptanedione	x		
129.126	$C_8H_{16}OH^+$	Octanal		x	
130.066	$C_9H_7NH^+$	N.I.	x		
131.037	$C_5H_6O_4H^+$	N.I.	x		
131.071	$C_6H_{10}O_3H^+$	Acetyloxy-butanone/ethanrdiol diacetate/oxopropoxy-propanone/ethyl-oxobutanoate	x		
131.105	$C_7H_{14}O_2H^+$	Heptanoic acid/Ethyl 2-methylbutanoate		x	
133.102	$C_{10}H_{12}H^+$	N.I.	x	x	
141.127	$C_9H_{16}OH^+$	Nonenal	x		
142.090	$C_7H_{11}O_2NH^+$	N.I.	x		
142.126	$C_8H_{15}ONH^+$	N.I.	x		
143.143	$C_9H_{18}OH^+$	Nonanal	x	x	

Table S2.3. Continued.

m/z	Sum formula	Tentative identification	Botanical origin [b]	Geographical origin	Botanical*Geographical origin
149.025	$C_8H_4O_3H^+$	N.I.			x
152.069	$C_8H_9O_2NH^+$	N.I.			x
156.141	$C_9H_{17}ONH^+$	N.I.	x		
159.070	$C_7H_{10}O_4H^+$	N.I.			x
159.140	$C_9H_{18}O_2H^+$	N.I.	x		x
164.072	$C_9H_9O_2NH^+$	N.I.		x	
165.092	$C_{10}H_{12}O_2H^+$	Methyl-cinnamate	x		x
165.163	$C_{12}H_{20}H^+$	N.I.	x		x
167.175	$C_{12}H_{22}H^+$	N.I.			x
169.193	$C_{12}H_{24}H^+$	N.I.	x		
175.165	$C_{10}H_{22}O_2H^+$	N.I.	x		
176.165	$C_9H_{21}O_2NH^+$	N.I.	x		
179.108	$C_{11}H_{14}O_2H^+$	N.I.		x	
185.157	$C_{11}H_{20}O_2H^+$	N.I.		x	x
185.190	$C_{12}H_{24}OH^+$	N.I.			
197.152	$C_{12}H_{20}O_2H^+$	N.I.	x		
198.128	$C_{14}H_{15}NH^+$	N.I.	x		
199.173	$C_{12}H_{22}O_2H^+$	N.I.	x	x	
222.225	$C_{15}H_{27}NH^+$	N.I.			
227.166	$C_{13}H_{22}O_3H^+$	N.I.	x		x
227.235	$C_{15}H_{30}OH^+$	N.I.		x	
233.224	$C_{12}H_{28}O_2N_2H^+$	N.I.		x	
235.207	$C_{12}H_{28}O_2N_2H^+$	N.I.	x		
237.255	$C_{17}H_{32}H^+$	N.I.			
239.101	$C_{11}H_{14}O_4N_2H^+$	N.I.		x	x
249.255	$C_{18}H_{32}H^+$	N.I.			x
266.176	$C_{15}H_{23}O_3NH^+$	N.I.		x	

Table S2.3. Continued.

m/z	Sum formula	Tentative identification	Botanical origin [b]	Geographical origin	Botanical*Geographical origin
269.281	$C_{18}H_{36}OH^+$	N.I.	x		
329.319	$C_{24}H_{40}H^+$	N.I.		x	
353.217	$C_{16}H_{32}O_8H^+$	N.I.		x	
356.071	$C_9H_{26}O_5Si_5H^+$	N.I.		x	
386.156	$C_{12}H_{36}O_4Si_5H^+$	N.I.	x		

[a] N.I.: Not identified

[b] The symbol x indicates if the mass peak is significant ($p < 0.05$) for Geographical origin, Botanical origin, interaction Geographical * Botanical origin.

Table S2.4. List of chocolate NS mass peaks tentatively identified and corresponding parameters showing significant differences according to the cocoa beans botanical origins during mastication and/or swallowing session. p-Values were obtained applying Kruskal-Wallis test followed by Mann-Whitney post-hoc.

m/z	Sum formula	Tentative identification with references	Chemical class	Parameters	Mean concentration (ppbv)							
					Mastication session			p value	Swallowing session			p value
					Criollo	Forastero	Trinitario		Criollo	Forastero	Trinitario	
41.039	$C_3H_5^+$	Alkylic fragment [92]	Fragment	Max	316.40	277.59	244.62	>0.05	534.07a	426.99ab	372.94b	0.005
43.017	$C_2H_3O^+$	Fragment (diverse origin) [6]	Fragment	Avg	–	–	–	–	115.68b	2.91a	6.29a	<10⁻³
				Max	–	–	–	–	548.26b	86.23a	72.19a	<10⁻³
				Area	–	–	–	–	5155.28b	2058.90a	1871.53a	<10⁻³
43.053	$C_3H_7^+$	Fragment (diverse origin) [6]	Fragment	Avg	7.10	4.24	3.70	>0.05	14.84b	7.28a	6.02a	0.001
				Max	29.61b	20.41a	21.52a	0.033	52.95b	29.56a	27.24a	<10⁻³
57.069	$C_4H_9^+$	**Butene**, propenal [5]		Avg	–	–	–	–	8.69b	3.29a	3.31a	0.005
				Max	–	–	–	–	33.19b	19.89a	18.73a	0.001
				Area	–	–	–	–	610.20b	499.76a	476.98a	<10⁻³
61.028	$C_2H_5O_2^+$	Acetic acid [4]	Acid	Max	–	–	–	–	426.21b	26.97a	24.37a	<10⁻³
				Area	–	–	–	–	3180.30b	600.97a	537.58a	<10⁻³
75.044	$C_3H_7O_2^+$	Propanoinc acid [1, 4, 5, 7]	Acid	Avg	61.57b	21.51b	19.37a	<10⁻³	156.32b	52.74a	46.60a	<10⁻³
				Max	261.70b	96.68a	86.46a	<10⁻³	484.06b	160.61a	135.23a	<10⁻³
				Area	1364.83b	648.32a	557.58a	<10⁻³	3505.91b	1367.32aa	1191.49a	<10⁻³
79.053	$C_6H_7^+$	Benzene [3]		Max	–	–	–	–	8.76b	2.11a	1.59a	<10⁻³
				Area	–	–	–	–	**139.04b**	**82.60a**	**64.70c**	**<10⁻³**

Table S2.4. Continued.

m/z	Sum formula	Tentative identification with references	Chemical class	Parameters	Mastication session			p value	Swallowing session			p value
					Criollo	Forastero	Trinitario		Criollo	Forastero	Trinitario	
81.035	$C_5H_5O^+$	Furan fragment [6]	Fragment	Area	–	–	–	–	48.49a	48.09a	39.76b	$<10^{-3}$
84.083	$C_5H_{10}N^+$	Pyridine	Pyridines	Max	2.78	2.15	2.20	>0.05	5.25b	3.92a	3.82b	0.017
85.028	$C_4H_5O_2^+$	2(5H)-Furanone [3]	Ketones	Area	8.92ab	10.16a	7.84b	$<10^{-3}$	9.99a	10.52a	8.54b	$<10^{-3}$
85.064	$C_5H_9O^+$	Methyl-butenal [2]	Aldehydes	Area	29.50ab	33.81a	28.55b	0.023	54.19	62.65	53.88	>0.05
87.044	$C_4H_7O_2^+$	2-3 Butanedione [2, 5]	Ketones	Avg	2.93b	0.89a	1.87ab	0.043	9.61b	4.01a	5.90a	$<10^{-3}$
				Max	17.53	9.88	11.42	>0.05	33.59b	17.42a	19.96a	$<10^{-3}$
				Area	166.02	151.08	134.70	0.033	325.39b	229.87a	232.58a	$<10^{-3}$
89.059	$C_4H_9O_2^+$	Methyl-propanoate/hydroxy-butanone [1]	Esters/Acids	Avg	17.13b	0.73a	0.30a	$<10^{-3}$	45.75b	3.33a	3.31a	$<10^{-3}$
				Max	76.44b	8.21a	9.41a	$<10^{-3}$	150.05b	13.45a	13.92a	$<10^{-3}$
				Area	437.48b	131.75a	179.78a	$<10^{-3}$	1084.34b	196.12a	254.12a	$<10^{-3}$
93.067	$C_7H_9^+$	Toluene, terpene fragment [4, 9]	Hydrocarbons	Avg	2.79b	1.10a	0.88a	$<10^{-3}$	6.52b	2.52a	2.17a	$<10^{-3}$
				Max	12.97b	6.82a	6.04a	$<10^{-3}$	21.29b	10.65a	9.18a	$<10^{-3}$
				Area	145.88b	127.11a	106.68c	$<10^{-3}$	241.48b	161.36a	141.88a	$<10^{-3}$
97.064	$C_6H_9O^+$	Dimethyl-furan [6]	Furans	Max	1.21	0.90	0.94	>0.05	1.97b	1.26a	1.24a	$<10^{-3}$
				Area	24.79a	25.85a	22.96a	0.033	34.52a	28.46ab	27.25b	0.015
98.061	$C_5H_8ON^+$	Dimethyl-oxazole [6]	Oxazoles	Avg	0.05b	0.0092a	0.0090a	0.016	0.28b	0.10a	0.10a	$<10^{-3}$
				Max	0.29b	0.18a	0.18a	0.001	0.60b	0.29a	0.30a	$<10^{-3}$
				Area	3.94	3.63	3.31	0.040	9.33b	5.59a	5.39a	$<10^{-3}$
99.044	$C_5H_7O_2^+$	Furfuryl alcohol/a-angelica lactone [6]	Pyrans/Furans	Area	17.75ab	19.46a	16.04b	$<10^{-3}$	21.91a	22.00a	18.44b	$<10^{-3}$
99.080	$C_6H_{11}O^+$	Hexenal/methyl-pentenone/2,4-hexadien-1-ol [2]	Aldehydes/Alcohol	Area	23.05ab	24.95a	19.47b	$<10^{-3}$	33.25a	32.67a	27.64b	$<10^{-3}$
100.076	$C_5H_{10}ON^+$	Ethylpyrrolidine/dimethylpyrrolidine [5]		Area	9.13ab	10.65a	7.70b	$<10^{-3}$	9.58a	10.48a	8.32b	$<10^{-3}$

Table S2.4. Continued.

| m/z | Sum formula | Tentative identification with references | Chemical class | Parameters | Mean concentration (ppbv) | | | | | | | |
| | | | | | Mastication session | | | | Swallowing session | | | |
					Criollo	Forastero	Trinitario	p value	Criollo	Forastero	Trinitario	p value
101.060	$C_5H_9O_2^+$	Dihydro-2-2methyl-furanone/Pentanedione/methyl-tetrahydrofuranone [5]	Furans/Ketones	Max	0.94	0.87	0.52	>0.05	1.90b	0.81a	0.14a	<10⁻³
107.050	$C_7H_7O^+$	Benzaldehyde [4]	Nitrogen compounds	Max	3.83	2.81	2.70	>0.05	7.91b	5.81a	5.63a	0.022
109.077	$C_6H_9N_2^+$	2-5 Dimetilpyrazine/ethylpyrazine [2, 5]	Pyrazines	Area	29.86	26.03	22.64	0.023	77.87a	61.56ab	59.41b	0.014
				Avg	—	—	—	—	0.56b	0.17a	0.26a	0.005
				Max	—	—	—	—	1.42b	0.88a	0.96a	0.010
				Area	—	—	—	—	35.31b	29.44a	28.16a	0.005
112.077	$C_6H_{10}N^+$	Trimethyl-oxazole/2-acetyl-1-pyrroline [1, 6]	Oxazoles	Avg	—	—	—	—	**0.20b**	**0.005a**	**0.05c**	**<10⁻³**
				Max	—	—	—	—	**0.44b**	**0.14a**	**0.20c**	**<10⁻³**
				Area	—	—	—	—	7.31b	3.80a	4.27a	<10⁻³
116.074	$C_5H_{10}O_2N^+$	Proline [7]		Area	1.89a	2.22b	1.85a	0.001	2.17	2.28	2.05	>0.05
118.066	$C_6H_8N^+$	1H-Indole [2, 5]	Pyrroles	Area	2.82a	2.85a	2.40b	0.003	4.18a	3.71ab	3.26b	0.002
123.092	$C_7H_{11}N_2^+$	Trimethylpyrazine/2-ethyl-3-methylpyrazine [2, 5]	Pyrazines	Avg	—	—	—	—	**0.62b**	**0.19a**	**0.29c**	**<10⁻³**
				Max	—	—	—	—	**1.33b**	**0.55a**	**0.75c**	**<10⁻³**
				Area	—	—	—	—	21.86b	13.81a	14.12a	0.000
125.095	$C_8H_{13}O^+$	Butylfuran/methyl-propylfuran/3,4,4-trimethyl-2-cyclopenten-1-one [2, 6]	Furans/Ketones	Area	11.70ab	13.75a	11.14b	0.005	13.55ab	15.18a	12.67b	0.001
127.039	$C_6H_7O_3^+$	Maltol/methyl-furoate [2, 6]	Pyrans/Furans	Area	2.74ab	3.05a	2.47b	<10⁻³	2.95a	3.12a	2.60b	<10⁻³
127.075	$C_7H_{11}O_2^+$	3-Ethyl-1-2-cyclopentanedione [6]	Carbonyls	Area	8.03ab	9.25a	7.85b	0.007	8.33a	8.99a	7.91a	0.037

Table S2.4. Continued.

m/z	Sum formula	Tentative identification with references	Chemical class	Parameters	Mastication session			p value	Swallowing session			p value
					Criollo	Forastero	Trinitario		Criollo	Forastero	Trinitario	
129.059	$C_6H_9O_3^+$	Furaneol [2, 5]	Alcohol	Area	3.74ab	4.19a	3.60b	0.024	3.91	4.08	3.58	>0.05
129.126	$C_8H_{17}O^+$	Octanal [5]	Aldehydes	Area	8.51ab	9.32a	8.32b	0.017	10.23a	9.98a	9.37a	0.018
131.071	$C_6H_{11}O_3^+$	Acetyloxy-butanone/ethanrdiol diacetate/oxopropoxy-propanone/ethyl-oxobutanoate [8]	Ester/Acids	Avg	–	–	–	–	0.12b	0.019a	0.02a	<10⁻³
				Max	–	–	–	–	0.29b	0.14a	0.14a	0.003
				Area	–	–	–	–	**5.98b**	**4.53a**	**3.77c**	**<10⁻³**
131.105	$C_7H_{15}O_2^+$	Heptanoic acid/Ethyl 2-methylbutanoate [1, 8]	Ester/Acids	Area	11.66ab	11.83a	8.95b	<10⁻³	16.89	15.66	13.29	>0.05
136.077	$C_8H_{10}ON^+$	Ethyl-dimethyl-pyrazine/1-(2-pyridyl)-1-propanone [2]	Pyridines	Area	1.77ab	1.83a	2.09b	0.009	1.97	1.86	2.29	>0.05
141.054	$C_7H_9O_3^+$	Furfuryl-acetate [8]	Pyrans/Furans	Area	2.13ab	2.36a	1.86b	<10⁻³	2.28ab	2.42a	2.04b	0.001
141.091	$C_8H_{13}O_2^+$	Propyl cyclopentendione/ethyl methyl cyclopentenedione [3]		Area	6.89ab	7.77a	6.09b	<10⁻³	6.75ab	7.16a	5.89b	0.001
141.127	$C_9H_{17}O^+$	Nonenal [5]	Aldehydes	Area	9.20ab	9.82a	7.91b	<10⁻³	10.27a	10.34a	8.89b	<10⁻³
145.123	$C_8H_{17}O_2^+$	Ethyl-hexanoic acid/C8 ester [8]	Ester/Acids	Area	10.50a	10.17a	7.56b	<10⁻³	18.25a	15.75ab	13.03b	0.003
153.127	$C_{10}H_{17}O^+$	E/E-2,4-decadienal [6]	Aldehydes	Area	6.52ab	7.16a	5.71b	<10⁻³	6.87ab	7.16a	6.12b	<10⁻³
155.143	$C_{10}H_{19}O^+$	Linalool [2]	Alcohol	Area	11.36ab	11.52a	9.37b	0.001	11.72	11.49	10.38	>0.05
159.140	$C_9H_{19}O_2^+$	Nonanoic acid/C9 ester [8]	Ester/Acids	Area	–	–	–	–	10.23a	9.16a	6.91b	<10⁻³
167.107	$C_{10}H_{15}O_2^+$	Ethyl-dimethoxy-benzene [8]	Esters	Area	4.79a	5.63a	5.85b	0.003	4.82b	5.17ab	5.81b	0.001
171.139	$C_{10}H_{19}O_2^+$	Linalool oxide/delta decenolactone [2, 5]	Furans	Area	4.55b	5.38a	4.23b	<10⁻³	4.74b	5.36a	4.43b	<10⁻³
173.151	$C_{10}H_{21}O_2^+$	Ethyloctanoate [6]	Esters	Area	–	–	–	–	4.22ab	4.87a	3.89b	0.016
182.083	$C_9H_{12}O_3N^+$	Tyrosine [7]		Area	–	–	–	–	0.34a	0.35a	0.28b	<10⁻³
285.276	$C_{18}H_{37}O_2^+$	Stearic acid [5]	Acids	Area	–	–	–	–	1.28ab	1.57a	0.88b	0.011

Mean concentration (ppbv)

The missing masses in the mastication phase did not show any statistical difference with the breathing session.
Different superscripts indicate significant differences Kruskal-Wallis and Mann Whitney-test (p < 0.05). Different colours indicate Mann–Whitney results. Blue colour shows which origin(s) are discriminated, light blue colour is the origin(s) to be compared with.
Bold font indicates values showing significant differences.
a References: [1] Frauendorfer & Schieberle (2008); [2] Afoakwa et al. (2008); [3] Romano et al. (2014); [4] Acierno et al. (2016); [5] Diab et al. (2014); [6] Yener et al. (2014); [7] Nolvachai et al. (2015); [8] Yener et al. (2016); [9] Özdestan et al. (2013).
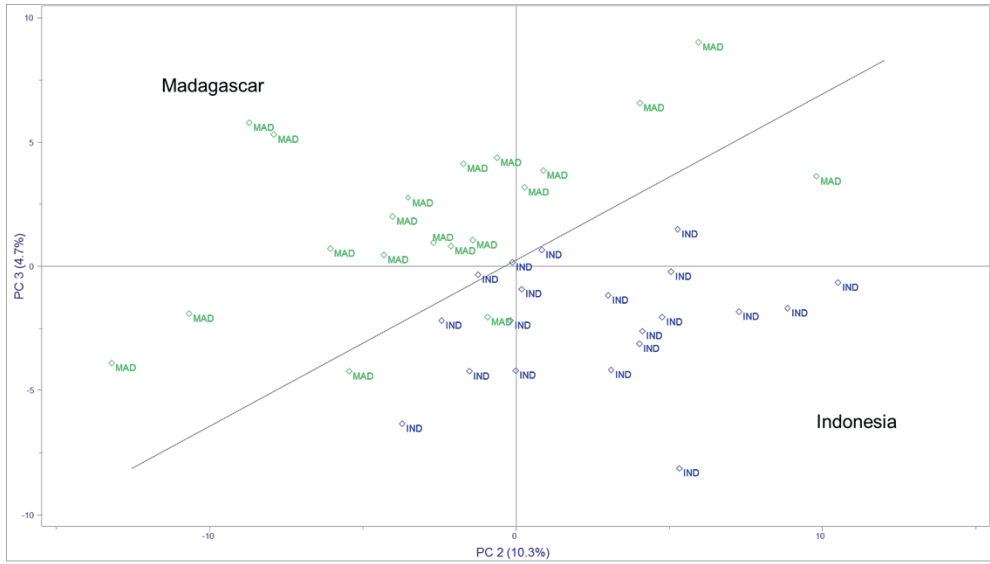
Figure S2.1. PCA scores plot of the mean-centred and normalized NS (area) data of the post-swallowing phase collected from subjects 10 when eating 'Criollo chocolates'. Chocolates made with beans from the same geographical origin are close together. IND: Indonesia: Mad: Madagascar
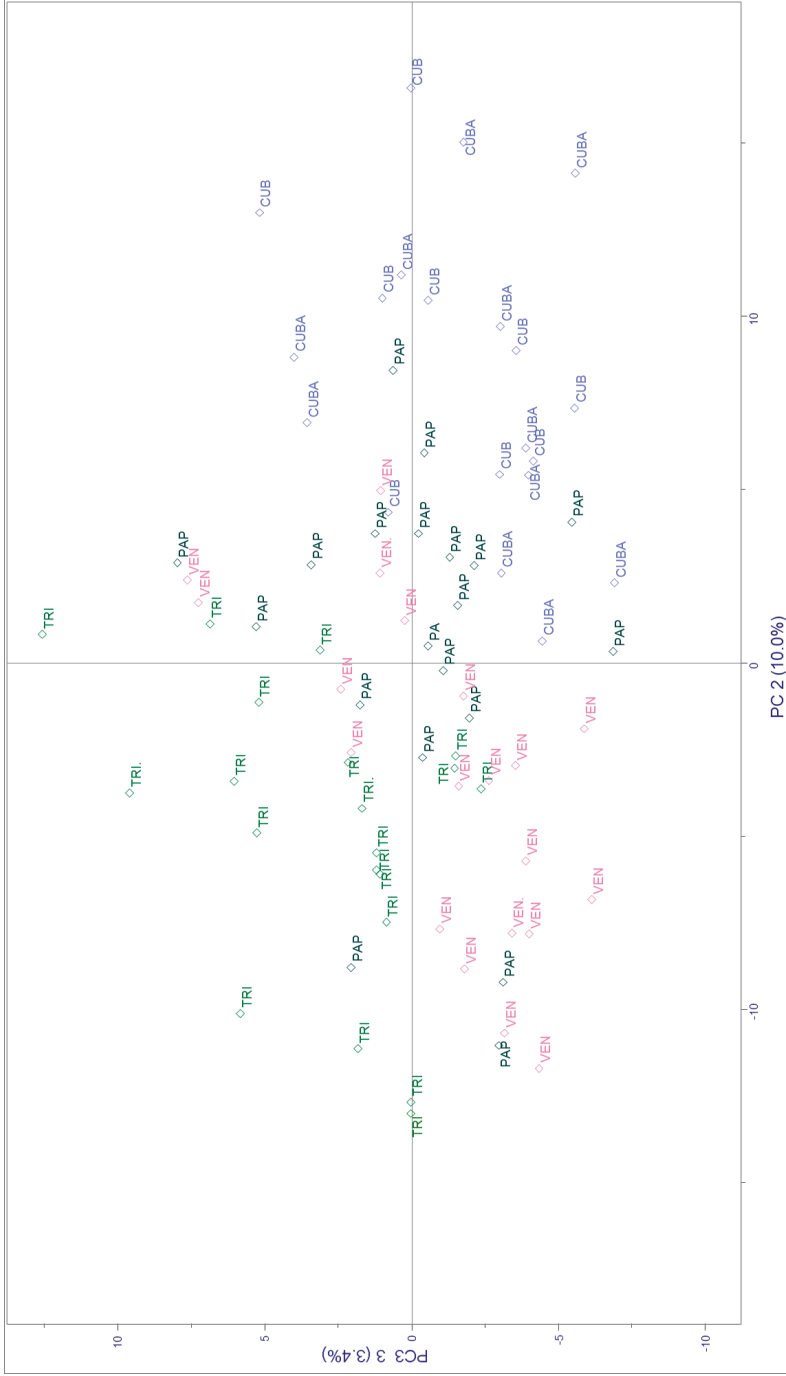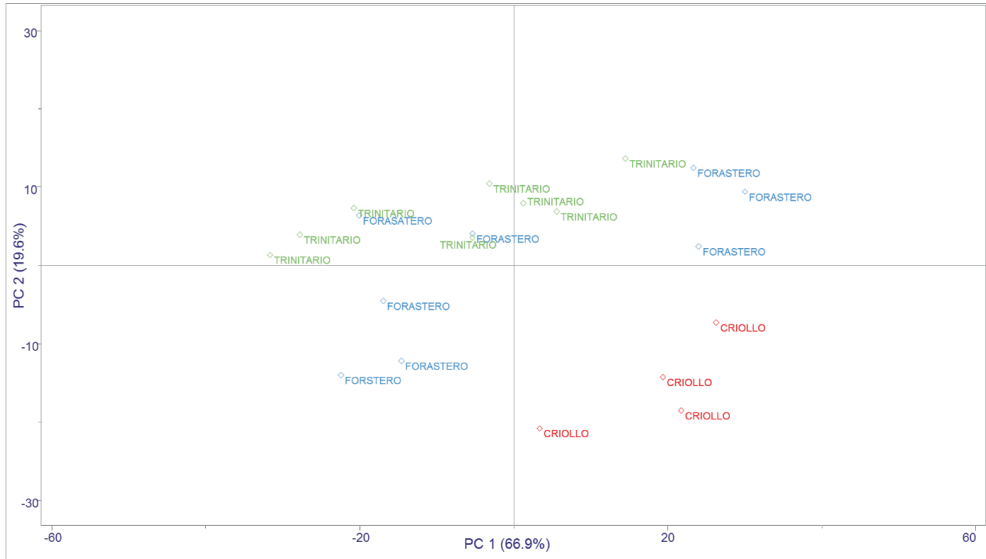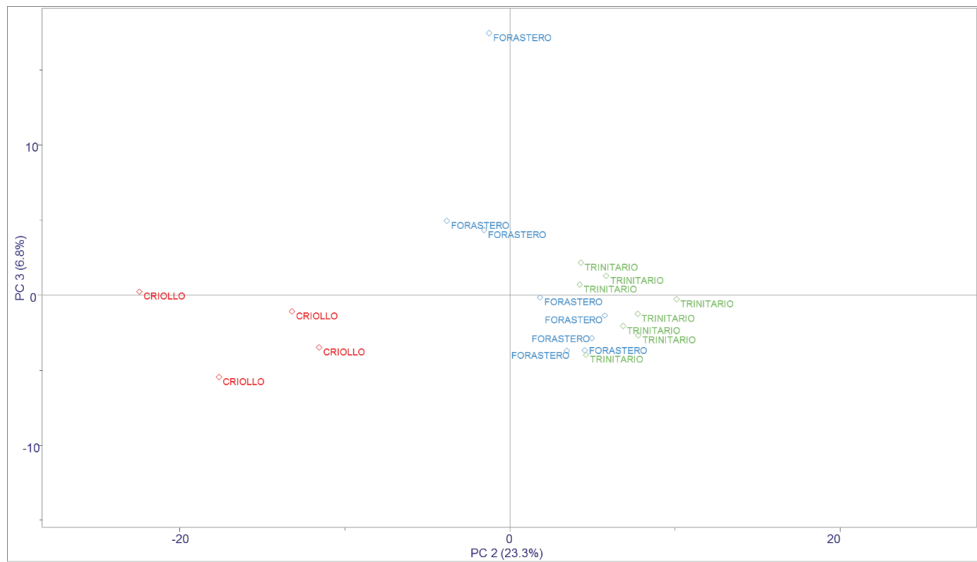
Figure S2.2. PCA scores plot of the mean-centred and normalized NS (area) data of the post-swallowing phase collected from subjects 10 when eating 'Trinitario chocolates'. Chocolates made with beans from the same geographical origin are close together. VEN:Venezuela; PAP: Papua Nuova Guinea; TRI:Trinidad; CUBA.

61

Figure S2.3. PCA scores plot of the mean-centred and normalized NS (*area*) data of the post-swallowing phase collected from subjects 1 when eating 10 different chocolates. A botanical trend is shown. Three clusters related to the three different botanical origin (Criolo, Forastero, Trinitario) are displayed.

Figure S2.4. PCA scores plot of the mean-centred and normalized NS (area) data of the post-swallowing phase collected from subjects 5 when eating 10 different chocolates. A botanical trend is shown. Three clusters related to the three different botanical origin (Criolo, Forastero, Trinitario) are displayed.

Chapter 3

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

Abstract

The intrinsic characteristics of chocolate and the complex technological process complicate the assessment of the typical features of this product and the verification of its authenticity. In this study, the influence of the botanical and geographical origin of the cocoa beans, as well as the impact of brand-related processing on the volatile organic compound (VOC) composition of the resulting chocolates was examined. A total of ninety dark chocolates available on the Dutch market were analysed using Proton-Transfer-Reaction-Mass Spectrometry (PTR-MS). The VOC profiles generated by PTR-MS (136 masses per sample) were used as fingerprints and investigated using chemometric tools to elucidate information on production factors of cocoa and subsequent processing, in the finished product. Principal component analysis (PCA) showed some clustering of the chocolates according to the botanical and geographical origins of beans as well as according to the brand. Partial least square discriminant analysis (PLS-DA) further discriminated the samples according to the three classes (botanical origin, geographical origin, brand) and the models with the best classification results were used to investigate the relevant masses for each class. PCA clustering and PLS-DA classification highlighted that chocolate profiles are strongly affected by the processing applied by the different brands. However, a reflection of the botanical and geographical origins of the beans was also mirrored in the VOC composition of the chocolates. PTR-Time of Flight-MS (PTR-ToF-MS) was used to tentatively identify the VOCs of the chocolates. These measurements allowed the identification of 36 spectrometric peaks which relate to the main classes of chocolate odorant compounds, in particular, aldehydes and pyrazines, products of Maillard reactions. Several compounds already present in unroasted beans were tentatively identified in the chocolates as well, such as acetic acid, methylpropanoic acid, 2- and 3-methylbutanoic acid, 2-phenylethanol, and tetramethylpyrazine. The results of this study emphasise the impact of the brand-related formulation and processing on VOC profiles of dark chocolates. However, using chemometrics, VOC reflection of the botanical origin and geographical origin of the beans in the chocolates was revealed, which may be useful for a future cocoa/chocolate traceability.

Keywords Chemometrics, Chocolate, Cocoa bean genetic group, Geographical origin, PTR-MS, Volatile compound fingerprint

3.1 Introduction

Food authentication aims to protect consumers from being sold an inferior product, with a false description, and honest traders from unfair competition. Common authenticity issues in food are: (1) to ensure composition and constituents of foods; (2) to guarantee the country, regional, and production provenance of a product; and (3) to protect from counterfeits. Developing or improving analytical methods able to investigate these parameters are necessary steps in food analyses (Dennis, 1998).

The complex characteristics of chocolate-related to the supply chain, the technology involved in the production, and the ingredients' composition complicate the assessment of the typical features of this product and the verification of its authenticity. Therefore, the discovery of markers related to e.g. the botanical and geographical origins of the beans would be a first step towards the development of a reliable method for authentication which in turn would underpin sustainable production and would help to preserve stakeholders' confidence (Saltini, Akkerman, & Frosch, 2013). Besides consumers, stakeholders include farmers, shipping organizations, processors, and distributors.

Mass spectrometry, spectroscopic and separation techniques have been applied to assess the food's geographical origin (Luykx & van Ruth, 2008) and species authentication (Bossier, 1999; Fajardo, González, Rojas, García, & Martín, 2010; Spaniolas, May, Bennett, & Tucker, 2006). Several studies in this area have been carried out on coffee (Downey, Briandet, Wilson, & Kemsley, 1997), wine (Liu, Cozzolino, Cynkar, Gishen, & Colby, 2006), fruit juice (Zielinski et al., 2014), and tea (Kovács et al., 2010). Chocolate authentication is especially focused on cocoa butter evaluation (Cheman, 2005; Dionisi et al., 2004; Ulberth & Buchgraber, 2003). Nevertheless, as stated before, other authentication issues are evenly relevant such as the assessment of the origin and verification of the brand.

Publications related to the geographical origin of chocolate and cocoa beans, and to the cocoa variety concern the differentiation of the products by their fatty acid profiles (Hernandez, Castellote, & Permanyer, 1991) and the examination of the volatile and non-volatile profiles (Afoakwa, Paterson, Fowler, & Ryan, 2008; Farah, Zaibunnisa, Misnawi, & Zainal, 2012; Hernandez & Rutledge, 1994; Jinap, Dimick, & Hollender, 1995). However, the majority of these studies aim to characterise and improve the aroma quality of cocoa and chocolate, and just a few of them regard chocolate authenticity (Caligiani, Cirlini, Palla, Ravaglia, & Arlorio, 2007; Cambrai et al., 2010).

Because chocolate is a complex matrix, it is difficult to pinpoint particular markers for its authenticity. Traditional techniques cannot fully satisfy the new needs of food authentication, as they just focus on specific markers or particular undesired compounds, which cannot characterise a product according to the origin or the production steps. Therefore, an analytical fingerprint approach may be more suitable. This methodology is a non-selective way of analysis and takes into account a complete spectrum or an image of the test material. Combining analytical techniques with statistical analyses, the fingerprint aims at having a more complete description of the product (Capuano & van Ruth, 2012). This approach has been applied for the authentication of different foods such as organic eggs (Tres & van Ruth, 2011), olive oil (Araghipour et al., 2008), and cheese (Biasioli et al., 2006) among many others.

The volatile organic compound (VOC) composition of chocolate is influenced by the genotype and origin of the cocoa beans, the agro-climatic condition of growing, and post-harvest fermentation, drying, and storage (Afoakwa et al., 2008; Camu et al., 2008; Jinap et al., 1995; Rodriguez-Campos, Escalona-Buendía, Orozco-Avila, Lugo-Cervantes, & Jaramillo-Flores, 2011). Therefore, each bean variety will result in a characteristic flavour that will impact on the final product, together with the manufacturing steps as roasting and conching (Counet, Callemien, Ouwerx, & Collin, 2002; Frauendorfer & Schieberle, 2008). By analysing the volatile profile, we aim to reveal the reflection of origin factors and processing in the final chocolate product.

Several studies showed that volatile compounds are preserved during cocoa beans fermentation and drying. In the final product, even after the thermal treatments, it is possible to find compounds developed before and after cocoa fermentation (Aculey et al., 2010; Frauendorfer & Schieberle, 2008). This is an important consideration that underlines the possibility to detect the VOCs related to the raw material in a chocolate bar.

This study aims to explore how the volatile profiles of dark chocolates are affected by the botanical and geographical origins of the cocoa beans used in the manufacturing and the different systems of processing applied by various brands, to extract information that may be useful for future authentication. The VOCs' investigation was carried out by High-sensitivity-Proton Transfer Reaction Mass Spectrometry (HS-PTR-MS). Headspace PTR-MS analysis requires no chemical pre-treatment of the sample and it is a sensitive (pptv, part per trillion by volume level) and fast technique (b1 min for a complete mass spectrum). PTR-MS is based on a "soft ionization" characterised by a proton transfer that generates protonated molecules with little or no fragmentation. Most commonly, H3O+ is used as a protonating agent. One of the most important advantages in using H3O+ is the possibility to use the samples headspace air as buffer gas since H3O+ does not react with the main natural components of air (Lindinger, Hansel, & Jordan, 1998). To further characterise the chocolates, VOCs were tentatively identified using PTR-time of flight-MS (PTR-ToF-MS). Ninety dark chocolates available on the Dutch market were analysed to compare products available to consumers.

3.2 Materials and methods

3.2.1 Sampling
Ninety dark chocolate bars, available on the Dutch market, were collected from retail outlets in 2013 and were analysed. The chocolates were collected considering the botanical and geographical origins of the cocoa beans, and the brand presented on the labels. According to this information, the set of samples was subdivided as shown in Table 3.1. Chocolate bars were stored at room temperature.

3.2.2 HS-PTR-MS analysis
Chocolate samples were powdered using an electrical grater for food and kept at 4 °C before analysis. For the measurements, 3.0 g of ground chocolate was weighed into clean and odourless flasks of 250 ml. The closed flasks were placed in a water bath at 40 °C for 30 min to equilibrate the samples with their headspace. Preliminary experiments showed that 30 min was sufficient for equilibration. The headspace of the powdered dark chocolate was measured by HS-PTR-MS (Ionicon Analytik G.m.b.H., Innsbruck, Austria). The PTR-MS conditions were as follows: drift pressure 2.20 mbar, inlet flow 60 ml/min, the reaction chamber and inlet temperature at 60 °C. The instrument was operated at an E/N (ratio of electric field strength across the reaction chamber, E, to buffer gas number density, N, within the chamber) of 119 Td (1 Td = 10^{-17} cm^2 V molecule^{-1}). For each sample, a mass range between 20 and 160 was measured using a dwell time of 0.2 s mass^{-1}. A blank (empty flask) was analysed before each sample. Five cycles per measurement were recorded and specifically the three last cycles for the blanks and the three in the middle for the samples were used for data analysis. The values obtained for the blank were subtracted from each sample and all values were corrected for transmission. HS-PTR-MS analysis was run in triplicate.

Table 3.1. Sample subdivision according to the properties of interest.

Group	Total number of samples	Subgroup[a]	Number of samples
Botanical origin	16		
		Criollo	6
		Forastero	4
		Trinitario	6
Geographical origin	90		
		Africa	15
		Asia	11
		Oceania	2
		South America	30
		Mixed origin	11
		Unknown origin	21
Brand	46		
		A	4
		B	4
		C	5
		D	6
		E	5
		F	14
		G	4
		H	4

[a] For each group, the total number of samples belonging and the subgroups with the related number of samples are specified.

3.2.2.1 HS-PTR-MS data evaluation and selection

In this study, the raw counts coming from the detector were converted into ppbv (part per billion by volume). The quality of cycles was evaluated checking the r^2 between the three chosen cycles. A value over 0.99 was obtained for all the samples. Subsequently, the average of the three cycles was calculated and then the average of the triplicate. Negative values were treated as zero. This all resulted in a matrix of 90 samples with 136 ion intensity values in ppbv.

3.2.2.2 HS-PTR-MS chemometrics

The chocolate VOC fingerprints were investigated using chemometric tools to extract information related to the botanical and geographical origin of the beans used and to the influence of the brand. To investigate the change in concentration of the predominant ions, m/z 33, 43 and 61, according to the three classes (botanical origin, geographical origin, brands), MANOVA was performed. MANOVA can assess independent variables, mass peaks intensity, to the significance of the effect on two or more metric dependents, in this case: factors botanical origin, samples, and repeats; factors geographical origin, samples, and repeats; factors brand, samples, and repeats. This allows a joint analysis rather than performing several univariate tests, avoiding multiple testing risk. Fisher's Least Significant Difference tests (LSD test) were carried out to determine significant differences among groups using XLSTAT 2014 3.02 (Addinsoft, Paris, France). A significance level of $p \leq 0.05$ was used throughout the study.

In addition, the total chocolate profile was statistically analysed using Pirouette 4.5 software (Infometrix, Seattle, USA). Principal Component Analysis (PCA) was applied to reduce the data dimensionality to enable the data visualisation and to find possible natural clusters, according to the chosen parameters (botanical origin, geographical origin, brands).

Partial Least Squares Discriminant Analysis (PLS-DA) was applied to classify the chocolate samples according to the beans' botanical origin (genetic group: Criollo, Forastero, Trinitario), to the beans' geographical origin (Africa, Asia, South-America), and the chocolates' brands. This statistical method reduces the number of variables used in the model, combining the variables in the data set to calculate factors that have the maximum correlation with the class (Tres & van Ruth, 2011). Both for PCA and PLS-DA, several data pre-processing and transforming methods were explored (mean-centering, autoscaling, normalization, log10 transformation) (Zielinski et al., 2014). The PLS-DA models were evaluated by exhaustive cross-validation.

3.2.3 PTR-ToF-MS analysis — VOCs' tentative identification

The volatile compounds of a set of samples were tentatively identified using PTR-ToF-MS. Four chocolates, one from Grenada (South America), one from Ecuador (South America), one from Vietnam (Asia), and one from Tanzania (Africa) were characterised. They were selected based on the PTR-MS results to provide a representative set of samples covering the variability encountered in the sample set. The chocolates are from different brands and have a cocoa percentage of 60, 70, 80 and 53 respectively. The samples were kept at 4 °C before analysis and ground to fine particle size. After grinding, 1 g of chocolate sample was weighed into 22 ml glass vials (Supelco, Bellefonte, USA) and 6 analytical replicates were prepared for each sample. Six vials containing lab air were prepared and employed as "blank".

The headspace measurements were performed by using a commercial PTR-ToF-MS 8000 instrument (Ionicon Analytik GmbH, Innsbruck, Austria). The instrumental conditions for the proton transfer reaction were the following: drift voltage 465 V, drift temperature 110 °C, drift pressure 2.30 mbar affording an E/N value of 120 Townsend (1Td = 10^{-17} cm^2 V $^{-1}$ s^{-1}). Sampling was performed with a flow rate of 40 ml/min in a mass range between 20 and 240 m/z. The mass resolution (m/ m) was at least 3600. Measurements were performed in an automated way by using a multipurpose GC automatic sampler (Gerstel GmbH, Mulheim am Ruhr, Germany) by connecting the automatic sampler syringe to the inlet of PTR-ToF-MS. The measurement cycle for each vial consisted of flushing the headspace of the sample with air for 1 min at 200 ml/min, then incubation for 30 min at 40 °C, and finally the measurement. The measurement order was randomized to avoid possible systematic memory effects. The sample headspace was measured by direct injection into the PTR-ToF-MS drift tube via a heated (110 °C) peek inlet. Each sample was measured for 30 s, at an acquisition rate of 1 spectrum/s. The thirty scans per sample were averaged and consequently, the average of the six replicates was taken into account for each sample for further mass spectral data analysis.

3.2.3.1 PTR-ToF-MS data analysis

Deadtime correction, internal calibration of mass spectral data and peak extraction were performed according to the procedures described in Cappellin et al. (2010) and Cappellin et al. (2011). The experimental m/z values were reported up to four decimals. In this paper, the VOC concentrations are presented in ppbv and the calculations from peak areas according to the formula described by Lindinger et al. (1998) using a constant reaction rate coefficient ($k_R = 2 \times 10^{-9}$ cm^3/s). The same procedure was applied to the six blank samples. The average signals of blanks were subtracted from the sample data set.

3.3 Results and discussion

3.3.1 HS-PTR-MS

The 90 chocolate samples were subjected to HS-PTR-MS analysis. The mass spectra of the 90 chocolates obtained by the PTR-MS, showed signals on most masses in the defined measurement range (20–160 amu). The spectra, used as fingerprints, comprised 136 masses after removal of O^{+2} (m/z 32) and water clusters (m/z 37). An example of a PTR-MS mass spectrum of chocolate produced with Criollo beans is presented in Figure 3.1.

For the sample set of the 90 chocolates, by far highest concentrations of the ions m/z 33, 43, and 61 were observed in all samples. For these masses, concentrations of at least three times the value of the other masses were observed. They are therefore considered the predominant masses of the chocolates. Masses 33, 43 and 61 were tentatively identified as methanol, fragment of diverse origin, and acetic acid respectively (Table 3.3). Because of their high concentration, these three masses can have a huge impact on the fingerprint analysis. For these reasons, to understand if the intensities of these mass peaks significantly differentiate the samples according to the cocoa botanical or geographical origin of the beans, or the brand, MANOVA was applied (Table 3.2) as a first step.

Masses 43 and 61 are related to acid acetic. However, taking into account that m/z 43 is a fragment of diverse origin, these two compounds were evaluated independently (Biasioli et al., 2004; Yeretzian, Jordan, & Lindinger, 2003). The MANOVA data related to the cocoa botanical origin show significant differences in mass peak intensities for the Criollo group compared to the other two genetic groups, i.e. higher intensities for mass peaks m/z 33, 43, and 61. Specifically, it has been shown that m/z 61, acetic acid, is one of the dominant compounds in the aroma of the unroasted cocoa beans of Criollo, and it is kept even after roasting (Frauendorfer & Schieberle, 2008). Furthermore, according to other studies chocolates produced from these cocoa beans have higher acidic tones, related to acetic acid, compared to Forastero and Trinitario beans (Holm, Astong, & Douglas, 1993; Jinap & Dimick, 1990; Saltini et al., 2013).
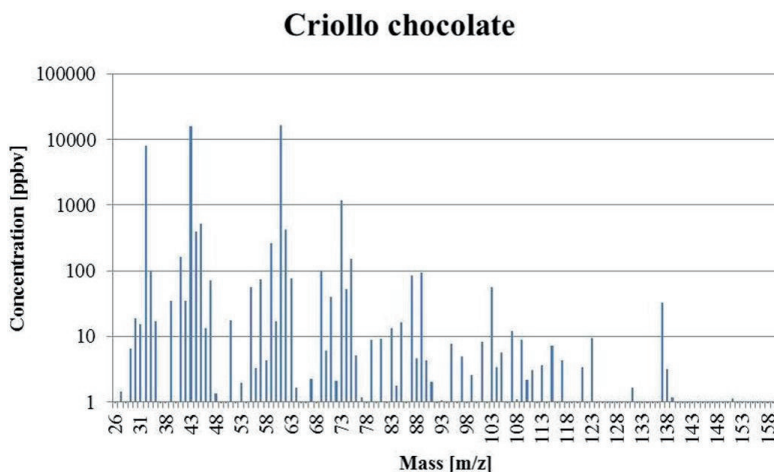
Figure 3.1. HS-PTR-MS mass spectrum of the volatile organic compounds of chocolate manufactured from Criollo cocoa beans.

Table 3.2. Comparison of the predominant mass peaks of the chocolates (mean across groups* ± SD)**.

Factor		Volatile compounds [ppbv]	
	m/z 33	m/z 43	m/z 61
Botanical origin n = 16)			
Criollo (n = 6 × 3)	8142 ± 7650[A]	15988 ± 14394[A]	16514 ± 12479[A]
Forastero (n = 4 × 3)	266 ± 79[B]	973 ± 440[B]	1636 ± 639[B]
Trinitario (n = 6 × 3)	418 ± 239[B]	1742 ± 1229[B]	2524 ± 1480[B]
Geographical origin (n = 56)			
Africa (n = 15 × 3)	1434 ± 1510[B]	1934 ± 1150[B]	3118 ± 1671[B]
Asia (n = 11 × 3)	785 ± 1127[B]	7278 ± 12557[A]	8079 ± 11115[A]
South America (n = 30 × 3)	3257 ± 5825[A]	6260 ± 5329[A]	8056 ± 5915[A]
Brands (n = 46)			
A (n = 4 × 3)	569 ± 210[CD]	2282 ± 1139[C]	4560 ± 1369[CD]
B (n = 4 × 3)	4422 ± 2246[B]	3358 ± 2143[C]	5000 ± 2846[C]
C (n = 5 × 3)	1373 ± 981[CD]	1504 ± 959[C]	2680 ± 1459[CD]
D (n = 6 × 3)	1766 ± 898[C]	1813 ± 532[C]	2984 ± 649[CD]
E (n = 5 × 3)	286 ± 27[D]	1917 ± 653[C]	2897 ± 933[CD]
F (n = 14 × 3)	350 ± 193[D]	1556 ± 982[C]	2440 ± 1314[D]
G (n = 4 × 3)	11881 ± 6672[A]	22095 ± 13874[A]	22139 ± 11327[A]
H (n = 4 × 3)	1967 ± 2371[C]	7229 ± 3014[B]	9847 ± 2566[B]

*Groups consist of samples of known botanical origin, known single geographical origin, and a minimum sample size of 4 per group.
**Different superscripts in a column within the cocoa botanical origin, geographical origin, and brand groups indicate significant differences (MANOVA and Fisher's LSD-test, p b 0.05).

Regarding the geographical origin, a sub-set of samples of known origin and a minimum sample set of four per group were used for the evaluation. For this reason, the attention was focused on the Asian, African and South American samples.

Significant differences for this group were observed for mass 33 between South America and the other continents. MANOVA did not show relevant differences for m/z 43 and 61 between chocolates made of Asian and South American beans but the headspace concentration of m/z 43 and 61 of chocolates from these continents are significantly different from those made of African beans. According to literature, Brazilian and Asian cocoa beans have a higher concentration of acetic and lactic acids levels compared to other origins (Jinap, 1991).

Regarding the brands, the three masses are particularly relevant for differentiating one brand, such as brand G. What is important to specify is that the chocolates related to brand G are manufactured with Criollo cocoa beans. Taking into account the means across the groups and the number of samples in Table 3.2, the concentration of m/z 61 is probably more related to the processing steps than to the beans. However, the value related to the Criollo beans is significantly higher compared to the other genetic groups and the geographical origin value, indicating a certain influence of this type of beans on the acetic acid concentration as well. A correlation between the acetic acid levels of cocoa and chocolate is likely, but further investigation and analysis are needed to confirm this hypothesis.

3.3.1.1 Cocoa botanical origin

To understand if the cocoa botanical origin affects the volatile profiles of the chocolates, a sub-set of the 90 samples, i.e. the samples for which the genetic groups of cocoa, Criollo (6), Forastero (4), and Trinitario (6) were known, were subjected to PCA.

Figure 3.2 shows the differentiation of three clusters related to the three genetic groups of cocoa included in our set of samples. Criollo samples are clustered apart from the Forastero and the Trinitario ones. Chocolates made with Trinitario beans are distributed between the other two classes. This is an interesting distribution as Trinitario cocoa beans are a hybrid of Criollo and Forastero so the spreading of these samples in the middle can be related to characteristics shared with the other two groups. The clear clustering in this PCA may be related to the unique profile of each bean variety, influenced by several factors as the species of cocoa, the location and conditions of growth, and the farming condition of the beans (Afoakwa et al., 2008). One sample of Criollo, not included in the cluster of this group, can be considered an outlier. The explanation of this distribution can be related to the fact that this Criollo sample belongs to the same brand of the Forastero samples with which it is grouped.

Subsequently, supervised classification was carried out. A PLS-DA model was built to classify the chocolates according to the kind of cocoa beans used, to maximise the separation between the classes. Leave one out cross-validated results showed that the PLS-DA model with 9 factors for Criollo class, 8 factors for Forastero class and 8 factors for Trinitario class, on the normalized and auto-scaled data provided good classification rates, with a total of 94% efficiency. The model was developed applying one orthogonal signal correction (OSC). Criollo and Trinitario samples were correctly classified while one sample of Forastero was allocated to an incorrect class (Trinitario). Even in this case, the sample belongs to the same brand. These results underline the impact of the brand on the final profile of the chocolates. However, the trend shown in the PCA related to the cocoa beans genetic group (Fig. 3.2) suggests that probably intrinsic features of the beans are still maintained, even after the processing. A similar characteristic was observed for wines (Granato, Katayama, & de Castro, 2011), and grape juices (Iora et al., 2015) even though these are not roasted foods.

From the PLS-DA important information on the contribution of the masses to the unique properties of the different groups was acquired, which is discussed in Section 3.3.3.

3.3.1.2 Geographical origin

        To examine differences related to the geographical origin of the beans used for the manufacturing of the chocolates, a chemometric analysis was applied to the chocolates' full volatile profiles obtained by HS-PTR-MS. PCA of the 90 chocolate samples revealed a tendency to group chocolates with single-origin (PCA not shown). The unknown samples and those made with mixed beans are spread out. Subsequently, the HS-PTR-MS data of samples of known single-origin were subjected to PCA, i.e. those from Africa (15), Asia (11), and South America (30) (Fig. 3.3). The two Oceania samples were left out due to their low number.

The PCA plot in Figure 3.3 shows three clusters of single-origin, i.e. for samples manufactured with beans from Africa, Asia and South America. A cluster characterised by a mix of samples with different origins is located in the dotted circle in the upper part of the plot. The presence of a mixed zone without particular variables related was underlined by Cambrai et al. (2010), too. The samples highlighted in the mixed zone belong to two distinct chocolate brands (13 chocolates of brand F, and 3 chocolates of brand G) indicating the interaction of the complex formulation and production process with the beans' origin. Nevertheless, the overlapping of the samples can be the result of the complex process of production and the blending techniques that strongly affect the volatile profile of the finished product, leading to a partial loss of the geographical origin features.

        In order to find the maximum separation between the geographical classes (Africa, Asia, South America), supervised statistical analysis (PLS-DA) was applied. A classification model was developed to classify the chocolates according to the origin of the beans. The model was evaluated by leave-5-out cross-validation.

        The results showed that the PLS-DA model with 4 factors for the Asia class and 5 factors for Africa and South America classes, on the normalized and auto-scaled data provided good classification rates with an average efficiency of 93%. The model was developed by applying one orthogonal signal correction (OSC). All the African and Asian samples were correctly classified. Four South American samples were not classified showing a prediction value lower than 0.5. These samples belong to brand F and it can be assumed that the no-classification condition is related to the influence of the brand on the chocolate geographical fingerprint.

        From the PLS-DA important information on the contribution of the masses to the unique properties of the different groups was acquired, which is discussed in Section 3.3.3.
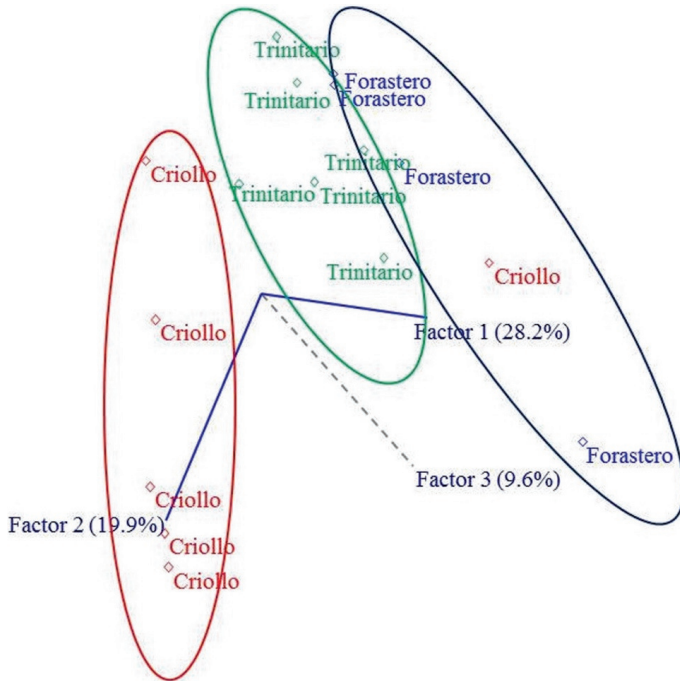
Figure 3.2. 3D PCA plot of normalized and auto-scaled HS-PTR-MS data of chocolates manufactured from specific cocoa bean genetic group.
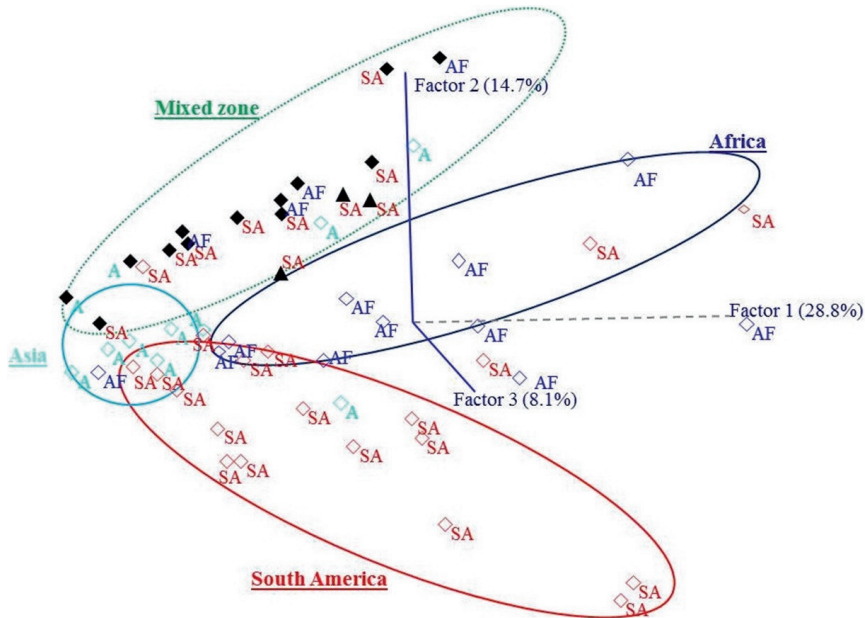
Figure 3.3. 3D PCA of normalized and auto-scaled HS-PTR-MS data of chocolates manufactured from beans from single geographical origins: Africa (AF), Asia (A), and South America (SA). A mixed zone is indicated by the dotted circle. Highlighted chocolates belong to two brands: Brand F (filled rhombus shape) and Brand G (filled triangle shape).

3.3.1.3 Brand: formulation and processing

The characteristics of the 8 brands were examined by PCA. For the brand analysis, significant differences in the sample number are showed for each group. The number of samples varied per group due to availability. The 3D PCA plot (Fig. 3.4) shows the grouping of the samples per brand. Looking at the brand distribution, in the plot, distinct clusters evidence a dissimilarity of brands A, E, F, G, and part of brand D from the others. Differences between the brands can be related to different ways of processing, such as the blend of ingredients used and the emphasis placed on different stages of the manufacturing process such as the temperature and the time of roasting (Hernandez & Rutledge, 1994) and length of conching (Afoakwa et al., 2008). On the other hand, overlapping conditions (brands B, C, H, and part of brand D) could underline a similarity within the production, including ingredients and methods applied. In some cases, possible differences within the same brand are even shown. For example, brand D is clustered in two different groups, brands A and brand E show samples out of the cluster and brand C is characterised by a large variability. The next step should be to investigate the single brand formulation and processing to support these results.

A PLS-DA was performed to classify the chocolate belonging to the same brand. Leave-one-out cross-validation results showed that the PLS-DA model on the normalized and auto-scaled data provided a good classification, with a total of 98% of efficiency. Different factors were used for the different brands: 5 factors for Brand A, 7 for Brand B, 5 for Brand C, 6 for Brand D, 6 for Brand E, 6 for Brand F, 6 for Brand G and 6 for Brand H. The model was developed by applying one OSC. Specifically, this PLS-DA classified all the chocolates in the right class, just one belonging to Brand A was not classified. The higher percentage of prediction, compared to the botanical origin and the geographical, underlines how the processing steps of the single brands have a distinct influence on the final volatile profiles of the chocolates.

From the PLS-DA important information on the contribution of the masses to the unique properties of the different groups was acquired, which is discussed in Section 3.3.3.

3.3.2 Tentative identification using PTR-ToF-MS

From the HS-PTR-MS analyses in Section 3.3.1 four representative chocolate samples, which covered the variation in the sample set, were selected for identification purposes. PTR-ToF-MS analysis was performed on them to tentatively identify the common headspace volatile compounds contributing to the unique traits of the dark chocolates in the sample set. The analysis of mass spectra resulted in extraction of 390 mass peaks in the headspace of chocolate samples. The ions that interfere on the final profile evaluation, O^+_2, NO^+ and water clusters and their isotope, were eliminated from the whole data matrix and the mass peaks with a concentration threshold of more than 1 ppbv were selected. The resulting data set consisted of 100 mass peaks. The accuracy of the PTR-ToF-MS in measuring the exact mass allowed the assignment of sum formulas. The PTR-ToF-MS results combined with literature data permitted a tentative identification of 36 masses/ions (Table 3.3). The majority of the literature related to the volatile compounds of chocolate refers to the key aroma compounds that play an important role in the flavour of beans and chocolate. For this reason, there is more information related to the key odorant compounds than for the non-odour active compounds.

The list of compounds specific for dark chocolate developed by Afoakwa et al. (2008) was taken into account. In order to have a better comparison, masses detected in beans (Diab, Hertz–Schünemann, Streibel, & Zimmermann, 2014; Frauendorfer & Schieberle, 2008) and cocoa powder (Bonvehí, 2005; Frauendorfer & Schieberle, 2006) were also considered. Moreover, several compounds were confirmed comparing these results with masses found in coffee studies applying the same instrument (Özdestan et al., 2013; Yener et al., 2014).

Mass peaks m/z 43.0183 and m/z 61.0284 were detected at saturation levels. They were dominating the headspace; for this reason, it was decided to include them in the list. These protonated masses can be tentatively assigned to a fragment of diverse origin (m/z 43.0183) and acetic acid (m/z 61.0284) (Yener et al., 2014). Specifically, m/z 43 and 61 together, are related to acetic acid (Biasioli et al., 2004; Yeretzian et al., 2003) as already mentioned in Section 3.3.1.
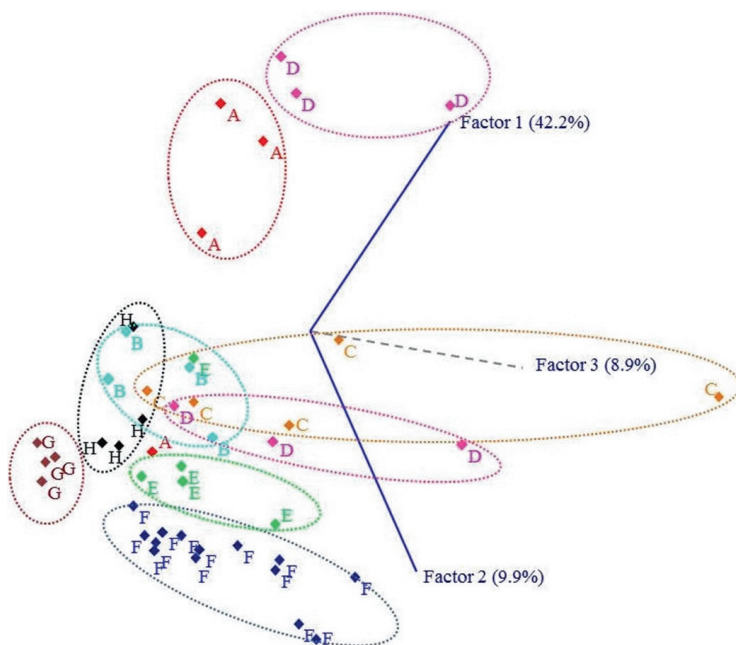
Figure 3.4. 3D PCA plot of normalized and auto-scaled HS-PTR-MS data showing the brand natural clusters.

The PTR-ToF-MS measurements allowed detection of compounds of the main classes of chocolate odorant compounds, in particular, aldehydes and pyrazines, products of Maillard reaction that takes place during the roasting of the cocoa beans. Among these compounds, we can distinguish Strecker aldehydes such as propenal, propanal, isobutanal, methylbutenal, 2/3-methylbutanal, 3-(methylthio) propionaldehyde (methional), benzaldehyde, and phenylacetaldehyde (Counet et al., 2002; Diab et al., 2014). Previous PTR-ToF-MS measurements related to cocoa beans were not able to differentiate methylbutanal and diacetyl, another important product of the non-enzymatic browning reaction (Diab et al., 2014). In this study, it was possible to tentatively assign these compounds to two different mass peaks m/z 87.0438 for diacetyl and 87.0804 for methylbutanal. Pyrroles, pyradine, furans and thiazoles are other heterocyclic compounds developed during Maillard reactions. Between these, it was possible to tentatively identify typical compounds of heated foods such as furan and methylfuran.

Table 3.3 lists several compounds already present in unroasted beans, too. During fermentation, not only aroma precursors such as amino acids, reducing sugars are formed but also volatile compounds such as alcohols, organic acids, aldehydes (Frauendorfer & Schieberle, 2008). Within these compounds, acetic acid, methylpropanoic acid, 2- and 3-methylbutanoic acid, and 2-phenylethanol were tentatively identified. As stated before, pyrazines are usually products of roasting but several studies showed the presence in the unroasted beans, too. For example, tetramethylpyrazine, related to mass peak 137.1049 in Table 3.2, is a typical compound of unroasted beans, synthetized during fermentation (Bonvehí, 2005; Reineccius, Keeney, & Weissberger, 1972). One of the compounds tentatively identified, hexanoic acid, is typical for milk chocolate (Schnermann & Schieberle, 1997) but was tentatively identified in the dark chocolate in the present study as well.

Table 3.3. Tentative identifications, sum formula from PTR-Tof-MS analysis (3 left-hand columns) as well as the indication of the most relevant compounds being impacted by the botanical and geographical origins, and brand, obtained by HS-PTR-MS and PLS-DA (right-hand columns).

	PTR-ToF-MS tentative identification		HS-PTR-MS and PLS-DA correlation Link		
m/z	Tentative identification with References [a]	Sum formula	Botanical origin	Geographical Origin	Brand
29	N.I.		Criollo /Forastero		
30	N.I.				Brand B
31	N.I		Forastero		Brand H
33.0339	Methanol[5]	CH_5O^+			Brand B
34	N.I.				Brand B
35	N.I.				Brand B
39	N.I.		Forastero	Africa	
40	N.I.				Brand C
41.0388	Alkylic fragment [4][5]	$C_3H_5^+$	Forastero	Africa	
43.0183	Fragment (diverse origin)[5]	$C_3H_2O_2^+$		Africa	Brand G
44	N.I.			Africa	Brand G
45.0338	Acetaldehyde [3][5]	$C_2H_5O^+$	Criollo/Trinitario		Brand F
46	N.I		Criollo/Trinitario		Brand F
47	N.I.		Criollo/Trinitario		
48	N.I.		Criollo/Trinitario		
51	N.I.				Brand G
52	N.I.				Brand G
53	N.I.			Africa	
55	N.I.				Brand H
57.0702	Butene/ Propenal [3][5]	$C_4H_9^+$ / $C_3H_5O^+$		Asia/South America	Brand E
58	N.I			Asia	Brand D/E
59.0493	Acetone/ Propanal [3]	$C_3H_7O^+$			
61.0284	Acetic acid [7]	$C_2H_5O_2^+$			Brand B/G
62	N.I.				Brand G
63	N.I.				Brand G
67	N.I.		Forastero	Africa	
68	N.I		Criollo/Trinitario		Brand C/H
69.0701	Furan [3][4][5]	$C_4H_5O^+$			
70	N.I.		Forastero		
71.0855	Terpene fragment[4][5]/ Pentene[3]/ Dihydrofuran[3]	$C_5H_{11}^+$/ $C_4H_7O^+$			
72.0535	Pyrrolidine [3]	$C_4H_{10}N^+$			Brand H
73.0649	Butanone[3][5]/ Isobutanal [1][3][5]	$C_4H_9O^+$			
75.0440	Propanoic acid [2][3]	$C_3H_7O_2^+$			
79	N.I.				Brand G
80.0419	Pyridine [1][2][4][5]	$C_5H_6N^+$			Brand B
81.0357	Pyrazine [1][2]/ Furan fragment [4][5]	$C_5H_5O^+$/ $C_4H_5N_2^+$			
82	N.I.				Brand H
83.0857	Methylfuran [3][4][5]	$C_5H_7O^+$			
84	N.I.				Brand B
85.0647	Methyl-butenal [1][3][4][5]	$C_5H_9O^+$			
85	N.I.				Brand H

Table 3.3. Continued.

	PTR-ToF-MS tentative identification		HS-PTR-MS and PLS-DA correlation Link		
m/z	Tentative identification with References [a]	Sum formula	Botanical origin	Geographical Origin	Brand
87.0438	2,3-Butandione (diacetyl)[1][4][5]	$C_4H_7O_2^+$			
87.0804	3-Methylbutanal[1][3]/ 2-Methylbutanal [1]	$C_5H_{11}O^+$			
89.0597	Methyl-propanoate [5]/2-Methylpropanoic acid [8]	$C_4H_9O_2^+$		South America	
90	N.I.			Asia/South America	
91	N.I.		Trinitario		
93.0688	Methylbenzene (Toluene) [1]	$C_7H_9^+$			Brand E
95.0525	Methylpyrazine [1][2][3][4][5]/ 2-pyridinamine [1]/ Phenol [1][3]	$C_5H_7N_2^+$/ $\underline{C_6H_7O^+}$		Africa	Brand D
97.0280	Furfural (furancarboxaldehyde) [1][2][3][5]	$C_5H_5O_2^+$	Criollo/Trinitario		
98	N.I.		Trinitario		
99	N.I.				Brand E
101.0590	Dihydro-2-methyl-3(2H)-furanone [1][5]/ 2,3 Pentanedione [2][5]	$C_5H_9O_2^+$			
102	N.I.			Asia	Brand E
103.0753	2/3-methylbuttanoic acid [2][4][8]/	$C_5H_{11}O_2^+$		Asia/South America	BrandA/F
104.0783	Benzonitrile [1][2]	$C_7H_6N^+$		Asia/South America	Brand B/E/F
105.0695	3-(methylthio)propionaldehyde (methional) [1][3]/ Styrene [3]	$C_4H_9OS^+$/ $C_8H_9^+$			Brand A/F/G
107.0493	Benzaldehyde [1][2]	$C_7H_7O^+$	Forastero		Brand B/C/D/E/G
107.0819	Ethenylpyrazine [1]/ Ethylbenzene [4]	$C_6H_7N_2^+$/ $C_8H_{11}^+$			
109.0750	2,5-Dimethylpyrazine[1][2]/ Ethylpyrazine[1][2][4][5]/2,3-Dimethylpyrazine[1][2]/ dimethylpyrazine[3][4][5]/ 4-methylphenol[1][2]/ benzyl alcohol [1][3]	$C_6H_9N_2^+$/ $\underline{C_7H_9O^+}$		Asia	Brand B/F
110	N.I.				Brand D
111	N.I.				Brand D
112	N.I.			Asia	Brand D/E/H
113	NI			Asia/South America	Brand D
114	N.I		Forastero/Trinitario		Brand C
115.1114	Heptanal [1]/ 2-heptanone [1][3]	$C_7H_{15}O^+$		South America	
116	N.I.			South America	Brand F
117.0895	Hexanoic acid [2][5][6]	$C_6H_{13}O_2^+$			
120	N.I.			Africa	
121.0615	Phenylacetaldehyde [1][2][3]/ Acetophenone [2][3]/ 2-ethenyl-6-methylpyrazine [1][3][4][[5]	$C_8H_9O^+$/ $\underline{C_7H_9N_2^{\pm}}$	Forastero		Brand A/F
122	N.I.		Forastero		Brand C/F/H
123.0908	2-ethyl-5(or 6)-methylpyrazine[1]/ Trimethylpyrazine[1][2][3][4][5]/ 2-ethyl-3-methylpyrazine [1][2]/ 2 phenylethanol [8]	$C_7H_{11}N_2^+$/ $\underline{C_8H_{11}O^+}$	Criollo	South America	Brand A
124	N.I.			South America	Brand A/F

Table 3.3. Continued.

	PTR-ToF-MS tentative identification		HS-PTR-MS and PLS-DA correlation Link		
m/z	Tentative identification with References [a]	Sum formula	Botanical origin	Geographical Origin	Brand
129	N.I.				Brand D
131.069	Ethyl valerate [2]/ Heptanoic acid [2]	$C_7H_{15}O_2^+$	Criollo		
137.1049	3(or 2),5-dimethyl-2(or 3)-ethylpyrazine [1]/ Tetramethylpyrazine [1][2][3]/ Ethyl-dimethylpyrazine [4][5]	$C_8H_{13}N_2^+$			
133	N.I.				Brand H
135	N.I.				Brand D/F
136	N.I.				Brand A
137	N.I.				Brand A
138	N.I.				Brand F
139	N.I.			Africa	Brand C
140	N.I.			Africa	Brand C/H
141	N.I.				Brand C
142	N.I.		Trinitario		Brand C
143.1422	Nonanal [1]	$C_9H_{19}O^+$			Brand A
147	N.I.				Brand E
151	N.I.				Brand A
156	N.I.				Brand A
157	N.I.				Brand C

Underlined compounds show a different formula.

N.I. (not identified) refers to the compounds not tentatively identified by PTR-ToF-MS.

[a] References: [1] Afoakwa et al. (2008); [2] Bonvehí (2005); [3] Diab et al. (2014); [4] Özdestan et al. (2013); [5] Yener et al. (2014); [6] Schnermann and Schieberle (1997); [7] Frauendorfer and Schieberle (2006); and [8] Frauendorfer and Schieberle (2008).

3.3.3 Cocoa botanical origin, cocoa geographical origin, and brand: characteristic VOCs

PLS-DA analysis (Sections 3.3.1.1, 3.3.1.2, 3.3.1.3) and PTR-ToF-MS tentative identification of the VOCs (Section 3.3.2) allowed a further investigation into the pinpointing of the most relevant compounds, mirroring the cocoa production (cocoa botanical and geographical origin) and chocolate processing factors (brand) in the chocolates. The correlation spectrum obtained from the PLS-DA models with 1 OSC of the auto-scaled and normalized data of HS-PTR-MS, was used to find the most relevant masses for the factors botanical origin, geographical origin, and the brands. The correlation spectrum is the resulting vector of the correlation values obtained comparing the different classes to each nominal mass. Specifically, each class variable is evaluated against the other, thus each mass will be correlated against 0 and 1. This allows the exclusion of poorly correlated masses. Evaluating the results of the correlation spectrum of the 10 highly correlated masses for each variable class, some of the masses related to specific classes were recognised according to the compounds tentatively identified in Table 3.3. Some of those masses are uniquely related to a certain factor, e.g. heptanal/2-heptanone to chocolates with beans originating from South America. However, often an interaction between botanical origin or geographical origin and brand is observed. Further investigations to reduce the interaction of these parameters are needed in chocolate authentication to determine the relevance of the impact of the various factors on the volatile composition of chocolates.

3.4 Conclusions

The current study showed that the VOC profiles of dark chocolates are strongly affected by the brand-related formulation and processing conditions. However, information related to the botanical and geographical origins of the beans is still available even after the complex steps of chocolate production. Further investigations to reduce the interaction of botanical origin/geographical origin/brand are needed to determine the relevance of each factor on the volatile composition of chocolates.

Acknowledgements

References

Aculey, P. C., Snitkjaer, P., Owusu, M., Bassompiere, M., Takrama, J., Nørgaard, L., ... Nielsen, D. S. (2010). Ghanaian cocoa bean fermentation characterized by spectroscopic and chromatographic methods and chemometrics. *Journal of Food Science, 75*(6), S300–S307.

Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2008). Flavor formation and character in cocoa and chocolate: A critical review. Critical Reviews in *Food Science and Nutrition, 48*(9), 840–857.

Araghipour, N., Colineau, J., Koot, A., Akkermans, W., Rojas, J. M. M., Beauchamp, J., ... van Ruth, S. M. (2008). Geographical origin classification of olive oils by PTR-MS. *Food Chemistry, 108*(1), 374–383.

Biasioli, F., Gasperi, F., Aprea, E., Endrizzi, I., Framondino, V., Marini, F., ... Märk, T. D. (2006). Correlation of PTR-MS spectral fingerprints with sensory characterisation of flavour and odour profile of "Trentingrana" cheese. *Food Quality and Preference, 17*(1–2), 63–75.

Biasioli, F., Gasperi, F., Odorizzi, G., Aprea, E., Mott, D., Marini, F., ... Tilmann, D. M. (2004). PTR-MS monitoring of odour emissions from composting plants. *International Journal of Mass Spectrometry, 239*(2–3), 103–109.

Bonvehí, J. S. (2005). Investigation of aromatic compounds in roasted cocoa powder. *European Food Research and Technology, 221*(1–2), 19–29.

Bossier, P. (1999). Authentication of seafood products by DNA patterns. *Journal of Food Science, 64*(2), 189–193.

Caligiani, A., Cirlini, M., Palla, G., Ravaglia, R., & Arlorio, M. (2007). GC–MS detection of chiral markers in cocoa beans of different quality and geographic origin. *Chirality, 19*(4), 329–334.

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*(3), 1478–1483.

Camu, N., De Winter, T., Addo, S. K., Takrama, J. S., Bernaert, H., & De Vuyst, L. (2008). Fer- mentation of cocoa beans: Influence of microbial activities and polyphenol concentrations on the flavour of chocolate. *Journal of the Science of Food and Agriculture, 88*(13), 2288–2297.

Cappellin, L., Biasioli, F., Fabris, A., Schuhfried, E., Soukoulis, C., Märk, T. D., & Gasperi, F. (2010). Improved mass accuracy in PTR-TOF-MS: Another step towards better compound identification in PTR-MS. *International Journal of Mass Spectrometry, 290,* 60–63.

Cappellin, L., Biasioli, F., Granitto, P. M., Schuhfried, E., Soukoulis, C., Costa, F., ... Gasperi, F. (2011). On data analysis in PTR-TOF-MS: From raw spectra to data mining. Sensors and Actuators B: *Chemical, 155*(1), 183–190.

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

Cheman, Y. (2005). Analysis of potential lard adulteration in chocolate and chocolate products using Fourier transform infrared spectroscopy. *Food Chemistry, 90*(4), 815–819.

Counet, C., Callemien, D., Ouwerx, C., & Collin, S. (2002). Use of gas chromatography–olfactometry to identify key odorant compounds in dark chocolate. Comparison of samples before and after conching. *Journal of Agricultural and Food Chemistry, 50*(8), 2385–2391.

Dennis, M. J. (1998). Recent developments in food authentication †. *Analyst, 123*(9), 151–156.

Diab, J., Hertz–Schünemann, R., Streibel, T., & Zimmermann, R. (2014). On-line measurement of volatile organic compounds released during roasting of cocoa beans. *Food Research International, 63,* 344–352.

Dionisi, F., Golay, P. -A., Hug, B., Baumgartner, M., Callier, P., & Destaillats, F. (2004). Triacylglycerol analysis for the quantification of cocoa butter equivalents (CBE) in chocolate: Feasibility study and validation. *Journal of Agricultural and Food Chemistry, 52*(7), 1835–1841.

Downey, G., Briandet, R., Wilson, R. H., & Kemsley, E. K. (1997). Near- and mid-infrared spectroscopies in food authentication: Coffee varietal identification. *Journal of Agricultural and Food Chemistry, 45*(11), 4357–4361.

Fajardo, V., González, I., Rojas, M., García, T., & Martín, R. (2010). A review of current PCR- based methodologies for the authentication of meats from game animal species. *Trends in Food Science & Technology, 21*(8), 408–421.

Farah, D. M. H., Zaibunnisa, A. H., Misnawi, J., & Zainal, S. (2012). Effect of roasting process on the concentration of acrylamide and pyrizines in roasted cocoa beans from different origins. *APCBEE Procedia, 4,* 204–208.

Frauendorfer, F., & Schieberle, P. (2006). Identification of the key aroma compounds in cocoa powder based on molecular sensory correlations. *Journal of Agricultural and Food Chemistry, 54,* 5521–5529.

Frauendorfer, F., & Schieberle, P. (2008). Changes in key aroma compounds of Criollo cocoa beans during roasting. *Journal of Agricultural and Food Chemistry, 56*(21), 10244–10251.

Granato, D., Katayama, F. C. U., & de Castro, I. A. (2011). Phenolic composition of South American red wines classified according to their antioxidant activity, retail price and sensory quality. *Food Chemistry, 129*(2), 366–373.

Hernandez, V., & Rutledge, D. N. (1994). Multivariate statistical analysis of gas chromato- grams to differentiate cocoa masses by geographical origin and roasting conditions. *Analyst, 119*(6), 1171–1176.

Hernandez, B., Castellote, A. I., & Permanyer, J. J. (1991). Triglyceride analysis of cocoa beans from different geographical origins. *Food Chemistry, 41*(3), 269–276.

Holm, C. S., Astong, J. W., & Douglas, K. (1993). The effects of the organic acids in cocoa on the flavour of chocolate*. *Journal of the Science of Food and Agriculture, 61*(1), 65–71.

Iora, S. R. F., Maciel, G. M., Zielinski, A. A. F., da Silva, M. V., Pontes, P. V. D. A., Haminiuk, C. W. I., & Granato, D. (2015). Evaluation of the bioactive compounds and the antioxidant capacity of grape pomace. *International Journal of Food Science & Technology, 50*(1), 62–69.

Jinap, S., Dimick, P. S., & Hollender, R. (1995). Flavour evaluation of chocolate formulated from cocoa beans from different countries. *Food Control, 6*(2), 105–110.

Jinap, S. (1991). Effect of roasting on acidic characteristics of cocoa beans. *Journal of the Science of Food and Agriculture, 54*(2), 317–321.

Jinap, S., & Dimick, P. S. (1990). Acidic characteristics of fermented and dried cocoa beans from different countries of origin. *Journal of Food Science, 55*(2), 547–550.

Lindinger, W., Hansel, A., & Jordan, A. (1998). Proton-transfer-reaction mass spectrometry (PTR-MS): On-line monitoring of volatile organic compounds at pptv levels. *Chemical Society Reviews, 27*(5), 347–375.

Liu, L., Cozzolino, D., Cynkar, W. U., Gishen, M., & Colby, C. B. (2006). Geographic classification of spanish and Australian tempranillo red wines by visible and near-infrared spectroscopy combined with multivariate analysis. *Journal of Agricultural and Food Chemistry, 54*(18), 6754–6759.

Luykx, D. M. A. M., & van Ruth, S. M. (2008). An overview of analytical methods for determining the geographical origin of food products. *Food Chemistry, 107*(2), 897–911.

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

Reineccius, G. A., Keeney, P. G., & Weissberger, W. (1972). Factors affecting the concentration of pyrazines in cocoa beans. *Journal of Agricultural and Food Chemistry, 20*(2), 202–206.

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

Saltini, R., Akkerman, R., & Frosch, S. (2013). Optimizing chocolate production through traceability: A review of the influence of farming practices on cocoa bean quality. *Food Control, 29*(1), 167–187.

Schnermann, P., & Schieberle, P. (1997). Evaluation of key odorants in milk chocolate and cocoa mass by aroma extract dilution analyses. *Journal of Agricultural and Food Chemistry, 45*(3), 867–872.

Spaniolas, S., May, S. T., Bennett, M. J., & Tucker, G. A. (2006). Authentication of coffee by means of PCR-RFLP analysis and lab-on-a-chip capillary electrophoresis. *Journal of Agricultural and Food Chemistry, 54*(20), 7466–7470.

Tres, A., & van Ruth, S. M. (2011). Verification of organic feed identity by fatty acid finger-printing. *Journal of Agricultural and Food Chemistry, 59*(16), 8816–8821.

Ulberth, F., & Buchgraber, M. (2003). Analytical platforms to assess the authenticity of cocoa butter. *European Journal of Lipid Science and Technology, 105*(1), 32–42.

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

Yeretzian, C., Jordan, A., & Lindinger, W. (2003). Analysing the headspace of coffee by proton-transfer-reaction mass-spectrometry. *International Journal of Mass Spectrometry, 224,* 115–139.

Zielinski, A. A. F., Haminiuk, C. W. I., Nunes, C. A., Schnitzler, E., van Ruth, S. M., & Granato, D. (2014). Chemical composition, sensory properties, provenance, and bioactivity of fruit juices as assessed by chemometrics: A critical review and guideline. Comprehensive Reviews in *Food Science and Food Safety, 13*(3), 300–316.

Chapter 4

Making cocoa origin traceable: Fingerprints of chocolates using Flow Infusion-Electro Spray Ionization-Mass Spectrometry

Abstract

   More and more attention is paid to sustainability in cocoa production. Tools that assist in making sustainable cocoa traceable are therefore welcome. In the present study, the applicability of Flow Infusion-Electrospray Ionization-Mass Spectrometry (FI-ESI-MS) to assess the geographical origin of cocoa beans used to produce chocolate has been tested. Fifty-seven dark chocolates available on the Dutch market were investigated taking into account the geographical origin of the beans: they originated from Africa (15), Asia (11) and South America (31). The FI-ESI-MS generated non-volatile profiles, were subjected to ANOVA selection (21 masses per sample). Those were subsequently used as fingerprints and investigated using chemometric tools to elucidate information on raw material origin. The Principal Component Analysis revealed a remarkable separation between African and Asian bean chocolates. No clear trend was observed for the chocolates manufactured from South American beans. The kNN results confirmed the possibility to separate African and Asian chocolates as well as these two continents together versus South America. The inability to separate the three continents at the same time and the weak differentiation of South American samples from the others is linked to brand-related factors, such as the formulation and industrial processing. Although diminishing of the influence of the brand-related factors will improve separation, the current approach is an important step in the geographical assessment of the raw material and the traceability of cocoa in its supply chain.

Keywords Chocolate, FI-ESI-MS, Fingerprint, Origin, Traceability

4.1 Introduction

When bulk products are involved in food manufacturing, traceability requires optimisation in terms of control strategies and analytical approaches which can verify the origin and characteristics of the raw materials. In the case of chocolate, the raw cocoa - through complex production lines - undergoes several transformations to create one or more final products. For instance, cocoa can be mixed with intermediate materials, such as cocoa mass, and transformed through physical/chemical processes. Currently, there are no traceability practices able to deal with these kinds of products (Comba, Belforte, Dabbene, & Gay, 2013; Dabbene, Gay, & Tortia, 2014).

As a result of growing concern regarding food quality and safety, an increasing effort in the implementation of traceability systems can be seen in recent years (Saltini, Akkerman, & Frosch, 2013). Traceability represents added value for consumers; for food producers, it is a tool to avoid damage to market share as well as guaranteeing that policy requirements are met (Dabbene et al., 2014; Hobbs, 2004).

The complexity of chocolate traceability can be linked to several aspects of the cocoa/chocolate supply chain, such as the length and number of actors involved in the trade network. While in general cocoa beans are produced in the Southern hemisphere, chocolate is predominantly manufactured in the Northern countries. The regulations applied in the EU countries are not applied at the beginning of the supply chain (Saltini et al., 2013); this explains the lack of data exchange between farmers and chocolate manufacturers (Saltini & Akkerman, 2012) that could result in potentially opportunist behaviour among the various actors for individual economic benefit.

Being able to extract information related to the raw material in a finished chocolate bar provides an important step in assessing the cocoa geographical origin and in tracing back through the supply chain. Improving the traceability ensures transparency along the supply chain and could also contribute to the sustainability of the product (Wognum, Bremmers, Trienekens, Van Der Vorst, & Bloemhof, 2011). The Dutch government, being the Netherlands one of the biggest cocoa processors and importers in the world, has committed to 100% certified sustainable cocoa by 2025 (Government of the Netherlands, 2014; CBI, 2016; Squicciarini & Swinnen, 2016). For these reasons, an enhancement in chocolate traceability is required.

Several analytical techniques have been used to extract information from the raw material in chocolate. The triglyceride profiles of cocoa butter were studied to detect the presence of any other vegetable butter (Simoneau, Hannaert, & Anklam, 1999; Ulberth & Buchgraber, 2003). Other researchers reported a link between raw material and finished product comparing cocoa beans and corresponding cocoa mass using gas chromatography (Hernandez & Rutledge, 1994), or comparing the near-infrared (NIR) spectra of raw and roasted beans, chocolate mass and finished chocolate (Davies et al., 1991). Nevertheless, these approaches were focused on quality and not directly on traceability. Recently, an increasing interest in investigating techniques able to support cocoa/chocolate traceability has been shown. NIR spectroscopy was tested for geographical discrimination of cocoa beans (Teye, Huang, Dai, & Chen, 2013); Proton Transfer Reaction-Mass Spectrometry was applied to link the volatile compounds of a chocolate bar with the geographical origin of cocoa beans used (Acierno, Yener, Alewijn, Biasioli, & van Ruth, 2016; Cambrai et al., 2010), and a method to trace back the final product was investigated comparing the multi-elemental composition of cocoa beans and chocolate by Inductively Coupled Plasma-Mass Spectrometry (Bertoldi, Barbero, Camin, Caligiani, & Larcher, 2016).

Traditional food quality and safety investigations are based on the determination of certain markers. A more comprehensive insight into the sample composition is possible by the extension of the analysis to a larger range of compounds. Recent research is focused on non-targeted analytical techniques able to produce a product fingerprint. This approach can be particularly useful in detection of attributes that are not necessarily correlated with quality but can be linked to traceability attributes such as country of origin (Luykx & van Ruth, 2008), organic (Tres, O'Neill, & van Ruth, 2011), kosher, halal (Rohman & Che Man, 2012), or "free-range" livestock (Alonso et al., 2008).

In this study, a non-targeted analytical approach was applied to the non-volatile compounds composition of 57 dark chocolates by Flow Infusion-Electrospray Ionization-Mass Spectrometry (FI-ESI-MS). FI-ESI-MS fingerprint has already been used to detect adulteration in different foods such as olive oil (Junia de O. Alves, Waldomiro B. Neto, Hery Mitsutake, Paulo S. P. Alves, 2010) and ham (González-Dominguez, García-Barrera, & Gõmez-Ariza, 2012). This approach can scan a large range of masses, including non- volatile and thermally labile compounds (Ho et al., 2003) and can be an alternative to laborious and time-consuming analytical procedures employing prior chromatographic separation of analytes (Draper, Lloyd, Goodacre, & Beckmann, 2012). ESI is able to detect the molecular mass of large molecules such as proteins, peptides, polysaccharides, triacylglycerides (Capuano & van Ruth, 2012). In the specific case of chocolate, ESI has been largely applied to study fatty acids (Perret, Gentili, Marchese, Sergi, & Caporossi, 2004), amino acids (Pätzold & Brückner, 2006) and polyphenols (Arlorio et al., 2008). The profile of these compounds depends on the characteristics of the cocoa beans such as cocoa genotype, agro-climatic condition of growing and cocoa geographical origins (Afoakwa, Paterson, Fowler, & Ryan, 2008; Brito et al., 2001; Caligiani, Cirlini, Palla, Ravaglia, & Arlorio, 2007; Camu et al., 2008; Rohsius, Matissek, & Lieberei, 2006). These differences in chemical composition are reflected in the quality characteristics of the final product (Counet, Ouwerx, Rosoux, & Collin, 2004; Jinap, Dimick, & Hollender, 1995; Saltini et al., 2013; Sukha, Butler, Umaharan, & Boult, 2007).

The aim of this study was to test the applicability of FI-ESI-MS to extract information on the origin of cocoa beans to find a link between raw material and final product when tracing back the cocoa/chocolate supply chain.

4.2 Materials and methods

4.2.1 Materials

Fifty-seven chocolates labelled with specific geographical origins were collected from the Dutch market. The set of samples comprised 15 chocolates made with beans from Africa (Af), 11 made with beans from Asia (As) and 31 made with beans from South America (SA). Each geographical group (Af, As, SA) included chocolates from different countries and different brands. All the information given by the brand such as bean geographical origin, bean botanical variety, cocoa percentage, and ingredient composition were collected. All chocolate was bought during May and June 2013 and stored at room temperature until the analyses in July 2013.

4.2.2 Methods

4.2.2.1 Sample preparation

Chocolate samples were ground using an electrical food grater and were stored at 4°C during the experiment. Five-hundred mg of chocolate sample was extracted in 5 ml of 100% ethanol (Emsure, Darmstadt, Germany). Samples were agitated head-over-head at 20°C for 1 h and centrifuged at 1500 rpm for 5 min at 4°C. The clear layers were collected and kept at 4°C for ~24 h. Afterwards, they were re-centrifuged at 1500 rpm for 5 min at 4°C. The new clear layers were collected and diluted 100-fold in 100% ethanol (Emsure, Darmstadt, Germany). Five-hundred ml of the resulting solution was filtered using a Single StEP Filter Vial (Thomson Instrument Company, Oceanside, CA, USA) equipped with a 0.45 mm filter base, and 1.0 ml was injected into the mass spectrometer.

4.2.2.2 FI-ESI-MS equipment

A single-stage Orbitrap Mass Spectrometer (Thermo Scientific, Bremen, Germany) was coupled to an Accela LC system (Thermo Scientific, Bremen, Germany). The injection part of the LC system auto sampler was connected directly to the HESI-II electrospray source of the mass spectrometer using ~75 cm PEEK tubing with an internal diameter of 0.13 mm. The samples were transferred to the mass spectrometer using a mixture of methanol/water/formic acid (90:10:0.1) at a flow of 200 ml/min. The HESI-II electrospray source was used in positive ionization mode with the following parameters: Spray voltage 3.5 kV; Sheath gas 15 arbitrary units (au); Auxiliary Gas 6 au; Heater Temp. 375°C; Capillary Temp. 300°C. The MS was used at a resolving power of 105, resulting in a scan rate of 1 Hz.

Mass spectra in the range of m/z 75e1500 were recorded. To minimize sample carry-over, a blank solvent was injected after every sample. Each sample was injected twice.

4.2.2.3 FI-ESI-MS data conversion

The raw data produced by the mass spectrometer were converted to the NetCDF format using the file converter build in the Xcalibur software (version 3.0, Thermo Scientific). Xcalibur was also used to control the mass spectrometer.

The NetCDF-files were then converted by in-house adaptation of the XCMS package using R 3.2.2 (the R Foundation for Statistical Computing). Peaks were found within each sample and the files pre-sorted into different classes suitable for grouping; the peak groups were generated taking into account an error of 5 ppm for the full mass range. In this phase, the signal-to-noise cut off was defined. Once the groups were created, peaks were aligned across all samples using overlapping m/z bins and calculating the smoothed peak distribution. After peak grouping, intensity values were produced in the peak groups that did not include peaks from every sample. For those missing samples, raw data are integrated into the peak group region.

4.2.3 Univariate analysis

4.2.3.1 ANOVA

The mass peaks were subjected to one-way ANOVA (IBM SPSS Statistics 23.0, IBM Corp., Armonk, NY, USA) to investigate and select masses showing statistical differences according to the geographical origins (Af, As, SA) of the beans used to produce the chocolates. For the entire data set mass normality was checked by Q-Q plot visualisation and homogeneity of variance by Levene's test. Differences between groups were assessed by one-way ANOVA, followed by Fisher's least significance difference (LSD) test. For the variables that presented non-homogenous variances, a robust test of equality of means (Welch's t-test) was performed to confirm the significant difference followed by a Games-Howell post-hoc test. A significance level of $p \leq 0.05$ was used throughout the study.

4.2.4 Multivariate analysis

Multivariate analysis was conducted using Pirouette 4.0 Software (Infometrix, Seattle, WA, USA) and R 3.2.2 (the R Foundation for Statistical Computing).

4.2.4.1 Principal Component Analysis (PCA)

PCA was used to graphically visualise the presence of any natural clustering in the data samples according to geographical origins of the beans. Prior to PCA, the data were auto-scaled to compensate for the different magnitudes of the variables.

4.2.4.2 k-Nearest Neighbour (kNN)

kNN was used to classify the samples according to the geographical origin (Af, As, SA).

kNN is a supervised pattern recognition approach, meaning that samples are assigned to a category. This method is based on the assumption that the closer the samples, the more likely they belong to the same category. First, a model is built and refined based on a training set, and later used to predict classes of new samples.

In this study, kNN was selected as it is robust and well-suited to a limited number of samples. The prediction of an unknown depends on the class of its k nearest neighbours; therefore an important part of the process was to determine an appropriate value for k. To test the robustness of the model a repeated random cross-validation was performed using R 3.2.2 (the R Foundation for Statistical Computing). A random selection of 80% of the samples was used to build the classification model while the other 20% was predicted; this random cross-validation was repeated 1000 times. A confusion matrix was built based on the 1000 random cross-validation. The rate of correct classification of the model was calculated and the single class (Af, As, SA) performance was taken into consideration.

4.2.5 Mass identification

A tentative formula was assigned to the masses selected by ANOVA using Thermo Scientific™ Xcalibur™ software. The formula was generated taking into account a mass error of 5 ppm and that the measured compounds are most likely protonated. Considering the typical chocolate compounds, the elements used to generate the sum formula were N in a range between 0 and 10, O in a range between 0 and 15, C in a range between 0 and 60, H in a range between 0 and 80. For the calculated formula with the smallest mass error, a compound was assigned using online databases such as ChemSpider (http://www.chemspider.com/Chemical-Structure. 1906.html) and taking into account literature on the compounds usually found in chocolate.

4.3 Results and discussion

The 57 dark chocolate samples were subjected to FI-ESI-MS analysis. The resulting FI-ESI-MS data were subsequently aligned and led to 305 masses in the range from 79 to 911 m/z that were detected and matched across the 57 samples. A total profile is shown in Figure 4.1.

4.3.1 Univariate analysis

4.3.1.1 Relevant masses for geographical origin – ANOVA

The final data set (57 samples*305 masses) was subjected to one-way ANOVA to detect statistically significant differences in mass intensities with respect to the origins of the beans of the chocolates. According to normal Q-Q plots, the masses were approximately normally distributed across all samples. The analysis of variance demonstrated that 21 peaks show significant differences according to the origin of the beans used to manufacturing the chocolates. A Welch's t-test was needed to confirm the significance of the difference of three masses as they did not show homogeneity of variance. ANOVA results are shown in Table 4.1.

Taking into account that we wanted to investigate potential distinctions between the three continents, seven comparisons were possible. For instance, it was possible to compare a continent versus the other two, a continent versus another and the three continents at the same time. The selected masses allowed five of the seven combinations. Evaluating the ANOVA and post-hoc results (Table 4.1), we distinguished three main groups of masses: masses differentiating Africa versus the other two continents (Table 4.1, comparison 1), masses differentiating Asia versus the other two continents (Table 4.1, comparison 2), and masses differentiating Africa versus South America (Table 4.1, comparison 5). Furthermore, ANOVA results highlighted one mass differentiating South America versus the other two continents and one mass differentiating Asia versus Africa (Table 4.1 comparison 3 and 4, respectively). No significant differences between Asia and South America were observed. Although several masses showed significant differences for the three geographical origins, none of them was able to statistically differentiate the three continents at the same time, according to the post-hoc tests.
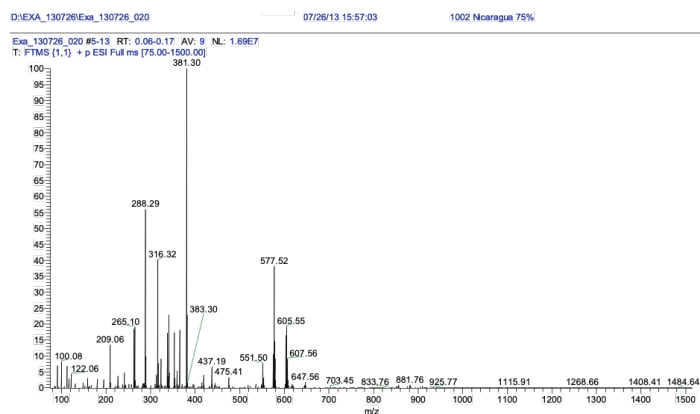
Figure 4.1. An example of a FI-ESI-MS spectrum of chocolate manufactured with beans from Africa.

Table 4.1. List of the FI-ESI-MS masses showing significant differences in intensities between chocolates manufactured with cocoa beans from different geographical origins (groups average intensity ± SD).

m/z	South America	Africa	Asia	p-value	Comparison
	Average intensity ±SD (x 10E+4)				
86.0966	10.9± 5.7b	6.6± 4.0a	11.1± 4.6b	0.021	1
132.1018	58.2± 29.9b	33.4± 18.4a	63.3± 20.9b	0.005	1
**166.0860	19.2± 9.4b	11.3± 4.6a	22.2± 8.9b	0.003	1
196.0908	24.7± 10.0b	17.5± 11.1a	27.9± 6.0b	0.020	1
197.1283	9.1± 6.3b	8.2± 7.1b	14.6± 5.8a	0.033	2
299.2576	9.5± 5.0a	13.7± 5.0b	10.2± 4.0ab	0.028	5
321.2395	14.3± 12.0a	24.8± 14.8b	16.6± 9.9ab	0.032	5
**413.3757	14.2± 9.1a	31.8± 22.8b	18.3± 13.8a	<0.001	1
496.3396	21.8± 6.8a	17.2± 7.6b	26.7± 8.3a	0.007	1
497.3431	5.1± 3.2a	2.6± 2.9b	7.0± 3.1a	0.003	1
524.3703	9.1± 4.8a	6.4± 4.2a	13.3± 5.1b	0.002	2
561.4870	4.4± 3.6a	9.1± 4.4b	6.2± 4.0ab	<0.001	5
563.5028	16.3± 8.9a	29.0± 12.0b	22.8± 9.9ab	<0.001	5
565.5189	9.8± 5.9a	14.5± 6.8b	14.6± 6.7b	0.026	3
591.4968	35.2± 10.1a	44.5± 11.7b	42.7± 13.0ab	0.019	5
592.4998	10.8± 6.9a	11.2± 9.0a	17.1± 5. 4b	0.048	2
617.5108	113.9± 33.6a	115.0± 25.4a	155.1± 58.3b	0.009	2
618.5143	40.7± 11.0a	42.5± 8.7a	53.4± 19.2b	0.019	2
639.4069	5.9± 3.6a	5.3± 4.0a	9.5± 3.3b	0.012	2
645.5424	77.6± 20.7a	74.9± 26.1a	113.2± 45.4b	<0.001	2
**853.7316	48.0± 18.1ab	30.8± 24.4a	53.5± 4.8b	0.004	4

Comparisons: 1) Af – As&SA; 2) A – Af&SA; 3) SA –Af&As; 4) Af – As; 5) Af- SA.
Af: Africa; As: Asia; SA: South America. Light blue colour shows which origin(s) are discriminated, the dark blue colour is the origin(s) to be compared with.
*Different superscripts indicate significant differences (ANOVA and Fisher's LSD-test, p < 0.05)
**Welch robust test of equality of means and Games Howell post hoc test were applied.

4.3.1.2 Relevant masses for geographical origin – tentative identification

Some of the discriminating masses were tentatively identified (Table 4.2). They include free amino acids (2e3), sterols (5) and diacylglycerols (8). It was possible to identify fatty acids but in a dimer form (6e7); formation of these compounds can be instrument-dependent (Pan, 2008).

Table 4.2. Sum formula and tentative identifications of the FI-ESI-MS masses selected by ANOVA in Table 4.1.

	m/z	Sum formula	Tentative identification with reference [a]
1	86.0966	$C_5H_{12}N^b$	Piperidine [2,3]
2	132.1018	$C_6H_{14}NO_2{}^b$	Isoleucine [4]
3	166.0860	$C_9H_{12}NO_2{}^b$	Phenylalanine [4]
4	299.2576	$C_{18}H_{35}O_3{}^b$	Ricinoleic acid [1]
5	413.3757	$C_{29}H_{49}O^b$	Stigmasterol [4]
6	561.4870	$C_{36}H_{65}O_4{}^b$	Dilinoleic acid
7	565.5189	$C_{36}H_{69}O_4{}^b$	Oleic acid, Dimer
8	591.4968	$C_{37}H_{67}O_5{}^b$	Diacylglycerol [5]

[a] References: [1] (Bódalo-Santoyo et al., 2005); [2] (Diab et al., 2014); [3] (Neurath, Dünger, Pein, Ambrosius, 1977); [4] (Nolvachai et al., 2015); [5] (Marty-Terrade & Marangoni, 2012).

Within the tentatively identified compounds, it is possible to recognise free amino acids and fatty acids that can be used to distinguish the geographical origin of the cocoa beans. This is in line with other studies, which reported the geographical influences on the free amino acid concentration in raw cocoa (Rohsius et al., 2006), and the fatty acid profile of chocolate (Torres-Moreno, Torrescasana, Salas-Salvadó, & Blanch, 2015) from different origins. On the other hand, there is no information about geographical differences and the other tentatively identified compounds, such as stigmasterol, piperidine, and diacylglicerol. Nevertheless, they could still represent a link with the raw material. For instance, stigmasterol is a plant sterol that can be found in the cocoa shell and cocoa butter (El-Saied, Morsi, & Amer, 1981; Lipp & Adam, 1998). Piperidine has been found in cocoa butter, plant origin food (Neurath, Dünger, Pein, Ambrosius, 1977), and in defatted cocoa beans (Diab, HertzeSchünemann, Streibel, & Zimmermann, 2014). Furthermore, diacylglycerol is one of the minor components of cocoa butter and the concentration of this compound can change during the industrial process; it can increase during alkalization or react differently according to the deodorization process (Marty-Terrade & Marangoni, 2012).

These findings are important to highlight that the non-volatile components of chocolate can be linked with the raw material although they can be influenced by the industrial processes such as roasting and mixing of ingredients. For instance, sugar and amino acids are used at certain stages of cocoa bean processing and this can make the evaluation of these compounds in the finished product difficult (Pätzold & Brückner, 2006).

A compound tentatively identified in this study that can be directly linked with the industrial production is ricinoleic acid (4), an emulsifier used to improve the viscosity of chocolate (Bódalo-Santoyo, Bastida-Rodríguez, Máximo-Martín, Montiel-Morte, & Murcia-Almagro, 2005).

4.3.2 Multivariate analysis

The non-volatile fingerprints of chocolate generated by FI-ESI- MS comprises information on various aspects, such as the raw materials but also the processing etc. (paragraph 4.3.1.2). In order to elucidate these signatures further, multivariate data analysis was carried out. The masses showing significant differences by ANOVA were used to create a new matrix for the multivariate data analysis (57 samples * 21 masses). In this study, a variable selection was preferred because the number of objects is relatively small compared to the number of masses and many of these could contain noisy information (Berrueta, Alonso-Salces, & Héberger, 2007).

4.3.2.1 Exploration of the data: principal component analysis

4.3.2.1.1 Cocoa geographical origin

The final data matrix (57*21) was auto-scaled to balance the variables and a PCA was performed to explore the data set; three principal components (PCs) were extracted. The score plots corresponding to PC1 versus PC2, and PC2 versus PC3 are shown in Figures 4.2 and 4.3, respectively. Using the ANOVA mass selection for PCA, a trend in the distribution of the samples can be visualised that was not shown by using the total FI- ESI-MS fingerprint (data not shown).

As presented in Figure 4.2a, the first two principal components display a separation between chocolates manufactured from African and Asian beans, spread in the opposite sides of the plot, while no clear trend is observed for the chocolates made with South American beans. The geographical information, particularly of African and Asian samples, has a higher impact on PC2 than on PC1. The distribution of the chocolates and the missing geographical information on PC1 can be explained by the impact of other parameters on this component. The assessment of the same PCA (Fig. 4.2a), taking into account chocolates belonging to the same brand (Fig. 4.2b), reveals a trend in chocolate distribution. Brands represented by at least two samples were displayed with a total of ten different brands. In this latter case, PC1 seems to be more affected by brand-related characteristics than by geographical origin. The signature of the brand originates from the raw materials but also from a similarity in formulation, processing, packaging, etc. A link between chocolates belonging to the same brand is assumed looking at Figure 4.2b; this chocolate distribution can be explained by the presence of brand (production process)-related masses (paragraph 3.1.2).

From these results, it is possible to deduce bean origin information from the non-volatile fingerprint but it is necessary to take into account the influence of the industrial processing as well. This is in agreement with a previous study showing the volatile profiles of chocolates affected by the brand-related formulation and processing conditions (Acierno et al., 2016).

Investigating further dimensions of the PCA, three clusters related to geographical origin are defined. In Figure 4.3a the score plot PC2 versus PC3 shows an improved geographical grouping compared to the previous PCA (Fig. 4.2a). It appears that the geographical information of the cocoa beans is mostly expressed in the higher dimensions.

The PCA loadings indicate which variables are most important and which contribute little to each PC. Figure 4.4 shows the loadings plot related to the PCA comparing PC2 and PC3 as it shows an improvement in the distribution of samples according to the geographical origin. Masses differentiating African and Asian chocolates, according to ANOVA selection, are in the opposite quadrants of the PCA. A group of masses are able to differentiate Africa versus the other two continents are close together in the upper right side of the plot indicating no influence for this group. This is due to these masses having a lower average intensity for African samples compared to the other two continents (Table 4.1). However, they can be considered markers for African chocolate because of the low intensity for this group.
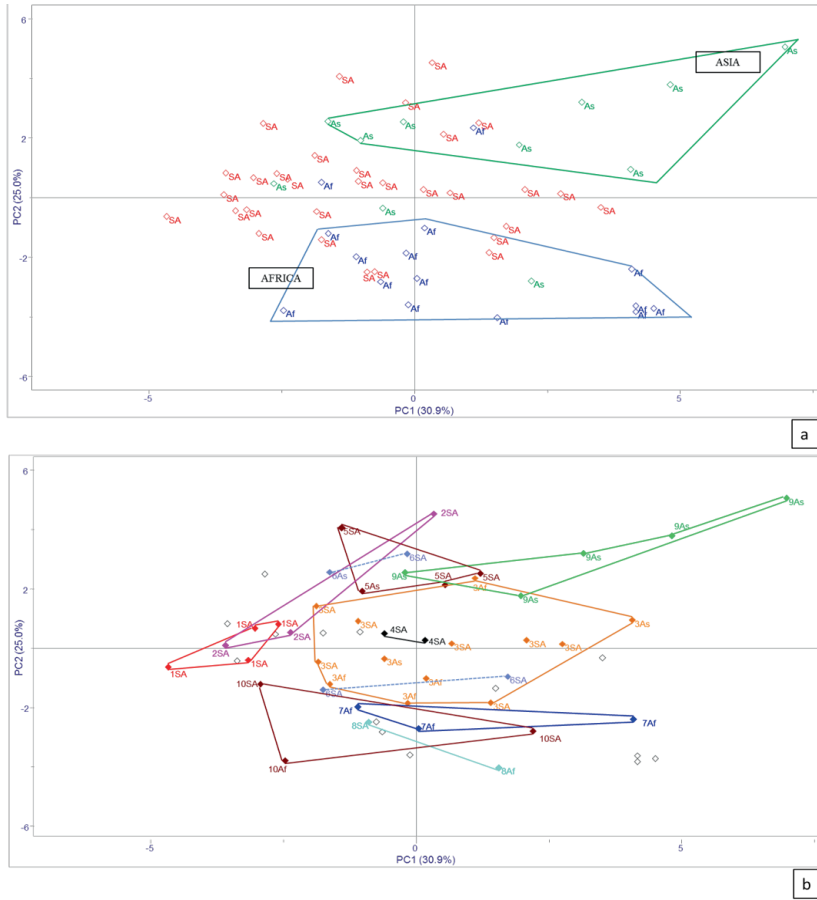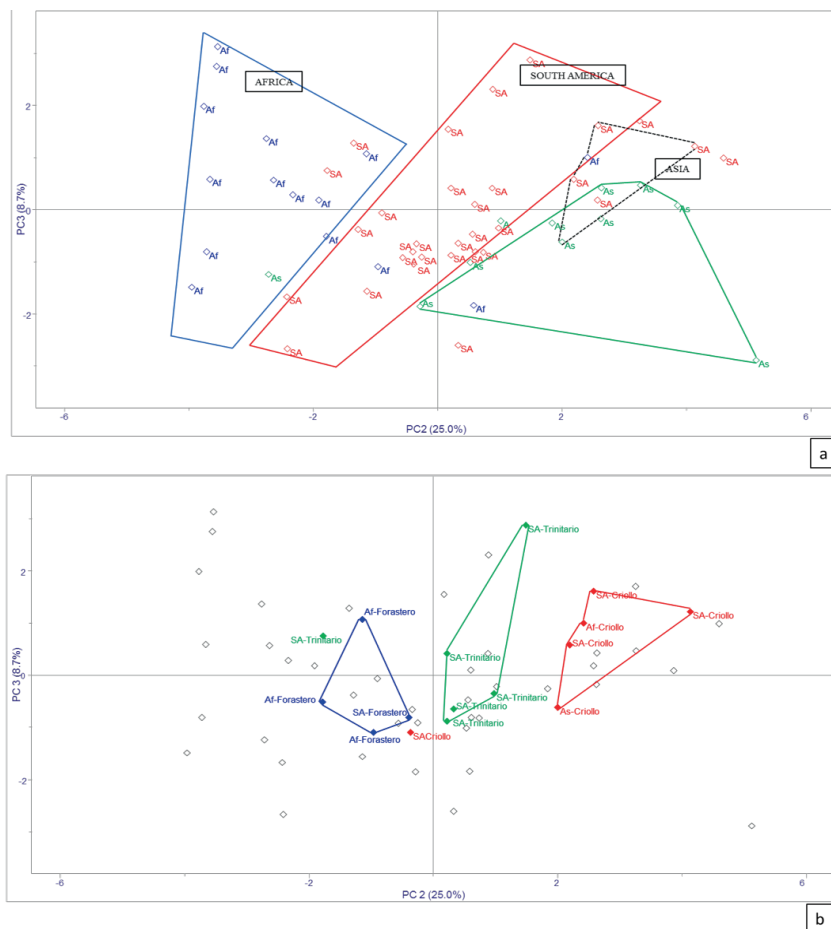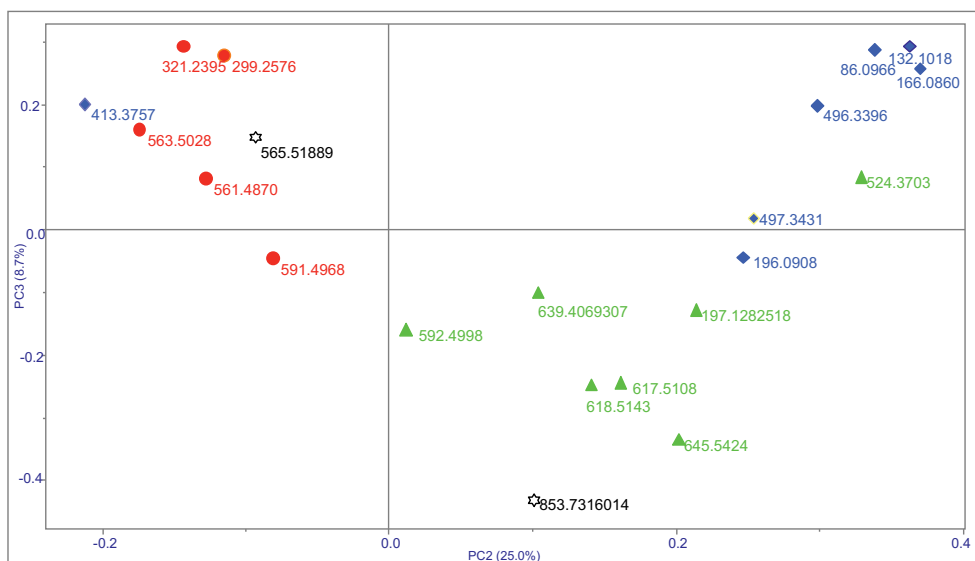
Figure 4.2. PCA score plots (PC1 vs PC2) of the auto-scaled FI-ESI-MS data of 57 chocolates manufactured with cocoa beans from various geographical and botanical origins. a) Plot showing clusters according to the geographical origin of the cocoa; b) Plot showing clusters according to the various brands (nr. 1e10). Empty symbols refer to samples for which there is only one chocolate representing a brand. Af: Africa; As: Asia; SA: South America.

As stated before, the beans' genotype can influence the final product (paragraph 3.1.2). The influence of the botanical origin of the cocoa beans is presented in Figure 4.3a and 4.3b: in this study, 16 chocolates for which the genetic groups of cocoa, Criollo (6), Forastero (4), and Trinitario (6) were known, were subjected to PCA. Figure 4.3b shows the same PCA as in Figure 4.3a with the samples now highlighted corresponding to the chocolates for which the botanical origin (Criollo, Forastero, Trinitario) was specified. The plot shows grouping according to the bean botanical origin. This result is predominant for Trinitario and Criollo beans and emphasises that the fingerprint comprises also a signature of the genetic component of the beans despite the relatively low number of samples per group.

Figure 4.3. PCA score plots (PC2 vs PC3) of the auto-scaled FI-ESI-MS data of 57 chocolates manufactured with cocoa beans from various geographical and botanical origins. a) Plot showing clusters according to the geographical origin of the cocoa; b) Plot showing clusters according to the botanical origin (Criollo, Forastero, Trinitario). The chocolates highlighted with the black dashed line in Figure 4.3a are produced from Criollo beans. Empty symbols are shown in Figure 4.3b for those samples for which the bean botanical origin is unknown. Af: Africa; As: Asia; SA: South America.

Figure 4.4. PCA loadings plot (PC2 vs PC3) of the auto-scaled FI-ESI-MS data of 57 chocolates manufactured with cocoa beans from various geographical origins. The different shapes and colours refer to the groups created by ANOVA results.

◇: masses able to differentiate Africa vs the rest of the continents; △ : masses able to differentiate Africa vs South America;

○ : masses able to differentiate Asia vs the rest of the continents; ✡ :mass able to differentiate South America vs All (565.5189), and Asia vs Africa (853.7316). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

It is obvious that botanical, geographical origin and/or brand-impact may be linked to each other in a few cases. Considering the results in Figure 4.3 it is possible to explain that a specific case of overlapping, such as the mixed area highlighted in Figure 4.3a (black dashed line = Criollo beans) can be related to the botanical origin of the beans used in manufacturing the chocolates.

The genetic cocoa groups generally used to produce chocolate are Criollo, Forastero, Trinitario and National (Saltini et al., 2013) and can be differentiated according to location of growing, post-harvesting condition such as length of fermentation, and quantitative content of certain compounds such as pyrazine, amino acids and polyphenols (Afoakwa, Paterson, & Fowler, 2007; Saltini et al., 2013). Some are shown relevant in the present study too.

Taking into account the distribution of the samples in Figures 4.3a and 4.3b, the masses in the upper right part of the loading plot (Fig. 4.4b) may be related to the chocolates made from Criollo beans. In fact for these samples, the average of the intensities of mass 86.0966, 132.1018, and 166.0860 are higher compared to the African, Asian and South American group (Table 4.1). Masses 132.1018 and 166.0860 are tentatively identified as amino acids, and may indicate a link with the Criollo chocolate as according to literature a higher concentration of amino acids is found in Criollo beans (Saltini et al., 2013).

4.3.2.2 Application: classification by k nearest neighbours models

A kNN classification method was applied to assess chocolate separation according to the cocoa beans geographical origins. The seven different comparisons (paragraph 3.1.1) were examined. The best k value was applied to each model (Table 4.3). kNN enables the separation between the three origins to be tested at the same time. The results showed that the 1000 x repeated random cross-validated model with k = 1 on the auto-scaled data provides 70% correct classification (Table 4.3).

Table 4.3. kNN classification of the auto-scaled FI-ESI-MS data of chocolates manufactured with cocoa beans of various geographical origins: correct classification rates (%) of the seven comparisons obtained by 1000 repeated random cross-validation.

Comparisons	*k* nearest neighbours (k)			
	k=1			
	Africa	Asia	South America	Total model
Africa-Asia-South America	60%	57%	80%	70%
	k=1			
	Africa	Asia and South America		Total model
Africa-(Asia South America)	60%	85%		78%
	k=3			
	Asia	Africa and South America		Total model
Asia-(Africa South America)	63%	93%		87%
	k=1			
	South America	Africa and Asia		Total model
South America-(Africa Asia)	80%	76%		78%
	k=3			
	Africa	Asia		Total model
Africa-Asia	86%	81%		84%
	k=1			
	Africa	South America		Total model
Africa-South America	60%	90%		80%
	k=3			
	Asia	South America		Total model
Asia-South America	63%	77%		73%

Table 4.3 shows that the rate of correct classification within the three continents indicates high percentages of correct classification for South American chocolates and relatively low percentage for African and Asian chocolates. The difficulties in discriminating the three continents at the same time can be related to similarities in composition between cocoa butters from several American origins, such as Costa Rica, the Dominican Republic and Mexico, and African cocoa butters (Chaiseri & Dimick, 1989). Another reason for the low percentage of discrimination could be related to the different climate conditions of cocoa growing. According to Chaiseri and Dimick (1989), more variability within South American and African samples is expected compared to an Asian one. This is because in Africa and South America cocoa is grown over a larger geographical area for which a stronger temperature variation exists. These differences could be reflected in the finished chocolate making the discrimination more difficult. The shared geographical characteristics and the different variability could explain the low percentage of correct classification for African and Asian chocolate and the impossibility to distinguish the three geographical origins at the same time.

To check the best separation within the chocolates, comparisons between both one continent and the remaining two as well as a single continent against another were examined. The classification of one continent versus the other two resulted in an improvement of the correct classification. The highest scores were obtained comparing the South American chocolates versus African and Asian together: showing 78% correct classification for the total model, and 80 and 76% for South America and Africa/Asia, respectively. The other comparisons revealed a good classification but a lower success rate per continent (Table 4.3).

Within the seven comparisons, the most successful classification was obtained in the separation between African and Asian beans comprising chocolates, they resulted in correct classification rates of 84% for the total model, and 86% and 81% for Africa and Asia, respectively.

Based on the current sample set and the variables selected, it appeared feasible to adequately differentiate chocolates manufactured from Asian and African cocoa beans, and both groups versus the chocolates produced from South American beans.

4.4 Conclusions

This study resulted in distinct FI-ESI-MS fingerprints for chocolates manufactured with cocoa beans from different geographical origins. ANOVA selection provided a range of masses that were used to find a link between the chocolates and the origins of their raw materials such as amino acids, fatty acids and sterols. Exploratory multivariate data analysis revealed the impact of the geographical origin of the cocoa, botanical origin of the cocoa as well as brand-related factors on the fingerprints. Classification by kNN confirmed the possibility of separating chocolates produced from African and Asian cocoa beans, as well as chocolates manufactured from beans from these two continents together versus those from South America. We were not able to sufficiently separate chocolates produced from beans of the three continents at the same time due to interference of the other factors, such climate growing-conditions and especially brand-related factors. This can be underlined by the compounds tentatively identified, such as ricinoleic acid and the PCA results where the brand influence was evident.

However, the ability to extract brand information means that FI- ESI-MS could be used to highlight differences in chocolates from different manufacturers. Future work, including cocoa bean analysis and studies on the impact of processing, will help to develop the methodology further for use in practice.

Acknowledgements

References

Acierno, V., Yener, S., Alewijn, M., Biasioli, F., & van Ruth, S. M. (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. http://doi.org/10.1016/j. foodres.2016.03.022.

Afoakwa, E. O., Paterson, A., & Fowler, M. (2007). Factors influencing rheological and textural qualities in chocolate e a review. *Trends in Food Science & Technology, 18*(6), 290-298. http://doi.org/10.1016/j.tifs.2007.02.002.

Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2008). Flavor formation and character in cocoa and chocolate: A critical review. *Critical Reviews in Food Science and Nutrition, 48*(9), 840-857. http://doi.org/10.1080/ 10408390701719272.

Alonso, R., Rodríguez-Estévez, V., Domínguez-Vidal, A., Ayora-Cañada, M. J., Arce, L., & Valcárcel, M. (2008). Ion mobility spectrometry of volatile compounds from Iberian pig fat for fast feeding regime authentication. *Talanta, 76*(3), 591-596. http://doi.org/10.1016/j.talanta.2008.03.052.

Arlorio, M., Locatelli, M., Travaglia, F., Coïsson, J.-D., Grosso, E. Del, Minassi, A., et al. (2008). Roasting impact on the contents of clovamide (N-caffeoyl-L-DOPA) and the antioxidant activity of cocoa beans (Theobroma cacao L.). *Food Chemistry, 106*(3), 967-975. http://doi.org/10.1016/j.foodchem.2007.07.009.

Berrueta, L. A., Alonso-Salces, R. M., & Héberger, K. (2007). Supervised pattern recognition in food analysis. *Journal of Chromatography A, 1158*(1-2), 196-214. http://doi.org/10.1016/j.chroma.2007.05.024.

Bertoldi, D., Barbero, A., Camin, F., Caligiani, A., & Larcher, R. (2016). Multielemental fingerprinting and geographic traceability of Theobroma cacao beans and cocoa products. *Food Control, 65*, 46-53. http://doi.org/10.1016/j.foodcont.2016.01.013.

Bódalo-Santoyo, A., Bastida-Rodríguez, J., Máximo-Martín, M. F., Montiel- Morte, M. C., & Murcia-Almagro, M. D. (2005). Enzymatic biosynthesis of ricinoleic acid estolides. *Biochemical Engineering Journal, 26*(2-3), 155-158. http:// doi.org/10.1016/j.bej.2005.04.012.

Brito, E. S. De, Garcı Pezoa, N. H., Gallao, M. I., Cortelazzo, A. L., Fevereiro, P. S., & Braga, M. R. (2001). Structural and chemical changes in cocoa (Theobroma cacao L) during fermentation, drying and roasting. *Journal of the Science of Food and Agriculture, 81*(2), 281-288.

Caligiani, A., Cirlini, M., Palla, G., Ravaglia, R., & Arlorio, M. (2007). GC-MS detection of chiral markers in cocoa beans of different quality and geographic origin. *Chirality, 19*(4), 329-334. http://doi.org/10.1002/chir.

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*(3), 1478-1483. http://doi.org/10.1021/jf903471e.

Camu, N., Winter, T. De, Addo, S. K., Takrama, J. S., Bernaert, H., & Vuyst, L. De (2008). Fermentation of cocoa beans: Influence of microbial activities and polyphenol concentrations on the flavour of chocolate. *Journal of the Science of Food and Agriculture, 88*(13), 2288-2297. http://doi.org/10.1002/jsfa.

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

CBI. (2016). CBI product Factsheet: Cocoa in The Netherlands. Retrieved from https://www.cbi.eu/ sites/default/files/market_information/researches/product-factsheet-netherlands-cocoa-2016.pdf/ Accessed 24.04.2017.

Chaiseri, S., & Dimick, P. S. (1989). Lipid and hardness characteristics of cocoa butters from different geographic Regions. *Journal of the American Oil Chemists Society, 66*(12), 1771-1776.

Comba, L., Belforte, G., Dabbene, F., & Gay, P. (2013). Methods for traceability in food production processes involving bulk products. *Biosystems Engineering, 116*(1), 51-63. http://doi.org/10.1016/j.biosystemseng.2013.06.006.

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

Dabbene, F., Gay, P., & Tortia, C. (2014). Traceability issues in food supply chain management: A review. *Biosystems Engineering, 120*, 65-80. http://doi.org/10. 1016/j.biosystemseng.2013.09.006.

Davies, A. M. C., Franklin, J. G., Grant, A., Griffiths, N. M., Shepherd, R., & Fenwick, G. R. (1991). Prediction of chocolate quality from near-infrared spectroscopic measurements of the raw cocoa beans. *Vibrational Spectroscopy, 2*(2-3), 161-172. http://doi.org/10.1016/0924-2031(91)85022-F.

Diab, J., HertzeSchünemann, R., Streibel, T., & Zimmermann, R. (2014). On-line measurement of volatile organic compounds released during roasting of cocoa beans. *Food Research International, 63*, 344-352. http://doi.org/10.1016/j. foodres.2014.04.047.

Draper, J., Lloyd, A. J., Goodacre, R., & Beckmann, M. (2012). Flow infusion electro-spray ionisation mass spectrometry for high throughput, non-targeted metabolite fingerprinting: A review. *Metabolomics, 9*(S1), 4-29. http://doi.org/ 10.1007/s11306-012-0449-x.

El-Saied, H. M., Morsi, M. K., & Amer, M. M. A. (1981). Composition of cocoa shell fat as related to cocoa butter. *Zeitschrift Für Ernährungswissenschaft, 20*(2), 145-151. http://doi.org/10.1007/BF02021260.

González-Dominguez, R., García-Barrera, T., & Gõmez-Ariza, J. L. (2012). Iberian ham typification by direct infusion electrospray and photospray ionization mass spectrometry fingerprinting. *Rapid Communications in Mass Spectrometry, 26*(7), 835-844. http://doi.org/10.1002/rcm.6155.

Government of the Netherlands (2014). Sustainable chocolate depends on fair price for farmers. Retrieved from https://www.government.nl/latest/news/2014/06/10/sustainable-chocolate-depends-on-fair-price-for-farmers/ Accessed 24.04.2017.

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

Hobbs, J. E. (2004). Information asymmetry and the role of traceability systems. *Agribusiness, 20*(4), 397-415. http://doi.org/10.1002/agr.20020.

Ho, C. S., Lam, C. W. K., Chan, M. H. M., Cheung, R. C. K., Law, L. K., Lit, L. C. W., ... Tai, H. L. (2003). Electrospray ionisation mass spectrometry: Principles and clinical applications. *The Clinical Biochemist. Reviews/Australian Association of Clinical Biochemists, 24*(1), 3-12.

Jinap, S., Dimick, P. S., & Hollender, R. (1995). Flavour evaluation of chocolate formulated from cocoa beans from different countries. *Food Control, 6*(2), 105-110. http://doi.org/10.1016/0956-7135(95)98914-M.

Junia de, O. Alves, Neto, Waldomiro B., Mitsutake, Hery, & Paulo S. P. Alves, R. A. (2010). Extra virgin (EV) and ordinary (ON) olive oils: Distinction and detection of adulteration (EV with ON) as determined by direct infusion electrospray ionization mass spectrometry and chemometric approaches. *Rapid Communications in Mass Spectrometry, 24*(13), 1875-1880.

Lipp, M., & Adam, E. (1998). Review of cocoa butter and alternative fats for use in chocolate-Part A. Compositional data. *Food Chemistry, 62*(1), 73-97.

Luykx, D. M. a. M., & van Ruth, S. M. (2008). An overview of analytical methods for determining the geographical origin of food products. *Food Chemistry, 107*(2), 897-911. http://doi.org/10.1016/j.foodchem.2007.09.038.

Marty-Terrade, S., & Marangoni, A. G. (2012). Impact of cocoa butter origin on crystal behaviour. In N. Garti, & N. R. Widlak (Eds.), *Cocoa butter and related compound*s (pp. 245-274). Urbana: AOCS Press.

Neurath, G. B., Dünger, M., Pein, F. G., Ambrosius, D., & Schreiber, O. (1977). Primary and secondary amines in the human environment. *Food and Cosmetics Toxicology, 15*(4), 257-282. http://doi.org/10.1136/bmj.2.3577.165-c.

Nolvachai, Y., Kulsing, C., Boysen, R. I., Matyska, M. T., Pesek, J. J., Marriott, P. J., & Hearn, M. T. W. (2015). Comparison of the performance of different silica hydride particles for the solid-phase extraction of non-volatile analytes from dark chocolate with analysis by gas chromatography-quadrupole mass spectrometry. *Food Chemistry, 174*, 434-439. https://doi.org/10.1016/j.foodchem.2014.10.083.

Pan, H. (2008). A non-covalent dimer formed in electrospray ionisation mass spectrometry behaving as a precursor for fragmentations. *Rapid Communications in Mass Spectrometry: RCM, 22*(22), 3555-3560. http://doi.org/10.1002/rcm.

Pätzold, R., & Brückner, H. (2006). Gas chromatographic determination and mechanism of formation of D-amino acids occurring in fermented and roasted cocoa beans, cocoa powder, chocolate and cocoa shell. *Amino Acids, 31*(1), 63-72. http://doi.org/10.1007/s00726-006-0330-1.

Perret, D., Gentili, A., Marchese, S., Sergi, M., & Caporossi, L. (2004). Determination of free fatty acids in chocolate by liquid chromatography with tandem mass spectrometry. *Rapid Communications in Mass Spectrometry: RCM, 18*(17), 1989-1994. http://doi.org/10.1002/rcm.1582.

Rohman, A., & Che Man, Y. B. (2012). Analysis of pig derivatives for halal authentication studies. *Food Reviews International, 28*(1), 97-112. http://doi.org/10.1080/87559129.2011.595862.

Rohsius, C., Matissek, R., & Lieberei, R. (2006). Free amino acid amounts in raw cocoas from different origins. *European Food Research and Technology, 222*(3-4), 432-438. http://doi.org/10.1007/s00217-005-0130-y.

Saltini, R., & Akkerman, R. (2012). Testing improvements in the chocolate traceability system: Impact on product recalls and production efficiency. *Food Control, 23*(1), 221-226. http://doi.org/10.1016/j.foodcont.2011.07.015.

Saltini, R., Akkerman, R., & Frosch, S. (2013). Optimizing chocolate production through traceability: A review of the influence of farming practices on cocoa bean quality. *Food Control, 29*(1), 167-187. http://doi.org/10.1016/j.foodcont.2012.05.054.

Simoneau, C., Hannaert, P., & Anklam, E. (1999). Detection and quantification of cocoa butter equivalents in chocolate model systems: Analysis of triglyceride profiles by high resolution GC. *Food Chemistry, 65*(1), 111-116.

Squicciarini, M. P., & Swinnen, J. (2016). The economics of chocolate: Introduction and overview. In M. P. Squicciarini, & J. Swinnen (Eds.), *The economics of chocolate* (pp. 1-10). Oxford: Oxford University Press.

Sukha, D. a., Butler, D. R., Umaharan, P., & Boult, E. (2007). The use of an optimised organoleptic assessment protocol to describe and quantify different flavour attributes of cocoa liquors made from Ghana and Trinitario beans. *European Food Research and Technology, 226*(3), 405-413. http://doi.org/10.1007/s00217- 006-0551-2.

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. http://doi.org/10.1016/j.saa.2013.05.063.

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. http://doi.org/10.1016/j.foodchem.2014.05.141.

Tres, A., O'Neill, R., & van Ruth, S. M. (2011). Fingerprinting of fatty acid composition for the verification of the identity of organic eggs. *Lipid Technology, 23*(2), 40-42. http://doi.org/10.1002/lite.201100084.

Ulberth, F., & Buchgraber, M. (2003). Analytical platforms to assess the authenticity of cocoa butter. European *Journal of Lipid Science and Technology, 105*(1), 32-42. http://doi.org/10.1002/ejlt.200390003.

Wognum, P. M., Bremmers, H., Trienekens, J. H., Van Der Vorst, J. G. A. J., & Bloemhof, J. M. (2011). Systems for sustainability and transparency of food supply chains - current status and challenges. *Advanced Engineering Informatics, 25*(1), 65-76. http://doi.org/10.1016/j.aei.2010.06.001.

Chapter 5

Sniffing out cocoa bean traits that persist in chocolates by PTR-MS, ICP-MS and IR-MS

Acierno, V., de Jonge, L., & van Ruth, S. M. (2020). Sniffing out cocoa bean traits that persist in chocolates by PTR-MS, ICP-MS and IR-MS. *Food Research International, 133(July 2020).*[1]

Abstract

Cocoa botanical and geographical origin and the primary processing steps applied by cocoa farmers at the beginning of the supply chain influence the characteristics of the cocoa beans. These features are carried along the supply chain as intrinsic markers up to the final products. These intrinsic markers could be used for tracking and tracing purposes. In this study, we examined the retention and loss of compositional signatures from cocoa beans to chocolates. Volatile, elemental and stable isotope signatures of cocoa beans of 10 different origins and 11 corresponding chocolates were determined by High Sensitivity Proton Transfer Reaction Mass Spectrometry (HS-PTR-MS), Inductively Coupled Plasma Mass Spectrometry (ICP-MS) and Isotope Ratio Mass Spectrometry (IR-MS), respectively. The volatile fingerprints provided mostly information on the origin and primary processing traits of the raw cocoa beans in the chocolates. Volatile compounds that are relevant markers include: acetic acid (m/z 61), benzene (m/z 79), pyridine (m/z 80), 2-phenylethanol (m/z 123), and maltol (m/z 127). On the other hand, the elemental and stable isotope characteristics are more indicative of the cocoa content and added ingredients. Possible elemental markers for geographical origin of the beans include Fe, Cr, and Cd. VOCs appear to be the most robust markers carried from cocoa beans to chocolates of the groups examined. This provides the potential for track and trace options of cocoa beans from farm to chocolates.

5.1 Introduction

The chain from cocoa beans to chocolate is fairly long and complex. Production dispersion of the batches can already take place in the first steps. The cocoa beans produced by small-sized farmers are merged into larger batches. Cocoa botanical and geographical origin and the primary processing applied at the beginning of the supply chain influences the finished chocolate. This information could be used to establish a compositional link between the raw materials and finished chocolate and possibly traceback or verify the origin of the cocoa beans in the supply chain.

A few studies searched for relationships between cocoa beans and chocolate characteristics. These studies focused mainly on quality assurance in terms of flavour development (Ascrizzi, Flamini, & Tessieri, 2017) or quality prediction (Davies, Franklin, Grant, Griffiths, Shepherd, & Fenwick, 1991). Other studies followed the evolution of volatile organic compounds (VOCs) (Braga, Oliveira, Hashimoto, Gama, Efraim, & Poppi, 2018), sugars, acids (Rodriguez-campos, Escalona-buendía, Contreras-ramos, & Orozco-avila, 2012) and amino acids (Pätzold & Brückner, 2006) along the chocolate production but they focused on optimizing production steps such as fermentation, drying, roasting and conching (Counet, Callemien, Ouwerx, & Collin, 2002).

A direct link with the raw material could be represented by the botanical and geographical information. So far, botanical and geographical origin of cocoa beans has been largely analysed. Volatile organic compounds (Qin et al., 2017), triglyceride composition (Hernandez, Castellote, & Permanyer, 1991) and elements (Bertoldi, Barbero, Camin, Caligiani, & Larcher, 2016) have been used to differentiate cocoa beans of different varieties and origins. A few studies detected the botanical and geographical origin of cocoa beans used to manufacture the chocolate into a finished chocolate bar (Acierno, Alewijn, Zomer, & van Ruth, 2018; Acierno, Yener, Alewijn, Biasioli, & van Ruth, 2016; Cambrai, Marcic, Morville, Sae Houer, Bindler, & Marchioni, 2010) indicating the possibility to extract raw material information from a finished product. However, the magnitude of changing of potential origin markers along the chocolate production has never been tested. Other parameters that could be used to track and trace the production are the primary processing steps, such as fermentation and drying, as they are characterising the final product too. These procedures could give cocoa beans origin information as they differ between regions, countries and/or continents (Acierno, Fasciani, Kiani, Caligiani, & van Ruth, 2019; Caligiani, Marseglia, Prandi, Palla, & Sforza, 2016; Caligiani, Palla, Acquotti, Marseglia, & Palla, 2014).

Mass spectrometry (MS) is usually applied to detect botanical and geographical information in food products (Luykx & van Ruth, 2008). Techniques such as Inductively Coupled Plasma-MS (ICP-MS) and Isotope Ratio-MS (IR-MS) could give information on factors both indicative of the geographical origin and the production such as the use of fertilisers, and seasonal and geological variations (Isabel et al., 2009; Serra et al., 2005). In the specific case of cocoa, multi-element screening has been used to identify macro-, micro- or trace-elements indicative both of cocoa and chocolate provenance (Bertoldi et al., 2016; Levi et al., 2014), and soil and cropping system reflection on dry cocoa beans (Araujo, Baligar, Loureiro, & Júnior, 2017). Only a few studies investigated stable isotopes for detection of the cocoa beans provenance. Diomande et al. (2015) used multi-compound isotope profiling to identify botanical and geographical information of fermented cocoa beans. Other analytical MS techniques have been studied to verify the origin of foods and the influence of the processing steps. These techniques include Proton Transfer Reaction –MS (PTR-MS). PTR-MS has been used to distinguish the provenance of several food products including cocoa and chocolate (Acierno et al., 2016; Acierno et al., 2019) and has also been used to study the impact of processing such as roasting (Yener, Navarini, Lonzarich, Cappellin, Märk, Bonn, Biasioli, 2016). Furthermore, the technique has been applied to assess the impact of different production systems in milk production (Liu, Koot, Hettinga, de Jong, & van Ruth, 2018). Therefore, PTR-MS seems a suitable technique for gathering information about the beginning of the production and for application at different stages of the supply chain.

In this study, we examined the retention and loss of compositional signatures from cocoa beans to chocolates. Volatile, elemental and isotopic fingerprints of cocoa beans (fermented and dried) and chocolates produced thereof have been evaluated to investigate the magnitude of similarity and dissimilarity between cocoa bean raw materials and finished chocolate products. Ten samples of cocoa beans of different origin and 11 corresponding chocolates were analysed for their volatile, elemental and isotopic signatures using High Sensitivity-PTR-MS (HS-PTR-MS), ICP-MS and IR-MS, respectively.

5.2 Material and methods

5.2.1 Samples

Ten fermented and dried cocoa bean samples from different geographical origins and eleven chocolate samples industrially manufactured from those beans were sourced from a single manufacturer (Table 5.1). From each bean sample one type of chocolate was manufactured, except for one cocoa bean sample from Trinidad (TT2) which was used to produce two chocolates with different cocoa contents ((TT2 (78% cocoa); TT2 (82% cocoa)). Chocolates generally varied in cocoa content. The samples were stored at room temperature (18-20 °C) until analysis. Cocoa beans and chocolates were ground a day prior to the start of the measurements. After grinding, they were stored in plastic bottles with screw caps at 4 °C.

Table 5.1. Cocoa beans and corresponding chocolates.

Cocoa beans		Chocolates	
Origin	Code	Cocoa content (w/w (%))	Code
Congo	CG	60	CG(60)
Mexico	MEX	34	MEX(34)
Ecuador	EC	72	EC(72)
Venezuela	YV	45	YV(45)
Costa Rica 1	CR1	48	CR1(48)
Costa Rica 2	CR2	48	CR2(48)
Vanuatu 1	VU1	44	VU1(44)
Vanuatu 2	VU2	44	VU2(44)
Trinidad 1	TT1	81	TT1(81)
Trinidad 2	TT2	78	TT2(78)
		82	TT2(82)

5.2.2 HS-PTR-MS analysis

One gram of powdered cocoa bean and three grams of ground chocolate were weighed into 250 ml flasks. Cocoa beans and chocolates were placed in a water bath at respectively 30 and 40 °C for 30 min to equilibrate the samples with their headspace. Due to compounds saturation in the cocoa beans headspace, lower quantity and lower water bath temperature have been used for the cocoa beans compare to the chocolates. The headspace of the powdered cocoa beans and chocolates was measured by HS-PTR-MS (Ionicon Analytik G.m.b.H., Innsbruck, Austria). The PTR-MS conditions were as follows: drift pressure 2.20 mbar, inlet flow 60 ml/min, the reaction chamber and inlet temperature at 60 ºC. The instrument was operated at an E/N (ratio of electric field strength across the reaction chamber, E, to buffer gas number density, N, within the chamber) of 119 Td (1 Td = 10^{-17} cm^2 V molecule^{-1}). For each sample, a mass range between 20 and 160 was measured using a dwell time of 0.2 s mass^{-1}. A blank (empty flask) was analysed before each sample. Five cycles per measurement were recorded and specifically the three last cycles for the blanks and the three cycles in the middle for the samples were used for data analysis. The values obtained for the blank were subtracted from each sample and all values were corrected for transmission. HS-PTR-MS analysis was run in triplicate.

5.2.3 ICP-MS

Cocoa beans and chocolate samples were pre-treated using acid digestion with a microwave oven (MARS express, CEM Corporation, Matthews, NC, USA). For the microwave digestion, 10 mL of concentrated nitric acid (69% Instra-analysed nitric acid, J.T. Baker, Phillipsburg, NJ, USA) was added to 0.5 g of sample in Teflon digestion vessels. The samples were then placed in the microwave oven at a temperature of 210°C. The digests were quantitatively transferred to 50 ml polypropylene (PP) centrifuge tubes (Greiner Bio-One, Frickenhausen, Germany) and diluted with de-ionized water to a final volume of 50 ml. The digests were diluted 5-fold and 200-fold before analysis with a final acid concentration of 2.8% v/v. Concentrations of 53 elements were determined using ICP-MS (NexION 300D, Perkin Elmer, Waltham, MA, USA). Samples were introduced into the ICP-MS using an SC-FAST autosampler (ESI, Omaha, NE, USA). The ICP-MS was equipped with a PFA-ST nebulizer, a quartz cyclonic spray chamber and a quartz torch with a sapphire injector. The ICP-MS was operated in collision mode (KED) with helium as collision gas for all elements, with a flow rate of 4.1 mL/min and an RPq of 0.25. Concentrations were determined using external calibration curves with rhodium as an internal standard.

5.2.4 IR-MS

Cocoa beans and chocolate samples were analysed for ^{13}C and ^{15}N enrichment by elemental analyses using an isotope ration mass spectrometer (Flash 2000 in combination with Delta V advantage, Thermo Scientific, Waltham MA USA). All the samples were analysed in duplicate. The relative ^{13}C and ^{15}N enrichment are expressed as the $^{13}C:^{12}C$ and $^{15}N:^{14}N$ ratio in the samples relative (in ‰) to the $^{13}C:^{12}C$ and $^{15}N:^{14}N$ ratio of the Vienna Pee Dee Belemnite (VPDB) international standard.

For ^{18}O and ^{2}H determination, cocoa beans and chocolate samples were analysed on a HEKAtech high-temperature elemental analyser coupled with a GV Instruments IsoPrime mass spectrometer. Samples were pyrolyzed at 1450 ºC, in silver capsules. All samples were analysed in triplicate. All results are reported with respect to Vienna Standard Mean Ocean Water (VSMOW), normalized to international standards IAEA-SO-5 and IAEA-SO-6 with reported values of +12.0‰ and -11.3‰ for δ18O values. The analytical precision for these measurements is 0.5‰.

5.2.4.1 Statistical analysis

5.2.4.1.1 Univariate analysis

For the entire data set normality within each group was checked by Shapiro-Wilk test and the assumption of equal variances was assessed by Levine's test based on the median. To measure similarity and dissimilarity in volatile, elemental and isotopic signatures between cocoa beans and corresponding chocolates, data were subjected to non-parametric tests. Spearman's Rho was applied to correlate data of cocoa beans and chocolates, Mann-Whitney tests were used to investigate significant differences between groups. The Mann-Whitney tests were followed by a rate of false discovery (RFD) correction. A significance level of $p \leq 0.05$ was used throughout the study. All the statistical analyses of data were conducted using IBM SPSS (Statistic 23.0, IBM Corp., Armonk, NY, USA).

5.2.4.1.2 Multivariate analysis

Principal component analysis (PCA) was applied to visually explore the presence of any natural clustering in the data using Pirouette 4.0 Software (Infometrix, Seattle, WA, USA). A correlation between the score values of cocoa beans and corresponding chocolates was measured with Pearson's correlation coefficient using IBM SPSS (Statistic 23.0, IBM Corp., Armonk, NY, USA).

5.2.4.1.3 Correlations between volatile, elemental and isotopic data

Correlation tables between volatile and elemental and isotopic data, based on Spearman's Rho values, were built using R 3.3.3 (R Foundation for Statistical Computing, Vienna, Austria).

5.3 Results and discussion

5.3.1 Signatures of volatile compounds of cocoa beans and corresponding chocolates

The 10 cocoa beans and the 11 corresponding chocolate samples were subjected to HS-PTR-MS analysis. The mass spectra of both cocoa beans and chocolates obtained by the PTR-MS showed signals on most masses in the defined measurement range (20–160 amu). The spectra, used as signatures, comprised 133 masses after removal of O_2^+ (m/z 32) and water clusters (m/z 37). The data were not normally distributed (Shapiro-Wilk test, $p \leq 0.05$), hence both significant correlations and significant differences were evaluated by non-parametric tests, i.e. the Spearman's Rho and Mann-Whitney tests, respectively.

Cumulated intensities of the volatile organic compounds (VOCs) of each cocoa bean sample and the corresponding chocolate revealed a significant correlation between the two data groups (Spearman's Rho = 0.59; $p \leq 0.05$) (Figure 5.1).

Cumulative VOC intensities of the chocolates were significantly lower (Mann-Whitney test, $p \leq 0.05$) than those of the cocoa beans. This also holds for the individual compounds, for which significant differences were noted for ~ 93% of the 133 masses detected. Rather more quantitative than qualitative differences in profile intensities between cocoa beans and chocolates were observed (Figure 5.2). For cocoa beans, highest intensities were observed for higher molecular weight compounds (m/z > 100). Within this group significantly higher intensities for cocoa beans were noted in comparison to the chocolates, e.g. for m/z 109 (4-methylphenol/benzyl alcohol), m/z 112 (trimethyloxazole), m/z 121 (2-ethenyl-6-methylpyrazine), m/z 123 (2-phenylethanol), m/z 125 (guaicol), m/z 129 (furaneol/ heptanedione/ octanal), m/z 131 (ethyl valerate/heptanoic acid), m/z 137(various monoterpenes), m/z 138 (pyridine carboxylic acid methyl ester), m/z 153 (E,E-2,4-decadienal), and m/z 159 (nonanoic acid) (Mann-Whitney test, $p \leq 0.05$; Table 5.2). The tentative identifications of the masses were based on our previous studies focused on the analysis of cocoa beans and chocolates using PTR-Time of Flight-MS (PTR-ToF-MS) (Acierno et al., 2016; Acierno et al., 2019).

The lower concentrations in the chocolate samples can be related to the influence of the chocolate processing and matrix effects which in turn alter the partitioning of VOCs from the product to the headspace (Ducki, Miralles-Garcia, Zumbé, Tornero, & Storey, 2008). For instance, concentrations of organic acids such heptanoic acid (m/z 131) and nonanoic acid (m/z 159) typically decrease during conching; in general the acid concentration is expected to be low in chocolate so the flavour is mild and pleasant (Ascrizzi et al., 2017). Terpenes (m/z 137, various monoterpenes) usually decrease during roasting (Yener et al., 2016). 4-Methylphenol (m/z 109) and guaicol (2-methoxyphenol) (m/z 125) increase during fermentation and drying, however, phenolic compounds are undesirable in the cocoa product so the low level in the final product is considered to add to a better chocolate flavour (Rodriguez-campos et al., 2012). Regarding the matrix effects, it is known that the headspace concentrations of volatiles are influenced by fat, sugar and additive contents (Romano, 2016). For instance, low concentrations of VOCs with caramel-like notes such as furaneol (m/z 129), 2-phenylethanol (m/z 123) and long aldehydes such as E/E-2,4-decadienal (m/z 153) depend on the fat content since a high lipid content reduces the volatility of lipophilic components (Afoakwa, Paterson, Fowler, & Ryan, 2009). Low concentration of 2-ethenyl-6-methylpyrazine (m/z 121) is mainly related to the fact that pyrazine, being a non-polar compound, tend to be trapped in the fat phase (Ziegleder, 1991).

Despite the difference in concentration, significant correlations (Spearman's Rho, $p \leq 0.05$) were established between mass intensities of the cocoa beans and chocolates (Figure 5.3).
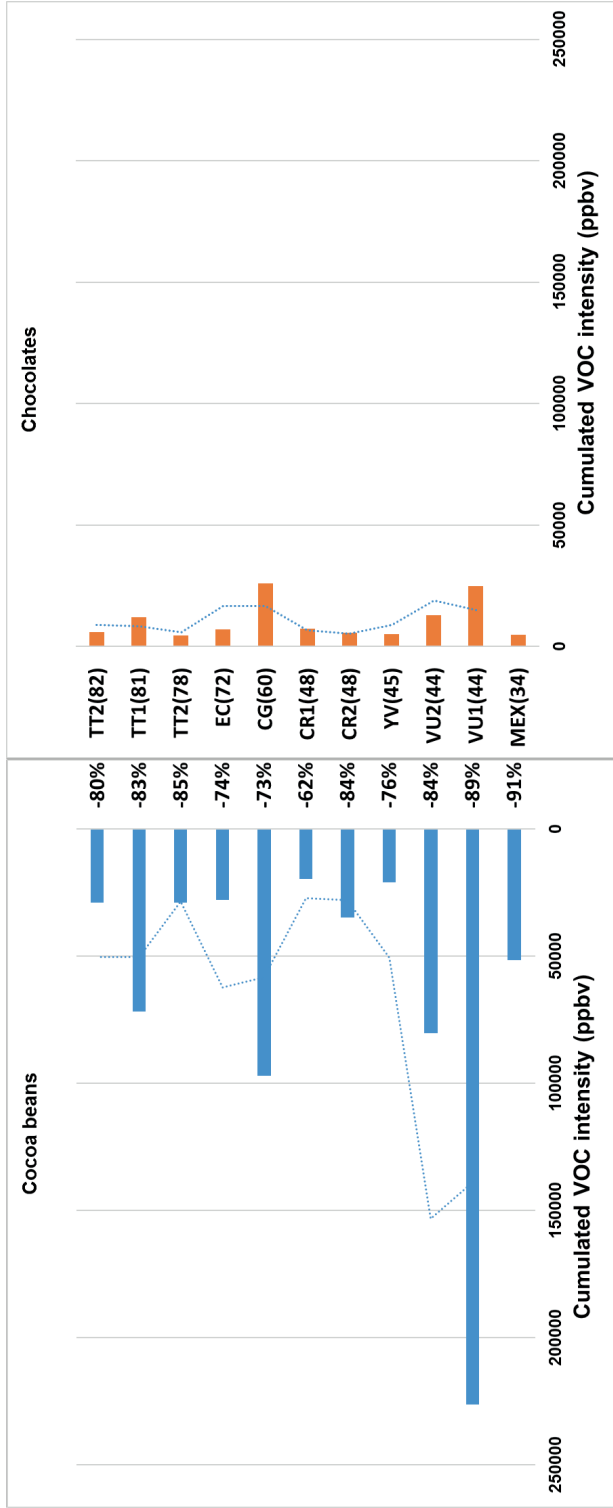
Figure 5.1. Cumulated VOC intensities of cocoa beans and corresponding chocolates and the relative intensity differences between cocoa beans to chocolates. MEX: Mexico; VU1 Vanuatu1; VU2: Vanuatu 2; YV: Venezuela; CR1: Costa Rica1; CR2: Costa Rica2; CG: Congo; EC: Ecuador; TT2: Trinidad 2; TT1: Trinidad1. The percentage in brackets indicates the relative cocoa content of the chocolates.
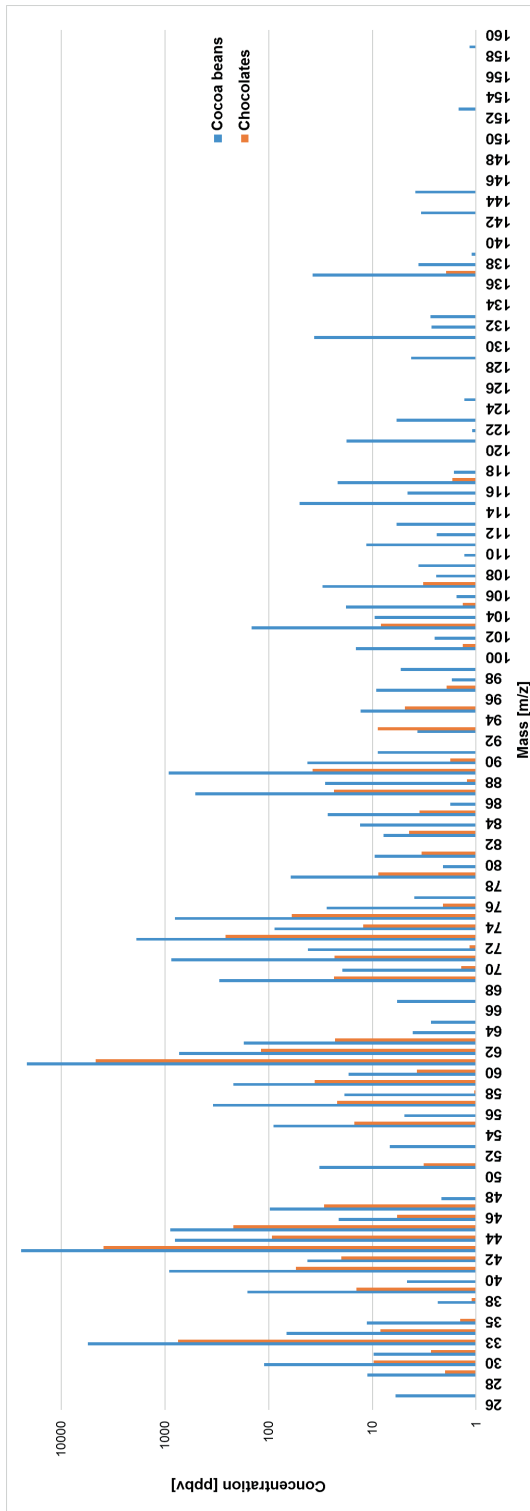
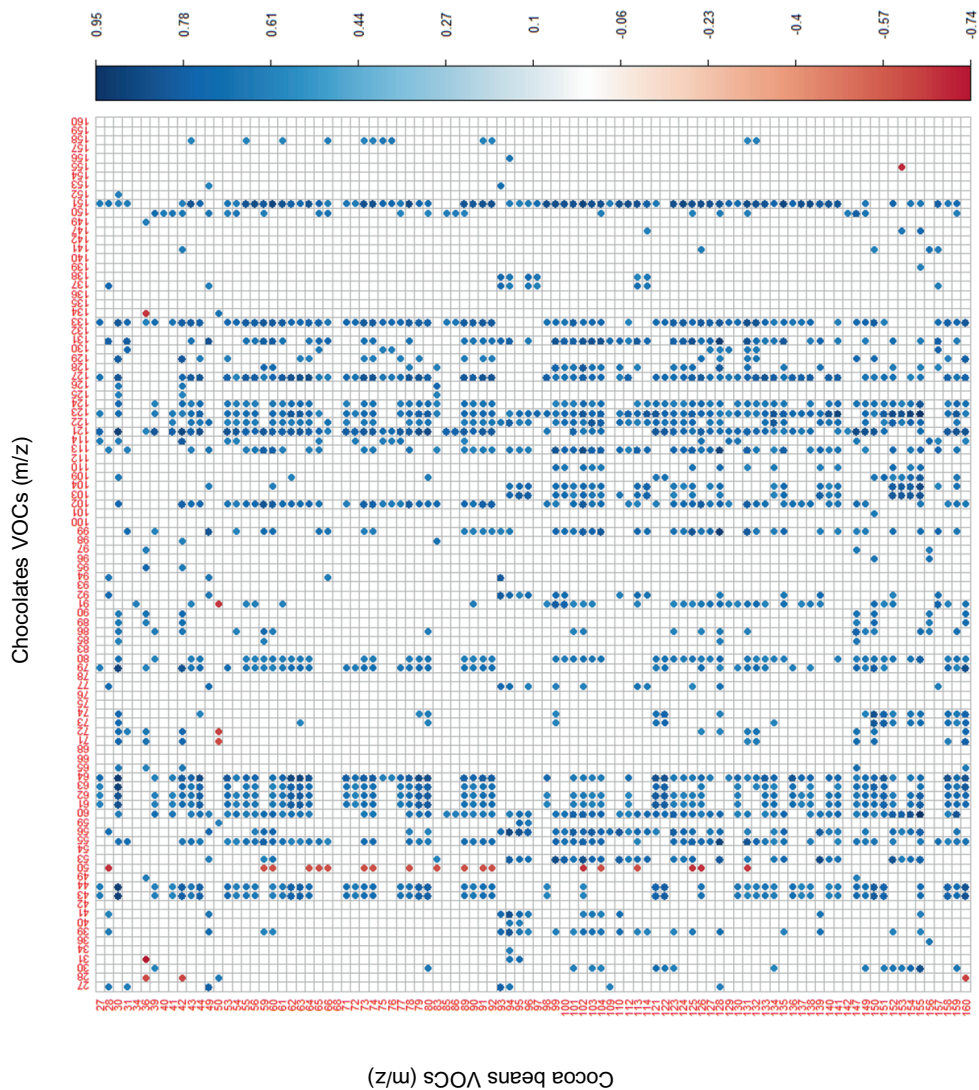**Figure 5.2.** HS-PTR-MS averaged profiles of cocoa beans and chocolates.

Figure 5.3. Significant correlations of VOC data of cocoa beans and corresponding chocolates (Spearman's Rho, p ≤ 0.05).

111

In Table 5.2 a tentative identification of the cocoa beans' masses that showed high correlation with chocolate masses (Spearman's Rho > 0.70; $p \leq 0.05$) is presented. Figure 5.3 shows that the majority of the masses noted for cocoa beans are correlated with certain masses determined in the chocolates, i.e. m/z 43, m/z 61, m/z 79 (benzene), m/z 80 (pyridine), m/z 99 (hexanal), m/z 113 (methylfurfuril alcohol), m/z 121, m/z 123, m/z 127 (maltol) and m/z 131. According to the tentative identifications in Table 5.2, these masses can be linked to compounds found in raw cocoa beans and/or related to an early stage of production, specifically during fermentation and drying. For instance, some of the highly correlated masses, tentatively identified as hexanal (m/z 99) and 2-phenylethanol (m/z 123), are already present in cocoa pulp and possibly influence the profile of fermented beans (Chetschik, Kneubu, Chatelain, Schlu, Bernath, & Hu, 2018). Acetic acid (m/z 61) can also be found in raw beans (Frauendorfer & Schieberle, 2008), however, its concentration, together with m/z 43 can be influenced by different geographical conditions of fermentation and drying (Acierno et al., 2019; Caligiani et al., 2016). 2-Ethenyl-6-methylpyrazine (m/z 121) has been defined as a technological marker related to hot-air roasting (Magagna et al., 2017). As stated before, pyrazine is generated during roasting. However, we have previously detected this compound in fermented and dried cocoa beans (Acierno et al., 2019). Heptanoic acid (m/z 131) has been found at certain conditions of fermentation times and drying temperature as well, including sun drying, artificial drying and convection oven (Rodriguez-Campos et al., 2012). Because of the correlation shown in Figure 5.3, knowing the different drying procedures applied at the growing location this mass could be further investigated as an origin marker (Beckett, 2008). Another interesting correlation was between maltol (m/z 127) in chocolates and the majority of the VOCs in cocoa beans. Maltol being a pyrone is released during drying and roasting via degradation of monosaccharides (Aprotosoaie, Luca, & Miron, 2016). It is usually produced during industrial drying and roasting and could be indicative of this kind of processes.

All the previous volatile compounds found in chocolate are correlated with the corresponding cocoa bean compounds, indicating the possibility to use VOCs for tracing back characteristics of cocoa in chocolate products along the supply chain. Specifically, compounds that are already present in non-fermented cocoa beans such as 2-phenylethanol can be used to directly link raw material and the finished chocolate. The compounds that are formed as a result of primary processes such as fermentation and drying could be used to track and trace the supply chain when the procedures applied at the location of cocoa beans origin are known.

Table 5.2. Tentative identifications and concentrations of volatile compounds of cocoa beans and corresponding chocolates, median values and relative differences.

m/z	Tentative identification	Sum formula	Cocoa beans concentration (ppbv) Median	Chocolates concentration (ppbv) Median	Chocolates-Cocoa beans difference (%) Median
43	Fragment (diverse origin)	$C_3H_7^+$	13291.84	2176.65	-84
59	Acetone/propanal	$C_3H_7O^+$	165.84	34.66	-77
61	Acetic acid	$C_2H_5O_2^+$	13538.21	3069.98	-79
71	Terpene fragment/pentene	C_5H^+	346.85	14.17	-97
73	2-Methylpropanal/Butanone	$C_4H_8O^+$	1056.88	185.53	-84
75	Propanoic acid/methyl-acetate/acetol	$C_3H_7O_2^+$	287.63	53.77	-88
81	Methylpentene/terpene fragment	$C_6H_9^+$	5.49	3.19	-55
85	2(5H)-furanone/	$C_4H_5O_2^+$/	20.93	3.05	-82
	methyl-butenal	$C_5H_9O^+$			
89	Methylpropanoate/2-methylpropanoic acid/hydroxybutanone/acetoin	$C_4H_9O_2^+$	399.44	29.00	-95
91	2,3- Butanediol	$C_4H_{11}O_2^+$	4.17	0.76	-84
93	Terpene fragment/Toluene	$C_7H_9^+$	2.65	1.10	-15
95	Phenol/methylpyrazine	$C_6H_7O^+$	9.64	4.56	-80
96	Dimethyl-pyrrole/ethyl-pyrrole	$C_6H_{10}N^+$	0.66	0.32	-64
97	Furfural/dimethyl-furan	$C_5H_5O_2^+$	7.41	1.80	-77
98	Dimethyl-oxazole	$C_5H_8ON^+$	1.05	0.18	-85
99	Hexenal/methyl-pentenone ;	$C_6H_{11}O^{+/}$	3.14	0.80	-78
	Furfuryl alcohol/a-angelica lactone	$C_5H_7O_2^+$			
101	Dihydro-2-methyl-3(2H)-furanone/2,3-	$C_5H_9O_2^+$	9.51	1.23	-88
	pentanedione/pentanedione/methyl-tetrahydrofuranone				
103	2,3-methylbutanoic acid/2-hydroxy-3-pentanone/	$C_5H_{11}O_2^+$	1.64	0.12	-95
	3-methylbutanoic acid				
104	Benzonitrile	$C_7H_6N^+$	118.57	7.50	-95
109	4-methylphenol/benzyl alcohol;	$C_7H_9O^+$;	7.75	0.43	-69
	2-5 dimethylpyrazine	$C_6H_9N_2^+$			
110	Formil-methylpyrrole/acetylpyrrole/2-acetylpyrrole	$C_6H_8ON^+$	3.49	0.74	-46
112	Trimethyloxazole	$C_6H_{10}ON^+$	0.61	0.34	-95
113	Methylfurfuril alcohol/	$C_6H_9O_2^+$/	1.31	0.08	-91
	Heptanone	$C_7H_{13}O^+$			
121	2-ethenyl-6-methylpyrazine/	$C_7H_9N_2^+$	4.92	0.44	-90
	6,7-dihydro-(5H)-cyclopentapyrazine				
123	2-phenylethanol /	$C_8H_{11}O^+$/	5.62	0.61	-86
	Trimethylpyrazine	$C_7H_{11}N_2^+$			
125	Guaiacol/methyl-benzenediol/furyl acetone;	$C_7H_9O_2^+$;	2.96	0.41	-54
	Butylfuran/methylpropylfuran/octadienone/	$C_8H_{13}O^+$			
	trimethylciclopentenone/alkylfurans/				
	E-2-octenal				

Table 5.2. Continued.

m/z	Tentative identification	Sum formula	Cocoa beans concentration (ppbv) Median	Chocolates concentration (ppbv) Median	Chocolates-Cocoa beans difference (%) Median
127	Maltol/methyl-furoate; 3-ethyl-1-2-cyclopentanedione	$C_6H_6O_3H^+$; $C_7H_{10}O_2H^+$	0.14	0.06	-59
129	Pentenylacetate/heptanedione; furaneol;octanal	$C_7H_{13}O_2^+$; $C_8H_8O_3H^+$; $C_8H_{16}OH^+$	0.62	0.26	-92
131	Ethyl valerate/heptanoic acid/C7 ester	$C_7H_{15}O_2^+$	2.10	0.28	-99
136	Butyl-pyridine/ethyl-propyl pyridine/ ethyl-dimethyl-pyrazine	$C_9H_{14}N^+$	14.36	0.20	-76
137	Various monoterpenes	$C_{10}H_{17}^+$	0.13	0.03	-86
138	Pyridinecarboxylic acid methyl ester	$C_7H_8O_2N^+$	9.98	1.75	-87
139	E,E-2,4-nonadienal/E-2-nonenal	$C_9H_{15}O^+$	1.04	0.20	-59
153	E/E-2,4-decadienal	$C_{10}H_{17}O^+$	0.89	0.30	-90
155	Linalool	$C_{10}H_{18}O^+$	1.40	0.12	-61
159	Nonanoic acid/C9 ester	$C_9H_{19}O_2^+$	0.24	0.08	-94

Significant relative differences between cocoa beans and chocolates are highlighted in blue (Mann–Whitney test, $p \leq 0.05$).

5.3.2 Signatures of elements of cocoa beans and corresponding chocolates

The contents of 24 macro-, micro-, and trace elements were determined in all cocoa bean and chocolate samples (Table 5.3; Table S5.1 and S5.2, Supplementary materials).

Table 5.3. Macro-, micro-, and trace elements (mg/kg) determined in cocoa beans and corresponding chocolates: median and relative differences.

	Cocoa beans (mg/kg)	Chocolates (mg/kg)	Chocolates-Cocoa beans difference* (%)
	Median	Median	Median
Na	57.00	550.00	746
Cr	0.30	0.23	-6
La	0.025	0.000	-25
Ce	0.03	0.01	-33
Mo	0.27	0.16	-34
Cs	0.01	0.01	-35
Ga	0.01	0.00	-38
Ti	0.99	0.38	-43
Y	0.01	0.00	-50
Ba	5.40	1.60	-62
Rb	18.00	6.80	-67
Zn	56.00	17.00	-67
Sr	9.70	3.10	-73
Fe	92.00	11.00	-74
Mg	3200.00	780.00	-75
Al	38.00	2.40	-77
Co	0.48	0.11	-77
Cu	25.00	5.50	-79
Cd	0.56	0.11	-80
Mn	27.00	6.40	-80
Ni	4.70	0.99	-81
As	0.03	0.01	-82
Pb	0.06	0.00	-83
V	0.08	0.00	-100

*Significant relative differences between cocoa beans and chocolates are highlighted in blue (Mann-Whitney test, $p \leq 0.05$).

The chocolates showed a significantly lower content for ~ 72% of the 24 elements in comparison to the cocoa beans. These results are in agreement with previous studies which showed that finished and semi-finished product (chocolate and cocoa butter) had the lowest level of trace elements compared to the cocoa beans and cocoa shells and/or cocoa nibs (Levi et al., 2014). Just the Na content was significantly higher in chocolates with low cocoa solid contents. A negative correlation between the chocolate cocoa content (Spearman's Rho = -0.81; p = 0.002) and Na content was measured. These results could be indicative of the influence of other ingredients on the Na content (the less cocoa content, the more extra ingredients added). Usually, most of the essential trace metals are found in the cocoa bean core while the highest level of toxic trace metals is found in the cocoa bean shells (Pb, Bi and As) (Levi et al., 2014). In our study, fermented and dried beans including the shell were analysed. Therefore, it is possible that some of the trace elements and metals are coming directly from the outer part of the beans. Nevertheless, the small amount of Pb and As found in the chocolate samples could result from small amounts of shells that are processed with the beans (Manton, 2010). The difference in element contents between cocoa bean and chocolate samples correlated well with the cocoa solids content (Spearman's Rho = 0.81; p < 0.001).

This correlation was also found in previous studies (Bertoldi et al., 2016; Levi, 2014). Taking into account the median value of the percentage of difference from chocolates to cocoa beans, as presented in Figure 5.4, it is evident that chocolates with a higher cocoa content differ less from their beans.
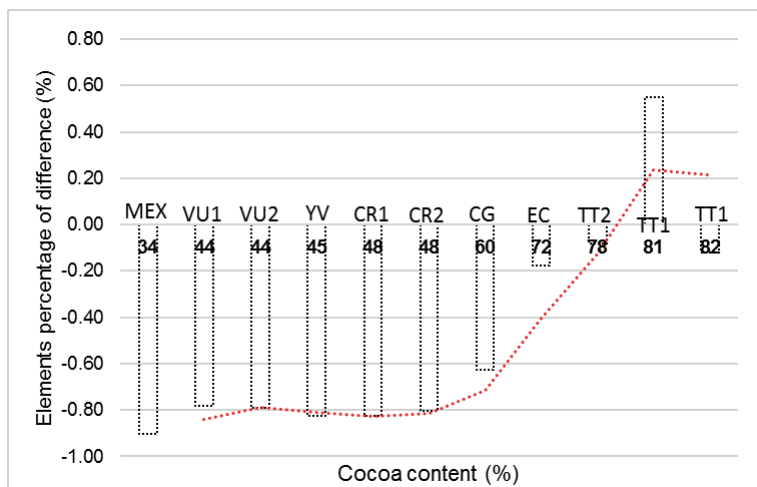
Figure 5.4. Relative differences in elemental composition between cocoa beans and corresponding chocolates. MEX: Mexico; VU1 Vanuatu1; VU2: Vanuatu 2; YV: Venezuela; CR1: Costa Rica1; CR2: Costa Rica2; CG: Congo; EC: Ecuador; TT2: Trinidad 2; TT1: Trinidad1.

Just Cd was not significantly correlated with the chocolate cocoa content (Spearman's Rho = 0.41; p = 0.20). This is in contrast with previous studies (Bertoldi et al., 2016), however, a significant correlation between Cd content in cocoa beans and corresponding chocolates has been measured (Spearman's Rho, p ≤ 0.05) (Figure 5.5). When correlating the elements contents of the cocoa beans and the chocolates, Fe and As contents in cocoa beans showed a significant correlation with the majority of the chocolate elements (Figure 5.5) which makes them an interesting marker for linking beans and chocolates. According to literature, Fe and As are related to the geographical origin of cocoa beans (Araujo et al., 2017; Bertoldi et al., 2016). Fe and As of cocoa beans are significantly correlated with Fe and As of the corresponding chocolates. Other elements that showed this direct correlation are Cr, Mo, and, as stated before, Cd. Furthermore, the latter has been found to be related to the geographical origin of cocoa beans before (Bertoldi et al., 2016).

Elemental fingerprints showed to be mainly informative of the cocoa content. Fe, As, Cr, Mo, and Cd were potentially able to link cocoa beans to chocolates and could be used for tracing back characteristics of cocoa in chocolate products along the supply chain as they are informative of cocoa origin. However, it is important to take into account that the heavy metals such as Fe, As, Cr and Cd are potentially subject to environmental contamination and their concentration can change over time (Bertoldi et al., 2016).
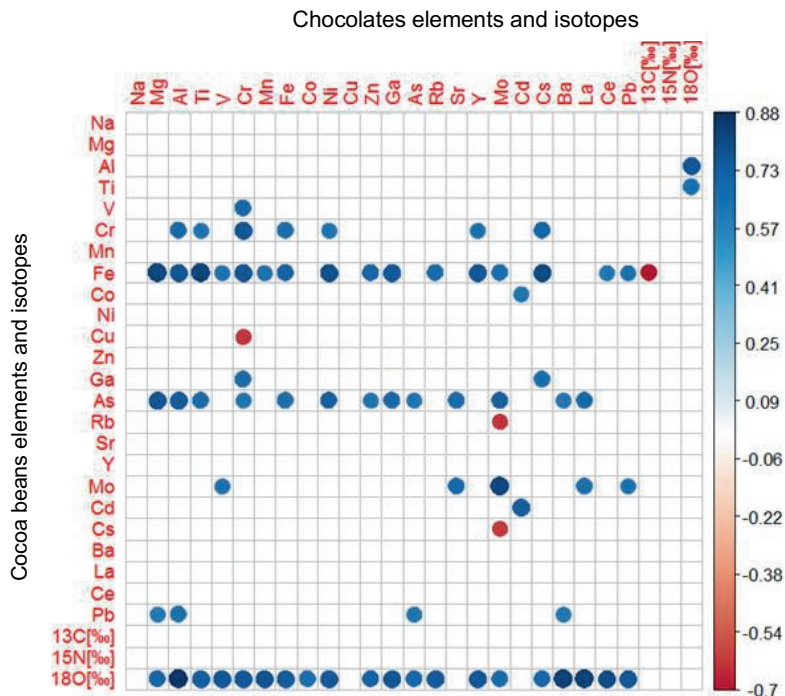
Figure 5.5. Significant correlations of elemental and isotopic data of cocoa beans and corresponding chocolates (Spearman's Rho, p ≤ 0.05).

5.3.3 Signatures of stable isotope ratios in cocoa beans and corresponding chocolates

The stable isotope ratio values of the various cocoa bean samples and chocolates varied considerably. They varied from -32.2 to -27.7‰ for $\delta^{13}C$, from 1.3 to 2.7‰ for $\delta^{15}N$, from -137.7 to -102.2‰ for δ^2H, and from 22.2 to 25.5‰ for $\delta^{18}O$. For the chocolates, they varied from -30.1 to -27.2‰ for $\delta^{13}C$, from 2.1 to 3.5‰ for $\delta^{15}N$, from -117.9 to -105.8‰ for δ^2H, and from 25.9 to 30.4‰ for $\delta^{18}O$ (Figure 5.6; Table S5.3 and S5.4, Supplementary materials).

The mean value of each stable isotope ratio is higher in chocolates compared to the cocoa beans except for δ^2H as is shown in Figure 5.6. The shift from beans to chocolates is most likely due to added ingredients. For instance, in the case of $\delta^{13}C$ the beans are included in the rage of C3 plant within -33 and -22‰ (Spangenberg, 2001). However, the possible enrichment in chocolate can be related to the ingredient that is added, e.g. the addition of oxidable compounds to cocoa butter as foreign fats. Evidence of this is that the chocolate $\delta^{13}C$ ratio is close to the one of cocoa butter equivalent made mainly from vegetable oils (-29.7‰ to -28.5‰) (Spangenberg, 2001). The $\delta^{13}C$ enrichment in chocolate samples can be also related to possible sugar added such as cane and/or corn sugar, derived from C4 plant (-23‰ to -6‰) (Spangenberg, 2001; Bostic et al., 2018). The influence of both fat and added sugars concertation on the carbon stable composition of other food product such milk has been proven in previous studies (Bostic et al., 2018).
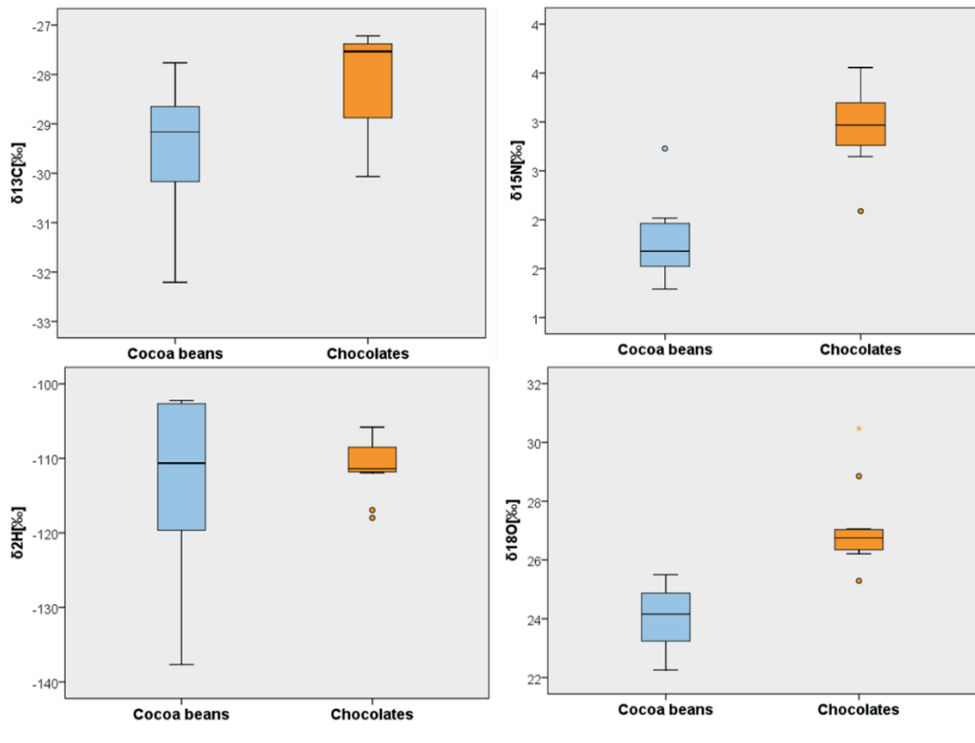
Figure 5.6. Boxplots of $\delta^{13}C$, $\delta^{15}N$, δ^2H, $\delta^{18}O$ of cocoa beans and chocolates.

The isotope ratios of cocoa beans and chocolates were significantly different for all the isotopes analysed except for δ^2H (Table 5.4).

Table 5.4. $\delta^{13}C$, $\delta^{15}N$, δ^2H and $\delta^{18}O$ stable isotope ratio values determined in cocoa beans and corresponding chocolates: median and relative differences.

	Cocoa Beans (‰)	Chocolates (‰)	Chocolates-Cocoa beans difference* (%)
	Median	Median	Median
$\delta^{13}C[‰]$	-29.16	-27.53	-3
$\delta^{15}N[‰]$	1.68	2.97	76
$\delta^2H[‰]$	-110.64	-111.38	-3
$\delta^{18}O[‰]$	24.16	26.75	12

*Significant relative differences between cocoa beans and chocolates are highlighted in blue (Mann-Whitney test, $p \leq 0.05$).

The relative standard deviation (RSD) of $\delta^{13}C$ (RSD cocoa beans: -4%; RSD chocolates: -4%) and $\delta^{18}O$ (RSD cocoa beans: 5%; RSD chocolates: 5%) both for cocoa beans and chocolates is in the same order of magnitude, highlighting a stable dispersion of the values in relation to their means; on the other hand, this is not valid for $\delta^{15}N$ (RSD cocoa beans: 22%; RDS chocolates: 13%). In the latter case, the variation within the cocoa beans samples is mostly related to the N supplied from the soil/fertilizers that is assumed to differ according to the production area and annual temperature and precipitation. However, the large shift between cocoa beans and chocolates $\delta^{15}N$, supported also by the high relative differences (76%) showed in Table 5.4, it is mostly related to the possible addition of protein source to the chocolate (i.e. milk powder or similar).

A trend within cocoa beans and corresponding chocolates can be noticed in Figure 5.7. This is mostly valid for $\delta^{15}N$, $\delta^{13}C$, $\delta^{18}O$ and mainly for the chocolates with high cocoa content (60-80%). This set of samples does not show evident information related to the origin (Figure 5.7). This could be related to the fact that in this study the whole beans (husk and nib) were analysed and previous studies showed that different part of the beans could have a different isotopic signature (Diomande et al., 2015). Furthermore, Figure 5.7 shows differences between cocoa beans from the same origin; this is particularly valid for the δ^2H of cocoa beans indicating a possible bean-specific differentiation related to the crop season and or weather condition. No significant correlations between stable isotope ratios of cocoa beans and chocolates were observed (Figure 5.5). Therefore, stable isotope signatures appear less sufficient for tracing back characteristics of cocoa in chocolate products along the supply chain. To accurately estimate the isotopic signature of multi-ingredient food products the quantity of each ingredient should be available as shown by Bostic et al. (2015).

5.3.4 Effect of cocoa contents on cocoa bean-chocolate relationships

The data of the various measurements were recalculated based on the cocoa contents of the chocolates ((measured value/cocoa content)*100). Some minor changes in individual VOCs and elements cocoa bean-chocolate correlations were observed (Figure S5.1 and S5.2, Supplementary materials).

Before and after correction for the cocoa content, the relative difference in cumulative VOC intensities between cocoa beans and chocolate samples did not show a significant correlation with the cocoa content (Spearman's Rho = 0.42; p = 0.19). After considering the cocoa contents, the majority of the VOCs in the beans still present significant correlations with those in the chocolates, e.g. for masses 43, 61, 79, 80, 113, 123, and 127. However, some differences are evident compared to the non-corrected data thus indicating a possible dilution effect due to the differences in cocoa contents. Interestingly, other masses such as m/z 73 (methylpropanal), m/z 91 (2-3 butanediol) and m/z 109 (4-methylphenol/ benzyl alcohol) show significant correlations with a higher number of cocoa bean masses compared to the non-corrected data (Figure 5.3).

After cocoa content correction, Fe and As show the same trend as previously. Cr, Fe, and Cd content of cocoa beans and corresponding chocolates are still significantly correlated (Spearman's Rho, p ≤ 0.05).

Before and after correcting for the cocoa content, the stable isotope ratios did not show a significant correlation between cocoa beans and corresponding chocolates (Spearman's Rho, p ≤ 0.05). Only $\delta^{13}C$, before the cocoa content correction, showed a significant negative correlation with the cocoa content (Spearman's Rho = -0.75; p < 0.001). The negative correlation indicates that the trend in Figure 5.6 might be partially due to the recipe (the less cocoa content, the more extra ingredients added).
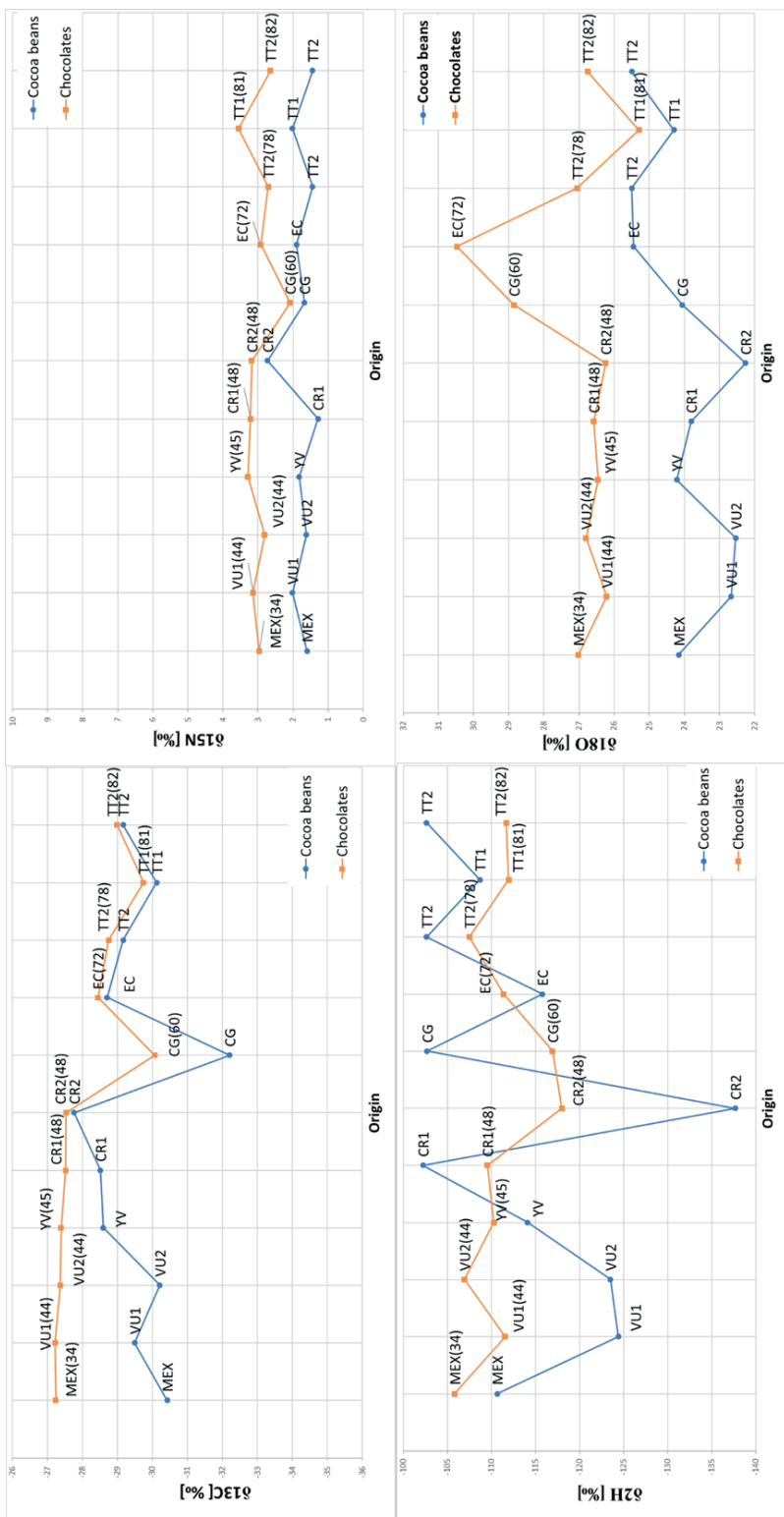
Figure 5.7. Stable isotope ratio values for cocoa beans and corresponding chocolates. The samples are ordered according to the cocoa content (%). Squared shape symbols indicate cocoa bean samples, round shape symbols indicate chocolate samples. MEX: Mexico; VU1 Vanuatu1; VU2: Vanuatu 2; YV: Venezuela; CR1: Costa Rica1; CR2: Costa Rica2; CG: Congo; EC: Ecuador; TT2: Trinidad 2; TT1: Trinidad2.

5.3.5 Comparison of volatile, elemental and isotopic signatures

In order to link cocoa beans and chocolates, a marker should be able to significantly differentiate beans and chocolates but should also show a consistent correlation between them. Considering the masses showing significant differences in concentration between cocoa beans and chocolates, the VOCs showed a larger relative difference between cocoa beans and chocolates (125 masses, median value ~-85%) compared to the elemental fingerprints (median value ~-77%) and isotope ratios (median value ~12%) (Table 5.3 and 5.4). The VOC fingerprint showed more variance among the cocoa beans samples (RSD: median value ~100%) compared to the chocolates (RSD median value ~52%). Since cocoa beans are a natural product high variation is expected. The elemental composition highlights more variance among the chocolate samples (RSD: median value ~72%) compared to the cocoa beans (RSD: median value ~37%); considering the behaviour of Na (paragraph 5.3.2), this could be related to possible extra ingredients added during processing. On the other hand, the relative variance of the isotope ratio values was similar for cocoa bean samples (RSD: median value ~5%) and chocolate samples (RSD: median value ~5%). These results highlight that the magnitude of dissimilarity between cocoa beans and chocolate samples tends to be higher for the VOCs and elemental composition compared to the isotopic ratios. In order to link cocoa beans and chocolates, a consistent correlation between markers and beans should also exist. Table 5.5 highlights that VOCs measured by using HS-PTR-MS are able to distinguish bean and chocolate samples according to the origin and show a consistent shift from beans to chocolates. This can be better visualised in the PCA in Figure 5.8 built with HS-PTR-MS results. Prior to PCA, the data were auto-scaled and normalized to balance the differences in VOC concentration between cocoa beans and chocolates.

Table 5.5. Evaluation of differences between cocoa beans and chocolates for volatiles, elements, and stable isotope data.

	% Difference	Consistent shift	Cocoa content influence
VOCs	**+** (~85)	**+**	**-**
Elements	**±** (~77)	**±**	**+**
Stable isotope	**-** (~12)	**-**	**±**

The variability within samples is mainly represented in PC1, while PC2 differentiates raw materials and the finished products. Along with PC1, connection points within the cocoa beans and the corresponding chocolates can be distinguished. The shift of the volatile signatures of the cocoa beans to corresponding chocolates is represented by the direction of the arrows. The latter underline a link between the cocoa beans and the corresponding chocolates which is supported by the significant correlations between the VOCs of cocoa beans and chocolates as shown in Figure 5.3 and 5.8. Considering the possible connection points on PC1, a significant correlation (Pearson's correlation coefficient = 0.65) between cocoa beans and chocolates has been determined using the score values of each sample on PC1. This further highlights the consistent shift of VOCs fingerprints from cocoa beans to chocolates. The trend in the PCA in Figure 5.8 is not visible for the elemental composition and the stable isotope ratio data nor exists a significant correlation between cocoa beans and chocolates (Pearson's correlation coefficient, $p \geq 0.05$).
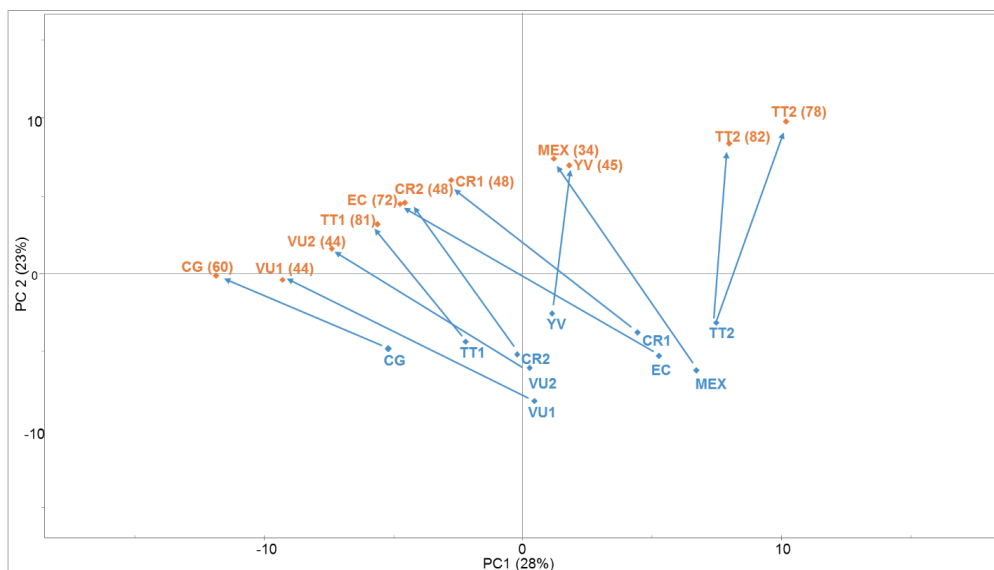
Figure 5.8. First two dimensions of Principal Component Analysis plot of normalized and auto-scaled HS-PTR-MS data of cocoa beans from different geographical origins and their corresponding chocolates samples. The numbers in brackets indicate the cocoa content (% w/w). The arrows indicate the shift in VOCs from cocoa beans to chocolates.
MEX: Mexico; VU1 Vanuatu1; VU2: Vanuatu 2; YV: Venezuela; CR1: Costa Rica1; CR2: Costa Rica2; CG: Congo; EC: Ecuador; TT2: Trinidad 2; TT1: Trinidad1.
Blue colour indicates cocoa beans samples, orange colour indicates chocolate samples.

When correcting for the cocoa content, the cocoa beans-chocolates scores correlation is still significant for VOCs (Pearson's correlation coefficient = 0.65). After the cocoa content correction, the correlations for the elemental data do not change. The correlations for the stable isotope ratios improve (Pearson's correlation coefficient = 0.37). However, the correlation is weaker than the correlation for the VOCs.

Figure 5.9 presents the selection of VOCs, elements that show significant correlations between beans and chocolates, both with and without cocoa content correction. These could be used as markers indicative of cocoa beans traits that significantly persist in the final product. Within the VOC group, it is possible to highlight compounds retained along with the production from both raw material and primary processing at the origin, while within the selected elements heavy metals appear informative of the geographical origin of the cocoa beans. M/z 79 and 123, together with Cr, Fe and Cd showed the highest direct correlation between cocoa beans and chocolates (Spearman's Rho > 0.70). Within the elements, Cd in particular, was able to statistically differentiate cocoa beans and chocolates (Mann-Whitney test, $p \leq 0.05$). At the same time, Cd shows a significant correlation between its contents in bean and corresponding chocolates (Spearman's Rho, $p \leq 0.05$) independently from the cocoa content. Apart from its toxicity, Cd could be informative of the soil at the growing location of the cocoa beans (Araujo et al., 2017). The traits presenting strong correlations between beans and chocolates could be used to track and trace beans along the supply chain and support the consideration of applying the concept of 'terroir' to chocolate as an indication of processing location and a growing environment (Engeseth & Fernando Ac Pangan, 2018; Sukha et al., 2014).
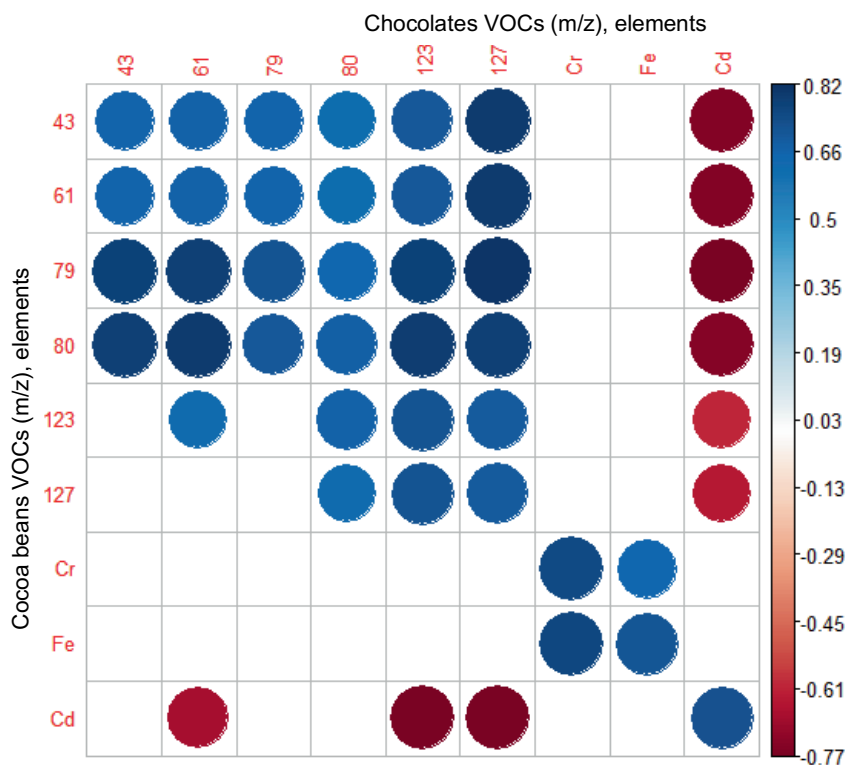
Figure 5.9. VOCs and elements showing a significant correlation between cocoa beans and corresponding chocolates (Spearman's Rho, $p \leq 0.05$).

5.4 Conclusions

When looking for intrinsic information carried from cocoa beans to chocolates, the volatile signatures generated by HS-PTR-MS revealed a constant shift between raw materials and the finished products as well as a high magnitude of dissimilarity between samples of different origins. The group of key VOC markers are represented by acetic acid (m/z 61), benzene (m/z 79), pyridine (m/z 80), 2-phenylethanol (m/z 123) and maltol (m/z 127). On the other hand, the elemental composition and isotopic data generally carried the information related to the cocoa content and ingredients added. Within the elemental group, potential geographical markers are Fe, Cr and Cd. However, we have to keep in mind that their concentrations are potentially susceptible to environmental contamination. According to univariate and multivariate correlations, VOCs appear to be establishing the most robust link between cocoa beans and corresponding chocolates. This provides potential use of track and trace options of cocoa beans from farm to chocolates. These results support the consideration of applying the concept of 'terroir' to chocolate and provides a scientific basis for cacao quality certification programmes.

Acknowledgements

References

Acierno, V., Alewijn, M., Zomer, P., & van Ruth, S. M. (2018). Making cocoa origin traceable: Fingerprints of chocolates using Flow Infusion - Electro Spray Ionization - Mass Spectrometry. *Food Control*, *85*. https://doi.org/10.1016/j.foodcont.2017.10.002

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

Acierno, V., Fasciani, G., Kiani, S., Caligiani, A., & van Ruth, S. M. (2019). PTR-QiToF-MS and HSI for the characterisation of fermented cocoa beans from different origins. *Food Chemistry*, *289* (September 2018), 591–602. https://doi.org/10.1016/j.foodchem.2019.03.095

Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2009). Matrix effects on flavour volatiles release in dark chocolates varying in particle size distribution and fat content using GC–mass spectrometry and GC–olfactometry. *Food Chemistry*, *113*(1), 208–215. https://doi.org/10.1016/j.foodchem.2008.07.088

Aprotosoaie, A. C., Luca, S. V., & Miron, A. (2016). Flavor Chemistry of Cocoa and Cocoa Products-An Overview. *Comprehensive Reviews in Food Science and Food Safety*, *15*(1), 73–91. https://doi.org/10.1111/1541-4337.12180

Araujo, Q. R. De, Baligar, V. C., Loureiro, G. A. H. D. A., & Júnior, J. O. D. S. (2017). Impact of soils and cropping systems on mineral composition of dry cacao beans. *Journal of soil science and plant nutritio*n, *17*(2), 410–428.

Ascrizzi, R., Flamini, G., Tessieri, C., & Pistelli, L. (2017). From the raw seed to chocolate: Volatile profile of Blanco de Criollo indifferent phases of the processing chain. *Microchemical Journal*, *133*, 474–479.

Beckett, S. (2008). Chococlate ingridients. In Beckett S. (Eds), *The science of chocolate* (pp. 11-38) Cambridge, UK: The Royal Society of Chemistry, 2008.

Bertoldi, D., Barbero, A., Camin, F., Caligiani, A., & Larcher, R. (2016). Multielemental fingerprinting and geographic traceability of Theobroma cacao beans and cocoa products. *Food Control*, *65*, 46–53. https://doi.org/10.1016/j.foodcont.2016.01.013

Bostic, J.N., Palafox, S.J., Rottmueller, M.E., & Jaren, A.H. (2015). Effect of baking and fermentation on the stable carbon and nitrogen isotope ratios of grain-based food. *Mass Spectrometry*, *29*, 937-947. doi.org/10.1002/rcm.7178

Bostic, J.N., Hagopian, W.M., & Jaren, A.H. (2018). Carbon and nitrogen stable isotopes in U.S. milk: Insight into production process. *Rapid comunication in mass spectrometry*, *32*, 561-566. doi: 10.1002/rcm.8069

Braga S., Oliveira L.F, JHashimoto J.C., Gama, M.R., Efraim, P., Poppi, R.J, & Augusto, F. (2018). Study of volatile profile in cocoa nibs, cocoa liquor and chocolate on production process using GC × GC-QMS. *Microchemical Journal, 141*, 353–361.

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

Caligiani, A., Palla, L., Acquotti, D., Marseglia, A., & Palla, G. (2014). Application of 1 H NMR for the characterisation of cocoa beans of different geographical origins and fermentation levels. *Food Chemistry*, *157*, 94–99. https://doi.org/10.1016/j.foodchem.2014.01.116

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*(3), 1478–1483. https://doi.org/10.1021/jf903471e

Chetschik, I., Kneubu, M., Chatelain, K., Schlu, A., Bernath, K., & Hu, T. (2018). Investigations on the aroma of cocoa pulp (Theobroma cacao L.) and its influence on the odor of fermented cocoa beans. *Journal of Agricultural and Food Chemistry*, 66, 2467-2472. https://doi.org/10.1021/acs.jafc.6b05008

Counet, C., Callemien, D., Ouwerx, C., & Collin, S. (2002). Use of gas chromatography-olfactometry to identify key odorant compounds in dark chocolate. Comparison of samples before and after conching. *Journal of Agricultural and Food Chemistry*, *50*(8), 2385–2391. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11929301

Davies, A. M. C., Franklin, J. G., Grant, A., Griffiths, N. M., Shepherd, R., & Fenwick, G. R. (1991). Prediction of chocolate quality from near-infrared spectroscopic measurements of the raw cocoa beans. *Vibrational Spectroscopy*, *2*(2–3), 161–172. https://doi.org/10.1016/0924-2031(91)85022-F

Diomande, D., Antheaume, I., Leroux, M., Lalande, J., Balayssac, S., Remaud, G. S., & Tea, I. (2015). Multi-element, multi-compound isotope profiling as a means to distinguish the geographical and varietal origin of fermented cocoa (Theobroma cacao L.) beans. *Food Chemistry*, *188*, 576–582. https://doi.org/10.1016/j.foodchem.2015.05.040

Ducki, S., Miralles-Garcia, J., Zumbé, A., Tornero, A., & Storey, D. M. (2008). Evaluation of solid-phase micro-extraction coupled to gas chromatography-mass spectrometry for the headspace analysis of volatile compounds in cocoa products. *Talanta*, *74*(5), 1166–1174. https://doi.org/10.1016/j.talanta.2007.08.034

Engeseth, N. J., & Pangan, M., F. (2018). Current context on chocolate flavor development — a review. *Current Opinion in Food Science*, *21*, 84–91. https://doi.org/https://doi.org/10.1016/j.cofs.2018.07.002

Frauendorfer, F., & Schieberle, P. (2008). Changes in key aroma compounds of Criollo cocoa beans during roasting. *Journal of Agricultural and Food Chemistry*, *56*(21), 10244–10251. https://doi.org/10.1021/jf802098f

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

Isabel, C., Maia, R., Miranda, M., Ribeirinho, M., Nogueira, J. M. F., & Ma, C. (2009). Stable isotope analysis for green coffee bean : A possible method for geographic origin discrimination. *Journal of Food Composition and Analysis*, *22*, 463–471. https://doi.org/10.1016/j.jfca.2008.06.010

Levi, R., Sela, H., Borojovich, E. J. C., Zakon, Y., Saphier, M., Nikolski, A., Gutflais, E., Lorber, A., & Karpas, Z. (2014). Trace elements in cocoa solids and chocolate : An ICPMS study. *Talanta*, *119*, 1–4. https://doi.org/10.1016/j.talanta.2013.10.048

Liu, N., Koot, A., Hettinga, K., Jong, J. De, & van Ruth, S. M. (2018). Portraying and tracing the impact of different production systems on the volatile organic compound composition of milk by PTR- (Quad) MS and PTR- (ToF) MS. *Food Chemistry*, *239*, 201–207. https://doi.org/10.1016/j.foodchem.2017.06.099

Luykx, D. M. a. M., & van Ruth, S. M. (2008). An overview of analytical methods for determining the geographical origin of food products. *Food Chemistry*, *107*(2), 897–911. https://doi.org/10.1016/j.foodchem.2007.09.038

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

Manton, W. I. I. M. (2010). Determination of the provenance of cocoa by soil protolith ages and assessment of anthropogenic lead Contamination by Pb / Nd and lead isotope ratios. *Journal of Agricultural and Food Chemistry*, *58*, 713–721. https://doi.org/10.1021/jf9029207

Pätzold, R., & Brückner, H. (2006). Gas chromatographic determination and mechanism of formation of D-amino acids occurring in fermented and roasted cocoa beans, cocoa powder, chocolate and cocoa shell. *Amino Acids*, *31*(1), 63–72. https://doi.org/10.1007/s00726-006-0330-1

Qin, X. W., Lai, J. X., Tan, L. H., Hao, C. Y., Li, F. P., He, S. Z., & Song, Y. H. (2017). Characterization of volatile compounds in Criollo, Forastero, and Trinitario cocoa seeds (Theobroma cacao L.) in China. *International Journal of Food Properties*, *20*(10), 2261–2275. https://doi.org/10.1080/10942912.2016.1236270

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

Romano A. (2017). Aroma release during in-mouth process. In Guichard, E., Salles, C., Morzel, M., & Le Bon A. (Eds), *Flavour: From food to perception* (pp. 235-265). Chichester, West Sussex, UK: Johnn Willey & Sons.

Serra, F., Guillou, C. G., Reniero, F., Ballarin, L., Cantagallo, M. I., Wieser, M., Iyer, S.S., Hèberger, K., & Vanhaecke, F. (2005). Determination of the geographical origin of green coffee by principal component analysis of carbon, nitrogen and boron stable isotope ratios, *Rapid Comunication in Mass Spectrometry*, *19*, 2111–2115. https://doi.org/10.1002/rcm.2034

Spangenberg, J. E., & Dionisi, F. (2001). Characterization of cocoa butter and cocoa butter equivalents by bulk and molecular carbon isotope analyses : Implications for vegetable fat quantification in chocolate. J*ournal of Agricultural and Food Chemistry*, *49*, 4271–4277.

Sukha, D. A., Butler, D. R., Comissiong, E. A., & Umaharan, P. (2014). The impact of processing location and growing environment on flavor in cocoa (Theobroma cacao L.) – Implications for " Terroir " and Certification – Processing Location Study. *International Society for Horticultural Science, 1047*, 255–262.

Yener, S., Navarini, L., Lonzarich, V., Cappellin, L., Märk, T. D., Bonn, G. K., & Biasioli, F. (2016). Monitoring single coffee bean roasting by direct volatile compound analysis with proton transfer reaction time-of-flight mass spectrometry. *Journal of Mass Spectrometry*, *51*(9), 690–697. https://doi.org/10.1002/jms.3825

Ziegleder, G. (1991). Composition of flavor extracts of raw and roasted cocoas. *Zeitschrift für Lebensmittel-Untersuchung und-Forschung*, *192*, 521–525.
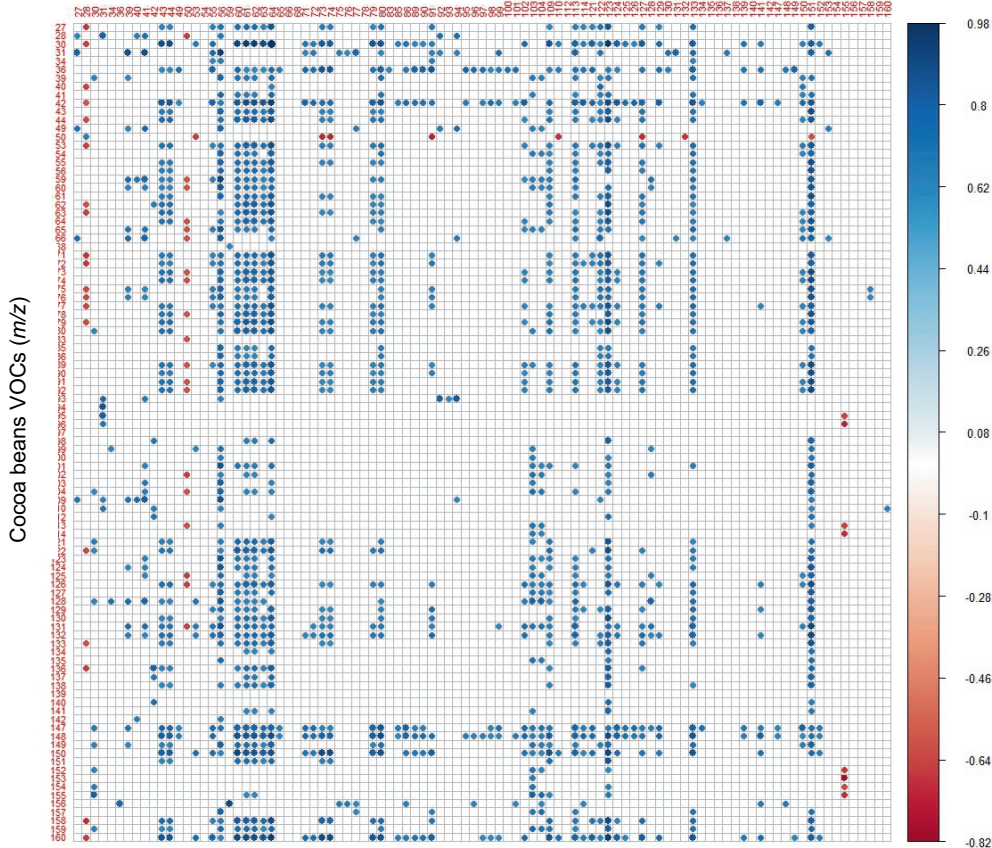
Figure S5.1. Significant correlation between VOCs values of cocoa beans and corresponding chocolates after cocoa content correction (Spearman's Rho, p ≤ 0.05).
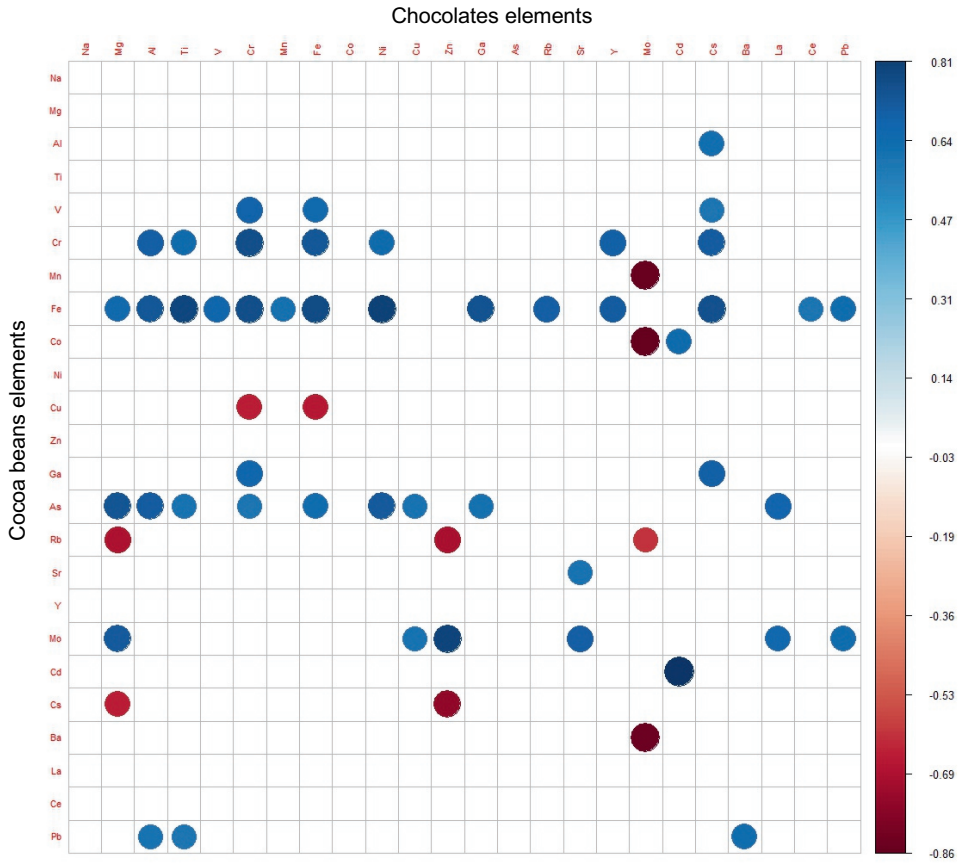
Figure S5.2. Significant correlation between elements values of cocoa beans and corresponding chocolates after cocoa content correction (Spearman's Rho, p ≤ 0.05).

Table S5.1. Macro-, micro-, and trace elements (mg/kg) determined in cocoa beans.

Cocoa beans	Na	Mg	Al	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	As	Rb	Sr	Y	Mo	Cd	Cs	Ba	La	Ce	Pb
CG	37	3000	110	2	0.23	1.4	27	140	0.48	4.5	21	50	0.04	0.03	55	4.9	0.02	0.09	0.1	0.13	3.7	0.06	0.11	0.06
MEX	67	3100	65	1.4	0.06	0.2	26	52	0.55	4.7	26	39	0.01	0.02	28	5.9	0.03	0.27	1.3	0.04	1.4	0.03	0.06	0.03
ECU	57	3100	65	3.3	0.16	0.95	17	76	0.42	0.38	20	72	0.01	0.02	18	11	0.01	0.54	0.59	0.01	5.4	0	0.02	0.09
YV	57	3000	63	2.4	0.25	1.2	35	95	0.45	6.2	20	51	0.03	0.05	25	16	0.03	0.18	0.6	0.07	11	0.06	0.12	0.16
CR1	56	4000	16	0.43	0.02	0.08	36	39	1.9	9	29	52	0	0	39	18	0	0.26	1.9	0.02	6.9	0	0	0
CR2	56	3200	8.2	0.39	0.01	0.08	41	33	0.84	0.86	32	61	0	0	29	7.6	0	0.06	0.52	0.04	6.6	0	0	0.02
VU1	45	3400	5.7	0.47	0.02	0.11	15	65	0.17	1.6	25	56	0	0.04	6.6	8.9	0	0.38	0.36	0	1.8	0	0	0.02
VU2	78	3400	44	2.6	0.15	0.26	16	92	0.22	2.2	22	81	0.01	0.02	12	9.7	0.01	0.2	0.2	0.01	1.4	0	0.01	0.09
TT1	47	3200	29	0.87	0.14	0.3	20	120	0.24	6.2	21	54	0.01	0.22	7.1	7.1	0.02	0.54	0.12	0.01	0.29	0.02	0.03	0.05
TT2	59	3200	38	0.99	0.08	0.82	27	120	0.69	8.1	26	56	0.01	0.11	17	10	0.01	0.29	0.56	0.01	5.6	0.02	0.06	0.11

CG: Congo; MEX: Mexico; VU1 Vanuatu1; VU2: Vanuatu 2; YV: Venezuela; CR1: Costa Rica1; CR2: Costa Rica2; CG: Congo; EC: Ecuador; TT2: Trinidad 2; TT1: Trinidad1.

Table S5.2. Macro-, micro-, and trace elements (mg/kg) determined in cocoa beans.

Chocolates	Na	Mg	Al	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	As	Rb	Sr	Y	Mo	Cd	Cs	Ba	La	Ce	Pb
CG(60)	47	1000	29	3	0.09	1.4	9.5	99	0.11	1.8	8.1	18	0.01	0	18	1.9	0.01	0.16	0.04	0.04	1.6	0	0.02	0.02
MEX(34)	760	540	1.7	0.27	0	0.11	2.8	4.5	0.05	0.4	2.5	13	0	0	4.6	1.2	0	0.12	0.13	0.01	0.33	0	0	0
ECU(72)	80	1800	41	1.9	0.13	2.2	14	160	0.44	2	15	40	0.01	0.03	17	6.6	0.01	0.23	0.22	0.03	3.9	0.02	0.04	0.05
YV(45)	690	750	2.4	0.26	0	0.23	6	11	0.08	0.99	4	16	0	0.01	6	3.1	0	0.12	0.11	0.01	2	0	0	0
CR1(48)	620	620	1.2	0.12	0	0.06	4.6	6.8	0.21	0.98	5.5	15	0	0	6.8	3.3	0	0.11	0.26	0	1.2	0	0	0
CR2(48)	550	580	1.4	0.22	0	0.09	6.4	8.7	0.11	0	5.1	16	0	0	5.7	1.8	0	0.08	0.08	0.01	1.3	0	0.01	0
VU1(44)	630	740	1.3	0.27	0	0.08	2.9	8.2	0.03	0.19	5.2	17	0	0	3.1	2.4	0	0.18	0.07	0	0.57	0	0	0
VU2(44)	660	780	1.5	0.38	0	0.16	3.2	8.8	0.05	0.42	4.6	17	0	0.01	4	2.2	0	0.14	0.04	0.01	0.53	0	0	0
TT1(81)	110	2100	61	1.9	0.21	4.5	19	260	0.58	5.5	21	39	0.02	0.04	19	7.2	0.01	0.36	0.07	0.04	3.3	0.03	0.07	0.08
TT2(78)	100	2100	42	4.4	0.11	1.4	21	120	0.49	5.3	20	40	0.02	0.03	20	7	0.01	0.19	0.29	0.03	4.7	0.04	0.08	0.04
TT2(82)	86	2000	41	4.5	0.1	1.3	22	110	0.49	5.4	20	42	0.01	0.01	20	7.1	0.01	0.19	0.32	0.03	4.7	0.03	0.07	0.04

The numbers in brackets indicate the cocoa content (% w/w).
CG: Congo; MEX: Mexico; VU1 Vanuatu1; VU2: Vanuatu 2; YV: Venezuela; CR1: Costa Rica1; CR2: Costa Rica2; CG: Congo; EC: Ecuador; TT2: Trinidad 2; TT1: Trinidad1.

Table S5.3. δ13C, δ15N, δ2H and δ18O stable isotope ratio values determined in cocoa beans.

	Stable isotope ratio ‰			
Cocoa beans	δ13C[‰]	δ15N[‰]	δ2H[‰]	δ18O[‰]
CG	-32.21	1.68	-102.70	24.06
MEX	-30.42	1.60	-110.64	24.16
ECU	-28.70	1.91	-115.78	25.45
YV	-28.60	1.83	-114.10	24.22
CR1	-27.76	2.73	-137.68	22.26
CR2	-28.52	1.29	-102.22	23.81
VU1	-29.50	2.02	-124.41	22.67
VU2	-30.21	1.62	-123.50	22.54
TT1	-30.13	2.02	-108.69	24.30
TT2	-29.16	1.44	-102.64	25.50

CG: Congo; MEX: Mexico; VU1 Vanuatu1; VU2: Vanuatu 2; YV: Venezuela; CR1: Costa Rica1; CR2: Costa Rica2; CG: Congo; EC: Ecuador; TT2: Trinidad 2; TT1: Trinidad1.

Table S5.4. δ13C, δ15N, δ2H and δ18O stable isotope ratio values determined in chocolates.

	Stable isotope ratio ‰			
Chocolates	δ13C[‰]	δ15N[‰]	δ2H[‰]	δ18O[‰]
CG(60)	-30.07	2.09	-116.94	28.85
MEX(34)	-27.24	2.97	-105.82	27.02
ECU(72)	-28.44	2.92	-111.38	30.47
YV(45)	-27.39	3.29	-110.28	26.46
CR1(48)	-27.53	3.18	-117.99	26.24
CR2(48)	-27.53	3.21	-109.54	26.58
VU1(44)	-27.22	3.15	-111.54	26.21
VU2(44)	-27.37	2.81	-106.94	26.8
TT1(81)	-29.75	3.56	-111.95	25.29
TT2(78)	-29.00	2.65	-111.68	26.75
TT2(82)	-28.76	2.71	-107.49	27.05

The numbers in brackets indicate the cocoa content (% w/w).

CG: Congo; MEX: Mexico; VU1 Vanuatu1; VU2: Vanuatu 2; YV: Venezuela; CR1: Costa Rica1; CR2: Costa Rica2; CG: Congo; EC: Ecuador; TT2: Trinidad 2; TT1: Trinidad1.

Chapter 6

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

Abstract

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

6.1 Introduction

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

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

Mass spectrometry (Qin et al., 2017), spectroscopic (Marseglia et al., 2016) and sensor techniques (Teye, Huang, Dai, & Chen, 2013) have been used to authenticate cocoa beans origin. Nevertheless, with the constant increase of authenticity challenges over the past decades (Lohumi, Lee, Lee, & Cho, 2015), the development of more new, fast and reliable analytical methods for geographical origin assessment is highly desirable. So far, looking for unique characteristics of cocoa beans originating from different continents, machine olfaction by of Proton Transfer Reaction-Mass Spectrometry (PTR-MS) for volatile compounds detection and machine vision by Hyperspectral Imaging (HSI) for spectral and spatial investigation has never been tested. The volatile compounds and spectral information have been widely used as sensitive and fast analytical techniques for authentication and quality analysis of food. PTR-Time of Flight-MS (PTR-ToF-MS) has been used to detect the geographical origin of several products such as coffee (Yener et al., 2014), saffron (Masi et al., 2016), olives (Masi, Romani, Pandolfi, Heimler, & Mancuso, 2015), and wines (Campbell-Sills et al., 2016). Moreover, this instrument has been used for chocolate analysis offering the possibility to compare cocoa and chocolate profiles. HSI has been successfully applied for the analysis of several food commodities (Lohumi et al., 2015) such as coffee beans (Cho, Bae, Cho, & Moon, 2017), tea (Zhao, Chen, Cai, & Ouyang, 2009) and turkey ham (Elmasry, Iqbal, Sun, Allen, & Ward, 2011).

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

6.2 Materials and methods

6.2.1 Materials

6.2.1.1 Samples

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

6.2.2 Methods

6.2.2.1 Machine olfaction: PTR-QiToF-MS

6.2.2.1.1 Instrumental conditions
      Cocoa beans samples were powdered using an electrical food grater. To prevent the grinder overheating, it was stopped every 15 s. The powder of each sample was sieved with a 630 microns (μm) mesh and kept at 4°C before analysis. For the measurement, 0.1 g of ground cocoa beans was weighed into clean and odourless flasks of 250 ml. The closed flasks were placed in a water bath at 25 °C for 30 min to equilibrate the samples with their headspace.
The headspace measurements were performed using a PTR-QiToF-MS (Ionicon Analytik G.m.b.H., Innsbruck, Austria). Ionisation was carried out under drift tube conditions of 900 ± 1 V, drift pressure of 3.80 mbar, temperature 60 °C, and E/N value of 120 Td. Data acquisition was carried out at 1 spectrum per second. For each sample, a mass range between 0.00 and 512.15 was measured using a dwell time of 0.1 s mass^{-1}. Sampling was performed at a flow rate of 60 ± 1 sccm. For all samples, the headspace was measured for 1 min and the measurements were executed in three replicates. The first 10 s were subtracted as blank to the next 40 s of acquisition and the last 10 s were not included in the analysis. The average of the three replicates was taken into account for each sample.

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

6.2.2.1.3 Statistical analysis

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

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

6.2.2.2 Machine vision: HSI

6.2.2.2.1 Instrumental conditions
      A Specim IQ hyperspectral line scan camera (Specim, Spectral Imaging Ltd.) was applied for HSI. Imaging was performed with exposure of 30ms, in the spectral range of 400–1000 nm, and with 2–3 nm spectral steps. A halogen-based lighting system was used to covers the full 400 to 1000nm range. White references were utilized to calibrate captured images.

The resulting image was a 3D data cube (X, Y, Z), two dimensions (X: 512 pixels and Y: 512 pixels) were utilized for spatial information and one for spectral bands (Z: 204 pixels). Each pixel represents a complete spectrum reflecting specific information against wavelengths. Before the measurements, homogeneity of the lighting system and the sensitivity of the camera on the scanned view were evaluated and confirmed.

6.2.2.2.2 Data treatment

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

6.2.2.2.3 Statistical analysis

6.2.2.2.3.1 Multivariate analysis: principal component analysis (PCA)

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

6.2.2.3 Machine olfaction and machine vision comparison

Correlations between VOCs and wavelengths were evaluated by Pearson's correlation coefficients (r). T-test was used to determine significant correlations ($p < 0.05$) (Liu et al., 2018). All the statistical analyses of data were conducted using R 3.3.3 (R Foundation for Statistical Computing, Vienna, Austria).
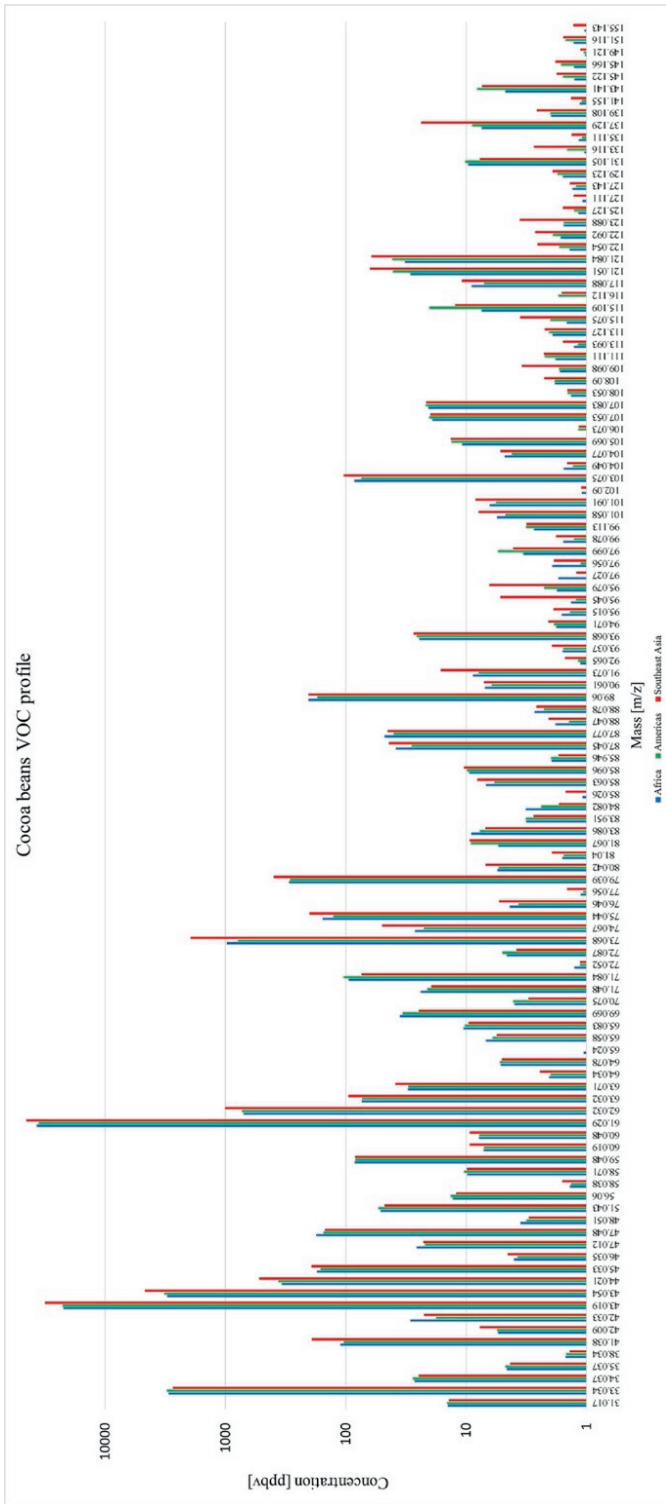
Figure 6.1. Mean PTR-QiToF-MS mass spectra of the volatile compounds of African, American and Southeast Asian cocoa beans. Note, the chart has a logarithmic scale.

6.3 Results and discussion

6.3.1 Machine olfaction: PTR-QiToF-MS

6.3.1.1 Qualitative geographical comparison

6.3.1.1.1 Preliminary comparison of VOC mass spectra
The cocoa beans from all 23 provenances were analysed by PTR-QiToF-MS. The headspace, measured in the range 0.00–521.15 m/z, amounted to 415 mass peaks overall. The dataset was filtered based upon average concentration, the 114 most abundant peaks (higher than 1 ppb) were retained and a final profile including masses from m/z 31.017 to 155.143 was obtained. The profiles of the averaged PTR-QiToF-MS mass spectra of Af, Am and As cocoa beans are presented in Figure 6.1.

The highest concentrations were observed for the ions m/z 33.033, 43.019, 43.054 and 61.028 in all samples. These masses were tentatively identified as methanol (33.033), a fragment of diverse origin (43.019, 43.054), and acetic acid (61.028). Figure 6.1 also shows m/z 73.068 at high concentration. This ion is tentatively identified as 2-methylpropanal, a flavour active compound with a strong chocolate character (Afoakwa, Paterson, Fowler, & Ryan, 2008). In Table 6.1, the averaged concentrations of the tentatively identified compounds are indicated for each continent. Acetic acid was also detected as the most abundant odorant in unroasted beans in another study (Frauendorfer & Schieberle, 2008). Moreover, masses 33.033, 43.019, and 61.028 were found to be predominant in previous studies focused on chocolates manufactured with cocoa beans from different origins analysed using PTR-QMS (Acierno, Yener, Alewijn, Biasioli, & van Ruth, 2016) and PTR-QiToF-MS (Acierno, Liu, Alewijn, Stieger, & van Ruth, 2019). Differently from chocolates, these three masses did not show significant differences between cocoa beans from the three different continents, however, the mass peaks averaged concentrations follow a similar trend both in cocoa beans and chocolates. The Southeast Asian samples showed the highest concentration of mass 43.019 and 61.028 both in beans and chocolates followed by the samples from the Americas. On the other hand, the Southeast Asian samples had the lowest concentration of mass 33.033 both in cocoa beans and chocolates. Taking into account that acid acetic contributes most to beans and chocolate acidity (Jinap & Dimick, 1990) the results of this study are in agreement with those of Counet, Ouwerx, Rosoux, and Collin (2004), that found dried and fermented beans from Southeast Asia and South Pacific to comprise higher acidity levels than those from South African origin due to the box of fermentation used, and possible application of rapid artificial drying. According to Figure 6.1 and Table 6.1, the samples from Southeast Asia tended to be higher in VOCs concentrations compared to the other continents. Our results suggested that the differences in VOC profiles of fermented and dried beans from different origins are quantitative rather than qualitative (Counet et al., 2004; Frauendorfer & Schieberle, 2008).

6.3.1.1.2 Tentative identification of VOCs released during fermentation and drying
The accuracy of the PTR-QiToF-MS in measuring the exact mass allowed the assignment of sum formulas. The PTR-QiToF-MS results combined with literature data permitted a tentative identification of 64 masses/ions (Table 6.1). The highlighted compounds in Table 6.1 were tentatively identified, however, because of the low concentration (< 1 ppb) they were not included in the analysis. The majority of the literature related to the volatile compounds of cocoa beans refers to the key aroma compounds that play an important role in the flavour of beans. The list of compounds developed by Huang and Barringer (2011) on cocoa and studies related to fermented and dry cocoa beans was taken into account. Moreover, several compounds were confirmed comparing these results with masses found in coffee studies applying the same instrument (Özdestan et al., 2013; Romano et al., 2014; Yener et al., 2014, 2015, 2016). Masses detected in chocolate (Acierno et al., 2016) were also considered.

The PTR-QiToF-MS measurements allowed detection of compounds of the main classes of cocoa odorant compounds, such as alcohols, acids, esters, aldehydes, pyridines, and pyrazines (Table 6.1). The cocoa beans analysed were fermented and dried. During these steps of production, the first aroma compounds are developed, which are predominantly aldehydes, alcohols, acids and acetates (Ziegleder, 2009).

It is important, however, to underline that differences in the aroma profile of unroasted (fermented and dried) and roasted bean are caused by more quantitative more than qualitative changes (Frauendorfer & Schieberle, 2008).

In this study, within the alcohols group methanol (m/z 33.034), 2, 3-butanediol (m/z 91.073), and 2-phenylethanol (m/z 123.088) were tentatively identified. The latter was also identified as the most odour-active compounds in fermented and unroasted cocoa beans by Frauendorfer and Schieberle (2008). It has been reported in cocoa fermentation with K. apiculata and S. cerevisiae var. chevalieri yeasts and described as desirable compounds for high-quality cocoa products (Rodriguez-Campos, Escalona-Buendía, Orozco-Avila, Lugo-Cervantes, & Jaramillo-Flores, 2011). High alcohols content is desirable, however, their concentration decreases during drying and roasting because of the high temperature (Aprotosoaie, Luca, & Miron, 2016).

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

Among the aldehydes, it was possible to identify acetaldehyde (m/z 45.033, from alanine), propanal (m/z 59.048), methyl-butenal (m/z 85.063), 2/3-methylbutanal (m/z 87.077, from isoleucine and leucine), benzaldehyde (m/z 107.053, from phenylalanine), and heptanal (m/z 115.109). Some of these compounds were identified as dominant odour-active volatiles in cocoa mass (Afoakwa et al., 2008). Aldehydes are formed by Strecker degradation of free amino acids during roasting. However, low concentrations of aldehydes may arise even during fermentation and drying (Aprotosoaie et al., 2016) as highlighted by the results in Table 6.1.

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

Pyrazines are usually products of roasting but several studies showed the presence in the unroasted beans, too (Frauendorfer & Schieberle, 2008). In this study, two pyrazines were identified: 2-ethenyl-6-methylpyrazine (m/z 121.084) and 6,7-dihydro-(5H)-cyclo-pentapyrazine (m/z 121.084). In Table 6.1, it is possible to highlight the presence of esters. This class of compounds, released during fermentation, is the second important class for cocoa and chocolate flavour after pyrazine and has been found both in unroasted and roasted cocoa beans (Aprotosoaie et al., 2016).

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

Table 6.1. Average of concentrations (ppbv) of tentatively identified mass peaks in the headspace of cocoa beans of different origins detected by PTR-QiToF-MS. Masses mean showing geographical significant difference according to Kruskal Wallis tests (p < 0.05) are highlighted.

m/z	Sum formula	Tentative identification	Chemical class	Average VOC concentrations (ppbv)±SD					
				Africa		Americas		Southeast Asia	
33.034	CH_5O^+	Methanol [3,4]	Alcohol	2962.9	± 713.6	3050.5	± 771.2	2724.5	± 735.9
41.038	$C_3H_5^+$	Alkylic fragment [3,6]	Fragment	110.9	± 51.9	104.5	± 47.8	190.9	± 277.3
43.019	$C_2H_3O^+$	Fragment (diverse origin) [4,7]	Fragment	22267.8	± 16113.7	22510.5	± 23284.8	31561.6	± 28765.4
43.054	$C_3H_7^+$	Fragment (diverse origin) [4,7]	Fragment	3037.8	± 2255.8	3232.7	± 3501.7	4680.6	± 4534.0
45.033	$C_2H_5O^+$	Acetaldehyde/ethanol [2,3,4]	Aldehyde/alcohol	173.7	± 45.0	161.8	± 55.6	193.2	± 94.3
59.048	$C_3H_7O^+$	Acetone/propanal [3,5]	Ketone/aldehyde	84.1	± 20.3	83.2	± 14.8	83.9	± 33.7
61.029	$C_2H_5O_2^+$	Acetic acid [3,4,5]	Acid/ester	37012.3	± 23610.2	35550.2	± 29906.5	44913.0	± 35949.8
67.053	$C_5H_7^+$	Terpene fragment [6,7]	Fragment	0.7	± 0.2	0.8	± 0.2	0.8	± 0.3
68.054	$C_4H_6N^+$	Pyrrole [4,5]	Pyrrole	0.3	± 0.1	0.3	± 0.1	0.4	± 0.2
71.048	$C_4H_7O^+$	Dihydrofuran/2-methylpropenal/3-buten-2-one/alcohol fragment [3,4]	Furan/aldehyde/ketone/fragment	24.0	± 15.1	21.0	± 10.5	19.5	± 11.4
71.084	$C_5H_{11}^+$	Terpene fragment/pentene [3,6]	Fragment/alkene	94.9	± 61.5	104.5	± 69.6	73.9	± 41.7
73.068	$C_4H_8O^+$	2-Methylpropanal (isobutanal)/Butanone[3,8]	Aldehydes/ketone	971.8	± 499.1	784.9	± 420.8	1933.2	± 1470.0
75.043	$C_3H_7O_2^+$	Propanoic acid/methyl-acetate/acetol [3,4,7]	Acid/ester/ketone	156.2	± 108.3	127.5	± 96.2	201.3	± 187.4
81.067	$C_6H_9^+$	Methylpentene/terpene fragment [4]	Alkene/fragment	5.4	± 3.1[a]	9.1	± 4.7[b]	9.4	± 10.6[ab]
85.026	$C_4H_5O_2+$	2(5H)-furanone [4]	Ketone	1.1	± 0.7	0.9	± 0.8	1.5	± 1.2
85.063	$C_5H_9O^+$	Methyl-butenal [3,5]	Aldehyde	6.8	± 3.6	5.8	± 3.8	8.1	± 5.7
87.045	$C_4H_7O_2^+$	2,3-butandione (diacetyl)/butanedione/butyrolactone [3,5,6]	Ketone/ester	38.2	± 23.8	28.3	± 17.5	44.1	± 23.7
87.077	$C_5H_{11}O^+$	3-methylbutanal/2-methylbutanal/methylbutanal[2,3,4,5]	Aldehyde/alcohol	47.9	± 24.5	40.0	± 23.1	45.0	± 21.0
89.060	$C_4H_9O_2^+$	Methylpropanoate/2-methylpropanoic acid/hydroxybutanone/acetoin [2,3,4,5]	Ester/acid/ketone	203.1	± 128.3	172.5	± 119.2	204.0	± 149.1

Table 6.1. Continued.

m/z	Sum formula	Tentative identification	Chemical class	Average VOC concentrations (ppbv)±SD		
				Africa	Americas	Southeast Asia
91.073	$C_4H_{11}O_2^+$	2,3- Butanediol [9]	Alcohol	8.8 ± 5.1	7.9 ± 5.4	16.2 ± 13.1
93.068	$C_7H_9^+$	Terpene fragment/Toluene [1,3,6]	Fragment	24.5 ± 6.2	25.8 ± 6.4	27.3 ± 8.0
95.045	$C_6H_7O^+$	Phenol/cyclohexadienone/1,4-cyclopentadiene-1-carbaldehyde	Alcohol/ketone/aldehyde	1.3 ± 0.6	1.2 ± 0.5	5.2 ± 8.3
96.081	$C_6H_{10}N^+$	Dimethyl-pyrrole/ethyl-pyrrole [4,5]	Pyrrole	0.3 ± 0.2	0.3 ± 0.1	0.7 ± 0.7
97.027	$C_5H_5O_2^+$	Furfural (furancarboxaldehyde) [3,5]	Aldehyde	1.7 ± 1.8	0.7 ± 0.5	1.2 ± 0.9
98.059	$C_5H_8ON^+$	Dimethyl-oxazole [5,8]	Pyrrole	0.2 ± 0.1	0.2 ± 0.1	0.2 ± 0.1
99.041	$C_5H_7O_2^+$	Furfuryl alcohol/a-angelica lactone [4,5]	Furan/ester	0.5 ± 0.3	0.4 ± 0.3	0.7 ± 0.4
99.078	$C_6H_{11}O^+$	Hexenal/methyl-pentenone [1,5]	Aldehyde/ketone	1.6 ± 0.8	1.3 ± 0.7	1.8 ± 1.0
101.058	$C_5H_9O_2^+$	Dihydro-2-methyl-3(2H)-furanone/2,3-pentanedione/pentanedione/methyl-tetrahydrofuranone [3,4,5]	Furan/ketone	5.5 ± 2.6	4.7 ± 3.1	7.9 ± 5.1
103.075	$C_5H_{11}O_2^+$	2,3-methylbutanoic acid/2-hydroxy-3-pentanone/3-methylbutanoic acid [3,4,5,6]	Acid/ketone	85.0 ± 58.7	74.4 ± 60.3	104.6 ± 85.7
104.049	$C_7H_6N^+$	Benzonitrile [3]	Nitrile	1.5 ± 0.6	1.3 ± 1.0	1.5 ± 0.9
105.069	$C_8H_9^+$	Styrene/phenyletanol fragment [1]	Alkene/fragment	10.8 ± 7.8	13.3 ± 11.8	13.5 ± 11.9
107.053	$C_7H_7O^+$	Benzaldehyde/benzyl alcohol [1,2]	Aldehyde/alcohol	19.0 ± 6.6	20.2 ± 8.8	19.9 ± 13.6
107.083	$C_8H_{11}^+$	Terpene fragment [1,7]	Fragment	20.6 ± 6.6	21.6 ± 9.1	21.5 ± 11.7
109.065	$C_7H_9O^+$	4-methylphenol/benzyl alcohol [3]	Alcohol	0.6 ± 0.5	0.8 ± 1.0	2.2 ± 2.4
110.060	$C_6H_8ON^+$	Formil-methylpyrrole/acetylpyrrole/2-acetylpyrrole [1,2,4]	Pyrrole	0.3 ± 0.3	0.3 ± 0.3	0.6 ± 0.5
111.046	$C_6H_7O_2^+$	2-acetylfuran/5-methylfurfural [4,5,6]	Furan	0.5 ± 0.4	0.3 ± 0.2	0.5 ± 0.3
111.078	$C_7H_{11}O^+$	2,3-dimethyl-2-cyclopente-1-one [4,5]	Ketone	0.8 ± 0.3	1.0 ± 0.5	1.2 ± 0.8
112.076	$C_6H_{10}ON^+$	Trimethyloxazole [4,5]	Furan	0.2 ± 0.2	0.2 ± 0.1	0.3 ± 0.3
113.053	$C_6H_9O_2^+$	Methylfurfuril alcohol/dimethylfuranone/methylciclopentan edione/cyclotene [4,5,7]	Furan/ketone	0.4 ± 0.2	0.3 ± 0.2	0.6 ± 0.3

Table 6.1. Continued.

				Average VOC concentrations (ppbv)±SD		
m/z	Sum formula	Tentative identification	Chemical class	Africa	Americas	Southeast Asia
113.093	$C_7H_{13}O^+$	Heptenone [1,2,4]	Ketone	1.3 ± 0.5	1.2 ± 0.6	1.6 ± 0.9
115.075	$C_6H_{11}O2^+$	4-methyltetrahydro-(2H)-pyran-2-one [4,5,7]	Ketone	1.5 ± 1.2	2.0 ± 2.1	3.6 ± 3.5
115.109	$C_7H_{15}O^+$	Heptanal/2-heptanone[1,2,3,5]	Aldehyde/ketone	7.4 ± 4.2[a]	20.2 ± 13.6[b]	12.4 ± 8.6[ab]
117.088	$C_6H_{13}O2^+$	Hexanoic acid/C6 ester [1,3,5]	Acid/ester	9.0 ± 6.7	7.1 ± 4.7	10.9 ± 8.6
121.084	$C_7H_9N_2^+$	2-ethenyl-6-methylpyrazine/6,7-dihydro-(5H)-cyclopentapyrazine [1,3,5]	Pyrazine	32.2 ± 28.9	40.9 ± 64.6	61.2 ± 66.1
123.088	$C_8H_{11}O^+$	2-phenylethanol [3,9]	Alcohol	1.5 ± 1.3	1.5 ± 1.2	3.6 ± 3.5
125.058	$C_7H_9O_2^+$	Guaiacol/methyl-benzenediol/furyl acetone [1]	Ether	0.5 ± 0.3	0.5 ± 0.4	0.9 ± 1.0
125.093	$C_8H_{13}O^+$	Butylfuran/methylpropylfuran/octadienone/trimethylciclopentenone/alkylfurans/E-2-octenal [2,4,5]	Furan/ketone/aldehyde	0.6 ± 0.3	0.6 ± 0.3	1.0 ± 1.0
126.090	$C_7H_{12}ON^+$	Acethyl-dimethylpyrrole/alkyloxazole [1,5]	Pyrrole	0.1 ± 0.1	0.1 ± 0.1	0.2 ± 0.1
129.088	$C_7H_{13}O2^+$	Pentenylacetate/heptanedione [4,5,7]	Ester/ketone	0.6 ± 0.3	0.7 ± 0.4	1.1 ± 0.8
131.105	$C_7H_{15}O_2^+$	Ethyl valerate/heptanoic acid/C7 ester [1,3,5]	Ester/acid	9.6 ± 8.9	10.1 ± 11.4	7.6 ± 8.2
136.113	$C_9H_{14}N^+$	Butyl-pyridine/ethyl-propyl pyridine [1,5]	Pyridine	0.3 ± 0.2	0.2 ± 0.1	0.6 ± 0.8
137.129	$C_{10}H_{17}^+$	Various monoterpenes [1]	Terpene	7.4 ± 11.2	8.9 ± 5.1	23.7 [2,4,5] ± 36.7
138.057	$C_7H_8O_2N^+$	Pyridinecarboxylic acid methyl ester [1,5]	Acid	0.1 ± 0.1	0.1 ± 0.1	0.3 ± 0.3
139.108	$C_9H_{15}O^+$	E,E-2,4-nonadienal/E-2-nonenal [2,4,5]	Aldehyde	2.0 ± 0.5	2.0 ± 0.7	2.6 ± 1.5
143.141	$C_9H_{19}O^+$	Nonanal [3]	Aldehyde	4.7 ± 3.3[a]	8.1 ± 4.6[b]	7.4 ± 7.3[ab]
145.122	$C_8H_{17}O_2^+$	Ethyl-exanoic acid/C8 ester [1,5]	Acid/ester	1.3 ± 0.9	1.6 ± 1.5	1.8 ± 1.7
148.078	$C_9H_{10}ON^+$	1-furfurylpyrrole [4,5]	Pyrrole	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.0
153.126	$C_{10}H_{17}O^+$	E,E-2,4-decadienal [5,6,7]	Aldehyde	0.4 ± 0.2	0.4 ± 0.2	0.7 ± 0.5
155.143	$C_{10}H_{18}O^+$	Linalool [8]	Alcohol	1.0 ± 0.4	0.9 ± 0.5	1.3 ± 0.6
*159.137	$C_9H_{19}O_2^+$	Nonanoic acid/C9 ester [1,5]	Acid/ester	0.9 ± 0.5	0.9 ± 0.6	1.2 ± 0.9

Table 6.1. Continued.

m/z	Sum formula	Tentative identification	Chemical class	Average VOC concentrations (ppbv)±SD		
				Africa	Americas	Southeast Asia
*166.092	$C_9H_{12}O_2N^+$	Methyl-pyrrolyl-butanedione [1,5]	Pyrrole	0.1 ± 0.1	0.1 ± 0.1	0.1 ± 0.1
*167.114	$C_{10}H_{15}O_2^+$	Ethyl-dimethoxy-benzene [1,5]	Ether	0.2 ± 0.1	0.2 ± 0.1	0.2 ± 0.1
*169.191	$C_{13}H_{25}^+$	Dodecane [2]	Alkane	0.6 ± 0.2	0.6 ± 0.2	0.7 ± 0.2
*191.160	$C_{13}H_{19}O^+$	Beta-damascenone [5]	Ketone	0.1 ± 0.1	0.1 ± 0.1	0.1 ± 0.1

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

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

*Compounds tentatively identified but with concentration lower than 1 ppb.
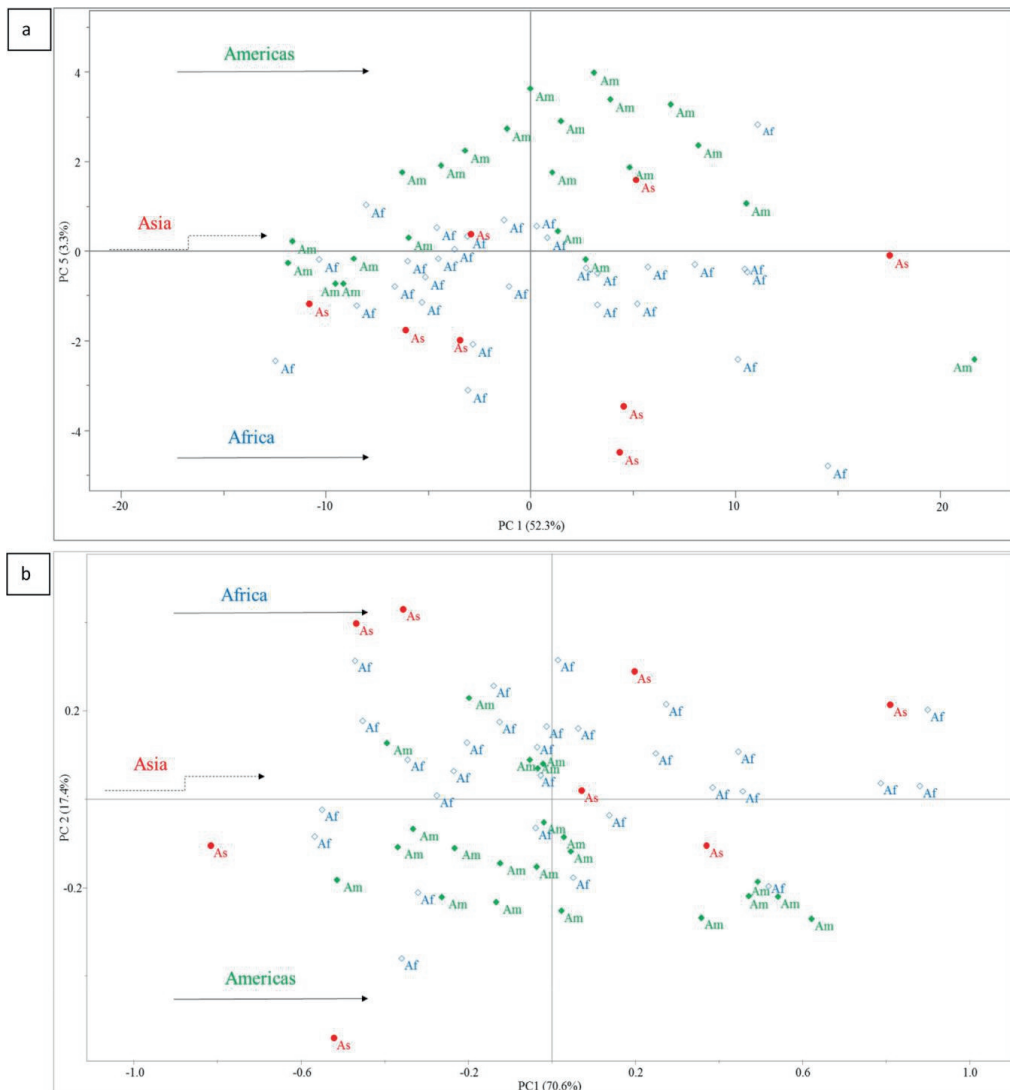
Figure 6.2. a) PCA of normalized and auto-scaled PTR-QiToF-MS data of cocoa beans from single geographical origins: Africa (Af), America (Am) and Southeast Asia (As). The arrows indicate the trend of the samples distribution. The broken arrow indicates the tendency to overlap of Asian samples. b) PCA of mean centred and SNV visual camera system data of cocoa beans from single geographical origins: Africa (Af), America (Am) and Southeast Asia (As). The arrows indicate the trend of the samples distribution. The broken arrow indicates the tendency to overlap of Asian samples.

6.3.1.2 Quantitative geographical comparison

6.3.1.2.1 Cocoa beans geographical origins: univariate comparison

      The final data set (59 samples×114 masses) was subjected to Kruskal-Wallis tests to detect significant differences in mass intensities with respect to the origins of the beans (Af, Am, As). The masses showing statistical differences ($p < 0.05$) according to the rank-based test are listed in Table 6.1. Shapiro Wilk tests were conducted to examine normality within the groups. The data were not normally distributed (Shapiro-Wilk test < 0.05), because of this, the significance of differences between the samples was evaluated by non-parametric tests Kruskall Wallis test and Mann-Whitney post-hoc.

      None of the masses in Table 6.1 showed distinct concentrations allowing discrimination of the three continents simultaneously. What stands out in Table 6.1 is the possibility to statistically differentiate from Americas and Africa while the means of the Southeast Asian samples lie between the other two continents. The higher variation of the Southeast Asian samples (Table 6.1) compared to the other continents highlights the difficulties to statistically differentiate these cocoa beans. This is consistent with the results of previous work on the same sample set (Marseglia et al., 2016), highlighting that Asian samples have intermediate characteristics between African and South American cocoa beans; a strong variability was observed inside the Asian group. These data are also consistent with the fact that, in some regions of Asia such as Indonesia, cocoa fermentation is short or even omitted (Rohsius, Matissek, & Lieberei, 2005), inducing a large differentiation caused by fermentation level, overpassing that of geographic origin. For example, in a previous work on peptide pattern of cocoa beans, samples from Indonesia (Flores-Indonesia and Sulawesi-Indonesia), also analysed in the current study, showed the same peptide pattern as unfermented slaty beans, indicating that they were poorly fermented, while other Asian samples as those from Papua New Guinea had the highest amount of peptides, indicating an extensive fermentation (Caligiani et al., 2016).

      Nevertheless, it is still possible to find information related to the continent but it is restricted to a limited number of masses (Table 6.1), similarly to the results of Hernandez and Rutledge (1994). Within the compounds in Table 6.1, the terpene fragment can be informative of the raw material as terpenes naturally occur in flowers and fruits. A study related to coffee beans confirmed that this compound could resist the roasting (Yener et al., 2016), indicating terpene as a probable geographical indicator for chocolate supply chain at least till the roasting step. Mass 115.109 tentatively identified as heptanal/2-heptanone was found to be also relevant in South American chocolate discrimination (Acierno et al., 2016), supporting the rank test results of the cocoa beans in this study and indicating this mass as a potential geographical indicator for products from Americas.

6.3.1.2.2 Cocoa beans geographical origins: multivariate comparison

      For further evaluation of the data, a PCA was performed using the full volatile profiles obtained by PTR-QiToF-MS. The investigation of the total volatile profiles can help in extracting more information (Capuano & van Ruth, 2012).

      The data matrix (59 samples × 114 masses) was auto-scaled and normalized prior to PCA; in Figure 6.2a the scores plot corresponding to PC1 versus PC5 is shown. In order to get geographical clusters, mass 43.019 and 61.028 were excluded. The distribution of the samples revealed a tendency to group cocoa beans according to their geographical origins (Figure 6.2a). As shown in Figure 6.2a, the variability within samples is mainly represented in PC1, while PC5 allows a geographical differentiation. A separation between samples from Africa and Americas is visible. On the other hand, a not clear trend is visible for the Southeast Asian cocoa beans: these do not form a separate cluster but are mostly grouped with African cocoa beans (Africa - Southeast Asia cluster). A similar trend was also found by Marseglia et al. (2016), analysing the same sample set using HR MAS ^{1}H NMR and HR ^{1}H NMR. Mass 43.019 and 61.028 were bringing a lot of variability within the samples reducing the visualisation of a geographical trend. In Figure 6.3 it is possible to notice the differences in concentration of masses 43.019 and 61.028 between samples in general and also between samples from the same country. Wide variation in acetic acid values in samples from the same country was found also by Jinap and Dimick (1990). The main explanation for mass 61.028 variation is the difference in fermentation conditions (Jinap & Dimick, 1990). High acidity can be partly due to long fermentation.

Looking at the concentration of mass 61.028 for the samples from Papua New Guinea and Indonesia (Flores-Indonesia and Sulawesi-Indonesia) it is possible to support what stated in paragraph 6.3.1.2.1 about the Southeast Asian variability. In fact, the higher concentration of acid acetic for the sample from Papa New Guinea can be an indication of well-fermented beans while the low acid acetic concentration for the Indonesian samples can be informative of unfermented beans. Long fermentation steps determine the higher release of flavour precursors (Afoakwa et al., 2008; Counet et al., 2004). In Figure 6.3 the fermented and dried samples with higher concentrations of m/z 61.029 show a higher total volatile headspace concentration compared to the other cocoa beans highlighting consistency with the hypothesis about Papua New Guinea samples. Counet et al. (2004) also reported that the samples from Papua New Guinea had been subjected to longer fermentation and presented a higher concentration of total aroma. A trend between m/z 61.029 and the total ion count (not including m/z 43.019 and m/z 61.029) is evident for all the cocoa beans (Figure S6.1, Supplementary materials). This is indicative of a possible link between acid acetic -and fermentation-flavour formation already before roasting.

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

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

Figure 6.4 shows the 50 highest loading values related to PC5. PC5 has been chosen as it is the one affected by the geographical information. Taking into account the sample distribution in Figure 6.2a, the cluster of beans from the Americas is linked with positive loadings on PC5 while the African-Southeast Asian one with negative loadings. The compounds showing statistical differences according to the origin in Table 6.1 are included in the most representative loadings values supporting an effective role in geographical differentiation. Specifically, heptanal/2- heptanone (m/z 115.109) and nonanal (m/z 143.141) can be confirmed as significant for South American samples. Several masses related to the beans from the Americas have been found relevant also for the classification of chocolate produced with South American beans (Acierno et al., 2016). Within them is it possible to underline some compounds tentatively identified as benzonitrile (m/z 104.049), benzaldehyde (m/z 107.053), heptanal/2-heptanone (m/z 115.109). Heptanal/2-hepta- none was the compound with a bigger role in the definition of the fifth PC in the positive range. Specifically, it was abundant in Peruvian samples. 2-heptanone has been found in beans fermented and dried at specific conditions (Rodriguez-campos, Escalona-buendía, Contreras-ramos, & Orozco-avila, 2012), for this reason, m/z 115.109 can be indicative of specific farming practice of this area/continent and indirectly linked with the provenance of the beans. On the other hand, the terpene fragment (m/z 107.083) can be indicative of geographical origin (Yener et al., 2016). Concerning the negative loading representative of the African – Southeast Asian samples, 2-methylpropanal (m/z 73.068) was the highest. Specifically, it was abundant in Madagascar and Papua New Guinea samples. Moreover, regarding the African cluster masses 41.038 and 45.033 were relevant for the classification of chocolate made from African beans (Acierno et al., 2016).
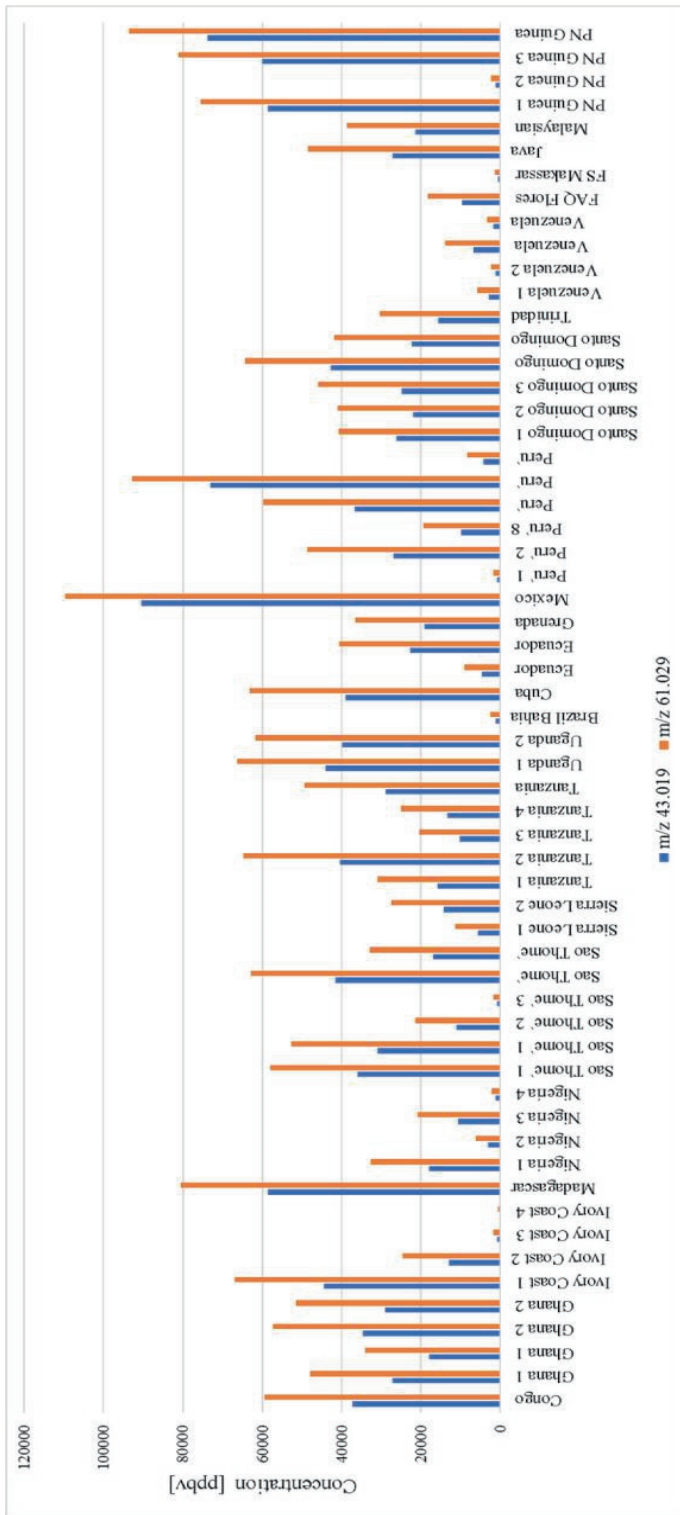
Figure 6.3. Concentrations of mass 61.029 and 43.019 for each sample.

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

Within the other compounds with negative loadings, 2/3 methylbutanal (m/z 87.077), 2 methylpropanoinc acid (m/z 89.060) and 2-phenylethanol (m/z 123.088) are reported as key aroma marker (Magagna et al., 2017). 3-Methylbutanal (m/z 87.077) and benzaldehyde (m/z 107.053) were found relevant by Hernandez and Rutledge (1994) to differentiate cocoa masses according to the geographical origin; which supports the influence of aldehydes on geographical distinction at PC5. Regarding 3- methylbutanal, Uganda and Tanzania showed the higher concentrations together with Madagascar ad Papua New Guinea as found in previous studies (Hernandez & Rutledge, 1994).

Therefore, the geographical information found by Hernandez and Rutledge (1994) could be further described as 3-methylbutanal relevant for the differentiation of African and Papua New Guinea samples from the rest and benzaldehyde (m/z 107.053) relevant for South American samples. Within the negative loading, 2-phenylethanol (m/z 123.088) can be found already in the pulp of the cocoa beans (Chetschik et al., 2018); it could represent a link with the raw material. Specifically, the Southeast Asian samples show higher concentrations. The cocoa beans from Southeast Asia showed high concentration for 2,3-butane- diol (m/z 91.073) as well; these masses (m/z 123.088 and 91,073) could be informative of this group even if it was difficult to differentiate these beans from the African ones. Within the African-Southeast Asian samples, is it important to highlight the monoterpenes (m/z 137.129) as possible geographical markers (Yener et al., 2016). Even if m/z 137.129 is relevant for the African -Southeast Asian cluster, it is mostly concentrated in the Asian beans from Java and Papa New Guinea. These results underline the possibility to characterise the Southeast Asian sample based on specific masses, however, more evidence is needed.
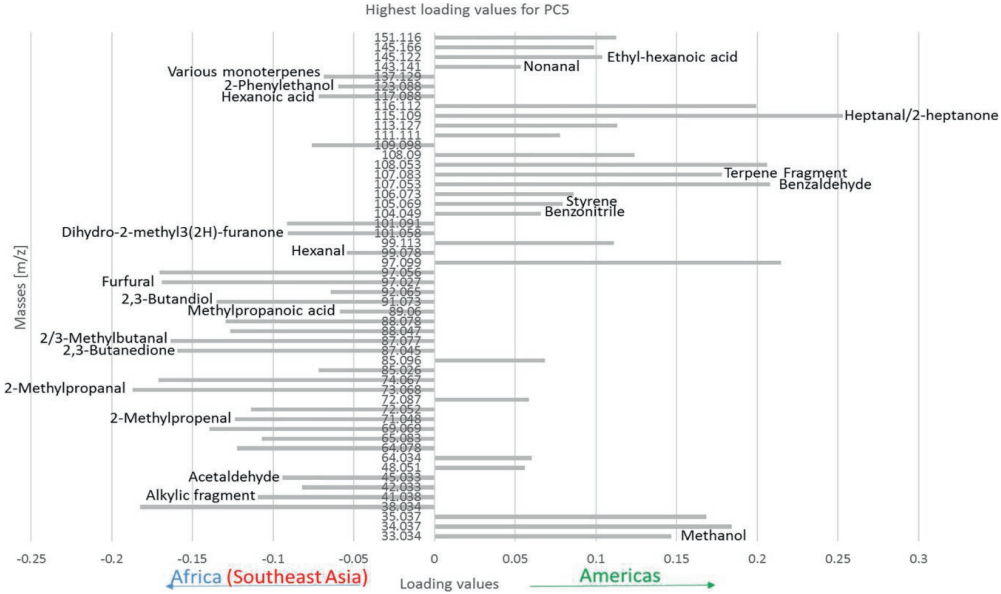
Figure 6.4. Fifty highest positive and negative loading values associated with PC5. Negative values are related to African samples while positive values are related to the samples from the Americas.

6.3.2 Machine vision: HSI

The cocoa beans samples were analysed by HSI and the spectra of each cocoa bean sample are shown in Figure 6.5a. Samples are coloured and highlighted according to the continent (Af, As, Am). Figure 6.5a shows that the spectra diverge after wavelength 625 nm. The patterns are fairly similar in the visible range, i.e. up to range 600–700 nm. This range reflects differences in colour, which obviously not differ with provenance. In the remaining Near Infrared range, patterns start to divert. In the 700–1000nm range, the third overtone region, differences in composition are reflected, especially those presented as O–H and C–H bonds. The groups of South American and Southeast Asian samples showing higher reflectance compared to the African cocoa bean group in the higher end wavelength range.

Spectroscopic data may not be interpreted directly because of the effect of a large number of factors, such as a high number of variables, slight scattering, instrumental drift, baseline shift, and slope variation, caused by differences in particle size and physical properties of the samples. Therefore, the use of chemometrics is required to amplify the information of interest and lessen the undesirable information in the spectra (Lohumi et al., 2015). In this study, PCA was used to explore the visual camera system data (Figure 6.3b). The matrix was mean centred and SNV (Standard Normal Variates) applied prior to PCA. In Figure 6.5b, the cocoa beans profiles are shown after data treatment.

As presented in Figure 6.2b, the first two principal components, which explain 88% of the total variability in the data, display a grouping of cocoa beans according to their geographic origin. PC2 is mainly affected by the geographical information while the variability within beans is mainly represented in PC1, probably affected by an unknown phenomenon such as lighting. A trend in the distribution of African and American cocoa beans is visible whereas a not so clear trend is shown by the cocoa beans from Southeast Asia.

The sample distribution of the spectroscopy data followed the same trend of the PTR-QiToF-MS, indicating a possible link between the two measurements. Nevertheless, also in this case, a similarity with the results showed by Marseglia et al. (2016), analysing the same sample set using HR MAS 1H NMR and HR 1H NMR, can be underlined. HR MAS 1H NMR fingerprint of cocoa beans combined with chemometrics discriminates African and American cocoa samples, based mainly on the fatty acids, acetate and saccharidic components. Moreover, HR 1H NMR data showed citrate and formiate as possible markers for African samples, while amino acids, caffeic acid, caffeine and epicatechin as possible markers for American cocoa samples. Taking into account the groups (CH3-CH2-CH) used by Caligiani, Acquotti, Cirlini, and Palla (2010), to identify these compounds, it is possible to relate them to the spectral range in Figure 6.5 and give a possible explanation to the similar trend in the sample distribution. Because of low evidence for the Southeast Asian samples the discussion is mainly focused on the African and American samples.

In Figure 6.5b the wavelengths are highlighted according to their contribution on PC2. Considering the sample distribution in Figure 6.2b, it is possible to state that the American cluster is linked with negative loading (highlighted area in Figure 6.5b) on PC2 and the African group with positive loadings (not highlighted area in Figure 6.5b). The wavelength range characteristic for the samples from the Americas includes the wavelength range related to CH3 groups (from ~700 nm to ~730 nm and from ~870 nm to ~910 nm); this group could be linked with particular organic compound groups, possibly fatty acids, etc. according to the NMR results. Amino acids may also play a role (RNH2: from ~770 nm to ~830 nm) in the distinction of the beans from the Americas. Moreover, sugars such as glucose and fructose may be reflected in the spectral data and are relevant in discriminating the African cocoa beans according to the NMR studies.

6.3.3 Machine olfaction and machine vision comparison

The machine olfaction and machine vision results were correlated (Pearson correlation tests). Spectral range 625–1000 nm and the majority of the VOCs showed significant correlations (Figure S6.2, Supplementary materials), although this was even more pronounced for the 800–1000 nm range and lower molecular weight VOCs. The VOCs are unlikely to alter the HSI profiles by themselves, due to their very low concentrations in the beans. However, differences in the major components in the beans, which are reflected in the HIS spectra, will certainly affect the formation of VOCs in the beans. For instance, the pronounced correlation of VOC with amine group wavelength range from ~770 nm to ~830 nm, can underline the significant role of amino acid in flavour development. Both VOCs and HSI spectra are closely related to local production circumstances and fermentation conditions. For instance, 2-methylpropenal (m/z 71.048), 2,3-butanedione (m/z 87.045), acetic acid (m/z 61.029) and the furan class (m/z 71.048) and their correlating ranges may link specifically to the local/country fermentation/drying processes. Although the correlation between the machine olfaction and vision measurements is interesting, there is most likely not a direct relationship but only an indirect one. Further research is required to examine this correlation, which is not the aim of the current study. The correlation shows though, that multiple linking characteristics allow distinction between the cocoa beans produced in different geographical areas.

6.4 Conclusions

As a whole, the results obtained in the present work point out the potential complementary use of machine olfaction by PTR-QiToF-MS and machine vision by HSI to have a fast screening of the cocoa beans geographical origin. The PTR-QiToF-MS and HSI fingerprints showed the same tendency in clustering South American and African samples. Southeast Asian samples showed great variability with both approaches. Most likely this variation is due to differences in fermentation levels. They appear to influence the metabolic profiles of cocoa extensively. Machine olfaction and machine vision characterisation provided a similar degree of sample separation. Significant correlations highlighted the possibility to use these techniques as complementary rapid tools, which may be further developed for use in practical settings. The cocoa beans volatile profiles measured using the PTR-QiToF-MS showed geographical similarities with the chocolate volatile profiles measured previously indicating the possibility to develop this technique further for assessments of the geographical origin along the whole cocoa-chocolate supply chain.
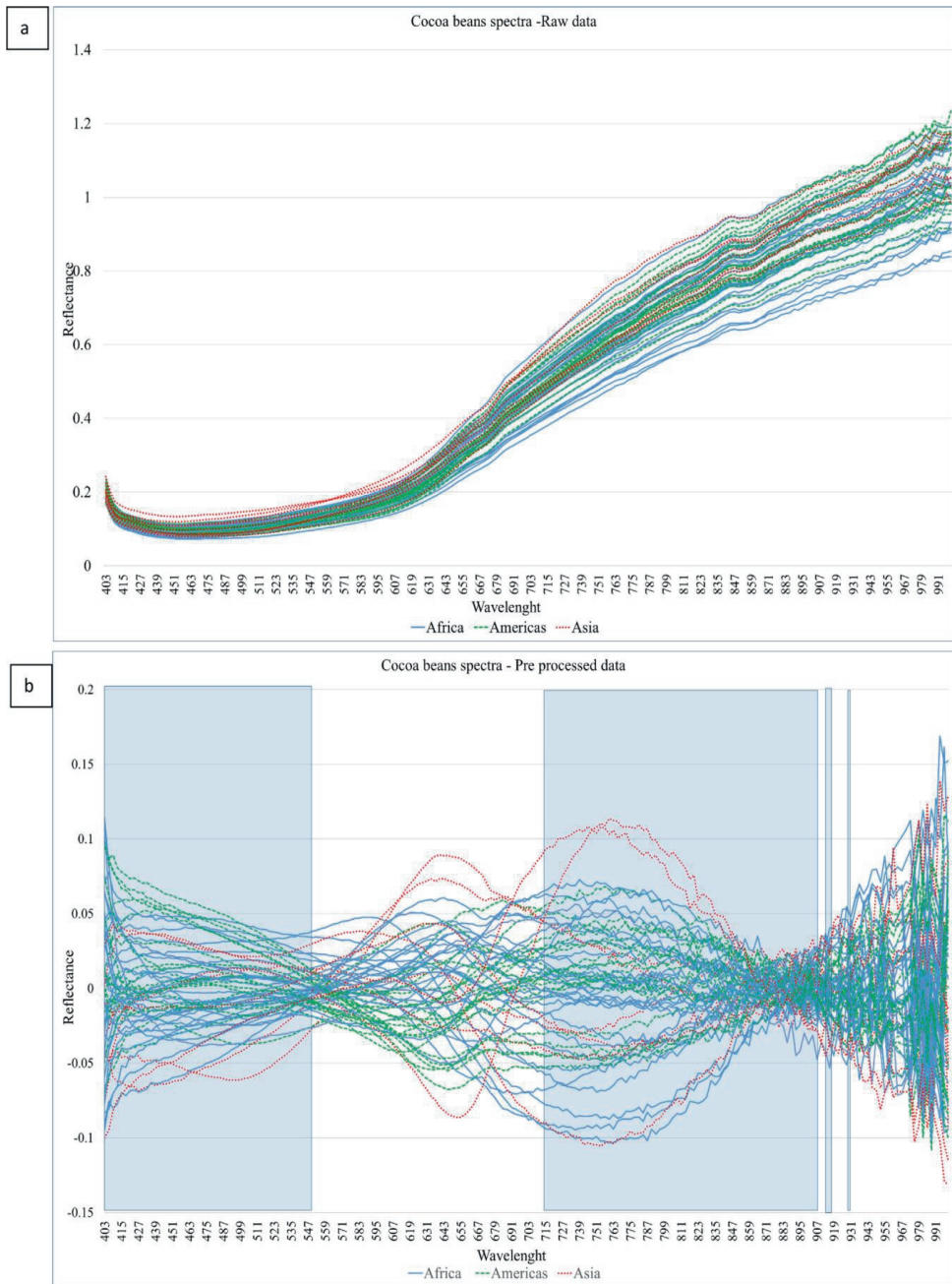
Figure 6.5. a) HSI spectra of the cocoa beans. Samples are coloured and underlined according to the continent of provenance: Africa, Americas, Asia. b) Cocoa beans HSI spectra after mean centring and SNV pre-processing. The highlighted areas correspond to the wavelengths relevant for the PCA distribution of the samples from Americas while the non-highlighted areas are important with regard to the distribution of the African samples.

Declarations of interests

None

References

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

Acierno, V., Yener, S., Alewijn, M., Biasioli, F., & van Ruth, S.M. (2016). Factors con- tributing to the variation in the volatile composition of chocolate: Botanical and geographical origins of the cocoa beans, and brand-related formulation and processing. *Food Research International, 84*, 86–95. https://doi.org/10.1016/j.foodres.2016. 03.022.

Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2008). Flavor formation and character in cocoa and chocolate: A critical review. *Critical Reviews in Food Science and Nutrition, 48*(9), 840–857. https://doi.org/10.1080/10408390701719272.

Aprotosoaie, A. C., Luca, S. V., & Miron, A. (2016). Flavor chemistry of cocoa and cocoa products-an overview. Comprehensive Reviews in *Food Science and Food Safety, 15*(1), 73–91. https://doi.org/10.1111/1541-4337.12180.

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

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

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

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

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

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

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

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

Frauendorfer, F., & Schieberle, P. (2008). Changes in key aroma compounds of Criollo cocoa beans during roasting. *Journal of Agricultural and Food Chemistry, 56*(21), 10244–10251. https://doi.org/10.1021/jf802098f.

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

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

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

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

Icco (International Cocoa Organization) (2012). Study on the costs, advantages and disadvantages of cocoa certification. Retrieved from https://www.icco.org/about-us/international-cocoa-agreements/cat_view/30-related-documents/37-fair-trade-organic-cocoa.html

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

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

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

Liu, N., Parra, H. A., Pustjens, A., Hettinga, K., Mongondry, P., & van Ruth, S. M. (2018). Evaluation of portable near-infrared spectroscopy for organic milk authentication. *Talanta, 184*(December 2017), 128–135. https://doi.org/10.1016/j.talanta.2018.02. 097.

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

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

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

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

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

Masi, E., Taiti, C., Heimler, D., Vignolini, P., Romani, A., & Mancuso, S. (2016). PTR-TOF- MS and HPLC analysis in the characterization of saffron (Crocus sativus L.) from Italy and Iran. *Food Chemistry, 192*, 75–81. https://doi.org/10.1016/j.foodchem.2015.06. 090.

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

Qin, X. W., Lai, J. X., Tan, L. H., Hao, C. Y., Li, F. P., He, S. Z., & Song, Y. H. (2017). Characterization of volatile compounds in Criollo, Forastero, and Trinitario cocoa seeds (Theobroma cacao L.) in China. *International Journal of Food Properties, 20*(10), 2261–2275. https://doi.org/10.1080/10942912.2016.1236270.

Rodriguez-campos, J., Escalona-buendía, H. B., Contreras-ramos, S. M., & Orozco-avila, I. (2012). Effect of fermentation time and drying temperature on volatile compounds in cocoa. *Food Chemistry, 132*(1), 277–288. https://doi.org/10.1016/j.foodchem.2011. 10.078.

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

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

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

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

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

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

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

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

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

Supplementary materials
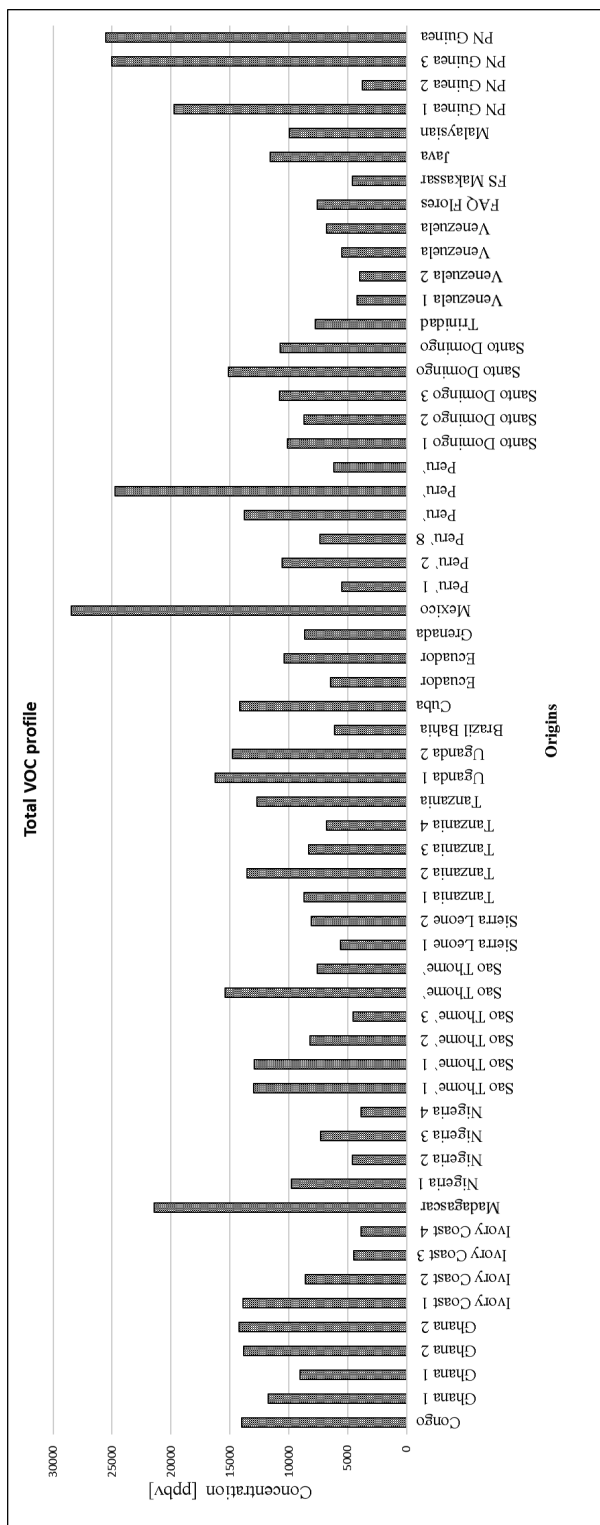
Figure S6.1. Total volatile content (sum of concentrations) per each single origin. Mass 43.019 and 61.029 are not included in the sum.
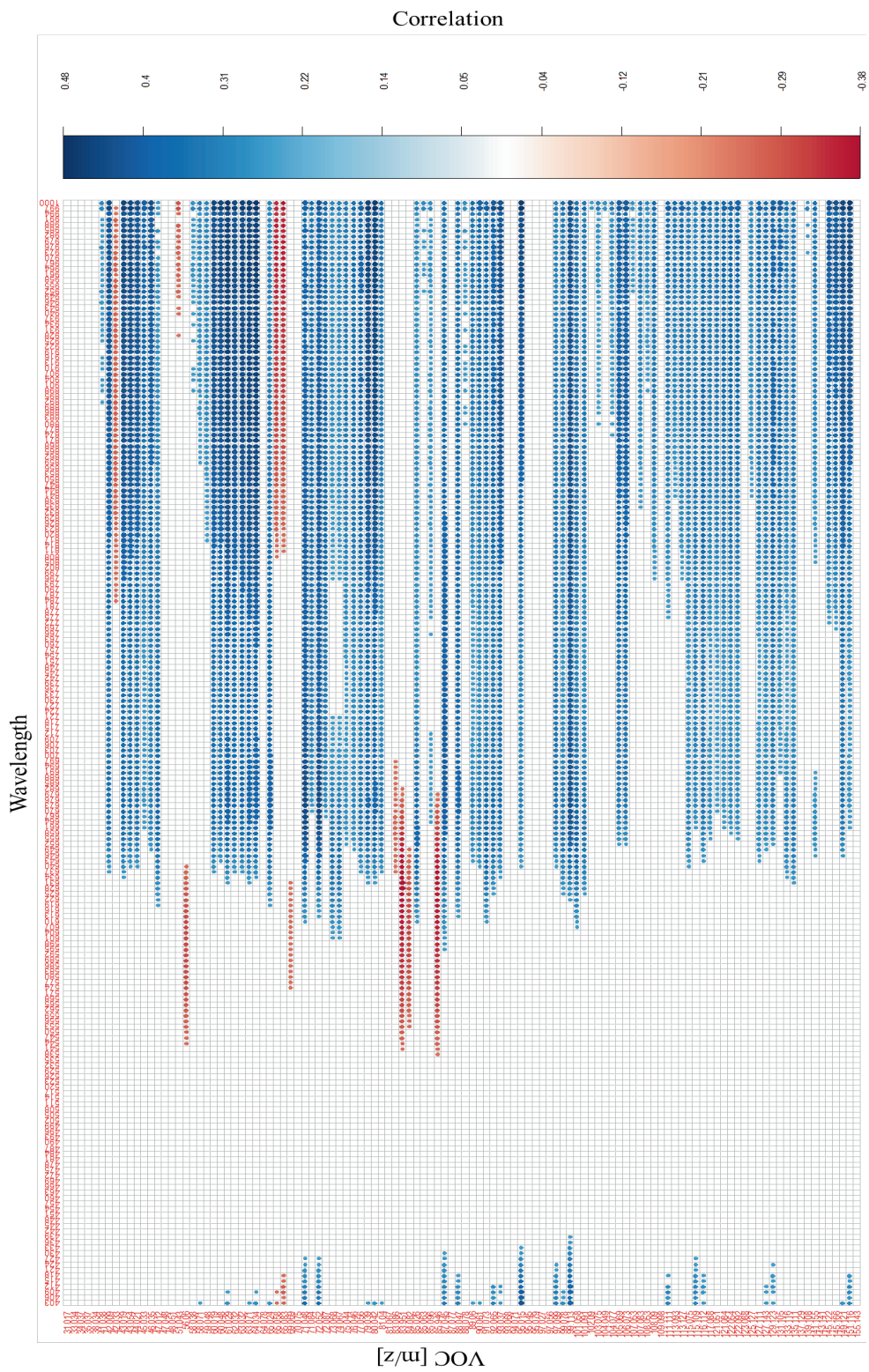
Figure S6.2. Significant Pearson correlations between volatile organic compounds (VOCs) and wavelengths.

Chapter 7

Discussion

As a result of the increasing interest in sustainability and demand of certifications in the chocolate sector and the emerging concerns regarding food quality and safety, a growing effort in the implementation of traceability systems to guarantee the origin of the raw material along the cocoa-chocolate supply chain has been seen in recent years.

In the current study, cocoa beans origin characteristics in terms of botanical and geographical origin along the cocoa-chocolate supply chain have been studied. Specifically, the cocoa beans botanical origin refers to the botanical varieties Criollo, Forastero and Trinitario, while the cocoa beans geographical origin refers to the production area such as region, country and/or continent. Compositional characteristics of cocoa beans were studied after fermentation and drying, after chocolate manufacturing and during chocolate consumption. This information can be used to underpin a link between the raw material and the finished product, which in turn can be used as a basis for future employment of traceability systems and enhanced certification support.

To investigate the cocoa-chocolate botanical and geographical origin information, markers related to intrinsic properties were used: volatile and non-volatile characteristics. The latter comprised small non-volatile compounds, elements, stable isotope ratios and hyperspectral profiles. For each characteristic, the persistence of the previously mentioned intrinsic markers from the raw material to the consumers' consumption will be discussed in the current chapter.

This chapter is subdivided according to the observed reflectance of the botanical and geographical origin traits along the cocoa-chocolate supply chain. Furthermore, the implications of the identified natural features of cocoa beans along the supply chain are discussed.

7.1 Cocoa beans' botanical characteristics reflected along the supply chain

7.1.1 Volatile characteristics
In this study, the volatile profiles as intrinsic characteristics showed the potential to track and trace raw material origin along the entire supply chain. The quantitative differences measured in the headspace of cocoa beans and chocolates (Chapter 5) and the inter-individual differences found in the nosespace (NS) profiles within the subjects (Chapter 2) did not influence the detection of raw material related VOCs in a chocolate bar and neither during chocolate consumption. Although cocoa beans were not tested in terms of botanical origin, in the current study VOCs appear to represent a more robust link between cocoa beans and corresponding chocolates compared to elemental and isotopic composition (Chapter 5). This link was highlighted in our study (Chapter 5) by compounds identified as fragment of diverse origin (m/z 43), acetic acid (m/z 61), benzene (m/z 79), pyridine (m/z 80) 2-phenylethanol (m/z 123) and maltol (m/z 127) using High Sensitivity-PTR-MS (HS-PTR-MS). These compounds can be linked with both raw material and primary processing, such as fermentation and drying (Frauendorfer & Schieberle, 2008), suggesting that intrinsic features of the cocoa beans are maintained even after secondary processing (winnowing, roasting, alkalization, grinding, conching, tempering, moulding and packing). According to the literature, some of the aforementioned compounds can be indicators of specific botanical origin. For instance, Qin et al. (2017) found high amount 2-phenylethanol in Trinitario cocoa beans, while a high amount of acetic acid was found in Criollo beans. The link between acetic acid and Criollo variety is also highlighted by our results when analysing chocolate (Chapter 3). According to other studies, chocolates produced from Criollo cocoa beans have higher acidic notes, related to acetic acid, compared to the chocolates produced with Forastero and Trinitario beans (Jinap, Selamat & Dimick, 1990; Saltini et al., 2013) supporting the results of our study. Even though, during chocolate consumption (Chapter 2), the dominance of acetic acid became less evident, the concentration of acetic acid in Criollo chocolates was still higher than Forastero and Trinitario chocolates. In our study, it was possible to detect acetic acid from the cocoa beans till the chocolate consumption identifying this compound as a potential botanical marker. Acetic acid showed the highest concentration also in other studies that focused on cocoa liquor (Rottiers et al., 2019) and chocolate (Deuscher et al., 2019) indicating to be a predominant compound along the supply chain.

Apart from the clear botanical trait of acetic acid, it was difficult to identify other markers persisting across the entire supply chain. Both in the headspace and in the NS profile of chocolate, it was possible to identify botanical markers but no link was found within them.

The difficulties in identifying individual markers when investigating chocolate headspace VOCs concentrations/compositions is most likely due to the interaction between cocoa botanical and/or geographical origin and the brand (i.e. formulation and processing) reducing the possibility to find markers linking the different steps along the supply chain (Chapter 3). However, within the 136 HS-PTR-MS masses detected in the chocolates in our study (Chapter 3), the masses identified as heptanoic acid and furfural appeared to be relevant for Criollo as well as for Criollo and Trinitario together, respectively. Heptanoic acid has been highlighted by Cevallos-Cevallos et al. (2018) as one of the main volatiles responsible for the differences between Criollo and Forastero cocoa beans from the same origin. Heptanoic acid was found to be a potential marker linking the cocoa beans and chocolates (Chapter 5), however, it was influenced by the cocoa content of the chocolates. This should be taken into consideration when analysing chocolate samples with different cocoa percentage. On the other hand, Magagna et al. (2017) described furfural as a technological marker in cocoa beans used for fine chocolate production. Being furfural generated via thermal processes, it is most probably indicative of the differences in roasting conditions between fine (Criollo and Trinitario) and bulk (Forastero) variety.

During chocolate consumption, the botanical information was consistently reflected in the NS profiles (Chapter 2) allowing a clear differentiation of Criollo chocolate from Forastero and Trinitario chocolates. When comparing chocolate headspace with chocolate NS profile, it was evident that compounds with higher m/z are becoming more dominant during eating (Chapter 2). From one side, this difference made it more difficult to link NS profile markers with cocoa and chocolate markers. On the other hand, the dominance of higher molecular weight volatile compounds during chocolate consumption (mastication and swallowing) gives the possibility to identify/find botanical markers that clearly distinguished Criollo chocolates from the others by their NS profiles (Chapter 2). Our results are in line with those reported by Sukha, Butler, Umaharan and Boult (2007), who investigated the possibility to distinguish cocoa botanical traits during consumption of cocoa liquor. Even if there were difficulties in differentiating Forastero and Trinitario NS profiles, the results of this thesis underline the ability of NS profiles to differentiate 'bulk' and 'fine' variety traits also after cocoa liquor processing and during chocolate consumption.

Despite the difficulties in identifying specific botanical markers along the supply chain, the possibility to find botanical information from cocoa beans to chocolate consumption highlights that the compositional botanical characteristics are a reflection of the global VOC profiles (fingerprints).

7.1.2 Non-volatile characteristics

In this study, the elemental composition and the stable isotope ratios of cocoa beans and chocolates were tested as non-volatile characteristics (Chapter 5).

Within the non-volatile compositional characteristics some origin markers were identified (Chapter 5). For example, Fe, Cr and Cd significantly correlated cocoa beans with the corresponding chocolates, indicating the possibility to use them as raw material markers in the final product. However, no literature indicating a possible link between these elements and the botanical origin of cocoa beans is available. Furthermore, the stable isotopes did not show a significant correlation between cocoa beans and chocolates (Chapter 5). According to Diomande et al. (2015), $\delta^{15}N$ was significantly higher for Forastero cocoa beans compared to Trinitario. However, the findings of the aforementioned study could be due to geographical differences more than botanical origin of the cocoa beans (Diomande et al., 2015) hence making it more difficult to use the stable isotopes for the cocoa botanical origin detection.

When following the non-volatile characteristics along the supply chain, the influence of the ingredients is evident (Chapter 4-5). A clear example of this is the trend of Na in chocolate (the lower the cocoa content, the higher the proportion of extra ingredients added). Regarding the stable isotopes, the shift from cocoa beans to chocolates is most likely due to added ingredients. Therefore, stable isotope signatures appear to be unsuitable for tracing the characteristics of cocoa in chocolate products back along the supply chain. To accurately estimate the isotopic signature of multi-ingredient food products, the quantity and the isotopic pattern of each ingredient should be determined separately as showed by Bostic, Palafox, Rottmueller and Jahren (2015).

The analysis of the chocolate small non-volatile fingerprint generated by Flow Infusion-Electrospray Ionization-MS (FI-ESI-MS) (Chapter 4) confirmed the influence of the brand and/or ingredients on the non-volatile composition. To extract botanical information from the chocolate samples, it is necessary to reduce the brand influence on the chocolate profiles (Chapter 4). We speculate that brand-related influences could have arisen from the utilization of different ingredients. For instance, sugar and amino acids are added at certain stages of cocoa bean processing and this makes the evaluation of these compounds in the finished product difficult (Pätzold & Brückner, 2006). Moreover, in this study, within the tentatively identified compounds, ricinoleic acid can be directly linked with industrial production. Indeed, this compound is an emulsifier used to improve the viscosity of chocolate (Bódalo-Santoyo, Bastida-Rodríguez, Máximo-Martín, Montiel-Morte & Murcia-Almagro, 2005).

Although industrial processing can influence non-volatile components of chocolate, it was still possible to identify relevant botanical markers (Chapter 4). Within the non-volatile compounds identified in chocolate fingerprints generated by FI-ESI-MS, masses 132.1018 and 166.0860 tentatively identified as isoleucine and phenylalanine, respectively, were mostly related to Criollo chocolates. A link between the Criollo chocolates and amino acids content can be associated with the differences in fermentation time within the three varieties. For instance, Criollo beans undergo a short fermentation process as well as instances of omitting fermentation all together compared to Forastero and Trinitario cocoa beans (Caligiani, Acquotti, Cirlini & Palla, 2010). According to Hue et al. (2016), the fermentation time has a significant effect on the amino acids concentration, therefore the differences in fermentation length within Criollo, Forastero and Trinitario could lead to different amino acid concentrations in cocoa beans and chocolate. However, more data about amino acids contents in cocoa beans are needed to confirm this.

The volatile (7.1.1) and non-volatile (7.1.2) markers of botanical origin could be merged by data fusion. For instance, the amount of flavour precursors such as amino acids identified in the finished product (Chapter 4) could support the quantity of certain volatile compounds and explain certain botanical traits and possibly link them with fermentation or roasting conditions providing a more in-depth picture of the chocolate's journey from farm to mouth. For instance, the amino acids identified in Chapter 4, isoleucine and phenylalanine, through the Strecker degradation, participate in the formation of amino acid specific aldehydes such as 2-methylbutanal and 2-phenylacetaldehyde that, in our study, have been found along the supply chain. However, due to the difficulties in extracting the botanical information from the non-volatile profiles caused by the brand/ ingredients influence, the applicability of the data fusion methods for volatile and non-volatile is reduced. Other studies tried to link flavour precursors and formed flavour compounds, but no clear quantitative correlation has been found (Frauendorfer & Schieberle, 2008; Tran et al., 2015). In these studies, a significant variation between cocoa botanical origins has been observed, supporting our results, but no relationship has been established among the flavour potential and the type of precursors generated. Hence, more research is needed to find the reasons that cause dissimilarities between the botanical origins in terms of flavour precursors (amino acids, peptides, reducing sugars) which are responsible for the flavour (Caligiani, Marseglia, Prandi, Palla & Sforza, 2016).

7.1.3 Summarising remarks for botanical characteristics

The possibility to differentiate the cocoa varieties has been largely tested in this thesis, although cocoa beans were not tested in terms of botanical origin. From the previous paragraphs (7.1.1 and 7.1.2), it is clear that the botanical characteristics affect both cocoa beans' and chocolates' headspace, consumers NS VOC profiles and the small non-volatile compound fingerprints of chocolate, indicating that intrinsic features of the beans are maintained after processing and even during consumption.

VOCs were found to be the most robust markers that were carried from cocoa beans to consumer products, revealing a constant shift/change between the raw materials and the finished products. The significant "volatile" link between cocoa beans and corresponding chocolates provides a good potential for track and trace options of cocoa beans from farm to chocolates. Potential markers (i.e. acetic acid) were detected to reveal botanical characteristics from raw material to chocolate. However, in general, it was difficult to identify a specific marker carried along the supply chain. When analysing the volatile botanical characteristics from cocoa beans to consumers, different markers where identified mainly because of the matrix changes and the interaction between cocoa origin traits and/or brand. The difficulties in identifying specific markers along the supply chain highlights the fact that the botanical characteristics are more reflected in the total volatile and non-volatile fingerprint.

7.2 Cocoa beans' geographical characteristics reflected along the supply chain

In this section, the geographical characteristics carried along the supply chain are discussed, focusing on cocoa beans, chocolates and during chocolate consumption. Both VOCs and non-volatile compounds have been investigated as potential markers. VOCs were followed from cocoa beans (Chapter 5-6) to chocolate (Chapter 3-5) and from chocolate to consumers (Chapter 2). The non-volatile compounds were studied to identify geographical information in cocoa beans (Chapter 5-6) and in finished chocolate after processing and flavour development (Chapter 4). We tested the possibility of using hyperspectral imaging to scan cocoa beans' origin traits (Chapter 6) and measured elemental and isotopic composition both in cocoa beans and chocolates (Chapter 5).

7.2.1 Volatile characteristics

Geographical origin defines the cocoa beans' volatile profile and final flavour potential (Counet, Callemien, Ouwerx & Collin, 2002; Liu et al., 2017). This has also been highlighted in our study by the significant correlation found between the volatile profile of cocoa beans and the chocolates (Chapter 5), indicating the possibility to extract raw material information. In this thesis, VOCs reflected the geographical characteristics from cocoa beans to consumers. However, we have observed that the geographical characteristics were often hindered by other factors along the supply chain. For instance, in cocoa beans, the influence of the geographical origin on the volatile profile was probably affected by the variability related to differences in primary production (fermentation and drying) within a country and between countries (Chapter 6). Similar to our findings, other studies also highlighted difficulties in identifying geographical information in cocoa beans, for example, Tran et al. (2015) encountered difficulties in discriminating the geographical origin of cocoa beans due to the more pronounced influence of the botanical origin on the volatile profile.

To reduce the above-mentioned botanical impact on cocoa beans, we focused on Forastero botanical origin when analysing the cocoa beans' geographical origin (Chapter 6). However, also in this case, it was difficult to have a clear geographical discrimination. African and American cocoa beans clusters were visible whereas a non-defined trend was observed for the cocoa beans originating from Southeast Asia. In a previous study where the same sample set was measured with High Resolution Magic Angle Spinning 1H Nuclear Magnetic Resonance (HR MAS 1H NMR) and 1H NMR, similar observations were made related to geographical origin discrimination, therefore supporting our findings (Marseglia, Acquotti, Consonni, Cagliani, Palla & Caligiani, 2016). The difficulties in differentiating the geographical origin can be related to the introduction of new growing regions as well as new hybrids and clone varieties (Tran et al., 2015) influencing the expression of the geographical traits. Moreover, we speculate that the impact of fermentation and drying on the volatile profile reduced the extraction of the geographical information. In our study, section 7.1.1, it was concluded that acetic acid in chocolate is most likely related to botanical traits more than to geographical traits (Chapter 3), as also stated by Tomlins, Baker, Daplyn and Adomako (1993). However, when one botanical origin was considered (Chapter 6), acetic acid caused a high variability within the cocoa samples and reduced the visualisation of geographical trends. The explanation of the variation in acetic acid is related to differences in fermentation conditions as stated by Jinap, Selamat and Dimick (1990) supporting our results. Different fermentation and drying procedures are applied depending on the climatic conditions or based on the traditional procedures applied in specific areas (Afoakwa, 2014a; Afoakwa, 2014b). These observations

suggest differences caused by fermentation level and drying conditions, overpassing that of geographic origin, confirming the results of Tran et al. (2015).

Reducing the influence of fermentation and/or drying related to acetic acid, it was possible to pinpoint markers for three geographical areas (Africa, Americas, Southeast Asia) (Chapter 6). Extracting geographical information along the supply chain allowed us to pinpoint geographical markers linking cocoa beans (Chapter 6) and chocolate (Chapter 3). However, these markers could not be highlighted during chocolate consumption (Chapter 2). Benzonitrile (m/z 104.049), benzaldehyde (m/z 107.053) and heptanal/2-heptanone (m/z 115.109), detected by PTR- Time of Flight-MS (PTR-ToF-MS) were relevant in differentiating both cocoa beans and chocolates from the Americans (Chapter 3 and Chapter 6). Regarding the African cocoa beans and chocolates, alkylic fragment (m/z 41.038) and acetaldehyde (m/z 45.033) detected by PTR-ToF-MS were identified as geographical markers (Chapter 3 and Chapter 6). The behaviour of 2-methylpropanoic acid and 2-phenylethanol is controversial. These compounds are indicative of African-Southeast Asian cocoa beans but in chocolates, they are indicative of South American samples. From one side, this result highlights the relevance of these compounds as markers along the supply chain and from the other side the difficulties in extracting geographical traits. Moreover, in this case the differences can be related to matrix changes during processing. For instance, 2-phenylethanol is influenced by the chocolate fat content (Afoakwa, Paterson, Fowler & Ryan, 2009). Depending on the fat content, this could alter the expression of geographical characteristics. However, the possibility to trace 2-phenylethanol along the supply chain and its ability to significantly link cocoa beans and chocolates (Chapter 5) make it a plausible marker. When linking raw material and the corresponding finished product (Chapter 5), some volatile compounds were found to correlate significantly. Among them maltol (m/z 127) has also been found by Rottiers et al. (2019) as relevant in discriminating South American cocoa liquor, adding this compound to the list of the markers relevant for geographical traits.

When analysing cocoa beans (Chapter 6) and chocolates (Chapter 3) more compounds were involved in the geographical characterisation (Chapter 3). However, it was not possible to consistently follow them along the supply chain (Chapter 5-6). This could be due to the above-mentioned matrix changes along the supply chain influencing the VOCs concentrations (Chapter 5) and/or to the interference between botanical geographical and brand characteristics highlighted during chocolate analysis (Chapter 3).

In chocolate, the geographical differences (Asia, Africa, South America) were evident (Chapter 3). This is supported by the geographical differentiation highlighted in cocoa liquor by Liu et al. (2017) and Counet et al. (2004) and in chocolate by Torres-Moreno, Tarrega, Costell and Blanch (2012). However, in Chapter 3, the proved interference with botanical characteristics and with brand in terms of processing and/or ingredient addition, lead to a partial loss of the geographical origin features. The interaction between the botanical and geographical traits was also highlighted by Rottiers et al. (2019) supporting our results. Interestingly, for several compounds Forastero chocolate samples overlapped with African chocolate samples and Criollo chocolate samples with South American chocolate samples (Chapter 3). This could be linked to the fact that the majority of Forastero is cultivated in Africa while Criollo is typically from South America (Afoakwa et al., 2008). However, a question arises if the geographical information could be covered by the botanical information if all the brands included in the analysis were specifying the botanical origin and not only the geographic origin of the cocoa beans used in the production.

Moving further along the supply chain, the geographical characteristics during chocolate consumption was covered by the botanical characteristic of chocolate (Chapter 2). In fact, a geographical trend is visible only when analysing the clusters related to the botanical origin separately in a two-step approach. Specifically, for Criollo and Trinitario chocolates, it is possible to notice that the NS profiles of chocolates made with beans from the same single-origin are located close to each other. We could speculate that the compounds exhaled during consumption are indicative of the botanical origin while the compounds concentrated in the chocolate headspace could provide more insights in the process and origin (Chapter 2). The difficulties in extracting geographical traits from the NS profile reduced the possibility to pinpoint geographical markers during consumption and thus the possibility to follow markers along the entire supply chain.

7.2.2 Non-volatile characteristics

The analysis of the non-volatile compositional characteristics highlighted a general geographical pattern for the hyperspectral fingerprints of cocoa beans (Chapter 6). The hyperspectral fingerprints of cocoa beans showed a geographical differentiation similar to the one highlighted by the VOC profiles (Chapter 6): African and American cocoa beans clusters were visible whereas a non-defined trend was observed for the cocoa beans originating from Southeast Asia. These results indicate a possible and unexpected link between the two measurements and confirm the difficulties in differentiating the Asian samples.

From the hyperspectral data of cocoa beans (Chapter 6), we can identify geographical markers supported by a previous study measuring the same set of samples with NMR (Caligiani et al., 2010). Fatty acids and amino acids play a role in the distinction of the cocoa beans from the Americas while sugars such as glucose and fructose are relevant in discriminating the African cocoa beans. Differences in amino acids and sugars can be possible evidences of different levels of fermentation. The possibility to discriminate cocoa beans' fermentation using HSI was proved in a recent study analysing the whole cocoa bean (Caporaso, Whitworth, Fowler & Fisk, 2018). However, it was not clear whether the differences in fermentation were related to the differences in colour (León-Roque, Abderrahim, Nuñez-Alejos, Arribas & Condezo-Hoyos, 2016) or nitrogen compounds (Hue et al., 2014).

The relevance of the amino acids in geographical differentiation of cocoa beans highlighted by the hyperspectral data (Chapter 6) could support the possibility to use amino acids detected in chocolates (Chapter 4) as potential origin markers, even if in section 7.1.2 we stated that they were most likely botanical markers. These findings point out the interaction between botanical and geographical origin and the difficulties in highlighting specific markers. According to our results, when reducing the brand influence on the non-volatile fingerprints of chocolate (Chapter 4), it was possible to improve the separation of the samples according to their geographical origins. Nevertheless, it was evident that the overlapping of the samples was related to the botanical influence. The compounds tentatively identified (Chapter 4) could help in solving the botanical and geographical interaction, however, also in literature there is no clear explanation of which characteristics are responsible for the non-volatile composition in cocoa and cocoa product. This makes it difficult to further explain some of our results and reduce the possibility to use the non-volatile profiles to track and trace the supply chain. Despite these limitations, based on the small non-volatile compound profiles of chocolates (Chapter 4) we could differentiate chocolates manufactured from Asian and African cocoa beans and both groups versus the chocolates produced from South American beans. The latter case is similar to the cocoa beans geographical trend described for volatile and hyperspectral profiles. However, the clear difference between African and Asian origins is a specific geographical trait most probably related to non-volatile characteristics of chocolate. The geographical trend highlighted by the non-volatile profiles can be explained by the spread of the geographical areas evaluated in this study but also by the possible similarities between the different origins. For instance, Chaiseri and Dimick (1989) observed similarities within the triacylglycerol composition of cocoa butter from Costa Rica, Dominican Republic, Mexico and Africa.

The investigation of geographical discrimination of chocolate in our study was further complicated by the influence of the added ingredients (Chapter 4). Evidence related to ingredient influence on non-volatile compositional traits was previously highlighted in section 7.1.2. This specifically affects the possibility to link cocoa beans and chocolate using the stable isotopic distribution without previously estimating the isotope signature of all chocolate ingredients. Despite the influence of the added ingredients, the analysis of the non-volatile compositional characteristics highlighted the possibility to use Cr, Fe and Cd as significant markers linking cocoa beans and the corresponding chocolates (Chapter 5). Differently from the botanical origin, Cr, Fe and Cd has been highlighted in literature as geographical markers even though they are susceptible to environmental contamination (de Araujo et al., 2017; Bertoldi et al., 2016). Apart from its toxicity, Cd could be informative of the soil in the growing location (de Araujo et al., 2017). Vanderschueren et al. (2019) also found a significant geographical difference in Cd concentrations when analysing chocolate. Interestingly, in the latter study about chocolate (Vanderschueren et al., 2019), a misclassification of Asian Pacific samples was shown.

Specifically, the misclassification of Asia Pacific samples was due to the large geographical heterogeneity and corresponding differences in soil compositions and climate within this group as it comprised a combination of islands and mainland origins. These results highlight the same difficulties in differentiating Southeast Asian samples found in Chapter 6 when analysing the VOCs and hyperspectral profiles of cocoa beans, supporting our results and indirectly linking cocoa beans and chocolates.

The results related to the non-volatile geographical characteristics of cocoa beans and chocolates, identified an incapability in sufficiently separating chocolates produced from beans of the three continents at the same time. However, we could speculate that these difficulties are related to the botanical origin of the cocoa beans (Chapter 2 and Chapter 4) and brand influence (Chapter 4). As mentioned for the volatile compounds, the majority of the analysed brands did not specify the botanical origin of the beans. This information could change the interpretation of the results.

According to results discussed in this section and in section (7.2.1), the interesting correlation found between VOCs and hyperspectral data of cocoa beans, it could be further implemented via the data fusion methods. This could help in extracting more information about geographical origin and primary processing. Moreover, the hyperspectral data could be linked with the volatiles' patterns of chocolate and NS profiles as well. The hyperspectral camera could be used to scan the cocoa beans and predict the chocolate and NS profile. There are no studies available on this topic, except Davies et al. (1991) who used the spectral data of cocoa beans to predict the quality of chocolate. These results highlight a possible applicability of hyperspectral imaging in linking cocoa beans and chocolate profiles. However, more indications about the possible link within VOCs and hyperspectral profile are needed.

Similar to the suggested data fusion approach, the volatile and the small non-volatile markers highlighted in chocolate (Chapter 3 and Chapter 5) could be combined and used in the same analysis. This is mainly because the flavour precursors such as amino acids, identified within the non-volatile detected could be used to justify the concentrations of certain VOCs. However, the volatile and non-volatile profiles showed a different trend towards geographical origin discrimination related to the different information carried by the volatile and non-volatile and the different matrix composition of cocoa beans, chocolates and during consumption. This might make it challenging and difficult to merge the data and empower the extraction of geographical information.

7.2.5 Summarising remarks on geographical characteristics

We extracted potential geographical markers along the cocoa supply chain. Some of them were specific for cocoa beans or chocolates. Some of these markers, specifically the volatile compounds and the elements, were able to link raw material and finished product's geographical traits.

The results of this section support the hypothesis that the geographical information is carried along the supply chain. However, it can be hidden by other factors such as the botanical characteristics and the changes in the product matrix (from beans to chocolate bar) by production steps. Knowing the changes and variations identified by our results can help in keeping track of the geographical information. Among the characteristics studied, VOCs appeared to be strong markers along all the whole supply chain. The elemental composition of cocoa beans and chocolates strongly correlated the raw material and the finished product's geographical origin. Specifically, we can pinpoint Fe, Cr and Cd as linking markers.

7.3 Implications and limitations of identified cocoa beans characteristics along the cocoa-chocolate supply chain and recommendations for future research

The possibility to detect the cocoa beans' origin traits in the finished chocolate bar and during consumption shows the feasibility to verify origin of cocoa beans and to follow the raw material traits along the supply chain. Knowing the farming practices, including fermentation and drying, the actors involved in the trading and the manufacturing conditions could help in further describing the volatile, non-volatile compositional characteristics evaluated in this thesis as a fingerprint that can reliably be utilised to track and trace the cocoa–chocolate supply chain. The production information and the information obtained by the analytical methods could be included in a blockchain and support traceability. For this kind of practical application, large databases need to be developed. Moreover, the compounds that are able to link cocoa beans and chocolates should be validated by repeating the measurements and taking into account more geographical origins, different seasons and years of production. The blockchain method combined with the appropriate analytical techniques could support the supply chain of specialty chocolates such as the single-origin chocolate. This process could increase traceability and confidence in the certification system.

Specifically, the volatile fingerprints seem to have the potential to be used as a powerful tool for tracing the cocoa origin. In this view, there is a prospective to combine this link with a traceability system to implement optimal parameters for cocoa beans processing to control their flavour potential.

One of the main issues we encountered is the interaction between botanical and geographical origin. In order to decrease or solve the interaction between botanical and geographical characteristics, it is recommended to study the robustness of the markers as identified in this thesis. For instance, furfural appeared to be relevant for Criollo chocolate in Chapter 3 but in Chapter 6 it was relevant for African and South-East Asian cocoa beans. Moreover, 2-phenylethanol has been found relevant for geographical characteristics of cocoa beans and chocolate in this thesis but is considered important for the cocoa botanical characteristics in the literature. It could be interesting to test the robustness of the identified markers and determine the relevance of the impact of a trait compared to the others. Despite this interaction, it was possible to extract botanical and geographical information. In the next session, we will discuss implementation and limitations related to botanical and geographical origin detection and how to utilize their information along the supply chain.

7.3.1 Botanical origin

To confirm the possibility to track and trace the botanical origin along the supply chain, it is suggested to evaluate the cocoa beans profile taking into account the origin and the farmer practices (primary processing). For instance, current knowledge could be improved by analysing cocoa beans from the different botanical origins but from the same geographical origin and subjected to the same conditions of fermentation and drying. Having the raw material information and comparing the cocoa beans with the corresponding chocolate, as has been done in Chapter 5, could provide more information about which volatile and non-volatile compounds persist in the chocolate and could further support the chocolate markers highlighted in this thesis. When analysing commercially available chocolates it is recommended to conduct a brand analysis and interview the producers to gain more information about the ingredients and production (secondary processing). For instance, more indications about the botanical origin are needed as the majority of the brands tend to specify the geographical origin and not the botanical one. This could be useful specifically when analysing the non-volatile profiles to explain the distribution of the samples and to better understand the possible impact of the botanical characteristics. Moreover, because of the strong influence of the brand on the non-volatile profile, specific information of the ingredient is also needed. However, the ability to extract brand information using the non-volatile profile could be used to guarantee the brand characteristics along the supply chain.

The collection of brand information in terms of ingredients added and secondary process conditions are recommended also when analysing a single brand. For instance, in Chapter 2, the NS profiles reflected the difference between bulk and fine varieties. However, having more information on

the processing steps applied for the different varieties could be useful in understanding if the differences shown by the NS profiles are related to the natural traits of the cocoa beans or the processing steps.

Despite the aforementioned limitations, according to the results of this thesis, the possibility to extract botanical origin along the supply chain may be important for the discrimination of bulk and fine variety. This differentiation could be specifically relevant in the case of bulk and fine varieties that are produced in the same country.

7.3.2 Geographical origin

This study demonstrated the evident possibility to extract information regarding cocoa beans origins in the context of fermentation and drying procedures. Furthermore, the impact of the processing location on the cocoa beans profile has been elucidated and exploited to gain additional origin information. Combining these data, it could be possible to further describe the origin of cocoa beans and in turn substantiate the cocoa terroir.

This model can provide a scientific basis for the cocoa quality and certification programs. In this thesis, Forastero beans from different origins have been evaluated. It is recommended to analyse volatile and non-volatile profiles of all the different cocoa varieties or the most common commercial ones, considering the highest variability for each group in terms of geographical origin to better understand the geographical information within a specific variety and between the different cocoa varieties. More information on the fermentation and drying such as time and temperature could help in identifying markers and explaining the distribution of the samples. This information could improve the link between raw material and the finished product in terms of origin and furthermore could improve transparency along the supply chain. The possibility to use the hyperspectral profiles to detect fermentation and/or drying information is a promising step in tracing cocoa along the supply chain. It is recommended to enlarge the database as suggested to confirm the trust ability of the results. Because of the highlighted botanical impact, specifically when analysing the non-volatile profile of chocolates, these results should be validated by repeating the analysis of commercially available chocolate in different years or seasons. On the other hand, the strong geographical link highlighted by the elemental characteristics can be validated by including additional origins together with the analysis of the environmental conditions such as climate and soil, is recommended to confirm the robustness of the markers as pinpointed in this thesis. Regarding the isotopic signature, it is highly recommended to get the isotopic information of each ingredient to increase the link raw material-finished product. Moreover, it is suggested to focus the attention on the single origins rather than a continental grouping of the samples.

It was more difficult to extract geographical characteristics during chocolate consumption. This can be related to a larger influence of botanical traits on the final profile but also to a matrix change during mastication and swallowing. Considering a possible comparison within NS profiles and sensory analysis, it is recommended to support the NS analysis with sensory evaluations to explore the extraction of geographical information further.

7.4 Final conclusions

The objective of this thesis was to explore the botanical and geographical origin characteristics of cocoa beans carried along with the cocoa–chocolate supply chain. Compositional characteristics of the botanical and geographical origins of cocoa beans after fermentation and drying, after manufacturing chocolate and during chocolate consumption were deciphered. The compositional characteristic of cocoa beans chocolates and chocolate during consumption were evaluated in terms of volatile and non-volatile characteristics.

This thesis showed that it is possible to extract cocoa beans origin, specifically it can be concluded that:

The **botanical traits** influence the volatile and non-volatile characteristics of chocolate and the profile of volatile compounds released during consumption of chocolates, which indicates that intrinsic features of the beans are retained after processing and they are even present during consumption.

The **geographical traits** influence the volatile and non-volatile characteristics along the supply chain from beans to chocolate consumption in addition to the botanical origin. This holds for small non-volatile compounds, as well as elements and general spectral traits. The geographical characteristics showed to be affected by the complex process of production of cocoa beans and chocolates and the blending techniques. However, these characteristics also provide opportunities to determine traits related to the growing environment and the processing steps at the growing locations. This all enriches the terroir characterisation of cocoa. The changes and variation pinpointed in this study can help to keep track of the geographical information.

According to the results of this thesis, the volatile characteristics are the most suitable for gathering information on the beginning of the production and for application at different stages of the supply chain from cocoa beans to chocolate. This information can complement the current traceability systems and reinforce certification systems to secure terroir added value.

References

Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2008). Flavor formation and character in cocoa and chocolate: a critical review. *Critical Reviews in Food Science and Nutrition*, *48*(9), 840–857. https://doi.org/10.1080/10408390701719272

Afoakwa, E. O., Paterson, A., Fowler, M., & Ryan, A. (2009). Matrix effects on flavour volatiles release in dark chocolates varying in particle size distribution and fat content using GC–mass spectrometry and GC–olfactometry. *Food Chemistry*, *113*(1), 208–215. https://doi.org/10.1016/j.foodchem.2008.07.088

Afoakwa, E. (2014a). Drying Techniques, Storage Practices and Trading Systems. In E. Afoakwa (Ed.), *Cocoa production and processing technology* (pp. 139–146). Boca Raton, FL, USA: Taylor & Francis Group

Afoakwa, E. (2014b). Post-harvest treatments and technologies of cocoa. In E. Afoakwa (Ed.), *Cocoa production and processing technology* (pp. 109-119). Boca Raton, FL, USA: Taylor & Francis Group

Bertoldi, D., Barbero, A., Camin, F., Caligiani, A., & Larcher, R. (2016). Multielemental fingerprinting and geographic traceability of Theobroma cacao beans and cocoa products. *Food Control*, *65*, 46–53. https://doi.org/10.1016/j.foodcont.2016.01.013

Bódalo-Santoyo, A., Bastida-Rodríguez, J., Máximo-Martín, M. F., Montiel-Morte, M. C., & Murcia-Almagro, M. D. (2005). Enzymatic biosynthesis of ricinoleic acid estolides. *Biochemical Engineering Journal*, *26*(2–3), 155–158. https://doi.org/10.1016/j.bej.2005.04.012

Bostic, J. N., Palafox, S. J., Rottmueller, M. E., & Jahren, A. H. (2015). Effect of baking and fermentation on the stable carbon and nitrogen isotope ratios of grain-based food. *Rapid Communications in Mass Spectrometry*, *29*(10), 937–947. https://doi.org/10.1002/rcm.7178

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

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

Caporaso, N., Whitworth, M. B., Fowler, M. S., & Fisk, I. D. (2018). Hyperspectral imaging for non-destructive prediction of fermentation index, polyphenol content and antioxidant activity in single cocoa beans. *Food Chemistry*, *258*(March), 343–351. https://doi.org/10.1016/j.foodchem.2018.03.039

Cevallos-Cevallos, J. M., Gysel, L., Maridueña-zavala, M. G., & Molina-miranda, M. J. (2018). Time-Related Changes in Volatile Compounds during Fermentation of Bulk and Fine-Flavor Cocoa (Theobroma cacao) Beans. *Journal of food quality, 2018*.

Chaiseri, S., & Dimick, P. S. (1989). Lipid and Hardness Characteristics of Cocoa Butters from Different Geographic Regions. *Journal of the American Oil Chemists' Society, 66*(12), 1771-1776.

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

Counet, C., Callemien, D., Ouwerx, C., & Collin, S. (2002). Use of gas chromatography-olfactometry to identify key odorant compounds in dark chocolate. Comparison of samples before and after conching. *Journal of Agricultural and Food Chemistry*, *50*(8), 2385–2391. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11929301

Davies, a. M. C., Franklin, J. G., Grant, a., Griffiths, N. M., Shepherd, R., & Fenwick, G. R. (1991). Prediction of chocolate quality from near-infrared spectroscopic measurements of the raw cocoa beans. *Vibrational Spectroscopy*, *2*(2–3), 161–172. https://doi.org/10.1016/0924-2031(91)85022-F

de Araujo, Q. R., Baligar, V. C., Loureiro, G. D. A., de Souza Júnior, J. O., & Comerford, N. B. (2017). Impact of soils and cropping systems on mineral composition of dry cacao beans. *Journal of soil science and plant nutrition, 17*(2), 410-428.

Deuscher, Z., Andriot, I., Sémon, E., Repoux, M., Preys, S., Roger, J. M., … Le Quéré, J. L. (2019). Volatile compounds profiling by using proton transfer reaction-time of flight-mass spectrometry (PTR-ToF-MS). The case study of dark chocolates organoleptic differences. *Journal of Mass Spectrometry*, *54*(1), 92–119. https://doi.org/10.1002/jms.4317

Diomande, D., Antheaume, I., Leroux, M., Lalande, J., Balayssac, S., Remaud, G. S., & Tea, I. (2015). Multi-element, multi-compound isotope profiling as a means to distinguish the geographical and varietal origin of fermented cocoa (Theobroma cacao L) beans. *Food Chemistry*, *188*, 576–582. https://doi.org/10.1016/j.foodchem.2015.05.040

Frauendorfer, F., & Schieberle, P. (2008). Changes in key aroma compounds of Criollo cocoa beans during roasting. *Journal of Agricultural and Food Chemistry*, *56*(21), 10244–10251. https://doi.org/10.1021/jf802098f

Hue, C., Gunata, Z., Breysse, A., Davrieux, F., Boulanger, R., & Sauvage, F. X. (2016). Impact of fermentation on nitrogenous compounds of cocoa beans (Theobroma cacao L.) from various origins. *Food Chemistry*, *192*, 958–964. https://doi.org/10.1016/j.foodchem.2015.07.115

Hue, C., Gunata, Z., Bergounhou, A., Assemat, S., Boulanger, R., Sauvage, F. X., & Davrieux, F. (2014). Near infrared spectroscopy as a new tool to determine cocoa fermentation levels through ammonia nitrogen quantification. *Food Chemistry*, *148*, 240–245. https://doi.org/10.1016/j.foodchem.2013.10.005

Jinap, Selamat, & Dimick, P. S. (1990). Acidic Characteristics of Fermented and Dried Cocoa Beans from Different Countries of Origin. *Journal of Food Science*, *55*(2), 547–550. https://doi.org/10.1111/j.1365-2621.1990.tb06806.x

León-Roque, N., Abderrahim, M., Nuñez-alejos, L., Arribas, S. M., & Condezo-hoyos, L. (2016). Talanta Prediction of fermentation index of cocoa beans (Theobroma cacao L.) based on color measurement and artificial neural networks. *Talanta*, *161*, 31–39. https://doi.org/10.1016/j.talanta.2016.08.022

Liu, M., Liu, J., He, C., Song, H., Liu, Y., Zhang, Y., … Su, X. (2017). Characterization and comparison of key aroma- active compounds of cocoa liquors from five different areas. *International Journal of Food Properties*, *20*(10), 2396–2408. https://doi.org/10.1080/10942912.2016.1238929

Magagna, F., Guglielmetti, A., Liberto, E., Reichenbach, S. E., Allegrucci, E., Gobino, G., … Cordero, C. (2017). Comprehensive Chemical Fingerprinting of High-Quality Cocoa at Early Stages of Processing: Effectiveness of Combined Untargeted and Targeted Approaches for Classification and Discrimination. *Journal of agricultural and food chemistry, 65*(30), 6329-6341.https://doi.org/10.1021/acs.jafc.7b02167

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

Pätzold, R., & Brückner, H. (2006). Gas chromatographic determination and mechanism of formation of D-amino acids occurring in fermented and roasted cocoa beans, cocoa powder, chocolate and cocoa shell. *Amino Acids*, *31*(1), 63–72. https://doi.org/10.1007/s00726-006-0330-1

Qin, X. W., Lai, J. X., Tan, L. H., Hao, C. Y., Li, F. P., He, S. Z., & Song, Y. H. (2017). Characterization of volatile compounds in Criollo, Forastero, and Trinitario cocoa seeds (Theobroma cacao L.) in China. *International Journal of Food Properties*, *20*(10), 2261–2275. https://doi.org/10.1080/10942912.2016.1236270

Rottiers, H., Tzompa Sosa, D. A., Van de Vyver, L., Hinneh, M., Everaert, H., De Wever, J., … Dewettinck, K. (2019). Discrimination of Cocoa Liquors Based on Their Odor Fingerprint: a Fast GC Electronic Nose Suitability Study. *Food Analytical Methods*, *12*(2), 475–488. https://doi.org/10.1007/s12161-018-1379-7

Saltini, R., Akkerman, R., & Frosch, S. (2013). Optimizing chocolate production through traceability: A review of the influence of farming practices on cocoa bean quality. *Food Control*, *29*(1), 167–187. https://doi.org/10.1016/j.foodcont.2012.05.054

Sukha, D. a., Butler, D. R., Umaharan, P., & Boult, E. (2007). The use of an optimised organoleptic assessment protocol to describe and quantify different flavour attributes of cocoa liquors made from Ghana and Trinitario beans. *European Food Research and Technology*, *226*(3), 405–413. https://doi.org/10.1007/s00217-006-0551-2

Tomlins, K. I., Baker, D. M., Daplyn, P., & Adomako, D. (1993). Effect of fermentation and drying practices on the chemical and physical profiles of Ghana cocoa. *Food chemistry, 46*(3), 257-263.

Torres-Moreno, Miriam, Tarrega, A., Costell, E., & Blanch, C. (2012). Dark chocolate acceptability: Influence of cocoa origin and processing conditions. *Journal of the Science of Food and Agriculture*, *92*(2), 404–411. https://doi.org/10.1002/jsfa.4592

Tran, P. D., Van de Walle, D., De Clercq, N., De Winne, A., Kadow, D., Lieberei, R., … Van Durme, J. (2015). Assessing cocoa aroma quality by multiple analytical approaches. *Food Research International*, *77*, 657–669. https://doi.org/10.1016/j.foodres.2015.09.019

Vanderschueren, R., Montalvo, D., De Ketelaere, B., Delcour, J. A., & Smolders, E. (2019). The elemental composition of chocolates is related to cacao content and origin: A multi-element fingerprinting analysis of single origin chocolates. *Journal of Food Composition and Analysis*, *83*(August), 103277. https://doi.org/10.1016/j.jfca.2019.103277

Summary

Summary

The assessment of cocoa's natural characteristics can be used to link cocoa beans and chocolate. Establishing a link between the raw material and the finished product is complicated not only because of complex characteristics of the supply chain but also because of the technology involved in the production and the composition of the ingredients. One may wonder why a link between raw material and the finished product is relevant. Nowadays, more and more attention is given to sustainability, pushing the cocoa manufacturers to take into account social, economic and environmental issues while producing. As a consequence, the demand in cocoa-importing countries for certified sustainable cocoa has increased considerably and is expected to show continuous growth over the next years. Moreover, more consumers have become increasingly aware of the existence of different cocoa varieties and their origins, and the market of specialty chocolates such as chocolates with single-origin bean, organic and fair-trade chocolate has largely grown in the last decades. The aforementioned trends in the chocolate sector and the growing concerns regarding food quality and safety led to a growing effort in the implementation of traceability systems. This trend has facilitated the need to verify and guarantee the origin of the cocoa beans along with the cocoa-chocolate supply chain and to establish a link between raw material and the finished product.

For these reasons, the main objective of this thesis is to discover cocoa beans' traits in terms of the botanical and geographical origin carried along the cocoa–chocolate supply chain. Compositional characteristics of botanical and geographical origin of cocoa beans after fermentation and drying, after manufacturing chocolate and during chocolate consumption were analysed. This information can be used to derive a link between raw material and the finished product, which in turn can be used to support traceability systems.

To investigate the cocoa-chocolate botanical and geographical origin information, markers related to intrinsic properties were used: volatile and non-volatile characteristics. The latter include small non-volatile compounds, elemental composition, isotope ratios and hyperspectral profiles (Chapter 2-6). The reflectance of the cocoa beans origin along the supply chain is investigated starting from the consumer perspective in Chapter 2 with the analysis of the nosespace (NS) profile of consumers eating chocolates manufactured from beans of different origins. Moving back along the supply chain, the possibility to extract botanical and geographical traits from both volatile and non-volatile profiles is further tested in chocolates available in the supermarket (Chapter 3 and Chapter 4). Chapter 5 compares cocoa beans and corresponding chocolates to extract raw material markers in a finished product analysing their volatile, elemental and isotopic compositional traits. Finally, the raw material analysis in Chapter 6 completes the assessment of cocoa beans origin traits along the supply chain.

Regarding the botanical traits carried along the supply chain, the volatile intrinsic characteristic showed the potential to track and trace raw material origin along the entire supply chain. It was possible to identify typical volatile compounds of the raw material in chocolate bar (Chapter 3) and chocolate during chocolate consumption (Chapter 2). VOCs were found to be the most robust markers that were carried from cocoa beans to consumer products, revealing a constant shift/change between the raw materials and the finished products (Chapter 5). Potential markers (i.e. acetic acid) were detected to reveal botanical traits from raw material to chocolate to chocolate consumption. However, in general, it was difficult to identify specific individual markers carried along the supply chain highlighting the fact that the botanical traits are more related to a general volatile pattern. The same is valid for the small non-volatile compounds analysed in chocolate (Chapter 4). Interestingly, when analysing the non-volatile compositional traits, the botanical trend explained the samples overlapping according to the geographical origin. This occurred despite the impact of the processing step or the ingredient added on the origin expression. The brand influence was confirmed in Chapter 5 when the elemental and stable isotope fingerprints of cocoa beans and chocolates were compared. Stable isotope signatures appear less sufficient for tracing the characteristics of cocoa in chocolate products without previously estimating the isotope signature of all the ingredients reducing the possibility to use them as origin markers. Our results support the botanical traits' influence as highlighted for the volatile and non-volatile profiles of chocolate and the volatile NS profile of chocolate during consumption and it indicates that intrinsic features of the beans are retained after processing and even during consumption.

Regarding the geographical origin traits carried along the supply chain, volatile compounds reflected the geographical trait from cocoa beans to consumers. Within the non-volatile compositional traits Cr, Fe and Cd appeared to be distinct geographical markers although they are susceptible to environmental contamination (Chapter 5). However, along the supply chain, the geographical traits were hidden by other factors. During chocolate consumption, geographical discrimination was possible only within each botanical group. In chocolate (Chapter 3 and Chapter 4), the geographical information was evidently contained by the volatile compounds (Chapter 3). However, we highlighted an interference with botanical and brand traits. This was particularly clear when analysing the small non-volatile compounds (Chapter 4). Only when reducing the influence of the brands, it becomes clear that the samples are spread according to the three origins. In Chapter 6, when reducing the botanical traits, the impact of the geographical origin on the volatile and hyperspectral profile of cocoa beans is influenced by the variability related to differences in production (fermentation and drying) within a country and between countries. Despite these interferences, a similar trend both in cocoa beans and in chocolates is visible for certain compounds, making them usable as cocoa-chocolate linking markers.

The evidence from this study suggests that volatile and non-volatile compositional traits of cocoa and cocoa products can be used to verify and/or follow origin traits along the supply chain and link raw material and the finished product. However, more research is needed on how to reduce botanical and geographical interaction to improve the characterisation of the product along the supply chain. Within the compositional traits that were measured, the volatile profile is the most suitable for gathering information at the beginning of the production and at different stages in the supply chain.

Acknowledgements

If you are reading this, it means that you have read all the previous chapters. After doing so, you probably realised that I could have not done this alone. Thankfully, there is this chapter to thank everyone.

First of all, I would like to express my gratitude to my supervisor **prof. dr. Saskia van Ruth** for giving me this opportunity. I know it is in not easy to deal with my "chaos" but somehow you trusted my abilities and gave me the freedom to grow as scientist in my own way. You helped me grow as a person and improved my scientific skills.

I could not write a single paper without the support of my daily supervisor **dr. Martin Alewijn**. Your enthusiasm in finding answers was contagious and you motivated me to work harder. Thanks for sharing your passion for statistics and data analysis. Now I love statistics! Thanks Saskia and Martin for your suggestions, criticism, and guidance throughout this period.

It was not easy in the beginning to move to another country and to adjust to a new life but luckily I have a big Italian family always ready to support me and to joke about my weak points, making everything funnier and less stressful. You are with many, **Mamma** and **Papà**, thanks for every single bag of pre-cooked rice, you saved my life a lot of times, **Mauro**, thanks for bringing me to Wageningen, **Ale**, thanks for joining me here, my amazing **Aunts** and **Uncles**, my "small" and "big" **Cousins** and all the family friends. You are my roots, you made me strong and remind me where I came from. I would like to express my gratitude to each one of you but I should write a second thesis and we are already at page 180. However, there is a special woman that needs a special thanks, my grandma. Grazie **Nonna** per prenderti ininterrottamente e instancabilmente cura di tutti noi. La tua forza e il tuo amore posso percepirli anche a km di distanza. Quello che sono lo devo anche a te.

Thanks for all my friends I grew up with that are still part of my life despite the distance. **Rori**, **Lilly**, **Ebby**, probably I could not finish my PhD without you waking me up every morning with your messages. To the Principesse dei Giardini **Anna**, **Carla** (and of course Lilly and Ebby), my crazy friends, thanks for letting me feel less crazy. **Gio** thanks for being always ready to support me and to pick up the phone every time it rang just once. **Davide**, thanks for sleeping on the sofa every time I came to visit you. **Silvy**, I am so happy to have you in my life. **Carmine** and **Marcella** without you Christmas will not make sense anymore. From now on I can come to Capri every weekend, be ready! **Clemente**, you are my secret paranymph. Many thanks to **Raffa**, **Vale P.**, **Benny, Gozzi**, **Andrea** (thanks for the cover), **Luigi**, **Marcella**, **Luciano** and **Angelo** as well. Thanks to all of you for calling me, texting me and lighten my mood up.

In the last years, a lot of things changed and I changed a lot. I became stronger and I learned to laugh when things became difficult. All of this thanks to you **PhD**. Thanks for the weekends in FQD and for the food intolerance. Yes, I did not have any before meeting you. However, these are only few of the things that I want to thank you for PhD. So, let's start from the beginning (a long time ago).

Thanks to you PhD, I spent some time in RIKILT (WFSR) where I met people that after all this time and besides some of them moved away, they are still part of my life. **Arie**, bello di mamma. I saw you growing up. I hope we will celebrate your 50th birthday with my grandma's lasagne. **Erika**, there are not enough words to thank you and I will look back with a smile thinking of us leaving as last in RIKILT. **Stefy** and **Victoria**, thanks for the time we spent together and the laughs. Thanks to you I started liking this place. **Patri P.**, thanks for the care and the attention. I hope to meet your children soon. **Patri L.**, I don't know if you should be in this group or in the housemate one. I think you are the only person with who I could work and live together. **Roby** my half Brazilian half Italian, I am so happy you moved to Italy and I can see you more often. **Eduardo,** thanks for supporting me during the Wednesday burn out sessions, you saved me! **Claudia**, we met in an uncomfortable situation, but you made my life very relaxed after that. **Marcia**, you are one of the first I met in Wageningen. Thanks for all the smiling faces every time you were passing by the office. I would also like to say thanks to my students **Anuja**, **Giorgia**, **Katharina** and **Victoria** (again) you made all this possible.

Acknowledgements

My PhD colleagues, **Isabelle**, **Gin**, **Gina**, **Stella** and **Hongwey** we shared so many things. If the office walls could speak they could tell amazing stories about six students that slowly became something more than colleagues. This is true also for you **Alfred**, thanks to you and **Thea** for your help and sorry **Aron** for taking time of your parents, they are amazing! Thanks to **Daniel**, **Sajad**, **Pedram** and **Ashraf** for all the coffee and tea breaks. Thanks to all of you for all the food you provided me with. I also deeply want to thank **Eric** and **Erika** for all the long talks about work and life. **Alex** for helping me during my lab and non-lab days. **Yannick,** thanks for the after work drinks and for trusting me in driving your car. **Erwin**, **Annemieke**, **Piet**, **Judith**, **Linda**, **Tjerk**, **Dave**, **Maikel** and **Michel** thanks to all of you for seeing me walking like an alien in the corridor and still giving me a smile. Thanks to my first office mate **Geert**, to **Yamine** and **Jenny** for finding always time for a little-big chat. My gratitude continues by acknowledging the help from **Irene**, **Gea**, **Jeannette**, **Leen** and **Robert**.

PhD you were split between RIKILT and FQD. FQD is like traveling around the world. I met people from different nationalities and religions, all having a peaceful lunch and drinks together. I deeply say thanks to all the past and present colleagues for sharing joyful and difficult moments: **Dylan, Elsa**, **Ruben**, **Renske**, **Grace**, **Arianne**, **Sara P.**, **Hanna**, **Burçe**, **Sara E.**, **Femke**, **Musti**, **Ayusta**, **Ita**, **Alim**, **Mohammed**, **Sydney**, **Zhijun**, **Lijiao**, **James**, **Chunyue**, **Faith**, **Ruth**, **Juliet**, **Shingai**, **Onu**, **Moheb**, **Tiantong**, **Qing**, **Li**, **Ling**, **Lintianxiang**, **Ana**, **Eduardo**, **Teresa**, **Erik**, **Pieter**, **Dario**, **Domenico**, **Ruud**, **Pieternel**, **Hein**, **Nicoletta**, **Anita, Catriona**, **Kasper**, **Femke**, **Lucia**, **Jonna**, **Andrijana**. How to forget the amazing time we had in Rome, Isabelle, Dylan, Ruben, Arianne, Jonna, Isabelle, Andrijana, Renske. I really enjoyed being your guide even if you should have not eaten the lasagne in that place. A special thanks to the PhD trip committee. It has been a pleasure to work with you. Next time we will collect more money! Hey PhD trip, thanks to you I met "friend-colleague" **Michi**. Thanks to all the amazing secretary-mum-sister-friend **Kimberley**, **Lysanne**, and **Corine**. FQD could not work without you and of course not without **Vincenzo**. Thanks for your advice and for being there even though officially we never worked together.

You PhD, you were a bit in Wageningen and a bit in Europe because of the Marie Curie project (PIMMS). **Nora** and **Sine**, the time we spent together during the PIMMS project was not just another "episode" of our life. After years, we are still together. I still remember the day we thought it was the last time we would meet and we were crying a lot. I do not want you to cry again so I will summarise our friendship with keywords: Innsbruck, Birmingham, Primark, Moscow mule, Project, Paris, I-have-a-certain-age, Scatman. Looking forward to add new keywords. Many more PIMMS contributed to making this project a chapter of my life: **Kos**, **Matteo**, **Jose**, **Ramon**, **Dušan**, **Andrea**, **Pritam**, **Raquel**, **Philip**, **Celia**, **Eva** and **Jackline**.

PhD, sometimes you have been tough on me, but luckily, I had amazing housemates where I could find shelter. **Tom**, thanks for your amazing food and gin tonics. While I was writing my PhD, you met my lovely **Vale M.**, got married and got an amazing daughter **Margherita** together. Vale, in the meanwhile I was finishing the PhD, you met my lovely housemate and got a baby (and got married). Thanks for letting me feel part of your family. **Polla**, I know I am noisy but you love my noise, thank you for standing me and knocking at my door every time I needed someone to speak with. **Camilo** you took her away from me but I am sure she is in good hands. **Gianky**, thanks for being my third brother and taking care of me. **Joao**, you probably were not annoyed by my noise because we are quite similar in this. Thanks for all the noise, the laughs and the advices you keep giving me. **Dafni**, even though you lived in the garage, you felt really close to me. My gratitude goes to **Kristina**, you are such as sweetheart, and also to **Louis** and **J.**, and the last one who entered the house **Nina**.

PhD, sometimes you were even more tough on me, but luckily, I had my friends or probably it is most appropriate to call them family. You are my heroes. Marvel or DC did not create yet a super hero with a power comparable to yours.

Carla from Avellino to Wageningen and back and return with love. **Dimi,** the first time we met you got a bit pissed at me but it is how love stories begin. Since then we became closer and closer. Thanks for supporting me every time and listening to all my crazy stories. Life has been a bit of a rollercoaster so far, but we rock it. Now I will have enough time to go whenever you want. Let's go to Paris with the boat! **Raffa**, you left but every time we are together it is like nothing changed, I am so happy that our friendship is alive despite the distance. **Edo**, daje e stra-ri-daje, I can't imagine my life

List of publications

Acierno, V., de Jonge, L., & van Ruth, S. M. (2020). Sniffing out cocoa bean traits that persist in chocolates by PTR-MS, ICP-MS and IR-MS. *Food Research International, 133*(July 2020).

Acierno, V., Liu, N., Alewijn, M., Stieger, M., & van Ruth, S. M. (2019). Which cocoa bean traits persist when eating chocolate? Real-time nosespace analysis by PTR-QiToF-MS. *Talanta, 195*, 676-682.

Acierno, V. Yener, S., Alewijn, M., Biasioli, F., & van Ruth, S. M. (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.

Acierno, V., Alewijn, M., Zomer, 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., Fasciani, G., Kiani, S., Caligiani, A., & van Ruth, S. M. (2019). PTR-QiToF-MS and HSI for the characterisation of fermented cocoa beans from different origins. *Food Chemistry, 289,* 591-602.

Boscaino, F., Acierno, V., Saggese, P., Cozzolino, R., Motta, C. M., & Sorrentino, A. (2014). Effectiveness of vacuum devices for home storage of rainbow trouts from game fishing lakes. *African Journal of Science and Research, 3*(2), 16-22.

Overview of completed training activities

Category A: Discipline specific activities	
PIMMS 1st Training Course: Hands on PTR-MS (2013). Ionicon, Innsbruck, Austria.	PIMMS 2nd Training Course (2013). University of Birmingham, Birmingham, UK.
Conference Sustainable-Food Summit (2013). Amsterdam, The Netherlands.	Workshop: Food Authenticity and Traceability (2013). Amsterdam, The Netherlands.
Poster - Conference Compositional Analysis of Lipids (2013). Ghent University, Ghent, Belgium.	Poster - Second International Congress on Cocoa coffee and tea (2013). Naples, Italy.
European Symposium on Advances in SIFT-MS (2014). Breda, The Netherlands.	Presentation - PIMMS Mid-term Review Meeting (2014). Ionicon, Innsbruck, Austria.
Training - PIMMS 3rd Annual Meeting (2014). University of Birmingham, Birmingham, UK.	World Cocoa Conference (2014). Amsterdam, The Netherlands.
Poster - Conference on Food Oral Processing (FOP2014): "Physics, Physiology and Psychology of Eating" (2014). Wageningen, The Netherlands.	Presentation - PIMMS 3rd ESR Training Summer School & Complementary Skills Training Workshop (2014). Juelich, Germany.
Poster - VLAG course "Advanced Food Analysis" (2015). VLAG, Wageningen, The Netherlands.	Poster/Presentation - Third International Congress on Cocoa Coffee and Tea (2015). Aveiro, Portugal.
Poster/Presentation - PIMMS 4th ESR Training Summer School (2015). Caen, France.	Presentation - 7th International PTR-MS Conference (2016). University Centre Obergurgl, Obergurgl, Austria.

Category B: General courses	
From PhD to valuable employee- When research meet industry (2013). VLAG, Wageningen, The Netherlands.	In-house PTR-MS training (2013). RIKILT, Wageningen, The Netherlands.
In-house multivariate statistics training (2013). RIKILT, Wageningen, The Netherlands.	ESR multivariate statistics course (2013). Fondazione Edmuch Mach, San Michele all'Adige, Italy.
VLAG PhD week (2013). VLAG, Baarlo, The Netherlands.	IASC-ERC summer school on CoDA (2014). Research Group on Statistical Method, Girona, Spain.
Course Chemiometric I (2014). Radboud Universiteit, Nijmegen, The Netherlands.	Writing for academic publication (2016), Wageningen University and Research, Wageningen, The Netherlands.

Category C: Optionals	
Preparation of research proposal (2013-2017). FQD, Wageningen, The Netherlands.	Expertise group/business unit meetings RIKILT (2013-2017). Research group, Wageningen, The Netherlands.
Colloquia (2013-2017). FQD, Wageningen, The Netherlands.	Project meetings with stakeholders (2013-2017). Research group, Wageningen, The Netherlands.
Coffee and Cocoa WUR research meeting (2013). Wageningen University and Research, Wageningen, The Netherlands.	Bachelor course - Chemometrics I (2014). Radboud Universiteit, Nijmegen, The Netherlands.
Committee - PhD trip (2016). Wageningen University and Research, Wageningen, The Netherlands.	

Teaching obligations
Mentoring undergraduate students (2013-2017). Wageningen University and Research, Wageningen, The Netherlands.
FQD-22306 Food Quality Analysis and Judgement (2016). Wageningen University and Research, Wageningen, the Netherlands.

About the author

Valentina Acierno was born in Avellino, Italy, on the 31st of July in 1985. She wanted to study archaeology but decided to study Marine Biology Productions, in which she obtained her Bachelor Degree in 2007. As a logical follow-up, she started her master in Marine Biology Productions of which she graduated cum laude in 2012. Her thesis was focused on the evaluation of the effects of storage environments on the quality of rainbow trout, *Oncorhynchus mykis*. Already during her master Valentina knew she wanted to do a PhD. As you would think, the topic of her PhD would be related to fish, but as we all know she decided to go in another direction and started a PhD on chocolate at Wageningen University and Research. As someone who does not like chocolate, she wrote quite a lot of articles about it. During her PhD she went for training courses to Birmingham, Innsbruck and Spain where she met some of her close friends. Doing a PhD did not only decrease her I-do-not-like-chocolate-perception, but it also made her a specialist in Proton Transfer Reaction–Mass Spectrometry. That's why now it's hard to imagine she switched from chocolate to spices for her job at EPOS in Nijkerk where she is working as a researcher in authenticity.

Acknowledgements of financial support

Cover design by Andrea Sessa