JACOBS UNIVERSITY

Comparison of chlorogenic acid and lipid profiles in green and roasted coffee beans

By

Sabur Olarotimi Badmos

A thesis submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy in Chemistry

Approved Dissertation Committee

Professor Dr. Nikolai Kuhnert
Professor of Chemistry, Jacobs University Bremen, Germany

Professor Dr. Gerd-Volker Röschenthaler
Professor of Chemistry, Jacobs University Bremen, Germany

Emeritus Professor Dr. Michael N. Clifford
Professor of Food Chemistry & Safety, University of Surrey
Guildford, United Kingdom

Date of Defense: **Wednesday 28th October, 2020**

Department of Life Sciences and Chemistry

Statutory Declaration

Family Name, Given/First Name	**BADMOS, Sabur Olarotimi**
Matriculation number	**20331354**
What kind of thesis are you submitting: Bachelor-, Master- or PhD-Thesis	**PhD Thesis**

English: Declaration of Authorship

I hereby declare that the thesis submitted was created and written solely by myself without an y external support. Any sources, direct or indirect, are marked as such. I am aware of the fact that the contents of the thesis in digital form may be revised with regard to usage of unauthori zed aid as well as whether the whole or parts of it may be identified as plagiarism. I do agree my work to be entered into a database for it to be compared with existing sources, where it wil l remain in order to enable further comparisons with future theses. This does not grant any rig hts of reproduction and usage, however.

The Thesis has been written independently and has not been submitted at any other University for the conferral of a PhD degree; neither has the thesis been previously published in full.

German: Erklärung der Autorenschaft (Urheberschaft)

Ich erkläre hiermit, dass die vorliegende Arbeit ohne fremde Hilfe ausschließlich von mir erst ellt und geschrieben worden ist. Jedwede verwendeten Quellen, direkter oder indirekter Art, si nd als solche kenntlich gemacht worden. Mir ist die Tatsache bewusst, dass der Inhalt der The sis in digitaler Form geprüft werden kann im Hinblick darauf, ob es sich ganz oder in Teilen um ein Plagiat handelt. Ich bin damit einverstanden, dass meine Arbeit in einer Datenbankein gegeben werden kann, um mit bereits bestehenden Quellen verglichen zu werden und dort auc hverbleibt, um mit zukünftigen Arbeiten verglichen werden zu können. Dies berechtigt jedoch nicht zur Verwendung oder Vervielfältigung.

Diese Arbeit wurde in der vorliegenden Form weder einer anderen Prüfungsbehörde vorgelegt noch wurde das Gesamtdokument bisher veröffentlicht.

5th October, 2020

……...

Date, Signature

“One looks back with appreciation to the brilliant teachers, but with gratitude to those who touched our human feelings”

“Education is a journey and not a destination”... Anonymous

Dedication

This research thesis is dedicated to my lovely wife and beloved children for their unflinching support, love, patience, sacrifice, and perseverance during the course of this historic journey…

Acknowledgement

All praises and adorations are due to Almighty Allah (SWT) for granting me the privilege to carry out this research successfully despite all the associated challenges along the path.

I would like to sincerely express my profound gratitude to my Ph.D advisor Professor Dr. Nikolai Kuhnert for his kindness, invaluable mentorship, fatherly support and encouragement throughout the period of my study. I have indeed benefited immensely from his knowledge and wealth of experience. I am also grateful to his wife Beatrix Tetzlaff for her kindness and her father Mr. Tetzlaff for providing some old coffee beans used in this research.

I would like to express my appreciation to Prof. Dr. Gerd-Volker Röschenthaler of Jacobs University Bremen, Germany and Emeritus Professor Dr. Michael N. Clifford, University of Surrey, United Kingdom for accepting to be part of my PhD thesis committee and the review.

I am grateful to Dr. Svenja Frischholz of Jacobs University Bremen for her kind assistance and support during my study. Equally, I acknowledge the financial support from Jacobs University Bremen and DAAD during the period.

Special appreciation to Dr. Roy D'Souza, Dr. Seung-Hun Lee and Dr. Martin Kangwa for their kind assistance and support with knowledge sharing through the years.

I am grateful to Anja Mueller and Britta Behrends for the technical assistance. Equally, same goes to my colleagues, the past and current members of Kuhnert research group for the knowledge sharing and support most importantly Dr. Roberto Megias, Dr. Abhinandan Shrestha, Dr. Mohammed Elsadig, Fariba Sabzi and all my research students particularly Maotian Fu. Likewise, I acknowledge the kind assistance and moral support from all my friends on Campus with special thanks to Kosisochukwu Nnoli and Febrina Desrianti.

Special thanks and appreciation to my German special family friends Hartmut Boeck, Rowsitha Boeck, Hermann Smidt and Doris Bendig for their love, kindness, unflinching support and for always been there for my family.

Furthermore, I appreciate Professor Olusegun George Ademowo, Professor Ganiyu Arinola, Professor Ahmed Adedeji and Dr. Misjudeen Raji for their kind assistance, good advice and moral support always. Also, special thanks to all my beloved brothers at home and away.

I would like to also express my heartfelt gratefulness to my siblings and in-laws for their love, kindness, motivation and moral support during the period most importantly Obafemi Badmos, Ismail Uthman, Hakeem Ayeni, Motunde Ogunjebe, Chris Saigbe and Jumoke Saigbe.

Without my parents Alhaji & Alhaja Badmos, parent-in-laws Alhaji & Alhaja Ayeni, there is no way I could have achieve this feat. It gives me joy and pleasure that you all witnessed this achievement. I thank and appreciate all your kindness, care, prayers, moral support and love.

Finally, I would like to express my profound appreciation to my wonderful and lovely wife Muslimah Badmos, and as well precious children Maryam, AbdulRahman, Aisha and Muhammad Badmos. Although, I have already dedicated this thesis to you all but I cannot do without mentioning that with your love, care, patience, understanding, perseverance and moral support we jointly accomplished this success together. I love you all!

Contents

List of Abbreviations

ABTS - 2,2'-azino-bis (3-ethylbenzothiazoline-6-sulphonic acid)

ACN – Acetonitrile

EtOH - Ethanol

ALA - alpha-linolenic acid (C18:3)

ANOVA – Analysis of variance

AR – Arabica

BIO – Biodynamic

CGA – Chlorogenic acid

CNS - Central nervous system

CONV – Conventional

CQA – Caffeoylquinic acid

DHA - docosahexaenoic acid (C22:6)

Di-CQA – Dicaffeoylquinic acid

DNA – Deoxyribonucleic acid

DPPH - 2,2-diphenyl-1-picrylhydrazyl

EB – Elution buffer

EPA - *Cis*-5,8,11,14,17-eicosapentaenoic acid (C20:5)

FAME - Fatty acid methyl esters

FFA –Free fatty acid

FID - Flame ionization detector

FQA – Ferruloylquinic acid

FRAP - Ferric ion reducing antioxidant power

GAE - Gallic acid equivalent

GC - Gas Chromatography

ICO – International Coffee Organization

ISO - International Organization for Standardization

IUPAC – International Union of Pure and Applied Chemistry

LC-ESI-MS – Liquid chromatography electrospray ionization mass spectrometry

LC-ESI-MS^n - Quadrupole ion-trap mass spectrometry

LC-MS - Liquid chromatography mass spectrometry

LDA – Least discriminant analysis

LDC - Least Developed Countries

LOD - Limit of detection

LOQ - Limit of quantification

MeOH - Methanol

MS-TOF – Mass spectrometry Time of Flight

NCA – National Coffee Association

ORG – Organic

OTA - Ochratoxin A

PCA – Principal component analysis

*p*CoQA - *p*-coumaroylquinic acids

PLS-DA – Partial least square discriminant analysis

POS - 1-palmitoyl-2-oleoyl-3-stearoyl-sn-glycerol

RB – Robusta

ROS - Reactive oxygen species

RRF - Relative response factor

RT – Room temperature

S.D – Standard deviation

SOL - 1-stearoyl-2-oleoyl-3-linoleoyl-glycerol

TAG – Triacylglycerol

Tandem-MS – Tandem mass spectrometry

TFA - *Trans* fatty acid

TMS – Tetramethylsilane

TPTZ - 2,4,6-Tris(2-pyridyl)-s-triazine

UHPLC –Ultra high performance liquid chromatography

UV-VIS – Ultraviolet visible light

16-*O*Me - 16-*O*-Methylcafestol

^{1}H-NMR – Proton Nuclear magnetic resonance

5-HMF - 5-hydroxylmethyl-furfural

ABSTRACT

Coffee production and consumption continuously increase despite the numerous barriers to its cultivation and global trade. The unique, refreshing and stimulating satisfactions associated with final coffee cup quality are among the reasons for the growing production and consumption rates. Coffee quality is a multidimensional feature, certainly influenced by the origin, plant variety, different forms of cultivation and post-harvest processing. Apparently, over 60% of the elements of coffee quality are determined by postharvest processing, which involves storage and roasting. This study aims to investigate the effects of long-term storage on some of the essential chemical constituents of the green and roasted coffee bean, and most importantly on the two main coffee varieties *Coffea arabica* and *Coffea canephora*.

Moreover, various dynamics such as environmental factors, duration of storage, coffee bean varieties and market conditions largely influence the optimal storage conditions. The viability and quality of preserving coffee beans in their active physiological system are highly affected by the different storage conditions. Environmental factors such as moisture, temperature, relative humidity and different air compositions (gases) are among the core dynamic forces of storage conditions.

The thesis investigation is sub-divided into seven chapters. Chapter **1** covers the introductory part, which encompasses the history and evolution of coffee, botanical classification and cultivation, economic importance, classes of compounds present in coffee, occurrence, dietary burden, biosynthesis and biological activities; and the aim of the project. Chapter **2** composes of profiling, identification and quantification of chlorogenic acids (CGAs) in green coffee; including the use of CGAs as a resourceful tool to discriminate between *Coffea arabica* and *Coffea canephora* by employing the use of nuclear magnetic resonance (NMR) and tandem mass spectrometry (MS/MS) for accurate profiling. From this study, LC-MS^n quantification of key CGAs in Arabica and Robusta coffee was achieved. Mono-CQAs showed no significant

variation among coffee varieties, while FQAs and Di-CQAs showed statistically significant variations if Robusta and Arabica beans were compared. Chapter **3** covers the use of CGAs characterisation in roasted coffee to classify coffee produced by different agricultural cultivation practices (organic, biodynamic and conventional) and from different geographical climatic regions using liquid chromatography coupled to mass spectrometry (LC-MS) systems and multivariate statistical analytical tools. From this investigation, it was observed that organic coffee contains less chlorogenic acids if compared to conventional coffee, and there was no significant difference in the coffees grown in different regions of Brazil. Chapter **4** comprises of the profiling and identification of triacylglycerol constituents with new compounds present in green coffee beans; and exploring the TAGs to discriminate Arabica from Robusta coffee based on MS/MS and multivariate statistical analysis. From this analysis, eight (8) new TAGs were identified including TAGs containing fatty acids with odd numbers of carbon atoms such as margaric (C17:0) and pentacosanoic (C25:0) acids. The identified TAGs coupled with multivariate analysis allowed the discrimination between Arabica coffee and coffees adulterated with Robusta coffees. Chapter **5** covers the multivariate discriminant analysis between *Coffea arabica* and C*offea canephora* green beans based on their fatty acids variations and continent of cultivation (geographical origin) by Gas Chromatography (GC). From this study part, discrimination among African, Asian, Central and South American green coffee beans on the basis of region of origin based on fatty acids analysis was achieved; while oleic, *cis*-5,8,11,14,17-eicosapentaenoic, eicosenoic, erucic, tetradecanoic, *Z*-10-pentadecenoic, and acids were observed to be higher in Robusta than Arabica coffees. Chapter **6** involves the investigation of ageing during storage, rate of degradation over time and stability of the chemical constituents in green and roasted coffee beans polyphenols and lipids by exploring all the analytical methods and tools used in Chapters 2 - 5, basically the LC-MS/MS, antioxidants estimation by ferric reducing antioxidant power (FRAP) assay, and the lipid constituents which are the triacylglycerol by LC-MS/MS and fatty acids by GC. From this investigation, Robusta

(87.76%) shows higher percentage degradation rate compared to Arabica (72.3%) in green coffee beans studied from year 2016 through 2020. While 3-CQA and 3,5-DiCQA were observed to be more depleted in two coffee varieties. Moreover, chapter **7** encompasses the inferences deducted from the study and conclusions.

Chapter 1

Introduction

1. INTRODUCTION

1.1 General overview of coffee and its chemical constituents

The interest in investigating coffee and its constituents has received lots of attention and expanded extensively within the last two to three decades. This is based on the increasing number of coffee consumers across the world. However, numerous studies have been published on the coffee bioactive constituents, the biological activities, economic gains, organoleptic properties, waste constituent's bioactivity, health benefits and the likes. However, this study intends to move beyond the aforementioned by investigating the differences in chemical profiles or metabolites of selected groups of samples including Arabica against Robusta, biodynamic against conventional agricultural practises, fresh against aged coffee beans.

1.2 The History and Evolution of Coffee

Until today, there is no specific record on how or when coffee was actually first discovered, nonetheless there are numerous folklores about its origin. Although history has it, that the wild coffee plant (*Coffea arabica*) is indigenous to Ethiopia where it was discovered around 850 A.D in Kaffa Mountain area of the country. However, due to the beans required and favourable topographical conditions, coffee is currently cultivated mainly in South America, Africa and Asia. According to some studies, coffee heritage can be traced back to centuries ago to the ancient coffee forests on the Ethiopian plateau in Africa; where it was reported that a shepherd known as Kaldi first discovered the potential of these beloved beans (Smith, 1985). From his observations, the goats usually become more excited and energetic after consuming berries from a specific tree such that they hardly slept at night. Thereafter, Kaldi reported his observations to the local monarch who prepared a drink from the berries to ascertain Kaldi's claim and the monarch found that the resulting drink kept him alert and awake for long hours of the evenings. From this, the knowledge of the energizing berries began to spread across the

cities and thus the globe. Later, coffee was then first exported out of Ethiopia to Yemen by some Somali merchants in the 15th century (Smith; 1985, Weinberg and Bealer; 2001).

Subsequently, the cultivation of coffee and trade commenced in the Arabian Peninsula in the Yemeni district of Arabia and spread to Makkah, Persia, Syria, Turkey and Egypt by the 16th century, then later to the Balkans (Southeast Europe), Italy and the rest of Europe, as well as Southeast Asia and followed by America (Meyers; 2005, NCA; 2020).

1.3 Botanical Classification

Coffee traces its origin to a genus of plants known as *Coffea*. It belongs to the plant family of *Rubiaceae*. Its flowers are visited for pollination by *Heliconius ismenius*, a species of butterflies in the family brush-footed butterflies. Within the coffee genus, there are over 500 genera and 6,000 species of tropical trees and shrubs. Also, there are nearly 100 coffee species in the world with the most commercialized being *Coffea arabica* (Arabica) and *Coffea canephora* (Robusta) coffee. Moreover, Arabica has the higher commercial value due to its aromatic superiority. Among these species, there are a variety of cultivars that can be grown and spread throughout the world, such as the Typica and Bourbon varieties (Arabica), but there are also some linked to their producing countries, such as Novo Mundo and Caturra (Brazil), Jimma and Harar (Ethiopia), and Villa Sarchi from Costa Rica (Toci et al; 2015 and NCA, 2020). In addition, *Coffea arabica spp*. varieties range from *Bourbon, Typica, Caturra, Mundo Novo, Tico, San Ramon* and *Jamaican Blue Mountain* among others.

Apart from *Coffea arabica* that was first discovered in Ethiopia, two other species of coffee, *Coffea liberica* and *Coffea robusta* (also referred to as *Coffea canephora*), and a cross-breed between *Coffea arabica* and *Coffea liberica*, were later also discovered in Africa. *Coffea robusta* is known for its resistance to disease, hence its specific appellation with the term Robusta derived from the word ‘robust’ (Smith, 1985).

Mainly, *Coffea arabica* and *Coffea canephora* var. *robusta*, commonly referred to as Arabica and Robusta, are the two major coffee varieties of highest agronomical importance, with the first accounting for approximately 70% of the world production (Alves et al, 2009).

1.4 Cultivation

Several steps are involved in transforming the coffee seed to the tasty drink. This usually takes a long journey starting from the point of cultivation, through beans transformation into cup beverage. Most importantly among these steps are cultivation, harvesting, cherries processing, drying, milling, shipping, cupping, roasting, grinding and brewing processes.

Coffee trees are mostly grown in the tropics by propagation and usually carried out by using seed. Moreover, budding, grafting and cutting methods are also employed. A more successful method is to raise seedlings in sheltered nurseries. At 6 - 12 months, seedlings are taken to fields, hardened and planted on contoured fields 2 - 3 m apart in 3 - 5 m rows. Holes are prepared by 40 x 40 x 40 cm and 4 seedlings placed in each. Plants may be shaded by taller trees or left unshaded. In addition, coffee is frequently intercropped with food crops, such as corn, beans or rice, during the first few years. Clean weed control is required throughout the entire season. Pruning is a common practice in some districts (Reed, 1976). The tree is pruned short to conserve energy and aid harvesting. The plant can grow up to 30 feet (9 meters) high and nicely covered with green, waxy leaves growing opposite each other in pairs. Coffee cherries usually sprout out along the branches. Since it breeds in an unceasing cycle, it is not unusual to see flowers, green and ripe fruit concurrently on a single coffee tree (Smith, 1985).

With regards to pollination system, *Coffea canephora* is self-sterile diploid species ($2n = 2x = 22$) and is reported to be primarily wind pollinated, but it is also expected to benefit from bees for effective outcrossing and fruit set. While *Coffea arabica* is naturally self-fertile tetraploid species ($2n = 4x = 44$). Studies have shown that cross pollination by bees causes a significant

increase in fruit set of not only the self-sterile, but also the self-fertile coffee species. The applied inference is that coffee yield may be improved by managing the fields properly for increased flower visits by bees (Klein et al, 2003, Geleta et al, 2012).

Typically, the trees grow best in rich soil, with mild temperatures, frequent rain and shaded sun, but different altitude depending on the variety. Its favourite locality is hill-sides at an elevation of about 1000 to 3000 feet above the sea level, though optimal altitude varies with proximity to the equator; hence its wide diffusion, extensive cultivation and large consumption. Mostly, it takes almost a year for a cherry to fully mature after the first flowering and roughly 5 years of growth to attain complete fruit production. The coffee plants can live up to 100 years and generally most productive between the ages of 7 and 20. Besides, appropriate care and good maintenance increases the trees output over the years, though depending on the variety. The average coffee tree produces 10 pounds of coffee cherry or 2 pounds of green beans annually (Van Der Vossen, 2005).

Specifically, the Arabica shrub typically grows between 2.5 - 4.5 m in height and requires temperature range between 15 °C and 24 °C with an annual rainfall of about 1200 - 2200 mm/year, while Robusta grows slightly taller at 4.5 - 6.5 m with a warmer temperature range of 18 °C – 36 °C and slightly more rainfall (2200 - 3000 mm/year) than Arabica. In terms of yield, Arabica produces less coffee per hectare than Robusta, making the cost of growing Arabica much higher. Arabica trees are pricey to nurture because the ideal terrain tends to be steep and access is quite difficult. Similarly, because Arabica trees are more disease-prone than Robusta, they require additional care and attention. C*offea arabica* trees produce a fine, mild and aromatic coffee. Its beans are flatter, more elongated than Robusta and with reduced caffeine quantity (FAO, 2016).

1.5 Coffee processing

It usually takes about 3 - 4 years on the average for the cultivated coffee trees to produce the fruit. The fruit is typically referred to as coffee cherry. A deep red coloured cherry signifies that the fruit is ripe and ready for harvest. Mostly the fruits are strip-picked from the tree branches by hand or machine (mechanised harvesting method). In addition, some farms usually employ the use of selective-picking method, which is mostly done by hand to selectively harvest the finer Arabica beans. Though, the method is labour-intensive and costly. Thereafter, the harvested coffee cherries are then transported to the processing plant.

Immediately after the coffee cherries arrive at the processing plant, the processes must commence to prevent fruit spoilage. Coffee is typically processed in two ways, either by dry (ancient method of processing coffee employed mostly in countries with limited water resources) or wet method.

The dry method involves the spreading of freshly picked cherries on a large verse of land with sun exposure to dry the cherries. To avoid rotten or damaged cherries, the cherries are usually raked and turn-over periodically, covered at night and during raining to prevent any contact with water. These steps are carried out for several weeks until the cherries moisture content is reduced to about 11% before storage (NCA, 2019). Subsequently, the dried flesh ie removed from the coffee cherry to obtain the bean.

In the wet method, pulp is removed from the cherry immediately after harvesting. With this system, the bean is dried only with the parchment skin. This is carried out by allowing the cherries to pass through a pulping machine, which separates the pulp and the skin from the coffee bean. Furthermore, the beans are passed via water channels with series of rotating drums thereby separating the beans based on weight and size. After sorting into different sizes, the beans are moved to a large fermentation tanks filled with water for about 48 hr to remove the

slickly mucilage outer-layer (parenchyma) attached to the parchment. During this process, naturally occurring enzymes will dissolve the outer-layer. This completes the fermentation procedure and beans are then rinsed via clean water channels and dried accordingly thereafter. Then the pulped and fermented beans are further sun-dried or machine-dried to reduce the moisture content to about 11% before storage. The dried beans are otherwise referred to as parchment coffee, stored in jute or sisal bags pending the time for exportation.

Before the coffee bean is exported, a milling procedure is carried out on the stored parchment coffee. The parchment coffee is treated with additional steps, which involve the removal of the exocarp, mesocarp and endocarp (entire dried husk) by the hulling machine. Furthermore, via polishing process, the silver skin on the beans are removed, followed by grading and sorting by size and weight using an air jet to separate light beans from the heavy ones. Similarly, at this stage the defective beans (undesirable colour or size, highly fermented, insect-spoiled, unshelled beans) are carefully removed to have the premium and high-quality coffee beans before the beans are exported. Thereafter, the milled beans otherwise known as green coffee beans are weighed into bags, loaded into shipping containers and shipped to different locations across the world (NCA, 2019).

Shortly after the bags of green coffee beans are delivered to the different coffee processing plants, the cupping process begins to evaluates the beans overall visual quality. This is carried out by collecting coffee samples from different batches and beans varieties, roasted and tasted daily. The cupper evaluates the brew to experience its flavour and aroma. This is a vital phase in determining the coffee beans grade and quality. For the coffee tasting, the coffee cupper sips a spoonful with a rapid breath. This is done with the aim of spraying the coffee uniformly over the cupper's tongue taste buds and weighs it on the tongue before sputtering it out.

After the cupping procedure, green beans are roasted at different temperatures and roasting duration to have beans with different degrees of flavour, aroma and taste such as light-roast,

deep-roast, expresso and the likes. This roasting step is usually carried out in the countries that import the coffee to ensure that freshly roasted beans are speedily delivered to the consumers. The last two coffee processing procedures involve grinding and brewing, which are mostly determined by the consumers based on their cup of coffee choice (NCA, 2019).

1.6 Economic importance

Coffee is among the most traded agricultural commodities in the world (Benitez et al, 2019). Coffee is of great economic importance to developing countries, including many Least Developed Countries (LDCs), and of considerable social impact in consuming countries. The annual export earnings typically exceed US$10 billion and it accounts for over three quarters of the total export earnings in some LDCs. It is estimated that approximately 125 million people around the equator axes largely depend on coffee for their means of livelihoods (FAO, 2019), with over 10.4 million tons which equates to 170 million bags (1 bag ~ 60 kg) are produced in over 70 countries, providing a livelihood for people both in the developing and developed countries annually. Among the consumers, coffee is a universal and popular beverage, with over US$50 billion in retail sales per year (ICO, 2019). In the twelve months ending June 2020, exports of Arabica equalled 79 million bags compared to 83 million bags in 2019; while Robusta exports corresponded to 49.35 million bags compared to 48.89 million bags (ICO, 2020). Invariably, Arabica accounts for nearly 75% of the world coffee production, while Robusta predominantly represents the remaining fraction (Mondego et al., 2011).

Besides consumption as a beverage, coffee is also commonly used as flavour in ice-cream, pastries, candies and liqueurs production to produce different varieties likewise distinct tastes for the consumers. Similarly, it is used as source of caffeine; dried ripe seeds are used as stimulant, nervine and diuretic on the central nervous system, kidneys, heart and muscles. Indonesians and Malaysians usually prepare infusions from its dried leaves (Duke, 1983).

Invariably, growing coffee demand provides employment and is an important monetary source in rural areas. Apart from this direct impact on livelihoods of farming families, coffee cultivation offers several other advantages. The cultivation of coffee helps in reducing soil erosion, also a useful carbon sink (particularly with plantations grown under shade trees), aids good water-shed management, helps to maintain a degree of biodiversity, predominantly in systems with mixed cropping on small family farms, and can provide a good habitat for many migrant birds and other animals. In addition, coffee can be stored and effectively used as savings for several months after drying. This is a common practice in many coffee producing countries (ICO, 2006).

1.7 Classes of compounds present in green coffee

Coffee contains several classes of chemical compounds. Carbohydrates, proteins, lipids, phenolics, organic acids, caffeine, trigonelline, water, fibre and few minerals constitute the notable classes of compounds present in the non-volatile fraction of coffee (Table 1.1). Most of these compounds have extensively been investigated in details over the last three decades. Although most constituents of *Coffea arabica* are also present in *Coffea canephora*, their relative proportions can differ considerably. Additionally, *Coffea canephora* contains a few more secondary metabolites compared to Arabica.

Furthermore, numerous studies have reported differences between the two coffee varieties with regards to their chemical constituents, explicitly chlorogenic acids (Badmos et al, 2019; Bicchi, Binello, Pellegrino, & Vanni, 1995), volatile compounds (Dirinck, Van Leuven & Dirinck, 2001), trigonelline and caffeine (Casal, Oliveira, Alves & Ferreira, 2000), amino acids (Casal, Alves, Mendes, Oliveira, & Ferreira, 2003), biogenic amines (Casal et al., 2004), triacylglycerols (Badmos et al, 2020), sterols (Cizkova et al, 2007; Carrera, León-Camacho, Pablos & González, 1998), diterpenes (Alves et al, 2003; Martín et al, 2001; Speer and Kölling-Speer, 2001; Shievano et al, 2014) and minerals (Martín, Pablos & González, 1999). Among

all the listed compounds present in green coffee, chlorogenic acids, caffeine, trigonelline, soluble fibre, and diterpenes from the lipid fraction are the most bioactive, and they notably contribute significantly to the beverage pleasant taste, flavour and aroma after roasting (Farah, 2012). However, naturally the progressive decline in the beans chemical constituents' physiological status usually leads to ageing.

Differentiating the two coffees based on geographical origins was also studied (Bicchi et al, 1995; Dirinck et al, 2001), although it is quite challenging, typically due to the influence of the beans' chemical constituents, which is not limited to climatic and agronomical practices, but likewise by the postharvest processing method (wet or dry), storage, distribution conditions and different roasting procedures. In the past, some studies reported a tentative discrimination between different postharvest processing technologies (Casal et al, 2003; Casal et al, 2004). Table 1.1 represents the different coffee main chemical constituents with their varied concentrations according to the two main coffee varieties.

Table 1.1 Chemical composition of green *Coffea arabica* and *Coffea canephora* beans.

	Concentration (g/100 g)*	
Component	***Coffea arabica***	***Coffea canephora***
Carbohydrates/fibre		
Sucrose	6.0-9.0	0.9-4.0
Reducing sugars	0.1	0.4
Polysaccharides	34-44	48-55
Lignin	3.0	3.0
Pectin	2.0	2.0
Nitrogenous compounds		
Proteins/peptides	10.0-11.0	11.0-15.0
Free amino acids	0.5	0.8-1.0
Caffeine	0.9-1.3	1.5-2.5
Trigonelline	0.6-2.0	0.6-0.7
Lipids		
Coffee oil (Triacylglycerols with unsaponifiables, sterols/tocopherols)	15.0-17.0	7.0-10.0
Diterpenes (free and esterified)	0.5-1.2	0.2-0.8
Minerals	3.0-4.2	4.4-4.5
Acids and esters		
Chlorogenic acids	4.1-7.9	6.1-11.3
Aliphatic acids	1.0	1.0
Quinic acid	0.4	0.4

* Content varies according to agricultural practices, climate, soil composition, cultivar (species) and methods of analysis (Farah, 2012).

1.7.1 Carbohydrates in green coffee

Carbohydrates are the major constituents of coffee and may account for about half of the green coffee bean dry weight (Trugo, 1985). The carbohydrates constituents of the coffee can be categorised into polysaccharides, oligosaccharides, disaccharides and monosaccharides. The insoluble and soluble polysaccharides contribute about 44% and 47% of dry matter in Arabica and Robusta green coffees respectively. About 9% in Arabica and 4% in Robusta of the coffee bean dry weight is sucrose; this higher content compared to Robusta contributes to the reason for the superior taste, pleasant aroma and flavour. Besides, simple sugars the likes of glucose, fructose, arabinose, mannose, and rhamnose; and some oligosaccharides like stachyose and raffinose have also been reported to be present in green coffee bean (Kolling-Speer and Speer, 2005). Besides, *Coffea canephora* (0.4%) contains more quantities of reducing sugars than Arabica (0.1%) coffee (Farah, 2012).

Moreover, sugars are the key precursors of Maillard reaction that occurs during coffee roasting as well as caramelization, which are quite essential in determining aroma and dark coffee colour formation (Zhang and Zhang, 2007). Arabinoglycans and galactoglycans are prominent soluble polysaccharides in coffee. Likewise, brewed coffee contains some amount of soluble fibres, which play key roles as substrate for probiotic microorganisms in human gut microbiota, because human body physiological enzymes lack the capacity to digest some insoluble polysaccharides such as cellulose, pectin, lignin and other indigestible materials, mostly considered as dietary fibres (Padayachee et al, 2015). The plant materials (complex carbohydrates) in diets are resistant to enzymatic digestion in humans, such as cellulose, hemicellulose, pectin, gums, mucilages and lignin etc. Besides coffee, dietary fibres are naturally present in other food sources such as cereals, fruits, vegetables, roots, legumes, apple-skin and nuts (Brglez et al, 2016).

1.7.2 Proteins in green coffee bean

Proteins, peptides and free amino acids constitute approximately 10% - 16% of the green coffee bean dry weight. The most abundant protein in the coffee bean is the 11S globular storage protein (Acuna et al, 1999). The total proteins and other nitrogenous constituents of coffee are observed to be marginally higher in Robusta (19%) than Arabica (13%). They act as precursors for volatile compounds formation such as furans, pyridines, pyrazines, pyrroles, aldehydes and melanoidins (Farah, 2012). These main nitrogenous food sources are essentially responsible for the coffee flavour, aroma and colour since they are actively involved in Maillard reaction. The free amino acids are largely degraded during roasting, which ultimately results in the aromatic characteristics noticeable in roasted coffee.

Generally, coffee is not a rich source of proteins because it lacks some of the basic essential amino acids required for normal human physiological functions. However, proteins bound amino acids have been observed to be more stable in roasting conditions than being in free-state as rightly reported in the study involving the formation of aroma compounds and metal-chelating agents in coffee brew (Macrae, 1985).

1.7.3 Lipids constituents of green coffee bean

Lipids constitute a larger percentage of the coffee dry weight, while the composition in the two main coffee varieties varied significantly with Arabica (15 - 18% d.m) and Robusta (8 - 10% d.m). The lipid fraction of coffee is composed mainly of triacylglycerols, sterols and tocopherols. Coffee triacylglycerol major constituents are fatty acids in proportions similar to those found in common edible vegetable oils. Moreover, coffee contains significant amount of diterpenes, identified to possess antimicrobial and anti-inflammatory properties such as cafestol, 16-*O*-Methylcafestol and kahweol, which influence the coffee cholesterol levels (Schievano et al, 2014). It is quite pertinent to note that 16-*O*-methylcafestol is present mainly

in Robusta compared to Arabica (Speer and Kölling-Speer, 2001), meaning this compound is a biomarker that allows for distinguishing between the two coffee types. However, recent investigations have demonstrated that some Arabica coffees also contain significant amounts of this biomarker (Gunning et al, 2018). This in effect means that the biomarker is not 100% precise and reliable to discriminate between Arabica and Robusta. Hence, there is need to develop a more reliable analytical method to differentiate the two-coffee species accurately. Further details on this will be discussed extensively in the subsequent chapters of this study on coffee lipids.

The relatively large unsaponifiable fraction is rich in diterpenes of the Kaurane family, mainly cafestol, kahweol and 16-*O*-methylcafestol (Speer & Kölling-Speer, 2006), all represented in Figure 1.1 and employed as biomarkers in lipids. Most of the lipids, the coffee oil, are located in the endosperm of green coffee beans only a small amount, the coffee wax is located on the bean outer layer (Wilson, Petracco, & Illy, 1997); and about 75% are triacylglycerol (Iriondo-Dehond et al, 2019; Jham, Nikolova-Damyavova, Viera, Natalino, & Rodrigues, 2003; Toci, Neto, Torres, & Farah, 2013).

Triacylglycerols make up over 95% of lipids present in diets and are generally found in fried foods, vegetable oil, butter, milk, fish, cheese and meats. Naturally occurring TAGs are mostly found in various foods such as avocados, olives, beans, nuts and cereals (Albuquerque et al, 2020). It consists of a glycerol molecule that is esterified with three fatty acids. *Coffea arabica* contains about 15 - 18% lipids, while *Coffea canephora* usually contains less than 10% (Speer & Kölling-Speer, 2006). Characteristically, linoleic (47.7%), palmitic (33.3%), stearic (7.3%) and oleic (6.6%) acids are the most abundant fatty acids present in the green coffee bean. Details of the fatty acids constituents of coffee are represented in Table 1.2 with their common structures provided in Figure 1.2.

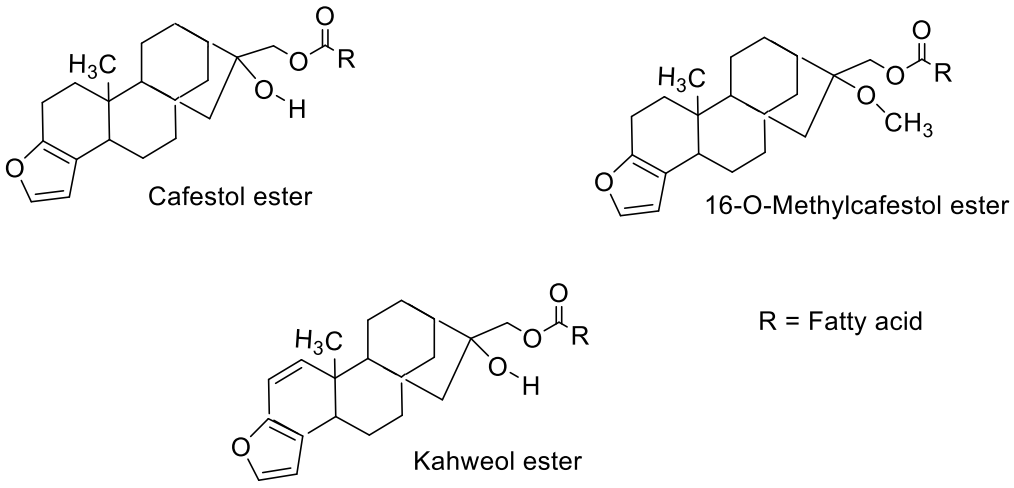


Figure 1.1 Structures of diterpenes in coffee

The genome of the coffee bean has recently been sequenced (Huang et al, 2020; Samson et al, 2007; Cheng et al, 2018; Tran et al, 2016; Acuna et al, 1999). The *Coffea arabica* has lesser genetic diversity and results in a higher quality beverage than *Coffea canephora* (Anthony et al, 2002), and thus most of this beverage consumers prefer the former to the latter due to its superior taste, aroma and flavour. Coffee lipid constituents essentially determine the drink's cup-quality in terms of taste, aroma and flavor. Biologically, lipids serve as one of the basic food constituents, essential in human diet for numerous roles (Lelyana, 2017). Also, lipid plays a crucial role in coffee crema formation (Folmer, Blank, & Hofmann, 2017). Moreover, this essential factor significantly contributes to consumers' choice of preferring Arabica to Robusta, invariably makes Arabica to be superior in terms of both demand and cost price (to be discussed extensively in subsequent chapter on lipids).

1.7.3.2 Triacylglycerol identification and nomenclature

For accurate identification of each of the TAGs, Tandem-MS in positive ion mode will be employed to investigate the TAGs fragmentation pattern, which will provide an unambiguous identification of the detailed TAGs constituents. The alphabetic letter 'M' will be used to represent the mass of an uncharged molecule, with the gain of a proton to create a cation denoted

by $[M+H]^+$. Integral masses or element symbols will be included within the brackets to indicate various other losses and gains, for example $[M+H-18]^+$ to indicate the loss of ammonium or $[M+K]^+$ to indicate the gain of a potassium ion. The quantity actually measured by a mass spectrometer will be denoted by the mass-to-charge ratio (*m/z*). The unsaturated fatty acids (UFA) and essential fatty acids (EFA) have been reported to have preference for *sn*-2- position, while saturated fatty acids (SFA) mostly preferred the *sn*-1- and the *sn*-3- positions (Cossignani et al, 2016). In addition, it is known that the regio-specificity of fatty acids in TAG is characteristic for the native oils and fats. Besides, it has been reported that UFA, most especially linoleic acid was preferably esterified with the secondary hydroxyl position in the glycerol of coffee TAG. As a result, it has been observed that fragmentation typically occurs at the MS^2 step preferentially from the *sn*-2- position (Folstar, 1985).

Figure 1.1B depicts the TAG fragmentation scheme. In identifying a specific TAG, the main TAG peak will have a mass denoted as *m/z* in the first window spectrum known as MS. This *m/z* will include the adduct which could be ammonium or potassium or sodium ion used as precursor ion. The next step would involve the fragmentation of the precursor ion into smaller fragments to be seen in MS^2. From this, two separate peaks would be observed in MS^2 with different masses. The first peak in the MS^2 would be *m/z* without the adduct ion $[M+H-18]^+$ and the second peak would be ($[M+H-18]^+$ - neutral loss). The neutral loss is typically a fatty acid. This means that the precursor ion has lost one fatty acid to produce a diacylglycerol (DAG) in MS^2. The mass of the neutral loss allows assignment of the first fatty acid. The first neutral loss occurs typically from *sn-1-* or *sn-3-* position. To corroborate this, previous studies revealed that the saturated acyl groups of fatty acid distribution occur exclusively most often in the *sn-1-* and *sn-3-* positions of the glycerol (Takagi & Ando, 1995; Smith, 2002). Thereafter, the DAG peak in MS^2 would further used as precursor ion and fragmented into two fatty acids, which would be seen in MS^3 spectrum window along with a monoglyceride at times.

Subsequently, the result would then be confirmed with available data on Lipid-MAPS. Further details are provided in section 3.2 of chapter 4 on lipid analysis.

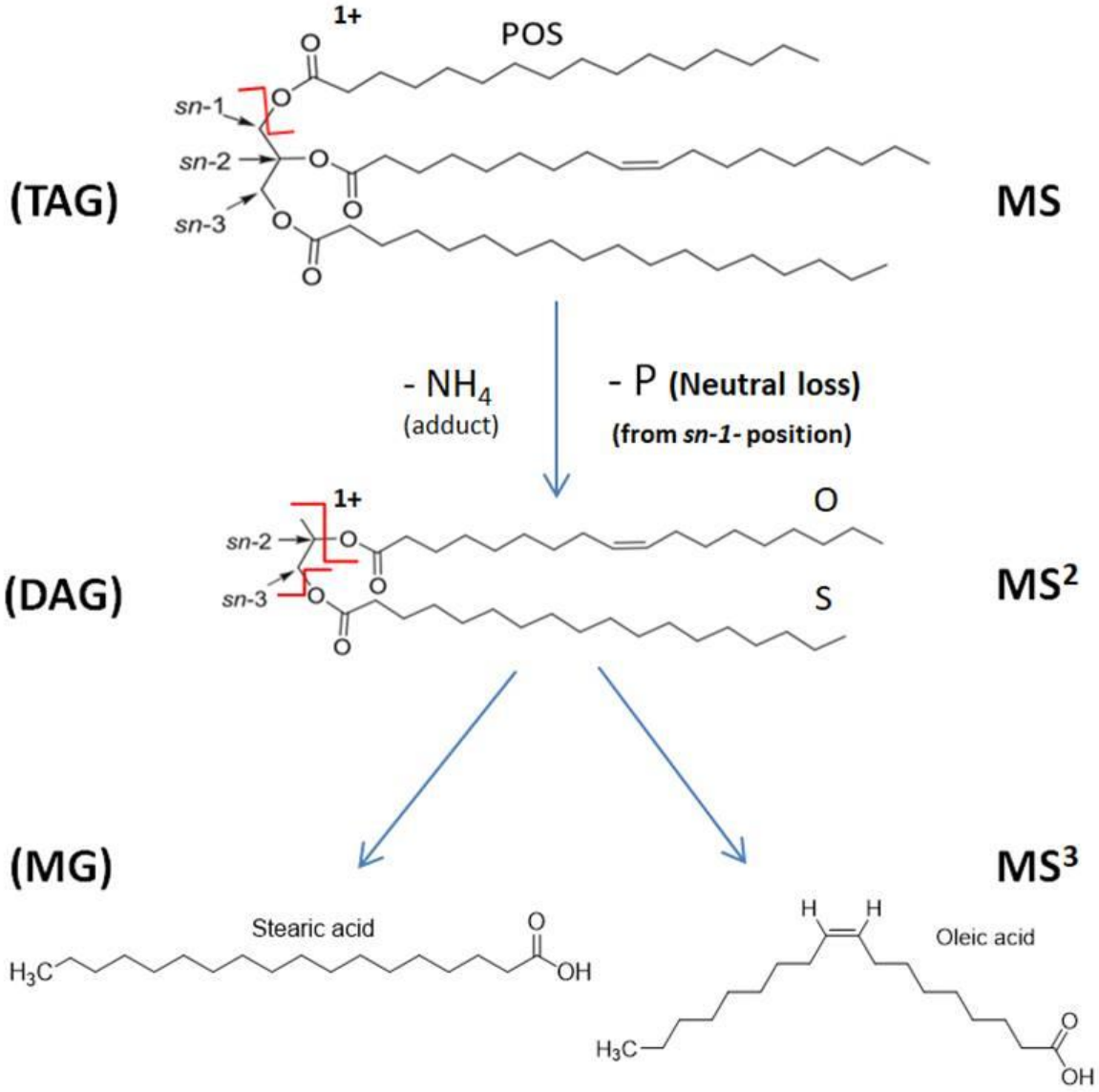


Figure 1.1B Scheme of a TAG (POS) fragmentation and identification pattern.

For the TAGs nomenclature, the typical abbreviation of three (3) fatty acid substituents was employed by using a single capital letter code to represent each fatty acid. For instance, SLL indicates 1-stearoyl-2-linoleoyl-3-linoleoyl-glycerol and POS signifies 1-palmitoyl-2-oleoyl-3-stearoyl-sn-glycerol (Fauconnot et al, 2004). The structure of POS (TAG) is represented in Figures 1.1B and 1.2; and detailed analysis of other TAGs are provided in the lipids chapter 4. However, majority of the fatty acids and/or any universal fatty acid have been reported to be

mostly less abundant in nature (Lísa & Holčapek, 2008), hence the current challenges with TAGs commercial standards unavailability.

1.7.3.3 Fatty acids

The esterification of fatty acids to fatty acid methyl esters is performed using an alkylation derivatization reagent. Methyl esters offer excellent stability, and provide quick and quantitative samples for GC analysis. GC is ideally suited for analysis of thermally stable volatile substances. Fatty acids are analysed as fatty acid methyl esters (FAME) because in their free and underivatized states, they are difficult to analyse due to their high polar compound properties, which tend to form hydrogen bonds, hence leading to adsorption challenges. Thus, reducing their polarity would make it easier for them to be analysed. The polar carboxyl functional groups must first be neutralized to distinguish between the slight differences exhibited by unsaturated fatty acids; this then allows the compounds proper separation by boiling point elution, and also by degree of unsaturation, position of unsaturation, and likewise the *Cis* vs. *Trans* configuration of unsaturation.

FAME procedure (method details provided in chapter 5) is commonly employed to derivatize the coffee oil samples before it is analysed on GC using fatty acids standards such as palmitic, linoleic, stearic and oleic acids, derivatized under similar conditions corresponding to the samples and analysed using gas chromatography (FAME method adopted from AOCS, 1989 and Firestone, 1995). In the FAME method, TAGs are hydrolysed under acidic conditions and converted into their volatile methyl esters. These are then separated by GC, identified and quantified using an authentic reference standard mixtures. Further details on the procedure are provided in section 2.5 of chapter 5 of this thesis.

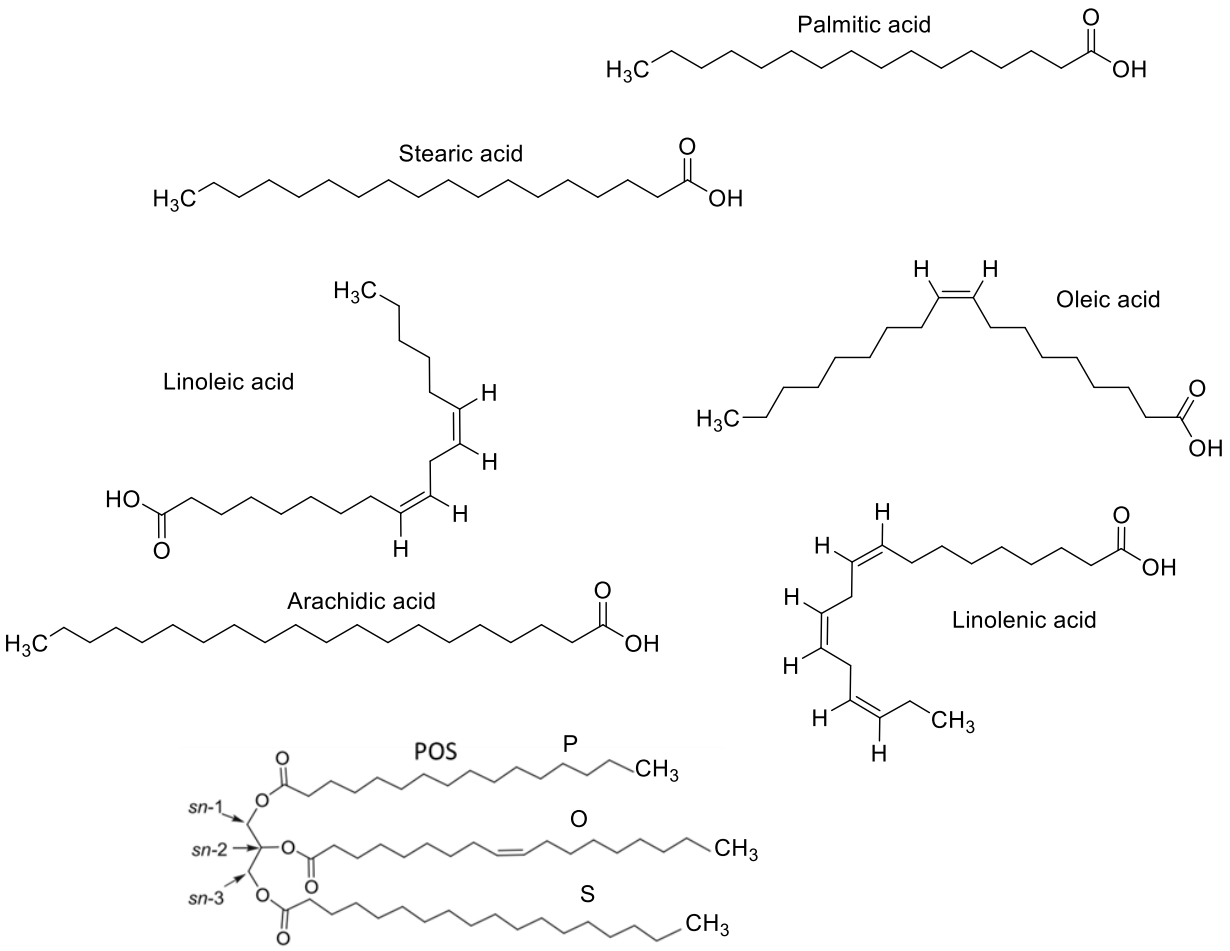


Figure 1.2 Structures of common fatty acids and a triacylglycerol in coffee

Table 1.2 Some of the fatty acid's constituents of coffee.

	RT (min)	Compound Name	Common Names
1	2.89	Methyl butyrate	Butyric acid
2	4.59	Methyl hexanoate	Caproic acid
3	6.35	Methyl octanoate	Octanoic acid (Caprylic acid)
4	7.96	Methyl decanoate	Decanoic acid
5	8.69	Methyl undecanoate	Undecanoic acid
6	9.41	Methyl laurate	Lauric acid
7	10.18	Methyl tridecanoate	Tridecanoic acid
8	11.07	Methyl myristate	Tetradecanoic acid
9	11.50	Methyl myristoleate	Myristoleic acid (C12:1)
10	12.12	Methyl pentadecanoate	Pentadecanoic acid
11	12.64	Methyl *cis*-10-pentadecenoate	(*Z*)-10-pentadecenoic acid
12	13.37	Methyl palmitate	Palmitic acid
13	13.82	Methyl palmitoleate	(9*Z*)-hexadecenoic acid (Palmitoleic acid)
14	14.85	Methyl heptadecanoate	Heptadecanoic acid (Margaric acid)
15	15.39	*Cis*-10-heptadecenoic acid methyl ester	*Cis*-10-heptadecanoic acid (Margaroleic acid; C17:1)
16	16.58	Methyl stearate	Stearic acid

17	16.87	*Trans*-9-elaidic acid methyl ester	Elaidic acid (C18:1)
18	17.06	*Cis*-9-oleic acid methyl ester	Oleic acid (C18:1)
19	17.55	Methyl linolelaidate	Linoleic acid (C18:2)
20	18.05	Methyl linoleate	Linoleic acid (C18:2)
21	18.66	Methyl linolenate	Linolenic acid (C18:3)
22	19.36	Methyl γ-linolenate	*Cis, cis, cis*-9,12,15-octadecatrienoic acid (C18:3)
23	20.70	Methyl arachidate	Arachidic acid
24	21.29	Methyl *cis*-11-eicosenoate	Eicosenoic acid (C20:1)
25	22.51	*Cis*-11,14-eicosadienoic acid	*Cis*-11,14-eicosadienoic acid (C20:2)
26	23.02	Methyl heneicosanoate	Heneicosylic acid (C21:0)
27	23.21	*Cis*-8,11,14-eicosatrienoic acid	*Cis*-8,11,14-eicosatrienoic acid (Dihomo-gamma-linolenic acid) (C20:3)
28	23.65	*Cis*-5,8,11,14-eicosatetraenoic acid	Arachidonic acid (C20:4)
29	24.06	*Cis*-11,14,17-eicosatrienoic acid	*Cis*-11,14,17-eicosatrienoic acid (C20:3)
30	25.24	*Cis*-5,8,11,14,17-eicosapentaenoic acid methyl ester	*Cis*-5,8,11,14,17-eicosapentaenoic acid (C20:5)
31	25.46	Methyl behenate	Behenic acid (C22:0)
32	26.14	Methyl erucate	Erucic acid (C22:1)
33	27.51	*Cis*-13,16-docosadienoic acid methyl ester	*Cis*-13,16-docosadienoic acid (C22:2)
34	27.96	Methyl tricosanoate	Tricosanoic acid (C23:0)
35	30.62	Methyl lignocerate	Lignoceric acid (C24:0)
36	31.22	Methyl nervonate	Nervonic acid (C24:1)
37	31.49	*Cis*-4,7,10,13,16,19-docosahexaenoic acid methyl ester	Docosahexaenoic acid (C22:6)

1.7.3.3 Sensory properties of lipids

Many of the significant flavour compounds from plants sources result from the enzymatic degradation of unsaturated fatty acids. The aerobic lipoxygenase cycle converts fatty acids containing a *cis*, *cis*-1,4-pentadiene system into hydroperoxy-octadecadienoic acids by regio- and enantioselective chemical reactions. Nevertheless, lipid is responsible for the rancid properties in fats, oils and lipid-containing foods. Nonetheless, they contribute immensely to the desirable flavours derived from vegetables like tomatoes, cucumbers, mushrooms and peas in addition to many deep-fat fried diets like fried potatoes and chicken (German, 1999; Zhang et al, 2015; Frankel, 1991).

Naturally, lipids contribute to the foods flavor and texture together. The role of lipids in flavour development is due to volatile oxidation products along with the taste of short-chain free fatty acids, and since the micro-structure of lipid emulsions promotes flavour release (Khrisanapant et al, 2019). Moreover, flavour, taste, aroma and texture are merged in the orbito-frontal cortex

in the human's brain to increase the general perception associated with a specific food and or its different degrees of attractiveness (Frankel, 1984; Small et al, 2007).

Hydroperoxides of unsaturated fatty acids formed by autoxidation are very unstable and its break down usually leads the formation of a variety of volatile flavour compounds as well as nonvolatile products. Among the volatiles released via the breakdown of the alkoxy radicals, aldehydes are the most significant flavour compounds. Aldehydes can be produced by splitting of the lipid molecules on either side of the radical. The products formed from these hydrolytic or metabolic reactions depend on the fatty acids' constituents, hydroperoxide isomers formed and the decomposition products stability. Temperature, time of heating and degree of autoxidation are variables, which affect thermal oxidation (Frankel, 1982). Numerous volatile aldehydes are formed by unsaturated fatty acids autoxidation reactions, some of the volatile products are enumerated in Table 1.2.1.

Generally, the flavors of aldehydes are described as green, metallic, beany and rancid, and are mainly responsible for the undesirable flavours in fats, oils and lipid-containing foods. Some aldehydes, particularly the unsaturated aldehydes, are very potent flavour compounds. It is quite important to mention that extremely low levels of several aldehydes contribute to the desirable flavours of many foods. Likewise, aliphatic ketones formed from lipids autoxidation contribute to the flavour of oils and food products. For instance, 2-Pentylfuran has been identified in many fats and oils and lipid containing foods such as spray-dried dairy products (Ho and Chen, 1994).

Table 1.2.1 Volatile aldehydes formed by autoxidation reactions of unsaturated fatty acids.

Fatty acid	Monohydroperoxides	Aldehydes formed
Oleate	8-OOH	2-undecenal
		Decanal
	9-OOH	2-decenal
		Nonanal
	10-OOH	Nonanal
	11-OOH	Octanal
Linoleate	9-OOH	2,4-decadienal

		3-nonenal
	13-OOH	Hexanal
	9-OOH	2,4,7-decatrienal
		3,6-nonadienal
	12-OOH	2,4-heptadienal
		3-hexenal
	13-OOH	3-hexenal
	16-OOH	Propanal
Arachidonate	8-OOH	2,4,7-tridecatrienal
		3,6-dodecadienal
	9-OOH	3,6-dodecadienal
	11-OOH	2,4-decadienal
		3-nonenal
	12-OOH	3-nonenal
	15-OOH	Hexanal
Eicosapentaenoate	5-OOH	2,4,7,10,13-hexadecapentaenal
		3,6,9,12-pentadecatetraenal
	8-OOH	2,4,7,10-tridecatetraenal
		3,6,9-dodecatrienal
	9-OOH	3,6,9-dodecatrienal
	11-OOH	2,4,7-decatrienal
		3,6-nonadienal
	12-OOH	3,6-nonadienal
	14-OOH	2,4-heptadienal
		3-hexenal
	15-OOH	3-hexenal
	18-OOH	Propanal

1.7.4 Water composition

The water composition of green coffee seed is between 8.5 - 12%. However, once it is beyond this level it results in microbial growth since moisture adversely affects the aroma and/or flavour quality as well as the health benefits. Equally, low moisture is known to produce cracks in the seeds and thereby decreases the seeds viability to germinate properly (Farah, 2004). Increased moisture content results in microbial growth associated with contamination of mycotoxins most notably Ochratoxin A (OTA). OTA is the foremost mycotoxin that has been detected in coffee bean, while Aflatoxin B1 is considered to be the most carcinogenic, recurrent and harmful among all aflatoxins (Marchese et al, 2018; Suarez-Quiroz et al, 2004). The OTA

presence in coffee beans can be due to several environmental conditions and/or processing conditions (Suarez-Quiroz et al, 2004; Nakajima et al, 1997). OTA is a common secondary metabolite produced by several toxigenic species of *Aspergillus* and *Penicillium*. Studies have shown that OTA has the potential to exhibit some carcinogenic, nephrotoxic and immunosuppressive properties (Le-Bars & Le-Bars, 2000; Hohler, 1998; IARC, 1993).

1.7.5 Caffeine content

Caffeine ($C_8H_{10}N_4O_2$) with a molecular mass (194) is a central nervous system (CNS) natural stimulant, and it belongs to the class of methylxanthine (Figure 1.3A). It is the world's most widely consumed psychoactive drug, mostly present in commonly consumed beverages such as coffee, tea, soft drinks, energy drinks and chocolates. It constitutes about 0.9 - 2.5% of the coffee dry weight. It accounts for over 10% of the perceived bitterness in coffee drink (Flament et al, 1968), and stimulates the CNS as an adenosine-receptor antagonist. This alkaloid is heat stable and its presence in Robusta is almost double in Arabica's concentration (dePaula and Farah, 2019).

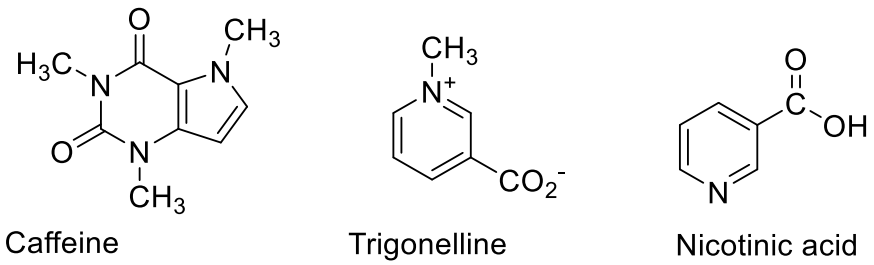


Figure 1.3 Chemical structures of caffeine, trigonelline and nicotinic acid.

While caffeine intake has been associated with high blood cholesterol levels, coronary diseases and cancer, other studies suggest that its consumption may lower the incidence of suicide and hepatic cirrhosis (Farah et al, 2006). Moderate level of caffeine consumption is typically associated with increased alertness, learning and memory capacity, body-building performance and imaginably improved humour, while higher dosage can result in undesirable effects in some

complex individuals as observed in disorders such as anxiety, tachycardia and insomnia during its half-life of approximately 2 - 6 hr after coffee ingestion (Clifford, 2000; Ogita et al, 2003). Generally, severe caffeine consumption is associated with some negative consequences on glucose tolerance and disposal, insulin sensitivity in slim-body, obese and Type-II diabetic humans; nevertheless, other coffee constituents have been reported with the potential of neutralizing these grave outcomes (Shearer et al; 2007).

Furthermore, acute caffeine consumption also raises the minerals levels in urinary excretion such as calcium (Ribeiro-Alves et al, 2003). Similarly, after long-term consumption of the beverage, most of these severe effects incline to disappear due to human metabolic adaptation mechanisms (Demirbag et al, 2006). Several metabolites resulting from caffeine breakdown in the body, particularly 1-methylxantine and 1-methylurate have demonstrated some levels of antioxidant activity in-vitro and in-vivo with an observation of consistent coffee iron-reducing capacity shown to be higher than that of decaffeinated coffee. The influence of antibacterial associated with regular coffee consumption against cariogenic microorganisms was reported to be higher than that of decaffeinated coffee (Lee, 2000; Antonio et al, 2010).

1.7.6 Trigonelline

Trigonelline ($C_7H_7NO_2$) is another naturally occurring, chemically heterogeneous, mostly alkaline, nitrogen-containing organic compound of secondary metabolism present in coffee. It is biologically derived from nicotinic acid enzymatic methylation process. Trigonelline (Figure 1.3B) adds to the bitterness of the coffee brew and has been reported as a precursor for the formation of diverse classes of volatile compounds during coffee roasting the likes of pyrroles and pyridines. However, some of these volatile compounds potent the capability of adding flavour to the drink upon roasting the bean (Flament et al, 1968). About 0.6 - 0.7% of this compound is present in Robusta, while Arabica contains more with about 2%. Moreover,

trigonelline demethylation during coffee roasting produces nicotinic acid, an essential vitamin B-complex constituent otherwise known as niacin (Trugo, 2003).

Niacin is also known as nicotinic acid (Figure 1.3C) is involved in protein, fat and carbohydrate metabolism. In coenzymes forms as NAD/NADP and their reduced forms $NADH^+$ H^+ / $NADPH^+$ H^+. Nicotinic acid is used as a hydrogen transmitter, meaning it is a reducing agent, also involved in the citrate (Tricarboxylic Acid Cycle) cycle and the respiratory chain in humans. It has an antioxidant effect and partakes in many physiological and enzymatic processes. Nicotinic acid is important for the regeneration of skin, muscles, nerves and DNA. Del Campo analysed and confirmed the presence of trigonelline and other compounds such as caffeine and 5-HMF (5-hydroxylmethyl-furfural) in soluble coffee by NMR spectroscopy (Del Campo et al, 2010).

1.7.7 Chlorogenic acids (CGAs)

CGAs are produced by many plants such as fruits, vegetables and medicinal plants with coffee and maté being the predominant rich dietary sources (Clifford, Kerimi & Williamson 2020). Coffee is the universal human major dietary source of acyl-quinic acids and consideration is given to their absorption and metabolism in the upper gastrointestinal tract and the colon, where the microbiota plays a key function in catabolism. Coffee possesses the most abundant concentration of CGAs of any species among the plant kingdom, with about 3.5 - 7.5% in Arabica and 7 - 14% in Robusta. Clifford and Ramirez-Martinez investigated the CGAs constituents in the two main coffee varieties and reported that green *Coffea canephora* beans grown in Sao Paulo and Santos (Brazilian regions) possessed a higher amount of CGAs when compared to green *Coffea arabica* beans cultivated in Africa regions specifically Ghana and Uganda (Clifford & Ramirez-Martinez; 1991).

CGAs represent a large family of esterified compounds present in green and roasted coffee. Characteristically, these classes of compounds are formed by esterification reaction between trans-cinnamic acids and quinic acid with the presence of axial hydroxyl groups on carbons -1 & 3, and equatorial hydroxyl groups on carbons -4 and 5. It has been observed that during processing, *Trans* isomers may be partly transformed into *cis*-forms (Clifford, 1999); and specifically during roasting, CGA's slowly decompose to form caffeic and quinic acid with about 50% of the original CGA being degraded in a medium roast (Liang & Kitts, 2015).

The word chlorogenic acids comprises of an enormous group of naturally-occurring compounds of which the majority are synthesised by the plants via esterification process of a C_6 - C_3 *trans*-hydroxycinnamic acid with 1L-(-)-quinic acid (Clifford & Abranko, 2017). The main hydroxycinnamates are caffeic, ferulic, p-coumaric and sinapic acids (Payen 1846a & 1846b; Robiquet et al, 1837; Rochleder & Liebigs, 1844).

Additionally, Clifford classified and defined CGAs as compounds formed when various quinic acid epimers, quinic acid methyl ethers, alky quinates, deoxyquinic acid, 2-hydroxyquinic acid, shikimic acid and its epimers, in addition to the analogous compounds are esterified with a hydroxybenzoic acid, hydroxyphenyl acetic acid or 3-(4'-hydroxyphenyl)-propionic acid (Clifford et al, 2017).

Quinic and caffeic acids have been associated with increased levels of astringency, and bitterness commonly seen in dark roasted coffees. CGAs production in the plant is initiated by several factors including responses to changes in the environments, plant stress and common pest infestation. In view of this, it is not unexpected that when grown in harsher conditions Robusta contains almost twice the concentration of CGA present in Arabica. CGA production also almost equals with caffeine; hence as CGA concentration increases so does the caffeine concentration.

CGAs ($C_{16}H_{18}O_9$) are primarily sub-divided according to the nature and number of cinnamic acids substituents and as well as the esterification position in the quinic acid cyclohexane ring (Clifford, 2000). The esters are preferentially formed with hydroxyl group located on carbon 5 along with those located on carbons 3 and 4. Besides, the less frequent esters may be formed with the -OH group located on carbon 1. The main sub-classes of CGAs in green coffee are caffeoylquinic acids, dicaffeoylquinic acids, feruloylquinic acids and less abundantly present are the *p*-coumaroylquinic acids and caffeoyl-feruloylquinic acids. Each of these subclasses consists of at least three key regio-isomers in addition to minor compounds, with the exception of the latter class, which contains six main isomers (Clifford et al, 2003; Farah et al, 2005, Jaiswal et al, 2010; Matei et al, 2012, Karaköse, Muller & Kuhnert, 2015). Among these classes, caffeoylquinic acids account for nearly 80% of the total CGAs content. In particular, 5-*O*-caffeoylquinic acid (5-CQA), the first to be identified, accounts for about 60% and thus the most studied isomer and with an available commercial standard. In view of this, 5-CQA is universally referred to as chlorogenic acid.

The coffee bean epidermal cell wall is enclosed by crystallized waxes, while CGAs, terpenes and 5-hydroxytryptamine derivatives (serotonin) are located in the cuticular layer. Basically, CGAs accumulate mostly in the cytoplasm of epidermal and parenchyma cells, however higher quantities are localised in the periplasmic membrane part of the cell wall. Similarly, caffeine is linked with CGAs to form a potassium chlorogenate complex and metallic salts in the cytoplasmic part of parenchyma cells (Dentan, 1987; Nestle, 1991; Bicho et al, 2013).

CGA are common dietary constituents in many plants aside coffee, like blueberries (*Vaccinium spp*.), cherries (*Prunus avium*), apples (*Malus pumila*), oregano (*Origanum vulgare*), spearmint (*Mentha spicata*), chicory (*Cichorium intybus*), aubergine (*Solanum melongena*) and sunflower (*Helianthus annus*) seeds (Clifford, 2000; Hermann, 1989; Manach et al, 2004; Crozier et al, 2006; Crozier et al, 2009), with high levels in globe artichoke (Schultz et al, 2004; Pandino et

al, 1990). In addition, the herbal tea mate infusion prepared from *Ilex paraguariensis* dry leaves contains substantial amounts of CQAs and DiCQAs (Clifford & Ramırez-Martınez, 1990).

Moreover, the bioavailability and metabolic fate of polyphenols in humans have been demonstrated to be absorbed and excreted to a large extent than numerous other dietary flavonoids and phenolics in nature (Stalmach et al, 2010). The main sources of cinnamates and their respective conjugates in human diets extracted from different data sources have previously been reported (Clifford, 1999). For instance, in cinnamates, caffeic acid sources are coffee beverage, blueberries, apples and ciders, *p*-coumaric acid (spinach, sugar beet fiber and cereal brans), ferulic acid (coffee beverage, citrus juices, sugar beet fiber and cereal brans) and sinapic acid (broccoli, kale, other leafy brassicas and citrus juices). While for conjugates, CQAs are mainly from coffee beverage, blueberries, apples and ciders, *p*-coumaric acid (sweet cherries), FQAs from coffee beverage, tartaric conjugates from spinach, lettuce, grapes and wines, malic conjugates (lettuce, spinach and legumes), rosmarinic acid (stuffing, culinary and mixed herbs) and cell wall conjugates from spinach, sugar beet fibre and cereal brans.

The hydroxycinnamic acid moiety of acyl-quinic acids is mostly found in the *Trans*-form. Some *cis*-isomers have also been identified; with cis-5-*O*-*p*-coumarylquinic acid presence in herbal aster (*Aster ageratoides*) of flower buds (Clifford et al, 2006), and similarly wide range of *cis*-isomers have been detected in other plant species (Kuhnert et al, 2010; Clifford et al, 2007; Jaiswal et al, 2013, Karaköse, Muller & Kuhnert, 2015; Jaiswal, Deshpande & Kuhnert, 2011; Clifford, 2017; Jaiswal & Kuhnert 2011). It has been proposed that the *cis*- derivatives possibly originates from plant tissues when *Trans* isomer are exposed to relatively strong UV radiation, which induces geometric isomerisation (Clifford et al, 2008; Karaköse et al, 2015).

1.7.8 Identification of CGAs

The mono-CQAs are usually identified based on their Tandem-MS in negative ion mode as [M-H] at *m/z* 353 as precursor ion, which yields fragmented ions with *m/z* 191 or 173, a characteristic feature of quinic and dehydrated quinic acid moieties respectively. The occurrence of these two fragmented ions in MS^2 and MS^3 indicate the presence of mono-acyl and di-acyl CGAs respectively; while in MS^4, they indicate the presence of tri-acyl CGAs. For example, Figure 1.4 represents the MS^2 and MS^3 of the mono-acyl 3-isomers (3-CQA, 3-FQA & 3-*p*CoQA) of the CGAs base peak ions. Most importantly, the retention time (min) and elution order on a reverse phase column aids in identifying the individual CGAs regio-isomers, while the parent ion *m/z* value reveals the chemical feature of the cinnamic acid moiety for instance *m/z* 353 represents caffeoylquinic acids, *m/z* 367 indicates feruloylqunic acids, *m/z* 337 is for *p*-coumaroylquinic acids and *m/z* 515 represents di-caffeoylquinic acids. It is quite important to note that the mono-acyl CGAs isomers elution order usually follows this pattern: 3-CQA (**1**) > 5-CQA (**3**) > 4-CQA (**2**), same order applies to the FQAs. Furthermore, in accordance with the hierarchical approach proposed by Clifford et al. (2005), the CQA isomers may be distinguished based on the intensities of the fragment ion signals at *m/z* 191, 179, and 173 in the MS^2 spectra with moderate fragmentation energy. The MS^n data of the mono-acyl and di-acyl CGAs are provided in Table 1.3.

The mono-acyl moieties being more polar usually elute before the di-acyl in the reverse phase column set-up. Water, methanol and acetonitrile are typically used for HPLC mobile phases coupled with diphenyl, C-18, C-8 and phenylhexyl columns as reverse stationary phases.

In addition, MS^n targeted studies can be carried out on an intact acyl substituent in negative ion mode when observed. For example, quinic acid moiety stabilizes the negative (-ve) charge thus it is mostly observed in negative ion mode. Moreover, the side chain of the acyl moiety intact

ion specifically stabilizes the positive (+ve) charge for it to be accurately observed in positive ion mode. In view of this, targeted MS^n experiments in positive ion mode also thus aid in the structural elucidation of the individual acyl moiety.

Furthermore, the high-resolution mass measurement using LC-TOF-MS is mostly explored in confirming the molecular formula of the respective CGAs with a mass error lower than five ppm being considered acceptable for publications standard (Jaiswal & Kuhnert, 2011; Jaiswal et al, 2010; Kuhnert et al, 2010; Jaiswal et al, 2011). Clifford and Kuhnert (2005) considered important characteristics such as hydrophobicity, fragmentation patterns and retention times to develop the hierarchical scheme to accurately identify CGAs. The schemes are used in this study for the CGAs regio-isomers assignment with the Tandem-MS spectra.

Some of the common CGAs in green coffee bean that has characterized in accordance with their regio-isomer levels, based on their fragmentation patterns from the Tandem-MS data and respective retention times are provided in Table 1.4, while Figure 1.5 represents the chemical structures of CGAs and their respective regio-isomers.

Table 1.3 The MS^n data of the mono- and di-acyl CGAs in negative ion mode.

S/N	CGA	MS^1	MS^2							MS^3					
		Parent ion	Base peak			Secondary peak				Base peak		Secondary peak			
			m/z	m/z	int	m/z	int	m/z	int	m/z	m/z	int	m/z	int	
1	**3-CQA**	353.1	190.9	178.5	50			134.9	7	85.3	127.0	71	172.9	67	
2	**4-CQA**	353.1	172.9	178.9	60	190.8	20	135.0	9	93.2	111.0	48			
3	**5-CQA**	353.2	190.0	178.5	5			135.0	15	85.2	126.9	66	172.9	27	
4	**3-FQA**	367.2	192.9	191.5	2	173.2	2			133.9	148.9	23			
5	**4-FQA**	367.2	172.9	192.9	16					93.1	111.5	44			
6	**5-FQA**	367.2	190.9	172.9	2					85.2	126.9	70			
7	**3-*p*CoQA**	337.1	162.9	190.0	5					118.9					
8	**4-*p*CoQA**	337.1	172.7							93.0	111.0	61			
9	**5-*p*CoQA**	337.2	190.9	162.9	5					85.2					

S/N	CGA	MS^1	MS^2							MS^3								MS^4		
		Parent ion	Base peak			Secondary peak				Base peak		Secondary peak						Base peak		Secondary peak
			m/z	m/z	int	m/z	int	m/z	int	m/z	m/z	int	m/z	int	m/z	int	m/z	m/z	int	
13	**3,4-DiCQA**	515.2	353.1	335.1	4	172.9	20			172.9	178.9	68	191.0	32	135.1	9	93.2	111.1	30	
14	**3,5-DiCQA**	515.2	353.1																	
15	**4,5-DiCQA**	515.2	353.1	335.1	2	172.9	6			172.9	178.9	76	190.9	9	135.0	19	93.1	111.0	20	

Figure 1.4 The MS^2 and MS^3 spectra of 3-mono acyl CGAs in negative ion mode.

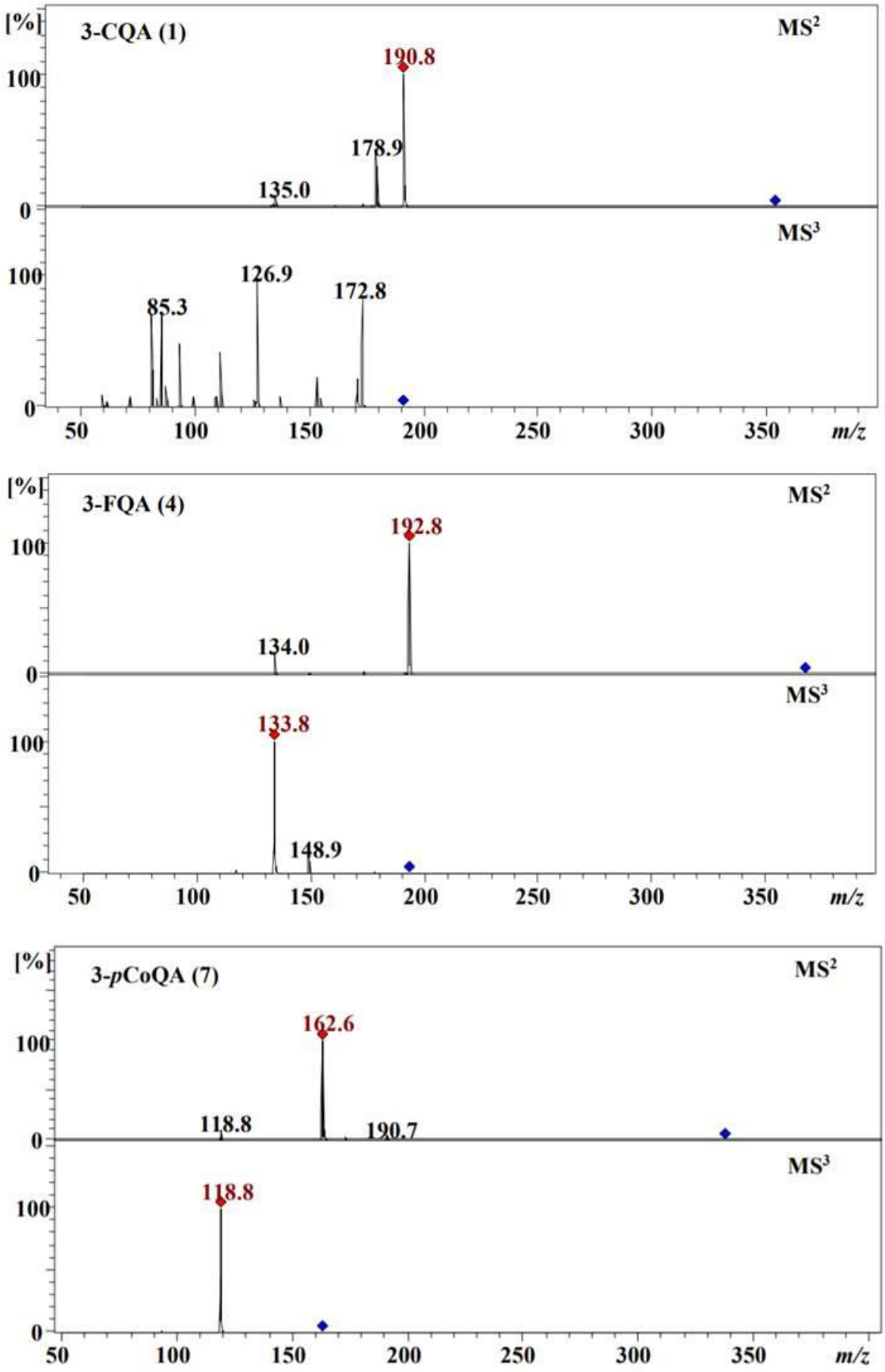
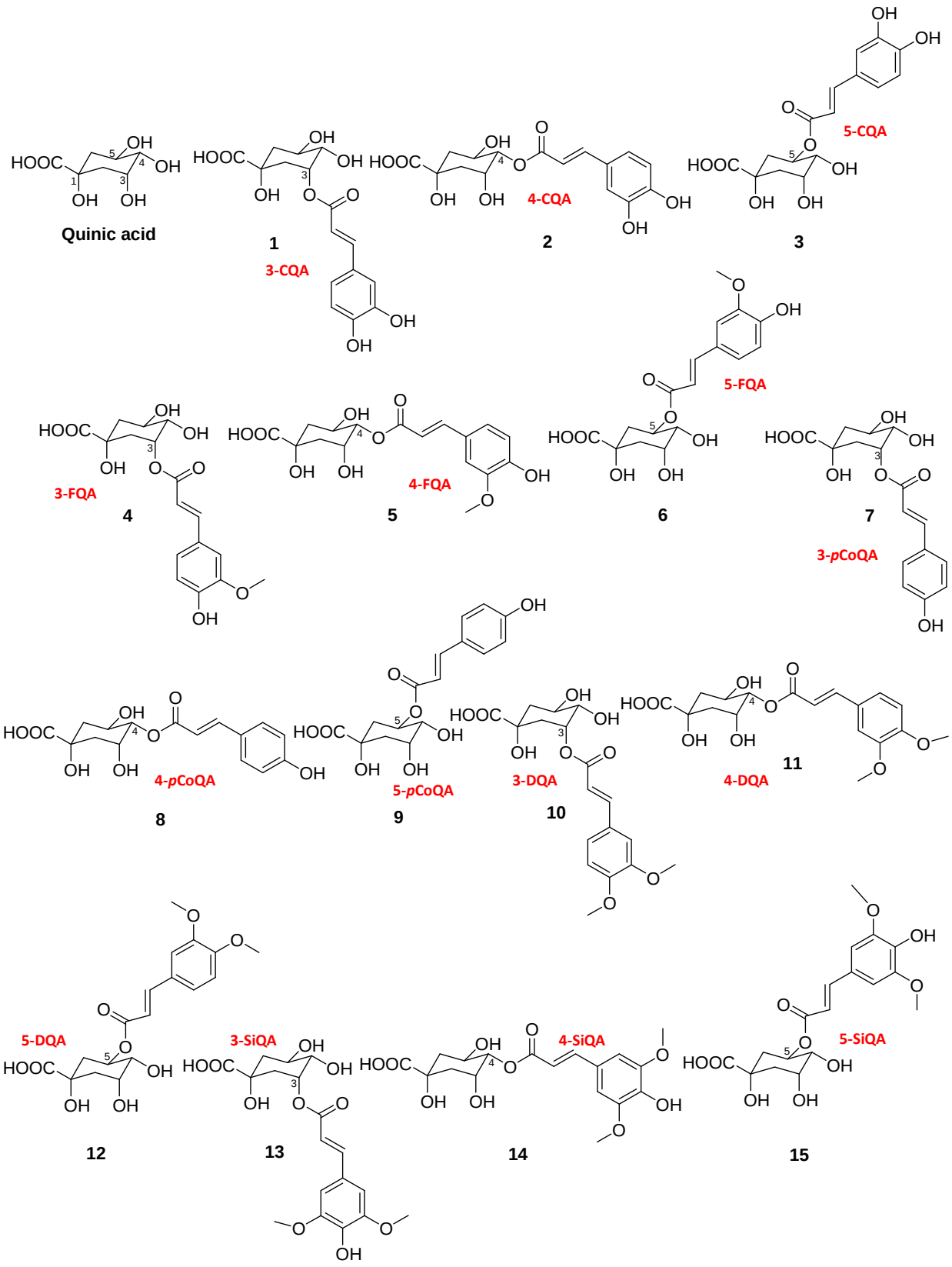


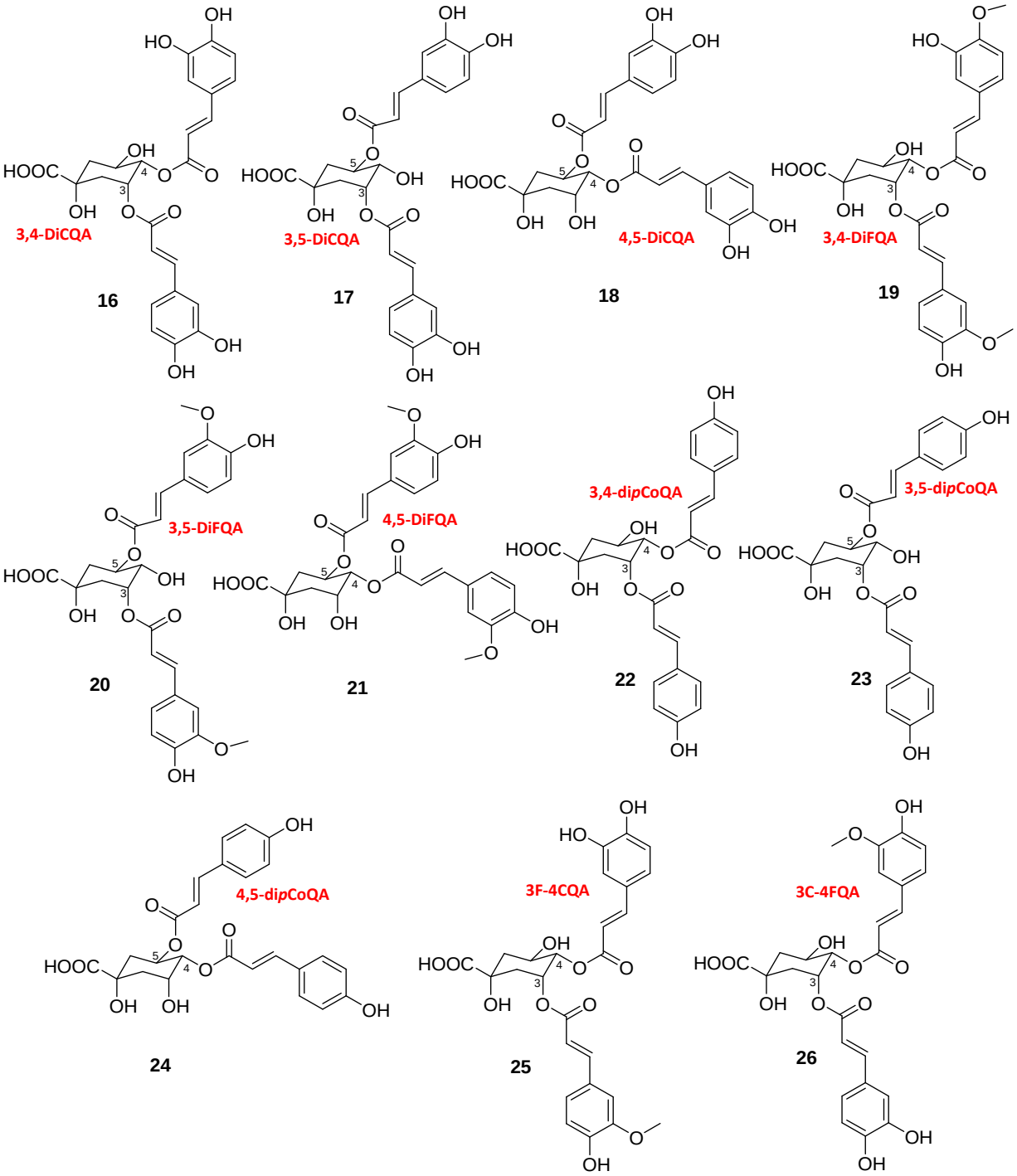
Table 1.4 List of identified CGAs with their respective molecular formula.

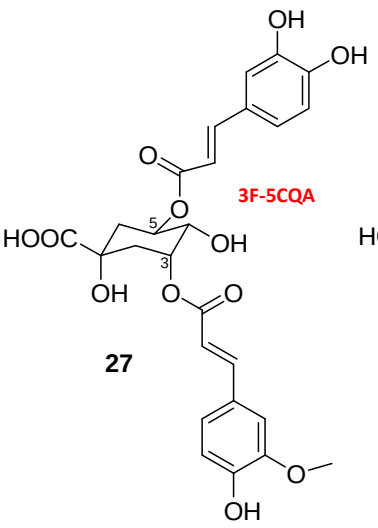
No.	CGA name	Abbreviation	Molecular formula
1	3-*O*-caffeoylquinic acid	3-CQA	$C_{16}H_{18}O_9$
2	4-*O*-caffeoylquinic acid	4-CQA	$C_{16}H_{18}O_9$
3	5-*O*-caffeoylquinic acid	5-CQA	$C_{16}H_{18}O_9$
4	3-*O*-feruloylquinic acid	3-FQA	$C_{17}H_{20}O_9$
5	4-*O*-feruloylquinic acid	4-FQA	$C_{17}H_{20}O_9$
6	5-*O*-feruloylquinic acid	5-FQA	$C_{17}H_{20}O_9$
7	3-*O*-*p*-coumaroylquinic acid	3-*p*CoQA	$C_{16}H_{18}O_8$
8	4-*O*-*p*-coumaroylquinic acid	4-*p*CoQA	$C_{16}H_{18}O_8$
9	5-*O*-*p*-coumaroylquinic acid	5-*p*CoQA	$C_{16}H_{18}O_8$
10	3-*O*-dimethoxycinnamoylquinic acid	3-DQA	$C_{18}H_{22}O_9$
11	4-*O*-dimethoxycinnamoylquinic acid	4-DQA	$C_{18}H_{22}O_9$
12	5-*O*-dimethoxycinnamoylquinic acid	5-DQA	$C_{18}H_{22}O_9$
13	3-*O*-sinapoylquinic acid	3-SiQA	$C_{18}H_{22}O_{10}$
14	4-*O*-sinapoylquinic acid	4-SiQA	$C_{18}H_{22}O_{10}$
15	5-*O*-sinapoylquinic acid	5-SiQA	$C_{18}H_{22}O_{10}$
16	3,4-di-*O*-caffeoylquinic acid	3,4-diCQA	$C_{25}H_{24}O_{11}$
17	3,5-di-*O*-caffeoylquinic acid	3,5-diCQA	$C_{25}H_{24}O_{11}$
18	4,5-di-*O*-caffeoylquinic acid	4,5-diCQA	$C_{25}H_{24}O_{11}$
19	3,4-di-*O*-feruloylquinic acid	3,4-diFQA	$C_{27}H_{28}O_{12}$
20	3,5-di-*O*-feruloylquinic acid	3,5-diFQA	$C_{27}H_{28}O_{12}$
21	4,5-di-*O*-feruloylquinic acid	4,5-diFQA	$C_{27}H_{28}O_{12}$
22	3,4-di-*O*-*p*-coumaroylquinic acid	3,4-di*p*CoQA	$C_{25}H_{24}O_{11}$
23	3,5-di-*O*-*p*-coumaroylquinic acid	3,5-di*p*CoQA	$C_{25}H_{24}O_{11}$
24	4,5-di-*O*-*p*-coumaroylquinic acid	4,5-di*p*CoQA	$C_{25}H_{24}O_{11}$
25	3-*O*-feruloyl-4-*O*-caffeoylquinic acid	3F-4CQA	$C_{26}H_{26}O_{12}$
26	3-*O*-caffeoyl-4-*O*-feruloylquinic acid	3C-4FQA	$C_{26}H_{26}O_{12}$
27	3-*O*-feruloyl-5-*O*-caffeoylquinic acid	3F-5CQA	$C_{26}H_{26}O_{12}$
28	3-*O*-caffeoyl-5-*O*-feruloylquinic acid	3C-5FQA	$C_{26}H_{26}O_{12}$
29	4-*O*-feruloyl-5-*O*-caffeoylquinic acid	4F-5CQA	$C_{26}H_{26}O_{12}$
30	4-*O*-caffeoyl-5-*O*-feruloylquinic acid	4C-5FQA	$C_{26}H_{26}O_{12}$
31	3-*O*-dimethoxycinnamoyl-4-*O*-caffeoylquinic acid	3D-4CQA	$C_{27}H_{28}O_{12}$
32	3-*O*-dimethoxycinnamoyl-5-*O*-caffeoylquinic acid	3D-5CQA	$C_{27}H_{28}O_{12}$
33	4-*O*-dimethoxycinnamoyl-5-*O*-caffeoylquinic acid	4D-5CQA	$C_{27}H_{28}O_{12}$
34	3-*O*-dimethoxycinnamoyl-4-*O*-feruloylquinic acid	3D-4FQA	$C_{27}H_{28}O_{12}$
35	3-*O*-dimethoxycinnamoyl-5-*O*-feruloylquinic acid	3D-5FQA	$C_{27}H_{28}O_{12}$
36	4-*O*-dimethoxycinnamoyl-5-*O*-feruloylquinic acid	4D-5FQA	$C_{27}H_{28}O_{12}$
37	3-*O*-p-coumaroyl-4-*O*-caffeoylquinic acid	3*p*Co-4CQA	$C_{25}H_{24}O_{11}$
38	3-*O*-caffeoyl-4-*O*-*p*-coumaroylquinic acid	3C-4*p*CoQA	$C_{25}H_{24}O_{11}$
39	3-*O*-*p*-coumaroyl-5-*O*-caffeoylquinic acid	3*p*Co-5CQA	$C_{25}H_{24}O_{11}$
40	3-*O*-caffeoyl-5-*O*-*p*-coumaroylquinic acid	3C-5*p*CoQA	$C_{25}H_{24}O_{11}$
41	4-*O*-caffeoyl-5-*O*-*p*-coumaroylquinic acid	4C-5*p*CoQA	$C_{25}H_{24}O_{11}$
42	4-*O*-*p*-coumaroyl-5-*O*-caffeoylquinic acid	4*p*Co-5CQA	$C_{25}H_{24}O_{11}$
43	3-*O*-*p*-coumaroyl-4-*O*-feruloylquinic acid	3*p*Co-4FQA	$C_{26}H_{26}O_{11}$
44	3-*O*-*p*-coumaroyl-5-*O*-feruloylquinic acid	3*p*Co-5FQA	$C_{26}H_{26}O_{11}$

C = caffeoyl; D = dimethoxycinnamoyl; F = feruloyl; *p*Co = *p*-coumaroyl; Si = sinapoyl;

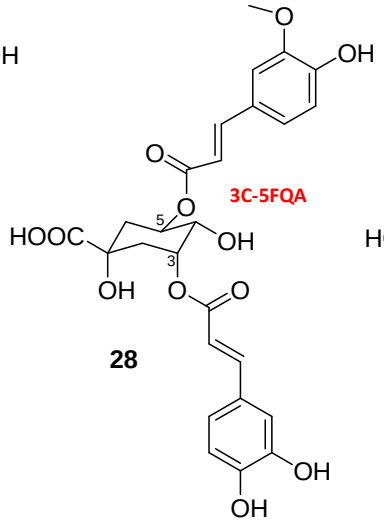
Figure 1.5 Structures of green coffee bean chlorogenic acids with IUPAC nomenclature.



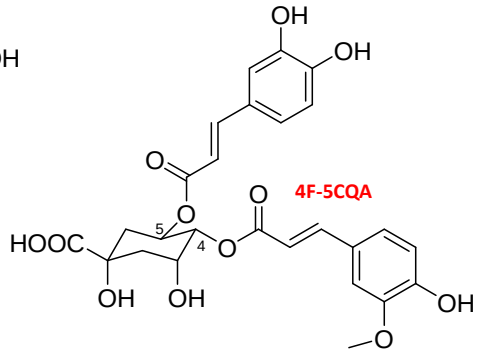

3,4-DiCQA
16
3,5-DiCQA
17
4,5-DiCQA
18
3,4-DiFQA
19
3,5-DiFQA
20
4,5-DiFQA
21
3,4-dipCoQA
22
3,5-dipCoQA
23
4,5-dipCoQA
24
3F-4CQA
25
3C-4FQA
26



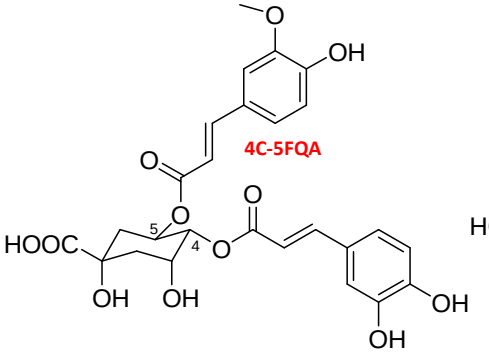
27



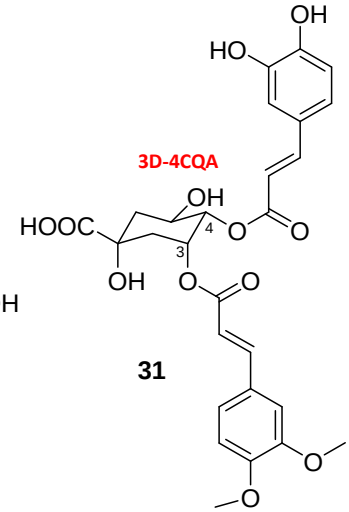
28



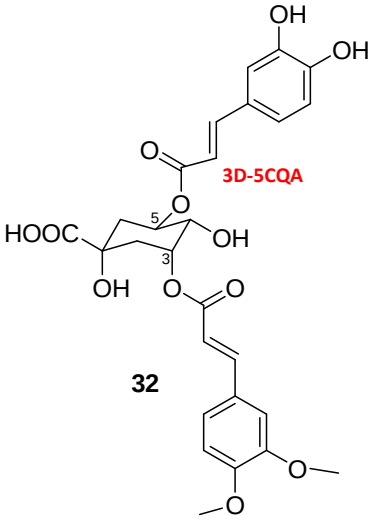
29



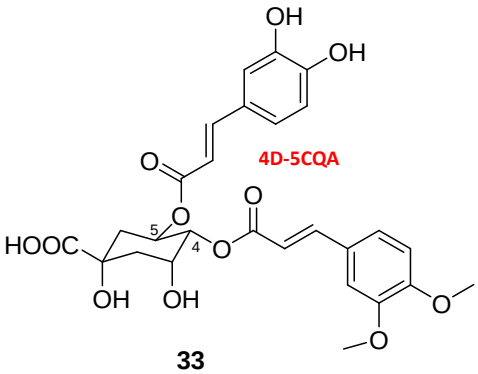
30



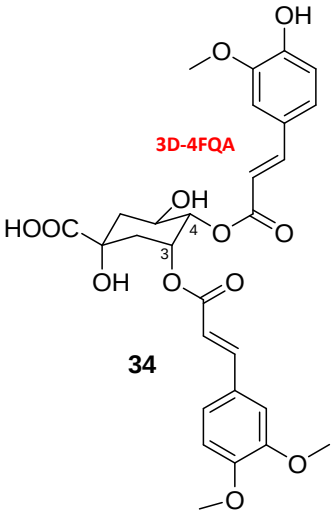
31



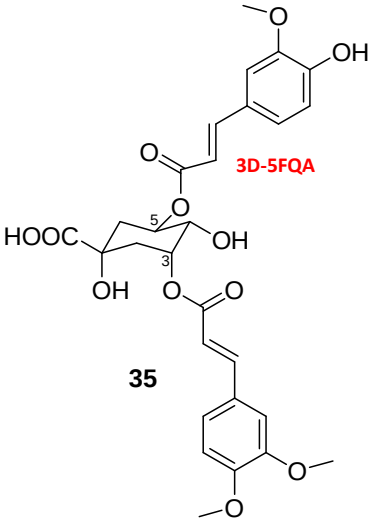
32



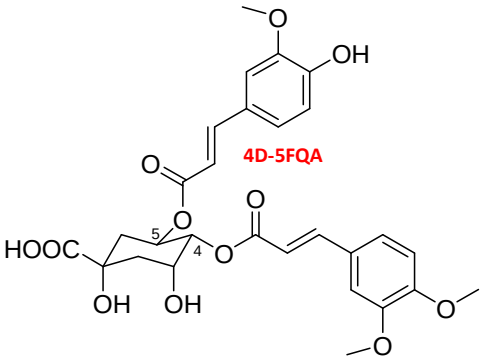
33



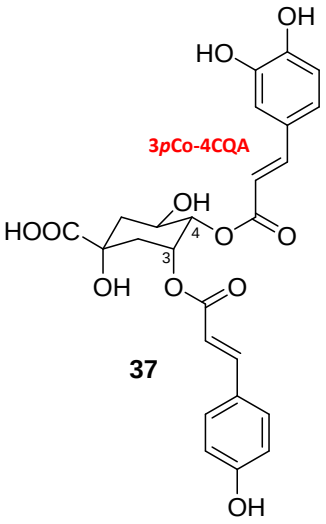
34



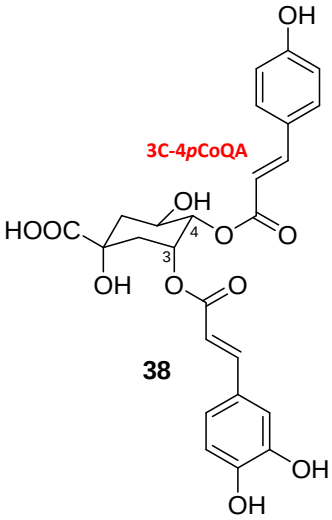
35



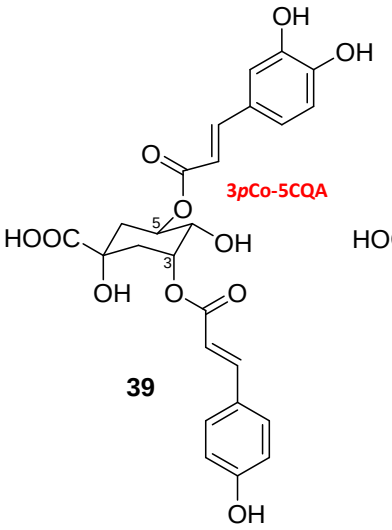
36



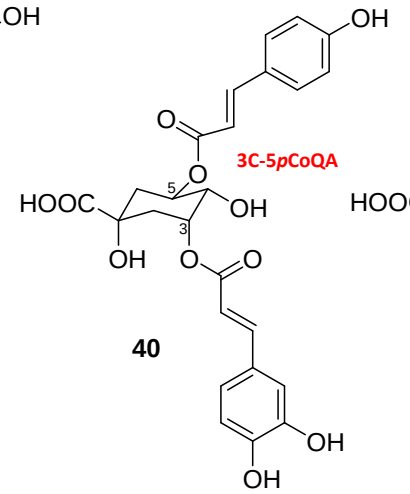
37



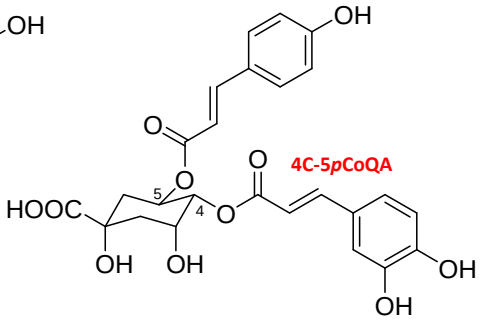
38



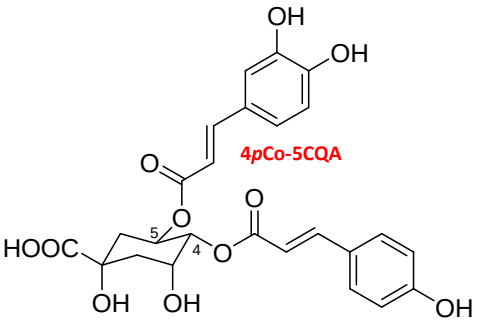
39



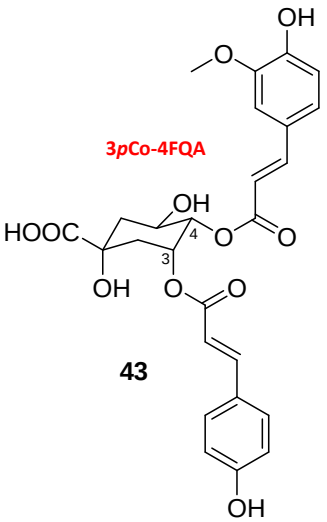
40



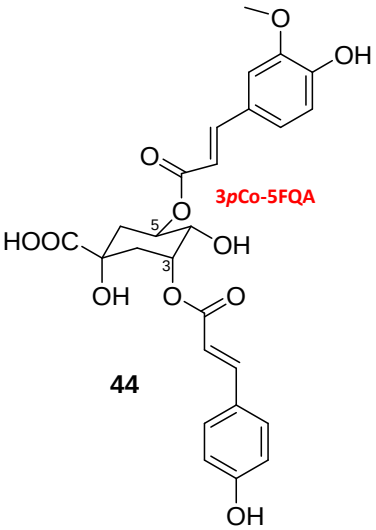
41



42



43



44

1.7.9 Coffee roasting and melanoidins formation

Mostly, coffee consumption is inspired by its satisfying aroma, flavour, taste and the resulting pleasant sensations as well as the physiological effects (Vignoli et al; 2014). The variations in the green coffee beans constituents and processing procedures influence the roasted coffee characteristics. These significant features in coffee brew, prepared by hot water extraction from roasted and ground coffee beans, are abundantly influenced by the roasting process (Schnker et al, 2002; Moon & Shibamoto, 2009; Kumazawa & Masuda; 2003).

The degree of coffee roasting is measured by temperature range and length of roasting (time in minutes). Typically, this is qualitatively evaluated from the resulting colour when the green bean is transformed into brown colour, and categorized as a light, medium or dark roast coffee (Somporn et al; 2011).

Naturally, green coffee beans are mainly composed of carbohydrates (59 - 61%), lipids (10 - 16%), proteins (10%) and chlorogenic acids (7 - 10%), with lower amounts of minerals (4%), aliphatic acids (2%), caffeine (1 - 2%), trigonelline (1%) and free amino acids (< 1%). However, after the bean roasting the coffee bean carbohydrates (38 - 42%), proteins (8%), chlorogenic acids (3 - 4%) and free amino acids are considerably reduced, while the lipids (11 - 17%), minerals (5%), aliphatic acids (3%), caffeine (1 - 2%) and trigonelline (1%) constituents remain relatively the same (Illy et al; 1995).

The bean roasting process usually affects the coffee chemical constituents and biochemical properties thus changes in the coffee physicochemical activities. Similarly, natural bioactive compounds such as polyphenols (antioxidants) might be partially or completely degraded, while other phenolic compounds are formed during roasting such as Maillard reaction products (Wang, Qian, & Yao, 2011). However, melanoidins formation seems to recompense for any potential decrease in antioxidants activity (Gniechwitz et al, 2008).

Maillard reaction is a significant phenomenon that occurs during coffee roasting. The coffee roasting process is a non-enzymatic browning reaction comprising of series of different chemical reactions between reducing sugars and compounds with a free amino group to form kinds of numerous products, which can be categorized as early-, intermediate- and last stage products, commonly refers to as melanoidins (Hodge; 1953). Figures 1.6 and 1.7 represent the simple scheme and reaction mechanism involved in Maillard reaction respectively.

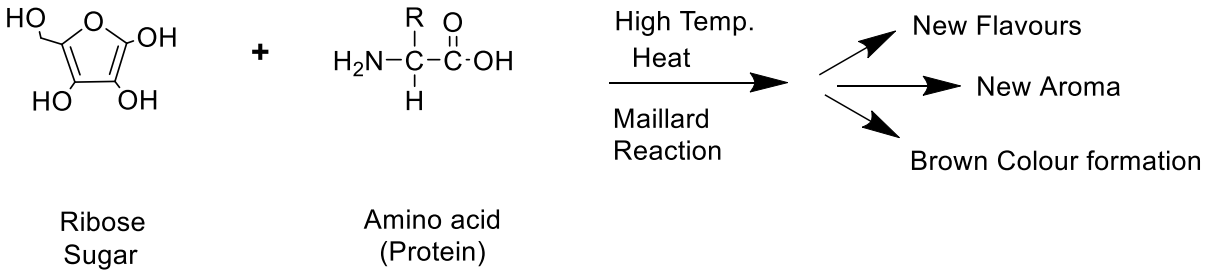


Figure 1.6 Scheme of Maillard reactions.

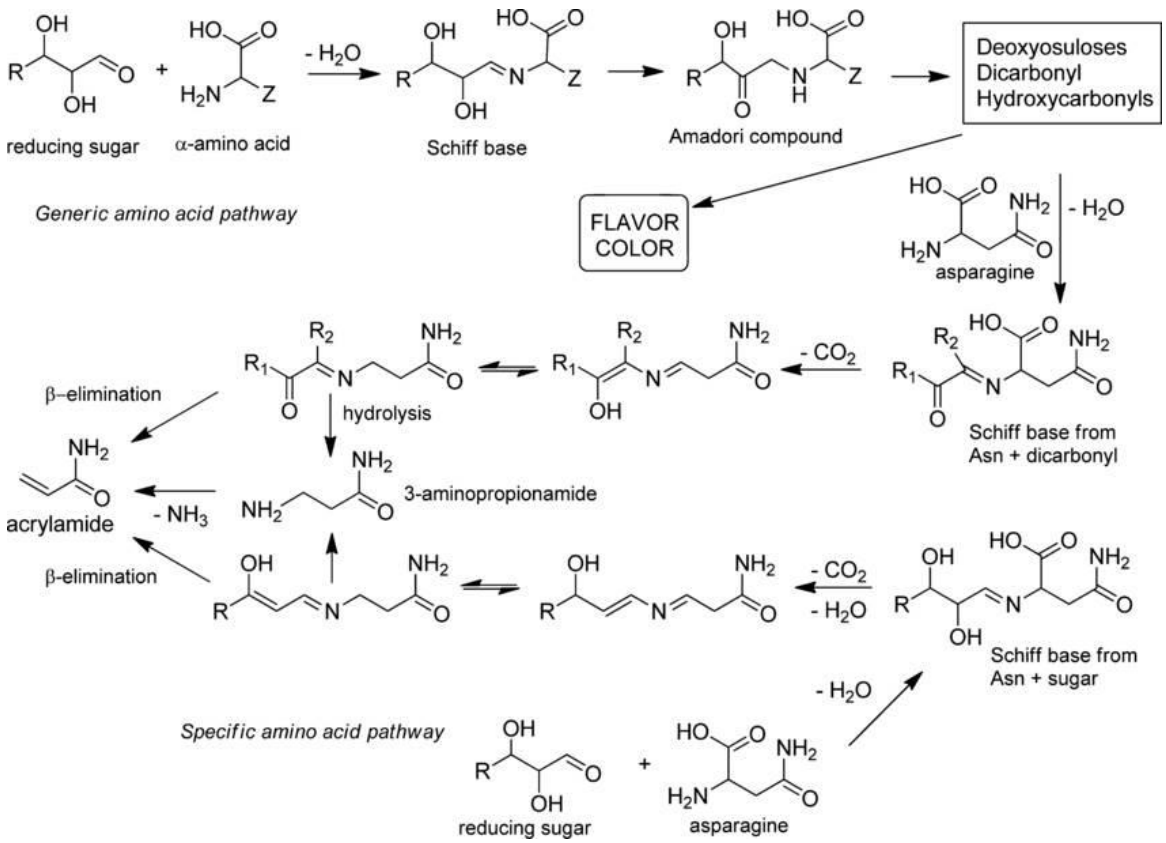


Figure 1.7 Maillard chemical reactions (Rizzi, 1997).

Similarly, the green coffee bean constituents undergo structural transformation and biochemical changes during roasting, which results in the formation of melanoidins (Rizzi, 1997; Martins et al, 2000). However, during roasting, there is a progressive destruction and transformation of CGAs with some 8 - 10% being lost for every 1% loss of the dry matter (Clifford, 1999). Likewise, CGAs are converted into CGA lactones during roasting by loss of a water molecule from the quinic acid moiety leading to the formation of an intramolecular ester bond as depicted in Figures 1.8 and 1.9. The non-volatile products of coffee roasting are generally referred to as melanoidins (Jaiswal et al, 2012) and classified as high molecular weight nitrogenous and brown-coloured compounds (Nunes & Coimbra, 2009; Borrelli et al, 2002), with examples shown in Figure 1.9.1. This development has been broadly reported to generate the pleasant aroma, colour and flavour observed in different coffees, equally contributes to the drink antioxidant activity via direct free radical scavenging activities, along with metal pro-oxidant sequestering activity (Liu & Kitts, 2011; Rufian-Henares & Morales, 2007; Liang & Kitts, 2014). The unique compounds produced commonly possess some biological activities in humans, and coffee-brew is one of the key sources of melanoidins in the human dietary intake. The human health effects of melanoidins are of great importance, for instance melanoidins are associated with antioxidant, antimicrobial, anti-hypertensive, anticariogenic, anti-glycative and anti-inflammatory properties (Moreira et al, 2012).

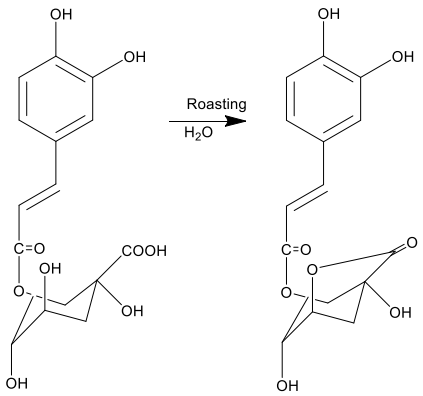


Figure 1.8 Formation of 1,5-γ-quinolactone from 3-*O*-caffeoylquinic acid during roasting.

Figure 1.9 illustrates the comprehensive scheme of the fate of chlorogenic acids during coffee raosting. Similarly, it shows that several chemical reactions such as acyl migration (Deshpande et al, 2014; Deshpande et al, 2016), lactonisation and dehydration (Jaiswal et al, 2012), *Trans-Cis* isomerization (Matei et al, 2016; Matei, Jaiswal & Kuhnert, 2012), epimerization (Jaiswal, Dickman & Kuhnert, 2012) and oxidation. These reactions characteristically take place during coffee roasting depending on the different conditions employed, which thus leads to the formation of several compounds such as the CGA lactones.

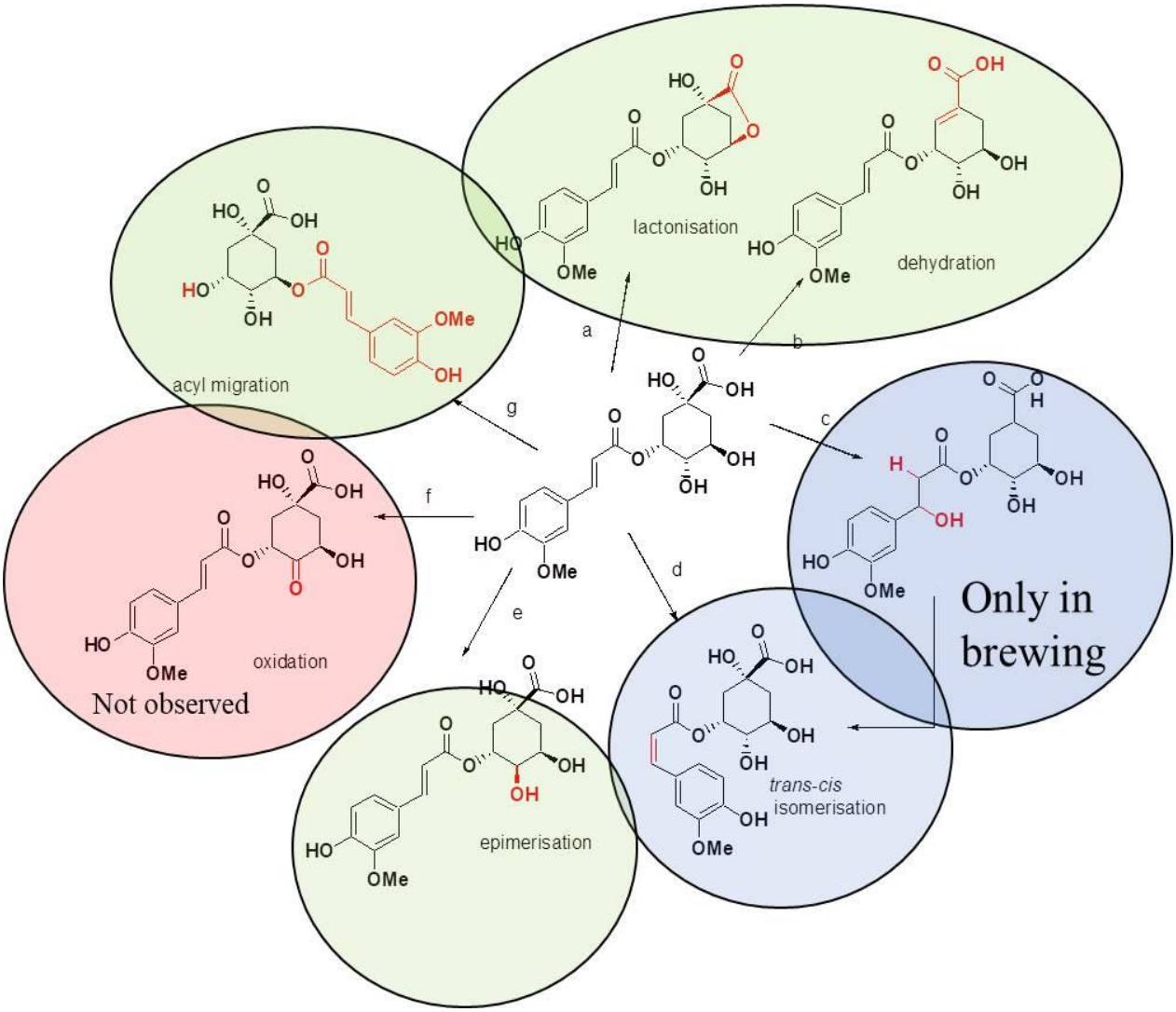


Figure 1.9 Chemical reactions during coffee roasting.

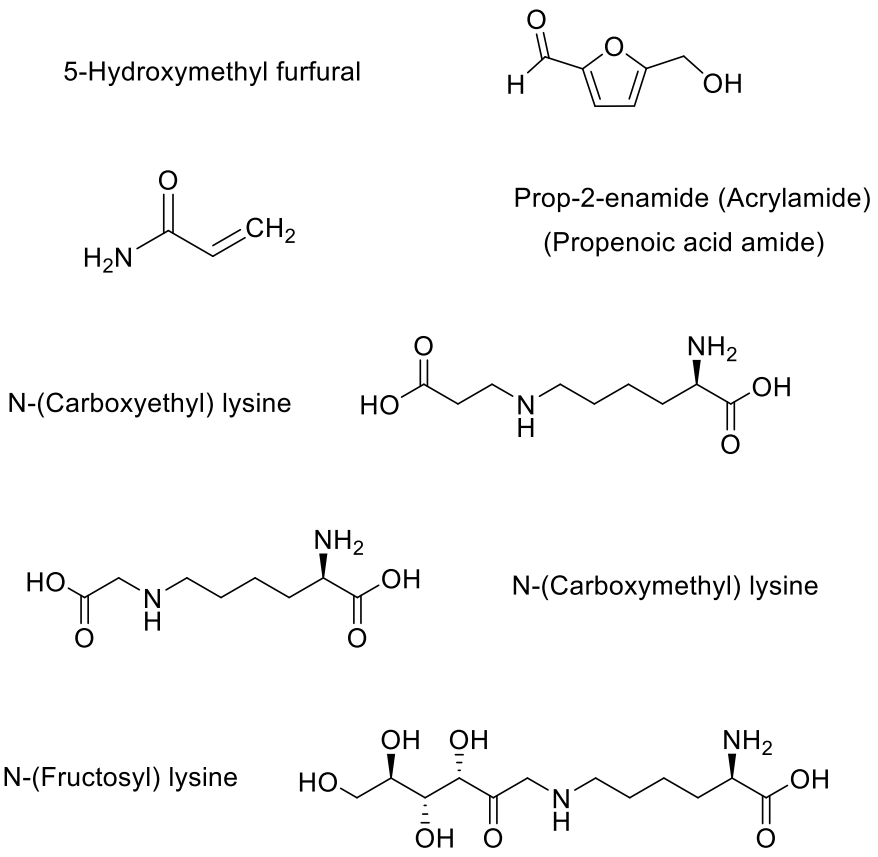


Figure 1.9.1 Products of Maillard reactions generated during coffee roasting.

Coffee is a rich source of dietary antioxidants (Clifford et al, 2006), nonetheless Somporn reported that a reduced level of antioxidant activity is associated with increased degree of coffee roasting (Somporn et al, 2011). Furthermore, some volatile heterocyclic compounds that are generated during this process similarly have been confirmed to possess antioxidants capacity (Yanagimoto et al, 2004).

In addition, due to the significant amount of CGA present in coffee beans, Cropley reported that coffee is a good source of antioxidants in human diets, which means that each cup of coffee consumed daily gives humans some health benefits (Cropley, 2012) apart from satisfaction derived from the drink, the brain alertness and staying awake tendency. Also, Fogliano and Morales reported that the universal dietary intakes of coffee melanoidins range from 5 - 40 mg/kg/day in 28 different countries investigated (Fogliano & Morales, 2011).

Additionally, recent studies revealed that coffee constituents can activate tissue antioxidant gene expression and protect it against gastrointestinal oxidative stress (Gunalan et al, 2012). In addition, contemporary works have linked coffee consumption with improved health status in humans (Huck et al, 2005; Prior et al, 2005). This is attributed to bioactive complexes presence in coffee (Trong et al, 2007; Tunnicliffe et al, 2008), with antioxidant properties (Gomez-Ruiz et al, 2008; Stadler et al, 1994). Nevertheless, this antioxidant hypothesis is still under discussion and recent evidence suggested that this role in human health is overestimated (Mikutis et al, 2013).

Furthermore, the antioxidant amount mostly depends on the method of preparation which could either be by roasting or fermentation. This is because there is a possibility that each of the processing methods influences the chemical components, likewise the rate of degradation, hence the variations in the antioxidants quantities (Duarte et al, 2005). Similarly, CGAs content in the coffee is dependent on the diverse species, the variety and the processing conditions of the coffee beans (Daglia et al, 2000; Moreira et al, 2005). Robusta has been widely reported to possess higher amount of antioxidant quantity compared to Arabica coffee (Clifford & Ramirez-Martinez, 1991; Vignoli et al, 2014; Daglia et al, 2000). Therefore, associated differences in green bean composition, roasting conditions and extraction procedures adopted for the coffee beverages preparation usually result in pronounced diversity of the chemical composition of the final product (Borrelli et al, 2002), which could contribute to the variances in the different coffee varieties biological activities.

Excessive production of free radicals chemically known as reactive oxygen species (ROS) can result into pathological conditions such as inflammation, premature aging disorders and several diseases like cancer, Parkinson's and Alzheimer's diseases, atherosclerosis, myocardial infarction, diabetes mellitus and chronic fatigue syndrome (Wilson et al, 2011). Over time,

humans and animals have systematically developed well-structured and complex antioxidant mechanisms to protect or prevent themselves from oxidative stress.

Nevertheless, excessive ROS production can be very severe and damaging to the human biological system. To ameliorate this pathology, humans consume various foods and drinks like coffee to reduce and/or purge the system of free radicals. Furthermore, the consumption of coffee in humans has been documented as one of the good sources of antioxidant to expunge the free radicals generated from various biochemical reactions occurring in the body system (Agudelo-Ochoa et al, 2016).

In addition, it has been revealed that green coffee possesses antioxidant potential against lipid peroxidation via in-vitro studies (Kroyer et al, 1989) and antineoplastic activity (Rosenberg, 1990). Also roasted coffee has been established to have mutagenic activity against sudden changes in gene expression, perhaps as a result of hydrogen peroxide formation attributed as a key provider to coffee genotoxicity in vitro (Miller et al, 1993) and antibacterial activity against a wide range of known bacteria.

There are various laboratory assays use in determining and quantifying the presence of antioxidants in a given food sample such as coffee. The universally used assays are DPPH [2,2-diphenyl-1-picrylhydrazyl], Folin-ciocalteau, ABTS [2,2'-azino-bis (3-ethylbenzothiazoli ne-6-sulphonic acid)] and FRAP [Ferric ion reducing antioxidant power] (Rigane et al, 2016).

In FRAP assay, a yellow coloured ferric reagent is reduced to ferrous ion at low pH leading to a coloured ferrous-tripyridyltriazine (TPTZ) deep blue-complex formation with an absorbance measurement at 593 nm (Zou et al, 2011; Huang et al, 2005). Besides, this assay could as well be employed in determining antioxidant capacity of human's biological samples such as the blood plasma and serum (Benzie, 1996).

1.8 Biological and pharmacological properties of coffee CGAs

CGAs abundantly present majorly in coffee or other plant sources are essential and biologically active dietary polyphenol, playing numerous significant and therapeutic functions most importantly as antidiabetic (Clifford et al, 2017; Van Dam, 2008; Yu et al, 2011; Williamson, 2020; Jiang, Zhang & Jiang, 2013; Marventano et al, 2016; Jacobs et al, 2014; Mellbye et al, 2017; Pimpley et al, 2020; Ludwig et al, 2014) and hepato-protective (Wijarnpreecha, Thongprayoon & Ungprasert, 2017; Salomone et al, 2017; Moreira et al, 2005; Shi et al, 2013; Pereira et al, 2003; Hemmerle et al, 1997; Natella et al, 2002) properties. In addition, studies have shown that cafestol possesses antidiabetic properties in induced-diabetic mice and it may contribute to the reduced risk of developing Type-II diabetes in coffee consumers, thus this compound has a potential role as an antidiabetic drug (Mellbye et al, 2017; Williamson, 2020).

Other known health functions are antibacterial (Karunanidhi et al, 2012; Bajko et al, 2016; Farzaei et al, 2015; Ayseli et al, 2016; Fu et al, 2016; Sousa et al, 2014; Naveed et al, 2018), antifungal (Sung & Lee, 2010), anti-inflammatory (Liu et al, 2017), anti-obesity (Meng et al, 2013; Cho et al, 2010; Jin et al, 2015; Huang et al, 2015), antiviral (Gamaleldin Elsadig Karar et al, 2016; Khan et al, 2005; Tamura et al, 2006; Chiang et al, 2002), anti-microbial (Venditti et al, 2015), anti-hypertensive (Wan et al, 2013; Zhao et al, 2012; Suzuki et al, 2002), possess inhibitory effect of multi-drug bacteria resistance (Fiamegos et al, 2011), protective effects on glucose and lipid metabolism hereditary disorders (Zhang et al, 2011), stimulates central nervous system (Snel & Lorist, 2011), act as free radicals scavenger with anti-oxidative potential (Santana-Gálvez et al, 2017). Furthermore, studies have shown that treatment with chlorogenic acid could protect against brain neurological degeneration associated with oxidative stress (Veljkovic et al, 2018). In addition, CGAs are currently being explored in the food and drugs industries as a natural precaution in food additive to substitute the synthetic antibiotics towards reducing medicinal cost (Naveed et al, 2018).

1.8.1 Coffee waste and by-products

A by-product is a secondary product obtained incidentally in the manufacturing process of the main product. More than 50% of the coffee berry is not used for green coffee commercialization and is discarded during processing (Esquivel & Jiménez, 2012). Coffee by-products comprise husks, skin and pulp, coffee mucilage, coffee parchment, coffee silver-skin and spent coffee grounds (Benitez et al, 2019). Coffee by-products could be applied as base or substrate for value adding applications. Currently, coffee by-product is applicable in biofuel, mushroom and fertilizer production, along with enzyme, dietary fibre, and bioactive compound extraction (Ballesteros, Teixeira & Mussatto, 2014; Janissen & Huynh, 2018).

1.8.2 Multivariate statistical analysis

Multivariate statistical analysis refers to the different statistical methods used in analyzing more complex and extensive datasets. It is associated with multivariate probability distributions, which involves observation and analysis of more than one statistical outcome variable at a time. It is very useful in addressing the conditions where multiple measurements are made on each experimental unit and the relations among these measurements.

This statistical method aids to obtain a summary or an overview of a dataset in a table. This is otherwise known as principal components analysis or factor analysis. With this, it is possible to identify the dominant patterns in the data, such as groups, classes, outliers, trends and so on. The patterns are mostly shown as two plots (a score and a loading plot). The score plot is a scatter plot, which contains the original data in a rotated coordinate system. While the values that the spectra have in the principal component (PC) coordinate system are called scores. The loading plot as the coefficients of the linear combination of the initial variables from which the principal components are constructed. In addition, from a numerical perspective, the loadings are equivalent to the variables coordinates divided by the square root of the

eigenvalue (each of a set of values of a parameter for which a differential equation has a non-zero solution) associated with the component. Thus, it illustrates how strongly each characteristic influences a principal component and this could be considered as the weights for each original variable when calculating the PC. Furthermore, multivariate statistical analysis could analyze the groups in a table, shows how the groups differ and to which individual group does the rows in the table belong. This is known as discriminant analysis. It is a method of group classification. This system establishes whether a set of variables can be used to distinguish between two or more groups of cases. Moreover, this analysis could be carried out using different analytical methods such as PCA, PLS-DA, LDA etc. (Brereton & Lloyd, 2018).

1.8.3 Principal Component Analysis (Unsupervised method)

Principal Component Analysis (PCA) is an analytical tool employed in analyzing datasets to visually summarize their main characteristics in form of plots. It is an unsupervised multivariate statistical tool, which could be defined as an orthogonal linear transformation that converts the data to a new coordinate system such that the greatest variance by some scalar projection of the data comes to lie on the first coordinate known as the first principal component (PC) and the second greatest variance on the second coordinate (Jolliffe, 2002).

PCA is used to reduce the dimensionality of a dataset, while retaining as much information as possible. The information referred to is known as the *variance*. The logic behind this is to generate uncorrelated artificial variables called *principal components* (PCs) that combine in a linear way the original (possibly correlated) variables such as metabolites, genes, etc (Haipeng & Huang, 2008). In using PCA, data dimension reduction is attained by projecting the data into space or area called the principal components (PC). This means that each sample is assigned a score on each new PC dimension - this score is calculated as a linear combination of the original variables to which a weight is applied. The weights of each of the original variables are stored in the so-called *loading vectors* associated to each PC. The

dimension of the data is reduced by projecting the data into the smaller sub-space spanned by the PCs, while capturing the largest sources of variation between samples. The principal components are obtained so that their variance is maximized. In view of this, the eigenvectors/eigenvalues of the variance-covariance matrix are calculated through singular value decomposition when the number of variables is very large (Jolliffe, 2005).

Moreover, the data are usually centered and sometimes scaled in the method. The latter is specifically advised in the case where the variance is not homogeneous across variables. The first PC is defined as the linear combination of the original variables that explains the greatest amount of variation. The second PC is then defined as the linear combination of the original variables that accounts for the greatest amount of the remaining variation subject of being orthogonal or uncorrelated to the first component. Subsequent components are defined correspondingly for the other PCA dimensions (Haipeng & Huang, 2008).

1.8.3 Partial least squares-discriminant analysis (Supervised method)

Partial least squares-discriminant analysis (PLS-DA) is one of the most famous classification procedures in chemometrics. It is a linear classification model and a multipurpose algorithm that can be used for predictive and descriptive modelling as well as for discriminative variable selection. It is a supervised multivariate statistical model. This approach combines dimensionality reduction and discriminant analysis into one algorithm and specifically applicable to modelling high-dimensional (HD) data (Barker & Rayens, 2003). This method has been broadly employed in “omics” related fields, such as metabolomics, proteomics, genomics (Boulesteix, & Strimmer, 2007) and in numerous other fields, which generates large amounts of data such as spectroscopy (Blekherman et al, 2011; Gromski et al, 2015).

In addition, PLS-DA is a chemometrics technique used in optimizing separation between different groups of samples, which is achieved by linking two data matrices X (raw data) and

Y (groups, classes, etc.). This method aims to maximize the covariance between the independent variables X (sample readings), which is the metabolomics data and the corresponding dependent variable Y (classes, groups), which is the targets to be predicted from a highly multidimensional data by finding a linear sub-space of the explanatory variables. This new subspace permits the prediction of the Y variable based on a reduced number of factors that is the PLS components otherwise refers to as latent variables. These factors describe the behavior of dependent variables Y and they span the subspace onto which the independent variables X are projected (Gromski et al, 2015).

The main advantage of this PLS-DA approach is the availability and handling of highly collinear and noisy data, which are very common outputs from metabolomics experiments (Want & Masson, 2011). It can also be applied to reduce the dimensionality of the data and facilitate biological interpretation, which could also improve the design of subsequent experiments. Moreover, PLS-DA does not adopt the data to fit a particular distribution and hence is more flexible than other discriminant algorithms like the linear discriminant analysis (LDA) (Lee et al, 2018). Furthermore, it has demonstrated great success in modelling high-dimensional datasets for diverse purposes, e.g. product authentication in food analysis, diseases classification in medical diagnosis and evidence analysis in forensic science (Pérez-Enciso & Tenenhaus, 2003).

1.8.4 Box plot

It is a descriptive non-parametric analytical method of showing variation in samples of a statistical population without any assumptions of the basic statistical distribution. The plot is a standardized system of presenting the dataset in a summarized form based on a five-number cores, which are the sample minimum, median, maximum, the first and third quartiles.

This method graphically depicts the groups of numerical data via their quartiles, which consists of two parts, a box and a set of whiskers. The plot may also have lines spreading from the boxes (*whiskers*) indicating variability outside the lower and upper quartiles. It can be drawn and represented either vertically or horizontally, while the spacings between the box different parts signify the degree of the data dispersion (spread) and skewness, and the outliers. In addition, it provides a visual means of estimating the various L-estimators, particularly the interquartile range (IQR), mid-hinge, range, mid-range and trimean. The minimum part represents the lowest data point excluding any outliers, with the maximum being the largest data point excluding any outliers and the median (Q_2) which is the 50th percentile indicates the middle value of the dataset. Furthermore, the first quartile (Q_1) is the 25th percentile also referred to as *lower quartile* (0.25) is the median of the lower half of the dataset. The third quartile (Q_3) 75th percentile otherwise known as the *upper quartile* (0.75) is the median of the upper half of the dataset and the IQR is the distance between the upper and lower quartiles (Dekking, 2005; Kronthaler, 2014; Mosler & Schmid, 2006).

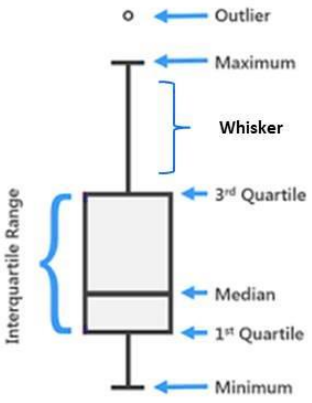


Figure 1.9.2 Typical example of a box-plot.

Heatmap is a graphical representation, which visually shows the relationships between two variables, with one variable plotted on each axis (y-axis and x-axis) by observing how cell colours change across each axis to observe any pattern in numerical value for one or both variables. The map is plotted from the Pearson correlation coefficients numerical data.

1.9 Aim of the Study

CGAs constitute an essential component of the human diet with daily intake of approximately 2 g. Similarly, triacylglycerols play key roles in several human body physiological systems and metabolic processes along with the fatty acids, which have some of it types as essential for normal body physiological functions. However, these listed essential food components have extensively been investigated in the past with the aid of a standard LC-MS^n method developed by the Kuhnert research group, which has been widely published and cited in numerous articles and books; likewise, the use of tandem mass spectrometry for triacylglycerol profiling and gas chromatography for fatty acids constituents of foods and beverages analysis and FRAP assay will be explored to investigate the following project aims:

i. Discriminate between Arabica and Robusta coffee using the LC-MS^n profile of CGAs and their regio-isomers variations.

ii. Develop and optimize extraction, separation and data analytical methods for coffee lipids constituents, most importantly the triacylglycerol.

iii. Identify new triacylglycerols with the optimized method.

iv. Discriminate between Arabica and Robusta coffee using LC-MS^n profile of the TAG composition extracted from lipids and likewise the fatty acids constituents by GC.

v. Explore the CGAs profile to classify coffee cultivated according to their different agricultural practices and their geographical origins with multivariate statistical analysis.

vi. Study the degradation rate in coffee over a period of time during storage.

vii. Study the antioxidant degradation in coffee considering the potential chemical changes during storage.

viii. Explore the use of chemometrics to statistically analysis the data acquired from the above itemized aims.

REFERENCES

1. **Abrankó**, L., & Clifford, M. N. (2017). An Unambiguous Nomenclature for the Acyl-quinic Acids Commonly Known as Chlorogenic Acids. *Journal of Agricultural and Food Chemistry, 65(18), 3602–3608*.
2. **Acuna**, R., Bassuner, R., Beilinson, V., Cortina, H., Cadena-Gomez, G., Montes, V., & Nielsen, N. C. (1999). Coffee seeds contain 11S storage proteins. *Physiologia Plantarum, 105(1), 122–131*.
3. **Albuquerque**, T. G., Nunes, M. A., Bessada, S. M. F., Costa, H. S., & Oliveira, M. B. P. P. (2020). Biologically active and health promoting food components of nuts, oilseeds, fruits, vegetables, cereals, and legumes. *Chemical Analysis of Food, 609–656.*
4. **Agudelo-Ochoa** G. M., Pulgarín-Zapata I. C., Velásquez-Rodriguez C. M., Duque-Ramirez M., Naranjo-Cano M., Quintero-Ortiz M. M., Lara-Guzmán O. J., Muñoz-Durango K. (2016). Coffee consumption increases the antioxidant capacity of plasma and has no effect on the lipid profile or vascular function in healthy adults in a randomized controlled trial *J Nutr.* 146(3):524-31.
5. **Alves**, R. C., Casal, S., Alves, M. R., & Oliveira, M. B. (2009). Discrimination between arabica and Robusta coffee species on the basis of their tocopherol profiles. *Food Chemistry, 114(1), 295–299.*
6. **Anthony**, F., Combes, M. C., Astorga, C., Bertrand, B., Graziosi, G., & Lashermes, P. (2002). The origin of cultivated *Coffea arabica L*. varieties revealed by AFLP and SSR markers. *Theoretical and Applied Genetics*, 104(5), 894-900.
7. **Antonio**, A. G., Moraes, R. S., Perrone, D., Maia, L. C., Santos, K. R. N., Iorio, N. L. P., Farah, A. (2010). Species, roasting degree and decaffeination influence the antibacterial activity of coffee against *Streptococcus mutans*. *Food Chem*., 118, 782–788.
8. **Ayseli**, M.T. & Ayseli, Y.İ. (2016). Flavours of the future: health benefits of flavour precursors and volatile compounds in plant foods, *Trends Food Sci. Technol*. 48: 69-77.
9. **Badmos**, S., Fu, M., Granato, D., & Kuhnert, N. (2020). Classification of Brazilian roasted coffees from different geographical origins and farming practices based on chlorogenic acid profiles. *Food Research International, 109218.*
10. **Badmos**, S., Lee, S. H., & Kuhnert, N. (2019). Comparison and quantification of chlorogenic acids for differentiation of green Robusta and Arabica coffee beans. *Food Research International, 108544.*

11. **Bajko** E., Kalinowska M., Borowski P., Siergiejczyk L., Lewandowski W., (2016). 5-O-caffeoylquinic acid: a spectroscopic study and biological screening for antimicrobial activity, *LWT-Food Sci. Technol*. 65 471–479.
12. **Ballesteros**, L. F., Teixeira, J. A., & Mussatto, S. I. (2014). Chemical, functional, and structural properties of spent coffee grounds and coffee silver skin. *Food and Bioprocess Technology*, 7(12), 3493–3503.
13. **Barker**, M., & Rayens, W. (2003). Partial Least Squares for Discrimination. *Journal of Chemometrics* 17 (3). Wiley Online Library: 166–73.
14. **Bekedam**, E. K., Loots, M. J., Schols, H. A., Van Boekel, M. A. J. S., & Smit, G. (2008). Roasting effects on formation mechanisms of coffee brew melanoidins. *Journal of Agricultural and Food Chemistry*, 56(16), 7138-7145.
15. **Bekedam**, E. K., Schols, H. A., Boekel, M. A. J. S., & Smit, G. (2006). High molecular weight melanoidins from coffee brew. *Journal of Agricultural and Food Chemistry*, 54, 7658–7666.
16. **Bekedam**, E. K., Schols, H. A., Boekel, M. A. J. S., & Smit, G. (2008). Incorporation of chlorogenic acids in coffee brew melanoidins. *Journal of Agricultural and Food Chemistry*, 56, 2055–2063.
17. **Benitez**, V., Rebollo-Hernanz, M., Hernanz, S., Chantres, S., Aguilera, Y., & Martin-Cabrejas, M. A. (2019). Coffee parchment as a new dietary fibre ingredient: Functional and physiological characterization. *Food Research International. 122, 105-113.*
18. **Benzie** I. F. F. & Strain J. J. (1996). Ferric reducing ability of plasma (FRAP) as a measure of antioxidant power: The FRAP assay. *Analytical Biochemistry* 239(1):70-76.
19. **Berrueta** L.A., Alonso-Salces R.M., Heberger K. (2007). Supervised pattern recognition in food analysis, *J. Chromatogr. A*, 1158: 196-214.
20. **Bicchi** C. P., Binello A. E., Pellegrino G. M., Vanni A. C. (1995) Characterization of green and roasted coffees through the chlorogenic acid fraction by HPLC-UV and principal component analysis. *Journal of Agricultural and Food Chemistry*, 43 pp. 1549-1555.
21. **Bicho**, N. C., Leitão, A. E., Ramalho, J. C., de Alvarenga, N. B., & Lidon, F. C. (2013). Identification of Chemical Clusters Discriminators of Arabica and Robusta Green Coffee *International Journal of Food Properties*, 16(4), 895–904.
22. **Blekherman**, G., Laubenbacher, R., Cortes, D. F., Mendes, P., Torti, F. M., Akman, S., Shulaev, V. (2011). Bioinformatics tools for cancer metabolomics. *Metabolomics, 7(3), 329–343.*

23. **Borrelli**, R.C., Visconti, A., Mennella, C., Anese, M., Fogliano, V. (2002). Chemical characterization and antioxidant properties of coffee melanoidins. *J. Agric. Food Chem*. 50, 6527–6533.

24. **Boulesteix**, A. L. & Strimmer K. (2007). Partial least squares: a versatile tool for the analysis of high-dimensional genomic data *Brief. Bioinf.*, 8 (2007), pp. 32-44.

25. **Brereton**, R. G., & Lloyd, G. R. (2018). Partial least squares discriminant analysis for chemometrics and metabolomics: How scores, loadings, and weights differ according to two common algorithms. *Journal of Chemometrics, 32(4), e3028.*

26. **Brglez** M., E., Knez H., M., Škerget, M., Knez, Ž. and Bren, U. (2016). Polyphenols: extraction methods, antioxidative action, bioavailability and anticarcinogenic effects. *Molecules, 21(7): 901.*

27. **Capuano**, E., & Fogliano, V. (2011). Acrylamide and 5-hydroxymethylfurfural (HMF): A review on metabolism, toxicity, occurrence in food and mitigation strategies. *LWT - Food Science and Technology, 44(4), 793–810.*

28. **Carrera** F., León-Camacho M., Pablos F., González A. G. (1998). Authentication of green coffee varieties according to their sterolic profile *Analytica Chimica Acta*, 370 pp.131-139.

29. **Casal**, S., Alves, M. R., Mendes, E., Oliveira, M. B. P. P., & Ferreira, M. A. (2003). Discrimination between arabica and Robusta coffee species on the basis of their amino acid enantiomers. *Journal of Agricultural and Food Chemistry,* 51(22), 6495-6501.

30. **Casal**, S., Mendes, E., Alves, M. R., Alves, R. C., Beatriz, M., Oliveira, P. P., & Ferreira, M. A. (2004). Free and conjugated biogenic amines in green and roasted coffee beans. *Journal of Agricultural and Food Chemistry, 52(20),* 6188–6192.

31. **Casal**, S., Oliveira, M. B. P. P., Alves, M. R., & Ferreira, M. A. (2000). Discriminate analysis of roasted coffee varieties for trigonelline, nicotinic acid, and caffeine content. *Journal of Agricultural and Food Chemistry, 48(8),* 3420–3424.

32. **Cheng**, B., Furtado, A., & Henry, R. J. (2018). The coffee bean transcriptome explains the accumulation of the major bean components through ripening. *Scientific Reports, 8(1).*

33. **Chiang**, L., Chiang, W., Chang, M., Ng, L., and Lin, C. (2002). Antiviral activity of Plantago major extracts and related compounds in vitro. *Antiviral Research,* 55(1),53-62.

34. **Cho**, A. S., Jeon, S. M., Kim, M. J., Yeo, J., Seo, K. I., Choi, M. S., & Lee, M. K. (2010). Chlorogenic acid exhibits anti-obesity property and improves lipid metabolism in high-fat diet-induced-obese mice. *Food and Chemical Toxicology*, 48(3), 937–943.

35. **Chong**, J., Wishart, D.S. and Xia, J. (2019). Using MetaboAnalyst 4.0 for comprehensive and integrative metabolomics data analysis. *Current protocols in bioinformatics* 68, e86. https://www.metaboanalyst.ca/MetaboAnalyst/home.xhtml. Accessed 9 November 2019.

36. **Cizkova**, H., Soukupova, V., Voldrich, M., & Sevcik, R. (2007). Differentiation of coffee varieties according to their sterolic profile. *Journal of Food and Nutrition Research*, 46, 28–34.

37. **Clifford** M. N., & Abranko, L (2017). *Researchgate*. https://www.researchgate.net/publication/312647842.

38. **Clifford** M. N., Ramirez-Martinez J. R. (1991) Phenols and caffeine in wet-processed coffee beans and coffee pulp. *Food Chem*. 40:35-42.

39. **Clifford**, M. N. (1999). Chlorogenic acids and other cinnamates - nature, occurrence and dietary burden. *Journal of the Science of Food and Agriculture*, 79(3), 362–372.

40. **Clifford**, M. N. (2000). Chlorogenic acids and other cinnamates-nature, occurrence, dietary burden, absorption and metabolism, *J. Sci. Food Agric*., 80, 1033–1043.

41. **Clifford**, M. N., & Ramirez-Martinez, J. R. (1990). Chlorogenic acids and purine alkaloids contents of Maté (*Ilex paraguariensis*) leaf and beverage. *Food Chemistry*, 35(1), 13–21.

42. **Clifford**, M. N., Johnston, K. L., Knight, S., & Kuhnert, N. (2003). Hierarchical scheme for LC-MS^n identification of chlorogenic acids. *Journal of Agricultural and Food Chemistry*, 51(10), 2900 - 2911.

43. **Clifford**, M. N., Johnston, K. L., Knight, S., Kuhnert, N. (2003). The characterisation by LC-MS^n of coffee seed caffeoylferuloylquinic acids. *J. Agric. Food Chem* 51, 2900–2911.

44. **Clifford**, M. N., Kerimi, A., & Williamson, G. (2020). Bioavailability and metabolism of chlorogenic acids (acyl-quinic acids) in humans. *Comprehensive Reviews in Food Science and Food Safety*. doi:10.1111/1541-4337.12518.

45. **Clifford**, M. N., Knight, S., Surucu, B., & Kuhnert, N. (2006). Characterization by LC-MS^n of four new classes of chlorogenic acids in green coffee beans: dimethoxycinnamoylquinic acids, diferuloylquinic acids, caffeoyl-dimethoxycinnamoylquinic acids, and feruloyl-dimethoxycinnamoylquinic acids. *Journal of Agricultural and Food Chemistry*, 54(6), 1957–1969.

46. **Clifford**, M. N., Zheng, W., & Kuhnert, N. (2006). Profiling the chlorogenic acids of aster by HPLC-MS^n. *Phytochemical Analysis, 17(6), 384–393.*
47. **Clifford**, M. N., Kirkpatrick, J., Kuhnert, N., Roozendaal, H., & Salgado, P. R. (2008). LC–MS^n analysis of the *cis* isomers of chlorogenic acids. *Food Chemistry*, 106(1), 379–385.
48. **Clifford**, M. N., Wu, W., Kirkpatrick, J., & Kuhnert, N. (2007). Profiling the Chlorogenic Acids and Other Caffeic Acid Derivatives of Herbal *Chrysanthemum* by LC−MS^n. *Journal of Agricultural and Food Chemistry*, 55(3), 929–936.
49. **Clifford**, M. N., Jaganath, I. B., Ludwig, I. A., & Crozier, A. (2017). Chlorogenic acids and the acyl-quinic acids: discovery, biosynthesis, bioavailability and bioactivity. *Natural Product Reports, 34(12), 1391–1421.*
50. **Cowie**, J. M. G., Arrighi, V. (2008). Polymers: chemistry and physics of modern materials (3rd edition) Scotland: *CRC Press*. pp: 9813-4.
51. **Cossignani**, L., Montesano, D., Simonetti, M. S., & Blasi, F. (2016). Authentication of *Coffea arabica* according to triacylglycerol stereospecific composition. *Journal of Analytical Methods in Chemistry,* 2016: 1 – 7.
52. **Crawford**, J. (1852). History of coffee. *Journal of the Statistical Society of London*, 15(1), 50.
53. **Cropley** V., Croft R., Silber B., Neale C., Scholey A., Stough C., Schmitt J. (2012). Does coffee enriched with chlorogenic acids improves mood and cognition after acute administration in healthy elderly? A pilot study *Psychopharmacology* 219(3):737-49.
54. **Crozier** A., Yokota T., Jaganath I. B. Marks S. C., Saltmarsh M. and Clifford M. N., (2006) Plant secondary metabolites. occurrence, structure and role in the human diet, A. Crozier, M. N. Clifford and H. Ashihara, *Wiley Blackwell Publishing*, Oxford, UK, pp. 208-302.
55. **Crozier**, A., Jaganath, I. B., & Clifford, M. N. (2009). Dietary phenolics: chemistry, bioavailability and effects on health. *Natural Product Reports, 26(8), 1001.*
56. **Daglia**, M., Papetti, A., Gregotti, C., Bertè, F., & Gazzani, G. (2000). In vitro antioxidant and ex vivo protective activities of green and roasted coffee. *Journal of Agricultural and Food Chemistry, 48(5), 1449–1454.*
57. **dePaula** and Farah. (2019). Caffeine Consumption through Coffee: Content in the Beverage, Metabolism, Health Benefits and Risks. *Beverages 5(2), 37.*

58. **Del Campo**, G., Berregi, I., Caracena, R., & Zuriarrain, J. (2010). Quantitative determination of caffeine, formic acid, trigonelline and 5-(hydroxymethyl) furfural in soluble coffees by ^{1}H-NMR spectrometry. *Talanta*, 81(1-2), 367–371.

59. **Dekking**, F. M. (2005). A Modern Introduction to Probability and Statistics. *Springer*. pp. *234*–238. ISBN 1-85233-896-2.

60. **Demirbag**, D., Ozdemir, F., Ture, M. (2006). Effects of coffee consumption and smoking habit on bone mineral density, *Rheumatol*. *Int*. 26, 530–535.

61. **Dentan**, E. (1987) The microscopic structure of the coffee bean. In: Coffee, Botany, Biochemistry and Production of Beans and Beverage; Clifford, M.N.; Wilson, K.C.; Eds.; *Croom Helm & Methuen Inc*: New York, 284–304.

62. **Deshpande**, S., Matei, M. F., Jaiswal, R., Bassil, B. S., Kortz, U., & Kuhnert, N. (2016). Synthesis, structure, and tandem mass spectrometric characterization of the diastereomers of quinic acid. *Journal of Agricultural and Food Chemistry, 64(38), 7298-7306.*

63. **Deshpande**, S., Jaiswal, R., Matei, M. F., & Kuhnert, N. (2014). Investigation of acyl migration in mono- and di-caffeoylquinic acids under aqueous basic, aqueous acidic and dry roasting conditions. *Journal of Agricultural and Food Chemistry, 62(37), 9160–9170.*

64. **Dirinck**, M. I., Van Leuven, I., & Dirinck, J. P. (2001). Analytical flavour characterisation and classification of Arabica and Robusta coffees from different origins. In *Proceedings of 19th ASIC Colloquium (Trieste)*. Paris.

65. **Duarte**, S. M. da S., Abreu, C. M. P. de, Menezes, H. C. de, Santos, M. H. dos, & Gouvêa, C. M. C. P. (2005). Effect of processing and roasting on the antioxidant activity of coffee brews *Ciência e Tecnologia de Alimentos, 25(2), 387–393.*

66. **Duke**, J. (1983). Handbook of energy crops. Accessed 15 April 2020. https://hort.purdue.edu/newcrop/duke_energy/Coffea_arabica.html

67. **Esquivel**, P., & Jiménez, V. M. (2012). Functional properties of coffee and coffee by-products. *Food Research International*, 46(2), 488–495.

68. **Fahy**, E., Subramaniam, S., Murphy, R. C., Nishijima, M., Raetz, C. R. H., Shimizu, T., Dennis, E. A. (2009). Update of the LIPID MAPS comprehensive classification system for lipids. Journal of Lipid Research, 50(SUPPL.), S9-S14.

69. **FAO** Report. (2016) *http://www.fao.org/docrep/006/y5143e/y5143e0v.html* Accessed 28. March 2020.

70. **FAO** Report. (2019). The role of agriculture in the development of least-developed countries and their integration into the world economy. http://www.fao.org/3/a-y3997e.pdf. Accessed 19. January 2020.

71. **Farah** A. (2012). Coffee constituents *Coffee: Emerging Health Benefits and Disease Prevention* Yi-Fang Chu (Editor). ISBN: 978-1-119-94989-3. *Wiley-Blackwell Online Library* 352 Pages 21-58.

72. **Farah**, A. (2004). Distribuicao nos graos, influencia sobre aqualidade da bebidae biodisponibilidade dos acidos clorogenicos do cafe. Instituto de Qu'ımica, Universidade Federal do Rio de Janeiro, RJ, *Doctorate Thesis*, Brazil.

73. **Farah**, A., de Paulis, T., Trugo, L. C., Martin, P. R. (2005) Effect of roasting on the formation of chlorogenic acid lactones. *J. Agric. Food Chem*., 53, 1505–1513.

74. **Farah**, A., de Paulis, T., Trugo, L. C., Martin, P. R. (2006). Chlorogenic acids and lactones in regular and water-decaffeinated Arabica coffee. *J. Agric. Food Chem*., 54, 374-381.

75. **Farzaei**, M. H. (2015). Role of dietary polyphenols in the management of peptic ulcer. *World Journal of Gastroenterology*, 21(21), 6499.

76. **Fauconnot**, L., Hau, J., Aeschlimann, J.-M., Fay, L.-B., & Dionisi, F. (2004). Quantitative analysis of triacylglycerol regioisomers in fats and oils using reversed-phase high-performance liquid chromatography and atmospheric pressure chemical ionization mass spectrometry. *Rapid Communications in Mass Spectrometry, 18(2), 218–224.*

77. **Fiamegos**, Y. C., Kastritis, P. L., Exarchou, V., Han, H., Bonvin, A. M. J. J., Vervoort, J., Tegos, G. P. (2011). Antimicrobial and efflux pump inhibitory activity of caffeoylquinic acids from *Artemisia absinthium* against Gram-Positive Pathogenic Bacteria. *PLoS ONE*, 6(4), e18127.

78. **Flament**, I., Gautschi, F., Winter, M., Willhalm, B., Stoll, M (1968). Les composants furanniques de l'arome cafe: quelques aspects chimiques et spectroscopiques. Proc. *3rd Coll. Int. Coffee Sci*. ASIC, 197–215. Paris.

79. **Fogliano** V., Morales F. J. (2011). Estimation of dietary intake of melanoidins from coffee and bread. *Food Funct*. 2:117–123.

80. **Folmer**, B., Blank, I., & Hofmann, T. (2017). Crema-formation, stabilization, and sensation. *The Craft and Science of Coffee*, 399-417.

81. **Folstar**, P. (1985). Lipids. In: Coffee, Clarke, R. J., Macrae, R., eds. *London: Elsevier Applied Science*, Vol. 1: Chemistry, pp. 203–222.

82. **Frankel**, E. N. (1983). Volatile lipid oxidation products *Progress in Lipid Research, 22(1), 1–33.*

83. **Frankel**, E. N. (1984). Lipid oxidation: Mechanisms, products and biological significance. *Journal of the American Oil Chemists' Society*, 61(12), 1908–1917.

84. **Frankel**, E. N. (1991). Review: Recent advances in lipid oxidation. *Journal of the Science of Food and Agriculture*, 54(4), 495–511.

85. **Fu**, L., Lu, W., & Zhou, X. (2016). Phenolic compounds and in vitro antibacterial and antioxidant activities of three tropic fruits: persimmon, guava and sweetsop. *BioMed Research International*, 2016, 1–9.

86. **Fuster**, M. D., Mitchell, A. E., Ochi, H., & Shibamoto, T. (2000). Antioxidative activities of heterocyclic compounds formed in brewed coffee. *Journal of Agricultural and Food Chemistry*, 48(11), 5600–5603.

87. **Gamaleldin Elsadig Karar**, M., Matei, M., Jaiswal, R., Illenberger, S., & Kuhnert, N. (2016). Neuraminidase inhibition of dietary chlorogenic acids and derivatives-potential antivirals from dietary sources *Food and Function*, 7(4), 2052–2059.

88. **German**, J. B. (1999). Food processing and lipid oxidation. *Impact of Processing on Food Safety*, 23–50.

89. **Geleta**, M., Herrera, I., Monzón, A., & Bryngelsson, T. (2012). Genetic diversity of Arabica coffee (*Coffea arabica L.*) in Nicaragua as estimated by simple sequence repeat markers. *The Scientific World Journal*, 2012, 1–11.

90. **Gniechwitz**, D., Reichardt, N., Ralph, J., Blaut, M., Steinhart, H., Bunzel, M. (2008). Isolation and characterisation of a coffee melanoidin fraction. *J. Sci. Food Agric*., 88, 2153–2160.

91. **Gomez-Ruiz** A, Ames M., Leake S. (2008). Antioxidant activity and protective effects of green and dark coffee components against human low-density lipoprotein oxidation, *Eur Food Res. Technol*., 22, 1017-1024.

92. **González**, A. G., Pablos, F., Martín, M. J., León-Camacho, M., & Valdenebro, M. S. (2001). HPLC analysis of tocopherols and triglycerides in coffee and their use as authentication parameters. *Food Chemistry, 73(1),* 93–101.

93. **Gromski**, P. S., Muhamadali, H., Ellis, D. I., Xu, Y., Correa, E., Turner, M. L., & Goodacre, R. (2015). A tutorial review: Metabolomics and partial least squares-discriminant analysis – a marriage of convenience or a shotgun wedding. *Analytica Chimica Acta, 879, 10–23.*

94. **Gunalan** G., Myla N., Balabhaskar R. (2012) *In vitro* antioxidant analysis of selected coffee bean varieties. *J. Chem. Pharm. Res.* 4:2126-2132.

95. **Gunning**, Y., Defernez, M., Watson, A. D., Beadman, N., Colquhoun, I. J., Le Gall, G., Kemsley, E. K. (2018). 16-*O*-methylcafestol is present in ground roast Arabica coffees: Implications for authenticity testing. *Food Chemistry*, 248, 52–60.

96. **Jacobs**, S., Kröger, J., Floegel, A., Boeing, H., Drogan, D., Pischon, T., Schulze, M. B. (2014). Evaluation of various biomarkers as potential mediators of the association between coffee consumption and incident type-2 diabetes in the EPIC-Potsdam Study. *The American Journal of Clinical Nutrition, 100(3), 891–900.*

97. **Jiang**, X., Zhang, D., & Jiang, W. (2013). Coffee and caffeine intake and incidence of type 2 diabetes mellitus: a meta-analysis of prospective studies. *European Journal of Nutrition, 53(1), 25–38.*

98. **Hameed**, A., Hussain, S. A., Ijaz, M. U., Ullah, S., Pasha, I., & Suleria, H. A. R. (2018). Farm to Consumer: Factors Affecting the Organoleptic Characteristics of Coffee. II: Postharvest Processing Factors. *Comprehensive Reviews in Food Science and Food Safety, 17(5), 1184–1237.*

99. **Hemmerle**, H., Burger, H. J., Below, P., Schubert, G., Rippel, R., Schindler, P. W., Herling, A. W. (1997). Chlorogenic acid and synthetic chlorogenic acid derivatives: novel inhibitors of hepatic glucose-6-phosphate translocase. *Journal of Medicinal Chemistry, 40(2),* 137–145.

100. **Hermann** K. (1989). Occurrence and content of hydroxycinnamic and hydroxybenzoic acid compounds in foods, *Crit. Rev. Food Sci. Nutr*., 28, 315-347.

101. **Ho**, C. T., & Chen, Q. (1994). Lipids in food flavors. *Lipids in Food Flavours, 2–14.*

102. **Hodge**, J. E. (1953). Dehydrated foods, chemistry of browning reactions in model systems. *Journal of Agricultural and Food Chemistry*, 1(15), 928–943.

103. **Hofmann**, T., Czerny, M., Calligaris, S., Schieberle, P. (2001). Model studies on the influence of coffee melanoidins on flavour volatiles of coffee beverages. *J. Agric. Food Chem.*, 49 (5), 2382– 2386.

104. **Hohler**, D. (1998). Ochratoxin A in food and feed: occurrence, legislation and mode of action. *Zeitschrift fur Ernahrungswissenschaft*, 37, 2–12.

105. **Huang**, K., Liang, X., Zhong, Y., He, W., & Wang, Z. (2014). 5-*O*-caffeoylquinic acid decreases diet-induced obesity in rats by modulating PPARα and LXRα transcription. *Journal of the Science of Food and Agriculture, 95(9), 1903–1910.*

106. **Huang**, L., Wang, X., Dong, Y., Long Y., Hao C., Yan L., Shi T. (2020). Resequencing 93 accessions of coffee unveil independent and parallel selection during Coffea species divergence. *Plant Mol. Biol.* 103, 51–61.

107. **Huck** C. W., Guggenbichler W., Bonn G. K. (2005). Analysis of caffeine, theobromine and theophylline in coffee by near infrared spectroscopy (NIRS) compared to HPLC coupled to mass spectrometry. *Analytica Chimica Acta*, 538, 195–203.

108. **IARC** (1993). Ochratoxin A. Monographs on the evaluation of carcinogenic risk to humans. Lyon, France, *IARC Working Group, World Health Organization*. pp:250–277.

109. **Illy** A., Illy E, Macrae R., Petracco M., Sondahl M. R., Valussi S. & Viani R., (1995). Espresso coffee: the chemistry of quality, *Academic Press*, London.

110. **Iriondo-Dehond**, A., Santillan Cornejo, F., Fernandez-Gomez, B., Vera, G., Guisantes-Batan, E., Alonso, S. G., Del Castillo, M. D. (2019). Bio-accesibility, metabolism and excretion of lipids composing spent coffee grounds. *Nutrients*, 11:6.

111. **Jaiswal** R., Patras M. A., Eravuchira P. J., Kuhnert N. (2010). Profile and characterization of the chlorogenic acids in green Robusta coffee beans by LC-MS^n: identification of seven new classes of compounds *J Agric Food Chem* 58(15):8722-37.

112. **Jaiswal**, R., and Kuhnert, N. (2010). Hierarchical scheme for liquid chromatography/multi-stage spectrometric identification of 3,4,5-triacyl chlorogenic acids in green Robusta coffee beans. *Rapid Communications in Mass Spectrometry, 24(15), 2283–2294.*

113. **Jaiswal**, R., Kuhnert, N. (2011). Identification and characterization of five new classes of chlorogenic acids in burdock (*Arctium lappa L.*) roots by liquid chromatography/tandem mass spectrometry. *Food Funct*. 2, 63−71.

114. **Jaiswal**, R., Matei, M. F., Ullrich, F., Kuhnert, N. (2011). How to distinguish between cinnamoylshikimate esters and chlorogenic acid lactones by liquid chromatography−tandem mass spectrometry. *J. Mass Spectrom*. 46, 933−942.

115. **Jaiswal**, R., Deshpande, S.; Kuhnert, N. (2011). Profiling the chlorogenic acids of *Rudbeckiahirta*, *Helianthus tuberosus*, *Carlinaacaulis* and *Symphyotrichum novae-angliae* leaves by LC-MS^n. *Phytochem. Anal.*, 22, 432−441.

116. **Jaiswal**, R., Matei, M. F., Golon, A., Witt, M., & Kuhnert, N. (2012). Understanding the fate of chlorogenic acids in coffee roasting using mass spectrometry based targeted and non-targeted analytical strategies. *Food & Function*, 3(9), 976.

117. **Jaiswal**, R., Matei, M. F., Ullrich, F., Kuhnert, N. (2011). How to distinguish between cinnamoylshikimate esters and chlorogenic acid lactones by liquid chromatography−tandem mass spectrometry. *J. Mass Spectrom*. 46, 933−942.

118. **Jaiswal**, R., Deshpande, S., Kuhnert, N. (2011). Profiling the chlorogenic acids of *Rudbeckiahirta*, *Helianthus tuberosus*, *Carlinaacaulis* and *Symphyotrichum novae-angliae* leaves by LC-MS^n. *Phytochem. Anal.*, 22, 432−441.

119. **Jaiswal**, R., Dickman, M. H., & Kuhnert, N. (2012). First diastereoselective synthesis of methyl caffeoyl- and feruloyl-muco-quinates. *Organic and Biomolecular Chemistry, 10(27), 5266.*

120. **Jaiswal**, R., Karaköse, H., Rühmann, S., Goldner, K., Neumüller, M., Treutter, D., & Kuhnert, N. (2013). Identification of Phenolic Compounds in Plum Fruits (*Prunus salicina L. and Prunus domestica L.*) by High-Performance Liquid Chromatography/Tandem Mass Spectrometry and Characterization of Varieties by Quantitative Phenolic Fingerprints. *Journal of Agricultural and Food Chemistry, 61(49), 12020–12031.*

121. **Janissen**, B., and Huynh, T. (2018). Chemical composition and value-adding applications of coffee industry by-products: A review. Resources, *Conservation and Recycling*, 128, 110-117.

122. **Jham**, G. N., Nikolova-Damyavova, B., Viera, M., Natalino, R., & Rodrigues, A. C. (2003). Determination of the triacylglycerol composition of coffee beans by reverse-phase high-performance liquid chromatography. *Phyt. Analysis*, 14(5), 310-314.

123. **Jin**, S., Chang, C., Zhang, L., Liu, Y., Huang, X., & Chen, Z. (2015). Chlorogenic acid improves late diabetes through adiponectin receptor signalling pathways in *db/db* mice. *PLOS ONE,* 10(4), e0120842.

124. **Jolliffe**, I. T. (2002). Principal component analysis, Series: Springer Series in Statistics (2nd ed.) *Springer*, NY, 2002, XXIX, 487 p.28 ISBN: 978-0-387-95442-4.

125. **Jolliffe**, I. T. (2005). Principal component analysis. *Wiley Online Library.*

126. **Karaköse**, H., Müller, A., & Kuhnert, N. (2015). Profiling and Quantification of Phenolics in *Stevia rebaudiana* Leaves. Journal of Agricultural and Food Chemistry, *63(41), 9188–9198.*

127. **Karaköse**, H., Jaiswal, R., Deshpande, S., & Kuhnert, N. (2015). Investigation of the Photochemical Changes of Chlorogenic Acids Induced by Ultraviolet Light in Model Systems and in Agricultural Practice with *Stevia rebaudiana* Cultivation as an Example. *Journal of Agricultural and Food Chemistry, 63(13), 3338–3347.*

128. **Karunanidhi**, A., Thomas, R., van Belkum, A., & Neela, V. (2013). In vitro antibacterial and antibiofilm activities of chlorogenic acid against clinical isolates of *Stenotrophomonas maltophilia* including the trimethoprim/sulfamethoxazole resistant strain. *BioMed Research International*, 2013, 1–7.

129. **Khan**, M. T. H., Ather, A., Thompson, K. D., & Gambari, R. (2005). Extracts and molecules from medicinal plants against herpes simplex viruses. *Antiviral Research, 67(2)*, 107–119.

130. **Khrisanapant** P., Kebede B., Leong S. Y., & Oey I. (2019). A comprehensive characterisation of volatile and fatty acid profiles of legume seeds. *Foods*, 8(12), 651.

131. **Klein**, A.M., Steffan-Dewenter, I., & Tscharntke, T. (2003). Bee pollination and fruit set of *Coffea arabica* and *C. canephora (Rubiaceae). American Journal of Botany,* 90(1), 153-157.

132. **Kolling-Speer**, L., Speer, K. (2005). The raw seed composition. In: Espresso coffee, the science of quality. Illy, A., Viani, R., eds. Italy: *Elsevier Academic Press*, pp. 148–178.

133. **Kroyer**, G. T., Kretschmer, L., Washuettl, J. (1989). Antioxidant properties of tea and coffee extracts. *Agric. Food Chem. Cons., Proc. Eur. Conf. Food Chem*., 5th, 2, 433- 437.

134. **Kronthaler**, F. (2014). Statistics applied. Data analysis is (not) an art. *Springer-Verlag*, Berlin Heidelberg. ISBN: 978-3-642-53739-4, p.38.

135. **Kumazawa**, K., & Masuda, H. (2003). Investigation of the changes in the flavour of a coffee drink during heat processing. *Journal of Agricultural and Food Chemistry*, 51(9), 2674–2678.

136. **Kuhnert**, N., Drynan, J. W., Obuchowicz, J., Clifford, M. N., Witt, M. (2010). Mass spectrometric characterization of black tea thearubigins leading to an oxidative cascade hypothesis for thearubigin formation. *Rapid Commun. Mass Spectrom*. 24, 3387−3404.

137. **Le-Bars**, J. & Le-Bars, P. (2000). Mycotoxigenesis in grains application to mycotoxic prevention in coffee. *Coffee Biotechnology and Quality*. Pp. 355–368. Sera, Netherlands: Kluwer Academic Publishers.

138. **Lee**, C. (2000). Antioxidant ability of caffeine and its metabolites based on the study of oxygen radical absorbing capacity and inhibition of LDL peroxidation. *Clin. Chim. Acta*., 295, 141–154. 42.

139. **Lee**, L. C., Liong, C.-Y., & Jemain, A. A. (2018). Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. *The Analyst, 143(15), 3526–3539.*

140. **Lelyana**, R. (2017). The functional benefit of coffee for being healthy and beautiful. *Journal of Nanomedicine & Nanotechnology*, 8(6).

141. **Liang**, N., & Kitts, D. (2014). Antioxidant property of coffee components: assessment of methods that define mechanisms of action. *Molecules, 19(11), 19180–19208.*

142. **Liang**, N., & Kitts, D. (2015). Role of chlorogenic acids in controlling oxidative and inflammatory stress conditions. *Nutrients*, 8(1), 16.

143. **Liu**, C. C., Zhang, Y., Dai, B. L., Ma, Y. J., Zhang, Q., Wang, Y., & Yang, H. (2017). Chlorogenic acid prevents inflammatory responses in IL-1β-stimulated human SW-1353 chondrocytes, a model for osteoarthritis. *Molecular Medicine Reports*, 16(2), 1369-1375.

144. **Liu**, Y. & Kitts, D.D. (2011). Confirmation that the Maillard reaction is the principle contributor to the antioxidant capacity of coffee brews. *Food Res. Int*., 44, 2418–2424.

145. **Lísa**, M., & Holčapek, M. (2008). Triacylglycerols profiling in plant oils important in food industry, dietetics and cosmetics using high-performance liquid chromatography-atmospheric pressure chemical ionization mass spectrometry. *Journal of Chromatography A*, 1198-1199(1-2), 115-130.

146. **Lísa**, M., Holčapek, M., & Boháč, M. (2009). Statistical evaluation of triacylglycerol composition in plant oils based on high-performance liquid chromatography - atmospheric pressure chemical ionization mass spectrometry data. *Journal of Agricultural and Food Chemistry*, 57(15), 6888-6898.

147. **Ludwig**, I. A., Clifford, M. N., Lean, M. E. J., Ashihara, H., & Crozier, A. (2014). Coffee: biochemistry and potential impact on health. *Food Function, 5(8), 1695–1717.*

148. **Macrae**, R. (1985). Nitrogenous components in Coffee, *Anonymous, Springer*; pp: 115-152.

149. **Manach**, C., Scalbert, A., Morand, C., Rémésy, C., & Jiménez, L. (2004). Polyphenols: food sources and bioavailability. *The American Journal of Clinical Nutrition, 79(5),* 727-747.

150. **Marchese**, S., Polo, A., Ariano, A., Velotto, S., Costantini, S., & Severino, L. (2018). Aflatoxin B1 and M1: Biological properties and their involvement in cancer development. *Toxins, 10(6), 214.*

151. **Martín** M. J., Pablos F., González A. G. (1999). Characterization of Arabica and Robusta roasted coffee varieties and mixture resolution according to their metal content *Food Chemistry*, 66, pp. 365-370

152. **Martín**, M. J., Pablos, F., González, A. G., Valdenebro, M. S., & León-Camacho, M. (2001). Fatty acid profiles as discriminant parameters for coffee varieties differentiation. *Talanta*, 54(2), 291–297.

153. **Martins**, S. I. F., Jongen, W. M., & van Boekel, M. A. J. (2000). A review of Maillard reaction in food and implications to kinetic modelling. *Trends in Food Science & Technology*, 11(9-10), 364-373.

154. **Marventano**, S., Salomone, F., Godos, J., Pluchinotta, F., Del Rio, D., Mistretta, A., & Grosso, G. (2016). Coffee and tea consumption in relation with non-alcoholic fatty liver and metabolic syndrome: A systematic review and meta-analysis of observational studies. *Clinical Nutrition, 35(6), 1269–1281*.

155. **Matei**, M. F., Jaiswal, R., & Kuhnert, N. (2012). Investigating the chemical changes of chlorogenic acids during coffee brewing: conjugate addition of water to the olefinic moiety of chlorogenic acids and their quinides *Journal of Agricultural and Food Chemistry*, 60(49), 12105–12115.

156. **Matei**, M. F., Jaiswal, R., Patras, M. A., & Kuhnert, N. (2016). LC-MS^n study of the chemical transformations of hydroxycinnamates during yerba maté *(Ilex paraguariensis*) tea brewing. *Food Research International, 90, 307–312.*

157. **Mellbye**, F. B., Jeppesen, P. B., Shokouh, P., Laustsen, C., Hermansen, K., & Gregersen, S. (2017). Cafestol, a bioactive substance in coffee, has antidiabetic properties in *KKAy* Mice. *Journal of Natural Products, 80(8), 2353–2359.*

158. **Meng**, S., Cao, J., Feng, Q., Peng, J., & Hu, Y. (2013). Roles of chlorogenic acid on regulating glucose and lipids metabolism: *A Review. Evidence-Based Complementary and Alternative Medicine*, 2013, 1–11.

159. **Meyers**, H. (2005). Suave molecules of Mocha - coffee, chemistry and civilization. http://newpartisan.squarespace.com/home/suave-molecules-of-mocha-coffee-chemistry-and-civilization.html Retrieved 15 April 2020.

160. **Miller**, E. G., Gonzales-sanders, A. P., Couvillon, A. M., Binnie, W. H., Sunahhara, G. I., Bertholet, R. (1993). ASIC, *15th Colloque, Montpellier; ASIC*: Paris, pp 420-425.

161. **Mikutis**, G., Karaköse, H., Jaiswal, R., LeGresley, A., Islam, T., Fernandez-Lahore, M., & Kuhnert, N. (2013). Phenolic promiscuity in the cell nucleus – epigallocatechin-gallate (EGCG) and theaflavin-3,3′-digallate from green and black tea bind to model cell nuclear structures including histone proteins, double stranded DNA and telomeric quadruplex DNA. *Food Funct., 4(2), 328–337.*

162. **Mondego**, J. M. C., Vidal, R. O., Carazzolle, M. F., Tokuda, E. K., Parizzi, L. P., Costa, G. G. L., Pereira, G. A. G. (2011). An EST-based analysis identifies new genes and reveals distinctive gene expression features of *Coffea arabica* and *Coffea canephora*. *BMC Plant Biology*, 11.

163. **Moon**, J. K., & Shibamoto, T. (2009). Role of roasting conditions in the profile of volatile flavor chemicals formed from coffee beans. *Journal of Agricultural and Food Chemistry*, 57(13), 5823–5831.

164. **Moreira**, A. S. P., Nunes, F. M., Domingues, M. R., & Coimbra, M. A. (2012). Coffee melanoidins: structures, mechanisms of formation and potential health impacts. *Food & Function*, 3(9), 903.

165. **Moreira**, D. P., Monteiro, M. C., Ribeiro-Alves, M., Donangelo, C. M., & Trugo, L. C. (2005). Contribution of chlorogenic acids to the iron-reducing activity of coffee beverages. *Journal of Agricultural and Food Chemistry, 53(5), 1399–1402.*

166. **Mosler**, K. & Schmid, F. (2006). Descriptive statistics and economic statistics. (3rd Ed.). *Springer-Verlag*, Berlin/Heidelberg. ISBN: 978-3-540-37459-6, p. 33.

167. **Nakajima**, M., Tsubouchi, H., Miyabe, M. & Ueno, Y. (1997). Survey of aflatoxin B1 and Occhratoxin A in commercial green coffee beans by high-performance liquid chromatography linked with immunoaffinity chromatography. *Food and Agricultural Immunology*, 9, 77–83.

168. **Natella** F., Nardini M., Gianetti I., Dattilo C., Scaccini C. (2002). Coffee drinking influences plasma antioxidant capacity in humans. *J. Agric. Food Chem*. 50:6211-6216.

169. **National Coffee Association** (2019). https://www.ncausa.org/about-coffee/10-steps-from-seed-to-cup. Accessed 05. June 2020.

170. **National coffee association** (2020). History of coffee. https://www.ncausa.org/About-Coffee/History-of-Coffee. Accessed 15 April 2020.

171. **Naveed**, M., Hejazi, V., Abbas, M., Kamboh, A. A., Khan, G. J., Shumzaid, M., XiaoHui, Z. (2018). Chlorogenic acid (CGA): A pharmacological review and call for further research. *Biomedicine & Pharmacotherapy,* 97, 67–74.

172. **Nestlé**. Coffee. (1991). Documentation Centre, Corporate Affairs Department, *Nestec Ltd*: Vevey, Switzerland, 1991; 1–60

173. **Nunes**, F. M., & Coimbra, M. A. (2009). Role of hydroxycinnamates in coffee melanoidin formation. *Phytochemistry Reviews*, 9(1), 171–185.

174. **Ogita**, S., Uefugi, H., Yamaguchi, Y., Koizumi, N., Sano, H. Producing decaffeinated coffee plants. *Nature* 2003, 423, 823.

175. **Padayachee**, A., Day, L., Howell, K., & Gidley, M. J. (2015). Complexity and health functionality of plant cell wall fibers from fruits and vegetables. *Critical Reviews in Food Science and Nutrition, 57(1), 59–81.*
176. **Pandino** G., Courts, F. L., Lombado S., Lombardo S., Mauromicale G., and Williamson G. (2010). *J. Agric. Food Chem.*, 58, 1026–1031.
177. **Payen** S., (1846). Comptes rendus hebdomadaires des s´eances de *l'Acad´emie des Sciences*, 23, 244–251.
178. **Payen**, S., (1846b). *Ann. Chem. Pharm*., 1846, 60, 286–294.
179. **Pereira,** A. S., Pereira, A. F. M., Trugo, L. C., Neto, F. R. A. (2003). Distribution of quinic acid derivatives and other phenolic compounds in Brazilian propolis. *Z. Naturforsch*. 58c: 590-593.
180. **Pérez-Enciso**, M. & Tenenhaus, M. (2003). Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach. *Human genetics*, 112(5-6): 581-592.
181. **Pimpley**, V., Patil, S., Srinivasan, K., Desai, N., & Murthy, P. S. (2020). The chemistry of chlorogenic acid from green coffee and its role in attenuation of obesity and diabetes. *Preparative Biochemistry & Biotechnology, 1–10.*
182. **Prior**, R. L., Wu, X., & Schaich, K. (2005). Standardized methods for the determination of antioxidant capacity and phenolics in foods and dietary supplements. *Journal of Agricultural and Food Chemistry, 53(10), 4290–4302.*
183. **Reed**, C. F. (1976). Information summaries on 1000 economic plants. *Typescripts submitted to the USDA.* Accessed 27 March 2020.
184. **Ribeiro-Alves**, M., Trugo, L. C., Donangelo, C. (2003). Use of oral contraceptives blunts the calciuric effect of caffeine in young adult women. *J. Nutr*., 133, 393–398.
185. **Rigane** G., Ghazghazi H., Aouadhi C., Salem R. B., Nasr Z. (2016). Phenolic content, antioxidant capacity and antimicrobial activity of leaf from *Pistacia atlantica Nat Prod Res* 3:14.
186. **Rizzi**, G. P. (1997). Chemical structure of coloured Maillard reaction products. *Food Reviews International*, 13(1), 1–28.
187. **Robiquet** P. J., Boutron-Charlard, A., Liebigs J., (1837). *Ann. Chem*., 23, 93–95.
188. **Rochleder**, F., Liebigs J., (1844). *Ann. Chem*., 1844, 50, 224–234.
189. **Rosenberg**, L. (1990). Coffee and tea consumption in relation to the risk of large bowel cancer: a review on epidemiological studies. *Cancer Lett*., 52, 163-171.

190. **Rufián-Henares**, J. A., & Morales, F. J. (2007). Functional properties of melanoidins: In vitro antioxidant, antimicrobial and antihypertensive activities. *Food Res. Int.*, 40, 995-1002.

191. **Salomone**, F., Galvano, F., & Li Volti, G. (2017). Molecular bases underlying the hepatoprotective effects of coffee. *Nutrients, 9(1), 85.*

192. **Samson**, N., Bausher, M. G., Lee, S.-B., Jansen, R. K., & Daniell, H. (2007). The complete nucleotide sequence of the coffee (Coffea arabica L.) chloroplast genome: organization and implications for biotechnology and phylogenetic relationships amongst angiosperms. *Plant Biotechnology Journal, 5(2), 339–353.*

193. **Santana-Gálvez** J., Cisneros-Zevallos L., Jacobo-Velázquez D. A. (2017). Chlorogenic acid: recent advances on its dual role as a food additive and a nutraceutical against metabolic syndrome, *Molecules* 22 (3), 358.

194. **Schenker**, S., Heinemann, C., Huber, M., Pompizzi, R., Perren, R., & Escher, R. (2002). Impact of roasting conditions on the formation of aroma compounds in coffee beans. *Journal of Food Science*, 67(1), 60–66.

195. **Schievano** E., Finotello C., De Angelis E., Mammi S., Navarini L. (2014). Rapid authentication of coffee blends and quantification of 16-*O*-methylcafestol in roasted coffee beans by nuclear magnetic resonance. *J Agric Food Chem*. 62(51):12309-14.

196. **Schütz**, K., Kammerer, D., Carle, R., & Schieber, A. (2004). Identification and quantification of caffeoylquinic acids and flavonoids from artichoke (*Cynara scolymus L.*) heads, juice, and pomace by HPLC-DAD-ESI/MS^n. *Journal of Agricultural and Food Chemistry, 52(13), 4090–4096.*

197. **Shearer**, J., Sellars, E., Farah, A., Graham, T. E., Wasserman, D. H. (2007). Effects of chronic coffee consumption on glucose kinetics in the conscious rat. Can. *J. Phys. Pharm.*, 85, 823–830.

198. **Shen** H. & Huang J. Z. (2008). Sparse principal component analysis via regularized low rank matrix approximation. *Journal of Multivariate Analysis* 99 (6). Academic Press: 1015–34.

199. **Shi**, H., Dong, L., Jiang, J., Zhao, J., Zhao, G., Dang, X., Jia, M. (2013). Chlorogenic acid reduces liver inflammation and fibrosis through inhibition of toll-like receptor 4 signalling pathway. *Toxicology*, 303, 107–114.

200. **Small**, D. M., Bender, G., Veldhuizen, M. G., Rudenga, K., Nachtigal, D., & Felsted, J. (2007). The role of the human orbitofrontal cortex in taste and flavour processing *Annals of the New York Academy of Sciences*, 1121(1), 136–151.

201. **Smith,** K. W. (2002). In Structured and Modified Lipids, Gunstone FD (ed). *Marcel Dekker*: New York, pp. 401–422.
202. **Smith**, R. F. (1985). A History of coffee. *Coffee*, 1-12.
203. **Snel**, J., & Lorist, M. M. (2011). Effects of caffeine on sleep and cognition. Human sleep and cognition part II - *Clinical and Applied Research, 105–117.*
204. **Somporn**, C., Kamtuo, A., Theerakulpisut, P., & Siriamornpun, S. (2011). Effects of roasting degree on radical scavenging activity, phenolics and volatile compounds of arabica coffee beans (*Coffea arabica L. cv. Catimor*). *International Journal of Food Science and Technology*, 46, 2287–2296.
205. **Sousa**, R., Munoz-Mahamud, E., Quayle, J., Dias da Costa, L., Casals, C., Scott, P., Guyot, A. (2014). Is asymptomatic bacteriuria a risk factor for prosthetic joint infection? *Clinical Infectious Diseases, 59(1), 41–47.*
206. **Speer** K., Kölling-Speer I. (2001). Lipids E.J. Clarke, O.G. Vitzthum (Eds.), Coffee: Recent developments, *Blackwell Sciences, London*, pp. 33-49.
207. **Speer**, K., & Kölling-Speer, I. (2006). The lipid fraction of the coffee bean. *Brazilian Journal of Plant Physiology*, 18(1), 201-216.
208. **Stadler**, R. H., Turesky, R. J., Müller, O., Markovic, J., & Leong-Morgenthaler, P. M. (1994). The inhibitory effects of coffee on radical-mediated oxidation and mutagenicity. Mutation *Research/Fundamental and Molecular Mechanisms of Mutagenesis, 308(2), 177–190.*
209. **Stalmach** A., Steiling H., Williamson G., Crozier A. (2010). Bioavailability of chlorogenic acids following acute ingestion of coffee by humans with an ileostomy. *Arch. Biochem. Biophys.*, 501:98–105.
210. **Suarez-Quiroz**, M., Gonzalez-Rios, O., Barel, M., Guyot, B., Schorr-Galindo, S., & Guiraud, J.-P. (2004). Study of ochratoxin A-producing strains in coffee processing. *International Journal of Food Science and Technology, 39(5), 501–507.*
211. **Sung** W. S., Lee D. G., (2010). Antifungal action of chlorogenic acid against pathogenic fungi, mediated by membrane disruption, *Pure Appl. Chem*. 82 (1) 219–226.
212. **Suzuki**, A., Kagawa, D., Ochiai, R., Tokimitsu, I., & Saito, I. (2002). Green coffee bean extract and its metabolites have a hypotensive effect in spontaneously hypertensive rats. *Hypertension Research,* 25(1), 99–107.
213. **Takagi**, T., & Ando, Y. (1995). Stereospecific analysis of monounsaturated triacylglycerol in cocoa butter. *Journal of the American Oil Chemists' Society, 72(10), 1203–1206.*

214. **Tamura**, H., Akioka, T., Ueno, K., Chujyo, T., Okazaki, K., King, P. J., & Robinson, W. E. (2006). Anti-human immunodeficiency virus activity of 3,4,5-tricaffeoylquinic acid in cultured cells of lettuce leaves. *Molecular Nutrition & Food Research, 50(4-5), 396–400.*

215. **Toci**, A. T., Farah, A., Pezza, H. R., & Pezza, L. (2015). Coffee Adulteration: More than Two Decades of Research. *Critical Reviews in Analytical Chemistry,* 46(2), 83-92.

216. **Toci**, A. T., Neto, V. J. M. F., Torres, A. G., & Farah, A. (2013). Changes in triacylglycerols and free fatty acids composition during storage of roasted coffee. *LWT - Food Science and Technology,* 50(2), 581-590.

217. **Tran**, H. T., Lee, L. S., Furtado, A., Smyth, H., & Henry, R. J. (2016). Advances in genomics for the improvement of quality in coffee. *Journal of the Science of Food and Agriculture, 96(10), 3300–3312.*

218. **Trugo**, L. C. (1985). Carbohydrates. In: Coffee, 1st edition, Clarke R. J., Macrae, R., eds. Essex: *Elsevier Applied Science Publishers*, Vol 1: Chemistry, p. 83.

219. **Trugo**, L. C. (2003). Coffee analysis. In: Encyclopedia of food science and Nutrition, 2nd edition, Caballero, B., Trugo, L. C., Finglas, P. M., eds. Oxford, UK: *Oxford Academic Press*, Vol. 2, p. 498.

220. **Truong**, V.D., McFeeters, R. F., Thompson, R. T., Dean, L. L., & Shofran, B. (2007). Phenolic acid content and composition in leaves and roots of common commercial sweet potato *(Ipomea batatas L.)* cultivars in the United States. *Journal of Food Science, 72(6),* C343–C349.

221. **Tunnicliffe** J. M., Shearer J., (2008). Coffee, glucose homeostasis, and insulin resistance: physiological mechanisms and mediators, *Appl. Physiol. Nutr. Metab*., 33, 1290–1300.

222. **Van Dam**, R. M. (2008). Coffee consumption and risk of type 2 diabetes, cardiovascular diseases, and cancer. *Applied Physiology, Nutrition, and Metabolism, 33(6), 1269–1283.*

223. **Van Der Vossen**, H. A. M. (2005). A critical analysis of the agronomic and economic sustainability of organic coffee production. *Experimental Agriculture,* 41(04): 449.

224. **Veljkovic**, E., Xia, W., Phillips, B., Wong, E. T., Ho, J., Oviedo, A., Peitsch, M. (2018). Other compounds from tobacco with potential impact on neurodegenerative diseases. Nicotine and other tobacco compounds in neurodegenerative and psychiatric diseases, *Overview of Epidemiological Data on Smoking and Preclinical and Clinical Data on Nicotine 83–97.*

225. **Vignoli**, J. A., Bassoli, D. G., & Benassi, M. T. (2012). Antioxidant activity of roasted and instant coffee: Standardization and validation of methodologies *Coffee Sci.* 7:65-78.

226. **Vignoli**, J. A., Viegas, M. C., Bassoli, D. G., & Benassi, M. de T. (2014). Roasting process affects differently the bioactive compounds and the antioxidant activity of arabica and Robusta coffees. *Food Research International, 61, 279–285.*

227. **Wan**, C. W., Wong, C. N. Y., Pin, W. K., Wong, M. H. Y., Kwok, C. Y., Chan, R. Y. K., Chan, S. W. (2012). Chlorogenic acid exhibits cholesterol lowering and fatty liver attenuating properties by up-regulating the gene expression of PPAR-α in hypercholesterolemic rats induced with a high-cholesterol diet. *Phytotherapy Research,* 27(4), 545–551.

228. **Wang**, H. Y., Qian, H., & Yao, W. R. (2011). Melanoidins produced by the Maillard reaction: Structure and biological activity. *Food Chemistry*, 128, 573–584.

229. **Want**, E. & Masson P. (2011). Processing and analysis of GC/LC–MS-based metabolomics data Methods *Mol. Biol*., 708: 277-298

230. **Weinberg**, B. A.; Bealer, B. K. (2001). The world of caffeine. *Routledge*. pp. 3-4. ISBN: 978-0-415-92723-9.

231. **Wijarnpreecha**, K., Thongprayoon, C., & Ungprasert, P. (2017). Coffee consumption and risk of nonalcoholic fatty liver disease. *European Journal of Gastroenterology & Hepatology*, 29(2), e8–e12.

232. **Williamson**, G. (2020). Protection against developing type 2 diabetes by coffee consumption: assessment of the role of chlorogenic acid and metabolites on glycaemic responses. *Food & Function.* doi:10.1039/d0fo01168a

233. **Wilson** K. M., Kasperzyk J. L., Rider J. R., Kenfield S., van Dam R. M., Stampfer M. J., Giovannucci E., Mucci L. A. (2011). Coffee consumption and prostate cancer risk and progression in the Health Professionals Follow-up Study *J. Natl Cancer Inst*. 103(11):876-84.

234. **Wilson**, A. J., Petracco, M., & Illy, E. (1997). Some preliminary investigations of oil biosynthesis in the coffee fruit and its subsequent re-distribution within green and roasted beans *In International Colloquium on the Chemistry of Coffee*, Paris: ASIC.

235. **Wold** S., Sjostrom M., Eriksson L. (2001). PLS-regression: a basic tool of chemometrics, *Chemometrics Intell. Lab. Syst*., 58: 109-130.

236. **Wu**, X., Beecher, G. R., Holden, J. M., Haytowitz, D. B., Gebhardt, S. E., & Prior, R. L. (2004). Lipophilic and hydrophilic antioxidant capacities of common foods in the United States. *Journal of Agricultural and Food Chemistry,* 52(12), 4026–4037.

237. **Yanagimoto**, K., Ochi, H., Lee, K. G., & Shibamoto, T. (2004). Antioxidative activities of fractions obtained from brewed coffee. *Journal of Agricultural and Food Chemistry, 52, 592–596.*

238. **Yu**, X., Bao, Z., Zou, J., & Dong, J. (2011). Coffee consumption and risk of cancers: a meta-analysis of cohort studies. *BMC Cancer, 11(1).*

239. **Zhang** L., Chang C., Liu, Y. Chen Z. (2011). Effect of chlorogenic acid on disordered glucose and lipid metabolism in *db/db* mice and its mechanism, *Acta Acad. Med. Sin*. 33 (3), 281–286.

240. **Zhang**, Q., Qin, W., Lin, D., Shen, Q., & Saleh, A. S. M. (2015). The changes in the volatile aldehydes formed during the deep-fat frying process. *Journal of Food Science and Technology*, 52(12), 7683–7696.

241. **Zhang**, Y., and Zhang, Y. (2007). Formation and reduction of acrylamide in Maillard reaction: A review based on the current state of knowledge. *Critical Reviews in Food Science and Nutrition, 47(5), 521–542.*

242. **Zhao**, Y., Wang, J., Ballevre, O., Luo, H., & Zhang, W. (2011). Anti-hypertensive effects and mechanisms of chlorogenic acids. *Hypertension Research,* 35(4), 370–374.

243. **Zou**, Y., Chang, S. K. C., Gu, Y., & Qian, S. Y. (2011). Antioxidant activity and phenolic compositions of lentil (*Lens culinaris var. Morton*) extract and its fractions. *Journal of Agricultural and Food Chemistry,* 59(6), 2268–2276.

Chapter 2

Coffee Arabica and Robusta CGA Comparison

Comparison and quantification of chlorogenic acids for differentiation of green Robusta and Arabica coffee beans

Sabur Badmos, Seung-Hun Lee, Nikolai Kuhnert

Adapted with permission from ***Food Res. Int., 2019 126, 108544***

Online published version available at: *https://doi.org/10.1016/j.foodres.2019.108544*

ABSTRACT

Coffea arabica and *Coffea canephora* (Robusta coffee) are the most commonly consumed coffee varieties globally. In this contribution, NMR was used to confirm the coffee authenticity and LC-ESI-MS^n technique was employed to profile and quantify the most abundant chlorogenic acid in 54 different samples of the two coffee varieties from diverse origins. Significant variations were observed for feruloyl quinic acids, dicaffeoyl quinic acids and 5-sinapoylquinic acid while the mono-caffeoyl quinic acids showed no variation when the two coffee varieties were compared. Additionally, isomer ratios were explored as a potential marker for coffee authenticity along with a thorough statistical evaluation of rather extensive data set. Robusta 5-CQA when compared with 3,4-DiCQA Robusta shows high positive correlation, similar high correlation coefficient was observed in 5-*p*CoQA Robusta when compared with 3-*p*CoQA as against the Arabica.

Keywords: Chlorogenic acids, Green coffee beans, 16-*O*-methylcafestol, Isomer ratios, Trihydroxylated chlorogenic acids, NMR, LC-MS^n, LDA.

1. INTRODUCTION

Chlorogenic acids (CGAs) are hydroxycinnamoyl esters of quinic acids. These plant secondary metabolites are ubiquitous in the plant kingdom. CGAs are among the most abundant dietary phenolics in an average diet (Grosso, Stepaniak, Topor-Madry, Szafraniec, & Pajak, 2014; Jaiswal & Kuhnert, 2010; Pandey & Rizvi, 2009; Pękal, Drózdz, Biesaga, & Pyrzynska, 2011).

Coffee is the main contributor to daily CGA intake with about 200 mg present in an average 200 ml cup (Svilaas et al, 2004). Daily intake is estimated at 1 - 2 g per capita per day (Grosso, Stepaniak, Topor-Madry, Szafraniec, & Pajak, 2014; Jaiswal & Kuhnert, 2010; Pandey & Rizvi, 2009; Pękal, Drózdz, Biesaga, & Pyrzynska, 2011). CGA structural chemistry is dominated by isomerism. The four non-equivalent hydroxyl substituents of quinic acids allow formation of four regio-isomeric CGAs and for multi-substituted derivatives a much larger number of regio-isomeric structures (Clifford, 1999). In coffee for example, all the three regio-isomers of dicaffeoyl quinic acids were observed and six of the theoretically possible regio-isomers of feruloyl quinic acid. In total, *Coffea arabica* produces between 45 and 50 different CGA derivatives, whereas its close relative *Coffea canephora* (Robusta coffee) produces 80 - 90 derivatives (Ky et al, 2001). During roasting of coffee beans, CGAs are intensively transformed into novel derivatives leading to a considerable reduction of concentration of CGAs originally present in the green coffee bean.

Commercially, due to food labelling requirements, the distinction between Arabica and Robusta coffee poses an important analytical task. Routinely in roasted coffees, 16-*O*-methylcafestol is used as a unique biomarker for *Coffea canephora* (Schievano, Finotello, De Angelis, Mammi, & Navarini, 2014). However, several groups have reported that CGAs can as well serve as a parameter allowing distinction between the two coffee varieties in green coffee beans (D'Amelio, Papamokos, Dreyer, Carloni, & Navarini, 2015; Hu et al, 2016; Rostagno, Celeghini, Debien, Nogueira, & Meireles, 2014). We have recently shown that Coffea

canephora produces a series of trihydroxylated CGAs, which are not present in *Coffea arabica* (Jaiswal, Matei, Subedi, & Kuhnert, 2014). Other research groups have consistently reported an increased CGA levels in Robusta if compared to Arabica coffee beans. Though with significant variation, quantitatively the total coffee CGAs is approximately 7.0 - 14.4% of dry-matter present in green Robusta and 4.0 - 8.4% in green Arabica beans (Farah & Donangelo, 2006). However, most data rely on the determination of sum parameters using the Folin-Ciocalteau or other antioxidant capacity assays, NMR-, IR- or Raman spectroscopic measurements of total CGA content. Moreover, which isomers of CGAs are present in increased quantities in Robusta coffee is currently unclear.

This contribution addresses three issues in chlorogenic acid chemistry. Firstly, we would like to add to the existing reliable quantitative information on coffee CGAs in green coffee beans. As a motivation, a survey of the literature shows that most quantitative data available refer to CGAs in the actual coffee beverage, which has suffered from intense chemical transformation due to thermal treatment, not in the green bean raw material. Data on green coffee beans include early work by Clifford (Clifford, 1999) and other two extensive papers by (Alonso-Salces, Serra, Remero, & Heberger, 2009; Babova, Occhipinti, & Maffei, 2016). However, the two latter contributions differ significantly in their findings, prompting a re-investigation along with an urgent alternative statistical evaluation of the data set. Secondly, we aim to compare quantitative CGA values from a large data set of Robusta and Arabica coffee beans with each other, in order to obtain a clear picture, which CGA derivatives are more abundant in Robusta coffee. This information seems to be largely missing from the literature.

Finally, we aim to investigate further in detail the CGAs isomer ratios in Robusta and Arabica coffee. The isomer ratio in CGAs constitutes in our view a neglected but potentially useful parameter in chemotaxonomy. For example, it is well documented that plants of the *Rosaceae* family (such as apple, apricot, pear, plum and peach plants) produce 3-substituted CGAs

preferentially, whereas in most other plant families the 5-isomer predominates (Möller & Herrmann, 1983; Upadhyay & Mohan Rao, 2013). Several studies have shown that one of the richest dietary sources of caffeoylquinic acids (CQAs) are the coffee beans (Gil & Wianowska, 2017). Additionally, we could recently show that the highly priced and sought-after Jamaica Blue Mountain Arabica coffee can be distinguished from less valuable alternatives based on its ratio of 3-CQA to 5-CQA (Lee, Jaiswal, & Kuhnert, 2016).

2. MATERIALS AND METHODS

2.1 Coffee beans

Green coffee beans from 20 different countries were used in this study. Arabica green coffee beans (40 samples from Brazil, Colombia, Nicaragua, Ethiopia, Honduras, Guatemala, Jamaica, Costa Rica, Rwanda, Laos, Kenya, China, Peru and Papua New Guinea) and green Robusta coffee beans (13 samples with 1 Arabica/Robusta blend sample from India, Vietnam, Ecuador, Uganda, Tanzania and Indonesia) were acquired from Deutscher Kaffeeverband, Hamburg and some local grocery shops in Bremen, Germany. Details of the coffee bean samples provided in supplementary data (Appendix Table A1).

2.2 Reagents

Acetonitrile used for HPLC sequence measurements was LC-MS grade (Acetonitrile for UHPLC gradient, AppliChem). Methanol, formic acid (used as a mobile phase modifier and as calibration buffer) were HPLC grade (both purchased from Sigma-Aldrich, Germany), while 5-CQA used as chlorogenic acid standard was purchased from PhytoLab.

2.3 Coffee beans methanolic extract

Methanolic extracts of fifty-four green coffee beans from different countries using 10 g each of the coffee bean samples were prepared. The beans were frozen using liquid nitrogen, ground properly using RAPIDO coffee grinder (TZS-FIRST, Austria) to super fine particlepowdered form and run on Soxhlet equipment (BUCHI B-811 Labortecnik System, Flawil, Switzerland) with 150 ml of 70% aqueous methanol. The Soxhlet setting used was continuous extraction mode for 8 h. This was followed by precipitation and removal of protein using Carrez reagents as previously described (Clifford, Johnston, Knight, & Kuhnert, 2003). The solvent was evaporated on rotary evaporator (Heidolph Laborota 4000-Efficient, Germany) at 50 °C and 120 rpm. Thereafter, the extract was stored at −80 °C for 1.5 h and lyophilized under 0.94 mbar

for 18 h using Christ-Alpha 1-4LSC freeze-dryer rotational-vacuum-concentrator (Osterode, Germany) to remove the aqueous phase completely from the extract. Extract was stored at −20 °C until required for LC-MS run. To prepare the LC-MS sample, extract was thawed at room temperature for about 1 h and re-dissolved in 70% methanol (50 mg/5 ml), filtered through a Chromafil PTFE membrane filter (Bremen, Germany) and used for LC-MS^n run to analysis the chlorogenic acids (CGA) content. Hesperitin (20 µl) was added to each prepared sample as internal standard.

2.4 NMR sample preparation and spectroscopy

Coffee bean samples were ground using RAPIDO coffee grinder (TZS-FIRST, Austria) in liquid nitrogen before used and 0.5 g of accurately weighed powder was dissolved in 1.5 ml of CDCl3 (99.96% deuterated chloroform with 0.05% (v/v) TMS, stabilized with silver) purchased from Sigma Aldrich, Germany. Samples vortex for 15 min at maximum speed and promptly filtered directly with a cotton wool into 5 mm precision glass NMR tubes and NMR spectrum acquired immediately. Steps carried out at 4 °C to minimize solvent evaporation (Schievano, Finotello, De Angelis, Mammi & Navarini, 2014). NMR spectroscopy carried out using Oxford AS400 instrument (England) coupled with JOEL STAC-MAN auto-sample changer (Germany). Spectrum was acquired using ^{1}H in deuterated chloroform with a pulse sequence, spectral width of 400 Hz, 32,768 data points and 16 scans to detect the presence of esterified 16-*O*-methylcafestol.

2.5 High resolution LC-ESI-MS/LC-MS^n

Separation was attained using a 150 × 3 mm i.d. column containing diphenyl 5 µm with a column guard with a diameter of 5 mm × 3 mm i.d. (Varian, Darmstadt, Germany). Water/formic acid (1000:0.005 v/v) was used as solvent A and methanol for solvent B, run at a total flow rate of 500 µL/min. Gradient profile was run from 10 to 70% B linearly in 60 min,

followed by 10 min isocratic and a return to 10% B at 90 and 10 min isocratic for re-equilibration.

The LC equipment (Agilent 1100 series, Karlsruhe, Germany) comprised a binary pump, an auto-sampler with a 100 µL loop, and a DAD detector with a light-pipe flow cell (recording at 320 and 254 nm and scanning from 200 to 600 nm). This was interfaced with an ion-trap mass spectrometer fitted with an ESI source (Bruker Daltonics HCT Ultra, Bremen, Germany) operating in full scan, auto MS^n mode to obtain fragment ion m/z. As required, MS^2, MS^3 and MS^4 fragment targeted experiments were performed with the focus mainly on compounds producing a parent ion at m/z 397, 559, 573, 587, 677, 691, 705 or 719. Tandem mass spectra were acquired in auto-MS^n mode (smart fragmentation) using a ramping of the collision energy. Maximum fragmentation amplitude was set to 1 V starting from 30% and ending at 200%. The MS operating conditions (negative mode) had been optimized using 5-*O*-caffeoylquinic acid with a capillary temperature of 365 °C, a dry gas flow rate of 10 L/min, and a nebulizer pressure of 10 psi. High-resolution LC-MS was carried out using the same HPLC equipped with a Micro-TOF Focus mass spectrometer (Bruker Daltonics, Bremen, Germany) fitted with an ESI source, and internal calibration was achieved with 10 mL of 0.1 M sodium formate solution injected through a six-port valve prior to each chromatographic run. Calibration was carried out using the enhanced quadratic mode (Jaiswal, Patras, Eravuchira, & Kuhnert, 2010).

2.6 HPLC coupled to MS

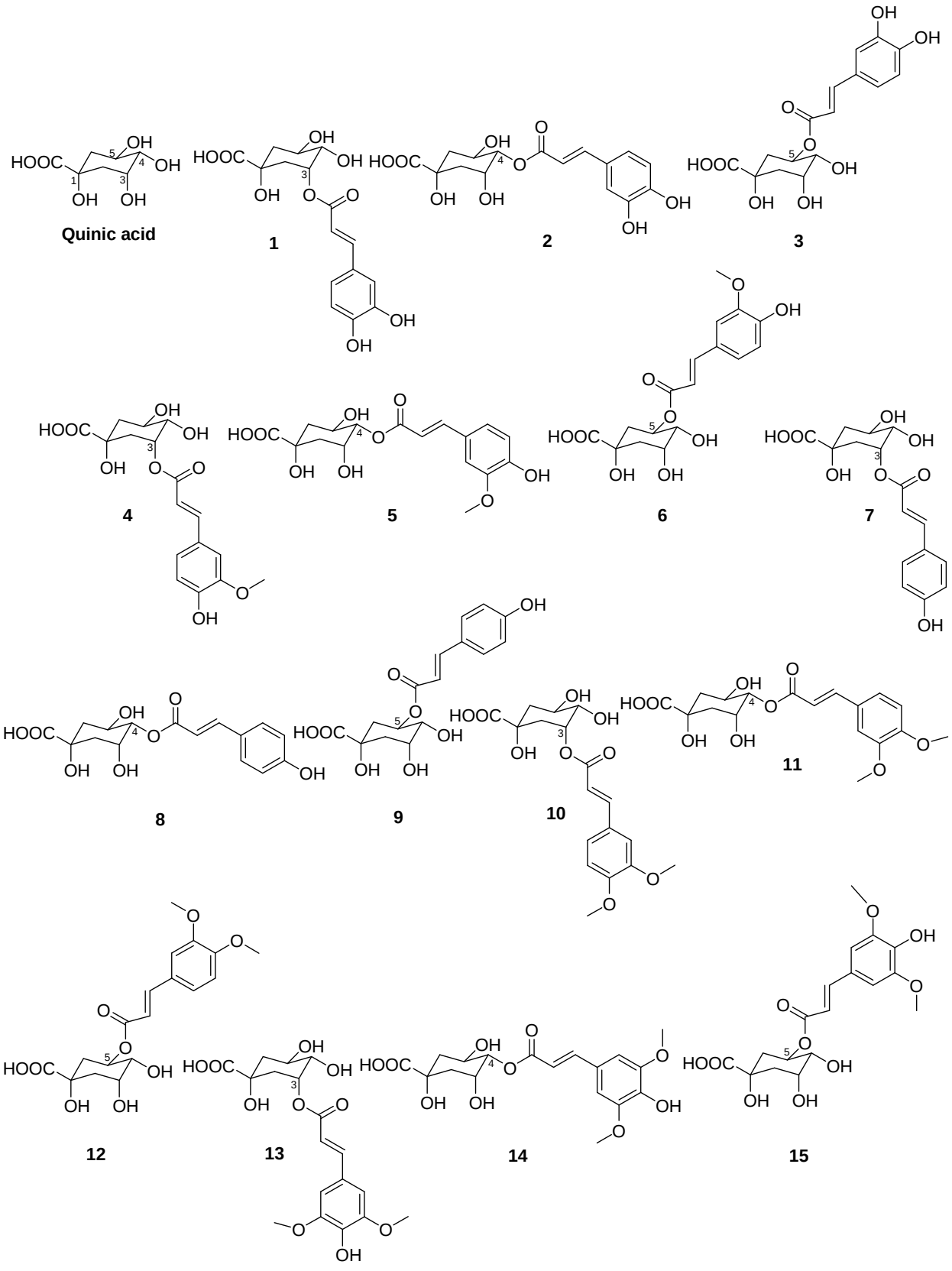
Separation was attained using a 150 × 3 mm i.d. column containing diphenyl 5 µm, with a column guard with a diameter of 5 mm × 3 mm i.d. (Varian, Darmstadt, Germany). Water/formic acid (1000:0.005 v/v) was used as solvent A and methanol for solvent B, run at a total flow rate of 500 µL/min. Gradient profile was run from 10 to 70% B linearly in 60 min, followed by 10 min isocratic and a return to 10% B at 90 and 10 min isocratic for re-equilibration. Compounds are identified by high-resolution mass spectrometry and tandem MS

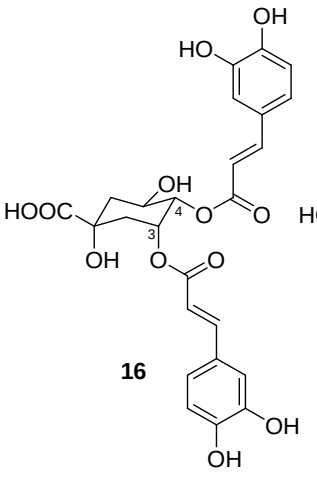
in the negative ion mode (Clifford, Knight, Surucu, & Kuhnert, 2006). Using the hierarchical scheme for LC-MS^n identification of CGAs, all compounds were assigned to regio-isomeric level. Chemical structures of key CGAs are shown in Fig. 1 with their respective details in Table 1, while the identified CGAs in this study are highlighted in the chromatogram provided in Fig. 2. For nomenclature of the CGAs we use the latest recommendations (Clifford, Jaganath, Ludwig, & Crozier, 2017).

2.7 Statistical evaluation

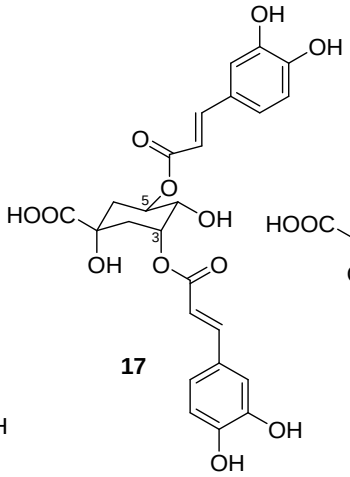
The mean, minimum and maximum values (n = 54) of all the Arabica and Robusta CGA were computed as shown in Table 2. These values were further used in calculating the skewness degree to determine the symmetry level of the data distribution and kurtosis to measure the data tailing degree relative to normal distribution towards determining the outliers' level in the data set evaluated. In addition, the p-values were computed and linear discriminate analysis (LDA) was carried out using the Orange 3.0 data mining software tool as previously reported by D'Souza et al. (2017) for data set with diverse variables degree.

Figure 1 Structures of green coffee bean chlorogenic acids with IUPAC nomenclature.

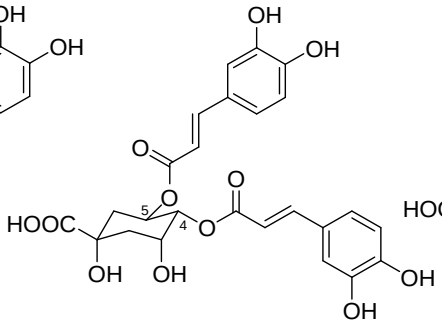




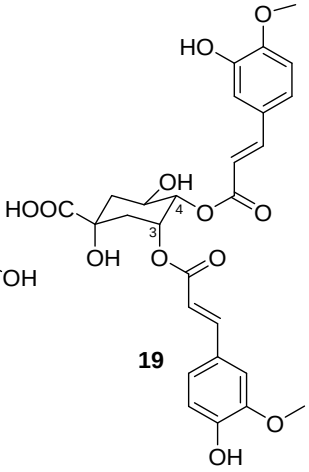
16



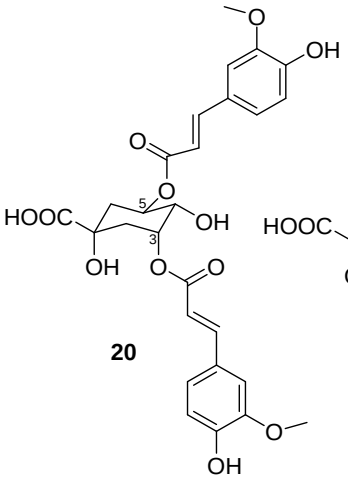
17



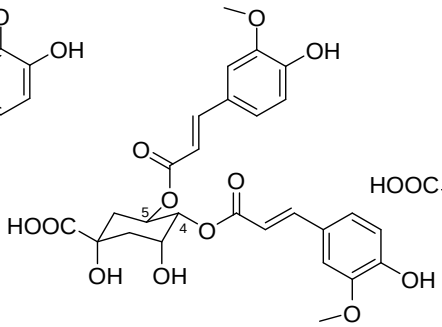
18



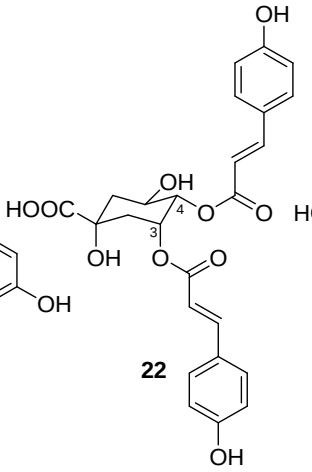
19



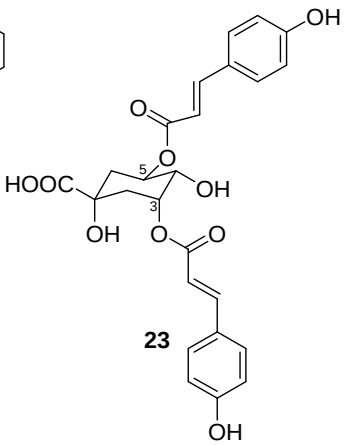
20



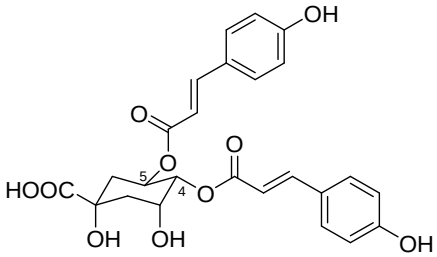
21



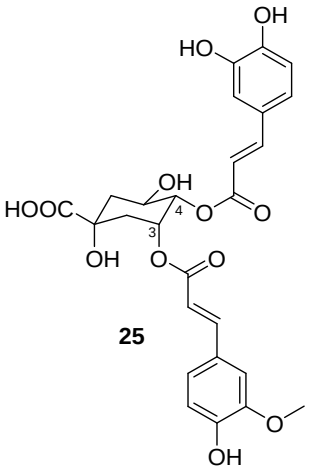
22



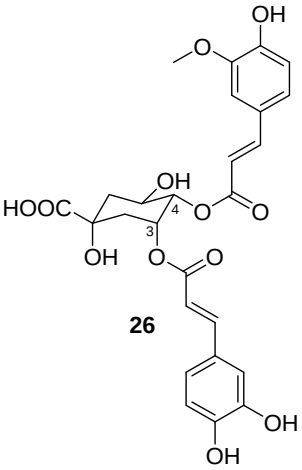
23



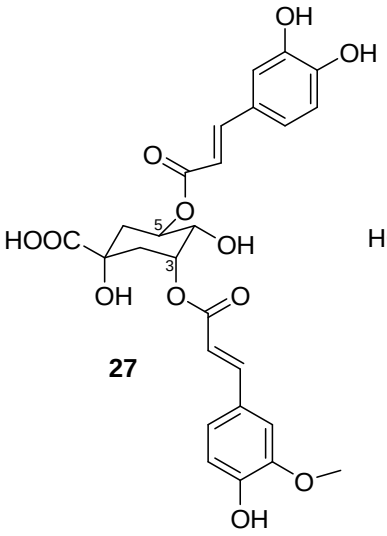
24



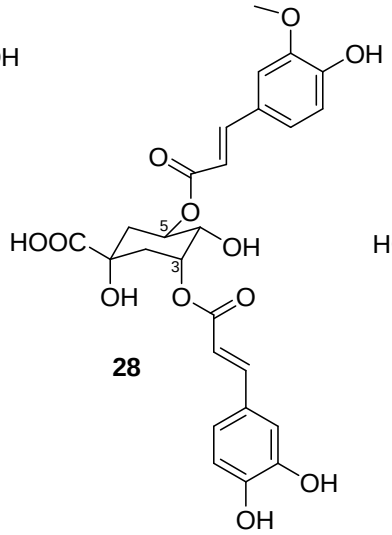
25



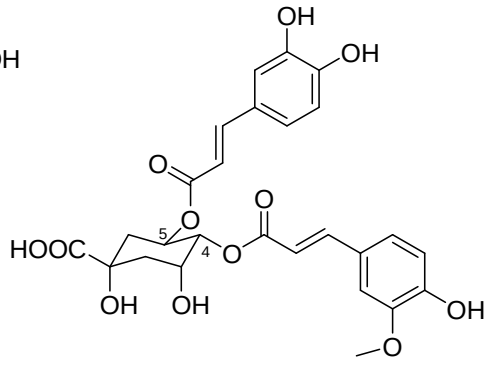
26



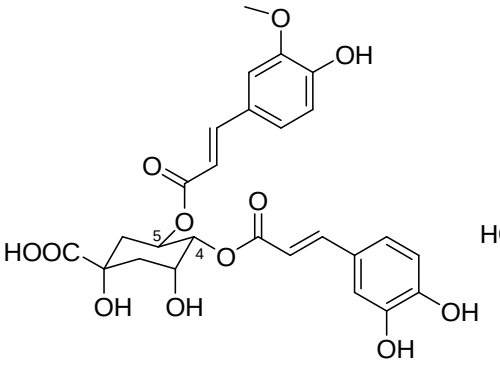
27



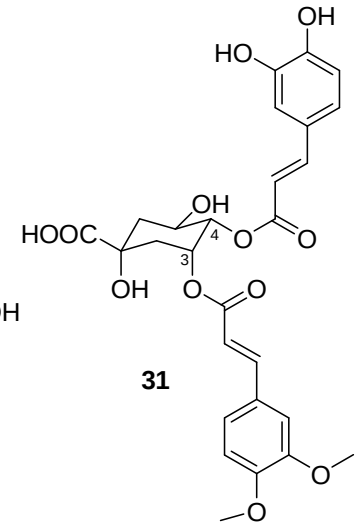
28



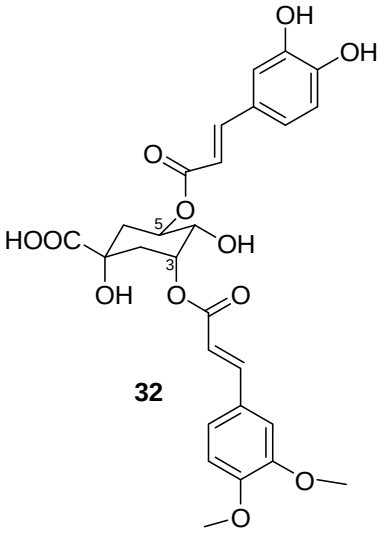
29



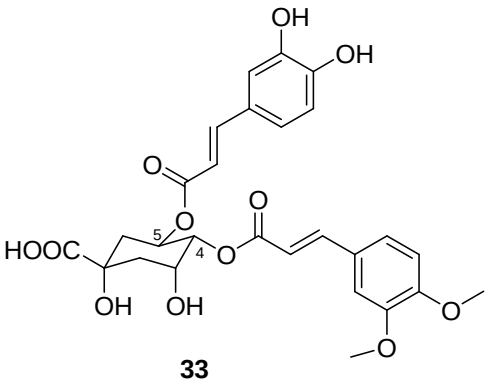
30



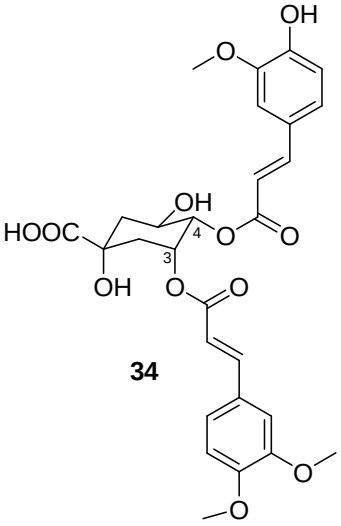
31



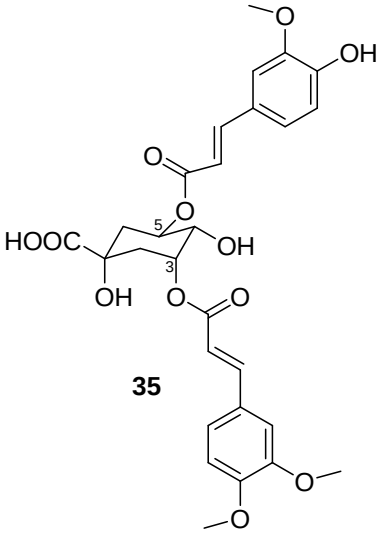
32



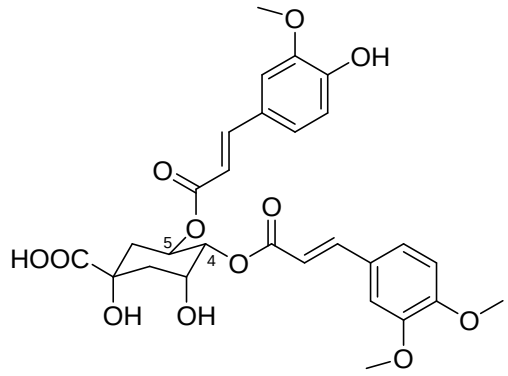
33



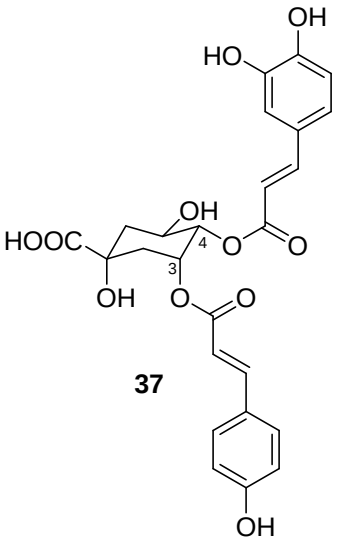
34



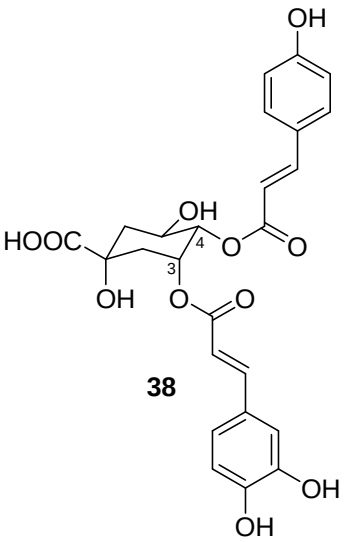
35



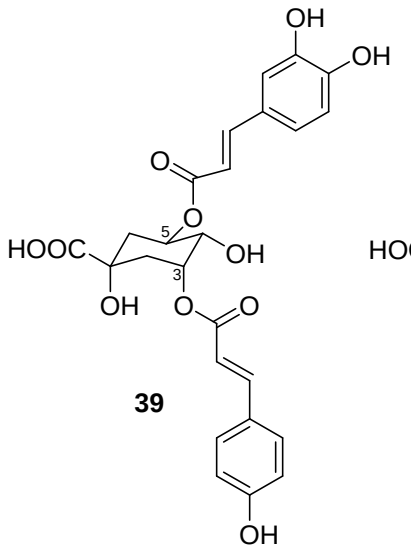
36



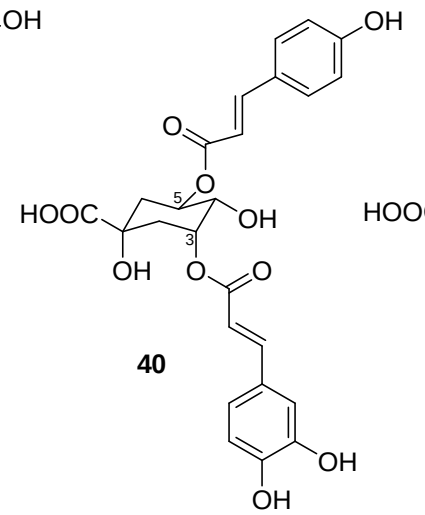
37



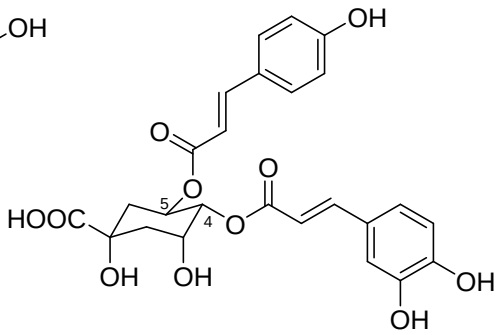
38



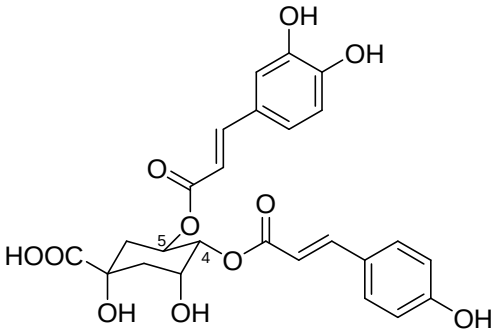
39



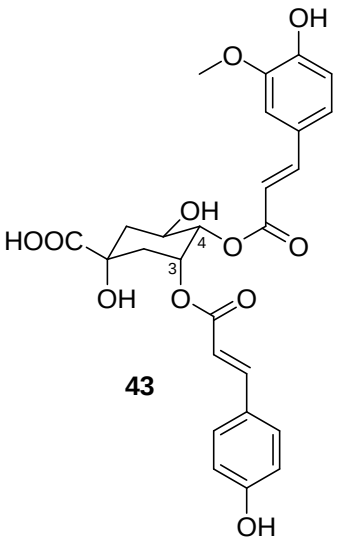
40



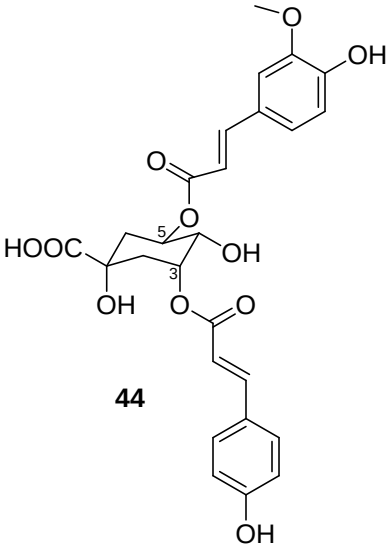
41



42



43



44

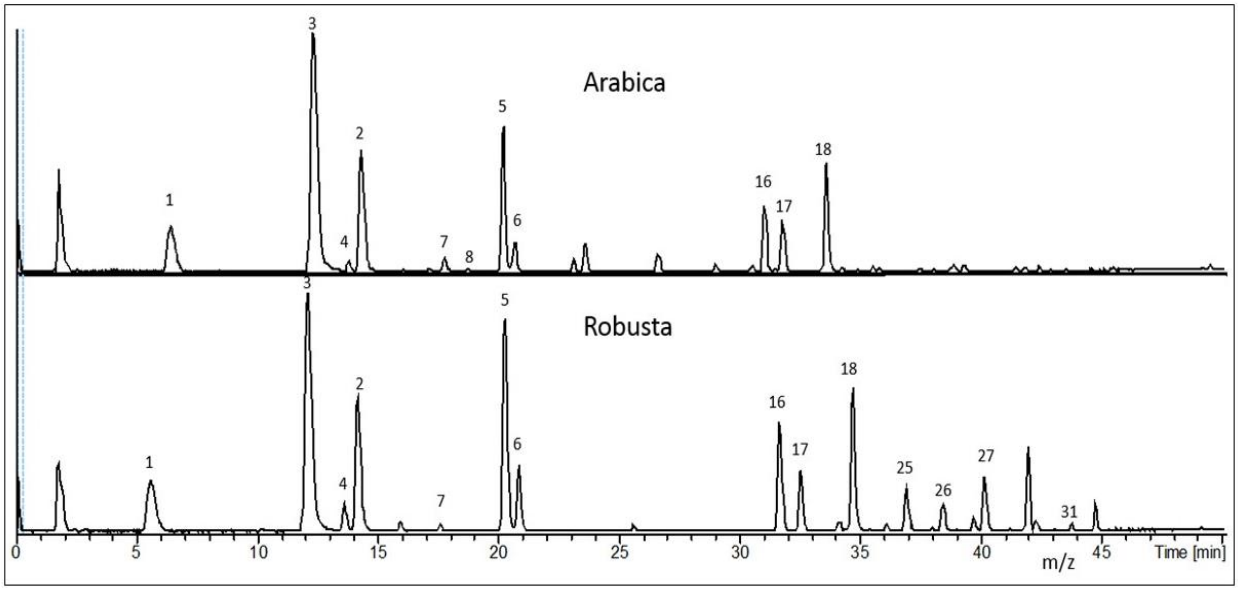
Table 1 High resolution mass (MS-TOF) data of identified CGAs, molecular formula and their parent ions (M-H).

No.	Name	Abbreviation	Mol. formula	Theor. *m/z* (M-H)	Exp. *m/z* (M-H)	Error (ppm)
1	3-*O*-caffeoylquinic acid	3-CQA	$C_{16}H_{18}O_9$	353.0878	353.0883	-0.5
2	4-*O*-caffeoylquinic acid	4-CQA	$C_{16}H_{18}O_9$	353.0878	353.0885	-1.3
3	5-*O*-caffeoylquinic acid	5-CQA	$C_{16}H_{18}O_9$	353.0878	353.0897	-3.4
4	3-*O*-feruloylquinic acid	3-FQA	$C_{17}H_{20}O_9$	367.0929	367.1042	-3.7
5	4-*O*-feruloylquinic acid	4-FQA	$C_{17}H_{20}O_9$	367.0929	367.1031	-0.6
6	5-*O*-feruloylquinic acid	4-FQA)	$C_{17}H_{20}O_9$	367.0929	367.1049	-2.3
7	3-*O*-*p*-coumaroylquinic acid	3-*p*CoQA	$C_{16}H_{18}O_8$	337.0929	337.0938	-0.8
8	4-*O*-*p*-coumaroylquinic acid	4-*p*CoQA	$C_{16}H_{18}O_8$	337.0929	337.0927	2.8
9	5-*O*-*p*-coumaroylquinic acid	5-*p*CoQA	$C_{16}H_{18}O_8$	337.0929	337.0925	2.1
10	3-*O*-dimethoxycinnamoylquinic acid	3-DQA	$C_{18}H_{22}O_9$	381.1191	381.1208	-2.5
11	4-*O*-dimethoxycinnamoylquinic acid	4-DQA	$C_{18}H_{22}O_9$	381.1191	381.1196	-2.9
12	5-*O*-dimethoxycinnamoylquinic acid	5-DQA	$C_{18}H_{22}O_9$	381.1191	381.1209	-2.6
13	3-*O*-sinapoylquinic acid	3-SiQA	$C_{18}H_{22}O_{10}$	397.1140	397.1128	3.1
14	4-*O*-sinapoylquinic acid	4-SiQA	$C_{18}H_{22}O_{10}$	397.1140	397.1147	-2.2
15	5-*O*-sinapoylquinic acid	5-SiQA	$C_{18}H_{22}O_{10}$	397.1140	397.1143	-4.2
16	3,4-di-*O*-caffeoylquinic acid	3,4-diCQA	$C_{25}H_{24}O_{12}$	515.1195	515.1189	0.7
17	3,5-di-*O*-caffeoylquinic acid	3,5-diCQA	$C_{25}H_{24}O_{12}$	515.1195	515.1179	3.4
18	4,5-di-*O*-caffeoylquinic acid	4,5-diCQA	$C_{25}H_{24}O_{12}$	515.1195	515.1177	3.2
19	3,4-di-*O*-feruloylquinic acid	3,4-diFQA	$C_{27}H_{28}O_{12}$	543.1508	543.1515	-0.4
20	3,5-di-*O*-feruloylquinic acid	3,5-diFQA	$C_{27}H_{28}O_{12}$	543.1508	543.1511	-0.8
21	4,5-di-*O*-feruloylquinic acid	4,5-diFQA	$C_{27}H_{28}O_{12}$	543.1508	543.1536	-2.6
25	3-*O*-feruloyl-4-*O*-caffeoylquinic acid	3F-4CQA	$C_{26}H_{26}O_{12}$	529.1351	529.1354	0.9
26	3-*O*-caffeoyl-4-*O*-feruloylquinic acid	3C-4FQA	$C_{26}H_{26}O_{12}$	529.1351	529.1357	-0.8
27	3-*O*-feruloyl-5-*O*-caffeoylquinic acid	3F-5CQA	$C_{26}H_{26}O_{12}$	529.1351	529.1369	-3.2
28	3-*O*-caffeoyl-5-*O*-feruloylquinic acid	3C-5FQA	$C_{26}H_{26}O_{12}$	529.1351	529.1359	-2.1
29	4-*O*-feruloyl-5-*O*-caffeoylquinic acid	4F-5CQA	$C_{26}H_{26}O_{12}$	529.1351	529.1353	0.4

30	4-*O*-caffeoyl-5-*O*-feruloylquinic acid	4C-5FQA	$C_{26}H_{26}O_{12}$	529.1351	529.1355	0.8
31	3-*O*-dimethoxycinnamoyl-4-*O*-caffeoylquinic acid	3D-4CQA	$C_{27}H_{28}O_{12}$	543.1508	543.1479	3.1
32	3-*O*-dimethoxycinnamoyl-5-*O*-caffeoylquinic acid	3D-5CQA	$C_{27}H_{28}O_{12}$	543.1508	543.1484	2.6
33	4-*O*-dimethoxycinnamoyl-5-*O*-caffeoylquinic acid	4D-5CQA	$C_{27}H_{28}O_{12}$	543.1508	543.1529	-2.8
34	3-*O*-dimethoxycinnamoyl-4-*O*-feruloylquinic acid	3D-4FQA	$C_{27}H_{28}O_{12}$	543.1508	543.1507	-3.9
35	3-*O*-dimethoxycinnamoyl-5-*O*-feruloylquinic acid	3D-5FQA	$C_{27}H_{28}O_{12}$	543.1508	543.1516	-1.3
36	4-*O*-dimethoxycinnamoyl-5-*O*-feruloylquinic acid	4D-5FQA	$C_{27}H_{28}O_{12}$	543.1508	543.1521	-3.3
37	3-*O*-p-coumaroyl-4-*O*-caffeoylquinic acid	3*p*Co-4CQA	$C_{25}H_{24}O_{11}$	499.1246	499.1229	3.2
38	3-*O*-caffeoyl-4-*O*-*p*-coumaroylquinic acid	3C-4*p*CoQA	$C_{25}H_{24}O_{11}$	499.1246	499.1243	-0.6
39	3-*O*-*p*-coumaroyl-5-*O*-caffeoylquinic acid	3*p*Co-5CQA	$C_{25}H_{24}O_{11}$	499.1246	499.1245	-0.9
40	3-*O*-caffeoyl-5-*O*-*p*-coumaroylquinic acid	3C-5*p*CoQA	$C_{25}H_{24}O_{11}$	499.1246	499.1244	-0.5
41	4-*O*-caffeoyl-5-*O*-*p*-coumaroylquinic acid	4C-5*p*CoQA	$C_{25}H_{24}O_{11}$	499.1246	499.1249	-4.2
42	4-*O*-*p*-coumaroyl-5-*O*-caffeoylquinic acid	4*p*Co-5CQA	$C_{25}H_{24}O_{11}$	499.1246	499.1247	-0.8
43	3-*O*-*p*-coumaroyl-4-*O*-feruloylquinic acid	3*p*Co-4FQA	$C_{26}H_{26}O_{11}$	513.1402	513.1389	2.4
44	3-*O*-*p*-coumaroyl-5-*O*-feruloylquinic acid	3*p*Co-5FQA	$C_{26}H_{26}O_{11}$	513.1402	513.1146	-2.0

C = caffeoyl; D = dimethoxycinnamoyl; F = feruloyl; *p*Co = *p*-coumaroyl.

Figure 2 The chromatograms representing the Arabica and Robusta CGA constituents' analysis on LC-ESI-QTOF-MS respectively.



3. RESULTS AND DISCUSSION

3.1. Confirmation of bean botany Arabica versus Robusta by ^{1}H-NMR

To ascertain correctness of bean botany and validate the results obtained from LC-MS analysis, 16-*O*-methylcafestol content in the 54 green coffee bean samples was additionally determined. While the International Organization for Standardization (ISO) on coffee standard uses LC-MS method (Clifford et al, 2006) to determine the 16-*O*-methylcafestol coffee oil constituent, which is solely present in Robusta coffee, but a simpler and more convenient ^{1}H-NMR method was used here (Schievano et al, 2014).

The coffee oil components were extracted using $CDCl_3$ (deuterated chloroform), which remarkably extracted all essential oil constituents principally the diterpenes (cafestol and kahweol). With the deuterated chloroform extraction, coffee lipid extracts were subsequently used to acquire NMR spectra using 16 scans to detect the presence of esterified 16-*O*-methylcafestol in each of the samples. The esterified 16-*O*-methylcafestol proton-21 ($-OCH_3$) has been reported to resonate as a singlet at position 3.17 ppm (Schievano et al, 2014). By visual inspection of the spectra, the protons signal at 3.17 ppm indicates presence of the esterified 16-*O*-methylcafestol in the coffee sample and hence it is identified as Robusta coffee, whereas its absence allows sample assignment as coffee Arabica (Fig. 3).

To understand this, identified mixtures or blends of Arabica and Robusta coffee mixtures with defined compositions were investigated as well. The signal at position 3.17 ppm, could be readily observed using 16 scans in the mixtures of Arabica/Robusta (50:50) and (60:40), but it was absent in (70/30) ratio as seen in supplementary data Fig. A1. For blends with lower Robusta concentration, no reliable authentication with this scan number was observed. From this observation, it could be inferred that the quantity or proportion of coffee Robusta mixed with coffee Arabica could affect the gold standard test of using esterified 16-*O*-methylcafestol

to differentiate between the two coffee varieties when mixed together with a degree ≤ 40% of coffee Robusta.

Figure 3 ^{1}H-NMR (400 Hz) in $CDCl_3$ spectrum of a confirmed coffee Robusta sample with the presence of 16-*O*-Methylcafestol lipid content at position 3.17 ppm on the left, while on the right is spectrum of Arabica indicating the absence of the biomarker at position 3.17 ppm.

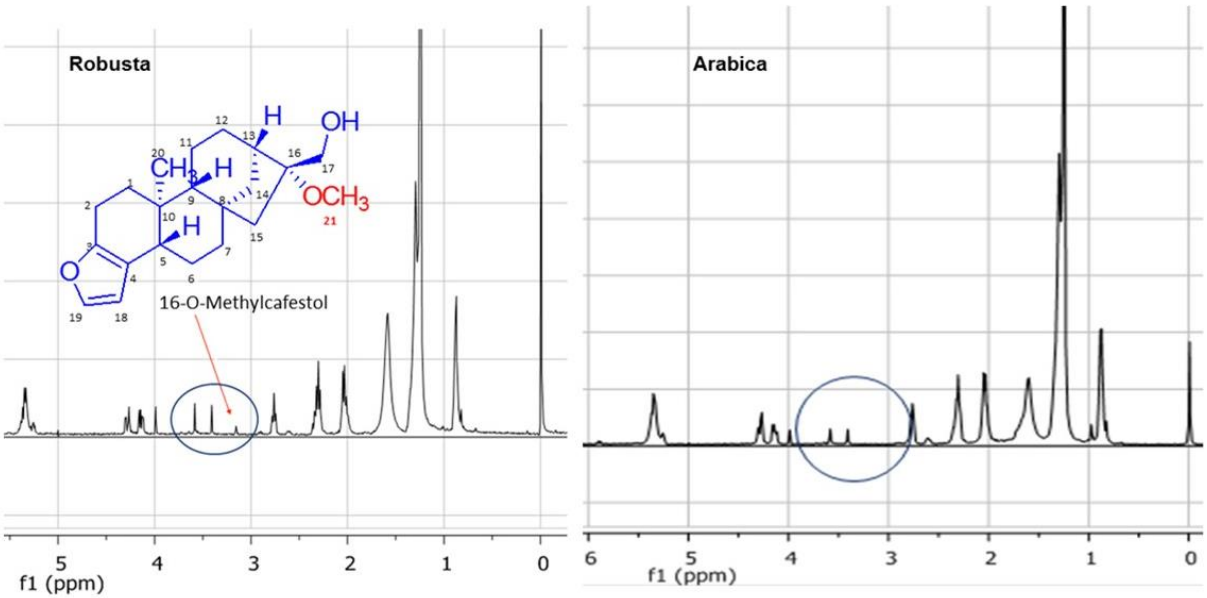


Table 2 Mean, minimum and maximum values of coffee Arabica and Robusta CGA in mg/100 g. Mean (n = 54).

	Arabica	Robusta	Arabica	Robusta	Arabica	Robusta
CQA	3-CQA		4-CQA		5-CQA	
Mean ± SD	3.64 ± 0.8	5.07 ± 1.9	5.28 ± 0.9	6.78 ± 2.5	37.88 ± 4.6	42.26 ± 7.5
Minimum	1.86	3.48	3.09	4.39	30.60	31.29
Maximum	5.68	10.07	7.55	13.21	47.22	50.17
Skewness	1.51	0.49	1.13	0.05	0.38	0.12
Kurtosis	4.6	0.08	2.53	0.01	0.78	0.16
DiCQA	3,4-diCQA		3,5-diCQA		4,5-diCQA	
Mean ± SD	1.92 ± 0.6	5.45 ± 1.1	2.73 ± 0.5	4.84 ± 1.3	2.32 ± 0.5	4.49 ± 0.9
Minimum	0.89	4.00	1.76	3.26	1.26	3.21
Maximum	4.14	6.70	4.02	6.67	3.57	6.03
Skewness	1.15	0.03	0.82	0.46	0.55	0.32
Kurtosis	3.04	1.95	0.96	−1.54	1.69	−1.81
FQA	3-FQA		4-FQA		5-FQA	
Mean ± SD	0.22 ± 0.1	0.61 ± 0.2	0.45 ± 0.1	1.14 ± 0.3	3.06 ± 0.5	8.71 ± 1.7
Minimum	0.00	0.34	0.21	0.64	1.98	6.42
Maximum	0.38	1.08	0.73	1.93	4.02	11.26
Skewness	0.68	1.15	0.69	0.19	1.12	0.73
Kurtosis	2.75	0.37	0.65	1.48	2.95	0.37
DiFQA	3,4-diFQA		3,5-diFQA		4,5-diFQA	
Mean ± SD	484,550	1,850,140	390,520	2,047,930	1,030,750	3,248,890
Minimum	55,610	1,417,470	100,470	875,140	420,900	2,538,050
Maximum	1,243,840	3,759,360	825,870	10,011,890	2,189,660	5,818,400
Skewness	1.03	1.62	0.73	2.47	1.08	1.37
Kurtosis	2.31	2.37	1.53	4.93	3.54	1.63
pCoQA	3-pCoQA		4-pCoQA		5-pCoQA	
Mean ± SD	133,190	44,240	2,822,260	1,483,630	583,190	340,480
Minimum	0	0	254,550	548,930	179,930	114,470
Maximum	2,611,780	383,120	5,680,690	2,113,700	1,479,540	1,229,170
Skewness	5.64	3.15	0.19	0.53	1.46	2.48
Kurtosis	4.08	6.21	1.74	0.26	4.21	5.13

3.2. LC-MS quantification of chlorogenic acids and respective regio-isomers

For selected CGAs available as authentic reference standards eight-point calibration curves were obtained using both HPLC-UV/VIS quantification at 320 nm and additionally calibration curves for LC-MS quantification in the negative ion mode were obtained. Analytical parameters determined, include LOD (limit of detection), LOQ (limit of quantification), RRF (relative response factor), p-values, Pearson correlation coefficients and standard deviations for all derivatives are given in the supplementary information (Tables 4 and 5). All full quantitative absolute CGA values are provided as well in the supplementary information, whereas data are presented in the main text as box plots providing a quick graphical overview over CGA distributions in the extensive data set.

3.3. Mono-caffeoyl quinic acids

The absolute values of the mono-caffeoyl quinic acids (CQAs) were determined and represented in the box-plot shown in Fig. 4(a) along with a typical chromatogram (Fig. 2). Although, prior studies have shown that the main CGA constituents of green coffee beans are mostly 5-CQA besides various diacyl-quinic acids and some triacyl-quinic acids (Clifford et al, 2017). In this study, the mono caffeoylquinic acids isomers under review are 3-CQA, 4-CQA and 5-CQA. The plot and statistical analysis indicate that 5-CQA is the predominant isomer in all samples compared to 3-CQA and 4-CQA. Levels of 3-CQA and 4-CQA are usually comparable at 50% of 5-CQA. There is no much significant difference in mono-CQA levels if Robusta and Arabica samples are compared. From the plot, the 5-CQA shows an average of (Arabica 37.9 mg/100 g; Robusta 42.3 mg/ 100 g), while the average was about (Arabica 5.3 mg/100 g; Robusta 6.8 mg/100 g) for the 4-CQA and (Arabica 3.6 mg/100 g; Robusta 5.1 mg/100 g) for 3-CQA. Variations observed in 5-CQA content are significantly higher in Arabica samples if compared to Robusta samples.

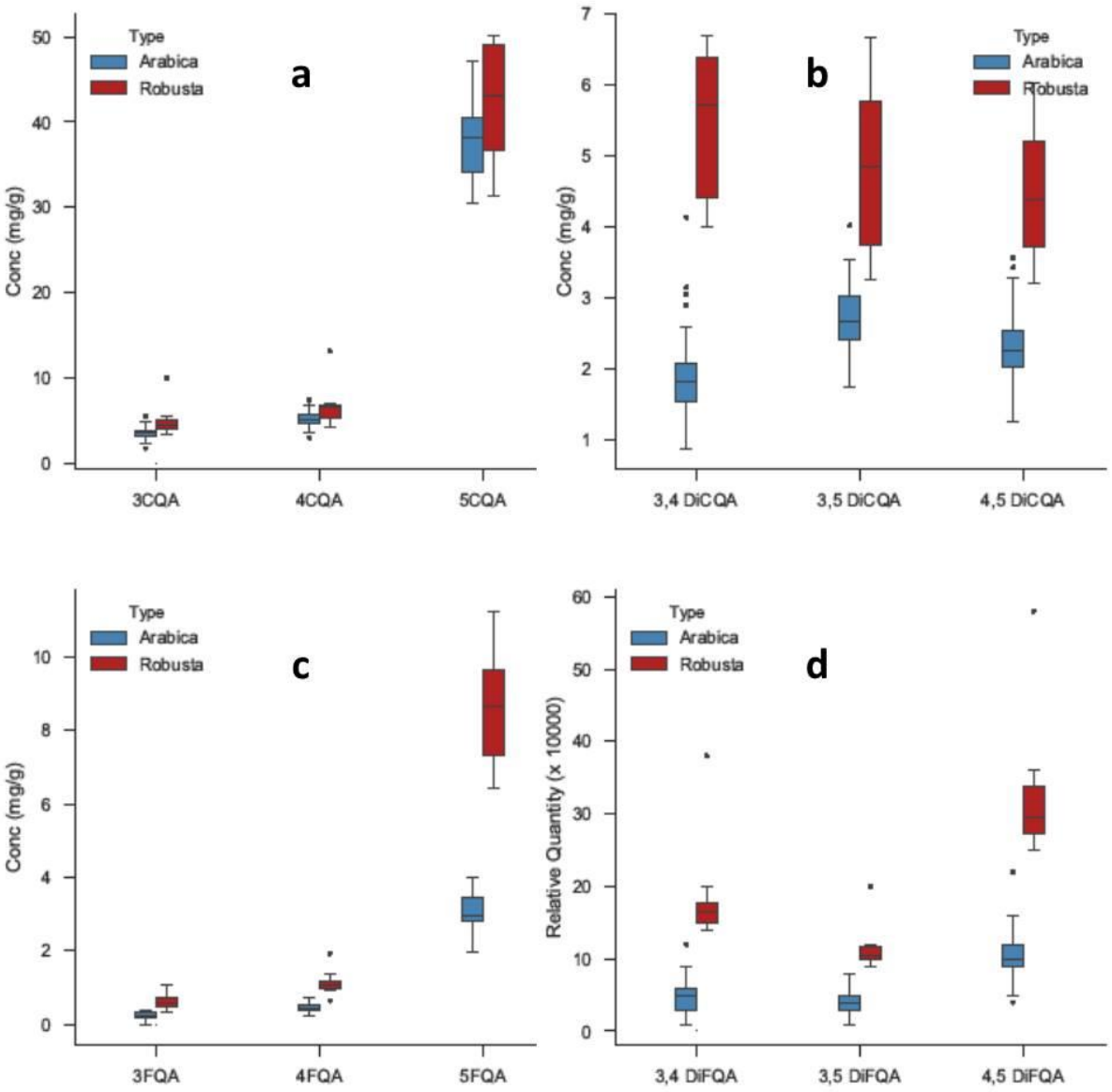


Figure 4 Box-plots of mono-caffeoylquinic acids (a), di-caffeoylquinic acids (b), mono-feruloylquinic acids (c) and di-feruloylquinic acids (d) content of coffee Arabica (blue) and Robusta (red) indicating the quantities of Robusta is higher in the two coffee samples in almost all the isomers. Mean (n = 54).

3.4. Dicaffeoyl quinic acids

All the three dicaffeoyl quinic acids were located in typical chromatogram (Fig. 2) and the derivatives were quantified with the resulting box plot in Fig. 4(b). Absolute levels of 3,4-DiCQA, 3,5-DiCQA and 4,5-DiCQA were found to be significantly higher in Robusta green coffee beans if compared to Arabica. The difference of a factor of around five is particularly significant for 3,4-DiCQA. From plot, Robusta (Rb) green beans DiCQA quantified gave 5.45 mg/100 g for 3,4-DiCQA, 4.84 mg/ 100 g for 3,5-DiCQA and 4.49 mg/100 g for 4,5-DiCQA; while Arabica (Ar) green beans has 1.92 mg/100 g for 3,4-DiCQA, 2.73 mg/100 g for 3,5-DiCQA and 2.32 mg/100 g for the 4,5-DiCQA.

Comparatively, while there is little significant difference among the isomers of a specific coffee bean either Arabica or Robusta, the Robusta quantities show a significant difference between some of the CGA quantities if compared to the Arabica coffee. Besides, the di-caffeoylquinic acid (DiCQA) indicates the quantity of 4,5-DiCQA Robusta is slightly elevated in the two-coffee species compared to 3,4-DiCQA and 3,5-DiCQA. It also shows that DiCQA is more in quantity in the Robusta than Arabica.

3.5. Mono-Feruloyl, feruloyl-caffeoyl quinic acids and –p-coumaroyl quinic acids

The feruloyl quinic acids contents of the coffee were also quantified using absolute values of the CGA from the UV peak values and the concentration of the respective regio-isomers. Fig. 4(c) indicates the mono-FQAs. From the plot, 5-FQA quantities were observed to be Robusta 8.7 mg/100 g and Arabica 3.1 mg/100 g; 4-FQA with an average of 1.1 mg/100 g for Robusta and 0.45 mg/100 g for Arabica and 3-FQA Rb is 0.6 mg/100 g and Ar 0.2 mg/100 g. This clearly indicates that the quantities of the 5-FQA are significantly higher in the two-coffee species compared to 3-FQA and 4-FQA. It also shows that Robusta are present more in FQA than Arabica samples. Furthermore, the DiFQAs were also quantified using the relative values and

their respective concentrations plotted in Fig. 4(d). The DiFQAs plot shows that the 4,5- DiFQA (Rb: 3.2×10^6 and Ar: 1.0×10^5) is significantly higher than 3,5-DiFQA (Rb: 2.1×10^6 and Ar: 3.9×10^5) and 3,4-DiFQA (Rb: 1.8×10^6 and Ar: 4.8×10^5). Since the quantities of 4,5-DiFQA quite differ and significantly higher in the two different coffee beans, it could be suggested as a tool to differentiate the two-coffee using this quantitation analytical method.

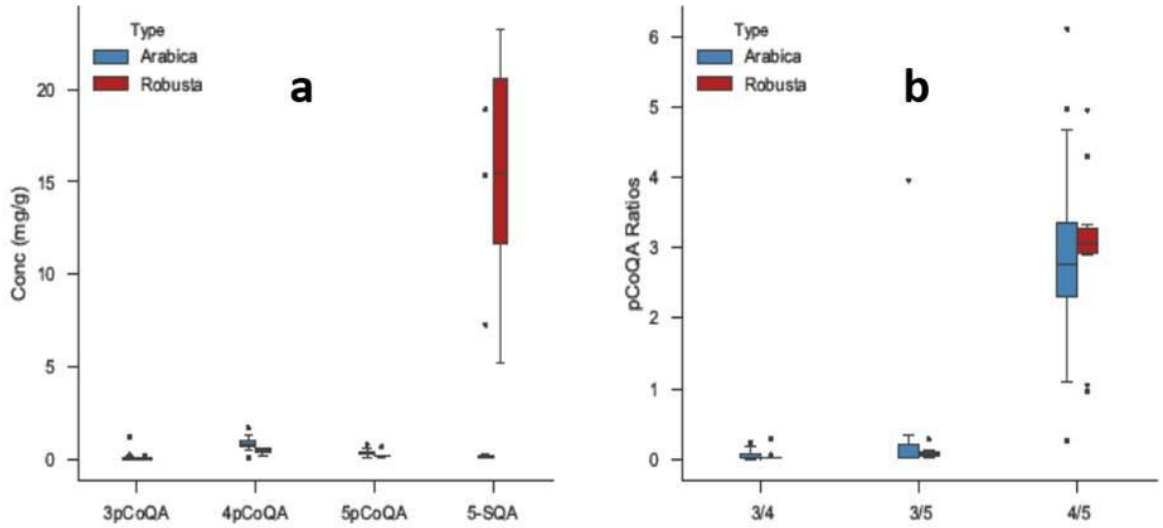


Figure 5 Box-plot of para-Coumaroylquinic acid and 5-Sinapoylquinic acid constituents of coffee Arabica (blue) and Robusta (red). Plot (a) indicates the concentrations (mg/g) of the *p*CoQA and 5-SQA in mg/100 g, while plot (b) represents the regio-isomer ratios of x to y for ratios of the three *p*CoQA isomers. Mean (n = 54).

The *p*CoQA acid constituents of the Arabica and Robusta coffee samples were quantified using the relative values (*p*CoQA only) of the CGA and their respective concentrations plotted as represented in Fig. 5(a). Three p-coumaroyl quinic acids isomers (3-*p*CoQA, 4-*p*CoQA and 5-*p*CoQA) were also identified in the chromatograms to differentiate Robusta coffee from Arabica coffee species. Furthermore, from same plot, 4-*p*CoQA has an average mean value of about 1.5×10^6 for Rb and 2.8×10^6 for Ar. compared to 5-*p*CoQA (Rb: 3.4×10^5 and Ar: 5.8×10^5) and 3-*p*CoQA (Rb: 4.4×10^4 and Ar: 1.3×10^5). This shows that the Arabica quantities are significantly higher in the three isomers compared to Robusta coffee.

Moreover, Clifford and others previously reported that Robusta coffee typically have a higher content in any CGA isomer if compared to Arabica, but with the exception of *p*CoQA, the prominent being an inclination for Arabica coffee to contain more 5-*p*CoQA than in Robusta (Clifford, Kellard, & Ah-sing, 1989). This fact was evidently observed in the three *p*CoQA isomers analysed, where Arabica was observed to be notably elevated than Robusta with about a factor of two, which is in contrast to quantities of both caffeoyl and feruloyl derivatives as seen in Fig. 4(a - d). Besides, this observation indicates that the 4-isomer is the most predominant regio-isomer if compared with 3-*p*CoQA and 5-*p*CoQA, and either in Arabica or Robusta. Furthermore, the *p*CoQA ratios were calculated and plotted as indicated in Fig. 5(b). From the ratio box-plots, the ratio of 4-*p*CoQA/5-*p*CoQA has a significantly elevated value (Rb: 5.92 and Ar: 5.32) compared to the ratios of 3-*p*CoQA/4-*p*CoQA (Rb: 0.02 and Ar: 0.25) and 3-*p*CoQA/5-*p*CoQA (Rb: 0.06 and Ar: 0.20).

3.6. Trihydroxylated chlorogenic acids

The presence of trihydroxylated cinnamoyl quinic acid derivatives was recently reported and exclusively observed in Robusta coffee. Furthermore, the triacyl-quinic acids have so far been found only in Robusta coffee beans (Jaiswal & Kuhnert, 2010). In this study, a more sensitive mass spectrometer was employed compared to previous work. As a result, trihydroxylated derivatives were found in 85% of all coffee samples analysed. However, if comparing relative quantities based on normalized peak areas Robusta samples show a ten to twenty folds increased level of these marker compounds. Additionally, for Robusta samples, multiple isomers for SQAs could be observed, whereas in Arabica samples typically only a single isomer could be observed to be above the limit of detection. Relative peak areas of 5-SQA are shown in Fig. 5a.

3.7. Inspection of characteristic isomer ratio parameters by comparing Robusta and Arabica CGA quantitative data

Isomer ratios of chlorogenic acids have been previously shown to allow distinction of Blue Mountain Jamaica Coffee from other coffee origins (Lee et al., 2016). However, several factors are known to influence the percentage compositions of the different CGAs constituents in green coffee beans. Among these are bean species, maturity level, country of origin, cultivation and climatic conditions, presence of defected beans, post-harvesting methods of processing by either wet or dry and washed or unwashed beans (De Menezes, 1994). de Menezes reported that the different maturity levels of coffee beans and the post-harvesting methods (wet or dry) have significant impact on the percentage ratio of the mono-CQA to di-CQA. It was observed in the study that the wet method usually present higher values compared to the dry method; and the inclusion of green unripe berries could negatively affect the beverage flavour due to the reduce ratios (De Menezes, 1994; Kleinwächter, Bytof & Selmar, 2014).

In addition, it has been observed that another post-harvesting stage which involves the cleaning of the beans (washed or unwashed) could considerably affects the coffee beans CGA levels, though there is no available evidence to this assertion as at present. Likewise, it has been reported that the presence of defective beans in samples could affect the percentage composition and different ratios of the various CGAs known to be abundantly present in green coffee beans (Kleinwächter et al, 2014; Upadhyay & Mohan Rao, 2013). Therefore, we decided to further investigate these parameters by calculating isomer ratios from the available quantification results. The ratios of the CQA and FQA constituents of the coffees are plotted in box-plot presented in Fig. 6 and Table 3.

However, the Robusta DiCQA isomers ratios mean 3,4/3,5 and 3,4/ 4,5 -DiCQA appear to be higher with 1.17 and 1.24 respectively compared to Arabica with 0.72 and 0.83, but with a slight difference in ratio 3,5/4,5 with 1.20 for Arabica and 1.07 for Robusta. This indicates that

the quantities of the Robusta ratios are slightly elevated here compared to the Arabica coffee and most significantly in DiCQA. Similar observation was seen in the DiFQA isomers where Robusta has an average ratio values of 1.45 for 3,4/3,5 ratio; 0.56 for 3,4/4,5 ratio and 0.67 for 3,5/4,5 ratio respectively compared to Arabica with 1.23 for 3,4/4,5 ratio; 0.45 for 3,4/3,5 ratio and 0.38 for 3,5/4,5 ratio. This implies that 3,4/3,5 ratio is slightly elevated in DiCQA Robusta and reduced in Arabica. From this, it could be inferred that Robusta quantities are higher in DiCQA and DiFQA than in Arabica coffee. In addition, the ratio values of 3/4 mono-CQA (Arabica 0.69 and Robusta 0.75) are more in quantity than the ratios of 3/5 (Arabica 0.10 and Robusta 0.12) and 4/5 (Arabica 0.14 and Robusta 0.16). Furthermore, similar observation was recorded for the ratio 3/4 mono-FQA (Arabica 0.48 and Robusta 0.53) are more in quantity than the ratios of 3/5 (Arabica 0.07 and Robusta 0.07) and 4/5 (Arabica 0.15 and Robusta 0.13).

Table 3 Mean, minimum and maximum values of coffee Arabica and Robusta CGA ratios. Mean (n = 54).

CQA	3/4	3/5	4/5	DiCQA	3,4/3,5	3,4/4,5	3,5/4,5
Arabica							
Min	0.53	0.05	0.10		0.41	0.48	0.87
Mean	0.69	0.10	0.14		0.72	0.83	1.20
Max	0.95	0.15	0.20		1.23	1.38	1.85
Robusta							
Min	0.65	0.09	0.12		0.92	1.10	0.94
Mean	0.75	0.12	0.16		1.17	1.24	1.07
Max	0.84	0.21	0.27		1.96	1.99	1.24
FQA	**3/4**	**3/5**	**4/5**	**DiFQA**	**3,4/3,5**	**3,4/4,5**	**3,5/4,5**
Arabica							
Min	0.00	0.00	0.10		0.54	0.13	0.22
Mean	0.48	0.07	0.15		1.23	0.45	0.38
Max	0.74	0.13	0.23		2.11	0.85	0.58
Robusta							
Min	0.32	0.04	0.10		0.16	0.53	0.34
Mean	0.53	0.07	0.13		1.45	0.56	0.67
Max	0.63	0.10	0.17		1.84	0.65	3.50
pCoQA	**3-pCoQA**	**4-pCoQA**	**5-pCoQA**		**3/4-pCoQA**	**3/5-pCoQA**	**4/5-pCoQA**
Arabica							
Min	0	254,550	179,930		0.00	0.00	0.50
Mean	133,190	2,822,260	583,190		0.25	0.20	5.32
Max	2,611,780	5,680,690	1,479,540		9.36	4.63	12.14
Robusta							
Min	0	548,930	114,470		0.00	0.00	1.72
Mean	44,240	1,483,640	340,480		0.02	0.06	5.92
Max	383,130	2,113,680	1,229,170		0.18	0.31	10.56

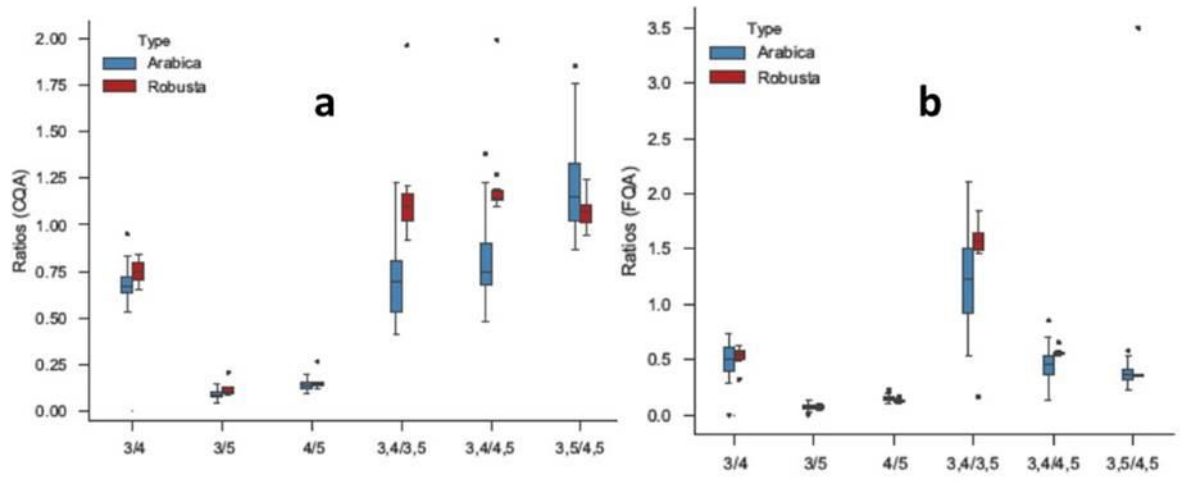


Figure 6 Box-plot of regio-isomer ratios of x to y for the mono-CQA, DiCQA, mono-FQA and DiFQA constituents of coffee Arabica (blue) and Robusta (red). Mean (n = 52) and expressed in mg/100 g. Plot (a) indicates the ratios of 3-CQA to 4-CQA, 3-CQA to 5-CQA, 4-CQA to 5-CQA, 3,4-DiCQA to 3,5-DiCQA, 3,4-DiCQA to 4,5-diCQA and 3,5-diCQA to 4,5-DiCQA, while plot (b) indicates the ratios for the FQA.

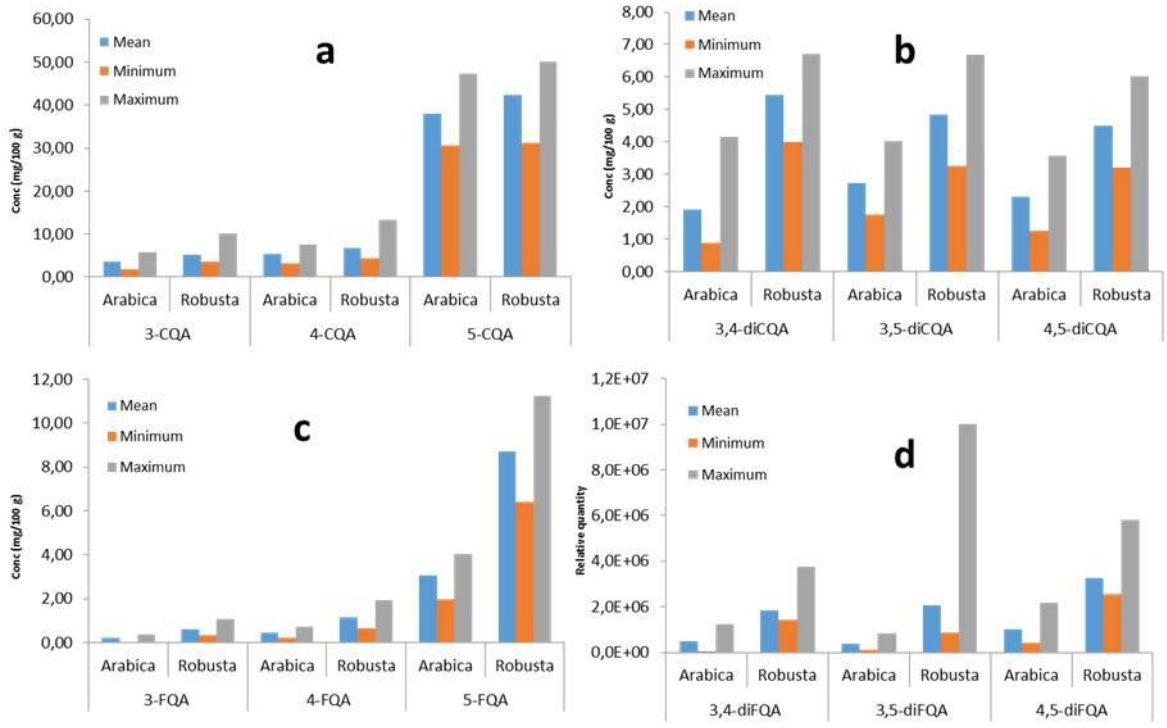


Figure 7 The four plots (a – d) represent the mean average, minimum and maximum values of Arabica and Robusta CQA and FQA with mean (n = 54). Charts (a – c) are expressed in mg/100 g, while (d) depicts the relative quantity of DiFQA. Plot (b) represents the DiCQA, while plot (c) represents the FQAs and plot (d) indicates relative quantity of the DiFQA.

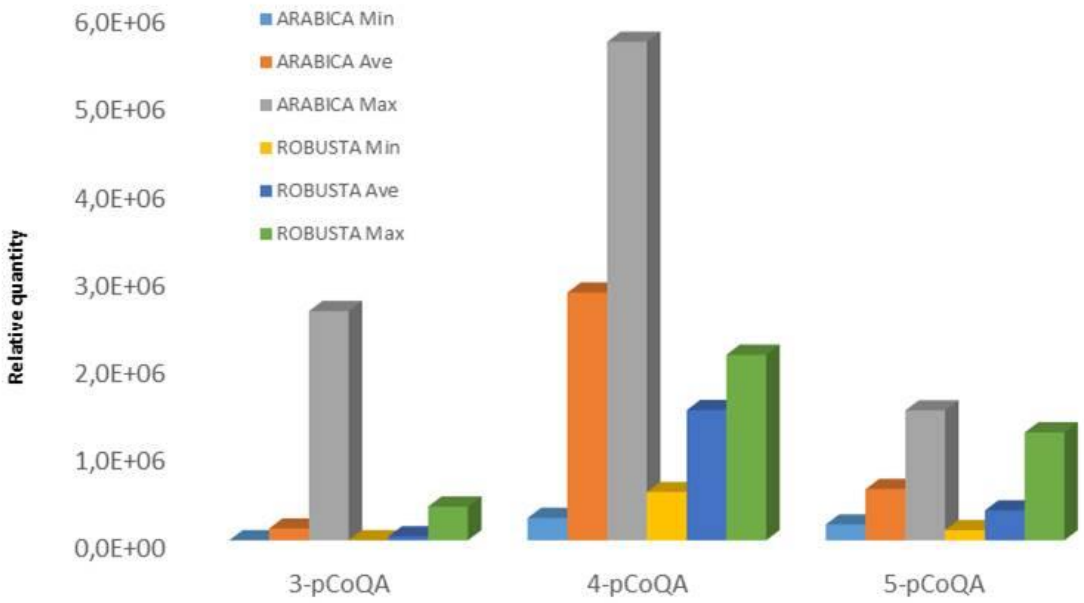


Figure 8 This histogram represents the mean, minimum and maximum values of Arabica and Robusta relative quantities of *p*CoQA. Mean (n = 54). This chart indicates that the 4-*p*CoQA values are significantly higher than in 3- *p*CoQA and 5-*p*CoQA.

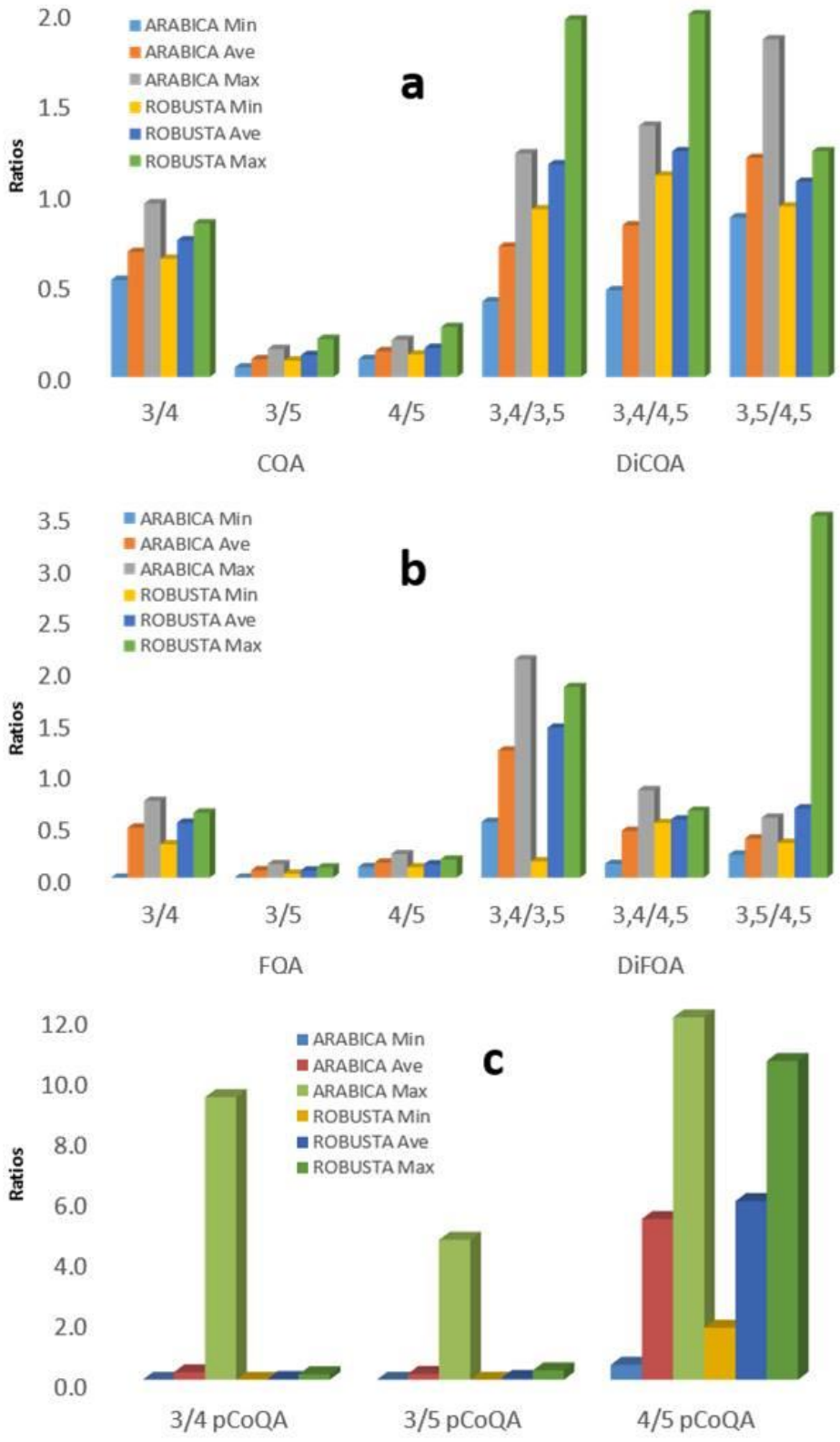


Figure 9 The plots (a – c) represent ratios of regio-isomers of x to y of the average mean, minimum and maximum of coffee Arabica and Robusta CGAs with mean (n = 54). Plot (a) indicates the mono-CQA and DiCQA values, while plot (b) shows the mono-FQA and DiFQA values and plot (c) representing the *p*CoQA values.

Table 4 P-values ($p < 0.05$) and Pearson correlation coefficient showing statistical significance of the comparison between the individual CGA 5-isomer and other coffee CGA isomers in coffee Robusta and Arabica.

Pearson Coefficient Arabica	3-CQA	4-CQA	5-CQA	3,4-DiCQA	3,5-DiCQA	4,5-DiCQA	3-FQA	4-FQA	5-FQA	3,4-DiFQA	3,5-DiFQA	4,5-DiFQA	3-pCoQA	4-pCoQA	5-pCoQA	5-SQA
3-CQA	1.00	0.87		0.63				0.75								
4-CQA	0.87	1.00		0.65				0.80								
5-CQA			1.00													
3,4-DiCQA	0.63	0.65		1.00		0.61		0.74	0.60							
3,5-DiCQA					1.00	0.59										
4,5-DiCQA				0.61	0.59	1.00			0.52							
3-FQA							1.00	0.55								
4-FQA	0.75	0.80		0.74			0.55	1.00	0.72	0.56						
5-FQA				0.60		0.52		0.72	1.00							
3,4-DiFQA								0.56		1.00	0.76	0.86			0.79	
3,5-DiFQA										0.76	1.00	0.85		0.69	0.66	
4,5-DiFQA										0.86	0.85	1.00		0.56	0.72	
3-pCoQA													1.00			
4-pCoQA											0.69	0.56		1.00	0.56	
5-pCoQA										0.79	0.66	0.72		0.56	1.00	
5-SQA																1.00

Pearson Coefficient Robusta	3-CQA	4-CQA	5-CQA	3,4-DiCQA	3,5-DiCQA	4,5-DiCQA	3-FQA	4-FQA	5-FQA	3,4-DiFQA	3,5-DiFQA	4,5-DiFQA	3-pCoQA	4-pCoQA	5-pCoQA	5-SQA
3-CQA	1.00	0.98					0.86	0.91	0.72	0.91		0.87	0.92		0.90	-0.64
4-CQA	0.98	1.00					0.80	0.90	0.73	0.91		0.86	0.91		0.91	-0.65
5-CQA			1.00	0.95		0.64			0.72							
3,4-DiCQA			0.95	1.00	0.66	0.68			0.69							
3,5-DiCQA				0.66	1.00	0.96										
4,5-DiCQA			0.64	0.68	0.96	1.00										
3-FQA	0.86	0.80					1.00	0.91	0.81	0.78		0.77	0.70		0.79	-0.78
4-FQA	0.91	0.90					0.91	1.00	0.83	0.88		0.87	0.80		0.89	-0.85
5-FQA	0.72	0.73	0.72	0.69			0.81	0.83	1.00					0.67		-0.86
3,4-DiFQA	0.91	0.91					0.78	0.88		1.00		0.99	0.96		0.99	-0.64

3,5-DiFQA											1.00					
4,5-DiFQA	0.87	0.86					0.77	0.87		0.99		1.00	0.93		0.97	-0.65
3-pCoQA	0.92	0.91					0.70	0.80		0.96		0.93	1.00		0.96	
4-pCoQA									0.67					1.00		-0.82
5-pCoQA	0.90	0.91					0.79	0.89		0.99		0.97	0.96		1.00	-0.70
5-SQA	0.64	0.65					0.78	0.85	0.86	-0.64		-0.65		-0.82	-0.70	1.00

P-values

Arabica	3-CQA	4-CQA	5-CQA	3,4-DiCQA	3,5-DiCQA	4,5-DiCQA	3-FQA	4- FQA	5-FQA	3,4-DiFQA	3,5-DiFQA	4,5-DiFQA	3-pCoQA	4-pCoQA	5-pCoQA	5-SQA
3-CQA	0.00	4E-13		1E-05				2E-08								
4-CQA	4E-13	0.00		7E-06				9E-10								
5-CQA			0.00													
3,4-DiCQA	1E-05	7E-06		0.00		3E-05		6E-08	5E-05							
3,5-DiCQA					0.00	6E-05										
4,5-DiCQA				3E-05	6E-05	0.00			0.0006							
3-FQA							0.00	0.0003								
4-FQA	2E-08	9E-10		6E-08			0.0003	0.00	2E-07	0.0002						
5-FQA				5E-05		0.0006		2E-07	0.00							
3,4-DiFQA								0.0002		0.00	1E-08	2E-12			1E-09	
3,5-DiFQA										1E-08	0.00	5E-12		9E-07	3E-06	
4,5-DiFQA										2E-12	5E-12	0.00		0.0002	2E-07	
3-pCoQA													0.00			
4-pCoQA											9E-07	0.0002		0.00	0.0002	
5-pCoQA										1E-09	3E-06	2E-07		0.0002	0.00	
5-SQA																2E-277

P-values

Robusta	3-CQA	4-CQA	5-CQA	3,4-DiCQA	3,5-DiCQA	4,5-DiCQA	3-FQA	4-FQA	5-FQA	3,4-DiFQA	3,5-DiFQA	4,5-DiFQA	3-pCoQA	4-pCoQA	5-pCoQA	5-SQA
3-CQA	4E-80	2E-06					0.002	0.0003	0.02	0.0003		0.001	0.0002		0.0004	0.05
4-CQA	2E-06	4E-80					0.006	0.0004	0.02	0.0003		0.002	0.0002		0.0003	0.04
5-CQA			4E-80	2E-05		0.04			0.02							
3,4-DiCQA			2E-05	4E-80	0.04	0.03			0.03							
3,5-DiCQA				0.04	4E-80	1E-05										
4,5-DiCQA			0.04	0.03	1E-05	4E-80										

3-FQA	0.002	0.01				4E-80	0.0003	0.004	0.01		0.01	0.02		0.001	0.01
4-FQA	0.0003	0.0004				0.0003	4E-80	0.003	0.001		0.001	0.005		0.001	0.002
5-FQA	0.02	0.02	0.02	0.03		0.0043	0.003	4E-80					0.04		0.001
3,4-DiFQA	0.0003	0.0003				0.008	0.001		4E-80		3E-08	9.9E-06		1E-07	0.05
3,5-DiFQA										4E-80					
4,5-DiFQA	0.001	0.002				0.01	0.001		3E-08		4E-80	0.0001		3E-06	0.05
3-pCoQA	0.0002	0.0002				0.0229	0.01		1E-05		0.0001	4.4E-80		9E-06	
4-pCoQA								0.04					4E-80		0.003
5-pCoQA	0.0004	0.0003				0.0066	0.001		1E-07		3E-06	8.5E-06		4E-80	0.02
5-SQA	0.05	0.04				0.01	0.002	0.002	0.05		0.045		0.003	0.02	4E-80

Table 5 Limit of Detection (LOD), Limit of Quantification (LOQ) and Relative Response Factor (RRF) values of the Mono- and Di-CQA computed from the calibration curve against the analysed samples. Internal standard (ISD) values were used for the RRF calculation.

	3-CQA	4-CQA	5-CQA	3,4-DiCQA	3,5-DiCQA	4,5-DiCQA
LOD	1.22	1.19	4.55	1.28	0.39	0.92
LOQ	3.67	3.56	13.66	3.85	1.16	2.77
RRF	0.99	0.99	1.04	1.01	1.01	0.99

3.8. Statistical analysis

In Table 4 and Figures 7, 8 and 9, the mean, minimum and maximum values of all the CGAs with their respective regio-isomers were computed and summarized to determine and allows comparison between the Arabica and Robusta coffee CGAs quantities.

3.9. Statistical analysis using P-values and Pearson correlation coefficient

A close inspection of the data distribution shows that all CGAs show a non-Gaussian distribution. Likewise, inspection of the uniformity of variance clearly indicates that the dataset is not homoscedastic for all the analytes investigated. These two observations clearly forbid a simple parametric test on the data as carried out in all previously published work, but rather requires non-parametric data treatment procedures such as using a logarithmic data treatment prior to Anova (Granato, de Araújo Calado & Jarvis, 2014).

Since the 5-isomer has been identified to be the most abundant of the CGAs usually in coffee, it is very important to consider its correlation with other CGAs present in the coffee. Table 4 statistically shows the p-values ($p < 0.05$) and Pearson correlation coefficient of each of the 5-isomer with respect to all other CGA isomers to determine the level of significance using t-test paired two samples for mean and most importantly compare Arabica to Robusta. This was carried out by comparing the 5-CQA isomer with all other CQAs (3-CQA, 4-CQA, 3,4-DiCQA, 3,5-DiCQA and 4,5-DiCQA), also 5-FQA isomer was compared with (3-FQA, 4-FQA, 3,4-DiFQA, 3,5-DiFQA and 4,5-DiFQA) and 5- *p*CoQA isomer compared with 3-*p*CoQA and 4-*p*CoQA. The values were computed (Table 4) and represented in Fig. 10 on histogram plot where 10(a) represents the 5-CQA compared with other CQAs, 10(b) shows the relationship between 5-FQA and other FQAs, and 10(c) represents 5-*p*CoQA in relation to 3- and 4-*p*CoQA.

From the values provided in Table 4 and Fig. 10, all the p-values for both the Arabica and Robusta coffees show different levels of significance using $p < 0.05$ as mentioned. This

indicates that the comparison between the 5-isomers and their respective CGA isomers are statistically significant. Furthermore, it was observed that the Arabica values are less compared to Robusta by a factor of two or more, indicating that the Robusta coffee probably contains more of the indicators that could use to differentiate between the two coffee varieties.

Considering the Pearson correlation coefficient as a statistical tool, the highlight of this is the Pearson correlation coefficient, which shows some disparities in terms of correlation among the CGAs and their regio-isomers when compared with their respective 5-isomers. Most significantly, the Robusta 5-CQA when compared with 3,4-DiCQA Robusta shows high positive correlation with a coefficient of 0.917 which is closer to +1 as against the Arabica which has a coefficient of 0.382. Similarly, the 5-*p*CoQA when compared with 3-*p*CoQA has 0.962 for the Robusta as against the Arabica with 0.221. From this, it could be inferred that the 3,4-DiCQA and 3-*p*CoQA could be possible useful tools for the development of an authentic and reliable biomarker for accurate discrimination of Robusta from Arabica.

3.10. Linear discriminant analysis model

Due to the low number of samples and the low representativeness of Robusta samples, we did not recommend the exclusion of this analysis to discriminate. Following the results above we decided to build a model based on linear discriminant analysis (LDA) allowing clear differentiation of Arabica and Robusta coffee beans. Based on the quantitative data we chose fifteen CGAs as main contributing features that have revealed significant variations between the two coffee varieties. A PCA identifies alternatively these fifteen CGAs among the 30 most important features. The resulting LDA results in a 100% discrimination with the plot shown in Fig. 11. Validation of the model was achieved through measurement of additional six samples (three Robusta and three Arabica), which were all correctly assigned.

Figure 10 Pearson correlation coefficient with a fixed substituent showing statistical significance of the comparison between individual 5-isomer CGA and other isomers in Robusta and Arabica green bean samples.

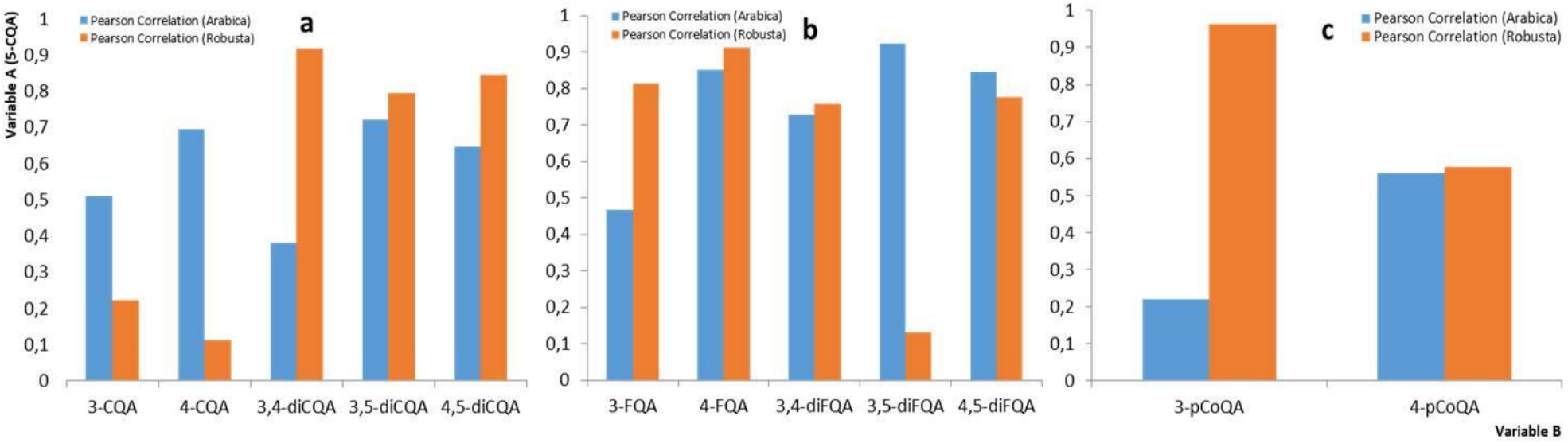
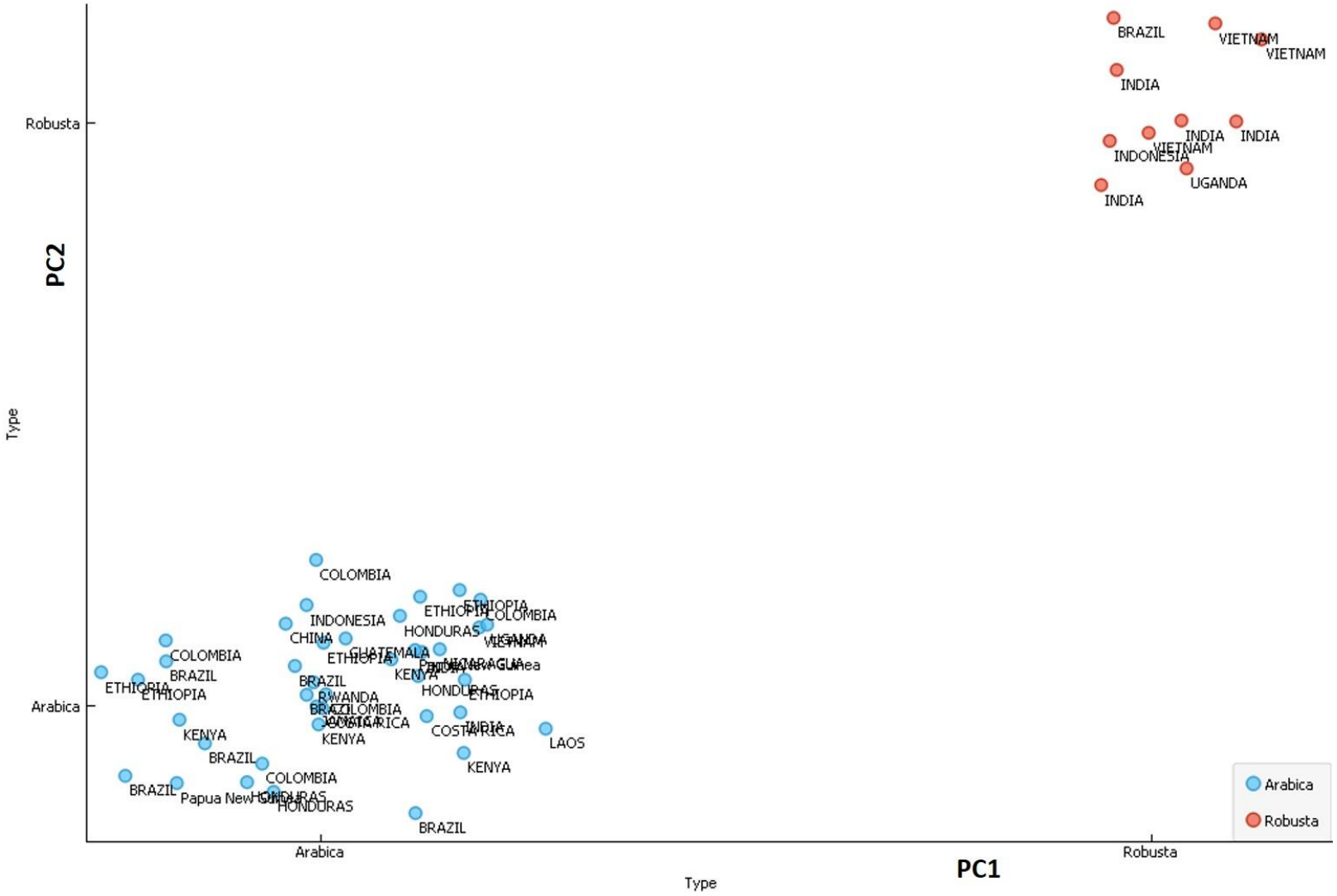


Figure 11 Linear discriminant analysis score plot for 100% differentiation between Robusta and Arabica beans based on twelve most contributing CGA.



4. CONCLUSION

Following NMR based validation of coffee variety authenticity, LC-MS based quantification of key chlorogenic acids in Arabica and Robusta coffee was carried out. While mono-caffeoyl quinic acids showed no significant variation among coffee varieties, FQAs and Di-CQAs showed statistically significant variations if Robusta and Arabica beans were compared. However, we suggest that such variations could be used for authentication purposes in only green coffee beans along with the established recently criticized 16-*O*-methylcafestol method (Gunning et al, 2018). Identification of key CGAs showing significant variations allows establishment of a reliable multivariate statistical method for coffee authentication only in green coffee beans. Statistical evaluation of the data furthermore shows a strong correlation of different classes of CGAs in all samples analysed. The observation that some CGA distribution in coffee shows neither a Gaussian nor homoscedastic behaviour clearly illustrates their role in plant protection with plants under stress presumably producing increase demand of CGAs.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

ACKNOWLEDGEMENT

The authors would like to thank Deutscher Kaffeeverband, Hamburg for providing the coffee bean samples used in this study, Dr. Roy D'Souza for the statistical analysis support and Ms. Anja Müller for the technical support.

APPENDIX A: SUPPLEMENTARY DATA

Supplementary data is available online at *https://doi.org/10.1016/j.foodres.2019.108544*

REFERENCES

Alonso-Salces, R. M., Serra, F., Remero, F., & Heberger, K. (2009). Botanical and geographical characterization of green coffee (*Coffea arabica* and *Coffea canephora*): Chemometric evaluation of phenolic and methylxanthine contents. *Journal of Agricultural and Food Chemistry*, 57(10), 4224–4235.

Babova, O., Occhipinti, A., & Maffei, M. E. (2016). Chemical partitioning and antioxidant capacity of green coffee (*Coffea arabica* and *Coffea canephora*) of different geographical origin. *Phytochemistry*, 123, 33–39.

Clifford, M. N. (1999). Chlorogenic acids and other cinnamates - nature, occurrence and dietary burden. *Journal of the Science of Food and Agriculture*, 79(3), 362–372.

Clifford, M. N., Jaganath, I. B., Ludwig, I. A., & Crozier, A. (2017). Chlorogenic acids and the acyl-quinic acids: Discovery, biosynthesis, bioavailability and bioactivity. *Natural Product Reports*, 34(12), 1391–1421.

Clifford, M. N., Johnston, K. L., Knight, S., & Kuhnert, N. (2003). Hierarchical scheme for LC-MS^n identification of chlorogenic acids. *Journal of Agricultural and Food Chemistry*, 51(10), 2900–2911.

Clifford, M. N., Kellard, B., & Ah-sing, E. (1989). Caffeoyltyrosine from green robusta coffee beans. *Phytochemistry*, 28(7), 1989–1990.

Clifford, M. N., Knight, S., Surucu, B., & Kuhnert, N. (2006). Characterization by LC-MS^n of four new classes of chlorogenic acids in green coffee beans: Dimethoxycinnamoylquinic acids, diferuloylquinic acids, caffeoyl-dimethoxycinnamoylquinic acids, and feruloyl-dimethoxy cinnamoylquinic acids. *Journal of Agricultural and Food Chemistry*, 54(6), 1957-1969.

D'Souza, R. N., Grimbs, S., Behrends, B., Bernaert, H., Ullrich, M. S., & Kuhnert, N. (2017). Origin-based polyphenolic fingerprinting of Theobroma cacao in unfermented and fermented beans. *Food Research International*, 99, 550–559.

D'Amelio, N., Papamokos, G., Dreyer, J., Carloni, P., & Navarini, L (2015) *Food Biophysics*, 10, 235.

De Menezes, H. C. (1994). The relationship between the state of maturity of raw coffee beans and the isomers of caffeoylquinic acid. *Food Chemistry*, 50(3), 293–296.

Farah, A., & Donangelo, C. M. (2006). Phenolic compounds in coffee. *Brazilian Journal of Plant Physiology*, 18(1), 23-36.

Gil, M., & Wianowska, D. (2017). Chlorogenic acids their properties, occurrence and analysis. *Annales* 72(1), 61.

Granato, D., de Araújo Calado, V. M., & Jarvis, B. (2014). Observations on the use of statistical methods in food science and technology. *Food Research International*, 55, 137–149.

Grosso, G., Stepaniak, U., Topor-Madry, R., Szafraniec, K., & Pajak, A. (2014). Estimated dietary intake and major food sources of polyphenols in the polish arm of the HAPIEE study. *Nutrition*, 30(11−12), 1398–1403.

Gunning, Y., Defernez, M., Watson, A. D., Beadman, N., Colquhoun, I. J., Le Gall, G., & Kemsley, E. K. (2018). 16-*O*-methylcafestol is present in ground roast arabica coffees: Implications for authenticity testing. *Food Chemistry*, 248, 52–60.

Hu, Y., Pan, Z. J., Liao, W., Li, J., Gruget, P., Kitts, D. D., & Lu, X. (2016). Determination of antioxidant capacity and phenolic content of chocolate by attenuated total reflectance-Fourier transformed-infrared spectroscopy. *Food Chemistry*, 202, 254–261.

Jaiswal, R., & Kuhnert, N. (2010). Hierarchical scheme for liquid chromatography/ multistage spectrometric identification of 3,4,5-triacyl chlorogenic acids in green robusta coffee beans. *Rapid Communications in Mass Spectrometry*, 24(15), 2283–2294.

Jaiswal, R., Matei, M. F., Subedi, P., & Kuhnert, N. (2014). Does roasted coffee contain chlorogenic acid lactones or/and cinnamoylshikimate esters? *Food Research International*, 61, 214–227.

Jaiswal, R., Patras, M. A., Eravuchira, P. J., & Kuhnert, N. (2010). Profile and characterization of the chlorogenic acids in green robusta coffee beans by LC-MS^n: Identification of seven new classes of compounds. *Journal of Agricultural and Food Chemistry*, 58(15), 8722–8737.

Kleinwächter, M., Bytof, G., & Selmar, D. (2014). Coffee beans and processing. *Coffee in health and disease prevention*, pp. 73–81.

Ky, C.-L., Louarn, J., Dussert, S., Guyot, B., Hamon, S., & Noirot, M. (2001). Caffeine, trigonelline, chlorogenic acids and sucrose diversity in wild *Coffea arabica L.* and *C. canephora* P. accessions. *Food Chemistry, 75(2), 223–230.*

Lee, S., Jaiswal, R., & Kuhnert, N. (2016). Analysis on different grades of highly-rated Jamaica blue mountain coffees compared to easily available originated coffee beans. *SCIREA Journal of Food* 1(2), 15–27.

Möller, B., & Herrmann, K. (1983). Quinic acid esters of hydroxycinnamic acids in stone and pome fruit. *Phytochemistry*, 22(2), 477–481.

Pandey, K. B., & Rizvi, S. I. (2009). Plant polyphenols as dietary antioxidants in human health and disease. *Oxidative Medicine and Cellular Longevity*, 2(5), 270–278.

Pękal, A., Drózdz, P., Biesaga, M., & Pyrzynska, K. (2011). Evaluation of the antioxidant properties of fruit and flavoured black teas. *European Journal of Nutrition*, 50(8), 681–688.

Rostagno, M. A., Celeghini, R. M. S., Debien, I. C. N., Nogueira, G. C., & Meireles, M. A. A. (2014). Phenolic compounds in coffee compared to other beverages. *Coffee in Health and Disease Prevention,* 1, 137–142.

Schievano, E., Finotello, C., De Angelis, E., Mammi, S., & Navarini, L. (2014). Rapid authentication of coffee blends and quantification of 16-*O*-methylcafestol in roasted coffee beans by nuclear magnetic resonance. *Journal of Agricultural and Food Chemistry*, 62(51), 12309–12314.

Svilaas, A., Sakhi, A. K., Andersen, L. F., Svilaas, T., Ström, E. C., Jacobs, D. R., Jr., Blomhoff, R. (2004). Intakes of antioxidants in coffee, wine, and vegetables are correlated with plasma carotenoids in humans. *Journal of Nutrition*, 134(3), 562–567.

Upadhyay, R., & Mohan Rao, L. J. (2013). An outlook on chlorogenic acids-occurrence, chemistry, technology, and biological activities. *Critical Reviews in Food Science and Nutrition*, 53(9), 968–984.

SUPPLEMENTARY DATA

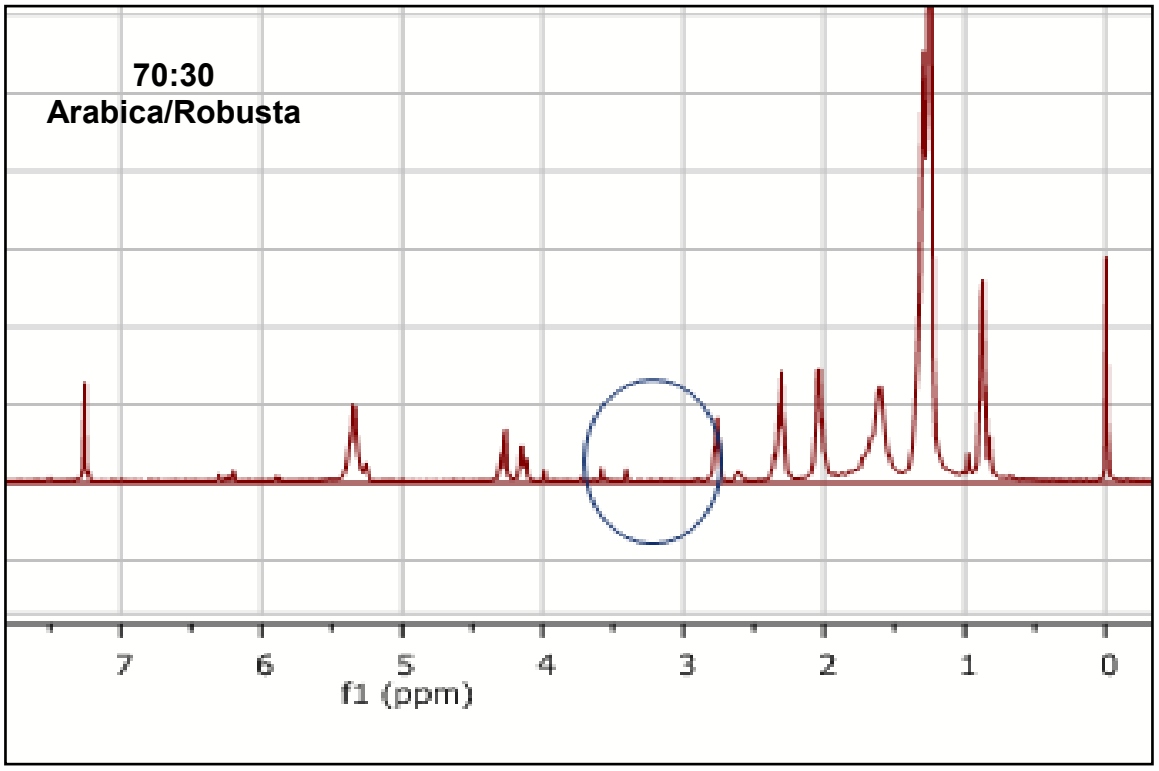


Figure A1 ^{1}H-NMR spectrum of Arabica mixed with Robusta in ratio 70/30, which clearly indicates that the biomarker 16-*O*-Methylcafestol lipid content is absent at position 3.17 ppm.

Table A1 List of the different coffee Arabica and Robusta bean samples used.

S/N	Label name	Coffee bean Samples	Species	Origin
1	H1	Brasil A	*C.arabica*	Brazil
2	H2	Brasil B	*C.arabica*	Brazil
3	H3	India Plantation A	*C.arabica*	India
4	H4	India Plantation AB	*C.arabica*	India
5	H6	Guatemala	*C.arabica*	Guatemala
6	H7	Honduras shg A	*C.arabica*	Honduras
7	H8	Colombia Excelso	*C.arabica*	Colombia
8	H9	Colombia Rio Negro	*C.arabica*	Colombia
9	H11	Ethiopian Djimmah	*C.arabica*	Ethiopia
10	H12	Ethiopian Sidamo	*C.arabica*	Ethiopia
11	H13	Ethiopian Yirgacheffe	*C.arabica*	Ethiopia
12	H14	PNG C grade Sigri	*C.arabica*	Papua New Guinea
13	H16	India Parchment AB	*C.canephora*	India
14	H17	India Parchment C	*C.canephora*	India

15	H18	Honduras shg B	*C.arabica*	Honduras
16	H19	Ethiopian Yirgacheffe Gelana Abaya	*C.arabica*	Ethiopia
17	H20	Colombia	*C.arabica*	Colombia
18	H21	Brasil Sertaozinho	*C.arabica*	Brazil
19	H31	Costa Rica	*C.arabica*	Costa Rica
20	H32	Vietnam Arabica	*C.arabica*	Vietnam
21	H33	Brasil Grinder	*C.arabica*	Brazil
22	H34	Uganda Arabica	*C.arabica*	Uganda
23	H35	Vietnam Rob. Gr.1	*C.canephora*	Vietnam
24	H36	Vietnam Rob. Gr.2	*C.canephora*	Vietnam
25	H37	Brasil Conillon Robusta	*C.canephora*	Brazil
26	H38	Vietnam Robusta	*C.canephora*	Vietnam
27	H39	Indonesian EK 1	*C.canephora*	Indonesia
28	H40	Indian washed Robusta	*C.canephora*	India
29	H41	Laos washed Arabica	*C.arabica*	Laos
30	BGC	Brasil Arabica	*C.arabica*	Brazil
31	H44	Ethiopian Limmu gr.2	*C.arabica*	Ethiopia
32	H45	Costa Rica	*C.arabica*	Costa Rica
33	H46	Laos Arabica	*C.arabica*	Laos
34	H47	Uganda Robusta	*C.canephora*	Uganda
35	H48	PNG Arabica	*C.arabica*	Papua New Guinea
36	H49	Kenya Arabica	*C.arabica*	Kenya
37	H50	Ethiopian Arabica	*C.arabica*	Ethiopia
38	H51	Ruanda Arabica	*C.arabica*	Rwanda
39	H52	Colombia Arabica	*C.arabica*	Colombia
40	H53	China Arabica	*C.arabica*	China
41	H54	Peru Arabica	*C.arabica*	Peru
42	BG	Brazil Green	*C.arabica*	Brazil
43	NGC	Nicaragua Green	*C.arabica*	Nicaragua
44	CGC	Colombia Green	*C.arabica*	Colombia
45	RIG	Indian Robusta Green	*C.canephora*	Indian
46	KGC	Kenya Green	*C.arabica*	Kenya
47	JGC	Jamaica Green	*C.arabica*	Jamaica
48	HBG	Honduras Green	*C.arabica*	Honduras
49	ECU	Ecuador Old Green	*C.canephora*	Ecuador
50	ECG	Ecuador Green	*C.canephora*	Ecuador
51	TZG	Tanzania Green	*C.canephora*	Tanzania
52	ING	Indonesia Green	*C.canephora*	Indonesia
53	GUG	Guatemala Green	*C.arabica*	Guatemala
54	ETG	Ethiopia Green	*C.arabica*	Ethiopia

Chapter 3

Classification of Brazilian Coffees by CGA Profiling

Classification of Brazilian roasted coffees from different geographical origins and farming practices based on chlorogenic acid profiles

Sabur Badmos, Maotian Fu, Daniel Granato, Nikolai Kuhnert

Adapted with permission from ***Food Res. Int., 2020 134, 109218***

Online published version available at: *https://doi.org/10.1016/j.foodres.2020.109218*

ABSTRACT

Sixty-seven roasted coffee samples from different regions of Brazil cultivated using organic, conventional and biodynamic farming practices were analyzed and quantified using high performance liquid chromatography coupled with mass spectrometry, and treated with supervised (PLS-DA) and unsupervised (PCA) multivariate statistical tools. The profile of the chlorogenic acids constituents were analyzed by high resolution and tandem mass spectrometry, which allowed the identification of mono- caffeoyl-, feruloyl-, para-Coumaroylquinic acids and their respective regio-isomers. This study provides a comprehensive analysis of absolute quantitative data set of chlorogenic acids constituents (CQA, FQA and *p*CoQA isomers) in Brazilian coffee beans produced from different regions of the country. Variations in the chlorogenic acids compositions were observed if organic and conventional roasted coffee beans were compared. The use of multivariate statistical tools allowed the identification of suitable biomarkers for determining significant differences between the three coffee agricultural practices, while coffees produced from the diverse geographical regions showed no significant difference.

***Keywords*:** Coffee, Mass spectrometry, Polyphenols, Chlorogenic acids, Multivariate statistics, Chemometrics, Organic farming, Conventional farming, Geographical origin.

1. INTRODUCTION

Coffee quality depends on a variety of combined factors including the botanical variety, topographical conditions, weather conditions linked to its geographical origin, agricultural practice, storage, export, transport and processing. Ultimately, these parameters define the sensory profile and nutritional properties of the coffee beverage and hence consumer preference. Higher quality coffee preferred by consumers achieves higher prices if compared to lower quality products. In turn the sensory profile, encompassing taste aroma and color as well as nutritional properties are defined by the chemical composition of the roasted coffee beans.

Chemical variations between different botanical varieties of coffee have been addressed on many occasions, in particular the difference between Arabica and Robusta coffee (Badmos, Lee & Kuhnert, 2019; Deshpande et al, 2014; Kurzrock & Speer, 2001).

In this contribution, we address chemical variations with respect to two additional parameters: Agricultural practice with organically grown coffee compared to conventionally grown coffee and secondly geographical origin of coffee beans within the world's largest producing country Brazil. Organically grown coffee is produced without the use of crop protection agents (pesticides, herbicides fungicides etc.) and depending on the certification agency omitting the use of synthetic fertilizers, added plant hormones or sewage sludge as fertilizers. Monitoring of compliance with organic production is in the hand of governmental and consumer associations certifying organic production.

Consumer demand for organically grown plant-based foods has dramatically increased, however in the case of coffee only 3% of global production is organically grown. Consumers associate organic coffee with high quality and added health benefits with respect to absence of toxic substances and possibly presence of food constituents promoting human health (Tajik,

Tajik, Mack & Enck, 2017). The organic farming is regulated both in European countries (i.e., EC 834/07 and EC 889/ 08) and in Brazil (law number 10.831 from 23rd December 2003) to ensure that consumers buy authentic products. As coffee has a considerable high price in the international market, especially organic coffee, this commodity needs to be constantly analyzed for authenticity since tracing its origin is an important task (Granato, de Magalhães Carrapeiro, Fogliano & van Ruth, 2016). Coffee is cultivated in several provinces of Brazil ranging from the Southern Amazonian region to the South of Brazil. However, perceived quality and hence prices among Brazilian coffee origins vary dramatically with coffee from the Southern provinces perceived as inferior thus achieving lower prices.

Secondly, we address the issue of geographical origins of coffee beans choosing Brazil as the largest global coffee producer (43.2%) and exporter. Brazil is the largest coffee producer, totaling 60.2 million sacks in 2018, in which 74% are *Coffea arabica* and 26% are *Coffea canephora*. The main producing states in Brazil are Minas Gerais (MG, 51.66% of the total volume produced), Espírito Santo (ES, 24.42%), São Paulo (SP, 9.97%) and Paraná (PR, 1.83%). From this amount, roughly 35 million sacks of green coffee are exported to many countries, especially USA, Germany, Italy, Japan, Belgium, Turkey, Canada, France, Spain and United Kingdom (Barros and Chanda, 2019).

Although some attempts have been made to trace the geographical origin and cultivation system of green and roasted coffee using different analytical approaches, it is still a challenge to be sure which method should be used for such differentiation (Babova, Occhipinti & Maffei, 2016; Cloete, Šmit, Minnis-Ndimba, Vavpetič, du Plessis, le Roux & Pelicon, 2019; Giraudo et al, 2019; Gutiérrez Ortiz et al, 2019; Peng, Zhang, Song Cai, Wang & Granato, 2019; Worku et al, 2019). Recently, Monteiro et al, (2018) attempted to differentiate Brazilian roasted coffee based on near infrared (NIR) spectroscopy and proton-transfer reaction-mass spectrometry but neither approaches seem to be useful to differentiate the geographical origin as correctly classification

rates were rather low (< 80%). On the other hand, using partial least squares-discriminant analysis (PLS-DA) to assess NIR data, 92% CONV and 84% ORG samples were correctly differentiated. In another analytical attempt to trace the origin of Brazilian roasted coffee (Peng et al, 2019), isotope ratio mass spectrometry coupled with supervised chemometrics were used, namely linear discriminant analysis, LDA, k-nearest neighbours, k-NN, and support vector machines, SVM. Results showed that neither the geographical location (e.g., PR, SP, MG, and ES) nor the cultivation systems (e.g., ORG and CONV) could be differentiated. Similarly, Cloete et al, (2019) analyzed the elemental concentration of roasted organic coffees from Ethiopia, Colombia, Honduras and Mexico and found that there was no significant association between elemental levels and geographical origin.

In this context, the most studied bioactive compounds in coffee are caffeine, trigonelline and chlorogenic acids. By definition, chlorogenic acids are hydroxycinnamic acid esters of quinic acid produced as secondary plant metabolites by the coffee plant in its seeds (Abrankó and Clifford, 2017). A typical Arabica green coffee bean contains around 45 different derivatives of chlorogenic acids as series of regio-isomers, while in a Robusta bean around 80 different derivatives have been reported (Jaiswal & Kuhnert, 2011), which in both cases roughly double during roasting by epimerization, elimination, acyl-migration and addition-elimination reactions during coffee processing (Jaiswal, Matei, Golon, Witt & Kuhnert, 2012; Kuhnert, Jaiswal, Matei, Sovdat & Deshpande, 2010; Matei, Jaiswal & Kuhnert, 2012). Chlorogenic acids have most notably been shown to be responsible for preventive effect in Type-II diabetes and act as hepatoprotective substances. Additionally, they have been shown to possess antiviral (Gamaleldin Elsadig Karar, Matei, Jaiswal, Illenberger & Kuhnert, 2016) and antibacterial (Naveed et al, 2018) properties.

Within this contribution, we investigate the chlorogenic acid profile by liquid chromatography coupled to mass spectrometry of sixty-seven (67) samples of roasted Arabica coffee from Brazil

from eight different provinces with a diversity of conventionally grown and organically grown coffee farms. The study is motivated by the assumption that chlorogenic acids are the key coffee constituents responsible for beneficial health effects. Similar exercises have been carried out using alternative analytical techniques and coffee analytes such as proton transfer mass spectrometry and isotope ratio mass spectrometry demonstrating a statistically significant difference between conventional and organically grown coffee along smaller variations based on geographical origin (Monteiro et al, 2018; Peng et al, 2019).

2. MATERIALS AND METHODS

2.1. Coffee samples

A total of 67 samples were investigated in this study, 22 were produced using the organic (ORG), 3 were planted using biodynamic (BIO) and 42 were by conventional (CONV) systems. ORG samples were certified by third parties certifying companies and all contained the seal of authenticity issued by the Brazilian Ministry of Agriculture, Livestock and Supply (MAPA). During the samples analysis, grams (g) of roasted coffee were used. This is accordance with the reported evidences that there is little or no variations with regards to the moisture content of roasted coffee as against the green coffee (Corrêa et al, 2016; de Oliveira, Corrêa, Reis & de Oliveira, 2015).

From these samples, 20 were produced in São Paulo (SP), 17 in Minas Gerais (MG), 6 in Espírito Santo (ES), 2 in Bahia (BA), 14 in Paraná (PR) and 8 coffees with different origins (blends including Rondonia [RO] samples). A certificate of geographic origin and cultivation system was issued by the Brazilian Association of Coffee Producers (ABIC). Comprehensive details of the coffee beans are provided in Table 1 of the supplementary information. Most coffees (63 out of 67) belonged to *Coffea arabica* and 4 were a blend between *C. arabica* and *C. canephora* with unknown proportion. Coffee beans were roasted each in a coffee roaster at 215 ± 5 °C/11 min and then milled to 60 Tyler mesh prior to analysis.

2.2. Chemicals and reagents

Acetonitrile (LC-MS grade) was used for the HPLC sequence measurements (Acetonitrile for UHPLC gradient, AppliChem). Methanol and formic acid (used as a mobile phase modifier and as calibration buffer) were likewise of HPLC grade (both purchased from Sigma Aldrich, Germany), while 5-CQA, 4-CQA, 3-CQA, 3,5-diCQA, 3,4-diCQA and 4,5-diCQA used as chlorogenic acid standards were purchased from PhytoLab (Vestenbergsgreuth, Germany). The

Sixty-seven different roasted coffee beans (10 g each) from diverse regions of Brazil listed above were ground to fine-powder. For each sample, 250 mg of the powdered bean were weighed and extracted with 70% methanol by sonication (with magnetic stirrer) for 15 min and vortex for another 15 min. The extract was filtered through a Chromafil PTFE membrane filter (Bremen, Germany) and used for LC-MS^n sequence run.

2.3. LC-ESI-TOF-MS (High resolution mass spectrometry) and quantification

The Agilent 1200 Series Liquid chromatography equipment (Bremen, Germany) consists of a binary pump, an auto-sampler with 100 µL capacity loop and a UV-Vis detector with a light-pipe flow cell was used for the compounds' separation and detection. The UV detector was set at 320 and 254 nm to measure the separated compounds. The compounds chromatographic separation was achieved using a 5 µm diphenyl column with an inner diameter of 250 × 3 mm (Varian, Darmstadt, Germany) and a column guard (inner diameter of 5 mm × 3 mm). This was connected to Bruker Daltonics micro-TOF Focus mass spectrometer (Bremen, Germany) equipped with an electrospray ionization source (ESI). Measurements were carried out in the negative ion mode. The internal calibration was attained by injecting 0.1 M sodium formate solution at 0.10 mL/min through the six-port valve using enhanced quadratic mode prior to each chromatographic run. Mass calibration was carried out using an enhanced quadratic fitting in a mass range from m/z 100 – 1200. Solvents A and B used were water/formic acid (1000:0.05 v/v) and methanol respectively. The flow rate of the solvents was set to 500 µL /min. However, linear gradient profile was used from 10% B to 70% B in 60 min and a further 10 min was assigned for the gradient to re-equilibrate from 90% B to 10% B for the subsequent run. The sample filtered extract (2 µL) was thereafter injected into the system for analysis. The Brukers' Hystar software was used for programming the system set-up.

Quantification of chlorogenic acids was carried out using the UV absorption at 320 nm in chromatograms. Eight-point calibration curves of authentic standards were acquired prior to quantification. Analytical parameters are provided in Table 2. Relative response factors were used for compounds where authentic standards were not available (Badmos et al, 2019; Clifford, 1999).

Table 2 Analytical parameters of HPLC-ESI-TOF-MS for the CGA detected in coffee beans.

Chlorogenic acid	3-CQA	4-CQA	5-CQA
LOD (mg/L)	1.22	1.19	4.55
LOQ (mg/L)	3.67	3.56	13.66
RRF	0.99	0.99	1.04
Analytical curve	(y = 5.51x - 6.0951)	(y = 5.4583x - 6.2424)	(y = 4.778x - 11.252)
R^2	0.9995	0.9997	0.9998

2.4. LC-ESI-MS^n (Quadrupole ion-trap mass spectrometry)

The Liquid chromatography Agilent 1100 series equipment (Karlsruhe, Germany) comprises of a binary pump, an auto sampler having a 100 µL capacity loop and a Diode Array Detector (DAD) with a range from 200 to 600 nm and the detector recorded at 254, 280 and 320 nm. The chromatographic separation was carried out using the same gradient method used in the LC-TOF analysis (2.3 above). The compounds separation was carried out with the aid of a 5 µm diphenyl column (inner diameter - 250 × 3 mm) with a column guard having an inner diameter of 5 mm × 3 mm (Varian, Darmstadt, Germany) and at 500 µL/min flow rate of solvent. The LC equipment was connected to an Ion-trap mass spectrometer fitted with an electrospray ionization source (HCT ultra, Bruker Daltonic, Bremen, Germany) operating in full scan auto-MS^n mode to obtain fragment ions. The Tandem mass spectra were acquired in Auto-MS^n negative ion mode (for smart fragmentation) with ramping of the collision energy and maximum fragmentation amplitude adjusted to 1 V within 30% to 200% range. The mass spectrometer operating conditions using negative mode was optimized using 5-caffeoylquinic

acid with a capillary temperature of 365 °C, nebulizer pressure of 10 psi and a dry gas flow-rate of 10 L/min as previously reported by (Jaiswal, Sovdat, Vivan & Kuhnert, 2010). The sample filtered extract (2 µL) was injected into the system. The Brukers' Agilent Chem-Station software was used for programming this set-up.

Compounds were identified using high-resolution and tandem mass spectrometry for the correct fragmentation patterns of each of the respective CGAs in negative ion mode (Clifford, Knight, Surucu & Kuhnert, 2006). Using the hierarchical scheme for LC-MS^n identification of CGAs, all compounds were assigned to regio-isomeric level. Besides, the chemical structures of key CGAs identified are shown in Fig. 1 (please see Chapter 2) with their respective details in Table 3. For the CGAs nomenclature, we adopted the latest recommendations (Badmos et al, 2019; Clifford, Jaganath, Ludwig & Crozier, 2017). While Fig. 1b depicts the HPLC-ESI-TOF-MS chromatogram of roasted coffee samples from Brazil in negative-ion mode.

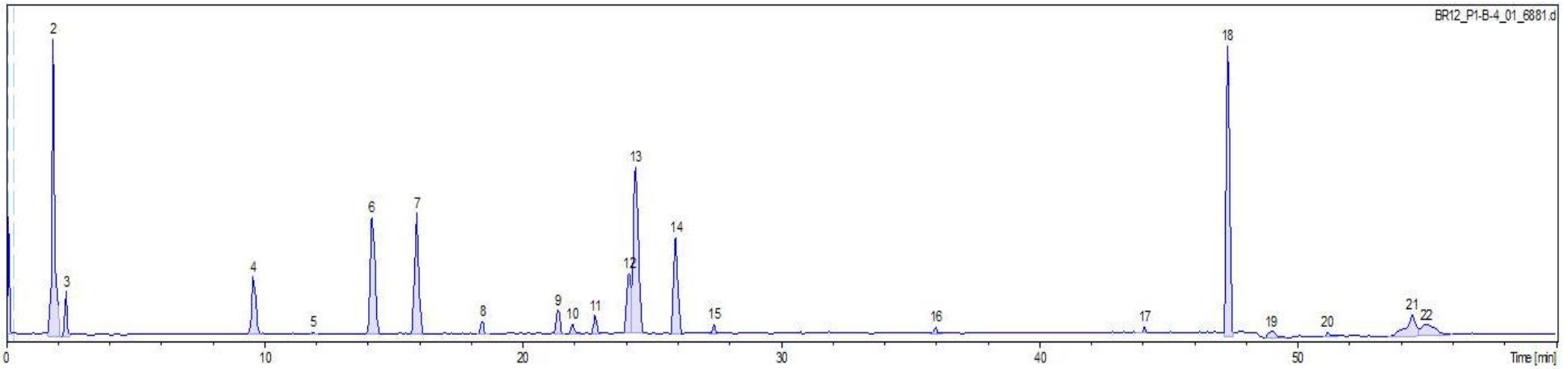
2.5. Statistical analyses

Kruskal-Wallis (otherwise known as One-way ANOVA on ranks) and Pairwise Wilcoxon Rank Sum tests were performed using the R built-in functions "kruskal.test" and "pairwise.wilcox.test" respectively with R version 3.6.3 software (R Core Team, 2019). Principal component analysis (PCA) and partial least square-discriminant analysis (PLS-DA) of the roasted coffees were carried out based on their CGAs profiles using *MetaboAnalyst 4.0* for Comprehensive and Integrative Metabolomics Data Analysis (Chong, Wishart & Xia, 2019). The pairwise comparisons analysis was performed on the variables based on the different cultivation systems and regions. In addition, in order to reduce the effects of Type-I error on the p-values ($p < 0.05$) the Benjamin Hochberg correction method was applied.

Table 3 High resolution mass (MS-TOF) of identified CGAs with their respective names, *m/z*, molecular formula and parent ions (M-H) data.

No.	Name	Abbreviation	Mol. formula	Theor. *m/z* (M-H)	Exp. *m/z* (M-H)	Error (ppm)
4	3-*O*-caffeoylquinic acid	3-CQA	$C_{16}H_{18}O_9$	353.0878	353.0886	-0.3
7	4-*O*-caffeoylquinic acid	4-CQA	$C_{16}H_{18}O_9$	353.0878	353.0884	-1.8
6	5-*O*-caffeoylquinic acid	5-CQA	$C_{16}H_{18}O_9$	353.0878	353.0899	-3.9
5	3-*O*-feruloylquinic acid	3-FQA	$C_{17}H_{20}O_9$	367.0929	367.1039	-3.1
9	4-*O*-feruloylquinic acid	4-FQA	$C_{17}H_{20}O_9$	367.0929	367.1027	-0.9
8	5-*O*-feruloylquinic acid	4-FQA)	$C_{17}H_{20}O_9$	367.0929	367.1052	-2.8
12	3-*O*-*p*-coumaroylquinic acid	3-*p*CoQA	$C_{16}H_{18}O_8$	337.0929	337.0933	-0.2
14	4-*O*-*p*-coumaroylquinic acid	4-*p*CoQA	$C_{16}H_{18}O_8$	337.0929	337.0936	2.2
13	5-*O*-*p*-coumaroylquinic acid	5-*p*CoQA	$C_{16}H_{18}O_8$	337.0929	337.0922	2.6

Figure 1b HPLC-ESI-TOF-MS chromatogram of roasted coffee samples from Brazil in negative-ion mode. Peak number 4 represents 3-CQA, 5. (3-FQA), 6. (5-CQA), 7. (4-CQA), 8. (5-FQA), 9. (4-FQA), 12. (3-*p*CoQA), 13. (5-*p*CoQA) and 14. (4-*p*CoQA).



3. RESULTS AND DISCUSSION

In this study, 67 green coffee beans were collected from different regions of Brazil and subjected to equal roasting conditions. We chose to analyze roasted material, rather than green coffee beans since this material is typically purchased by consumers and carries all properties defining coffee quality as perceived by the consumers. To comprehensively determine the variations in the roasted coffee produced from different regions of Brazil, we followed previous extraction and workup protocols for chlorogenic acids profiling method. Thereafter, the extracts were analyzed on LC-MS^n in negative ion mode. Compounds were assigned based on high-resolution mass spectrometry measurements to reveal molecular formulae and subsequently fragmented on tandem-MS for structural elucidation as previously described (Clifford, Johnston, Knight & Kuhnert, 2003; Clifford et al, 2006; Clifford et al, 2017; Jaiswal, Patras, Eravuchira & Kuhnert, 2010). Quantitative analysis was carried out as previously described using authentic standards if available or accepted relative response factors from UV-VIS data. The limit of quantification, limit of detection, relative response factor (RRF), R^2 (coefficient of determination) and calibration curve values were calculated. The corresponding analytical parameters of HPLC-ESI-TOF-MS for the chlorogenic acids detected in coffee beans are provided in Table 2.

The resulting LC-MS^n data were analyzed and subjected to multivariate statistical analysis using data with m/z range of 100-1200. This multivariate statistical analysis allows for the exploration of relationships between many different types of attributes. Thus, this tool possesses the ability to reduce data dimensionality, identify similarities and as well discriminate samples with a large data set without losing any vital information therein. Consequently, PCA and PLS-DA were employed in this regard. PCA is an unsupervised multivariate statistical analysis tool focusing on maximized correlation, while PLS-DA is a supervised method because the supervised algorithm works best for labelled data.

3.1. Statistical means and standard deviation of the roasted coffees

The mean, standard deviation, minimum and maximum values (n = 67) of all the organic, biodynamic and conventional coffee samples were computed and represented in Table 4; while the p-values ($p < 0.05$) were indicated in Table 5. The CQA values indicated that the 5-CQA and 4-CQA are significantly higher than the 3-CQA. Furthermore, it was observed from the plots (Figs. 3a and 3b) that there is a significant difference ($p < 0.05$) between the means of organic and conventional samples as shown in Table 5. In the FQAs, 5-FQA values were observed to be higher than the 4-FQA values. It is worthy to mention here that 3-FQA was scantly detected in many of the samples and thus it was excluded from the analyzed data set to avoid huge errors and/or variations.

As for para-coumaroylquinic acids, the 5-*p*CoQA shows higher values when compared to the 3-*p*CoQA and 4-*p*CoQA isomers. With this, it evidently shows that the 5-*p*CoQA could be used as a good biomarker to differentiate the organic from the conventional samples with respect to their cultivation patterns.

In summary, the statistical data shows that 4-CQA, 5-FQA and 5-*p*CoQA could be used to differentiate between organic and conventional samples based on their chlorogenic acids constituents using quantitative analysis. While no significant difference was observed between organic and biodynamic cultivated samples.

3.2. PCA of the roasted coffee samples

Principal component analysis (PCA) was employed for general classification of the coffee samples using *MetaboAnalyst 4.0* data analysis software (Chong, Wishart & Xia, 2019). PCA is an unsupervised statistical tool which detects and classifies variances in a specific data set and also provides an overview of the diversity in the samples under study. From this, two plots were generated which are scores and loading plots. The scores plot indicates the relationship

among the samples with a quantitative value to determine the variance; while the loading plot specifically identifies the key variables responsible for the variances observed in the samples data set and the variables' influence for the separation of the groups.

After the data set were treated with this statistical method, PCA scores plot shown in Fig. 2a shows a good separation between the coffee samples cultivated from different regions of Brazil with a score of 97.2%. This score is above the acceptable score range (40 – 50%). For the significant compounds identified in the loading plot (Fig. 2b), the variable compounds considered in the range of m/z 100 - 1200 were 3-CQA, 4-CQA, 5-CQA, 4-FQA, 5-FQA, 3-*p*CoQA, 4-*p*CoQA and 5-*p*CoQA. These eight compounds constitute the key biomarkers as the principal components that contribute to the variances in the sample data set.

Table 4 The mean average, standard deviation, minimum and maximum values of the CGAs constituents of the roasted coffee beans, mean (n=67) expressed in mg/g.

Compound (mg/g)		**3-CQA**			**4-CQA**			**5-CQA**			**4-FQA**	
Cultivation	**Bio**	**Org**	**Conv**	**Bio**	**Org**	**Conv**	**Bio**	**Org**	**Conv**	**Bio**	**Org**	**Conv**
Mean±SD	0.38±0.3	1.40±2.5	4.21±2.7	0.87±0.8	2.95+5.0	8.26±5.6	0.84±0.7	3.27±5.7	9.12±6.3	0.06±0.06	0.31±0.4	0.83±0.5
Minimum	0.10	0.06	0.10	0.26	0.13	0.03	0.29	0.13	0.07	0.01	0.01	0.01
Maximum	8.31	9.73	10.26	18.94	19.28	20.40	17.16	21.89	22.96	1.33	1.61	2.04
		5-FQA			**3-*p*CoQA**			**4-*p*CoQA**			**5-*p*CoQA**	
	Bio	**Org**	**Conv**	**Bio**	**Org**	**Conv**	**Bio**	**Org**	**Conv**	**Bio**	**Org**	**Conv**
Mean±SD	0.16±0.1	0.79±1.4	2.37±1.52	0.10±0.08	0.37±0.4	0.97±0.6	0.16±0.1	0.42±0.4	0.96±0.7	1.37±1.1	5.02±7.1	13.25±8.7
Minimum	0.06	0.04	0.04	0.03	0.02	0.03	0.03	0.04	0.03	0.52	0.25	0.14
Maximum	3.29	4.79	6.11	1.87	1.67	3.34	3.21	1.51	3.70	27.03	27.81	31.97

Table 5a P-values of the different cultivation systems of coffee

3-CQA	Bio	Conv	**5-FQA**	Bio	Conv
Conv	0.02	n/a	Conv	0.02	n/a
Org	0.48	0.001	Org	0.80	0.001
4-CQA			**3-pCoQA**		
Conv	0.02	n/a	Conv	0.04	n/a
Org	0.25	0.003	Org	0.63	0.01
5-CQA			**4-pCoQA**		
Conv	0.01	n/a	Conv	0.01	n/a
Org	0.32	0.001	Org	0.13	0.01
4-FQA			**5-pCoQA**		
Conv	0.03	n/a	Conv	0.01	n/a
Org	0.53	0.001	Org	0.17	0.001

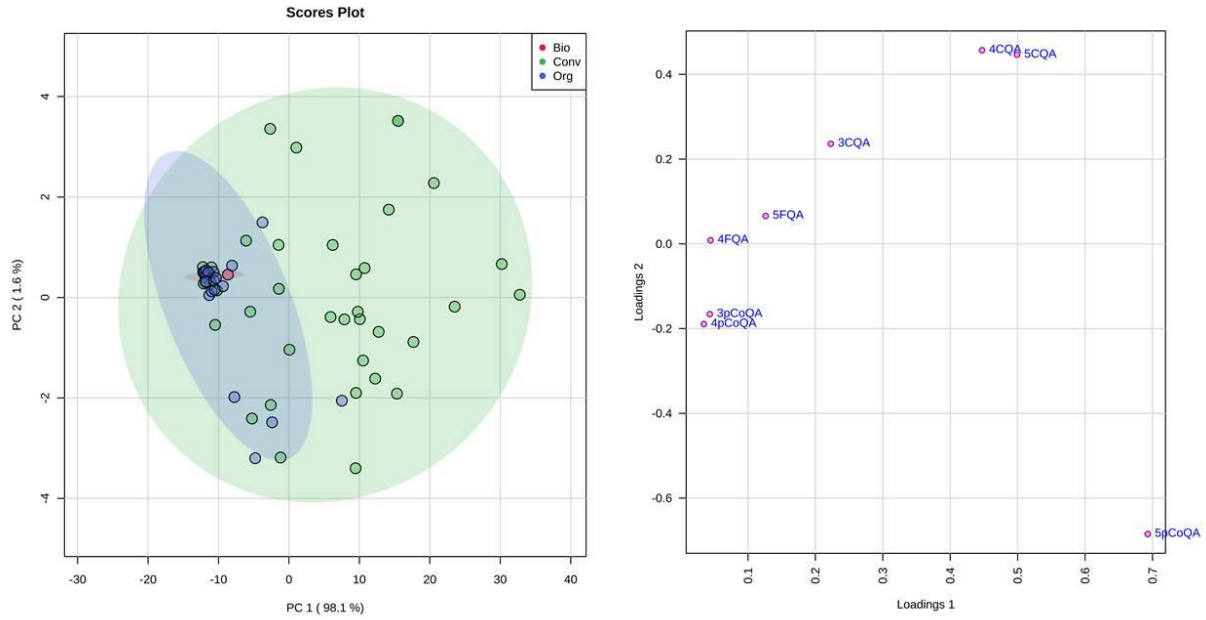


Figure 2 PCA score and loading plots of identified compounds using *m/z* range of 100 – 1200 with Pareto scaling and without any transformation. With colour-code of red dots for Bio, green dots for conventional and blue dots for organic cultivation systems.

In addition, box-plots were generated for each of the identified compounds as shown in Fig. 3a. From the box-plots, it was observed that the quantities of all the eight compounds were significantly higher ($p < 0.05$) in the conventionally cultivated samples when compared to organic and biodynamic cultivated samples. Furthermore, 5-CQA and 4-CQA were observed to be significantly higher than 3-CQA in all the three cultivation systems. This observation is quite surprising, since CGAs are assumed to act as pest defense compounds and the absence of pesticide would be expected to result in higher levels of CGAs. The p-values for the different compounds' comparisons were provided in Table 5a.

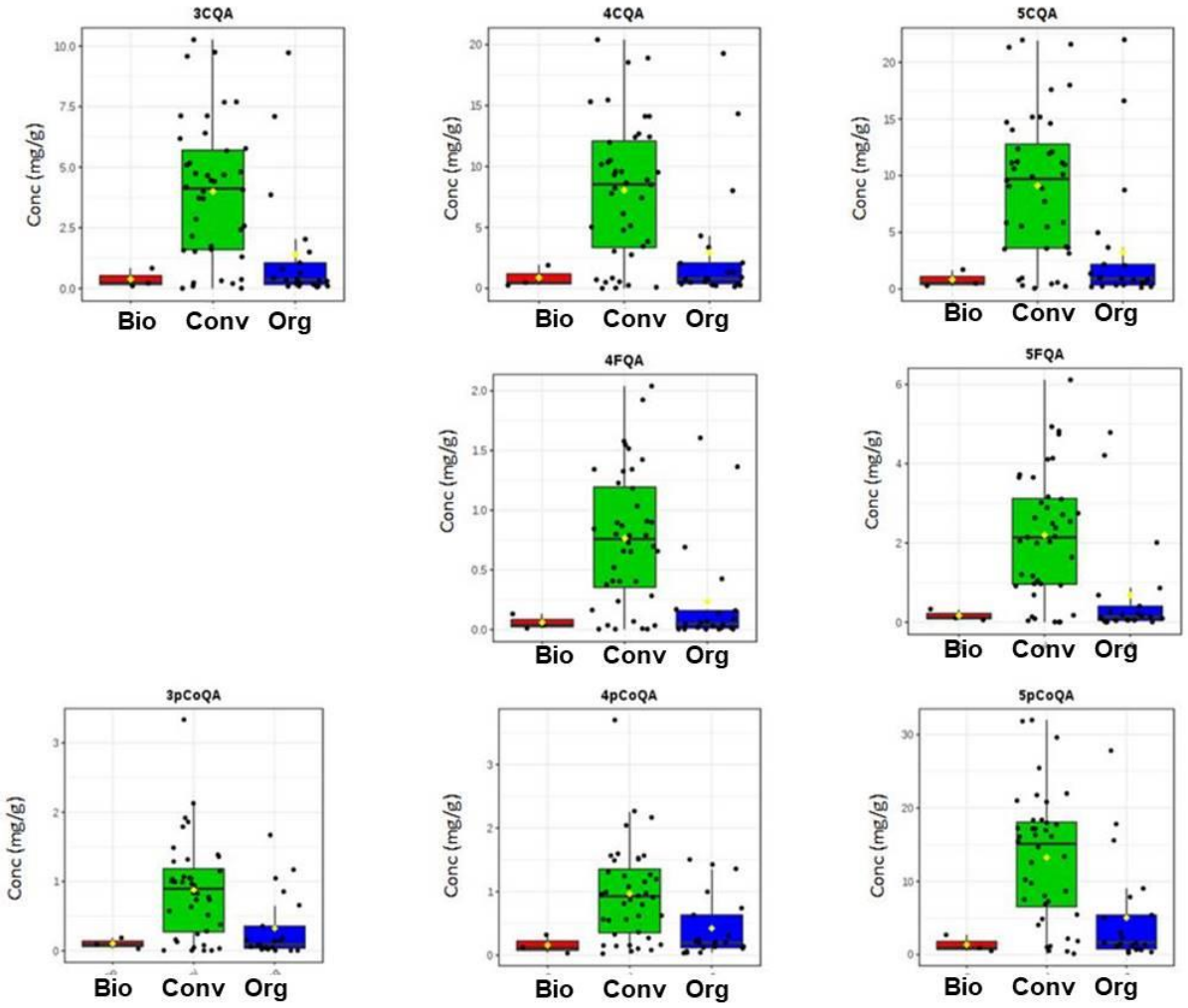


Figure 3a Box-plots of the identified variables (mono-CQAs, mono-FQAs and *p*-CoQAs) contents of the bio, conventional and organic roasted coffee beans and their quantities in mg/g with mean average (n=67).

Moreover, as for the FQAs, 3-FQA was not detected in most of the samples. While 4-FQA was observed to be significantly higher than the quantity of 5-FQA in all the conventional samples, followed by organic and biodynamic cultivated samples respectively. Also, from the box-plots, the 5-*p*CoQA was significantly higher than both 4- and 3- isomers of *p*CoQA.

Table 5b P-values of the coffee cultivated from different States in Brazil.

3-CQA	BA	ES	MG	PR	PR + ES + RO	SP	SP + MG
ES	1.00	n/a	n/a	n/a	n/a	n/a	n/a
MG	0.53	0.79	n/a	n/a	N/A	N/A	N/A
PR	0.69	0.60	0.24	N/A	n/a	n/a	n/a
PR + ES + RO	0.74	0.79	0.75	0.65	n/a	n/a	n/a
SP	0.18	0.62	0.67	0.09	0.25	n/a	n/a
SP + MG	0.20	0.24	**0.03**	0.33	0.10	**0.02**	n/a
SP + MG + PR	0.17	0.11	0.06	0.17	0.15	0.06	0.10
4-CQA	**BA**	**ES**	**MG**	**PR**	**PR + ES + RO**	**SP**	**SP + MG**
ES	1.00	n/a	n/a	n/a	n/a	n/a	n/a
MG	0.50	0.85	N/A	n/a	n/a	n/a	n/a
PR	0.62	0.57	0.18	n/a	n/a	n/a	n/a
PR + ES + RO	0.75	0.82	0.75	0.44	n/a	n/a	n/a
SP	0.18	0.59	0.66	0.06	0.27	n/a	n/a
SP + MG	0.20	0.31	**0.05**	0.44	0.10	**0.01**	n/a
SP + MG + PR	0.17	0.11	0.06	0.27	0.15	0.06	0.10
5-CQA	**BA**	**ES**	**MG**	**PR**	**PR + ES + RO**	**SP**	**SP + MG**
ES	1.00	n/a	n/a	n/a	n/a	n/a	n/a
MG	0.53	0.91	n/a	n/a	n/a	n/a	n/a
PR	0.70	0.70	0.22	n/a	n/a	n/a	n/a
PR + ES + RO	0.74	0.82	0.64	0.82	#N/A	n/a	n/a
SP	0.18	0.64	0.70	0.09	0.24	n/a	n/a
SP + MG	0.20	0.22	**0.05**	0.22	0.09	**0.02**	n/a
SP + MG + PR	0.15	0.10	0.06	0.09	0.13	0.06	0.09
4-FQA	**BA**	**ES**	**MG**	**PR**	**PR + ES + RO**	**SP**	**SP + MG**
ES	1.00	n/a	n/a	n/a	n/a	n/a	n/a
MG	0.56	0.60	n/a	n/a	n/a	n/a	n/a
PR	1.00	0.72	0.25	n/a	n/a	n/a	n/a
PR + ES + RO	0.16	0.39	0.45	0.16	n/a	n/a	n/a
SP	0.16	0.45	0.56	0.14	0.68	n/a	n/a
SP + MG	0.22	0.35	0.06	0.47	0.14	0.06	n/a
SP + MG + PR	0.16	0.39	0.14	0.60	0.16	0.16	1.00
5-FQA	**BA**	**ES**	**MG**	**PR**	**PR + ES + RO**	**SP**	**SP + MG**
ES	1.00	n/a	n/a	n/a	n/a	n/a	n/a
MG	0.67	0.82	n/a	n/a	n/a	n/a	n/a
PR	0.82	0.70	0.27	n/a	n/a	n/a	n/a
PR + ES + RO	0.17	0.82	0.95	0.44	n/a	n/a	n/a
SP	0.27	0.67	0.67	0.15	0.44	n/a	n/a
SP + MG	0.17	0.27	0.13	0.44	0.15	0.15	n/a
SP + MG + PR	0.17	0.15	0.15	0.44	0.17	0.20	0.67
3-pCoQA	**BA**	**ES**	**MG**	**PR**	**PR + ES + RO**	**SP**	**SP + MG**
ES	0.82	n/a	n/a	n/a	n/a	n/a	n/a
MG	0.70	0.62	n/a	n/a	n/a	n/a	n/a
PR	1.00	0.72	0.12	n/a	n/a	n/a	n/a

PR + ES + RO	0.12	0.34	0.95	0.05	n/a	n/a	n/a
SP	0.10	0.21	0.64	**0.01**	0.70	n/a	n/a
SP + MG	0.12	0.28	0.05	0.21	0.10	**0.01**	n/a
SP + MG + PR	0.12	0.10	0.09	0.41	0.12	0.05	0.65
4-pCoQA	**BA**	**ES**	**MG**	**PR**	**PR + ES + RO**	**SP**	**SP + MG**
ES	1.00	n/a	n/a	n/a	n/a	n/a	n/a
MG	0.78	0.95	n/a	n/a	n/a	n/a	n/a
PR	0.78	1.00	0.80	n/a	n/a	n/a	n/a
PR + ES + RO	0.11	0.33	**0.05**	**0.05**	n/a	n/a	n/a
SP	0.05	0.65	0.12	0.06	0.06	n/a	n/a
SP + MG	0.34	0.68	**0.05**	0.08	0.07	**0.001**	n/a
SP + MG + PR	0.11	0.33	**0.05**	0.07	0.10	**0.05**	0.25
5-pCoQA	**BA**	**ES**	**MG**	**PR**	**PR + ES + RO**	**SP**	**SP + MG**
ES	1.00	n/a	n/a	n/a	n/a	n/a	n/a
MG	0.72	0.73	n/a	n/a	n/a	n/a	n/a
PR	0.77	0.84	0.16	n/a	n/a	n/a	n/a
PR + ES + RO	0.12	0.84	0.95	0.24	n/a	n/a	n/a
SP	0.15	0.62	0.62	**0.05**	0.36	n/a	n/a
SP + MG	0.12	0.28	**0.03**	0.11	0.09	**0.01**	n/a
SP + MG + PR	0.12	0.10	0.05	0.09	0.12	0.05	0.10

However, the individual chlorogenic acids (CQAs, FQAs and *p*CoQAs) were analyzed with PCA to see if there could be any difference with respect to the cultivation systems. The PCA plots in Fig. 4 show that the biodynamic and organic farming systems clustered together, while conventional were observed to spread mostly outside the organic or biodynamic clusters. Since organic and biodynamic agriculture are very similar to an extent, this result is expected, thus the difference in the separation observed in the three plots. Hence, it could be inferred that 5-CQA, 4-CQA and 5-*p*CoQA could be used as biomarkers to categorize the samples in accordance with their respective cultivation patterns.

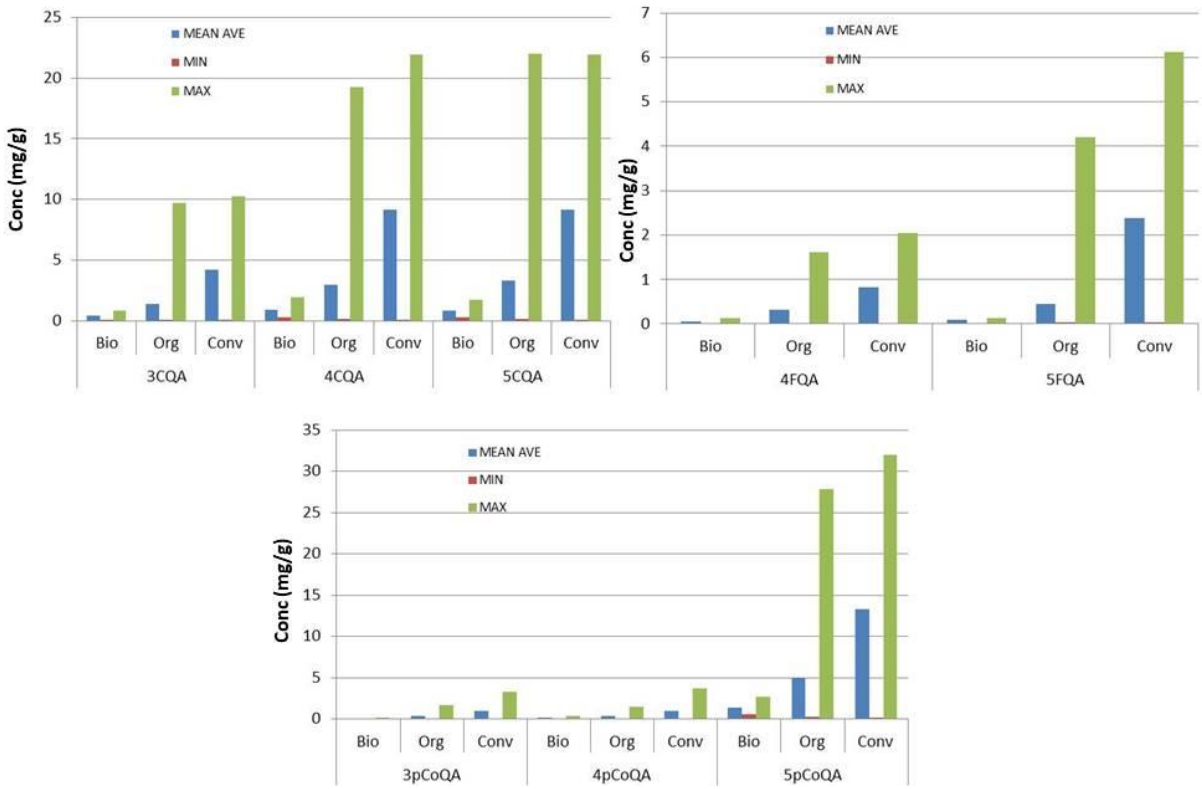


Figure 3b The plots represent the mean average, minimum and maximum values of the mono-CQAs, mono-FQAs and *p*-CoQAs constituents of the roasted coffee beans with mean (n = 67) expressed in mg/g.

3.3. PLS-DA of the roasted coffee samples

Partial least square discriminant analysis (PLS-DA) as a supervised multivariate statistical method, was also applied to the data set as shown in Fig. 5a with the important marker compounds shown in Fig. 5b plot respectively. This method basically finds the fundamental relations existing between the variables using multidimensional variances.

The PLS-DA plot in Fig. 5a represents the CGA variations on the sample's differences with respect to their different means of cultivation. It was observed that there was a clear and good separation between organic/biodynamic and conventional samples with about 98% variance. Likewise, the FQAs and *p*CoQAs were also well separated like the CQAs. This shows that there are significant differences between the samples based on their different system of cultivation

considering the use of chlorogenic acids as a biomarker to differentiate between the different patterns of cultivation. The top three important features (compounds) that characterized the cultivation patterns to show the differences majorly are 5-*p*CoQA, 5-CQA and 4-CQA. These three compounds have above 1.0 VIP scores compared to other marker compounds and the three features were found to be higher in conventional samples when compared with organic and/or biodynamic samples. However, 3-CQA and 5-FQA respectively have about 0.6 and 0.4 VIP scores compared to 3-pCoQA, 4-*p*CoQA and 4-FQA with about 0.2 VIP scores. It could thus be inferred that 4-FQA, 5-FQA and 5-*p*CoQA are good biomarkers to discriminate between organic and conventional samples.

Figure 4 PCA of mono-CQAs, mono-FQAs and *p*CoQAs constituents of the roasted coffee beans using *m/z* range 100 - 1200 based on the three different cultivation systems (organic, conventional and biodynamic systems).

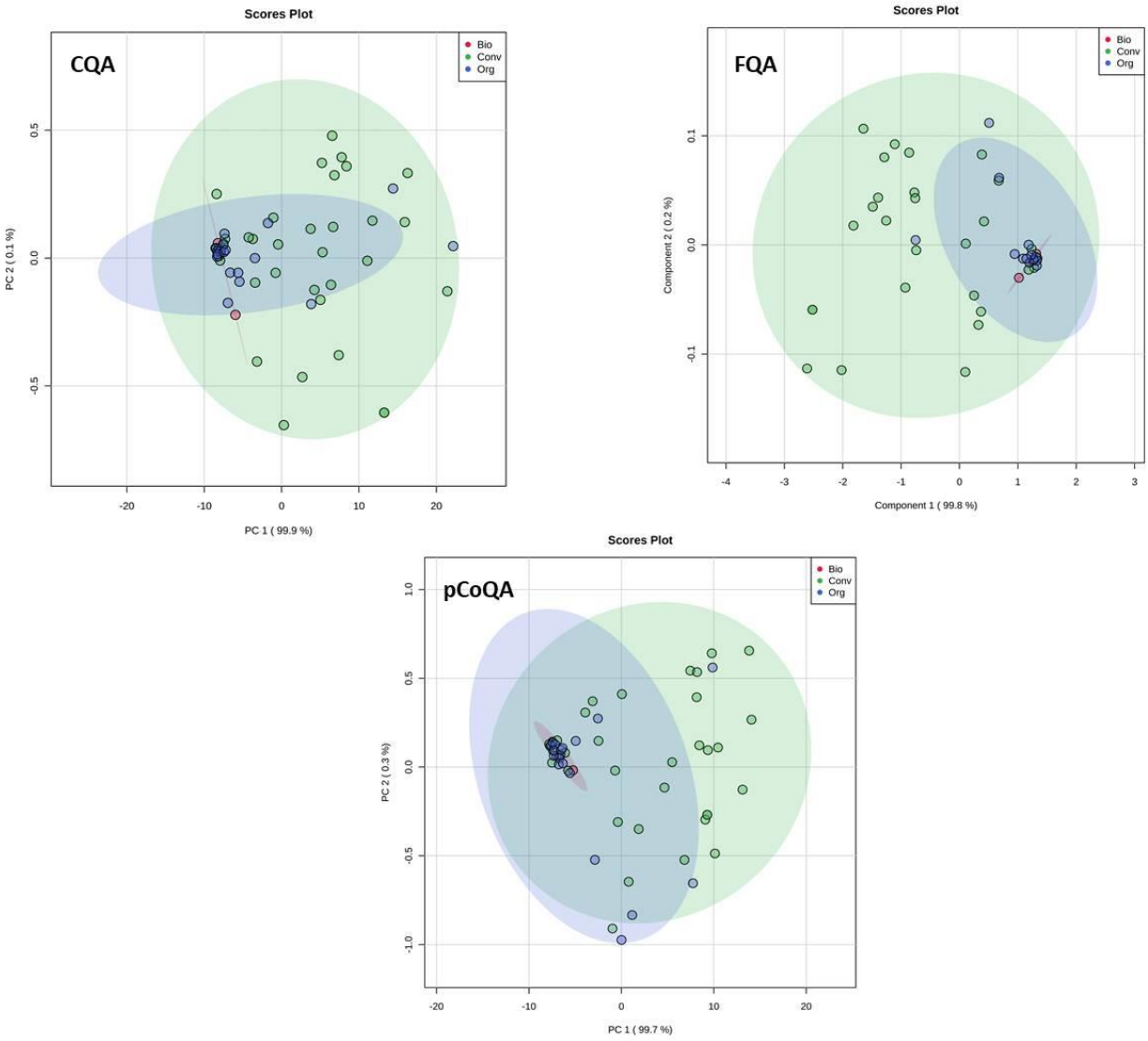
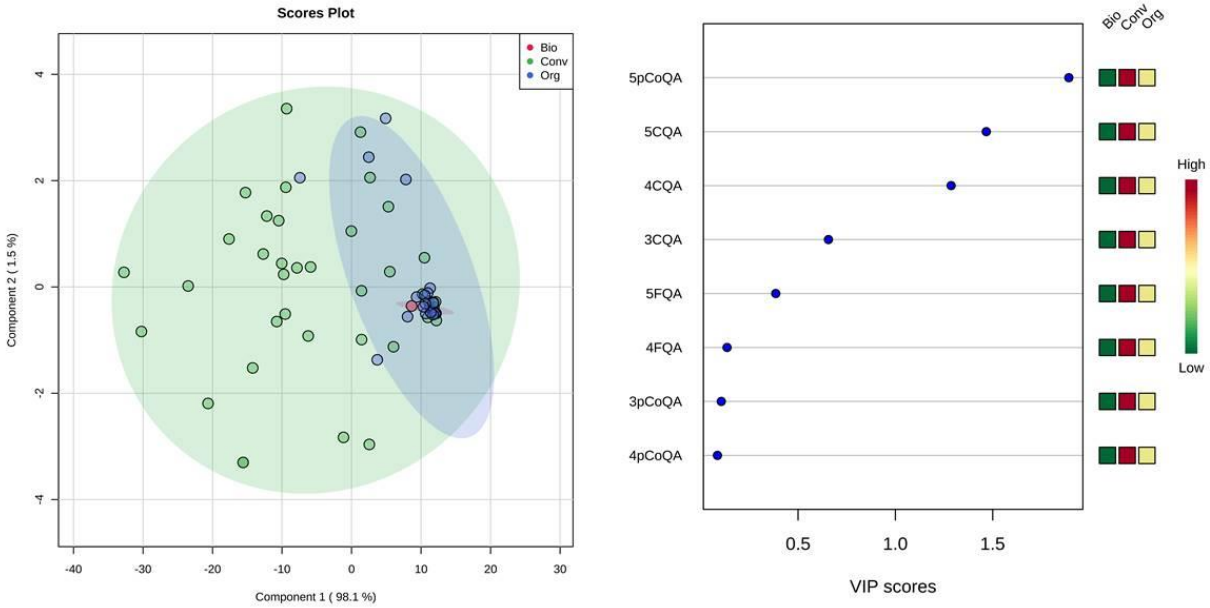


Figure 5 PLS-DA plot and important features of the coffee samples based on the different agricultural practices (organic, conventional and biodynamic) using *m/z* range 100 - 1200.



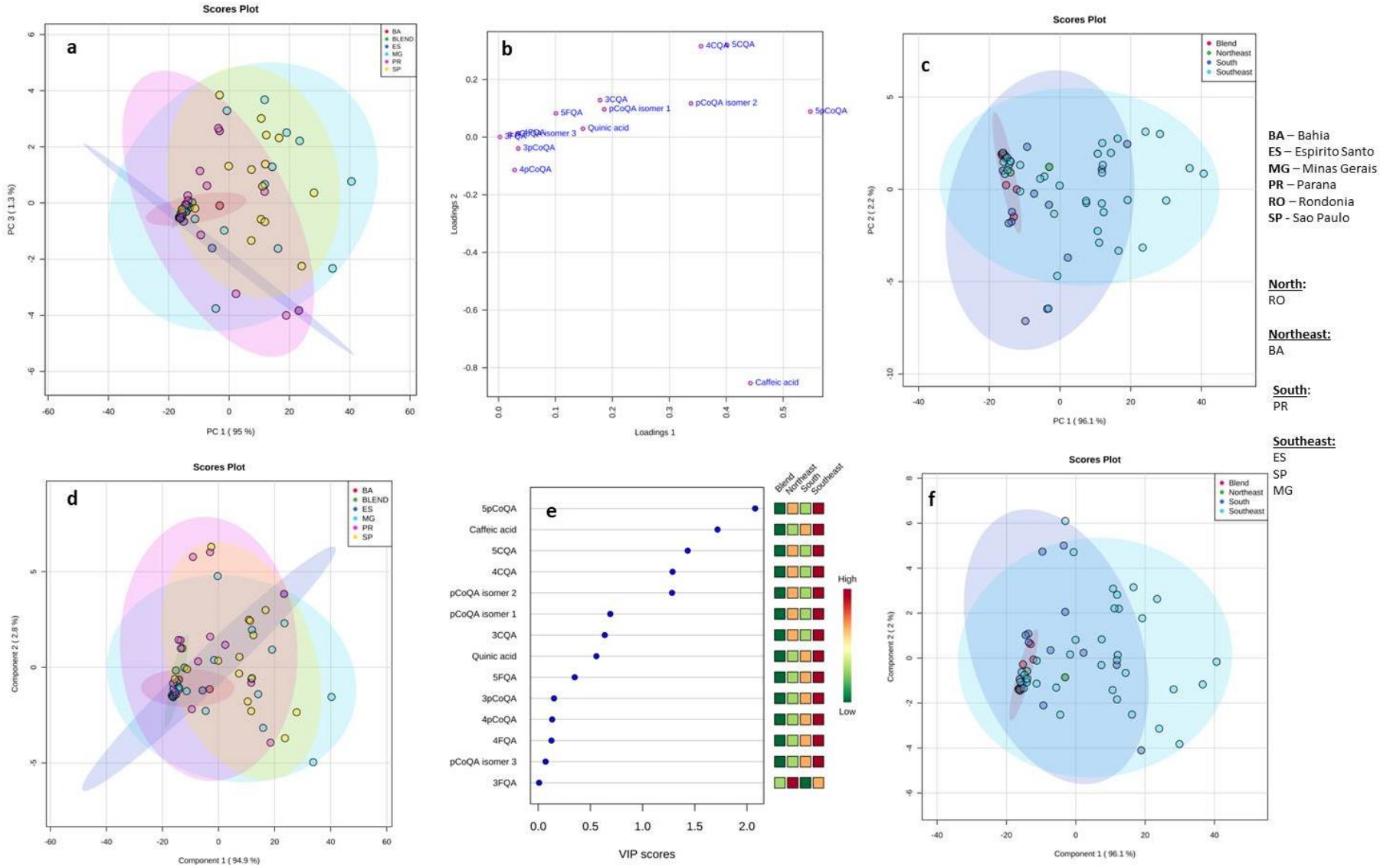
3.4. Discrimination between the coffees based on States and regions

The coffee samples were also investigated considering the different States and regional sources. The BA and RO samples are from the Northern States while the other four (ES, SP, MG and PR) are the Southern States. The data set were treated using PCA and PLS-DA. The plots in Fig. 6(a–f) represent the samples from all the regions together including the loading plots and important features (top ten compounds). The PCA in plot 6a shows the samples from the different States while the plot 6c indicates the samples analyzed based on regions. However, the observed plots show that there is no difference between the samples irrespective of the States or regions. Plot 6b represents the loading plot with the important marker compounds. As well, statistical analysis shown in Table 5b further supports these observations with the obtained p-values ($p < 0.05$).

In addition, PLS-DA was also used in analyzing the data set for any degree of variance as shown in plots 6d and 6f with the important compounds represented in plot 6e. The observations from

the PLS-DA plots for both the States and regions were not different from the PCA plots. Likewise, the p-values ($p < 0.05$) from the Kruskal-Wallis test indicates no significant statistical difference in the chemical composition between the coffees planted in the Northern provinces compared to Southern regions of the country if Northeast and/or Northern samples were compared with Southeast and/or Southern samples.

Figure 6 PCA score, loading plot, PLS-DA plot and important features of the coffee samples based on States and regions.



4. CONCLUSION

Considering the use of LC-MS^n quantification of key chlorogenic acids in organic, biodynamic and conventional coffees, the contents of 4-FQA, 5-FQA and 5-pCoQA showed statistically significant variations ($p < 0.05$) if organic roasted coffees were compared with conventional roasted coffees. Hence, we propose that the variations in these three key chlorogenic acids could be used in authenticating the method of cultivation (organic and conventional) of coffees in Brazil and also in many other countries where there are challenges in differentiating between organic and conventional cultivated coffee beans. Interestingly, organic coffee contains less chlorogenic acids if compared to conventional coffee. Hence, the consumer must realize that organic coffee appears to be associated with a reduction of compounds responsible for beneficial health effects. Moreover, no significant variation ($p < 0.05$) could be observed for coffees grown in different regions of Brazil. It might be concluded that based on the chlorogenic acids profiles the price variations in Brazilian coffees from North and South is not justified.

ACKNOWLEDGEMENT

The authors appreciate Anja Müller and Mauricio Moreno Zambrano for the technical and statistical analysis support respectively.

APPENDIX A. Supplementary material

Supplementary data is available online at *https://doi.org/10.1016/j.foodres.2020.109218*

REFERENCES

Abrankó, L., & Clifford, M. N. (2017). An unambiguous nomenclature for the acyl-quinic acids commonly known as chlorogenic acids. *Journal of Agricultural and Food Chemistry*, 65(18), 3602–3608.

Babova, O., Occhipinti, A., & Maffei, M. E. (2016). Chemical partitioning and antioxidant capacity of green coffee (*Coffea arabica* and *Coffea canephora*) of different geographical origin. *Phytochemistry*, 123, 33–39.

Badmos, S., Lee, S., & Kuhnert, N. (2019). Comparison and quantification of chlorogenic acids for differentiation of green robusta and arabica coffee beans. *Food Research International*, 126, 108544.

Barros, S., & Chanda B. (2019). United states department of agriculture coffee GAIN report. *Brazil Coffee Annual Report 2019* (No. BR19006). http://www.usdabrazil.org.br/pt-br/reports/coffee-annual2019.pdf. Accessed 02 November 2019.

Chong, J., Wishart, D.S. and Xia, J. (2019). Using MetaboAnalyst 4.0 for comprehensive and integrative metabolomics data analysis. *Current protocols in bioinformatics* 68, e86. https://www.metaboanalyst.ca/MetaboAnalyst/home.xhtml. Accessed on 10 November 2019.

Clifford, M. N. (1999). Chlorogenic acids and other cinnamates -nature, occurrence and dietary burden. *Journal of the Science of Food and Agriculture*, 79(3), 362–372.

Clifford, M. N., Jaganath, I. B., Ludwig, I. A., & Crozier, A. (2017). Chlorogenic acids and the acyl-quinic acids: Discovery, biosynthesis, bioavailability and bioactivity. *Natural Product Reports*, 34(12), 1391–1421.

Clifford, M. N., Johnston, K. L., Knight, S., & Kuhnert, N. (2003). Hierarchical scheme for LC-MS^n identification of chlorogenic acids. *Journal of Agricultural and Food Chemistry*, 51(10), 2900–2911.

Clifford, M. N., Knight, S., Surucu, B., & Kuhnert, N. (2006). Characterization by LC-MS^n of four new classes of chlorogenic acids in green coffee beans: Dimethoxycinnamoylquinic acids, diferuloylquinic acids, caffeoyl-dimethoxycinnamoylquinic acids, and feruloyl-dimethoxy cinnamoylquinic acids. *Journal of Agricultural and Food Chemistry*, 54(6), 1957-1969.

Cloete, K. J., Šmit, Ž, Minnis-Ndimba, R., Vavpetič, P., du Plessis, A., le Roux, S. G., & Pelicon, P. (2019). Physico-elemental analysis of roasted organic coffee beans from Ethiopia, Colombia, Honduras and Mexico using X-ray micro-computed tomography and external beam particle induced X-ray emission. *Food Chemistry*: X, 2.

Corrêa, P. C., de Oliveira, G. H. H., de Oliveira, A. P. L. R., Vargas-Elías, G. A., Santos, F. L., & Baptestini, F. M. (2016). Preservation of roasted and ground coffee during storage part 1: Moisture content and repose angle. [Conservação de café torrado e moído durante o armazenamento. Parte 1: Teor de água e ângulo de repouso]. Revista Brasileira De Engenharia *Agricola E Ambiental*, 20(6), 581–587.

de Oliveira, A. P. L. R., Corrêa, P. C., Reis, E. L., & de Oliveira, G. H. H. (2015). Comparative study of the physical and chemical characteristics of coffee and sensorial analysis by principal components. *Food Analytical Methods*, 8(5), 1303–1314.

Deshpande, S., El-Abassy, R. M., Jaiswal, R., Eravuchira, P., Von Der Kammer, B., Materny, A., & Kuhnert, N. (2014). Which spectroscopic technique allows the best differentiation of coffee varieties: Comparing principal component analysis using data derived from CD-, NMR- and IR-spectroscopies and LC-MS in the analysis of the chlorogenic acid fraction in green coffee beans. *Analytical Methods*, 6(10), 3268–3276.

Gamaleldin Elsadig Karar, M., Matei, M., Jaiswal, R., Illenberger, S., & Kuhnert, N. (2016). Neuraminidase inhibition of dietary chlorogenic acids and derivatives-potential antivirals from dietary sources. *Food and Function*, 7(4), 2052–2059.

Giraudo, A., Grassi, S., Savorani, F., Gavoci, G., Casiraghi, E., & Geobaldo, F. (2019). Determination of the geographical origin of green coffee beans using NIR spectroscopy and multivariate data analysis. *Food Control*, 99, 137–145.

Granato, D., de Magalhães Carrapeiro, M., Fogliano, V., & van Ruth, S. M. (2016). Effects of geographical origin, varietal and farming system on the chemical composition and functional properties of purple grape juices: A review. *Trends in Food Science and Technology*, 52, 31-48.

Gutiérrez-Ortiz, A. L., Berti, F., Solano Sánchez, W., Navarini, L., Colomban, S., Crisafulli, P., & Forzato, C. (2019). Distribution of p-coumaroylquinic acids in commercial coffea spp. of different geographical origin and in other wild coffee species. *Food Chemistry*, 286, 459–466.

Jaiswal, R., & Kuhnert, N. (2011). Identification and characterization of two new derivatives of chlorogenic acids in arnica (*Arnica montana L.*) flowers by high-performance liquid chromatography/tandem mass spectrometry. *Journal of Agricultural and Food Chemistry*, 59(8), 4033–4039.

Jaiswal, R., Matei, M. F., Golon, A., Witt, M., & Kuhnert, N. (2012). Understanding the fate of chlorogenic acids in coffee roasting using mass spectrometry based targeted and non-targeted analytical strategies. *Food and Function*, 3(9), 976–984.

Jaiswal, R., Patras, M. A., Eravuchira, P. J., & Kuhnert, N. (2010). Profile and characterization of the chlorogenic acids in green robusta coffee beans by LC-MS^n: Identification of seven new classes of compounds. *Journal of Agricultural and Food Chemistry*, 58(15), 8722–8737.

Jaiswal, R., Sovdat, T., Vivan, F., & Kuhnert, N. (2010). Profiling and characterization by LC-MS^n of the chlorogenic acids and hydroxycinnamoylshikimate esters in maté (*Ilex paraguariensis*). *Journal of Agricultural and Food Chemistry*, 58(9), 5471–5484.

Kuhnert, N., Jaiswal, R., Matei, M. F., Sovdat, T., & Deshpande, S. (2010). How to distinguish between feruloyl quinic acids and isoferuloyl quinic acids by liquid chromatography/tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, 24(11), 1575–1582.

Kurzrock, T., & Speer, K. (2001). Diterpenes and diterpene esters in coffee. *Food Reviews International*, 17(4), 433–450.

Matei, M. F., Jaiswal, R., & Kuhnert, N. (2012). Investigating the chemical changes of chlorogenic acids during coffee brewing: Conjugate addition of water to the olefinic moiety of chlorogenic acids and their quinides. *Journal of Agricultural and Food Chemistry*, 60(49), 12105–12115.

Monteiro, P. I., Santos, J. S., Alvarenga Brizola, V. R., Pasini Deolindo, C. T., Koot, A., Boerrigter-Eenling, R., & Granato, D. (2018). Comparison between proton transfer reaction mass spectrometry and near infrared spectroscopy for the authentication of Brazilian coffee: A preliminary chemometric study. *Food Control*, 91, 276–283.

Naveed, M., Hejazi, V., Abbas, M., Kamboh, A. A., Khan, G. J., Shumzaid, M., & XiaoHui, Z. (2018). Chlorogenic acid (CGA): A pharmacological review and call for further research. *Biomedicine and Pharmacotherapy*, 97, 67–74.

Peng, C. -., Zhang, Y. -., Song, W., Cai, H. -., Wang, Y., & Granato, D. (2019). Characterization of Brazilian coffee based on isotope ratio mass spectrometry (δ13C, δ18O, δ2H, and δ15N) and supervised chemometrics. *Food Chemistry*, 297, 124963.

R Core Team, (2019). A language and environment for statistical computing. *R foundation for statistical computing*, Vienna, Austria. https://www.R-project.org/. Accessed 06 December 2019.

Tajik, N., Tajik, M., Mack, I., & Enck, P. (2017). The potential effects of chlorogenic acid, the main phenolic components in coffee, on health: A comprehensive review of the literature. *European Journal of Nutrition*, 56(7), 2215–2244.

Worku, M., Upadhayay, H. R., Latruwe, K., Taylor, A., Blake, W., Vanhaecke, F., & Boeckx, P. (2019). Differentiating the geographical origin of Ethiopian coffee using XRF- and ICP-based multi-element and stable isotope profiling. *Food Chemistry*, 290, 295–307.

SUPPLEMENTARY DATA

Table 1 Details of the roasted coffee samples (conventional, organic and biodynamic) cultivated from different regions.

Samples	Class	Region	Location
1	Bio	PR	PR
2	Org	SP + MG	BLEND
3	Org	PR	PR
4	Bio	SP	SP
5	Org	PR	PR
6	Org	SP + MG + PR	BLEND
7	Org	SP + MG	BLEND
8	Org	MG	MG
9	Org	SP + MG	BLEND
10	Org	SP + MG	BLEND
11	Org	SP	SP
12	Org	SP + MG	BLEND
13	Org	SP	SP
14	Org	BA	BA
15	Org	SP + MG	BLEND
16	Bio	ES	ES
17	Org	MG	MG
18	Org	MG	MG
19	Conv	MG	MG
20	Conv	MG	MG
21	Conv	PR	PR
22	Conv	MG	MG
23	Conv	PR	PR
24	Conv	PR	PR
25	Conv	BA	BA
26	Conv	PR	PR
27	Org	PR	PR
28	Conv	SP	SP
29	Conv	SP	SP
30	Conv	SP	SP
31	Conv	SP	SP
32	Conv	SP	SP
33	Conv	SP	SP
34	Conv	SP	SP
35	Conv	MG	MG
36	Conv	MG	MG
37	Conv	MG	MG
38	Conv	MG	MG
39	Conv	MG	MG
40	Conv	MG	MG
41	Conv	MG	MG
42	Conv	ES	ES
43	Conv	ES	ES

44	Conv	SP	SP
45	Conv	PR + ES + RO	BLEND
46	Conv	SP	SP
47	Conv	PR	PR
48	Org	SP	SP
49	Org	SP	SP
50	Conv	SP	SP
51	Org	SP	SP
52	Conv	SP	SP
53	Org	MG	MG
54	Org	MG	MG
55	Conv	PR	PR
56	Conv	SP	SP
57	Conv	PR	PR
58	Conv	SP	SP
59	Conv	MG	MG
60	Conv	PR	PR
61	Conv	ES	ES
62	Conv	PR	PR
63	Conv	SP	SP
64	Conv	PR	PR
65	Conv	MG	MG
66	Org	ES	ES
67	Conv	ES	ES

*São Paulo (SP), Minas Gerais (MG), Espírito Santo (ES), Bahia (BA), Paraná (PR) and Rondonia (RO). Likewise, ORG for organic, CONV for conventional and BIO for biodynamic.

Chapter 4

Multivariate discriminant analysis of Arabica and Robusta green coffee beans based on triacylglycerol profiles using HPLC-ESI mass spectrometry

1. INTRODUCTION

Coffea arabica and *Coffea canephora* (Robusta) are the most consumed coffee varieties worldwide. Arabica accounts for approximately 75% of the world coffee production, while Robusta represents the remaining fraction (Mondego et al, 2011). *Coffea arabica* has lesser genetic diversity and results in a higher quality beverage than *C. canephora* (Anthony et al, 2002), and thus most of this beverage consumers prefer the former to the latter due to its superior taste, aroma and flavor resulting in higher prices for Arabica coffee. Coffee lipid constituents essentially determine the drink's cup-quality in terms of taste, aroma and flavor. Also, it plays a crucial role in coffee crema formation (Folmer et al, 2017), and serve as one of the basic foods' constituents, essential in human diet for numerous roles (Lelyana, 2017).

The lipid fraction of coffee is composed mainly of triacylglycerols, sterols and tocopherols. The Coffee triacylglycerol major constituents are fatty acids in proportions similar to those found in common edible vegetable oils. The relatively large unsaponifiable fraction is rich in diterpenes of the kaurane family, mainly cafestol, kahweol and 16-*O*-methylcafestol (Speer & Kölling-Speer, 2006). Most of the lipids in the coffee oil are located in the endosperm of green coffee beans with a small amount, the coffee wax is located on the bean outer layer (Wilson et al., 1997); and about 75% are triacylglycerol (Iriondo-Dehond et al, 2019; Jham et al, 2003; Toci et al, 2013).

In animals, triacylglycerol serves primarily as an energy store, while they function in plants mainly as a carbon reserve in seeds, which are used by humans as vegetable oils. It essentially maintains body temperature, protect vital organs, regulate hormones, transmitting nerve impulses, memory storage and tissue structure. Among its components are omega-3 and omega-6 essential fatty acids (polyunsaturated fatty acids) which regulate cholesterol, blood clotting and inflammation in the joints, tissues and bloodstream (Heldt & Piechulla, 2011).

Characteristically, linoleic (47.7%), palmitic (33.3%), stearic (7.3%) and oleic (6.6%) acids are the most abundant fatty acids present in the green coffee bean. *Coffea arabica* contains about 15 - 18% lipids, while *Coffea canephora* usually contains less than 10% (Speer & Kölling-Speer, 2006). Moreover, this essential factor significantly contributes to consumers' choice of preferring Arabica to Robusta, invariably makes Arabica superior in terms of both demand and cost price.

Moreover, the oil extracted from unroasted coffee beans is known as green coffee oil. It has been increasingly used in the food and pharmaceutical industries, as well as in cosmetics, due to its composition (Oliveira, Franca, Mendonça, & Barros-Júnior, 2006). The major compounds of the green coffee oil are triacylglycerols and free fatty acids, majorly linoleic and palmitic acids (Wagemaker, Carvalho, Maia, Baggio & Guerreiro Filho, 2011). The latter are present in the stratum corneum and contribute to the skin barrier and improve hydration (Patzelt et al, 2012). Essentially, the hydrolysis of the triacylglycerol produces free fatty acids, which when oxidized could affect the coffee cup-quality.

Recent study also established the potential of the green coffee oil to absorb UVB (Type-B Ultraviolet) radiation as well as protective properties, which vary within the different coffee plant species; since UVB has harmful effects on human skin with it rays causing skin burns, most skin cancers, prematurely skin aging (Wagemaker, Silva, Leonardi & Maia Campos, 2015) and directly damage DNA. Thus, the green coffee oil helps in preventing and/or reducing these various hazardous effects.

The green coffee beans of the Arabica and Robusta varieties can be distinguished by their size, shape and colour, but the roasting process eliminates these macroscopic aspects (Wei & Tanokura, 2015). Furthermore, since the main fraud involving coffee is the undeclared addition of Robusta to Arabica variety, there are important economic reasons to demand warranties on the authenticity of coffee species, even if the identification and the quantification of Arabica in

roasted and minced coffee blends are very challenging (Pustjens, Weesepoel & van Ruth, 2016). Besides, due to consumers' preference to Arabica compared to Robusta it is quite pertinent to differentiate Arabica from Robusta coffees. In view of this, the German standard method DIN 10779 is used in the determination of 16-*O*MC in roasted coffee beans to detect C. canephora in blends (Schievano, Finotello, De Angelis, Mammi & Navarini, 2014). The 16-*O*-MC is a coffee lipid constituent mainly found in Robusta but recent studies have otherwise shown that small amounts of esterified 16-*O*-methylcafestol were also found in Arabica coffees (Gunning et al, 2018). Hence, this shows that the use of this marker compound is not 100% accurate and/or reliable to differentiate between the two coffee varieties for consumer protection and likewise the global coffee industry.

Furthermore, identification of key chlorogenic acids showing significant variations allows establishment of a reliable multivariate statistical method for coffee authentication only in green coffee beans (Badmos, Lee & Kuhnert, 2019; Jaiswal & Kuhnert, 2011; Abranko & Clifford, 2017; Clifford, Johnston, Knight & Kuhnert, 2003; Jaiswal, Patras, Eravuchira & Kuhnert, 2010).

In addition, before now several techniques such as infra-red, nuclear magnetic resonance and mass spectrometry have been employed in differentiating between the two coffee varieties (Barbin, Felicio, Sun, Nixdorf & Hirooka, 2014; Briandet, Kemsley & Wilson, 1996; Dias & Benassi, 2015; El-Abassy, Donfack & Materny, 2011), but despite these efforts identifying a flawless biomarker is still a challenge.

Moreover, within the last decade the literature has devoted much interest in studying more about TAGs with well over 75,000 publications with chemistry related articles, but it has been observed that those dealing with analytical methods, specifically in coffee are far less (Indelicato et al, 2017). Besides, recent studies now give more attention with specific interest in TAGs analysis towards establishing a more reliable biomarker to discriminate the two

coffees; and another reason could be associated with increasing care given to diet and its effect on human health.

Despite the numerous publications on coffee lipids, information concerning TAGs is still very limited. The method of TAGs analysis in coffee involves several challenges, such as HPLC column type, cumbersome extraction method, choice of efficient extraction solvent with less toxicity or minimal lethal dose, duration of LC gradient sequence for complete TAGs components good separation; and most importantly unavailability of established commercial TAGs standards for precise identification.

In view of this, some of these challenges have therefore been addressed in this study using a short column that allows better separation of the individual TAGs to militate against the problem of co-elution. Furthermore, the duration of LC gradient run was reduced to less than a quarter of the time (25 min) compared to what was used in previous studies. In addition, with the use of molecular ion fragmentation and a newly developed sequence gradient method for proper identification of TAGs, with this new and more TAGs were identified. Hence, the study aims to establish a more reliable food authentication method and biomarkers as required in the coffee producing industry using the coffee triacylglycerol constituents.

Before now, several studies have established that the coffee varieties TAGs composition is similar but the relative percentage of the TAGs in distinct varieties can vary; and among the previously identified TAGs, which used using mostly HPLC, Ag-TLC and RP-TLC methods were LLL, PLnLn, OLL, PLL, PPLn, OOL, SLL, POL, PPL, ALL, OOO, PLS, POO, POLn, StLL, StLLn, StOL and SSLn (Cossignani, Montesano, Simonetti & Blasi, 2016; González, Pablos, Martín, León-Camacho & Valdenebro, 2001; Jham et al, 2003; Nikolova-Damyanova, Velikova & Jham, 1998; Segall, Artz, Raslan, Jham & Takahashi, 2005).

Nevertheless, this study observed that some other important TAGs present in the green coffee bean are yet to be identified and reported. Besides, this study went further to investigate the TAGs differences in the two coffee varieties compared to previous reports.

2. MATERIALS AND METHODS

2.1 Chemicals and reagents

All reagents and solvents were of analytical or chromatographic grade. Acetonitrile (LC-MS ultra-gradient grade), chloroform (HPLC grade), isopropanol (HPLC grade), heptane (HPLC grade) were purchased from Carl Roth (Karlsruhe, Germany); Ethanol (LC-MS grade) was purchased from Merck (Darmstadt, Germany) and distilled to achieve ~99.7% purity before use. Formic acid and ammonium formate (LC-MS grade) were purchased from Sigma-Aldrich (Steinheim, Germany) and and the internal standard - Benzyl-dimethyl-hexadecyl ammonium chloride produced from Sigma Life Science, Germany.

2.2 Coffee beans

Green coffee beans (*Coffea arabica*) 3 pairs from each of these different countries, Arabica (Kenya, Brazil, Ethiopia, Peru, Guatemala, Nicaragua, China, Laos, Costa Rica, Honduras, Rwanda, Colombia, Papua New Guinea, Rwanda and Jamaica) and Robusta (Ecuador, Uganda, Vietnam, India, Tanzania and Indonesia) acquired from local grocery shops in Bremen and around Germany were used in this study. The authors of this study chose and decided to use green coffee beans because roasted coffee beans cannot be relied upon due to loss of some lipid fractions during roasting.

2.3 Coffee oil extraction

The green beans were ground to extract the lipid fraction and analysed to determine the TAG components. The coffee green cherry beans were frozen using liquid nitrogen and ground properly using RAPIDO coffee grinder (TZS-FIRST, Austria) to super fine particle-powdered form. The ground bean (10 g) was extracted with 150 mL of heptane using Soxhlet (BUCHI B-811 extraction system, Flawil, Switzerland) with a continuous mode setting for 12 h. The

solvent was removed using rotatory evaporator (Laborata 4000-Efficient, Heidolph, Germany) under reduced pressure at 45 °C, 120 rpm and the extracted oil was stored at -32 °C until when needed for analysis. The extracted oil was dissolved in Ethanol/DCM of ratio (8:2) with a concentration of 0.1 mg/mL for HPLC-MS analysis. An amount of 10 µL of internal standard using 0.1 mg/mL concentration was added to each of the sample. This was observed to have a retention time of 1.7 min and *m/z* (360.36).

2.4 Preparation of salt clusters (calibration solution)

The calibration solution used in calibrating the instrument was prepared with a mixture of 12.5 mL of ddH_2O, 12.5 mL of isopropanol, 50 µL of formic acid and 250 µL of 1M NaOH. All chemical used were of HPLC analytical grade.

2.5 Chromatographic separation of triacylglycerol

The TAGs molecular species were separated using Agilent HPLC equipment (1200 series) from Agilent Technologies, Germany. A Poroshell 120 EC-C18 column with 3 x 150 mm and particle size 4 µm was used for the chromatographic separation sequence at temperature of 25 °C. The mobile phase has acetonitrile as solvent A with 0.01% formic acid; ethanol was used as solvent B with the addition of 0.01% formic acid and 10 mM/L ammonium formate. The formic acid was added to the solvents to optimize ionization, while ammonium formate was added for TAGs adduct formation for proper identification during data analysis. The flow rate used for the mobile phase was 0.5 mL/min and sample injection volume of 2 µL. The sequence gradient run for the column was set as follows: column solvent pre-run A/B (v/v, 40:60) for 3 min, this was followed by a linear gradient to solvent B at 100 for 24 min and ended with isocratic gradient elution for solvent B (60) for 10 min. Thereafter, the column was re-equilibrated for 3 min with the initial run of solvent A/B at ratio 40:60 before running the next sample sequence. A blank containing only the mobile phase was run in between each sample chromatographic

sequence. The whole column effluent (0.5 mL/min) was directly connected to the mass spectrometer.

2.6 LC-ESI-TOF-MS (High resolution mass spectrometry)

High-resolution LC-MS was carried out using the same HPLC equipped with a Micro-TOF Focus mass spectrometer (Bruker, Bremen, Germany) fitted with an ESI source as the detector. The instrument was programmed to operate using a capillary voltage of 4.5 kV, nebulizing gas pressure of 2 bars with a drying gas temperature of 220 °C, drying gas flow rate of 10 L/min and ESI mass spectra measured in positive ion mode using a range of *m/z* 200 – 1200. The internal calibration was achieved with 10 mL of 0.1 M sodium formate solution injected through a six-port valve prior to each chromatographic run. Calibration was carried out using the enhanced quadratic mode. Analysis of TAGs was carried out by LC-ESI-TOF-MS. Data acquisition and processing was executed using Bruker Data analysis software 4.2 system (Bremen, Germany). Molecular formulae suggestions were accepted with a mass error < 5 ppm. The percentages (relative abundance) of each TAG separated by HPLC were determined based on the detector's peak area response normalized with the internal standard (ISD). The ISD was prepared by dissolving 10 mg of the compound in 10 mL of ethanol and diluted to 0.1 mg/mL before use.

2.7 LC-ESI-MS^n (Quadrupole ion-trap mass spectrometry)

The Liquid chromatographic Agilent 1260 series equipment (Karlsruhe, Germany) comprises of a binary pump, an auto sampler having a 100 µL capacity loop and a Diode Array Detector (DAD) with a range from 200 to 600 nm and the detector recorded between 254 and 320 nm. Full scan mass spectra were recorded in the range *m/z* 200 - 1200, operating in positive ion mode. The chromatographic separation was carried out using the same gradient method used in the 2.4 above. The chromatographic compounds separation was carried out with the aid of a

Poroshell 120 EC-C18 column with 3 x 150 mm and particle size of 4 µm at 35 °C using a solvent flow rate 500 µL/min. The LC equipment was connected to an Ion-trap mass spectrometer fitted with an electrospray ionization source (Bruker Daltonics, Bremen, Germany) operating in full scan auto-MS^n mode to obtain fragment ions. As required, MS^2 and MS^3 fragment-targeted experiments were performed. Tandem mass spectra were acquired in auto-MS^n mode (smart fragmentation) using a ramping of the collision energy. Helium was used as a collision gas for collision-induced dissociation (CID). Maximum fragmentation amplitude was set to 1 V starting from 30% and ending at 200%. MS operating conditions (positive mode) with a capillary voltage of 4.5 kV and temperature of 365 °C, a dry gas flow rate of 10 L/min, and a nebulizer pressure of 10 psi. This method was adapted from our Lab well-established set-up for lipids analysis as previously reported (Sirbu, Corno, Ullrich & Kuhnert, 2018). The Brukers' Agilent Chemstation software was used for programming this set-up. Likewise, data acquisition and processing were executed using Bruker Data analysis software 4.2 system (Bremen, Germany).

2.8 Data analysis reproducibility and repeatability

To verify reproducibility and confirm repeatability of the TAGs analysed, the same coffee sample was analysed on three different occasions with triplicates starting from extraction with heptane on Soxhlet through solvent evaporation, analysis on HPLC coupled with mass spectrometry and Tandem mass spectrometry. The same data was used to confirm the instrument precision and column accuracy. However, the column used was rinsed properly for about an hour before each analytical run.

2.9 Statistical analysis

Analysis of variance (ANOVA) was used to study the differences in the TAGs contents of the two coffees (Arabica against Robusta). Data processing and analysis were carried out

using Python for P-values, ANOVA significance and Pearson correlation coefficient. All statistical tests were conducted at the 95% confidence level ($p < 0.05$). Principal component analysis (PCA) and partial least square-discriminant analysis (PLS-DA) of the green coffee bean were carried out based on their triacylglycerol profiles using *MetaboAnalyst 4.0* for Comprehensive and Integrative Metabolomics Data Analysis (Chong, Wishart and Xia, 2019). Lipid-MAPS Lipidomics Gateway tool was used to confirm the masses of the TAGs by comparing the experimental mass with theoretical mass (already identified TAGs) available in the online database (Fahy et al, 2009).

3. RESULTS AND DISCUSSION

The main objective of this study is to profile the complete triacylglycerol constituents of the two main coffee types and explore the outcome to differentiate the two coffee varieties Arabica and Robusta using chemometrics tools based on LC-MS profiles. The triacylglycerol constituents of Arabica and Robusta green coffee beans have been analysed in several studies (Cossignani et al, 2016; Jham et al, 2003; Nikolova-Damyanova et al, 1998; Segall et al, 2005), from the studies about eighteen (18) TAGs were identified. However, this study explored the use of a chromatographic Poroshell 120 EC-C18 column with a novel gradient elution developed here coupled with the use of Tandem mass spectrometry not employed previously. In view of this, new triacylglycerols were identified and combined with the previously known, multivariate discriminant analysis was applied to differentiate between the two main coffee varieties.

Green coffee beans from twenty (20) different origins representing different botanical origins including Arabica from Kenya, Brazil, Ethiopia, Peru, Guatemala, Nicaragua, China, Honduras, Rwanda, Colombia, Papua New Guinea, and Jamaica; and Robusta from Ecuador, Uganda, Vietnam, India, Tanzania and Indonesia were selected for this study (see Table 1 for the details). The total lipids content was extracted using an optimized Soxhlet extraction in continuous mode for 12 h with heptane. The gravimetric data of the total lipids extracted for each bean sample is provided in Table 1. In agreement with previous reports, Arabica beans (13.17 – 16.68%) show significantly higher total lipid values compared to Robusta beans (8.56 – 11.05%) when 10 g of powdered coffee was extracted with 150 mL of heptane.

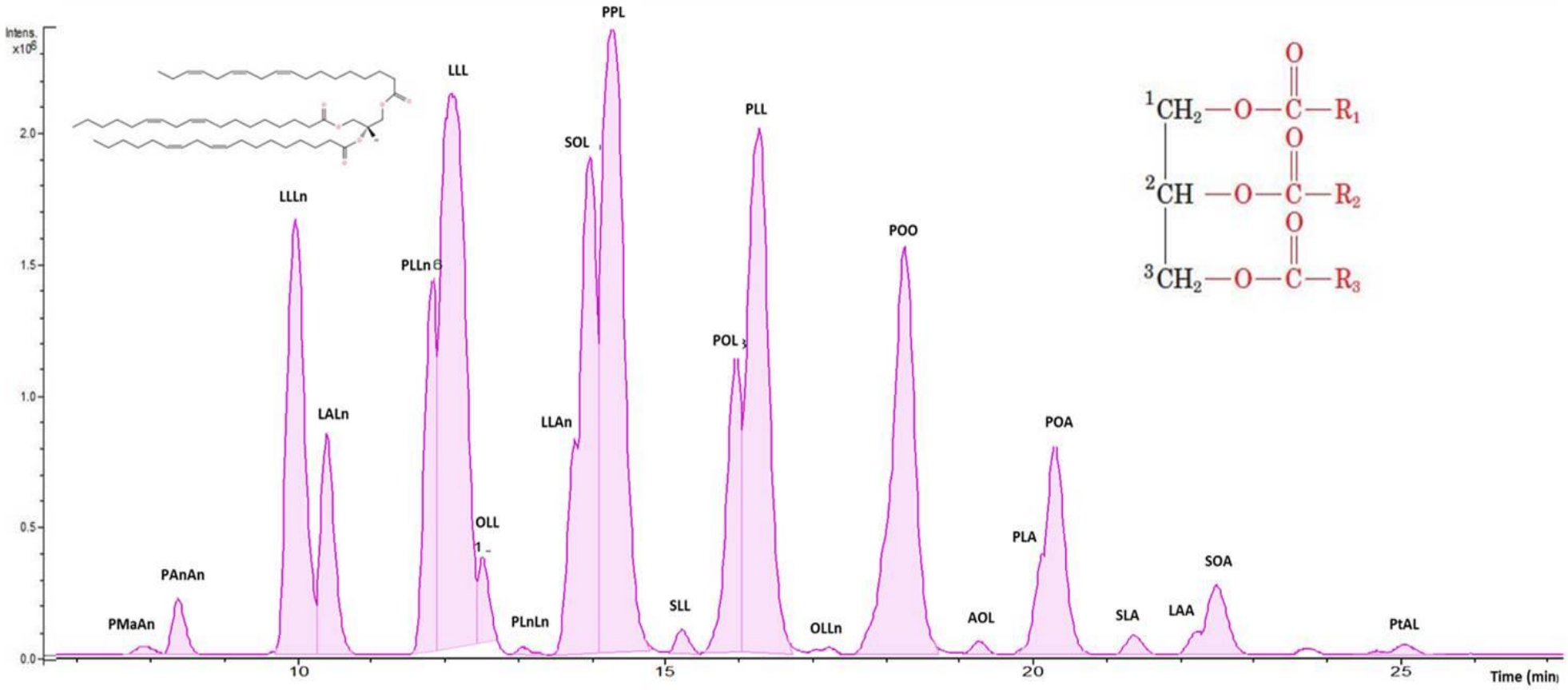
Moreover, for the triacylglycerol analysis an optimized reversed-phase HPLC method was developed using electrospray ionization mass spectrometry for detection. After optimisation, ethanol demonstrated to be quite efficient for the good separation of the TAGs from a reversed

Poroshell C18 stationary phase column with less co-elution. Additionally, acetonitrile proved to elute the TAGs much earlier than other solvents like methanol and likewise the addition of ammonium formate enabled triacylglycerol ionization in ESI (Electrospray Ionization). Figure 1 revealed the identified TAGs at different elution time from 6.2 min to 26 min. Furthermore, the Poroshell column eluted all the TAGs in less than half an hour compared to previously reported studies where it was observed than most of the TAGs identified were completely eluted in 60 min to 120 min or even longer. For the TAGs nomenclature, the typical abbreviation of three (3) fatty acid substituents was employed by using a single capital letter code to represent each fatty acid. For instance, SOL indicates 1- stearoyl-2-oleoyl-3-linoleoyl-glycerol as shown in Table 2 for the major fatty acids and/or any universal fatty acid which are mostly less abundant in nature (Lísa & Holčapek, 2008).

3.1 Percentage composition of Arabica and Robusta coffee oil extract

In this study, after grinding coffee beans, which was initially freeze-dried with Liquid Nitrogen, 10 g of each ground coffee bean varieties was extracted with 150 mL of Heptane for 12 h using Soxhlet extractor. The extracting solvent was evaporated on Rotavap and 10 mg of the extract was dissolved in 100 mL of the LC-MS mobile phase, which is ACN/EtOH (v/v; 20/80) to give a concentration of 0.1 mg/mL of the extract and for LC-MS analysis. Thereafter, the weighted averages of the extracted coffee oil for all the coffee samples were determined. Table 1 depicts details of the respective coffee oil percentage average weights. The mean for the Arabica was estimated to be 1.69 g (16.68%), while Robusta was 0.86 g (8.62%). This shows that there is significant difference between the two coffee oil qualities. This fact also buttresses the point that differences in the coffee lipids composition could contribute to the different aroma, taste, flavour and significant health for humans.

Figure 1 HPLC-MS chromatogram of crude heptane extract of green Arabica coffee in positive ion mode with the TAGs details provided in Table 3.



In agreement with the previous published data on coffee TAGs, this study re-analysed the Arabica coffee lipid extract using Micro-TOF LC-MS and fragmentation on Tandem-MS for unambiguous identification of the detailed TAGs constituents. A Poroshell 120 EC-C18 column was used for the LC run to achieve good TAGs separation with reduced co-elution of compounds, reduced sequence gradient period from the common 100 min run to 37 min. The extracted oil samples described in 3.1 above were analyzed and over 28 peaks were observed in the chromatogram with different retention time, out of which 26 were identified as TAGs using the Tandem-MS fragmentation analysis.

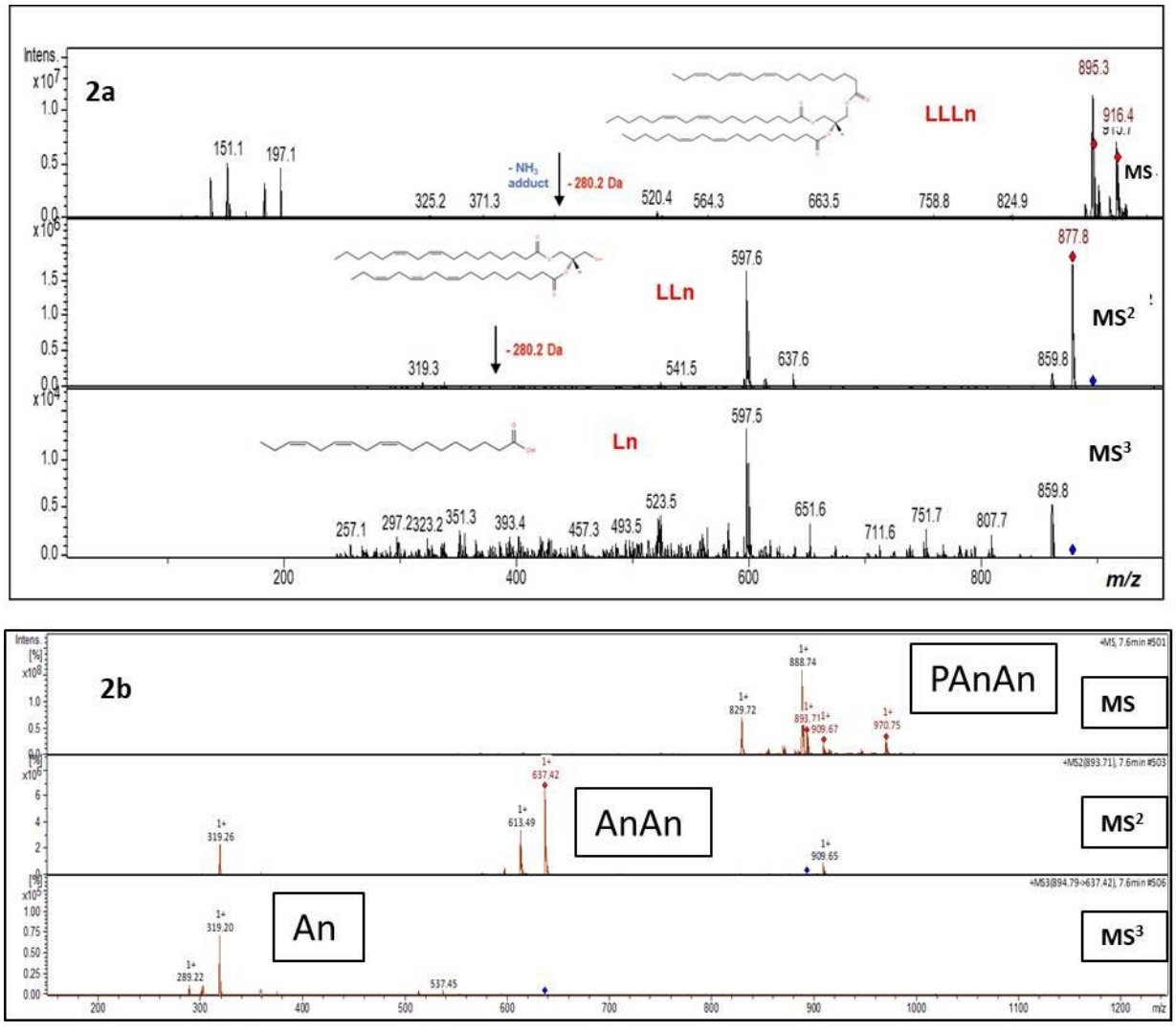


Figure 2 Using an ion-trap mass spectrometry in positive ion mode (a) Tandem mass spectra of ammonium adduct ion $[C_{18}H_{32}O_2{:}C_{18}H_{32}O_2{:}C_{18}H_{30}O_2 + 17]^+$ of TAG LLLn showing the fragmented ions corresponding to the losses of neutral fatty acids and ammonium adduct ions

(L, L, and Ln). (b) Tandem mass spectrum of TAG PAnAn showing the fragmented ions corresponding to the losses of Palmitic acid and diacylglycerol fragmented into two molecules of Arachidonic acids (P, An, and An).

In addition, Figure 1 shows the LC-MS chromatograms of coffee lipid TAG constituents and Figure 2 provide details of the fragmentation pattern of one of the TAGs using Tandem-MS and confirmed with Lipid MAPS online database of known TAGs. While Table 3 indicates TAGs of coffee bean details with their chemical formulae, carbon skeletons, *m/z* values and ammonium adduct, and confirmed by Tandem MS and confirmed Lipid MAPS online tool.

3.2 Identification of triacylglycerol

The tandem-MS fragmentation patterns of the TAGs showed that most of the peaks identified have ammonium adduct $[M + 18]^+$ in the parent ion which has previously been reported (Hvattum, 2001; Sandra, Medvedovici, Zhao & David, 2002; Segall et al, 2005). The details of TAGs of the cherry coffee bean with their chemical formulae, carbon skeletons, *m/z* values and ammonium adduct using tandem MS and confirmed with Lipid MAPS online tool database are provided in Table 2 of the supplementary information section, while a discussion on TAGs structure assignment is presented in the following segment.

3.2.1 Structure assignment of the identified TAGs

From the chromatogram in Figure 1, 28 peaks were eluted, out of which 26 were identified and confirmed to be TAGs with respect to their fragmentation patterns. Some of the peaks (TAGs) fragmentation patterns with details were explained below, while others are provided in the supplementary information section. From the tandem MS in Figure 2, the first peak that was identified has a precursor ion *m/z* 888.75 (RT of 6.4 min) had a loss of NH_4+ adduct give 871.71. The 871.71 yielded palmitic acid (256.2 and molecular formula: $C_{16}H_{31}O_2$) as neutral loss to

give DAG 615.49. The 615.49 was further fragmented into margaric acid (270.2 - $C_{17}H_{34}O_2$) and arachidonic acid (306.5 - $C_{20}H_{32}O_2$). Thus, the TAG was identified as **PMaAn**. The diacylglycerol (DAG) of margaric acid and arachidonic acid combination was also confirmed by mass spectrometry analysis of long chain fatty acids reported by (Murphy & Axelsen, 2011). Likewise, previous study has shown the presence of margaric acid in coffee by gas chromatography (Carisano and Gariboldi, 1964).

The next peak with a precursor ion *m/z* 894.7 (RT of 6.8 min) yields a neutral loss of palmitic acid (256.2) which gives DAG of 637.42 in MS^2. While the 637.42 was further fragmented into two (2) molecules of a hydroxylated form of Arachidonic acid (318.2), this hydroxyl fatty acid is otherwise known as *9-hydroxyl-2Z,5E,7Z,11Z,14Z-Eicosapentaenoic acid*. *9-hydroxyl-arachidonic acid* belongs to alternative class of fatty acids called 'Hydroxyl fatty acids' with molecular formula ($C_{20}H_{30}O_3$) similar to hydroxyl fatty acids previously reported (Sirbu et al, 2018). Hence, the TAG was named **PAnAn** with fragmentation pattern details shown in Fig 3.

Furthermore, one of the TAGs has a precursor ion *m/z* 895.3 (RT of 10.4 min) with the loss of NH_4^+ to give 877.8 and yielded Linoleic acid (280.2 - $C_{18}H_{32}O_2$) as neutral loss to give DAG of *m/z* 597.6 in MS^2. The diacylglycerol 597.6 was further fragmented into Linoleic (280.2) and Linolenic (278.3 - $C_{18}H_{30}O_2$) acids. Thus, the TAG was identified as **LLLn**. This TAG is known and has been reported in several publications (Jham et al, 2003; Segall et al, 2005).

A study by Cossignani reported that the unsaturated fatty acids (UFA) and essential fatty acids (EFA) have preference for *sn*-2- position, while saturated fatty acids (SFA) preferred mostly the *sn*-1- and the *sn*-3- positions (Cossignani et al, 2016). Though, it is known that the regio-specificity of fatty acids in TAG is characteristic for the native oils and fats. Besides, Folstar in 1985 observed that UFA, especially linoleic acid, was preferably esterified with the secondary hydroxyl position in glycerol of coffee TAG In view of this, it has previously been observed

that fragmentation usually occurs at the MS^2 step preferentially from the *sn-2*- position (Folstar, 1985).

Besides, one TAG with *m/z* 862.75 (RT of 14.6 min) yields linolenic acid (278.3 - $C_{18}H_{30}O_2$) and NH_4^+ adduct to give 589.5. The 589.5 was fragmented further into oleic acid (282.2 - $C_{18}H_{34}O_2$) and linoleic acid (280.2 - $C_{18}H_{32}O_2$). Hence, this TAG was identified as **OLLn**.

However, from the 26 TAGs identified eight (8) are being reported for the first time via this study. They are PMaAn, PAnAn, LLAn, OLLn, PtAL, LAA, SOA and SAAn. Moreover, it was observed that two (2) of these new TAGs consist of odd number carbon saturated fatty acids precisely margaric (C17:0) and pentacosanoic (C25:0) acids. However, these two saturated fatty acids have been reported in several other analysed plant oils, while pentacosanoic (C25:0) acid is yet to be reported in coffee (Holčapek, Lísa, Jandera & Kabátová, 2005; Lísa, Holčapek & Boháč, 2009; Rueda et al, 2014; Yu, Du, Yuan & Hu, 2016). Also, due to incomplete fragmentation two more TAGs were not fully elucidated.

Table 3 Triacylglycerol identified from green Arabica coffee bean lipid extract. TAG abbreviation details are provided in Table 2 of the appendix.

Number	RT (min)	M+NH4	Experimental mass	Theoretical mass	Error (ppm)	Carbon Skeleton	TAG Abbreviation	Molecular Formula
1	6.4	888.8015	870.7671	870.7676	2.3	TG (16:0/17:0/20:3)	PMaAn	$C_{56}H_{102}O_6$
2	6.8	909.6755	893.7154	893.8139	0.9	TG (16:0/20:4/20:4)	PAnAn	$C_{59}H_{101}O_6$
3	7.9	926.5904	908.6982	908.7833	1.1	TG (18:2/18:3/20:0)	LALn	$C_{59}H_{104}O_6$
4	8.4	949.8962	932.2085	n/a	-	TG (18:0/-/20:3)	S-A	Incomplete fragmentation
5	9.6	904.4972	886.8125	886.4872	0.3	TG (18:2/18:2/20:4)	LLAn	$C_{59}H_{108}O_6$
6	9.8	896.7702	878.7358	878.7363	0.1	TG (18:2/18:2/18:2)	LLL	$C_{57}H_{98}O_6$
7	10.4	895.3545	877.7201	877.7236	0.7	TG (18:2/18:2/18:3)	LLLn	$C_{57}H_{96}O_6$
8	11.9	870.7545	852.7201	852.7207	0.5	TG (16:0/18:2/18:3)	PLLn	$C_{55}H_{96}O_6$
9	12.3	898.7858	880.7514	880.752	1.4	TG (18:1/18:2/18:2)	OLL	$C_{57}H_{100}O_6$
10	13.6	900.8015	882.7671	882.7676	0.6	TG (18:0/18:2/18:2)	SLL	$C_{57}H_{102}O_6$
11	13.9	876.8015	858.7671	858.767	0.1	TG (16:0/18:1/18:1)	POO	$C_{55}H_{102}O_6$
12	14.2	848.7702	830.7358	830.7368	1.8	TG (16:0/16:0/18:2)	PPL	$C_{53}H_{98}O_6$
13	14.6	872.7702	854.7358	854.7363	0.7	TG (16:0/18:2/18:2)	PLL	$C_{55}H_{98}O_6$
14	15.3	868.7389	850.7045	850.7058	0.9	TG (16:0/18:3/18:3)	PLnLn	$C_{55}H_{94}O_6$
15	15.8	872.7702	854.7358	854.7364	0.3	TG (16:0/18:1/18:2)	POL	$C_{55}H_{98}O_6$
16	17.1	896.7702	878.7357	878.7363	2.1	TG (18:1/18:2/18:3)	OLLn	$C_{57}H_{98}O_6$
17	18.3	902.8171	884.7827	884.7841	0.7	TG (18:0/18:1/18:2)	SOL	$C_{57}H_{104}O_6$
18	19.4	908.6588	890.8764	n/a	-	TG (18:2/20:0/25:0)	PtAL	Unusual TAG
19	19.9	930.8484	912.814	912.8146	1.3	TG (18:1/18:2/20:0)	AOL	$C_{59}H_{108}O_6$
20	20.4	904.8328	886.7984	886.7989	0.6	TG (16:0/18:2/20:0)	PLA	$C_{57}H_{106}O_6$
21	21.4	932.8641	914.8297	914.8302	0.3	TG (18:0/18:2/20:0)	SLA	$C_{59}H_{110}O_6$
22	22.5	906.8484	888.814	888.8146	0.9	TG (16:0/18:1/20:0)	POA	$C_{57}H_{108}O_6$
23	23.8	953.7709	936.3887	936.8146	3.4	TG (18:0/20:0/20:4)	SAAn	$C_{61}H_{108}O_6$
24	24.7	925.494	908.3644	n/a	-	TG (16:1/-/20:0)	P-A	Incomplete fragmentation
25	25.2	960.8954	942.8619	942.8615	1.4	TG (18:2/20:0/20:0)	LAA	$C_{61}H_{114}O_6$
26	25.2	934.8797	916.8453	916.8459	1.1	TG (18:0/18:1/20:0)	SOA	$C_{59}H_{112}O_6$

* n/a means unavailable in the online lipids database.

3.3 Discriminating between Arabica and Robusta coffee triacylglycerol

Since it has been established that the coffee oil is very rich in triacylglycerol (about 75% of total coffee lipids) and due to the TAGs contributions to the coffee aroma, taste and flavour; it is pertinent to look into the differences between the triacylglycerol compositions of the two coffee varieties towards determining the possible influence of TAGs on the physicochemical properties of the different coffee variety. Although, some studies have investigated the differences between Arabica and Robusta considering their TAGs variations (Cossignani et al, 2016; González et al, 2001; Speer & Kölling-Speer, 2006) but despite this, none of these previous studies reported or identified any of these newly identified TAGs and also explored the use of multivariate statistical analysis to differentiate between the two coffees. Moreover, to discriminate between the TAGs from the two coffee (Arabica against Robusta), the acquired peak intensities from LC/MS/MS were normalised with the internal standard and subjected to statistical tools using principal component analysis (PCA) and partial least squares-discriminant analysis (PLS-DA). In view of this, the unsupervised PCA and supervised PLS-DA classification methods were applied together, for the first time to discriminate *C. arabica* and *C. canephora* based on their TAGs compositions.

3.4 Principal component analysis of Arabica against Robusta coffee

The extracted coffee oil (TAGs) analysed on LC/MS/MS was subjected to PCA which is an unsupervised multivariate statistical tool. This tool is employed to reduce the total number of measured variables and further explain largest possible proportion of the variables' variances. Figure 4 shows the PCA and loading plots generated from the data set of the analysed Arabica and Robusta coffee samples. However, it was observed that there is less separation between the two coffee varieties with this statistical tool. Invariably, it could be said that using PCA for this

discrimination might not be that effective analytically. This is because the variables were distributed unevenly with less separation between Arabica and Robusta designated samples.

3.5 Partial least squares-discriminant analysis of Arabica against Robusta coffee

The PLS-DA mostly employed in discriminative variable selection of large data set. This supervised tool models high-dimensional datasets for classification and diverse purposes to achieve reliable and valid outcomes. This versatile algorithm was deployed for the acquired LC/MS/MS datasets and two plots were generated. In Figure 5, the left plot represents the PLS-DA scores while the right illustrates the loading plot of important features that differentiate the two coffees. From the loading plot, the top ten (10) marker compounds among the identified TAGs are SOL, LAA, PLLn, SLL, PtAL, POL, POO, PLnLn, PPL and SLA, and more or less contribute significantly to the differentiation between Arabica and Robusta coffee varieties. However, from the PLS-DA plot it could be observed that the Arabica coffee was well separated from the Robusta coffee with about 29.3% variance. This is completely different from the plot generated using PCA and despite the fact that all the coffee samples used are from diverse origins with different climatic and cultivation systems. This indicates that the use of triacylglycerol analysis on LC/MS/MS in conjunction with PLS-DA statistical tool could be used in discriminating between the two topmost and most consumed coffee varieties. This further shows that this analytical procedure could be employed in food industrial to reduced foods and beverages adulteration.

However, from the loading plot in Figure 5 it could be seen that some TAGs have higher scores in Robusta than in Arabica, and vice versa. For instance, TAGs like LAA, SLL, PtAL, POO, PPL and SLA are higher in Robusta compared to Arabica; while the plot also shows that SOL, PPLn, POL and PLnLn are higher in Arabica compared to Robusta. This further illustrates that

triacylglycerols profiling could be used in discriminating between Arabica and Robusta coffee to minimize the common coffee adulteration in the global markets and industries.

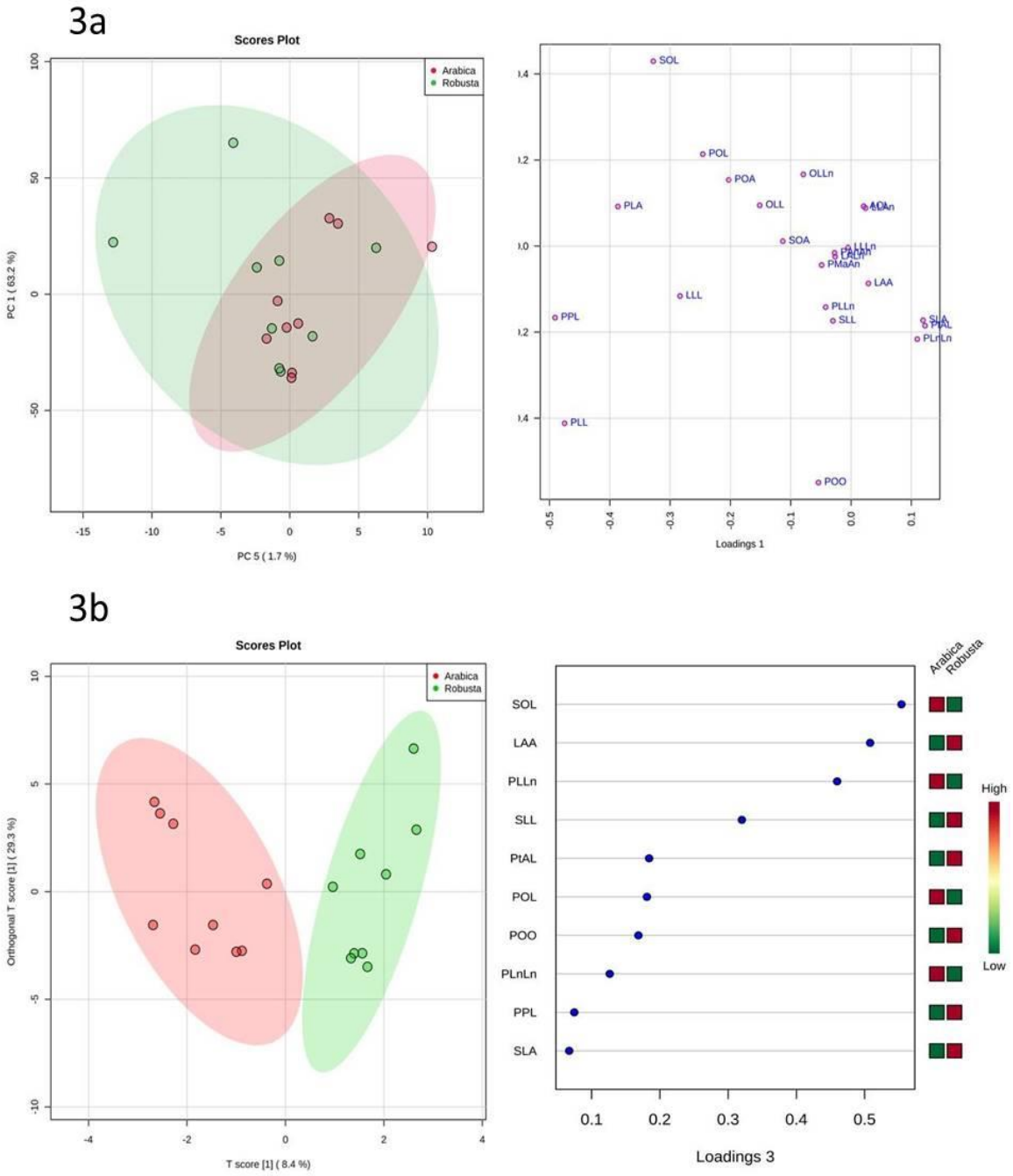


Figure 3 (a) PCA and loading plots of the Arabica (red dots) and Robusta (green dots) coffee samples. Points represent the coffee samples in two different clusters in the PCA plot. (b) Scores plot of PLS-DA model showing two clusters and the important features on the right. Each dot represents a coffee sample. The important features plot presents the influence of the variables for the separation of the groups and shows important biomarkers for the distinction.

3.6 Relative quantification of Arabica and Robusta coffee TAGs

Appendix B represents boxplots of the TAGs for Arabica and Robusta coffees. The relative abundance (%) of each of the TAGs were determined from extracted LC\MS\MS data and plotted as boxplots. Generally, the plots indicate that the quantities of TAGs vary irrespective of the coffee variety (Arabica or Robusta). The TAG with the lowest abundance is SLL (Stearic-Linoleic-Linoleic acid), followed by LLLn, PtAL and LLAn respectively (SLL < LLLn < PtAL < LLAn). However, among these four TAGs with the exception of LLAn, Robusta is observed to be more than Arabica in all the four, while LLAn in Arabica was shown to be higher than Robusta. Moreover, among the total TAGs identified it was observed that Arabica was higher than Robusta in seven (7) of the TAGs, while Robusta was higher than Arabica in eleven (11) of the total TAGs. Furthermore, amongst all PLL was observed to be significantly higher than others followed by PPL, POL and LLL (PLL > PPL > POL > LLL). What was understood to be common among these four (4) most abundant TAGs is the presence of an essential fatty acid known as Linoleic acid (C18:2) and also palmitic acid (C16:0). Interestingly, these two fatty acids have notably been reported to be the most abundant fatty acids in coffee and followed by Oleic acid (C18:1) (Cossignani et al, 2016; Dong, Tan, Zhao, Hu & Lu, 2015; Speer & Kölling-Speer, 2006). Furthermore, PLA, AOL and PLnLn indicated no significant difference between the Arabica and Robusta quantities.

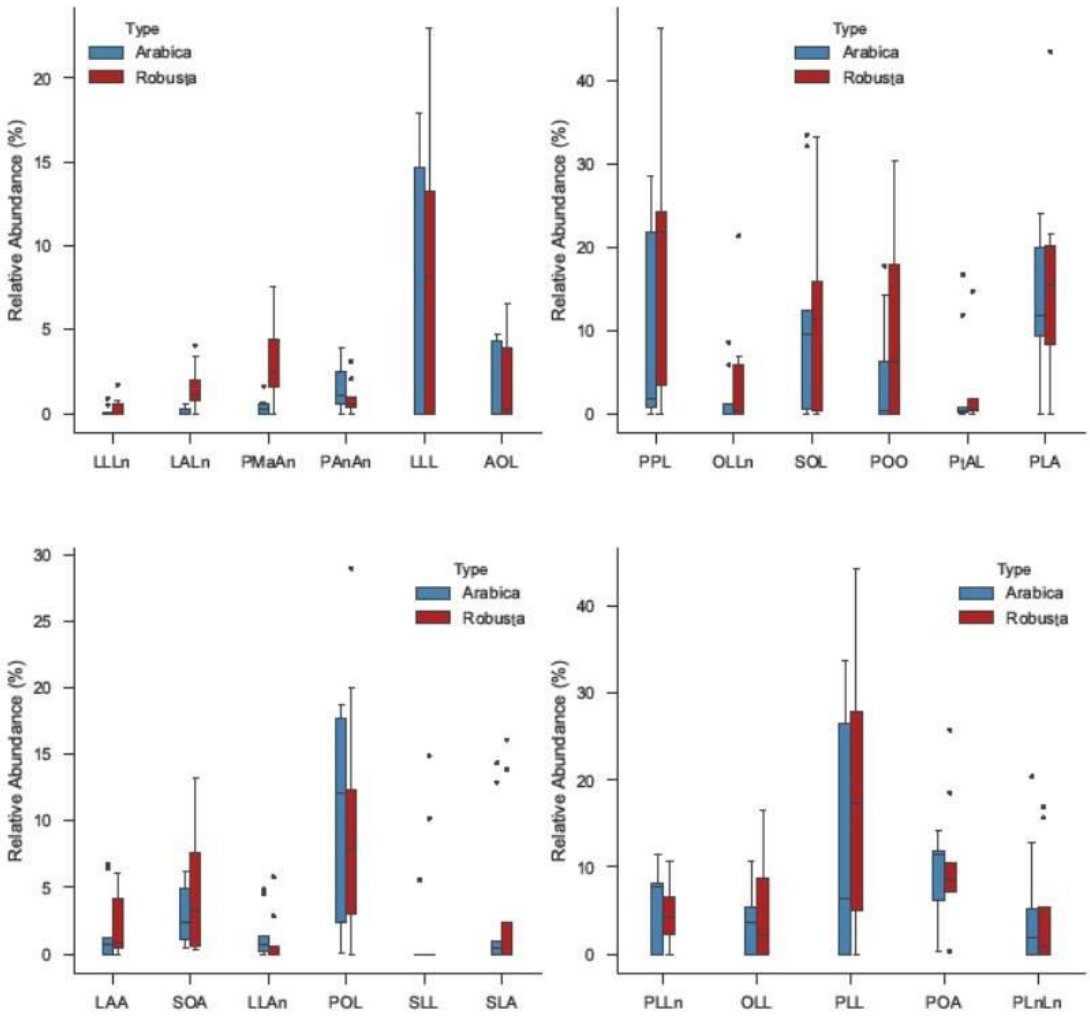


Figure 4 Boxplots of relative quantities of Arabica and Robusta green coffee TAGs. Boxes represent interquartile range (IQR) and median values are depicted as horizontal lines in the boxes.

3.7 Statistical significance

The p-values in shown Table 4 (see appendix section) indicate that there is significant difference ($p < 0.05$) between the Arabica and Robusta coffee TAGs when compared. Furthermore, the scatter plot in Figure 5 (see appendix section) shows that there is difference between the percentage abundance of some of the TAGs when Arabica samples were compared to the Robusta samples.

4. CONCLUSION

In this study, LC-MS coupled with ion-trap Tandem-MS has shown to be a proficient analytical technique for oil analysis especially for coffee lipids, where precise data are necessary in order to evaluate the compositional and structural properties. The HPLC sequence gradient method used with the short Poroshell chromatographic column reveals more TAGs with improved chromatographic resolution with respect to previous studies when applied to the analysis of coffee TAGs. Furthermore, this method has allowed not only the complete characterization of the TAG composition of coffee oil with new TAGs identification, but likewise enabled the separation and identification of mono- and diacylglycerols fragments in a single run together with TAGs. This method is faster with less sequence duration to allow an accurate TAGs separation and reduced the common co-elution effects. It equally provides excellent chromatographic resolution and broad retention window together with tandem mass spectra which enabled positive identification of TAGs containing fatty acids with odd numbers of carbon atoms such as margaric (C17:0) and pentacosanoic (C25:0) acids.

In this contribution, using Tandem MS for the first time in coffee lipid analysis existing assignment could be confirmed and TAGs previously not reported were identified. Furthermore, the results achieved from the discrimination of Arabica from Robusta coffees evidently indicate that analysis of TAGs using LC-MS coupled with Tandem-MS and explained by multivariate statistical analysis procedure can be considered as a more valid approach in differentiating 100% Arabica coffee from coffees adulterated with Robusta or any other coffee variety.

In addition, nine (9) TAGs (LAA, SLL, PtAL, POO, PPL, OLLn, LALn, PMaAn and SLA) were identified as marker compounds for Robusta coffee and occurring in significantly higher concentration when compared to Arabica coffee.

ACKNOWLEDGEMENT

The authors are grateful to Deutscher Kaffeeverband, Germany for providing some of the green coffee samples.

REFERENCES

Abrankó, L., & Clifford, M. N. (2017). An Unambiguous Nomenclature for the Acyl-quinic Acids Commonly Known as Chlorogenic Acids. *Journal of Agricultural and Food Chemistry, 65(18), 3602–3608*.

Anthony, F., Combes, M. C., Astorga, C., Bertrand, B., Graziosi, G., & Lashermes, P. (2002). The origin of cultivated *Coffea arabica L*. varieties revealed by AFLP and SSR markers. *Theoretical and Applied Genetics*, 104(5), 894-900.

Badmos, S., Lee, S., & Kuhnert, N. (2019). Comparison and quantification of chlorogenic acids for differentiation of green Robusta and arabica coffee beans. *Food Research International*, 126, 108544.

Barbin, D. F., Felicio, A. L. S. M., Sun, D., Nixdorf, S. L., & Hirooka, E. Y. (2014). Application of infrared spectral techniques on quality and compositional attributes of coffee: An overview. *Food Research International*, 61, 23-32.

Briandet, R., Kemsley, E. K., & Wilson, R. H. (1996). Discrimination of arabica and Robusta in instant coffee by Fourier transform infrared spectroscopy and chemometrics. *Journal of Agricultural and Food Chemistry*, 44(1), 170-174.

Carisano, A. and Gariboldi, L. (1964). Gas chromatographic examination of the fatty acids of coffee oil. *Journal of the Science of Food and Agriculture, 15(9), 619–622.*

Chong, J., Wishart, D.S. and Xia, J. (2019). Using *MetaboAnalyst* 4.0 for comprehensive and integrative metabolomics data analysis. *Current Protocols in Bioinformatics* 68, e86. https://www.metaboanalyst.ca/MetaboAnalyst/home.xhtml/ Accessed 10 November 2019.

Clifford, M. N., Johnston, K. L., Knight, S., Kuhnert, N. (2003) The characterisation by LC-MS^n of coffee seed caffeoylferuloylquinic acids. *J. Agric. Food Chem* 51, 2900–2911.

Cossignani, L., Montesano, D., Simonetti, M. S., & Blasi, F. (2016). Authentication of *Coffea arabica* according to triacylglycerol stereospecific composition. *Journal of Analytical Methods in Chemistry*, 2016: 1 – 7.

Dias, R., & Benassi, M. (2015). Discrimination between arabica and Robusta coffees using hydro soluble compounds: Is the efficiency of the parameters dependent on the roast degree? *Beverages*, 1(3), 127-139.

Dong, W., Tan, L., Zhao, J., Hu, R., & Lu, M. (2015). Characterization of fatty acid, amino acid and volatile compound compositions and bioactive components of seven coffee (*Coffea robusta*) cultivars grown in Hainan province, china. *Molecules*, 20(9), 16687-16708.

El-Abassy, R. M., Donfack, P., & Materny, A. (2011). Discrimination between arabica and Robusta green coffee using visible micro Raman spectroscopy and chemometric analysis. *Food Chemistry*, 126(3), 1443-1448.

Fahy, E., Subramaniam, S., Murphy, R. C., Nishijima, M., Raetz, C. R. H., Shimizu, T., Dennis, E. A. (2009). Update of the LIPID MAPS comprehensive classification system for lipids. *Journal of Lipid Research*, 50 (Suppl.), S9-S14.

Folmer, B., Blank, I., & Hofmann, T. (2017). Crema-formation, stabilization, and sensation. *The Craft and Science of Coffee*, 399-417.

Folstar, P. (1985). Lipids in coffee chemistry in: Clarke R.J., Macrae R. (eds) coffee. *Springer*, Dordrecht, London, 1, 203-222.

González, A. G., Pablos, F., Martín, M. J., León-Camacho, M., & Valdenebro, M. S. (2001). HPLC analysis of tocopherols and triglycerides in coffee and their use as authentication parameters. *Food Chemistry*, 73(1), 93-101.

Gunning, Y., Defernez, M., Watson, A. D., Beadman, N., Colquhoun, I. J., Le Gall, G., Kemsley, E. K. (2018). 16-*O*-methylcafestol is present in ground roast arabica coffees: Implications for authenticity testing. *Food Chemistry*, 248, 52-60.

Heldt, H., & Piechulla, B. (2011). Lipids are membrane constituents and function as carbon stores. *Plant Biochemistry*, 359–398.

Holčapek, M., Lísa, M., Jandera, P., & Kabátová, N. (2005). Quantitation of triacylglycerols in plant oils using HPLC with APCI-MS, evaporative light-scattering, and UV detection. *Journal of Separation Science*, 28(12), 1315-1333.

Hvattum, E. (2001). Analysis of triacylglycerols with non-aqueous reversed-phase liquid chromatography and positive ion electrospray tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, 15(3), 187-190.

Indelicato, S., Bongiorno, D., Pitonzo, R., Di Stefano, V., Calabrese, V., Indelicato, S., & Avellone, G. (2017). Triacylglycerols in edible oils: Determination, characterization, quantitation, chemometric approach and evaluation of adulterations. *Journal of Chromatography A*, 1515, 1-16.

Iriondo-Dehond, A., Santillan Cornejo, F., Fernandez-Gomez, B., Vera, G., Guisantes-Batan, E., Alonso, S. G., Del Castillo, M. D. (2019). Bioaccesibility, metabolism, and excretion of lipids composing spent coffee grounds. *Nutrients*, 11(6).

Jaiswal, R.; Kuhnert, N. (2011) Identification and characterization of five new classes of chlorogenic acids in burdock (*Arctium lappa L.*) roots by liquid chromatography/tandem mass spectrometry. *Food Funct*. 2, 63−71.

Jaiswal R, Patras M. A., Eravuchira P. J., Kuhnert N. (2010). Profile and characterization of the chlorogenic acids in green Robusta coffee beans by LC-MS^n: identification of seven new classes of compounds *J Agric Food Chem* 58(15):8722-37.

Jham, G. N., Nikolova-Damyavova, B., Viera, M., Natalino, R., & Rodrigues, A. C. (2003). Determination of the triacylglycerol composition of coffee beans by reverse-phase high-performance liquid chromatography. *Phytochemical Analysis*, 14(5), 310-314.

Lelyana, R. (2017). The functional benefit of coffee for being healthy and beautiful. *Journal of Nanomedicine & Nanotechnology*, 8(6).

Lísa, M., & Holčapek, M. (2008). Triacylglycerols profiling in plant oils important in food industry, dietetics and cosmetics using high-performance liquid chromatography-atmospheric pressure chemical ionization mass spectrometry. *Journal of Chromatography A*, 1198-1199 (1-2), 115-130.

Lísa, M., Holčapek, M., & Boháč, M. (2009). Statistical evaluation of triacylglycerol composition in plant oils based on high-performance liquid chromatography - atmospheric pressure chemical ionization mass spectrometry data. *Journal of Agricultural and Food Chemistry*, 57(15), 6888-6898.

Mondego, J. M. C., Vidal, R. O., Carazzolle, M. F., Tokuda, E. K., Parizzi, L. P., Costa, G. G. L., Pereira, G. A. G. (2011). An EST-based analysis identifies new genes and reveals distinctive gene expression features of *Coffea arabica* and *Coffea canephora*. *BMC Plant Biology*, 11.

Murphy, R. C., & Axelsen, P. H. (2011). Mass spectrometric analysis of long-chain lipids. *Mass Spectrometry Reviews*, 30(4), 579-599.

Nikolova-Damyanova, B., Velikova, R., & Jham, G. N. (1998). Lipid classes, fatty acid composition and triacylglycerol molecular species in crude coffee beans harvested in Brazil. *Food Research International*, 31(6-7), 479-486.

Oliveira, L. S., Franca, A. S., Mendonça, J. C. F., & Barros-Júnior, M. C. (2006). Proximate composition and fatty acids profile of green and roasted defective coffee beans. *LWT - Food Science and Technology*, 39(3), 235-239.

Patzelt, A., Lademann, J., Richter, H., Darvin, M. E., Schanzer, S., Thiede, G., Hauser, M. (2012). In vivo investigations on the penetration of various oils and their influence on the skin barrier. *Skin Research and Technology*, 18(3), 364-369.

Pay, E. (2009). The market for organic and fair-trade coffee: The state of agricultural commodity markets. (Food and Agriculture Organization of The United Nations). Rome, Italy: *FAO*. Accessed 20 February 2020. http://www.fao.org/fileadmin/templates/organicexports/docs/Market_Organic_FT_Coffee.pdf

Pustjens, A. M., Weesepoel, Y., & van Ruth, S. M. (2016). Food fraud and authenticity: Emerging issues and future trends. Innovation and Future *Trends in Food Manufacturing and Supply Chain Technologies*, 3-20.

R Core Team. (2019). A language and environment for statistical computing. *R foundation for statistical computing*, Vienna, Austria. https://www.R-project.org/. Accessed 6 Dec. 2019.

Rueda, A., Seiquer, I., Olalla, M., Giménez, R., Lara, L., & Cabrera-Vique, C. (2014). Characterization of fatty acid profile of argan oil and other edible vegetable oils by gas chromatography and discriminant analysis. *Journal of Chemistry*, 2014.

Sandra, P., Medvedovici, A., Zhao, Y., & David, F. (2002). Characterization of triglycerides in vegetable oils by silver-ion packed-column supercritical fluid chromatography coupled to mass spectroscopy with atmospheric pressure chemical ionization and coordination ion spray. *Journal of Chromatography A*, 974(1-2), 231-241.

Schievano, E., Finotello, C., De Angelis, E., Mammi, S., & Navarini, L. (2014). Rapid authentication of coffee blends and quantification of 16-O-methylcafestol in roasted coffee beans by nuclear magnetic resonance. *Journal of Agricultural and Food Chemistry*, 62(51), 12309-12314.

Segall, S. D., Artz, W. E., Raslan, D. S., Jham, G. N., & Takahashi, J. A. (2005). Triacylglycerol composition of coffee beans (*Coffea canephora P*.) by reversed phase high-performance liquid chromatography and positive electrospray tandem mass spectroscopy. *Journal of Agricultural and Food Chemistry*, 53(25), 9650-9655.

Sirbu, D., Corno, M., Ullrich, M. S., & Kuhnert, N. (2018). Characterization of triacylglycerols in unfermented cocoa beans by HPLC-ESI mass spectrometry. *Food Chemistry*, 254, 232-240.

Speer, K., & Kölling-Speer, I. (2006). The lipid fraction of the coffee bean. *Brazilian Journal of Plant Physiology*, 18(1), 201-216.

Toci, A. T., Neto, V. J. M. F., Torres, A. G., & Farah, A. (2013). Changes in triacylglycerols and free fatty acids composition during storage of roasted coffee. *LWT - Food Science and Technology*, 50(2), 581-590.

Wagemaker, T. A. L., Carvalho, C. R. L., Maia, N. B., Baggio, S. R., & Guerreiro Filho, O. (2011). Sun protection factor, content and composition of lipid fraction of green coffee beans. *Industrial Crops and Products*, 33(2), 469-473.

Wagemaker, T. A. L., Silva, S. A. M., Leonardi, G. R., & Maia Campos, P. M. B. G. (2015). Green *Coffea arabica L:* Seed oil influences the stability and protective effects of topical formulations. *Industrial Crops and Products*, 63, 34-40.

Wei, F., & Tanokura, M. (2015). Chemical changes in the components of coffee beans during roasting. *Coffee in Health and Disease Prevention*, 83-91.

Wilson, A. J., Petracco, M., & Illy, E. (1997). Some preliminary investigations of oil biosynthesis in the coffee fruit and its subsequent re-distribution within green and roasted beans. Paper presented at the 17th International Colloquium on the Chemistry of Coffee, *ASIC*, Paris, France. 92-99.

Yu, S., Du, S., Yuan, J., & Hu, Y. (2016). Fatty acid profile in the seeds and seed tissues of *Paeonia L. species* as new oil plant resources. *Scientific Reports*, 6.

SUPPLEMENTARY DATA

Table 1 Percentage composition of lipids extracted from the respective green coffee samples.

S/N	Coffee origin	Bean variety	Amount of extracted lipids (%)
1	Brazil	Arabica	16.57
2	China	Arabica	13.17
3	Colombia	Arabica	16.68
4	Ecuador	Robusta	9.54
5	Ethiopia	Arabica	13.69
6	Guatemala	Arabica	13.89
7	Honduras	Arabica	14.32
8	India	Robusta	8.62
9	Indonesia	Robusta	9.91
10	Jamaica	Arabica	14.17
11	Kenya	Arabica	14.82
12	Nicaragua	Arabica	14.33
13	Papua New Guinea	Arabica	13.81
14	Peru	Arabica	16.24
15	Rwanda	Arabica	13.85
16	Tanzania	Robusta	8.56
17	Costa Rica	Arabica	14.37
18	Laos	Arabica	13.49
19	Uganda	Robusta	11.05
20	Vietnam	Robusta	9.68

Table 2 Coffee fatty acids with their chemical characteristics, symbols and molecular masses.

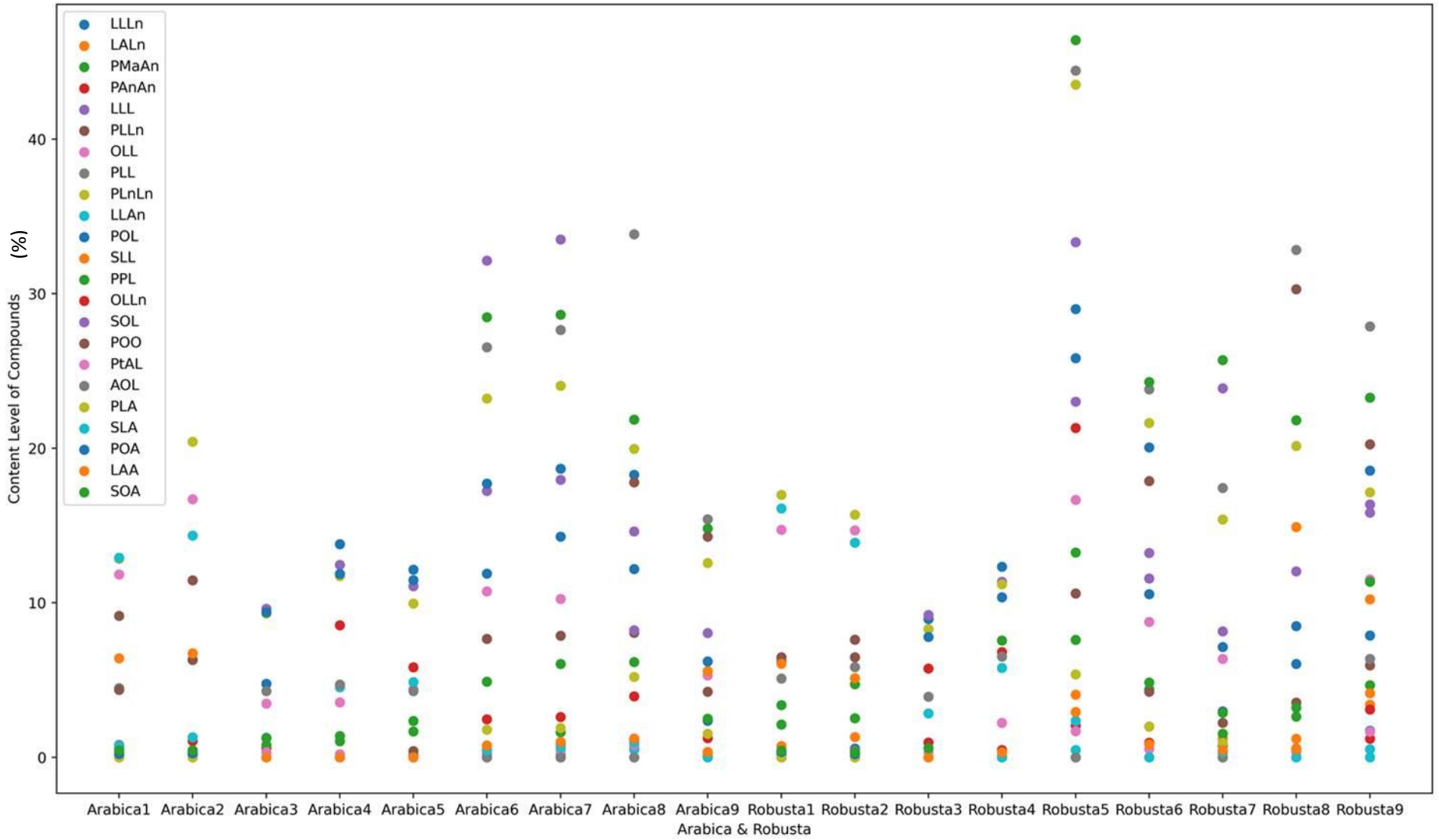
Fatty Acid Common Name	Symbol	CN:DB	R-COOH	Theor. Mass	Expt. Mass
Capric	C	C10:0	$C_{10}H_{20}O_2$	172.1463	172.2201
Lauric	L	C12:0	$C_{12}H_{24}O_2$	200.1776	200.2313
Myrisitc	M	C14:0	$C_{14}H_{28}O_2$	228.2089	228.2421
Pentacosanoic	Pt	C15:0	$C_{15}H_{30}O_2$	242.2246	242.2254
Palmitic	P	C16:0	$C_{16}H_{32}O_2$	256.2402	256.2021
Margaric	Ma	C17:0	$C_{17}H_{35}O_2$	270.2559	270.2541
Stearic	S	C18:0	$C_{18}H_{36}O_2$	284.2715	284.2023
Oleic	O	C18:1	$C_{18}H_{34}O_2$	282.2559	282.2212
Linoleic	L	C18:2	$C_{18}H_{32}O_2$	280.2402	280.2031
Linolenic	Ln	C18:3	$C_{18}H_{30}O_2$	278.2246	278.2114
Nonadecanoic	N	C19:0	$C_{19}H_{38}O_2$	298.2872	298.2432
Arachidic	A	C20:0	$C_{20}H_{40}O_2$	312.3028	312.2112
Arachidonic	An	C20:4	$C_{20}H_{32}O_2$	304.2402	304.2624
Behenic	B	C22:0	$C_{22}H_{44}O_2$	340.3341	340.2422
Lignoceric	Lg	C24:0	$C_{24}H_{48}O_2$	368.3654	368.2671

APPENDIX

Table 4 P-values ($p < 0.05$) of Arabica against Robusta green coffee TAGs.

	OLLn	PLA	OLL	SOA	SOL	POL	POA	LALn	PMaAn	PAnAn	PLL	LLL	PPL	PLLn	LLLn	SLL	POO	LAA	PLnLn	PtAL	SLA	LLAn	AOL
OLLn	NA	0.02	0.07	0.03	0.10	0.01	0.00	0.06	0.06	0.98	0.40	0.34	0.24	1.00	0.30	0.49	0.15	0.56	0.45	0.30	0.37	0.11	0.37
PLA	0.02	NA	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.02	0.00	0.00	0.00	0.39	0.65	0.56	0.72	0.11	0.04	0.01	0.01	0.36	0.47
OLL	0.07	5E-05	NA	0.00	2E-05	2E-03	0.00	0.01	0.00	0.06	0.01	0.00	0.00	0.38	0.07	0.98	0.57	0.42	0.10	0.06	0.06	0.38	0.95
SOA	0.03	5E-05	0.00	NA	6E-03	1E-03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.30	0.11	0.40	0.57	0.63	0.14	0.07	0.07	0.78	0.52
SOL	0.10	0.00	0.00	0.01	NA	0.00	0.00	0.24	0.18	0.06	0.03	0.00	0.00	0.62	0.12	0.35	0.07	0.15	0.07	0.03	0.04	0.79	0.96
POL	0.01	0.00	0.00	0.00	0.00	NA	0.00	0.16	0.08	0.03	0.01	0.00	0.00	0.68	0.80	0.47	0.65	0.06	0.05	0.01	0.01	0.67	0.97
POA	0.00	0.00	0.00	0.00	0.00	0.00	NA	0.01	0.02	0.02	0.00	0.00	0.00	0.98	0.33	0.59	0.93	0.07	0.01	0.00	0.00	0.71	0.27
LALn	0.06	0.01	0.01	0.00	0.24	0.16	0.01	NA	0.00	0.25	0.00	0.01	0.00	0.20	0.11	0.19	0.17	0.37	0.89	0.84	0.79	0.12	0.84
PMaAn	0.06	0.00	0.00	0.00	0.18	0.08	0.02	4E-10	NA	0.33	0.00	0.00	0.00	0.11	0.23	0.30	0.22	0.34	0.90	0.99	1.00	0.07	0.42
PAnAn	0.98	0.02	0.06	0.00	0.06	0.03	0.02	3E-01	0.33	NA	0.00	0.00	0.01	0.03	0.10	0.65	0.31	0.95	0.52	0.27	0.25	0.13	0.58
PLL	0.40	0.00	0.01	0.00	0.03	0.01	0.00	4E-03	0.00	0.00	NA	4E-10	4E-09	0.03	0.25	0.12	0.06	0.73	0.41	0.19	0.19	0.01	0.06
LLL	0.34	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	NA	2E-10	0.06	0.11	0.28	0.24	0.51	0.20	0.08	0.08	0.03	0.18
PPL	0.24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	NA	0.09	0.16	0.38	0.36	0.48	0.24	0.10	0.11	0.02	0.08
PLLn	1.00	0.39	0.38	0.30	0.62	0.68	0.98	0.20	0.11	0.03	0.03	0.06	0.09	NA	0.55	0.81	0.58	0.00	0.00	0.02	0.02	0.01	0.00
LLLn	0.30	0.65	0.07	0.11	0.12	0.80	0.33	0.11	0.23	0.10	0.25	0.11	0.16	0.552	NA	0.25	0.64	0.51	0.64	0.87	0.74	0.18	0.51
SLL	0.49	0.56	0.98	0.40	0.35	0.47	0.59	0.19	0.30	0.65	0.12	0.28	0.38	0.81	0.25	NA	0.00	0.81	0.28	0.43	0.34	0.33	0.78
POO	0.15	0.72	0.57	0.57	0.07	0.65	0.93	0.17	0.22	0.31	0.06	0.24	0.36	0.58	0.64	0.00	NA	0.77	0.91	0.92	0.79	0.08	0.43
LAA	0.56	0.11	0.42	0.63	0.15	0.06	0.07	0.37	0.34	0.95	0.73	0.51	0.48	0.00	0.51	0.81	0.77	NA	3E-07	7E-08	2E-07	2E-01	2E-01
PLnLn	0.45	0.04	0.10	0.14	0.07	0.05	0.01	0.89	0.90	0.52	0.41	0.20	0.24	0.00	0.64	0.28	0.91	3E-07	NA	3E-12	1E-11	2E-01	6E-02
PtAL	0.30	0.01	0.06	0.07	0.03	0.01	0.00	0.84	0.99	0.27	0.19	0.08	0.10	0.02	0.87	0.43	0.92	7E-08	3E-12	NA	5E-15	3E-01	2E-01
SLA	0.37	0.01	0.06	0.07	0.04	0.01	0.00	0.79	1.00	0.25	0.19	0.08	0.11	0.02	0.74	0.34	0.79	2E-07	1E-11	5E-15	NA	0.31	0.12
LLAn	0.11	0.36	0.38	0.78	0.79	0.67	0.71	0.12	0.07	0.13	0.01	0.03	0.02	0.01	0.18	0.33	0.08	2E-01	2E-01	3E-01	0.31	NA	0.00
AOL	0.37	0.47	0.95	0.52	0.96	0.97	0.27	0.84	0.42	0.58	0.06	0.18	0.08	0.00	0.51	0.78	0.43	2E-01	6E-02	2E-01	0.12	0.00	NA

Figure 5 Scatter plot of TAGs (ANOVA) extracted from Arabica and Robusta green coffees. Each colour dot represents the different TAGs in samples.



Chapter 5

Multivariate discriminant analysis between *Coffea arabica* and *Coffea canephora* green beans based on fatty acid variations and geographical origin by Gas Chromatography

Chapter 5: CLASSIFICATION OF COFFEE BY GEOGRAPHICAL ORIGIN

1. INTRODUCTION

Coffee is an evergreen plant belonging to the *Rubiaceae* family and happens to be one of the most consumed beverages in the world (Monente et al, 2015). Being among the topmost consumed drinks, the coffee beverage colour is associated with some intrinsic sensory features (such as flavour, taste, aroma and freshness) along with commercial availability, rational price and further lately due to the functional health benefits as established by numerous numbers of epidemiological or clinical studies (Xie et al, 2015; Walton et al, 2013; Liu et al, 2016; Buscemi et al, 2016). In view of this, coffee is abundantly available worldwide and among the largest traded agricultural commodities, providing a livelihood for people both in the developing and developed countries (ICO, 2019).

Amongst the consumers, coffee is a universal and popular beverage, with over US$50 billion in retail sales annually (ICO, 2019). In April 2020, the world coffee exports with global shipments reached approximately 72.78 million bags. Exports from Africa increased by 7% to 7.66 million bags and from Asia and Oceania rose by 0.6% to 23.62 million bags in October 2019 to April 2020. During the same period, shipments from Central America and Mexico declined by 4.9% to 8.77 million bags and from South America by 8.6% to 32.74 million bags (ICO, 2020); with *Coffea arabica* and *Coffea canephora* being the most produced and consumed coffees globally.

Coffee is commonly grown in Brazil, Vietnam, Colombia, Indonesia, Ethiopia, Honduras, India, Uganda, Mexico and Guatemala among many others. These are among the top 10 coffee producing countries in the world according to International Coffee Organization (ICO, 2019), and research studies have shown that the topographic conditions of the producing region highly affect the chemical compounds and quality traits of coffee. Most importantly, climate, altitude and shade play significant roles through temperature, light and water availability during the

ripening period. Rainfall and sunshine distributions have a strong influence on flowering, bean expansion and ripening. For instance, chlorogenic acids and lipids content have been found to increase with elevation in *Coffea arabica* (Clifford, 1985; Bertrand et al, 2003; Silva et al, 2005; Charrier & Berthaud, 1985; Bertrand et al, 2005a; Leroy et al, 2006). The impact of soil types has also been studied and reported that the most acidic coffees are cultivated on rich volcanic soils (Harding et al, 1987).

Moreover, besides tea, coffee has gained much attention despite its caffeine content restricting consumption, and contains a relatively large amount of other bioactive compounds, predominantly phenolics, which exhibit a significant amount of antioxidant (Santana-Gálvez et al, 2017), anti-diabetic (Clifford et al, 2017; Pimpley et al, 2020; Ludwig et al, 2014; Wilson et al, 2011), hepato-protective (Wijarnpreecha, Thongprayoon & Ungprasert, 2017; Shi et al, 2013), anti-hypertensive (Wan et al, 2013) and anti-inflammatory (Liu et al, 2017) properties; and also protect against brain neurological degeneration associated with oxidative stress (Veljkovic et al, 2018) when consumed as part of regular human diet.

However, it has been reported that green coffee possesses antioxidant potential against lipid peroxidation via in-vitro studies (Kroyer et al, 1989) and antineoplastic (inhibition of a neoplasm ‘tumour’ development) activity (Rosenberg, 1990). Also roasted coffee has been established to have mutagenic activity against sudden changes in gene expression, perhaps as a result of hydrogen peroxide formation attributed to a key provider to coffee genotoxicity in-vitro (Miller et al, 1993) and antibacterial activity against a wide range of known bacteria (Bajko et al, 2016; Farzaei et al, 2015; Naveed et al, 2018).

Nevertheless, the consumption of coffee has been reported to be associated with some health disadvantages due to the presence of some carcinogenic substances in the coffee brew, which mostly occurs during coffee roasting process. Epidemiological studies have shown that frequent

coffee consumption may leads to a slight decrease in cancer incidence, a phenonmenon commonly refers to as the ‘‘*coffee paradox*’’ (Van Dam, 2008; Yu et al, 2011; Williamson, 2020). The roasting process is a vital phase in coffee production, which promotes the color, flavor and aroma formation, otherwise known as the key physiognomies in determining the coffee quality. Likewise, the roasting process usually leads to the generation of many undesirable compounds such as the key carcinogens like acrylamide and 5-HMF (Capuano & Fogliano, 2011), and polycyclic aromatic hydrocarbons (PAH) (Oosterveld et al, 2003). Among the resulting PAH compounds are acenaphthylene, acenaphthene, fluorene, phenanthrene, anthracene, 2-methyl-phenanthrene, 2-methyl-anthracene, 2,4-dimethyl phenanthrene, fluoranthene, 1,2-dimethyl phenanthrene, pyrene, methyl-pyrene, benz[a]anthracene, chrysene, benzo[b]fluoranthene, benzo[k]fluoranthene, benzo[e]pyrene, benzo[a]pyrene, perylene, benzo[g,h,i]-perylene, dibenz[a,l]-pyrene, dibenz[a,e]-pyrene, dibenz[a,i]-pyrene, dibenz[a,h] and many more (Orecchio et al, 2009); but these compounds generally contribute insignificant quantities to the human daily intake of carcinogenic PAHs.

Among several coffee components, lipids constitute the most essential components of the beans, which consists of triglycerides, wax and unsaponifiable materials most notably the terpenes (Tango, 1971). The oil content in *Coffea arabica* is about 16% or more, while *Coffea canephora* contains about 10% (Clifford, 1985) or less. The fatty acid composition of green beans is determined by many factors, including the bean species and climatic conditions during cultivation and growth development (Alves et al, 2003; Villareal et al, 2009).

Moreover, it has been reported that green coffee beans are abundantly rich in unsaturated fatty acids the likes of oleic, linoleic and linolenic (18:3n-3) acids (Dong et al, 2015), which are otherwise known as essential fatty acids that the human physiological body cannot specifically synthesis. To an extent, fatty acids composition has also been explored in differentiating Arabica from Robusta coffee (Toci et al, 2013; Romano et al, 2014).

In addition, lipids play an important role in the quality of several important plant crops such as soybeans, cacao, etc., aside being important flavour complexes in coffee (Akiyama et al, 2004). Despite the available information on the composition of coffee triacylglycerols (TAGs) (Nikolova-Damyanova et al, 1998; Jham et al, 2001; 2003; 2005; 2006; 2007; Gonzalez et al, 2001; Segall et al, 2005), it has been hypothesized that hydrolysis of TAGs typically results in the release of free fatty acids, which are oxidized to produce off-flavour (Multon et al, 1973; Wajda and Walczyk, 1978; Foumey et al, 1982; Spadone et al, 1990; Speer et al, 1993) in any given coffee species. Hence, it is quite pertinent to further investigate extensively the role of lipids in coffee quality, since TAGs and its hydrolysis results may affect or lower coffee quality (Kaufmann and Hamsagar, 1962; Nikolova-Damyanova et al, 1998).

Before now, some studies have examined the fatty acids constituents in the coffee localised to some countries like from Brazil (Luisa et al; 2015; Nikolova-Damyanova et al, 1998), Ethiopia (Mehari et al; 2019), Colombia (Hurtado-Benavides et al, 2016) and China (Dong et al, 2015), but detailed information on the variations in the fatty acids altogether from all the regions of the world (Central America, South America, East Africa, West Africa and Asia) where coffee is mainly cultivated is yet to be investigated. Although, Martin profiled fatty acids in coffees from some regions of the world to differentiate Arabica from Robusta coffees, but the study did not categorise the coffee specifically according to the regions and based on the fatty acid contents (Martin et al, 2001). Also, similar study was carried out by D'Souza et al, (2017) where lipid constituents of cocoa butter were explored in profiling and classifying cocoa cultivated from different geographical origins using lipids constituents.

The main objective of food security is to aid accessibility of adequate, safe and nutritious food by the population to meet the diet and health requirements (FAO, 1996). Since, the quality of each coffee species varies with its cultivars; geographical origin and climatic environment coupled with the growing recognition and significance of green coffee beans to potential health

benefits for the consumers and global economic importance, its unique composition and properties have attracted great attention most recently (Dziki et al, 2015). Hence, this study aims to investigate the fatty acids constituents of coffee from all the main regions where coffee are produced across the globe and provide more information on the variations and differences. Furthermore, it intends to explore the information to discriminate between the two main coffee varieties based on the fatty acids composition and cultivation origin.

2. MATERIALS AND METHODS

2.1 Chemicals and reagents

All reagents were of analytical grade. Acetonitrile (LC-MS ultra-gradient grade), chloroform (HPLC grade), methanol (HPLC grade), heptane (HPLC grade) were purchased from Carl Roth (Karlsruhe, Germany); while methanol (MeOH), potassium hydroxide (KOH) and sulphuric acid (H_2SO_4) were purchased from Sigma-Aldrich (Steinheim, Germany).

2.2 Coffee beans

Fifty-nine (59) green coffee beans produced from sixteen (16) countries and four (4) geographical locations across the world were used in this study. The coffee beans originated from South America (Brazil, Colombia, and Peru), Central America (Costa Rica, Guatemala, Honduras, Jamaica and Nicaragua), Africa (Ethiopia, Kenya, Rwanda, Tanzania and Uganda) and Asia (India, Indonesia, and Vietnam) were purchased from local grocery shops in Bremen and around Germany were used in this study. Details of the respective green coffee beans are provided in Table 1. The authors of this study selected and categorically used green coffee beans as a result of the fact that roasted coffee beans cannot be much relied upon owing to loss of some lipid fractions during roasting.

Table 1 Green coffee bean from different geographical origins.

	Green Coffee Bean	Bean Variety	Geographical Origin	Sample Quantity
1	Brazil	*Coffea arabica*	South America	7
2	Colombia	*Coffea arabica*	South America	5
3	Costa Rica	*Coffea arabica*	Central America	2
4	Ethiopia	*Coffea arabica*	Africa	5
5	Guatemala	*Coffea arabica*	Central America	2
6	Honduras	*Coffea arabica*	Central America	3
7	India	*Coffea canephora*	Asia	6
8	Indonesia	*Coffea canephora*	Asia	3
9	Jamaica	*Coffea arabica*	Central America	3
10	Kenya	*Coffea arabica*	Africa	3
11	Nicaragua	*Coffea arabica*	Central America	3
12	Peru	*Coffea arabica*	South America	6

13	Rwanda	*Coffea arabica*	Africa	2
14	Tanzania	*Coffea canephora*	Africa	5
15	Uganda	*Coffea canephora*	Africa	2
16	Vietnam	*Coffea canephora*	Asia	2

2.3 Coffee oil extraction

Details as described in section 2.4 of Chapter four.

2.4 Coffee bean variety validation by Nuclear magnetic resonance

The different coffee beans are firstly validated to determine the variety of each of the bean. Nuclear magnetic resonance using method previously reported (Schievano et al, 2014). The presence of a bio-marker compound 16-*O*-methylcafestol in Robusta coffee was employed in categorizing the bean into two main classes as Arabica or Robusta coffee. The coffee bean samples were ground using RAPIDO coffee grinder (TZS-FIRST, Austria) in liquid nitrogen to the finest particle size. Then 0.5 g of weighed powder was dissolved in 1.5 mL of $CDCl_3$ (99.96% deuterated chloroform with 0.05% (v/v) TMS, stabilized with silver) procured from Sigma Aldrich, Germany. The mixture was vortex for 15 min at maximum speed and swiftly filtered into 5 mm precision glass NMR tubes with the aid of a cotton wool. Thereafter, NMR spectrum was acquired. The procedure was carried out at 4 °C to minimize solvent loss (Schievano et al, 2014).

The NMR spectroscopy performed using Oxford AS400 instrument (England) joined with JOEL STAC-MAN auto-sample injector (Germany). The spectra were acquired using ^1H in deuterated chloroform with a pulse sequence, 16 scans, spectral width of 400 Hz and 32,768 data points to detect the presence of esterified 16-*O*-methylcafestol at 3.12 ppm.

2.5 Derivatization of Fatty acid by Fatty acids methyl ester (FAME)

The FAME was prepared by dissolving 200 mg of extracted oil was dissolved in 500 µL of chloroform in a 10 mL glass vial, a clean magnetic stirrer was added and vortex for 10 sec to

get the oil completely dissolved in the solvent. Thereafter, 700 µL of 10N $KOH_{(aq)}$ was added to saponify the mixture (alkaline hydrolysis) and 5.3 mL of methanol (methylation). The mixture was incubated for 1 hr at 55 ºC with stirring and intermittent shaking in between. During the incubation, saponification process took place in the presence of acidic catalyst (MeOH) to cleave the ester bonds and allowed the release of fatty acids from the glycerol moiety. Afterwards, the mixture was cooled down to room temperature in a water bath for about 5 min, then 580 µL of H_2SO_4 was added and incubated for another 1 hr at 55 ºC. Then the mixture was cooled down to room temperature in a water bath to room temperature and 2 mL of heptane was added and vortex for 2 min. The resulting mixture was transferred into 15 mL Falcon tube and centrifuged at 3000 rpm for 5 min. Then 1.5 mL of the organic upper layer (heptane phase) was transferred into a clean 10 mL glass vial and 500 mg of activated molecular sieve was added. The vial was allowed to stand for about 10 min before it was transferred into GC vial. The extract was passed under Nitrogen gas before covering the vial and store at 4 ºC for further analysis. Subsequently, the extract was diluted with chloroform with ratio 1:20 by adding 50 µL of extract to 950 µL of chloroform. Fatty acid standards, palmitic, linoleic, stearic and oleic acids were dissolved in chloroform (5 mg/mL) and derivatized under similar conditions corresponding to the samples. The samples were then analysed on gas chromatography (FAME method adopted from AOCS, 1989 and Firestone, 1995).

2.6 Gas chromatography

The GC analysis was carried out using a gas chromatograph Shimadzu GC-2010-Plus coupled with an AOC-20i sample auto-injector and equipped with a flame ionization detector (FID-2010 Plus). A DB-23 column made of fused silica capillary tubing (length-30 m, inner diameter - 0.25 mm, film thickness - 0.25 µm with 7 inch) was used. The temperature setting was programmed to hold at 50 ºC for 1 min; ramped to 170 ºC at 15 ºC/min, then further elevated to 210 ºC at 2 ºC/min and held for 5 min; total run time was 34 min. An injector (split 1/10)

was used with chloroform as injection solvent with a pressure of 149 kPa at 250 °C and FID detector at 250 °C with a purge flow rate of 3 mL/min. Helium was used as the carrier gas with air-flow of 400 mL/min and an injection volume of 1 µL. Peak identification was made by comparing standards of methylated fatty acids analysed on a DB-23 column. The final content of fatty acids is given as the percentage of dry matter (% d.m.). Details of the fatty acids quantified based on peak area normalization method (Christie, 1989) are listed in Table 2.

2.7 Statistical analysis

Analysis of variance (ANOVA) was used to study the variances in the fatty acid contents of coffees based on geographical origins. Data processing and analysis were carried out using Python for P-values ANOVA significance, Pearson correlation coefficient and heatmap. All statistical tests were conducted at the 95% confidence level ($p < 0.05$).

The data acquired from GC Lab solutions software was further subjected to multivariate statistical analyses using *MetaboAnalyst 4.0* software for comprehensive and integrative metabolomics Data Analysis. Discrimination among samples was performed using principal component analysis (PCA) and partial least square discriminant analysis (PLS-DA) based on the fatty acids chemical compositions across the regions and the different beans variety.

Table 2 Details of the fatty acids constituents in coffee based on the DB-23 GC column measured as fatty acid methyl esters following esterification.

	RT (min)	Compound Name	Common Name	Molecular formula	Molecular mass
1	2.88	Methyl butyrate	Butyric acid	$C_4H_8O_2$	88.1
2	4.59	Methyl hexanoate	Caproic acid	$C_6H_{12}O_2$	116.1
3	6.35	Methyl octanoate	Octanoic acid (Caprylic acid)	$C_8H_{16}O_2$	114.2
4	7.96	Methyl decanoate	Decanoic acid	$C_{10}H_{20}O_2$	172.3
5	8.69	Methyl undecanoate	Undecanoic acid	$C_{11}H_{22}O_2$	186.3
6	9.41	Methyl laurate	Lauric acid	$C_{12}H_{24}O_2$	200.3
7	10.18	Methyl tridecanoate	Tridecanoic acid	$CH_3(CH_2)_{11}COOH$	214.4
8	11.07	Methyl myristate	Tetradecanoic acid	$C_{14}H_{28}O_2$	228.3
9	11.49	Methyl myristoleate	Myristoleic acid (C12:1)	$C_{14}H_{26}O_2$	226.3
10	12.12	Methyl pentadecanoate	Pentadecanoic acid	$C_{15}H_{30}O_2$	242.4
11	12.64	Methyl *cis*-10-pentadecenoate	(Z)-10-Pentadecenoic acid	$C_{15}H_{28}O_2$	240.4
12	13.37	Methyl palmitate	Palmitic acid	$C_{16}H_{32}O_2$	256.4
13	13.82	Methyl palmitoleate	(9Z)-hexadecenoic acid (Palmitoleic acid)	$C_{16}H_{30}O_2$	254.4
14	14.85	Methyl heptadecanoate	Heptadecanoic acid (Margaric acid)	$C_{17}H_{34}O_2$	270.5
15	15.39	*Cis*-10-Heptadecenoic acid methyl ester	*Cis*-9-heptadecanoic acid (Margaroleic acid; C17:1)	$C_{17}H_{32}O_2$	268.4
16	16.58	Methyl stearate	Stearic acid	$C_{18}H_{36}O_2$	284.4
17	16.87	*Trans*-9-Elaidic acid methyl ester	Elaidic acid (C18:1)	$C_{18}H_{34}O_2$	282.5
18	17.06	*Cis*-9-Oleic acid methyl ester	Oleic acid (C18:1)	$C_{18}H_{34}O_2$	282.5
19	17.55	*Trans*-Linolelaidic acid methyl ester	Linolelaidic acid (Methyl *Trans*, *Trans*-9,12-octadecadienoate; C18:2)	$C_{19}H_{34}O_2$	280.4
20	18.05	Methyl linoleate	Linoleic acid (C18:2)	$C_{18}H_{32}O_2$	280.4
21	18.66	Methyl linolenate	Linolenic acid (C18:3)	$C_{18}H_{30}O_2$	278.4
22	19.36	Methyl γ-linolenate	*Cis, cis, cis*-9,12,15-Octadecatrienoic acid (C18:3)	$C_{18}H_{30}O_2$	278.4
23	20.70	Methyl arachidate	Arachidic acid	$C_{20}H_{40}O_2$	312.5
24	21.29	Methyl *cis*-11-eicosenoate	Eicosenoic acid (Gondoic acid; C20:1)	$C_{20}H_{38}O_2$	310.5
25	22.51	*Cis*-11,14-Eicosadienoic acid	*Cis*-11,14-Eicosadienoic acid (C20:2)	$C_{20}H_{36}O_2$	308.5
26	23.02	Methyl heneicosanoate	Heneicosylic acid (C21:0)	$C_{21}H_{42}O_2$	326.5
27	23.21	*Cis*-8,11,14-Eicosatrienoic acid	*Cis*-8,11,14-Eicosatrienoic acid (Dihomo-gamma-linolenic acid; C20:3)	$C_{20}H_{34}O_2$	306.5
28	23.65	*Cis*-5,8,11,14-Eicosatetraenoic acid	Arachidonic acid (C20:4)	$C_{20}H_{32}O_2$	304.5

29	24.06	*Cis*-11,14,17-Eicosatrienoic acid	*Cis*-11,14,17-Eicosatrienoic acid (C20:3)	$C_{20}H_{34}O_2$	306.5
30	25.24	*Cis*-5,8,11,14,17-Eicosapentaenoic acid methyl ester	*Cis*-5,8,11,14,17-Eicosapentaenoic acid (C20:5)	$C_{20}H_{30}O_2$	302.5
31	25.46	Methyl behenate	Behenic acid (C22:0)	$C_{22}H_{44}O_2$	340.5
32	26.14	Methyl erucate	Erucic acid (Cis-13-docosenoic acid: C22:1)	$C_{22}H_{42}O_2$	338.5
33	27.51	*Cis*-13,16-Docosadienoic acid methyl ester	*Cis*-13,16-Docosadienoic acid (C22:2)	$C_{23}H_{42}O_2$	350.6
34	27.96	Methyl tricosanoate	Tricosanoic acid (C23:0)	$C_{23}H_{46}O_2$	354.6
35	30.62	Methyl lignocerate	Lignoceric acid (C24:0)	$C_{24}H_{48}O_2$	368.6
36	31.22	Methyl nervonate	Nervonic acid (C24:1)	$C_{24}H_{46}O_2$	366.6
37	31.49	*Cis*-4,7,10,13,16,19-Docosahexaenoic acid methyl ester	Docosahexaenoic acid (Cervonic acid: C22:6)	$C_{22}H_{32}O_2$	328.5

3. RESULTS AND DISCUSSION

Several studies have been carried out coffees based on geographical origin most importantly Arabica coffees since it is most sort after between the two top coffee varieties and the consumer most preferred choice (Marquetti et al, 2016; Mehari et al, 2019). Also, numerous studies have investigated the fatty acids constituents of coffee (Alves et al, 2003; Villareal et al, 2009; Dong et al, 2015; Toci et al, 2013; Romano et al, 2014; Clifford, 1985; Luisa et al; 2015; Nikolova-Damyanova et al, 1998; Mehari et al; 2019; Hurtado-Benavides et al, 2016; Martin et al, 2001), but none of these studies was able to extensively studied and classified coffee beans from all the coffee beans producing regions based on the coffee fatty acids constituents and likewise distinguishingly differentiate the beans based on bean variety using multivariate statistical analysis. In the light of this, this study aims to expand the scope of this significant investigation to the four main coffee producing geographical origins, namely East Africa, Asia, Central and South Americans. In view of this, green coffee beans produced from these four continents specifically Brazil, Colombia, Costa Rica, Ethiopia, Guatemala, Honduras, India, Indonesia, Jamaica, Kenya, Rwanda, Nicaragua, Peru, Tanzania, Uganda and Vietnam were used in this study due to the fact that they are highly sourced for in the globally coffee market be it Arabica or Robusta coffee variety.

Thereafter, the coffee beans were ground and their lipid constituents extracted using heptane by Soxhlet method and extraction solvent evaporated. The extracted coffee oil was then derivatized by fatty acid methyl ester (FAME) standard method and analysed on GC to have the detailed fatty acid profiles of each of the coffees from different countries, classify the beans according to their respective source of origin and further investigate the comparison of the bean varieties (Arabica vs Robusta).

3.1 Bean variety authentication by ^{1}H-NMR

NMR spectroscopy was initially used in screening the samples to confirm the authenticity of their varietal species before the rapid approach of analysing the fatty acids constituents by gas chromatography.

Green coffee beans oil was extracted from the different samples using Soxhlet. Each of the individual bean variety was validated using NMR spectroscopy. Using (Schievano et al, 2014) method, 0.5 g of the extracted coffee oil was dissolved in 1.5 mL of deuterated chloroform ($CDCl_3$), vortex for 15 min and filtered into NMR tube with the aid of a cotton-wool. The spectrum was acquired with proton NMR using 16 scans and the results showed that 16-*O*-methylcafestol was present in Robusta at 3.17 ppm as observed in the Robusta sample (Fig. 1) while the peak was absent in the Arabica samples. This procedure was used in confirming the variety authenticity of all the samples before categorically classifying each into Arabica or Robusta coffee groups accordingly.

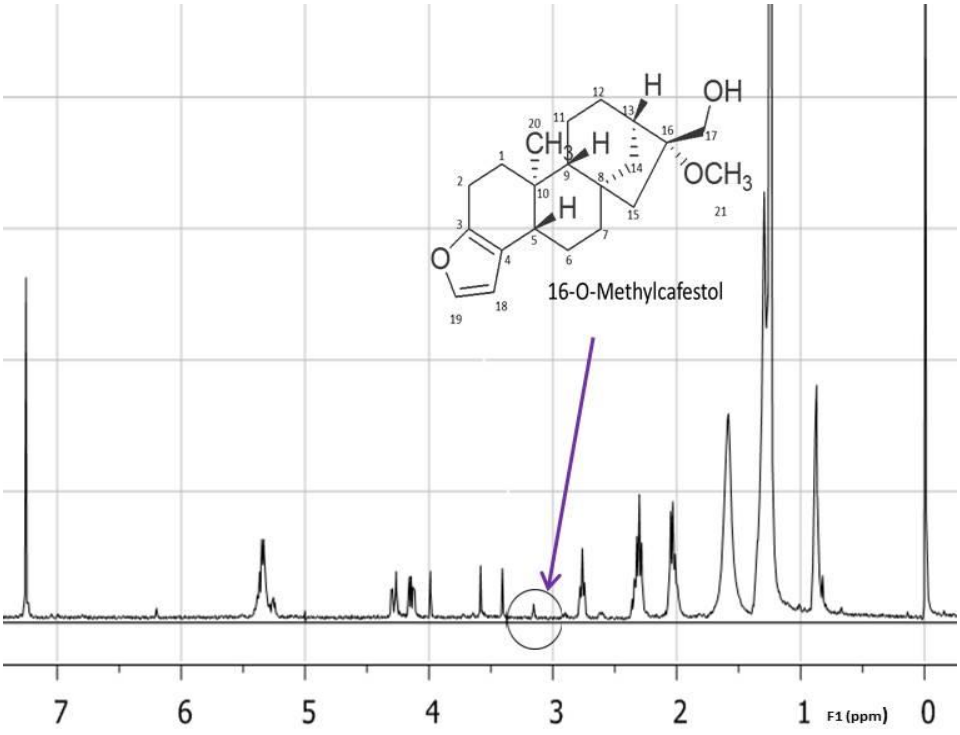


Figure 1 ^{1}H-NMR spectrum of 16-*O*-methylcafestol present in Robusta coffee at 3.17 ppm

3.2 Detection and quantification of fatty acids by Gas chromatography

The extracted coffee oil was derivatized by fatty acid methyl ester (FAME) procedure as explained in section 2.5 above. Thereafter, the derivatized samples were analysed on GC and spectra were generated for each of the samples which were further analysed using Lab Solutions software (Shimadzu Incorporations). Each peak in the chromatogram represents a specific fatty acid as listed in Table 2 with different peak intensity, retention time and chemical properties.

Figure 2 represents a typical chromatogram of the analysed coffee samples. The fatty acids peaks and their respective retention times can be observed. The retention time ranged from approximately 2.8 min to 31.5 min. Table 3 shows means of the fatty acids contents (% d.m.), beans country/geographical origins and the bean variety (Arabica or Robusta).

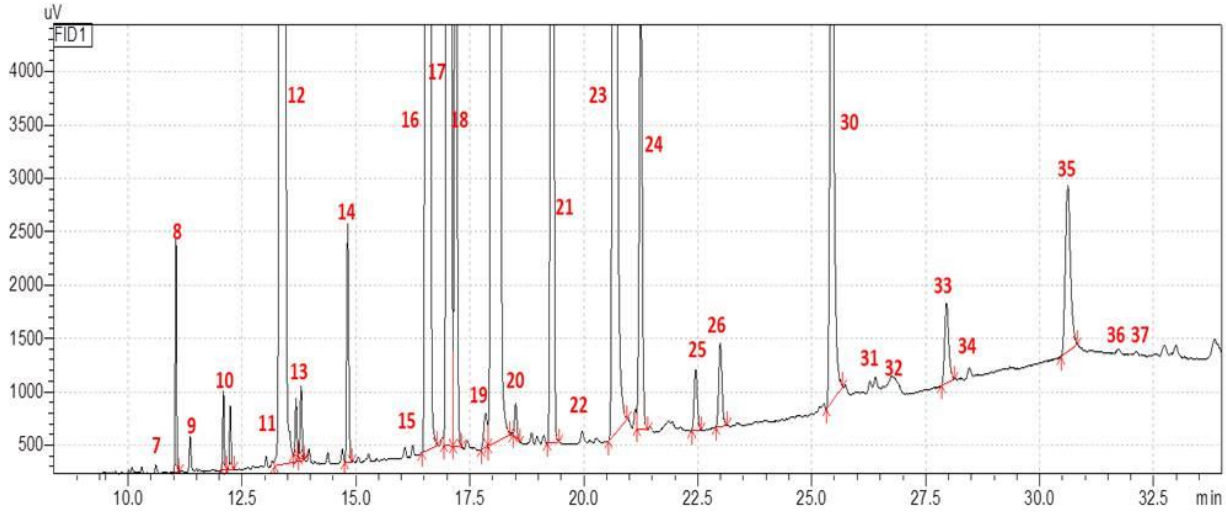


Figure 2 Gas chromatogram of coffee fatty acids. Peaks annotated according to Table 2 and correspond to a specific fatty acid.

From the GC chromatogram, 32 out of 37 analysed fatty acids were detected in majority of the coffee samples. However, linoleic (C18:2), palmitic (C16:0), stearic (C18:0) and oleic (C18:1) acids show higher contents compared to fatty acids. Similarly, arachidic (C20:0) and *Cis,cis,cis*-9,12,15-octadecatrienoic (C18:3) acids indicate values less than 5% but higher than 1% on the

average; except for the *Cis,cis,cis*-9,12,15-octadecatrienoic acid (C18:3) quantities in Robusta which is approximately 50% of the Arabica. The fatty acids constituents acquired from the green coffee in this study correlates with the values previously observed (Martin et al, 2001; Bertrand et al, 2008; Jham et al, 2008a; Joët et al, 2010a and Luisa et al, 2015).

3.3 Trans-fatty acids and Essential fatty acids in coffee

Some studies have reported the presence of *Trans*-oleic acid (C18:1t) in specialty roasted coffee (Alves et al, 2003; Luisa et al, 2015). However, this study also observed that green coffee contains more than one *Trans* fatty acid and most importantly an elaidic acid with two double bonds within its Carbon-18 aliphatic chain called *Trans*-linolelaidic acid (C18:2). Besides, this is observed to be present in green coffee not roasted coffee. Moreover, one of the attributes of *Trans*-fatty acids is the high melting point properties when compared with the *cis* isomer form of the compound (Hagemann et al, 1975). The Danish government issued an order (Order no. 160), which came into operation on 31 March 2003. From the decree, Section 3 states that the content of *trans* fatty acids in the oils and fats covered by the order must not exceed 2 g/100 g of oil or fat. This decree has now been adopted as a general European regulation. From the products that are claimed to be free from fatty acids, the content of *trans* fatty acids in the finished product shall be less than 1 g/100 g of the individual oil or fat (Dijkstra et al, 2008).

Moreover, *Trans*-elaidic acid is an omega-9 *Trans* fatty acid (TFA) and a geometric isomer of oleic acid; while *Trans*-linolelaidic acid is an omega-6 *Trans* fatty acid and a geometric isomer of linoleic acid. It is found in partially hydrogenated vegetable oils. Oleic acids are the most common Omega-9 fatty acid and the most common monounsaturated fatty acid in the diet. Omega-9 fatty acids are not strictly essential, meaning they can be synthesized by the human body. In fact, omega-9 fats are the most abundant fats in most cells in the human body. Likewise, it is quite necessary to state that the three (3) essential fatty acids were detected in

green coffee with this method though in a minute quantities. They are alpha-linolenic acid (C18:3) [ALA], *Cis*-5,8,11,14,17-eicosapentaenoic acid (C20:5) [EPA] and docosahexaenoic acid (C22:6) [DHA]. Figure 3 depicts structures of the *trans*-fatty acids.

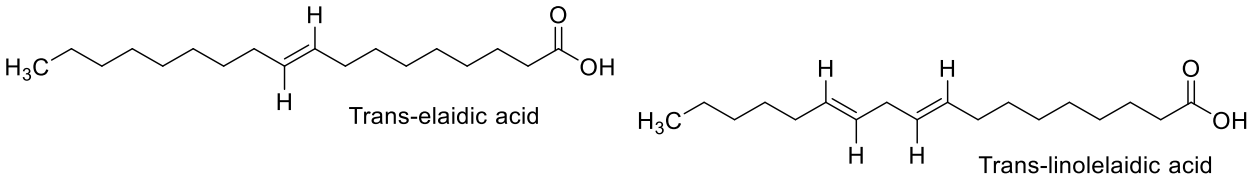


Figure 3 *Trans* Fatty acids

Table 3 represents the mean+s.d of fatty acid variations detected in all the coffee beans across the different geographical origins under study, which are South America (Arabica only), Central America (Arabica only), Africa (Arabica and Robusta) and Asia (Robusta only). The data was further analysed and denoted on box-plots to explain the variations among the regions.

Table 3 Variations (mean+s.d) of the fatty acid constituents of coffee from different geographical origins.

	Regions	South America (Arabica)			Central America (Arabica)					Africa (Arabica)			Africa (Robusta)		Asia (Robusta)		
	Countries	Brazil	Colombia	Peru	Costa Rica	Guatemala	Honduras	Jamaica	Nicaragua	Ethiopia	Kenya	Rwanda	Tanzania	Uganda	India	Indonesia	Vietnam
	Compounds	(n=7)	(n=5)	(n=6)	(n=2)	(n=2)	(n=3)	(n=4)	(n=4)	(n=4)	(n=3)	(n=2)	(n=4)	(n=2)	(n=6)	(n=3)	(n=2)
1	Butyric acid	0.03±0.8	0.03±0.2	0.03±0.01	0.04±0.01	0.03±0.04	0.03±0.08	0.03±0.04	0.03±0.01	0.03±0.02	0.03±0.04	0.03±0.01	0.03±0.01	0.03±0.02	0.03±0.07	0.03±0.3	0.03±0.01
2	Caproic acid	0.00±0.0	0.02±0.1	0.00±0.0	0.03±0.02	0.03±0.01	0.00±0.0	0.00±0.0	0.00±0.0	0.00±0.0	0.00±0.0	0.00±0.0	0.00±0.0	0.00±0.0	0.04±0.07	0.00±0.0	0.02±0.04
3	Octanoic acid	0.02±0.2	0.09±0.01	0.02±0.01	0.41±0.7	0.41±0.5	0.00±0.0	0.02±0.04	0.00±0.0	0.03±0.02	0.00±0.0	0.02±0.01	0.02±0.07	0.00±0.0	0.16±0.7	0.00±0.0	0.00±0.0
4	Decanoic acid	0.02±0.1	0.04±0.02	0.02±0.02	0.29±1.03	0.29±1.18	0.02±0.04	0.02±0.03	0.02±0.05	0.03±0.07	0.00±0.0	0.00±0.0	0.00±0.0	0.00±0.0	0.08±0.03	0.00±0.0	0.00±0.0
5	Undecanoic acid	0.03±0.1	0.03±0.01	0.03±0.01	0.06±0.05	0.04±0.07	0.00±0.0	0.03±0.09	0.03±0.01	0.03±0.06	0.02±0.04	0.03±0.08	0.03±0.03	0.02±0.01	0.03±0.11	0.02±0.06	0.02±0.04
6	Lauric acid	0.03±0.2	0.02±0.01	0.02±0.01	0.02±0.06	0.02±0.01	0.00±0.0	0.03±0.0	0.03±0.12	0.02±0.03	0.02±0.03	0.02±0.01	0.03±0.04	0.02±0.07	0.02±0.01	0.02±0.03	0.03±0.04
7	Tridecanoic acid	0.00±0.0	0.02±0.02	0.00+0.0	0.02+0.02	0.02+0.05	0.00+0.04	0.00±0.0	0.00±0.0	0.00±0.0	0.00±0.0	0.00±0.0	0.00±0.0	0.00±0.0	0.01+0.0	0.00±0.0	0.00±0.0
8	Tetradecanoic acid	0.07±0.2	0.07±0.01	0.07±0.04	0.06±0.04	0.07±0.06	0.06±0.09	0.07±0.1	0.07±0.1	0.07±0.04	0.07±0.3	0.07±0.15	0.08±0.2	0.07±0.08	0.09±0.06	0.10±1.3	0.10±1.1
9	Myristoleic acid (C12:1)	0.03±0.1	0.02±0.04	0.02±0.01	0.00±0.0	0.00±0.0	0.00±0.0	0.02±0.1	0.02±0.03	0.02±0.05	0.00±0.0	0.02±0.7	0.03±0.05	0.02±0.08	0.02±0.07	0.03±0.01	0.04±0.08
10	Pentadecanoic acid	0.03±0.1	0.03±0.02	0.03±0.01	0.03±0.07	0.03±0.6	0.02±0.04	0.03±0.02	0.03±0.09	0.03±0.07	0.03±0.04	0.03±0.02	0.03±0.08	0.03±0.01	0.03±0.05	0.04±0.06	0.04±0.08
11	(Z)-10-Pentadecenoic acid	0.02±0.1	0.02±0.03	0.02±0.02	0.02±0.05	0.02±0.04	0.02±0.01	0.03±0.02	0.02±0.08	0.02±0.08	0.02±0.01	0.02±0.01	0.04±0.02	0.04±0.05	0.04±0.08	0.05±0.01	0.04±0.03
12	Palmitic acid	35.31±1.3	35.66±1.8	36.73±1.5	35.03±1.6	34.97±1.9	37.02±1.7	35.58±1.2	34.71±1.1	37.34±1.7	36.81±0.9	37.50±0.5	34.91±0.8	33.42±2.3	35.11±1.2	34.50±1.8	34.72±1.8
13	Palmitoleic acid	0.03±0.2	0.04±0.01	0.03±0.01	0.04±0.02	0.03±0.02	0.03±0.07	0.03±0.01	0.03±0.06	0.03±0.03	0.03±0.07	0.03±0.08	0.02±0.05	0.02±0.01	0.03±0.07	0.03±0.09	0.03±0.1
14	Margaric acid	0.10±0.01	0.09±0.02	0.10±0.07	0.10±0.01	0.10±0.03	0.10±0.05	0.10±0.08	0.10±0.2	0.10±0.4	0.09±0.04	0.10±0.07	0.07±0.06	0.07±0.08	0.08±0.01	0.07±0.01	0.08±0.03
15	Stearic acid	6.84±1.8	7.01±1.2	6.18±1.9	7.52±2.2	7.46±1.6	5.91±1.3	7.61±1.9	6.95±1.5	6.66±2.3	6.50±1.1	6.86±1.5	6.82±1.9	6.63±1.2	6.40±1.7	6.48±1.4	6.96±1.1
16	*Trans*-Elaidic acid (C18:1)	0.02±0.1	0.01±0.0	0.00±0.0	0.02±0.01	0.02±0.04	0.00±0.0	0.00±0.0	0.00±0.0	0.00±0.0	0.00±0.0	0.00±0.0	0.03±0.01	0.00±0.0	0.02±0.06	0.00±0.0	0.00±0.0
17	Oleic acid (C18:1)	8.17±1.5	8.19±1.7	6.33±1.1	7.06±1.2	7.15±1.3	6.51±2.7	7.38±1.1	7.51±1.8	8.09±1.9	7.12±0.8	8.21±1.3	13.05±1.8	13.95±1.4	10.71±2.8	9.60±2.1	10.31±0.1
18	*Trans*-Linolelaidic acid (C18:2)	0.40±0.1	0.43±0.6	0.48±0.3	0.37±0.5	0.38±1.8	0.41±1.1	0.39±1.2	0.42±1.3	0.40±1.9	0.40±1.3	0.40±2.3	0.31±2.6	0.32±1.4	0.31±1.2	0.31±1.9	0.32±0.9
19	Linoleic acid (C18:2)	43.77±2.2	42.59±1.9	44.99±1.7	41.78±1.1	43.13±1.6	45.34±1.9	42.95±2.3	44.74±1.8	42.50±1.2	44.55±2.3	41.82±2.2	39.75±1.2	39.11±1.1	41.31±0.2	42.77±1.1	41.56±1.4
20	Linolenic acid (C18:3)	0.02±0.1	0.02±0.03	0.02±0.01	0.02±0.03	0.02±0.1	0.00±0.0	0.02±0.4	0.02±0.2	0.03±0.08	0.00±0.0	0.00±0.0	0.00±0.0	0.00±0.0	0.02±0.01	0.02±0.01	0.02±0.03
21	*Cis,cis,cis*-9,12,15-Octadecatrienoic acid (C18:3)	1.24±0.7	1.26±0.5	1.18±0.5	0.91±0.01	1.07±0.02	1.04±0.07	1.11±0.5	1.24±0.9	1.20±1.2	1.17±1.1	1.23±0.6	0.48±1.6	0.66±1.8	0.73±1.2	0.72±1.5	0.59±1.8
22	Arachidic acid	2.37±1.9	2.52±1.7	2.25+1.6	2.94±1.3	2.76±1.5	2.08±0.9	2.76±0.1	2.62±1.8	1.95±2.6	1.93±0.3	1.98±1.7	2.23±1.3	2.17±0.8	2.68±0.2	3.16±0.5	3.01±0.2
23	Eicosenoic acid (C20:1)	0.28±0.4	0.28±0.3	0.28±0.3	0.28±0.2	0.27±0.8	0.29±1.1	0.29±0.6	0.28±0.1	0.26±0.02	0.25±0.9	0.26±0.7	0.40±1.4	0.40±1.8	0.34±0.07	0.34±1.6	0.33±0.01
24	*Cis*-11,14-Eicosadienoic acid (C20:2)	0.03±0.1	0.04±0.01	0.05±0.01	0.04±0.2	0.04±0.8	0.04±0.02	0.04±0.04	0.04±0.09	0.03±0.01	0.04±0.01	0.03±0.05	0.04±0.08	0.04±0.01	0.05±0.06	0.05±0.01	0.05±0.02
25	Heneicosylic acid (C21:0)	0.06±0.01	0.06±0.03	0.05±0.07	0.07±0.01	0.07±0.08	0.05±0.01	0.07±0.01	0.07±0.04	0.04±0.01	0.04±0.03	0.04±0.01	0.06±0.02	0.05±0.09	0.07±0.04	0.09±0.01	0.08±0.03
26	*Cis*-5,8,11,14,17-Eicosapentaenoic acid (C20:5)	0.00±0.0	0.04±0.01	0.04±0.02	0.06±0.01	0.06±0.03	0.04±0.05	0.03±0.01	0.00±0.0	0.02±0.01	0.00±0.0	0.02±0.04	0.11±0.6	0.00±0.0	0.10±0.8	0.08±0.01	0.12±0.03
27	Behenic acid (C22:0)	0.57±0.6	0.62±0.4	0.48±1.3	0.71±1.6	0.66±1.2	0.40±1.8	0.58±0.1	0.63±1.1	0.46±0.8	0.43±0.2	0.47±1.0	0.39±0.7	0.30±0.4	0.41±1.1	0.55±0.3	0.52±0.7
28	Erucic acid (C22:1)	0.02±0.01	0.03±0.02	0.06±0.04	0.10±1.1	0.10±0.1	0.05±0.7	0.06±0.01	0.00±0.0	0.04±0.2	0.02±0.9	0.04±0.07	0.16±1.3	0.03±0.01	0.15±0.2	0.13±0.5	0.18±1.2
29	Tricosanoic acid (C23:0)	0.07±0.5	0.07±0.01	0.06±0.02	0.09±0.02	0.08±0.05	0.07±0.03	0.08±0.01	0.08±0.01	0.06±0.02	0.06±0.07	0.05±0.01	0.07±0.01	0.06±0.03	0.08±0.01	0.10±1.3	0.08±1.4

30	Lignoceric acid (C24:0)	0.19±0.3	0.22±0.4	0.20±1.3	0.23+1.4	0.21±0.4	0.18±0.1	0.20±0.7	0.22±1.3	0.18±1.6	0.17±0.8	0.18±1.1	0.21±1.8	0.19±2.1	0.22±0.07	0.27±1.1	0.23±0.1
31	Nervonic acid (C24:1)	0.06±0.04	0.00±0.0	0.06±0.04	0.00±0.0	0.00±0.0	0.00±0.0	0.06±0.01	0.00±0.0	0.00±0.0	0.00±0.0	0.00±0.0	0.05±0.01	0.00±0.01	0.00±0.0	0.00±0.0	0.00±0.0
32	Docosahexaenoic acid (C22:6)	0.09±0.02	0.00±0.0	0.04±0.07	0.00±0.0	0.00±0.0	0.00±0.0	0.08±0.03	0.00±0.0	0.00±0.0	0.00±0.0	0.00±0.0	0.08±0.5	0.00±0.0	0.08±0.1	0.00±0.0	0.00±0.0

*Fatty acids in percentage of dry matter (% d.m) and n represents number of samples from each country.

3.4 Variations in the fatty acids present in the analysed coffees

The boxplots in Figures 4 - 6 depict the variations in the fatty acids among coffees from the different geographical origins. Although, it was observed that majority of the fatty acids do not show significant difference among the same variety ($p < 0.05$) and little difference when the two bean varieties are compared as previously reported in many studies considering the respective values obtained.

From box-plot 4, it was observed that fatty acids such as linoleic, *Trans*-linolelaidic, margaric, palmitic, *Cis, cis, cis*-9,12,15-octadecatrienoic acid (C18:3), palmitoleic (C18:1) and behenic (C22:0) acids were observed to be higher in Arabica samples, invariably the coffees from African Arabica, South and Central America geographical origins when compared to the Robusta coffee from Asia and Africa Robusta coffees. Moreover, it was observed that linolenic, lignoceric and tricosanoic acids amounts were reduced in the African varieties both Arabica and Robusta compared to coffees from other geographical origins. Furthermore, the quantities in stearic and myristoleic acids indicate little differences in the coffees across all the regions.

In Figure 5 box-plots, most of the fatty acids indicate that the coffees from Africa RB and Asian continents are higher compared to the Arabica coffees from Africa, Central and South America regions. Likewise, from these plots, it was observed that *Cis*-5,8,11,14,17-eicosapentaenoic (C20:5), tetradecanoic, *Z*-10-pentadecenoic, oleic, eicosenoic and erucic acids were higher in the Robusta than Arabica coffee samples. While there was no difference in the levels of lauric, butyric and undecanoic acids from all the different regions.

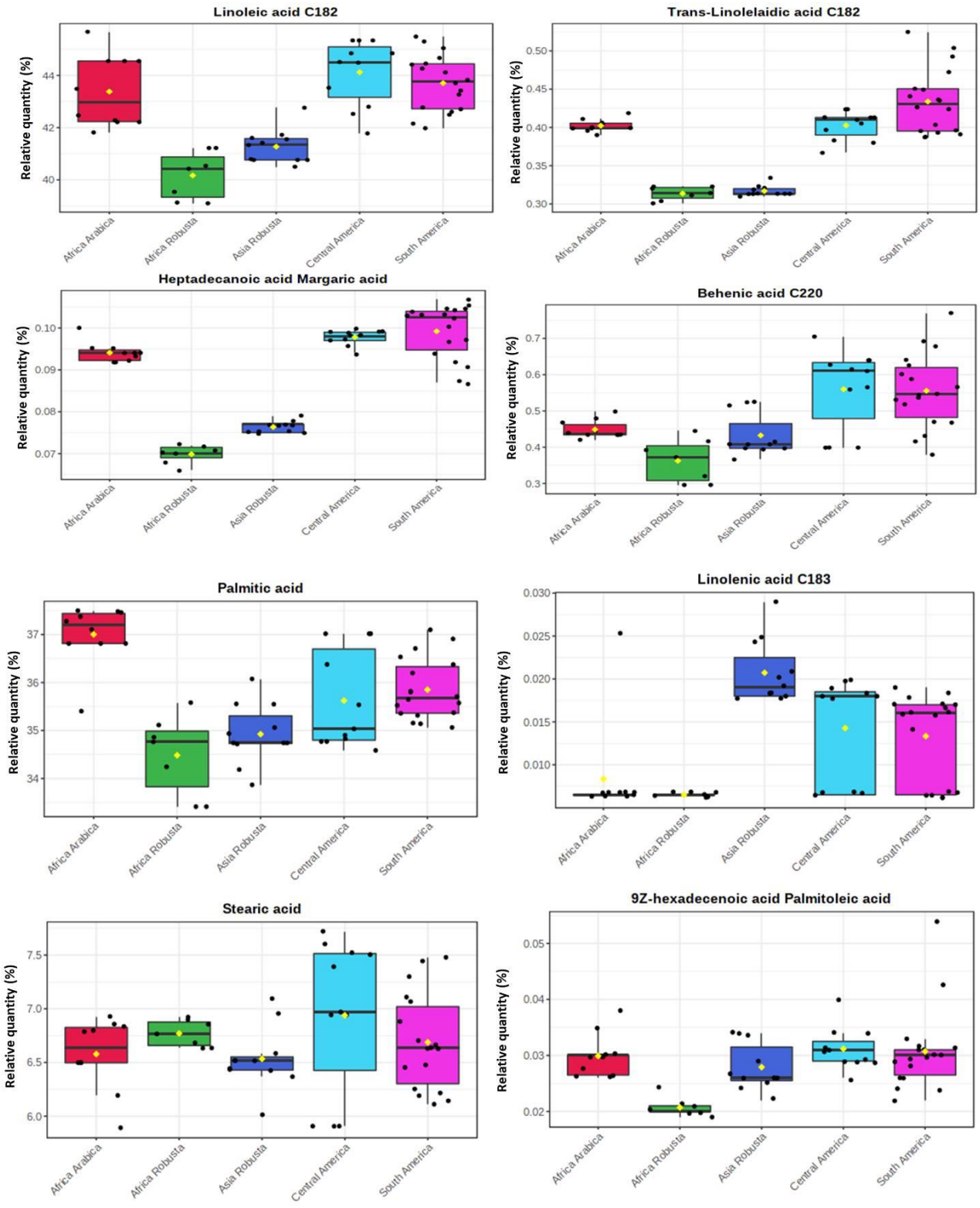


Figure 4 Boxplots of coffee bean fatty acids. Red boxes indicate African Arabica, green for African Robusta, blue for Asian Robusta, light-blue for Central American Arabica and purple represent South American Arabica coffee beans.

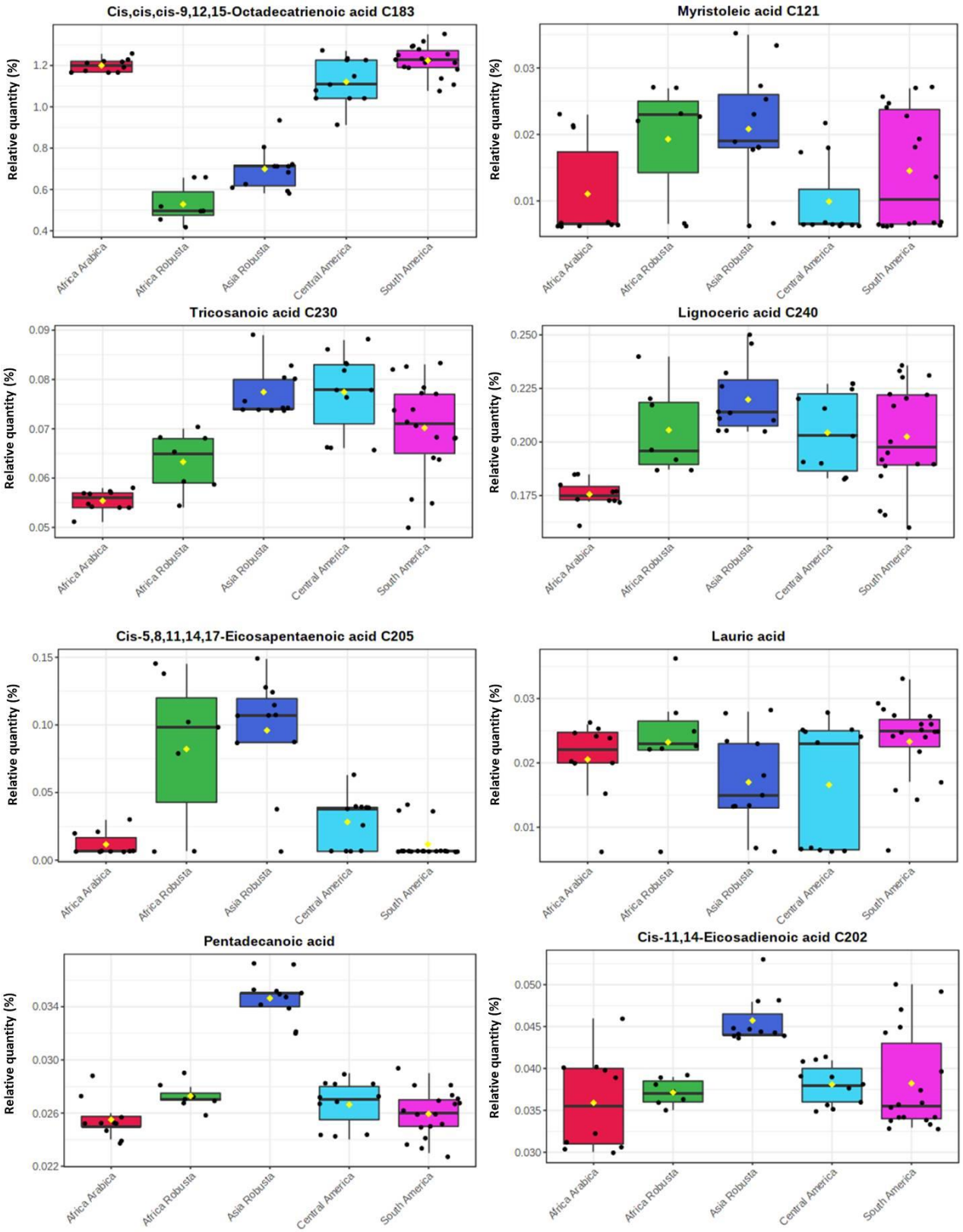


Figure 5 Boxplots of coffee bean fatty acids. Red boxes indicate African Arabica, green for African Robusta, blue for Asian Robusta, light-blue for Central American Arabica and magenta represent South American Arabica coffee beans.

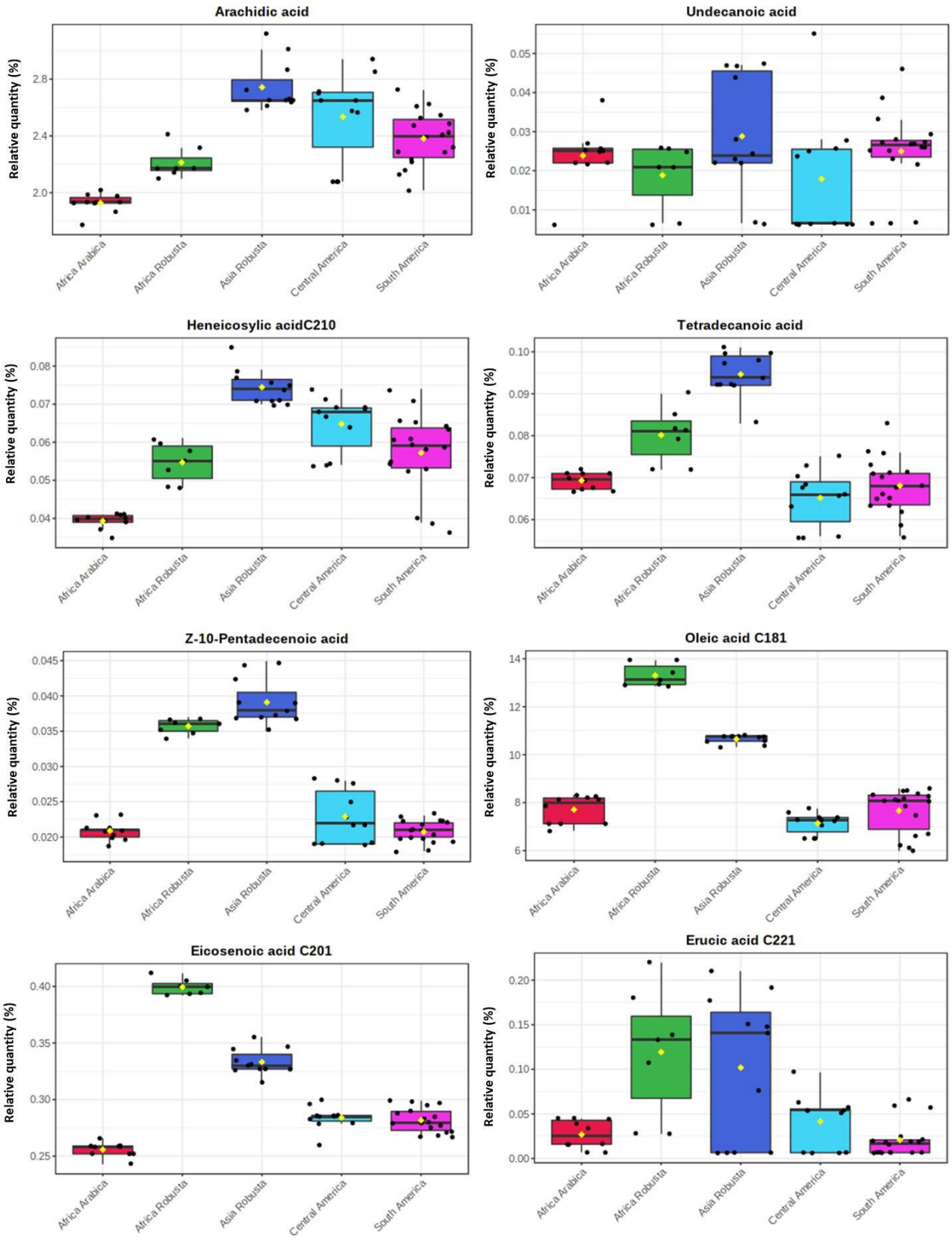


Figure 6 Boxplots of coffee bean fatty acids. Red boxes indicate African Arabica, green for African Robusta, blue for Asian Robusta, light-blue for Central American Arabica and magenta represent South American Arabica coffee beans.

3.5 Multivariate statistical analysis

Moreover, this study went further to explore the use of multivariate statistical analysis tools to investigate potential differences among the fatty acids from the same bean variety, likewise Arabica against Robusta coffee. In addition, the data was applied most importantly by using green bean and not roasted to observe the differences in the unprocessed green beans towards categorising the coffees geographically. Principal component analysis (PCA) and partial least square discriminant analysis (PLS-DA) were further applied to categorise and discriminate the analysed coffee using the fatty acids constituents into different topographical origins. PCA is an unsupervised tool while PLS-DA is a supervised multivariate statistical analysis tool.

Basically, principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables (entities each of which takes on various numerical values) into a set of values of linearly uncorrelated variables called principal components. The PCA has been gaining attractiveness as a tool to bring out strong patterns from complex biological datasets. It captures the essence of the data in a few principal components, which convey the most variation in the dataset.

3.6 Discrimination between Arabica and Robusta

PLS-DA was applied to further understand the differences between the two coffee varieties (Arabica against Robusta). Figure 7 shows that the PLS-DA score plot of Arabica coffee samples (Africa Arabica, Central and South American Arabica) clustered together on the right side of the plot, while the Robusta coffee was observed to have two distinct clusters on the left (Asian Robusta in purple clusters and African Robusta in light-blue clusters). This plot clearly distinguished between the two coffee varieties with 30.8% in components 1 and 16.2% in components 2. With this clear separation, it could be said that the different coffee fatty acids were able to discriminate between the two coffee varieties with the aid of multivariate statistical analytical tool. Most notably and as observed in Figure 4 boxplots, linoleic, *Trans*-linolelaidic,

margaric, palmitic, *Cis, cis, cis*-9,12,15-Octadecatrienoic acid, palmitoleic and behenic acids were observed to be higher in Arabica compared to Robusta coffees. Previous studies have demonstrated that oleic acid based on concentration, has been found to be a suitable discriminant marker for East (Asian) coffees (Mehari et al, 2019).

Likewise, in this study oleic, *cis*-5,8,11,14,17-Eicosapentaenoic, tetradecanoic, *Z*-10-pentadecenoic, eicosenoic and erucic acids were observed to be higher in the Robusta than Arabica coffees. These fatty acids variations could be said to account for the discrete separation conspicuously seen in the PLS-DA score plot in Figure 7. Furthermore, the increased levels of pentadecanoic, *Trans*-elaidic, caproic, decanoic and octanoic acids in Asian Robusta coffees could be said to account for the distinct separation between Robusta coffees of African and Asian origins. In addition, coffees of African origin irrespective of the variety shows low levels in linolenic acids compared to other geographical origins coffees (Figure 4 boxplots) potentially contributed to the African coffees clustering towards the upper part of the plot, while coffees from other regions clustered more towards the middle and downward sides of the score plot.

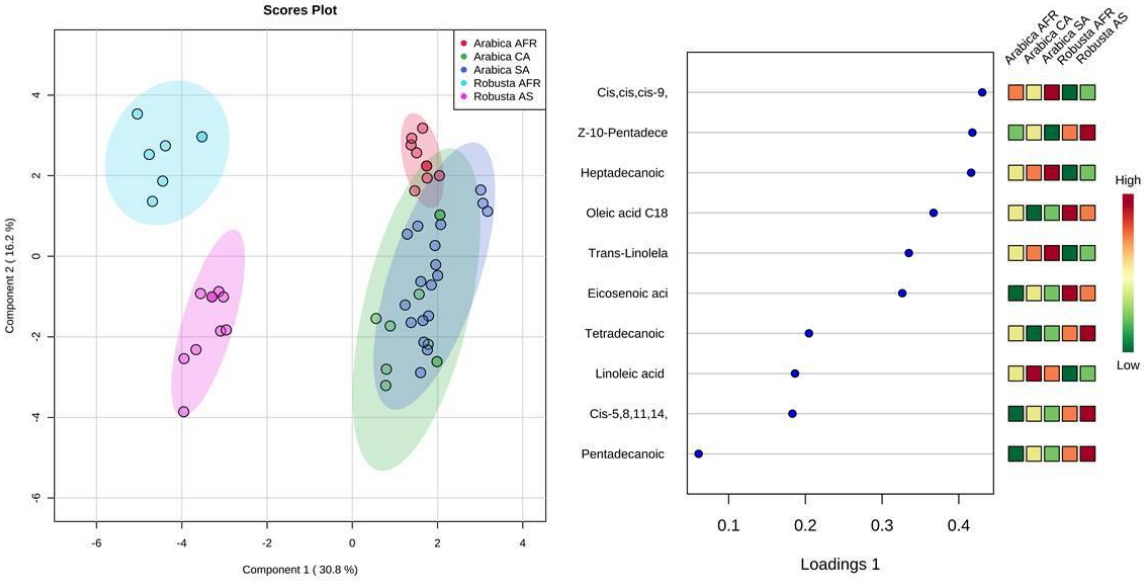


Figure 7 PLS-DA score plot and most important feature (right) of Arabica and Robusta coffees. Score plot of PLS-DA model showing 5 clusters. Each dot represents a coffee sample and the important feature plot shows the most important biomarkers for the distinction.

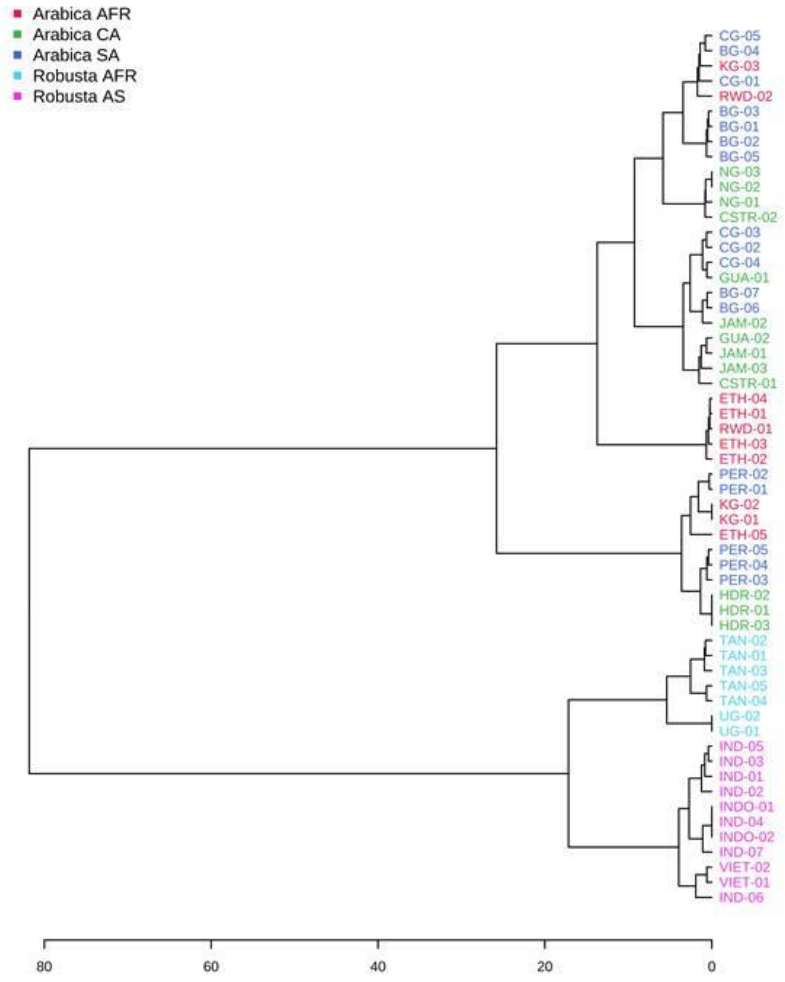


Figure 8 Hierarchical clustering analysis of coffee fatty acid based on bean variety and region. AFR (African), CA (Central American), SA (South American) and AS (Asian).

3.6 PCA and PLS-DA discrimination based on geographical origins

The plot on the left in Figure 9 represents the PCA score of the fatty acids' profiles from the analysed coffee from different topographical origins. The PCA score plot of the first two PCs of a data set of the coffee fatty acid profiles provide a map of how the regions relate to each other. The first component explains 82.6% of the variation, and the second component 10.2%; coloured by coffee fatty acids of the respective geographical origins. It could be observed from the plot that the Arabica coffee samples clustered on the left side of the plot with the big light blue circle showing the Central America, magenta circle represents the South America and the long red cluster for the Africa coffees.

In Figure 9 the plot on the right represents the PLS-DA score plot of the coffee fatty acid profiles from the different geographical regions. The PLS-DA maximizes the co-variances between X (data) and Y (group). The variance displayed in the PLS-DA score plot above is the explained variance for X. It is a supervised statistical tool compared to PCA. It was observed from the plot that the Arabica coffees clustered together on the upper part of plot with the African (red), Central (light-blue) and South (magenta) American coffees. While the Robusta coffees from Africa (green) and Asia (blue) coffees clustered but separately in the lower part of the plot.

Similarly, it was observed that the geographical distribution indicates that the American coffees that are basically Arabica only clustered with the African Arabica coffees, while the Robusta coffee beans from Asia and Africa were distinctly separated as shown in the plot. From the plot, component 1 has 81.2% while component 2 has 10.3%. These plots demonstrated that coffee fatty acids profile in conjunction with multivariate statistical analysis could be used in determining the origin of a particular coffee and likewise distinctly differentiate between the two coffee varieties.

These analytical tools combined will aid the global coffee processing industries in fraud detection and prevention. Figure 10B indicates the loading plot scores of the important compounds used in differentiating the coffees based on origin. Among the top 10 fatty acids are oleic, linoleic, palmitic, arachidic, behenic, stearic, erucic, *Trans*-linoledate and *Cis, cis, cis*-9,12,15-octadecatrienoic acids. Moreover, oleic and erucic acids were observed to be higher in the Robusta coffees compared to the Arabica coffees. While linoleic, palmitic, stearic, behenic, *Trans*-linoledate and octadecatrienoic acids have higher amount in the Arabica coffees compared to the Robusta.

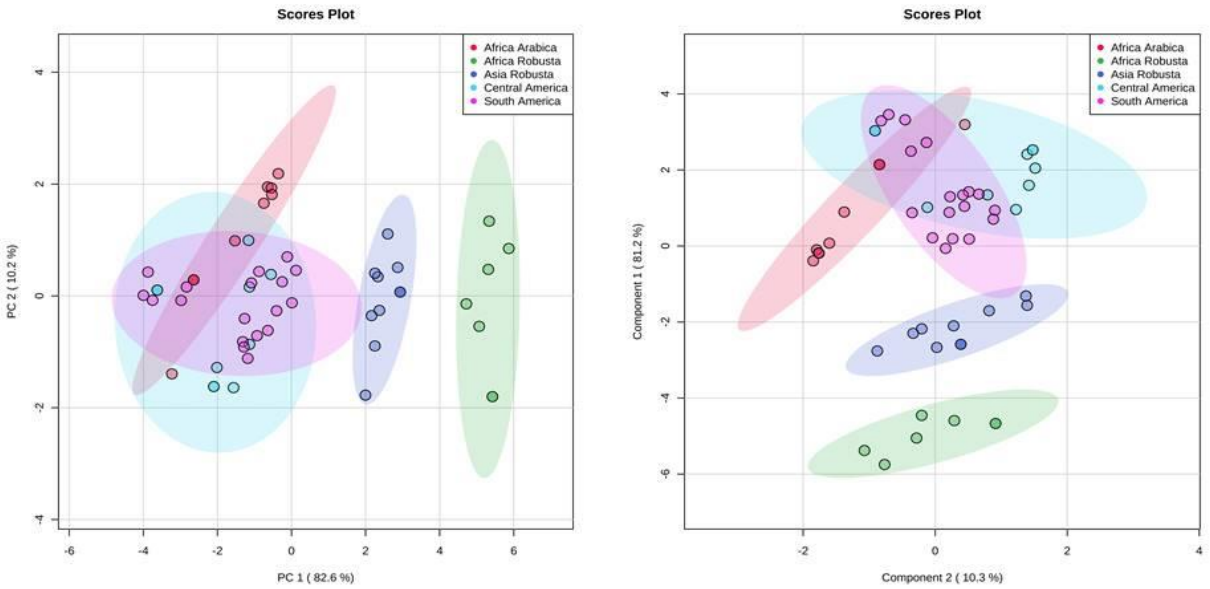


Figure 9 PCA and PLS-DA score plots of Arabica and Robusta green coffees based on origin. Points represent the coffee samples in five different clusters in the PCA and PLS-DA plots.

In addition, the 3-dimensional synchronised plot was used to further demonstrate the differences among the coffees based on their source of origin and bean variety. Figure 10a shows the 3-dimensional plot of the fatty acids profile. The coefficients by which the original variables are multiplied to obtain the PC are called loadings, whose numerical values illustrate the similarities of each variable when compared to that component, the distribution of samples for different origins can be visualized in PCA two or three-dimensional (2D or 3D) plots defined by PCs (Gao et al, 2013).

In Figure 10A, the Robusta coffees from Asia and Africa clustered in the component 2 segment of the plot, with the African coffees clustering separately much to the left region while the Asian coffees were observed to cluster in the same component but on the right. The Arabica coffees were seen to cluster separately on a different component on the right. The American Arabica coffees clustered on the same plain while the African Arabica coffees were observed to cluster on a different plain. The components 1 indicated 81.2%, components 2 has 10.3% and components 3 with 6.5%.

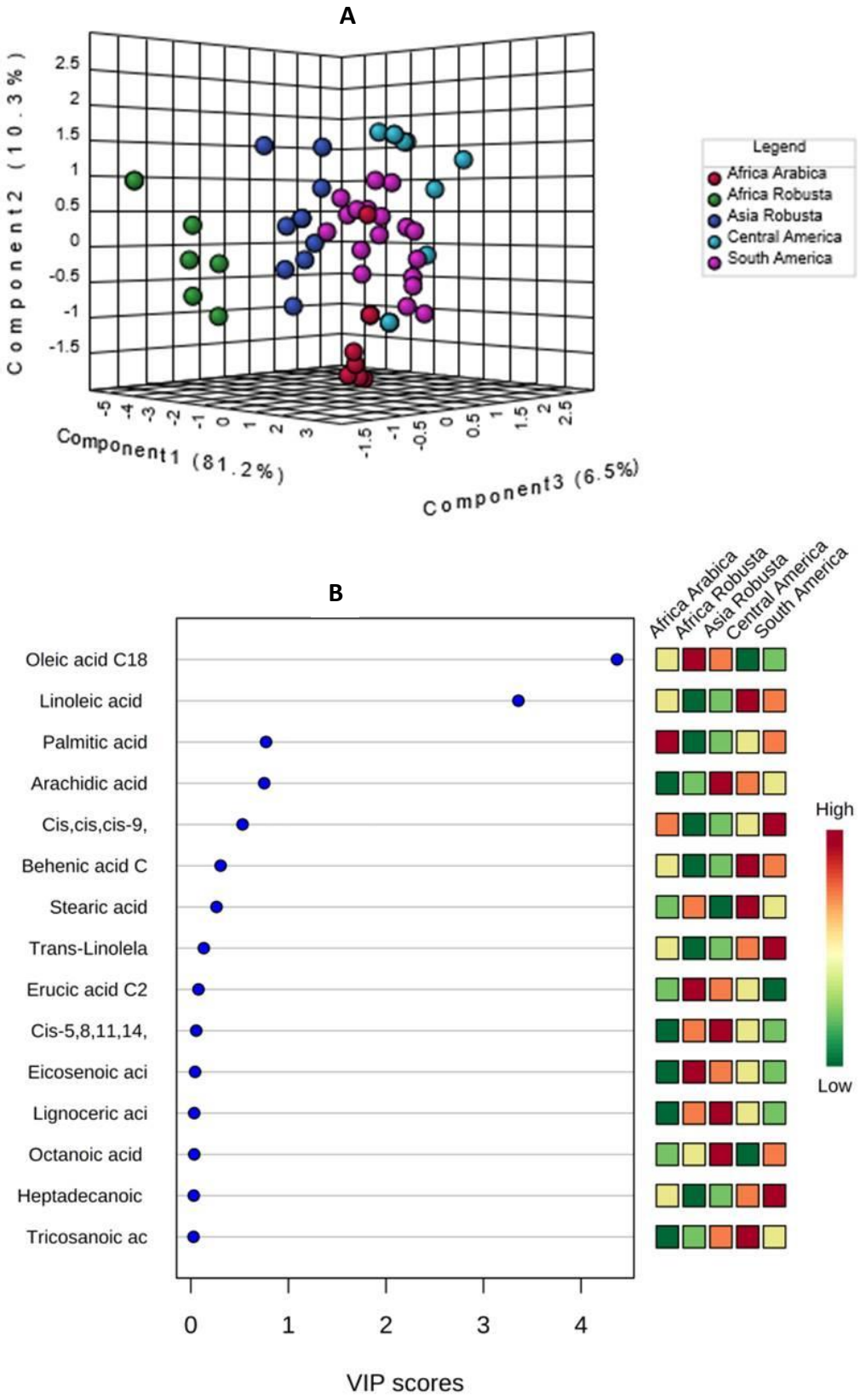


Figure 10 The synchronised 3-dimensional and important feature plots of the analysed coffee. In 3D plot, red dots indicate African Arabica, green (African Robusta), Asian Robusta (blue), Central American Arabica (light-blue) and magenta (South American Arabica) coffees.

Furthermore, the important feature plot in Figure 10 signifies that the coffee fatty acids differentiated the coffees with variations based on their source of origin. For instance, oleic and linoleic acids have strong influence on all the coffees with a VIP score between 3 and 4, while other fatty acids have weak influence on the components irrespective of the coffee origins. In addition, oleic, arachidic, erucic, eicosenoic, lignoceric, tetradecanoic, *Z*-10-pentadecenoic, octanoic and *cis*-5,8,11,14-eicosapentaenoic acids were observed to be very high in Robusta coffee beans irrespective of the origin compared to the Arabica coffees. Likewise, arachidic and octanoic acids quantities were higher in Asian coffees and these two fatty acids could be distinctly be explored to differentiate Asian coffees from other coffees.

Moreover, linoleic, palmitic, behenic, stearic, *trans*-linolelaidic, heptadecanoic and tricosanoic acids were observed to be very high in the Arabica coffees compared to the Robusta coffees. Additionally, linoleic, palmitic, behenic, *trans*-linolelaidic, heptadecanoic and tricosanoic acids were observed to be higher in the South/Central American coffees compared to African and Asian coffees. Most importantly, the South/Central American coffees contains higher amount of the *trans*-linolelaidic acid compared to other regions. This suggests that these fatty acids could be used in differentiating South American coffees from coffee beans originated from other regions. But the African coffees most important the Robusta variety contains more oleic, erucic and eicosanoic acids more than other coffees irrespective of the origin and bean varieties.

Moreover, palmitic acid promotes healthy skin (Fuchs & Green, 1981), cell regulation (Duester, 2008) and maintenance of the immune system (Ertesvag et al, 2002; Mora et al, 2008). This fatty acid is observed to be very high in the Africa and South/Central American Arabica coffees, this suggest that nutritionally the Arabica coffees from these two regions are of more benefits for humans to consume compared to others.

3.7 Statistical significance

The p-values data in Table 4 indicates that there is significant difference ($p < 0.05$) between the different coffees produced from the five geographical coffee origins across the world when the fatty acid constituents are compared.

4. CONCLUSION

In this study, for the first time the possibility of discriminating among African, Asian, Central and South American green coffee beans on the basis of region of origin based on fatty acids analysis was demonstrated. With the aid of chemometrics, oleic, *cis*-5,8,11,14,17-eicosapentaenoic, tetradecanoic, z-10-pentadecenoic, eicosenoic and erucic acids were observed to be higher in the Robusta than Arabica coffees; thus, they were identified as the most discriminating marker compounds among the different beans. In addition, palmitic acid could be used to differentiate Arabica coffees from Africa and South/Central American from other coffees. While South/Central American coffees contains higher amount of the *trans*-linolelaidic acid compared to other regions. By exploiting the fatty acid compositional data in conjunction with different multivariate statistical analysis tools provided a classification model with recognition and prediction abilities of 82.6% and 81.2% respectively at region level. With the varied market prices and typically influenced significantly based on coffees geographical origin, this present study outcomes can be applied to help authenticate and determine the coffee production regions or sources of different coffee beans at the level of continent of origin.

REFERENCES

Albert J. D., Hamilton, R. J., Hamm W. (2008). *Trans* Fatty Acids ISBN: 1405156910, *Wiley-Blackwell Publisher*, Oxford UK.

Alves, M. R., Casal, S., Oliveira, M. B. P. P., Ferreira, M. A. (2003). Contribution of FA profile obtained by high-resolution GC/chemometric techniques to the authenticity of green and roasted coffee varieties. *J. Am. Oil Chem.* Soc. 80, 511–517.

Bajko, E., Kalinowska, M., Borowski, P., Siergiejczyk, L., Lewandowski, W. (2016). 5-*O*-caffeoylquinic acid: a spectroscopic study and biological screening for antimicrobial activity, *LWT-Food Sci. Technol.* 65, 471–479.

Bertrand B., Guyot B., Anthony F., Lashermes P. (2003) Impact of *Coffea canephora* gene introgression on beverage quality of *C. arabica*. *Theor. Appl. Genet*. 107:387-394.

Bertrand B., Etienne H., Cilas C., Charrier A., Baradat P. (2005a) *Coffea arabica* hybrid performance for yield, fertility and bean weight. *Euphytica* 141:255-262.

Bertrand B., Villarreal D., Laffargue A., Posada H., Lashermes P., Dussert S. (2008). Comparison of the effectiveness of fatty acids, chlorogenic acids, and elements for the chemometric discrimination of coffee (*Coffea arabica L.*) varieties and growing origins. *J. Agric. Food Chem*. 56:2273-80.

Buscemi, S., Marventano, S., Antoci, M., Cagnetti, A., Castorina, G., Galvano, F., Mistretta, A. (2016). Coffee and metabolic impairment: An updated review of epidemiological studies. *NFS Journal, 3, 1–7.*

Capuano, E., & Fogliano, V. (2011). Acrylamide and 5-hydroxymethylfurfural (HMF): A review on metabolism, toxicity, occurrence in food and mitigation strategies. *LWT - Food Science and Technology, 44(4), 793–810.*

Charrier A., Berthaud J. (1985) Botanical classification of coffee. In: Clifford MN, Wilson KC (eds), Coffee: botany, biochemistry and productions of beans and beverage, pp.13-47. *Avi Publishing Company*, Westport, Connecticut, USA.

Christie W. W. (1989). Gas chromatography and lipids: A practical guide. First. W.W. Christie ed. *Pergamon Press*, Ayr, Scotland.

Clifford, M. N. (1985). Chemical and physical aspects of green coffee and coffee products. In: Clifford, M. N., Willson, K.C. (Eds.), Coffee: Botany, biochemistry and production of beans and beverage, vol. 13. *Avi Publishing*, Westport, Connecticut, pp. 305–374.

Clifford, M. N., Jaganath, I. B., Ludwig, I. A., & Crozier, A. (2017). Chlorogenic acids and the acyl-quinic acids: discovery, biosynthesis, bioavailability and bioactivity. *Natural Product Reports, 34(12), 1391–1421.*

Cossignani, L., Montesano, D., Simonetti, M. S., & Blasi, F. (2016). Authentication of *Coffea arabica* according to triacylglycerol stereospecific composition. *Journal of Analytical Methods in Chemistry, 2016, 1–7.*

Dong, W., Tan, L., Zhao, J., Hu, R., & Lu, M. (2015). Characterization of fatty acid, amino acid and volatile compound compositions and bioactive components of seven coffee (*Coffea robusta*) cultivars grown in Hainan Province, China. *Molecules*, 20(9), 16687–16708.

D'Souza, R. N., Grimbs, S., Behrends, B., Bernaert, H., Ullrich, M. S., & Kuhnert, N. (2017). Origin-based polyphenolic fingerprinting of *Theobroma cacao* in unfermented and fermented beans. *Food Research International, 99, 550–559.*

Duester, G. (2008). Retinoic acid synthesis and signaling during early organogenesis. *Cell, 134(6), 921–931.*

Dziki, D., Gawlik-Dziki, U., Pecio, L., Różyło, R., Świeca, M., Krzykowski, A., Rudy, S. (2015). Ground green coffee beans as a functional food supplement - preliminary study. *LWT-Food Sci. Technol*. *63*, 691–699.

Ertesvag, A., Engedal, N., Naderi, S., & Blomhoff, H. K. (2002). Retinoic acid stimulates the cell cycle machinery in normal t cells: involvement of retinoic acid receptor-mediated IL-2 secretion. *The Journal of Immunology, 169(10), 5555–5563.*

FAO Report (1996). Rome declaration on world food security (World Food Summit). Rome: Food and Agriculture Organization. *FAO* Accessed 04 April 2020.

Farzaei M. H., Abdollahi Rahimi M. R., (2015). Role of dietary polyphenols in the management of peptic ulcer, *World J. Gastroenterol.* 21 (21) 6499.

Fuchs, E., & Green, H. (1981). Regulation of terminal differentiation of cultured human keratinocytes by vitamin A. *Cell, 25(3), 617–625.*

Gao, B.; Lu, Y.J.; Chen, P.; Yu, L.L. (2013) Differentiating organic and conventional sage by chromatographic and mass spectrometry flow injection fingerprints combined with principal component analysis. *J. Agric. Food Chem*., 61, 2957–2963

Gressler, V., Yokoya, N. S., Fujii, M. T., Colepicolo, P., Filho, J. M., Torres, R. P., Pinto, E. (2010). Lipid, fatty acid, protein, amino acid and ash content in four Brazilian red algae species. *Food Chem*., 120, 585–590.

Hagemann, J. H., Tallent, W. H., Barve, J. A., Ismail, I. A. & Gunstone, F. D. (1975) Polymorphism in single acid triglycerides of positional and geometric isomers of octadecenoic acid. *J Am Oil Chem Soc* 52, 204– 207.

Hurtado-Benavides, A., Dorado A., D., & Sánchez-Camargo, A. del P. (2016). Study of the fatty acid profile and the aroma composition of oil obtained from roasted Colombian coffee beans by supercritical fluid extraction. *The Journal of Supercritical Fluids, 113, 44–52.*

International Coffee Organisation (2019) Annual Review 2019: Addressing the coffee price crises. http://www.ico.org/documents/cy2019-20/annual-review-2018-19-e.pdf Accessed 13 May 2020.

International Coffee Organisation (2020) http://www.ico.org/ trade statistics of April 2020, Accessed on 08 June 2020.

Jham G. N., Berhow M. A., Manthey L. K., Palmquist D. A., Vaughn S. F. (2008a). The use of fatty acid profile as a potential marker for Brazilian coffee (*Coffea arabica L*.) for corn adulteration. *J. Braz. Chem. Soc*. 19:1462-1467.

Joët T., Laffargue A., Descroix F., Doulbeau S., Bertrand B., Kochko A. De, Dussert S. (2010a). Influence of environmental factors, wet processing and their interactions on the biochemical composition of green Arabica coffee beans. *Food Chem*., 118:693-701.

Kroyer, G. T., Kretschmer, L., Washuettl, J. (1989). Antioxidant properties of tea and coffee extracts. Agric. Food Chem. Consum., *Proc. Eur. Conf. Food Chem*., 5th, 2, 433- 437.

Leroy, T., Ribeyre, F., Bertrand, B., Charmetant, P., Dufour, M., Montagnon, C., Pot, D. (2006). Genetics of coffee quality. *Brazilian Journal of Plant Physiology, 18(1), 229–242.*

Liu, C. C., Zhang, Y., Dai, B. L., Ma, Y. -J., Zhang, Q., Wang, Y., & Yang, H. (2017). Chlorogenic acid prevents inflammatory responses in IL-1β-stimulated human SW-1353 chondrocytes, a model for osteoarthritis. *Mol. Medicine Reports, 16(2), 1369–1375.*

Liu, Q. P., Wu, Y. F., Cheng, H. Y., Xia, T., Ding, H., Wang, H., Xu, Y. (2016). Habitual coffee consumption and risk of cognitive decline/dementia: A systematic review and meta-analysis of prospective cohort studies. *Nutrition, 32(6), 628-636.*

Ludwig, I. A., Clifford, M. N., Lean, M. E. J., Ashihara, H., & Crozier, A. (2014). Coffee: biochemistry and potential impact on health. *Food Function, 5(8), 1695–1717.*

Luisa, P. F., Flavio, M. B., Fabiana, C. R., Gerson, S. G., Jose, H. da S. T., & Marcelo, R. M. (2015). Fatty acid profiles and parameters of quality of specialty coffees produced in different Brazilian regions. *African Journal of Agricultural Research, 10(35), 3484–3493.*

Mao, Z. X., Fu, H., Nan, Z. B., Wan, C. G. (2015). Fatty acid, amino acids and mineral composition of four common vetch seeds on Qinghai-Tibetan plateau. *Food Chem* 171:13-18.

Marquetti, I., Link, J. V., Lemes, A. L. G., Scholz, M. B. dos S., Valderrama, P., & Bona, E. (2016). Partial least square with discriminant analysis and near infrared spectroscopy for evaluation of geographic and genotypic origin of arabica coffee. *Computers and Electronics in Agriculture, 121, 313–319.*

Martín, M. J., Pablos, F., González, A. G., Valdenebro, M. S., & León-Camacho, M. (2001). Fatty acid profiles as discriminant parameters for coffee varieties differentiation. *Talanta, 54(2), 291–297.*

Mehari B., Redi-Abshiro M., Chandravanshi B. S., Combrinck S., McCrindle R., Atlabachew M. (2019). GC-MS profiling of fatty acids in green coffee (*Coffea arabica L.*) beans and chemometric modelling for tracing geographical origins from Ethiopia *J Sci Food Agric*.; 99: 3811–3823.

Miller, E. G., Gonzales-sanders, A. P., Couvillon, A. M., Binnie, W. H., Sunahhara, G. I., Bertholet, R. ASIC, 15th Colloque, Montpellier; *ASIC*: Paris, 1993; pp 420-425.

Monente, C., Ludwig, I. A., Irigoyen, A., de Peña, M. P., Cid, C. (2015). Assessment of total (free and bound) phenolic compounds in spent coffee extracts. *J. Agric. Food Chem., 63*, 4327-4334.

Mora, J. R., Iwata, M., & von Andrian, U. H. (2008). Vitamin effects on the immune system: vitamins A and D take centre stage. *Nature Reviews Immunology, 8(9), 685–698.*

Naveed, M., Hejazi, V., Abbas, M., Kamboh, A. A., Khan, G. J., Shumzaid, M., XiaoHui, Z. (2018). Chlorogenic acid (CGA): A pharmacological review and call for further research. *Biomedicine & Pharmacotherapy, 97, 67–74.*

Nikolova-Damyanova, B., Velikova, R., & Jham, G. N. (1998). Lipid classes, fatty acid composition and triacylglycerol molecular species in crude coffee beans harvested in Brazil. *Food Research International, 31(6-7), 479–486.*

Orecchio, S., Ciotti, V. P., & Culotta, L. (2009). Polycyclic aromatic hydrocarbons (PAHs) in coffee brew samples: Analytical method by GC-MS, profile, levels and sources. *Food and Chemical Toxicology, 47(4), 819–826.*

Oosterveld, A., Voragen, A. G. J., Schols, H. A. (2003). Effect of roasting on the carbohydrate composition of *Coffea arabica* beans. *Carbohydrates Polymers* 54, 183–192.

Pimpley, V., Patil, S., Srinivasan, K., Desai, N., & Murthy, P. S. (2020). The chemistry of chlorogenic acid from green coffee and its role in attenuation of obesity and diabetes. *Preparative Biochemistry & Biotechnology, 1–10.*

Romano, R., Santini, A., Grottaglie, L. L., Manzo, N., Visconti, A., Ritieni, A. (2014). Identification markers based on fatty acid composition to differentiate between roasted Arabica and Canephora (Robusta) coffee varieties in mixtures. *J. Food Compo. Anal*., 35:1-9.

Rosenberg, L. (1990) Coffee and tea consumption in relation to the risk of large bowel cancer: a review on epidemiological studies. *Cancer Lett*., 52, 163-171.

Rui Alves, M., Casal, S., Oliveira, M. B. P. P., & Ferreira, M. A. (2003). Contribution of FA profile obtained by high-resolution GC/chemometric techniques to the authenticity of green and roasted coffee varieties. *Journal of the American Oil Chemists So*ciety, 80(6), 511–517.

Santana-Gálvez, J., Cisneros-Zevallos L., Jacobo-Velázquez, D.A. (2017). Chlorogenic acid: recent advances on its dual role as a food additive and a nutraceutical against metabolic syndrome, *Molecules* 22: (3) 358.

Silva E. A., Mazzafera P., Brunini O., Sakai E., Arruda F. B., Mattoso L. H. C., Carvalho C. R. L., Pires R. C. M. (2005) The influence of water management and environmental conditions on

the chemical composition and beverage quality of coffee beans. *Braz. J. Plant Physiol*. 17:229-238.

Shi, H., Dong, L., Jiang, J., Zhao, J., Zhao, G., Dang, X., Jia, M. (2013). Chlorogenic acid reduces liver inflammation and fibrosis through inhibition of toll-like receptor 4 signalling pathway. *Toxicology*, 303, 107–114.

Tango, J.S., (1971). Utilizac¸ ão industrial do café e dos seus subprodutos (in Portuguese), Boletim do Instituto de Tecnologia de Alimentos – *ITAL*, p. 28.

Toci, A. T., Neto, V. J. M. F., Torres, A. G., Farah, A. (2013). Changes in triacylglycerols and free fatty acids composition during storage of roasted coffee. *LWT-Food Sci. Technol*., 50: 581-590.

Van Dam, R. M. (2008). Coffee consumption and risk of type 2 diabetes, cardiovascular diseases, and cancer. *Applied Physiology, Nutrition, and Metabolism, 33(6), 1269–1283.*

Van der Vossen, H. A. M. (1985) Coffee selection and breeding. In: Clifford C and Wilson KC (eds.), Coffee: botany, biochemistry and production of beans and beverage, pp.46-68. *Avi Publishing Company*, Westport, Connecticut, USA.

Veljkovic, E., Xia, W., Phillips, B., Wong, E. T., Ho, J., Oviedo, A., Peitsch, M. (2018). Other compounds from tobacco with potential impact on neurodegenerative diseases. Nicotine and other tobacco compounds in neurodegenerative and psychiatric diseases, *Overview of Epidemiological Data on Smoking and Preclinical and Clinical Data on Nicotine 83–97.*

Villareal, D., Laffargue, A., Posada, H., Bartrand, B. (2009). Genotypic and environmental effects on coffee (*Coffea arabica L*.) bean fatty acid profile: Impact on variety and origin chemometric determination. *J. Agric. Food Chem.*, 57, 11321–11327.

Walton, H., Masterton, G., & Hayes, P. (2013). An epidemiological study of the association of coffee with chronic liver disease. *Scottish Medical Journal, 58(4), 217–222.*

Wan, C.-W., Wong, C. N. Y., Pin, W. K., Wong, M. H. Y., Kwok, C. Y., Chan, R. Y. K., Chan, S. W. (2012). Chlorogenic acid exhibits cholesterol lowering and fatty liver attenuating properties by up-regulating the gene expression of PPAR-α in hypercholesterolemic rats induced with a high-cholesterol diet. *Phytotherapy Research, 27(4), 545–551.*

Wijarnpreecha, K., Thongprayoon, C., & Ungprasert, P. (2017). Coffee consumption and risk of nonalcoholic fatty liver disease. *European Journal of Gastroenterology & Hepatology*, 29(2), e8–e12.

Williamson, G. (2020). Protection against developing type 2 diabetes by coffee consumption: assessment of the role of chlorogenic acid and metabolites on glycaemic responses. *Food & Function.* doi:10.1039/d0fo01168a

Wilson K. M., Kasperzyk J. L., Rider J. R., Kenfield S., van Dam R. M., Stampfer M. J., Giovannucci E., Mucci L. A. (2011). Coffee consumption and prostate cancer risk and progression in the Health Professionals Follow-up Study *J. Natl Cancer Inst*. 103(11):876-84.

Xie, Y., Qin, J., Nan, G., Huang, S., Wang, Z., & Su, Y. (2015). Coffee consumption and the risk of lung cancer: an updated meta-analysis of epidemiological studies. *European Journal of Clinical Nutrition, 70(2), 199–206.*

Yu, X., Bao, Z., Zou, J., & Dong, J. (2011). Coffee consumption and risk of cancers: a meta-analysis of cohort studies. *BMC Cancer, 11(1).*

Zou, Y., Chang, S. K. C., Gu, Y., & Qian, S. Y. (2011). Antioxidant activity and phenolic compositions of lentil *(Lens culinaris var. Morton)* extract and its fractions. *Journal of Agricultural and Food Chemistry,* 59(6), 2268–2276.

Table 4 P-values of coffee cultivated from different geographical origins.

	Pentadeca	Tetradeca	Z-10-Penta	Myristolei	Oleic acid	Eicosenoic	Cis-5,8,11,	Erucic acid	Undecano	Caproic ac	Octanoic a	Decanoic a	Linolenic a	Tricosanoi	Lignoceric	Arachidic a	Heneicosy	Palmitic ac	Heptadeca	Cis,cis,cis-(	Trans-Lino	Linoleic ac	Butyric aci	Nervonic a	Docosahe	Trans-Elai	Cis-11,14-I	Lauric acid	Stearic aci	Behenic ac	Tridecanoi	9Z-hexade
Pentadeca	NA	0	2.22E-15	0.001	0.000	0.001	0.000	0.002	0.017	0.001	0.173	0.040	0.000	0.000	0.000	1.4E-08	1.5E-09	7.6E-05	3.6E-07	1.0E-06	2.2E-07	4.6E-05	9.0E-01	1.5E-01	9.0E-01	0.053	0.000	0.891	0.777	0.535	0.969	0.969
Tetradeca	0	NA	0	0.000	0.000	0.000	0.000	0.000	0.191	0.079	0.941	0.849	0.004	0.013	0.000	1.2E-04	1.5E-05	7.9E-04	1.1E-10	1.3E-08	7.1E-09	3.8E-09	8.6E-01	2.3E-01	8.3E-01	0.393	0.001	0.644	0.596	0.157	0.247	0.064
Z-10-Penta	2.22E-15	0	NA	0.001	0.000	0.000	0.000	0.000	0.898	0.152	0.568	0.989	0.026	0.030	0.001	5.4E-04	1.9E-05	2.5E-07	0.0E+00	0.0E+00	6.7E-16	6.7E-10	9.8E-01	9.4E-01	2.4E-01	0.230	0.001	0.932	0.856	0.001	0.177	0.001
Myristolei	0.001	0.000	0.001	NA	0.001	0.005	0.004	0.018	0.027	0.969	0.915	0.534	0.019	0.029	0.003	6.4E-03	5.2E-03	3.8E-02	3.1E-02	3.0E-03	1.2E-03	1.9E-04	5.7E-02	5.1E-02	3.3E-02	0.249	0.734	0.018	0.413	0.826	0.274	0.118
Oleic acid	0.000	0.000	0.000	0.001	NA	0	1.35E-06	2.93E-05	0.962	0.482	0.630	0.650	0.884	0.993	0.054	2.9E-01	6.1E-02	2.9E-07	0.0E+00	4.4E-16	2.2E-16	0.0E+00	5.1E-01	9.5E-01	3.8E-01	0.217	0.733	0.597	0.700	0.000	0.337	0.000
Eicosenoic	0.001	0.000	0.000	0.005	0	NA	3.74E-09	8.08E-08	0.676	0.800	0.891	0.685	0.986	0.408	0.004	9.9E-02	1.8E-02	3.0E-07	2.2E-16	0.0E+00	3.5E-10	1.9E-10	9.0E-01	7.1E-01	2.8E-01	0.248	0.058	0.926	0.691	0.000	0.449	0.000
Cis-5,8,11,	1.41E-07	6.08E-08	1.10E-10	3.67E-03	1.35E-06	3.7E-09	NA	6.66E-16	0.163	0.007	0.030	0.031	0.068	0.038	0.001	1.7E-03	1.4E-04	1.7E-02	6.7E-10	0.0E+00	2.6E-09	1.8E-09	8.2E-01	9.3E-01	6.3E-01	0.004	0.018	0.668	0.761	0.031	0.715	0.151
Erucic acid	1.81E-03	1.43E-04	8.48E-07	1.77E-02	2.93E-05	8.1E-08	6.66E-16	NA	0.833	0.592	0.323	0.659	0.205	0.102	0.002	2.1E-02	5.0E-03	1.1E-01	2.6E-07	2.4E-11	3.4E-06	7.6E-07	5.2E-01	8.6E-01	4.6E-01	0.384	0.194	0.921	0.801	0.081	0.420	0.020
Undecano	1.68E-02	1.91E-01	8.98E-01	2.68E-02	9.62E-01	6.8E-01	0.163	0.833	NA	3.47E-08	1.33E-08	7.98E-09	4.30E-01	8.85E-02	1.00E-01	3.8E-02	2.2E-01	3.8E-01	5.9E-01	7.1E-01	7.5E-01	3.5E-02	3.4E-02	2.8E-03	1.1E-03	0.022	0.790	0.001	0.221	0.055	0.000	0.000
Caproic ac	5.98E-04	7.94E-02	1.52E-01	9.69E-01	4.82E-01	8.0E-01	0.007	0.592	3.47E-08	NA	9.31E-08	1.68E-13	1.13E-01	8.59E-02	4.11E-01	1.2E-02	3.3E-02	7.2E-02	1.4E-01	6.7E-02	5.2E-02	6.1E-02	5.0E-01	4.5E-01	4.1E-01	0.000	0.298	0.642	0.514	0.795	0.001	0.000
Octanoic a	1.73E-01	9.41E-01	5.68E-01	9.15E-01	6.30E-01	8.9E-01	0.030	0.323	1.33E-08	9.31E-08	NA	8.88E-16	6.09E-02	3.09E-02	1.00E-01	6.2E-03	5.8E-02	2.8E-01	8.8E-01	2.3E-01	6.7E-01	1.1E-01	5.7E-01	4.2E-01	3.8E-01	0.000	0.925	0.830	0.189	0.050	0.000	0.000
Decanoic a	3.96E-02	8.49E-01	9.89E-01	5.34E-01	6.50E-01	6.9E-01	0.031	0.659	7.98E-09	1.68E-13	8.88E-16	NA	0.105	0.044	0.296	1.4E-02	4.8E-02	1.8E-01	7.4E-01	2.1E-01	2.4E-01	3.4E-01	7.2E-01	3.6E-01	3.2E-01	0.000	0.485	0.557	0.475	0.197	0.000	0.000
Linolenic a	1.52E-05	3.56E-03	2.57E-02	1.87E-02	8.84E-01	9.9E-01	0.068	0.205	4.30E-01	1.13E-01	0.061	0.105	NA	6.02E-09	6.61E-05	1.7E-10	1.0E-09	1.0E-02	8.0E-01	4.5E-01	1.0E-01	5.1E-01	2.4E-01	5.5E-01	2.8E-01	0.123	0.042	0.708	0.063	0.019	0.952	0.562
Tricosanoi	1.61E-05	1.34E-02	3.01E-02	2.92E-02	9.93E-01	4.1E-01	0.038	0.102	8.85E-02	8.59E-02	0.031	0.044	6.02E-09	NA	1.78E-15	0.0E+00	0.0E+00	4.2E-05	9.2E-01	2.7E-01	9.5E-02	4.1E-01	8.0E-01	2.2E-01	5.2E-01	0.773	0.332	0.880	0.000	0.000	0.349	0.092
Lignoceric	1.78E-06	2.13E-05	8.74E-04	2.73E-03	5.44E-02	3.8E-03	0.001	0.002	1.00E-01	4.11E-01	0.100	0.296	6.61E-05	1.78E-15	NA	2.28E-13	2.71E-13	1.29E-04	3.5E-02	7.0E-03	2.7E-02	6.0E-03	6.4E-01	2.5E-02	1.6E-01	0.711	0.080	0.710	0.017	0.000	0.617	0.452
Arachidic a	1.42E-08	1.19E-04	5.37E-04	6.37E-03	2.93E-01	9.9E-02	0.002	0.021	3.84E-02	1.23E-02	0.006	0.014	1.69E-10	0	2.28E-13	NA	0	1.13E-05	0.232	0.023	0.008	0.032	0.677	0.736	0.994	0.355	0.108	0.623	0.000	0.000	0.248	0.148
Heneicosy	1.45E-09	1.47E-05	1.85E-05	5.16E-03	6.10E-02	1.8E-02	0.000	0.005	2.18E-01	3.32E-02	0.058	0.048	1.02E-09	0	2.71E-13	0	NA	3.34E-07	0.067	0.004	0.000	0.013	0.558	0.521	0.689	0.203	0.115	0.814	0.002	0.005	0.762	0.722
Palmitic ac	7.61E-05	7.90E-04	2.48E-07	3.81E-02	2.90E-07	3.0E-07	0.017	0.112	3.80E-01	7.17E-02	0.277	0.183	0.010	0.000	0.000	0.000	3.34E-07	NA	0.000	0.000	0.000	0.002	0.824	0.669	0.129	0.023	0.210	0.733	0.019	0.895	0.616	0.142
Heptadeca	3.61E-07	1.12E-10	0.00E+00	3.14E-02	0.00E+00	2.2E-16	0.000	0.000	5.92E-01	1.40E-01	0.876	0.738	0.801	0.917	0.035	0.232	0.067	0.000	NA	0	1.98E-12	3.64E-13	0.430	0.389	0.848	0.210	0.026	0.631	0.472	0.000	0.920	0.008
Cis,cis,cis-(	1.01E-06	1.31E-08	0.00E+00	3.02E-03	4.44E-16	0.0E+00	0.000	0.000	7.07E-01	6.73E-02	0.234	0.206	0.450	0.269	0.007	0.023	0.004	0.000	0	NA	1.78E-15	7.58E-12	0.786	0.987	0.609	0.058	0.019	0.645	0.516	0.000	0.831	0.002
Trans-Lino	2.19E-07	7.14E-09	6.66E-16	1.17E-03	2.22E-16	3.5E-10	0.000	0.000	7.49E-01	5.16E-02	0.669	0.237	0.103	0.095	0.027	0.008	0.000	0.000	1.98E-12	1.78E-15	NA	4.38E-12	0.994	0.878	0.283	0.028	0.668	0.702	0.332	0.001	0.502	0.003
Linoleic ac	4.63E-05	3.78E-09	6.75E-10	1.85E-04	0.00E+00	1.9E-10	0.000	0.000	3.55E-02	6.06E-02	0.115	0.336	0.513	0.409	0.006	0.032	0.013	0.002	3.64E-13	7.58E-12	4.38E-12	NA	0.373	0.660	0.874	0.106	0.393	0.118	0.033	0.070	0.806	0.088
Butyric aci	9.03E-01	8.62E-01	9.82E-01	5.73E-02	5.06E-01	9.0E-01	0.817	0.519	3.37E-02	5.00E-01	0.565	0.724	0.240	0.797	0.643	0.677	0.558	0.824	4.30E-01	7.86E-01	0.994	0.373	NA	0.001	1.05E-09	0.014	0.077	0.599	0.989	0.808	0.792	0.557
Nervonic a	1.46E-01	2.34E-01	9.45E-01	5.10E-02	9.47E-01	7.1E-01	0.930	0.863	2.81E-03	4.50E-01	0.416	0.365	0.547	0.215	0.025	0.736	0.521	0.669	3.89E-01	9.87E-01	0.878	0.660	0.001	NA	0	0.065	0.961	0.127	0.448	0.437	0.518	0.035
Docosahe	9.00E-01	8.30E-01	2.43E-01	3.32E-02	3.84E-01	2.8E-01	0.632	0.458	1.05E-03	4.11E-01	0.376	0.324	0.276	0.520	0.160	0.994	0.689	0.129	8.48E-01	6.09E-01	0.283	0.874	0.000	0	NA	0.003	0.524	0.427	0.364	0.272	0.482	0.025
Trans-Elai	5.32E-02	3.93E-01	2.30E-01	2.49E-01	2.17E-01	2.5E-01	0.004	0.384	2.20E-02	3.18E-06	0.000	0.000	0.123	0.773	0.711	0.355	0.203	0.023	2.10E-01	5.85E-02	0.028	0.106	0.014	0.065	0.003	NA	0.229	0.344	0.624	0.424	0.005	0.242
Cis-11,14-I	1.85E-05	6.25E-04	6.75E-04	7.34E-01	7.33E-01	5.8E-02	0.018	0.194	7.90E-01	2.98E-01	0.925	0.485	0.042	0.332	0.080	0.108	0.115	0.210	2.61E-02	1.87E-02	0.668	0.393	0.077	0.961	0.524	0.229	NA	0.038	9.08E-06	1.74E-02	6.07E-01	4.14E-01
Lauric acid	8.91E-01	6.44E-01	9.32E-01	1.77E-02	5.97E-01	9.3E-01	0.668	0.921	1.41E-03	6.42E-01	0.830	0.557	0.708	0.880	0.710	0.623	0.814	0.733	6.31E-01	6.45E-01	0.702	0.118	0.599	0.127	0.427	0.344	0.038	NA	2.32E-03	6.73E-02	6.73E-01	7.29E-01
Stearic aci	7.77E-01	5.96E-01	8.56E-01	4.13E-01	7.00E-01	6.9E-01	0.761	0.801	2.21E-01	5.14E-01	0.189	0.475	0.063	0.000	0.017	0.000	0.002	0.019	4.72E-01	5.16E-01	0.332	0.033	0.989	0.448	0.364	0.624	0.000	0.002	NA	1.08E-06	5.57E-01	3.96E-01
Behenic ac	5.35E-01	1.57E-01	1.03E-03	8.26E-01	4.46E-05	1.1E-04	0.031	0.081	5.46E-02	7.95E-01	0.050	0.197	0.019	0.000	0.000	0.000	0.005	0.895	3.65E-06	3.27E-05	0.001	0.070	0.808	0.437	0.272	0.424	0.017	0.067	1.08E-06	NA	0.099	0.000
Tridecanoi	9.69E-01	2.47E-01	1.77E-01	2.74E-01	3.37E-01	4.5E-01	0.715	0.420	6.09E-06	1.23E-03	0.000	0.000	0.952	0.349	0.617	0.248	0.762	0.616	9.20E-01	8.31E-01	0.502	0.806	0.792	0.518	0.482	0.005	0.607	0.673	0.557	0.099	NA	6.73E-10
9Z-hexade	9.69E-01	6.39E-02	1.44E-03	1.18E-01	1.21E-04	1.1E-04	0.151	0.020	6.44E-06	1.02E-04	0.000	0.000	0.562	0.092	0.452	0.148	0.722	0.142	8.27E-03	1.91E-03	0.003	0.088	0.557	0.035	0.025	0.242	0.414	0.729	0.396	0.000	6.73E-10	NA

Table 5 Pearson correlation coefficient values (r = 1) of coffee cultivated from different geographical origins.

	Pentadeca	Tetradeca	Z-10-Penta	Myristolei	Oleic acid	Eicosenoic	Cis-5,8,11,	Erucic acid	Undecano	Caproic ac	Octanoic a	Decanoic a	Linolenic a	Tricosanoi	Lignoceric	Arachidic a	Heneicosy	Palmitic ac	Heptadeca	Cis,cis,cis-9	Trans-Lino	Linoleic ac	Butyric aci	Nervonic a	Docosahe	Trans-Elai	Cis-11,14-I	Lauric acid	Stearic aci	Behenic ac	Tridecanoi	9Z-hexade
Pentadeca	1	0.88	0.82	0.42	0.46	0.42	0.63	0.40	0.31	0.44	0.18	0.27	0.53	0.53	0.58	0.66	0.69	-0.50	-0.61	-0.59	-0.62	-0.51	-0.02	-0.19	-0.02	0.26	0.53	-0.02	0.04	-0.08	0.01	0.01
Tetradeca	0.88	1	0.88	0.50	0.63	0.55	0.64	0.48	0.17	0.23	0.01	0.03	0.38	0.32	0.53	0.48	0.54	-0.43	-0.73	-0.66	-0.67	-0.68	-0.02	-0.16	0.03	0.11	0.44	0.06	0.07	-0.19	-0.15	-0.24
Z-10-Penta	0.82	0.88	1	0.43	0.81	0.78	0.73	0.59	0.02	0.19	-0.08	0.00	0.29	0.28	0.43	0.44	0.53	-0.62	-0.87	-0.85	-0.83	-0.70	0.00	-0.01	0.16	0.16	0.43	0.01	0.02	-0.42	-0.18	-0.41
Myristolei	0.42	0.50	0.43	1	0.41	0.37	0.38	0.31	0.29	0.01	-0.01	-0.08	0.31	0.29	0.39	0.35	0.36	-0.27	-0.28	-0.38	-0.42	-0.47	-0.25	-0.26	-0.28	-0.15	-0.05	0.31	0.11	0.03	-0.15	-0.21
Oleic acid	0.46	0.63	0.81	0.41	1	0.90	0.59	0.52	0.01	0.09	-0.06	-0.06	-0.02	0.00	0.25	0.14	0.25	-0.61	-0.88	-0.83	-0.84	-0.85	-0.09	-0.01	0.12	0.16	0.05	0.07	0.05	-0.51	-0.13	-0.48
Eicosenoic	0.42	0.55	0.78	0.37	0.90	1	0.68	0.64	-0.06	0.03	-0.02	-0.05	0.00	0.11	0.37	0.22	0.31	-0.61	-0.84	-0.89	-0.71	-0.72	-0.02	0.05	0.14	0.15	0.25	0.01	-0.05	-0.49	-0.10	-0.49
Cis-5,8,11,	0.63	0.64	0.73	0.38	0.59	0.68	1	0.83	0.19	0.35	0.29	0.28	0.24	0.27	0.42	0.40	0.48	-0.31	-0.70	-0.85	-0.69	-0.69	0.03	0.01	0.06	0.37	0.31	-0.06	-0.04	-0.28	0.05	-0.19
Erucic acid	0.40	0.48	0.59	0.31	0.52	0.64	0.83	1	-0.03	-0.07	0.13	0.06	0.17	0.22	0.39	0.30	0.36	-0.21	-0.62	-0.74	-0.57	-0.60	0.09	0.02	0.10	0.12	0.17	0.01	0.03	-0.23	0.11	-0.30
Undecano	0.31	0.17	0.02	0.29	0.01	-0.06	0.19	-0.03	1	0.65	0.66	0.67	0.11	0.23	0.22	0.27	0.16	-0.12	-0.07	-0.05	-0.04	-0.28	-0.28	-0.39	-0.42	0.30	-0.04	0.41	0.16	0.25	0.56	0.55
Caproic ac	0.44	0.23	0.19	0.01	0.09	0.03	0.35	-0.07	0.65	1	0.63	0.79	0.21	0.23	0.11	0.33	0.28	-0.24	-0.20	-0.24	-0.26	-0.25	-0.09	-0.10	-0.11	0.57	0.14	-0.06	0.09	0.03	0.41	0.49
Octanoic a	0.18	0.01	-0.08	-0.01	-0.06	-0.02	0.29	0.13	0.66	0.63	1	0.83	0.25	0.28	0.22	0.36	0.25	-0.15	-0.02	-0.16	-0.06	-0.21	0.08	-0.11	-0.12	0.58	0.01	-0.03	0.17	0.26	0.61	0.50
Decanoic a	0.27	0.03	0.00	-0.08	-0.06	-0.05	0.28	0.06	0.67	0.79	0.83	1	0.22	0.27	0.14	0.32	0.26	-0.18	-0.04	-0.17	-0.16	-0.13	0.05	-0.12	-0.13	0.63	0.09	-0.08	0.10	0.17	0.61	0.50
Linolenic a	0.53	0.38	0.29	0.31	-0.02	0.00	0.24	0.17	0.11	0.21	0.25	0.22	1	0.68	0.50	0.72	0.70	-0.34	0.03	-0.10	-0.22	-0.09	0.16	0.08	0.15	0.20	0.27	-0.05	0.25	0.31	0.01	0.08
Tricosanoi	0.53	0.32	0.28	0.29	0.00	0.11	0.27	0.22	0.23	0.23	0.28	0.27	0.68	1	0.82	0.90	0.92	-0.51	-0.01	-0.15	-0.22	-0.11	-0.03	-0.17	-0.09	0.04	0.13	0.02	0.46	0.59	0.13	0.22
Lignoceric	0.58	0.53	0.43	0.39	0.25	0.37	0.42	0.39	0.22	0.11	0.22	0.14	0.50	0.82	1	0.79	0.79	-0.48	-0.28	-0.35	-0.29	-0.36	-0.06	-0.29	-0.19	-0.05	0.23	0.05	0.31	0.45	0.07	0.10
Arachidic a	0.66	0.48	0.44	0.35	0.14	0.22	0.40	0.30	0.27	0.33	0.36	0.32	0.72	0.90	0.79	1	0.93	-0.54	-0.16	-0.30	-0.34	-0.28	0.06	-0.05	0.00	0.12	0.21	0.07	0.49	0.45	0.15	0.19
Heneicosy	0.69	0.54	0.53	0.36	0.25	0.31	0.48	0.36	0.16	0.28	0.25	0.26	0.70	0.92	0.79	0.93	1	-0.61	-0.24	-0.38	-0.48	-0.32	0.08	-0.09	0.05	0.17	0.21	-0.03	0.40	0.37	0.04	0.05
Palmitic ac	-0.50	-0.43	-0.62	-0.27	-0.61	-0.61	-0.31	-0.21	-0.12	-0.24	-0.15	-0.18	-0.34	-0.51	-0.48	-0.54	-0.61	1	0.48	0.48	0.57	0.39	-0.03	-0.06	-0.20	-0.30	-0.17	-0.05	-0.31	-0.02	-0.07	0.20
Heptadeca	-0.61	-0.73	-0.87	-0.28	-0.88	-0.84	-0.70	-0.62	-0.07	-0.20	-0.02	-0.04	0.03	-0.01	-0.28	-0.16	-0.24	0.48	1	0.88	0.77	0.78	0.11	0.12	-0.03	-0.17	-0.29	0.06	0.10	0.57	0.01	0.34
Cis,cis,cis-9	-0.59	-0.66	-0.85	-0.38	-0.83	-0.89	-0.85	-0.74	-0.05	-0.24	-0.16	-0.17	-0.10	-0.15	-0.35	-0.30	-0.38	0.48	0.88	1	0.83	0.75	0.04	0.00	-0.07	-0.25	-0.31	0.06	0.09	0.52	0.03	0.40
Trans-Lino	-0.62	-0.67	-0.83	-0.42	-0.84	-0.71	-0.69	-0.57	-0.04	-0.26	-0.06	-0.16	-0.22	-0.22	-0.29	-0.34	-0.48	0.57	0.77	0.83	1	0.76	0.00	0.02	-0.14	-0.29	-0.06	0.05	-0.13	0.42	0.09	0.39
Linoleic ac	-0.51	-0.68	-0.70	-0.47	-0.85	-0.72	-0.69	-0.60	-0.28	-0.25	-0.21	-0.13	-0.09	-0.11	-0.36	-0.28	-0.32	0.39	0.78	0.75	0.76	1	0.12	0.06	-0.02	-0.21	0.11	-0.21	-0.28	0.24	-0.03	0.23
Butyric aci	-0.02	-0.02	0.00	-0.25	-0.09	-0.02	0.03	0.09	-0.28	-0.09	0.08	0.05	0.16	-0.03	-0.06	0.06	0.08	-0.03	0.11	0.04	0.00	0.12	1	0.44	0.70	0.32	0.23	-0.07	0.00	-0.03	0.04	-0.08
Nervonic a	-0.19	-0.16	-0.01	-0.26	-0.01	0.05	0.01	0.02	-0.39	-0.10	-0.11	-0.12	0.08	-0.17	-0.29	-0.05	-0.09	-0.06	0.12	0.00	0.02	0.06	0.44	1	0.84	0.24	-0.01	0.20	0.10	-0.10	-0.09	-0.28
Docosahe	-0.02	0.03	0.16	-0.28	0.12	0.14	0.06	0.10	-0.42	-0.11	-0.12	-0.13	0.15	-0.09	-0.19	0.00	0.05	-0.20	-0.03	-0.07	-0.14	-0.02	0.70	0.84	1	0.38	0.09	0.11	0.12	-0.15	-0.09	-0.29
Trans-Elai	0.26	0.11	0.16	-0.15	0.16	0.15	0.37	0.12	0.30	0.57	0.58	0.63	0.20	0.04	-0.05	0.12	0.17	-0.30	-0.17	-0.25	-0.29	-0.21	0.32	0.24	0.38	1	0.16	-0.13	-0.07	-0.11	0.36	0.16
Cis-11,14-I	0.53	0.44	0.43	-0.05	0.05	0.25	0.31	0.17	-0.04	0.14	0.01	0.09	0.27	0.13	0.23	0.21	0.21	-0.17	-0.29	-0.31	-0.06	0.11	0.23	-0.01	0.09	0.16	1	-0.27	-0.55	-0.31	-0.07	-0.11
Lauric acid	-0.02	0.06	0.01	0.31	0.07	0.01	-0.06	0.01	0.41	-0.06	-0.03	-0.08	-0.05	0.02	0.05	0.07	-0.03	-0.05	0.06	0.06	0.05	-0.21	-0.07	0.20	0.11	-0.13	-0.27	1	0.39	0.24	0.06	-0.05
Stearic aci	0.04	0.07	0.02	0.11	0.05	-0.05	-0.04	0.03	0.16	0.09	0.17	0.10	0.25	0.46	0.31	0.49	0.40	-0.31	0.10	0.09	-0.13	-0.28	0.00	0.10	0.12	-0.07	-0.55	0.39	1	0.59	0.08	0.11
Behenic ac	-0.08	-0.19	-0.42	0.03	-0.51	-0.49	-0.28	-0.23	0.25	0.03	0.26	0.17	0.31	0.59	0.45	0.45	0.37	-0.02	0.57	0.52	0.42	0.24	-0.03	-0.10	-0.15	-0.11	-0.31	0.24	0.59	1	0.22	0.49
Tridecanoi	0.01	-0.15	-0.18	-0.15	-0.13	-0.10	0.05	0.11	0.56	0.41	0.61	0.61	0.01	0.13	0.07	0.15	0.04	-0.07	0.01	0.03	0.09	-0.03	0.04	-0.09	-0.09	0.36	-0.07	0.06	0.08	0.22	1	0.70
9Z-hexade	0.01	-0.24	-0.41	-0.21	-0.48	-0.49	-0.19	-0.30	0.55	0.49	0.50	0.50	0.08	0.22	0.10	0.19	0.05	0.20	0.34	0.40	0.39	0.23	-0.08	-0.28	-0.29	0.16	-0.11	-0.05	0.11	0.49	0.70	1

Chapter 6

Investigation of changes in the polyphenols and lipid profiles of coffee beans stored for over a decade

1. INTRODUCTION

Coffee being among world topmost consumed beverages is predominantly stimulated by its pleasant aroma, flavour and taste, exciting sensations and physiological effects (Vignoli et al, 2014). Similarly, coffee has extensively been reported to possess numerous health benefits (Santana-Gálvez et al, 2017; Naveed et al, 2018; Venditti et al, 2015; Meng et al, 2013) apart from the stimulating features that the consumers derived from it. *Coffea arabica* and *Coffea canephora* are the most commonly cultivated coffee species in the diverse topographical regions of the world, while Arabica is the most preferred due to its superior sensory properties.

The beverage is abundantly rich in numerous classes of bioactive compounds, most importantly polyphenols, lipids, sugars, amino acids, caffeine, trigonelline, other chemical compounds with their respective metabolites, which are generated during processing such as roasting. Most significantly, coffee is a rich source of phenolic compounds, particularly the chlorogenic acids (CGAs) and their degradation products like caffeic, ferulic and coumaric acids. In addition, it is the major source of CGAs in human diet and has been revealed as an efficient antioxidant (Xu, Hu & Liu, 2012). Essentially, these compounds contribute to the total polyphenolic dietary intake and are beneficial to consumer's health (Benigno et al, 2018). While triacylglycerols and free fatty acids are the main lipids constituents (Nikolova-Damyanova et al, 1998), these compounds in different chemical forms contribute immensely to the drink's unique taste, flavour and aroma.

The coffee bean profile and bioactive compounds contents are largely influenced by several factors such as method of cultivation, harvesting, drying and/or fermentation, farm settlement storage mechanisms and types, transportation (shipping) systems and the bean variety. It is known that the chemical composition of green coffee beans depends on the genotypes and geographic area of origin, as well as the cultivation practices, maturation and post-harvest conditions, particularly storage (Smith, 1989).

Furthermore, due to the logistics involved in coffee agriculture, harvesting, storage, shipping, processing and warehousing before it is eventually made available to consumers; coffee beans usually travel round the world in shipping containers passing through different climatic zones for a long period of time. These chains of events do influence the coffee chemical constituents, the organoleptic characteristics and thus affects the coffee cup quality.

Additionally, in some cases the green beans are stored for an average of 1 - 2 years or even more before it is eventually shipped to processing locations due to fluctuations in harvesting periods, farm settlements localization, distance to the main port cities and transportation logistics. Besides, the bean long shelf life at the processing warehouse also influences the bean quality thus the coffee cup physicochemical properties are affected.

Presently, Europe and United States of America do not cultivate coffee largely due to the soil topography and climatic conditions required for coffee cultivation; but the two continents are the world largest consumers of coffee. Secondly, it usually takes months before the green coffee arrives the coffee processing factories in these two regions from the point of harvesting through warehousing, shipping and delivery, for the bean's complete transformation into cup drink. These factors significantly play some key roles on the bean ageing processes.

Likewise, since the green bean usually travels through different planet's hydrosphere (Pacific, Atlantic, Indian, Southern and Arctic oceans) depending on the origin of cultivation and final destination, this factor contributes to the coffee quality in a way. Similarly, the differences in bean variety, chemical components, percentages and ratios of bioactive compounds are also affected during farm storage after harvesting and processing factories warehouse storage with the different storage conditions before the final processes are carried out on the coffee bean.

However, naturally the progressive decline in the beans chemical constituents' physiological status usually leads to ageing. This mostly occurs during storage, when ground and roasted

coffee loses its aroma and fresh flavour as a result of lipid oxidation and degradation of some of the aroma compounds (Clarke, 1986), and thus the coffee quality decreases (Vila et al, 2005). Coffee lipid degradation can take place via two different mechanisms simultaneously: acylglycerol hydrolysis and oxidation. Lipolysis releases free fatty acids (FFA), which are more prone to oxidation than esterified fatty acids, particularly long-chain unsaturated FA (Belitz et al, 2004).

Król investigated the content of polyphenols in coffee beans considering roasting degrees, origin and storage effect for about a year and reported insignificant decrease in the phenolic content of the stored coffees (Król et al, 2020). However, this study mainly focused on the phenolic content which invariably cannot absolutely be used to conclude on the effects of storage on coffee considering the different classes of bioactive constituents present in coffee. In view of this, there is insufficient information concerning the effects of coffee storage for a long time, how it affects the chemical constituents and its stability on the shelf.

Hence this study aims to investigate the effects of different storage conditions on the phenolic components of coffee including the lipids' constituents exploring different storage conditions for coffees stored for about a decade. This will reveal more information on the stability of the different chemical compounds present in the bean and how this consequently influences the coffee quality over time. In addition, this study plans to examine the impact of different storage conditions (temperatures) over time compared to room temperature, and rate of degradation of the main compounds in the green and roasted coffees.

2. MATERIALS AND METHODS

2.1 Coffee samples

A total of 61 samples were investigated in this study, 51 were green coffee beans while ten (10) were roasted beans. The coffee beans were produced from different regions of the world. From the total number of samples, 41 green coffee beans were supplied to the Laboratory by Deutscher Kaffeeverband, Hamburg while the rest were purchased from local groceries stores in Bremen. Comprehensive details of the coffee beans are provided in Table 1 of the supplementary information of Chapter two. The coffees samples were mainly *Coffea arabica* (40) and *Coffea canephora* (17). The studied roasted coffee beans were roasted each in coffee roaster at 230 ± 5 ºC/11 min and then milled to 60 Tyler mesh prior to analysis. The presence of coffee oil biomarker 16-*O*-methylcafestol was explored in confirming each of the bean sample botanical variety with the aid of nuclear magnetic resonance (NMR) as previously described (Speer & Kölling-Speer, 2006; Schievano et al, 2014; Badmos et al, 2019).

The coffee beans (10 g each) were ground to fine-powder. For each sample, 200 mg of the powdered bean were weighed and extracted with 70% methanol by sonication (with magnetic stirrer) for 15 min and vortex for another 15 min. The extract was filtered through a Chromafil PTFE membrane filter (Bremen, Germany) and used for LC-MS^n sequence run.

2.2 Chemicals and reagents

Details as described in section 2.2 of Chapter three.

2.3 Preparation of salt clusters (calibration solution)

Details as described in section 2.4 of Chapter four.

2.4 LC-ESI-TOF-MS (High resolution mass spectrometry) and quantification for CGAs

Details as described in section 2.3 of Chapter three.

2.5 LC-ESI-MS^n (Quadrupole ion-trap mass spectrometry) for CGAs

Details as described in section 2.4 of Chapter three.

2.6 Coffee oil extraction

Details as described in section 2.3 of Chapter four.

2.6.1 Chromatographic separation of triacylglycerol

Details as described in section 2.5 of Chapter four.

2.6.2 LC-ESI-TOF-MS (High resolution mass spectrometry) for TAGs

Details as described in section 2.5 of Chapter four.

2.6.3 LC-ESI-MS^n (Quadrupole Ion-trap mass spectrometry) for TAGs

Details as described in section 2.6 of Chapter four.

2.7 Derivatization of Fatty acid by Fatty acids methyl ester (FAME)

Details as described in section 2.5 of Chapter five.

2.7.1 Gas chromatography

Details as described in section 2.6 of Chapter five.

2.8 Spectrophotometric estimation of total phenolic content (FRAP assay)

The antioxidant assay was carried out using extraction of 200 mg of ground coffee in 5 mL of 70% methanol the extraction buffer (EB), sonicated for 15 min and vortex for another 15 min. The extract was filtered using PTFE membrane filter and diluted by adding 4 mL of EB to 1 mL of the filtered extract. Gallic acid (1.2 mg dissolved in 2 mL of EB) and Trolox (2.2 mg dissolved in 2 mL of EB) were prepared in serial dilution and used as standards. The amount of antioxidant present in each sample was determined using the gallic acid equivalent (GAE) estimation. The standards were purchased from Sigma-Aldrich, Germany. The FRAP reagent

was prepared mixing Iron (III) chloride ($FeCl_3$) 27.3 mg dissolved in 5 mL of H_2O, TPTZ (2,4,6-Tris(2-pyridyl)-s-triazine) and 15.93 mg dissolved in 5 mL of 40 mM of HCl). Acetate Buffer was prepared using 500 mL each of 0.5 N of acetic acid and 0.5 N sodium acetate. Then 185 mL acetic acid and 15 mL sodium acetate were combined and diluted to 400 mL with a confirmed pH of 3.6.

The FRAP reagent solution was freshly prepared by mixing 10 mL acetate buffer, 1 mL of TPTZ solution and 1 mL of $FeCl_3$ solution together. The resulting solution was mixed thoroughly and used for the assay. The extracted samples were pipetted (10 µL) with the standards as unknowns into a 96-well plate, 200 µL of FRAP reagent solution was added to each sample well and gently shake for 30 sec. The plate was incubated in the dark for 10 min at 37 °C and absorbance at 593 nm was measured using a spectrophotometric microplate reader.

2.9 Statistical analysis

Analysis of variance (ANOVA) was used to study the changes in contents of coffees during storage (Arabica against Robusta). Data processing and analysis were carried out using Python for p-values, ANOVA significance, Pearson correlation coefficient and heatmap. All statistical tests were conducted at the 95% confidence level ($p < 0.05$). Principal component analysis (PCA) and partial least square-discriminant analysis (PLS-DA) of the coffee bean were carried out on the CGAs, TAGs and fatty acids profiles using *MetaboAnalyst 4.0* for Comprehensive and Integrative Metabolomics Data Analysis (Chong, Wishart, and Xia, 2019). Lipid-MAPS Lipidomics Gateway tool was used for the TAGs identification (Fahy et al, 2009).

3. RESULTS AND DISCUSSION

The main objective of this study is to investigate the ageing process and degree of degree of the phenolics and lipids constituents of green and roasted coffee stored with different conditions over a period of time. In addition, the differences in the stored Arabica and Robusta coffee beans would be examined and compared. Before now, Krol et al (2020) investigated the polyphenol contents of roasted coffee beans stored for a year. However, the study outcome is not sufficient to conclude on the effects of storage conditions on coffee beans. In view of this, this study intends to comprehensively investigate the two main coffee bean varieties stored with different storage conditions for about a decade. Multivariate statistical analysis would be applied to the study outcomes to determine the different variations and similarities present in the studied samples.

Green (51) and roasted (10) coffee beans produced from different geographical locations across the world were used in this study. The green coffee beans were produced from the year 2009 to 2016, while the roasted coffee beans were produced from 1990 to 2016. The first study group will comprises of 47 (out of 51) green coffee beans and the 10 roasted coffee beans, which were stored in the Lab at room temperature. From this first group, the CGAs, antioxidant capacity, TAGs and fatty acids contents were profile annually for four (4) years starting from 2016 till 2020.

Moreover, the second study group consists of 10 green coffee beans, which were stored with different gaseous phase conditions both at room temperature (RT) and 4 ºC for a period of two (2) years. Out of this, 5 were Arabica and 5 were Robusta green coffee beans. The samples were not ground until when needed for analysis. The first category of samples (Arabica and Robusta) in this group were stored with carbon dioxide (CO_2) differently at RT and 4 ºC with one control group (without CO_2 at RT). The second group was stored with Nitrogen (N_2) only differently

at RT and 4 ºC. The third was stored with N_2 & H_2O differently at RT and 4 ºC. While the fourth group was stored with atmospheric air differently at RT and 4 ºC. Furthermore, the CGA constituents of this groups were profiled to examine the effect of the different gaseous phase and storage conditions on the beans over a period of two (2) years.

Moreover, the chlorogenic acids, antioxidant estimation, triacylglycerols and fatty acids constituents of the investigated coffee samples were analysed. LC-ESI-TOF-MS (High resolution mass spectrometry) was employed in identifying and quantifying the respective chlorogenic acids and triacylglycerols present in the coffee samples. LC-ESI-MS^n (Quadrupole ion-trap mass spectrometry) was used to determine the fragment ions (MS^2 and MS^3) data for both the CGA and TAG. FRAP antioxidant assay was used in estimating the total phenolic contents of the coffee. While the fatty acids constituents of the lipid fraction were derivatized by FAME and analysed on GC to determine the individual fatty acids and relative quantities using peak areas compared to the standards of the methylated fatty acids analysed.

3.1. Confirmation of bean botany Arabica versus Robusta by ^{1}H-NMR

To establish the bean variety and validate the LC-MS analysis results, 16-*O*-methylcafestol (16-*O*Me) content of coffee bean samples was analysed. Instead of using the LC-MS method designed by the International Organization for Standardization on coffee standard analysis (Clifford et al, 2006) to determine the 16-*O*Me coffee oil content exclusively present in Robusta coffee, a less laborious and more convenient ^{1}H-NMR method was employed in this regard (Schievano et al, 2014) as described in section 3.1 of Chapter 2.

The NMR results show that 16-*O*Me is present in all the coffee bean samples labelled as Robusta while the Arabica coffee bean do not possess this biomarker. NMR spectra of the two different coffee varieties are presented in the chapter two appendix section.

3.2. LC-MS quantification of chlorogenic acids and respective regio-isomers

Eight-point calibration curves were acquired using both HPLC-UV/VIS quantification at 320 nm for selected CGAs available as authentic reference standards and the calibration curves for LC-MS quantification in the negative ion mode was also obtained. The analytical parameters including LOD (limit of detection), LOQ (limit of quantification), RRF (relative response factor) were determined for all derivatives as presented in Table 5 of Chapter 2 of this thesis. Likewise, all the full quantitative absolute CGA values were provided in the supplementary information, while data are presented in the main text as box-plots and bar charts providing a quick graphical overview of the CGA distributions in the extensive data set; followed by multivariate statistical analysis.

3.3 CGA analysis

In this study, the coffee extracts were analysed using HPLC method as previously reported (Jaiswal et al, 2010; Badmos et al, 2019). With the aid of this method, chromatograms were acquired using a high resolution QTOF-ESI mass detector. The coffee extract produced a typical chromatogram, in which chlorogenic acids previously reported in green coffee beans were readily identified according to their *m/z* ratio and fragmentation pattern (Clifford et al, 2003; Clifford & Jarvis, 1998; Clifford et al, 2005; Clifford et al, 2006; Wei et al, 2012; Clifford, Knight, Surucu & Kuhnert, 2006). All the CGAs shown in Table 1 could be identified according to their high-resolution mass data in the negative ion mode and using tandem LC-MS^n with a mass error under 5 ppm for the correct fragmentation patterns, confirming the molecular formulas of the CGAs as previously assigned and all compounds assigned according to regio-isomeric level. The latest recommendations for CGAs nomenclature was adopted (Badmos, Lee & Kuhnert, 2019; Clifford, Jaganath, Ludwig & Crozier, 2017; Abranko & Clifford, 2017).

3.3.1 Mono-acyl and di-acyl CGA in green coffee beans

The CGA constituents most importantly 5-CQA, which is an antioxidant decreases with time during storage, due to possible involvement of radical scavenging actions, for instance lipid and protein oxidation (Rendon et al, 2014). In view of this, the mono-acyl and di-acyl CGA constituents of the green coffee bean stored from 2016 through 2020 were analysed to investigate the rate of annual degradation of the respective compounds. Figure 1.5 of Chapter 1 of this thesis represented the chemical structures of the CGAs extracted from the coffee bean samples. For this study, 47 different Arabica (33) and Robusta (14) green coffee bean samples produced from the year 2009 through 2020 were investigated. However, the analysis started in the year 2016 and the samples were later analysed once a year from 2017, 2018, 2019 and 2020 respectively. The average mean, standard deviation, p-values ($p < 0.05$) and Pearson correlation coefficient scores ($r = 1$) of the CGAs for two different bean varieties were computed and the data interpretation by analysis of variance (ANOVA) as presented in Table 1. Figure 1 represents the rate of degradation profile of mono-CQA (3-CQA, 4-CQA and 5-CQA) extracted from the Arabica and Robusta green beans stored over time measured in mg/g.

The CQA, di-CQA, and FQA represent about 98% of total CGAs in coffee (Clifford & Staniforth, 1977). Regarding CQAs, it has been observed and reported that 5-CQA is the most abundant compound present both in Arabica and Robusta green coffees, representing approximately 77% of total CQA composition. Furthermore, di-CQA were the second most abundant of the class with the total di-CQA corresponding to about 12 - 20% of total CGAs, and followed by the FQAs representing about 4 - 8% of the total CGA composition (Clifford & Staniforth, 1977; Bicho et al, 2013; Badmos et al, 2019). In view of this, CQA, Di-CQA and FQA were the CGAs profile used in this study for the green coffee beans analysis.

For the mono-CQA, 5-CQA was observed to be the most abundant CGA in the green coffee bean for all the samples analysed, followed by 4-CQA and 3-CQA (5-CQA > 4-CQA > 3-CQA). Furthermore, the 5-CQA of the Robusta green coffee beans (mean±sd; 44.14±7.3) was observed to be higher compared to Arabica beans 5-CQA (37.01±3.5), and similar observations was made in all the analysed coffee as previously reported on coffee CGA profiling and analysis (Bicho et al, 2013; Badmos et al, 2019). Equally, the year 2016 CGAs profile analysis results show that Robusta values were higher than Arabica beans in the caffeoylquinic, feruloylquinic and dicaffeoylquinic acids as presented in Table 1. Similar observations were recorded for the years 2017, 2018, 2019 and 2020. The data presented in Table 1 depicts the quantitative values of all the analysed CGAs in the two coffee varieties from the year 2016 to 2020 in chronological order.

Similar trend was observed in the FQAs, with 5-FQA (8.81±1.5), 3-FQA (1.77±2.6) higher than the 4- isomer (1.10±0.5) in the Robusta coffee (5-FQA > 3-FQA > 4-FQA) when compared with the CQAs; except for the Arabica FQAs where 4-FQA was observed to be higher than the 3-FQA, with 5-FQA (3.27±3.5) > 4-FQA (0.52±0.1) > 3-FQA (0.26±0.2).

However, among the DiCQA the 3,5- isomer (2.51±0.6) was higher than 3,4-DiCQA (2.18±0.5) and 4,5- (2.25±0.5) isomers (3,5-DiCQA > 4,5-DiCQA > 3,4-DiCQA) in Arabica beans, but in Robusta the pattern was different with 3,4- isomer (5.89±1.1) being higher than 3,5- (5.06±1.5) and 4,5- (4.65±1.2) isomers (3,4-DiCQA > 3,5-DiCQA > 4,5-DiCQA), this as well corresponds to previous studies (Bicho et al 2013; Badmos et al, 2019).

Likewise, the p-values and Pearson correlation coefficient (r) scores were computed using one-way ANOVA as described in Chapters 2 and 3. The p-values ($p < 0.05$) show that there is a significant difference with regards to difference between year 2016 and 2017, 2017 and 2018, 2018 and 2019, likewise 2019 and 2020 respectively in all the Arabica green coffee beans. Similar trend was observed in the Robusta samples except for the 4-CQA (0.17) and 3-FQA

(0.23) which show no level of significance ($p < 0.05$). In addition, there is a significant difference between the rate degradation of the three FQAs in Arabica and Robusta coffees when compared. The compound variables were found to be strongly correlated, and in some it was moderately correlated in both coffees between years 2016 and 2020 as shown in the Table, though some were observed to be negatively correlated notably in 3-FQA and 4-FQA for both coffees, while negative correlation was observed only in Robusta Di-CQA isomers.

3.3.2 Rate of degradation of mono-acyl and di-acyl CGA in green coffee beans

The CGA data in Table 2 shows the percentage rate of degradation (%) of the CGAs extracted and commuted from the green coffee beans from year 2016 to 2020. The percentage values show that virtually all the key CGAs under consideration in this study were depleted gradually from the year 2016 to 2020 at different rates. The 'old' represents the first year (2016) of analysis, while 'new' (2020) stands for the last year of analysis. Equation below was used in calculating the degradation rate of the corresponding compounds in the stored green coffee beans over time.

(i) $$\textbf{Percentage change (\%)} = \frac{\textbf{New - Old}}{\textbf{Old}} \times 100$$

(ii) $$\textbf{Percentage change (\%)} = \frac{\textbf{2020 - 2016}}{\textbf{2016}} \times 100$$

The highest percentage degradation rate was observed in 5-CQA compared to other compounds. Also, Robusta (87.76%) shows higher percentage degradation rate when compared to Arabica (72.3%) samples between the year 2016 and 2020. Furthermore, 3-CQA quantities were observed to be more depleted in two coffee varieties with slight increase in Arabica (65.61%) compared to Robusta (62.25%). Among the mono-CQA, 4-CQA has the lowest percentage degradation rate with higher rate observed in Arabica (48.10%) compared to Robusta (46.44%). Besides, among the Di-CQA, the 3,5- isomer was observed to be more depleted compared to

3,4- and 4,5- isomers in both coffee varieties. The Robusta 3,5- isomer (64.62%) was more depleted compared to Arabica (59.36%). Also, 4,5- isomer in Robusta has higher degradation rate (21.51%) compared to Arabica (17.78%), while contrast observation was seen in 3,4- isomer with higher rate in Arabica (57.34%) compared to Robusta (52.12%).

The 3-FQA was more depleted in both coffee varieties, followed by 5-FQA and 4-FQA in the Robusta coffee. But the trend was different in Arabica (3-FQA > 4-FQA > 5-FQA).

From the percentage rate of degradation data, it could be concluded that there are variations with regards to the rate of degradation of the different class of CGAs and also the 3 positions hydrolyse fastest (Deshpande et al, 2014) when bean varieties CGA constituents are compared. However, most importantly the 5-CQA was more observed to be more depleted than others, followed by 3-CQA and 3,5-DiCQA, 3,4-DiCQA, 4-CQA and 3-FQA respectively (5-CQA > 3-CQA > 3,5-DiCQA > 3,4-DiCQA > 4-CQA > 3-FQA). In addition, Figures 1 and 2 represent the boxplots of the varied quantities of the mono- and diacyl-CQAs in the two coffee varieties respectively from year 2016 through 2020.

Table 1 Mean+s.d, p-values (< 0.05) and Pearson correlation coefficient (r = 1) of the CGA extracted from green coffee bean from 2016 to 2020.

Year	Mean±SD (mg/g)	3-CQA	4-CQA	5-CQA	3-FQA	4-FQA	5-FQA	3,4-DiCQA	3,5-DiCQA	4,5-DiCQA
2016	Arabica (n = 33)	4.71±1.7	6.32±1.5	37.01±3.5	0.26±0.2	0.52±0.1	3.27±0.5	2.18±0.5	2.51±0.6	2.25±0.5
	Robusta (n = 14)	5.59±2.6	7.58±3.3	44.14±7.3	1.77±2.6	1.10±0.5	8.81±1.5	5.89±1.1	5.06±1.5	4.65±1.2
2017	Arabica (n = 33)	3.42±0.9	4.78±1.7	27.78±3.8	0.02±0.0	0.03±0.0	0.25±0.1	1.46±0.4	2.18±0.4	1.64±0.6
	Robusta (n = 14)	3.55±0.4	10.27±5.6	32.15±4.6	0.10±0.1	0.14±0.2	1.66±2.3	4.05±0.8	4.00±1.5	3.04±0.9
2018	Arabica (n = 33)	2.76±0.7	3.16±0.4	21.92±2.2	0.41±0.1	0.98±0.3	6.91±1.8	2.19±0.6	2.39±0.8	4.39±1.0
	Robusta (n = 14)	2.66±0.5	9.20±5.6	24.78±2.1	1.28±0.6	1.32±1.1	3.02±0.9	7.09±0.4	3.95±1.4	8.68±1.8
2019	Arabica (n = 33)	2.12±0.3	4.67±1.1	13.63±2.8	0.25±0.1	0.57±0.2	4.14±1.1	1.36±0.5	1.49±0.5	2.68±0.7
	Robusta (n = 14)	3.36±1.5	6.57±2.5	16.65±2.2	0.70±0.5	1.62±1.0	8.28±1.4	4.43±1.2	2.75±0.5	5.79±1.0
2020	Arabica (n = 33)	1.62±0.3	3.28±0.5	10.25±1.1	0.16±0.1	0.38±0.1	2.82±0.5	0.93±0.3	1.02±0.3	1.85±0.4
	Robusta (n = 14)	2.11±0.7	4.06±1.3	11.66±2.8	0.44±0.3	1.05±0.7	5.77±1.3	2.82±0.8	1.79±0.4	3.64±0.8
(p < 0.05)	Arabica	0.004	0.02	0.0001	0.0002	0.0001	0.0001	0.002	0.001	0.0002
	Robusta	0.01	**0.17**	0.0001	**0.23**	0.049	0.0001	0.0001	0.002	0.0001
Pearson Corr. Coeff.	**(2016 – 2020)**	**3-CQA**	**4-CQA**	**5-CQA**	**3-FQA**	**4-FQA**	**5-FQA**	**3,4-DiCQA**	**3,5-DiCQA**	**4,5-DiCQA**
(r = 1)	Arabica	0.01	0.10	0.31	-0.61	-0.22	0.15	0.11	0.84	0.61
	Robusta	0.11	0.09	0.67	-0.17	-0.49	0.98	-0.82	-0.32	-0.92

Table 2 Percentage rate of degradation (%) of CGA compounds extracted from the green coffee bean from 2016 to 2020.

% Change (2016 - 2020)	3-CQA	4-CQA	5-CQA	3-FQA	4-FQA	5-FQA	3,4-DiCQA	3,5-DiCQA	4,5-DiCQA
Arabica	65.61	48.10	72.30	38.46	26.92	13.76	57.34	59.36	17.78
Robusta	62.25	46.44	87.76	75.14	4.55	34.51	52.12	64.62	21.51

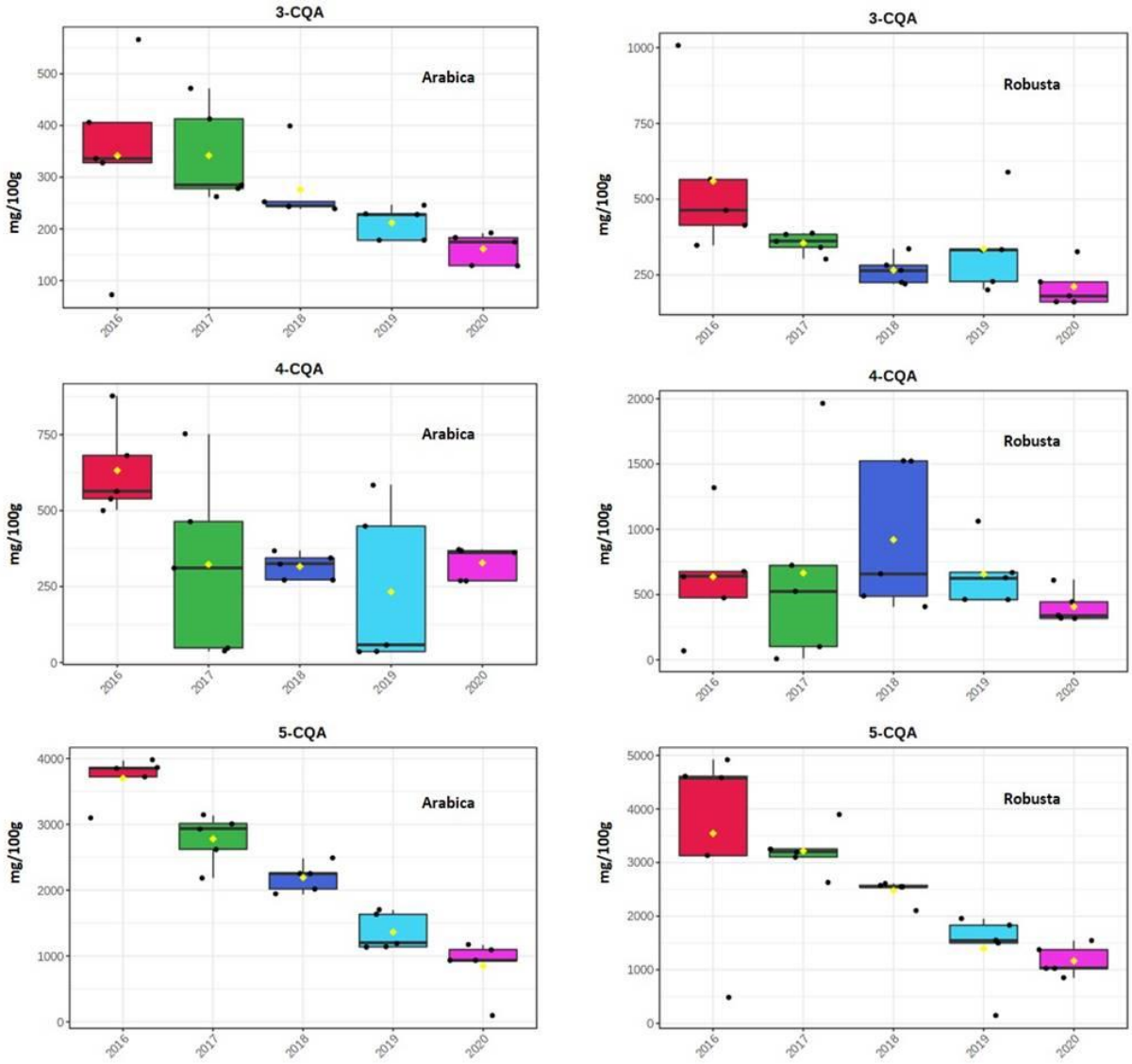


Figure 1 Boxplots of mono-CQA in Arabica and Robusta green coffee from 2016 to 2020.

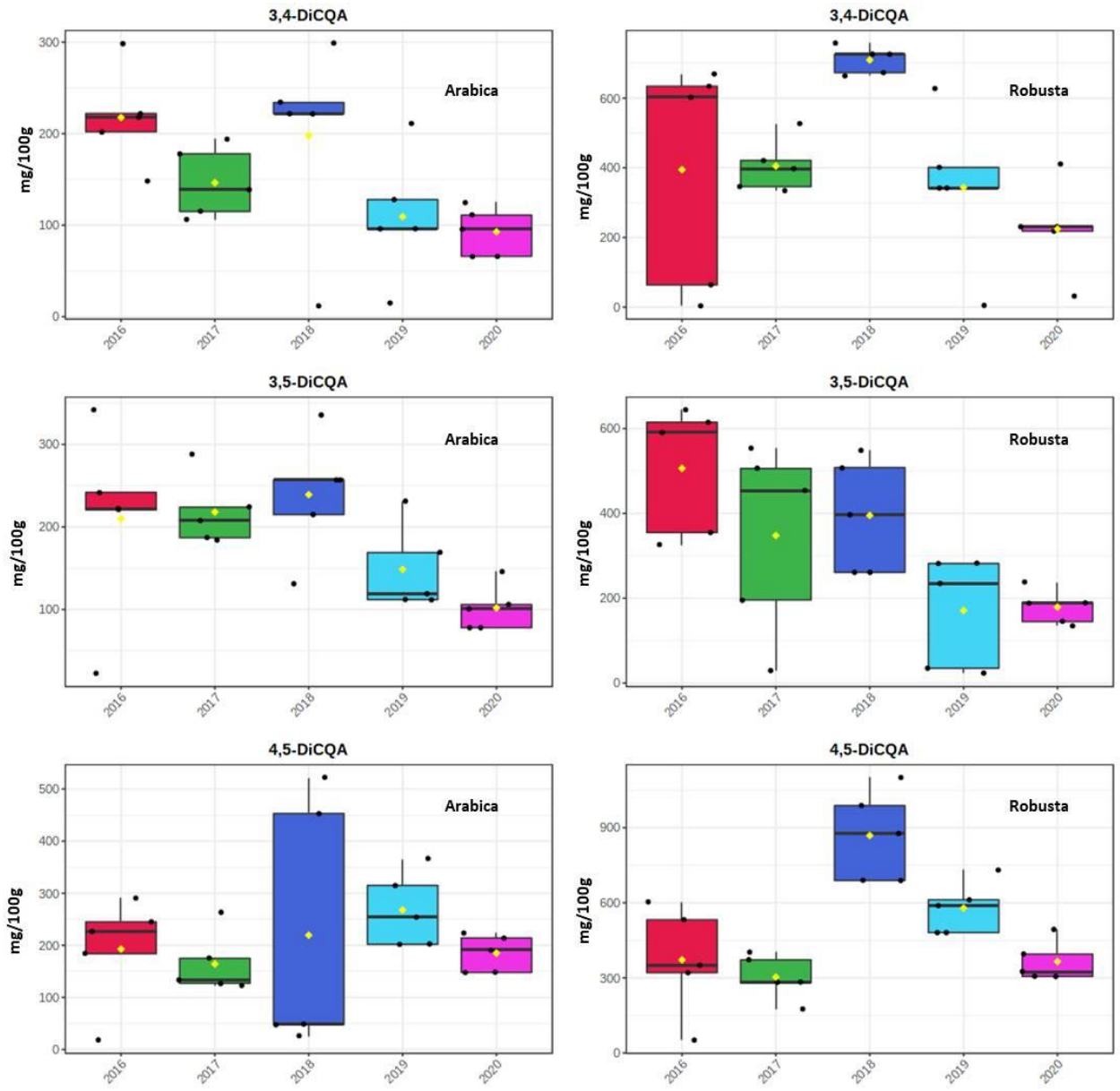


Figure 2 Boxplots of Di-CQAs in Arabica and Robusta green coffee from year 2016 to 2020.

The CQA and DiCQA data were further represented on bar chart to compare the coffee varieties and years across the studied years on one graphical representation as shown in Figures 3 and 4. The bar chart shows that there was gradual decrease of the compounds quantities from 2016 through 2020 as initially seen in the boxplots.

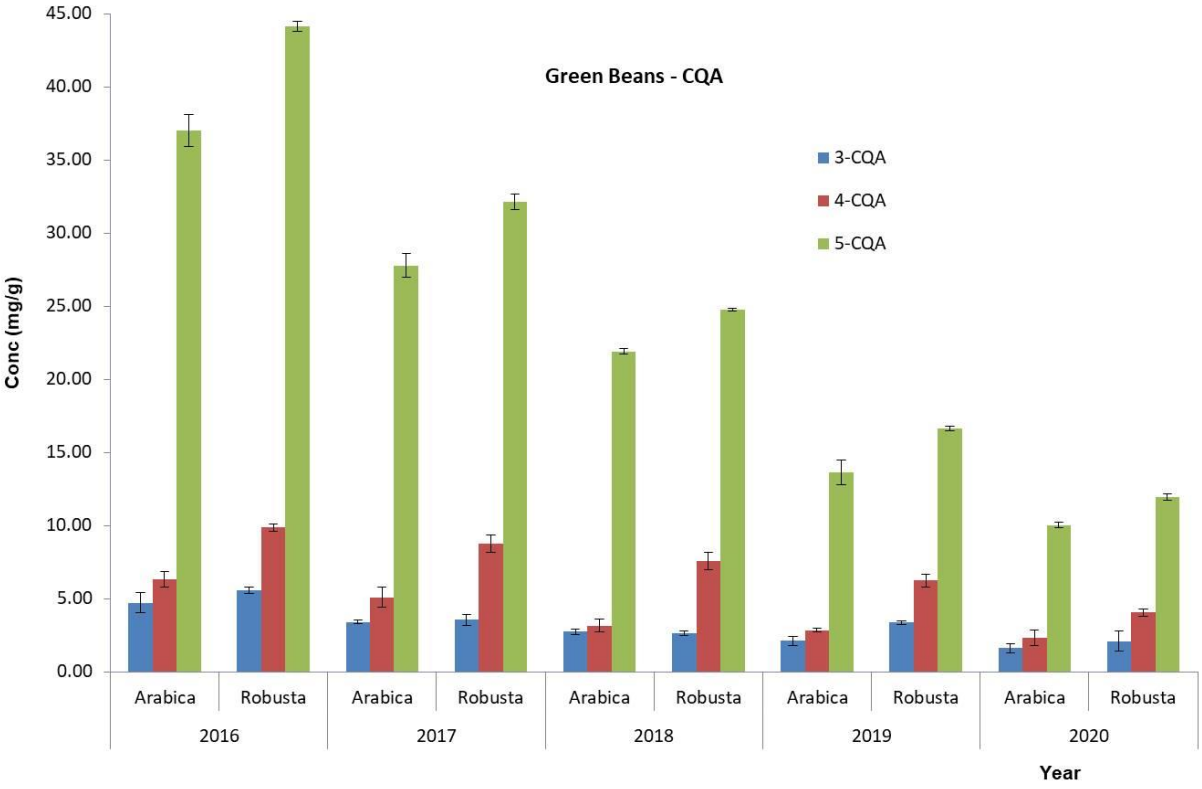


Figure 3 Mono-CQA extracted from stored green coffee beans from year 2016 through 2020.

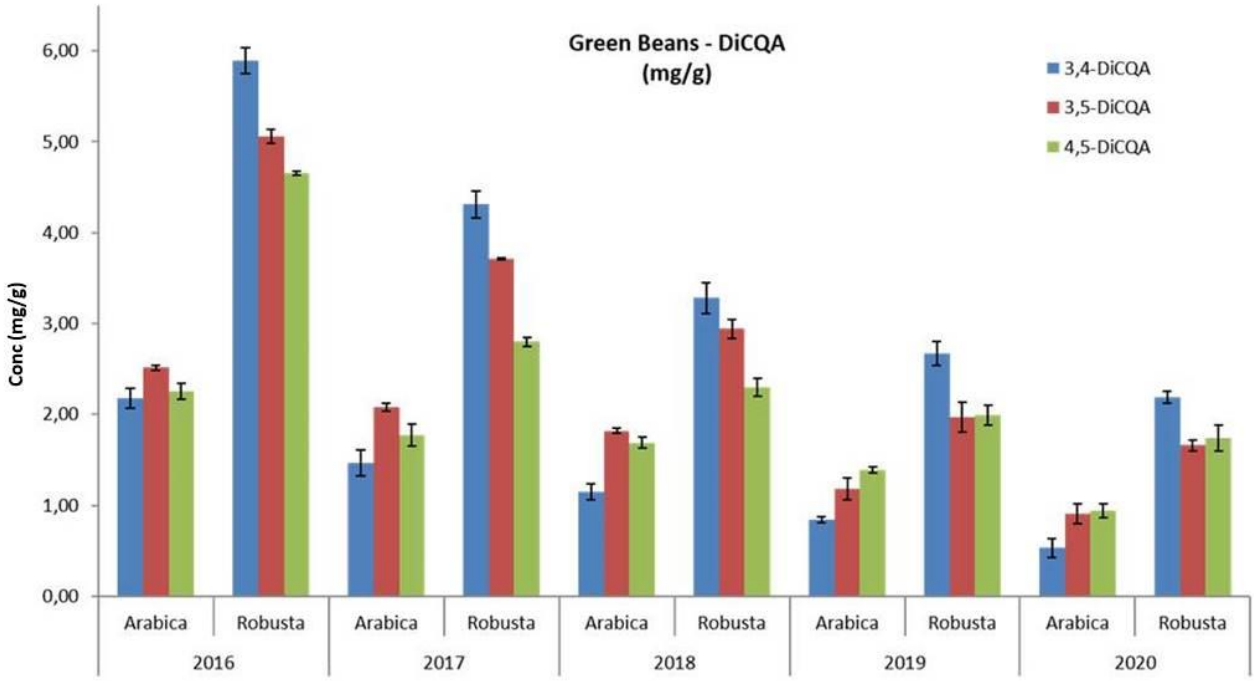


Figure 4 Di-CQA extracted from stored green coffee beans from year 2016 through 2020.

3.3.3 PCA of green coffee bean CGAs

Principal component analysis (PCA) was employed for the coffee CQAs using *MetaboAnalyst 4.0* data analysis software (Chong, Wishart & Xia, 2019). PCA is an unsupervised statistical tool which detects and classifies variances in a specific data set and also provides an overview of the diversity in the samples under study. From this, two plots were generated which are scores and loading plots. The mono- and di-CQAs quantitative data were analysed using PCA to see if the samples could be separated based on each year of analysis. The PCA plot in Figure 5 shows that the year 2018 was separated from other years distinctly in the middle of the plot, while years 2016 and 2017 overlapped, similar trend was observed between years 2019 and 2020 with a score of 81%. The mono-CQAs, 3,4- and 3,5- DiCQA isomers were observed to be the top 5 important features (compounds) as seen in the loading plot. The mono-CQAs and 3,5-DiCQA were observed to depleted from year 2016 through 2020.

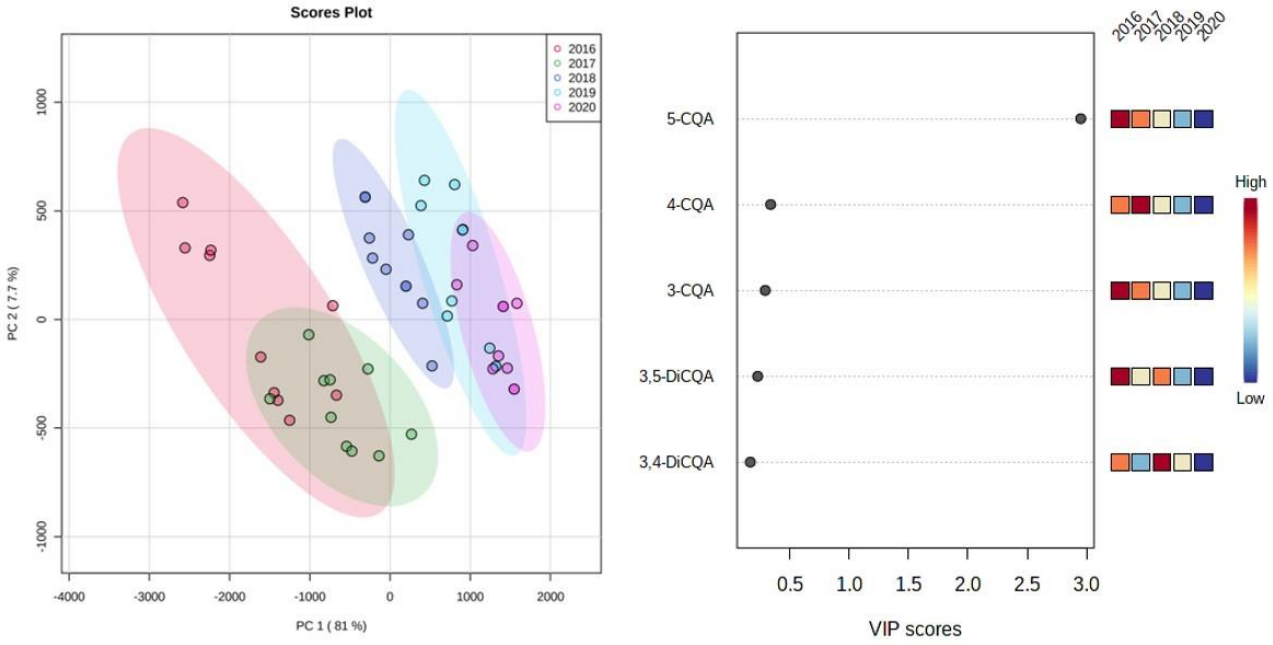


Figure 5 PCA scores and most important features of green coffee CGA from 2016 through 2020 using *m/z* range of 100 - 1200 with Pareto scaling and without any transformation.

Similarly, the FQA constituents of the coffee samples were graphically represented on a bar chart as shown in Figure 6. The 5-FQA was observed to be higher than the other two isomers in both Arabica and Robusta green coffee beans. Furthermore, it was observed that 4-FQA was higher than 3-FQA in Arabica across the years while a different trend was observed in Robusta with 3-FQA seen to be higher than 4-FQA. Moreover, the three compounds were observed to decrease significantly ($p < 0.05$) from year 2016 through 2020 irrespective of the coffee variety.

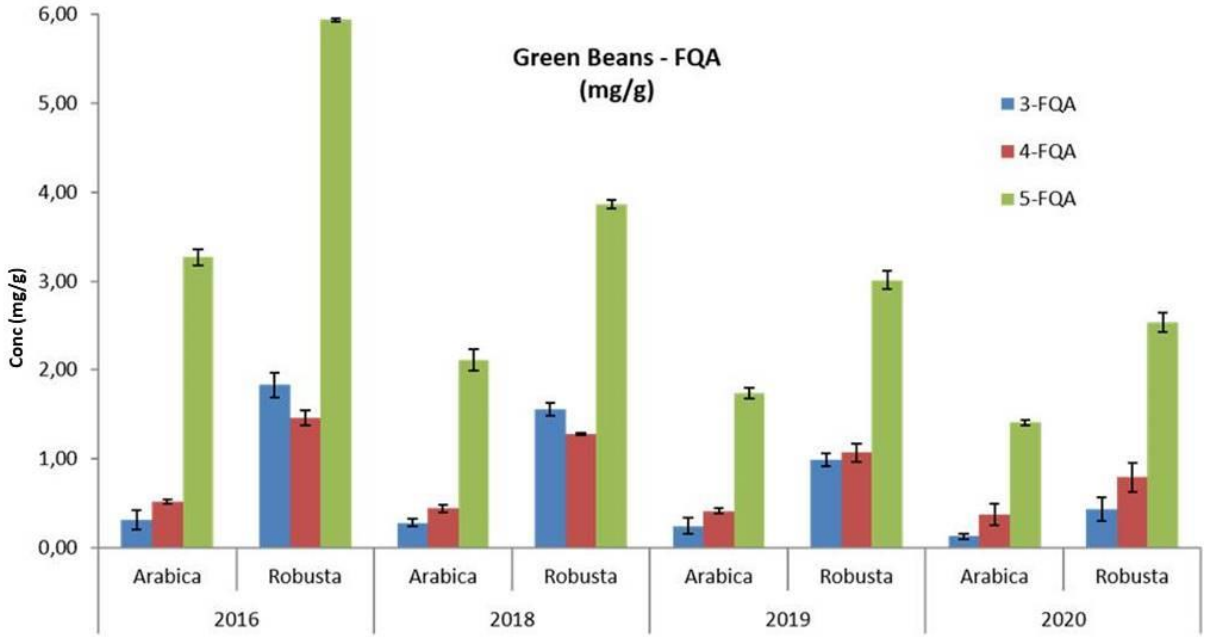


Figure 6 Mono-FQA extracted from stored green coffee beans from year 2016 through 2020.

3.3.4 Mono-acyl and di-acyl CGA in roasted coffee beans

During roasting green coffee beans act like a mini-bioreactor when exposed to high temperature (200 - 250 °C), which influences several biochemical reactions that occur during roasting such as Maillard reactions, Strecker degradation, pyrolysis (decomposition by heat) etc. This results in the production of over 1,000 different aroma compounds (Nebesny et al, 2007). However, coffee aroma is formed as a result of a complex mixture of these volatile compounds. Any

instabilities could affect the aromatic compounds profile (Nagaraju & Bhattacharya, 2010; Charles-Bernard et al, 2005).

The key coffee compositional primary and secondary metabolites are usually degraded during roasting to produce volatile or non-volatile compounds required for the development of distinctive coffee characteristics such as sweetness, flavour, aroma-body, bitterness, pigmentation, astringency etc. (Farah, 2005; Toci & Farah, 2008).

In this study, some roasted coffee samples were also studied for about two (2) years from the year 2018 through 2020. Two (2) commercially roasted coffee samples produced in 1990 and 2000 respectively were explored in this study and a replicate of each coffee produced by the same Companies (same pack and labels) were purchased from groceries store in Bremen in 2018 for comparison. In addition, three each of Arabica and Robusta green coffee beans were roasted as described in section 2.1 above and studied along with the two commercial roasted coffee beans. It is quite pertinent to mention that the two old coffees (1990 and 2000) were with factory seal before the commencement of this study, the samples were only opened for the first time in 2018. Thereafter, the beans were weighed out of the packs and ground accordingly. Invariably, ten (10) roasted coffee samples were used in this regard.

Furthermore, the CGA extraction was carried out using 70% MeOH, filtered and use for LC-MS measurements and antioxidant estimation. The chromatograms were acquired and data analysed as described above. CQAs, Di-CQAs, FQAs, *p*CoQAs and CGA lactones were detected in the roasted coffees. The identified CGA constituents were subsequently quantified.

In the roasted coffees, the three (3) isomers of mono-CQAs were observed to be depleted over time. Also, the 5-CQA being the most abundant CQA showed a significant decrease ($p < 0.05$) in all the samples over the period under study, and likewise, the 4-CQA and 3-CQA isomers. Table 3 shows the mean, standard deviation and p-values ($p < 0.05$) of the quantified CGA.

However, the quantities of the mono-CQAs in the two old coffees (1990 & 2000) were observed to significantly higher than the 2018 controls and newly roasted coffees bought in year 2011 and 2014. However, there is no significant difference between the Di-CQA and FQAs. For instance, shortly after opening the year 1990 coffee pack for the first time and being exposed to atmospheric air, the 5-CQA has a value of 15.97 mg/g compared to year 2018 coffee (9.43 mg/g), while the year 2000 coffee has 5-CQA (17.55 mg/g) compared to the 2018 control (13.09 mg/g). However, when the 5-CQA laboratory roasted coffee samples (2011 & 2014) were compared, there is a significant difference in their quantities.

Moreover, the significant phenomenon observed here is that the old coffees (1990 & 2000) CQAs were depleted at a faster rate compared to other coffees, which probably indicates that there is a possibility that when the coffee was first exposed to air the compounds undergoes degradation more quickly compared to when it was sealed.

Table 3 Mean+s.d and p-values ($p < 0.05$) of the CGA extracted from roasted coffee bean from 2018 through 2020.

		3-CQA	4-CQA	5-CQA	3,4-DiCQA	3,5-DiCQA	4,5-DiCQA	3-FQA	4-FQA	5-FQA	1-CQL	2-CQL	3-CQL	4-CQL
2018	DALL1990 (n=1)	10.14	13.18	15.97	1.37	0.67	2.06	1.34	2.42	4.23	1.72	0.99	0.86	0.00
	DALL2018 (n=1)	5.77	8.91	9.43	0.99	0.38	1.54	1.29	1.63	2.32	0.78	0.60	0.58	0.14
	JCB2000 (n=1)	11.13	15.96	17.55	1.77	0.68	2.31	1.47	2.67	4.35	1.38	0.70	0.64	0.00
	JCB2018 (n=1)	8.02	12.23	13.09	1.46	0.46	2.05	1.24	2.34	4.13	0.89	0.70	0.53	0.22
	Brazil (n=3)	2.62±0.2	5.41±0.6	15.86±0.8	1.34±0.1	1.04±0.1	2.03±0.2	0.00±0.0	0.30±0.0	2.56±0.2	0.60±0.1	0.46±0.1	0.22±0.04	0.00±0.0
	Indian (n=3)	3.21±0.3	8.32±0.4	21.57±1.6	2.11±0.2	1.49±0.1	3.08±0.2	0.00±0.0	1.12±0.1	4.74±0.3	0.23±0.1	0.62±0.1	0.49±0.0	0.00±0.0
2019	DALL1990 (n=1)	6.30	8.51	9.81	0.93	0.31	1.50	0.77	1.38	2.44	0.94	0.53	0.52	0.00
	DALL2018 (n=1)	5.12	8.66	9.08	0.77	0.28	1.22	0.69	1.53	1.82	0.80	0.56	0.43	0.21
	JCB2000 (n=1)	9.11	10.35	15.66	1.12	0.46	1.83	1.48	2.64	3.70	1.25	0.67	0.49	0.00
	JCB2018 (n=1)	5.97	8.93	9.69	1.04	0.37	1.54	0.99	1.71	3.18	0.70	0.54	0.53	0.17
	Brazil (n=3)	2.11±0.1	3.05±0.2	11.03±0.6	1.30±0.1	0.68±0.1	1.84±0.2	0.00±0.0	0.26±0.01	1.73±0.2	0.36±0.01	0.31±0.1	0.27±0.1	0.00±0.0
	Indian 2014 (n=3)	2.78±0.3	4.79±0.3	17.14±1.9	1.49±0.2	1.06±0.1	1.97±0.3	0.00±0.0	0.93±0.1	3.61±0.4	0.13±0.01	0.58±0.02	0.39±0.02	0.00±0.0
2020	DALL1990 (n=1)	5.24	7.11	7.98	0.87	0.26	1.24	0.59	1.06	1.81	0.73	0.40	0.34	0.00
	DALL2018 (n=1)	4.81	8.04	8.90	0.73	0.25	1.03	0.67	1.27	1.65	0.70	0.51	0.41	0.00
	JCB2000 (n=1)	6.45	7.88	10.71	0.79	0.33	1.18	1.22	2.07	2.93	1.06	0.53	0.44	0.00
	JCB2018 (n=1)	4.57	6.91	7.47	0.61	0.30	0.95	0.74	1.35	2.47	0.48	0.41	0.32	0.15
	Brazil (n=3)	1.77±0.2	2.02±0.1	6.29±0.5	0.55±0.01	0.19±0.3	0.91±0.2	0.00±0.0	0.17±0.01	0.96±0.3	0.22±0.03	0.15±0.02	0.09±0.0	0.00±0.0
	Indian (n=3)	2.20±1.5	3.13±2.4	12.46±5.7	1.01±0.6	0.49±0.8	1.48±1.6	0.00±0.0	0.60±1.3	2.96±4.8	0.01±0.0	0.34±1.6	0.06±0.4	0.00±0.0

2018/2020	**3-CQA**	**4-CQA**	**5-CQA**	**3,4-DiCQA**	**3,5-DiCQA**	**4,5-DiCQA**	**3-FQA**	**4-FQA**	**5-FQA**	**3-*p*CoQA**	**4-*p*CoQA**	**5-*p*CoQA**	**1-CQL**	**2-CQL**	**3-CQL**
($p < 0.05$) Arabica	0.94	0.77	0.59	0.62	0.59	0.79	0.92	0.79	0.69	0.07	0.44	0.39	0.82	0.43	0.66

Table 4 Percentage rate of degradation (%) of CGA compounds extracted from the roasted coffee bean from 2018 to 2020.

Coffee	Year	Bean Variety	3-CQA	4-CQA	5-CQA	3,4-DiCQA	3,5-DiCQA	4,5-DiCQA	3-FQA	4-FQA	5-FQA	3-pCoQA	4-pCoQA	5-pCoQA	1-CQL	2-CQL	3-CQL	4-CQL
D-I	1990	Arabica	48.30	46.02	50.02	36.69	61.27	39.83	56.06	56.24	57.18	52.73	57.99	53.62	57.59	59.67	59.85	Tr
D-II	2018	Arabica	16.65	9.78	5.61	26.26	34.21	33.12	47.91	21.90	29.02	Tr	3.53	8.23	10.46	15.33	29.31	Tr
J-I	2000	Arabica	42.07	50.61	38.97	55.37	51.21	48.92	17.17	22.56	32.64	12.74	23.41	22.85	23.34	24.47	31.48	Tr
J-II	2018	Arabica	43.03	43.50	42.90	58.22	35.05	53.56	40.56	42.37	40.35	45.77	44.27	35.91	45.72	41.79	40.25	32.42
B-I	2011	Arabica	32.40	62.58	60.34	59.01	81.73	55.17	Tr	43.33	62.48	Tr	Tr	Tr	63.33	67.39	59.09	Tr
IN-I	2014	Robusta	31.44	62.38	42.23	52.13	67.11	52.01	Tr	46.43	37.55	Tr	Tr	Tr	95.59	45.21	87.67	Tr

Tr: Trace amount detected.

The bar chart in Figure 7 represents the varied quantities of the mono-CQA in the roasted coffee from year 2018 through 2020. It was observed that the Robusta (Indian 2014) samples has the highest amount of the 5-CQA compared to other coffees that are Arabica. Furthermore, it could be said that the factory roasting conditions (different parameters) and the use of preservatives have an influence on the chlorogenic acids contents and amounts when the four (4) factory roasted (1990, 2000 & 2018) coffees are compared with the laboratory roasted (2011 & 2014) samples. But despite this, the CGAs in the roasted coffees were observed to be degraded over time, and this is more conspicuous in the laboratory roasted coffee samples. The reason for the reduced amounts of the CGAs in the laboratory roasted coffees could be because the samples have been exposed to atmospheric air for some years before roasting compared to the factory roasted coffee samples. Figures 8, 9, 10 and 11 represent the bar charts of the FQA, *p*CoQA, CGA lactones and Di-CQA respectively. The charts show that the compounds were depleted over time, but not significantly ($p < 0.05$) as shown in Table 3 data with the p-values.

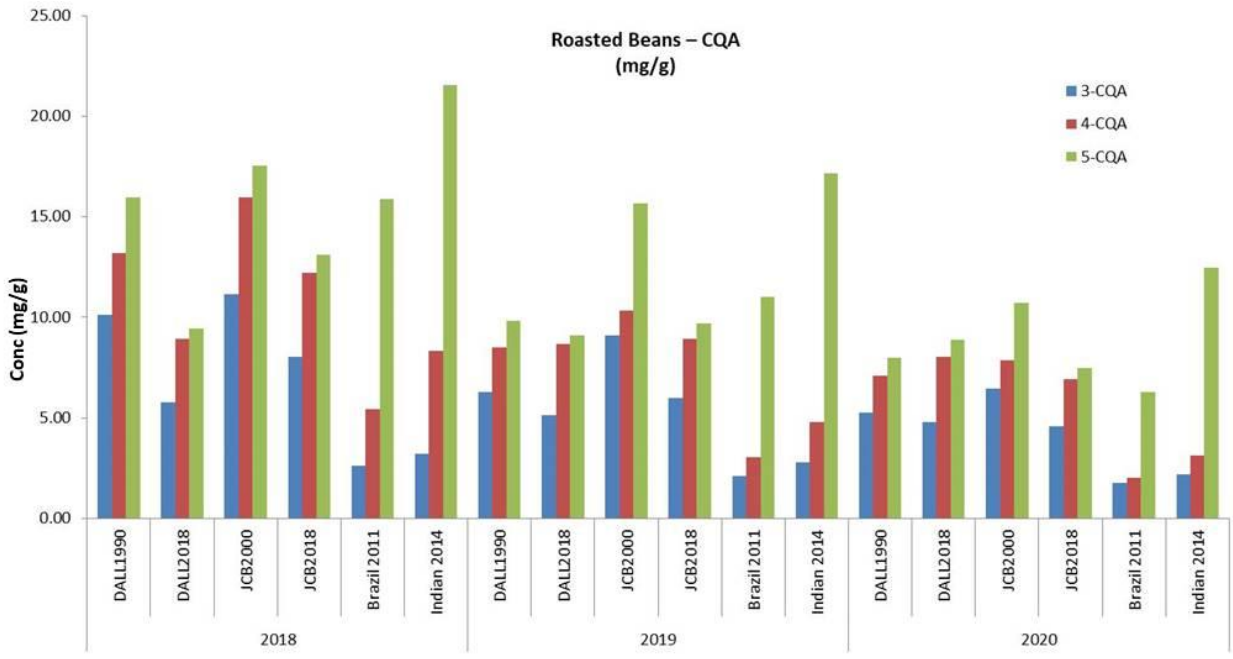


Figure 7 Mono-CQA extracted from stored roasted coffee from year 2018 through 2020. DALL, JCB and Brazil are Arabica while Indian is Robusta.

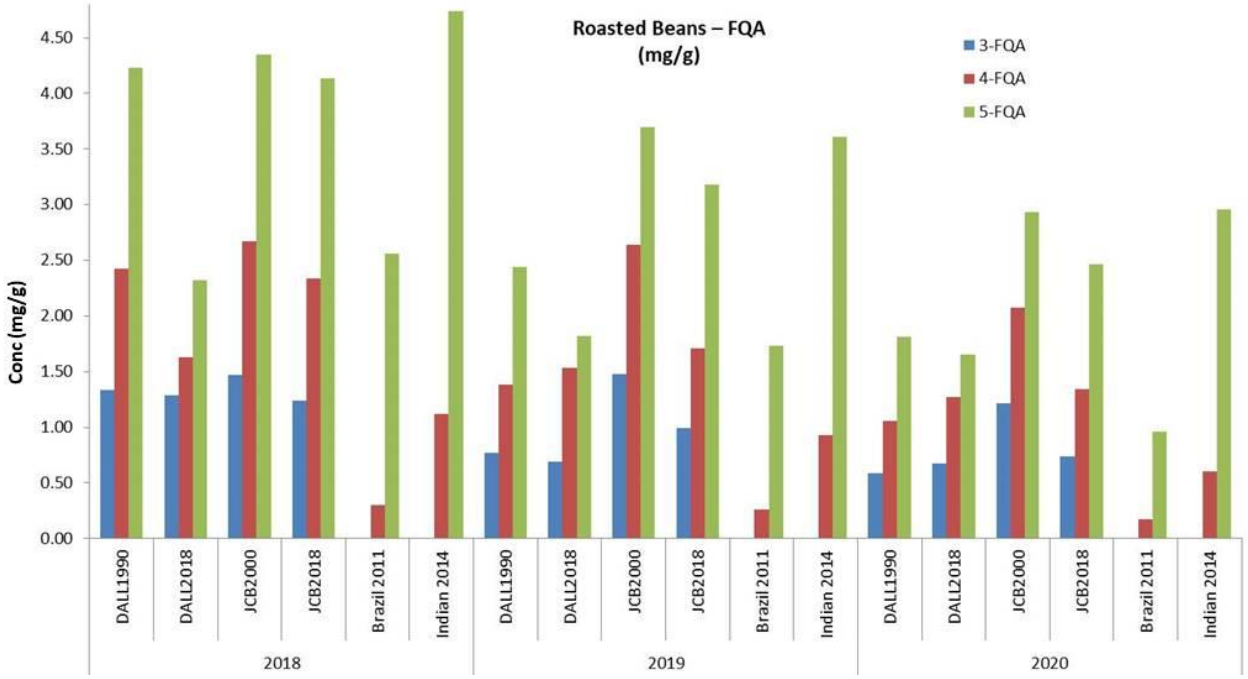


Figure 8 Mono-FQA extracted from stored roasted coffee beans from year 2018 to 2020. DALL, JCB and Brazil are Arabica while Indian is Robusta.

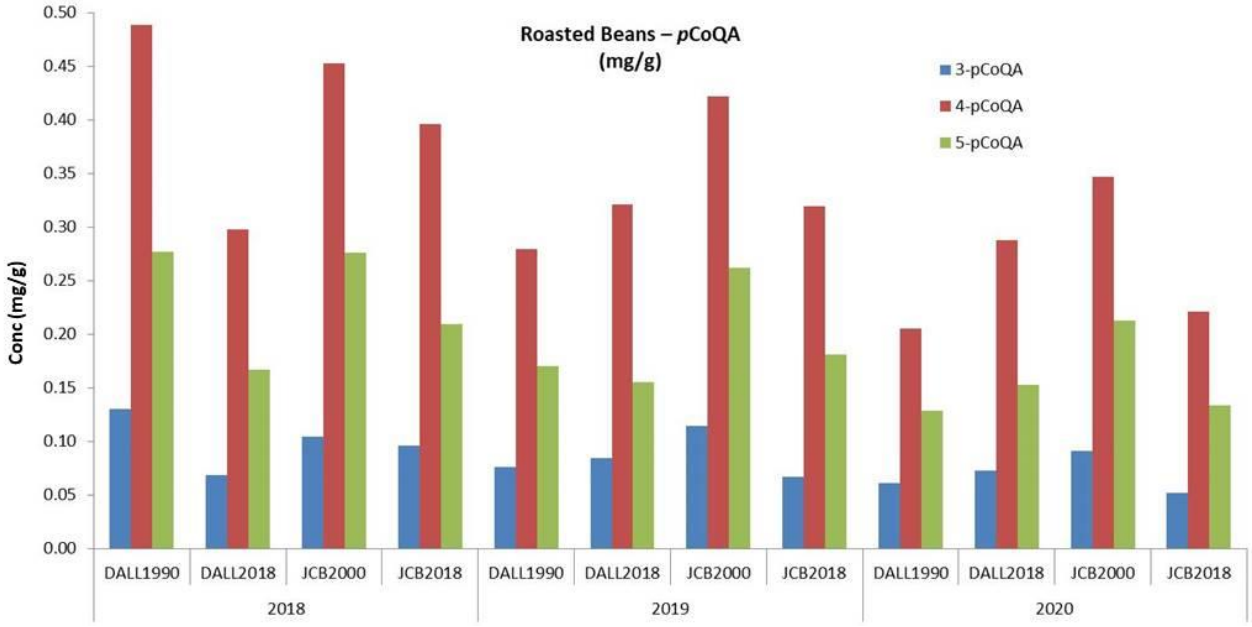


Figure 9 *p*CoQA extracted from stored roasted coffee beans from year 2018 through 2020. DALL, JCB and Brazil are Arabica while Indian is Robusta.

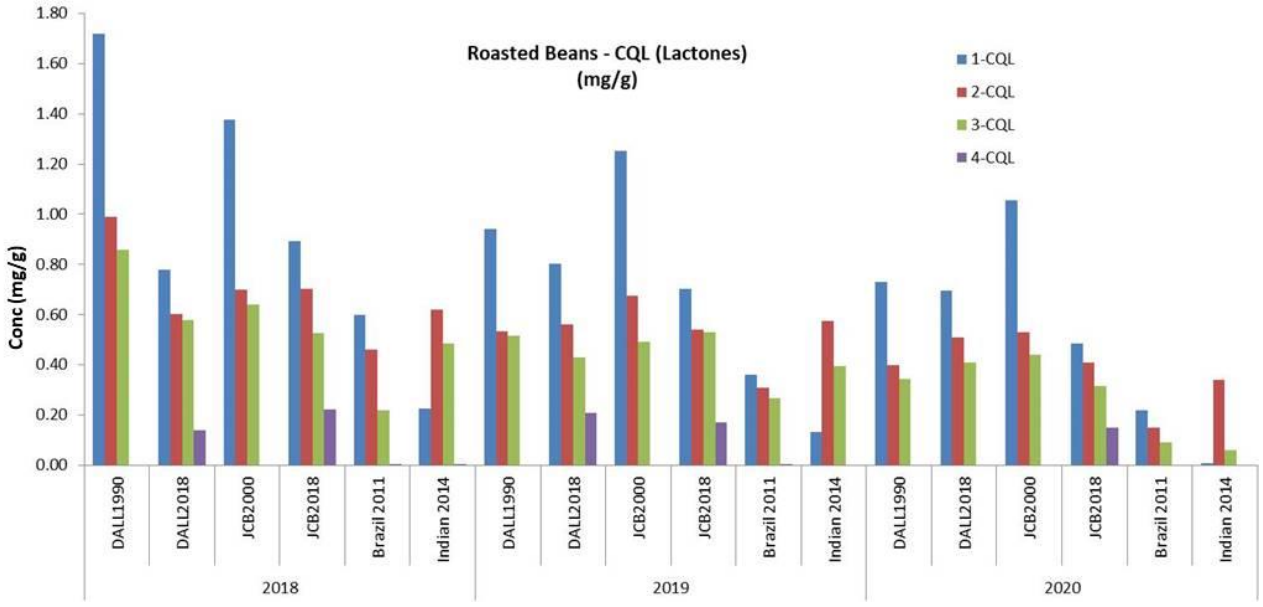


Figure 10 CQL (lactones) extracted from stored roasted coffee beans from year 2018 to 2020. DALL, JCB and Brazil are Arabica while Indian is Robusta.

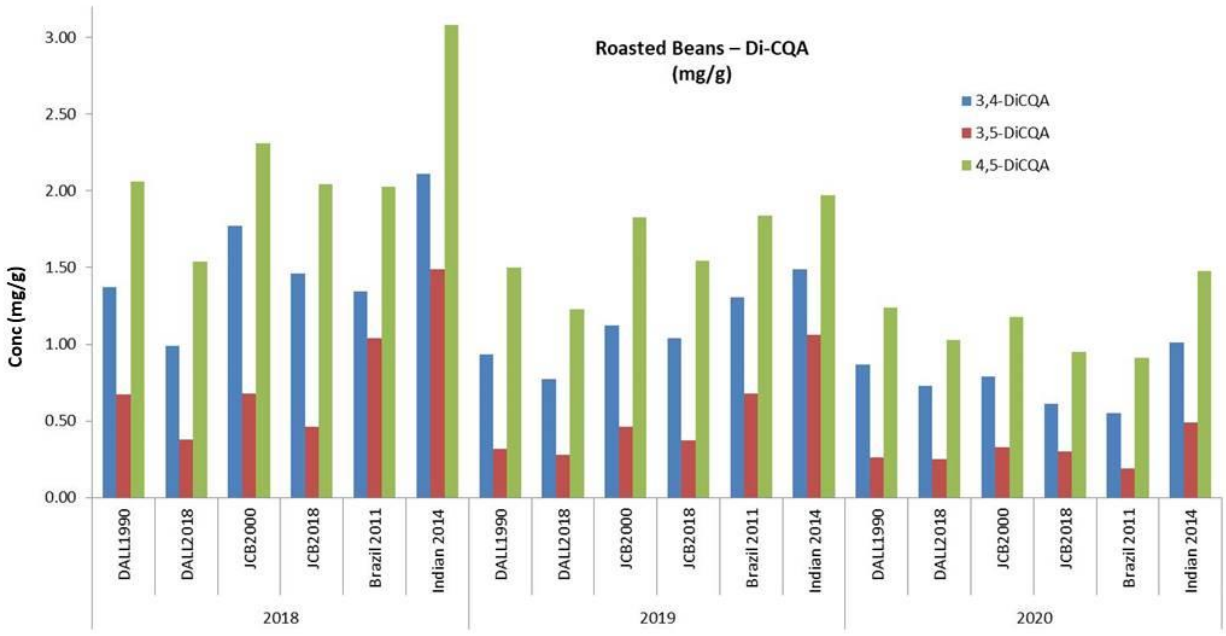


Figure 11 Di-CQA extracted from stored roasted coffee beans from year 2018 through 2020. DALL, JCB and Brazil are Arabica while Indian is Robusta.

The percentage rate of degradation was calculated using the formulae below and computed for the roasted samples as shown in Table 3.

$$\textbf{Percentage change (\%)} = \frac{(2018) - (2020)}{(2020)} \times 100$$

The data in the rate of degradation Table 4 shows that the CGA compounds in the 1990 roasted coffee were depleted by over 50%. The most degraded compound among all is 3,5-Di-CQA (61.27%) followed by the CGA lactones, FQAs and CQAs respectively. Similar trend was observed in other coffees. However, it could be suggested that most of the CGA compounds are averagely stable despite the long years of storage and for the fact that nearly all the compounds of interest were observed to be present after three decades (30 years) of storage when compared to other coffee samples.

3.3.5 Green coffee beans stored with carbon dioxide

Green coffee beans were stored with carbon dioxide for a period of two (2) years differently at room temperature (RT) and 4 ºC to investigate the effect of the gas on the bean chemical constituents in the absence of normal atmospheric air. Basically, the main atmospheric air components are nitrogen and oxygen. Both constitute 99% of clean and dry air, while other components are water vapor and trace gases with approximately 0.25% and 0.97% of the atmosphere by mass respectively. Besides the atmospheric air could also contains pollutants (such as sulphur gases, chloride etc) which influences the main air components mixture and ratios. Invariably, the exclusion of the nitrogen, oxygen, water vapour, trace gases and pollutant could possibly influence the green beans chemical components and varying concentration.

In this study, the coffee bean was weighed into a round-bottom flask, followed by vacuum and connected with carbon dioxide supply before the supply tap is closed, to ensure that it is only carbon dioxide and coffee bean that are present in the round-bottom flask. The CGA contents

of the beans were analysed before the start of the experiment which serves as control. Thereafter, the samples (Arabica and Robusta groups) were further grouped into two storage conditions which are RT and 4 ºC. After the duration of storage, the CGA contents of the samples were analysed and compared with the control groups. Figure 12 depicts the mono-CQA contents of the green coffee stored with carbon dioxide for two (2) years. It was observed that the 5-CQA constituents of the two coffee varieties were higher than 3-CQA and 4-CQA in control groups. Moreover, the concentration of the 5-CQA decreases by a factor of 4 after the two years duration of storage when compared with the controls.

Furthermore, a slight decrease was observed with 3-CQA and 4-CQA isomers when compared with the control. This suggests that 5-CQA depletion was influenced by the carbon dioxide, whereas the gas has little or no effects on 3-CQA and 4-CQA. So, it can be deduced that carbon dioxide would be among the gases that affects 5-CQA contents of coffee CGA during storage. However, the chart shows that carbon dioxide has little or no influence on the coffee bean when bean variety is considered.

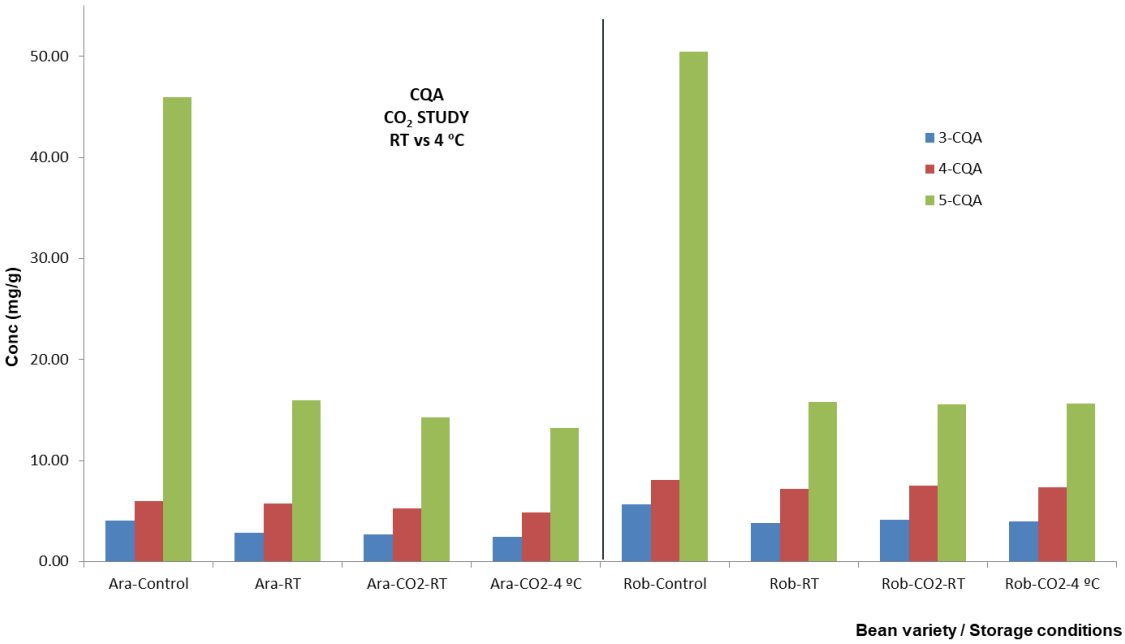


Figure 12 Mono-CQA extracted from Arabica and Robusta green coffee beans stored with CO_2 differently at room temperature and 4 ºC for a period of two (2) years.

Likewise, the Di-CQA content of the coffee samples stored with CO_2 were analysed and data represented in Figure 13. It was observed that there are differences between the Arabica and Robusta coffees when compared with regards to each of the Di-CQA isomers. In the Arabica control, 3,5-DiCQA level was observed to be higher than the other two isomers that is 3,4- and 4,5- isomers; however, after the storage period the 3,5-DiCQA decreases in both coffees stored at RT and 4 ºC when compared with the control. Furthermore, it was observed that 3,4-Di-CQA levels decreases with storage time in both storage temperature conditions when compared with the control. Whereas, the 4,5-Di-CQA levels remained fairly the same but with slight decrease in the 4 ºC stored coffee.

Moreover, in the Robusta coffee beans, the levels of 4,5-Di-CQA remained relatively the same, while the level of 3,5-Di-CQA was observed to decreased by over 50% when compared with the control. Also, a decrease was also observed with the 3,4-Di-CQA levels in both temperature storage conditions when compared with the control. It could be suggested that CO_2 has an effect on both 3,5-Di-CQA and 3,4-Di-CQA levels, which more pronounced in the former, while the gas has little or no effect on 4,5-Di-CQA.

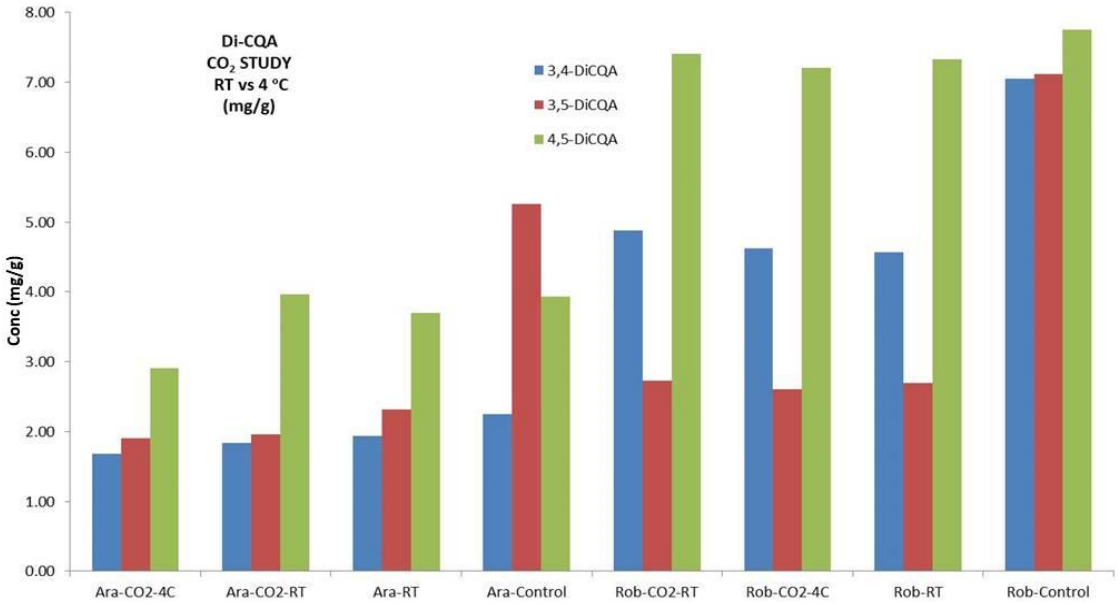


Figure 13 Di-CQA extracted from Arabica and Robusta green coffee beans stored with CO_2 differently at room temperature and 4 ºC for a period of two (2) years.

The *p*CoQA isomers were also analysed in coffees stored with CO_2 and data shown in Figure 14. In the Arabica coffees, a slight decrease was observed in 4-*p*CoQA and 5-*p*CoQA levels, while the levels of the 3-*p*CoQA decreased by about 50% when compared with the controls. Whereas in the Robusta, 3-*p*CoQA was observed to be virtually depleted when compared with the control. The levels of 4-*p*CoQA remained relatively the same except for a slight decrease observed in sample stored at RT. Also, the 5-*p*CoQA levels was observed to remained almost the same level when compared with the control.

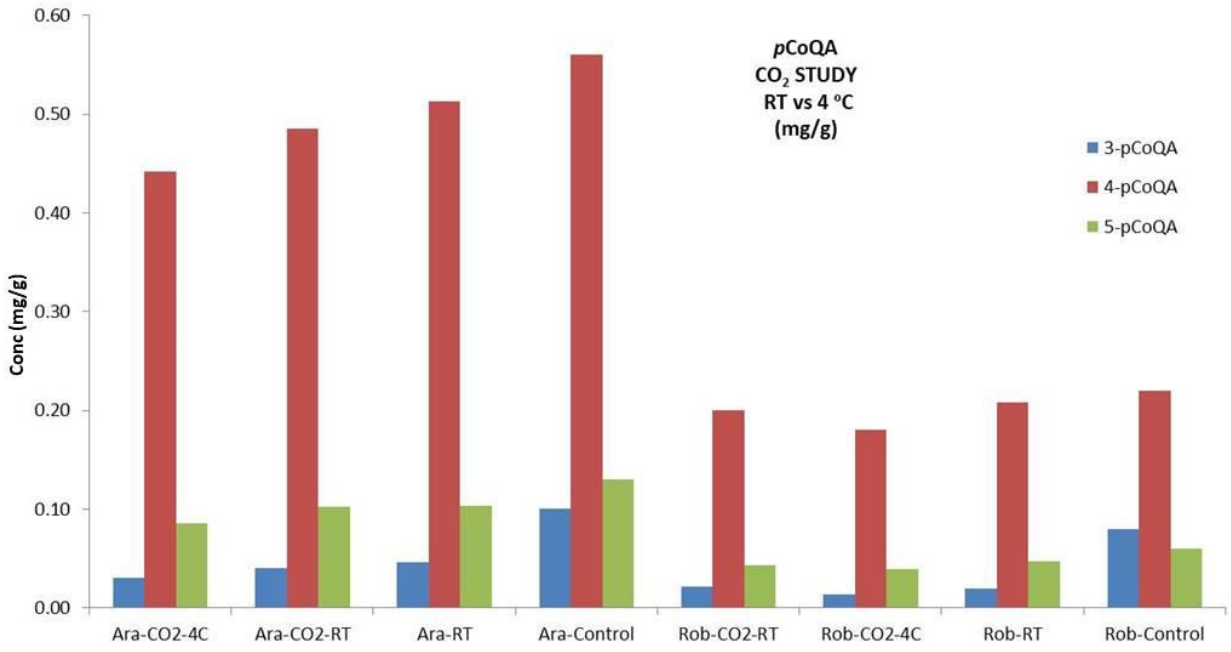


Figure 14 Mono-*p*CoQA isomers extracted from Arabica and Robusta green coffee beans stored with CO_2 differently at room temperature and 4 ºC for a period of two (2) years.

3.3.6 Green coffee beans stored with Nitrogen, atmospheric air and H_2O

In this study, the coffees were grouped into three sets stored differently with nitrogen only, atmospheric air only, and nitrogen with water only at RT and 4 ºC respectively. The samples were then analysed after the two years storage.

Figure 15 represents the coffees stored with CQAs. From the bar chat, it was observed that the mono-CQAs were more depleted in the Arabica coffees than in the Robusta coffees and when

compared with the control. Most importantly, the 5-CQA was observed to be higher than 3- and 4- isomers in the controls; and it decreased with a factor of about four (4) after the storage period when compared with the controls in both Arabica and Robusta coffee varieties. Furthermore, it was observed that 4-CQA remained the same in the Arabica group stored with nitrogen only compared to second group stored with nitrogen with H_2O which decreased significantly both at RT and 4 ºC. It could be suggested that the high depletion observed with the nitrogen and water group was as a result of hydrolysis reaction. Moreover, in the Robusta coffee groups, the 3-CQA levels remained relatively the same in both groups stored with at RT and 4 ºC. While a slight increase was observed in 4-CQA in all the groups when compared with controls. In addition, nitrogen only group at 4 ºC was slightly depleted when compared with the controls.

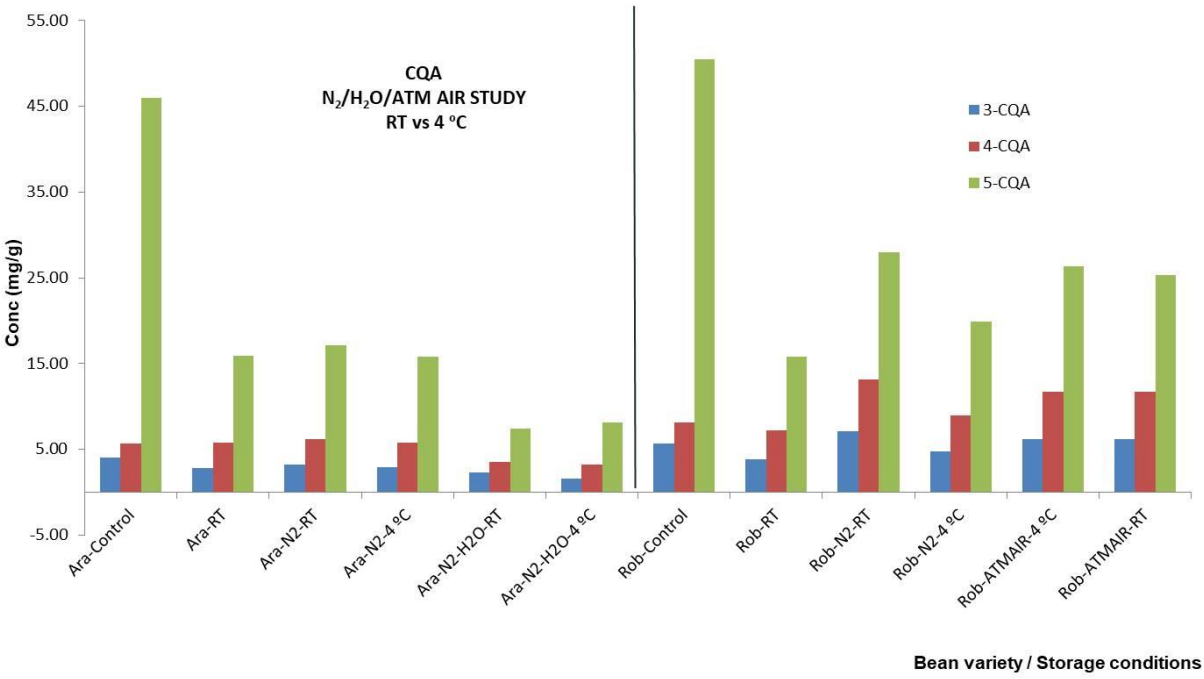


Figure 15 Mono-CQA extracted from Arabica and Robusta green coffee stored with N_2 only, N_2 with H_2O and atmospheric air differently at RT and 4 ºC for a period of two (2) years.

Furthermore, the Di-CQAs were also analysed and presented in the bar chart in Figure 16. The Robusta groups presented higher concentrations when compared with the Arabica groups

generally. Moreover, the 4,5-Di-CQA was observed to be increased in the nitrogen only and the atmospheric groups when compared with the controls. Besides, the 3,4-Di-CQA was also increased in the two groups when compared with the controls. While the control group was observed to be decreased by nearly 50% in 3,5-Di-CQA and a moderate decrease level was seen with the 3,4-Di-CQA both with the control and nitrogen only group.

Moreover, in the Arabica groups, the 3,5-Di-CQA was observed to be higher than 3,4- and 4,5-Di-CQA in the control, while it was seen with about 50% decrease after the storage period time as shown in Figure 16. In addition, the 3,4-Di-CQA remained relatively the same in the nitrogen only groups when compared with the control group. Whereas all the three isomers in the nitrogen with H_2O group was drastically depleted when compared with the control group. It could also be suggested that nitrogen with H_2O has significant effects on the three isomers both at RT and 4 ºC storage conditions probably due to hydrolysis.

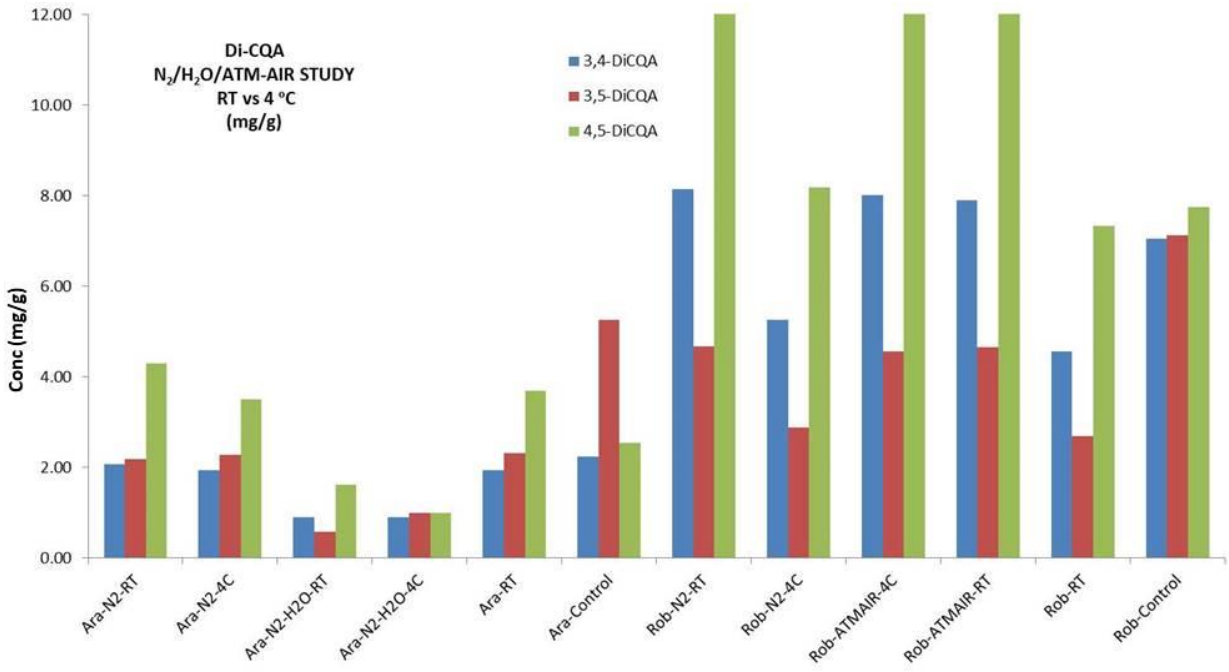


Figure 16 Di-CQA extracted from Arabica and Robusta green coffee beans stored with N_2 only, N_2 with H_2O and atmospheric air differently at room temperature and 4 ºC for 2 years.

The *p*CoQA isomers were also analysed to see the effect of nitrogen only, nitrogen with water (in Arabica) and nitrogen only and atmospheric air only (in Robusta) at RT and 4 ºC. Figure 17 depicts the bar chart of the *p*CoQA levels before and after the two years of different storage conditions. The levels of 4-pCoQA remained relatively the same in the nitrogen only and atmospheric air only groups when compared with the controls. The group with nitrogen and water were observed to be decreased when compared with other groups and the controls. 3-pCoQA was observed to be completely depleted in the Robusta stored with nitrogen only and stored at 4 ºC. It could be said that nitrogen with H_2O has a profound effect on the three *p*CoQA isomers when compared with the controls and other groups.

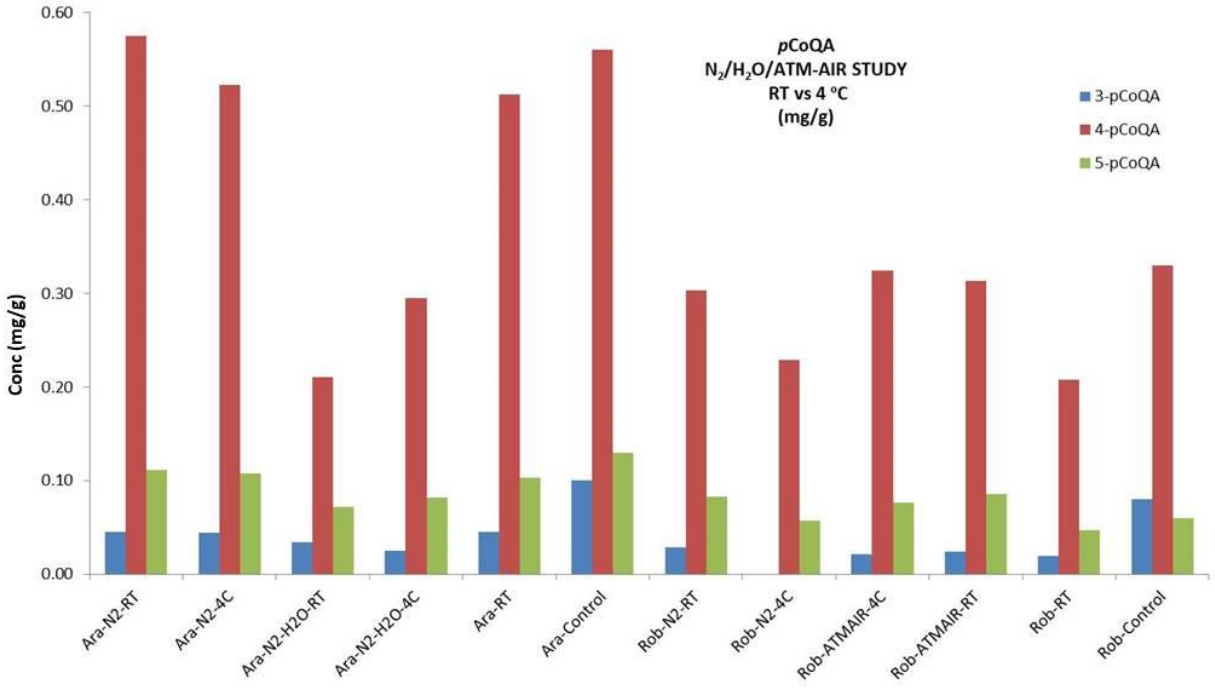


Figure 17 *p*CoQA extracted from Arabica and Robusta green coffee beans stored with N_2 only, N_2 with H_2O and atmospheric air differently at room temperature and 4 ºC for 2 years.

3.3.7 Statistical analysis of CGAs

Pearson's correlation was utilized in this study to see if there is a linear relationship between the CGAs and the research hypothesis here would be how does one score affects the other in a certain way. The correlation is generally affected by the variables size and r values.

Furthermore, Figures 18 and 19 with Tables 5 and 6 represent the heatmaps and Pearson correlation coefficient values of the green coffee beans mono-CQA, Di-CQA and FQA for the Arabica and Robusta. The Pearson values (r = 1) show that most of the CGA were positively correlated across the years from 2016 to 2020, while some were negatively correlated.

In addition, heatmap were plotted to visually study the relationships between the variables (CGAs) and the years of study (2016 – 2020), with one variable plotted on each axis. This was carried out by observing how the cell colors change across each axis and observe if there are any patterns in value for one or both variables. From the heatmap plots for both the Arabica and Robusta coffees, it was observed that as the years passes by from 2016 to 2020 the amount of the CGA constituents in the two coffees were gradually depleted. This invariably corroborates the numeric data (r = 1) acquired from the Pearson correlation coefficient tables for both coffee types.

Table 5 Pearson correlation coefficient (r = 1) of green Arabica CGAs 2016 to 2020.

Arabica	Year	3-CQA	4-CQA	5-CQA	3-FQA	4-FQA	5-FQA	3,4-DiCQA	3,5-DiCQA	4,5-DiCQA
Year	1	-0,--786	-0,556	-0,960	0,014	0,093	0,155	-0,581	-0,723	-0,006
3-CQA	-0,786	1	0,803	0,829	0,193	0,005	-0,102	0,653	0,487	0,086
4-CQA	-0,556	0,803	1	0,61	0,02	-0,14	-0,23	0,44	0,25	-0,09
5-CQA	-0,960	0,829	0,613	1	0,031	-0,040	-0,096	0,641	0,763	0,066
3-FQA	0,014	0,193	0,019	0,031	1	0,921	0,877	0,677	0,289	0,818
4-FQA	0,093	0,005	-0,145	-0,040	0,921	1	0,984	0,623	0,354	0,916
5-FQA	0,155	-0,102	-0,232	-0,096	0,877	0,984	1	0,540	0,313	0,911
3,4-DiCQA	-0,581	0,653	0,439	0,641	0,677	0,623	0,540	1	0,807	0,714
3,5-DiCQA	-0,723	0,487	0,247	0,763	0,289	0,354	0,313	0,807	1	0,487
4,5-DiCQA	-0,006	0,086	-0,094	0,066	0,818	0,916	0,911	0,714	0,487	1

Figure 18 Heatmap from Pearson correlation coefficient of green Arabica CGA 2016 to 2020.

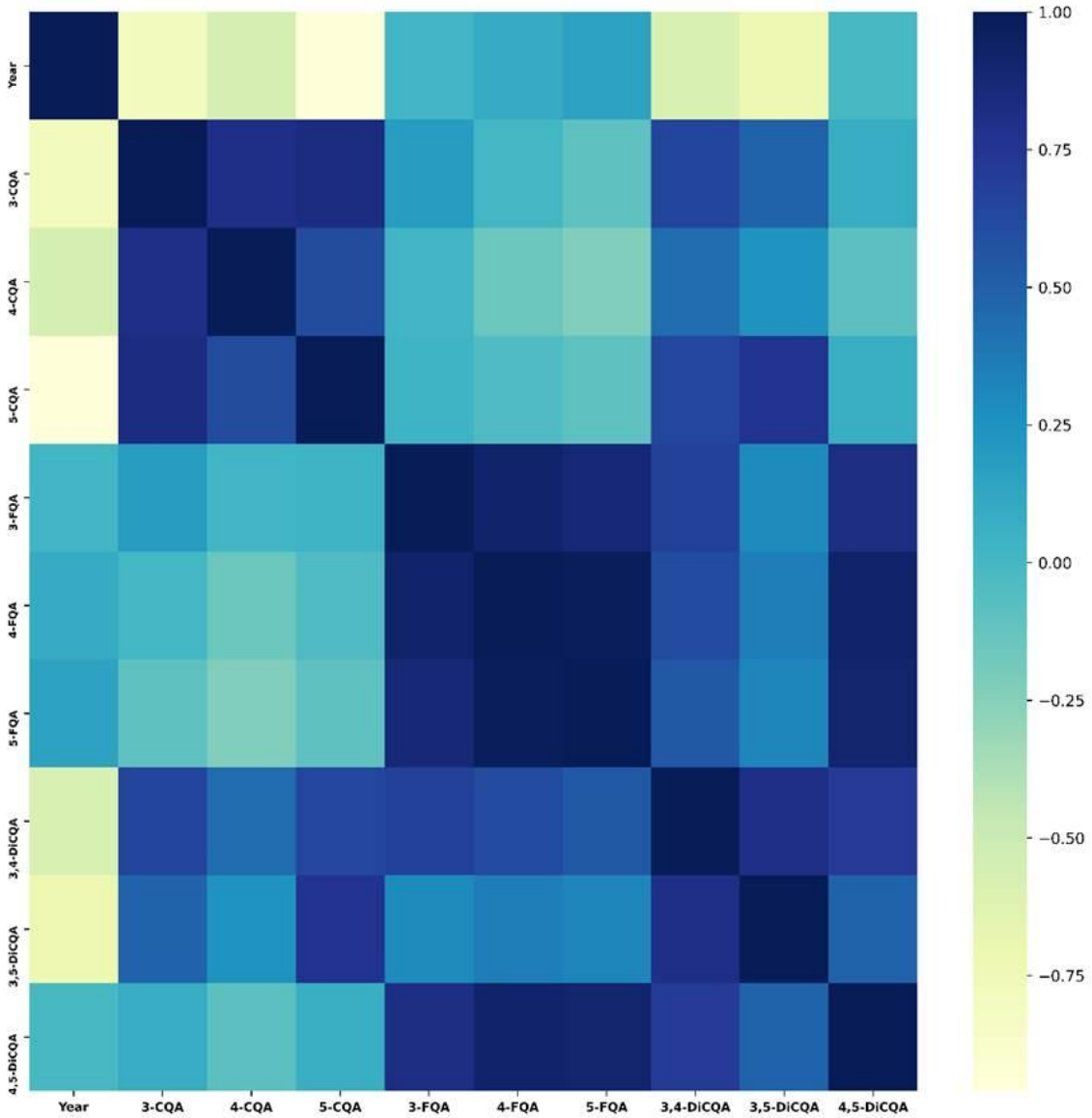
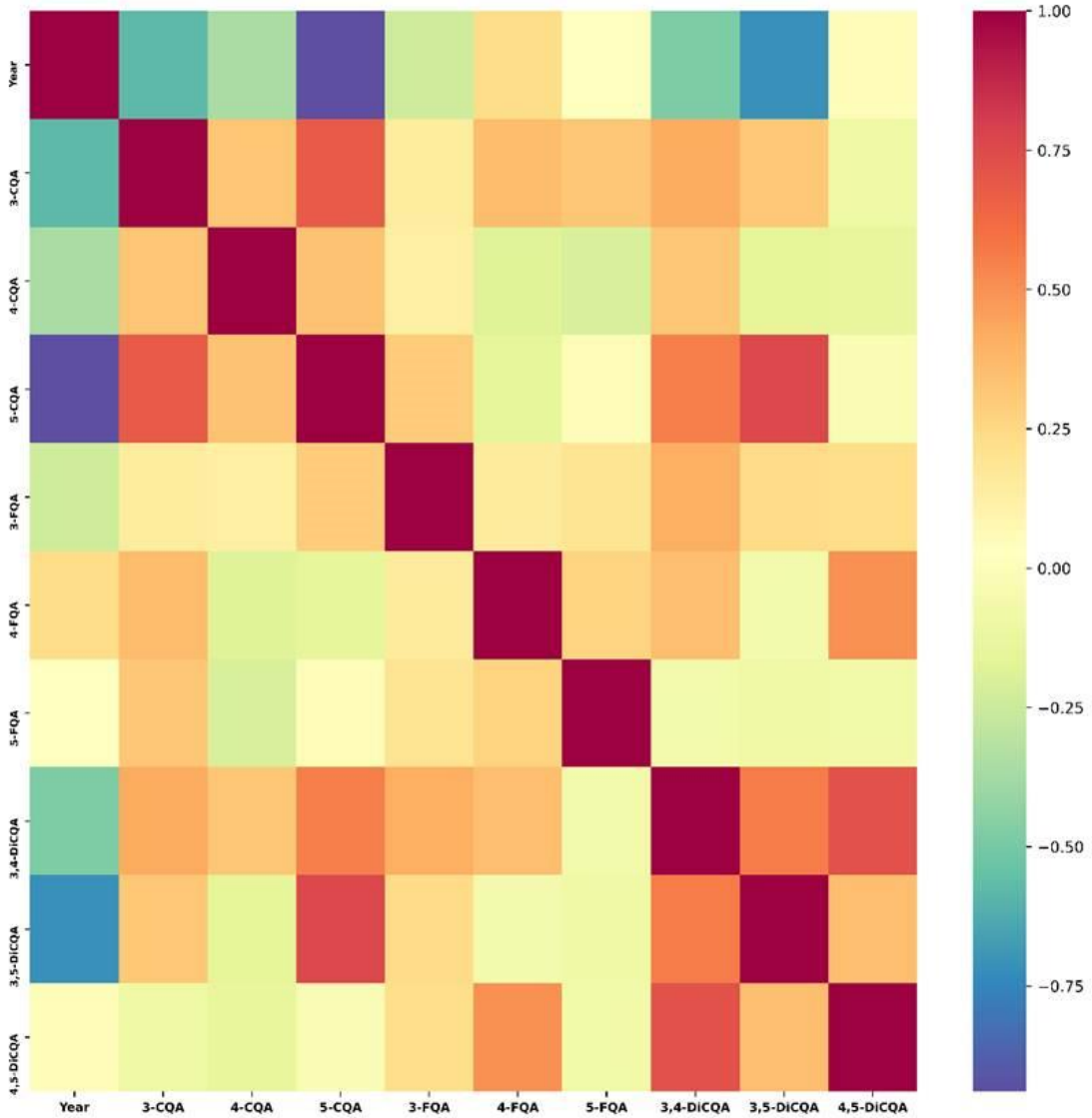


Table 6 Pearson correlation coefficient (r = 1) of Robusta CGAs.

Robusta	Year	3-CQA	4-CQA	5-CQA	3-FQA	4-FQA	5-FQA	3,4-DiCQA	3,5-DiCQA	4,5-DiCQA
Year	1	-0,582	-0,363	-0,939	-0,238	0,230	0,024	-0,481	-0,716	0,047
3-CQA	-0,582	1	0,330	0,687	0,150	0,360	0,322	0,424	0,314	-0,094
4-CQA	-0,363	0,330	1	0,339	0,136	-0,174	-0,210	0,323	-0,164	-0,139
5-CQA	-0,939	0,687	0,339	1	0,301	-0,148	0,048	0,554	0,758	-0,026
3-FQA	-0,238	0,150	0,136	0,301	1	0,155	0,197	0,414	0,241	0,220
4-FQA	0,230	0,360	-0,174	-0,148	0,155	1	0,270	0,354	-0,056	0,505
5-FQA	0,024	0,322	-0,210	0,048	0,197	0,270	1	-0,061	-0,091	-0,070
3,4-DiCQA	-0,481	0,424	0,323	0,554	0,414	0,354	-0,061	1	0,566	0,721
3,5-DiCQA	-0,716	0,314	-0,164	0,758	0,241	-0,056	-0,091	0,566	1	0,351
4,5-DiCQA	0,047	-0,094	-0,139	-0,026	0,220	0,505	-0,070	0,721	0,351	1

Figure 19 Heatmap from Pearson correlation coefficient of green Robusta CGA 2016 to 2020



3.3.8 Antioxidant estimation by FRAP assay

Reactive oxygen species (ROS) accumulation during oxidative stress usually leads to disintegration of cell membranes and structures, through which numerous enzymes via their activities can basically bind with their substrates, which ultimately results in the formation of compounds that affect both the colour and cup quality of coffee (Scheidig et al., 2007). In view of this, the antioxidant capacity of the green coffee beans stored from 2016 through 2020 were determined by FRAP assay. As stated above, 5-CQA is the major antioxidant constituent in the coffee bean, and it decreases with time during storage, due to radical scavenging activities involving lipid and protein oxidation (Rendon et al, 2014). As a result of this, it is expected that a decrease in 5-CQA components of the beans will leads to decrease in the antioxidant capacity of the green and roasted coffee bean. Figure 20 represents the gradually depletion of the antioxidant contents of both Arabica and Robusta coffees during storage over time. It could be suggested that the antioxidant contents of the green coffee decreases as each year passes by.

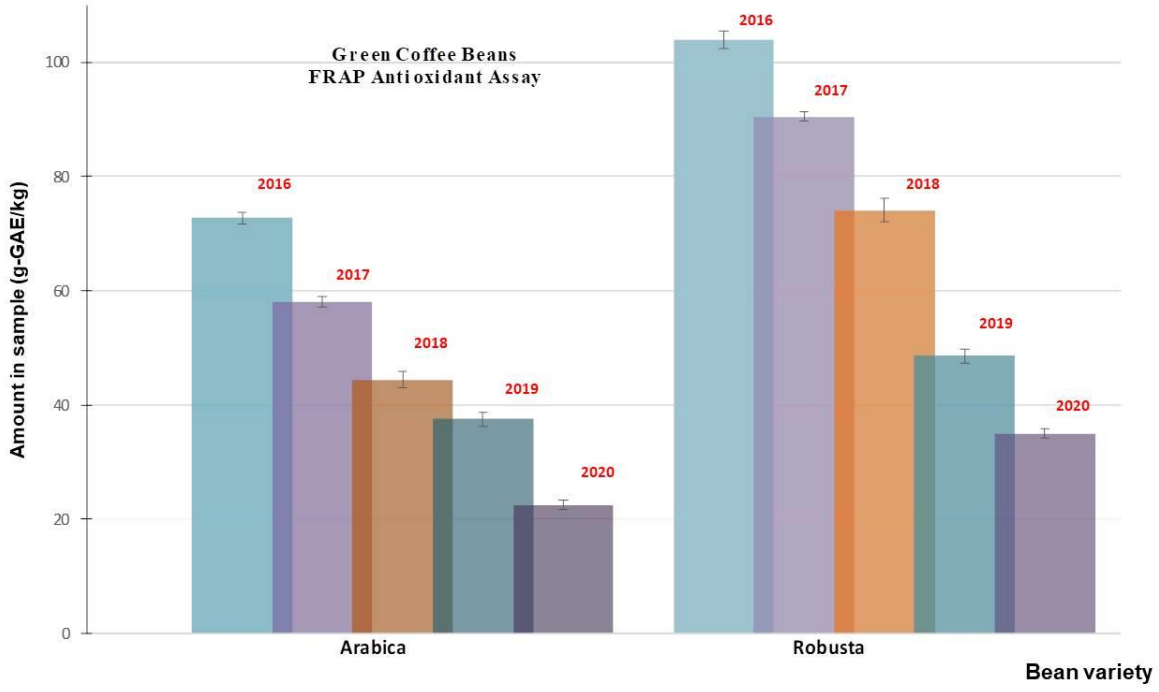


Figure 20 Antioxidant capacity based on FRAP assay (g-GAE/kg) of stored green Arabica and Robusta coffee (2016 to 2020).

Furthermore, Figure 21 shows the antioxidant capacity of the roasted coffee beans of the old (1990 & 2000) coffees estimated from year 2016 through 2020. A steady decrease was observed over time in all the coffees (both Arabica and Robusta coffee beans).

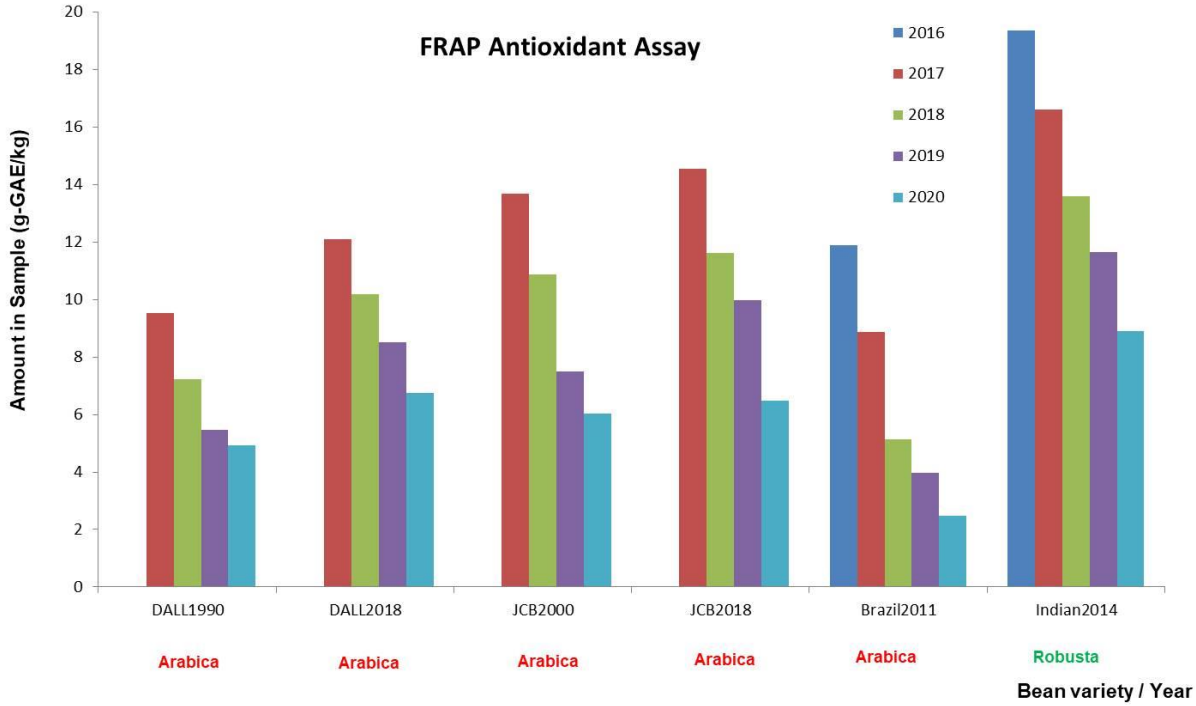


Figure 21 Antioxidant capacity based on FRAP assay (g-GAE/kg) of stored roasted coffees analysed from year 2016 - 2020. DALL, JCB and Brazil are Arabica while Indian is Robusta.

3.3.9 Triacylglycerols (TAGs) analysis

The total lipids content of the green coffee beans was extracted using an optimized Soxhlet extraction in continuous mode for 12 h with heptane as described in section 2.3 of Chapter 4 of this thesis. The oil constituents were separated and analysed using LC-MS methods described in sections 2.5 – 2.6 of Chapter 4 and the TAGs identification was carried using the LC-ESI-MS^n method with the details provided in section 2.7 of Chapter 4. From the identification, 26 TAGs were identified (according to Table 3 of Chapter 4), and 20 of the identified TAGs with significant quantities present in all the green beans were explored for this ageing study. Table 7 depicts mean $\pm$ SD of the TAGs extracted from green coffee bean from

2016 to 2020. From the computed mean values, it was observed that nearly all the Arabica green beans TAGs were higher than the Robusta TAGs.

Furthermore, the amount of each TAG decreases as each year passes by from 2016 through 2020. This phenomenon would be as a result of lipids degradation by lipolysis and oxidation processes as previously observed by Vila et al (2005). In addition, Ortala identified the impact of storage time on bean lipids and highlighted that storage temperature is the main cause of hydroperoxides in low-moisture coffee beans (Ortala et al, 1998). Although, coffee lipid degradation usually occurs by two different concurrent mechanisms, acyl-glycerol hydrolysis and oxidation, while lipolysis releases free fatty acids, which are more susceptible to oxidation than esterified fatty acids, predominantly the long-chain unsaturated fatty acids. The lipids auto-oxidation occurs via a complex process involving several reactions, but another study has shown that the rate of autooxidation is mostly affected by the fatty acids composition, degree of unsaturation, presence and activity of antioxidants in food substrates and the oxygen partial pressure (Belitz et al, 2004). This suggests that the degradation observed in the TAGs are as a result of these numerous processes taking place during the green beans storage over time. Further studies show that coffees stored under vacuum and low oxygen pressure confirmed lipid oxidation process due to the initial presence of free radicals generated during post-harvesting procedures (Baesso et al, 1990; Morrice et al, 1993).

Moreover, Figure 22 represents the boxplots of the TAGs degradation patterns from 2016 through 2020. Most significantly, the plots indicate that the TAGs decreased across the years till year 2020 except for TAG PtAL in Arabica and LLLn (both Arabica and Robusta), where it was observed that there was no significant change in the two TAGs. In addition, the rate of degradation in the two coffee varieties was observed and the TAGs values show that there no significant differences between Arabica and Robusta ($p < 0.05$) but the Arabica TAGs (left side of the plot) were observed to be more stable than the Robusta TAGs (right side of the plot).

From these observations, it could be said that the TAGs extracted from stored green coffee beans for a long time decreases insignificantly over time irrespective of the bean variety and lipids fractional constituents.

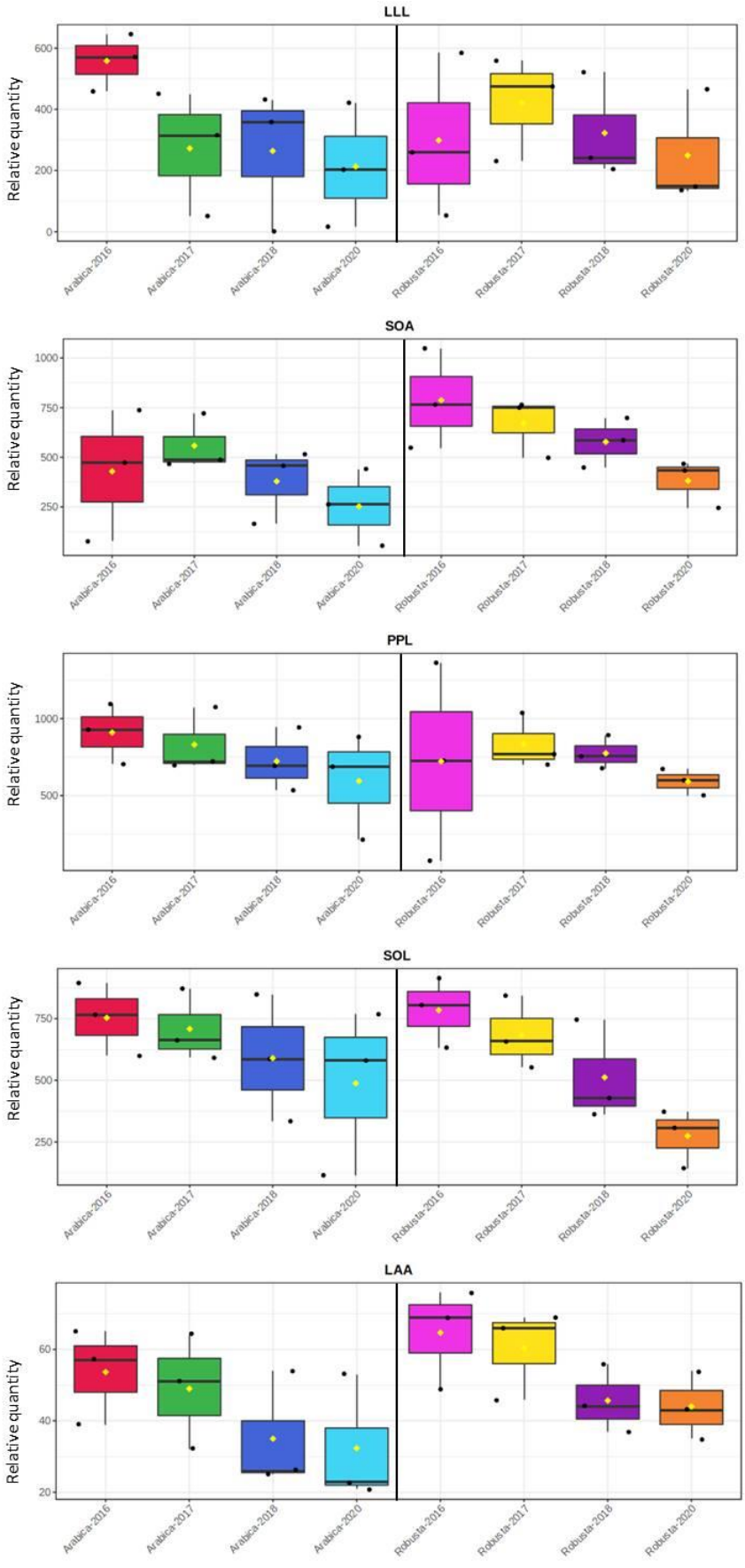
Table 7 Mean+s.d of the TAGs extracted from green coffee bean from 2016 to 2020.

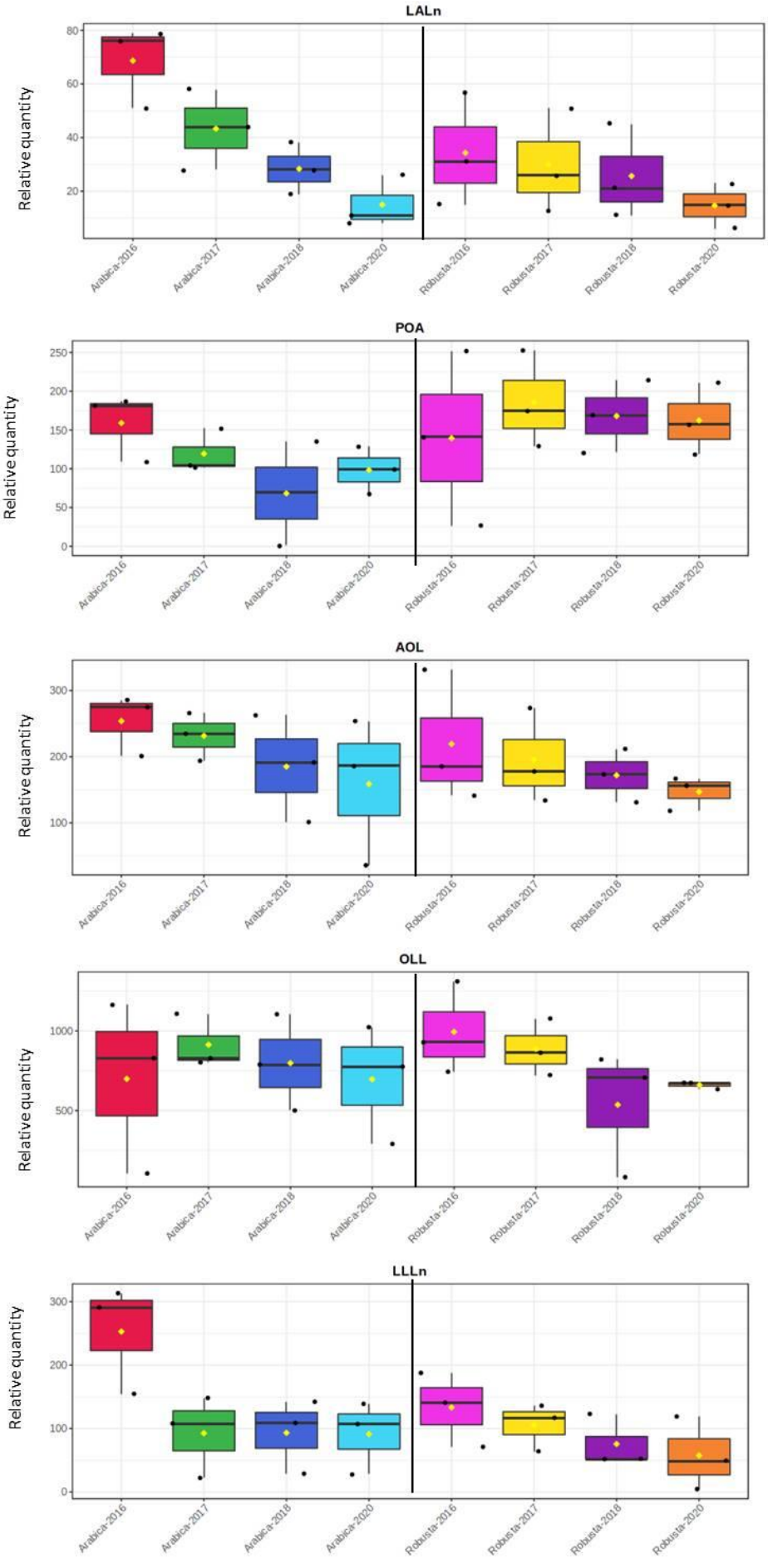
TAGs (Mean+s.d)	(n=10)	Arabica 2016	Arabica 2017	Arabica 2018	Arabica 2020	(n=10)	Robusta 2016	Robusta 2017	Robusta 2018	Robusta 2020
LALn		0.68±0.2	0.43±0.1	0.28±0.1	0.15±0.1		0.35±0.2	0.30±0.2	0.26±0.2	0.13±0.1
LLL		5.58±0.9	4.25±4.0	2.97±1.7	2.13±2.0		4.58±1.7	4.21±1.7	3.23±1.7	2.50±1.9
LLLn		2.53±0.9	1.59±1.6	0.93±0.6	0.91±0.6		1.33±0.6	1.06±0.4	0.76±0.4	0.73±0.4
LLAn		0.26±0.0	0.21±0.1	0.17±0.1	0.11±0.1		0.44±0.2	0.35±0.1	0.21±0.0	0.12±0.0
PLLn		3.59±0.9	2.02±2.8	2.13±1.0	1.22±1.9		4.21±1.2	3.62±0.8	3.21±0.8	2.21±2.5
PLL		12.43±2.9	9.02±9.6	9.64±7.4	8.29±8.8		11.43±1.7	10+76±1.1	10.07±0.7	9.16±8.6
OLL		10.14±1.7	8.13±8.7	7.98±5.0	6.97±6.7		9.95±2.9	8.88±1.8	7.83±0.7	6.61±6.2
PLnLn		0.43±0.1	0.37±0.1	0.33±0.1	0.25±0.1		0.39±0.1	0.36±0.1	0.31±0.1	0.21±0.1
OLLn		1.26±0.5	0.79±0.4	0.40±0.2	0.22±0.2		0.74±0.5	0.51±0.2	0.36±0.1	0.20±0.1
POO		5.55±1.2	4.08±4.3	4.26±2.4	3.36±3.9		6.07±1.4	5.12±1.5	4.50±1.6	3.51±0.3
POL		2.96±1.1	2.32±2.5	2.10±1.3	1.73±1.3		2.77±0.9	2.42±0.7	1.92±0.6	1.61±1.7
PPL		9.09±2.0	6.31±6.1	7.23±5.1	5.94±5.4		9.56±3.5	8.36±1.8	7.74±1.1	5.91±5.9
PtAL		0.38±0.0	0.29±0.0	0.27±0.0	0.17±0.0		1.11±0.4	0.73±0.1	0.58±0.1	0.26±0.1
AOL		2.54±0.5	1.32±1.4	1.85±1.8	1.59±1.1		2.19±1.0	1.95±0.7	1.72±0.4	1.47±1.3
PLA		3.35±0.9	2.70±2.6	2.25±1.0	1.89±1.3		4.38±1.2	3.97±1.0	3.12±0.4	2.07±2.8
SOL		7.53±1.5	5.08±5.5	5.90±3.6	4.88±4.4		7.84±1.4	6.85±1.5	5.12±2.0	2.74±1.2
POA		1.59±0.4	1.19±1.3	1.01±0.3	0.98±0.3		2.17±0.7	1.86±0.6	1.68±0.5	1.62±1.5
SOA		6.60±1.6	4.58±4.4	3.79±1.9	2.53±2.9		7.87±2.5	6.70±1.5	5.77±1.3	3.82±2.2
LAA		0.54±0.1	0.49±0.2	0.35±0.2	0.32±0.2		0.65±0.1	0.60±0.1	0.46±0.1	0.44±0.1
SAAn		1.83±0.6	1.33±1.3	0.81±0.5	0.62±0.4		1.78±0.7	1.48±0.4	1.10±0.2	0.81±0.3

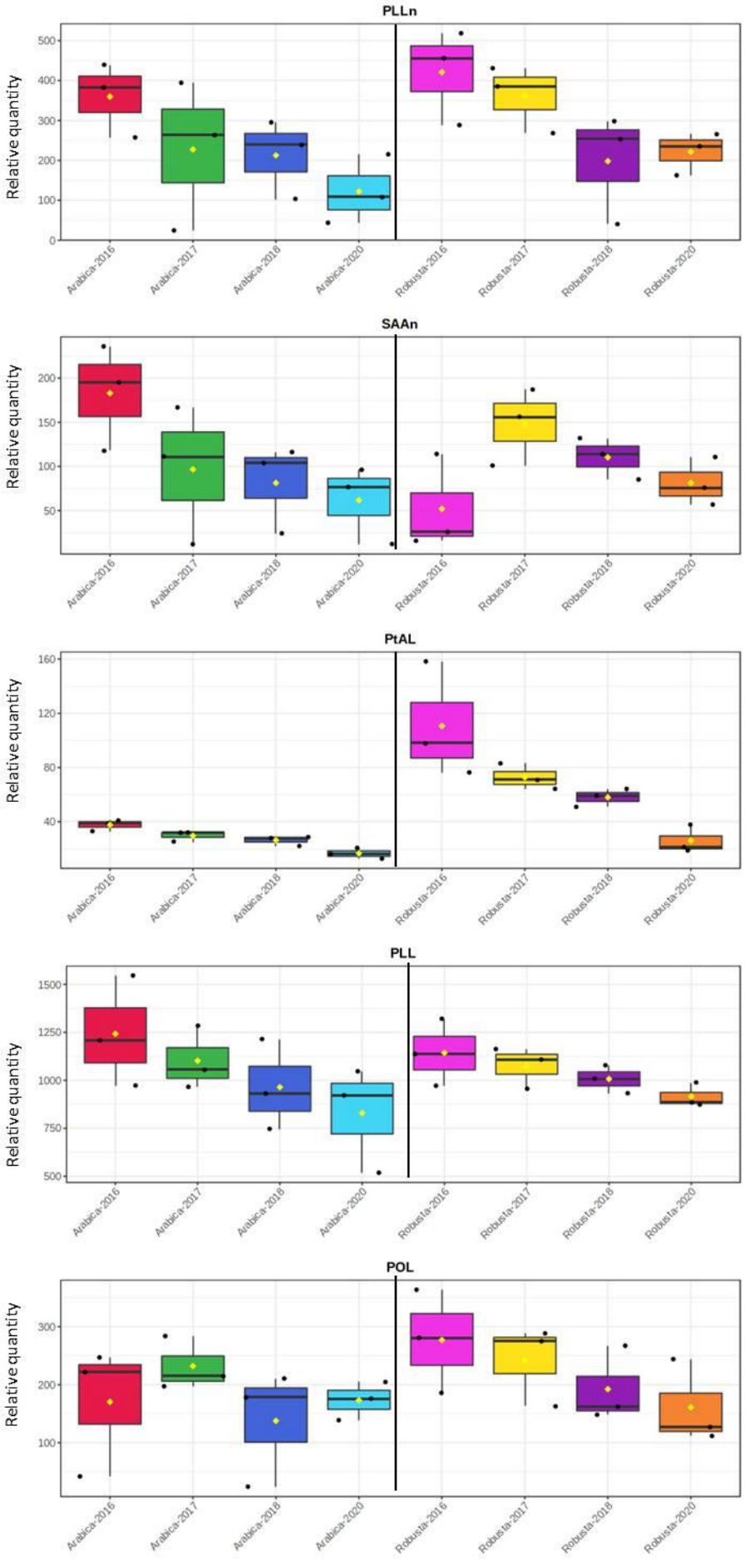
Table 8 P-values ($p < 0.05$)

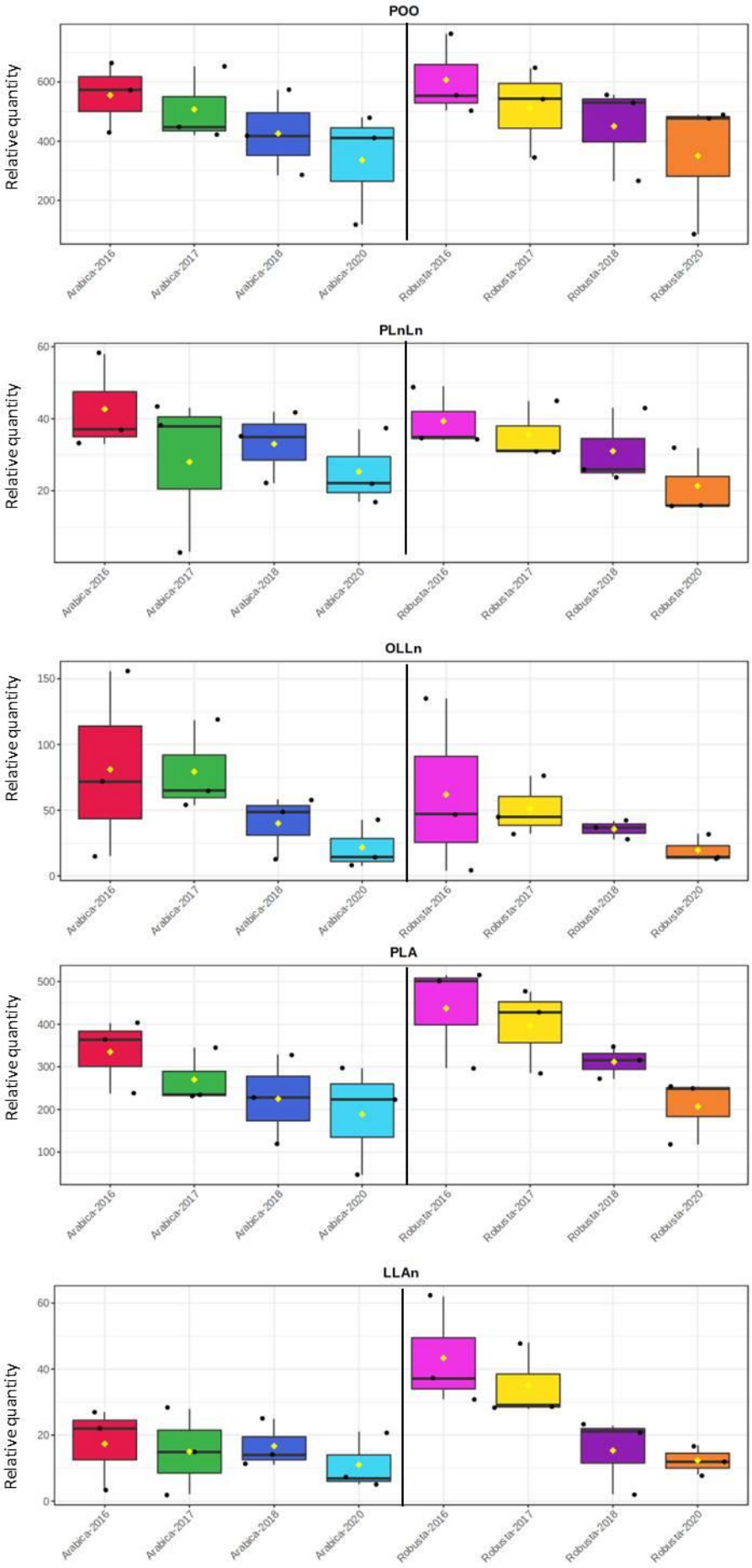
Arabica	2016	2017	2018	2020	Robusta	2016	2017	2018	2020
2016	1	0.588	0.304	0.096	2016	1	0.591	0.311	0.307
2017	0.588	1	0.552	0.189	2017	0.591	1	0.512	0.946
2018	0.304	0.552	1	0.386	2018	0.311	0.512	1	0.503
2020	0.096	0.189	0.386	1	2020	0.307	0.946	0.503	1

Figure 22 Boxplots of TAGs extracted from Arabica and Robusta green beans; with Arabica 2016 – 2020 on the left, while Robusta 2016 – 2020 on the right.




LALn
POA
AOL
OLL
LLLn
Relative quantity
Arabica-2016
Arabica-2017
Arabica-2018
Arabica-2020
Robusta-2016
Robusta-2017
Robusta-2018
Robusta-2020


PLLn
SAAn
PtAL
PLL
POL
Relative quantity
Arabica-2016
Arabica-2017
Arabica-2018
Arabica-2020
Robusta-2016
Robusta-2017
Robusta-2018
Robusta-2020


POO
PLnLn
OLLn
PLA
LLAn
Relative quantity
Arabica-2016
Arabica-2017
Arabica-2018
Arabica-2020
Robusta-2016
Robusta-2017
Robusta-2018
Robusta-2020

3.3.9.1 Principal component analysis of green bean TAGs

The PCA plot is an unsupervised dimensionality reduction algorithm that aids data reduction dimensions and shows clusters of samples with a minimum loss of information. Moreover, it explains the variance-covariance structure of a set of variables via linear combinations and it does not discard any samples or characteristics (variables). Alternatively, it reduces the overwhelming number of dimensions by representing the variables on principal components (PCs) based on their linear combinations. It also adopts the directions with the largest variances are the most "significant" (most principal). Figure 23 depicts the PCA plot and important features in the loading plot of the stored Arabica and Robusta green coffee TAGs. In the plot, the PC1 axis is the first principal direction along which the samples show the largest variation, while PC2 axis is the second most important direction and it is orthogonal to PC1 axis. While the corresponding PC plot values are numbers that signify the degree of variances in the data along the principal component. This means that a larger PC value represents a large amount of the variance in the dataset. The PC1 shows a variance of 56.6% and 15.8% for the PC2. This degree of variation indicates that there is a significant variation between the year 2016 starting coffee samples and the 2020 coffee samples. Likewise, some degree of variations was observed when Arabica and Robusta coffee samples were compared. This is shown in the loading plot where Robusta 2016 TAGs were observed to be almost completely different from the Arabica 2016 TAGs with respects to the colour codes in the loading plot. Usually, the loading plot identifies which variables have the largest effect on each component. The plot loadings range from 0.1 to 1. Moreover, loadings close to 1 indicate that the variable strongly influences the component, while loadings close to 0 indicate that the variable has a weak influence on the component. In view of this, PLL, SOL, POL and POO have a weak influence on the components; while PtAL, LLAn, PLA, SOA and LAA have strong influence on the components.

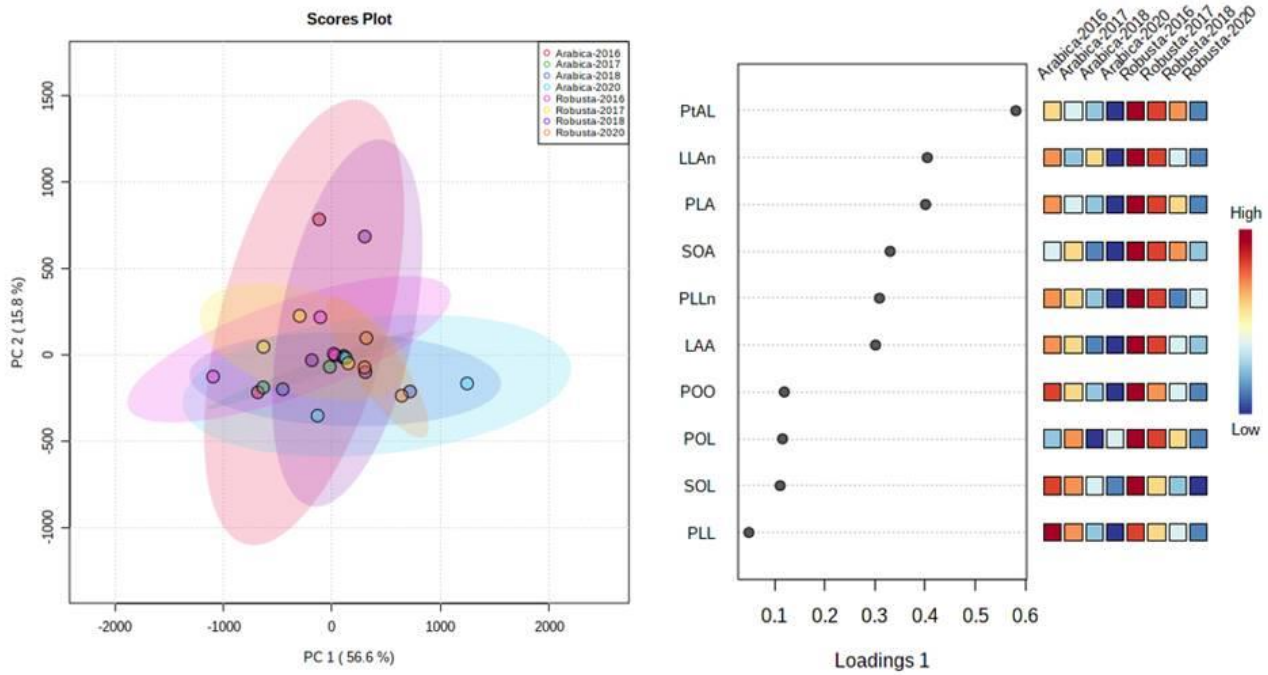


Figure 23 PCA and loading plots of Arabica against Robusta TAGs from 2016 to 2020.

3.3.9.2 Statistical significance in TAGs degradation

The Tables 9 and 10 represent the Pearson correlation coefficient values of TAGs extracted from Arabica and Robusta green coffee beans stored from 2016 till 2020 respectively, while Figures 24 and 25 show the heatmaps with the correlation between the individual TAGs and the years from 2016 and 2020. Likewise, from the heatmaps, it was observed that as the period of storage moves from 2016 through 2020 the TAGs quantity decreases, meaning that the TAGs are correlated as each year passes by.

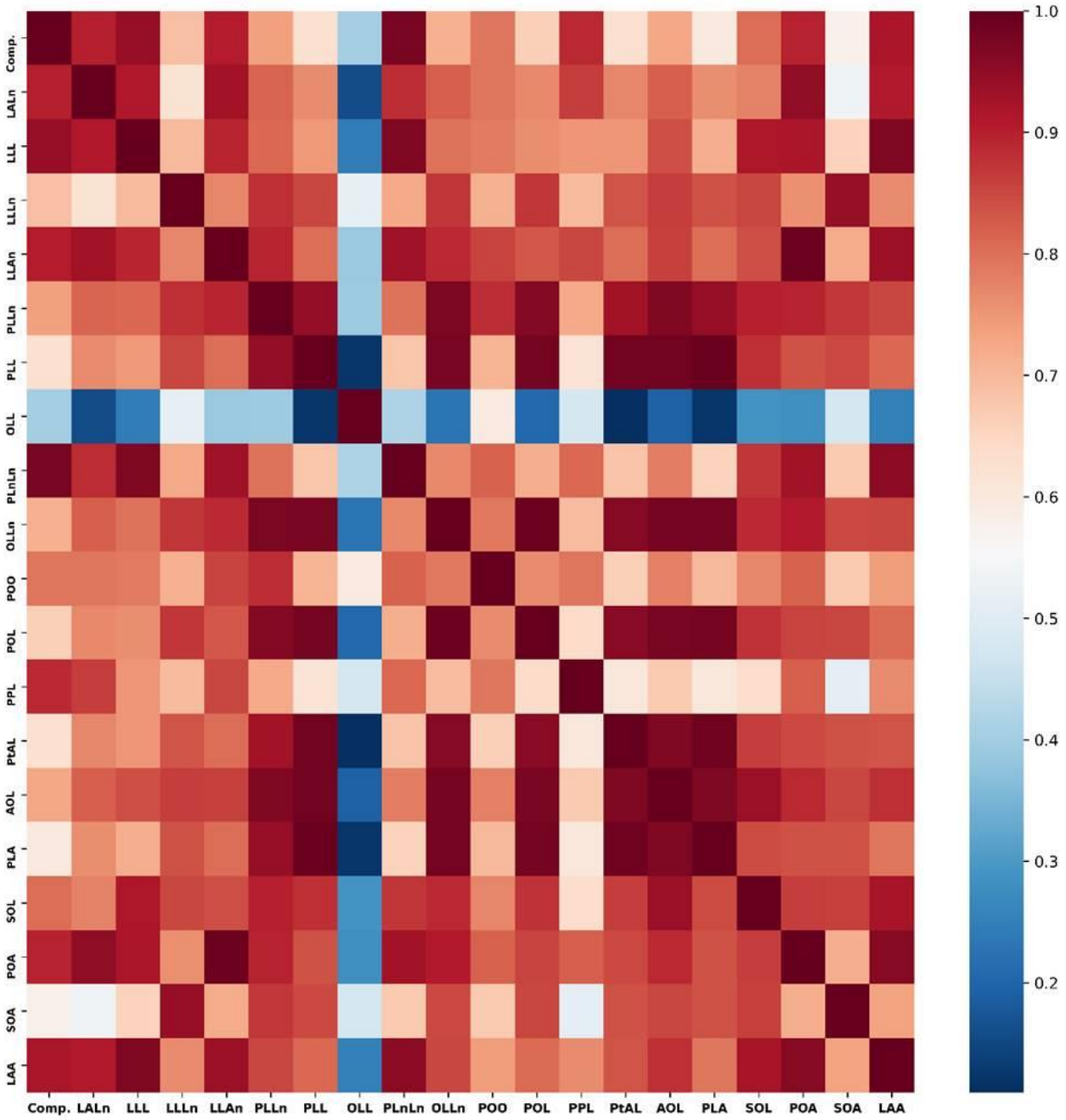
Table 9 Pearson correlation coefficient of TAGs extracted from Arabica green coffee beans stored from 2016 till 2020.

	LALn	LLL	LLLn	LLAn	PLLn	PLL	OLL	PLnLn	OLLn	POO	POL	PPL	PtAL	AOL	PLA	SOL	POA	SOA	LAA	SAAn
LALn	1	0,900	0,943	0,689	0,903	0,739	0,625	0,405	0,978	0,715	0,794	0,662	0,886	0,627	0,728	0,599	0,805	0,898	0,577	0,918
LLL	0,900	1	0,912	0,618	0,928	0,814	0,765	0,157	0,884	0,821	0,794	0,768	0,862	0,771	0,822	0,760	0,774	0,950	0,534	0,907
LLLn	0,943	0,912	1	0,695	0,892	0,812	0,748	0,242	0,968	0,797	0,786	0,761	0,751	0,753	0,842	0,718	0,914	0,920	0,659	0,966
LLAn	0,689	0,618	0,695	1	0,772	0,880	0,853	0,514	0,725	0,871	0,714	0,873	0,697	0,835	0,864	0,837	0,853	0,759	0,944	0,767
PLLn	0,903	0,928	0,892	0,772	1	0,896	0,804	0,391	0,932	0,887	0,855	0,830	0,853	0,804	0,859	0,805	0,841	0,990	0,721	0,936
PLL	0,739	0,814	0,812	0,880	0,896	1	0,945	0,393	0,796	0,973	0,882	0,963	0,725	0,928	0,967	0,944	0,902	0,897	0,871	0,851
OLL	0,625	0,765	0,748	0,853	0,804	0,945	1	0,122	0,680	0,978	0,707	0,981	0,617	0,983	0,985	0,995	0,880	0,838	0,848	0,810
PLnLn	0,405	0,157	0,242	0,514	0,391	0,393	0,122	1	0,418	0,230	0,595	0,207	0,477	0,110	0,192	0,123	0,290	0,283	0,475	0,252
OLLn	0,978	0,884	0,968	0,725	0,932	0,796	0,680	0,418	1	0,767	0,819	0,716	0,811	0,682	0,782	0,657	0,873	0,929	0,672	0,952
POO	0,715	0,821	0,797	0,871	0,887	0,973	0,978	0,230	0,767	1	0,789	0,992	0,694	0,960	0,980	0,980	0,887	0,906	0,849	0,853
POL	0,794	0,794	0,786	0,714	0,855	0,882	0,707	0,595	0,819	0,789	1	0,766	0,793	0,659	0,778	0,699	0,772	0,817	0,671	0,740
PPL	0,662	0,768	0,761	0,873	0,830	0,963	0,981	0,207	0,716	0,992	0,766	1	0,640	0,956	0,976	0,981	0,877	0,855	0,851	0,807
PtAL	0,886	0,862	0,751	0,697	0,853	0,725	0,617	0,477	0,811	0,694	0,793	0,640	1	0,606	0,671	0,607	0,635	0,821	0,512	0,764
AOL	0,627	0,771	0,753	0,835	0,804	0,928	0,983	0,110	0,682	0,960	0,659	0,956	0,606	1	0,970	0,986	0,863	0,847	0,840	0,836
PLA	0,728	0,822	0,842	0,864	0,859	0,967	0,985	0,192	0,782	0,980	0,778	0,976	0,671	0,970	1	0,972	0,941	0,889	0,854	0,882
SOL	0,599	0,760	0,718	0,837	0,805	0,944	0,995	0,123	0,657	0,980	0,699	0,981	0,607	0,986	0,972	1	0,844	0,838	0,837	0,792
POA	0,805	0,774	0,914	0,853	0,841	0,902	0,880	0,290	0,873	0,887	0,772	0,877	0,635	0,863	0,941	0,844	1	0,863	0,860	0,922
SOA	0,898	0,950	0,920	0,759	0,990	0,897	0,838	0,283	0,929	0,906	0,817	0,855	0,821	0,847	0,889	0,838	0,863	1	0,717	0,961
LAA	0,577	0,534	0,659	0,944	0,721	0,871	0,848	0,475	0,672	0,849	0,671	0,851	0,512	0,840	0,854	0,837	0,860	0,717	1	0,736
SAAn	0,918	0,907	0,966	0,767	0,936	0,851	0,810	0,252	0,952	0,853	0,740	0,807	0,764	0,836	0,882	0,792	0,922	0,961	0,736	1

Table 10 Pearson correlation coefficient of TAGs extracted from Robusta green coffee beans stored from 2016 till 2020.

	LALn	LLL	LLLn	LLAn	PLLn	PLL	OLL	PLnLn	OLLn	POO	POL	PPL	PtAL	AOL	PLA	SOL	POA	SOA	LAA	SAAn
LALn	1	0,882	0,744	0,303	0,791	0,612	0,438	0,187	0,335	0,570	0,727	0,306	0,421	-0,156	0,692	0,494	0,901	0,678	0,602	0,561
LLL	0,882	1	0,928	0,483	0,880	0,676	0,606	0,227	0,594	0,677	0,932	0,531	0,556	0,076	0,633	0,545	0,911	0,764	0,630	0,764
LLLn	0,744	0,928	1	0,645	0,898	0,794	0,762	0,340	0,793	0,723	0,977	0,697	0,720	0,317	0,670	0,577	0,896	0,846	0,705	0,889
LLAn	0,303	0,483	0,645	1	0,758	0,865	0,933	0,782	0,903	0,605	0,707	0,865	0,948	0,811	0,815	0,823	0,391	0,825	0,763	0,878
PLLn	0,791	0,880	0,898	0,758	1	0,869	0,836	0,493	0,791	0,677	0,916	0,740	0,845	0,382	0,865	0,729	0,815	0,903	0,733	0,890
PLL	0,612	0,676	0,794	0,865	0,869	1	0,957	0,767	0,881	0,800	0,835	0,871	0,867	0,663	0,903	0,826	0,692	0,967	0,890	0,934
OLL	0,438	0,606	0,762	0,933	0,836	0,957	1	0,788	0,955	0,744	0,815	0,946	0,905	0,808	0,832	0,823	0,576	0,915	0,842	0,954
PLnLn	0,187	0,227	0,340	0,782	0,493	0,767	0,788	1	0,718	0,744	0,458	0,827	0,689	0,825	0,671	0,909	0,265	0,748	0,809	0,712
OLLn	0,335	0,594	0,793	0,903	0,791	0,881	0,955	0,718	1	0,751	0,837	0,967	0,900	0,817	0,686	0,754	0,551	0,883	0,747	0,961
POO	0,570	0,677	0,723	0,605	0,677	0,800	0,744	0,744	0,751	1	0,805	0,808	0,595	0,521	0,584	0,821	0,714	0,879	0,839	0,844
POL	0,727	0,932	0,977	0,707	0,916	0,835	0,815	0,458	0,837	0,805	1	0,776	0,748	0,406	0,696	0,680	0,853	0,898	0,746	0,934
PPL	0,306	0,531	0,697	0,865	0,740	0,871	0,946	0,827	0,967	0,808	0,776	1	0,843	0,859	0,667	0,827	0,508	0,883	0,777	0,934
PtAL	0,421	0,556	0,720	0,948	0,845	0,867	0,905	0,689	0,900	0,595	0,748	0,843	1	0,694	0,829	0,780	0,507	0,859	0,703	0,880
AOL	-0,156	0,076	0,317	0,811	0,382	0,663	0,808	0,825	0,817	0,521	0,406	0,859	0,694	1	0,466	0,649	0,050	0,589	0,579	0,683
PLA	0,692	0,633	0,670	0,815	0,865	0,903	0,832	0,671	0,686	0,584	0,696	0,667	0,829	0,466	1	0,797	0,637	0,854	0,834	0,784
SOL	0,494	0,545	0,577	0,823	0,729	0,826	0,823	0,909	0,754	0,821	0,680	0,827	0,780	0,649	0,797	1	0,531	0,862	0,893	0,830
POA	0,901	0,911	0,896	0,391	0,815	0,692	0,576	0,265	0,551	0,714	0,853	0,508	0,507	0,050	0,637	0,531	1	0,763	0,710	0,725
SOA	0,678	0,764	0,846	0,825	0,903	0,967	0,915	0,748	0,883	0,879	0,898	0,883	0,859	0,589	0,854	0,862	0,763	1	0,876	0,958
LAA	0,602	0,630	0,705	0,763	0,733	0,890	0,842	0,809	0,747	0,839	0,746	0,777	0,703	0,579	0,834	0,893	0,710	0,876	1	0,855
SAAn	0,561	0,764	0,889	0,878	0,890	0,934	0,954	0,712	0,961	0,844	0,934	0,934	0,880	0,683	0,784	0,830	0,725	0,958	0,855	1

Figure 24 Heatmap of Arabica green coffee TAGs across the years from 2016 to 2020.

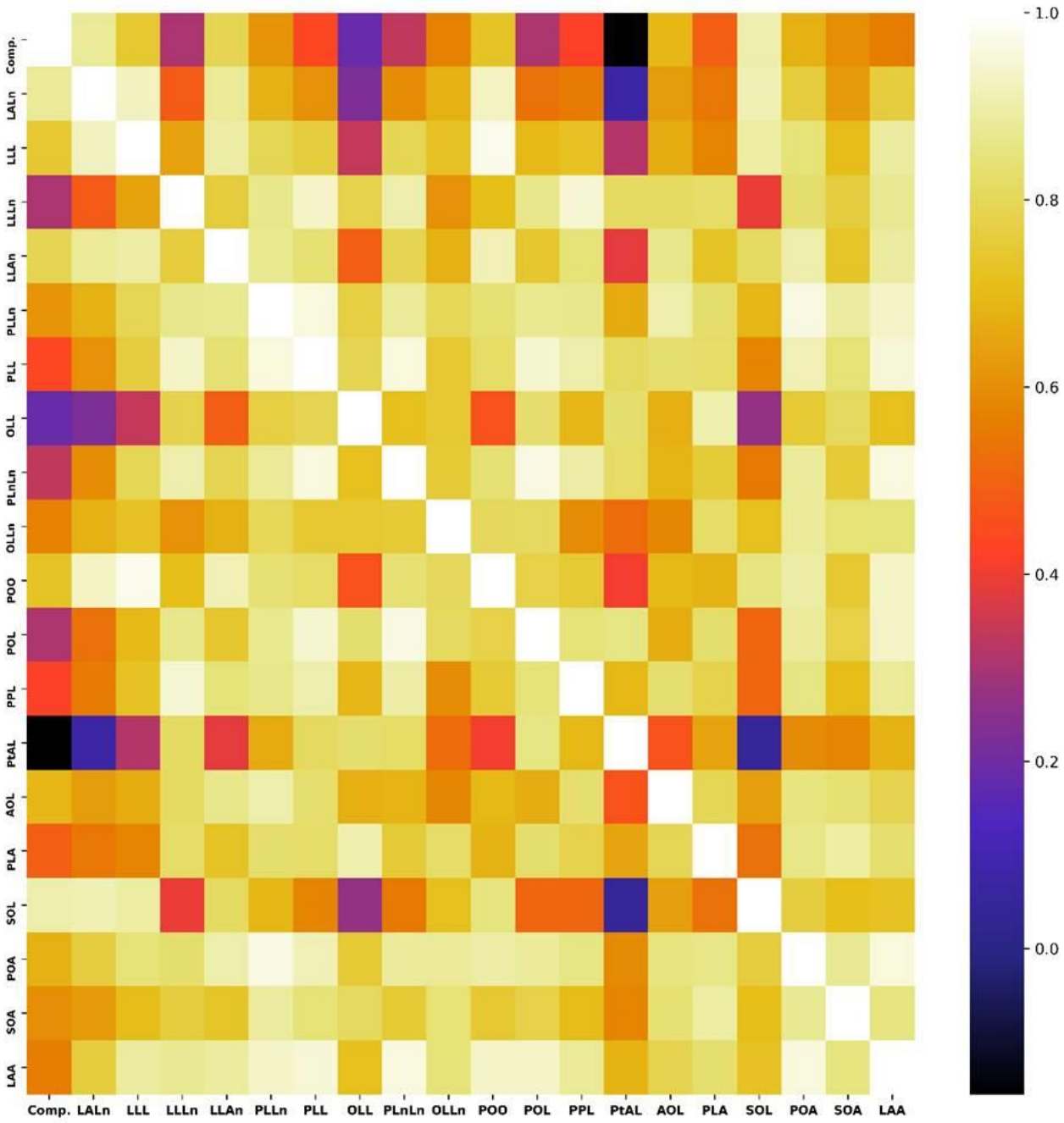


3.3.9.3 Heatmap of the coffee TAG

Figures 24 and 25 represent the heatmaps of the pearson correlation coeffiecient values of the TAGs from Arabica and Robusta in cell colour changes as the years goes by from 2016 to 2020. From the heatmap plots for both the Arabica and Robusta coffees, it was observed that as the

years passes by from 2016 to 2020 the amount of the TAGs constituents in the two coffees were gradually depleted. In addition, it was observed that TAGs in Arabica were more stable when compared to Robusta TAGs.

Figure 25 Heatmap of Robusta green coffee TAGs across the years from 2016 to 2020.



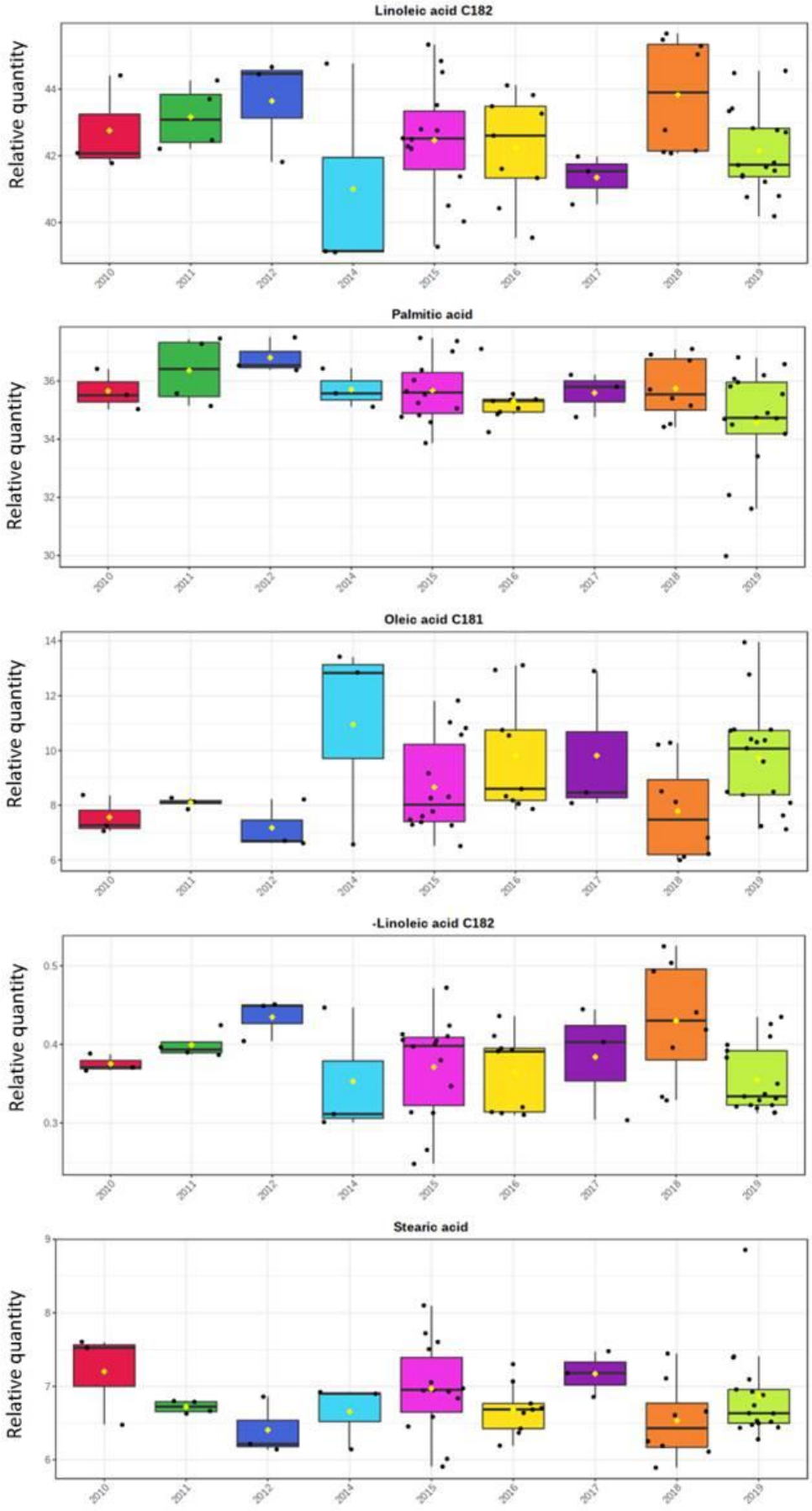
3.3.10 FAME analysis of the ageing coffee beans

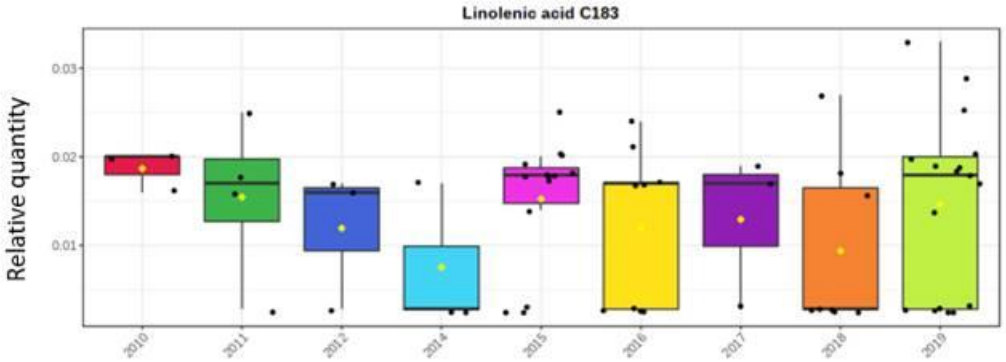
The coffee oil extracted from the green beans from year 2010 till 2019 were analysed to profile the fatty acids constituents. The extracted coffee oil was derivatized by fatty acid methyl ester (FAME) procedure as explained in section 2.5 of Chapter 5. Thereafter, the derivatized samples were analysed on GC and spectra were generated for each of the samples which were further analysed using Lab Solutions software (Shimadzu Incorporations). Each peak in the chromatogram corresponds to a specific fatty acid as listed in Table 2 (Chapter 5) with different peak intensity and retention time. The fatty acid profiles were explored to investigate the ageing process and rate of degradation in the green coffee beans.

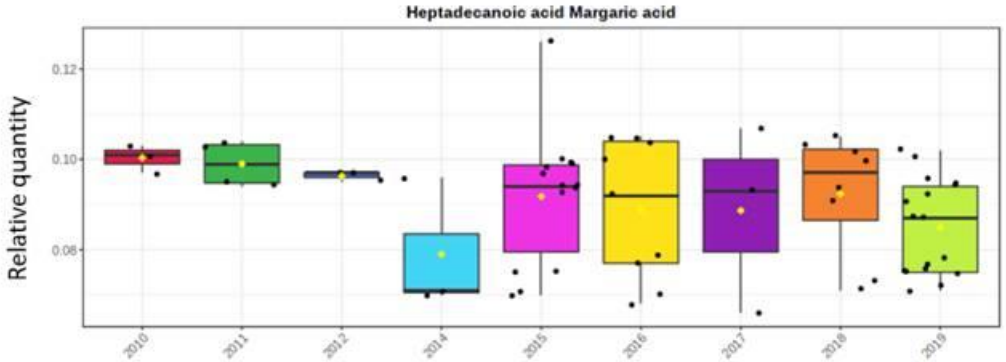
3.3.10.1 Box-plot of fatty acids

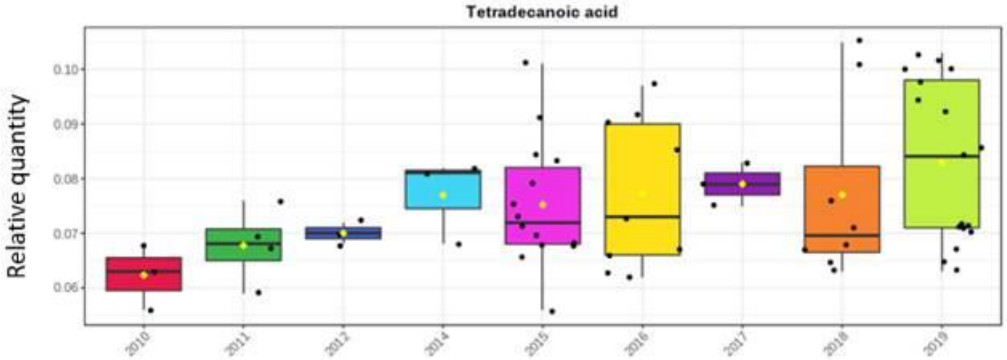
Figure 26 represents the box-plots of the fatty acids extracted from the green Arabica and Robusta coffee stored from the year 2010 to 2019. The variations among the stored coffees fatty acids in Arabica and Robusta coffees were represented for each year as a box on the plot and the palmitic acid was observed to be very stable over time irrespective of the bean variety and year. Oleic acid was observed to be more depleted in the year 2010 to 2012 when compared to other fatty acids, while other fatty acids were observed to be depleted over time.

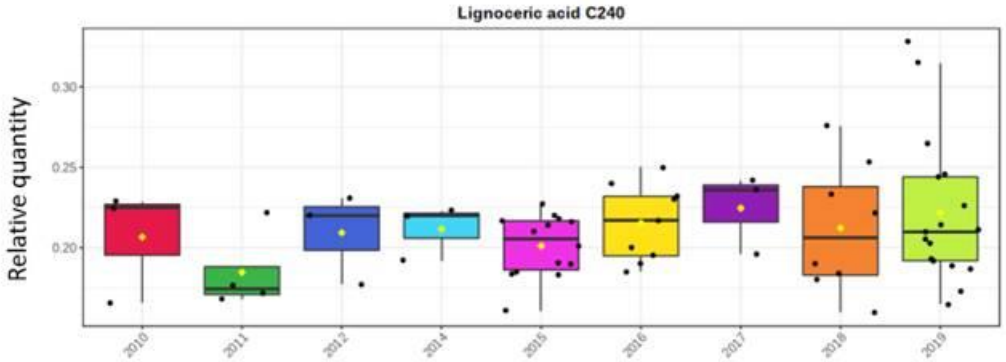
Figure 26 Boxplots of the fatty acids extracted from green coffee beans from 2010 to 2020.

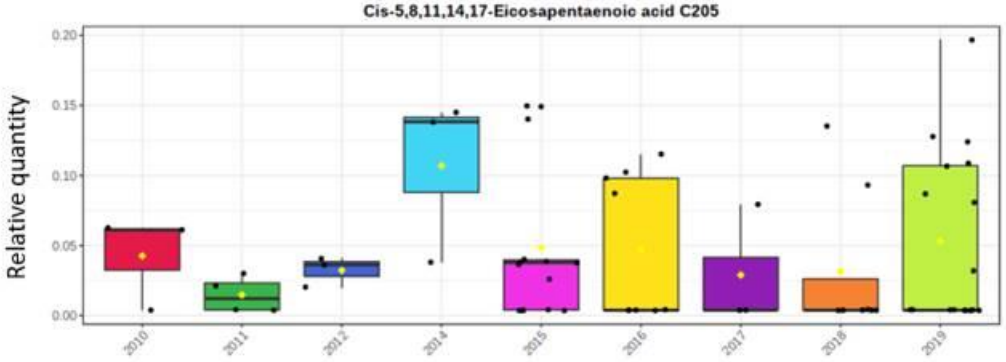



Linolenic acid C183
Relative quantity
0.03
0.02
0.01
2010
2011
2012
2014
2015
2016
2017
2018
2019


Heptadecanoic acid Margaric acid
Relative quantity
0.12
0.10
0.08
2010
2011
2012
2014
2015
2016
2017
2018
2019


Tetradecanoic acid
Relative quantity
0.10
0.09
0.08
0.07
0.06
2010
2011
2012
2014
2015
2016
2017
2018
2019


Lignoceric acid C240
Relative quantity
0.30
0.25
0.20
2010
2011
2012
2014
2015
2016
2017
2018
2019


Cis-5,8,11,14,17-Eicosapentaenoic acid C205
Relative quantity
0.20
0.15
0.10
0.05
0.00
2010
2011
2012
2014
2015
2016
2017
2018
2019

3.3.10.2 Partial least square discriminant analysis of fatty acids in green coffee

The PLS-DA was empolyed in examining the differences in the fatty acid profiles of the coffees for each from 2010 to 2019. Figure 27 represents the PLS-DA plot and important features of the fatty acids with the different colours codes. It was observed from the plot (63.7%) that there is no significant differences among the fatty acids constituents of the coffee for the period of storage, except for some of the year 2019 fresh samples with the yellow codes in the plot which were found outside the circles compared to others. Moreover, this observation corresponds to a previous study on fatty acids stored for about 6 months (Vila et al, 2005) where it was reported that there was no significant change in the fatty acid profile as observed for the entire storage duration.

Furthermore, the loading plot with the important features was studied to determine the influence of each of the variable (fatty acid) on the components. From this, the VIP scores on the plot which range from 0 to 4 determine how strong or weak is the influence of the fatty acids on the component. The loadings close to 4 indicate that the variable strongly influences the component, while loadings close to 0 indicate that the variable has a weak influence on the component. In view of this, linoleic, lignoceric, eicosenoic, octanoic, stearic, arachidic and alpha-linoleic acids were observed to have a weak influence on the components; while oleic and palmitic acids have strong influence on the components. In addition, it could be observed from the colour codes that the concentration of the different fatty acids varied significantly across each year under study.

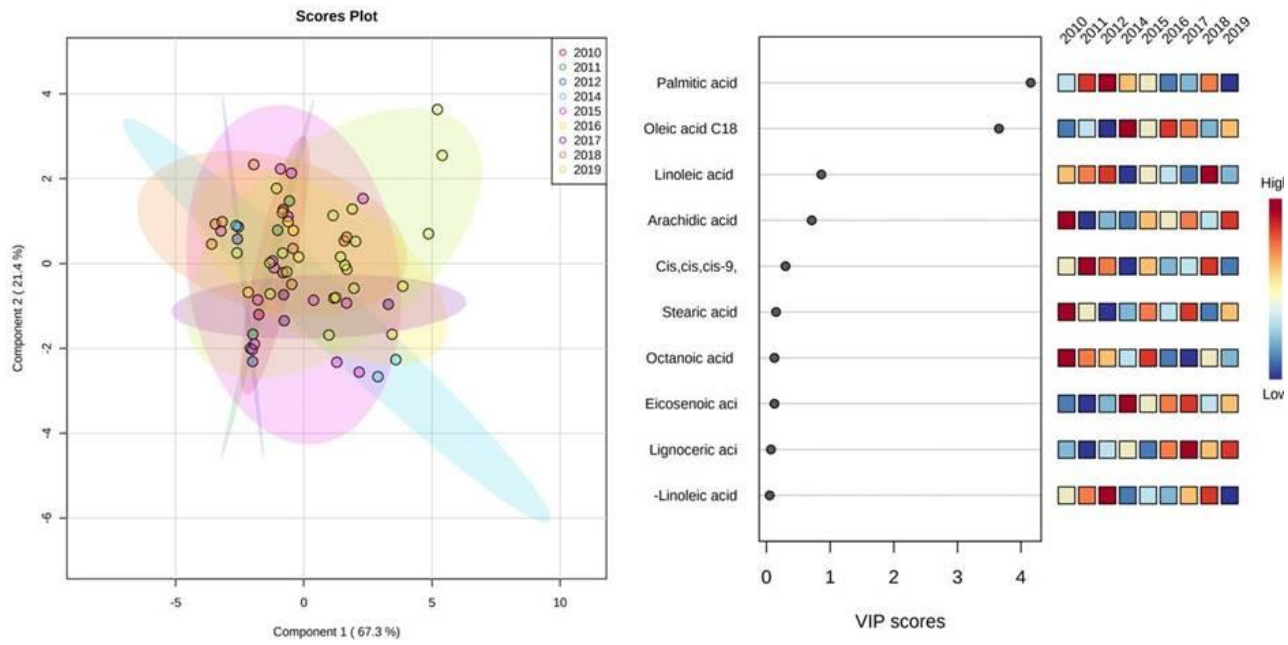


Figure 27 PLS-DA analysis and important features of green coffee fatty acids (2010 – 2020).

The PCA was also explored to investigate differences in the fatty acids' components of the stored coffee. Figure 28 shows the PCA and 3-dimensional (3-D) plots generated from the fatty acid analysed. The PC1 shows a variance of 73.2% and 16.2% for the PC2. From the plot, it was observed that there is no significant difference among all the years under study. Moreover, the 3-D plot was also considered to have a clearer view of the fatty acids pattern with more dimensions. The 3-D plot likewise show a similar pattern observed in the PLS-DA and PCA plots, with no clear separation among the samples analysed in different years, hence there is no significant difference among the coffees with respect to the years when compared.

In addition, the heatmap of the fatty acids was also generated to observe the degradation rate across the years. Figure 29 represents the heatmap of the different fatty acids, which shows that nearly all the fatty acids show a negative correlation across the years from 2010 to 2020. Also, Table 11 in the supplementary information section represents the correlation values of the fatty acids extracted from green beans from 2016 to 2020.

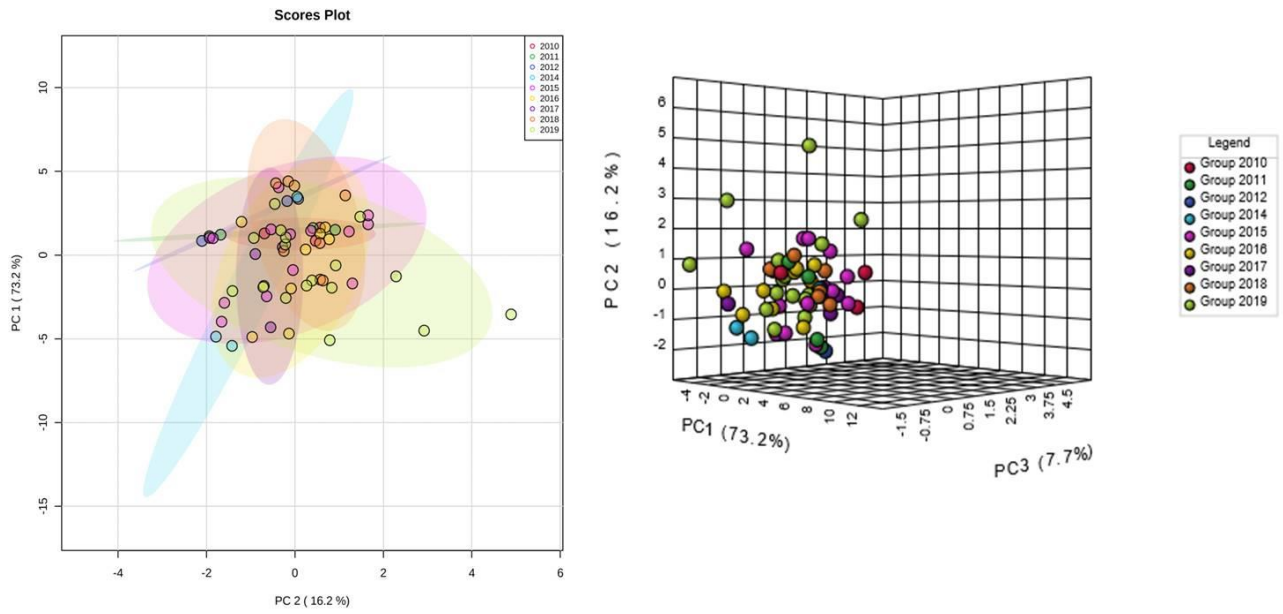
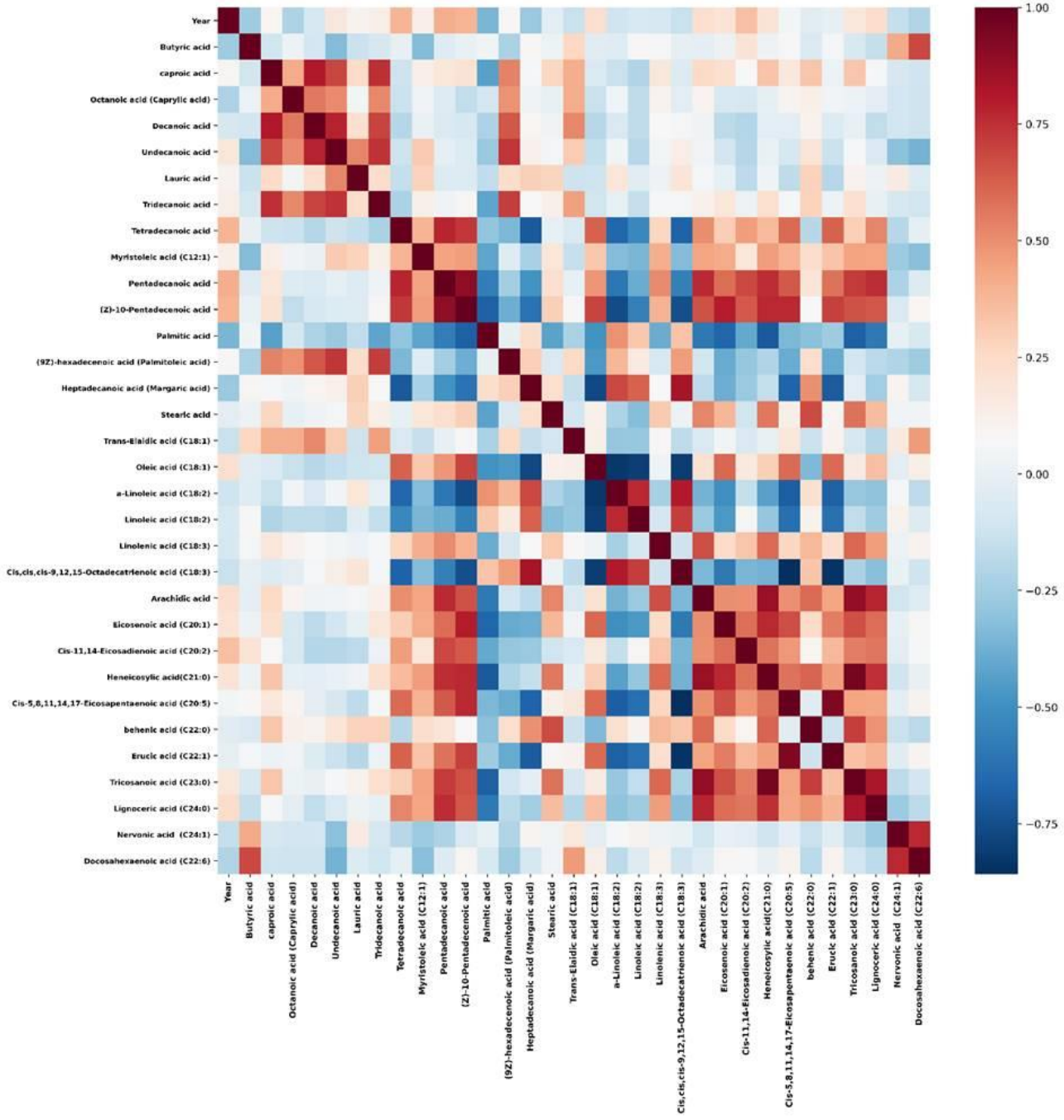


Figure 28 PCA and 3-dimensional plots of the fatty acids analysis (2010 to 2020). The colour dots represent the different years of analysis.

3.3.10.3 Heatmap of the fatty acids

Figure 29 represents the heatmap of the pearson correlation coeffiecient values with colour codes for the fatty acids extracted from the coffees from the year 2010 through 2020. This shows the relationship between each of the fatty acids as the years goes. The plot shows that there the amount of the respective fatty acids decreases as the year goes by from 2010 to 2020.

Figure 29 Heatmap of Pearson correlation coefficient of fatty acid analysis (2010 to 2020).



In summary, this study investigates the change in lipid and CGA profiles during coffee storage. Coffee storage and associated change in quality is an important issue for the logistic companies and coffee producers. In a series of experiments carried out as depicted above, samples were aged in different time domains and their chemical constituents compared. Samples were

additionally aged under defined conditions such as temperature variations and application of inert gases. In general, a decline of CGA levels with longer storage was observed. While under protecting conditions, it was observed that CGA levels can remain constant for decades. For different CGA derivatives different half-life times were observed and thus a hierarchy of stability could be established. This hierarchy is in good agreement with previous studies on CGA general reactivity.

4. CONCLUSION

The highest percentage degradation rate was observed in 5-CQA compared to other compounds. Also, Robusta (87.76%) shows higher percentage degradation rate when compared to Arabica (72.3%) samples between the year 2016 and 2020. Furthermore, 3-CQA quantities were observed to be more depleted in two coffee varieties with slight increase in Arabica (65.61%) compared to Robusta (62.25%).

Among the mono-CQA, 4-CQA has the lowest percentage degradation rate with higher rate observed in Arabica (48.10%) compared to Robusta (46.44%). Besides, among the Di-CQA, the 3,5- isomer was observed to be more depleted compared to 3,4- and 4,5- isomers in both coffee varieties. The Robusta 3,5- isomer (64.62%) was more depleted compared to Arabica (59.36%). Also, 4,5- isomer in Robusta has higher degradation rate (21.51%) compared to Arabica (17.78%), while contrast observation was seen in 3,4- isomer with higher rate in Arabica (57.34%) compared to Robusta (52.12%). The 3-FQA was more depleted in both coffee varieties, followed by 5-FQA and 4-FQA in the Robusta coffee. But the trend was different in Arabica (3-FQA > 4-FQA > 5-FQA).

Moreover, the 5-CQA was more depleted than others, followed by 3-CQA and 3,5-DiCQA, 3,4-DiCQA, 4-CQA and 3-FQA respectively (5-CQA > 3-CQA > 3,5-DiCQA > 3,4-DiCQA > 4-CQA > 3-FQA).

In the roasted coffees, the three (3) isomers of mono-CQAs were observed to be depleted over time. The 5-CQA being the most abundant CQA showed a significant decrease ($p < 0.05$) in all the samples over the period under study, and likewise, the 4-CQA and 3-CQA isomers. Table 3 represent the mean, standard deviation and p-values ($p < 0.05$) of the quantified CGA.

However, the quantities of the mono-CQAs in the two old factory roasted coffees (1990 & 2000) were observed to be significantly higher than the 2018 controls and newly roasted coffees

bought in year 2011 and 2014. But there is no significant difference between the Di-CQA and FQAs.

Moreover, the significant phenomenon observed here is that the old coffees (1990 & 2000) CQAs were depleted at a faster rate compared to other coffees, which probably indicates that there is a possibility that when the coffee was first exposed to air the compounds undergoes degradation more quickly compared to when it was sealed.

The FQA, *p*CoQA, CGA lactones and Di-CQA in roasted coffees were depleted over time, but not significantly ($p < 0.05$) when compared to the green coffees.

The data in the rate of degradation Table 4 shows that the CGA compounds in the 1990 roasted coffee were depleted by over 50%. The most degraded compound among all is 3,5-DiCQA (61.27%) followed by the CGA lactones, FQAs and CQAs respectively. Similar trend was observed in other coffees. However, it could be suggested that most of the CGA compounds are averagely stable despite the long years of storage and for the fact that nearly all the compounds of interest were observed to be present after three decades (30 years) of storage when compared to other coffee samples.

The FRAP assay shows that there was a gradual depletion in the antioxidant contents of both Arabica and Robusta coffees during storage over time, which was observed in both the green and roasted coffees.

The TAGs decreased across the years till year 2020 except for TAG PtAL in Arabica and LLLn (both Arabica and Robusta), where it was observed that there was no significant change in the two TAGs. In addition, the rate of degradation in the two coffee varieties was observed and the TAGs values show that there no significant differences between Arabica and Robusta. From these observations, it could be said that the TAGs extracted from stored green coffee beans for

a long time decreases over time irrespective of the bean variety and lipids fractional constituents.

The variations among the stored coffees fatty acids in Arabica and Robusta coffees were represented indicate that the palmitic acid was very stable over time irrespective of the bean variety and year. Oleic acid was observed to be more depleted in the year 2010 to 2012 when compared to other fatty acids, while other fatty acids were observed to be significantly depleted over time.

REFERENCES

Abrankó, L. & Clifford, M. N. (2017). An unambiguous nomenclature for the acyl-quinic acids commonly known as chlorogenic acids. *Journal of Agricultural and Food Chemistry, 65(18), 3602–3608*.

Baesso, M. L., Correa Da Silva E., Vargas H., Cortez J. G., and Pelzl J. (1990). Use of electron spin resonance for the determination of staling of roast coffee in polyethylene bag packs, *Z. Lebensm. Untersuch. Forsch*. 191:24–27

Belitz, H.D., Grosch W., and Schieberle P., (2004). Lipids in food chemistry, edited by H.D. Belitz, W. Grosch, and P. Schieberle, *Springer-Verlag*, Heidelberg, pp. 157-242.

Benigno M. E., Fong L. E., Biju D., Muharram A., Davis I. M., Vela K. O., Rios D., Osorio-Camacena E., Kaur B., Rojas S., Forester S. C. (2018) The impact of the roast levels of coffee extracts on their potential anticancer activities. *J Food Sci* 83(4):1125–113.

Berrueta L.A., Alonso-Salces R.M., Herberger K. (2007). Supervised pattern recognition in food analysis, *J. Chromatogr. A*, 1158: 196-214.

Charles-Bernard, M., Roberts, D. D., & Kraehenbuehl, K. (2005). Interactions between volatile and non-volatile coffee components. 2. Mechanistic study focused on volatile thiols. *Journal of Agricultural and Food Chemistry, 53(11), 4426–4433*.

Chong, J., Wishart, D.S. and Xia, J. (2019). Using *MetaboAnalyst* 4.0 for comprehensive and integrative metabolomics data analysis. *Current Protocols in Bioinformatics* 68, e86. https://www.metaboanalyst.ca/MetaboAnalyst/home.xhtml/ Accessed 10 November 2019.

Christie WW (1989). Gas chromatography and lipids: A practical guide. First. W.W. Christie ed. *Pergamon Press*, Ayr, Scotland.

Clarke, R.J., (1986). Coffee: in handbook of food and beverage stability: chemical, biochemical, microbiological, and nutritional aspects, edited by G. Charalambous, *Academic Press*, London.

Clifford, M. N., Staniforth. P. S. (1977) A critical comparison of six spectrophotometric methods for measuring chlorogenic acids in green coffee beans. In: *Huitième Colloque Scientifique International sur le Café;* Abidjan, *ASIC*: Paris, 109–114.

Clifford, M. N.; Jarvis, T. (1988) The chlorogenic acids content of green Robusta coffee beans as a possible index of geographic origin. *Food Chem*., 29: 291–298.

Clifford, M. N.; Johnston, K. L.; Knight, S.; Kuhnert, N. (2003) Hierarchical scheme for LC-MS^n identification of chlorogenic acids. *J. Agric. Food Chem*., 51: 2900–2911.

Clifford, M. N.; Knight, S.; Kuhnert, N. (2005) Discriminating between the six isomers of dicaffeoylquinic acid by LC-MS^n. *J. Agric. Food Chem*.,53, 3821–3832.

Clifford, M. N.; Knight, S.; Surucu, B.; Kuhnert, N. (2006) Characterization by LC-MS^n of four new classes of chlorogenic acids in green coffee beans: dimethoxycinnamoyl quinic acids, diferuloylquinic acids, caffeoyl-dimethoxycinnamoyl quinic acids and feruloyl-dimethoxy cinnamoyl quinic acids. *J. Agric. Food Chem*., 54:1957–1969.

Clifford, M. N.; Marks, S.; Knight, S.; Kuhnert, N. (2006) Characterization by LC-MS^n of four new classes of p-coumaric acid-containing diacyl chlorogenic acids in green coffee beans. *J. Agric. Food Chem*., 54: 4095–4101.

Deshpande, S., Jaiswal, R., Matei, M. F., & Kuhnert, N. (2014). Investigation of acyl migration in mono- and di-caffeoylquinic acids under aqueous basic, aqueous acidic and dry roasting conditions. *Journal of Agricultural and Food Chemistry, 62(37), 9160–9170.*

Fahy, E., Subramaniam, S., Murphy, R. C., Nishijima, M., Raetz, C. R. H., Shimizu, T., Dennis, E. A. (2009). Update of the LIPID MAPS comprehensive classification system for lipids. *Journal of Lipid Research*, 50(SUPPL.), S9-S14.

Farah, A. (2005). Distribuicao nos graos, importancia na qualidade da bebidae biodisponibilidade dos acidos clorogenicos do cafe. *Ph.D. Thesis*, UFRJ.

Hameed, A., Hussain, S. A., Ijaz, M. U., Ullah, S., Pasha, I., & Suleria, H. A. R. (2018). Farm to Consumer: Factors Affecting the Organoleptic Characteristics of Coffee. II: Postharvest Processing Factors. *Comprehensive Reviews in Food Science and Food Safety*, 17(5), 1184–1237.

Hečimović I, Belščak-Cvitanović A, Horžić D, Komes D (2011) Comparative study of polyphenols and caffeine in different coffee varieties affected by the degree of roasting. *Food Chem* 129:991–1000

Jaiswal, R., Patras, M. A., Eravuchira, P. J., & Kuhnert, N. (2010). Profile and characterization of the chlorogenic acids in green Robusta coffee beans by LC-MS^n: Identification of seven new classes of compounds. *Journal of Agricultural and Food Chemistry*, 58(15), 8722–8737.

Król, K., Gantner, M., Tatarak, A., Hallmann E. (2020) The content of polyphenols in coffee beans as roasting, origin and storage effect. *Eur Food Res Technol* 246, 33–39.

Lee M., McGeer E. G., McGeer P. L. (2016) Quercetin, not caffeine, is a major neuroprotective component in coffee. *Neurobiol Aging* 46:113–123

Meng, S., Cao, J., Feng, Q., Peng, J., & Hu, Y. (2013). Roles of Chlorogenic Acid on Regulating Glucose and Lipids Metabolism: *A Review. Evidence-Based Complementary and Alternative Medicine, 2013, 1–11.*

Morrice, A. E., Deighton N., Glidwell S. M., and Goodman B. A. (1993) Free radical scavenging reactions in coffee, in proceedings of the 15th international colloquium on the chemistry of coffee (Montpellier, France), *ASIC*, Paris, pp. 644–649.

Nagaraju, V. D., & Bhattacharya, S. (2010). Roasting green coffee beans using spouted bed roaster: Changes in physical characteristics. *Journal of Food Science & Technology*, 47(6), 674–677.

Naveed, M., Hejazi, V., Abbas, M., Kamboh, A. A., Khan, G. J., Shumzaid, M., XiaoHui, Z. (2018). Chlorogenic acid (CGA): A pharmacological review and call for further research. *Biomedicine & Pharmacotherapy*, 97: 67–74.

Nebesny, E., Budryn, G., Kula, J., & Majda, T. (2007). The effect of roasting method on headspace composition of Robusta coffee bean aroma. *European Food Research & Technology*, 225 9–19.

Nikolova-Damyanova, B., Velikova R., and Jham G. N., (1998). Lipid classes, fatty acid composition and triacylglycerol molecular species in crude coffee beans harvested in Brazil, *Food Res. Int*. 31:479–486.

Nunes F. M., & Coimbra M. A. (2007) Melanoidins from coffee infusions. Fractionation, chemical characterization, and effect of the degree of roast. *J Agric Food Chem* 55(10):3967–3977.

Ortala, M. D., Gutierrez, C. L., Chiralt, A., & Fito, P. (1998). Kinetic study of lipid oxidation in roasted coffee. *Food Science and Technology International*, 4, 67–73.

Rendon, M. Y., Salva, T. J. G., & Bragagnolo, N. (2014). Impact of chemical changes on the sensory characteristics of coffee beans during storage. *Food Chemistry*, 147, 279–286.

Rojast, J. (2004). Green coffee storage. In J N. Wintgens (Ed.), Coffee: Growing, processing, sustainable production. Weinheim: *Wiley-VCH* Verlag GmbH & Co. KGaA (pp. 733–750).

Santana-Gálvez, J., Cisneros-Zevallos L., Jacobo-Velázquez D. A. (2017). Chlorogenic acid: recent advances on its dual role as a food additive and a nutraceutical against metabolic syndrome, *Molecules* 22 (3) 358.

Scheidig, C., Czerny, M., & Schieberle, P. (2007). Changes in key odorants of raw coffee beans during storage under defined conditions. *Journal of Agricultural and Food Chemistry*, 55, 5768–5775.

Schievano, E., Finotello, C., De Angelis, E., Mammi, S., & Navarini, L. (2014). Rapid authentication of coffee blends and quantification of 16-*O*-Methylcafestol in roasted coffee beans by nuclear magnetic resonance. *Journal of Agricultural and Food Chemistry*, 62(51), 12309–12314.

Sirbu, D., Corno, M., Ullrich, M. S., & Kuhnert, N. (2018). Characterization of triacylglycerols in unfermented cocoa beans by HPLC-ESI mass spectrometry. *Food Chemistry*, 254, 232-240.

Smith, A. W. (1989) Introduction. In: Coffee, I: Chemistry; Clark, R.J.; Macrae, R.; Eds.; *Elsevier Applied Science Publishers Co Ltd*: London and New York, 1–41.

Speer, K., & Kölling-Speer, I. (2006). The lipid fraction of the coffee bean. *Brazilian Journal of Plant Physiology*, 18(1), 201–216.

Tfouni S. A. V., Serrate C. S., Carreiro L. B., Camargo M. C. R., Teles C. R. A., Cipolli K. M. V. A. B., Furlani R. P. Z. (2012) Effect of roasting on chlorogenic acids, caffeine and polycyclic aromatic hydrocarbons levels in two *Coffea* cultivars: *Coffea arabica cv. Catuaí Amarelo* IAC-62 and *Coffea canephora* cv. Apoatã. Int *J Food Sci Technol* 47:406–415

Toci, A. T., & Farah, A. (2014). Volatile fingerprint of Brazilian defective coffee seeds: Corroboration of potential marker compounds and identification of new low-quality indicators. *Food Chemistry*, 153, 298–314.

Venditti, A., Bianco, A., Frezza, C., Conti, F., Bini, L. M., Giuliani, C., Maggi, F. (2015). Essential oil composition, polar compounds, glandular trichomes and biological activity of *Hyssopus officinalis subsp. aristatus* (Godr.) Nyman from central Italy. *Industrial Crops and Products*, *77, 353–363.*

Vila M. A., Andueza S., Paz de Peña M., & Cid, C. (2005) Fatty acid evolution during the storage of ground, roasted coffees. *Journal of the American Oil Chemists Society*, 82(9), 639-646.

Wei F., Furihata K., Koda M., Hu F., Kato R., Miyakawa T., and Tanokura M., (2012) 13C NMR-Based metabolomics for the classification of green coffee beans according to variety and origin. *Journal of Agricultural and Food Chemistry***,** *60* (40): 10118-10125.

Wold S., Sjostrom M., Eriksson L. (2001). PLS-regression: a basic tool of chemometrics, *Chemometrics Intell. Lab. Syst*., 58: 109-130.

Xu, J. G., Hu, Q. P., & Liu, Y. (2012). Antioxidant and DNA-protective activities of chlorogenic acid isomers. *Journal of Agricultural and Food Chemistry*, 60, 11625–11630.

Table 11 Pearson correlation coefficient (r = 1) of fatty acids extracted from green beans (2016 – 2020).

	Year	Butyric acid	caproic acid	Octanoic acid (Caprylic acid)	Decanoic acid	Undecanoic acid	Lauric acid	Tridecanoic acid	Tetradecanoic acid	Myristoleic acid (C12:1)	Eicosenoic acid (C20:1)	Cis-11,14-Eicosadienoic acid (C20:2)	Heneicosylic acid (C21:0)	Cis-5,8,11,14,17-Eicosapentaenoic acid (C20:5)	behenic acid (C22:0)	Erucic acid (C22:1)	Tricosanoic acid (C23:0)	Lignoceric acid (C24:0)	Nervonic acid (C24:1)	Docosahexaenoic acid (C22:6)
Year	1	-0,258	0,089	-0,213	-0,083	0,177	0,103	0,136	0,384	0,124	0,198	0,351	0,200	0,048	-0,031	-0,010	0,176	0,238	-0,152	-0,208
Butyric acid	-0,258	1	-0,104	0,017	-0,097	-0,329	-0,127	-0,086	0,003	-0,335	0,026	0,201	0,028	0,075	-0,062	0,052	-0,078	-0,147	0,425	0,688
caproic acid	0,089	-0,104	1	0,418	0,813	0,689	0,250	0,753	-0,124	0,135	0,220	0,065	0,338	0,166	0,327	0,016	0,331	0,063	-0,087	-0,110
Octanoic acid (Caprylic acid)	-0,213	0,017	0,418	1	0,559	0,511	0,045	0,520	-0,130	0,029	-0,078	-0,085	-0,011	0,096	0,124	0,040	0,020	-0,018	-0,098	-0,120
Decanoic acid	-0,083	-0,097	0,813	0,559	1	0,783	0,220	0,698	-0,195	0,010	-0,170	-0,203	-0,025	-0,028	0,102	-0,058	0,004	-0,154	-0,097	-0,123
Undecanoic acid	0,177	-0,329	0,689	0,511	0,783	1	0,524	0,741	-0,124	0,315	-0,102	-0,191	0,008	-0,083	0,194	-0,130	0,059	-0,061	-0,306	-0,370
Lauric acid	0,103	-0,127	0,250	0,045	0,220	0,524	1	0,232	-0,151	0,283	-0,006	-0,175	0,035	-0,165	0,284	-0,185	0,109	0,022	0,144	-0,058
Tridecanoic acid	0,136	-0,086	0,753	0,520	0,698	0,741	0,232	1	-0,224	0,129	0,182	0,017	0,218	0,062	0,288	0,019	0,219	-0,007	-0,081	-0,103
Tetradecanoic acid	0,384	0,003	-0,124	-0,130	-0,195	-0,124	-0,151	-0,224	1	0,387	0,299	0,461	0,367	0,599	-0,184	0,618	0,291	0,523	-0,187	-0,009
Myristoleic acid (C12:1)	0,124	-0,335	0,135	0,029	0,010	0,315	0,283	0,129	0,387	1	0,414	0,181	0,458	0,395	0,237	0,338	0,443	0,436	-0,266	-0,316
Pentadecanoic acid	0,410	-0,039	0,181	-0,058	-0,062	-0,015	-0,052	0,031	0,775	0,446	0,590	0,681	0,757	0,638	0,187	0,579	0,719	0,757	-0,216	-0,045

(Z)-10-Pentadecenoic acid	0,386	0,011	0,208	-0,157	-0,084	-0,056	-0,056	0,081	0,727	0,462	0,799	0,639	0,769	0,770	0,069	0,706	0,662	0,649	-0,073	0,091
Palmitic acid	-0,354	0,038	-0,434	-0,095	-0,214	-0,273	-0,160	-0,418	-0,295	-0,419	-0,666	-0,401	-0,707	-0,344	-0,313	-0,260	-0,684	-0,592	-0,002	-0,074
(9Z)-hexadecenoic acid (Palmitoleic acid)	0,072	-0,223	0,533	0,485	0,650	0,737	0,235	0,714	-0,344	-0,045	-0,395	-0,287	-0,207	-0,387	0,227	-0,398	-0,096	-0,181	-0,208	-0,261
Heptadecanoic acid (Margaric acid)	-0,276	0,079	0,063	0,033	0,090	0,121	0,291	0,068	-0,707	-0,230	-0,384	-0,277	-0,148	-0,670	0,494	-0,699	-0,025	-0,248	0,100	-0,084
Stearic acid	-0,024	0,015	0,269	-0,004	0,035	0,057	0,278	0,126	-0,007	0,182	0,380	-0,106	0,562	0,127	0,674	0,082	0,575	0,359	0,035	0,037
Trans-Elaidic acid (C18:1)	-0,137	0,264	0,407	0,406	0,515	0,308	-0,118	0,453	-0,065	-0,144	0,038	-0,062	-0,004	0,179	-0,126	0,108	-0,087	-0,194	0,163	0,465
Oleic acid (C18:1)	0,225	-0,032	-0,065	-0,132	-0,184	-0,147	-0,111	-0,105	0,621	0,320	0,606	0,159	0,291	0,597	-0,335	0,594	0,150	0,351	-0,030	0,135
α-Linoleic acid (C18:2)	-0,112	-0,038	-0,137	0,038	-0,067	0,036	0,187	-0,005	-0,662	-0,333	-0,496	-0,148	-0,432	-0,694	0,223	-0,684	-0,281	-0,303	0,058	-0,173
Linoleic acid (C18:2)	-0,077	0,074	-0,210	-0,163	-0,164	-0,184	-0,034	-0,148	-0,530	-0,351	-0,463	0,081	-0,285	-0,628	0,102	-0,621	-0,155	-0,262	0,101	-0,047
Linolenic acid (C18:3)	-0,087	0,054	0,178	0,109	0,071	0,046	0,018	0,058	0,271	0,408	0,248	0,332	0,597	0,271	0,372	0,230	0,606	0,456	0,027	0,108
Cis,cis,cis-9,12,15-Octadecatrienoic acid (C18:3)	-0,137	-0,010	-0,052	-0,027	0,051	0,135	0,193	0,054	-0,676	-0,330	-0,595	-0,353	-0,368	-0,857	0,326	-0,841	-0,220	-0,317	0,013	-0,117

Arachidic acid	0,224	-0,022	0,257	0,096	0,031	0,036	0,052	0,119	0,501	0,438	0,501	0,497	0,873	0,498	0,596	0,439	0,889	0,779	-0,108	-0,043
Eicosenoic acid (C20:1)	0,198	0,026	0,220	-0,078	-0,170	-0,102	-0,006	0,182	0,299	0,414	1	0,581	0,772	0,672	0,251	0,547	0,669	0,588	-0,007	0,084
Cis-11,14-Eicosadienoic acid (C20:2)	0,351	0,201	0,065	-0,085	-0,203	-0,191	-0,175	0,017	0,461	0,181	0,581	1	0,573	0,473	0,092	0,375	0,544	0,565	-0,027	0,059
Heneicosylic acid(C21:0)	0,200	0,028	0,338	-0,011	-0,025	0,008	0,035	0,218	0,367	0,458	0,772	0,573	1	0,575	0,615	0,460	0,958	0,749	-0,112	0,003
Cis-5,8,11,14,17-Eicosapentaenoic acid (C20:5)	0,048	0,075	0,166	0,096	-0,028	-0,083	-0,165	0,062	0,599	0,395	0,672	0,473	0,575	1	-0,033	0,932	0,435	0,429	-0,022	0,101
behenic acid (C22:0)	-0,031	-0,062	0,327	0,124	0,102	0,194	0,284	0,288	-0,184	0,237	0,251	0,092	0,615	-0,033	1	-0,122	0,713	0,486	-0,133	-0,177
Erucic acid (C22:1)	-0,010	0,052	0,016	0,040	-0,058	-0,130	-0,185	0,019	0,618	0,338	0,547	0,375	0,460	0,932	-0,122	1	0,341	0,397	-0,033	0,095
Tricosanoic acid (C23:0)	0,176	-0,078	0,331	0,020	0,004	0,059	0,109	0,219	0,291	0,443	0,669	0,544	0,958	0,435	0,713	0,341	1	0,827	-0,169	-0,099
Lignoceric acid (C24:0)	0,238	-0,147	0,063	-0,018	-0,154	-0,061	0,022	-0,007	0,523	0,436	0,588	0,565	0,749	0,429	0,486	0,397	0,827	1	-0,266	-0,175
Nervonic acid (C24:1)	-0,152	0,425	-0,087	-0,098	-0,097	-0,306	0,144	-0,081	-0,187	-0,266	-0,007	-0,027	-0,112	-0,022	-0,133	-0,033	-0,169	-0,266	1	0,775
Docosahexaenoic acid (C22:6)	-0,208	0,688	-0,110	-0,120	-0,123	-0,370	-0,058	-0,103	-0,009	-0,316	0,084	0,059	0,003	0,101	-0,177	0,095	-0,099	-0,175	0,775	1

Chapter 7

General Conclusions

GENERAL CONCLUSIONS

From this study investigations involving the use of an established LC-MS method for the profiling and quantification of CGAs, estimation of antioxidant capacity by FRAP assay, also profiling and identification of TAGs with the aid of an optimized LC-MS method coupled with a Tandem-MS and a standard FAME method to profile fatty acids, the variations in green and roasted coffees, fresh and aged coffees, biodynamic and conventionally cultivated coffee beans of both Arabica and Robusta coffee varieties were evaluated using multivariate statistical methods. Hence, the following key observations were made and conclusions drawn:

Chapter 2

i. LC-MS^n quantification of key chlorogenic acids in Arabica and Robusta coffee, mono-CQAs showed no significant variation among coffee varieties,while FQAs and Di-CQAs showed statistically significant variations if Robusta and Arabica beans were compared.
ii. It could be suggested that these variations could be used for authentication purposes in only green coffee beans along with the established recently critiqued 16-*O*-methylcafestol method.
iii. Identification of key CGAs showing significant variations which allows establishment of a reliable multivariate statistical method for coffee authentication only in green coffee beans.
iv. Data statistical evaluation shows a strong correlation of different classes of CGAs in all samples analysed.

Chapter 3

i. With LC-MS^n quantification of key CGAs in organic, biodynamic and conventional coffees, the 4-FQA, 5-FQA and 5-*p*CoQA contents showed statistically significant

variations ($p < 0.05$) if organic roasted coffees were compared with conventional roasted coffees.

ii. The variations in the key three chlorogenic acids could be used in authenticating the method of cultivation (organic and conventional) of coffees in Brazil and also in many other countries where there are challenges in differentiating between organic and conventional cultivated coffee beans.

iii. Organic coffee contains less chlorogenic acids if compared to conventional coffee.

iv. There was no significant difference in coffees grown in different regions of Brazil.

Chapter 4

i. Optimized LC-MS method coupled with ion-trap Tandem-MS has shown to be a proficient analytical technique for oil analysis especially coffee lipids.

ii. Optimized method enabled the separation and identification of mono- and diacylglycerols fragments in a single run together with TAGs.

iii. A faster method with less sequence duration (37 min) was achieved, which allows an accurate TAGs separation and also reduced the common co-elution challenges.

iv. Eight new TAGs were identified including TAGs containing fatty acids with odd numbers of carbon atoms such as margaric (C17:0) and pentacosanoic (C25:0) acids.

v. The identified TAGs and multivariate analysis could be employed in differentiating Arabica coffee from coffees adulterated with Robusta or other coffee varieties.

Chapter 5

i. Discrimination among African, Asian, Central and South American green coffee beans on the basis of region of origin based on fatty acids analysis was achieved.

ii. Oleic, *cis*-5,8,11,14,17-eicosapentaenoic, eicosenoic, erucic, tetradecanoic, *Z*-10-pentadecenoic, and acids were observed to be higher in Robusta than Arabica coffees.

iii. Palmitic acid could be used to differentiate Arabica coffees from Africa and South/Central American coffees from Asia.

iv. South/Central American coffees contains higher amount of the *trans*-linolelaidic acid compared to other regions.

v. Fatty acid compositions with multivariate statistical analysis provided a classification model for coffee with distinct and accurate recognition and prediction at region level.

vi. Fatty acids could be employed in authenticating coffee production regions or sources.

Chapter 6

i. The highest percentage degradation rate was observed in 5-CQA when compared to other compounds.

ii. Robusta (87.76%) shows higher percentage degradation rate when compared to Arabica (72.3%) samples between the year 2016 and 2020.

iii. 3-CQA quantities were observed to be more depleted in two coffee varieties with slight increase in Arabica (65.61%) compared to Robusta (62.25%).

iv. Among the mono-CQA, 4-CQA has the lowest percentage degradation rate with higher rate observed in Arabica (48.10%) compared to Robusta (46.44%).

v. Among the Di-CQA, the 3,5- isomer was observed to be more depleted compared to 3,4- and 4,5- isomers in both coffee varieties. The Robusta 3,5- isomer (64.62%) was more depleted compared to Arabica (59.36%).

vi. The 4,5- Di-CQA isomer in Robusta has higher degradation rate (21.51%) compared to Arabica (17.78%), while contrast observation was seen in 3,4- isomer with higher rate in Arabica (57.34%) compared to Robusta (52.12%).

vii. The 3-FQA was more depleted in both coffee varieties, followed by 5-FQA and 4-FQA in the Robusta coffee. But the trend observed in Arabica was different (3-FQA > 4-FQA > 5-FQA).

viii. Moreover, the 5-CQA was more depleted than others, followed by 3-CQA and 3,5-DiCQA, 3,4-DiCQA, 4-CQA and 3-FQA respectively (5-CQA > 3-CQA > 3,5-DiCQA > 3,4-DiCQA > 4-CQA > 3-FQA).

ix. In the roasted coffees, the three (3) isomers of mono-CQAs were observed to be depleted over time. The 5-CQA showed a more significant decrease ($p < 0.05$) in all the samples compared to 3-CQA and 4-CQA.

x. The mono-CQAs in the two old factory roasted coffees (1990 & 2000) were observed to be significantly higher than the 2018 controls and newly roasted coffees (2011 and 2014). But, there is no significant difference between the Di-CQA and FQAs.

xi. The old roasted coffees (1990 and 2000) CQAs were rapidly depleted at a faster rate compared to other coffees with over 50% degradation level. The most degraded compound among all is 3,5-DiCQA (61.27%) followed by the CGA lactones, FQAs and CQAs respectively.

xii. The FQA, *p*CoQA, CGA lactones and Di-CQA in roasted coffees were depleted over time, but not significantly ($p < 0.05$) when compared to the green coffees.

xiii. It could be suggested that most of the CGAs were averagely stable despite the long years of storage and for the fact that nearly all the compounds of interest were observed to be present after three decades (30 yrs) of storage.

xiv. The FRAP assay shows that there was a gradual depletion in the antioxidant contents of both Arabica and Robusta coffees during storage over time, which was observed in both the green and roasted coffees.

xv. The TAGs decreased across the years till year 2020 except for TAG PtAL in Arabica and LLLn (both Arabica and Robusta), where it was observed that there was no significant change in the two TAGs.

xvi. The rate of degradation of the TAGs values shows no significant differences between Arabica and Robusta coffees.

xvii. It could be said that the TAGs extracted from green coffee beans stored for a long time decreases over time irrespective of the bean variety and lipids fractional constituents.

xviii. The variations among the stored coffees fatty acids in Arabica and Robusta coffees indicate that the palmitic acid was very stable over time irrespective of the bean variety and year.

xix. Oleic acid was more depleted in the year 2010 to 2012 when compared to other fatty acids, while other fatty acids were observed to be significantly depleted over time.

Inconclusion, the various analytical methods explored in profiling CGAs, TAGs and fatty acids along with antioxidant estimation in this thesis proved to be proficient and appropriate in the investigation of ageing process and rate of degradation of chemical compounds in aged green and roasted coffees. The results obtained from the different studies and chemical compounds profiles revealed that coffee quality degrades slowly over time and could remain very stable for a long period of time. Thus, it could be concluded that the coffee quality to a large extent is not affected when coffee is stored under protective conditions for many years.