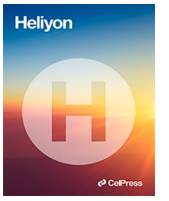
Heliyon 9 (2023) e21402

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Review article

Metabolomics as a tool for geographic origin assessment of roasted and green coffee beans

Claudia de León-Solis *, Victoria Casasola, Tania Monterroso

Instituto de Investigaciones Químicas, Biológicas, Biomédicas y Biofísicas, Mariano Gálvez University, 3ª Avenida 9-00 zona 2, 01002, Interior Finca El Zapote, Ciudad de Guatemala, Guatemala

ARTICLE INFO

Keywords:
Food metabolomics
Coffee beans metabolomic profiling
Geographic origin assignment
Instrumental techniques
Multivariate data analysis

ABSTRACT

Coffee is widely consumed across the globe. The most sought out varieties are Arabica and Robusta which differ significantly in their aroma and taste. Furthermore, varieties cultivated in different regions are perceived to have distinct characteristics encouraging some producers to adopt the denomination of origin label. These differences arise from variations on metabolite content related to edaphoclimatic conditions and post-harvest management among other factors. Although sensory analysis is still standard for coffee brews, instrumental analysis of the roasted and green beans to assess the quality of the final product has been encouraged. Metabolomic profiling has risen as a promising approach not only for quality purposes but also for geographic origin assignment. Many techniques can be applied for sample analysis: chromatography, mass spectrometry, and NMR have been explored. The data collected is further sorted by multivariate analysis to identify similar characteristics among the samples, reduce dimensionality and/or even propose a model for predictive purposes.

This review focuses on the evolution of metabolomic profiling for the geographic origin assessment of roasted and green coffee beans in the last 21 years, the techniques that are usually applied for sample analysis and also the most common approaches for the multivariate analysis of the collected data. The prospect of applying a wide range of analytical techniques is becoming an unbiased approach to determine the origin of different roasted and green coffee beans samples with great correlation. Predictive models worked accurately for the geographic assignment of unknown samples once the variety was known.

1. Introduction

Consumption of coffee is widespread across the globe. According to the International Coffee Organization (2022) "world coffee exports amounted to 11.11 million bags in June 2022". The complexity of coffee flavor arises from the combination of different factors that can be classified into three categories: environmental conditions, genetic resource and management [1]. Environmental conditions are usually specific for this crop, topography and climate should be optimal to ensure quality of coffee [2]. It is well documented how the composition of coffee beans varies according to the altitude at which they were grown [3–5]. In general, coffee grown in higher altitudes has a higher market price. Temperature has a key role in seed development which finally affects the sensory profile of the roasted coffee by increasing the concentration of volatile compounds that are related to off-flavors [6]. Regarding the genetic

* Corresponding author.
E-mail address: cdeleon@umg.edu.gt (C. de León-Solis).

https://doi.org/10.1016/j.heliyon.2023.e21402
Received 19 June 2023; Received in revised form 2 October 2023; Accepted 20 October 2023
Available online 30 October 2023

resource, as expected, coffee varieties differ in their metabolite composition and most notably Arabica brews are more acidic than Robusta due to their higher concentration of citric, malic, chlorogenic and quinic acids [7]. Management encompasses first agronomic management: irrigation, pest control, fertilization and harvesting strategies influence the quality of the bean regarding its size and ripeness that will ultimately affect its composition [1]. In fact, immature coffee cherries are characterized by having a lower amount of sucrose than those that were harvested at their optimal ripeness. Other undesirable compounds might also be present in larger amounts in the brewed product [8]. On the contrary, coffee cherries collected at their ideal point of maturity will exhibit greater amounts of sugars, decreasing the proportion of chlorogenic acid and increasing that of organic acids. A high concentration of fumaric, tartaric and oxalic acids is directly related to high-quality coffee since these acids have the property of enhancing the flavor and being responsible for providing a greater sensation of juiciness and granting a greater number of notes to the cup profile [9]. Management also relates to post-harvest handling. Arguably the most important factor is roasting where the volatiles that are produced are key to attribute coffee its quality and value. Chemically speaking, these volatiles belong to different families of compounds but the most significant to coffee flavor are pyrazines and sulfur-containing molecules. Other volatiles related to this characteristic of coffee are alcohols, aldehydes, furans, furanones, ketones and phenols just to name a few. Non-volatile compounds might also be involved: alkaloids such as caffeine which is related to bitterness and chlorogenic acids which are related to coffee's astringency are just some examples [7]. The variation of concentration of these compounds explains why coffee flavor is such a complex attribute. Different markers were identified in green and roasted coffee beans processed in three ways: natural, fully-washed and honey, thus evidencing the impact of post-harvest handling [10]. Grinding contributes as well to the perceived aroma due to the larger surface area that releases more volatiles [11]. Even the preparation method has an influence on the coffee aroma and it's been found that pressurized methods tend to extract less compounds than brewing [12].

Finally, although sensory analysis of a cup of coffee —based on visual impressions (crema, color and volume), aroma, flavor (acidity, fruitiness and roast, among other sensory differences) and body [13] is usually required, researchers are suggesting more reproducible assessments [5,14] factoring in the influence of enviromental factors [15] and linking directly coffee quality to specific geographic location [16]. Fig. 1 illustrates how some common influencing factors give different metabolomic profiles that affect coffee quality. Thus metabolomics rises as a promising tool for geographic origin assignment.

Carbohydrates, organic and amino acids with tannins, terpenes and flavonoids among others constitute the metabolome of any species. Metabolites are elements or small molecules with molecular weights lower than 1000 Da present in species that may vary between populations, regions and species among other factors. Monitoring these variations allows the identification of metabolic fingerprints or biomarkers that could potentially be used for species differentiation, origin assignment and variety authentication. This field is known as metabolomics. Metabolomics is an omics science, along with genomics, transcriptomics and proteomics where

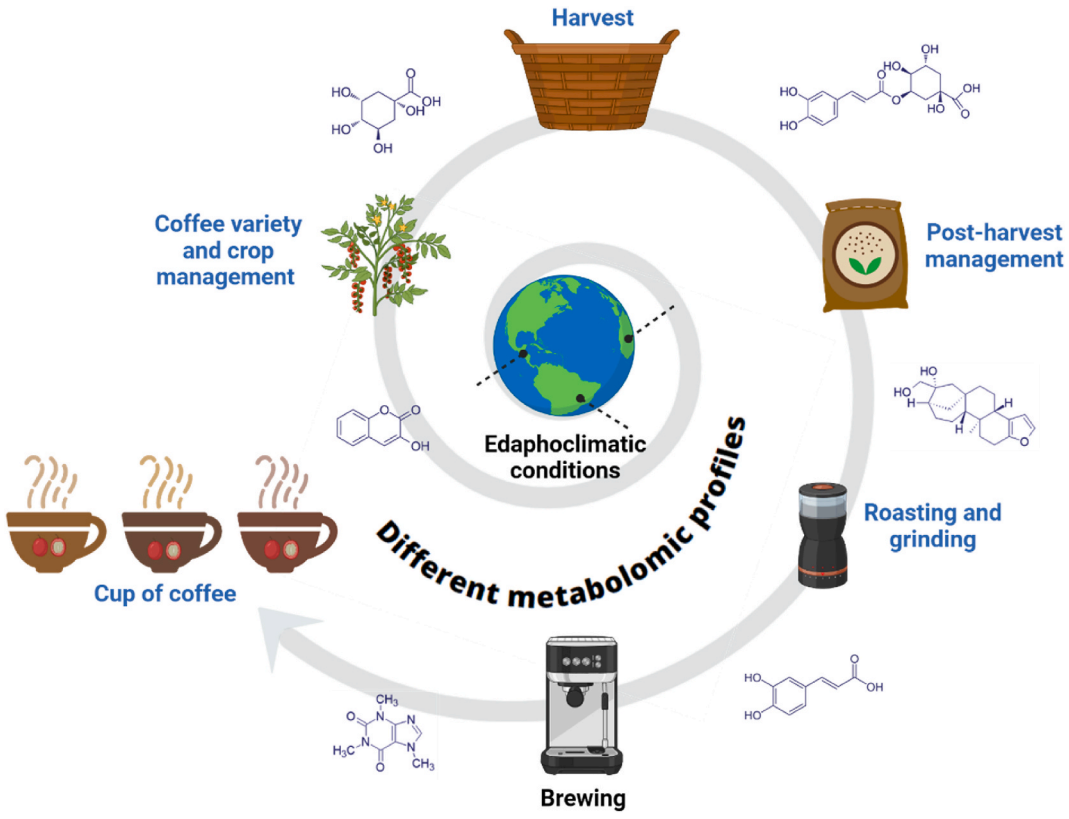


Fig. 1. Influencing factors on metabolomic profile key to quality of coffee. Created with BioRender.com.

Analytical Chemistry, Bioscience and Informatics intersect [17]. Through different analytical techniques, metabolites are identified and quantified, then this data may be processed through a multivariate analysis to group samples into blocks that share some characteristics. A model can also be generated with the data to assign an unknown sample to either group. This approach has been successful in the discrimination of different products according to their country of origin: spices such as saffron [18] and oregano [19]; edible plant products such as oranges [20], hazelnuts [21], celery [22] and argan [23]; beverages like cocoa [24] and green tea [25]; grains like wheat [26], rice [27] and maize [28]; and meat products like beef [29] and shrimp [30], have been successfully categorized by the identification of key biomarkers that set them apart. Differentiation by regions that are close by could potentially be more difficult but it has been achieved for Chinese licorice [31], Egyptian carob [32], Italian extra-virgin olive oil [33,34] and wine [35]. This has opened the field for authentication of products with Protected Denomination of Origin labels such as cheese [36,37]. Most recently, it was suggested that metabolomics techniques could be applied for the geographical indication (GI) registration process in Brazil [38]. Therefore this approach should be consider for a highly valuable product such as coffee.

Although other reviews on this topic have been published [39], they have focused on different coffee products while we have kept it to roasted and green coffee beans which are the main trade for many producing countries [40].

2. Analytical techniques used for metabolomic profiling of coffee

Metabolomic profiling is carried out through a wide variety of techniques available. A single sample can potentially be processed by different procedures, each giving unique information that is characteristic of that sample [41,42]. Fig. 2 shows the most frequently reported techniques involved in this particular omic science, namely liquid chromatography - mass spectrometry (LC/MS) [42–45]; gas chromatography - mass spectrometry (GC/MS) [46–49] and nuclear magnetic resonance (NMR) [50,51]. Others have also been used for this specific purpose such as ultraviolet spectrometry (UV) [8], infrared spectrometry [52–54] and Raman spectrometry [53,54]. Specifically, chromatography has many variants when coupled with different detectors such as quadrupole [48], triple quadrupole, time of flight [45] or another high-resolution mass spectrometry (MS) [43]. Also NMR allows a wide variety of experiments ranging from simple 1H spectrum to bi-dimensional correlations all with a single NMR tube [50,51]. The data collected has distinct degrees of precision but overall it has allowed the identification of metabolites and the characterization of samples by comparison with existing and home-made libraries [43,46,55]. Each technique will be discussed in more detail below.

2.1. Gas chromatography

Gas chromatography (GC) has proven to be a reliable and robust analytical tool that enables the observation and comparison of results obtained from analyzing the metabolomic profiles differences between samples of the same type (clinical, industry,

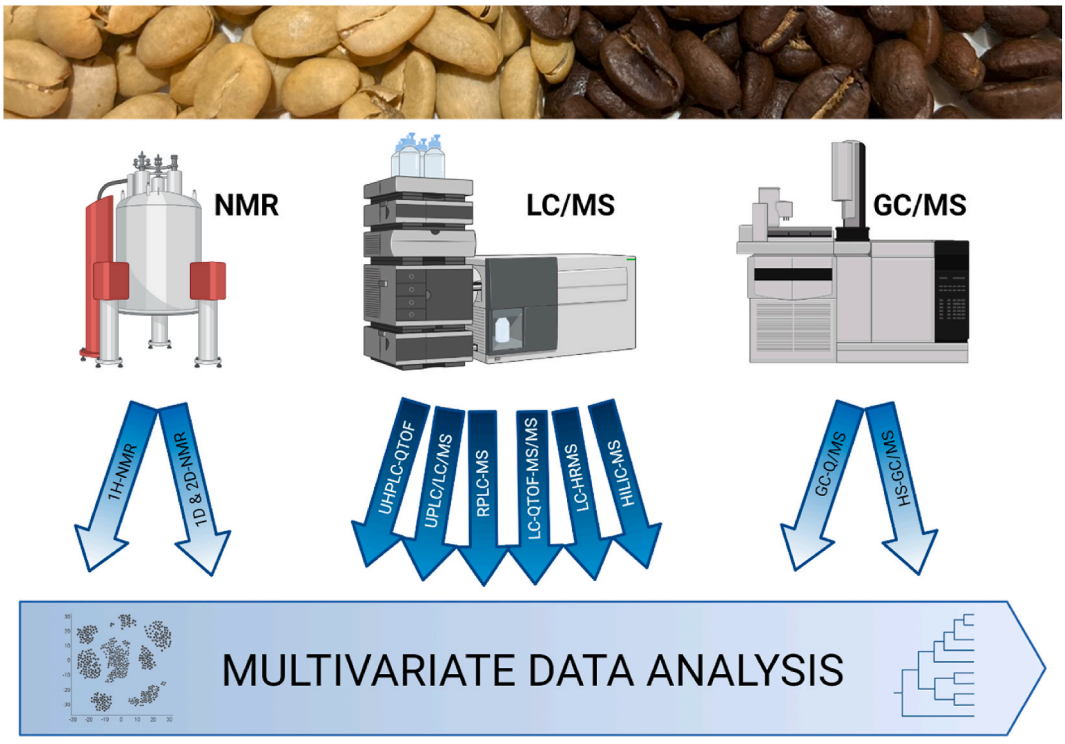


Fig. 2. Most frequently used techniques in metabolomics. Created with BioRender.com.

agricultural, etc) [56].

This technique is widely used for the identification and quantification of volatile and semi-volatile metabolites by usually coupling it with a mass detector [46,47,57] although there are multiple examples where a flame ionization detector (FID) was employed depending on the approach selected for the study [58,59]. For targeted approaches, where few metabolites are analyzed, GC coupled either with FID or MS has been useful in the coffee industry, it has been helpful for the metabolomic profiling of coffee seeds [48]. More specifically, the GC analysis of thermolabile coffee diterpenes has been carried out to establish their relationship with cup quality [60]. Although this strategy is key for well documented matrices or particular metabolites of interest [56], non-targeted approaches are favored since they provide the largest amount of information and are usually the first step prior to a targeted approach [61]. Untargeted approaches usually involve applying different extraction procedures [62] to ensure the analysis of the largest number of metabolites with different physicochemical properties, especially regarding their polarity, solubility and boiling point [56]. Sensory quality in ground coffee has also been evaluated by the identification of metabolite markers related to coffee quality [55]. Finally, it's been used to assess the solid state fermentation in Arabica coffee beans [46]. A summary of GC techniques and multivariate data analysis used for metabolomic profiling of coffee products are presented in Table 1.

2.1.1. Volatile metabolites

Volatile organic compounds (VOCs) are of particular interest in metabolomic profiling of coffee since these species are responsible for the distinctive aroma [63,64]. The most useful technique for the sampling of these molecules is solid phase microextraction (SPME): a fiber is exposed to the volatile portion of the coffee sample and then the compounds are subsequently desorbed and injected into a gas chromatographer. This specific field has been coined volatome or volatolome [65–67] and it has been useful for the comparison of capsule-brewed espresso coffees in Italy [68], monitoring changes in the volatile compounds of Robusta coffee beans during drying [57], the distinction between decaffeinated and regular coffee [49] and the prediction of Arabica coffee quality through its volatile composition [63] just to name a few examples.

2.1.2. Non-volatile metabolites

Although GC is usually reserved for volatile and semi-volatiles compounds, previous derivatization of the less volatile products can be achieved, rendering them suitable for this technique. This was useful for the differentiation of green coffee samples that were cultivated at different altitudes and had different post-harvest processes [47]. Derivatization by oximation and trimethylsilylation allowed the GC/MS analysis of amino acids such as lysine and glycine, and sugars such as sorbose and fructose that were identified as markers for the post-harvesting process. Regarding altitude, inositol and serotonin were the metabolites of interest since they showed a positive and negative correlation respectively. Silylation was also the derivatization process of choice for the GC/MS analysis of

Table 1
Summary of GC analytical techniques and multivariate data analysis used for metabolomic profiling of coffee products.

Analytical techniques	Abbreviation	Sample	Metabolites	Multivariate Data analysis	Reference
Head Space-Solid Phase Microextraction - Two-dimensional Gas Chromatography fast Quadrupole Mass Spectrometry	HS-SPME - GCxGC-qMS	Roasted and ground coffee	Volatiles	PCA and CA	[61]
Pulsed Split - Gas Chromatography - Flame Ionization Detector - Mass Spectrometry	*PS*-GC/FID & *PS*-GC/MS	Green coffee beans oil extract	Non-volatiles	Not reported	[60]
Gas Chromatography - Flame Ionization Detector	GC/FID	Mucilage, pulp and endosperm of coffee cherries	Volatiles and non-volatiles	PCA	[58]
Gas Chromatography - Flame Ionization Detector	GC/FID	Ground green coffee beans	Non-volatiles	HCA	[59]
Gas Chromatography/Mass Spectrometry	GC/MS	Roasted and ground coffee	Non-volatiles	PCA	[46]
Head Space-Solid Phase Microextraction- Gas Chromatography/Mass spectrometry	HS-SPME-GC/MS	Coffee capsules	Volatiles	PCA and PLS-DA	[68]
Head Space-Solid Phase Microextraction- Gas Chromatography/Mass spectrometry	HS-SPME-GC/MS	Roasted and ground coffee	Volatiles	HCA and OPLS-DA	[55]
Head Space-Solid Phase Microextraction- Gas Chromatography/Mass spectrometry	HS-SPME-GC/MS	Ground green coffee beans, roasted and ground coffee	Volatiles and non-volatiles	Not reported	[64]
Head Space-Solid Phase Microextraction- Gas Chromatography - Time Of Flight - Mass spectrometry	HS-SPME-GC-TOF/MS	Ground green coffee beans	Volatiles	HCA	[57]
Gas Chromatography/Mass Spectrometry	GC/MS	Ground green coffee beans	Non-volatiles	PCA and OPLS	[47]
Head Space-Solid Phase Microextraction - Gas Chromatography/Mass spectrometry	HS-SPME-GC/MS	Roasted and ground coffee	Volatiles	PLS and GA-SVR	[63]
Gas Chromatography/Mass Spectrometry	GC/MS	Roasted and ground coffee	Volatiles and non-volatiles	PCA and HCA	[48]
Head Space-Solid Phase Microextraction- Gas Chromatography - Mass spectrometry	HS-SPME-GC/MS & E-nose	Roasted and ground coffee	Volatiles	PLS-DA	[57]
Head Space-Solid Phase Microextraction- Two Dimensional Gas Chromatography - Time Of Flight - Mass spectrometry	HS-SPME-GCxGC-TOF/MS	Coffee capsules	Volatiles	PCA and PLS-DA	[49]

non-volatile metabolites such as fatty acids, organic acids and sugars for the differentiation of twenty coffee samples according to their genotype, roasting degree and blending [48]. Oximation and trimethylsilylation were also helpful to analyze low molecular weight metabolites such as amino acids and sugars in Indonesian coffee, produced by different fermentation isolates to improve its quality [46].

2.2. Liquid chromatography

For non-volatile metabolites and especially those that might decompose in GC conditions, liquid chromatography (LC) is one of the best alternatives. It also provides a wider range of information due to the nature of the separation, especially related to the variants of the stationary phase [69]. It is usually coupled with mass detectors [43,70] but for targeted approaches, UV or diode array detectors (DAD) could be used too [71,72].

Some applications involved monitoring coffee roasting and its effect by determining the changes in the metabolomic profile of the samples. Potential biomarkers were identified as responsible for the roasting process by a hydrophilic interaction chromatography-mass spectrometry (HILIC/MS) based metabolomic approach [42]. LC coupled to high resolution mass spectrometry (LC-HRMS) proved to be suitable for the metabolomic fingerprinting of nine *Coffea* species leaves to potentially assign their botanical origin as well as their sampling period [73]. Furthermore, ultra performance liquid chromatography - quadrupole time of flight (UPLC-QTOF) was helpful in the assessment of three preparation techniques for coffee: boiling, pour over and cold-brew. Metabolomic profiling of each brewed product allowed the identification of potential markers for each method and showed more differences between those requiring higher temperatures and cold-brew [45]. In the same vein, coffee extracts obtained by traditional Italian methods were analyzed by UPLC-QTOF to acquire their metabolomic profile and identify markers characteristic of each extraction technique [44]. It was thus possible to determine similarities between them along with their differences related to caffeine content and vegetal aroma among other parameters, and even to infer that the extraction process was more important than the coffee species for the cup profile.

A summary of LC techniques and multivariate data analysis used for metabolomic profiling of coffee products is presented in Table 2.

Table 2
Summary of LC analytical techniques and multivariate data analysis used for metabolomic profiling of coffee products.

Analytical techniques	Abbreviation	Sample	Metabolites	Multivariate Data analysis	Reference
High Performance Liquid Chromatography	HPLC	Ground green coffee beans	Non-volatiles	Not reported	[8]
High Performance Liquid Chromatography	HPLC	Green coffee beans extracts	Non-volatiles	Not reported	[72]
Ultra performance liquid chromatography Mass Mass	UPLC-MS/MS	Mucilage, pulp and endosperm of coffee cherries	Volatiles and non-volatiles	PCA	[58]
High Performance Liquid Chromatography - Diode Array Detection	HPLC-DAD	Ground green coffee beans	Non-volatiles	PLS-DA	[71]
Reverse Phase - Liquid Chromatography - Mass Spectrometry	RP-LC-MS	Roasted and ground coffee	Non-volatiles	PCA and PLS-DA	[70]
Liquid Chromatography-High Resolution Mass Spectrometry Liquid Chromatography-Quadrupole - Time of Flight	LC-HRMS/LC-QTOF	Coffee leafs	Non-volatiles	PCA and PLS-DA	[73]
Ultra-High Pressure Liquid Chromatography - Quadrupole Time Of Flight - Mass Spectrometry	UHPLC-QTOF/MS	Coffee brews	Non-volatiles	PCA and HCA	[45]
Ultra performance Liquid Chromatography Mass/Mass	UPLC-MS/MS	Ground green coffee beans	Non-volatiles	HCA	[59]
High performance Liquid Chromatography - Mass/Mass	HPLC-MS/MS	Ground green coffee beans	Non-volatiles	HCA	[59]
High Performance Liquid Chromatography	HPLC	Roasted and ground coffee	Non-volatiles	Not reported	[82]
Hydrophilic Interaction Chromatography - Mass Spectrometrý	HILIC - MS	Roasted and ground coffee	Non-volatiles	PCA, HCA and PLS-DA	[42]
Ultra-High Pressure Liquid Chromatography - Quadrupole Time Of Flight - Mass Spectrometry	UHPLC-QTOF/MS	Roasted and ground coffee	Non-volatiles	HCA and OPLS-DA	[55]
Ultra Performance Liquid Chromatography - Triple Quadrupole - Mass Spectrometry	UPLC-QQQ/MS	Coffee brews	Non-volatiles	PCA and HCA	[45]
Rapid Resolution Liquid Chromatography - Electrospray Ionization - Quadrupole Time Of Flight - High Resolution Mass Spectrometer	RRLC-ESI-QTOF/HRMS	Leafs and fruits of coffee	Non-volatiles	PCA and PLS-DA	[43]
Ultra-High Pressure Liquid Chromatography - Quadrupole Time Of Flight - Mass Spectrometry	UHPLC-QTOF/MS	Coffee brews	Volatiles and non-volatiles	PCA, HCA and OPLS-DA	[44]

2.3. NMR

This technique has emerged as one of the most promising high-throughput tools due to the lack of sample preparation that avoids time-consuming purification processes that are common for these complex mixtures. It also has many advantages related to inherent characteristics of the technique such as low quantity of solvents needed, rapid analysis, robustness and reproducibility [74]. Due to its nature, NMR is more suitable for untargeted approaches and signal assignment for metabolites is achieved by comparison with databases and previously reported information [48,75]. It has been applied successfully for the authentication of roasted and ground coffee [76] as well as instant coffee according to manufacturer [77], to monitor the roasting process of Brazilian coffee beans [78], successfully differentiating organic from conventionally farmed coffee [79].

It has even been tested as an artificial tongue to predict sensations of roasted coffee bean products with good correlations [51].

Regarding targeted approaches, some studies have been reported for the quantification of metabolites such as caffeine and other organics acids in commercial coffee samples in Brazil [80] and the Middle East region [48].

Although most of these studies rely on proton spectra, whether it be mono or bidimensional correlations, especially for untargeted approaches, the potential for the analysis of other nuclei is still an open field [74].

A summary of NMR techniques and multivariate data analysis used for metabolomic profiling of coffee products is presented in Table 3.

2.4. Other techniques

Although chromatographic techinques and NMR are the most widely used techniques in this field, other less common ones have been successfully applied in metabolomics as shown on Table 4. The application of direct spectroscopic techniques such as UV, infrared or Raman spectroscopy for metabolomic approaches yield robust results in a shorter amount of time but they only can identify a few compounds per run contrary to the larger profiles obtained by chromatography for example. These are particularly appropriate for fingerprinting [41].

UV spectroscopy is a non-destructive, economical, and facile technique that has been used to identify selected coffee components as phenolic and chlorogenic acids, methylxanthines and antioxidants among others [41,52]. It has also been a fast and easy method to determine and quantify adulterations components in coffee [81] and has been a complementary analytical tool in other studies carried out for the discrimination of specialty and traditionally roasted commercial coffee [82] and for the evaluation of antioxidant activity related to phenolic compounds and melanoidin in Robusta and Arabica coffee extracts [83] to name a few.

Like UV, infrared spectroscopy is considered not very expensive, nondestructive and an easy analytical technique to evaluate coffee quality features such as sensory properties [14] and degree of roasting [84], to discriminate roasted coffee defectives by Fourier-transform infrared spectroscopy (FTIR) [85] or to determine other specific metabolites as chlorogenic acids in Arabica green coffee beans [52].

Regarding Raman spectroscopy, no sample preparation nor pre-treatment is needed, and it requires just a small amount of sample [86]. This technique has been used to discriminate between two species of coffee (Arabica and Robusta) based on their Kahweol content both in green and roasted coffee beans [54] and also based on the comparison of their chlorogenic acid and lipid content [53].

3. Chemometric tools

Due to the amount of data collected through targeted and untargeted studies, multivariate data analysis is required in order to

Table 3
Summary of NMR analytical techniques and multivariate data analysis used for metabolomic profiling of coffee products.

Analytical techniques	Abbreviation	Sample	Metabolites	Multivariate Data analysis	Reference
Proton Nuclear Magnetic Resonance	1H NMR	Instant coffee	Volatiles	PCA and LDA	[77]
Proton Nuclear Magnetic Resonance	1D and 2D NMR	Coffee brews	Non-volatiles	Not reported	[80]
Proton Nuclear Magnetic Resonance	1H NMR	Roasted and ground coffee	Non-volatiles	PCA and OPLS	[51]
Proton Nuclear Magnetic Resonance	1H NMR	Roasted and ground coffee	Non-volatiles	PCA	[76]
Proton Nuclear Magnetic Resonance	1H NMR	Ground green coffee beans	Non-volatiles	PCA	[50]
Proton Nuclear Magnetic Resonance	1H NMR	Roasted and ground coffee	Volatiles and non-volatiles	PCA, PLS-DA and OPLS-DA	[79]
Proton Nuclear Magnetic Resonance	1D and 2D NMR	Roasted and ground coffee, instant coffee	Non-volatiles	PCA and OPLS-DA	[75]
Proton Nuclear Magnetic Resonance	1D and 2D NMR	Roasted and ground coffee	Non-volatiles	PCA	[78]
Proton Nuclear Magnetic Resonance	1D and 2D NMR	Roasted and ground coffee	Volatiles and non-volatiles	PCA and HCA	[48]

Table 4
Summary of other analytical techniques and multivariate data analysis used for metabolomic profiling of coffee products.

Analytical techniques	Abbreviation	Sample	Metabolites	Multivariate Data analysis	Reference
Fourier Transform Raman Spectroscopy	FT-IR	Green and roasted coffees	Non-volatiles	PCA	[54]
Ultraviolet–Visible Spectroscopy	UV-VIS	Green coffee beans	Non-volatiles	Not reported	[8]
Raman spectroscopy	Raman	Green coffee beans	Non-volatiles	PCA	[53]
Near Infrared Spectroscopy	NIR	Coffee brews	Non-volatiles	PLS	[14]
Fourier Transform Infrared Spectroscopy	FT-IR	Roasted and ground coffee	Volatiles and non-volatiles	PCA	[84]
Ultraviolet–Visible Spectroscopy	UV-VIS	Green coffee beans	Non-volatiles	PCA and HCA	[52]
Ultraviolet–Visible Spectroscopy	UV-VIS	Roasted and ground coffee	Non-volatiles	PCA and PLS	[83]
Fourier Transform Infrared Spectroscopy	FT-IR	Roasted and ground coffee	Non-volatiles	PCA and PLS	[83]
Ultraviolet–Visible Spectroscopy	UV-VIS	Roasted and ground coffee	Not specified	PCA and PLS	[81]
Raman spectroscopy	Raman	Ground green coffee beans	Not specified	PCA and PLS-DA	[86]
High Performance Anion Exchange Chromatography - Pulsed Amperometric Detection	HPAEC - PAD	Ground green coffee beans	Non-volatiles	HCA	[59]
Ultraviolet–Visible Spectroscopy	UV-VIS	Roasted and ground coffee	Non-volatiles	Not reported	[82]
Fourier Transform Infrared Spectroscopy	FT-IR	Roasted and ground coffee	Non-volatiles	Not reported	[82]

retrieve patterns that can ultimately lead to predictive models. Fig. 3 shows the most common chemometric tools used for multivariate analysis of the metabolomic profiling of coffee-related samples. Unsupervised pattern recognition is usually the first step to ordinate the data and reduce the dimensionality to more manageable clusters. The most common unsupervised methods used in metabolomics are principal component analysis (PCA) [32,42,45,49,59,87–89] and hierarchical cluster analysis (HCA) [42,48,90,91].

PCA is usually the first step for exploratory data analysis to reduce the dimensionality [92] and explain the maximum variation between samples [93]: in this scenario, uncorrelated new variables arise from mathematical simplification of the originals, also known as principal components (PC). These PCs account for the total variance of the original variables, obtained by linear combination [92] and the first components bear most of the information.

In some cases, good cluster separation can be obtained without data set separation [45,49,70] but in others, a previous data set separation is required: this was especially true for the volatilome analysis of four typologies of coffee where the PCA was applied to compare encapsulated and non-encapsulated samples for each one [89].

On the other hand, HCA aims to classify samples hierarchically based on similarity or dissimilarity and it is represented by

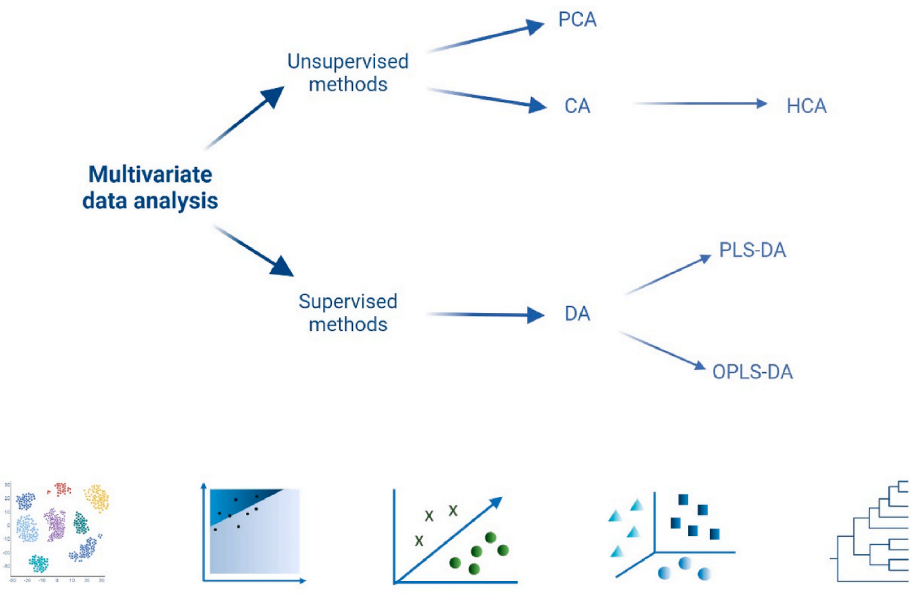


Fig. 3. Multivariate data analysis used in metabolomics. Created with BioRender.com.

Table 5
Major findings in coffee metabolomics studies for origin assessment.

Analytical techniques	Abbreviation	Sample	Multivariate Data analysis	Statistical Analysis Findings	Chemical Findings	Reference
Inductively Coupled Argon Plasma Atomic Emission Spectrometer	ICP AES	Roasted coffee beans	PCA and CDA	The statistical analysis differentiated samples from the three major geographic regions of coffee production in the world: Indonesia, East Africa, and Central/ South American.	It was found that Cu, Na, Mn, and Fe have some discriminating power with the geographic regions tested, but one cannot determine origin with these elements alone. Another important finding of the element concentration distribution is that no region is responsible for all of the high or low concentrations.	[99]
Liquid Chromatography - Mass Spectrometry, Gas Chromatography - Flame Ionization Detector	LC - MS, GC-FID	Roasted and ground coffee	PCA	Statistical analysis differentiated coffee samples from the regions studied: Asia, South America and Africa.	It was found that samples from Asia and South America were distinct in GC–FID outputs and amines. Samples from South America were different from the others in that outputs of overall GC–FID were very high and amines were very low. Samples from Africa were different from the others in protein (rather low) and monosaccharide (rather high) content.	[100]
Proton Nuclear Magnetic Resonance	1H NMR	Roasted and ground coffee	PCA and OPLS-DA	The statistical analysis differentiated the coffee samples of the regions studied: America, Africa and Asia.	It was found that American roasted coffee samples were characterized by fatty acids chains, the African samples by chlorogenic acids and lactate, while the Asian ones by acetate and trigonelline.	[101]
Unidimensional Proton and Carbon 13 Nuclear Magnetic Resonance	1D 1H and 13C NMR	Ground green coffee beans	PCA and OPLS-DA	The statistical analysis demonstrated that the greatest significance differentiation of coffee samples was related to the species: arabica and robusta.	It was found that the significantly different metabolites captured by PCA and OPLS-DA models were sucrose, citrate, malate, trigonelline, caffeine, choline, 5-CQA, 4-CQA, 3-CQA, acetic acid, L-Ala, L-Asn, L-Glu and γ-aminobutyric acid.	[102]
Direct Infusion Electrospray Ionization Fourier Transform Ion Cyclotron Resonance Mass Spectrometry	ESI FT-ICR MS	Ground green coffee beans	PCA and PLS-DA	The statistical analysis differentiated coffee samples from the regions studied: Londrina and Mandaguari regions of Brazil.	It was found that two compounds with m/z values of 695.20442 and 833.51922 were indicated by multivariate analyses as very important for discriminating the coffee cultivars. However, they could not be fully characterized. The most important compounds, responsible for discriminations in all PLS-DA analyses were: the unidentified $[M - H]$ of m/z 695.20442, atractyloside analogue II, sucrose, and the chlorogenic acids CQA and diCQA.	[107]
Gas Chromatography - Mass Spectrometry	GC - MS	Roasted and ground coffee	PCA and OPLS-DA	The statistical analysis differentiated commercial Kopi Luwak, commercial regular coffee, and counterfeit coffee.	It was found that the discriminant marker candidates were identified and quantitated against six authentic standards (malic acid, citric acid, glycolic acid, pyroglutamic acid, caffeine, and inositol) at various concentrations.	[109]
Gas Chromatography-Quadrupole - Mass Spectrometry	GC-Q - MS	Ground green coffee beans	PLS-DA	The statistical analysis differentiated between *Coffea arabica* L. genotypes: Mundo Novo and Bourbons, and betweeen coffee	It was found that the features that were most influential in differentiating genotype were: 5-CQA, oxalic acid, galactinol, nicotinic acid, caffeine, and caffeic acid (Bourbon) and myo-inositol, quinic acid, malic acid,	[108]

(*continued on next page*)

Table 5 (*continued*)

Analytical techniques	Abbreviation	Sample	Multivariate Data analysis	Statistical Analysis Findings	Chemical Findings	Reference
				producing municipalities in Brazil	fructose, and D-glucose (Mundo Novo).	
High Performance Liquid Chromatography -Diode Array Detector	HPLC-DAD	Green coffee beans	PCA-LDA	The statistical analysis demonstrated a moderate classification efficiency in differentiate the geographical origin of the coffee beans from the major production regions of Ethiopia using alkaloids.	It was found that using alkaloids, in conjunction with other chemical constituents such as phenolic compounds has potential, for the construction of classification models and generation of databases useful for the geographical origin discrimination of Ethiopian green coffee beans.	[110]
Ultra performance Liquid Chromatography Mass Spectrometry	UPLC-MS	Green coffee beans	PCA-LDA	The statistical analysis differentiated the geographical origin of the coffee beans from the major production regions and sub-regions of Ethiopia using chlorogenic acids.	They were identified 3 compounds (3-caffeoylquinic acid, 3,4-dicaffeoylquinic acid, 3,5-dicaffeoylquinic acid and 4,5-dicaffeoylquinic acid) as the most discriminating compounds for the authentication of the various regional and sub-regional green coffee beans of Ethiopia.	[40]
Inductively Coupled Plasma – Optical Emission Spectroscopy	ICP OES	Green coffee beans	PCA-LDA	The statistical analysis provided a reliable prediction model for the three major producing regions of coffee of Ethiopia using elemental analysis.	It was found that the elements P, Mn, S, Cu, and Fe were the most discriminating elements. Mg, P, S, Ca, Mn, Fe, Cu, Ba, Si, and K were determined as representative major and minor elements in coffee that are simple to determine, which is ideal for routine analysis. These elements have also been shown to be suitable for discriminating between Arabica and Robusta varieties, as well as for tracing the geographical origin of coffee beans.	[111]
Ultra-High Pressure Liquid Chromatography - Quadrupole Time Of Flight - High Resolution Mass Spectrometry	UHPLC-QTOF - HRMS	Ground green coffee beans	PCA and PLS-DA	The statistical analysis differentiated the green coffee samples from 5 geographical regions of Colombia.	They were identified 13 biomarkers (8 of them tentatively elucidated). The markers selected to create the discrimination model were 1-*o*-sinapoylglucose, 3-hydroxysuberic acid, *N*-acetyl-L-phenylalanine, 5-caffeoyl-metylquinic acid (5-ferulic acid trans), caffeoyl alcohol, 5-caffeoylquinic acid (5-CQAcis), 5-caffeoyl-methylquinic acid (5-ferulic acid cis), palmitic acid, and 5 more with feature names.	[112]
Gas Chromatography - Mass Spectrometry	GC-MS	Ground green and roasted coffee beans	PCA	The statistical analysis differentiated the coffee of western, central and eastern regions of Indonesia.	It was found that metabolites showing higher concentration in Sulawesi, Papua, Flores and Sumatra samples were glycerol, glucuno-1,5-lactone, gluconic acid and sorbitol. A clear distinction in galactitol and galactinol concentration between all samples from eastern part of Indonesia and western and middle part of Indonesia was also observed.	[113]
Gas Chromatography - Mass Spectrometry	GC-MS	Ground green coffee beans	PCA and LDA	The statistical analysis differentiated the coffee of northwest, west, east and south regions of Ethiopia, based on fatty acid composition.	It was found that oleic, linoleic, palmitic, stearic and arachidic acids were the most discriminating compounds among the production regions.	[114]

(*continued on next page*)

Table 5 (*continued*)

Analytical techniques	Abbreviation	Sample	Multivariate Data analysis	Statistical Analysis Findings	Chemical Findings	Reference
Ultra-High Pressure Liquid Chromatography - Quadrupole Exactive - Mass Spectrometry	UHPLC-QE-MS	Ground green coffee beans	PCA and OPLS-DA	The statistical analysis differentiated the coffee from 18 regions.	They were considered ten different families of compounds as potential markers of the coffee beans: 3- hydroxycoumarin, 4,5-di-*O*-caffeoylquinic, cryptochlorogenic acid, palmitic amide, linoleamide, arachidic acid, petroselinic acid, trehalose, L-glutamic acid, L-malic acid.	[103]
Ultra-High Pressure Liquid Chromatography - Electro Spray Ionization - High Resolution Mass Spectrometry	UHPLC-ESI -HRMS	Roasted and ground coffee	PCA and OPLS-DA	The statistical analysis differentiated the coffee of two species and five botanical varieties collected during the Brazilian International Conference of Coffee Tasters" (2022 Edition).	It was found that Caffeine, DIMBOA-Gl, roemerine, and cajanin were determined as chemical markers for robusta samples, and toralactone, cnidilide, LysoPC(18:2 (9Z,12Z)), Lysophosphatidylcholine (16:0/0:0), and 2,3-Dehydrosilybin for arabica samples.	[104]
Ultraviolet–Visible Spectroscopy, High Performance Liquid Chromatography - Diode Array Detection - Mass Mass, Gas Chromatography - Mass Spectrometry, Gas Chromatography - Flame Ionization Detector, Polymerase Chain Reaction - Restriction Fragment Length Polymorphism	UV/VIS, HPLC-DAD–MS/MS, GC–MS, GC-FID, PCR-RFLP	Ground green coffee beans	PCA	The statistical analysis differentiated the two species, arabica and robusta, according to their geographical origin.	It was found that Robusta accessions were confirmed to possess a higher antioxidant activity due to the high content of total phenolic compounds and caffeine when compared to Arabica.	[106]

dendrograms. Usually, HCA is preceded by PCA to reduce dimensionality but there are reported instances where the contrary is performed [48]. In that matter, HCA takes the scattergram obtained from PCA to build a dendrogram that allows the identification of clusters [90]. This has been useful for the evaluation of coffee samples that had different roasting degrees and were conclusive for differences in their metabolic profiles [42]. It was also carried out to assess heterogeneity regarding species, processing and additives present in commercially available coffee in the Middle East compared to four authenticated coffee samples [48]. As a final example, highly heritable metabolites were identified to be used as markers for the selection of Robusta seedlings that could potentially provide the best coffee cup quality [91].

Unsupervised methods as described above are usually followed by supervised methods where the groups are known a priori and the aim is to devise rules which can allocate previously unclassified objects or individuals into these groups in an optimal fashion [94].

The most common supervised methods used in metabolomics are partial least squares discriminant analysis (PLS-DA) and orthogonal partial least squares discriminant analysis (OPLS-DA) [95].

PLS-DA is an algorithm that combines dimensionality reduction with discriminant analysis in a flexible manner by not fitting the data into a specific distribution [95,96]. Dimensionality reduction is achieved by variable selection instead of the linear combination performed in PCA. Model validation methods allow the construction of a predictive model and can be grouped into three categories: (a) internal methods such as cross-validation [41,42,70], (b) external testing and (c) optional methods such as permutation tests [70,79]. Selection of either method depends on the size of the dataset [96]. PLS-DA has been used to discriminate samples depending on their roasted degree: low, medium and dark. Pairwise models were constructed whose aim was to obtain information about which metabolites were affected by the roasting process. The models were cross-validated and strengthened the robustness of the constructed models [42,70].

OPLS-DA is a special case of multilinear regression whose aim is to maximize both cross-covariance matrices and group discriminability. It consists of two independent processes: (a) discovery of hidden components for predictor variables and (b) data fitting through a projection matrix [97]. For this purpose, orthogonal variations are filtered from the variations of the data that are effective for the prediction of quantitative responses making it easier to interpret the models obtained [98]. In regard to regression prediction results, OPLS-DA is similar to PLS-DA modeling. The principal advantage of the former consists in the easier interpretation, specifically for multi-class cases [98]. It has also been used to identify markers of sensory quality in ground coffee [55], to differentiate organic from conventional coffee [79], to assess the impact of extraction methods in coffee brews [45] even evidencing that espresso preparations exhibit a distinctive chemical profile [44], and to evaluate coffee beans according to their cultivation altitudes, origins and post-harvest processing, being the latter the dominant factor influencing the final metabolite composition [47].

4. Application of metabolomics for origin assignment of coffee beans

Due to the effect edaphoclimatic conditions have on metabolite content in coffee, it was imperative to determine if it was possible to discriminate regions of origin using metabolomic profiling. As shown on Table 5, the research to differentiate coffee's geographic origins has evolved, from differentiating roughly between continents to discriminate between regions of the same country and most recently to identify potential markers for a specific coffee cultivar. Below are compiled, to the best of our knowledge, the studies that have focused on country and region assignment using metabolomics as a tool, as well as the attempts for authentication purposes where sensory analysis could fall short.

4.1. Studies focusing on discrimination by countries

The first metabolomic profile to differentiate geographic origins of coffee was reported in 2002 and focused on elements present in coffee samples from the three major producing regions: Indonesia, East-Africa and Central/South America. A total of 160 samples from these regions were analyzed by inductively coupled plasma atomic emission spectroscopy (ICP-AES) to quantify eighteen elements from which twelve are considered essential nutrients: K, Mg, Ca, Na, Cr, Mn, Fe, Cu, Zn, Mo, P and Co. The results showed that individually none of these elements could potentially be used as biomarkers to discriminate between regions. Still, it was highlighted that the highest concentrations of copper and sodium were found in Costarrican coffee; Colombian coffee had the highest zinc concentration; and Guatemalan coffee, the highest amounts of calcium and sulfur. Contrary, coffee from Panama had the lowest aluminum on average and Ethiopian coffee had the lowest concentrations of iron and magnesium. Coffee from Kenya had the highest concentrations of manganese and the lowest of potassium and sulfur. The same tendency was found in coffee from Sulawesi regarding magnesium (highest) and copper and calcium (lowest). Finally Sumatran coffee was the one that most differed from the others with the lowest zinc, manganese, potassium, and sodium content and the highest for iron and aluminum. The data was then submitted to multivariate data analysis for pattern recognition. Both PCA and canonical discriminant analysis (CDA) were performed to reduce variables and discriminant functional analysis and neural networks allowed building a classification model. The best visuals were obtained with CDA where the regions could be separated as shown in Fig. 4. The statistical analysis indicated that the three major regions had patterns and the model predicted the origin of the samples with a 70–86 % successful rate [99].

The first report on metabolomics for origin identification of coffee beans focused on small molecules was carried out in South Korea in 2010. On the basis that flavor and taste are influenced by the environment, the metabolite profiling of 21 samples from three distinctive regions (Asia, South America and Africa) were performed using LC/MS and GC/FID. The untargeted approach showed that coffees from the same region shared chromatographic patterns. For the targeted approach, carbohydrates and amines were quantified by LC/MS, monosaccharides were analyzed by GC/FID and proteins were determined with a protein-assay kit. All the data were combined in a PCA from which three PCs were selected to explain 64.83 % of the samples's diversity. PC1 corresponded to monosaccharide and early outcome of GC/FID quantification, PC2 to monosaccharide and later outcomes of GC/FID quantification and finally, PC3 to monosaccharide and high negative weights to GC/FID quantification. Fig. 5 shows the schematic 3D plot of the coffee samples along these PCs. It was concluded that Asian coffees are more affected by protein content, South American by volatiles and carbohydrates, and African by monosaccharides [100].

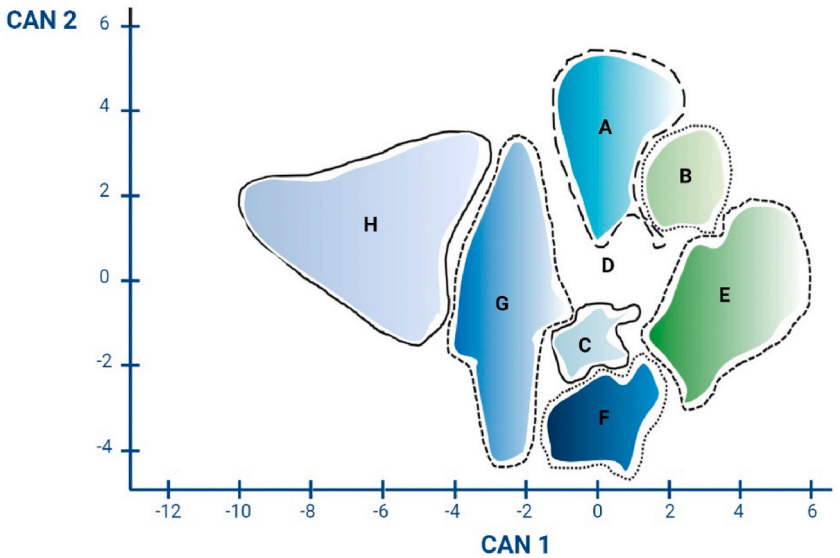


Fig. 4. Schematic representation of the patterns found by region in the plot of scores of first and second canonical functions for 160 coffee samples, adapted from Anderson (2002). Region A: Colombia, region B: Costa Rica, region C: Ethiopia, region D: Guatemala, region E: Kenya, region F: Panama, region E: Sulawesi and region H: Sumatra. Created with BioRender.com.

Later on, in 2012, NMR spectroscopy was employed for the first time to determine the metabolite profile of coffee beans and relate it to their geographic origin. Forty samples from Asia, Africa and America were ground and extracted at room temperature in buffered deuterated water (D_2O). The proton nuclear magnetic resonance (^1H NMR) spectra were recorded and then aligned for bucket integration on the formic signal at 8.424 ppm. PCA and OPLS-DA were carried out with the data collected. It was thus possible to determine that chlorogenic acids and lactate are responsible for African sample differentiation, acetate and trigonelline for Asian samples and fatty acids chains for those of American origin. This method has the advantage that it needs no or little derivatization and the experimental times are extremely fast compared to others [101].

A similar study was carried out in Japan in 2012, this time using green coffee instead of roasted coffee beans. Arabica coffee beans from Brazil, Colombia, Guatemala, and Tanzania along with Robusta coffee beans from Indonesia and Vietnam were supplied by a local vendor. Metabolites were extracted with D_2O at higher temperatures with less probability to lose compounds that give coffee its characteristic aroma. The supernatants were buffered and spiked with 4-dimethyl-4-silapentane-1-sulfonate as an internal standard. Both (1D) ^1H and ^{13}C NMR spectra were recorded at 500 and 125.65 MHz, respectively, and the latter was reduced into 1 ppm spectral buckets. The data were analyzed by PCA and OPLS-DA was applied giving a good separation of samples according to the geographical origin (Fig. 6). Although the most differences arose from the distinctive species, applying the model to the Arabica data set showed that Guatemalan coffee contains higher quantities of caffeine; Tanzanian coffee higher content of sucrose, acetic acid, and trigonelline; higher concentrations of caffeoylquinic acids (CQAs), citrate, and sucrose for Colombian coffee; and finally, higher levels of amino acids for Brazilian coffee [102].

Another report came from China where eigtheen samples of coffee beans from different countries in Asia, Africa, Oceania, North and South America were analyzed by an UPLC-Q-Exactive Orbitrap/MS method (UHPLC-QE-MS). Samples were extracted with a mixture of acetonitrile, methanol and water, vortexed, incubated at −40 °C (−233.15 K) and centrifuged. The supernatant was then taken into the UPLC instrument for an untargeted analysis. The data collected was processed with an in-house program and the multivariate analysis was performed with SIMCA 14.0 software. The samples could be separated in clusters by PCA although there was some sample overlapping. The supervised analysis OPLS-DA allowed the identification of sixteen potential biomarkers for the discrimination of the samples according to the continent of origin: 3-hydroxycoumarin, 4,5-Di-*O*-caffeoylquinic acid, linoleamide, palmitic amide, L-glutamic acid, D-aspartic acid and L-phenylalanine were more abundant in Asian coffees; cryptochlorogenic acid (4-*O*-caffeoylquinic acid) and arachidic acid, in Oceanian samples; alanine, in North American coffee; and the relative quantification of organic acid such as succinic acid and L-malic acid was higher both in North American and South American samples. Furthermore, comparison of Yunnan coffee to samples from other origins returned seventeen metabolites that could potentially be used as biomarkers for authentication purposes [103].

The metabolomic profiles of twenty one samples of three varieties of Arabica and two varieties of Robusta from Brazil and Mexico were evaluated by Ultra-High Pressure Liquid Chromatography - Electro Spray Ionization - High Resolution Mass Spectrometry (UHPLC-ESI-HRMS). Samples were extracted with a methanol-water (7:3) mixture, sonicated and centrifuged. The supernatant was extracted with heptane and the aqueous layer was filtered before injection into the instrument. A total of thirty three compounds were identified using different confidence levels. Multivariate statistical analysis on ProteoWizard software preceded by pre-processing the data by ionization mode, included PCA and OPLS-DA, the latter allowing to discriminate the samples by species and variety and the metabolomic fingerprinting showed good correlations to assess what the authors called the terroir effect [104], the interactions between the physical environment and coffee cultivars [105]. This approach is potentially useful for the added value of specialty coffees

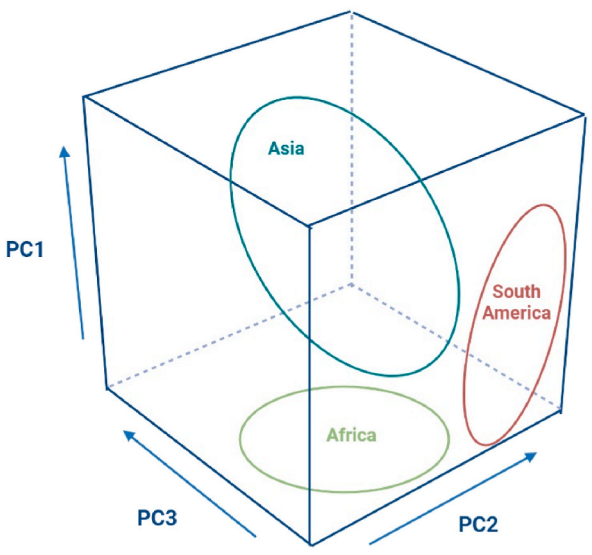


Fig. 5. Schematic representation of the 3D plot of the coffee samples, adapted from Choi (2010). Created with BioRender.com.

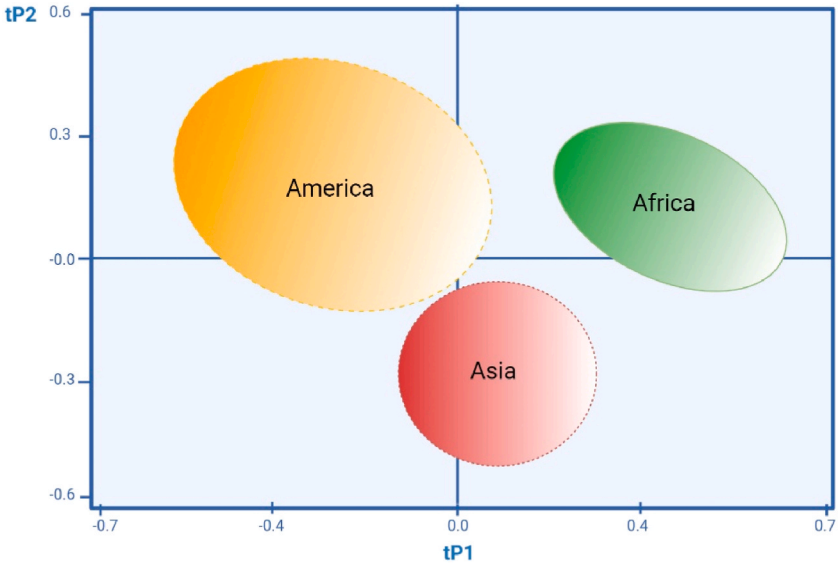


Fig. 6. Schematic representation of the patterns obtained by score and loading plots of OPLS DA performed by considering all roasted coffee samples, adapted from Wei (2012). Created with BioRender.com.

[104].

Most recently, five samples of green coffee Robusta and fifteen of Arabica from ten different countries underwent a battery of instrumental analysis (UV/Vis, HPLC-DAD–MS/MS, GC/MS, and GC-FID) as well as molecular (Polymerase Chain Reaction - Restriction Fragment Length Polymorphism, PCR-RFLP) fingerprinting in order to differentiate them by geographical origin using their flavonoid profile among other compound families. In total, thirty two compounds were identified, twenty eight flavonoids among them which were the main focus of this study due to the limited information available on them in the literature. Statistical análisis consisted of ANOVA followed by the Tukey–Kramer HSD test using using SPSS v. 28 software. Although PCA of the instrumental data revealed differences between varieties, discrimination by geographical origin was only achieved by the DNA-fingerprinting [106].

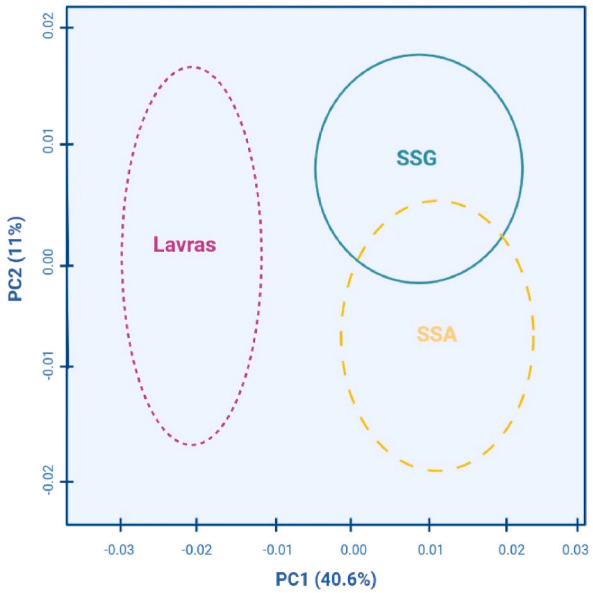


Fig. 7. Schematic representation of the patterns found from the PCA score plot of principal component 1 (40.6 % of the total variability) and principal component 2 (11 % of the total variability) of the metabolite profile differentiating green coffee harvested from three different origins: Lavras, Santo Antônio do Amparo (SAA) and São Sebastião da Grama (SSG), adapted from Da Silva (2014). Created with BioRender.com.(For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

4.2. Studies focusing on discrimination by regions within a country

As stated previously it is well known that coffees from different countries usually have different chemical composition but this is less evident for coffees that are grown in regions that are closer to each other.

Twenty three green coffee samples from eight different cultivars that had the same edaphoclimatic conditions in two different regions of Brazil were analyze with direct infusion electrospray ionization Fourier transform ion cyclotron resonance (ESI FT-ICR MS). Samples were ground, extracted with methanol in an ultrasonic bath and diluted with a mixture of methanol/water (1:1) prior to MS analysis. Metabolite identification was accomplished based on comparison of the *m/z* values and MS/MS data obtained by ESI FT-ICR MS in negative-ion mode with a homemade library and standards, identifying twenty metabolites. PCA and PLS-DA analysis was performed on twenty three samples with forty seven variables and showed that coffees from the Londrina region had higher levels of sucrose and feruloylquinic acid, whereas Mandaguari region provided higher levels of atractyloside analogue II. Apart from these compounds, others that were key for discrimination of these coffee samples were chlorogenic acids caffeoylquinic acid and di-caffeoylquinic acid [107].

In 2014, thirty six samples of Arabica L. genotypes (Mundo Novo and Bourbons) grown in the Brazilian municipalities of Lavras, Santo Antônio do Amparo (SAA), and São Sebastião da Grama (SSG) were subjected to metabolomic profiling in order to identify markers that could help differentiate them and assign their origin. Samples were extracted and derivatized to be analyzed by GC-Q/MS which provided forty four metabolites as potential biomarkers. The two PCs responsible for 51.6 % of the variations present in the samples were selected for the PCA analysis (Fig. 7). Both regions had similar profiles with SAA exhibiting higher levels of oxalic acid, malic acid, D-glucose, fructose, D-sorbitol, and galactinol. Conversely, Lavras showed higher levels of quinic acid, caffeine, and 5-caffeoylquinic acid (5-CQA), and SSG displayed higher levels of citric acid and glutamic acid. The genotypes were successfully differentiated. The metabolites that were used as markers were 5-CQA, oxalic acid, galactinol, nicotinic acid, caffeine, and caffeic acid for Bourbon and myo-inositol, quinic acid, malic acid, fructose, and D-glucose for Mundo Novo [108].

4.3. Studies with authentication purposes

Among coffee connoisseurs, the variety that really piques their interest is the Indonesian Kopi Luwak. Made from berries that have been eaten and excreted by the Asian palm civet, this coffee holds the title of being the most expensive coffee in the world. With such a standard to hold, it becomes evident how important it is to have methods to determine its authenticity and protect consumers from counterfeit products. As aforementione earlier, metabolomics techniques have risen as powerful tools for chemical profiling in different matrices which could be useful for authentication purposes and assigning geographical origin. In 2013, the first report on metabolite biomarkers that could potentially help differentiate civet coffee from other varieties was published. Samples consisted of Kopi Luwak, commercial Kopi Luwak, commercial regular coffee, fake coffee, and coffee blend. Multimarker profiling was achieved by gas chromatography quadrupole mass spectroscopy (GC-Q/MS) which provided discriminant markers for the differentiation of Kopi Luwak from other varieties and quantification of these metabolites was then performed by GC/MS. Multivariate analysis involved PCA to classify samples, followed by OPLS-DA and significance analysis of microarrays/metabolites (SAM) for the identification of discriminant marker candidates. Citric acid, malic acid and the inositol/pyroglutamic acid ratio were the selected markers. The model was tested with good results for differentiation against regular coffee and counterfeits and was acceptable for adulterated Kopi Luwak [109].

A targeted approach was examined in 2016 where the alkaloid content of ninety nine Ethiopian green coffee bean samples of eight varieties was determined. Alkaloids were extracted with boiling water and centrifuged. Precipitation of colloidal material such as polysaccharides was achieved with an aqueous lead acetate solution and the supernatants were filtered prior to their injection onto an High Pressure Liquid Chromatography - Diode Array Detector (HPLC-DAD) system. Concentrations were determined against calibration curves of each alkaloid (caffeine, theobromine and trigonelline) and linear discriminant analysis (LDA) of these results gave moderate correlations regarding the classification (75 %) and the prediction (74 %) abilities of the models. Nonetheless, alkaloid content still exhibits a good potential for further use as a discrimination parameter for geographical origin [110].

Simultaneously, phenolic compound content was also evaluated on 100 samples of Ethiopian green beans belonging to eight different varieties (Harar, Jimma, Kaffa, Wollega, Sidama, Yirgachefe, Benishangul and Finoteselam) for authentication purposes. Extraction was achieved by shaking with aqueous methanol, followed by centrifugation and retrieval of the supernatant. Polymeric components were removed with Carrez reagent I and II and the centrifuged and filtered supernatant was analyzed by ultra performance liquid chromatography - time of flight - mass spectrometry (UPLC-TOF-MS). The main phenolic compounds found in these samples were chlorogenic acids, eight of which were quantified with calibration curves. Using SIMCA13 (Umetrics, Sweden) and SPSS 20 (IBM Corp, USA), discriminant analysis was performed showing that 3-caffeoylquinic acid and 4,5-dicaffeoylquinic acid were suitable as markers for the determination of geographical origin of green coffee beans from Northwest and East growing regions of Ethiopia and a combination of other chlorogenic acids were useful for the South and West regions. The model had good prediction abilities for both regional and sub-regional level proving to be a convenient tool to identify fraudulent products [40].

Since elemental composition depends to some degree on environmental conditions, it could be assessed as a discriminant for geographic origin. Contrary to organic compounds which can undergo changes during post-harvest processing, elemental composition in coffee beans remains relatively stable. On this basis, forty nine coffee bean samples −Harar, Jimma, Kaffa, Wollega, Sidama, and Yirgachefe varieties− from different Ethiopian regions were digested in a mixture of H_2O_2 and HNO_3 using a microwave-assisted method. Inductively coupled plasma – optical emission spectroscopy (ICP-OES) analysis of the samples allowed the determination of nine elements (Mg, P, S, Ca, Mn, Fe, Cu, Ba and Si) and flame atomic emission spectrophotometer (FAES) was used for the

determination of potassium. Trends were visualized by PCA (performed on the statistic software package SPSS 20) suggesting that the best descriptors for these regions were the P, S, Mn, Si, Cu, and Fe content. PC1 and PC2 accounted for 57 % of the variance of the data. A LDA model was proposed which had good regional and sub-regional classification and prediction abilities, 98 % and 92 % of success rate, respectively, necessary for the authentication of some of the best coffee beans in Ethiopia [111].

Colombian coffee is one of the most renowned coffees in the world, it has been registered as "Protection of Geographical Indications granted by the European Commission", it is sold under the "Café de Colombia" trademark, thus it becomes evident how important is to ensure the origin. In 2018, forty one samples of green coffee harvested in experimental stations located in Caldas, Antioquia, Huila, Valle del Cauca and Santander were subjected to an untargeted metabolomic analysis in order to identify metabolite markers which could potentially identify the origin of Colombian coffee. Two types of extraction were performed to retrieve polar and semi-polar compounds and these extracts underwent three chromatographic separations under a HPLC-QTOF MS system. The data was carried into a PCA followed by PLS-DA to develop a model that could discriminate between groups analyzed. Thirteen markers were selected for the prediction model, from which eight were elucidated by accurate mass: 5-CQA, caffeic acid, ferulic acid, *p*-coumaric acid and feruloylquinic acid isomers. The models showed a 94 % classification accuracy which makes metabolomics a promising approach for origin discrimination [112].

Another example of applying metabolomics to the discrimination of geographical origin was published in 2019. As a coffee producer, Indonesia is known for its Arabica coffee, being the second largest exporter, and its specialty coffees. The demand for high quality products increases the interest in having better assays to ensure the origin of these sought after goods. Indonesian Specialty Arabica (Bali, Flores, Java, Papua, Sulawesi, Sumatra, Sumbawa) and Fine Robusta (Java Sulawesi, Sumatra) coffee samples were subjected to a non-targeted metabolome analysis in order to determine markers that could differentiate between major cultivation areas in 2019. Green and roasted coffee beans were extracted and derivatized prior to GC/MS analysis. Data collected was processed with MetAlign software which was later carried into a PCA. Even if the model allowed the differentiation of each coffee variety, the determination of their origin was possible only if each variety was analyzed independently. Markers for coffee originating from Sulawesi, Papua, Flores and Sumatra were glycerol, glucuno-1,5-lactone, gluconic acid and sorbitol. Other interesting markers were galactitol and galactinol which allowed differentiating samples from the eastern part of Indonesia and the western and middle part [113].

Fatty acids are another group of compounds that can be targeted to predict geographical origin. In this Ethiopian report from 2019, 100 samples of the four major producing areas, Northwest (varieties Benishangul and Finoteselam), West (varieties Jimma, Kaffa and Wollega), East (variety Harar) and South (varieties Sidama and Yirgachefe), were subjected to lipid extraction, followed by fatty acid derivatization to methyl ethers prior to GC/MS analysis. Eight fatty acids (myristic, palmitic, linoleic, stearic, oleic, arachidic, behenic and lignoceric acids) were quantified using calibration curves with the corresponding standards and three more were quantified by relation to the internal standard (hypogeic, margaric and gondoic acids). Distribution patterns were found with PCA for the fatty acid composition in these regional green coffee beans performing one-way ANOVA on SIMCA 13 (Umetrics, Sweden) and SPSS 20 (IBM Corp, USA). Finally LDA was applied which yielded a model with good recognition and prediction abilities (95 % and 92 %, respectively) [114].

Fig. 8 provides a summary of metabolomics origin assessment over time.

5. Conclusions

The introduction of metabolomics for the analysis of different types of produce has expanded the possibilities to have more accurate parameters to evaluate not only their quality but also their origin and authenticity. Coffee being a high-demanded product that is consumed all over the world, would definitely benefit from having assays that could ensure these aspects that ultimately determine its price in the market. As its been demonstrated, geographical origin can be determined not only by differentiation of continents, but even by regions within the same country which according to edaphoclimatic differences, might vary in their components and give different notes on the final brew. Different techniques are available for this purpose: among the most relevant are chromatography -either gas or liquid-coupled with mass detection, NMR, ICP and HRMS. All of them give different information but overall, they provide the necessary data to design models that can help assign the origin of unknown samples.

Funding

We thank Unversidad Mariano Gálvez de Guatemala, Guatemala, to support the authors in the preparation of this review. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethics declarations

Review and approval by an ethics committee was not needed for this study because the research reviewed does not involve live vertebrates and higher invertebrates.

Informed consent was not required for this study because the research reviewed does not involve live vertebrates and higher invertebrates.

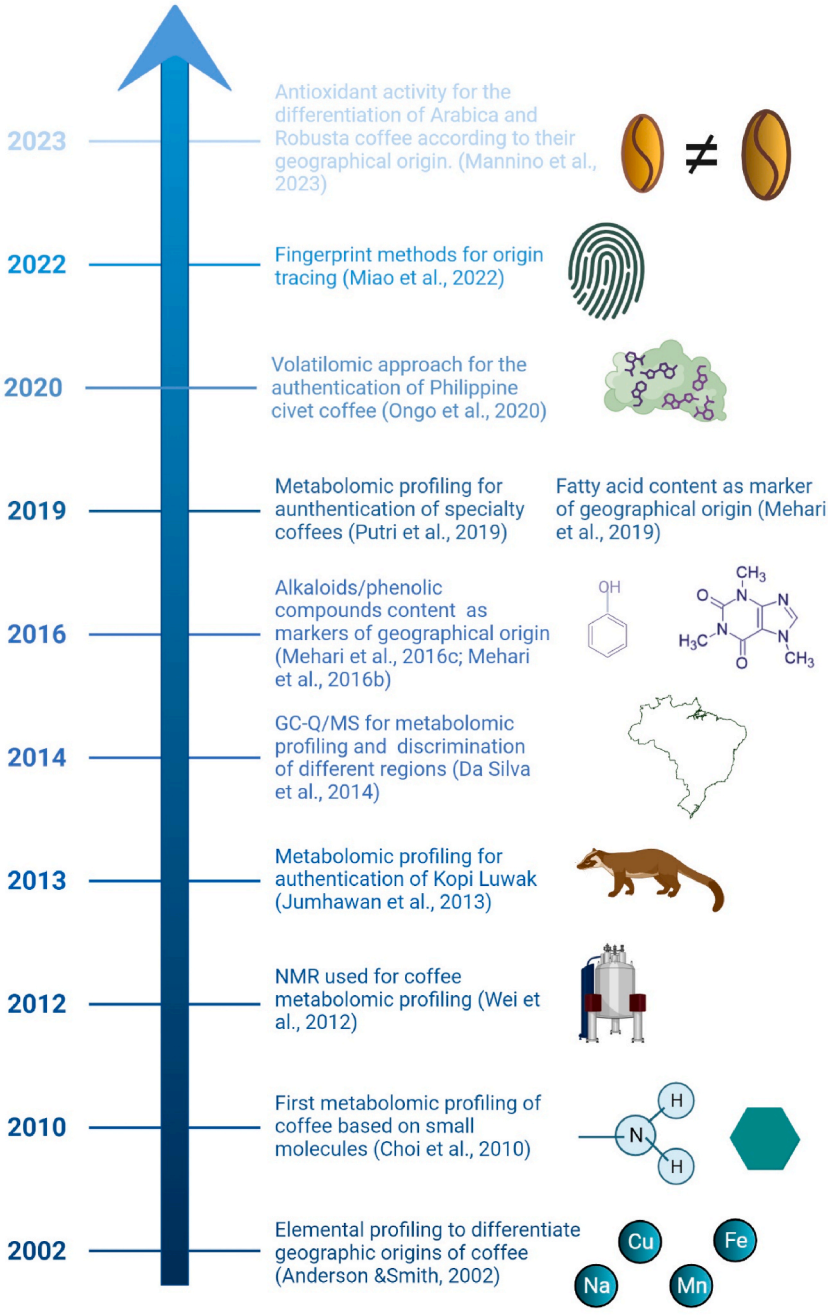


Fig. 8. Summary of metabolomics for origin assessment over time.

Data availability

No data was used for the research described in this review.

CRediT authorship contribution statement

Claudia de León-Solis: Conceptualization, Investigation, Supervision, Writing – original draft, Writing – review & editing. **Victoria Casasola:** Investigation, Visualization, Writing – original draft, Writing – review & editing. **Tania Monterroso:** Investigation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Elizabeth Solórzano (Instituto de Investigaciones Químicas, Biológicas, Biomédicas y Biofísicas, Universidad Mariano Gálvez) for her insight and discussion regarding topics of multivariate data analysis and the review in general.

References

[1] A.S. Bosselmann, K. Dons, T. Oberthur, C.S. Olsen, A. Ræbild, H. Usma, The influence of shade trees on coffee quality in small holder coffee agroforestry systems in Southern Colombia, Agric. Ecosyst. Environ. 129 (2009) 253–260, https://doi.org/10.1016/j.agee.2008.09.004.
[2] A. Hameed, S.A. Hussain, H.A.R. Suleria, Coffee Bean-Related" Agroecological Factors Affecting the Coffee (2018) 1–67, https://doi.org/10.1007/978-3-319-76887-8_21-1.
[3] J. Avelino, B. Barboza, J.C. Araya, C. Fonseca, F. Davrieux, B. Guyot, C. Cilas, Effects of slope exposure, altitude and yield on coffee quality in two altitude terroirs of Costa Rica, Orosi and Santa María de Dota, J. Sci. Food Agric. 85 (2005) 1869–1876, https://doi.org/10.1002/jsfa.2188.
[4] B. Bertrand, P. Vaast, E. Alpizar, H. Etienne, F. Davrieux, P. Charmetant, Comparison of bean biochemical composition and beverage quality of Arabica hybrids involving Sudanese-Ethiopian origins with traditional varieties at various elevations in Central America, Tree Physiol. 26 (2006) 1239–1248, https://doi.org/10.1093/treephys/26.9.1239.
[5] T. Oberthür, P. Läderach, H. Posada, M.J. Fisher, L.F. Samper, J. Illera, L. Collet, E. Moreno, R. Alarcón, A. Villegas, H. Usma, C. Perez, A. Jarvis, Regional relationships between inherent coffee quality and growing environment for denomination of origin labels in Nariño and Cauca, Colombia, Food Pol. 36 (2011) 783–794, https://doi.org/10.1016/j.foodpol.2011.07.005.
[6] B. Bertrand, R. Boulanger, S. Dussert, F. Ribeyre, L. Berthiot, F. Descroix, T. Joët, Climatic factors directly impact the volatile organic compound fingerprint in green Arabica coffee bean as well as coffee beverage quality, Food Chem. 135 (2012) 2575–2583, https://doi.org/10.1016/j.foodchem.2012.06.060.
[7] W.B. Sunarharum, D.J. Williams, H.E. Smyth, Complexity of coffee flavor: a compositional and sensory perspective, Food Res. Int. 62 (2014) 315–325, https://doi.org/10.1016/j.foodres.2014.02.030.
[8] K. Ramalakshmi, I.R. Kubra, L.J.M. Rao, Physicochemical characteristics of green coffee: comparison of graded and defective beans, J. Food Sci. 72 (2007) 333–337, https://doi.org/10.1111/j.1750-3841.2007.00379.x.
[9] S.E. Yeager, M.E. Batali, J.X. Guinard, W.D. Ristenpart, Acids in coffee: a review of sensory measurements and meta-analysis of chemical composition, Crit. Rev. Food Sci. Nutr. (2021) 1–27, https://doi.org/10.1080/10408398.2021.1957767.
[10] Y. Yulianti, D.R. Adawiyah, D. Herawati, D. Indrasti, N. Andarwulan, Detection of markers in green beans and roasted beans of Kalosi-enrekang arabica coffee with different postharvest processing using LC-MS/MS, Int J Food Sci 2023 (2023) 1–12, https://doi.org/10.1155/2023/6696808.
[11] N. Bhumiratana, K. Adhikari, E. Chambers, Evolution of sensory aroma attributes from coffee beans to brewed coffee, LWT - Food Sci. Technol. (Lebensmittel-Wissenschaft -Technol.) 44 (2011) 2185–2192, https://doi.org/10.1016/j.lwt.2011.07.001.
[12] G. Caprioli, M. Cortese, G. Sagratini, S. Vittori, The influence of different types of preparation (espresso and brew) on coffee aroma and main bioactive constituents, Int. J. Food Sci. Nutr. 66 (2015) 505–513, https://doi.org/10.3109/09637486.2015.1064871.
[13] A.N. Gloess, B. Schönbächler, B. Klopprogge, L. D'Ambrosio, K. Chatelain, A. Bongartz, A. Strittmatter, M. Rast, C. Yeretzian, Comparison of nine common coffee extraction methods: instrumental and sensory analysis, Eur. Food Res. Technol. 236 (2013) 607–627, https://doi.org/10.1007/s00217-013-1917-x.
[14] J.S. Ribeiro, M.M.C. Ferreira, T.J.G. Salva, Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy, Talanta 83 (2011) 1352–1358, https://doi.org/10.1016/j.talanta.2010.11.001.
[15] M. Creydt, M. Fischer, Omics approaches for food authentication, Electrophoresis 39 (2018) 1569–1581, https://doi.org/10.1002/elps.201800004.
[16] M.U. Markos, Y. Tola, B.T. Kebede, O. Ogah, Metabolomics: a suitable foodomics approach to the geographical origin traceability of Ethiopian Arabica specialty coffees, Food Sci. Nutr. (2023) 1–13, https://doi.org/10.1002/fsn3.3434.
[17] Putri Sastia Prama, Fukusaki Eiichiro, Mass Spectrometry-Based Metabolomics A Practical Guide, CRC Press, Boca Ratón, 2015.
[18] A. Zalacain, S.A. Ordoudi, E.M. Díaz-Plaza, M. Carmona, I. Blázquez, M.Z. Tsimidou, G.L. Alonso, Near-infrared spectroscopy in saffron quality control: determination of chemical composition and geographical origin, J. Agric. Food Chem. 53 (2005) 9337–9341, https://doi.org/10.1021/jf050846s.
[19] A. Massaro, A. Negro, M. Bragolusi, B. Miano, A. Tata, M. Suman, R. Piro, Oregano authentication by mid-level data fusion of chemical fingerprint signatures acquired by ambient mass spectrometry, Food Control 126 (2021), https://doi.org/10.1016/j.foodcont.2021.108058.
[20] R. Díaz, O.J. Pozo, J.V. Sancho, F. Hernández, Metabolomic approaches for orange origin discrimination by ultra-high performance liquid chromatography coupled to quadrupole time-of-flight mass spectrometry, Food Chem. 157 (2014) 84–93, https://doi.org/10.1016/j.foodchem.2014.02.009.
[21] S. Klockmann, E. Reiner, R. Bachmann, T. Hackl, M. Fischer, Food fingerprinting: metabolomic approaches for geographical origin discrimination of hazelnuts (corylus avellana) by UPLC-QTOF-MS, J. Agric. Food Chem. 64 (2016) 9253–9262, https://doi.org/10.1021/acs.jafc.6b04433.
[22] H. Lau, A.K.C. Laserna, S.F.Y. Li, 1H NMR-based metabolomics for the discrimination of celery (Apium graveolens L. var. dulce) from different geographical origins, Food Chem. 332 (2020), https://doi.org/10.1016/j.foodchem.2020.127424.
[23] M. Kharbach, J. Viaene, H. Yu, R. Kamal, I. Marmouzi, A. Bouklouze, Y. Vander Heyden, Secondary-metabolites fingerprinting of Argania spinosa kernels using liquid chromatography–mass spectrometry and chemometrics, for metabolite identification and quantification as well as for geographic classification, J. Chromatogr. A 1670 (2022), https://doi.org/10.1016/j.chroma.2022.462972.
[24] K. Hori, T. Kiriyama, K. Tsumura, A liquid chromatography time-of-flight mass spectrometry-based metabolomics approach for the discrimination of cocoa beans from different growing regions, Food Anal. Methods 9 (2016) 738–743, https://doi.org/10.1007/s12161-015-0245-0.
[25] J.E. Lee, B.J. Lee, J.O. Chung, H.N. Kim, E.H. Kim, S. Jung, H. Lee, S.J. Lee, Y.S. Hong, Metabolomic unveiling of a diverse range of green tea (Camellia sinensis) metabolites dependent on geography, Food Chem. 174 (2015) 452–459, https://doi.org/10.1016/j.foodchem.2014.11.086.
[26] H. Zhao, B. Guo, Y. Wei, B. Zhang, Near infrared reflectance spectroscopy for determination of the geographical origin of wheat, Food Chem. 138 (2013) 1902–1907, https://doi.org/10.1016/j.foodchem.2012.11.037.
[27] R. Ch, O. Chevallier, P. McCarron, T.F. McGrath, D. Wu, L. Nguyen Doan Duy, A.P. Kapil, M. McBride, C.T. Elliott, Metabolomic fingerprinting of volatile organic compounds for the geographical discrimination of rice samples from China, Vietnam and India, Food Chem. 334 (2021), https://doi.org/10.1016/j.foodchem.2020.127553.
[28] D. Schütz, E. Achten, M. Creydt, J. Riedl, M. Fischer, Non-targeted LC-MS metabolomics approach towards an authentication of the geographical origin of grain maize (Zea mays L.) samples, Foods 10 (2021) 2160, https://doi.org/10.3390/foods10092160.
[29] M.T. Osorio, A.P. Moloney, O. Schmidt, F.J. Monahan, Multielement isotope analysis of bovine muscle for determination of international geographical origin of meat, J. Agric. Food Chem. 59 (2011) 3285–3294, https://doi.org/10.1021/jf1040433.
[30] N.S. Chatterjee, O.P. Chevallier, E. Wielogorska, C. Black, C.T. Elliott, Simultaneous authentication of species identity and geographical origin of shrimps: untargeted metabolomics to recurrent biomarker ions, J. Chromatogr. A 1599 (2019) 75–84, https://doi.org/10.1016/j.chroma.2019.04.001.
[31] Y. Lu, G. Yao, X. Wang, Y. Zhang, J. Zhao, Y.J. Yu, H. Wang, Chemometric discrimination of the geographical origin of licorice in China by untargeted metabolomics, Food Chem. 380 (2022), https://doi.org/10.1016/j.foodchem.2022.132235.

[32] M.A. Farag, D.M. El-Kersh, A. Ehrlich, M.A. Choucry, H. El-Seedi, A. Frolov, L.A. Wessjohann, Variation in Ceratonia siliqua pod metabolome in context of its different geographical origin, ripening stage and roasting process, Food Chem. 283 (2019) 675–687, https://doi.org/10.1016/j.foodchem.2018.12.118.

[33] S. Ghisoni, L. Lucini, F. Angilletta, G. Rocchetti, D. Farinelli, S. Tombesi, M. Trevisan, Discrimination of extra-virgin-olive oils from different cultivars and geographical origins by untargeted metabolomics, Food Res. Int. 121 (2019) 746–753, https://doi.org/10.1016/j.foodres.2018.12.052.

[34] A. Da Ros, D. Masuero, S. Riccadonna, K.B. Bubola, N. Mulinacci, F. Mattivi, I. Lukić, U. Vrhovsek, Complementary untargeted and targeted metabolomics for differentiation of extra virgin olive oils of different origin of purchase based on volatile and phenolic composition and sensory quality, Molecules 24 (2019) 2896, https://doi.org/10.3390/molecules24162896.

[35] E. López-Rituerto, F. Savorani, A. Avenoza, J.H. Busto, J.M. Peregrina, S.B. Engelsen, Investigations of la Rioja terroir for wine production using 1H NMR metabolomics, J. Agric. Food Chem. 60 (2012) 3452–3461, https://doi.org/10.1021/jf204361d.

[36] G.P. Danezis, A.C. Pappas, E. Tsiplakou, E.C. Pappa, M. Zacharioudaki, A.S. Tsagkaris, C.A. Papachristidis, K. Sotirakoglou, G. Zervas, C.A. Georgiou, Authentication of Greek protected designation of origin cheeses through elemental metabolomics, Int. Dairy J. 104 (2020), https://doi.org/10.1016/j.idairyj.2019.104599.

[37] G. Rocchetti, L. Lucini, A. Gallo, F. Masoero, M. Trevisan, G. Giuberti, Untargeted metabolomics reveals differences in chemical fingerprints between PDO and non-PDO Grana Padano cheeses, Food Res. Int. 113 (2018) 407–413, https://doi.org/10.1016/j.foodres.2018.07.029.

[38] M.M. Artêncio, A.L.L. Cassago, J. de M.E. Giraldi, S.I.D. Pádua, F.B. Da Costa, One step further: application of metabolomics techniques on the geographical indication (GI) registration process, Bus. Process Manag. J. 28 (2022) 1093–1116, https://doi.org/10.1108/BPMJ-12-2021-0794.

[39] F.S. Aurum, T. Imaizumi, T. Manasikan, D. Praseptiangga, K. Nakano, Coffee origin determination based on analytical and nondestructive approaches -A systematic literature review, Reviews in Agricultural Science 10 (2022) 257–287, https://doi.org/10.7831/ras.10.0_257.

[40] B. Mehari, M. Redi-Abshiro, B.S. Chandravanshi, S. Combrinck, M. Atlabachew, R. McCrindle, Profiling of phenolic compounds using UPLC-MS for determining the geographical origin of green coffee beans from Ethiopia, J. Food Compos. Anal. 45 (2016) 16–25, https://doi.org/10.1016/j.jfca.2015.09.006.

[41] M.A. Farag, A. Zayed, I.E. Sallam, A. Abdelwareth, L.A. Wessjohann, Metabolomics-based approach for coffee beverage improvement in the context of processing, brewing methods, and quality attributes, Foods 11 (2022), https://doi.org/10.3390/foods11060864.

[42] R. Pérez-Míguez, M. Castro-Puyana, E. Sánchez-López, M. Plaza, M.L. Marina, Untargeted HILIC-MS-based metabolomics approach to evaluate coffee roasting process: contributing to an integrated metabolomics multiplatform, Molecules 25 (2020) 887, https://doi.org/10.3390/molecules25040887.

[43] A. Montis, F. Souard, C. Delporte, P. Stoffelen, C. Stévigny, P. Van Antwerpen, Targeted and untargeted mass spectrometry-based metabolomics for chemical profiling of three coffee species, Molecules 27 (2022) 3152, https://doi.org/10.3390/molecules27103152.

[44] F. Vezzulli, G. Rocchetti, M. Lambri, L. Lucini, Metabolomics combined with sensory analysis reveals the impact of different extraction methods on coffee beverages from coffea arabica and coffea canephora var, Robusta, Foods 11 (2022), https://doi.org/10.3390/foods11060807.

[45] L. Xu, F. Lao, Z. Xu, X. Wang, F. Chen, X. Liao, A. Chen, S. Yang, Use of liquid chromatography quadrupole time-of-flight mass spectrometry and metabolomic approach to discriminate coffee brewed by different methods, Food Chem. 286 (2019) 106–112, https://doi.org/10.1016/j.foodchem.2019.01.154.

[46] P. Aditiawati, D.I. Astuti, J.A. Kriswantoro, S.M. Khanza, Kamarisima, T. Irifune, F. Amalia, E. Fukusaki, S.P. Putri, GC/MS-based metabolic profiling for the evaluation of solid state fermentation to improve quality of Arabica coffee beans, Metabolomics 16 (2020) 57, https://doi.org/10.1007/s11306-020-01678-y.

[47] F. Amalia, P. Aditiawati, Yusianto, S.P. Putri, E. Fukusaki, Gas chromatography/mass spectrometry-based metabolite profiling of coffee beans obtained from different altitudes and origins with various postharvest processing, Metabolomics 17 (2021) 69, https://doi.org/10.1007/s11306-021-01817-z.

[48] A. Zayed, A. Abdelwareth, T.A. Mohamed, H.A. Fahmy, A. Porzel, L.A. Wessjohann, M.A. Farag, Dissecting coffee seeds metabolome in context of genotype, roasting degree, and blending in the Middle East using NMR and GC/MS techniques, Food Chem. 373 (2022), https://doi.org/10.1016/j.foodchem.2021.131452.

[49] Y. Zou, M. Gaida, F.A. Franchina, P.H. Stefanuto, J.F. Focant, Distinguishing between decaffeinated and regular coffee by HS-SPME-GC×GC-TOFMS, chemometrics, and machine learning, Molecules 27 (2022) 1806, https://doi.org/10.3390/molecules27061806.

[50] P.O. Sandusky, Introducing undergraduate students to metabolomics using a NMR-based analysis of coffee beans, J. Chem. Educ. 94 (2017) 1324–1328, https://doi.org/10.1021/acs.jchemed.6b00559.

[51] F. Wei, K. Furihata, T. Miyakawa, M. Tanokura, A pilot study of NMR-based sensory prediction of roasted coffee bean extracts, Food Chem. 152 (2014) 363–369, https://doi.org/10.1016/j.foodchem.2013.11.161.

[52] A.E. Terrile, G.G. Marcheafave, G.S. Oliveira, M. Rakocevic, R.E. Brunsd, I.S. Scarminio, Chemometric analysis of UV Characteristic profile and infrared fingerprint variations of coffea arabica green beans under different space management treatments, J. Braz. Chem. Soc. 27 (2016) 1254–1263, https://doi.org/10.5935/0103-5053.20160022.

[53] R.M. El-Abassy, P. Donfack, A. Materny, Discrimination between Arabica and Robusta green coffee using visible micro Raman spectroscopy and chemometric analysis, Food Chem. 126 (2011) 1443–1448, https://doi.org/10.1016/j.foodchem.2010.11.132.

[54] A.B. Rubayiza, M. Meurens, Chemical discrimination of arabica and robusta coffees by fourier transform Raman spectroscopy, J. Agric. Food Chem. 53 (2005) 4654–4659, https://doi.org/10.1021/jf0478657.

[55] G. Rocchetti, G.P. Braceschi, L. Odello, T. Bertuzzi, M. Trevisan, L. Lucini, Identification of markers of sensory quality in ground coffee: an untargeted metabolomics approach, Metabolomics 16 (2020), https://doi.org/10.1007/s11306-020-01751-6.

[56] D.J. Beale, F.R. Pinu, K.A. Kouremenos, M.M. Poojary, V.K. Narayana, B.A. Boughton, K. Kanojia, S. Dayalan, O.A.H. Jones, D.A. Dias, Review of recent developments in GC–MS approaches to metabolomics-based research, Metabolomics 14 (2018), https://doi.org/10.1007/s11306-018-1449-2.

[57] K. Zhang, J. Cheng, Q. Hong, W. Dong, X. Chen, G. Wu, Z. Zhang, Identification of changes in the volatile compounds of robusta coffee beans during drying based on HS-SPME/GC-MS and E-nose analyses with the aid of chemometrics, LWT 161 (2022), https://doi.org/10.1016/j.lwt.2022.113317.

[58] F. De Bruyn, S.J. Zhang, V. Pothakos, J. Torres, C. Lambot, A.V. Moroni, M. Callanan, W. Sybesma, S. Weckx, L. De Vuyst, Exploring the impacts of postharvest processing on the microbiota and metabolite profiles during green coffee bean production, Appl. Environ. Microbiol. 83 (2017), https://doi.org/10.1128/AEM.02398-16.

[59] S.J. Zhang, F. De Bruyn, V. Pothakos, J. Torres, C. Falconi, C. Moccand, S. Weckx, L. De Vuyst, Following coffee production from cherries to cup: microbiological and metabolomic analysis of wet processing of Coffea arabica, Appl. Environ. Microbiol. 85 (2019), https://doi.org/10.1128/AEM.02635-18.

[60] F.J.M. Novaes, S.S. Oigman, R.O.M.A. De Souza, C.M. Rezende, F.R. De Aquino Neto, New approaches on the analyses of thermolabile coffee diterpenes by gas chromatography and its relationship with cup quality, Talanta 139 (2015) 159–166, https://doi.org/10.1016/j.talanta.2014.12.025.

[61] C. Cordero, E. Liberto, C. Bicchi, P. Rubiolo, S.E. Reichenbach, X. Tian, Q. Tao, Targeted and non-targeted approaches for complex natural sample profiling by GC×GC-qMS, J. Chromatogr. Sci. 48 (2010) 251–261, https://doi.org/10.1093/chromsci/48.4.251.

[62] X. Duportet, R.B.M. Aggio, S. Carneiro, S.G. Villas-Bôas, The biological interpretation of metabolomic data can be misled by the extraction method used, Metabolomics 8 (2012) 410–421, https://doi.org/10.1007/s11306-011-0324-1.

[63] B.Z. Agnoletti, G.S. Folli, L.L. Pereira, P.F. Pinheiro, R.C. Guarçoni, E.C. da Silva Oliveira, P.R. Filgueiras, Multivariate calibration applied to study of volatile predictors of arabica coffee quality, Food Chem. 367 (2022), https://doi.org/10.1016/j.foodchem.2021.130679.

[64] C. Wang, J. Sun, B. Lassabliere, B. Yu, S.Q. Liu, Coffee flavour modification through controlled fermentations of green coffee beans by Saccharomyces cerevisiae and Pichia kluyveri: Part I. Effects from individual yeasts, Food Res. Int. 136 (2020), https://doi.org/10.1016/j.foodres.2020.109588.

[65] Y.Y. Broza, P. Mochalski, V. Ruzsanyi, A. Amann, H. Haick, Hybride Volatolomik und der Nachweis von Krankheiten, Angew. Chem. 127 (2015) 11188–11201, https://doi.org/10.1002/ange.201500153.

[66] M.K. Das, S.C. Bishwal, A. Das, D. Dabral, A. Varshney, V.K. Badireddy, R. Nanda, Investigation of gender-specific exhaled breath volatome in humans by GCxGC-TOF-MS, Anal. Chem. 86 (2014) 1229–1237, https://doi.org/10.1021/ac403541a.

[67] M. Phillips, R.N. Cataneo, A. Chaturvedi, P.D. Kaplan, M. Libardoni, M. Mundada, U. Patel, X. Zhang, Detection of an extended human volatome with comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry, PLoS One 8 (2013), https://doi.org/10.1371/journal.pone.0075274.

[68] V. Lolli, A. Acharjee, D. Angelino, M. Tassotti, D. Del Rio, P. Mena, A. Caligiani, Chemical characterization of capsule-brewed espresso coffee aroma from the most widespread Italian brands by HS-SPME/GC-MS, Molecules 25 (2020), https://doi.org/10.3390/molecules25051166.

[69] S. Forcisi, F. Moritz, B. Kanawati, D. Tziotis, R. Lehmann, P. Schmitt-Kopplin, Liquid chromatography-mass spectrometry in metabolomics research: mass analyzers in ultra high pressure liquid chromatography coupling, J. Chromatogr. A 1292 (2013) 51–65, https://doi.org/10.1016/j.chroma.2013.04.017.

[70] R. Pérez-Míguez, E. Sánchez-López, M. Plaza, M. Castro-Puyana, M.L. Marina, A non-targeted metabolomic approach based on reversed-phase liquid chromatography–mass spectrometry to evaluate coffee roasting process, Anal. Bioanal. Chem. 410 (2018) 7859–7870, https://doi.org/10.1007/s00216-018-1405-z.

[71] S. De Luca, E. Ciotoli, A. Biancolillo, R. Bucci, A.D. Magrì, F. Marini, Simultaneous quantification of caffeine and chlorogenic acid in coffee green beans and varietal classification of the samples by HPLC-DAD coupled with chemometrics, Environ. Sci. Pollut. Control Ser. 25 (2018) 28748–28759, https://doi.org/10.1007/s11356-018-1379-6.

[72] A.P. Craig, C. Fields, N. Liang, D. Kitts, A. Erickson, Performance review of a fast HPLC-UV method for the quantification of chlorogenic acids in green coffee bean extracts, Talanta 154 (2016) 481–485, https://doi.org/10.1016/j.talanta.2016.03.101.

[73] F. Souard, C. Delporte, P. Stoffelen, E.A. Thévenot, N. Noret, B. Dauvergne, J.M. Kauffmann, P. Van Antwerpen, C. Stévigny, Metabolomics fingerprint of coffee species determined by untargeted-profiling study using LC-HRMS, Food Chem. 245 (2018) 603–612, https://doi.org/10.1016/j.foodchem.2017.10.022.

[74] D. Kumar, Nuclear magnetic resonance (NMR) spectroscopy for metabolic profiling of medicinal plants and their products, Crit. Rev. Anal. Chem. 46 (2016) 400–412, https://doi.org/10.1080/10408347.2015.1106932.

[75] N. Villalón-López, J.I. Serrano-Contreras, D.I. Téllez-Medina, L. Gerardo Zepeda, An 1H NMR-based metabolomic approach to compare the chemical profiling of retail samples of ground roasted and instant coffees, Food Res. Int. 106 (2018) 263–270, https://doi.org/10.1016/j.foodres.2017.11.077.

[76] M.V. De Moura Ribeiro, N. Boralle, H. Redigolo Pezza, L. Pezza, A.T. Toci, Authenticity of roasted coffee using1H NMR spectroscopy, J. Food Compos. Anal. 57 (2017) 24–30, https://doi.org/10.1016/j.jfca.2016.12.004.

[77] A.J. Charlton, W.H.H. Farrington, P. Brereton, Application of 1H NMR and multivariate statistics for screening complex mixtures: quality control and authenticity of instant coffee, J. Agric. Food Chem. 50 (2002) 3098–3103, https://doi.org/10.1021/jf011539z.

[78] R.P. Alves, N.R.A. Filho, L.M. Lião, I.S. Flores, Evaluation of the metabolic profile of arabica coffee via NMR in relation to the time and temperature of the roasting procedure, J. Braz. Chem. Soc. 32 (2021) 123–136, https://doi.org/10.21577/0103-5053.20200162.

[79] R. Consonni, D. Polla, L.R. Cagliani, Organic and conventional coffee differentiation by NMR spectroscopy, Food Control 94 (2018) 284–288, https://doi.org/10.1016/j.foodcont.2018.07.013.

[80] L.A. Tavares, A.G. Ferreira, Análises quali- e quantitativa de cafés comerciais via ressonância magnética nuclear, Quim. Nova 29 (2006) 911–915, https://doi.org/10.1590/S0100-40422006000500005.

[81] D. Suhandy, M. Yulia, The use of partial least square regression and spectral data in UV-visible region for quantification of adulteration in Indonesian palm civet coffee, Int J Food Sci 2017 (2017), https://doi.org/10.1155/2017/6274178.

[82] M.B. Abreu, G.G. Marcheafave, R.E. Bruns, I.S. Scarminio, M.L. Zeraik, Spectroscopic and chromatographic fingerprints for discrimination of specialty and traditional coffees by integrated chemometric methods, Food Anal. Methods 13 (2020) 2204–2212, https://doi.org/10.1007/s12161-020-01832-1.

[83] M.F. Kurniawan, N. Andarwulan, N. Wulandari, M. Rafi, Metabolomic approach for understanding phenolic compounds and melanoidin roles on antioxidant activity of Indonesia robusta and arabica coffee extracts, Food Sci. Biotechnol. 26 (2017) 1475–1480, https://doi.org/10.1007/s10068-017-0228-6.

[84] N. Wang, L.T. Lim, Fourier transform infrared and physicochemical analyses of roasted coffee, J. Agric. Food Chem. 60 (2012) 5446–5453, https://doi.org/10.1021/jf300348e.

[85] A.P. Craig, A.S. Franca, L.S. Oliveira, Discrimination between defective and non-defective roasted coffees by diffuse reflectance infrared Fourier transform spectroscopy, LWT 47 (2012) 505–511, https://doi.org/10.1016/j.lwt.2012.02.016.

[86] G.F. Abreu, F.M. Borém, L.F.C. Oliveira, M.R. Almeida, A.P.C. Alves, Raman spectroscopy: a new strategy for monitoring the quality of green coffee beans during storage, Food Chem. 287 (2019) 241–248, https://doi.org/10.1016/j.foodchem.2019.02.019.

[87] B.S. Everitt, G. Dunn, Applied multivariate data analysis C1, in: B.S. Everitt, G. Dunn (Eds.), Applied Multivariate Data Analysis, first ed., John Wiley & Sons, Ltd., 2001, pp. 1–8, https://doi.org/10.1002/9781118887486.ch1.

[88] M.A. Farag, A.M. Otify, A.M. El-Sayed, C.G. Michel, S.A. ElShebiney, A. Ehrlich, L.A. Wessjohann, Sensory metabolite profiling in a date pit based coffee substitute and in response to roasting as analyzed via mass spectrometry based metabolomics, Molecules 24 (2019), https://doi.org/10.3390/molecules24183377.

[89] G. Greco, E. Núñez-Carmona, M. Abbatangelo, P. Fava, V. Sberveglieri, How coffee capsules affect the volatilome in espresso coffee, Separations 8 (2021), https://doi.org/10.3390/separations8120248.

[90] B.S. Everitt, G. Dunn, Applied multivariate data analysis C6, in: B.S. Everitt (Ed.), Applied Multivariate Data Analysis, first ed., John Wiley & Sons, Ltd., 2001, pp. 125–160, https://doi.org/10.1002/9781118887486.ch6.

[91] R. Gamboa-Becerra, M.C. Hernández-Hernández, Ó. González-Ríos, M.L. Suárez-Quiroz, E. Gálvez-Ponce, J.J. Ordaz-Ortiz, R. Winkler, Metabolomic markers for the early selection of coffea canephora plants with desirable cup quality traits, Metabolites 9 (2019), https://doi.org/10.3390/metabo9100214.

[92] B.S. Everitt, G. Dunn, Applied multivariate data analysis C3, in: B.S. Everitt (Ed.), Applied Multivariate Data Analysis, John Wiley & Sons, Ltd., 2001, pp. 48–73, https://doi.org/10.1002/9781118887486.ch3.

[93] F.H. Long, Multivariate analysis for metabolomics and proteomics data, in: Proteomic and Metabolomic Approaches to Biomarker Discovery, Elsevier Inc., 2013, pp. 299–311, https://doi.org/10.1016/B978-0-12-394446-7.00019-4.

[94] B.S. Everitt, G. Dunn, Applied multivariate data analysis C11, in: B.S. Everitt, G. Dunn (Eds.), Applied Multivariate Data Analysis, John Wiley & Sons, Ltd., 2001, pp. 248–270, https://doi.org/10.1002/9781118887486.ch11.

[95] R.G. Brereton, G.R. Lloyd, Partial least squares discriminant analysis: taking the magic away, J. Chemom. 28 (2014) 213–225, https://doi.org/10.1002/cem.2609.

[96] L.C. Lee, C.Y. Liong, A.A. Jemain, Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps, Analyst 143 (2018) 3526–3539, https://doi.org/10.1039/c8an00599k.

[97] B.W. Chen, Novel kernel orthogonal partial least squares for dominant sensor data extraction, IEEE Access 8 (2020) 36131–36139, https://doi.org/10.1109/ACCESS.2020.2974873.

[98] M. Bylesjö, M. Rantalainen, O. Cloarec, J.K. Nicholson, E. Holmes, J. Trygg, OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification, J. Chemom. 20 (2006) 341–351, https://doi.org/10.1002/cem.1006.

[99] K.A. Anderson, B.W. Smith, Chemical profiling to differentiate geographic growing origins of coffee, J. Agric. Food Chem. 50 (2002) 2068–2075, https://doi.org/10.1021/jf011056v.

[100] M.Y. Choi, W. Choi, J.H. Park, J. Lim, S.W. Kwon, Determination of coffee origins by integrated metabolomic approach of combining multiple analytical data, Food Chem. 121 (2010) 1260–1268, https://doi.org/10.1016/j.foodchem.2010.01.035.

[101] R. Consonni, L.R. Cagliani, C. Cogliati, NMR based geographical characterization of roasted coffee, Talanta 88 (2012) 420–426, https://doi.org/10.1016/j.talanta.2011.11.010.

[102] F. Wei, K. Furihata, M. Koda, F. Hu, R. Kato, T. Miyakawa, M. Tanokura, 13C NMR-based metabolomics for the classification of green coffee beans according to variety and origin, J. Agric. Food Chem. 60 (2012) 10118–10125, https://doi.org/10.1021/jf3033057.

[103] Y. Miao, Q. Zou, Q. Wang, J. Gong, C. Tan, C. Peng, C. Zhao, Z. Li, Evaluation of the physiochemical and metabolite of different region coffee beans by using UHPLC-QE-MS untargeted-metabonomics approaches, Food Biosci. 46 (2022), https://doi.org/10.1016/j.fbio.2022.101561.

[104] M.M. Artêncio, A. Luis, L. Cassago, R.K. Silva, F. Batista, D. Costa, Untargeted Metabolomic Approach Based on UHPL-ESI-HRMS to Investigate Metabolic Proles of Different Coffea Species and Terroir, 2023, https://doi.org/10.21203/rs.3.rs-2828021/v1.

[105] S.D. Williams, B.J. Barkla, T.J. Rose, L. Liu, Does coffee have terroir and how should it Be assessed? Foods 11 (2022) https://doi.org/10.3390/foods11131907.

[106] G. Mannino, R. Kunz, M.E. Maffei, Discrimination of green coffee (coffea arabica and coffea canephora) of different geographical origin based on antioxidant activity, high-throughput metabolomics, and DNA RFLP fingerprinting, Antioxidants 12 (2023), https://doi.org/10.3390/antiox12051135.

[107] R. Garrett, E.M. Schmidt, L.F.P. Pereira, C.S.G. Kitzberger, M.B.S. Scholz, M.N. Eberlin, C.M. Rezende, Discrimination of arabica coffee cultivars by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry and chemometrics, LWT - Food Sci. Technol. (Lebensmittel-Wissenschaft -Technol.) 50 (2013) 496–502, https://doi.org/10.1016/j.lwt.2012.08.016.
[108] J.H. Da Silva Taveira, F.M. Borém, L.P. Figueiredo, N. Reis, A.S. Franca, S.A. Harding, C.J. Tsai, Potential markers of coffee genotypes grown in different Brazilian regions: a metabolomics approach, Food Res. Int. 61 (2014) 75–82, https://doi.org/10.1016/j.foodres.2014.02.048.
[109] U. Jumhawan, S.P. Putri, Yusianto, E. Marwani, T. Bamba, E. Fukusaki, Selection of discriminant markers for authentication of asian palm civet coffee (Kopi Luwak): a metabolomics approach, J. Agric. Food Chem. 61 (2013) 7994–8001, https://doi.org/10.1021/jf401819s.
[110] B. Mehari, M. Redi-Abshiro, B.S. Chandravanshi, M. Atlabachew, S. Combrinck, R. McCrindle, Simultaneous determination of alkaloids in green coffee beans from Ethiopia: chemometric evaluation of geographical origin, Food Anal. Methods 9 (2016) 1627–1637, https://doi.org/10.1007/s12161-015-0340-2.
[111] B. Mehari, M. Redi-Abshiro, B.S. Chandravanshi, S. Combrinck, R. McCrindle, Characterization of the cultivation region of Ethiopian coffee by elemental analysis, Anal. Lett. 49 (2016) 2474–2489, https://doi.org/10.1080/00032719.2016.1151023.
[112] D.E. Hoyos Ossa, R. Gil-Solsona, G.A. Peñuela, J.V. Sancho, F.J. Hernández, Assessment of protected designation of origin for Colombian coffees based on HRMS-based metabolomics, Food Chem. 250 (2018) 89–97, https://doi.org/10.1016/j.foodchem.2018.01.038.
[113] S.P. Putri, T. Irifune, Yusianto, E. Fukusaki, GC/MS based metabolite profiling of Indonesian specialty coffee from different species and geographical origin, Metabolomics 15 (2019), https://doi.org/10.1007/s11306-019-1591-5.
[114] B. Mehari, M. Redi-Abshiro, B.S. Chandravanshi, S. Combrinck, R. McCrindle, M. Atlabachew, GC-MS profiling of fatty acids in green coffee (Coffea arabica L.) beans and chemometric modeling for tracing geographical origins from Ethiopia, J. Sci. Food Agric. 99 (2019) 3811–3823, https://doi.org/10.1002/jsfa.9603.