Evaluating the influence of germination as a post-harvest treatment on green and roasted South African coffee beans (*Coffea arabica*)

by

Stephanie Leigh Phelps

Thesis presented in partial fulfilment of the requirements for the degree of

Master of Food Science

at

Stellenbosch University

Department of Food Science, Faculty of AgriSciences

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the NRF.

Supervisor: Dr. Paul James Williams

Co-supervisor: Dr. Janine Colling

December 2021

Stellenbosch University https://scholar.sun.ac.za

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third-party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2021

Copyright © 2021 Stellenbosch University
All rights reserved

Stellenbosch University https://scholar.sun.ac.za

Summary

Coffee is popular worldwide for its diversity in flavour as well as its beneficial health factors. The coffee bean is produced from *Coffea* species, of which there are 103 different species. The chosen method for post-harvest processing plays an essential role in the development of the coffee bean's flavour and aroma. Aside from the hybrid "semi-washed" process, no other processing has been commercially established in the last few decades. Germination has been proposed as a novel, post-harvest treatment for coffee bean processing, specifically for South African coffee beans. Although research has focused on germination and its effects in terms of sprouting seeds, no research has been conducted regarding intentional germination during coffee processing.

In this research, liquid chromatography coupled to a mass spectrometer (LCMS) was used to analyze the various chemical components of South African coffee beans (*C. arabica*, Catuai). For the first study, caffeine, chlorogenic acid, trigonelline, and nicotinic acid were identified as key compounds that play a role in coffee's flavour profile. These compounds were therefore evaluated to determine if germination influenced their concentration levels. Analysis of Variance (ANOVA) results revealed that the interaction effect of production stage and treatment were not significant ($p < 0.05$) for all the compounds studied. The main effect of treatment also did not reveal significant differences ($p < 0.05$) for all the compounds (caffeine ($p = 0.48$), chlorogenic acid ($p = 0.27$), trigonelline ($p = 0.28$), and nicotinic acid ($p = 0.44$)). The low p-values of chlorogenic acid and trigonelline suggest that perhaps some significance could be observed. However, further sampling a second population would be required to support this since the current research had a sampling size of 24 and only one coffee variety. The influence of germination on the key compounds was not observed for either green or roasted coffees and therefore it is assumed that the flavour profile would remain the same between control and germinated coffee.

The second study focused on the entire phenolic profile of the coffee beans to determine if other compounds were being influenced. Principal Component Analysis (PCA) and Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) were the chosen chemometric techniques to analyze the data matrices. Normalization, pareto scaling, and automated minor peak-shift alignments were applied as pre-processing techniques to eliminate unwanted variations. However, neither PCA nor OPLS-DA could distinguish significant differences ($p < 0.05$) between the control and germinated coffee samples.

Stellenbosch University https://scholar.sun.ac.za

Although the sampling size was sufficient for statistical analysis, the small sample set, from only two harvest years, impacted the reproducibility of the multivariate data analysis. Thus, it is recommended to gain more samples from different harvest years to determine if germination influences chemical composition.

The results observed in this study reveal the first evaluation of a South African origin coffee bean and the first study of germination as novel, post-harvest treatment. Although the results suggest that germination has no influence on coffee's flavour profile, more research should be conducted to include samples from different origins, different species, and different harvest years.

Stellenbosch University https://scholar.sun.ac.za

List of abbreviations

ANOVA	Analysis of variance
CGA	Chlorogenic acid
CID	Collision-induced dissociation
DAD	Diode array detector
DW	Dry weight basis
EMRT	Exact Mass/Retention Time
ESI	Electrospray ionization
GC	Gas chromatography
ICL	Isocitrate lyase
IQR	Interquartile range method
LC	Liquid chromatography
LOD	Limit of detection
LOQ	Limit of quantification
m/z	Mass-to-charge ratio
MS	Mass spectrometry/spectrometer
MVA	Multivariate analysis
NAD	Nicotinamide adenine dinucleotide
OPLS-DA	Orthogonal partial least squares discriminant analysis
PCA	Principal component analysis
Q-TOF	Quadruple time-of-flight
RT	Retention time
UV	Ultraviolet
VIP	Variable importance in projection

Stellenbosch University https://scholar.sun.ac.za

This thesis is dedicated to my parents.

Thank you for your unconditional love and support, for without which none of this would have been possible. To my father, for your continuous sage advice and incredible dad-jokes throughout this process. To my mother, for always encouraging me towards the finish line and giving the greatest hugs.

Stellenbosch University https://scholar.sun.ac.za

Acknowledgments

I wish to express my sincere appreciation to the following persons and institutions:

To my supervisor, Dr. Paul Williams, for your guidance and support throughout this study. It was not an easy project and you helped me align onto the right path when I wandered off on a seemingly good tangent.

To my co-supervisor, Dr Janine Colling. I am so grateful for your presence in this research. You provided different perspectives during times of struggle, and it brought a lot of clarity and calmness when I needed it most.

To Dylan Cummings from Beaver Creek Coffee Estate for providing the initial idea behind the research project and also the wonderful coffee beans.

This research would not have been possible without the financial support of the National Research Foundation of South Africa (Grant nr. 116307 to P.J. Williams; Grant nr. 132887 to S.L. Phelps).

To Dr. Marietjie Stander, Mr. Malcolm Taylor, and Mr Erick Van Schalkwyk for your assistance and advice regarding the LCMS analyses (Mass Spectrometry unit, Central Analytical Facilities).

To Prof. Martin Kidd for the statistical data analysis (Centre for Statistical Consultation. Department of Statistics and Actuarial Sciences, University of Stellenbosch).

To my dear family and friends, too many to name but you know who you are. Thank you to all those who encouraged me from start to finish, those who spent many hours in coffee shops working with me, and those who lent an open ear for when I needed it. You are all truly appreciated!

Stellenbosch University https://scholar.sun.ac.za

Preface

This thesis is presented as a compilation of five chapters followed by a sixth Supplementary Notes chapter. Each chapter is introduced separately and is written according to the style of the International Journal of Food Science and Technology.

Chapter 1	General Introduction
Chapter 2	Literature review A review of *Coffea* spp. production methods, chemical composition, and the use of LCMS to evaluate overall coffee quality.
Chapter 3	Materials and Methods
Chapter 4	Results and Discussion
Chapter 5	General conclusions and recommendations
Chapter 6	Supplementary Notes

Stellenbosch University https://scholar.sun.ac.za

Table of Contents

Declaration **ii**

Summary **iii**

List of abbreviations **v**

Acknowledgments **vii**

Preface **viii**

Table of Contents **ix**

1 General Introduction 1

2 A literature review of coffee's production methods, chemical composition, and the use of LCMS to evaluate coffee quality. 9

2.1 Introduction 9

2.1.1 Coffee's origin and history 9

2.1.2 World production and consumption 10

2.1.3 South African coffee market 11

2.2 Botany 12

2.2.1 Species classification 12

2.2.2 Characteristic features 12

2.2.3 Ecological cultivation requirements 14

2.2.4 Geographical distribution 15

2.3 Processing of coffee 16

2.3.1 Harvesting 16

2.3.2 Post-harvest methods 17

2.3.2.1 Wet processing 17

2.3.2.2 Dry processing 18

2.3.2.3 Semi-dry processing 18

Stellenbosch University https://scholar.sun.ac.za

2.3.2.4 Germination 18
2.3.3 Roasting 20

2.4 Chemical components 20
2.4.1 Nitrogenous compounds 21
2.4.2 Carbohydrates 22
2.4.3 Chlorogenic acid and Organic acids 22
2.4.4 Trace compounds 23

2.5 Quality of coffee 23
2.5.1 Defining quality 23
2.5.2 Governing factors 25

2.6 Analytical techniques 25
2.6.1 Instrumentation 26
2.6.2 Data analysis 28
2.6.3 Summary 29

3 Methods & Materials 34

3.1 Materials 34
3.1.1 Chemicals 34
3.1.2 Samples 34
3.1.3 Sampling procedure 34

3.2 Methodology 35
3.2.1 Sample preparation 35
3.2.2 Extraction 35
3.2.3 Liquid Chromatography-Mass Spectrometry analysis 36

3.3 Targeted approach 36
3.3.1 Chromatogram processing 36
3.3.2 Data analysis 37

3.4 Untargeted approach 38
3.4.1 Chromatogram processing 38
3.4.2 Data analysis 38

Stellenbosch University https://scholar.sun.ac.za

4 Results & Discussion 41

4.1 Method Selection 41

4.2 Targeted approach 42

4.2.1 Peak identification 42

4.2.2 Calibration parameters 44

4.2.3 Outlier Detection 45

4.2.4 Analysis of Variance 47

4.2.4.1 Caffeine concentration 47

4.2.4.2 Chlorogenic acid quantification 49

4.2.4.3 Trigonelline and Nicotinic acid quantification 50

4.2.5 Influence on quality 54

4.2.6 Summary 54

4.3 Untargeted approach 56

4.3.1 Chromatogram transformation 56

4.3.2 Principal Component Analysis (PCA) 57

4.3.3 Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) 63

4.3.4 Summary 67

5 General conclusions and recommendations 74

5.1 General conclusions 74

5.2 Recommendations for future research 77

6 Supplementary Notes 80

Stellenbosch University https://scholar.sun.ac.za

1 General Introduction

The coffee bean, as well as the variety of products that are produced from it, is an exported commodity with a high demand. In the 2019/20 period, the International Coffee Organization (2020a) reported that 10.1 billion kg of coffee had been consumed worldwide. Europe contributed the largest proportion (33.4%) to this total, whereas Africa ranked 5th with a mere 6.6% (International Coffee Organization, 2020a). Although South Africa contributed less than 0.5% to the global consumption, there has been a steady increase in its own coffee market over the years due to the rapid development in coffee culture (International Coffee Organization, 2020b; van der Merwe & Maree, 2016).

Of the 103 *Coffea* species currently reported in the world, primarily the arabica (*Coffea arabica* L.) and robusta (*Coffea canephora* Pierre) species are cultivated for commercial use (Kathurima *et al.,* 2009). Aside from quality differences between these two species, the differences in characteristic attributes such as aroma or mouthfeel can be a result of the post-harvest treatment that was applied. Generally, coffee beans are either obtained via the "dry" or the "wet" processing method (Bytof *et al.,* 2000; Tarzia *et al.*, 2010). In the last two decades, a third process has developed known as the "semi-dry" treatment (Bytof *et al.,* 2000; Tarzia *et al.*, 2010; Teketay, 1999), but no further processing methods have been established since. These treatments each have varying influences on the quality of the coffee bean produced.

There is difficulty in defining quality as a trait, since the attributes within coffee have likely evolved over time and the definition can vary between the farmer and consumer levels (Leroy *et al.*, 2006). The biochemical content can be related to factors such as taste and aroma, which can be used as a basis for defining quality (Kathurima *et al.,* 2009). To date, coffee research has focused on the effects of post-harvest influences on both arabica and robusta species, specifically looking at the most abundant compounds such as carbohydrates, chlorogenic acids, and caffeine. Further insight can be gained from studying alternative treatments, such as *germination,* as a potential novel post-harvest method for the production of coffee beans by focusing on the chemical composition changes and quality aspects.

Stellenbosch University https://scholar.sun.ac.za

As mentioned previously, the post-harvest treatments are generally one of two options (the dry or wet method). During either process, the fundamental goal is to reduce the moisture content to between 10% and 12% by means of drying (Bytof *et al.*, 2000; Selmar *et al.*, 2006). Once properly dried, the beans are ready to go through a de-hulling machine and finally the roasting process, which will develop the much-loved coffee aromas. For the "dry" method, the fruits are not required to be sorted and after harvesting are immediately spread out on covered ground and allowed to sun-dry. This natural drying process can last up to 25 days (Gonzalez-Rios *et al.*, 2007). Alternatively, forced airflow from air dryers can also be used to dry the cherries and can take up to 9 days when performed in a closed environment (Tarzia *et al.*, 2010). The drying method causes the skins of the fruit to wither and detach from the seed, or commonly referred to as the bean. The dried matter can then pass through a dehulling machine that removes the husk of the bean by means of abrasive action. The second method, classified as a "wet" treatment, uses only mature coffee cherries for production (Bytof *et al.*, 2000). Prior to being soaked, the flesh of the fruit is removed by means of a depulper resulting in coffee beans (seeds) which are surrounded by a mucilage layer (Tarzia *et al.,* 2010). The soaking stage allows for the fermentation of the mucilage which greatly improves the overall quality (Bytof *et al.,* 2000; Tarzia *et al.*, 2010; Teketay, 1999). Lastly, the beans are left out to dry followed by the dehulling step to remove the parchment layer as opposed to the husk layer in the "dry" method (Tarzia *et al.*, 2010). Another treatment often used in smaller scale coffee productions has been adapted from the "wet" method and is known as "semi-dry" processing. The cherry flesh is depulped, as in the wet method, but then the fermentation stage is omitted and the seeds are immediately submitted to drying (Bytof *et al.,* 2000; Tarzia *et al.*, 2010). As one would expect, the beans produced by this intermediary process have properties that lie in between the dry and wet treatments (Bytof *et al.*, 2000).

The process of germinating coffee seeds has been widely studied in terms of the changing metabolic activities that occur as well as the resulting chemical and physical properties of the seed (Bytof *et al.*, 2007; Kim *et al.*, 2018; Selmar *et al.* 2006; Waters *et al.*, 2017). Germination is activated at various points throughout the post-harvest treatment and depends on the method which was used (Selmar *et al.*, 2006). However, this activation is an *unintentional* side effect due to a metabolic stress and/or the change during embryogenesis (Bytof *et al.*, 2007; Selmar *et al.*, 2006). At the time of writing, no

Stellenbosch University https://scholar.sun.ac.za

studies have been found that consider the effects of germination as an intentional post-harvest method. If an intentional germination step was to be implemented, it could be seen as an extension of the fermentation stage that occurs during the wet treatment, or as an entirely new step during the dry processing treatment. The influence of the new proposed post-harvest treatment on the quality of coffee requires further investigation.

"Coffee quality" is extremely complex to define. It is governed by two factors: the inherent properties resulting from environmental and genetic variation (physical and chemical properties), and the processing parameters from harvesting to delivery (Sylvian, 1958). External traits of coffee beans such as size, shape, uniformity, and so forth are collectively characterized as the physical properties. Chemical aspects include aroma, flavour and body which are dependent on the biochemical processes that take place within the bean (Selmar *et al.*, 2000; Sylvian, 1958). It is therefore important to identify which compounds are responsible for certain key characteristics in coffee and how they evolve during the processing of coffee. The non-volatile compounds that are found in coffee partake in various reactions that ultimately have the most defining impact on the end quality. The volatile substances from coffee are also important for coffee profiling but are generally by-products of the non-volatile substances. Volatile components in coffee have been previously studied (Eroz Poyraz *et al.*, 2016; Sunarharum, 2016) and will not be discussed further. The various non-volatile components have also been studied in terms of their role within coffee and the influence on quality.

Compounds that potentially play key roles in different coffees are grouped into alkaloids, phenolics, organic acids, and carbohydrates. Trigonelline occurs in many plants and the coffee plant is no exception. This nitrogenous substance belongs to the alkaloid family and contributes a bitter taste to a brewed coffee, with minimal influence towards the final aroma. Studies have shown that there is a reduction in trigonelline during post-harvest processing, especially when the wet method is applied to either robusta or arabica coffee species (Farah, 2012; Selmar *et al.*, 2000). The degradation during roasting produces volatile compounds like pyrroles and pyridines, as well as nicotinic acid (Oestreich-Janzen, 2010). Chlorogenic acids (CGAs) are a large family of phenolic compounds with numerous possible substitutions on the cyclohexane ring to form a range of isomers and/or epimers (Farah, 2012; Oestreich-Janzen, 2010). The influence of post-harvest processing on CGAs has been studied and results show that

Stellenbosch University https://scholar.sun.ac.za

both wet and dry methods result in changes in CGA content. This could also be because of the different maturation stages of fruits used during the dry process compared to the wet process (Selmar *et al.*, 2000). Total CGAs as well as subgroups thereof are all significantly changed during the roasting process of coffee, which can create a unique flavour profile (Oestreich-Janzen, 2010). In brewed coffee, CGAs are responsible for the astringency, bitterness and acidity of coffee (Farah, 2012). Caffeine belongs to the alkaloid class of compounds. Unlike most constituents in coffee, caffeine is thermostable and therefore unaffected by the roasting stage (Oestreich-Janzen, 2010; Selmar *et al.*, 2000). Contrary to popular belief, caffeine plays a very small role in conferring bitterness to beverages even though it is commonly used as a 'bitter' standard in sensorial analysis (Oestreich-Janzen, 2010). Nicotinic acid, or niacin, is the degradation product of trigonelline which is produced during roasting (Oestreich-Janzen, 2010; Selmar *et al.*, 2000). The compound plays a secondary role in coffee in terms of sensory attributes and instead offers benefits for human health and nutrition as a vitamin source.

Other constituents that are found in coffee include the free amino acids, proteins and peptides. The free amino acid profile and protein content can play a role in aroma and cup quality as they are responsible for carrying important compounds via reaction pathways. They are affected by roasting, but not by any of the postharvest processes (Selmar *et al.*, 2000). The carbohydrates represent the largest fraction of the dry coffee matter (Bradbury, as cited by Selmar *et al.*, 2000). The various subgroups (low and high molecular compounds) are involved in numerous reactions from the time of harvesting to the green bean stage and finally the roasted beans. Most notable is their role during the roasting stage where they undergo the Maillard reaction and caramelization to form flavour compounds, as well as contribute to colour via production of melanoidins (Selmar *et al.*, 2000). Last are the lipid constituents which are widely affected by climatic and environmental factors (Villarreal *et al.*, 2009) and are therefore a tedious group to study in terms of coffee classification and comparison since coffee production occurs in various countries and under different cultivation conditions. Triacylglycerols, esters of diterpene alcohols, and fatty acids are the largest contributing fractions to the lipid content (Oestreich-Janzen, 2010). Compounds within coffee contribute in different ways to coffee's profile and various techniques have been developed to identify and quantify them.

Stellenbosch University https://scholar.sun.ac.za

High Performance Liquid Chromatography (HPLC) is a popular analytical technique in industry for qualitative and quantitative purposes (Stander *et al.*, 2019). The fundamentals of HPLC analysis are based on the separation of compounds in a liquid form because of their polarity and interaction with a mobile phase. The benefit of combining HPLC with spectral detection methods is that vast amounts of additional information about a sample can be derived. In a targeted approach, HPLC can be used to detect and quantify specific compounds of interest but it can also be used to compare the overall chemistry of samples by means of an untargeted approach. Alonso-Salces *et al.* (2009) studied the structures of various phenolic acids in coffee by means of an advanced liquid chromatography technique that was combined with a photodiode array detector, electrospray ionization, collision-induced dissociation and tandem mass spectrometer (LC-DAD/ESI-CID-MS/MS). The study identified ten compounds that had never been reported before in coffee. Similarly, another study only used ultra-performance liquid chromatography (UPLC) coupled to a quadruple time-of-flight mass spectrometry (QTOF-MS) to analyse coffee bean extracts both quantitatively and qualitatively (O'Driscoll, 2014). The QTOF-MS detector provided additional information (fragmentation ions) that facilitated the identification of ten important compounds in the coffee (O'Driscoll, 2014). The coupling of more detectors to a standard HPLC unit can provide large amounts of additional data, but meaningful analysis and processing is required to properly interpret the results.

The data that is generated is usually in the form of a complex matrix and requires analyses of a specialized nature. Chemometrics is a multivariate data analysis method used to tackle such matrices by means of applying complex statistical and mathematical functions (Lavine & Workman, 2013). Before multivariate data analysis can be performed, pre-processing is performed on the data to remove variation that is unrelatedto the differences between sample groups (Lavine & Workman, 2013; Rabatel *et al.*, 2020). Each pre-processing technique will correct for specific variation observed in data. For example, normalisation is a common tool used to correct data points when the quantity or concentration of compounds for a product vary between samples (Rabatel *etal.*, 2020). There are numerous pre-treatment methods available and widely studied (Verboven *et al.*, 2012; Lavine & Workman, 2013; Rabatel *et al.*, 2020). Multivariate analysis, after correct pre-processing, can provide information pertaining to similarities and/or differences between samples. Principal component analysis is a

Stellenbosch University https://scholar.sun.ac.za

popular choice used to quickly observe trends within sample sets (Lavine & Workman, 2013). Chemometric analysis of data obtained from HPLC with hyphenated techniques can provide valuable insight into the chemical composition of samples.

The aim of this study was to investigate the effect of germination as an intentional post-harvest treatment during coffee processing via two objectives. The first was a targeted approach by quantifying chemical changes in four specific compounds: caffeine, chlorogenic acid, trigonelline, and nicotinic acid. Secondly, an untargeted analysis was conducted to compare overall chemical changes which might have occurred because of the new post-harvest treatment.

Stellenbosch University https://scholar.sun.ac.za

References

Alonso-Salces, R. M., Guillou, C. & Berrueta, L. A. (2009). Liquid chromatography coupled with ultraviolet absorbance detection, electrospray ionization, collision induced dissociation and tandem mass spectrometry on a triple quadruple for the on-line characterization of polyphenols and methylxanthines in green coffee beans. *Rapid Communications in Mass Spectrometry*, **23**(3), 363-383.

Bytof, G., Knopp, S.-E., Kramer, D., Breitenstein, B., Bergervoet, J. H. W., Groot, S. P. C. & Selmar, D. (2007). Transient occurrence of seed germination processes during coffee post-harvest treatment. *Annals of Botany*, **100**, 61-66.

Bytof, G., Selmar, D. & Schieberle, P. (2000). New aspects of coffee processing: How do the different post harvest treatments influence the formation of potential flavour precursors? *Journal of Applied Botany*, **74**, 131-136.

Eroz Poyraz, I., Ozturk, N., Kiyan, H. T. & Demirci, B. (2016). Volatile compounds of *Coffea* Arabica L. green and roasted beans. *Anadolu University Journal of Science and Technology C – Life Science and Biotechnology*, **5**(1), 31-35.

Farah, A. (2012). Coffee constituents. In: *Coffee: Emerging Health Effects and Disease Prevention*, 1st ed. (edited by Y.-F. Chu). Pp. 21-58. USA: Wiley-Blackwell Publishing Ltd.

Gonzalez-Rios, O., Suarez-Quiroz, M.L., Boulanger, R., Barel, M., Guyot, B., Guiraud, J.-P. & Schorr-Galindo, S. (2007). Impact of "ecological" post-harvest processing on coffee aroma: II. Roasted coffee. *Journal of Food Composition and Analysis*, **20**(3-4), 297-307.

International Coffee Organization. (2020a). *World coffee consumption*. [Internet document]. URL http://www.ico.org/prices/new-consumption-table.pdf. Accessed 13/10/2020.

International Coffee Organization. (2020b). *Crop year production*. [Internet document]. URL http://www.ico.org/prices/po-production.pdf. Accessed 13/10/2020.

Kathurima, C. W., Gichimu, B. M., Kenji, G. M., Muhoho, S. M. & Boulanger, R. (2009). Evaluation of beverage quality and green bean physical characteristics of selected Arabica coffee genotypes in Kenya. *African Journal of Food Science*, **3**(11), 365-371.

Kim, Y., Kim, Y. & Jhon, D.-Y. (2018). Changes of the chlorogenic acid, caffeine, γ-aminobutyric acid (GABA) and antioxdant activities during germination of coffee bean (*Coffea arabica*). *Emirates Journal of Food and Agriculture*, **30**(8), 675-680.

Lavine, B. K. & Workman, J. (2013). Chemometrics. *Analytical Chemistry*, **85**(2), 705-714.

Stellenbosch University https://scholar.sun.ac.za

Leroy, T., Ribeyre, F., Bertrand, B., Charmetant, P., Dufour, M., Montagnon, C., Marraccini, P. & Pot, D. (2006). Genetics of coffee quality. *Brazilian Journal of Plant Physiology*, **18**(1), 229-242.

O'Driscoll, D. J. (2014). Analysis of coffee bean extracts by use of ultra-performance liquid chromatography coupled to quadruple time-of-flight mass spectrometry. *MethodsX*, **1**, 264-268.

Oestreich-Janzen, S. (2010). Chemistry of coffee. *Comprehensive Natural Products II: Chemistry and Biology*, **3**, 1085-1117.

Rabatel, G., Marini, F., Walczak, B. & Roger, J.-M. (2020). VSN: Variable sorting for normalization. *Journal of Chemometrics*, **34**, 1-16.

Selmar, D., Bytof, G., Knopp, S.-E. & Breltenstein, B. (2006). Germination of coffee seeds and its significance for coffee quality. *Plant Biology*, **8**, 260-264.

Stander, M. A., Joubert, E. & De Beer, D. (2019). Revisiting the caffeine-free status of rooibos and honeybush herbal teas using specific MRM and high resolution LC-MS methods. *Journal of Food Composition and Analysis*, **76**, 39-43.

Sunarharum, W. B. (2016). *The compositional basis of coffee flavour*. PhD Thesis. University of Queensland, Brisbane.

Sylvian, P. G. (1958). Ethiopian coffee: its significance to world coffee problems. *Economic Botany*, **12**(2), 111-139.

Tarzia, A., dos Santos Scholz, M. B. & de Oliveira Petkowiez, C. L. (2010). Influence of the postharvest processing method on polysaccharides and coffee beverages. *International Journal of Food Science and Technology*, **45**, 2167-2175.

Teketay, D. (1999). History, botany and ecological requirements of coffee. *Walia*, 20, 28-50.

van der Merwe, K. & Maree, T. (2016). The behavioral intentions of specialty coffee consumers in South Africa. *International Journal of Consumer Studies*, **40**, 501-508.

Verboven, S., Hubert, M. & Goos, P. (2012). Robust preprocessing and model selection for spectral data. *Journal of Chemometrics*, **26**(6), 282-289.

Villarreal, D., Laffargue, A., Posada, H., Bertrand, B., Lashermes, P., and Dussert, S. (2009). Genotypic and environmental effects on coffee (*Coffea arabica* L.) bean fatty acid profile: Impact on variety and origin chemometric determination. *Journal of Agricultural and Food Chemistry*, **57**, 11321–11327.

Stellenbosch University https://scholar.sun.ac.za

2 A literature review of coffee's production methods, chemical composition, and the use of LCMS to evaluate coffee quality.

The following literature review will discuss coffee with regards to its background, botany, processing methods, chemical composition, quality, and analysis techniques. The introductory paragraphs will briefly discuss the origin and history of coffee as well as discuss production and consumption in a global context as well as within a South African context. An overview of the coffee plant's characteristics and ecological cultivation requirements will follow afterwards. Clarity will be provided on the processing steps involved in coffee production, with emphasis on the post-harvest methods. The chemical composition of coffee will also be highlighted. Quality parameters of coffee are important for industry and consumers and shall be discussed in relation to important chemical compounds. The review will conclude with a section on analytical techniques typically used to analyze coffee as well as the different approaches for data analysis.

2.1 Introduction

2.1.1 Coffee's origin and history

Coffee is an ancient commodity, dating back more than 1 000 years. There are a number of conflicting stories of where coffee was first discovered. The most popular and most likely apocryphal story is that of an Ethiopian goat-herder who witnessed his goats eating red berries from nearby bushes and soon after they began moving with intense energy (Smith, 1985; Teketay, 1998). Alternatively, another legend is that of a sick and dying Mohammed who was visited in his dreams by the Angel Gabriel. The Angel carried with him a dark beverage and told Mohammed of its healing benefits (Smith, 1985; Teketay, 1998). However there is no physical evidence that either of these tales took place. The first-ever written documentation of coffee's medicinal features was dated 1000 AD by a writer called Avicenna who named the drink *bunchum* or "black liquid" (Luttinger & Dicum, 2006).

There also appears to be a divide in literature about where the first cultivation of coffee began. According to some, coffee originated from Arabia, hence the species name

Stellenbosch University https://scholar.sun.ac.za

'Arabica' (Oestreich-Janzen, 2010). For some time, there were heavy restrictions on the trade of coffee plants and coffee beans were made infertile but the growing interest in the plant resulted in the smuggling of seedlings and beans. On the other hand, Crawford (1985) and Teketay (1998) both reported that, although named after Arabia, the plant is in fact native to Ethiopia but quickly found its way to Arabia soon after first cultivation. Further evidence for this is that wild *Coffea* Arabica plants are still seen in Ethiopia today whereas only commercialized plants are found in Arabia. From Arabia, it is assumed that the plant next appeared in India and then its cultivation quickly expanded to surrounding areas like Sri Lanka and Java (Oestreich-Janzen, 2010). By the mid-nineteenth century, coffee had been introduced to the United States and the establishment of the New York Coffee exchange occurred. The explosion of coffee's expansion, and subsequently its popularity, continued during the 20th and 21st century. One can see the increase in the amounts of Brazilian, German, Colombian and African blends being imported throughout America, Europe and the rest of the world (Luttinger & Dicum, 2006; Teketay, 1998) as well as the development of instant coffees, espresso's, and specialty blends.

2.1.2 World production and consumption

The International Coffee Organization (2020a) reports on the production and consumption of coffee for various countries during each year of production. In 2019, production from the top five producing countries equated to 7.26 million tons of coffee (Table 2.1.1). Brazil remained the top producing country but still experienced an 11% drop from its previous year. Although Indonesia ranked 4th and saw a 16.5% increase in production from 2018, this does not rival the capacity at which Brazil and Vietnam produce coffee beans. In the 2019/20 period, the International Coffee Organization reported that a total of 10.1 billion kg of coffee beans had been consumed worldwide (International Coffee Organization, 2020b).

Stellenbosch University https://scholar.sun.ac.za

Table 2.1.1. Coffee production (metric tons) in 2019 for the top five coffee producing countries (International Coffee Organization, 2020a).

Ranked	Country	Crop year production (tons)	Change since 2018
1st	Brazil	3.5 million	-10.9%
2nd	Vietnam	1.8 million	-1.7%
3rd	Columbia	840 000	1.7%
4th	Indonesia	660 000	16.5%
5th	Ethiopia	462 000	2.1%

2.1.3 South African coffee market

South Africa contributed less than 0.5% to the global consumption over the last five years. However, there has been a steady increase in coffee demand over the years due to the rapid development in coffee culture (International Coffee Organization, 2020b; van der Merwe & Maree, 2016). Most coffee beverages in South Africa are produced from beans that have been imported from numerous countries like Brazil, Vietnam, and Columbia with a greater support for African countries such as Ethiopia, Rwanda, Kenya, and Uganda. A small portion of South Africa is suitable for crop production with the first documented plantation appearing at a Tea Estate near Tzaneen in Limpopo (Department of Agriculture, Forestry, and Fisheries, 2012). The provinces identified (Table 2.1.2) are located along the eastern shores and experience high levels of rainfall and humidity necessary for the growth of coffee. The coffee produced in South Africa is not typically exported for retail, and rather aimed at the local and tourist markets (van der Merwe & Maree, 2016)

Table 2.1.2. Province and districts where coffee is commonly cultivated in South Africa (Department of Agriculture, Forestry, and Fisheries, 2012).

Province	District
KwaZulu-Natal	South and North coast
Mpumalanga	Barberton
	Hazyview
	Bosbokrand
Eastern Cape	East London
Limpopo	Grenshoek Tea Estate

Stellenbosch University https://scholar.sun.ac.za

2.2 Botany

2.2.1 Species classification

The tropical coffee plant is a member of the Rubiaceae family, genus *Coffea* L. The most important economic species are that of *Coffea arabica* (arabica) providing 80% of the coffee market, followed by *Coffea canephora* (robusta) supplying approximately 20% of the world's coffee, and lastly *Coffea liberica* with less than 1% of the global production (Farah, 2012; Mussatto *et al.*, 2011). The three mentioned species form part of the 100 identified species belonging to *Coffea* spp. (Oestreich-Janzen, 2010; Teketay, 1998).

Subspecies within a species are typically referred to as either "varieties" or "cultivars". However, these two terms should not be used interchangeably as they represent different breeding conditions of a coffee plant. The term "variety" should be used when the coffee plant subspecies has been bred through natural selection because of environmental fluctuations (Farah, 2012). "Cultivar" is the term given when intentional cross breeding to obtain desired attributes is used and therefore would not typically appear in the wild (Farah, 2012).

2.2.2 Characteristic features

Numerous factors can be used to differentiate between the main *Coffea* species, such as root system, leaves, flowers, fruits, height, and canopy diameter to name a few. Only the key features will be discussed further, namely height parameters, leaf shape and size, and fruits.

The difference between the terminology of a "shrub" and "tree" is often regarding the height of the plant. Mature *C. arabica* plants can grow to a height of 4 to 6 m and therefore are termed a shrub, whereas the *C. canephora* and *C. liberica* are both coffee trees and grow up to heights of 12 m and 18 m, respectively (Oestreich-Janzen, 2010). However, in plantations they are often trimmed back to 3 m for ease of harvesting (DAFF, 2012).

The leaves of the coffee plant are sometimes used to easily differentiate between robusta and arabica species. Robusta leaves are larger than that of arabica with a length of 20-35 cm, and a width of 8-15 cm with a rubbery outer layer whereas the arabica leaf is half the length (10-15 cm) and width (4-6 cm) and is shinier in appearance (Oestreich-Janzen, 2010). In comparison, *C. liberica* has the largest leaves of all the

Stellenbosch University https://scholar.sun.ac.za

coffee species, nearly twice the dimensions of arabica (Teketay, 1998). Figure 2.2.1 illustrates the visual differences between the three mentioned species' leaves as well as the variation within subspecies.

Figure 2.2.1. Visual differences between the leaves of the coffee species *C. liberica*, *C. canephora*, and *C. arabica* (source: https://gcrmag.com/wp-content/uploads/2020/09/gcr-wcr-arabica-lrg.jpg).

A variety of colours exist for the coffee fruits which is highly dependent on the maturation stage of the fruit. The colours generally range from green (unripe) through to red and purple (ripe) to black (overripe), however some varieties appear light yellow or white (Teketay, 1998). All fruits are a fleshy berry containing two seeds or "beans" within and are deemed the essential part of the plant (Figure 2.2.2) (DAFF, 2012). During the dry processing method, the mesocarp is called the "husk" once removed during dehulling. However, during the wet processing method, the mesocarp is termed the "parchment" once removed during dehulling. Depending on the species, and sometimes subspecies, the pulp of the berry can either be bitter and unpalatable or it could have a sweet taste (Teketay, 1998). This fact generally goes together with the size of the fruit, for instance the larger the berry the sweeter the taste, and vice versa (Teketay, 1998). Coffee trees are harvested when they reach five years of age and every year thereafter if conditions are favourable (Mussatto *et al.*, 2010).

Stellenbosch University https://scholar.sun.ac.za

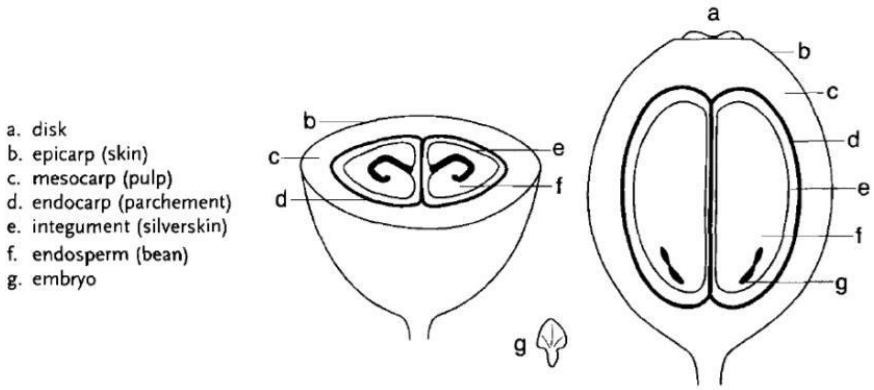


Figure 2.2.2. Transverse and longitudinal cross-sections of a coffee berry (Adapted from J. N. Wintgens, 2004).

For cultivation purposes, these three attributes play an important role for determining the yield of coffee fruit (DAFF, 2012; Teketay 1998). If a coffee tree is allowed to grow too tall, then less energy of the plant is given to producing good quantity and quality fruits. The physical properties of the coffee leaves do not influence the quality of the fruits produced. However, the leaves can provide the first signs of disease within a plant and therefore need to be monitored (Oestreich-Janzen, 2010). Coffee leaf rust, American leaf spot, and *Cercospora* spot are diseases commonly affecting coffee plants and, if left uncontrolled, can lead to lower yield of fruit (DAFF, 2012). For obvious reasons, the fruit's quality during cultivation is important. The fruits are also susceptible to disease and pest infestations, as well as scavenging by birds and insects (DAFF, 2012).

2.2.3 Ecological cultivation requirements

Environmental conditions which coffee crops are subjected to, are required to be maintained to ensure annual harvesting of good quality crops. Not all climates are suitable for coffee production and therefore the following discussion will outline the conditions that various coffee species require in order to flourish.

The differences between arabica and robusta coffees are not only seen in the final products, but also in terms of their requirement for specific climate conditions. *C. arabica* requires rainfall of *ca.* 1 520 to 2 280 mm each year, however that range extends from below *ca.* 760 mm to well over 2 540 mm depending on the species and location of the plantation (Oestreich-Janzen, 2010; Teketay, 1998). For *C. robusta*, the rainfall range is more standard between species and is from *ca.* 1 000 to 2 500 mm

Stellenbosch University https://scholar.sun.ac.za

annually (Oestreich-Janzen, 2010; Teketay, 1998). South African coffees are sensitive to water levels and require 1 500 mm rainfall per year with preferably a dry winter to promote flowering (DAFF, 2012).

Temperature is another aspect to be considered when trees are planted. Arabica coffees are tolerable of the lower temperature ranges (15-24 °C) whereas robusta coffees prefer warm temperatures (18-36 °C) (Oestreich-Janzen, 2010; Teketay, 1998). Low temperatures and frost can negatively influence the plant's growth and, in severe cases, inhibit growth (DAFF, 2012). Development and ripening of coffee fruit requires warmer temperatures often during summer times (DAFF, 2012; Oestreich-Janzen, 2010). The average temperature range for South African coffees is on the lower side, 12 to 26 °C, which is ideal for arabica plantations (DAFF, 2012). Temperature is directly influenced by altitude, in that for every rise in 1 000 m there is a loss of 6 °C (Teketay, 1998). For this reason, arabica plantations are found in the highland areas, for example Southwest Ethiopia, Kenya, and Brazil. Robusta plantations grow in lowlands such as in Java and Uganda (Oestreich-Janzen, 2010; Teketay, 1998).

A further ecological requirement of coffee plants includes the quality of soil that the cultivations are propagated and grown in. Coffee is produced in many countries across most of the continents, and so long as the soil is fertile and suitable climate conditions are "provided" the plants will grow well. However, Teketay (1998) reported that adequate drainage is also needed in order to prevent over-collection of water which is in line with another study (DAFF, 2012) that states that sandy-loam is the best suited soil type for growing coffee. Teketay (1998) further details the exceptions to normal parameters of soil types as well as discusses effect of wind, humidity, and cloud cover.

2.2.4 Geographical distribution

Many countries that produce and export coffee are located within the tropical regions surrounding the equator. This is commonly termed the "coffee belt" (Oestreich-Janzen, 2010) and is illustrated in Figure 2.2.3 for the top 36 countries (ICO, 2020a). Small-scale farmers, such as those in South Africa who do not fall within the belt region, are still capable of producing coffee if ecological requirements are suitable (see Ecological cultivation requirements above).

Stellenbosch University https://scholar.sun.ac.za

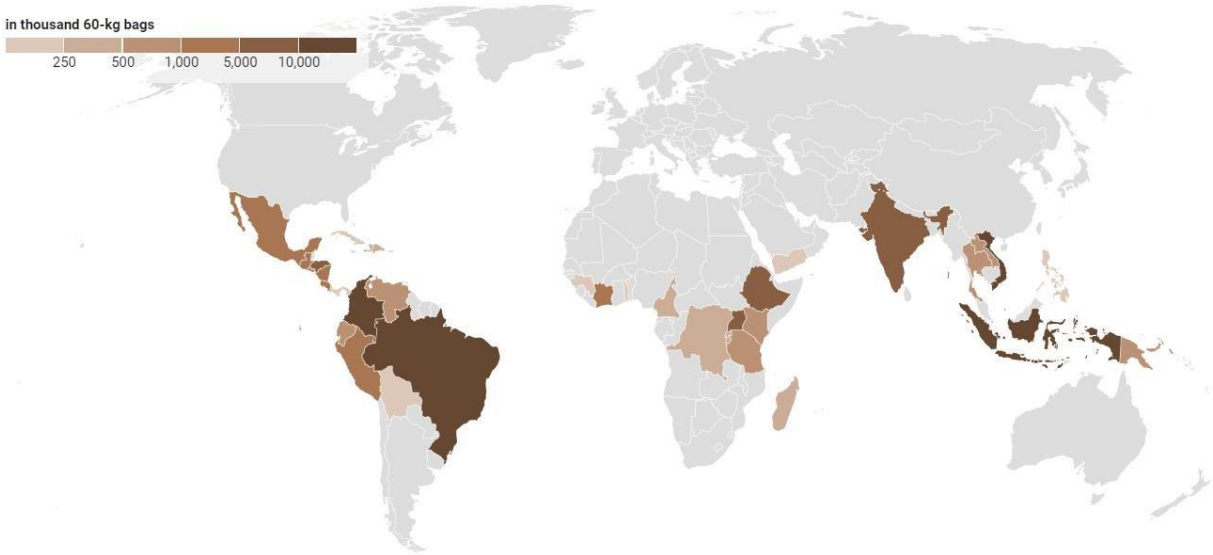


Figure 2.2.3. Locations of the major coffee producing countries during the 2019 production year (Data sourced from International Coffee Organization, 2020a; Map created with Datawrapper).

2.3 Processing of coffee

Coffee is the second most traded commodity in the world, after the trade of crude oil, and its production supplies more than 400 billion brewed cups annually (Mussatto et al., 2011). Its widespread popularity has motivated manufacturers to refine the process, and therefore it is important to understand the steps involved in the production of coffee, from berries to roasted beans.

2.3.1 Harvesting

The first stage in coffee processing begins with the harvesting of the fruits, which is generally done five years after planting (Mussatto *et al.*, 2011). Due to the nature in which coffee cherries mature, they are not always at the same stage of maturity when harvested. The consequence of this is that there are differing proportions of immature, ripe, and overripe fruit on the same tree (de Melo *et al.*, 2019; Waters *et al.*, 2017). There are typically three ways in which harvesting can occur, either by hand-picking, stripping or mechanically picking the cherries. Hand-picked coffee fruits are known to produce better quality coffees because workers select only the ripe cherries and limit the inclusion of immature or overripe fruit as well as sticks, insects and foreign materials (de Melo *et al.*, 2019). However, this method is equally time-consuming as it is labour intensive and is generally done for coffees that require a high final standard. Stripping is a more common method where all the cherries on a branch are

Stellenbosch University https://scholar.sun.ac.za

simultaneously removed and placed in baskets or on sheets below the trees (de Melo *et al.*, 2019). There is some exclusion of foreign material and unacceptable fruits by quick, visual inspection, however this is a quick process to gather the cherries and still produce coffee of acceptable qualities. Similar to stripping, mechanical picking has also grown in popularity due to the low labour costs and fast harvesting times. Mechanical picking relies on the vibration of tree branches to knock off fruits (Mussatto *et al.*, 2011; de Melo *et al.*, 2019) but this will also include the immature or overripe cherries as well as undesirable matter such as bird's nests and loose sticks or branches. Therefore, it is in the farmer's best interest to apply some degree of sorting after the harvesting stage to minimize unwanted materials.

2.3.2 Post-harvest methods

After harvesting, coffee processing can commence. There are two dominant methods used, the dry method which produces "natural coffees" and the wet method which produces "washed coffees" (Oestreich-Janzen, 2010; Selmar *et al.*, 2000; Waters *et al.*, 2017). In the 1990s, a combination of the two was also developed called the semi-dry method and results in a "pulped natural coffee" (Waters *et al.*, 2017). Typically, arabica beans are produced via the wet processing method in countries such as Colombia and Kenya and the state of Hawaii (Schwan *et al.*, 2012). In contrast, robusta coffees are usually processedwith the dry method, in areas such as Indonesia and Yemen due to the limited rainfall and dry, sunny days. Regions that deviate slightly from the normal practices are Brazil, Ethiopia, Haiti and Paraguay which process both Robusta and Arabica coffee by meansof the wet method (Schwan *et al.*, 2012).

2.3.2.1 Wet processing

The wet processing technique requires coffee fruits that are at full maturity because the skin and pulp can easily be removed from the cherry. The cherries are first depulped in order to remove the pulp layer surrounding the coffee bean. The beans are placed in tanks where the mucilage layer (a sticky outer layer) is degraded by natural microbial fermentation and followed by a washing step (de Melo *et al.*, 2019; Oestreich-Janzen, 2010; Tarzia *et al.*, 2010). There are inconsistent reports of the required time period for fermentation. Oestreich-Janzen (2010) reported that between 18 and 36 hours was required, whereas Waters *et al.* (2017) states that 24 to 48 hours is sufficient. Ideally, the producers need to assess the level of mucilage removed and decide on a timeframe most appropriate for the bean cultivar being processed. However, the time should

Stellenbosch University https://scholar.sun.ac.za

preferably be kept to a minimum to avoid microbial spoilage or over-fermentation which can lead to off flavours in the coffee. Once washed, the coffee seeds are dried either naturally under the sun or by means of forced airflow to a final moisture content between 10% to 12% (de Melo *et al.*, 2019; Tarzia *et al.*, 2010; Waters *et al.*, 2017).

2.3.2.2 Dry processing

Fruits from all stages of maturation are used during the dry processing method, which is a much simpler process that requires less labour and machinery (Vincent, 1987). Once the whole fruits are harvested, they are immediately dried in the same manner as the wet method, either exposed to the sun or placed in air dryers with forced airflow. One study described a prior wash-and-sort method to separate the different cherries by density and therefore kept cherries of similar maturity together (Schwan *et al.*, 2012). This, however, would negate the supposed ease of the process and would suggest a wet process was followed and possible fermentation could have occurred. The length of drying is also dependent on the producer as well as the conditions of drying. Forced airflow can take between 3 and 9 days (Oestreich-Janzen, 2010), whereas sun drying is between 10 and 25 days (Waters *et al.*, 2017; Schwan *et al.*, 2012). The important factor is that the final moisture content should drop to between 10% to 12% (de Melo *et al.*, 2019; Tarzia *et al.*, 2010; Waters *et al.*, 2017). Due to the type of processing steps in the dry method, the dried seeds are dehulled to remove the husks. Whereas during the wet method, the seeds are dehulled to remove the parchment layer. In either case, this renders the green beans ready for roasting.

2.3.2.3 Semi-dry processing

The semi-dry method is an alternative way of processing coffee beans. Instead of adhering completely to the dry method, which subjects whole fruits to drying, it follows the wet method initially by pulping the fruit to remove the cherry skin and pulp, but then is submitted to drying with the mucilage layer still surrounding the beans (Selmar *et al.*, 2006). This method is generally used for specialty coffees and/or farmers with the necessary machinery who wish to reduce their water usage associated with the fermentation stage. The resulting coffee quality lies somewhere between that of dry processed coffee and wet processed coffee, as the name would suggest.

2.3.2.4 Germination

There is a lack of research on germination as an intentional post-harvest treatment for coffee beans due to the simple reason that coffee producers do not apply this "extended

Stellenbosch University https://scholar.sun.ac.za

fermentation" step in their processing. It should be noted that research has been conducted on metabolic germination that occurs naturally during any of the processing methods (Bytof *et al.*, 2007; Selmar *et al.*, 2006). However, this germination does not proceed to the same extent as what is being proposed. As mentioned earlier, during the wet processing method, fermentation is meant to be kept as short as possible to prevent over fermentation and or microbial spoilage. A coffee farm in Port Edward, South Africa, has recently explored "extending" the fermentation stage which reportedly led to a final brewed coffee that was sweeter in taste (Cummings, D. 2019, Owner, Beaver Creek Coffee Estate, Port Edward, South Africa, personal communication, 19 March). In order to consider coffee bean germination as a post-harvest treatment during the processing of coffee, one should consider the origin of the germination process.

Selmar *et al.* (2006) proposed that germination occurs both at the seed infancy stage when planted into the ground, but also after the cherries have been harvested for processing. Once the fruit has reached maturity on the tree, the metabolic activity changes are no longer due to maturation but instead attributed to germination (Bewley & Black, 1994; Selmar *et al.*, 2006). However, this is a stress response of the coffee seeds and, in order to differentiate between natural and desired germination, this should be classified as metabolic germination. Whereas in Port Edward, desired germination is occurring whereby the seeds are allowed to germinate after the mucilage layer has been degraded. Metabolic germination is based on the seeds' response to the immediate environment. For instance, while the coffee fruit is still ripe on the trees, germination is inhibited due to the combination effect of abscisic acid and the osmotic potential found in the cherry pulp surrounding the seed (Bewley & Black, 1994). However, by de-pulping the cherry, and after the addition of water and oxygen, the seed undergoes both fermentation as well as initial germination reactions. After the fermentation period in wet coffee processing, the seeds are spread out to dry. The drying stage causes a different stress response, whereby the seed goes into dormancy to preserve itself (Bewley & Black, 1994) and thus preventing the germination process from continuing. If this germination continued, isocitrate lyase expression would occur which is a key enzyme during germination. This can be measured by the increase in β-tubulin, a marker for cell division or elongation (Bytof *et al.*, 2007).

The induced or desired germination is achieved by washing the mucilage residue from the fermented beans, thus removing majority of the yeast, bacteria and

Stellenbosch University https://scholar.sun.ac.za

filamentous fungi (Pereira *et al.*, 2016) responsible for fermentation. The washed beans are stored in a high humidity environment and monitored daily (Cummings, D. 2019, Owner, Beaver Creek Coffee Estate, Port Edward, South Africa, personal communication, 19 March). At the time of writing, the germination period was allowed to proceed for five days (if sprouting occurs then germination has been exceeded) after which drying commenced as per the normal method.

2.3.3 Roasting

The three distinct organoleptic properties of coffee (flavour, aroma and colour) all develop during the roasting stage of processing via complex chemical and biological reactions (Mussato *et al.*, 2011). Green coffee is placed in sealed units within the roaster and then subjected to temperatures reaching as high as 240 °C (Gonzalez-Rios *et al.,* 2007a). The time-temperature combination is important, and can result in either a light, medium or dark roasted coffee. The heat transferred to the beans will begin the transformation of various precursor compounds into products that have become synonymous with coffee. The dominating processes are the cascade of condensation, degradation and oxidative polymerization steps associated with the Maillard reaction (Oestreich-Janzen, 2010) that produce the characteristic nutty/roasted aroma of coffee. Different degrees of roasting will follow the same reaction pathways but instead produce varying amounts of the final products (ketones, furans, and pyrazines). The degree of roasting and the influences on chemical composition of coffee have been previously reported (Baggenstoss *et al.*, 2008).

2.4 Chemical components

The chemical composition of coffee beans has been extensively reported (Buffo & Cardelli-Freire, 2004; Gonzalez-Rios *et al.*, 2007a; Higdon & Frei, 2006; Oestreich-Janzen, 2010; Sunarharum, 2016). The presence of certain compounds before and after the roasting step as well as their respective concentrations, can be related to beneficial or undesirable sensory traits in the final brewed cup of coffee. The key sensory parameters for coffee are aroma and flavour, and are linked to volatile and non-volatile components, respectively. The chemical components in coffee beans can be categorized into these two groups.

Non-volatile components are responsible for the flavour or taste of the coffee. These compounds can differ in concentration levels due to the different effects of pre- and

Stellenbosch University https://scholar.sun.ac.za

post-harvest influences. The compound concentration levels also differ between the green and roasted stages of the beans. Nonetheless, these components can be grouped into different classes, namely: nitrogenous compounds, carbohydrates, chlorogenic acids, organic acids, lipids, and other trace compounds (Buffo & Cardelli-Freire, 2004; Oestreich-Janzen, 2010; Ribeiro *et al.*, 2009). These classes contain important precursor compounds that take part in various reactions during roasting to produce the characteristic aroma of coffee.

2.4.1 Nitrogenous compounds

The nitrogenous group is home to three well-known compounds, caffeine, trigonelline, and amino acids. Caffeine (1, 3, 7-trimethylxanthine) is naturally found in coffee beans and is famous for its stimulatory effects on the nervous system and brain (Higdon & Frei, 2006). Notably, the quality of a coffee beverage is often associated with its caffeine content and therefore becomes an important aspect when grading coffee. In terms of its contribution to a brewed cup of coffee it is believed to add to the overall strength and body and is only partially responsible for the bitterness associated with coffee (Oestreich-Janzen, 2010; Sunarharum, 2016). According to Oestreich-Janzen (2010), during the roasting process the caffeine content remains relatively unchanged. However, this is contradictory to a study by Bayle and Adamu (2019) who found that the caffeine content changed depending on the time-temperature combination that was used. A trend was observed where the content increased until a point (175 °C) and was then followed by a decrease in caffeine content. The difference in the outcomes of the two studies is due to their stated aims. Oestreich-Janzen (2010) was comparing content levels between that of green and roasted coffee, whereas Bayle and Adamu (2019) were studying the effects of roasting at different temperatures on the caffeine levels. Although this is evidence that roasting plays a role in the concentration levels of caffeine, if not for other compounds as well, the change in caffeine levels were insignificant for Bayle and Adamu (2019). Therefore, it could be considered negligible in the broader context of this work.

Trigonelline (N-methylpyridinium-3-carboxylate) plays a role in aromatic development of coffee whereby, during roasting, thermal degradation of the compound leads to the production of pyridines and pyrroles (Oestreich-Janzen, 2010; Ribeiro *et al.*, 2009). One of its degradation products, nicotinic acid, is an important vitamin (B3 or niacin) which contributes to the nutritional value of coffee (Buffo & Cardelli-Freire,

Stellenbosch University https://scholar.sun.ac.za

2004). Unlike caffeine, the trigonelline concentration can undergo a drastic reduction during roasting, and is influenced by the temperature of the roast (Oestreich-Janzen, 2010). At lower roasting temperatures approximately 60% can be lost whereas at higher temperatures up to 90% can be lost to degradation (Oestreich-Janzen, 2010). Trigonelline is therefore thermally labile.

Amino acids play a vital role in the Maillard reaction where the amine group undergoes a condensation reaction with the carboxyl group of reducing sugars to form aldehydes and ketones (Sunarharum, 2016; Thaler, 1978). The products impart the brown colour to the roasted coffee beans but also carry the volatile components, thus contributing to the aroma. Amino acids, along with an alpha-dicarbonyl, also partake in the Strecker degradation reaction which produces numerous compounds, such as aldehydes, pyrazines, pyrroles and pyridines (Ribeiro *et al*, 2009), all contributing in different ways to the volatile profile of coffees.

2.4.2 Carbohydrates

In green coffee beans, polysaccharides are the dominating group of carbohydrates. Thaler (1978) claimed that mannan and galactan were the predominant compounds, present in coffee at equal proportions, and that araban and glucan were present at 1-3% of the extracts. A more recent paper, with improved analytical methods, reported that arabinogalactan type II and galactomannans were the two predominant types (Oosterveld *et al*., 2003). These polysaccharides are responsible for the viscosity of brewed coffee (Buffo & Cardelli-Freire, 2004) but also undergo various reactions during roasting, such as thermal and pyrolytic degradation to produce furans and carboxylic acids, and ketones respectively (Oestreich-Janzen, 2010; Ribeiro *et al.*, 2009). More importantly, simple carbohydrates can react with proteins and amino acids in a significant reaction called the Maillard reaction. This reaction results in the brown-coloured melanoidin products, and also nitrogen and sulfur containing compounds, which are linked to coffee flavour (Sunarharum, 2016).

2.4.3 Chlorogenic acid and Organic acids

Chlorogenic acids (CGAs) are considered a family of esters derived from the parent structure compounds of quinic acid and cinnamic acid. These parent compounds can have varying substitutions at the aromatic ring to form numerous isomers and epimers (Oestreich-Janzen, 2010). 5-O-caffeoylquinic acid is the most abundant CGA and has been attributed to the astringency of coffee beverages (Buffo & Cardelli-Freire, 2004). It

Stellenbosch University https://scholar.sun.ac.za

is common practice to refer to 5-O-caffeoylquinic acid as "chlorogenic acid". The result due to degradation of CGAs is the production of phenolic compounds (Ribeiro *et al.*, 2009). A detailed analysis of green robusta coffee was conducted by Jaiswal *et al.* (2010) which identified and confirmed 69 chlorogenic acids as well as classifying seven new CGA classes in coffee samples.

Although chlorogenic acid is also an organic acid, it is usually discussed separately from the group due to its abundance in coffee and the sheer number found. Aside from chlorogenic acids, the other organic acids found in coffee appear in much lower proportions. Citric acid and malic acid are the next abundant organic acids but only account for 1.3% and 0.3% of the dry weight (Oestreich-Janzen, 2010). It should be noted that in terms of total titratable acidity, chlorogenic acid only contributes 8% whereas citric acid accounts for 20% with the largest amount being for acetic acid, 25% (Oestreich-Janzen, 2010). Some organic acids are produced during the roasting process which causes carbohydrate compounds to degrade and produce acetic, formic, glycolic, and lactic acid (Oestreich-Janzen, 2010; Sunarharum, 2016). Whereas ascorbic acid breaks down during roasting to create furan compounds (Sunarharum, 2016).

2.4.4 Trace compounds

Other trace compounds found in both green and roasted coffee include the mineral composition with potassium forming most of the content. However, manganese, iron, and copper are also present in reasonable amounts (10-50 ug/g, 15-40 ug/g, and 2-5 ug/g, respectively) (Buffo & Cardelli-Freire, 2004; Sunarharum, 2016). They are often defined as catalytic compounds in biochemical reactions because they assist in the release of compounds required for flavour development (Sunarharum, 2016).

2.5 Quality of coffee

2.5.1 Defining quality

The coffee plant and fruit, like all species, has evolved since it was first discovered, and therefore defining quality needs to adapt as well. Belay *et al.* (2016) discussed that there is variation in quality parameters between the different levels of the supply chain. The different levels include the farmer, the exporter/importer, the roaster, and the consumer. In some instances, the farmer, exporter, and roaster are one entity and therefore quality standards are easily achieved. A farmer's key concerns for quality would include price and production efficiency, whereas the exporter/importer would

Stellenbosch University https://scholar.sun.ac.za

assess lack of defects and weight of a shipment (Belay *et al.*, 2016). The moisture content, price and origin would be important at the roasting level, and lastly, consumers are motivated by taste and flavour along with health implications and pricing (Belay *et al.*, 2016).

Sylvian (1958) concluded that quality was characterized by physical and chemical attributes of the coffee beans after roasting. Oetreich-Janzen (2016) stated that "good quality coffee flavour has a pleasant sensation, a balanced combination of flavour, body, and aroma in the absence of faults". The latter definition is quite broad and does not include the preference of the individual experiencing the coffee. Defining 'coffee quality' can further be a challenge because each individual will have his or her own preference for how they drink coffee. The present study proposes that instead of defining coffee quality as a whole, one should identify key individual components within coffee and relate them to descriptive sensory attributes. For example, a high concentration of sugar is likely to correlate to a sweet coffee flavour. Whether a person enjoys sweet coffee or not, is therefore subjective. Sensorial analysis of coffee beverages has been studied in terms of both aroma and flavour (Sunarharum, 2016). Previously identified compounds (section 2.4) have different contributions to the overall quality of coffee and have been summarized in Table 2.5.1.

Table 2.5.1. Identified key compounds and the influence on quality attributes of coffee.

Compound	Aroma	Flavour	Mouth-feel	Health benefits
Caffeine[a,b,d,f,g]		✓		✓
Trigonelline[b,d,e,f,g]	✓			
Polysaccharides[a,b,d,e,f,g,h]	✓	✓	✓	
Chlorogenic acids[a,b,c,d,g]		✓		✓
Nicotinic acid[a,b,c,f]		✓		✓
Lipids[a,b,d,e,f]		✓	✓	

[a]Buffo & Cardelli-Freire (2004), [b]Farah (2012), [c]Higdon & Frei (2006), [d]Oestreich-Janzen (2010), [e]Ribeiro *et al.* (2009), [f]Selmar *et al.* (2006), [g]Sunarharum (2016), [h]Tarzia *et al.* (2010).

Stellenbosch University https://scholar.sun.ac.za

2.5.2 Governing factors

Of all the quality aspects mentioned, two main factors govern them: 1) inherent features of the beans and 2) the handling of the beans from harvesting to market (Sylvian, 1958). Physical aspects of a coffee bean include the external traits such as shape, size and uniformity and the chemical nature includes aroma, flavour and body (Sylvian, 1958). Genetic composition is not easily changed, and careful consideration should be given to the species and cultivar types as well as their desired climate conditions. An easier manner to control quality is adapting the processing treatment for maximized beneficial gains. The effect of the post-harvest treatment selected for coffee processing is well known and has been previously investigated. Arabica coffees are well suited for the wet method of processing, and the Robusta species is generally processed with the dry method. There is a consensus among studies that the wet method produces coffees of higher quality with a full, rich aroma, pleasant acidity and less body compared to that of dry methods which result in full, fruitier body (Mussatto *et al.*, 2011; Selmar *etal.*, 2000; Tarzia *et al.*, 2010). The products of the semi-dry method exhibit attributes that are intermediary to wet and dry methods. A reason for this is the lack of a fermenting stage and the remaining pulp on the bean which prevents the polysaccharides from being degraded to the full extent (Farah, 2012). Oestreich-Janzen (2010) suggested that the variation observed between the different modes of processing is as a result of different time durations of the chemical reactions taking place. This statement is supported by the findings of Bytof *et al.* (2007) who identified the highest levels of isocitrate lyase (ICL) – an enzyme found during fermentation – in wet processed beans two days after commencing the treatment compared to dry processed beans which recorded highest levels of ICL only on the seventh day after processing. Roasting is an essential stage in coffee processing and has an important influence on the quality parameters because the organoleptic properties (flavour, colour, aroma) develop here (Mussatti *et al.*, 2011).

2.6 Analytical techniques

Over the years, different analytical techniques have been developed to examine coffee seeds and beverages in terms of chemical composition, quality control and sensory aspects (Alonso-Salces *et al.*, 2009; Barbin *et al.*, 2014; Correia *et al.*, 2018; Perrone *et al.*, 2008; Sunarharum, 2014). Techniques range from simple liquid chromatography-mass spectrometry (LCMS) to more complex, hyphenated systems such as liquid

Stellenbosch University https://scholar.sun.ac.za

chromatography coupled to a photodiode array detector with an electrospray ionization and mass spectrometer (LC-DAD/ESI-MS) but the type of technique used is dependent on the aim of the research. Furthermore, the complexity of the data that is generated requires multivariate data analysis to accurately interpret and report the results (Lavine & Workman, 2013).

2.6.1 Instrumentation

Chromatographic techniques are widely used to study and identify different compounds within a sample, including at the trace level (Nolvachai *et al.,* 2017). The principle of chromatography is the separation of compounds from one another based on their affinity for either a moving or stationary phase. Liquid chromatography (LC) and gas chromatography (GC) are the most abundantly used techniques for non-volatile and volatile compound analysis, respectively. Liquid Chromatography and Gas Chromatography methods have both been used to analyze food composition, each having advantages for certain food groups. For instance, GC is mostly applied for oils, fatty acids, and aroma component detection whereas LC is useful for detecting inorganic salts, proteins, and polysaccharides (Lehotay & Hajslova, 2002). Gas Chromatography is a well-established technique which has been used for assessing quality parameters for food aromas in various products, e.g. coffee, strawberries, and rice (Du *et al.*, 2011; Mahattanatawee & Rouseff, 2014; Pua *et al.*, 2020). However, GC analysis has slowly become overshadowed by LC analyses.

Liquid Chromatography methods have become more widely accepted for the majority of food analyses in publications since the start of the century, even for traditional GC applications (Lehotay & Hajslova, 2002). This is because it can separate all organic compounds irrespective of volatility or polarity characteristics (Lehotay & Hajslova, 2002) and over the years these instruments have only become more sophisticated. The separation of various types of compounds can be achieved by using different solvents and/or columns that are best suited for the compounds of interest. Typically, two solvents are used for separation: 1) LC grade water (ultrapure water) with 0.1% acid and 2) an organic solvent. Popular organic solvents include methanol and acetonitrile (Figure 2.6.1). Methanol is classified as a polar-protic solvent because it is able to form hydrogen bonds either with itself or other compounds due to the presence of a hydrogen connected to an electronegative atom (O-H in Figure 2.6.1A). Whereas acetonitrile is classified as a polar-aprotic solvent because it cannot form

Stellenbosch University https://scholar.sun.ac.za

hydrogen bonds (Figure 2.6.1B). Formic acid or acetic acid are generally used as the acid of choice for LC methods because it provides a source of protons which is suitable for ionizing analytes, as well as improving the resolution of peaks. The coupling of additional detectors to LC devices can also provide clearer, more informative data about a compound.

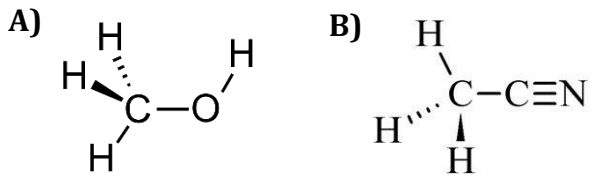
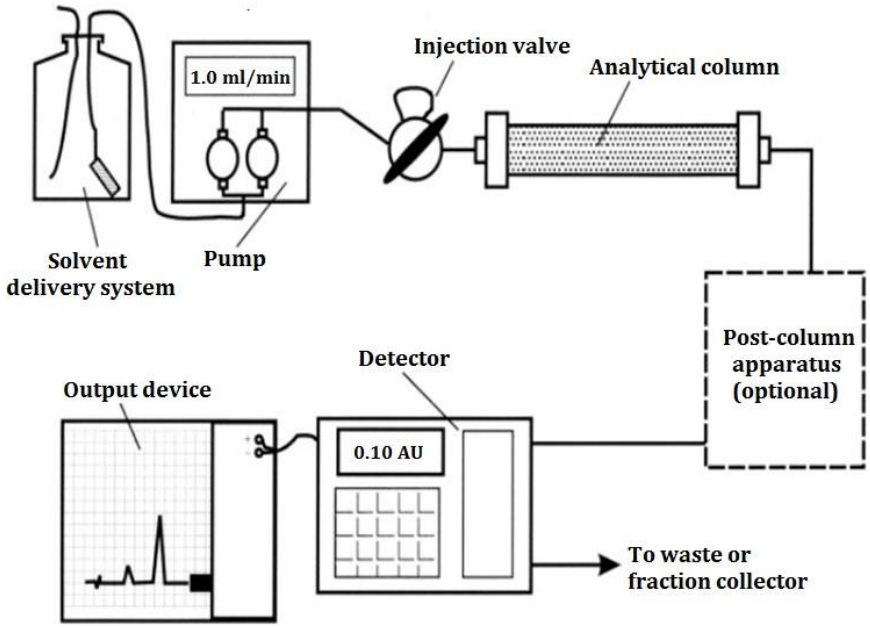


Figure 2.6.1. Structural illustration of the compounds A) methanol and B) acetonitrile to illustrate the presence and absence of an electronegative atom, respectively.

Spectrometry is a different mode for the detection of analytes that is simple, fast and capable of analyzing a series of analytes at one given time for a sample (Barbin *et al.*, 2014). This is advantageous as it negates the need for multiple sample preparations. Spectrometry techniques involved the application of spectroscopy, which is based on the fundamental principle that light of a certain wavelength will interact uniquely with different bonds in a functional group of a molecule (Correia *et al.*, 2018). The most popular type is mass spectrometry (MS) due to its ability to obtain structural information of compounds (Hites, 2016). The quantitative and qualitative information of complex systems, such as coffee, can be obtained from chromatographic techniques coupled to detectors with increased selectivity and sensitivity, such as mass spectrometry or photo diode array detectors (Nolvachai *et al.*, 2017). A study by Alonso-Salces *et al.* (2009) made use of multiple detectors (LC-DAD/ESI-CID-MS/MS) to examine various polyphenols and methylxanthines in coffee extracts to elucidate compound structures for undiscovered analytes in coffee. The complementary information allowed the authors to confirm the presence of ten new, natural compounds in the coffee samples. In comparison, research by Perrone *et al.* (2008) was solely focused on simultaneously quantifying target compounds known in coffee compared to previous studies that individually quantified the same analytes. This simultaneous method had a short run- time (6 min) and high accuracy (within 15% unity) (Perrone *et al.*, 2008). The instrument used was a simplified LC system coupled to an MS detector

Stellenbosch University https://scholar.sun.ac.za

which provideda rapid analysis, often necessary in industry. The basic setup for an LC instrument can be seen in Figure 2.6.2 with the option of additional devices indicated. Examples of post-column apparatuses include mass spectrometers and quadruple time-of-flight devices. The application range of chromatography and spectrometry is endless and has been applied within various industries, including pharmaceutical, biochemistry and food technology. However, with such information-rich data being produced,



appropriate dataanalysis techniques are also required.

Figure 2.6.2. The basic schematic of a liquid chromatography device (adapted from LaCourse, 2017).

2.6.2 Data analysis

Data analysis involves applying statistical or logical approaches in order to represent data in a manner that is easy for evaluation. Before data analysis can be applied, pre-processing the data is necessary in order to remove unwanted variation. For example, light scattering, background noise or sample outliers may negatively impact or skew results. Once the data have been 'cleaned', the next phase is to undergo multivariate analysis (MVA). Numerous studies have shown the outcomes of using different MVA techniques for qualitative and quantitative analysis (Verboven *et al.*, 2012; Calvini *et al.*, 2015; Liu *et al.*, 2015). Principal component regression and partial least squares regression are most often used for developing robust models for quantitative purposes because they are capable of reducing large amounts of data into fewer variables while

Stellenbosch University https://scholar.sun.ac.za

still accurately predicting quality attributes (Verboven *et al*., 2012; Liu *et al.*, 2015). Furthermore, for qualitative analysis the techniques available are based on computer intelligence that are designed to predict future relationships from a training dataset. These qualitative techniques include Gaussian mixture models, K-means, principal component analysis, K-nearest neighbours, support vector machine, and orthogonal partial least squares discriminant analysis (Caporaso *et al.*, 2018; Smrke *et al.*, 2015; Verboven *et al.*, 2012). A key difference within these techniques are whether they are supervised, i.e. require prior knowledge of the dataset attributes, or unsupervised, i.e. no prior knowledge of the data (Liu *et al.*, 2015). However, the core principle of these techniques is to observe trends in the data that can be linked to key features (Liu *et al.*, 2015). Comparisons of different pre-processing methods have been researched (Verboven *et al.*, 2012). Selection of the appropriate pre-processing method is subjective to the type of dataset that is being analyzed and not one pre-processing is perfect for all conditions. Therefore, sufficient time is required to properly assess all the possible techniques, as mentioned earlier, in order to not misrepresent the results.

2.6.3 Summary

Coffee is a popular commodity, both as a beverage and as an item of interest for researchers. Studies have focused on three main aspects: 1) effect of different post-harvesting methods; 2) the chemical composition of coffee; and 3) the quality of coffee based on chemical and sensorial attributes. Other than the establishment of the semi-dry processing method in the 1990s, no other post-harvest treatment has been recorded in recent times. Germination has been noted to occur in coffee beans during processing as an unintentional reaction. The aim of the proposed project is to determine if germination has an effect as an *intentional* post-harvest treatment and what chemical changes are occurring.

Stellenbosch University https://scholar.sun.ac.za

References

Baggenstoss, J., Poisson, L., Kaegi, R. Perren, R & Escher, F. (2008). Coffee roasting and aroma formation: application of different time-temperature conditions. *Journal of Agricultural and Food Chemistry*, **56**, 5836-5846.

Barbin, D. F., Felicio, A. L. d.-S. M., Sun, D.-W., 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.

Bayle, E. A. & Adamu, D. B. (2019). Chemical characterization of green coffee beans and determining the effect of roasting temperature on the content of caffeine. *Chemistry and Materials Research*, **11**(8), 14-21

Belay, S., Mideksa, D., Gebrezgiabher, S. & Seifu, W. (2016). Factors affecting coffee (*Coffea Arabica* L.) quality in Ethiopia: A review. *Journal of Multidisciplinary Scientific Research*, **4**(1), 22-28.

Bewley, J. D. & Black, M. (1994). Seeds. Physiology of Development and Germination. New York, London: Plenum Press.

Buffo, R. A. & Cardelli-Freire, C. (2004). Coffee flavour: an overview. *Flavour and Fragrance Journal*, **19**, 99-104.

Bytof, G., Knopp, S.-E., Kramer, D., Breitenstein, B., Bergervoet, J. H. W., Groot, S. P. C. & Selmar, D. (2007). Transient occurrence of seed germination processes during coffee post-harvest treatment. *Annals of Botany*, **100**, 61-66.

Correia, R. M., Tosato, F., Domingos, E. Rodrigues, R. R. T., Aquino, L. F. M., Filgueiras, P. R., Lacerda, V. Jr. & Romão, W. (2018). Portable near infrared spectroscopy applied to quality control of Brazilian coffee. *Talanta*, **176**, 59-68.

Crawford, J. (1852). History of coffee. *Journal of the Statistical Society of London*, **15**(1), 50-58.

de Melo Pereira, G. V., de Carvalho Neto, D. P., Magalhães Júnior, A. I., Vásquez, Z. S., Medeiros, A. B. P., Vandenberghe, L. P. S. & Soccol, C. R. (2019). Exploring the impacts of postharvest processing on the aroma formation of coffee beans – A review. *Food Chemistry*, **272**, 441-452.

Department of Agriculture, Forestry and Fisheries. (2012). Production Guideline: Coffee (Coffea spp.). Pp. 1-18.

Du, X., Plotto, A., Baldwin, E. & Rouseff, R. (2011). Evaluation of volatiles from two subtropical strawberry cultivars using GC-Olfactometry, GC-MS odor activity values, and sensory analysis. *Journal of Agricultural and Food Chemistry*, **59**, 12569-12577.

Stellenbosch University https://scholar.sun.ac.za

Gonzalez-Rios, O., Suarez-Quiroz, M. L., Boulanger, R., Barel, M., Guyot, B., Guiraud, J.-P. & Schorr-Galindo, S. (2007a). Impact of "ecological" post-harvest processing on coffee aroma: II. Roasted coffee. *Journal of Food Composition and Analysis,* **20**(4), 297-307.

International Coffee Organization. (2020a). *Crop year production*. [Internet document]. URL http://www.ico.org/prices/po-production.pdf. Accessed 13/10/2020.

International Coffee Organization. (2020b). *World coffee consumption*. [Internet document]. URL http://www.ico.org/prices/new-consumption-table.pdf. Accessed 13/10/2020.

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**, 8722-8737.

LaCourse, W. R. (2017). HPLC Instrumentation. In: *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*. Elsevier.

Lavine, B. K. & Workman, J. (2013). Chemometrics. *Analytical Chemistry*, **85**(2), 705-714.

Lehotay, S. & Hajslova, J. (2002). Application of gas chromatography in food analysis. *Trends in Analytical Chemistry*, **21**, 686-697.

Luttinger, N. & Dicum, G. (2006). The Coffee Book: anatomy of an industry from crop to the last drop. The New Press, New York.

Mahattanatawee, K. & Rouseff, R. L. (2014). Comparison of aroma active and sulfur volatiles in three fragrant rice cultivars using GC-Olfactometry and GC-PFPD. *Food Chemistry*, **154**, 1-6.

Mussatto, S. I., Machado, E. M. S., Martins, S. & Teixeira, J. A. (2011). Production, Composition, and Application of Coffee and its Industrial Residues. Food and Bioprocess Technology, **4**, 661-672.

Nolvachai, Y., Kulsing, C. & Marriott, P. J. (2017). Multidimensional gas chromatography in food analysis. *Trends in Analytical Chemistry*, **96**, 124-137.

Oestreich-Janzen, S. (2010). Chemistry of coffee. In: *Comprehensive Natural Products II: Biology and Chemistry,* 3rd ed. (edited by M. Lew & L. Hung-Wen). 1085-1117. Oxford: Elsevier.

Oosterveld, A., Voragen, A. G. J. & Schols, H. A. (2003). Effect of roasting on the carbohydrate composition of Coffea Arabica beans. *Carbohydrate Polymers*, **54**, 183-192.

Pereira, G. V., Soccol, V. T. & Soccol, C. R. (2016). Current state of research on cocoa and coffee fermentations. *Current Opinion in Food Science*, **7**, 50-57.

Stellenbosch University https://scholar.sun.ac.za

Perrone, D., Donangelo, C. M. & Farah, A. (2008). Fasti simultaneous analysis of caffeine, trigonelline, nicotinic acid and sucrose in coffee by liquid chromatography-mass spectrometry. *Food Chemistry*, **110**, 1030-1035.

Pua, A., Lau, H., Liu, S. Q., Tan, L. P., Goh, R. M. V., Lassabliere, B., Leong, K.-C., Sun, J., Cornuz, M. & Yu, B. (2020). Improved detection of key odourants in Arabica coffee using gas chromatography-olfactometry in combination with low energy electron ionization gas chromatography-quadrupole time-of-flight mass spectrometry. *Food Chemistry*, **302**, 1-16.

Ribeiro, J. S., Augusto, F., Salva, T. J. G., Thomaziello, R. A. & Ferreira, M. M. C. (2009). Prediction of sensory properties of Brazilian Arabica roasted coffees by headspace solid phase microextraction-gas chromatography and partial least squares. *Analytica Chimica Acta*, **634**, 172-179.

Schwan, R. F., Silva, C. F. & Batista, L. R. (2012). Coffee fermentation. In: *Handbook of Plant-Based Fermented Food and Beverage Technology,* 2nd ed. (edited by Y. H. Hui & E. Özgül Evranuz). 677-690. United States of America: CRC Press.

Selmar, D., Bytof, G., Knopp, S. E. & Breitenstein, B. (2006). Germination of coffee seeds and its significance for coffee quality. *Plant Biology*, **8**, 260-264.

Selmar, D., Kleinwachter, M. & Bytof, G. (2000). New aspects of coffee processing: How do the different post harvest treatments influence the formation of potential flavour precursors? *Journal of Applied Botany*, **74**, 131-136.

Silvarola, M. B., Mazzafera, P. and Fazuoli, L. C. (2004). A naturally decaffeinated arabica coffee. *Nature*, **249**, 826.

Smith R.F. (1985) A History of Coffee. In: *Coffee*. (edited by M. N. Clifford & K. C. Willson). 1-12. Boston, MA: Springer.

Stander, M. A., Van Wyk, B.-E., Taylor, M. J. C. & Long, H. S. (2017). Analysis of phenolic compounds in Rooibos tea (Aspalathus linearis) with a comparison of flavonoid-based compounds in natural populations of plants from different regions. Journal of Agricultural and Food Chemistry, **65**, 10270-10281.

Sunarharum, W. B., Williams, D. J. & Smyth, H. E. (2014). Complexity of coffee flavor: a compositional and sensory perspective. *Food Research International*, **62**, 315-325.

Tarzia, A., dos Santos Scholz, M. B. & de Oliveira Petkowicz, C. L. (2010). Influence of the postharvest processing method on polysaccharides and coffee beverages. *International Journal of Food Science and Technology*, **45**, 2167-2175.

Teketay, D. (1999). History, botany and ecological requirements of coffee. *Walia*, **20**, 28-50.

Stellenbosch University https://scholar.sun.ac.za

Thaler, H. (1978). The chemistry of coffee extraction in relation to polysaccharides. *Food Chemistry*, **4**, 13-22.

van der Merwe, K. & Maree, T. (2016). The behavioral intentions of specialty coffee consumers in South Africa. *International Journal of Consumer Studies*, **40**, 501-508.

Vincent, J. C. (1987). Green coffee processing. In: *Coffee*, 2nd ed. (edited by R. J. Clarke & R. Macrae). 1-33. Springer, Dordrecht.

Waters, D. M., Arendt, E. K. & Moroni, A. V. (2017). Overview on the mechanisms of coffee germination and fermentation and their significance for coffee and coffee beverage quality. *Critical Reviews in Food Science and Nutrition*, **57**(2), 259-274.

Stellenbosch University https://scholar.sun.ac.za

3 Methods & Materials

3.1 Materials

3.1.1 Chemicals

HPLC-grade methanol, formic acid (analytical grade) and the caffeine standard (99% purity) were generously supplied by the Mass Spectrometry Unit (Central Analytical Facility, University of Stellenbosch) and originally obtained from Sigma-Aldrich. The reference standards for chlorogenic acid (98% purity), nicotinic acid (99% purity), and trigonelline (research chemical) were purchased from Protea Laboratory Solutions (Midrand, Johannesburg, South Africa). HPLC-grade water was prepared using a Millipore water purification system (Merck, Germany).

Each reference standard was individually prepared to a concentration of 1 000 ppm. A 250 ppm stock solution containing all the standards was then prepared, and subsequent dilutions made to the following concentrations (ppm): 5, 10, 20, 50, 100, and 200. The solvent used throughout was a standardized 50% (v/v) methanol-water with 0.1% (v/v) formic acid mixture.

3.1.2 Samples

Green and medium roasted coffee beans from six different batches were obtained from Beaver Creek Coffee Estate (Port Edward, Durban, South Africa). The roasting and germination processing parameters cannot be revealed for confidentiality reasons. The germination treatment was performed as part of the coffee post-harvest process. All samples were of the same species and varietiy (*C. arabica*, Catuai). Harvesting and processing occurred during two periods, May until August in 2019 and June until August in 2020. All coffee beans were subjected to the "wet processing method" because it is most commonly used for commercial coffees around the world. The samples were stored in a -18°C freezer until required.

With regard to the sample coding information (Table S1), codes were applied to the samples in order to provide a shorthand for the harvest, batch number, production stage and treatment but otherwise in no particular order. The sample codes were used for easy identification purposes by the author and for easy referral in the discussions.

3.1.3 Sampling procedure

Coffee samples were taken from two harvest years (2019 and 2020). Three batches

Stellenbosch University https://scholar.sun.ac.za

from each year were acquired, therefore six batches in total. Each batch consisted of two production stages, green and roasted beans. Furthermore, each production stage of beans was divided into a control and germinated (treated) beans. Figure 3.1.1 graphically illustrates the above-mentioned sampling procedure.

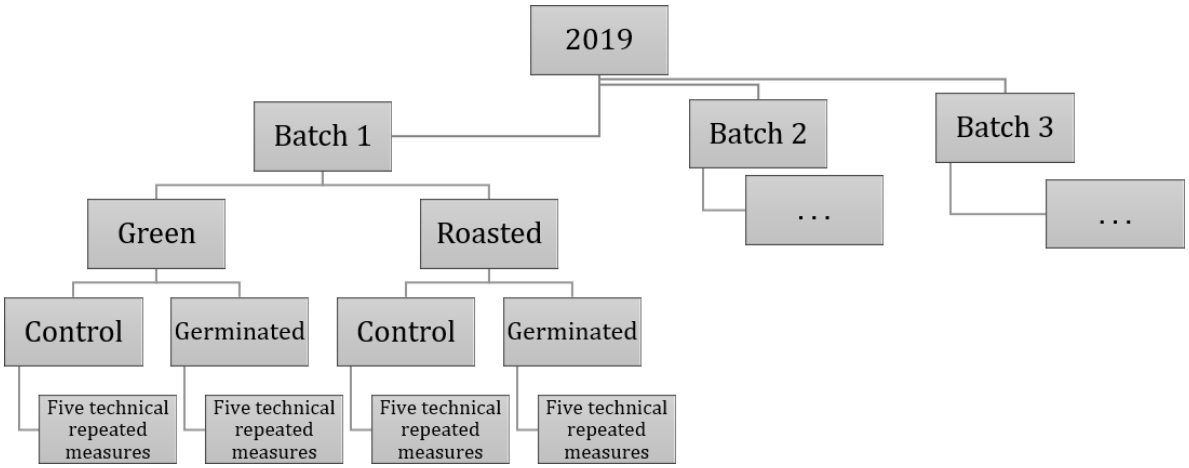


Figure 3.1.1. Illustration of the sampling procedure. The 2nd and 3rd batches followed the outline set out by batch one. The 2020 harvest followed the outline of the 2019 harvest for the 4th, 5th, and 6th batches.

3.2 Methodology

3.2.1 Sample preparation

Liquid nitrogen was used to flash-freeze the coffee beans prior to grinding in a MM400 Mixer Mill (Retsch, Germany). The frequency was set at 29 Hz and the duration to 5 min. On occasion, if more grinding was required then 2 min intervals at the same frequency was used until a fine powder was achieved.

3.2.2 Extraction

Approximately 1 g of ground sample was weighed out into 50 mL cylindrical centrifuge tubes and 15 mL of the solvent was added. The containers were placed in a sonication bath for 30 min without heat and then allowed to stand for a further 1.5 h outside of the sonication bath. After the 2 h extraction period, each container was briefly vortexed. Thereafter, 2 mL of the liquid was transferred to a 2 mL Eppendorf tube and centrifuged at 14 000 rpm for 5 min (Hermle Z 160 M, Lasec). Next, 1 mL of supernatant was transferred to a 2 mL glass vial, sealed with a screw cap and stored at 4°C. This method was repeated five (5) times per sample as technical repeated measurements.

Stellenbosch University https://scholar.sun.ac.za

3.2.3 Liquid Chromatography-Mass Spectrometry analysis

The experimental procedure used here was based on a study by Stander *et al.* (2017) which analyzed phenolic compounds in rooibos tea samples. High-resolution analysis was achieved with a Waters Acquity ultra-performance liquid chromatograph (UPLC) connected to a Waters Synapt G2 Quadrupole time-of-flight (QTOF) mass spectrometer (MS) with an Acquity photo diode array (PDA) detector in front of the MS (Waters, Milford, MA, USA). The electrospray ionization (ESI) parameters, as described by Stander *et al.* (2017), were setup as follows: ESI in the positive and then negative mode with a cone voltage of 15 V, the desolvation temperature set to 275°C, and Nitrogen used as the desolvation gas at 650 L/h. However, modification to the resolution mode was made in order to scan from 130 *m/z* to 1 500 *m/z* to detect trigonelline which has a *m/z* of 138, which was not investigated in the rooibos tea samples.

MS^E is a more powerful and efficient mode of acquiring data. It is capable of recording the exact-mass information for precursor and fragment ions while simultaneously providing the quantitative data for each component detected (Waters Corporation, 2011). For this study, the MS^E mode consisted of two channels. The first was a low collision energy (4 V) which produced the precursor ions and no fragmentation information. The second channel was a collision energy ramp (40 V to 100 V) that produced maximum information from the fragmented ions (Waters, 2011). Separation was achieved on an Acquity UPLC HSS T3, 2.1 × 150 mm, 1.8 µm column. The column parameters were as follows: 2 µL injection volume; 0.1% formic acid in water (solvent A) and acetonitrile containing 0.1% formic acid (solvent B) as the mobile phase; 0.3 mL/min flow rate; and the column temperature was maintained at 55°C. The gradient started at 100% solvent A for 1 min and changed to 28% solvent B over 22 min in a linear manner. It then went to 40% solvent B over 50 s and a wash stepof 1.5 min at 100% solvent B, followed by re-equilibration to initial conditions for 4 min. The total run time was 29.5 min.

3.3 Targeted approach

3.3.1 Chromatogram processing

MassLynx software (V4.1, 2010, Waters, Milford, USA), specifically the application manager "TargetLynx", was used to correct for the shifts in peak retention time. The method was setup to quantify the four key compounds (caffeine, chlorogenic acid,

Stellenbosch University https://scholar.sun.ac.za

trigonelline, and nicotinic acid). It is important to note, that the nature of the TargetLynx program requires that all major peaks be identified by their Exact Mass/Retention Time pairs (EMRTs) for the method to work optimally. Signal intensities were automatically integrated to generate their concentrations for each peak.

Caffeine, trigonelline, and nicotinic acid are best detected when scanned in the positive mode, whereas chlorogenic acid and most other phenolic compounds in coffee are detected in the negative mode. The monoisotopic masses for caffeine, chlorogenic acid (5-caffeoylquinic acid), trigonelline, and nicotinic acid are expected to be 195 *m/z*, 353 *m/z*, 138 *m/z*, and 124 *m/z*, respectively (Angelino *et al.*, 2018). Information pertaining to the ultraviolet (UV) absorbance of a compound was obtained via a photodiode array (PDA) detector. The UV wavelengths for the detection of caffeine, chlorogenic acid (5-caffeoylquinic acid), trigonelline, and nicotinic acid are previously reported as 273 nm, 325 nm, 264 nm, and 261 nm, respectively (Rodrigues & Bragagnolo, 2013). The assignment of a peak to one of the four compounds under investigation was based on three parameters: 1) the retention time 2) the characteristics of the UV information and 3) MS spectra compared to the reference standards. The quantification of each compound was determined using external standards to generate a six-point calibration curve. All quantitative results were expressed as mg/g dry weight basis for caffeine, chlorogenic acid, and trigonelline. Nicotinic acid was expressed as ug/g dry weight basis.

3.3.2 Data analysis

Descriptive statistics of the dataset were computed which included the mean, standard deviation, and the standard error. In addition, a test for outliers was performed. Further analysis by means of analysis of variance (ANOVA) provided information regarding the F-ratio, probability (*p*-value), and a post-hoc correlation table.

The average of a group is defined as the sum of all the values for each sample divided by the total number of samples and represents the center of a numerical data set. This is important because it can summarize large data sets into one value. The standard deviation represents the variability of an individual data point to the mean of the sample group, while the standard error is the amount of variability of the sample mean from the population mean. A standard deviation value that is close to zero indicates the data points are not spread out but instead localized around the mean.

Stellenbosch University https://scholar.sun.ac.za

Outliers are data points that do not conform to the general trend of the data; in other words their standard deviation is a certain distance away from the mean as set by a threshold. The Interquartile Range Method (IQR) defines limits that are a certain factor of the IQR below the 25th percentile and above the 75th percentile. The factor value that is used varies between literatures, but common practice is to use 1.5 for "general" outliers or a factor of 3 for the "extreme" outliers. The IQR was used to detect outliers within the data set and visually assessed by means of boxplots. All descriptive statistics were performed in Microsoft Excel (Version 16.0, 2016).

The data was also imported into R statistical software (R Development Core Team, 2015) and the lmer Package was used for analysis (Kuznetsova *et al.*, 2017). A mixed model analysis of variance (ANOVA) was used to investigate main effects and the interaction effects for each compound of interest, to determine significant differences (α=0.05) in the sample set (24 samples x two production stages x two treatments). Kenward-Rogers approximation was used for determining the degrees of freedom, along with a confidence interval of 95%. Kenward-Rogers approximation provides more accurate *p*-values, however it is severely affected by computational time for observations greater than 2 500 (Kuznetsova *et al.*, 2017). The small data set (n = 24) of the current research project is satisfactory.

3.4 Untargeted approach

3.4.1 Chromatogram processing

The "MarkerLynx" application manager in the MassLynx software (V4.1, 2010, Waters, Milford, USA) was used to correct for shifts in retention time of peaks. MarkerLynx transformed the data to retention time/mass pairs with each peak allocated to its corresponding signal intensity. The data that was produced is in the form of a complex matrix where the rows represent the different samples, and the columns represent the dependent variables.

3.4.2 Data analysis

Data matrices require multivariate techniques to be analyzed. Multivariate data analysis often involves the reduction of large data sets into a few key components (substantially less than the initial number of variables) that explains the largest portion of variation observed between variables while still preserving as much information as possible. Principal Component Analysis (PCA) is an example of multivariate data analysis used for

Stellenbosch University https://scholar.sun.ac.za

exploratory analysis. The PCA scores plot can illustrate which classes or groups are similar, while the loadings plot indicates why they are different (Steiman, 2003). However, the loadings plot cannot explain which of the classes or groups are of interest. Discriminate analysis (DA) methods can also be used to examine the differences between groups of a sample, such Partial Least Squares (PLS-DA) and Orthogonal Partial Least Squares (OPLS-DA) (Steiman, 2003; Worley & Powers, 2016). These two methods both produce models that can predict group membership. However, OPLS-DA is more ideal compared to PLS-DA because of its ability to adequately explain the data in the simplest model otherwise known as parsimony. Whereas PLS-DA will form complex models when uncorrelated variance in an experimental group is observed (Worley & Powers, 2016). In should be noted that PLS and OPLS analyses find separations based upon class membership assigned to the samples which could lead to unreliable interpretation of results. Therefore, the supervised models need to be cross-validated.

The MarkerLynx software was used to pre-process the data. Normalization, pareto-scaling, and peak alignment were the pre-processing techniques chosen. Principal component analysis was performed on all the variables detected in the coffee samples, i.e. the green-control, green-germinated, roasted-control, and roasted-germinated datasets. Afterwards, OPLS-DA was computed to investigate possible differences between the same category groups. Tentative assignment of the important peaks identified in the OPLS-DA was based on parameters similar to the targeted approach. In other words, the retention time (where possible), the characteristics of the UV information (wavelength), and MS spectra (parent and daughter ions) can be compared with published literature and open-source databases to tentatively predict the identity of the compounds.

Stellenbosch University https://scholar.sun.ac.za

References

Angelino, D., Tassotti, M., Brighenti, F., Del Rio, D. & Mena, P. (2018). Niacin, alkaloids and (poly)phenolic compounds in the most widespread Italian capsule-brewed coffees. *Scientific Reports*, **8**, 1-11.

Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, **82**(13), 1-26.

R Development Core Team. (2015). A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL. http://www.R-project.org/.

Rodrigues, N. P. & Bragagnolo, N. (2013). Identification and quantification of bioactive compounds in coffee brews by HPLC-DAD-MS^n. *Journal of Food Composition and Analysis*, **32**, 105-115.

Stander, M. A., Van Wyk, B.-E., Taylor, M. J. C. & Long, H. S. (2017). Analysis of phenolic compounds in Rooibos tea (*Aspalathus linearis)* with a comparison of flavonoid-based compounds in natural populations of plants from different regions. *Journal of Agricultural and Food Chemistry*, **65**, 10270-10281.

Steiman, S. R. (2003). *Method development for green coffee analysis and its possible application for group discrimination and correlation of green coffee chemistry with cupping quality.* Master of Science in Horticulture Thesis, University of Hawai'i, United States of America.

Waters. (2011). *An overview of the principles of MS^E: The engine that drives MS performance.* MA, USA: Waters Corporation.

Worley, B. & Powers, R. (2016). PCA as a practical indicator of OPLS-DA model reliability. *Current Metabolomics*, **4**(2), 97-103

Stellenbosch University https://scholar.sun.ac.za

4 Results & Discussion

4.1 Method Selection

A test-run of the reference standards was performed to confirm which electrospray ionization (ESI) mode would be most appropriate for detecting the desired compounds – with consideration for both the targeted and the untargeted approach. Initially, all three modes of detection, ESI positive, ESI negative, and photon diode array (PDA), were investigated. In Figure 4.1.1C, caffeine, trigonelline, and nicotinic acid could be clearly identified in the ESI+ mode. Whereas the chlorogenic acid peak was identified in the ESI– mode (Figure 4.1.1B). All four compounds were also identified by the PDA detector (Figure 4.1.1A). However, studies have reported that other coffee phenolic compounds are mostly identified in the ESI– mode and ESI+ provides little additional information (Angelino *et al*., 2018). Therefore, only the combined ESI– and PDA methods were used for detecting the four compounds and other phenolic compounds in the coffee in order to reduce analysis time because the ESI modes each took 25 minutes. Quantification of caffeine, trigonelline, and nicotinic acid was achieved by UV-peak integration. Although the chlorogenic acid peak was also observed with the PDA, better peak detection was seen in the ESI– mode.

The caffeine standard had a retention time of 12.12 min with an absorbance maximum at 273 nm using the PDA detector. The trigonelline standard had a retention time of 1.64 min and a maximum absorbance at 230 nm. The nicotinic acid standard had a retention time of 2.72 min and an absorbance maximum at 261 nm. The chlorogenic acid standard had a [M-H]-[a] of 353 *m/z* and a fragment ion of 191. The fragmentation ion of 191 *m/z* is key for identifying 5-caffeoylquinic acid, commonly called chlorogenic acid, specifically because 3-caffeoylquinic acid and 4-caffeoylquinic acid also have [M-H]- of 353 *m/z* but the fragmentation patterns are unique for each. The fragmentation pattern MS spectra of chlorogenic acid is seen in Figure 4.1.2. The 707 *m/z* ion fragment is also important as this indicates two chlorogenic acid molecules bound together and protonated. Simply,

$$353\ m/z + 353\ m/z = 706 + H^{+} = 707\ m/z$$

[a] Molecular ions observed in negative mode of mass spectra are usually represented as M- or [M-H]- when depronotated, and M+ or [M+H]+ when protonated.

Stellenbosch University https://scholar.sun.ac.za

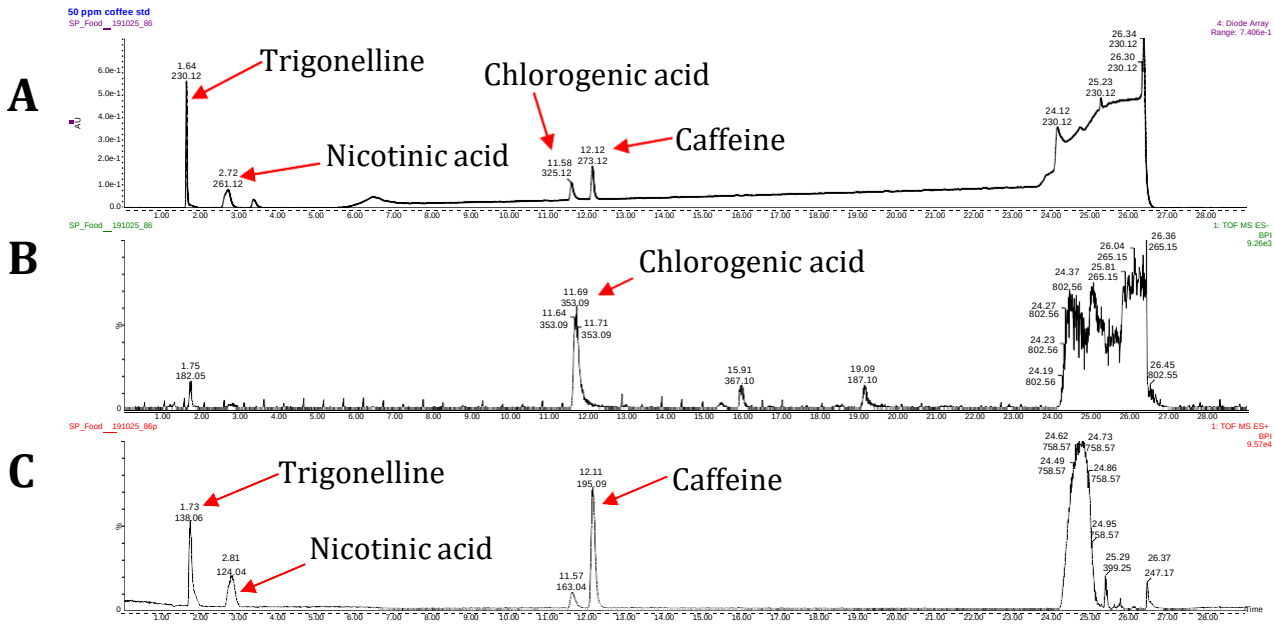


Figure 4.1.1. Separation of caffeine, chlorogenic acid, trigonelline, and nicotinic acid in a 50 ppm reference standard solution of 50% methanol with 1% formic acid. (A) UV chromatogram from a PDA detector in front of the MS, (B) base peak intensity chromatogram in ESI negative mode, and (C) base peak intensity chromatogram in ESI positive mode.Scales are not comparable.

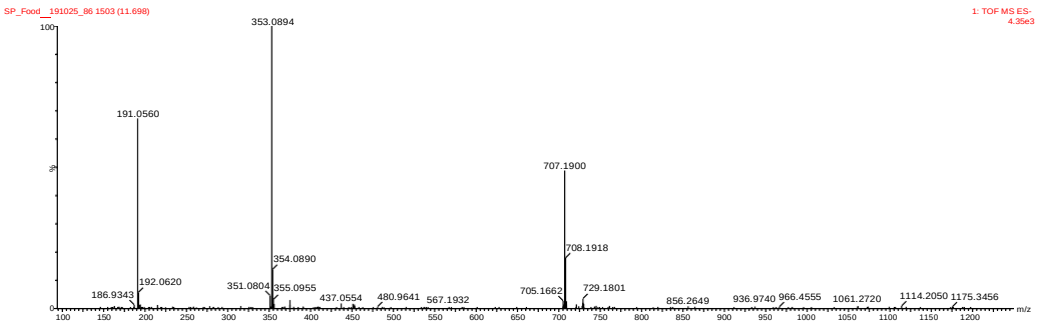


Figure 4.1.2. ESI negative ion mode MS/MS spectrum of the fragmentation pattern for 5-caffeoylquinic acid.

4.2 Targeted approach

4.2.1 Peak identification

A typical MS and UV chromatogram is seen for the 2019 harvest season in Figure 4.2.1 and Figure 4.2.2, respectively. The MS and UV chromatograms for the 2019 germinated coffee (Figure S1, Figure S2) and the 2020 control coffee (Figure S3, Figure S4) and germinated coffee (Figure S5, Figure S6) can be seen in the supplementary notes. Peaks

Stellenbosch University https://scholar.sun.ac.za

were identified by comparing the *m/z*, retention time, and fragmentation data to the reference standards.

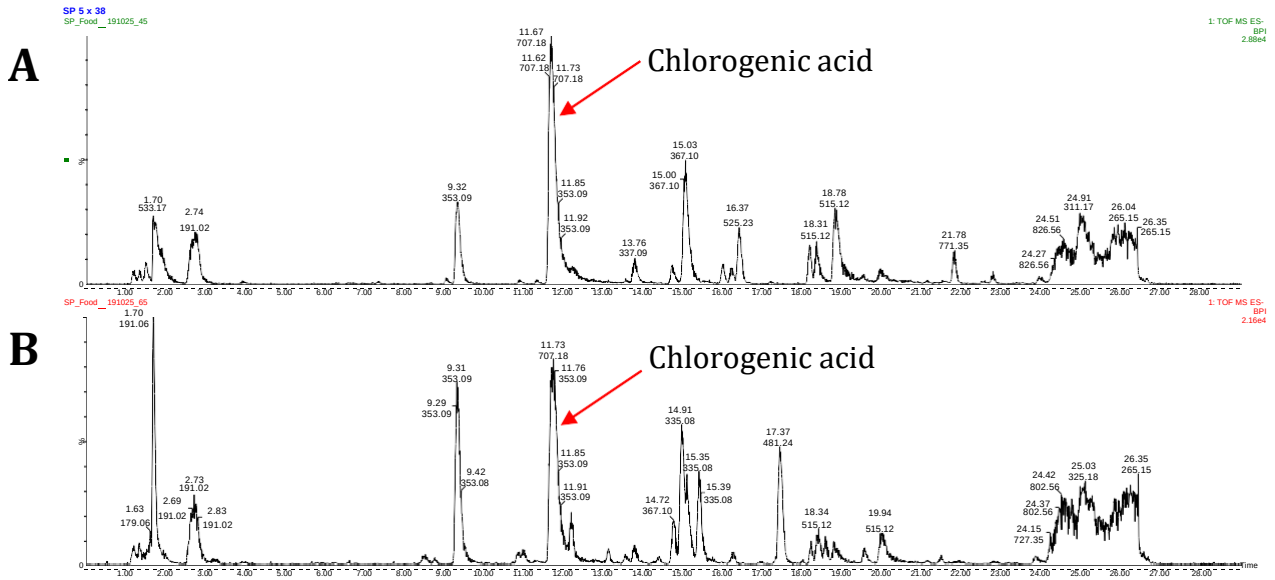


Figure 4.2.1. Typical base peak intensity chromatogram for the 2019 coffee harvest season showing (A) green coffee control and (B) roasted coffee control in the ESI negative mode to illustrate chlorogenic acid peak identification. Scales are not comparable.

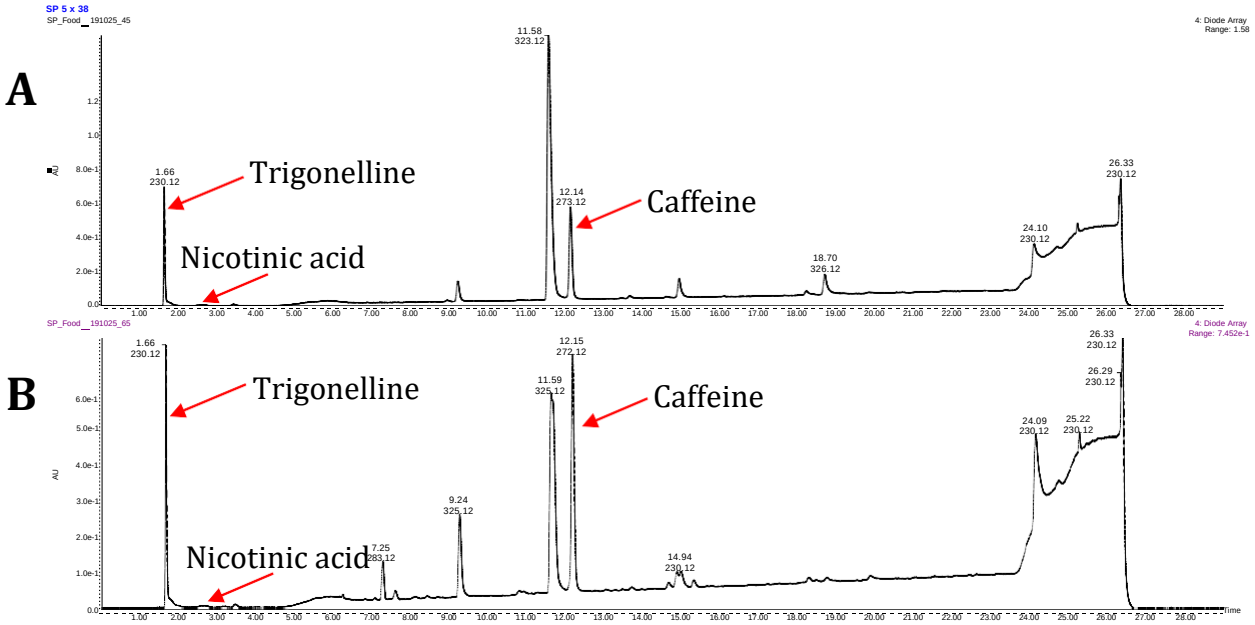


Figure 4.2.2. Typical UV chromatogram for the 2019 coffee harvest season showing (A) green coffee control and the (B) roasted coffee control from the PDA detector to illustrate peak identification of trigonelline (230 nm), nicotinic acid (261 nm), and caffeine (272 nm). Scales are not comparable.

Stellenbosch University https://scholar.sun.ac.za

Similar peaks were observed in the chromatogram for the 2020 harvest season, apart from differences in peak retention times. Most of the peak shifts occurred during the second gradient change from 100% solvent A to 28% solvent B. This means the early eluting compounds, trigonelline and nicotinic acid, will have consistent retention times but the later compounds, caffeine and chlorogenic acid, will have different retention times. The peak shifts were likely due to instrumental drift which occurs slowly over time. The same instrument was used for analysis, but the two analyses were performed one year apart which meant the instrumental drifts were larger. Therefore, the older the column becomes and/or the more it is used will have an influence on the ability to separate the compounds. This was unavoidable as the chemical composition of the coffee samples deteriorate over time and therefore required immediate analysis after harvesting and processing. After observing the peak shifts, the standards were re-analyzed in order to confirm the "new" retention times – the standards "new" retention agreed with that of the shifted compounds. The peak shifts did not affect quantification of the targeted compounds because the TargetLynx application (V4.1, 2010, Waters, Milford, USA) required the retention times of each compound to be manually added into the software and therefore was easily adjusted.

4.2.2 Calibration parameters

The integration of peak intensities requires calibration curves of the different reference standards. Table 4.2.1 provides the parameters used for each calibration curve as well as the Limit of Detection (LOD) and Limit of Quantification (LOQ). All the calibration curves were linear and forced through the origin. The chlorogenic acid curve only had five concentration points because one point was detected as an outlier and therefore removed. This explains the marginally lower coefficient of determination (R^2) observed for the chlorogenic acid curve ($R^2 = 0.9367$) compared to caffeine ($R^2 = 0.9990$), trigonelline ($R^2 = 0.9975$), and nicotinic acid ($R^2 = 0.9998$). The LOD was determined as three times the ratio of the standard deviation for the y-responses to the slope of the calibration curve. Whereas the LOQ was determined as ten times this ratio. The LOD values for caffeine (1.09 ug/L) and nicotinic acid (1.04 ug/L) were comparable with findings by Perrone *et al.* (2008) but the LOD for trigonelline (9.19 ug/L) was nearly two times higher than the amount reported by Perrone *et al.* (2008). The chlorogenic acid LOD (12.68 ug/L) was found to be almost half the amount reported by Angelino *et al.* (2018).

Stellenbosch University https://scholar.sun.ac.za

Table 4.2.1. Calibration parameters and limits of detection (LOD) and quantification (LOQ) for the concentration curves prepared in 50% methanol with 0.1% formic acid.

Compound	n[a]	R^{2b}	LOD (ug/L)	LOQ (ug/L)
Caffeine	6	0.9990	1.09	3.30
CGA[c]	5	0.9367	12.68	38.41
Trigonelline	6	0.9975	9.19	27.86
Nicotinic acid	6	0.9998	1.04	3.16

[a] The number of points used in the calibration curve

[b] The coefficient of determination

[c] CGA: chlorogenic acid

4.2.3 Outlier Detection

It is common practice to determine outliers before performing any type of analysis on a data set. This is done to determine if any observations deviate from the general trend of the sample group and to investigate if human error, instrumental error or natural variation might have occurred (Dawson, 2011; Hodge & Austin, 2004). The interquartile range method (IQR) was used to calculate outliers which can be easily visualized with boxplots (Figure 4.2.3). The upper and lower limits were defined as the 1.5 times the range between the 1st quartile and the 3rd quartile, respectively. These limits are displayed as the "whisker" portion of the boxplot. For this study, observations that are outside these limits were represented by circles on the plot and required further investigation.

No outliers were identified for the chlorogenic acid (CGA), trigonelline, and nicotinic acid data. However, five data points were identified as outliers in the caffeine concentrations (Figure 4.2.3A) indicated as spheres on the plots. Upon further examination, the five data points all belonged to samples collected during the 2020 harvest season. All of the outliers were classed as "green" coffee beans; however three samples were from control groups and the remaining two were from the germinated groups. The values of the five outliers ranged from 33.70 mg/g to 38.70 mg/g, which is approximately two times the average content for green coffee samples (17.21 mg/g). Initial assumptions about outliers might lead one to think that they should be removed.

Stellenbosch University https://scholar.sun.ac.za

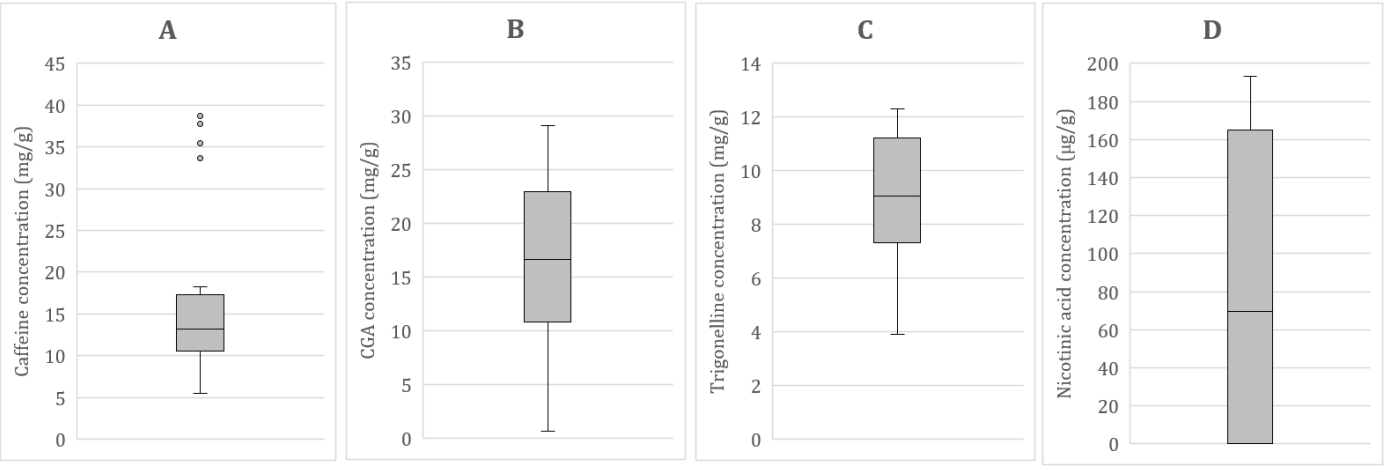


Figure 4.2.3. Boxplot representations for the concentrations of (A) caffeine, (B) chlorogenic acid (CGA), (C) trigonelline, and (D) nicotinic acid in Coffea arabica beans (n=24). Lower and upper "whiskers" were defined as 1.5 times the interquartile range.

Instead, the data points suggest that sudden spikes in concentration levels are natural occurrences and these samples should therefore remain as representatives in the dataset. Also, it is possible that the germination treatment was not properly controlled for those few samples which led to the larger caffeine levels. Germination as a post-harvest method has not been reported on and therefore this remains an unsupported deduction. Furthermore, the outliers were identified from the 2020 harvest season and might in fact represent differences between harvest years. A study by Zhou *et al.* (2019) reported that polyphenol content of Merlot and Pinot Noir grapes were significantly affected by different harvest times. Although grapes and wines were the focus of the aforementioned research, the authors also reported phenolic changes in the seeds of the grapes between different harvest times (Zhou *et al.*, 2019). The seeds of the grapes could be likened to the seeds (or beans) of the coffee cherry and the reasoning could be tentatively applied to the coffee samples of the current research paper.

Additionally, Dawson (2011) stated that when a sample size is small then the IQR can be unrepresentative of the population and is more likely to detect outliers because the quartiles will include high variation. Theoretically, there is a 0.8% probability of encountering an outlier in a data set (Dawson, 2011). The coffee sampling size of the present objective totaled n = 24, which would mean that the five detected outliers represented 20% of the data. Unfortunately, the small sample size does not allow for proper deductions to be made about the larger coffee population and the removal of the

Stellenbosch University https://scholar.sun.ac.za

outliers could drastically affect the results. For these reasons, it was decided to retain the outlier values in the data set as they potentially form a valid part of the population and reflect important variation between the coffee samples.

4.2.4 Analysis of Variance

The nature of the experimental design (two production stages x two treatments) requires the analysis of the main effects of the independent variables (treatment and production stage) as well the observed interaction effects between the two independent variables. This is done to determine which (if any) were significantly different. Analysis is performed on the mean ($\bar{x}$) concentrations and reported in terms of mean ± the standard deviation on a dry weight basis (DW). The units for caffeine, chlorogenic acid, and trigonelline are in mg/g DW and nicotinic acid is reported as µg/g DW.

4.2.4.1 Caffeine concentration

The average concentration for caffeine between the control and treated samples at each production stage can be seen in Figure 4.2.4 below. The average concentration for green-control, green-germinated, roasted-control, and roasted-germinated coffee beans were determined to be 24.90±13.65 mg/g DW, 19.79±13.06 mg/g DW, 12.22±2.3 mg/g DW, and 11.94±2.3 mg/g DW respectively. The descriptive statistics can be found in the supplementary notes (Table S2).

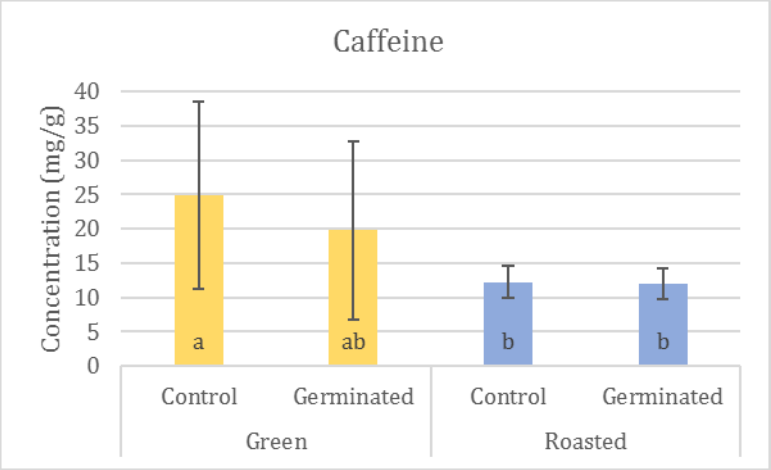


Figure 4.2.4. Two-way mixed model analysis of variance (ANOVA) for the main effects between sample group treatments (control vs. germinated) and production stage for the mean concentrations of caffeine. Different letters in the bar's bases indicate significant differences (p = 0.05). Vertical bars denote standard deviation.

Stellenbosch University https://scholar.sun.ac.za

A two-way mixed model ANOVA (2 production stage x 2 treatment) performed on the caffeine concentrations revealed that the main effect for production stage was significant $F(1,15) = 7.58$, $p = 0.01$. Thus, there was an overall difference in the concentrations for green coffee ($\bar{x}$ = 22.35 mg/g DW) compared to roasted coffee ($\bar{x}$ = 12.08 mg/g DW). The main effect for treatment was not significantly different, $F(1,15) = 0.52$, $p = 0.48$ and therefore the caffeine concentration for the control group ($\bar{x}$ = 18.56 mg/g DW) did not differ from the germinated coffees ($\bar{x}$ = 15.87 mg/g DW). When analyzing the interaction effect between production stage and treatment, no significant difference was observed, $F(1,15) = 0.42$, $p = 0.53$. Closer examination of the means indicated that there was a decrease in caffeine concentration for the green coffee from the control stage ($\bar{x}$ = 24.9 mg/g DW) to the germinated group ($\bar{x}$ = 19.79 mg/g DW), and the same trend was seen for the roasted coffee from the control group ($\bar{x}$ = 12.22 mg/g DW) to the germinated group ($\bar{x}$ = 11.94 mg/g DW). As mentioned earlier, these differences are not significant. Therefore, based on the results of this study it can be said that the change in caffeine concentration for a coffee sample is dependent on the production stage, i.e. being green or roasted, and not due to germination.

Caffeine is widely reported as not being affected by roasting (Oestreich- Janzen, 2010; Sunarharum 2016). The results of this study are not in agreement with the prior statement, and furthermore the concentration range for caffeine was not in the same range as previous reports (Moon *et al.*, 2009; Sunarharum 2016). The present findings saw a decrease in caffeine levels after roasting which is contrary to the reports by Moon *et al.* (2009) who saw an increase after roasting for Colombian, Ethiopian, Guatemalan, Mexican, Nicaraguan, and Papuan coffees. To the authors best knowledge, this is the first study reporting on caffeine concentrations in South African produced coffee. It is proposed that the South African climate provides different conditions from that of typical coffee regions and could therefore have played a role.

It was seen that caffeine content decreased significantly after roasting. The final, roasted concentrations were similar between control and germinated treatments. For this reason, it is postulated that regardless of initial compound content, the roasting process will degrade caffeine until its end-point is reached. Different degrees of the roasting (e.g. light, medium, dark, French) are applied throughout literature because this is not a standardized procedure but rather based on the requirements of the farmer, roaster, or

Stellenbosch University https://scholar.sun.ac.za

consumer (Belay *et al.*, 2016). Therefore, perhaps the roasting temperature-time combination is an important factor when controlling concentration levels. This is outside the scope of the current study but is still important to consider.

4.2.4.2 Chlorogenic acid quantification

The chlorogenic acid isomer that is most abundant in coffee is 5-caffeoylquinic acid and has been referred to as "chlorogenic acid" throughout literature. Likewise, this study refers to 5-caffeoylquinic acid interchangeably with "chlorogenic acid" (CGA). The mean concentration of chlorogenic acid for green-control samples was 22.92±6.6 mg/g DW, for green-germinated coffee it was 17.13±10.59 mg/g DW, for roasted-control coffee it was 14.59±6.23 mg/g DW, and for roasted-germinated samples it was 13.94±5.8 mg/g DW (Figure 4.2.5). Similar to the trend seen in caffeine concentrations, chlorogenic acid concentrations decreased in both green and roasted coffees when the germination process was applied at the post-harvest treatment stage.

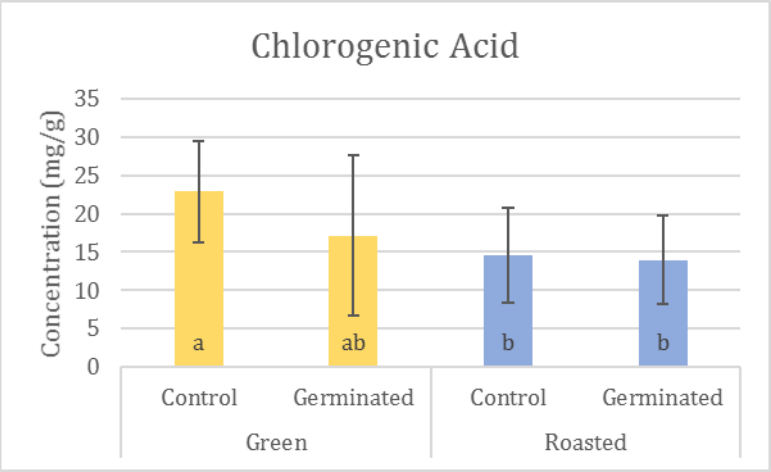


Figure 4.2.5. Two-way mixed model analysis of variance (ANOVA) for the main effects between sample group treatments (control vs. germinated) and production stage for the mean concentrations of Chlorogenic Acid. Different letters in the bar's bases indicate significant differences (p = 0.05). Vertical bars denote standard deviation.

The main effects for production stage and treatments were both not significant, $F(1,15) = 4.25$, $p = 0.06$ and $F(1,15) = 1.34$, $p = 0.27$, respectively. Thus, there was no overall difference in the concentrations for green coffee ($\bar{x}$ = 20.02±8.9 mg/g DW) compared to roasted coffee ($\bar{x}$ = 14.27±5.7 mg/g DW), as well as for the control treatment ($\bar{x}$ = 18.76±7.5 mg/g DW) compared to the germinated treatment ($\bar{x}$ = 15.53±8.3 mg/g

Stellenbosch University https://scholar.sun.ac.za

DW). It should be noted that the production stage probability value ($p = 0.06$) was very small and almost within the boundaries of the 95% confidence interval. Since green and roasted coffees are known as being visually and chemically different, this value for all intents and purposes indicates significant differences between the two groups. Should a larger sample size be used in any future work, then one could expect this to be significant.

Analyzing the interaction effect between production stage and treatment for CGA, no significant difference was observed, $F(1,15) = 0.85$, $p = 0.37$. This result implies that the production stage has no effect on the outcome of the treatment, and vice versa. However, the probability value was marginally low ($p = 0.37$) and so perhaps if a larger sampling size was obtained then, one might have seen an interaction effect. In other words, the CGA concentration for a given treated coffee sample *might* be subject to variation based on the production stage. A bigger decrease in CGA concentration is seen for the green coffee from control($\bar{x}$ = 22.92 mg/g DW) to germinated ($\bar{x}$ = 17.13 mg/g DW) compared to the decrease in roasted samples from control ($\bar{x}$ = 14.59 mg/g DW) to germinated ($\bar{x}$ = 13.94 mg/g DW) beans.

Chlorogenic acid was expected to drop in concentration after roasting based on previous reports that indicated that CGA is degraded to produce quinic acid along with the corresponding lactone-products (Bennat *et al.*, 1994). As assumed, CGA decreased during roasting for both treatment groups. However, both degraded to approximately the same end-point even though they had a 5-mg difference in their "green" concentrations. The results agree with the postulation suggested for caffeine that regardless of initial compound content, the roasting process will degrade CGA until the reaction reaches its end-point. To confirm this would require studying the reaction kinetics of the roasting process which was outside the scope of the research.

4.2.4.3 Trigonelline and Nicotinic acid quantification

Trigonelline and nicotinic acid are compounds involved in the same pyridine nucleotide cycle (Ashihara, 2008). Trigonelline is reported as being present in higher concentrations in green coffee beans and levels are reduced during the roasting stage. During roasting trigonelline is broken down to produce nicotinic acid as well as other volatile products. For this reason, they will be discussed in tandem with one another.

The mean concentrations for trigonelline were found to be present in smaller concentration ranges compared to published works. The green-control and green-

Stellenbosch University https://scholar.sun.ac.za

germinated coffee samples had mean concentrations of 7.85±0.6 mg/g DW and 6.42±1.8 mg/g DW, respectively. The roasted-control and roasted-germinated coffees had mean concentrations of 10.75±0.6 mg/g DW and 11.15±1.1 mg/g DW, respectively. Figure 4.2.6 illustrates the same decreasing trend after the germination treatment applied for green coffee, however an increase after roasting was seen for both control and germinated coffees. This is an unexpected outcome because, as mentioned earlier, trigonelline is typically degraded and not produced when thermally processed.

In terms of nicotinic acid concentrations, the mean values for green-control, green-germinated, roasted-control, and roasted-germinated were determined as 98.19±71.9 µg/g DW, 102.54±89.4 µg/g DW, 98.02±70.0 µg/g DW, and 49.61±74.1 µg/g DW respectively (Figure 4.2.7). It is important to note the high standard deviation values. This is because of the large range of values obtained for nicotinic acid. Excluding the seven samples that measured 0.0 µg/g DW, the concentration range for nicotinic acid was 30.65 µg/g DW to 193.17 µg/g DW.

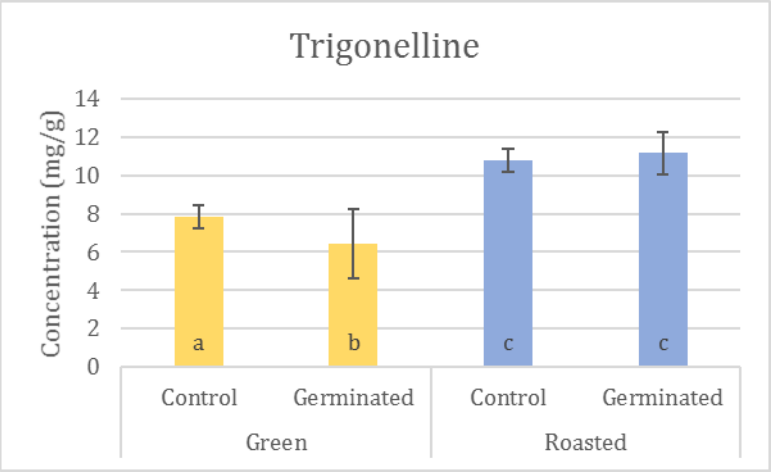


Figure 4.2.6. Two-way mixed model analysis of variance (ANOVA) for the main effects between sample group treatments (control vs. germinated) and production stage for the mean concentrations of Trigonelline. Different letters in the bar's bases indicate significant differences (p = 0.05). Vertical bars denote standard deviation.

Stellenbosch University https://scholar.sun.ac.za

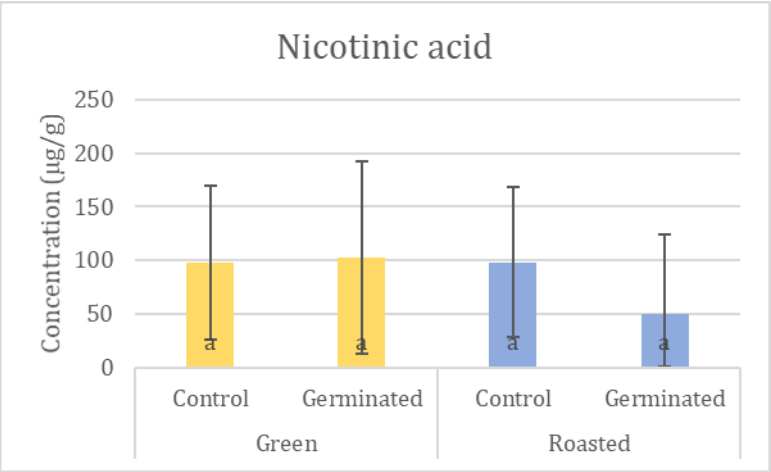


Figure 4.2.7. Two-way mixed model analysis of variance (ANOVA) for the main effects between sample group treatments (control vs. germinated) and production stage for the mean concentrations of Nicotinic Acid. Different letters in the bar's bases indicate significant differences (p = 0.05). Vertical bars denote standard deviation.

Trigonelline mean concentrations were also subjected to a mixed model ANOVA. The main effect for production stage was found to be significant, $F(1,15) = 69.62$, $p < 0.01$, but the main effect for treatments was not significant, $F(1,15) = 1.27$, $p = 0.28$. However, the p-value for treatment effects was quite low ($p = 0.28$) and warranted a further investigation into the post-hoc information (Table 4.2.2). This revealed significant differences between all pairwise groups except between roasted-control and roasted-germinated ($p = 0.55$). This is in line with the current study's findings of the above-mentioned compounds where no significant difference was seen between treatments after the roasting process.

Table 4.2.2. Post-hoc information after a two-way mixed model ANOVA for trigonelline in Arabica coffees. Values are representative of the p-value outcome.

Production stage	Treatment	Green-control	Green-germinated	Roasted-control	Roasted-germinated
Green	Control	1			
Green	Germinated	0.04	1		
Roasted	Control	< 0.01	< 0.01	1	
Roasted	Germinated	< 0.01	< 0.01	0.55	1

Stellenbosch University https://scholar.sun.ac.za

The interaction effect for trigonelline between production stage and treatment was $F(1,15) = 3.98$, $p = 0.06$ which indicates no significant difference was observed even though the *p*-value fell just outside the confidence interval. Although statistically insignificant, the very low p-value should not be ignored because in future work, with a larger sample size, there could be an interaction effect between treatment and production stage. This would therefore require further investigation into the main effects. As previously mentioned, the same postulation is made for trigonelline in that the roasting process for this study allowed for the same reaction end-point to be reached regardless of initial concentrations.

The nicotinic acid concentrations revealed that the main effect for production stage was not significant $F(1,15) = 0.9$, $p = 0.36$. Thus, for the first time, there was no overall difference in the concentrations for green coffee groups compared to roasted coffee groups. The main effect for treatments was not significant, $F(1,15) = 0.62$, $p = 0.44$. Upon examination of the interaction effect between production stage and treatment for nicotinic acid concentration, the resulting difference was not significant, $F(1,15) = 0.88$, $p = 0.36$.

As previously mentioned, trigonelline and nicotinic acid are linked in their production cycles, generally in an inverse relationship. It was assumed the trigonelline content would decrease after roasting, and a proportional amount of nicotinic acid would be produced (Oestreich-Janzen, 2010). However, the opposite appears to have occurred. For trigonelline, a significant increase for both control and germinated treatments after roasting was observed. Whereas nicotinic acid stayed unchanged except for a decrease in the germinated samples after roasting. Interestingly, the roasted-germinated decrease in nicotinic acid could perhaps be correlated to the increase in trigonelline because of the known relationship between the two compounds.

There is no current literature that accounts for the increase in trigonelline *after roasting* and this observation remains a puzzling phenomenon. According to Ashihara (2008), trigonelline is stored in the seeds and converted to nicotinic acid when NAD[b] synthesis is activated. This would explain the decrease observed in trigonelline content for the green-germinated samples of this study. It is further postulated that perhaps trigonelline was used for synthesis of a different compound in the pyridine nucleotide

[b] Nicotinamide Adenine Dinucleotide (NAD) is a coenzyme found in many living cells and functions as an electron acceptor during different metabolic pathways.

Stellenbosch University https://scholar.sun.ac.za

cycle, such as nicotinamide (Ashihara, 2008). This, however, requires investigation of compound biosynthesis and is beyond the scope of the current research. A more practical solution for this artefact could be that possibly the extraction procedure for trigonelline in the green coffee beans was not optimal. Similar papers also used aqueous methanol as the extraction solvent however various extraction times have been used. Extraction periods range from 20 min (Farah *et al.*, 2005) to 5 hours (Jaiswal *et al.*, 2012) to allowing the material to soak overnight (Stander *et al*., 2017).

Another point of discussion is the very low concentrations for nicotinic acid before and after roasting. According to literature, nicotinic acid is generally recorded in very low quantities or not at all in green samples, and for roasted coffee the values were detected on the lower end compared to other compounds (Casal *et al.*, 1998; Perrone *et al.*, 2008). The present study found that the average nicotinic acid concentrations were in fact in the range reported by Casal *et al.* (1998).

4.2.5 Influence on quality

The findings presented above revealed that germination did not have an effect on the concentrations of four key compounds in coffee. These compounds each have contributing roles in the aroma and flavour of coffee. Therefore, the main aroma and/or flavour profile of the control and germinated beans would not be distinguishable, and so the perceived quality attributes would also not change. Due to the lack of research on South African coffee species, it is possible that the key compound's concentrations can vary from that of commercial coffee species, typically Brazilian or Ethiopian. This could suggest that germination can influence the composition of other compounds not mentioned in this study and it is recommended that a full chemical profile of South African coffee beans be established.

4.2.6 Summary

The results of the two-way mixed model ANOVA revealed that the interaction effect of production stage and treatment were not significant ($\alpha = 0.05$) for all the compounds studied (caffeine ($p = 0.53$), chlorogenic acid ($p = 0.37$), trigonelline ($p = 0.06$), and nicotinic acid ($p = 0.36$)). The main effect of treatment did not reveal significant differences ($\alpha = 0.05$) for caffeine ($p = 0.48$), chlorogenic acid ($p = 0.27$), trigonelline ($p = 0.28$), and nicotinic acid ($p = 0.44$) for the coffees. These results indicate that germination as an intentional post-harvest treatment has no influence on the key compounds in coffee. The low *p*-values of chlorogenic acid and trigonelline suggest that perhaps some

Stellenbosch University https://scholar.sun.ac.za

significance could be observed. Additional sampling groups would be required to support this that could include different harvest years, various *Coffea* species and varieties, other coffees of different origins, different germinating conditions, or different roasting conditions (time, temperature). The influence of germination on the key compounds was not observed in the roasted coffees and therefore it is assumed the aroma and flavour profiles would remain the same between control and germinated coffee.

However, coffee matrices are vastly different between species depending on their geographical origin and are inherently very complex. For this reason, it is necessary to also perform an untargeted analysis in order to gain a broader understanding of the changes that are possibly occurring during germination. This might provide insight into other lesser-known compounds in coffee that are being affected and that potentially influence the quality.

Stellenbosch University https://scholar.sun.ac.za

4.3 Untargeted approach

4.3.1 Chromatogram transformation

Similar chromatograms were observed for green coffee from the 2019 and 2020 harvest seasons, except for differences in peak retention times (Figure 4.3.1). The peak shifts likely occurred for of two reasons: 1) the long period of time between the two harvest seasons can allow for instrumental drifts to occur which is the natural ageing of the HPLC column and thus effects its ability to properly separate compounds. The time between analyses is unavoidable as the chemical composition of the coffee samples degrade over time and therefore require immediate analysis, i.e. the 2019 samples could not be stored for a year. 2) Alternatively, different peak retention times could also mean that different types of compounds are present which could be related to seasonal variation. The fragmentation data acquired from the HPLC analysis revealed that the same compounds were present (square boxes in Figure 4.3.1) but at different retention times and therefore it was deduced that instrumental drifts caused the differences. The peak shifts were between 1 min and 5 min apart for the 2019 and 2020 samples. Althoughthere are several techniques to overcoming misaligned peaks or peak shifting (Korifi *et al.*, 2014), they require a deep understanding of the various algorithms used as well as the software required to perform them. Selecting the appropriate method is also time consuming. The MarkerLynx application used for this study was able to correct for minor shifts in peaks (± 0.5 min). "Minor shifts" is when the same compound in different samples from the two seasons has slightly different retention times. However, the software could not correct for the large peak shift variation seen between the 2019 and 2020 coffee seasons. Thus, the decision was made to analyze the two harvest seasons separately and compare the results and trends afterwards.

Principal Component Analysis is an unsupervised data analysis technique that attempts to determine trends between samples based on no prior knowledge of the groupings and to observe similarities or differences in the data. Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) is a supervised analysis and requires assignment of samples to a class prior to analysis. After computation, the differences that are observed are between the two (or more) groups assigned to the samples. For these reasons, it was decided to use both PCA and OPLS-DA to determine if differences between

Stellenbosch University https://scholar.sun.ac.za

samples existed, specifically for the control and germinated beans. If differences were observed, the compounds which were causing the differences were identified.

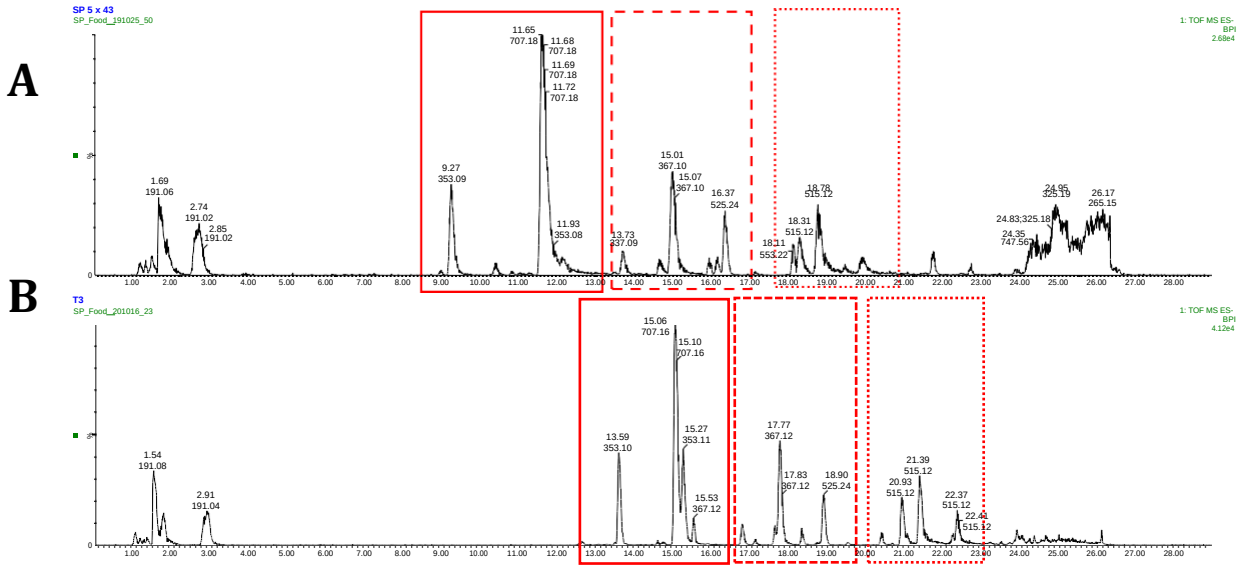


Figure 4.3.1. Typical chromatograms for germinated, green coffee from the (A) 2019 harvest season and (B) 2020 harvest season. Retention time shifts are observed on the horizontal axis, and similar peak patterns are identified by red boxes with different outlines. Horizonal axis was retention time and vertical axis was peak intensity.

4.3.2 Principal Component Analysis (PCA)

PCA can be used to identify trends in a data set which illustrate similarities or differences between groups or classes (Steiman, 2003). PCA was initially performed on the 2019 and 2020 harvest data sets, and the scores plots are seen in Figure 4.3.2. Clustering is primarily seen based on the production stage of the beans, i.e. green or roasted, for both harvest years. Obvious differences between green and roasted beans are confirmed via visual aspects of the two classes but also by their chemical composition. Therefore, it was expected that variation between the production stages would cause the most separation. The five green, germinated replicates on the right-side of Figure 4.3.2A which clustered together were identified as all belonging to sample “G” from the 2019 harvest. Similarly in the 2020 harvest, four green, control replicates were spread to the left-side away from the rest of the samples and were identified as belonging to sample “P” and “R” (Figure 4.3.2B).

Stellenbosch University https://scholar.sun.ac.za

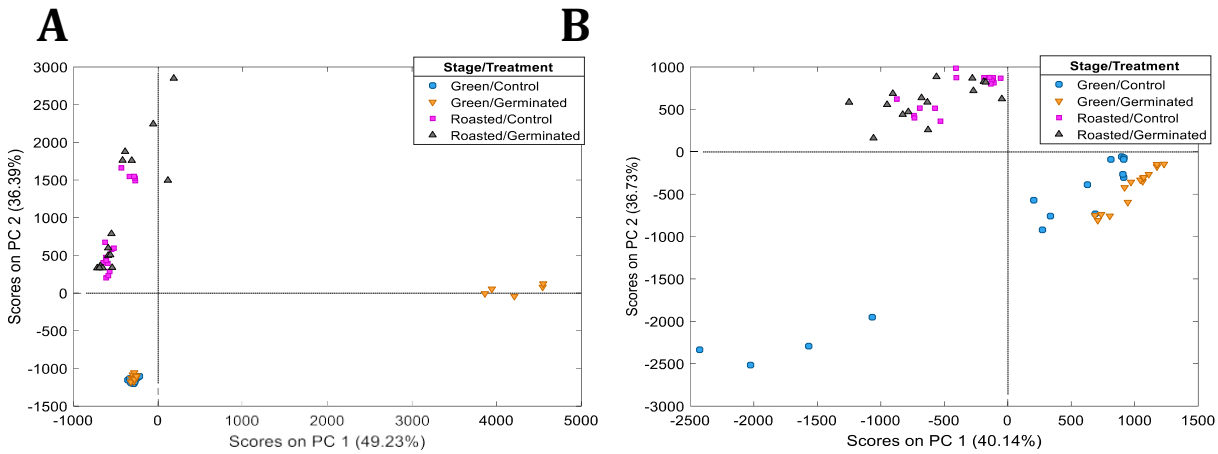


Figure 4.3.2. PC score plots for the (A) 2019 harvest season and (B) 2020 harvest season of coffee (*C. arabica*). The colouring is based on the production stage/treatment combination. There are five technical repeats for each sample.

Two deductions can be made regarding these irregular samples. The first, is that the chemical composition of the "G", "P", and "R" samples were different because they were processed in a different manner from the rest of the beans. The second reason could be that the manner in which the beans were ground and extracted for analysis was not performed optimally. It is likely that both deductions had an influence. Every effort was made to communicate consistent processing and germination throughout the production. Since the coffee production was performed externally by the coffee estate, consistency cannot be guaranteed. There is also evidence to suggest that the extraction method was also inconsistent for samples "G", "P", and "R". This is because the data points in the PC score plot represent replicates of a sample and therefore each replicate should have the same chemical composition and be clustered close together. Since the extraction process was identical for each sample, it was deduced that the grinding stage prior to extraction was not ideal. The grinding step only involved a visual assessment of the particle size and therefore proper homogeneity was not guaranteed. It is recommended that a particle sizing step be integrated after the grinding process to ensure uniform particle sizes, such as using different mesh-plates that allow particles of a certain size to pass through. The aim of this study was to determine if germination as a treatment causes differences in coffee's chemical composition, thus it was decided to observe trends between the treatments only and not the production stage (green or roasted beans).

Stellenbosch University https://scholar.sun.ac.za

Principal component analysis was computed separately for the green and roasted data sets for each harvest year. To reiterate, the two harvest years could not be combined because of retention time shifts observed in the chromatograms (Figure 4.3.1 above). The initial attempt of PCA for the 2019 green data set (Figure S7) illustrates the outlier sample "G" that separated differently. The reason for this group's difference was due to irregular chromatographic responses of four compounds which was approximately 1000 times larger than the other samples from that year and production stage. The identities of the compounds are unknown and fall outside the scope of the immediate research, however the retentiontime and *m/z* pair were as follows: 1) 16.37 min, *m/z* 525.23; 2) 16.38 min, *m/z* 481.24; 3) 18.14 min, *m/z* 553.22; 4) 19.48 min, *m/z* 181.09. These compounds account for 82.97% of the variation observed along Principal Component one (PC1) in theinitial PC scores plot.

The data points were classed as outliers, excluded and PCA was recalculated (Figure 4.3.3). PC1 explained 45.70% and PC2 explained 16.94% of the variation between the samples. No clear trends were observed – the control and germinated samples were interspersed. Figure 4.3.3A illustrates that the control and germinated groups were similar in chemical composition for the 2019 green coffee beans, and so germination did not appear to have an influence. Figure 4.3.3B further examines the scores plots based on sample identification. Each sample was analyzed by means of five technical repeats. Therefore, the poor groupings of the sample replicates (Figure 4.3.3B) support the idea that the sample grinding method was not optimal since the technical repeats did not cluster together. Pinelo *et al.* (2007) studied the effect of different ground coffee particle sizes during extraction of phenolic compounds and concluded that the smallest particle size (125 μm) allowed for maximum surface contact with the solvent and therefore enhanced the phenolic content. Thus, the lack of a proper particle size assessment could have caused an inconsistent extraction process and affected the concentration values.

Stellenbosch University https://scholar.sun.ac.za

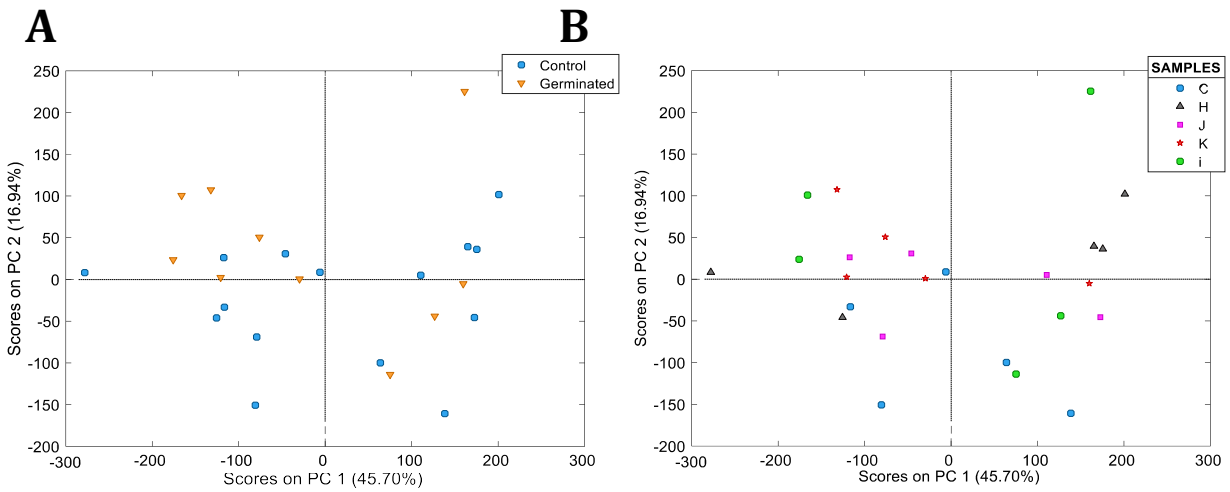


Figure 4.3.3. PC score plots for the 2019 green coffee (*C. arabica*) data. (A) Samples are coloured by the treatment applied. (B) Samples are coloured by the sample code. Each sample includes five technical repeats. The samples labelled C, H, and J are control coffee samples. The samples labelled K and i are germinated coffee samples.

In comparison, the score plots of the green beans from the 2020 harvest revealed slight groupings based on the treatment applied (Figure 4.3.4). The technical repeats belonging to samples "P" and "R" are control samples and separated different from the other control repeats and sample (sample "Q"). As previously discussed, the insufficient grinding method that was used could have impacted the extractability of the compounds and caused these repeats to cluster separately. On the other hand, these could also represent a larger variation within control coffee samples compared to that of germinated coffee samples. As can be observed, there was tight clustering of the germinated samples in the upper-right quadrant, whereas most of the control samples were spread out away from the germinated samples. The separation was observed when plotting PC1 (73.16%) versus PC4 (3.73%). This would suggest that chemical differences between control and germinated beans was achieved but is overshadowed by other variation within the samples. Further sampling and analysis would be required to produce more robust results.

Stellenbosch University https://scholar.sun.ac.za

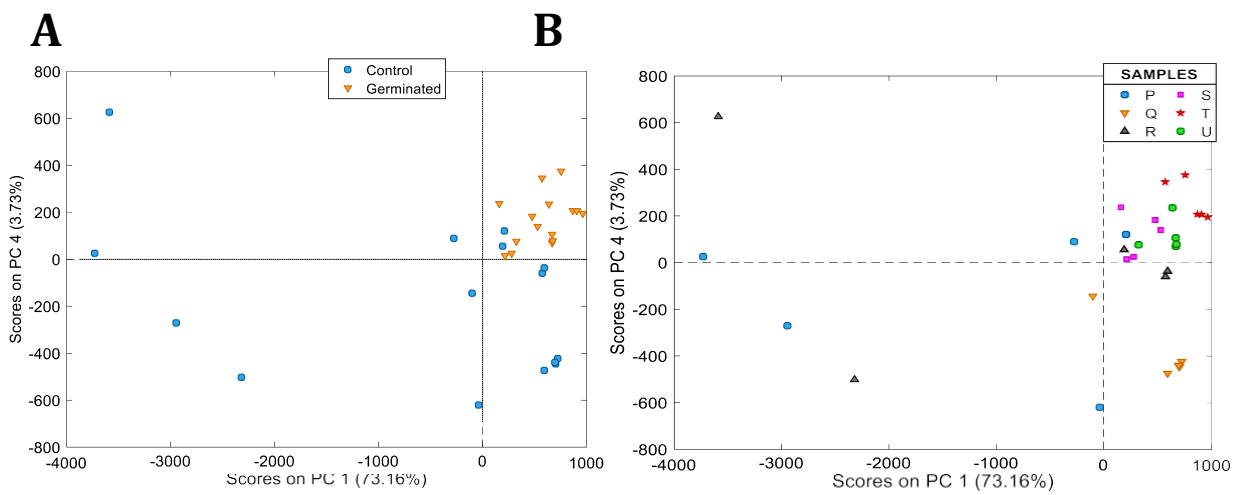


Figure 4.3.4. PC score plots for the 2020 green coffee (*C. arabica*) data. (A) Samples are coloured by the treatment applied. (B) Samples are coloured by the sample code. Each sample includes five technical repeats. The samples labelled P, Q, and R are control coffee samples. The samples labelled S, T, and U are germinated coffee samples.

The results of PCA for the green coffee beans provides conflicting information. The 2019 data suggests that germination did not have an impact on the chemical composition of the samples. Whereas the 2020 data implies that germinating the beans resulted in chemically similar profiles compared to the control samples which had more chemical variation. It was deduced that a poor grinding procedure was applied that did not adequately homogenize the samples which led to inconsistent compound extractions. Furthermore, roasting is a necessary step in coffee production and should be studied in tandem with the treatment of germination to determine if any chemical composition changes occurred that carry through the roasting process. It was expected that similar trends would be observed in the roasted beans.

PCA was also calculated for the roasted coffee beans data from the two harvest seasons. This was done to confirm the assertion that similar trends were expected as deduced from the PCA results of the green coffee samples. The PC score plot for the 2019 roasted coffee data revealed two cluster groups (Figure 4.3.5). Similar to the 2019 green data, the groups were not based on the treatment applied and is supported by the mixed clusters of control and germinated samples (blue and orange dots). Evaluation of the clusters based on sample code information, revealed that better groupings of the technical repeats was observed compared to the green data of the same year (Figure

Stellenbosch University https://scholar.sun.ac.za

4.3.5B). This suggests that partially consistent extractions were obtained for the replicate analyses. Perhaps, since roasted coffee beans are softened after the roasting process they are easier to grind down compared to green coffee beans which are extremely hard and brittle. This would support the idea that better clusters of the technical repeats was observed after roasting compared to the green data PC score plot. A further interesting observation was noted whereby samples "N" and "O" were grouped separately from the other clusters. Both samples "N" and "O" originated from a similar batch cycle and so this would suggest that the processing for those samples was not performed in the same manner as the other batches. Thus, their chemical composition may be different from the other batches.

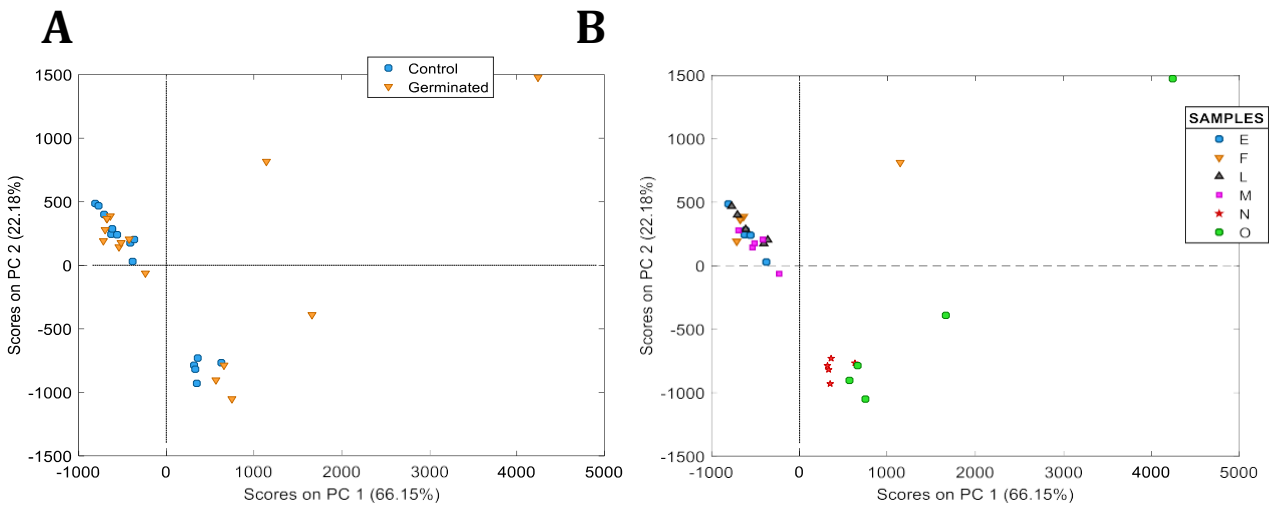


Figure 4.3.5. PC score plots for the 2019 roasted coffee (*C. arabica*) data. (A) Samples are coloured by the treatment applied. (B) Samples are coloured by the sample code. Each sample includes five technical repeats. The samples labelled E, L, and N are control coffee samples. The samples labelled F, M, and O are germinated coffee samples.

A similar trend was observed for the 2020 roasted data. Unlike the 2020 green data, two cluster groups formed that were not distinguishable based on the treatment applied (Figure 4.3.6A). Separation was mostly seen along PC1 which explained 64.17% of the variation, but also a bit of separation along PC2 which explained 13.06% variation. In Figure 4.3.6B, the groupings of the technical repeated measures for each sample was similar to the previous green coffee score plots (Figure 4.3.3B, Figure 4.3.4B) because no tight clusters were seen. This is unlike the 2019 roasted data (Figure 4.3.5B). It is strongly believed that the grinding process was not comprehensive enough which led to

Stellenbosch University https://scholar.sun.ac.za

inconsistent chemical profiles after extraction. As previously mentioned, a particle sizing assessment should be incorporated to ensure uniform particle sizes are obtained prior to extraction. This will allow for effective contact of the coffee particle surface area with the solvent.

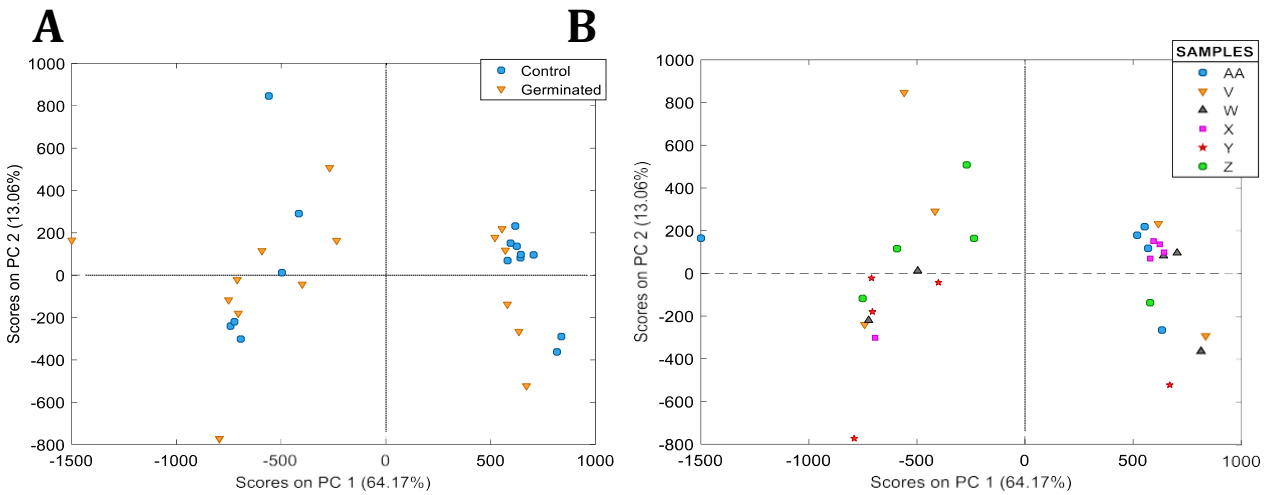


Figure 4.3.6. PC score plots for the 2020 roasted coffee (*C. arabica*) data. (A) Samples are coloured by the treatment applied. (B) Samples are coloured by the sample code. Each sample includes five technical repeats. The samples labelled V, W, and X are control coffee samples. The samples labelled Y, Z, and AA are germinated coffee samples.

Application of a supervised analysis method, such as pre-selecting the data according to the treatment applied, can allow for identification of differences between control and germinated samples. Orthogonal partial least squares discriminant analysis is a type of supervised analysis method which can be used to determine the driving factors of separation between control and germinated coffee beans in this study are.

4.3.3 Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA)

OPLS-DA was used to visualize trends that were specific to the treatment that was applied, as well as determine which compounds were contributing to the differences. Typically, in literature, PCA highlights trends in the data which can then be further observed in the OPLS-DA results (Li *et al.*, 2008; Zhang *et al.*, 2015). However, since the PCA for the current research did not reveal satisfactory differences between control and germinated samples, the outcome of the OPLS-DA should be interpreted with caution.

The leave-*n*-out cross-validation approach was used. This process holds one observation out of the model and performs analysis (OPLS-DA in this case) on the

Stellenbosch University https://scholar.sun.ac.za

remaining samples in order to predict the one that was held back. It then repeats this process for all the observations. All the models had 30 observations except for the Green 2019 model which had 25 observations. This is because of the five outliers identified in sample "G" that were removed. This method of validation was selected because of the small sample size in this study. The outcome of the OPLS-DA model can be seen in Table 4.3.**1**.

Table 4.3.1. OPLS-DA parameter results of four different models.

Model	Observations	X Variables	R^2Y (cum.)	Q^2Y (cum.)
Green 2019	25	33	0.8	0.25
Green 2020	30	38	0.2	0.025
Roasted 2019	30	37	0.7	0.35
Roasted 2020	30	20	0.86	0.28

The output parameters for the OPLS-DA models include the coefficient of determination (R^2Y) and the goodness of prediction (Q^2Y). The former represents how much variation is explained by the model while the latter explains the predictive power or performance (Zhang *et al.*, 2015) which in this case is based on cross-validation. The model indicates that a moderate separation was achieved for the treatment based on the R^2Y values. With the exception of the Green 2020 model, which achieved a low R^2Y value (0.2) indicating that the model could only explain or fit 20% of the data. The Green 2020 model therefore could not find a relationship between the measured variables (chemical compounds) and the class labels. The extremely low Q^2Y values also suggests that all the models are unreliable and would not be able to distinguish future samples based on the treatment, and that the model has overfitted the data. Ideally, the Q^2Y value should not be less than 0.4 (Worley & Powers, 2013). Although undesirable, these model values were to be expected as the sample size was small because the different harvest years were kept separate. Perhaps, if the peak alignment of the chromatograms had been achieved then the model parameters could have improved.

Although the model is unreliable in terms of future predictions for South African coffees, the model can still provide information as to which compounds were contributing to the variation for this specific research. This information is found from the model's corresponding S-plot which is also produced from OPLS-DA (Figure 4.3.7). Each point represents a retention time-*m/z* (RT-*m/z*) pairing for an ion detected in the LCMS

Stellenbosch University https://scholar.sun.ac.za

analysis. The variables circled in green had a variable importance in projection (VIP) of larger than 2.5 which was based upon previous literature (Bao *et al.*, 2016). Thus, indicating the important metabolites for separating the control and treated coffee groups. In total, 18 ion RT-*m/z* pairs were selected based on the above criteria. The tentative identification of 14 compounds was based on comparisons between the current LCMS information and that of published literature. The mass-to-charge ratio and retention time for each variable was used (Table 4.3.2).

The majority of compounds identified belonged to the chlorogenic acid family and were specifically derived from caffeoylquinic acid. The caffeoylquinic acid was observed in both green and roasted coffees for both harvest years. In each model, the presence of a compound with *m/z* of 353 (compounds 1, 5, 8, 11, and 16) indicates that isomeric structures of this compound are present. Isomers are compounds with the same chemical formula but which have different arrangements of atoms (Scheschkewitz, 2016). These isomers may therefore have chemical and/or physical properties which are different from one another (Scheschkewitz, 2016). This would explain the different elution times of the "same" compound during LCMS analysis as this method involves compounds interacting with a stationary phase based on their chemical properties. The study by del Pilar Fernandez-Poyatos *et al.* (2019) only identified the compounds as caffeoylquinic acids. In comparison, the work by Jaiswal *et al.* (2010) further identified the structures based on their fragmentation patterns. The MS^E fragments of *m/z* 191, 179, and 135 are important for characterizing Neochlorogenic acid (3-*O*-caffeoylquinic acid), Cryptochlorogenic acid (4-*O*-caffeoylquinic acid), and Chlorogenic acid (5-O-caffeoylquinic acid) (Willems *et al.*, 2016). Neochlorogenic acid produces the fragment ions with *m/z* 191 and 179 only, Cryptochlorogenic acid produces all three daughter fragments, and Chlorogenic acid produces only the *m/z* 191 fragment (Willems *et al.*, 2016). Compound 12 was identified as Quinic acid (*m/z* 191) and compound 17 was also a Quinic acid but with the loss of a water (H_2O) molecule (*m/z* 173). Both compounds 12 and 17 were identified in the roasted coffees which was expected because during the roasting process the chlorogenic acid molecules are degraded. These compounds would not be ideal for distinguishing between control and germinated coffees because their presence would be found in all *Coffea* spp regardless.

Compounds 9, 10, and 15 with a *m/z* of 707 were assumed to be dimers of caffeoylquinic acid. In other words, two caffeoylquinic acid structures bound to each

Stellenbosch University https://scholar.sun.ac.za

other via hydrogen bonds. Similarly, compounds 2 and 3 with a *m/z* of 705 are also dimers of caffeoylquinic acid except they are assumed to have undergone dehydrodimerization (del Pilar Fernandez-Poyatos *et al.*, 2019). This process is the removal of a hydrogen atom to form a radical which becomes the new location of the bond. These compounds are similar to the X-*O*-caffeoylquinic acids mentioned earlier. However, the presence of these compounds as well as the *m/z* 353 are not suitable for playing a role as a biomarker because they are present sporadically between the different production stages and harvest years. In this case, an ideal biomarker should be selective for the specific treatment applied.

Both compounds 6 and 14 had a parent ion of *m/z* 515 and produced daughter ions of *m/z* 353 and 179. This corresponded to the compound named 3,5-dicaffeoylquinic acid which, as the name suggests, is a caffeoylquinic acid derivative. This compound has been previously reported in *Coffea* spp. (Clifford *et al.*, 2003). Since this is not a typical caffeoylquinic acid derivative, this could potentially be important for separating control and germinated beans. Further studies would be required to confirm this claim, especially since the compound was higher in the *2019 germinated green* beans but then instead higher in the *control* group for the *roasted* beans of the same harvest year.

To the authors best knowledge, the four unknown compounds (4, 7, 13, and 18) have not been previously reported in *Coffea* spp. Compound 4 had a parent ion at *m/z* 525 which yielded daughter ions at *m/z* 481 and 119. Compound 7 had a $[M-H]^-$ ion at *m/z* 771 and produced fragments at *m/z* 727 and 360. Compound 13 ($[M-H]^-$ ion at *m/z* 335) is thought to perhaps be a chlorogenic acid derivate because of the daughter ion at *m/z* 191 (quinic acid). Compound 18 had a parent ion at *m/z* 963 and yielded fragment ions at *m/z* 591 and *m/z* 545. Compounds 4, 13, and 18 had larger response values for the germinated beans compared to the control beans. This should perhaps be addressed in future work as they may be specific to the germination treatment and may be of valuable use in measuring the influence of germination on coffee beans. Compound 7 should not be ruled out as unimportant even though it was higher in the control coffee groups, because it could be unique to the South African coffee variety.

The OPLS-DA S-plot revealed interesting compounds for the separation of control and germinated coffee beans (Figure 4.3.7B, D, E, H). It is vital to reiterate that the goodness of prediction (Q^2Y) parameters were less than satisfactory, and future studies might find that other compounds separate the two groups. However, since South African

Stellenbosch University https://scholar.sun.ac.za

coffee varieties have not been previously reported on (at the time of writing) this study provides some clarity on the chemical composition of these coffees and gives a starting point for future identifications.

4.3.4 Summary

The results from PCA and OPLS-DA revealed no trends in the data to suggest that germination had a significant effect on the chemical composition of *C. arbaica* beans from South Africa. These results agree with the results observed in the targeted approach section (section 4.2). Clusters in PCA suggest that the grinding method was not optimal and future work should include a particle size assessment to ensure optimum compound extraction.

The OPLS-DA models were adequate for the current dataset but would not be suitable for future sample classifications. The OPLS-DA S-plot's revealed interesting compounds for the separation of control and germinated coffee beans. The Q^2Y parameters were less than satisfactory, and future studies might find other compounds that separate the two groups. This study provides some clarity on the chemical composition of South African coffees and gives a starting point for future compound identifications.

Stellenbosch University https://scholar.sun.ac.za

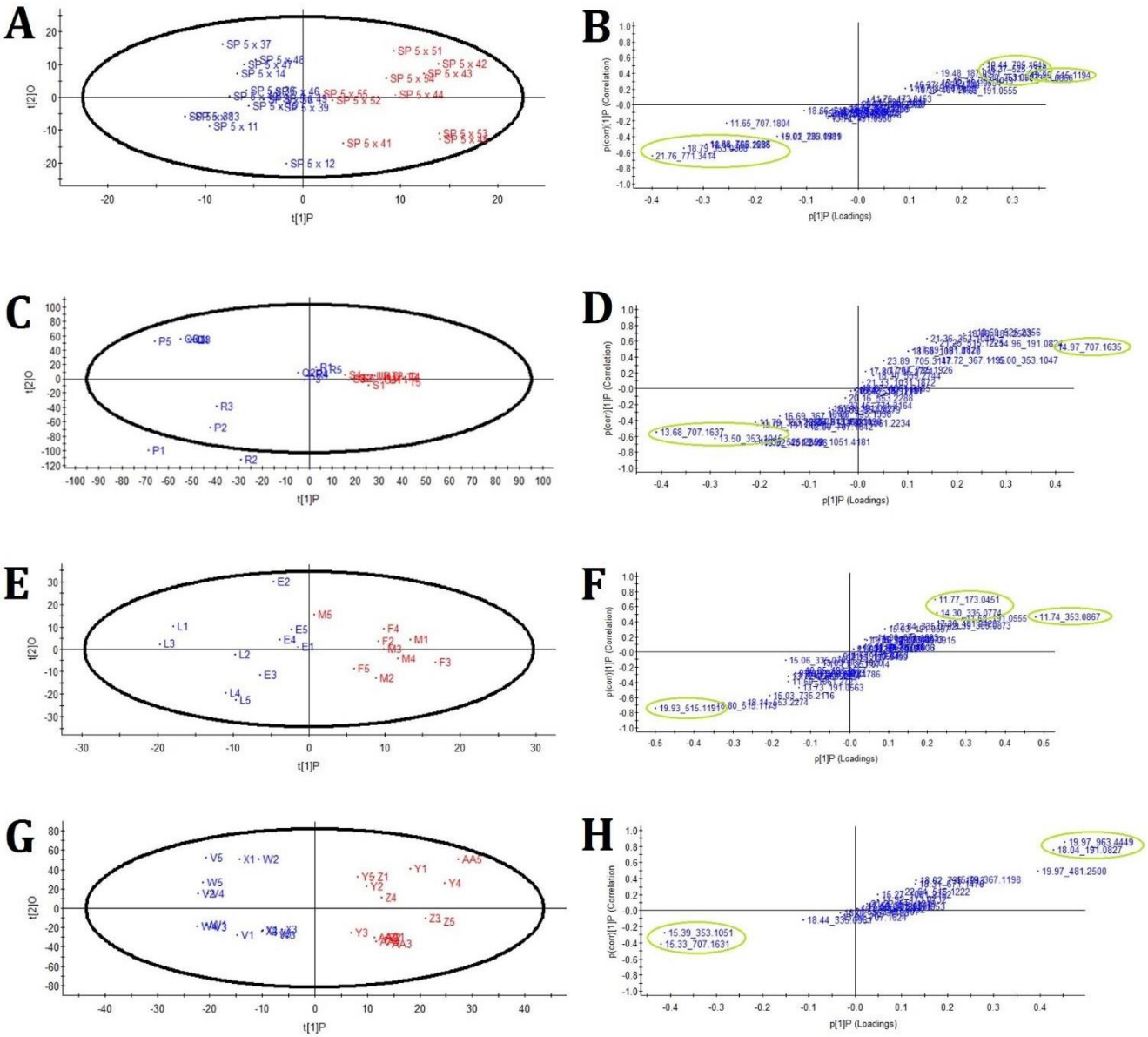


Figure 4.3.7. OPLS-DA score plot and the corresponding S-plot of HPLC-QTOF-MS profiling data of coffee bean extracts. (A,B) OPLS-DA score plot and S-plot for the green, 2019-harvest coffee beans. (C,D) OPLS-DA score plot and S-plot for the green, 2020-harvest coffee beans. (E,F) OPLS-DA score plot and S-plot for the roasted, 2019-harvest coffee beans. (G,H) OPLS-DA score plot and S-plot for the roasted, 2020-harvest coffee beans. Score plot: control group = blue, and germinated group = red; black ellipse indicates 95% confidence interval. S-plot: variables with VIP ≥ 2.5 are circled in green; control group = -1, and germinated group = +1.

Stellenbosch University https://scholar.sun.ac.za

Table 4.3.2. The tentative identification of compounds via HPLC-QTOF-MS analysis of coffee extracts.

No.	OPLS-DA Model	Primary ID	Retention time (min)	$[M-H]^-$ (m/z)	MS^E fragments (m/z)	C to G ratio[a]	Potential identification(s)
1	Green 2019	9.30_353.0867	9.3	353.0867	191, 179, 135	↓	Caffeoylquinic acid[b]; 4-*O*-caffeoylquinic acid[c]
2		10.44_705.1645	10.44	705.1645	513, 339, 229, 191	↓	Caffeoylquinic acid dehydrodimer[b]
3		11.66_705.1638	11.66	705.1638	513, 339, 229, 191	↑	Caffeoylquinic acid dehydrodimer[b]
4		16.37_525.2340	16.37	525.234	481, 119	↓	Unidentified
5		18.79_353.0868	18.79	353.0868	191, 179	↑	Caffeoylquinic acid[b]; 3-*O*-caffeoylquinic acid[c]
6		19.95_515.1194	19.95	515.1194	353, 179	↓	3,5-dicaffeoylquinic acid[b]
7		21.76_771.3414	21.76	771.3414	727, 360	↑	Unidentified
8	Green 2020	13.50_353.1045	13.5	353.1045	191	↑	Caffeoylquinic acid[b]; 5-*O*-caffeoylquinic acid[c]
9		13.68_707.1637	13.68	707.1637	353, 191	↑	Caffeoylquinic acid dimer[b]
10		14.97_707.1635	14.97	707.1635	353, 191	↓	Caffeoylquinic acid dimer[b]
11	Roasted 2019	11.74_353.0867	11.74	353.0867	191	↓	Caffeoylquinic acid[b]; 5-*O*-caffeoylquinic acid[c]
12		11.77_173.0451	11.77	173.0451		↓	Quinic acid less H_2O[d]
13		14.30_335.0774	14.3	335.0774	191	↓	Unidentified
14		19.93_515.1191	19.93	515.1191	353, 179	↑	3,5-dicaffeoylquinic acid[b]

Stellenbosch University https://scholar.sun.ac.za

15		15.33_707.1631	15.33	707.1631	353, 191	↑	Caffeoylquinic acid dimer[b]
16	Roasted 2020	15.39_353.1051	15.39	353.1051	191, 179	↑	Caffeoylquinic acid[b]; 3-*O*-caffeoylquinic acid[c]
17		18.04_191.0827	18.04	191.0827	110	↓	Quinic acid[d]
18		19.97_963.4449	19.97	963.4449	591, 545	↓	Unidentified

[a] C: control, G: germinated

[b] del Pilar Fernandez-Poyatos *et al.* (2019)

[c] Jaiswal *et al.* (2010)

[d] Simirgiotis *et al.* (2015)

Stellenbosch University https://scholar.sun.ac.za

References

Ashihara, H. (2008). Trigonelline (*N*-methylnicotinic acid) biosynthesis and its biological role in plants. *Natural Product Communications*, **3**(9), 1423-1428.

Bao, J.-L., Liu, F., Zhang, C., Wang, K., Xuejing, J., Wang, X., Chen, M., Li, P., Su, H., Wang, Y., Wan, J. & He, C. (2016). Anti-melanoma activity of *Forsythiae Fructus* aqueous extract in mice involves regulation of glycerophospholipid metabolisms by UPLC/Q-TOF MS-based metabolomics study. *Scientific Reports,* **6**(1), 39415.

Belay, S., Mideksa, D., Gebrezgiabher, S. & Seifu, W. (2016). Factors affecting coffee (*Coffea Arabica* L.) quality in Ethiopia: A review. *Journal of Multidisciplinary Scientific Research*, **4**(1), 22-28.

Bennat, C., Engelhardt, U. H., Kiehne, A., Wirries, F.-M. & Maier, H. G. (1994). HPLC analysis of chlorogenic acid lactones in roasted coffee. *Zeitschrift für Lebensmitteluntersuchung und-Forschung A*, **199**, 17-21.

Casal, S., Oliveira, M. B. & Ferreira, M. A. (1998). Development of an HPLC/Diode-Array detector method for simultaneous determination of trigonelline, nicotinic acid, and caffeine in coffee. *Journal of Liquid Chromatography & Related Technologies*, **21**(20), 3187-3195.

Cromer, M. R., Mai, H. T., Richards, K. M., Smith, R. E. & Tran, K. (2014). *Determination of chlorogenic acid in green coffee beans and dietary supplements labeled green coffee extract using LC-MS/MS*. Lenexa, Kansas: U.S. Food and Drug Administration.

Dawson, R. (2011). How significant is a boxplot outlier? *Journal of Statistics Education*, **19**(2), 1-13.

del Pilar Fernandez-Poyatos, M., Ruiz-Medina, A., Zengin, G. & Llorent-Martinez, E. J. (2019). Phenolic characterization, antioxidant activity, and enzyme inhibitory properties of *Berberis thunbergia* DC. leaves: a valuable source of phenolic acids. *Molecules*, **24**(22), 4171.

Farah, A., de Paulis, T., Trugo, L. C. & Martin, P. R. (2005). Effect of roasting on the formation of chlorogenic acid lactones in coffee. *Journal of Agricultural and Food Chemistry*, **53**, 1505-1513.

Hodge, V. J. & Austin, J. (2004). A survey of outlier detection methodologies. *Artificial Intelligence Review*, **22**, 85-126.

Stellenbosch University https://scholar.sun.ac.za

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 target and non-targeted analytical strategies. *Food & Function*, **3**(9), 976-984.

Jaiswal, R., Patras, M. A., Eravuchira, P. J. & Kuhnert, N. (2010). Profile and characterization of the chlorogenic acid 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.

Korifi, R., Le Dreau, Y. & Dupuy, N. (2014). Comparative study of the alignment method on experimental and simulated chromatographic data. *Journal of Separation Science*, **37**(22), 3276-3291.

Li, M., Wang, B., Zhang, M., Rantalainen, M., Wang, S., Zhou, H., Zhang, Y., Shen, J., Pang, X., Zhang, M., Wei, H., Chen, Y., Lu, H., Zuo, J., Su, M., Qiu, Y., Jia, W., Xiao, C., Smith, L.M., Yang, S., Holmes, E., Tang, H., Zhao, G., Nicholson, J.K., Li, L. & Zhao, L. (2008). Symbiotic gut microbes modulate human metabolic phenotypes. *Proceedings of the National Academy of Sciences*, **105**(6), 2117-2122.

Moon, J.-K., Yoo, H. S. & Shibamoto, T. (2009). Role of roasting conditions in the level of chlorogenic acid content in coffee beans: correlation with coffee acidity. *Journal of Agricultural and Food Chemistry*, **57**, 5365-5369.

Oestreich-Janzen, S. (2010). Chemistry of coffee. In: *Comprehensive Natural Products II: Biology and Chemistry,* 3rd ed. (edited by M. Lew & L. Hung-Wen). 1085-1117. Oxford: Elsevier.

Perrone, D., Donangelo, C. M. & Farah, A. (2008). Fast simultaneous analysis of caffeine, trigonelline, nicotinic acid and sucrose in coffee by liquid chromatography-mass spectrometry. *Food Chemistry*, **110**(4), 1030-1035.

Pinelo, M., Tress, A. G., Pedersen, M., Arnous, A. & Meyer, A. S. (2007). Effect of cellulases, solvent type and particle size distribution on the extraction of chlorogenic acid and other phenols from spent coffee grounds. *American Journal of Food Technology*, **2**(7), 641-651.

Scheschkewitz, D. Heavyweight isomer brings stability. *Nature Chemistry*, **8**(11), 993-995.

Simirgiotis, M. J., Benites, J. Areche, C. & Sepulveda, B. (2015). Antioxidant capacities and analysis of phenolic compounds in three endemic *Nolana* species by HPLC-PDA-ESI-MS. *Molecules*, **20**(6), 11490-11507.

Stellenbosch University https://scholar.sun.ac.za

Stander, M. A., Van Wyk, B.-E., Taylor, M. J. C. & Long, H. S. (2017). Analysis of phenolic compounds in rooibos tea (*Aspalathus linearis*) with a comparison of flavonoid-baed compounds in natural populations of plants from different regions. *Journal of Agricultural and Food Chemistry*, **65**, 10270-10281.

Steiman, S. R. (2003). *Method development for green coffee analysis and its possible application for group discrimination and correlation of green coffee chemistry with cupping quality.* Master of Science in Horticulture Thesis, University of Hawai'i, United States of America.

Sunarharum, W. B., Williams, D. J. & Smyth, H. E. (2014). Complexity of coffee flavor: a compositional and sensory perspective. *Food Research International*, **62**, 315-325.

Willems, J. L., Khamis, M. M., Saeid, W. M., Purves, R. W., Katselis, G., Low, N. H. & El-Aneed, A. (2016). Analysis of a series of chlorogenic acid isomers using differential ion mobility and tandem mass spectrometry. *Analytica Chimica Acta*, **933**, 164-174.

Worley, B. & Powers, R. (2013). Multivariate analysis in Metabolomics. *Current Metabolomics*, **1**(1), 92-107.

Zhang, L., Li, M., Zhan, L., Lu, X., Liang, L., Su, B., Sui, H., Gao, Z., Li, Y., Liu, Y., Wu, B. & Liu, Q. (2015). Plasma metabolomic profiling of patients with diabetes-associated cognitive decline. *Public Library of Science: One*, **10**(5), 1-15.

Zhou, Y., Su, P., Yin, H., Dong, Z., Yang, L. & Yuan C. (2019). Effects of different harvest times on the maturity of polyphenols in two red wine grape cultivars (*Vitis vinifera* L.) in Qingtongxia (China). *South African Journal of Enology and Viticulture*, **40**(1), 1-12.

Stellenbosch University https://scholar.sun.ac.za

5 General conclusions and recommendations

5.1 General conclusions

The aim of the present research was to determine whether the use of germination, as a novel post-harvest treatment during coffee production, had an influence on the chemical composition of a South African coffee variety (*Coffea arabica* Catui). The analysis included a targeted approach of key compounds identified from literature, followed by an untargeted approach to gain information regarding other compounds that are influencing the coffee's profile.

Coffee is a popular choice of beverage for individuals and is consumed on a daily basis worldwide. Unsurprisingly, the coffee plant as well as the beverage have been the focus of research studies for nearly a century. Specifically, in the last few decades there has been an increased interest in the effects of post-harvest treatments on the quality of coffee (Bytof *et al.*, 2000; Gonzalez-Rios *et al.,* 2007). The germination of coffee seeds has also been studied (Selmar *et al.*, 2006; Waters *et al.*, 2017), however not in the capacity where it is used as an intentional post-harvest treatment method.

The focus of the initial scientific literature review was the identification of key compounds which are important for coffee characteristics and quality as well as gain an understanding of the current post-harvest treatments and their effects on coffee production. There are well over 1 000 compounds identified in coffee beans (Eroz Poyraz *et al.*, 2016). However, only a handful have been recognized as playing defining roles in the aromatic and flavour quality of the coffee. In terms of aroma development, chlorogenic acid is involved in numerous pathways that result in the production of volatile components, and it is also important for flavour as it imparts astringency (Buffo & Cardelli-Freire, 2004). Trigonelline was also identified as an important compound that contributes toward aroma of coffee products. After roasting, the compound continues to break down to produce volatile compounds such as pyridines and pyrroles that are generally linked to fish-like and nutty aromas, respectively (Oestreich-Janzen, 2010). Coffee beverages also have numerous health benefits (Higdon & Frei, 2006). Nicotinic acid (or niacin / vitamin B3) is found in coffee and linked to lowering cholesterol levels in the body (Higdon & Frei, 2006). Since nicotinic acid is produced from trigonelline during the roasting stage of coffee, it serves

Stellenbosch University https://scholar.sun.ac.za

as an important compound to investigate in terms of health benefits as well as chemical composition changes. An additional health benefit of coffee is stimulation of the nervous system and improved brain functioning which is attributed to the caffeine content (Higdon & Frei, 2006; Oestreich-Janzen, 2010). Caffeine is notably linked to all coffee beverages and is an important aspect when grading coffee. It is also responsible for the strength and body of a brew, and only partially associated to the bitterness (Oestreich-Janzen, 2010). Although the four chosen compounds were indeed important for coffee flavour, there was a missed opportunity to also evaluate the carbohydrate content of the coffees especially in terms of mono- and polysaccharide content. These components are integral during the roasting stage when the Maillard reaction occurs and therefore might have been influenced differently because of the germination process. However, numerous sugars are present in coffees and a comprehensive analysis would have required an in-depth investigation which was beyond the scope of this study.

The next objective was to determine if the germination treatment had an influence on the concentrations of the four key compounds selected. The plant material preparation included a grinding and extraction stage. A MM400 Mixer Mill (Retsch, Germany) was used to finely grind the green and roasted beans. Green coffee beans are extremely firm and required more force to obtain the desired particle size, compared to roasted coffees which are softer and easily ground down. Particle size uniformity was not assessed according to a standard, instead only a visual assessment of "fineness" was performed. This proved unsatisfactory when evaluating the PCA results which revealed that the technical repeated measures for a sample was chemically different, i.e. the replicate measures did not cluster together. The effect of particle size on sufficient extraction has been previously discussed (Pinelo *et al.*, 2007) and it was found that a size of 125 µm allows for maximum surface contact with the solvent. Thus, it is recommended that a particle size assessment should be performed in future work.

The phenolic compounds were extracted using a 50% methanol-water with 1% formic acid solvent followed by a 30 min sonication period without applying heat. The chromatographic data results showed well resolved peaks and were ideal for quantifying the necessary target compounds. Overall, similar trends were observed in the works of Moon *et al.* (2009), Sunarharum (2016), and Jeska-Skowron *et al.* (2020) with regards to the chosen compounds of interest and their respective concentrations.

Stellenbosch University https://scholar.sun.ac.za

The average concentrations reported in this study for green coffee beans were generally not in line with previous reports. For instance, Moon *et al.* (2009) and Sunarharum (2016) both reported chlorogenic acid content of 50 mg/g whereas this study observed an average of 22.92 mg/g DW. As for caffeine content, the works mentioned above reported concentrations of 12 mg/g, in comparison to the current research which found an average of 24.90 mg/g DW. This could be due to the fact that the current research used South African coffee beans compared to the likes of Brazilian, Ethiopian and Guatemalan coffees. This is the first reporting of caffeine, chlorogenic acid, trigonelline, and nicotinic acid in arabica coffees from South Africa.

The last objective also focused on the potential influence of germination as a post-harvest treatment but in terms of an untargeted approach by means of PCA and OPLS-DA. Pre-processing multivariate data was identified as an important step prior to analysis of variance. This is performed to remove unwanted variation and is generally specific to the type of variation, i.e. normalizing a dataset corrects for differences in overall concentrations between samples. Peak alignment of the chromatographic data was necessary for the different harvest year datasets. Unfortunately, due to limitations in the MarkerLynx software package, the chromatograms could not be corrected for peak shifts. This resulted in separate analyses that further reduced the sample size for each production and treatment group.

PCA and OPLS-DA are best performed with larger datasets because they are able to produce robust models that can be utilized in predictive algorithms. It is suggested that additional green and roasted coffee samples from different harvest years and batches be obtained that also undergo germination in order to increase the dataset. Furthermore, an improved method for processing peak alignments in the chromatograms can be achieved. Although the OPLS-DA models were not deemed reliable for predictive work in future studies, certain compounds were still identified as important. Overall, the family of chlorogenic acid compounds were prominent in separating control and germinated coffees with 12 of the 18 metabolites recognized as a caffeoylquinic acid derivative. In agreement with similar suggestions made by Sunarharum (2016), it is recommended that internal standards of different chlorogenic acid isomers be used. This may assist in profiling the South African coffees as well as confirm the specific compounds identified from the OPLS-DA results. Internal standards of chlorogenic acid

Stellenbosch University https://scholar.sun.ac.za

isomers could include, but are not limited to, 3-/4-/5-caffeoylquinic acid and dicaffeoylquinic acids.

5.2 Recommendations for future research

The improvement of methodologies and analysis techniques to study coffee beans are crucial for future studies because better results and greater knowledge can benefit its production process and overall coffee quality. Research should be focused on studying the complexity of coffee chemistry and the challenges that come with studying it because there are diverse chemical compositions for the same coffee species (arabica versus robusta) but also for coffee from different origins (Ethiopian versus South African) and varieties/cultivars.

The germination treatment for the current project was performed as part of the coffee post-harvest process which occurred externally to the experimental procedure. In-person evaluation of the production premises revealed that the processing is not done under scientific conditions. General hygiene practices are maintained for food safety reasons, but perfect consistency between batches is not necessary at the market level. Therefore, variation in the stages of soaking, drying, germinating, and roasting could have had an impact on the chemical composition of the coffee beans. A recommendation for future studies would be to perform the germination treatment and the roasting process as an in-house experiment in order to maintain experimental consistency and limit any outside factors influencing the outcome. This could also be expanded into exploring approaches to optimize the germination treatment and evaluate its influence on the chemical composition.

Quality parameters of coffee are not only found in the non-volatile fraction of coffees but also the volatile fractions. Therefore, it is advised that follow-up research should include an analysis of the volatile profile for the control and germinated coffees to determine if aroma compounds are affected by the proposed new post-harvest treatment. This focus could also be done in a targeted and untargeted approach, similar to the current research. Gas chromatography and other advanced techniques (e.g. olfactometry) can be utilized to compare chemical and sensory parameters (Sunarharum, 2016). This would be essential in thoroughly analyzing the effect of germination on volatile fractions with respect to overall coffee quality.

Stellenbosch University https://scholar.sun.ac.za

The reports of the current study were indicative but not conclusive and therefore it is recommended that an in-depth profiling of South African coffees be performed. This will provide a comprehensive foundation for South African coffees as well as allow for easier comparisons with coffees of other origins. This could be performed on both green and roasted coffee varieties grown in South Africa in order to determine "before and after" characteristics of the coffees. This could be followed up with a metabolomics approach to identify biomarkers within South African coffees in order to distinguish it from coffees of other origins. Such approaches could include ligand-binding assays or flow cytometry. However, these can be expensive analyses to run and preparatory research should be done first.

In conclusion, the study of germination as a novel, post-harvest treatment process on a South African coffee variety provided new insights. The importance of additional samples was emphasized during the targeted approach because *p*-values were low enough to suggest a trend but not enough to be classified as significant ($p = 0.05$). These trends could be clarified by including samples from different harvest years or different origins for example. Furthermore, the multivariate data analysis did not produce the anticipated models needed to accurately distinguish between future samples. Instead, it provides insight into the compounds that were identified in the present samples. These both can provide a platform for future research opportunities to grow from, with a specific focus on South African varieties.

Stellenbosch University https://scholar.sun.ac.za

References

Buffo, R. A. & Cardelli-Freire, C. (2004). Coffee flavour: an overview. *Flavour and Fragrance Journal*, **19**, 99-104.

Bytof, G., Selmar, D. & Schieberle, P. (2000). New aspects of coffee processing: How do the different post harvest treatments influence the formation of potential flavour precursors? *Journal of Applied Botany*, **74**, 131-136.

Eroz Poyraz, I., Ozturk, N., Kiyan, H. T. & Demirci, B. (2016). Volatile compounds of Coffea Arabica L. green and roasted beans. *Anadolu University Journal of Science and Technology C – Life Sciences and Biotechnology*, **5**(1), 31-35.

Gonzalez-Rios, O., Suarez-Quiroz, M.L., Boulanger, R., Barel, M., Guyot, B., Guiraud, J.-P. & Schorr-Galindo, S. (2007). Impact of "ecological" post-harvest processing on coffee aroma: II. Roasted coffee. *Journal of Food Composition and Analysis*, **20**(3-4), 297-307.

Higdon, J. V. & Frei, B. (2006). Coffee and health: A review of recent human research. *Critical Reviews in Food Science and Nutrition*, **46**, 101-123.

Jeska-Skowron, M., Frankowski, R. & Zgola-Grzeskowiak, A. (2020). Comparison of methylxantines, trigonelline, nicotinic acid and nicotinamide contents in brews of green and processed Arabica and Robusta coffee beans – Influence of steaming, decaffeination and roasting processes on coffee. *LWT – Food Science and Technology*, **125**, 109344.

Moon, J.-K., Yoo, H. S. & Shibamoto, T. (2009). Role of roasting conditions in the level of chlorogenic acid content in coffee beans: correlation with coffee acidity. *Journal of Agricultural and Food Chemistry*, **57**, 5365-5369.

Oestreich-Janzen, S. (2010). Chemistry of coffee. In: *Comprehensive Natural Products II: Biology and Chemistry,* 3rd ed. (edited by M. Lew & L. Hung-Wen). 1085-1117. Oxford: Elsevier.

Pinelo, M., Tress, A. G., Pedersen, M., Arnous, A. & Meyer, A. S. (2007). Effect of cellulases, solvent type and particle size distribution on the extraction of chlorogenic acid and other phenols from spent coffee grounds. *American Journal of Food Technology*, **2**(7), 641-651

Stellenbosch University https://scholar.sun.ac.za

6 Supplementary Notes

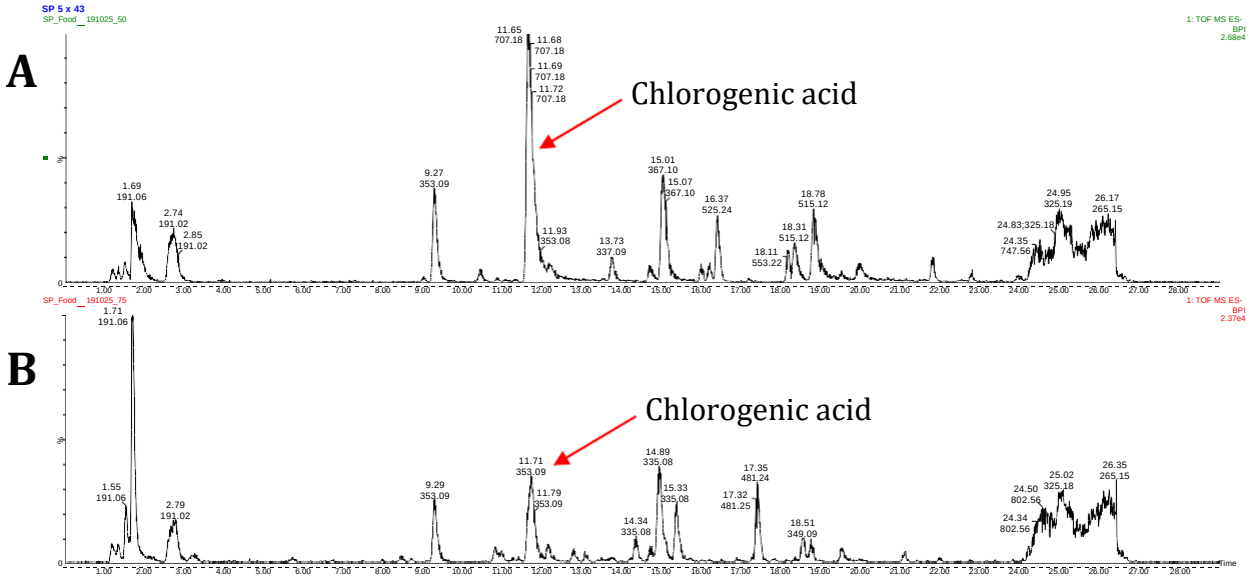


Figure S1. Typical base peak intensity chromatogram for the 2019 coffee harvest season. (A) Green, germinated coffee and (B) roasted, germinated coffee in ESI negative mode to illustrate chlorogenic acid peak identification. Horizonal axis was retention time and vertical axis was peak intensity.

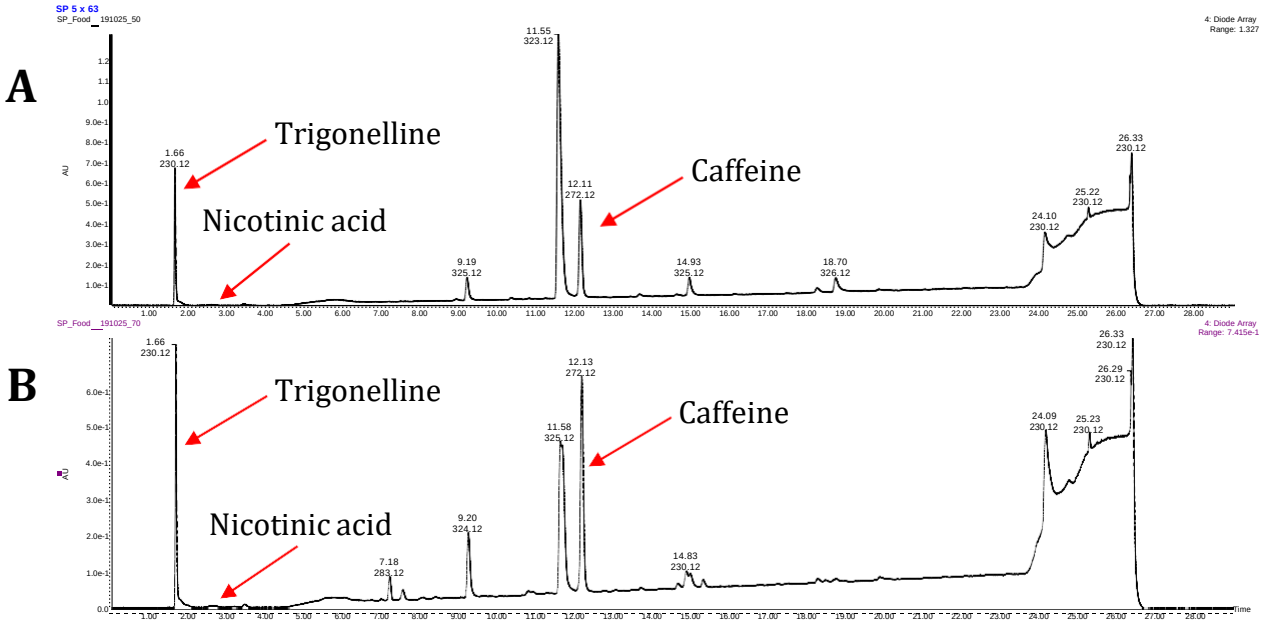


Figure S2. Typical UV chromatogram for the 2019 coffee harvest season. (A) Green, germinatedcoffee and (B) roasted, germinated coffee from the PDA detector to illustrate peakidentification of trigonelline (230 nm), nicotinic acid (261 nm), and caffeine (272 nm). Horizonal axis was retention time and vertical axis was peak intensity.

Stellenbosch University https://scholar.sun.ac.za

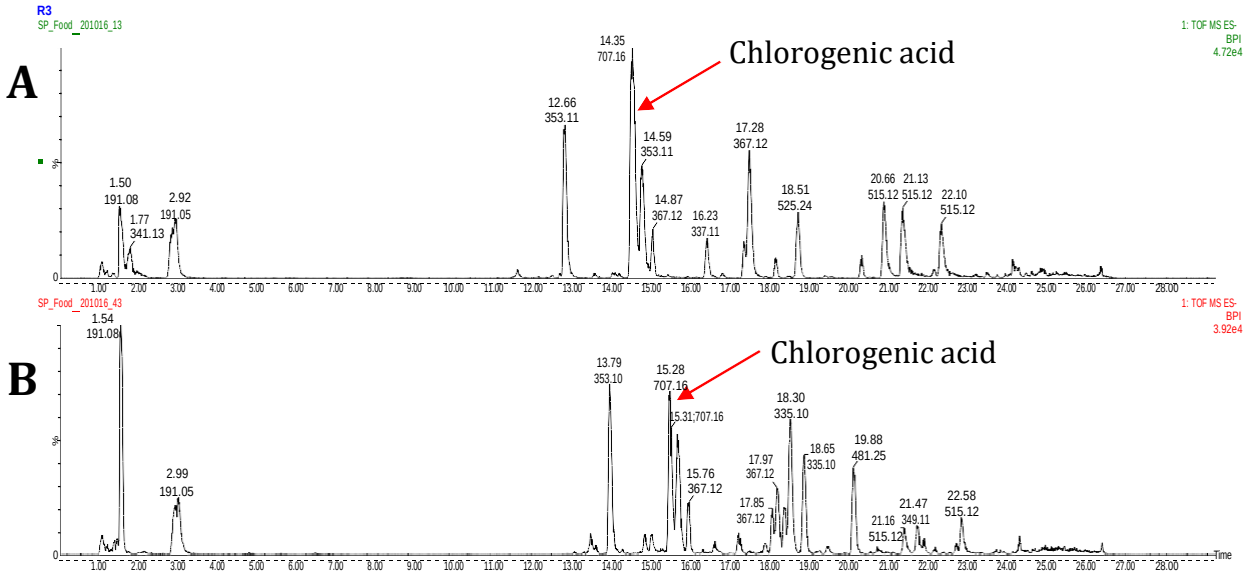


Figure S3. Typical base peak intensity chromatogram for the 2020 coffee harvest season. (A) Green, control coffee and (B) roasted, control coffee in ESI negative mode to illustrate chlorogenic acid peak identification. Horizonal axis was retention time and vertical axis was peak intensity.

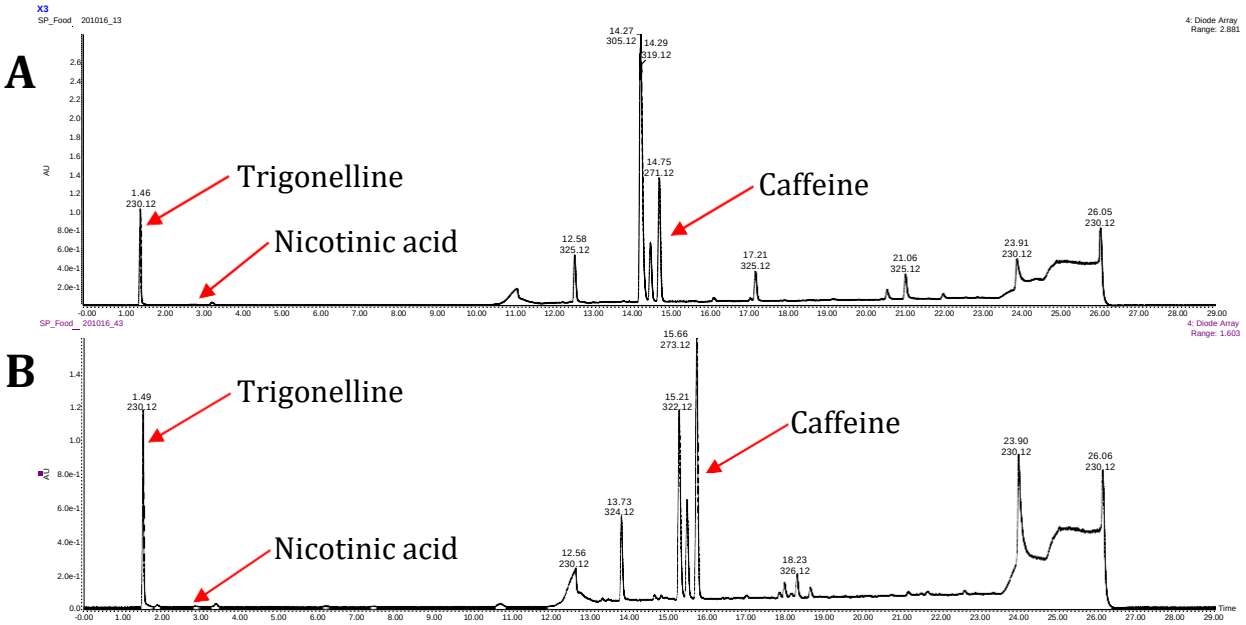


Figure S4. Typical UV chromatogram for the 2020 coffee harvest season. (A) Green, control coffee and (B) roasted, control coffee from the PDA detector to illustrate peak identification of trigonelline (230 nm), nicotinic acid (261 nm), and caffeine (272 nm). Horizonal axis was retention time and vertical axis was peak intensity.

Stellenbosch University https://scholar.sun.ac.za

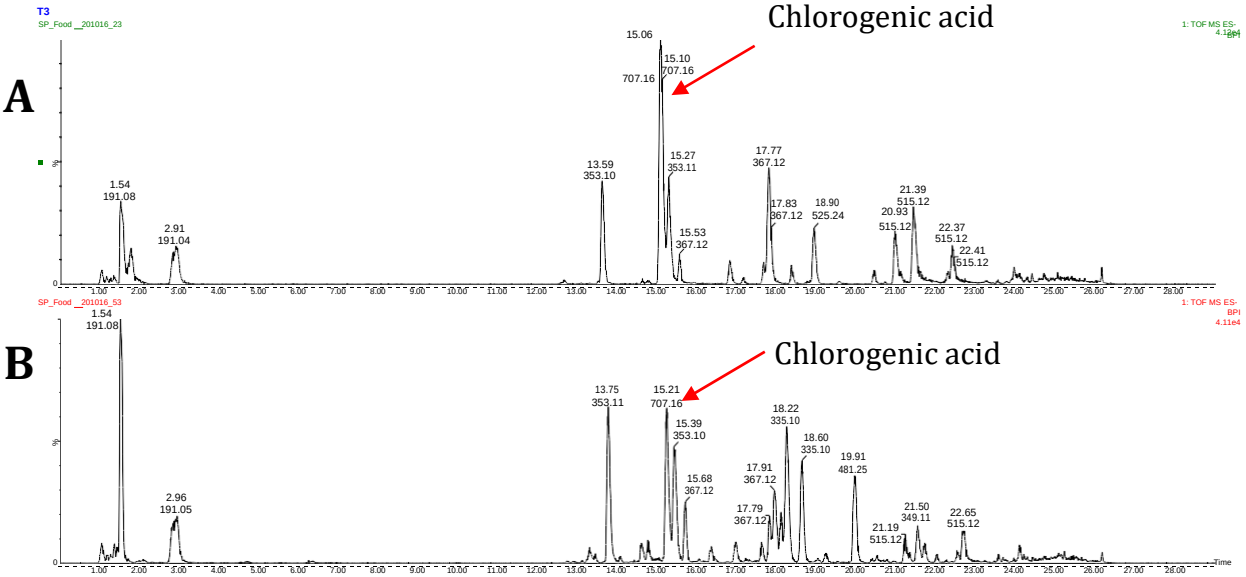


Figure S5. Typical base peak intensity chromatogram for the 2020 coffee harvest season. (A) Green, germinated coffee and (B) roasted, germinated coffee in ESI negative mode to illustrate chlorogenic acid peak identification. Horizonal axis was retention time and vertical axis was peak intensity.

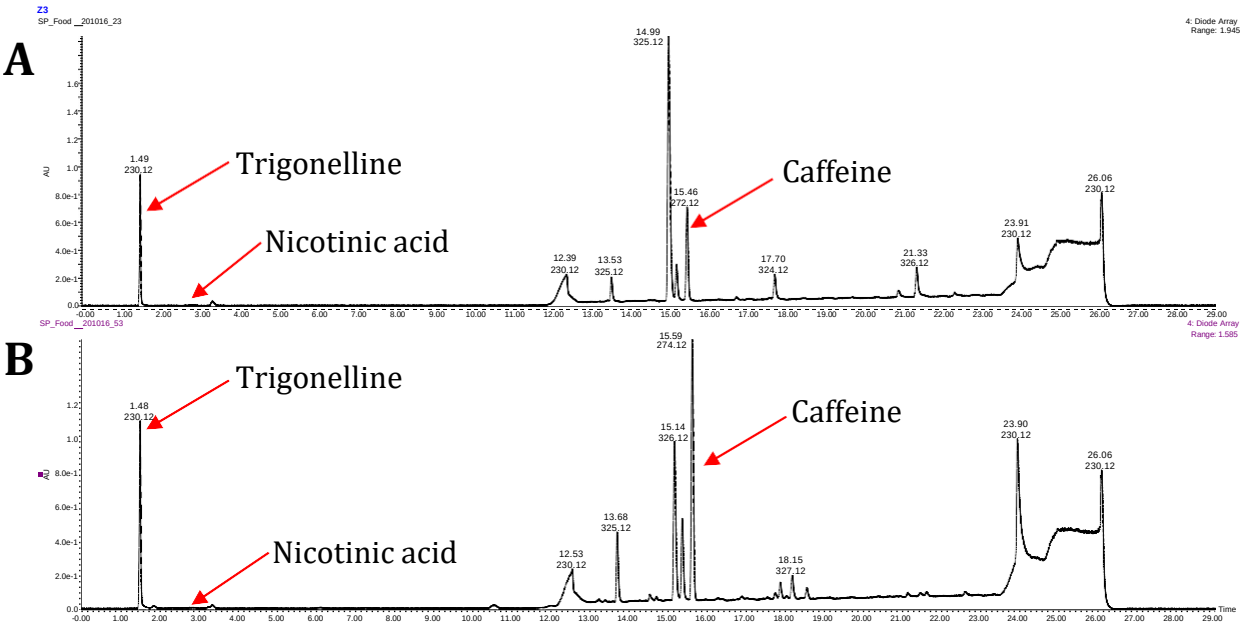


Figure S6. Typical UV chromatogram for the 2020 coffee harvest season. (A) Green, germinatedcoffee and (B) roasted, germinated coffee from the PDA detector to illustrate peakidentification of trigonelline (230 nm), nicotinic acid (261 nm), and caffeine (272 nm). Horizonal axis was retention time and vertical axis was peak intensity.

Stellenbosch University https://scholar.sun.ac.za

Table S2. Mean data of the measured concentrations for targeted compounds in green and roasted South African coffee beans (*C. arabica*).

Harvest Year	Batch	Stage	TMT[b]	Caffeine		Chlorogenic Acid[a]		Trigonelline		Nicotinic Acid		Sample Code
				Mean ±SD[c] (mg/g)	RSD[d] (%)	Mean ±SD (mg/g)	RSD (%)	Mean ±SD (mg/g)	RSD (%)	Mean ±SD (µg/g)	RSD (%)	
2019	1	Green	C[e]	12.94 ±0.48	3.77	29.06 ±0.43	1.50	7.69 ±0.25	3.26	119.20 ±13.1	11.06	C
2019	1	Roasted	C	13.49 ±1.01	7.52	21.82 ±0.79	3.62	10.52 ±0.83	7.91	64.0 ±35.3	55.16	E
2019	1	Roasted	G[f]	13.33 ±0.93	7.04	21.76 ±0.76	3.54	9.79 ±0.52	5.36	43.2 ±35.4	82.00	F
2019	1	Green	G	5.47 ±0.14	2.67	0.62 ±0.23	37.35	3.89 ±0.12	3.30	0.00	0.00	G
2019	2	Green	C	12.10 ±0.73	6.08	29.15 ±0.96	3.32	7.12 ±0.29	4.09	0.00	0.00	H
2019	2	Green	G	11.51 ±0.32	2.81	28.38 ±0.45	1.59	7.09 ±0.20	2.83	0.00	0.00	i
2019	2	Roasted	C	14.77 ±0.43	2.92	23.29 ±0.47	2.04	11.30 ±0.35	3.11	0.00	0.00	L
2019	2	Roasted	G	13.82 ±0.19	1.45	20.83 ±0.29	1.43	11.35 ±0.17	1.50	0.00	0.00	M
2019	3	Green	C	12.38 ±0.39	3.21	28.59 ±0.61	2.17	7.19 ±0.25	3.56	98.36 ±7.34	7.47	J
2019	3	Green	G	12.08 ±0.78	6.48	28.78 ±0.94	3.29	7.12 ±0.45	6.38	75.34 ±38.0	50.47	K
2019	3	Roasted	C	14.58 ±0.70	4.83	9.23 ±0.37	4.02	9.70 ±0.47	4.93	193.17 ±33.8	17.50	N
2019	3	Roasted	G	14.42 ±0.74	5.16	7.97 ±0.19	2.48	9.64 ±0.37	3.92	190.65 ±16.3	8.58	O
2020	4	Green	C	38.70 ±3.50	9.06	16.28 ±0.96	5.92	8.34 ±0.42	5.04	30.65 ±61.3	200.00	P
2020	4	Green	G	37.81 ±1.51	4.01	17.42 ±0.51	2.94	8.49 ±0.21	2.51	191.13 ±25.5	13.38	S
2020	4	Roasted	C	10.36 ±1.02	9.92	11.48 ±0.57	5.00	11.01 ±0.97	8.83	62.37 ±76.3	122.47	V
2020	4	Roasted	G	10.42 ±0.99	9.56	10.81 ±0.48	4.47	11.65 ±1.08	9.27	63.81 ±85.8	134.51	Y
2020	5	Green	C	37.87 ±0.98	2.61	17.45 ±0.33	1.94	8.35 ±0.30	3.66	150.49 ±30.4	20.22	Q
2020	5	Green	G	18.20 ±0.88	4.84	12.38 ±0.32	2.65	4.29 ±0.22	5.33	168.75 ±23.2	13.80	T

Stellenbosch University https://scholar.sun.ac.za

2020	5	Roasted	C	10.13 ±0.25	2.53	10.99 ±0.30	2.73	10.49 ±0.21	2.09	154.29 ±18.0	11.69	W
2020	5	Roasted	G	8.83 ±0.61	6.92	10.43 ±0.62	6.01	12.16 ±0.68	5.67	0.00	0.00	Z
2020	6	Green	C	35.42 ±1.05	2.97	16.98 ±0.32	1.92	8.40 ±0.31	3.76	190.44 ±22.0	11.58	R
2020	6	Green	G	33.70 ±1.40	4.17	15.16 ±0.52	3.43	7.64 ±0.33	4.44	180.02 ±39.1	21.75	U
2020	6	Roasted	C	9.96 ±0.65	6.60	10.75 ±0.37	3.47	11.48 ±5.71	49.75	114.19 ±90.7	79.43	X
2020	6	Roasted	G	10.81 ±0.64	5.99	11.82 ±0.49	4.15	12.29 ±6.15	50.04	0.00	0.00	AA

[a] General term given for 5-caffeoylquinic acid

[b] TMT: treatment

[c] SD: standard deviation

[d] RSD: relative standard deviation

[e] C: control

[f] G: germinated

All mean values were obtained from five technical repeated measures; data rounded to two decimal points only.

Stellenbosch University https://scholar.sun.ac.za

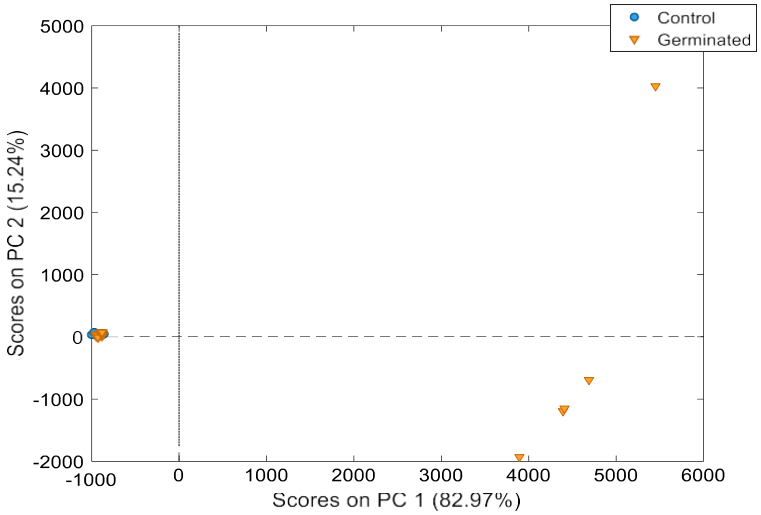


Figure S7. Initial PC score plot for the 2019 green coffee (*C. arabica*) data. Samples are coloured by the treatment applied. The five points clustered away from the main group are technical repeated measures for sample "G" (germinated coffee).