岐阜大学機関リポジトリ
Gifu University Institutional Repository

Mass Spectrometry-Based Lipidomic Profiling of Indonesian Coffee Beans for Origin Determination

メタデータ	言語: English 出版者: 公開日: 2022-12-12 キーワード (Ja): キーワード (En): 作成者: FAWZAN SIGMA AURUM メールアドレス: 所属:
URL	http://hdl.handle.net/20.500.12099/88950

Mass Spectrometry–Based Lipidomic Profiling of Indonesian Coffee Beans for Origin Determination

（質量分析に基づく脂質プロファイリングによるインドネシア産コーヒー豆の原産地判別）

2022

The United Graduate School of Agricultural Science,

Gifu University

Science of Biological Production

(Gifu University)

FAWZAN SIGMA AURUM

Mass Spectrometry–Based Lipidomic Profiling of Indonesian Coffee Beans for Origin Determination

（質量分析に基づく脂質プロファイリングによるインドネシア産コーヒー豆の原産地判別）

FAWZAN SIGMA AURUM

別紙様式第３号（第４条，第６条関係）　Form No.3

学　位　論　文　要　旨
DISSERTATION SUMMARY

氏　　名
Name　FAWZAN SIGMA AURUM

題　　目
Title of Dissertation
Mass Spectrometry–Based Lipidomic Profiling of Indonesian Coffee Beans for Origin Determination

学位論文要旨(Dissertation Summary)

Coffee has attracted consumers worldwide for its unique sensorial properties. The unique flavor of coffee is affected by numerous factors. The biochemical properties associated with geographical features are one of the essential aspects that may modulate distinct sensorial profiles of coffee. Among the factors influencing the hedonic preference of coffee consumers, aroma and mouthfeel are by far the most important. The aroma is associated with volatile compounds, while the mouthfeel is generally based on lipid components.

Lipids are biochemical compounds that are substantially present in coffee beans. However, lipids in coffee have not been comprehensively studied thus far and have not been used to differentiate the geographical origin of coffee. Indonesian coffee offers unique flavor characters, and its value in international trade is increasing. Nonetheless, the biochemical compounds data on Indonesian coffee has not been established, not to mention the lipid profile itself. This study aimed to investigate the applicability of lipid profiling for use in coffee origin authentication of Indonesian coffee. Based on the literature, this is the first study on comprehensive lipid profiling of superior coffee produced in various regions of Indonesia.

In this study, coffees produced in six different provinces of Indonesia, from different harvest years and different regions were obtained and grouped into training and validation sets for the creation of a region-discriminating model. The green coffee beans from six locations were roasted separately at the same temperature and time to obtain medium roasted beans. The roasted beans were individually ground and lipids were extracted using methyl tert-butyl ether (MTBE). Lipid extracts from roasted coffee were subjected to high-performance liquid chromatography coupled with triple-quadrupole mass spectrometry (LC–MS/MS). The lipid compounds were separated in a C-18 reversed-phase chromatographic column. Mass spectrometry separation was performed based on the multiple reaction monitoring (MRM) targeting 953 lipid features. This study used a different sample set to build up the discriminant model based on their lipidomic profile, followed by validation analysis. The obtained data were analyzed using the multivariate

approach, including partial least-squares discriminant analysis (PLS-DA), principal component analysis (PCA), and clustering analysis.

The LC–MS/MS analysis tentatively identified 85 lipid species from five global lipid classes, such as neutral lipids, sphingolipids, sterol, glycerophospholipids, and glyceroglycolipids. The PLS-DA model exhibited an accuracy of 90%–100% in discriminating the origins of coffee based on receiver operating characteristics–area under the curve analysis. The selection of important lipid features for each coffee origin was determined based on the Variable Importance in Projection (VIP) score > 0.9 and p-value < 0.05 based on the PLS-DA model. Based on this benchmark, 38 lipid species were assigned as the discriminant features of the six coffee origins. Overall, the discriminant analysis showed promising results for separating the coffee origins using the lipid profiles.

Furthermore, the validation sample set was extracted separately and subjected to lipidomic analysis using the LC–MS/MS employing MRM mode to confirm the discrimination capacity of the important features. The obtained dataset was then subjected to an unsupervised data exploration using PCA to observe the natural classification of the samples. A heatmap was visualized to illustrate the similarities between samples using a tree-structured cluster. The results showed that both PCA pattern and heatmap demonstrate natural discrimination of coffee samples based on their origins.

Conclusively, the results of this research provide solid evidence for the applicability of lipidomics profiling using LC-MS/MS for the origin discrimination of coffee. Moreover, this study might benefit the coffee industry by establishing an advanced method for determining the origin of coffee.

学　位　論　文　要　旨
DISSERTATION SUMMARY

氏　　名
Name　FAWZAN SIGMA AURUM

題　　目
Title of Dissertation
Mass Spectrometry–Based Lipidomic Profiling of Indonesian Coffee Beans for Origin Determination
（質量分析に基づく脂質プロファイリングによるインドネシア産コーヒー豆の原産地判別）

学位論文要旨(Dissertation Summary)

コーヒーは、そのユニークな官能的特性で世界中の愛好家を魅了している。コーヒーの独特な風味は、多くの因子によって構成されるが，産地に由来する生化学成分の違いは，特に強く影響を与える。また、喫飲時の嗜好性に影響を与える要因の中で、香りと口当たりは極めて重要である。香りは揮発性化合物と関連しており、口当たりは一般的に脂質成分に基づいている。

脂質は、コーヒー豆に非常に多く存在する生化学成分である。しかし、コーヒーに含まれる脂質成分の包括的研究はこれまでに殆どなく、コーヒーの産地判別に利用されることはなかった。特に、インドネシア産コーヒーは独特の風味を有しているため、国際取引価値が高まっている。しかしながら、インドネシア産コーヒーの生化学成分に関する知見は充分ではなく、脂質プロファイルについての研究は皆無である。そこで本研究では、コーヒーの産地認証における脂質プロファイリングの適用性について検討することを目的とした。インドネシアの各地域で生産されるプレミアムコーヒーを対象とした包括的な脂質プロファイリングに関する最初の研究として位置づけられる。

研究では、インドネシアの 6 つの異なる州にて生産されるコーヒーを、異なる収穫年、異なる地域から入手し、それらを産地判別モデル作成のためのトレーニングセットと検証セットにグループ化した。サンプルのコーヒー生豆は、同一温度・時間で焙煎し、中煎り豆として分析に供した。焙煎豆を粉砕し、メチル tert-ブチルエーテル（MTBE）を用いて脂質を抽出した。焙煎コーヒーからの脂質抽出物は、高速液体クロマトグラフィー・トリプル四重極質量分析計（LC-MS/MS）によって網羅的に解析した。脂質成分は、C-18 逆相クロマトグラフィーカラムで分離し、953 種類の脂質分子をターゲットとした多重反応モニタリング（MRM）に基づいて分析した。異なるサンプルセットの包括的脂質プロファイルに基づいて判別モデルを構築し、検証解析を行った。得られたデータは、部分最小二乗判別分析（PLS-DA）、主成分分析（PCA）、クラスタリング分析などの多変量解析によって分析した。

LC-MS/MS 分析により、中性脂質、スフィンゴ脂質、ステロール、グリセロリン脂質、グリセロ糖脂質の 5 種類の脂質クラスから 85 種類の脂質分子種がアノテーションされた。脂質プロファイルより得られた PLS-DA モデルは、コーヒーの産地判別に 90%-100%の精度を有することが ROC-AUC 分析により示された。各コーヒー産地の重要な脂質分子を PLS-DA モデルに基づく Variable Importance in Projection（VIP）スコアが 0.9 以上で且つ、p 値が 0.05 以下を基準に選択したところ、38 種類の脂質が 6 種類のコーヒー原産地を判別するためのマーカー分子として示された。全体として、脂質プロファイルを用いた判別分析は、コーヒーの原産地を分離するために有用な結果を示した。さらに、検証用サンプルセットを別途対象とし、MRM モードを用いた LC-MS/MS によりリピドームの分析を行い、選択された産地判別マーカーによる産地識別能力を確認した。検証用データセットを PCA や階層的クラスター解析による教師なし分類に供し、サンプル間の類似性を可視化した。その結果、PCA パターン、クラスター解析ともに、コーヒーサンプルの産地判別が可能であることが示された。

以上のことから、LC-MS/MS を用いた包括的脂質プロファイリングはコーヒーの産地判別に適用可能であることが示された。さらに、本研究はコーヒーの産地判別のための高度な手法を確立したもので、得られた成果はコーヒー産業に大いに貢献するものと考えられる。

Table of Contents

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

1.1. General Background

Coffee is the most consumed hot beverage worldwide. Geographically, coffee-producing countries are in the equatorial zone. Interestingly, coffee-consuming countries are not included in the production area. Countries with a developed economy, such as the European Union countries, United States, Japan, Russian Federation, and Republic of Korea, are the top coffee importers. The global sales of coffee has exhibited an increasing trend for the past decade (International Coffee Organization, 2020). In recent years, consumers prefer single-origin coffee. Single-origin coffee offers a unique flavor because it is obtained from a single farm or plantation in a specific region and is not blended with coffee from other origins.

Several factors affect the flavor and character of coffee, such as the roasting condition (Dias et al., 2014), product packaging method (Cincotta et al., 2020), pre- and post-harvest practice (da Silva Oliveira et al., 2021), varieties of coffee cultivated, environmental features (Bodner et al., 2019), and the portion of undergrade or defective beans (J. R. Santos & Rodrigues, 2020; N. Yang et al., 2016). However, the biological interaction between the genetic aspects and surrounding geographical environment can be an essential factor in the flavor and character of coffee. This interaction may lead to the distinction of a coffee phenotype as a biological entity which, in turn, contributes to the distinctive and diverse flavor formation, which is reflected in the biochemical profile.

Biochemical compounds are responsible for the sensorial profile and specific notes formed during the developmental stage of coffee cherry fruit. This is followed by

post-harvest treatments all the way to cup serving. These compounds are essential in influencing the hedonic preference of the coffee consumer. It is common practice in a coffee community to conduct a cupping test in order to evaluate the sensorial attributes of the coffee, especially for single-origin specialty coffee. The test method is internationally standardized by the Specialty Coffee Association of America, and the test is conducted by a certified panelist (Pereira et al., 2017). However, this method of testing is arguably subjective.

Single-origin coffee has been changing market dynamics. However, the coffee industry has not yet established critical indicators to claim coffee originality related to its unique biochemical substance. In the market, the price of single-origin coffee is determined by its grade and unique flavor, which, in turn, is based on the growing region (Mehari et al., 2016a; Putri & Fukusaki, 2018). The price disparity between *Coffea arabica* (Arabica coffee) and *Coffea canephora* (Robusta coffee), as well as higher- and lower-grade quality coffee, may stimulate food fraud and the practice of adulteration. Therefore, the coffee industry needs a robust method to authenticate the origin of coffee, in which one of the most effective strategies is to create a database of the biochemical profile and descriptors of coffee that is on the market.

Several works were completed on coffee origin authentication using various biochemical compounds, including caffeine, chlorogenic acid, total phenolic, and triglyceride compositions, as the basic chemical descriptor in coffee using high-performance liquid chromatography (HPLC) and or gas chromatography–mass spectrometry (GC/MS) (Carrera et al., 1998; González et al., 2001; Martín et al., 2001; Martín et al., 1998). These studies successfully distinguished the Arabica and Robusta varieties but were not able to reveal the geographical origin classification.

Other studies used a volatile compound (Ongo et al., 2020; Risticevic et al., 2008; Zambonin et al., 2005), volatile and carbohydrate (Choi et al., 2010), phenolic compounds and chemical profile (Mehari et al., 2016b; Monteiro et al., 2019), and alkaloid profile (Mehari et al., 2016a) to distinguish the different geolocations of each coffee. However, only a few studies pay attention to the most abundant biochemical compound in the coffee bean, which remains intact during storage and after being roasted: the lipid constituent (Anese et al., 2000; Speer & Kölling-Speer, 2006). The lipid constituent is not involved in the Maillard reaction during the roasting process. Yet, it may derive a hydrophobic compound that influences the flavor and provides fat-soluble vitamins that facilitate the organoleptic parameters of coffee (Selmar et al., 2014).

A limited number of studies employed specific lipid class profiling to discriminate the origin of coffee, especially the fatty acid group (Dong et al., 2015; Mehari et al., 2019; Romano et al., 2014). Another study used a triacylglycerol (TAG) profile to distinguish between Arabica and Robusta blends (Cossignani et al., 2016). These studies indicate that lipids can be used as a discriminant marker in coffee. However, lipids consist of five global classes, including sphingolipids, sterol, polar glycerophospholipids, polar glyceroglycolipids, and neutral glycerolipids (Tarazona et al., 2015). A more comprehensive lipid profile using the five classes of lipids has not been studied for the geographical origin classification of coffee.

In this study, six Indonesian coffee origins were selected to represent the major producing regions in Indonesia. In addition to the increasing market value, these coffee offer unique flavor characters. Therefore, they are labeled with a geographical indication by the Indonesian government (DJKI Kementerian Hukum dan HAM, 2022).

The coffee used in this study included Gayo, Mandheling, and Lampung, which all came from Sumatera Island. Sumatera is one of the largest islands in Indonesia that accounts for 70% of coffee production in Indonesia (Directorate General of Estate, 2020). Gayo and Mandheling coffee, originating from the Aceh province and North Sumatera, respectively, are the most demanded varieties for the export market (Damayanti & Setiadi, 2019). Lampung coffee is a Robusta variety with large productivity and economic value (Rosiana, 2020). Coffee from Kintamani in Bali Island is famous for its touristic name. Toraja coffee from Sulawesi Island has been exported to Japan, the USA, and Australia as a premium coffee. As a representation of the East Indonesian region, coffee from Wamena was selected owing to its unique sensorial properties.

1.2. Objective of the Study

Currently, the study on biochemical compounds of Indonesian coffee is developing, however, several data shortages are yet to be explored. Therefore, all parties, including academia, government, and industry, need to establish accurate origin assurance to convince the global market. To date, much of the coffee studies in lipidomics reported on the specific class of lipid. Based on the literature, comprehensive lipid profiling studies of Indonesian coffee have not been found.

Therefore, this main aim of this study was to investigate the applicability of lipid profiling by LC-MS/MS for determining the origin of valuable Indonesian coffee. To achieve this aim, several steps were done which are divided into several chapters in this dissertation. The second chapter of this dissertation contains the literature review of the most recent techniques for coffee authentication, origin determination, and adulteration

detection from both analytical and nondestructive approaches that have been published in the literature (Aurum et al., 2022a). The review was done following the systematic literature review (SLR) using Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2009).

In chapter three, the methodology for analyzing and observing associated parameters related to lipid metabolites profiling based on Liquid Chromatography Mass Spectrometry (LC-MS/MS) technique are extensively discussed. This chapter also mentions the method for identifying the variable important based on the analysis to find the significant discriminant marker of coffee from each geographical origin followed by model validation using the coffee samples from different harvest years. Furthermore, several multivariate data analyses followed by a model performance evaluation are employed in the present study to obtain a robust model. Chapter four describe the results of the study and its substantial discussion pertaining to lipidomics profiling of Indonesian coffee for geographical determination (Aurum et al., 2022b). To the best of the authors' knowledge, this is the first study reporting a comprehensive lipid profiling of Indonesian coffee.

CHAPTER 2

LITERATURE BACKGROUND

2.1. Systematic Literature Review on Coffee Origin Determination

Generally, coffee consumers have a preferred or favorite coffee origin. This consumer choice is related to sensory properties, such as aroma and mouthfeel. Coffee has a unique organoleptic profile associated with its growth geographical location. Single-origin coffee refers to coffee cultivated in a specific microclimate and typically sourced from a certain geographical place, such as a farm or multiple farms, or plantations, or a region within the same country. This type of consumption has been increasing all over the globe (Wilson et al., 2012).

However, high-quality single-origin coffee is prone to misleading labels, false declarations, and fraudulent practices to increase their economic profit. Recently, the Federal Food Safety and Veterinary Office of Switzerland reported a falsely declared "100% Arabica" coffee substituted by the cheaper Robusta coffee (Federal Food Safety and Veterinary Office, 2019).

Coffee fraud may imply, for instance, counterfeiting high-quality and specialty coffee beans with lower quality or defective beans. Another possibility is the falsification of geographical origin information (Toci et al., 2016). As an indicator of product and process quality, product origin shows increasing importance for the business and for informing consumers' purchasing decisions.

In response to consumer demands for authenticity of coffee origin, various strategies encompassing a broad range of technology and scientific techniques have been applied to assure this point. In the past decade, studies pertaining to coffee origin authentication, determination and classification were done utilizing near-infrared (NIR)

(Bona et al., 2017; Okubo & Kurata, 2019; Scholz et al., 2014), Fourier Transform Mid-infrared (FT-MIR) (Mendes et al., 2022), Terahertz Spectroscopy (S. Yang et al., 2021), e-nose or e-tongue sensors (Domínguez et al., 2014; Flambeau et al., 2017), and UV-visible spectroscopy (Suhandy & Yulia, 2017). Moreover, several studies with similar purposes were conducted by gas (Dong et al., 2015; Mehari et al., 2019; S. P. Putri et al., 2019) or liquid chromatography (Aurum et al., 2022b; Badmos et al., 2020; Mehari et al., 2016b) coupled to mass spectrometry, nuclear magnetic resonance (NMR) (de Moura Ribeiro et al., 2017; Happyana et al., 2020b), and polymerase chain reaction (PCR) (Combes et al., 2018; Ferreira et al., 2016; Hamdouche et al., 2016).

These techniques can be categorized into nondestructive and analytical approaches. The nondestructive technologies allow a rapid analysis and are less laborious, considerably saving costs, and require little or no disruption of the biochemical potency of the sample (Faith Ndlovu et al., 2022). On the other hand, the advantage of using an analytical approach in the geographical determination of coffee is associated with the possibility of analyzing important markers indicative of its origin (Thorburn Burns et al., 2017).

Nondestructive equipment produces large number of spectral signals that can be count as variables, therefore it is attractive in their data processing and chemometric analysis. Machine learning and deep learning algorithms are often used to interpret the data. Nevertheless, the data processing of analytical techniques can be more advanced when coupled to the bioinformatics aspect to reveal the important compound markers. Nondestructive methods require a large samples dataset to build the initial model. In addition, data interpretation can be difficult. Therefore, other studies use analytical approaches, such as chromatography and mass spectrometry, or a combination.

Regarding the capability of the analytical method to identify coffee markers, the sensorial properties of coffee heavily rely on its biochemical compound content. Its unique taste is affected by numerous factors. From the very beginning is the environment where it is planted, followed by the coffee cherry growth stage, the local farmers' postharvest tradition and then successively up to the method of brewing the coffee for serving in the cup.

Among the factors influencing the hedonic preference of coffee consumers, aroma and mouthfeel are by far the most important. The aroma is associated with volatile compounds, while the mouthfeel is generally based on its lipids components. Both compounds can be analyzed using a metabolomics approach. According to Fiehn (2002), metabolomics is the comprehensive analysis of global metabolites of a biological system. Similarly, the comprehensive study of lipid compounds is commonly called Lipidomics (Watson, 2006). Lipidomics is a developing research field supported by the improvement of various analytical methods, especially mass spectrometry and bioinformatics (Fahy et al., 2011).

Most studies applying chromatography and mass spectrometry aim to explore the metabolite profile and identify the key flavors of coffee identifying different origin locations. Furthermore, numerous studies show that each coffee origin is characterized by distinct biochemical compounds.

This review aimed to investigate the most recent techniques for coffee authentication, origin determination, and adulteration detection from both analytical and nondestructive approaches. Despite the existence of previous similar reviews, this is the first systematic literature review (SLR) following Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2009).

2.2. Protocol and Eligibility Criteria

The objective of this SLR is to offer a thorough description of the latest findings in the latest updated research context and the scope for research questions for future study. This study aimed to deliver an accurate scientific report and avoid bias; therefore, this study adopts SLR methodology by Moher et al. (2009). Figure 1 presents the study flow, which began by identifying studies in literature databases according to certain search words, followed by several screening steps. The article title was the first screened, followed by the abstract. Finally, the full text was rigorously studied to be included in the primary literature.

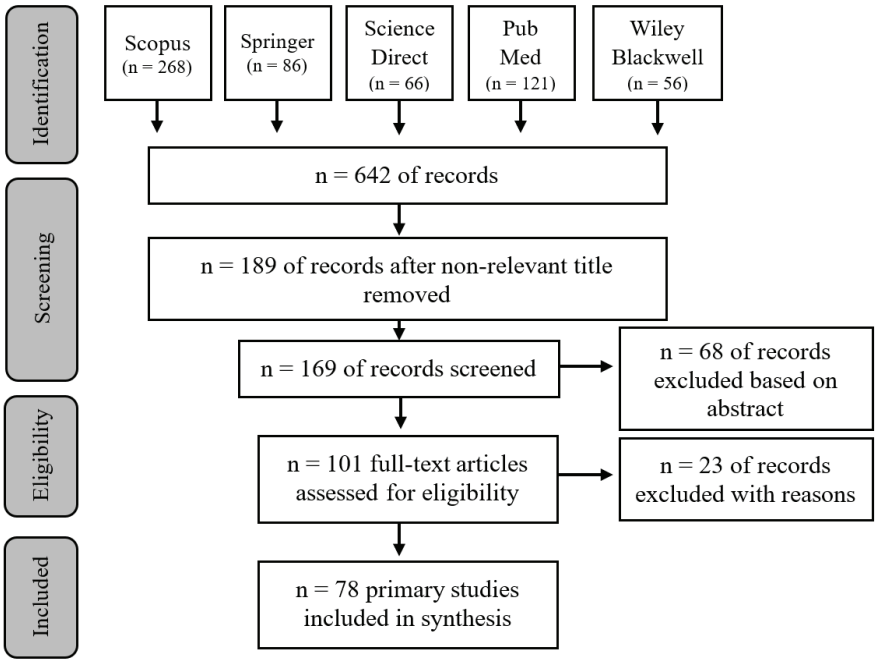


Figure 1: PRISMA flow chart for primary literature screening.

The PICO (Population, Intervention, Comparison, and Outcome) framework was used to define the inclusion criteria (Table 1). The PICO framework is a model for conducting a reference search that divides a formulated research question into four

distinct components: the population of interest, the applied intervention, the comparison or controls, and the measured outcome.

2.3. Information Sources and Search Strategy

Articles were obtained from well-established databases, i.e., Scopus, Springer, ScienceDirect, PubMed, and Wiley-Blackwell. Each database uses a different style of search syntax and operators. The keywords or syntax used are indicated in Table 2. The search strategy yielded several number of articles (n), which is indicated in Figure 1. The syntax limited the search for peer-reviewed publications (original and review papers), in English language, journal articles, and published within the past 10 years, from 2012 onwards. These records were exported to Mendeley reference manager (Ver 1.19.8), following the removal of duplicated studies and irrelevant article types.

Table 1: PICO Summary in this review

Framework	Criteria
Population (P)	All coffee varieties in the Coffea (genus)
Intervention (I)	Geographical authentication, origin determination
Comparison (C)	Destructive vs. Nondestructive approaches
Outcomes (O)	Comparison of both approaches

2.4. Data Extraction

Various data were extracted from the final list of included studies, namely, the publication year, techniques or instrumental approach used, country of origin of coffee

samples, classification model algorithm and performance evaluation, feature selection, associated information on coffee processing, and key findings.

Table 2: Syntax and keywords for database search

Database		Keywords and syntax
ScienceDirect	:	[(coffee OR coffea) AND (geographic OR origin OR region OR country) AND (authentication OR determination OR discrimination)] year 2012–2022
Scopus	:	TITLE-ABS-KEY [(coffee OR coffea) AND (geographic* OR region*) AND (authenticat* OR origin OR provenance) AND (determin* OR discriminat*)) AND PUBYEAR > 2011 AND (LIMIT-TO (PUBSTAGE, "final")) AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "ch") OR LIMIT-TO (DOCTYPE, "re")) AND (LIMIT-TO (LANGUAGE, "English")]
Springer	:	with all of the words: coffee geographic* origin country with at least one of the words: authenticat* discriminat* determin* classif* where the title contains: coffee year: 2012–2022
PubMed	:	"coffee OR coffea" in Title and "(origin OR region* OR geographic* OR country) AND (trac* AND authenticat* OR determin* OR discriminat* OR classification)" in Abstract
Wiley-Blackwell	:	"coffee OR coffea" in Title and "(origin OR region* OR geographic* OR country) AND (trac* AND authenticat* OR determin* OR discriminat* OR classification)" in Abstract Year 2012–2022

2.5. General Overview of the Literature

Screening results using the PRISMA approach are indicated in Figure 1. The search strategy identified a total of 642 records. Then, each record was rigorously assessed for eligibility, after which 78 original research papers were retrieved to be reviewed in detail. The number of studies using analytical and nondestructive methods was 55 and 23, respectively. In addition to the original papers, seven review articles on similar topics to the current study are listed and briefly discussed in Table 3. Generally, said reviews discussed and identified common methodologies to determine coffee's

geographical origin, adulteration, and fraudulent practices. However, none of the articles performed a systematic review. In contrast, the present study performed a more detailed and structured assessment of the most updated research in coffee origin classification, determination, and authentication. The importance, advantages, and features of this review compared to existing reviews are shown in Table 3.

2.6. Analytical Approaches

Numerous studies on biochemical coffee profiling based on broad range metabolome analysis have been conducted. The key information of the 55 studies on coffee origin determination, authentication, and adulteration using the analytical approaches are exhaustively summarized and listed according to publication date in Table 4. Early studies on coffee classification used Nuclear Magnetic Resonance (NMR)-based fingerprinting and elemental analysis using Inductively Coupled Plasma Mass Spectrometry (ICP-MS)-based. Using NMR approaches, coffee green beans from different countries (Wei et al., 2012) and roasted coffee from several continents were classified (Consonni et al., 2012), as well as quantification of adulteration of coffee varieties (Arabica and Robusta) (Cagliani et al., 2013). In addition, Arana et al. (2015) employed NMR to distinguish Colombian coffee from that from other origins. These studies found that the NMR spectra of coffee samples showed significant resonance from caffeine, sugar compounds, chlorogenic acids, fatty acids, and amino acids. In addition, recent research using the NMR approach also found that lipids, acetic acid, lactic acid, and quinine were discriminative compounds for several Indonesian coffees (Happyana et al., 2020a, 2020b).

Table 3: Previously published reviews on similar topics to the current article

Authors (year)	Aim of the review	Adoption of SLR	Number of Articles	Number of Journals	Interval Time	Classification variables
Toci et al. (2016)	Reviewing three main general methods (physical, chemical, and biological) for the determination of coffee adulteration.	No	30	Not explicitly indicated	1984–2014 (Selective only)	- Global methods classification (Physical, Chemical, Biological) - Equipment/Analytical method - Authors - Year - Types of adulterants
Thorburn Burns et al. (2017)	Reviewing all kinds of method for the determination of coffee adulteration materials, geographical origin, and genotype. And suggesting the appropriate approach for the determination.	No	30	Not explicitly indicated	Not explicitly indicated	- Purpose of determination - Identified or employed markers - Equipment/Analytical method
Martins et al. (2018)	Reviewing analytical approach specifically for liquid and gas chromatography in coffee fraud studies. And explaining the economic importance of coffee.	No	Not explicitly indicated	Not explicitly indicated	1988–2018 (The last 30 years from publication date)	- Types of biochemical analytes - Methods of fraud detection - Fraud types (e.g., geographical authenticity, adulteration of undeclared plant materials, coffee variety substitution)
Wang et al. (2020)	Reviewing chemical profiling both targeted and nontargeted method, including the nondestructive approach for the detection of adulteration focusing on brewed coffee.	No	Not explicitly indicated	Not explicitly indicated	Not explicitly indicated	- Grouped based on methodology and analytical instrument
Thorburn Burns and Walker (2020)	assessed for the identification of the most common materials used to adulterate coffee by dilution, to establish the geographic origins, the genotypes of beans, and to assess the authenticity of kopi luwak coffee. Also, this	No	Not explicitly indicated	Not explicitly indicated	1820–2018	- Types of adulteration materials - Analytical Methods of fraud detection - Identified and employed markers - Geographical origin - States of beans (e.g., green beans or

	report also provides a historical overview that looks at studies across period of time.					roasted beans)
dos Santos and Boffo (2021)	Reviewing general aspects of coffee biochemical attributes, extraction methods, and the existing bioactive compounds. And comparing the analytical techniques employed for characterization and quantification of chemical composition of coffee associated with its quality, origin, and adulteration detection.	No	Not explicitly indicated	Not explicitly indicated	2010–2020 (The last 20 years from the publication date)	- Analytical method and chemometrics data processing, as well as the combination of the analytical methods.
Perez et al. (2021)	General review of the chromatography, spectroscopy, and single-nucleotide polymorphism-based approaches that have been employed to discriminate between the dominant coffee species Arabica and Robusta, geographical origin, and adulteration.	No	Not explicitly indicated	Not explicitly indicated	Not explicitly indicated	- Analytical method - Biochemical compounds - Coffee origin - States of beans (e.g., green beans or roasted beans) - Coffee variety or its blends - Advantages and disadvantages - Multivariate analysis
The current article	This review assesses the most recent studies focusing on the comparison of analytical and nondestructive approaches for coffee geographical origin determination and authentication. This study uses a more systematic structure. A brief and clear comparison parameter to compare the selected literature that was rigorously studied employing PRISMA framework. This is the first work of SLR on coffee authentication studies which focus and emphasize on methodology and its data processing or multivariate or	Yes	78	37	2012–2022 (the last 10 years)	- Analytical methods/technique - Number of classes of origin - Number of samples per class - Countries of origin - Algorithm for classification - Classification model performance evaluation - Model prediction performance - Important feature selection method - Associated information regarding coffee production/processing steps - Important finding of the studies

ICP is often used for fingerprinting the elemental compounds and or isotope ratios of coffee samples. Green and roasted coffee beans analyzed by ICP-MS and ICP-Emission Spectroscopy showed negligible differences in the elemental composition. Furthermore, harvest year and degree of ripeness were nonsignificant (Valentin & Watling, 2013). Another study employing ICP-optical emission spectrometry (OES) found that metal element content could discriminate the coffee origin from different countries in South America (Muñiz-Valencia et al., 2013), the inter-Mexican region (Muñiz-Valencia et al., 2014), cross-continental samples (Carter et al., 2015), Ethiopian coffee from 11 different regions (Habte et al., 2016) and different postharvest process (Mehari et al., 2016b), and Jamaican coffee against non-Jamaican (Antoine et al., 2016). In agreement with Valentin and Watling (2013), Habte et al. (2016) confirmed that harvest year did not significantly influence coffee's elemental compounds.

The more recent studies in Table 4 use chromatography combined with mass spectrometry. Gas Chromatography-Mass Spectrometry (GC/MS) volatile compound profiling served to determine the coffee origin among several countries on different continents; this study used postharvest process as a classification variable (Caporaso et al., 2018), and also civet coffee (*kopi luwak*) discrimination against non*Luwak* coffee from the Philippines (Ongo et al., 2020). Other volatile metabolomic approaches were applied for authentication by analyzing variation in its roasting levels (Abdelwareth et al., 2021; Demianová et al., 2022). Overall, volatile profiling found that pyrazines, furans, and other aromatic hydrocarbons influenced the coffee origin classification. GC/MS untargeted metabolomics profiling was used to determine coffee from various places in Indonesia, finding that the metabolome profile of green Arabica coffee beans differed from Robusta beans, as well as, the differentiation of roasted coffee beans from

various island (Putri et al., 2019). Recently, the same researchers indicated that the postharvest process is the most discriminative aspect in coffee, followed by geographical origin (Amalia et al., 2021).

The volatile compounds of coffee from 7 different cultivars in Hainan (China) were profiled using headspace solid-phase microextraction (SPME) GC/MS, with unsatisfactory results for the differentiation of green coffee beans. Yet, combining several analysis including fatty acids, amino acids, and proteins, the study could successfully classify the Robusta Hainan coffee samples (Dong et al., 2015). Volatile profiling was not effective in discriminating the origin of raw green coffee because of the lack of aroma at this stage. However, several studies could discriminate raw green coffee beans. For instance, a study using Ultra Performance Liquid Chromatography-Mass Spectrometry (UPLC-MS) to determine phenolic (Mehari et al., 2016b) and alkaloid compound (Mehari et al., 2016a) profiles could classify Ethiopian coffee (east, northwest, west, and south regions). As well as Yemeni (Mohammed et al., 2019) and Ethiopian (Endaye et al., 2020), coffee green beans were classified using elemental analysis by ICP–OES.

Table 4: List of studies employing analytical approaches

Ref	Publication	Technique	Number of class origin/number of sample	Countries of coffee origin	Classification model and performance evaluation	Availability of feature selection method	Associated information in coffee processing (postharvest, roasting, etc.)	Important finding
(Wei et al., 2012)	J. Agric. Food Chem.	^{13}C-NMR	6 origins (60 total samples)	Brazil, Columbia, Guatemala, Tanzania, Indonesia, Vietnam.	PCA and OPLS-DA evaluated by R^2X (goodness of fit) and Q^2 (predictability)	Guided selection based on the S-plot from OPLS-DA.	Different year of harvest.	The general ^{1}H-NMR fingerprints between the Arabica and the Robusta samples were not significantly different. The selected major features were sucrose, caffeine, chlorogenic acids, choline, amino acids, organic acids, and trigonelline.
(Consonni et al., 2012)	Talanta	^{1}H NMR data	23 countries divided into 3 classes (continent); total 40 samples.	Cape Verde, Ethiopia, Kenya, Malawi, Saint Helena, Tanzania, Brazil, Colombia, Costa Rica, El Salvador/Brazil, Galapagos, Guatemala, Hawaii, Honduras, Jamaica, Nicaragua/ Guatemala, Peru, Mix South America, India, Indonesia, Nepal, Yemen.	OPLS-DA with cross validation and evaluated by Sensitivity and Specificity.	Selected based on S-plot	Different roasting process and powder size	The use of OPLS-DA models on 1H-NMR data resulted in a clear discrimination of samples based on their origin. The primary components defining coffee from America: fatty acids and chlorogenic acids; African was lactate and Asian samples were acetate and trigonelline. OPLS-DA overall goodness of fit (R^2Y) of 81.5% and an overall cross-validation coefficient (Q^2Y) of 69.7% were achieved for the classification of samples based on the continent of origin. Overall specificity of the model was 100%, while sensitivity ranged from 88.89% to 100%.
(Garrett et al., 2013)	LWT	Direct-infusion electrospray ionization Fourier-transform ion cyclotron resonance mass spectrometry (ESI FT-ICR MS)	2 classes (cultivars) and 2 sub classes (growing region)	Brazil, (Cultivars: Sarchimor and Catuaí) cultivated in Londrina and Mandaguari regions.	PLS-DA evaluated by R^2 and Q^2 which is validated by LOOCV.	VIP score > 1 from PLS-DA algorithm	Green beans	The study identified 20 important compounds: Caffeic acid, Ferulic acid, Quinic acid, Caffeoylquinic acid, Feruloylquinic acid, diCaffeoylquinic acid, Feruloylcaffeoylquinic acid, Palmitic acid, Linoleic acid, Oleic acid, Stearic acid, Arachidic acid, Behenic acid, Sucrose, Atractyloside II, Carboxyatractyloside II, Atractyloside III, Carboxyatractyloside III, Atractyloside I, Carboxyatractyloside I. Coffee from different cultivars and geographical origin can be distinguished based on these 20 compounds using PLS-DA.
(Caglia ni et al., 2013)	Talanta	^{1}H-NMR	15 classes (from 0% to 100% Arabica coffee)	Africa, America, Asia (not explicitly mentioned)	OPLS with model performance parameter of $R^{2\,=}$ 0.998 on test dataset and $Q^{2\,=}$ 97.0% (cross validation).	Selected several important variables but not explicitly mentioning the methods.	Roasted coffee	Identified several compounds: acetate, chlorogenic acids, caffeine, quinic acids, trigonelline, 2-furyl methanol, N-methyl pyridine, and formiate.

(Sun et al., 2013)	J. Radioanal. Nucl. Chem.	Photon activation analysis (PAA)	3 classes	Columbia, Guatemala, Hawaii	Canonical Scores Plot, not explicitly mentioning the validation method.	Selected several important variables but not explicitly mentioning the methods.	Coffee sample from one of the origins were washed process and dried process.	More than 30 elements were found in coffee samples: Na, Mg, Ca, Sc, Ti, Cr, Mn, Fe, Co, Ni, Zn, Ga, As, Se, Br, Rb, Y, Zr, Mo, Ag, Cd, Sn, Sb, Te, I, Ba, Ce, Tl, Pb, U. And Fe, Sn, and Rb are the most important discriminant compounds based on the statistical analysis.
(Valentin & Watling, 2013)	Food Chemistry	Inductively coupled plasma mass spectrometry (ICP-MS) and inductively coupled plasma emission spectroscopy (ICP-ES).	3 classes based on continents, 3 classes from the same continent, 5 classes from the same country	Kenya, Ethiopia, Uganda, Indonesia, India, East Timor, Australia, Papua New Guinea, Cuba, Dominican Republic, Costa Rica, Guatemala, Colombia, Brazil, Peru.	LDA with up to 100% correct cross validation.	Forward stepwise method from LDA algorithm.	Roasting process, coffee cherry ripeness level, harvest time.	Most discriminative elements for regional separation Ca, Ti, Mn, Co, Ni, Se, Rb, Sr, Mo, Cs, and Ba. Harvest year, degree of ripeness, and roasting process had little influence on the elemental composition of the samples.
(Muñiz-Valencia et al., 2013)	Food Anal. Methods	Inductively coupled plasma optical emission spectrometry (ICP–OES)	3 classes (total 46 samples)	Brazil, Colombia, Mexico.	LDA and SIMCA. Model performance of LDA (overall Sensitivity 97%; specificity 99%), and SIMCA (Sensitivity 94%; Specificity 99%)	Forward stepwise method from LDA algorithm	N/A	Ca, Cu, Fe, K, Mg, Mn and Na were the most discriminant variables.
(Muñiz-Valencia et al., 2014)	J. Food Compos. Anal.	Inductively coupled plasma optical emission spectrometry (ICP–OES)	4 classes (total 51 samples)	Mexican coffee from Chiapas, Colima, Oaxaca and Veracruz regions.	LDA and Artificial Neural Networks (ANN). LDA overall model performance (Sensitivity 81% and Specificity 94%), ANN overall model performance (Sensitivity 93% and Specificity 98%)	Forward stepwise method from LDA algorithm	N/A	ICP–OES can be employed for region-based same country determination of coffee. Ca, K, Mn, Mg, Na, and Zn were the elements that can be employed for several Mexican coffee differentiations.

(Liu et al., 2014)	Food Chem.	High-resolution inductively coupled plasma mass spectrometer (HR-ICP-MS)	4 classes (total 14 samples)	Taiwan, Ethiopia, Tanzania, Malawi, Rwanda, Uganda, El Salvador, Guatemala, Puerto Rico, Jamaica, Colombia, Brazil, Papua New Guinea, Indonesia.	PCA and Pearson Correlation. The article is not informing the model validation method.	Selecting several important variables yet does not clearly mentioning the method.	N/A	Elemental analysis cannot be employed to classify coffee geographical origin based on PCA model. On the other hand, the isotope ratio of boron and strontium may discriminate the origin of coffee in the continental level or smaller region.
(Link et al., 2014)	Food Res. Int.	Density (free fall method) and chemical analysis performed using several techniques, e.g., caffeine (spectrophotometric), CGA [refers to Clifford and Wight (1976)], total tannins [refers to AOAC (1990)], total sugar (Somgyi and Nelson reagent), protein and lipids [AOAC (1990)]	2 global classes (modern genotypes and traditional cultivars) (total 54 samples)	Brazil (grown in Paraná), 14 genotypes.	Self–organizing map (SOM). Evaluated by mean quantization error and trained with 7000 epochs.	Visual observation based on a clear separation of the samples on the maps.	N/A	Several chemical analyses coupled with SOM data processing and modeling can be used to classify different genotypes of coffee from Brazil.
(Yener et al., 2014)	J. Mass Spectrom.	Proton-Transfer-Reaction-Time of Flight-Mass Spectrometry (PTR-ToF-MS)	3 classes (108 samples)	Brazil, Ethiopia, and Guatemala	PLS-DA. Cross validation was used to calibrate the model, followed by validation from the different batch of sample.	Variables Importance in Projection (VIP) > 1.5 was used to select the most influential markers.	Different batch was used for validation	The provenance of the coffee beans shows significant effect on the volatile composition of roasted and ground coffee. N-heterocycles like pyrazines (e.g., 3-methylpyrazine), pyrroles (methyl pyrrole), pyrazole, and furans (3-penthlyfuran) are the most determinative compounds.

(Arana et al., 2015)	Food Chem.	^{1}H-NMR	2 global classes (1. Colombian coffee 2. Non-Colombian coffee); 160 total samples.	Colombia, Brazil, Ecuador, Peru, Hawaii, Costa Rica, Dominican Rep, El Salvador, Guatemala, Honduras, Jamaica, Mexico, Nicaragua, Panama, Uganda, Togo, Tanzania, Ethiopia, Ivory Coast, Cameron, China, India, Indonesia, Vietnam	PLS-DA. Cross validation was done to reduce over fitting when constructing the model. Dataset was divided into training (80%) and validation (20%). Model goodness was evaluated by R^2 and Q^2 analysis. Model sensitivity 95% and specificity 97%.	Selection based on the loading scores of PLS-DA. Not clearly mentioning the algorithm used for the selection.	Two years consecutive sample collection.	^{1}H-NMR untargeted fingerprinting may be used to differentiate coffees from Colombia and with the other origins from several continents.
(Oliveira et al., 2015)	Food Chem.	High-Resolution Continuum Source Atomic Absorption Spectrometry (HR-CS-AAS)	5 global classes (39 samples)	Kenya, Papua New Guinea, Timor, Mussulo, Colombia, India, Brazil, Honduras, Guatemala, Cuba, Mexico, China, Ethiopia	Canonical Discriminant Analysis (CDA)	Synthesized from CDA model, yet not clearly mentioning the algorithm for selection.	Extracted with espresso coffee serving method	Elemental analysis based on Ca, Mg, Na, K, P, Fe, and Mn using HR-CS-AAS was able to be used to classify coffee from different continental origin. In addition, the significant chemical descriptors (i.e., Mn and Ca) were able to be used to classify coffee based on country of origin. Both were achieved using CDA model.
(Dong et al., 2015)	Molecules	HS-SPME/GC-MS, LC-MS	7 classes (total 21 samples)	China (Hainan Province), from 7 different cultivars.	PCA, Hierarchical Cluster Analysis (HCA)	Based on loading plot of PCA.	N/A	Fatty Acid, Amino Acid, and 77 volatile elements and sensorial attributes were used for classification of cultivars. Volatile compounds cannot be used to differentiate green beans of Arabica against Robusta coffee. Combination
(Carter et al., 2015)	J. Agric. Food Chem.	Isotope Ratio Mass Spectrometer (IRMS), ICP–OES, GC/ISQ-MS	7 classes (total 54 samples)	Region-based classification, i.e., Africa, Australia, Central America, Indonesia, India, Papua New Guinea, and South America	Principal Component Discriminant Analysis and Pearson correlation. Validation was done using jackknife approach leave one out for cross validation.	Stepwise Discriminant Analysis (DA) using Mahalanobis distance	N/A	PCA model cannot classify coffee from different region using elemental composition. Stepwise DA algorithm selected Ca, Ti, Fe, Ni, Zn, (elemental composition) and δ^{2}H, δ^{13}C, δ^{18}O (isotopic data) as the major discriminant variables.

(Habte et al., 2016)	Food Chem.	ICP–OES, ICP-mass spectrometry (ICP-MS), and direct mercury analyzer (DMA)	11 classes (129 total samples)	Ethiopia (11 major producing regions)	LDA (cross validation was implemented for model calibration). PCA model was used to determine sample separation.	Forward stepwise method and PCA score plot.	For certain sampling location, two different harvest years were used for building the LDA model.	Micro and trace elements analyzed by ICP–OES or ICP-MS from the Ethiopian coffee can be used for geographical discrimination. LDA model shows better classification compared to PCA model.
(Mehari et al., 2016c)	Anal. Lett.	ICP–OES	3 global classes (49 total samples)	Ethiopia (East: Harar; South: Sidama, Yirgachefe; West: Wollega, Kaffa, Jimma)	LDA. Model assessment was done by LOOCV. Data were divided into training and validation set. Model performance was evaluated by class prediction ability on the unlabeled validation dataset. Accuracy of 92% and 79% were achieved for predicting growing region and variety, respectively.	Several important elements were selected based on LDA loading score. Canonical function coefficients value was used to calculate the loading contribution.	Washed and natural/dry process	ICP–OES was able to be used for elemental analysis of Ethiopian coffee, following its regional classification. Elements of P, Mn, S, Cu, and Fe were the most important variables for classification. Score plot of the three dimensional LDA canonical function was able to be used to classify the subregional level coffee origins.
(Antoine et al., 2016)	J. Radioanal. Nucl. Chem.	Instrumental neutron activation analysis (INAA) and ICP–OES	2 global classes (total of 24 samples)	(Jamaican and Non-Jamaican)	Agglomerative Hierarchical Clustering (AHC) and PCA. Not clearly mentioning the validation method.	Several elements were claimed to be more influential which was observed based on their concentrations.	Roasted, roasted and powdered, instant/soluble coffee.	INAA and ICP–OES were utilized to analyze elemental profile of 16 elements (Al, Br, Ca, Cr, Cu, Fe, K, Mg, Mn, Na, Ni, P, Rb, Sc, Sr, and Zn) for further classification of Jamaican vs. other origins. Brewed coffee shows different elemental profile compared to roasted or powdered coffee. Interclassification of several Jamaican origins was not able to be done using this method.
(Hamdouche et al., 2016)	Food Control	Microbial community by PCR-DGGE	2 classes of geographical origin	Cameroon (Bafoussam and Dschang)	PCA and Ascending Hierarchical Classification (AHC). Not mentioning the validation method.	Important feature microbial DNA markers were selected based on the existence of the microorganism in samples, which is shown by AHC dendrogram.	Dry and wet process	Coffee origin cannot be discriminated based on its bacterial or fungal community. The diversity of microbial community is more closely associated with the postharvest process, i.e., wet or dry method.

(Mehari et al., 2016b)	J. Food Compos. Anal.	Phenolic compounds by UPLC-DAD and QTOF MS	4 classes (100 samples)	Ethiopia (East, Northwest, South, West)	PCA and LDA. PCA Model performance was evaluated by R^2 (0.75) and Q^2 (0.50) values. LDA model was used for origin prediction using the LOOCV. LDA recognition and prediction abilities of 91% and 90%, respectively, at regional level.	Important features selected by 3 consecutive steps, i.e., PCA loading score plot, significance different in ANOVA, and boxplot.	N/A	Polyphenolic compounds were found to be discriminative for classification of green. 3-O-caffeoylquinic; 3,5-O-dicaffeoylquinic; 4,5-O-dicaffeoylquinic acids; and 3,4-O-dicaffeoylquinic acid are discriminant compounds for Ethiopian coffee from several location. Several coffees can also be classified using the ratio comparison of the important markers.
(Ferreira et al., 2016)	Food Chem.	DNA marker - Real-time PCR	3 classes (based on quality)	South America, Central America, and Asia grouped into 3 quality levels: Gourmet (highest), traditional ground roast coffee, and traditional soluble coffee.	Comparison of the mean Ct and quantification of % adulterant in coffee mixture.	N/A	Ground roast and soluble instant coffee	Detection of commercial coffee DNA adulterated with corn, barley, rice.
(Mehari et al., 2016a)	Food Anal. Methods	Alkaloids composition– HPLC	4 classes (99 total samples)	Ethiopia (from several regions, i.e., East, Northwest, West, South)	LDA model calibrated by LOOCV. Dataset was divided into 75% training and 25% validation set.	N/A	Green beans	LDA model showed moderate prediction and classification ability based on the 4 selected alkaloids compounds (caffeine, theobromine, theophylline and trigonelline). PCA model of the samples showed overlap score plot between several samples from different origins.
(de Moura Ribeiro et al., 2017)	J. Food Compos. Anal.	^{1}H-NMR fingerprint	2 global classes (total 31 samples)	Brazil (Arabica blend) and Adulterants (Corn, Soy, Barley, and coffee husks)	PCA without clear information on calibration and validation steps.	Loading plot observation.	N/A	Important spectral was found in the region from 5.1 to 9.5 ppm. Most of the pure coffee samples form a group cluster in the PCA model. However, several pure coffee samples overlaps with adulterants (barley, corn and soybean), and also some other samples clustered with coffee husks in the PCA score plot.
(Gamboa-Becerra et al., 2017)	Food Anal. Methods	Direct-injection electrospray (DIESI) - MS and low-temperature plasma ionization (LTP) - MS	2 classes of species; 6 classes of geographical origin; 2 classes of postharvest process (100 total samples)	Mexico (Chiapas, Veracruz, Puebla, Oaxaca) Vietnamese coffee, and unknown origin.	Random Forest with 500 decision trees, dataset divided into training (70%), validation (15%), and test (15%).	Mean decreased Gini and mean decreased accuracy.	Dry and wet postharvest process. Roasted and Lyophilized coffee.	DIESI-MS detected several important compounds for the differentiation of geographical origin: m/z 60.473 (trimethylamine), m/z 91.227 (lactic acid), m/z 115.176 (glycolic acid), m/z 265.168 (pentadecanoic acid) and ions m/z 681.361 and m/z 758.574. LTP-MS is a potential method for rapid classification of coffee from different origin, since it needs negligible sample preparation.

(Combes et al., 2018)	Food Control	DNA-based authentication - High-Resolution Melting–PCR	7 classes	Robusta (Cameroon and Vietnam) and Arabica (India, Brazil and Guatemala). Adulteration admixture of Robusta on Arabica were 1%, 5%, 10%, 25%, and 50%.	Adulteration was detected by comparing the Genotypes Confidence Percentage using pair-wise comparison followed by Wilcoxon Rank Test to indicate significant different.	N/A	Comparing the green and roasted coffee.	Classification of adulterated coffee was achieved using HRM analysis employing several primer pairs, especially for the nonroasted green bean coffee. Detection of Robusta adulteration in an Arabica coffee was up to 1% limit. Authentication of roasted coffee is problematic due to the DNA damage during the roasting.
(Sezer et al., 2018)	Food Chem.	Laser-Induced Breakdown Spectroscopy (LIBS) for fingerprinting and ICP–OES for elemental analysis.	7 classes (total 58 samples)	Coffee from 21 different countries ensembled as one group for authentic coffee samples. Adulterations were corn, chickpea, wheat, and durum wheat	PCA was used for discrimination of pure coffee against adulterated samples. Samples divided into training (2/3) and validation (1/3) dataset. PLS was used to predict the adulteration percentages.	Significant features were evaluated based on spectral fingerprint.	Roasted coffee	LIBS spectral fingerprint analysis, as a simple method for elemental profile showed potential identification of adulteration in coffee. LIBS emission lines consist of organic molecules (C, N, O, and H) and inorganic compounds (Ca, Mg, Na, K, P, Fe and Cu). The most discriminative elements for coffee differentiation against the adulterated were K, P, Mg and Ca.
(Monteiro et al., 2018)	Food Control	Volatile organic compounds (PTR-MS) and nondestructive (NIRS)	2 classes of cultivation methods, 4 classes of geographical origin. (Total 45 samples)	Brazil (Organic and conventional); (Regional: Minas Gerais, São Paulo, Paraná, Espírito Santo, and Bahia)	PLS-DA, SIMCA, k-NN, PCA-kNN, LDA-kNN, SVM, LDA-SVM. Validation was done in two steps, i.e., cross validation and splitting dataset into training (70%) and validation (30%).	Feature selection algorithm derived from each model for classification.	Comparing coffee from the organic and conventional farming system.	Overall model performance of PTR-MS and NIRS data were comparable. Both methods reached 85% accuracy for predicting the organic or conventional farming. Among several classification model, PLS-DA and LDA-kNN achieved the best performance. PTR-MS data coupled to PLS-DA showed classification rates 69% higher than of NIRS 61% for the prediction of geographical origin of 5 regions in Brazil in the calibration dataset.

(Hoyos Ossa et al., 2018)	Food Chem.	Untargeted metabolomics–UHPLC-QToF HRMS	5 classes (41 total samples)	Colombia (Naranjal, Rosario, Gigante, Sirena, and San Antonio)	PLS-DA and DD-SIMCA. Data was cross validated for PLS-DA and acceptance plot was used in DD-SIMCA model. Validation of the model was done by using external dataset. 94% of sample was properly classified.	Filtering important feature was done by selecting the first top 50 variables from the VIP score coupled with the VIP score > 1.	Coffee from different harvest year was used for validation.	The findings indicate that UHPLC-QToF MS-based metabolomics is an appropriate method for developing green bean coffee origin discrimination techniques. Based on MS^2 experiment, 13 biomarkers were discovered (8 of which were tentatively elucidated, i.e., 1-O-Sinapoylglucose, 3-Hydroxysuberic acid, N-Acetyl-L-Phenylalanine, 5-Caffeoyl-Methylquinic acid, Caffeoyl alcohol, 5-Caffeoylquinic acid, 5-Caffeoyl-Methylquinic acid, Palmitic acid).
(Toci et al., 2018)	Food Sci. Biotechnol.	^{1}H-NMR fingerprint	4 classes (19 total samples)	Brazil (Minas Gerais, Bahia, São Paulo, and Paraná)	PCA-DA with cross validation.	Discriminant analysis using the most influential variables from the NMR data.	Roasted coffee	Important compounds for classification identified were trigonelline, formic acid, caffeine, n-methylpyridine7 CQAs, catechol. The classification performance of DA was cross validated resulting in 33.3% up to 66.7% correct origin prediction.
(Caporaso et al., 2018)	Food Res. Int.	Volatile aroma compounds–SPME–GC/MS	4 classes of continents; 2 classes of variety (250 total samples)	Brazil, Colombia, Costa Rica, Ethiopia, Guatemala, Honduras, India, Kenya, Mexico, Nicaragua, Rwanda, Uganda, and Vietnam.	LDA and Multiple Layer Perceptron (MLP). LDA was cross validated. MLP as a neural network-based algorithm segmented the training sample set as 90% and test set as 10%.	MLP used neural network feature extractor. LDA used algorithm which is derived from the parent model.	Postharvest process, i.e., wet and dry methods.	Single-origin coffee volatile compounds data can be employed for developing a reliable classification models. LDA model indicates 95.97% correct classification for geographical origin of coffee samples, and up to 82.3% to differentiate their postharvest process. The alternative MLP algorithm model performance was comparable to the LDA accuracy. Pyrazines compounds, 3-ethylpyridine, acetoxyacetone, guaiacol, ethylpyrazine, and 2-furanmethanol were volatiles that are effective for coffee geographical origin discrimination.
(Mohammed et al., 2019)	Microchem. J.	Elemental analysis ICP–OES	2 global classes (27 total samples)	Yemeni (16 regions) and Ethiopia (11 regions)	PCA and HAC without obvious information on validation method.	Important features were selected based on the PCA loading plot.	Green coffee beans	Yemeni coffee is strongly characterized by Ca-content in comparison to Ethiopian coffee. Macro element K, Ca, and Mg were strongly associated with Yemeni origin, while Na is strongly associated with Ethiopian coffee.
(Worku et al., 2019)	Food Chem.	Multielement analysis using wavelength-dispersive X-ray fluorescence spectrometry (XRF) and ICP.	4 classes (103 total samples)	Ethiopia (Harar, Southeast, Southwest, Northwest)	LDA with 10-fold cross validation.	Several features were selected as important yet there is no unambiguous information about the feature selection algorithm.	Different cropping season and harvest year.	The overall classification accuracy after cross validation for the XRF-multielements and $\delta^{13}C$ values was about 89%. XRF can be applied as a direct measurement of solid samples without digestion steps as in ICP method. ICP-based techniques and stable isotope ratios showed classification accuracy up to 80%, yet the LDA plot was not showing clear separation in 2 out of 4 of the sample classes. Incorrect classification of origin is higher in ICP -based method compared to XRF-based.

(Schipilliti et al., 2019)	Food Anal. Methods	Carbon isotope ratio ($\delta^{13}C$) of caffeine and $\delta^{13}C$ of the whole volatile fraction using (GC-C-IRMS).	8 classes (320 total samples)	Vietnam, Brazil, Cameroon, India, El Salvador, Ethiopia, India (Monsonato), India (Robusta Monsonato)	Comprehensive Isotopic Data Evaluation (CIDE) with several criteria to obtain a robust model. Data were visualized by plotting a 2D model of the selected features.	Important features were selected based on CIDE requirements, e.g., to obtain valid 13C(y) value, the y compound must be selected among well-isolated (chromatographically pure) compounds of low sensory importance.	Comparing green and roasted coffee.	This study found the discrimination between the carbon isotope ratio of caffeine extract of green and roasted coffee from different geographical origin. CIDE approach is not applicable in green beans. Data of volatile compounds of $\delta^{13}C$ as *x*-axis using CIDE approach works well to separate the roasted coffee from several country of origins (including the coffee varieties) when coupled to the $\delta^{13}C$ of caffeine as *y*-axis.
(Mehari et al., 2019d)	J. Sci. Food Agric.	Fatty acids profiling using GC-MS	4 classes (100 total samples)	Ethiopia (Northwest, west, east, south regions)	LDA with LOOCV for building up the calibration model. Dataset divided into training (74%) and validation (25%).	Discriminant features were selected from the significance values from ANOVA followed by their respective boxplot variations.	In the subregional level, the coffee sample was harvested from different area.	Oleic, gondoic, arachidic, and stearic acids were primary features to classify the Ethiopian coffee from four regions. Based on LDA model, the overall true classification rate obtained 96% for the four regions. In the subregional level, involving 8 primary locations, the overall proportion of correct classification is 92%.
(Putri et al., 2019)	Metabolomics	Untargeted metabolomics of hydrophilic compounds–GC/MS	4 global classes (Robusta, Arabica, Green bean, Roasted bean); 7 classes of origin of Arabica coffee; 3 classes of origin of Robusta coffee. (31 total samples)	Indonesia (Temanggung, Bondowoso, Toraja, Lampung, Toraja, Mandheling, Aceh, Andungsari, Malang, Bali, Bima, Blue Flores, Bajawa, Pogapa, Kerinci, and Kaliselogiri.	PCA without indicating the validation method.	Important features were selected based on the observation of PCA loading plot.	Comparing green beans and roasted beans, as well as using the coffee from different harvest year.	The untargeted metabolomics GC/MS approach found 64 compounds indicating cluster area of origin of the 4 global classes from 16 samples. Also, it was found 53 compounds for both Arabica and Robusta classes. Glycerol, glucuno-1,5-lactone, gluconic acid, and sorbitol were among the metabolites with higher concentrations in the coffee from Sulawesi, Papua, Flores, and Sumatra than in any of the other samples studied. Galactitol and galactinol concentrations differed significantly across samples from the eastern and western parts of Indonesia.

(Monteiro et al., 2019)	J. Food Sci.	Physicochemical properties and antioxidant activity coupled to chemometrics approach	2 classes of cultivation method; 8 classes of geographical origin (45 total samples)	Brazil (Organic and conventional cultivated coffee); (Minas Gerais, São Paulo, Paraná, Espírito Santo, and the blends)	PLS-DA, LDA, DD-SIMCA, SVM, and k-NN. Samples were cross validated and then divided into training (75%) and validation (25%) for PLS-DA and LDA analysis.	LDA–stepwise algorithm.	Roasted coffee	Using the chemical composition and antioxidant activity values, all model/algorithms showed remarkable performance for classifying the coffee cultivation methods (conventional and organic) of Brazilian coffee. However, the specific producing region was not classified properly. Most discriminative parameters were caffeine, quercetin-3-rutinoside, the Folin–Ciocalteu reducing capacity, total soluble solids, and antioxidant capacity (DPPH), especially for separating the single-origin Arabica against the blends.
(Peng et al., 2019)	Food Chem.	Stable isotope–IRMS	2 classes of cultivation system; 6 classes of origin (67 total samples)	Brazil (organic and conventional cultivation system) and (São Paulo, Minas Gerais, Paraná, Espírito Santo, Bahia, their blends.	LDA, kNN, SVM. Data were cross validation 20 times. Training dataset uses 70%, test dataset uses 30% of the samples. ROC curve was used to evaluate SVM model. The Chebyshev distance metric was used for kNN classification.	N/A	Roasted beans	From four isotopes ratio$\delta^{13}C$, $\delta^{18}O$, $\delta^{2}H$, and $\delta^{15}N$ values, the organic and conventional coffee beans showed no significant different, except in $\delta^{15}N$ ratio. All classification algorithms were not satisfactory for determining the coffee geographical origin using the four isotopes ratio–IRMS based. The maximum satisfactory of model was achieved by SVM with accuracy for São Paulo origin (75%), and LDA accuracy was 71% for Minas Gerais origin.
(Badmos et al., 2019)	Food Res. Int.	Identification of 16-O-methylcafestol by NMR and identification of CGA profile by LC-ESI-HRMS	2 classes (54 total samples)	Arabica (Brazil, Colombia, Nicaragua, Ethiopia, Honduras, Guatemala, Jamaica, Costa Rica, Rwanda, Laos, Kenya, China, Peru, and Papua New Guinea) and Robusta (Robusta blend sample from India, Vietnam, Ecuador, Uganda, Tanzania, and Indonesia) varieties.	LDA with validation dataset of (3 Arabica and 3 Robusta samples).	The discriminant features were selected from the LDA model supported with PCA.	Green beans	16-O-methylcafestol identified in Robusta yet it was not found in Arabica. In addition, this study found that the 16-O-methylcafestol in Arabica and Robusta blends were only detected in the threshold of < 40% Robusta mixture. Feruloyl quinic acids and several di-O-caffeoylquinic acids showed accuracy to discriminate Robusta against Arabica. This study found 15 important CGAs compounds that are discriminative for Arabica vs. Robusta green beans from 20 different countries and continents.

(Ongo et al., 2020)	Food Res. Int.	Volatile compounds–SPME–GC/MS	8 classes (8 total samples)	Philippines (Arabica civet and noncivet coffee); (Robusta civet and noncivet)	PCA and cluster analysis (dendrogram) without clear information on validation method	Observation of PCA loading score.	Civet and noncivet coffee	Several discriminative volatiles metabolites responsible for the classification of civet and noncivet of both Arabica and Robusta coffees were acetic acid, furfural, 5-methylfurfural, 2-formylpyrrole, maltol, phenol and 4-ethyl-guaiacol. Overall, using the volatile compounds profile of the coffee samples, this study was able to discriminate the origin and postharvest process of coffee from the Philippines.
(Bitter et al., 2020)	Food Chem.	Trace elements–ICP-MS	(53 total samples)	21 countries (Brazil, Burundi, Colombia, Costa Rica, El Salvador, Ethiopia, Guatemala, Hawaii, Hawaii–Kona, Honduras, India, Indonesia, Kenya, Mexico, Nicaragua, Panama, PNG, Peru, Rwanda, Tanzania, Vietnam, Yemen)	Biplot projection of certain elements from different origin. Classification employs “One vs. Others” concept.	Comparing the combination of element ratio abundance from one country to the others.	Roasted beans	Classification of coffee origins from certain origin may be easier to be conducted than others using this technical approach. The ration of Mn, Fe, and Rb were most frequently useful for origin discrimination. In country Brazilian sample was able to be discriminated by using ratios of Ce/Dy and Ce/Nd. To distinguish Brazilian origin against other countries ratios of Ce/Yb and La/Er. Guatemalan coffee can be classified by biplot of Mn/Sr and U/Yb against the samples from other countries. This study use specific elements ratio biplot to distinguish coffee origins, yet not all origins were classified in this study.
(Núñez et al., 2020)	Foods	Untargeted fingerprinting–HPLC-UV detector	5 classes for origins, varieties, and roasting degree; 2 classes for nearer distance of origins. (306 total samples)	Samples obtained as a commercial Nespresso® coffee, i.e., Ethiopia, Brazil, Central–South America, India, Uganda, Colombia, India, Nicaragua, Indonesia, and the blends. As well as Vietnam and Cambodia from different market.	PLS-DA for geographical origin classification and PLSR for adulteration detection. Data were cross validated (venetian blind). Data were split into training (70%) and validation (30%) set.	N/A	Roasted coffee beans with various darkness levels, but not clearly mentioning the temperature of roasting.	PLS-DA model showed samples discrimination for the known geographical origin, except the Central–South American coffee which is overlapped with other data plot. Separation of pure Robusta, Arabica, and their blends were observed in PLS-DA model. the separation was also observed in the PLS-DA score plots of coffee with different roasting degrees. Adulterated coffee with mixture of different origins and composition of Arabica–Robusta, using this method PLSR showed good performance for predicting the percentage of adulteration.
(Enday e et al., 2020)	Biol. Trace Elem. Res.	Elemental analysis - ICP–OES	4 classes (120 total samples)	Ethiopia ((West Gojjam, East Gojjam, Awi, and Bahir Dar Especial Zones)	PCA with cross validation of the model and LDA with dataset split into training (80 samples) and validation (40 samples) for prediction.	Features were selected based on PCA loading plot.	Green coffee beans	The most discriminative and important elements were Ca, Mg, K, Na, in PCA model. LDA provides classification model with an overall 94.2% accuracy and 93.4% prediction ability of the production zone of the coffee samples.

(Badmos et al., 2020)	Food Res. Int.	Chlorogenic Acids profile - HPLC-ESI-TOF-MS	7 classes of origin; 3 classes of farming system (67 total samples)	Brazil. Farming system (biolodynamic, organic, conventional); Origin (São Paulo, Minas Gerais, Espírito Santo, Bahia, Paraná.	PLS-DA without any information about validation method	VIP score derived from PLS-DA	Roasted beans	Several CGA compounds were significantly characterizing the differences in the farming system, i.e., 5-pCoQA, 5-CQA, and 4-CQA. The determination of geographical origin of Brazilian coffee grown in different location was not possible employing this methodology based on the PLS-DA model.
(Happyana et al., 2020b)	Indones. J. Chem.	^{1}H-NMR fingerprint	2 classes (12 total samples)	Indonesia (Aceh and Lampung)	OPLS-DA with cross validation and permutation for 200 times.	S-Plot derived from the OPLS-DA model algorithm.	Coffee from the same origin yet provided by different supplier	Quinic acid is the main compound to discriminate coffee from Lampung. Lipid is the most important compound for Aceh coffee.
(Happyana et al., 2020a)	Curr. Res. Nutr. Food Sci.	^{1}H-NMR fingerprint	4 classes (24 total samples)	Indonesia. (Gayo–Sumatra, Preanger–Java, Bajawa–Flores, and Toraja–Sulawasi)	PLS-DA and OPLS-DA with cross validation and 200 permutations. Model performance evaluated by R^2X and Q^2Y.	S-plot derived from the OPLS-DA model algorithm.	Roasted beans	The distinctive metabolites of Gayo–Sumatra coffee were discovered to be lipids, acetic acid, and lactic acid. Quinide was discovered to be the most significant marker for Bajawa–Flores coffee. Meanwhile, Toraja–Sulawesi coffee has a balanced chemical makeup, indicating a well-balanced taste.
(Amalia et al., 2021)	Metabolomics	Metabolite profile–GC/MS	6 classes of postharvest process (Dry, Honey, Washed coupled to 2 different altitudes); 4 classes of altitudes. (23 total samples)	Indonesia. Bajulmati, Kalibendo, Aceh Gayo, Andungsari, Bodowoso, Manglayang, Flores.	OPLS-DA. The data was divided into training and validation. RMSE Estimate and RMSECV were used for model evaluation.	VIP score derived from OPLS-DA algorithm	Roasted beans, Green beans with different postharvest process, altitude	Postharvest process was suggested to be the primary discrimination of coffee metabolite profile, followed by geographical origin and altitude of plantation. Glutamic acid and galactinol were important metabolites for washed and honey process. Glycine, lysine, sorbose, fructose, glyceric acid, and glycolic acid were important for dry process coffee.
(Bosmali et al., 2021)	LWT	Internal Transcribed Spacer region 2 (ITS2)-based marker–DNA Barcoding High-Resolution Melting (HRM) analysis.	2 global classes for species discrimination (18 total samples)	Thailand and commercial coffee from Greek market	Confidence value (%), i.e., square root of correlation coefficient (R). .	DNA marker species specific is decided.	Roasted coffee and brewed coffee in several methods.	Based on HRM analysis, the melting curve profiles are only associated to the coffee species content regardless the coffee brewing methods. This study claims to be the first species-specific HRM analysis showed satisfactory results for the direct authentication of the brewed product.

(Mehari et al., 2021)	Int. J. Food. Prop.	The total polyphenol content - Folin–Ciocalteu reagent. Total Flavonoid - Spectrophotometry	4 classes (100 total samples)	Ethiopia (Northwestern, Southern, Western, and Eastern regions)	Statistical comparison ANOVA.	The significant different of mean concentration of the compounds.	Green beans	It has been shown that the polyphenol content of green Arabica coffee beans varies depending on their geographical origins. The majority of polyphenols were chlorogenic acids, while flavonoids were only present in minimal amounts.
(Zhu et al., 2021)	LWT	Protein, lipid, sucrose, total phenolics (TPC), and total titratable acidity as well as fatty acid profile–GC-FID	8 classes (50 total samples)	Brazil, Colombia, Ethiopia, Guatemala, Honduras, Indonesia, Kenya, China	PLS-DA. The goodness of fit of the model was $R^2X = 0.568$, $R^2Y = 0.27$	VIP score from PLS-DA algorithm.	Green beans from different harvest years.	Samples from Kenya and Ethiopia were overlapped and cannot be separated in the PLS-DA model. Important variables in this study were total lipid content, C24:0, C22:0, C18:3, C17:0, C18:0, C20:0, C16:0, protein, C18:1, and C18:2.
(Hung et al., 2021)	J. Food Process. Preserv.	Fatty acid–Gas Chromatography	4 classes (74 total samples)	Vietnam, Brazil, Colombia, El Salvador, Fros, Malan, and Mandheling	Cluster analysis, LDA, and Neural Network model. No clear information about the model validation.	Statistical significant and coefficient of variance of each compounds.	Green and roasted beans	Fatty acid species of C18:1, C18:2, and C18:3 were strong compounds for coffee classification of both roasted and green to differentiate the Arabica and Robusta type.
(Núñez, Saurina, et al., 2021)	Food Control	Untargeted fingerprinting–HPLC-FLD	11 classes of adulterated coffee.	Colombia, Ethiopia, India, Indonesia, Nicaragua, Vietnam and Cambodia.	PLS-DA with cross validation and dividing data into calibration and prediction groups.	Observation of the signal profile of HPLC.	Roasted beans.	The wavelength for excitation 310 nm and emission 410 nm were satisfactory to be used as fingerprint descriptor for further determination of adulteration in coffee (origin admixtures).
(Abdel wareth et al., 2021)	Food Chem.	Volatile aroma compounds (HS-SPME–GC/MS)	10 classes (10 samples)	Brazil (Minas Gerais) and commercial coffee from the market	HCA and OPLS-DA using cross validation. Model performance was evaluated by Q^2, R^2, and *p*-values	Score-plot of OPLS-DA	Several roasting levels were compared	Robusta coffee showed strong volatile markers (pyrazines, furans, and aromatic hydrocarbons compounds) which were found to be less abundant in Arabica. Several volatile compounds were effective to discriminate the brewed coffee from different origins, such as, eugenol for Brazilian roasted Robusta coffee decoction method against the Arabica type, also terpinyl acetate and octyl acetate for several coffee brews.
(Núñez, Martínez, et al., 2021)	J. Sci. Food Agric.	Untargeted fingerprinting - HPLC-FLD	6 classes of country of origins; 2 classes of coffee types, 4 classes of roasting degree (186 total samples)	Colombia, Ethiopia, India, Indonesia, Nicaragua, Vietnam and Cambodia.	PLS-DA with cross validation (venetian blind) and dividing data into calibration (70%) and prediction (30%) groups.	The important features were the range of chromatographic windows, observed from the loading plot of PLS-DA.	Roasted beans with several levels of darkness.	Several important chromatogram ranges for classification of origin were, from 2 to 4.5 min, 8–27 min and 36.5–38 min segments. Also, 2–5 min and 8–27 min segments were important specifically for nearer distance origin (Cambodia and Vietnam). Overall classification rate was 100% for both calibration and prediction steps.

(Demianová et al., 2022)	Food Control	Soluble compound characters and Volatile compounds (SPME–GC/MS)	3 global classes (continents) (23 total samples)	Africa, Central America, South America	LDA with LOOCV followed by model application on the prediction dataset. Confusion matrix of CV probabilities was used to evaluate model performance.	N/A	Green beans	LDA model showed promising results for classifying African, South American, and Central American coffees based on both volatile and chemical compounds. The confusion matrix showed that 71%–100% correct classification based on volatile profile, slightly higher than the classification using the soluble chemical compounds profile.
(Miao et al., 2022)	Food bioscience	Untargeted fingerprint - UHPLC-QE-MS	5 global classes of origins and 3 classes of cultivars (18 total samples)	Asia (China: 6 origins, Indonesia: 1 origin), North America: 4 origins, South America: 2 origins, Africa: 4 origins, Oceania: 1 origin)	OPLS-DA without information about model validation. Model was evaluated based on R^2X, R^2Y, and Q^2	S-plot and VIP score derived from OPLS-DA	Green beans	Untargeted fingerprint was able to discriminate coffee samples based on the cultivars with 15 potential markers, i.e., 3-hydroxycoumarin, quinic acid, 4,5-di-o-caffeoylquinic acid, cryptochlorogenic acid (4-O-caffeoylquinic acid), palmitic amide, linoleamide, arachidic acid, 16-methylheptadeca- noic acid, ethyl oleate, 13S-hydroxyoctadecadienoic acid, petroselinic acid, 8,9-DiHETrE, L-malic acid, trehalose, L-glutamic acid. Additionally, 10 compounds were important for continent-based classification, i.e., 3-hydroxycoumarin, 4,5-di-O-caffeoylquinic, cryptochlorogenic acid, palmitic amide, linoleamide, arachidic acid, petroselinic acid, trehalose, L-glutamic acid, L-malic acid.

Other authentication method employed various techniques. For instance, photon activation analysis (PAA) using a radioanalytical method in the elemental analysis was employed for classifying three South American coffee beans and for distinguishing washed and natural process coffee (Sun et al., 2013). Proton-Transfer-Reaction (PTR-ToF-MS) showed different volatile compound profiles in coffee from Brazil, Ethiopia, and Guatemala (Yener et al., 2014). High-Resolution Continuum Source Atomic Absorption Spectrometry could classify the espresso extracted coffee based on its elemental profile (Oliveira et al., 2015). Other studies applied direct-injection electrospray-MS for fingerprinting and low-temperature plasma ionization-MS for rapid analysis (Gamboa-Becerra et al., 2017). The carbon isotope ratio ($\delta^{13}C$) of caffeine and that of the whole volatile fraction have been analyzed using GC-Carbon Isotope Ratio-MS (Schipilliti et al., 2019). Laser-Induced Breakdown Spectroscopy (LIBS), a new technique to detect and quantify coffee adulterants (chickpeas, maize, and wheat), could identify < 0.6% adulterations in coffee (Sezer et al., 2018).

A few studies used a DNA-based approach, Polymerase Chain Reaction (PCR)−Denaturing Gradient Gel Electrophoresis (DGGE) to understand the microbial community existing in coffee from different origins and processing. The study found that geographical origin has little effect on microbial diversity (Hamdouche et al., 2016). In addition, Ferreira et al. (2016) used DNA markers of adulterants such as corn, barley, and rice to quantify the percentage of noncoffee contents by Real-time PCR. Combes et al. (2018) employed high-resolution melting (HRM)–PCR to identify adulterated coffee in both green and roasted beans, resulting in a 1% threshold for adulterants content detection. Recently, the HRM method was applied to brewed Thailand samples,

showing promising results for the detection of Arabica–Robusta admixtures in brewed coffee (Bosmali et al., 2021).

Interestingly, a couple of studies performed untargeted fingerprinting analysis using HPLC without MS. High Performance Liquid Chromatography (HPLC)-UV fingerprinting was used to classify coffee samples from several countries and continents with varied roasting levels (Núñez et al., 2020). The same group used HPLC-Fluorescence Detection (FLD) to identify an admixture of coffee from different origins (Núñez, Saurina, et al., 2021), and classify coffee origin based on countries, variety (Arabica and Robusta), and roasting degree (Núñez, Martínez, et al., 2021). HPLC-FLD achieved a richer chromatogram fingerprint than HPLC-UV.

With respect to analytical approaches, several studies used specific chemical compounds such as antioxidant compounds (Monteiro et al., 2019), total phenolic compounds, protein, and total lipids (Zhu et al., 2021) as variables for coffee origin classification. However, these studies did not find a significant origin classification when data were modeled with multivariate analysis, as confirmed by Alnsour et al. (2022).

Generally, various analytical methods used to determine the geographical origin of coffee from numerous countries, establish authentication methods, and detect adulteration with noncoffee materials or addition of lower value substances. Overall, NMR is a high-throughput analytical device. This method requires little sample preparation, and can separate substances based on their NMR fingerprint. In comparison to MS-based approaches, NMR has limited sensitivity. Given its effectiveness and superior separation capabilities, numerous coffee research uses chromatography coupled to MS which has been proven for its reproducibility.

Moreover, high-resolution MS offers accurate mass measurements and can lead to the prediction of empirical formulas for unidentified compounds. This tool is widely employed when coupled with GC, and supported by fragmentation libraries (commercial and open-source) that assist metabolite identification. However, the method requires chemical derivatization and cannot be used for larger, nonvolatile substances. New HPLC methods have increased its separation efficiency, and together with MS permits the identification of substances without chemical derivatization. Additionally, the automated sampling facilitates the daily assessment of several samples. In terms of data comparison, it is difficult because of uniform ionization and fragmentation.

2.7. Nondestructive Approaches

The 23 selected articles on this topic are shown in Table 5. During the last 10 years, numerous nondestructive approaches for coffee origin determination, authentication, and variety assessment of the raw green, roasted, and brewed coffee samples have been extensively done. Most of the studies used spectroscopy-based methods. Spectroscopy is a fast-growing technique due to its speed, simplicity, safety, and ability to examine several characteristics simultaneously without requiring lengthy sample preparation (Barbin et al., 2014). Particularly, spectroscopic procedures in the visible, near, and midinfrared regions are rapid, almost chemical-free, inexpensive, and sample-processing-free techniques widely used to predict the chemical composition of coffee, making them suitable for routine application.

Several studies on this topic use single equipment, but most applied combined approaches to achieve their goals. Infrared (IR)-based technology is the most employed

method in nondestructive analysis. It is categorized into three regions, i.e., near IR (NIR) from 0.77 to 2.5 μm (corresponds to wavenumber = 13000–4000 cm^{-1}); mid-IR(MIR) from 2.5 to 15 μm (4000–400 cm^{-1}), and far IR >25 μm (<400 cm^{-1}) (Wang et al., 2020). This technique can create a spectral "fingerprint" of coffee samples by direct measurement. However, for coffee authentication or origin determination studies, only a certain range of wavelengths is meaningful. For instance, a study using FTIR coupled to a specific detector selected wavelengths ranging between 600 and 1000 cm^{-1} for differentiation of two cultivation systems of roasted coffee (organic and conventional) (Gordillo-Delgado et al., 2012), while other studies using MIR employed spectral features between 2970 and 3600 cm^{-1} to classify Robusta and Arabica coffee (Assis et al., 2018).

The studies listed in Table 5 used NIR for numerous purposes, e.g., differentiation of modern and traditional coffee cultivars from Brazil (Scholz et al., 2014), regional classification of Brazilian coffee samples (Marquetti et al., 2016), origin determination of coffee from Cuba, Ethiopia, Indonesia (Bali, Java, and Sumatra), Tanzania, and Yemen (Mendez et al., 2022) comparison between South American and Asian coffee (Giraudo et al., 2019), and to detect impurities (Corn, Rice, Barley, Soybeans, Coffee husks) in Arabica roasted coffee samples as well as to separate South and Central American coffee samples (de Carvalho Couto et al., 2021). All the above studies used >100 total samples. The number of samples is probably one of the good parameters for obtaining robust classification in NIR approaches.

Table 5: List of studies employing the nondestructive approach

Ref	Publication	Technique	Number of class origin/number of sample	Countries of coffee origin	Classification model and performance evaluation	Availability of feature selection method	Associated information in coffee processing (postharvest, roasting, etc)	Important finding
(Ongo et al., 2012)	Procedia Engineering	E-nose and GC/MS	8 classes	Philippines (Kalinga, Asipulo, Cordillera, and South Cotabato)	PCA and dendrogram without information about model validation method.	NA	Civet coffee and normal (control) coffee	The volatile chemical composition of civet coffee is nearly identical to that of its controls, but at differing amounts. E-nose data showed clear separation on PCA plot and dendrogram for civet and noncivet coffee samples. The finding was supported by GC-MS headspace analysis.
(Gordillo-Delgado et al., 2012)	J. Sci Food Agric.	FTIR-photo-acoustic spectrometry	2 classes (Organic and conventional); (60 total samples)	Colombia *C. Arabica* ("Caturra," "Castilla," and "Tipica")	PCA without information about model validation method.	PC scores from the PCA model	Roasted coffee	The method can be used to indicate the cultivation system difference. Spectra in the range between 600 and 1000 cm^{-1} were used for differentiation of the two cultivation systems (organic and conventional)
(Domínguez et al., 2014)	Sensors	Voltametric sensor - Electronic Tongue (ET)	3 classes (42 total samples)	Mexico	LDA and SVM evaluated by Sensitivity and Specificity. Prediction accuracy: 87.5% (LDA); 97.5% (SVM).	NA	Organic and nonorganic, and the growing altitude of the plantation.	ET can be employed in a simple and rapid way to distinguish samples from different origin and cultivation practice (organic vs. conventional) as well as altitude of the plantation of the Mexican coffee.
(Scholz et al., 2014)	J. Near-Infrared Spectrosc.	Nondestructive - NIR	3 classes (N samples = 254–427, selected randomly)	Ethiopian coffee accession grown in Brazil and several modern and traditional cultivars from Brazil.	Modified PLS using cross validation	No variable selection	N/A	This study found the ability of NIRs for predicting several quality parameters (Caffeine, CGA, Sucrose, total sugars, protein, lipids, phenolic compounds) and further to determine the cultivars of coffee.
(Lopetcharat et al., 2016)	J. Food Eng.	Cyclic voltametric electronic tongue (ET)	8 classes (no information about the number of total samples)	Thailand (Doitung, Doichang, Chiang Rai, Loei), Indonesia, and Vietnam	PCA (no information about the model validation)	N/A	Civet coffee, weasel coffee, and wet process coffee	ET can be used to differentiate coffee from different country of origin as well as processing technique (civet vs. weasel vs. wet-processed coffee)

(Medina et al., 2017)	Int. J. Anal. Chem.	Attenuated Total Reflectance Mid-Infrared (ATR-mIR), Near Infrared (NIR), and ^{1}H-NMR	2 global classes (total 97 samples)	Colombia against other origins (Arabica group: Guatemala, Peru, Brazil, Costa Rica, and Panama. Robusta group: Vietnam, India, Uganda, Indonesia, Togo, Tanzania, Ivory Coast, and Cameroon)	PCA evaluated by Receiver Operating Characteristics (ROC)–Area Under the Curve and Q^2 values.	N/A	N/A	Three methodologies to discriminate Colombian against other coffee origins were compared. ATR-mIR indicated better classification of the coffee species and country of origin, compared to NIR and ^{1}H-NMR. oPLS-DA model was the best model for classifying ATR-mIR and ^{1}H-NMR, while PLS-DA was outstanding for NIR data.
(Knysak, 2017)	Food Sci. Technol.	Electronic nose with Ultra-Fast GC- Flame Ionization Detectors (FID)	4 classes (not clearly mentions the number of sample)	Nepal, Colombia, Vietnam, Uganda	PCA without clear information on calibration and validation steps.	N/A	N/A	Electronic nose coupled to GC-FID proven to be useful in classification of coffee from several countries. Each coffee origin identified with distinct volatile compounds.
(Bona et al., 2017)	LWT	(Non-destructive) NIRS and FTIR	4 classes (74 samples)	Brazil (Paranavaí, Cornelio Procopio, Mandaguari, Londrina)	SVM with 10-fold cross validation. Dataset was divided into training (2/3) and test (1/3) group.	Stable feature selected as support vector in SVM algorithm.	Different harvest year and genotypes.	NIRS performs better accuracy for geographical origin classification compared to FTIR. NIRS method coupled with SVM model reach 100% specificity and sensitivity in AUROC algorithm.
(Botelho et al., 2017)	Food Control	Fluorescene spectroscopy	4 classes (110 total samples)	Brazil (Cerrado Mineiro, Matas de Minas, Norte de Minas, e Sul de Minas)	PARAFAC, NPLS-DA, UPLS-DA. f-scores was used to indicate accuracy of classification. Model were cross validation, followed by dividing dataset into training (2/3) and test (1/3) data.	VIP score inherited from NPLS-DA and UPLS-DA	Samples harvested in different years	Fluorescence spectroscopy was able to be used for classification of Brazilian coffee from 4 regions with a simple sample preparation. UPLS-DA model was the most accurate compared to PARAFAC and NPLS-DA for classification based on f-score value. UPLS-DA outperformed with f-score range from 0.82 to 1.00 in training dataset.
(Flambeau et al., 2017)	Food Sci Biotechnol	Nondestructive e-nose and e-tongue	10 classes (growing area) (50 total samples)	Rwanda (Gekenke, Rulindo, Kamonyi 1 and 2, Rwamagana, Rutsiro, Rusizi, Rubavu) Ethiopia (Yirgacheffe), Brazil (Cerrado)	PCA and Discriminant Factorial Analysis. Model validation was not clearly indicated in the article.	Important features were selected by observing the peaks with discrimination power ≥95%.	Two different roasting temperatures was applied using the same coffee origin for sample references.	E-nose peak data was able to be used for bourbon Rwanda coffee discrimination against the non-Rwanda. Several Bourbon samples from different provinces in Rwanda cluster to each other and form two distinct groups in PCA plot. E-tongue peak data was not significant for discrimination due to similar profiles among the samples.

(Suhandy & Yulia, 2017)	Int. J. Food Sci.	Nondestructive UV-Visible Spectrophotometer	2 classes (98 total samples)	Indonesian civet coffee (*kopi luwak* Lampung) against adulterated kopi luwak.	PLSR with cross validation. Samples divided into training (58) and test (40). The Ratio Prediction to Deviation value and Range Error Ratio were used to evaluate the model performance.	Observation of the loading weight values. The larger the absolute the more important and significant the wavelength feature for classification.	Roasted coffee	A simple preparation of sample was applicable for the authenticity determination of Indonesian civet coffee against its adulterated one using UV-Visible spectrophotometer. The UV-Vis spectral data was processed by PLSR model with several preprocessing steps, the model was satisfactory for separating the original civet coffee from it counterfeit.
(Assis et al., 2018)	Food Anal. Methods	Nondestructive Midinfrared (4000–800 cm^{-1}) - ATR–FTIR spectroscopy	3 global classes (light, medium, dark roast coffee) (120 total samples)	Brazil. Arabica (State of Minas Gerais); Robusta (State of Espírito Santo)	Ordered predictors selection (OPS), interval PLS, successive projections algorithm (SPA), and Genetic Algorithm (GA). Cross validation was done for all models. Dataset were divided into calibration (70%) and validation (30) sets.	Important features were selected using the algorithm derived from each model, i.e., OPS, SPA, GA, iPLS.	Coffee roasted in light, medium, and dark mode were compared for model calibration and validation.	The best variable selection results were obtained with discrete methods, OPS and GA. The region of 3000 and 2800 cm^{-1} reported to be important information associated with hydrocarbon or carboxylic acid. Spectrum of 3600 and 2970 cm^{-1} was fit to be included in the robust model yet it was not informative for individual roasting levels. A robust model that was built based on all roasting degrees performed better prediction independent of roasting degree.
(Obeidat et al., 2018)	J. Appl. Spectrosc.	ATR–FTIR (600–4000 cm^{-1})	5 classes (48 total samples)	Brazil, Colombia, Ethiopia, Kenya, and Yemen.	PCA with LOOCV.	Important features were obtained by observing the PCA loading plot.	Green coffee beans	Important spectral visual comparison was indicated in the region 1775–1500 cm^{-1} and 3030–2750 cm^{-1} which illustrate differentiation between coffee origins.
(Makimori & Bona, 2019)	Food Anal. Methods	Nondestructive - E-nose	6 classes (53 total samples)	Brazil (commercial coffee samples)	Common Dimension Analysis (ComDim) and LDA with LOOCV.	Feature extraction was done based on the program and setting of e-nose equipment.	Different degree of roasting of instant commercial coffee without indicating the roasting method.	All models achieved sensitivity and specificity value of 100%. E-nose transient signal showed differentiation among the samples, therefore it can be used for discrimination of coffee origin when coupled with chemometrics approach. This method is applicable for aromatic quality control in instant coffee industry.
(Mendes et al., 2019)	Foods	Nondestructive - NIRS	5 classes	Cuba, Ethiopia, Indonesia (Bali, Java, and Sumatra), Tanzania, and Yemen	SIMCA with cross validation followed by building a model from the training dataset. The test dataset was used to validate the model performance.	Important features selected based on the absorption spectra after data normalization	Green beans	Using the nondestructive NIRS coupled to SIMCA model, the classification rate reaches 70%. However, several coffee samples from different origin were overlapped as showed in the score-plot PCA model. NIR spectra between 1850 and 1950 as well as 2000 and 2500 were expressing the existence of C=O and H_2O absorption.

(Giraudo et al., 2019)	Food Control	Nondestructive - FT-NIR spectrometer	2 classes of continents; 5 classes of countries (191 total samples)	America (Brazil, Honduras, Guatemala, Colombia, Costa Rica, Nicaragua); Asia (India, Vietnam, Indonesia)	PLS-DA with cross validation for calibration data. Dataset were divided into training (75%) and validation (25%). Further validation were done by conducting the same methodology in the different laboratory.	The most discriminative subset of NIR spectra were selected by interval PLS-DA algorithm.	Green coffee beans	iPLS-DA algorithm selected 90 important spectral features from the NIR data in the range of (9018–8871 cm^{-1}, 8632–8177 cm^{-1}, and 6009–5940 cm^{-1}) for country-based classification. The algorithm was selected 40 important spectra for continent-based classification. The results of prediction showed efficiency from 98% to 100% correct classification for continent cluster. Country-wise classification achieved 95.9% for training dataset, and 94.6% for validation dataset obtained from different laboratory.
(Bilge, 2020)	J. Food Sci. Technol.	Physicochemical properties, antioxidant, and spectral properties (UV-Vis and Fluorescence Spectroscopy).	3 classes of geographical origins; 3 classes of roasting methods; 2 classes of particle size	(Peru, Colombia, and Brazil)	PCA with cross validation evaluated by RMSEC for the calibration and RMSECV for the cross validation.	Derived from the PCA model.	Samples were roasted using several darkness levels, i.e., light city, city, and full city. The coffee was powdered in fine and coarse particle size. Several brewing methods were used, i.e., hot (French press and chemex methods) and cold brew methods.	Geographical origin of coffee showed no significant effect on the differences of antioxidant activity and total phenolic compounds. However, UV-Vis spectral data varies for coffee from different geographical origin. Roasting levels able to cluster the coffee samples in PCA plot, independent from the geographical origin. The fluorescence spectral data documented that the geographical origin, particle size and roasting degree had less influence on the formed molecules compared to brewing method.
(Marek et al., 2020)	Sensors	Volatile compound–E-nose and GC-MS	5 classes	Brazil, Ethiopia, Guatemala, Costa Rica, and Peru	PCA without information about validation method.	Loading plot of PCA model.	Roasted beans	The E-nose volatile compounds analysis was consistent with the results GC-MS analysis for the discrimination of coffee from the particular origin. Pyridine and 2-oxoproponal were abundant in samples from Brazil and Peru. Butan-2-one, 2-methylpirimidine, and 4.6-dimethylpyrimidine were high in other 3 samples.
(Arrieta et al., 2020)	Int. J. Technol.	Non-destructive cyclic voltammetry mini electronic tongue	5 classes	Colombia (Cauca, Risaralda, Cesar, Quindío, and Antioquia)	PCA without information about the validation method.	N/A	Coffee suspension	The nondestructive approach, the electronic tongue sensor array was able to discriminate coffee from different region of Colombia origin, sufficiently.
(Couto et al., 2022)	Foods	Nondestructive NIR	3 global classes	Coffee samples from Brazil, Honduras, Colombia, Vietnam, Cameroon. Adulterants: Corn, Rice, Barley, Soybeans, Coffee husks	PCA with cross validation. The data was divided into training and validation. RMSEC and RMSECV were used for model evaluation.	N/A	Roasted beans	NIR spectroscopy can distinguish pure Arabica coffee samples from contaminated ones, including mix of Robusta coffees or coffee husks. The identification of the adulterant in the sample was only possible for single or double adulterations at concentrations of < 10%. NIR spectroscopy also shown promise for the geographic classification of Arabica coffees (South and Central America).

(S. Yang et al., 2021)	Front. Nutr	Nondestructive Terahertz (THz) spectroscopy	3 classes (96 total samples)	Kenya, Kilimanjaro, Yunnan (China).	Convolutional Neural Network (CNN), PCA-LDA, GA-LDA, PCA-SVM, and GA-SVM. Dataset were cross validation followed by splitting samples into training and validation with various combinations.	Features were selected using the derived from the dimensional reduction algorithm, e.g., PCA and GA.	One of the samples were roasted in 3 different levels (light, medium, dark).	THz amplitudes of the samples from several origins showed differentiation. Most convincing classification results were showed by deep learning CNN model, e.g., 90.0% Accuracy, 90.5% Sensitivity, and 95% Specificity in the prediction set after feature extraction using GA.
(Mendes et al., 2022)	Curr. Res. Nutr. Food Sci.	Nondestructive - Fourier-Transform Mid-Infrared (FT-MIR)	4 classes (36 samples)	Brazil (Minas gerais from Cerrado, Matas, North, South regions)	PLS-DA with LOOCV (24 samples) and validation (12 samples).	Loading plot from the MIR spectra.	Green beans	PLS-DA model showed $R^2X = 0.892$, $R^2Y = 0.659$; $Q^2Y = 0.494$, RMSEP = 0.182387. Also, it shows 100% of both sensitivity and specificity. Coffee origin class prediction using MIR spectra obtained correct classification rate from 83% to 100%.
(Robert et al., 2022)	Food Chem.	Nondestructive Fluorescene spectroscopy	4 classes (200 total samples)	Brazil (Ariquemesm, Alta Floresta D'Oeste, Ouro Preto do Oeste, Porto Velho	LDA, QDA, RDA, MDA, SVM (linear), SVM (polynomial), SVM (radial base function), RF, GBM, ANN, k-NN, LVQ, and optimized LVQ. Data were cross validated and divided into training (75%) and validation (25%).	Spectral selection based on PCA model.	Green beans	Direct solid sample analysis was enable rapid and simple coffee origin determination using this technique. Spectral data process was done by combining the data array from the sensors and principal components of the PCA scores with subsequent fusion. Pareto data scaling improve the accuracy of the prediction. SVM polynomial model showed the highest accuracy 0.97 with both specificity and sensitivity maximum to 1 for determining the Brazilian coffee geographical origin.

Other noninvasive methods listed in our study are the combination of several tools. For instance, a study compared the effectiveness of Attenuated Total Reflectance Mid-Infrared (ATR-MIR), NIR, and ^{1}H-NMR. ATR-MIR led to better classification of coffee species and country of origin than NIR and ^{1}H-NMR to distinguish Colombian coffee from counterfeit beans (Medina et al., 2017).

Furthermore, voltametric sensor technology, particularly the electronic nose (E-nose) and electronic tongue (ET), was used in several studies. Some studies used the sensor together with other analytical approaches to confirm sensor data. For example, the E-nose and GC/MS were used to determine the origin of civet coffee (the Philippines) against regular coffee from the same region (Ongo et al., 2012) and to discriminate against coffee from Brazil, Ethiopia, Guatemala, Costa Rica, and Peru (Marek et al., 2020). As well as E-nose and GC/FID which was used in classifying coffee samples from several countries (Knysak, 2017). The E-nose and GC were used to assess the volatile aroma compounds of coffee samples. Meanwhile, several ET sensors were used to distinguish samples from different origins and cultivation practices (organic vs. conventional) as well as the altitude of t Mexican coffee plantations (Domínguez et al., 2014).

On this subtopic, one study in Table 5 used a UV-Vis spectrometer to distinguish Indonesian civet coffee (*kopi luwak Lampung*) from adulterated *kopi luwak* (Suhandy & Yulia, 2017). Recently, Terahertz spectroscopy was used to discriminate coffee samples from Kenya, Kilimanjaro, and Yunnan (China). One of the samples was roasted at three different levels (light, medium, and dark). The result of model classification was satisfactory (S. Yang et al., 2021).

Overall, coffee origin determination, authentication against contaminated samples, and in-country origin determination can be performed by simple and rapid, nondestructive approaches. Most of the studies used a lot of samples with multiple replications. Furthermore, the spectral data from these approaches were processed with multivariate analysis or employing machine learning modeling.

2.8. Multivariate Model and Data Analysis

Prior to data analysis, several aspects of data collection are essential to obtain robust classification and coffee origin determination or authentication. The number of samples and representation of biological samples from a certain geographical origin follows a rule of thumb, especially when applying a multivariate analysis or machine learning model. Using highly representative samples for creating a calibration model will be more accurate in predicting the studied class. Therefore, this review indicates not only the number of samples used but also the model performance evaluation method.

Comparing the classification model performance among studies is difficult due to technical bias, such as HPLC or GC/MS methodology or NIR setup. A study using both analytical and nondestructive approaches on the same samples is required for a valuable comparison. A study by Monteiro et al. (2018) (Table 4) may illustrate this comparison, where volatile organic compounds profiling (based on PTR-MS) and nondestructive (based on NIR) was used to classify organic vs. conventional coffee farming. The data was processed with several machine learning models using cross validation, followed by validation by an external dataset. In the model both method can be compared for the percentage of accuracy.

Spectral data from analytical platforms like NIR, MIR, NMR, chromatography, and MS, contain molecular information that can serve as fingerprints for coffee origin, species, and types. So that spectral data can be useful for coffee authentication, statistical analysis are frequently required to reduce data dimensionality, such as the identification of spectrum regions relevant to quality parameters, pattern recognition, and detecting outliers. Unsupervised exploratory approaches, including PCA, factorial analysis, Soft Independent Modeling of Class Analogy (SIMCA), and cluster analysis, are frequently employed for this purpose. The main aim of unsupervised modeling is to explore the natural sample grouping (Kotu & Deshpande, 2019). This model is not aimed at finding important variables in large data, such as in metabolomics. However, numerous studies used this model and arbitrarily selected the important compounds by observing the loading plot.

Other studies used better data modeling, such as supervised machine learning. Several algorithms can provide insights into feature selection, such as Linear Discriminant Analysis (LDA) (Mehari, et al., 2016a; Muñiz-Valencia et al., 2013; Valentin & Watling, 2013), Partial Least Square (PLS)-Discriminant Analysis (DA) (Hoyos Ossa et al., 2018; Yener et al., 2014), PLS Regression (Núñez et al., 2020), Orthogonal PLS-DA (Miao et al., 2022), [29], [32], Random Forest (RF) (Gamboa-Becerra et al., 2017), support vector machine (SVM) (Monteiro et al., 2018, 2019; Peng et al., 2019), and k-nearest neighbour (k-NN) (Monteiro et al., 2018; Peng et al., 2019). The model used in all studies is indicated in Tables 4 and 5. These algorithms have been adopted to handle metabolomic datasets in a supervised method. In the supervised model, the class of observation is already decided for the classification of the sample data. Therefore, the subjectivity of the human interference in the mathematical

formulation can be considered not the best fit for the grouping from random set of data, yet the result of the model might be as satisfactory (Nocairi et al., 2005). Moreover, model resulted in the DA can be more stable (James et al., 2013).

Table 6: General comparison of the advantages and limitations between analytical and nondestructive approaches

Method	Advantages	Limitations
Analytical (in general)	• Proper technique for elucidation of the biochemical fingerprint of food materials. • Golden standard to find the information of important biochemical compounds which drive sample discrimination. • In certain approaches, it can be interpreted immediately real time, employing simple statistical analysis.	• Labor intensive since it requires trained operator to comprehend the advanced technology. • High investment for purchasing the whole system and certain components. • Requires complex sample preparation and pretreatments.
Nondestructive (in general)	• Minimum sample preparation (e.g., does not require extraction, derivatization, and other pretreatments). • Relatively simple and rapid operation. • May accurately be implemented when the training model have been developed, and • In certain case, can be used along with the different	• Could not be used to discover the important biomarker of a product. • Need a large number of samples to create an accurate determination. • Could not be interpreted immediately to indicate the discrimination factor since the spectral patter might be similar, and thus it should be proceeded to the chemometrics

	production steps to provide real-time information.	data analysis to find the discriminative wavelength.

Another important part of data analysis is cross validation (CV). Several studies used CV to reduce bias and avoid overfitting (Tables 4 and 5), whereas others split data into calibration/training datasets and validation/test samples. Furthermore, normally a DA uses root mean square error to indicate the error level of a model, as shown in Amalia et al. (Amalia et al., 2021). Yet, other models prefer to use R^2X and Q^2Y to evaluate the goodness of fit of their model (Happyana et al., 2020b). Important features for discrimination are normally selected based on the model algorithm capacity. For example, the PLS family model uses a Variables Importance in Projection (VIP) score or S-plot to indicate the influence of different variables in the classification. An exhaustive list of models and evaluation parameters, as well as the feature selection methods can be found in Tables 4 and 5.

Table 7: Several topics for future research on coffee origin authentication

No	Ideas for future prospective studies
1.	Authentication of coffee based on untargeted metabolomics analysis have not been studied comprehensively especially for the samples involving the variability of origins coupled to the variations of degree of roasting as well as the different harvest seasons.
2.	Volatile compounds and lipid fraction investigation in association with geographical origin of coffee with variation on their altitudes and postharvest processing.
3.	As a future direction, LC/MS represents a potential tool for targeting nonvolatile metabolites in brews, including primary and secondary metabolites.
4.	Untargeted HPLC-UV or HPLC-FLD have not been thoroughly studied, specifically for the coffee green beans and roasted bean of the highly valued

coffee such as *kopi luwak* and Monsoonal processed coffee.

5. Numerous studies identified potential markers compounds for geographical origin determination. However, the confirmation of the selected markers after long storage has never been found in the literature.

Finally, a general comparison of the advantages and disadvantages of analytical and nondestructive approaches is illustrated in Table 6. Furthermore, future research ideas on coffee origin classification and/or authentication are indicated in Table 7.

CHAPTER 3

METHODOLOGY

3.1. Materials

Indonesian coffee, both *Coffea arabica* (Arabica coffee) and *Coffea canephora* (Robusta coffee), were partially obtained from the University of Lampung (Indonesia), local coffee farmers, and a trusted distributor from each geographical origin in Indonesia. As presented in Figure 2, the roasted beans from six districts across four different islands were selected. A sample harvested in 2019 was used to build up the discriminant model based on the comprehensive analysis of lipid species. Coffee from the harvest year of 2019 or 2020 from different harvesting locations was used to validate the model by measuring the important lipid species selected by the model (details in Table 8).

Figure 2: Map of Indonesia depicting the coffee origins used in this study

The green bean was roasted by a 500-g capacity roaster machine (WE 600I, Indonesia) using a controlled temperature of 195 °C for 12 min, followed by fan-assisted cooling for 10 min. The roasted coffee beans were ground using a home grinder (Tiamo, DongGuan Co., Ltd., China), and the powder was sieved with a stainless-steel

testing sieve with a sieve opening of 425 μm (Tokyo Screen, Ltd, Tokyo, Japan). The coffee samples were stored inside sealed aluminum foil at −80 °C for future use.

Table 8: Geographical origin, harvest year and assigned dataset for the coffee sample

No	Origin	Harvest Year	Region	Island	Type	Dataset
1	Gayo Aceh	2019	Atu Lintang	Sumatera	Arabica	Training
2	Gayo Aceh	2020	Bener meriah	Sumatera	Arabica	Training
3	Gayo Aceh	2019	Bener meriah	Sumatera	Arabica	Validation
4	Kintamani	2019	Belantih	Bali	Arabica	Training
5	Kintamani	2019	Catur	Bali	Arabica	Training
6	Kintamani	2020	Catur	Bali	Arabica	Validation
7	Lampung	2019	Tanggamus	Sumatera	Robusta	Training
8	Lampung	2019	Tanggamus	Sumatera	Robusta	Training
9	Lampung	2019	Margoyoso	Sumatera	Robusta	Validation
10	Mandheling	2019	Simpang Banyak	Sumatera	Arabica	Training
11	Mandheling	2019	Panyabungan	Sumatera	Arabica	Training
12	Mandheling	2020	Panyabungan	Sumatera	Arabica	Validation
13	Toraja	2019	Sapan	Sulawesi	Arabica	Training
14	Toraja	2019	Pulu-pulu	Sulawesi	Arabica	Training
15	Toraja	2020	Pulu-pulu	Sulawesi	Arabica	Validation
16	Wamena	2019	Baliem	Papua	Arabica	Training
17	Wamena	2019	Mimika	Papua	Arabica	Training
18	Wamena	2019	Baliem	Papua	Arabica	Validation

The chemical reagents used in this study were of LC-MS and HPLC grades. The LC-MS-grade chemicals of chloroform ($CHCl_3$), 2-isopropanol (IPA), acetonitrile (MeCN), methanol (MeOH), methyl *tert*-butyl ether (MTBE) were purchased from Fujifilm Wako Pure Chemical (Osaka, Japan). A 20-mm ammonium acetate solution was prepared from HPLC-grade ammonium acetate (Sigma-Aldrich, St. Louis, USA) using high-purity water (resistivity > 18MΩ cm) provided by a Millipore Direct-Q 3UV system (Merck KGaA, Darmstadt, Germany). A mixture of 20 µg/mL of 1,2-diheptadecanoyl-sn-glycero-3-phosphocholine (PC 17:0/17:0) and 1,2-diheptadecanoyl-sn-glycero-3-phosphoethanolamine (PE 17:0/17:0) (Avanti Polar Lipids, Inc., USA) was used as an internal standard (IS). PC 17:0/17:0 was dissolved in MeOH, and PE 17:0/17:0 was dissolved in a mixture of $CHCl_3$: MeOH: H_2O (6:4:0.5, v/v/v).

3.2. Lipid Extraction

Lipids were extracted from coffee samples as described by Matyash et al. (2008), with a slight modification. A homogenized 50-mg powder sample was added to 300-µL MeOH, 1000-µL MTBE, and 50-µL IS followed by a vortex for 1 min. The solution was incubated in a water-bath shaker (Personal 11, Taitec, Japan) for 1 h at 25 °C at 100 rpm. Phase separation was triggered by adding 250 µL of H_2O followed by centrifugation at 20000 × *g* for 5 min at 10 °C (Model 1720, Kubota, Japan). The separated aqueous solution was withdrawn and transferred into a new tube and subjected to a centrifugal evaporator system (EYELA CVE 2100, Tokyo Rikakikai Co. Ltd., Japan) for 4 h to obtain a dried lipid residue.

3.3. Liquid Chromatography–Electrospray Ionization–Triple Quadrupole Mass Spectrometry

The lipid in the samples was analyzed using a series of HPLC system coupled with a triple quadrupole mass spectrometry with an electrospray ionization (ESI) source (Q-TRAP 4500, AB-Sciex, Framingham, MA, USA). This system consists of an autosampler (SIL-20AC, Shimadzu, Kyoto, Japan), two high-pressure gradient pumps (LC-20AD, Shimadzu, Kyoto, Japan), a column oven (CTO-20A, Shimadzu, Kyoto, Japan), a reversed-phase chromatographic column (Cadenza CD-C18, particle size 3 μm, 150 × 2 mm, Imtakt, Kyoto, Japan), and a communication bus module (CBM-20A, Shimadzu, Kyoto, Japan).

The HPLC separation used binary flow pumping on IPA/MeOH/20-mM ammonium acetate (1:3:7, v/v/v) with the addition of 0.1% acetic acid (mobile phase A) and IPA/MeCN/20-mM ammonium acetate (7:3:1, v/v/v) with the addition of 0.1% acetic acid (mobile phase B). The time program elution was started at 40% of B in the first minute. The gradient increased to 80% of B in 3 min; then, it was held for 3 min, followed by 95% of B, which was held for 2 min. It ended with the isocratic elution at 100% of B for 14 min. The column then equilibrated at 40/60 (B/A) for 1.8 min before the next sample was evaluated. The total flow rate of the mobile phase was 0.35 mL/min. The column oven temperature was set to 40 °C.

Lipid extraction (10 μL), in which the dried lipid was reconstituted *via* the mobile phase B, was injected into HPLC and then subjected to the following tandem MS experiment: the acquisition batch of samples were analyzed in the randomized order with five replications and three injections. A pooled quality control (QC) sample

(consisting of all samples of the mixed solution) was sequentially injected every five samples during the run.

Mass spectrometry separation was performed on the basis of the multiple reaction monitoring (MRM) by the triple quadrupole mass spectrometer as previously studied (Aurum et al., 2022b) based on Tarazona et al. (2015) and Abhyankar et al. (2018). The MRM transitions for the targeted lipid species are listed in Tables S2 and S3 of the published article. A total of 953 lipid features were targeted and divided into five LC-MS/MS acquisition blocks. Each block consists of around 190 targeted molecules to obtain enough dwell time and data points in the chromatogram. The electrospray ionization (ESI) was applied for ionization using the TurboV ion source, which was operated in both the positive and negative modes with a source temperature of 300 °C. The positive ESI mode setup was as follows: ion spray, 5500 V; curtain gas, 30 kPa; collision gas, 9 kPa; ion source gas 1 (sheath gas), 50 kPa; and ion source gas 2 (drying gas), 80 kPa. The declustering potentials, entrance potential, collision energy, and collision exit potential (CXP) were different depending on the target metabolites. The negative ESI was conducted at –4500 V in ion spray voltage, whereas other parameters remained the same as the positive one.

3.4. Pre-data Processing

The raw peak intensity data (*.wiff) from the Analyst® software (SCIEX, Framingham, MA, USA) was extracted using the Marker View™ 1.2.1 software (SCIEX, Framingham, MA, USA), followed by a transformation into a comma delimited (CSV) file for further data processing. The peaks in MRM chromatograms were extracted in the following parameters: Gaussian smoothing, 1.5 points; noise

percentage, 50%; baseline subwindow, 1 min; peak-splitting factor, 4 points. The peaks were filtered with a minimum intensity of 1500 cps and minimum signal/noise ratio of 300 with a minimum of 5000 peaks. Pre-processing of the data, such as the elimination of the non-informative peaks, QC-based robust LOESS (locally estimated scatterplot smoothing) signal correction (QC-RLSC) (Dunn et al., 2011), the normalization by IS and sample weight, was performed using the in-house script written in R (version 4.0.3) (R Core Team, 2020) and run in RStudio (Rstudio, 2020). LOESS smoothing was applied to the pooled QC data with a smoother span value of 0.5 and spline interpolation. In the lipid feature filtering, a minimum threshold of 20% RSD based on QC values was used. The raw data was normalized by the smoothed QC values, followed by normalizing it by the value of the ratio of the IS (average divided by the IS of the sample). The peaks from the positive ESI mode data were normalized by IS PC 17:0/17:0, whereas the negative ESI mode data were normalized by IS PE 17:0/17:0. Cleaning of the non-informative data and any other incompatible values was performed simultaneously. The zero value in the detected lipid species was replaced with the means of the replication in the same sample. Normalization by weight of each sample was performed, followed by auto-scaling of the dataset, which was applied prior to the further multivariate analysis.

3.5. Statistical Analysis

The overall detected lipid species data were grouped into the associated head group. A statistical analysis of the overall data was conducted to understand the significant difference in the intensity of the lipid subclass among the coffee origins using the nonparametric ANOVA, Kruskal–Wallis test, because they do not meet the

normality assumption. The data was further subjected to a *post hoc* test using the Dunn method to find the differences with a probability level of 5%. The statistical test was conducted in R (version 4.0.3) (R Core Team, 2020) using the FSA package (Ogle et al., 2021).

3.6. Multivariate Data Analysis

In the multivariate data analysis, the mixOmics R package version 6.15.0 was used (Rohart et al., 2017). The partial least squares discriminant analysis (PLS-DA) was employed to classify coffee samples into the known groups of origins and predict the class of the test dataset samples (Indahl et al., 2007). In this study, the multivariate regression was used to build the model. Also, in the PLS-DA algorithm, the variable important in projection (VIP) can indicate the key lipid species that drive such a classification. Calibration of the model was performed using 5-fold cross-validation and was repeated 20 times to reduce bias on the model. A Hierarchical Cluster Analysis (HCA) dendrogram was built to support the supervised data analysis measured by Euclidian distance and Ward's method in "cluster" R package Ver 2.1.2.

The selection of potential markers for each coffee origin was determined based on the VIP score > 0.9 and *P*-value < 0.05 (Lim et al., 2018; Thammarat et al., 2018). They were selected from the loading of the chosen components. The number of components is determined by the lowest overall misclassification or balanced error rate and soft-thresholding penalization of the selected features extracted from the PLS-DA algorithm. The ability of the model to classify the coffee origin was evaluated using the area under the curve–receiver operating characteristics (AUC-ROC) curve (Fawcett,

2006). Subsequently, each coffee origin was designated with fewer than 10 important lipid species.

3.7 Validation of the Potential Marker

The validation of the potential marker from the important lipid species of each coffee origin was completed in a different experiment time. Validation of the model was performed using a separated sample set, as presented in Table 8. The coffee from either harvest year or the village of cultivation in the same harvest year was used for the validation sample set. After determining the most important lipid species from each coffee origin based on the model created using the PLS-DA (Section 3.6.), the validation sample set was further subjected to the lipid extraction step (Section 3.2.), followed by LC-MS/MS analysis (Section 2.3), in which only the specific lipids of interest for each coffee origin were targeted. The validation dataset was then subjected to an unsupervised data exploration using the principal component analysis (PCA) to observe the natural classification of the samples. A heatmap was visualized to illustrate the similarities between samples using a tree-structured cluster. The lipid features and the associated coffee origins were reordered based on a hierarchical clustering method to explore informative correlation patterns. A dendrogram was calculated based on the Euclidian distance of the *n*-dimensional plane, and the clustering analysis was conducted based on the *complete-linkage* algorithm that computes the least similar bits of a cluster.

CHAPTER 4

RESULTS AND DISCUSSION

4.1. Comprehensive Analysis of Lipid Species in Indonesian Coffee

This study tentatively identified 85 lipid species from five major classes of lipids analyzed by LC-MS/MS from all six coffee origins in the training sample set after data filtering, as presented in Table 9. Several lipid features were eliminated due to inconsistency and instability of the peak intensity based on the QC samples within the threshold limit of 20% RSD. In this study, the neutral glycerolipid class was outnumbered compared with the sphingolipids and sterol lipid classes. Neutral lipids in the subclass of TAG were detected with 22 species, seven species of diacylglycerol (DAG), and one species of monoacylglycerol (MAG). Ceramides (Cer) and glycoceramides (GlcCer) from the sphingolipid class were detected with three and one species, respectively.

The sterol lipid group is dominated by acylated steryl glycosides (ASG) and steryl esters (SE) with the number of species being seven and five, respectively. The polar glycerophospholipid class contributes 26 species, and the polar glyceroglycolipid class shares 12 species. Phosphatidylinositol (PI), phosphatidylethanolamine (PE), and phosphatidylglycerol (PG) are the major lipid subclasses in polar glycerophospholipids, with the number of species being four, eight, and nine, respectively. In addition, several lipid species categorized as sulfoquinovosyldiacylglycerols (SQDG), sulfoquinovosylmonoacylglycerols (SQMG), and monogalactosyldiacylglycerol (MGDG) subclasses were also detected in the coffee samples.

Figure 3 presents the overall lipid feature of Indonesian coffee from six major producing areas. The bar graph expresses the mean abundance of each lipid class, in

which the statistical test indicates that 15 out of 16 lipid groups were statistically significantly different ($P < 0.05$) among the coffee origins. For instance, coffee from Gayo and Toraja are both from the same *Coffea arabica* species, yet several lipid classes exhibit significant differences between the two geographical origins, such as in the PE, PG, PI, and SQDG groups. However, both origins share several lipid class features in DAG, LGPL, and TAG.

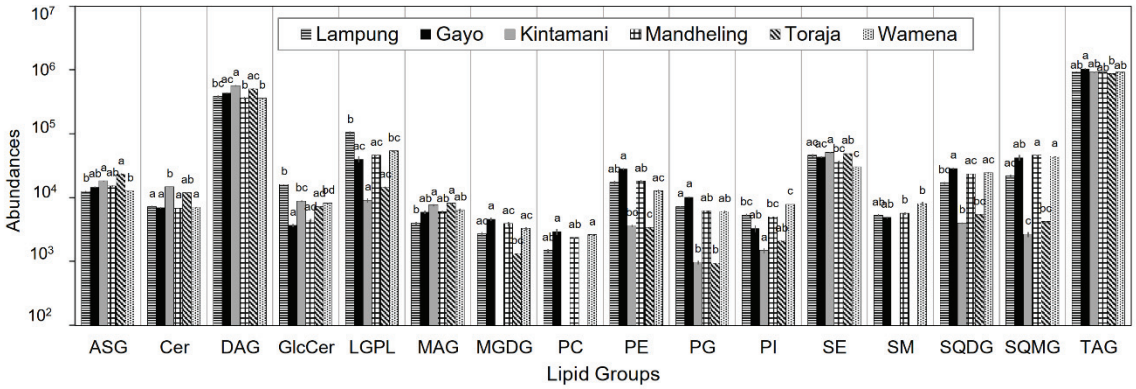
Table 9: Number of tentative lipid species detected in coffee samples by LC-MS/MS

Lipid Class (Positive ESI)	Number of species	Lipid Class (Negative ESI)	Number of species
Sphingolipids		**Polar glycerophospholipids**	
Ceramides (Cer)	3	Phosphatidylinositols (PI)	4
Glycoceramides (GlcCer)	1	Phosphatidylcholine (PC)	1
Sphingomyelin (SM)	1	Phosphatidylglycerols (PG)	9
Sterol lipids		Phosphatidylethanolamine (PE)	8
Acylated steryl glycosides (ASG)	7	Lysoglycerophospholipids (LGPL)	4
Steryl esters (SE)	5	**Polar glyceroglycolipids**	
Neutral glycerolipids		Monogalactosyldiacylglycerols (MGDG)	2
Triacylglycerol (TAG)	22	Sulfoquinovosyldiacylglycerols (SQDG)	8
Diacylglycerol (DAG)	7	Sulfoquinovosylmonoacylglycerols (SQMG)	2
Monoacylglycerol (MAG)	1		
Total number of species	47	Total number of species	38

Note: Lysoglycerophospholipids (LGPL) class includes Lyso-Phosphatidylinositols (LPI), Lyso-Phosphatidylcholine (LPC), Lyso-Phosphatidylglycerols (LPG)

On the other hand, several lipid groups may not be detected in all of the samples or were detected but were below the threshold value of 20% RSD, such as SM and PC, which are not available in Kintamani and Toraja, as both are from central Indonesia, as well as MGDG, which is not present in Kintamani. Furthermore, TAG was observed in all six coffee origins with similar abundance. Hence, it statistically demonstrated the same value for all origins. TAG is naturally present in coffee, since it may account for about 75% of the lipids in crude coffee, as previously studied by Nikolova-Damyanova et al. (1998). Therefore, both compounds were shown to be among the highest in abundance and were plentiful in the species detected in the coffee samples.

Figure 3: Overall lipid profile of Indonesian coffee from the six origins obtained from



the LC-MS/MS analysis, grouped based on 16 lipid class head groups (the different letters indicate significantly different P-values < 0.05, n = 5)

Several studies on the identification and quantification of lipid groups in coffee were conducted based on the FA content (Bertrand et al., 2008; Martín et al., 2001; Mehari et al., 2019d; Romano et al., 2014). TAG and DAG were also elucidated in the other studies (Cossignani et al., 2016; Toci et al., 2013). Recently, a study found that

lysophosphatidylcholine (LPC), PC, and PI were detected in Brazilian coffee from the species of *Coffea arabica* (Silva et al., 2020).

4.2. Supervised Data Exploration

In this study, the PLS-DA algorithm as a supervised analysis was used to develop the discriminative model to classify the coffee samples. Figure 4 presents the PLS-DA sample plot of the training dataset of the coffee samples from all six origins. The PLS-DA model illustrates promising results for the separation of the class of origin.

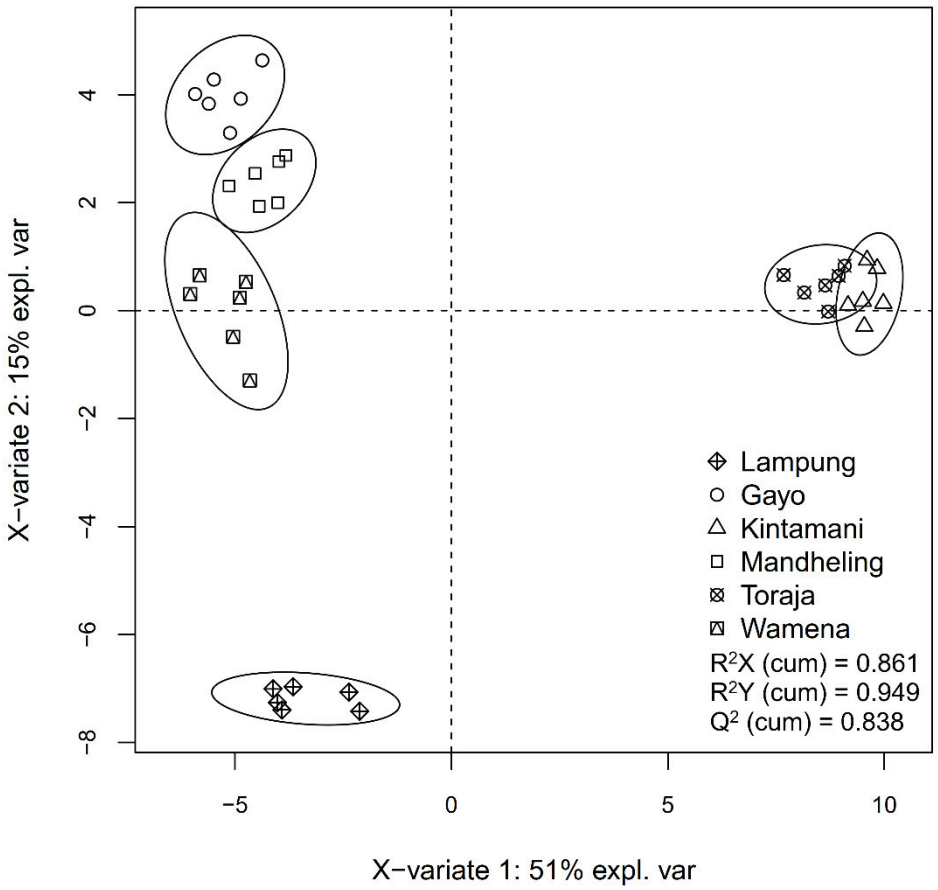


Figure 4: PLS-DA sample plot of the comprehensive lipid profile with all tentative detected lipid species in the training dataset

It shows total explainable variations of R^2X (cum) = 0.861, R^2Y (cum) = 0.949, and predictive capability Q^2 = 0.838 (values near to 1 shows better performance) built by a total of 10 x-variates. In this plot, all samples from the same origins are clustered together. The coffee samples are separated based on each origin, with a confidence ellipse of 95%. In Figure 4, discrimination can be found in either Kintamani or Toraja against the Gayo, Mandheling, Wamena, and Lampung origins on X-variate 1 (x-axis). Differentiation of Gayo, Mandheling, Wamena, and Lampung can be seen from X-variate 2 (y-axis). The sample plot of Lampung origin independently creates its own cluster since it may contain unique lipid profiles compared to other samples. Moreover, the Lampung coffee used in this study is a *C. canephora* Robusta variety, thus it is natural to depict different lipid profiles.

Gayo, Mandheling, and Wamena which are Arabica coffees independently assemble their clusters in the top left corner. These coffee samples are unique in their taste and characters, originating from different islands. Gayo indicates a strong body taste compared with other coffee, while Toraja coffee taste is more balance between the body and acidity (Happyana et al., 2020a). On the positive side of X-variate 1 of the PLS-DA sample plot, the origins of Toraja and Kintamani are separated with a very slight eclipse. In this case, these two coffee origins may share a few similar features. Although they come from different islands, both are in the central part of Indonesia. Overall, the PLS-DA sample plot indicates separation among the different origins. To support these findings an HCA dendrogram was built to show the correlation between the samples, as seen in Figure 5.

In PLS-DA, the X-variate, components, or latent variables indicate the amount of covariation explained. Contrary to PCA, the component in the PLS-DA maximizes

the sample co-variance between the response (y) and the linear combination of the predictor variable (x) by developing an array of orthogonal components. The PLS-DA performance is indicated by the balanced error rate (BER), which is the average proportion of the misclassified samples weighted by the number of samples in each class, as presented in Figure 6 (Lê Cao et al., 2011).

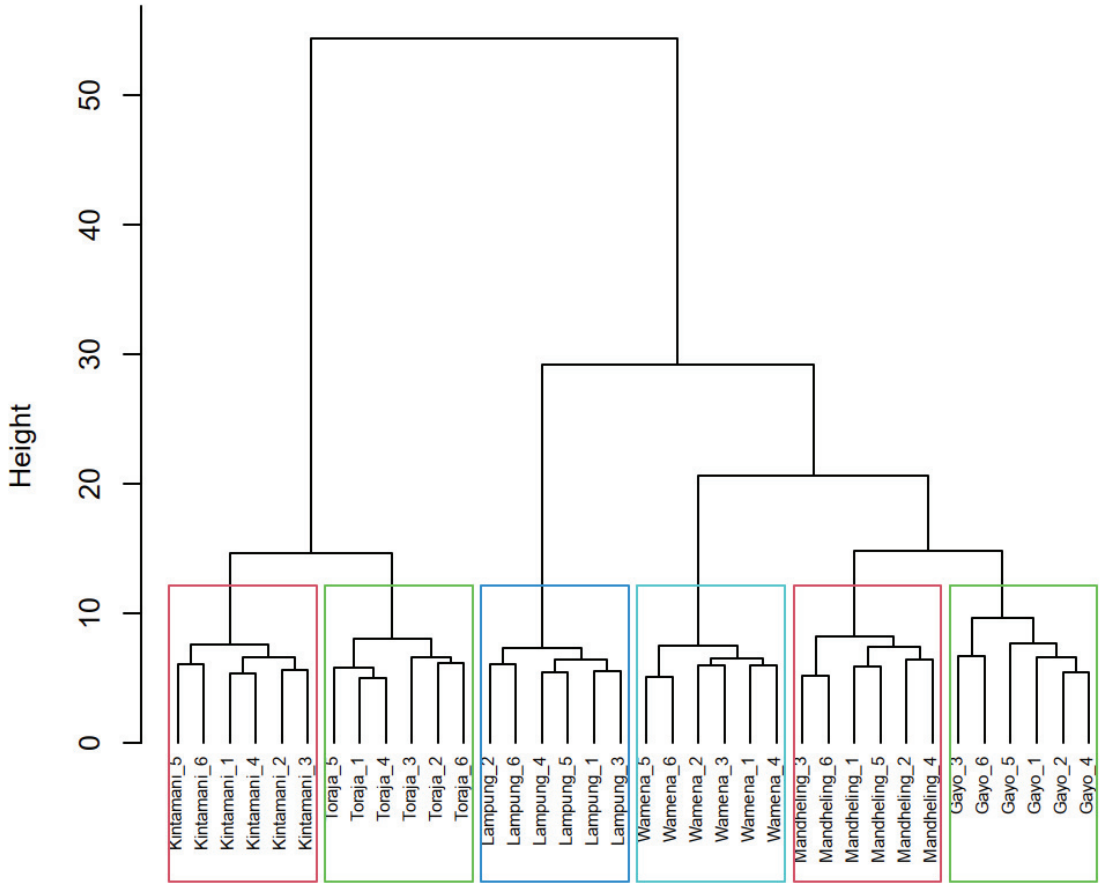


Figure 5: Hierarchical Cluster Analysis (HCA) Dendrogram of coffee samples from the training dataset illustrating the correlation between the origins. Each coffee sample clusters to its origins. Kintamani and Toraja assemble nearer distance, as well Gayo and Mandheling are in the nearby cluster linking to Wamena, while Lampung coffee forms a different cluster. HCA used Ward's algorithm and squared Euclidean distance measure criterion.

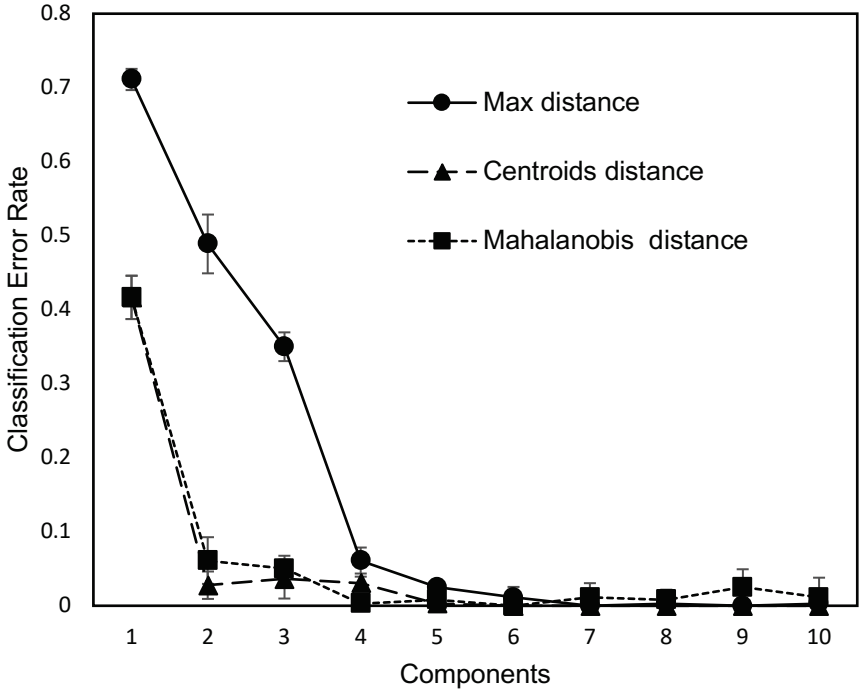


Figure: 6 Balanced error rate of the PLSDA model which illustrates the classification error rate of each class, lower value indicates better accuracy of the model prediction.

To elucidate the contribution of lipid species that drive such discrimination in the PLS-DA model, Figure 7 presents the pyramid bar plots in order of increasing importance (the bottom part is the most important and strongest contribution). Moreover, because using only two dimensions of the component will not be informative, four dimensions of the component are presented to point out the separation. The number of dimensions was selected based on the error rate (Figure 6) which illustrates that four components achieved the lowest classification error rate while avoiding overfitting of the model. The color of the bar in Figure 7 shows a sample class (coffee origin) with the maximum average loading weight value.

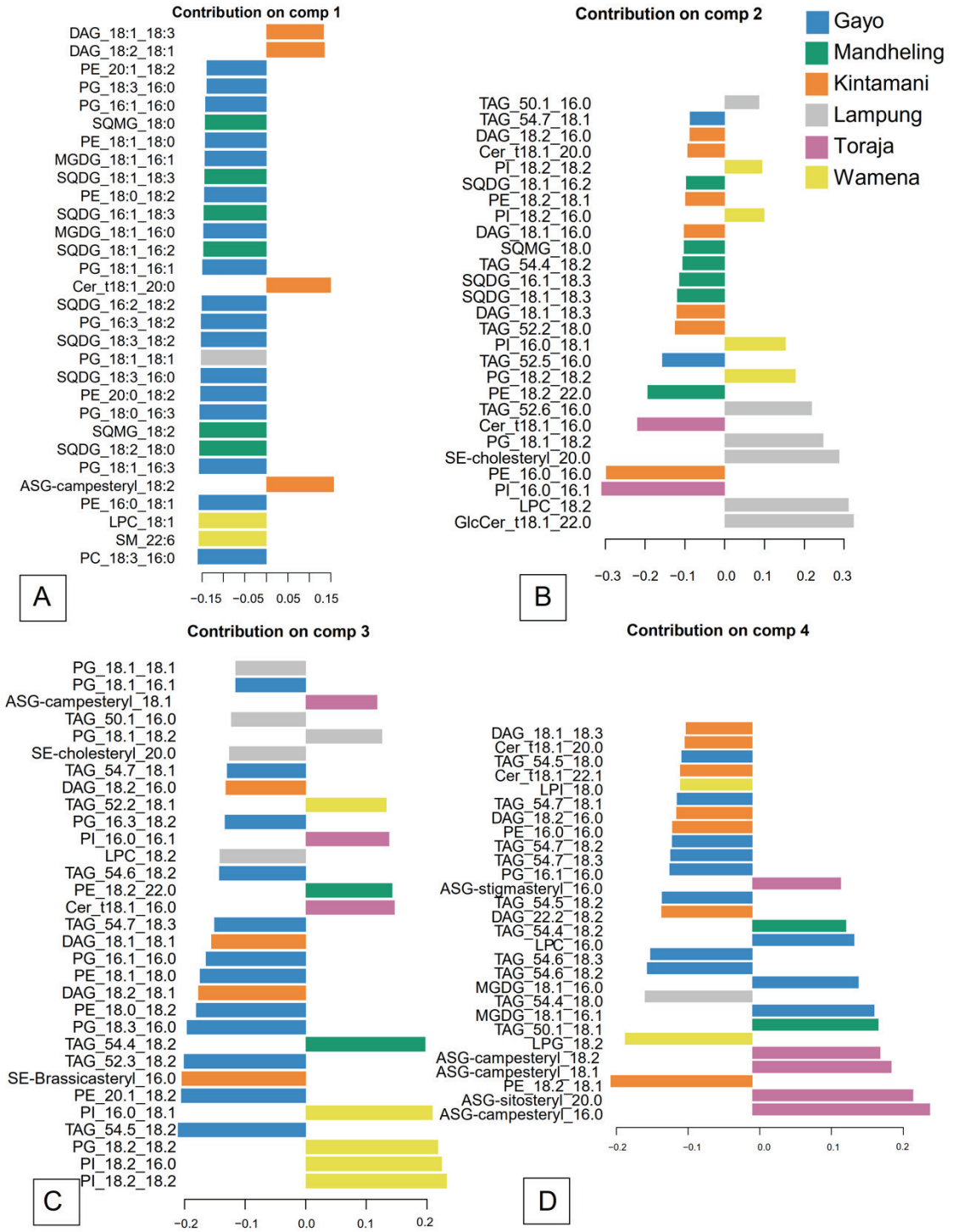


Figure 7: Bar plot of the loading of components 1 (A), 2 (B), 3 (C), and 4 (D). Lipid species contribution in each component as regression coefficient based on the PLS-DA model

The lipid feature was determined to be discriminative for the origin of coffee if the normalized mean value in that specific coffee was higher than the others. For instance, Figure 7A presents component 1 loading (51% expl. var) that showed that PC 18:3_16:0 is associated with the Gayo origin, indicating a negative coefficient. SM 22:6 and LPC 18:1 are both associated with Wamena. On the other hand, ASG-campesteryl 18:2 shows the highest value for Kintamani in a positive coefficient. This indicates that the mentioned features are important for the differentiation of the associated coffee originating elsewhere. Moreover, the opposite position in the y-axis of the bar plot indicates that these important lipids in each coffee origin led to discrimination between classes. Overall, the point of view of component 1 expresses discrimination between Kintamani against Mandheling, Wamena, Gayo, and Lampung.

Discrimination of the Lampung origin coffee against all other samples can be found in the loading of component 2 (15.1% expl. var) (Figure 7B), since a high loading coefficient of GlcCer t18:1_22:0, LPC 18:2, and SE-cholesteryl 20:0 is indicated by the Lampung coffee in positive coefficient. On the other hand, PI 16:0_16:1 and Cer t18:1_16:0 contributed to the discrimination of the origin of Toraja coffee. Thus, these species are notable for both coffee origins. Furthermore, Wamena is strongly characterized by several PI species in component 3 (8.3% expl. var) (Figure 7C). Component 4 (5.5% expl. var) (Figure 7D) indicates several loading weights; however, several loadings in this component overlap with the other components. Thus, the loadings in the earlier components can be selected as stronger marker candidates.

The loading weights demonstrate the magnitude of each predictor's (lipid species) contribution to the determination of the sample class (coffee origin) on each model component. The higher the absolute value of the loading, the more significant the

contribution is in explaining the coffee's origin class. However, it is important to note that each component in the model has a different loading weight range.

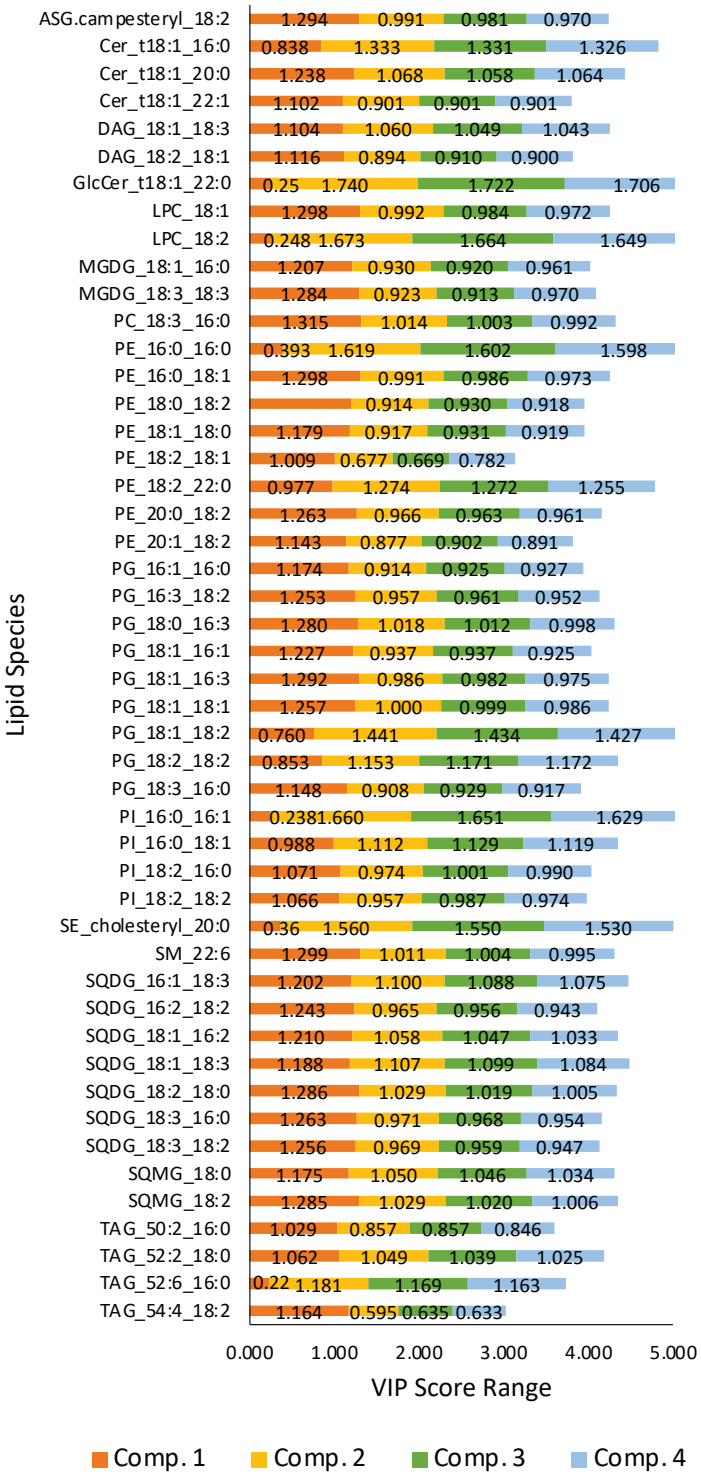


Figure 8: Variable important in projection (VIP) score consists of the lipid species with the score > 0.9 from component 1 to 4 based on PLS-DA model

Thus, the loading weight of the four components in the model coffee from different origins generally indicates different lipid metabolites. In agreement with the loading weight, the VIP score of each significant species associated with coffee origins indicates a value > 0.9 and a *P*-value < 0.05 in the univariate analysis (Figure 8). Furthermore, narrowing down the number of important lipid species of each class is more insightful for understanding the potential marker of each coffee origin.

Therefore, after getting an accurate model using PLS-DA, it is necessary to tune the number of components to lessen the computational time and reduce the required resources. Tuning was performed using the sparse mode of the PLS-DA algorithm to obtain the most discriminative lipid species, which may lead to the selection of the potential marker of each coffee origin (Lê Cao et al., 2011). The validity of the classification of coffee samples based on the variable importance of each component was evaluated using AUC-ROC, as presented in Table 10.

Table: 10 Classification performance rate based on area under the receiver operating characteristic (ROC) curve before and after tuning the number of variables in each component on PLS-DA

Origin comparison	AUC (Initial)	AUC (Tuned)	P-value (Initial)	P-value (Tuned)	Classification error rate (Initial)	Classification error rate (Tuned)
Gayo vs Other(s)	0.998	1.000	.00236	.00006	0.11	0.00
Kintamani vs Other(s)	1.000	1.000	.00006	.00006	0.00	0.00
Lampung vs Other(s)	1.000	0.938	.00006	.00505	0.00	0.17
Mandheling vs Other(s)	0.907	1.000	.00761	.00006	0.19	0.00
Toraja vs Other(s)	0.938	0.990	.0024	.00210	0.17	0.11
Wamena vs Other(s)	1.000	1.000	.00005	.00005	0.00	0.00

The initial AUC for Kintamani, Lampung, and Wamena versus other coffee origins is equal to 1, whereas that for Gayo, Mandheling and Toraja is less than 1. After tuning the model, the AUC of Gayo and Mandheling was increased to 1, and that of Toraja was increased to 0.99. However, the AUC of Lampung decreased to 0.93 after tuning. Overall, based on the AUC, the model classification error can be reasonably used to determine the geographical origin of coffee. Theoretically, the ROC curve expresses the capability of the sparse PLS-DA to distinguish among the coffee origins with maximum ideal performances equal to 1 (Fawcett, 2006). This classification performance may not be absolutely accurate or in full agreement with the PLS-DA performance due to different thresholds. Furthermore, this model is elaborated to find the potential lipid marker of each coffee origin.

4.3. Validation of the Potential Lipid Markers as Origin Discriminator

To determine the discriminant feature of each coffee origin in this study, the lipid species in the loading plot must demonstrate the essential contribution in each component indicated by a VIP score larger than 0.9, which is statistically significant. Based on this benchmark, 38 lipid species were assigned as the discriminant features of the six coffee origins, as presented in Table 11. The important lipid species used as discriminator compounds from each location may have a different number to satisfy the statistical model qualification. Categorically, the Kintamani origin is characterized by nine lipid species, Lampung by seven, Gayo by seven, Mandheling by nine, Toraja by seven, and Wamena by six features. A few coffee origins share the same prospective discriminant compound, such as Kintamani and Toraja, which share ASG-campesteryl 18:2. In addition, coffee from Gayo and Lampung share PG 18:1_18:2 as the important

variable. This circumstance is likely due to Sumatra Island being the same origin of Gayo and Lampung in western Indonesia.

Table 11. List of the important putative lipid candidate markers of each coffee origin

Origin	Lipid Species	Precursor Ion	Fragment Ion	Retention Time (min)	VIP score	p-value
Kintamani	DAG 18:1_18:3	634.5	335.3	13.1	1.1043	2.12E-11
	DAG 18:2_18:1	636.6	337.3	14.4	1.1163	1.60E-11
	TAG 52:2_18:0	876.8	575.5	29.0	1.0622	6.30E-09
	PE 18:2_18:1	740.5	279.2	9.1	1.0090	8.10E-04
	PE 16:0_16:0	690.5	255.2	7.1	1.6189	2.38E-18
	Cer t18:1_20:0	610.6	262.3	11.2	1.3330	1.11E-31
	Cer t18:1_22:1	636.6	262.3	10.4	1.1022	3.05E-09
	ASG.campesteryl_18:2	842.7	383.4	18.5	1.2940	1.24E-08
	TAG 50:2_16:0	848.8	575.5	30.7	1.0293	1.61E-06
Toraja	Cer t18:1_16:0	554.5	262.3	12.0	1.3508	1.08E-41
	Cer t18:1_22:1	636.6	262.3	10.4	1.1022	3.05E-09
	TAG 50:2_16:0	848.8	575.5	30.7	1.2940	1.61E-06
	PI 16:0_16.1	807.5	255.2	9.2	1.6602	8.03E-19
	ASG.campesteryl_18:2	842.7	383.4	18.5	1.0060	1.24E-08
	MGDG 18:3_18:3	833.5	277.2	10.5	1.2070	1.74E-08
	Cer t18:1_20:0	610.6	262.3	11.2	1.3330	1.11E-31
Mandheling	TAG 54:4_18:2	900.8	603.5	30.7	1.1644	4.65E-06
	PE 18:2_22:0	798.6	279.2	15.8	1.2742	8.36E-18
	SQDG 18:2_18:0	845.5	279.2	13.6	1.2856	5.48E-21
	SQDG 18:1_18:3	841.5	281.2	13.7	1.1878	7.53E-17
	SQDG 16:1_18:3	813.5	253.2	12.2	1.2020	9.73E-17
	SQDG 18:1_16:2	815.5	281.2	13.4	1.2099	2.14E-15
	PG 18:1_16:1	745.5	283.2	12.1	1.2270	1.17E-14
	PG 18:0_16:3	743.5	249.2	11.5	1.2799	1.54E-22
	SQMG 18:2	583.3	283.3	5.2	1.2854	6.02E-21
Lampung	TAG 52:6_16:0	868.7	595.5	24.7	1.1813	5.20E-04
	GlcCer t18:1_22:0	816.7	280.3	14.1	1.7402	3.19E-23
	SE.cholesteryl_20:0	698.7	369.4	25.4	1.5596	9.58E-12
	LPC_18:2	578.3	279.2	11.2	1.6734	4.78E-29
	PG 18:1_18:2	771.5	279.2	13.0	1.4409	5.96E-20
	PG 18:1_18:1	773.5	281.2	14.2	1.2565	6.20E-23
	PE 18:1_18:0	744.6	281.2	11.0	1.1791	5.93E-16
Wamena	PG 18:2_18:2	769.5	279.2	11.7	1.1721	6.04E-16
	PI 16:0_18.1	835.5	255.2	10.4	1.1285	2.96E-18
	PI 18:2_16:0	833.5	279.2	9.5	1.0712	2.14E-17
	PI 18:2_18:2	857.5	279.2	10.7	1.0656	2.69E-17
	SM 22:6	775.6	184	13.1	1.2987	1.33E-25
	LPC_18:1	580.4	281.2	11.5	1.2983	1.28E-20
Gayo	PG 18:1_18:2	773.5	279.2	13.0	1.4409	5.96E-20
	MGDG 18:3_18:3	833.5	277.2	10.5	1.2070	7.96E-20
	PG 18:1_16:3	741.5	249.2	10.0	1.2917	8.78E-26

Origin	Lipid Species	Precursor Ion	Fragment Ion	Retention Time (min)	VIP score	p-value
	PG 18:1_16:1	745.5	283.2	12.1	1.2270	1.17E-14
	PG 18:0_16:3	743.5	249.2	11.6	1.2799	1.54E-22
	PE 16:0_18:1	716.5	255.2	12.8	1.2981	7.33E-23
	PC 18:3_16:0	814.6	277.2	11.5	1.3152	4.30E-24

Furthermore, to examine the efficacy of the selected potential markers, the validation sample set was extracted separately to obtain the lipid compounds. It was then subjected to analysis using the LC-MS/MS MRM mode to confirm the discrimination capacity of the selected species. The discrimination is presented in Figure 9 as a PCA plot and in Figure 10 as a heatmap. To evaluate the separation among the samples, PCA was used as an unsupervised multivariate analysis. The PCA pattern demonstrates natural discrimination of coffee samples based on their origins, employing <10 lipid species for each coffee origin. In this plot, the PCA algorithm finds the directions of the sample points without referring to the class labels with the explainable variant of 55% in PC1 and 22% in PC2, as well as the total sum of variation in X explained by the model (R^2X) = 0.993, and predictive ability (Q^2) = 0.836 based on cross-validation, built on 10 PCs. The PCA plot gives insight into the selected features of each sample, which can be used as a prospective discrimination marker of the origin of coffee.

Figure 10 presents a heatmap coupled with a dendrogram to illustrate the cluster of the samples against the associated important lipid species. The y-axis in Figure 10 shows that the samples from the same origin indicate the same pattern of associated variables. A hierarchical relationship among the samples is illustrated by the length of the lines on the left side of the heatmap. The samples from the same origin were connected in the nearest distance measured using the Euclidean formula. At the x-axis,

the correlated lipid species tended to construct a nearer distance. In Table 11, the selected species from each origin organized the tiles on the heatmap and showed a substantial contribution, as indicated by the orange tiles. For instance, coffee from Wamena was clearly discriminated by several PI and PG groups with a positive correlation.

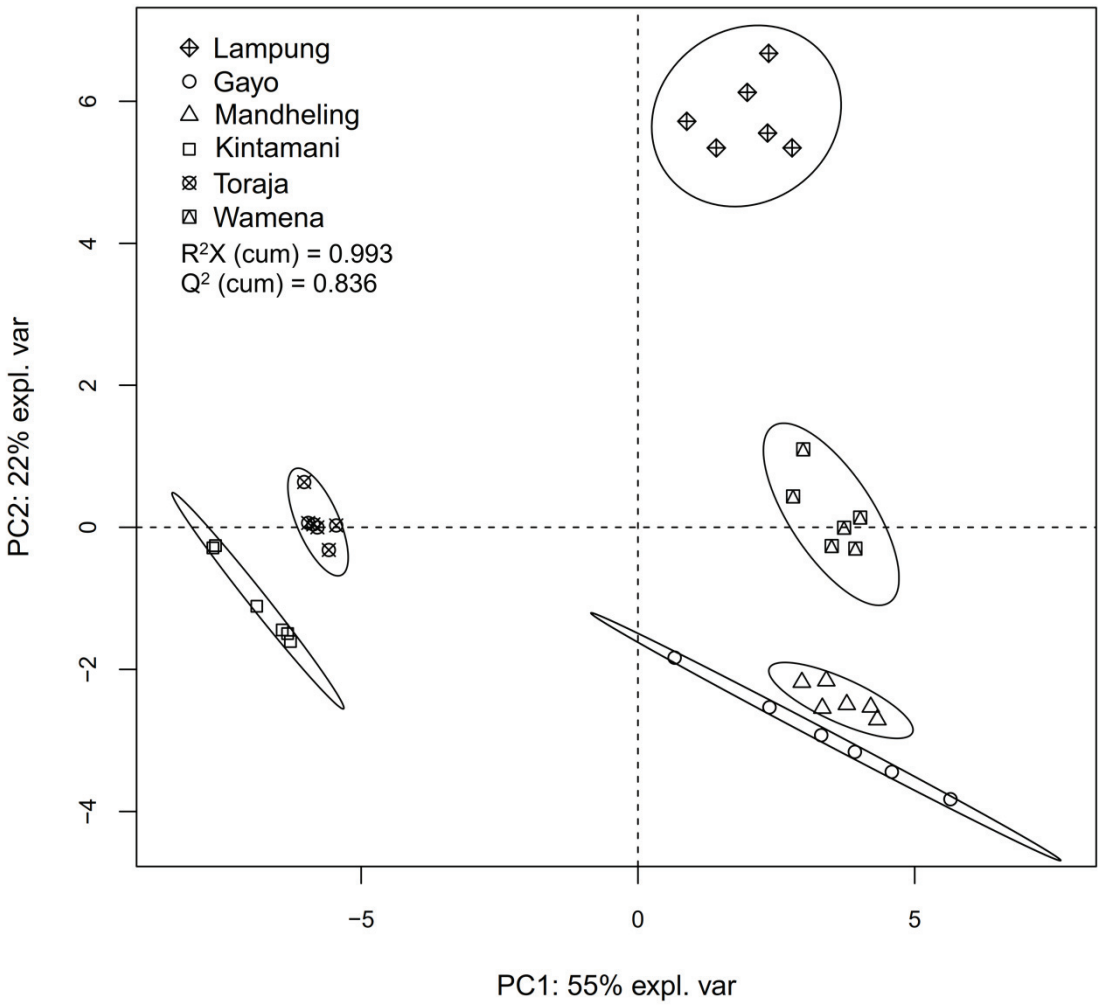


Figure 9: Principal component analysis (PCA) score plot to illustrate the natural grouping of the validation sample

On the other hand, the species that characterize Wamena negatively correlated with other origins, such as in Toraja and Kintamani, indicated by the blue tiles. The

lipid group of PI was found in Brazilian coffee by Silva et al. (2020), yet PG was not found in that particular study.

Kintamani and Toraja coffee exhibits a tentative lipid profile of Cer t18:1_20:0 and Cer t18:1_22:1. Ceramides in plants are usually found as 4-hydroxy-8-sphingadenin (t18:1), which have been studied to reduce Alzheimer's disease (Eguchi et al., 2020). This is the first report that suggests the presence of lipid feature believed to be Ceramides in coffee, although it requires further study for confirmation.

Toraja coffee also shares most of the same features with Kintamani, as presented in Figure 10, yet Kintamani shows higher DAG 18:2_18:1 and DAG 18:1_18:3, as well as several TAG feature intensities useful in distinguishing between the two. It should be noted that even though Kintamani and Toraja are from different islands, they are both in central Indonesia and are from the Arabica variety. According to Speer and Kölling-Speer (2006), the acyglycerol DAG and TAG are found in the endosperm of coffee beans. In addition, TAG is an aroma enhancer found in roasted coffee beans (Flament, 2002).

Gayo coffee was potentially described with glycerophospholipids owing to its strong abundance in PE 18:1_18:0, PG 18:0_16:3, and PE 16:0_18:1. In addition, MGDG and SQDG from the group of glyceroglycolipids were also present. Lampung coffee was significantly influenced by TAG 52:6_16:0, GlcCer t18:1_22:0, SE-cholesteryl 20:0, and LPC 18:2. The Mandheling coffee was strongly associated with an abundance of TAG 54:4_18:2, which is significant compared with other samples, as well as several SQDG features, which have the potential to be selected as discriminant features.

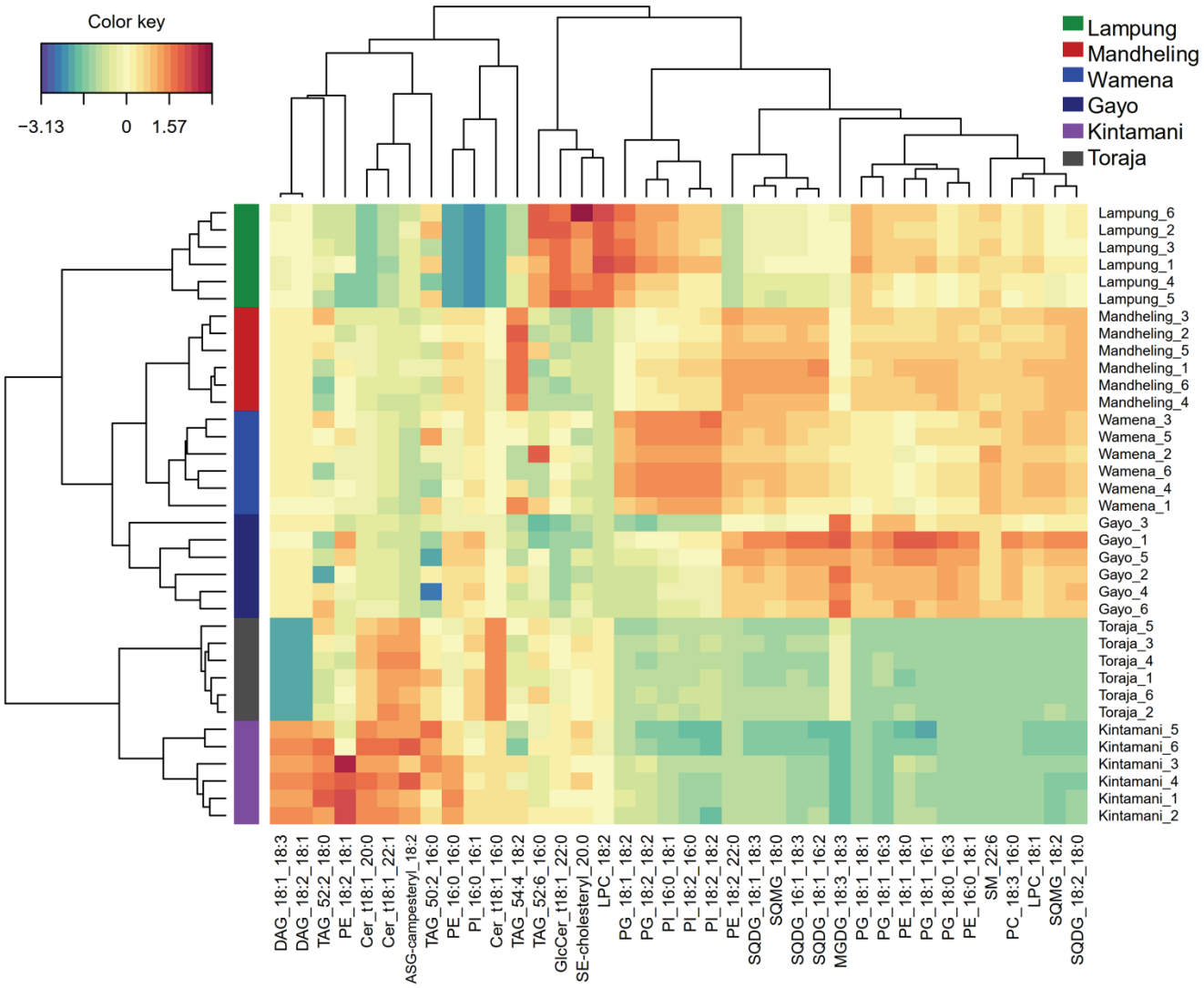


Figure: 10 Heatmap and dendrogram to illustrate clustering of the validation sample from each geographical origin

Overall, both the PCA (Figure 9) and heatmap (Figure 10) support that the selected features of each origin are dependable for the accurate discrimination of coffee samples. A clear separation was illustrated in both figures. Lastly, this study confirms that the lipid profile of coffee is a useful chemical descriptor that can be used for the discrimination of origin. Moreover, the results of this study will be advantageous, because this is the first lipidomic profiling of Indonesian coffee that is exported

worldwide. The methodological approach, multivariate model, and results of this study are valuable information for the industrial application of lipid profile identification, especially in the application of LC-MS/MS for coffee authentication.

CHAPTER 5

CONCLUSION

The classification of coffee expanded over time, earlier studies still classify and identify coffee types (Arabica–Robusta), yet currently numerous studies were done employing diverse origins of coffee. However, the in-country geographical classification remains an issue since coffee within nearby cultivation areas cannot be separated. Lipidomic profiling using LC-MS/MS to discriminate the geographical origin of coffee from several regions in Indonesia showed promising results. With the aid of a multivariate analysis, the lipid feature demonstrates the possibility to discriminate coffee origins, which leads to finding potential markers. The selected lipid species assigned to each coffee origin demonstrate reliability as the geographical origin descriptor. This study will be valuable for coffee industries to determine the coffee origin based on the lipid profile of coffee. This research may also enrich the metabolite database of single-origin coffee. In addition, the proposed bioactive lipid features might be a useful information for further study.

PUBLICATIONS

Aurum, Fawzan Sigma., Imaizumi, Teppei., Thammawong, Manasikan., Suhandy, Diding., Praseptiangga, Danar., Tsuta, Mizuki., Nagata, Masayasu., Nakano, Kohei. 2022. Lipidomic profiling of Indonesian coffee to determine its geographical origin by LC- MS/MS. European Food Research and Technology. https://doi.org/10.1007/s00217-022-04098-5

Aurum, Fawzan Sigma., Imaizumi, Teppei., Thammawong, Manasikan., Preseptiangga, Danar., Nakano, Kohei. 2022. Coffee Origin Determination Based on Analytical and Nondestructive Approaches–A Systematic Literature Review. Reviews in Agricultural Science. https://doi.org/g/10.7831/ras.10.0_257

ACKNOWLEDGEMENT

First of all, sincere praise and gratitude to Allaah God the Almighty, we seek His help, forgiveness, grace and mercy so that I can finish this thesis. Salutation is always devoted to the Prophet Muhammad and whoever is constantly following his path until the end of time.

I would like to thank my supervisor, Prof. Kohei Nakano, for his patience during the process of research experiment and publication writing. And for his countless times for giving me excellent guidance and advice for the problem-solving that encompasses the success of my PhD finalization. Also, I would like to thank the MEXT Scholarship- the Japanese Government for the financial support and for allowing me to pursue my PhD at Gifu University. My highest appreciation for IAARD, The Ministry of Agriculture Indonesia, for the permission to study and support.

I also thank all co-supervisors, Dr. Danar Praseptiangga, Dr. Mizuki Tsuta, Dr. Teppei Imaizumi, and Dr. Thammawong Manasikan, for not only strengthening our knowledge in the current field of study and related science but also for always supporting us in any situation during all stages of my study.

And for Dr. Diding Suhandy, thank you for always being available to give suggestions and involved in an exciting discussion.

To all my friends in Postharvest Engineering Laboratory, we have made such an unforgettable memory. To Madam Nakano, thank you for making a friendly environment both inside and outside academic activities.

To my family member in Indonesia, my mother, father and sister and also my parents-in-law, my appreciation and prayer for them will never end. And because of their prayer, God the Almighty has always blessed me in all my life processes.

My deepest gratitude also to my wife, Zoraya Amalia, and my daughter Laksmi Asshoffa who have been through a very hard life during my PhD study. They always give me total support and motivation. To realize that they are always here to stand by me is the highest relief I have ever experienced. There are no words to express my great thank and love for them.

REFERENCES

Abdelwareth, A., Zayed, A., & Farag, M. A. (2021). Chemometrics-based aroma profiling for revealing origin, roasting indices, and brewing method in coffee seeds and its commercial blends in the Middle East. *Food Chemistry*, *349*, 129162. https://doi.org/10.1016/j.foodchem.2021.129162

Abhyankar, V., Kaduskar, B., Kamat, S. S., Deobagkar, D., & Ratnaparkhi, G. S. (2018). Drosophila DNA/RNA methyltransferase contributes to robust host defense in aging animals by regulating sphingolipid metabolism. *Journal of Experimental Biology*, *221*(22), 187989. https://doi.org/10.1242/jeb.187989

Alnsour, L., Issa, R., Awwad, S., Albals, D., & Al-Momani, I. (2022). Quantification of Total Phenols and Antioxidants in Coffee Samples of Different Origins and Evaluation of the Effect of Degree of Roasting on Their Levels. *Molecules*, *27*(5), 1–10. https://doi.org/10.3390/molecules27051591

Amalia, F., Aditiawati, P., Yusianto, Putri, S. P., & Fukusaki, E. (2021). Gas chromatography/mass spectrometry-based metabolite profiling of coffee beans obtained from different altitudes and origins with various postharvest processing. *Metabolomics*, *17*(7), 69. https://doi.org/10.1007/s11306-021-01817-z

Anese, M., De Pilli, T., Massini, R., & Lerici, C. R. (2000). Oxidative stability of the lipid fraction in roasted coffee. *Italian Journal of Food Science*, *12*(4), 457–462.

Antoine, J. M. R., Hoo Fung, L. A., & Grant, C. N. (2016). Geographic determination of the growing origins of Jamaican and international coffee using instrumental neutron activation analysis and other methods. *Journal of Radioanalytical and Nuclear Chemistry*, *309*(2), 525–534. https://doi.org/10.1007/s10967-015-4666-4

Arana, V. A., Medina, J., Alarcon, R., Moreno, E., Heintz, L., Schäfer, H., & Wist, J. (2015). Coffee's country of origin determined by NMR: The Colombian case. *Food Chemistry*, *175*, 500–506. https://doi.org/10.1016/j.foodchem.2014.11.160

Arrieta, A. A. A., Nuñez, Y. E., Mendoza, J. M. M., Núñez, Y. E., & Mendoza, J. M. M. (2020). Mini-electronic Tongue Used to Discriminate between Coffee Samples of Different Geographical Origin. *International Journal of Technology*, *11*(2), 288–298. https://doi.org/10.14716/ijtech.v11i2.3225

Assis, C., Oliveira, L. S., & Sena, M. M. (2018). Variable Selection Applied to the Development of a Robust Method for the Quantification of Coffee Blends Using Mid Infrared Spectroscopy. *Food Analytical Methods*, *11*(2), 578–588. https://doi.org/10.1007/s12161-017-1027-7

Aurum, F. S., Imaizumi, T., Manasikan, T., Praseptiangga, D., & Nakano, K. (2022a). Coffee Origin Determination Based on Analytical and Nondestructive Approaches-A Systematic Literature Review. *Reviews in Agricultural Science*, *10*, 257–287. https://doi.org/10.7831/ras.10.0_257

Aurum, F. S., Imaizumi, T., Manasikan, T., Suhandy, D., Praseptiangga, D., Tsuta, M., Nagata, M., & Nakano, K. (2022b). Lipidomic profiling of Indonesian coffee to determine its geographical origin by LC–MS/MS. *European Food Research and Technology*. https://doi.org/10.1007/s00217-022-04098-5

Badmos, S., Fu, M., Granato, D., & Kuhnert, N. (2020). Classification of Brazilian roasted coffees from different geographical origins and farming practices based on chlorogenic acid profiles. *Food Research International*, *134*, 109218. https://doi.org/10.1016/j.foodres.2020.109218

Badmos, S., Lee, S.H.H., & Kuhnert, N. (2019). Comparison and quantification of chlorogenic acids for differentiation of green Robusta and Arabica coffee beans. *Food Research International*, *126*, 108544. https://doi.org/10.1016/j.foodres.2019.108544

Barbin, F. D., Felicio, A. L. de S. M., Sun, D., Nixdorf, S. L., & Hirooka, E. Y. (2014). Application of infrared spectral techniques on quality and compositional attributes of coffee : An overview. *Food Research International*, *61*, 23–32. https://doi.org/10.1016/j.foodres.2014.01.005

Bertrand, B., Villarreal, D., Laffargue, A., Posada, H., Lashermes, P., & Dussert, S. (2008). Comparison of the effectiveness of fatty acids, chlorogenic acids, and elements for the chemometric discrimination of coffee (Coffea arabica L.) varieties and growing origins. *Journal of Agricultural and Food Chemistry*, *56*(6), 2273–2280. https://doi.org/10.1021/jf073314f

Bilge, G. (2020). Investigating the effects of geographical origin, roasting degree, particle size and brewing method on the physicochemical and spectral properties of Arabica coffee by PCA analysis. *Journal of Food Science and Technology*, *57*(9), 3345–3354. https://doi.org/10.1007/s13197-020-04367-9

Bitter, N. Q., Fernandez, D. P., Driscoll, A. W., Howa, J. D., & Ehleringer, J. R. (2020). Distinguishing the region-of-origin of roasted coffee beans with trace element ratios. *Food Chemistry*, *320*, 126602. https://doi.org/10.1016/j.foodchem.2020.126602

Bodner, M., Morozova, K., Kruathongsri, P., Thakeow, P., & Scampicchio, M. (2019). Effect of harvesting altitude, fermentation time and roasting degree on the aroma released by coffee powder monitored by proton transfer reaction mass spectrometry. *European Food Research and Technology*, *245*(7), 1499–1506. https://doi.org/10.1007/s00217-019-03281-5

Bona, E., Marquetti, I., Link, J. V., Makimori, G. Y. F., da Costa Arca, V., Guimarães Lemes, A. L., Ferreira, J. M. G., dos Santos Scholz, M. B., Valderrama, P., & Poppi, R. J. (2017). Support vector machines in tandem with infrared spectroscopy for geographical classification of green arabica coffee. *LWT - Food Science and Technology*, *76*, 330–336. https://doi.org/10.1016/j.lwt.2016.04.048

Bosmali, I., Lagiotis, G., Stavridou, E., Haider, N., Osathanunkul, M., Pasentsis, K., & Madesis, P. (2021). Novel authentication approach for coffee beans and the brewed beverage using a nuclear-based species-specific marker coupled with high resolution melting analysis. *LWT*, *137*, 110336. https://doi.org/10.1016/j.lwt.2020.110336

Botelho, B. G., Oliveira, L. S., & Franca, A. S. (2017). Fluorescence spectroscopy as tool for the geographical discrimination of coffees produced in different regions of Minas Gerais State in Brazil. *Food Control*, *77*, 25–31. https://doi.org/10.1016/j.foodcont.2017.01.020

Burns, D. T., & Walker, M. J. (2020). Critical Review of Analytical and Bioanalytical Verification of the Authenticity of Coffee. *Journal of AOAC INTERNATIONAL*, *103*(2), 283–294. https://doi.org/10.5740/jaoacint.18-0392

Cagliani, L. R., Pellegrino, G., Giugno, G., & Consonni, R. (2013). Quantification of Coffea arabica and Coffea canephora var. robusta in roasted and ground coffee blends. *Talanta*, *106*, 169–173. https://doi.org/10.1016/j.talanta.2012.12.003

Caporaso, N., Whitworth, M. B., Cui, C., & Fisk, I. D. (2018). Variability of single bean coffee volatile compounds of Arabica and robusta roasted coffees analysed by SPME-GC-MS. *Food Research International*, *108*, 628–640. https://doi.org/10.1016/j.foodres.2018.03.077

Carrera, F., León-Camacho, M., Pablos, F., & González, A. G. (1998). Authentication of green coffee varieties according to their sterolic profile. *Analytica Chimica Acta*, *370*(2–3), 131–139. https://doi.org/10.1016/S0003-2670(98)00303-1

Carter, J. F., Yates, H. S. A., & Tinggi, U. (2015). Isotopic and Elemental Composition of Roasted Coffee as a Guide to Authenticity and Origin. *Journal of Agricultural and Food Chemistry*, *63*(24), 5771–5779. https://doi.org/10.1021/acs.jafc.5b01526

Choi, M. Y., Choi, W., Park, J. H., Lim, J., & Kwon, S. W. (2010). Determination of coffee origins by integrated metabolomic approach of combining multiple analytical data. *Food Chemistry*, *121*(4), 1260–1268. https://doi.org/10.1016/j.foodchem.2010.01.035

Cincotta, F., Tripodi, G., Merlino, M., Verzera, A., & Condurso, C. (2020). Variety and shelf-life of coffee packaged in capsules. *LWT - Food Science and Technology*, *118*, 108718. https://doi.org/10.1016/j.lwt.2019.108718

Combes, M.-C., Joët, T., & Lashermes, P. (2018). Development of a rapid and efficient DNA-based method to detect and quantify adulterations in coffee (Arabica versus Robusta). *Food Control*, *88*, 198–206. https://doi.org/10.1016/j.foodcont.2018.01.014

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

Cossignani, L., Montesano, D., Simonetti, M. S., & Blasi, F. (2016). Authentication of Coffea arabica according to Triacylglycerol Stereospecific Composition. *Journal of Analytical Methods in Chemistry*, *2016*, 1–7. https://doi.org/10.1155/2016/7482620

Couto, C. de C., Freitas-Silva, O., Oliveira, E. M. M., Sousa, C., & Casal, S. (2022). Near-infrared spectroscopy applied to the detection of multiple adulterants in roasted and ground arabica coffee. *Foods*. https://doi.org/10.3390/foods11010061

da Silva Oliveira, E. C., da Luz, J. M. R., de Castro, M. G., Filgueiras, P. R., Guarçoni, R. C., de Castro, E. V. R., da Silva, M. de C. S., Pereira, L. L., Catarina, E., José, O., Rodrigues, M., Gomes, M., Roberto, P., Rogério, F., Guarçoni, C., Vinicius, E., & Castro, R. De. (2021). Chemical and sensory discrimination of coffee: impacts of the planting altitude and fermentation. *European Food Research and Technology*. https://doi.org/10.1007/s00217-021-03912-w

Damayanti, T., & Setiadi, H. (2019). The Influence of Certificaton of Gayo Coffee Geographical Indication Against Value Added of Coffee in Gayo Highlands, Aceh. *IOP Conference Series: Earth and Environmental Science*, *338*(1). https://doi.org/10.1088/1755-1315/338/1/012028

de Carvalho Couto, C., Freitas-Silva, O., Morais Oliveira, E. M., Sousa, C., & Casal, S. (2021). Near-Infrared Spectroscopy Applied to the Detection of Multiple Adulterants in Roasted and Ground Arabica Coffee. *Foods*, *11*(1), 61. https://doi.org/10.3390/foods11010061

de Moura Ribeiro, M. V., Boralle, N., Redigolo Pezza, H., Pezza, L., & Toci, A. T. (2017). Authenticity of roasted coffee using 1H NMR spectroscopy. *Journal of Food Composition and Analysis*, *57*, 24–30. https://doi.org/10.1016/j.jfca.2016.12.004

Demianová, A., Bobková, A., Lidiková, J., Jurčaga, L., Bobko, M., Belej, Ľ., Kolek, E., Poláková, K., Iriondo-DeHond, A., & Dolores del Castillo, M. (2022). Volatiles as chemical markers suitable for identification of the geographical origin of green Coffea arabica L. *Food Control*, *136*, 108869. https://doi.org/https://doi.org/10.1016/j.foodcont.2022.108869

Dias, R. C. E., de Faria-Machado, A. F., Mercadante, A. Z., Bragagnolo, N., & Benassi, M. de T. (2014). Roasting process affects the profile of diterpenes in coffee. *European Food Research and Technology*, *239*(6), 961–970. https://doi.org/10.1007/s00217-014-2293-x

Directorate General of Estate. (2020). *Coffee Production by Province in Indonesia 2019-2021*. https://ditjenbun.pertanian.go.id/?publikasi=buku-statistik-perkebunan-2019-2021 (Accessed date: 24 December 2021)

DJKI Kementerian Hukum dan HAM. (2022). *E-Indikasi Geografis*. Listing Indikasi Geografis Terdaftar. https://ig.dgip.go.id/ (Accessed date: 23 December 2021)

Domínguez, R. B., Moreno-Barón, L., Muñoz, R., & Gutiérrez, J. M. (2014). Voltammetric Electronic Tongue and Support Vector Machines for Identification of Selected Features in Mexican Coffee. *Sensors*, *14*(9), 17770–17785. https://doi.org/10.3390/s140917770

Dong, W., Tan, L., Zhao, J., Hu, R., & Lu, M. (2015). Characterization of fatty acid, amino acid and volatile compound compositions and bioactive components of seven coffee (Coffea robusta) cultivars grown in Hainan Province, China. *Molecules*, *20*(9), 16687–16708. https://doi.org/10.3390/molecules200916687

dos Santos, H. D., & Boffo, E. F. (2021). Coffee beyond the cup: analytical techniques used in chemical composition research—a review. *European Food Research and Technology*, *247*(4), 749–775. https://doi.org/10.1007/s00217-020-03679-6

Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., Brown, M., Knowles, J. D., Halsall, A., Haselden, J. N., Nicholls, A. W., Wilson, I. D., Kell, D. B., & Goodacre, R. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols*, *6*(7), 1060–1083. https://doi.org/10.1038/nprot.2011.335

Eguchi, K., Mikami, D., Sun, H., Tsumita, T., Takahashi, K., Mukai, K., Yuyama, K., & Igarashi, Y. (2020). Blood-brain barrier permeability analysis of plant ceramides. *PLoS ONE*, *15*, 1–12. https://doi.org/10.1371/journal.pone.0241640

Endaye, M., Atlabachew, M., Mehari, B., Alemayehu, M., Mengistu, D. A., & Kerisew, B. (2020). Combining Multi-Element Analysis with Statistical Modeling for Tracing the Origin of Green Coffee Beans from Amhara Region, Ethiopia. *Biological Trace Element Research*, *195*(2), 669–678. https://doi.org/10.1007/s12011-019-01866-5

Fahy, E., Cotter, D., Sud, M., & Subramaniam, S. (2011). Lipid classification, structures and tools. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, *1811*(11), 637–647. https://doi.org/10.1016/j.bbalip.2011.06.009

Faith Ndlovu, P., Samukelo Magwaza, L., Zeray Tesfay, S., & Ramaesele Mphahlele, R. (2022). Destructive and rapid non-invasive methods used to detect adulteration of dried powdered horticultural products: A review. *Food Research International*, *157*, 111198. https://doi.org/10.1016/j.foodres.2022.111198

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Ferreira, T., Farah, A., Oliveira, T. C., Lima, I. S., Vitório, F., & Oliveira, E. M. M. (2016). Using Real-Time PCR as a tool for monitoring the authenticity of commercial coffees. *Food Chemistry*, *199*, 433–438. https://doi.org/10.1016/j.foodchem.2015.12.045

Fiehn, O. (2002). Metabolomics - The link between genotypes and phenotypes. *Plant Molecular Biology*, *48*(1–2), 155–171. https://doi.org/10.1023/A:1013713905833

Flambeau, K. J., Lee, W.J., & Yoon, J. (2017). Discrimination and geographical origin prediction of washed specialty Bourbon coffee from different coffee growing areas in Rwanda by using electronic nose and electronic tongue. *Food Science and Biotechnology*, *26*(5), 1245–1254. https://doi.org/10.1007/s10068-017-0168-1

Flament, I. (2002). *Coffee Flavor Chemistry*. John Wiley &Sons, Ltd.

Federal Food Safety and Veterinary Office. (2019). *Checking of coffee labelling: three false declarations in Switzerland*. https://www.blv.admin.ch/blv/en/home/dokumentation/nsb-news-list.msg-id-75501.html (Accessed date: 13 December 2021)

Gamboa-Becerra, R., Montero-Vargas, J. M., Martínez-Jarquín, S., Gálvez-Ponce, E., Moreno-Pedraza, A., & Winkler, R. (2017). Rapid Classification of Coffee Products by Data Mining Models from Direct Electrospray and Plasma-Based Mass Spectrometry Analyses. *Food Analytical Methods*, *10*(5), 1359–1368. https://doi.org/10.1007/s12161-016-0696-y

Garrett, R., Schmidt, E. M., Pereira, L. F. P., Kitzberger, C. S. G., Scholz, M. B. S., Eberlin, M. N., & Rezende, C. M. (2013). Discrimination of arabica coffee cultivars by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry and chemometrics. *LWT - Food Science and Technology*, *50*(2), 496–502. https://doi.org/10.1016/j.lwt.2012.08.016

Giraudo, A., Grassi, S., Savorani, F., Gavoci, G., Casiraghi, E., & Geobaldo, F. (2019). Determination of the geographical origin of green coffee beans using NIR spectroscopy and multivariate data analysis. *Food Control*, *99*, 137–145. https://doi.org/10.1016/j.foodcont.2018.12.033

González, A. G., Pablos, F., Martín, M. J., León-Camacho, M., & Valdenebro, M. S. (2001). HPLC analysis of tocopherols and triglycerides in coffee and their use as authentication parameters. *Food Chemistry*, *73*(1), 93–101. https://doi.org/10.1016/S0308-8146(00)00282-X

Gordillo-Delgado, F., Marín, E., Cortés-Hernández, D. M., Mejía-Morales, C., & García-Salcedo, A. J. (2012). Discrimination of organic coffee via Fourier transform infrared-photoacoustic spectroscopy. *Journal of the Science of Food and Agriculture*, *92*(11), 2316–2319. https://doi.org/10.1002/jsfa.5628

Habte, G., Hwang, I. M., Kim, J. S., Hong, J. H., Hong, Y. S., Choi, J. Y., Nho, E. Y., Jamila, N., Khan, N., & Kim, K. S. (2016). Elemental profiling and geographical differentiation of Ethiopian coffee samples through inductively coupled plasma-optical emission spectroscopy (ICP-OES), ICP-mass spectrometry (ICP-MS) and direct mercury analyzer (DMA). *Food Chemistry*, *212*, 512–520. https://doi.org/https://doi.org/10.1016/j.foodchem.2016.05.178

Hamdouche, Y., Meile, J. C., Nganou, D. N., Durand, N., Teyssier, C., & Montet, D. (2016). Discrimination of post-harvest coffee processing methods by microbial ecology analyses. *Food Control*, *65*, 112–120. https://doi.org/10.1016/j.foodcont.2016.01.022

Happyana, N., Hermawati, E., Syah, Y. M., & Hakim, E. H. (2020a). Discrimination of the Indonesian roasted arabica coffees using 1H NMR-based metabolomics. *Current Research in Nutrition and Food Science*, *8*(2), 479–488. https://doi.org/10.12944/CRNFSJ.8.2.13

Happyana, N., Hermawati, E., Syah, Y. M., & Hakim, E. H. (2020b). Metabolite profile evaluation of Indonesian roasted robusta coffees by 1H NMR technique and chemometrics. *Indonesian Journal of Chemistry*, *20*(4), 850–857. https://doi.org/10.22146/IJC.46492

Hoyos Ossa, D. E., Gil-Solsona, R., Peñuela, G. A., Sancho, J. V., & Hernández, F. J. (2018). Assessment of protected designation of origin for Colombian coffees based on HRMS-based metabolomics. *Food Chemistry*, *250*, 89–97. https://doi.org/10.1016/j.foodchem.2018.01.038

Hung, Y. C., Lee, F. S., & Lin, C. I. (2021). Classification of coffee bean categories based upon analysis of fatty acid ingredients. *Journal of Food Processing and Preservation*, *45*(9), 1–17. https://doi.org/10.1111/jfpp.15703

International Coffee Organization. (2020). World coffee consumption. https://www.ico.org/trade_statistics.asp?section=Statistics (Accessed date: 6 Dec 2021).

Indahl, U. G., Martens, H., & Næs, T. (2007). From dummy regression to prior probabilities in PLS-DA. *Journal of Chemometrics*, *21*(12), 529–536. https://doi.org/10.1002/cem.1061

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning. In *Synthesis Lectures on Mathematics and Statistics* (Vol. 103, Issue 4). Springer New York. https://doi.org/10.1007/978-1-4614-7138-7

Knysak, D. (2017). Volatile compounds profiles in unroasted Coffea arabica and Coffea canephora beans from different countries. *Food Science and Technology*, *37*(3), 444–448. https://doi.org/10.1590/1678-457x.19216

Kotu, V., & Deshpande, B. (2019). Introduction. In *Data Science* (2nd Ed, pp. 1–18). Elsevier. https://doi.org/10.1016/B978-0-12-814761-0.00001-0

Lê Cao, K. A., Boitard, S., & Besse, P. (2011). Sparse PLS discriminant analysis: Biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics*, *12*(June). https://doi.org/10.1186/1471-2105-12-253

Lim, D. K., Mo, C., Lee, J. H., Long, N. P., Dong, Z., Li, J., Lim, J., & Kwon, S. W. (2018). The integration of multi-platform MS-based metabolomics and

multivariate analysis for the geographical origin discrimination of Oryza sativa L. *Journal of Food and Drug Analysis*, *26*(2), 769–777. https://doi.org/10.1016/j.jfda.2017.09.004

Link, J. V., Guimarães Lemes, A. L., Marquetti, I., dos Santos Scholz, M. B., & Bona, E. (2014). Geographical and genotypic segmentation of arabica coffee using self-organizing maps. *Food Research International*, *59*, 1–7. https://doi.org/10.1016/j.foodres.2014.01.063

Liu, H.C., You, C.F., Chen, C.Y., Liu, Y.C., & Chung, M.T. (2014). Geographic determination of coffee beans using multi-element analysis and isotope ratios of boron and strontium. *Food Chemistry*, *142*, 439–445. https://doi.org/10.1016/j.foodchem.2013.07.082

Lopetcharat, K., Kulapichitr, F., Suppavorasatit, I., Chodjarusawad, T., Phatthara-Aneksin, A., Pratontep, S., & Borompichaichartkul, C. (2016). Relationship between overall difference decision and electronic tongue: Discrimination of civet coffee. *Journal of Food Engineering*, *180*, 60–68. https://doi.org/10.1016/j.jfoodeng.2016.02.011

Makimori, G. Y. F., & Bona, E. (2019). Commercial Instant Coffee Classification Using an Electronic Nose in Tandem with the ComDim-LDA Approach. *Food Analytical Methods*, *12*(5), 1067–1076. https://doi.org/10.1007/s12161-019-01443-5

Marek, G., Dobrzański, B., Oniszczuk, T., Combrzyński, M., Ćwikła, D., & Rusinek, R. (2020). Detection and differentiation of volatile compound profiles in roasted coffee arabica beans from different countries using an electronic nose and GC-MS. *Sensors*, *20*(7). https://doi.org/10.3390/s20072124

Martín, M. J., Pablos, F., González, A. G., Valdenebro, M. S., & León-Camacho, M. (2001). Fatty acid profiles as discriminant parameters for coffee varieties differentiation. *Talanta*, *54*(2), 291–297. https://doi.org/10.1016/S0039-9140(00)00647-0

Martins, V. D. C., Godoy, R. L. D. O., Gouveâ, A. C. M. S., Santiago, M. C. P. D. A., Borguini, R. G., Braga, E. C. D. O., Pacheco, S., & Nascimento, L. D. S. D. M. (2018). Fraud investigation in commercial coffee by chromatography. *Food Quality and Safety*, *2*(3), 121–133. https://doi.org/10.1093/fqsafe/fyy017

Martín, M. J., Pablos, F., & González, A. G. (1998). Discrimination between arabica and robusta green coffee varieties according to their chemical composition. *Talanta*, *46*(6), 1259–1264. https://doi.org/10.1016/S0039-9140(97)00409-8

Matyash, V., Liebisch, G., Kurzchalia, T. V., Shevchenko, A., & Schwudke, D. (2008). Lipid extraction by methyl- tert -butyl ether for high-throughput lipidomics. *Journal of Lipid Research*, *49*(5), 1137–1146. https://doi.org/10.1194/jlr.D700041-JLR200

Medina, J., Caro Rodríguez, D., Arana, V. A., Bernal, A., Esseiva, P., & Wist, J. (2017). Comparison of Attenuated Total Reflectance Mid-Infrared, Near Infrared, and 1 H-Nuclear Magnetic Resonance Spectroscopies for the Determination of Coffee's Geographical Origin. *International Journal of Analytical Chemistry*, *2017*, 1–8. https://doi.org/10.1155/2017/7210463

Mehari, B., Chandravanshi, B. S., Redi-Abshiro, M., Combrinck, S., McCrindle, R., & Atlabachew, M. (2021). Polyphenol contents of green coffee beans from different regions of Ethiopia. *International Journal of Food Properties*, *24*(1), 17–27. https://doi.org/10.1080/10942912.2020.1858866

Mehari, B., Redi-Abshiro, M., Chandravanshi, B. S., Atlabachew, M., Combrinck, S., & McCrindle, R. (2016a). Simultaneous Determination of Alkaloids in Green Coffee Beans from Ethiopia: Chemometric Evaluation of Geographical Origin. *Food Analytical Methods*, *9*(6), 1627–1637. https://doi.org/10.1007/s12161-015-0340-2

Mehari, B., Redi-Abshiro, M., Chandravanshi, B. S., Combrinck, S., Atlabachew, M., & McCrindle, R. (2016b). Profiling of phenolic compounds using UPLC-MS for determining the geographical origin of green coffee beans from Ethiopia. *Journal of Food Composition and Analysis*, *45*, 16–25. https://doi.org/10.1016/j.jfca.2015.09.006

Mehari, B., Redi-Abshiro, M., Chandravanshi, B. S., Combrinck, S., & McCrindle, R. (2016c). Characterization of the Cultivation Region of Ethiopian Coffee by Elemental Analysis. *Analytical Letters*, *49*(15), 2474–2489. https://doi.org/10.1080/00032719.2016.1151023

Mehari, B., Redi-Abshiro, M., Chandravanshi, B. S., Combrinck, S., McCrindle, R., & Atlabachew, M. (2019d). GC-MS profiling of fatty acids in green coffee (Coffea arabica L.) beans and chemometric modeling for tracing geographical origins from Ethiopia. *Journal of the Science of Food and Agriculture*, *99*(8), 3811–3823. https://doi.org/10.1002/jsfa.9603

Mendes, G. de A., de Oliveira, M. A. L., Rodarte, M. P., de Carvalho dos Anjos, V., & Bell, M. J. V. (2022). Origin geographical classification of green coffee beans (Coffea arabica L.) produced in different regions of the Minas Gerais state by FT-MIR and chemometric. *Current Research in Food Science*, *5*, 298–305. https://doi.org/10.1016/j.crfs.2022.01.017

Miao, Y., Zou, Q., Wang, Q., Gong, J., Tan, C., Peng, C., Zhao, C., & Li, Z. (2022). Evaluation of the physiochemical and metabolite of different region coffee beans by using UHPLC-QE-MS untargeted-metabonomics approaches. *Food Bioscience*, *46*, 101561. https://doi.org/10.1016/j.fbio.2022.101561

Mohammed, F., Guillaume, D., Dowman, S., & Abdulwali, N. (2019). An easy way to discriminate Yemeni against Ethiopian coffee. *Microchemical Journal*, *145*, 173–179. https://doi.org/10.1016/j.microc.2018.10.039

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Monteiro, P. I., Santos, J. S., Alvarenga Brizola, V. R., Pasini Deolindo, C. T., Koot, A., Boerrigter-Eenling, R., van Ruth, S., Georgouli, K., Koidis, A., & Granato, D. (2018). Comparison between proton transfer reaction mass spectrometry and near infrared spectroscopy for the authentication of Brazilian coffee: A preliminary chemometric study. *Food Control*, *91*(10), 276–283. https://doi.org/10.1016/j.foodcont.2018.04.009

Monteiro, P. I., Santos, J. S., Rodionova, O. Y., Pomerantsev, A., Chaves, E. S., Rosso, N. D., & Granato, D. (2019). Chemometric Authentication of Brazilian Coffees Based on Chemical Profiling. *Journal of Food Science*, *84*(11), 3099–3108. https://doi.org/10.1111/1750-3841.14815

Muñiz-Valencia, R., Jurado, J. M., Ceballos-Magaña, S. G., Alcázar, Á., & Hernández-Díaz, J. (2014). Characterization of Mexican coffee according to mineral contents by means of multilayer perceptrons artificial neural networks. *Journal of Food Composition and Analysis*, *34*(1), 7–11. https://doi.org/10.1016/j.jfca.2014.02.003

Muñiz-Valencia, R., Jurado, J. M., Ceballos-Magaña, S. G., Alcázar, A., & Reyes, J. (2013). Geographical Differentiation of Green Coffees According to Their Metal Content by Means of Supervised Pattern Recognition Techniques. *Food Analytical Methods*, *6*(5), 1271–1277. https://doi.org/10.1007/s12161-012-9538-8

Nikolova-Damyanova, B., Velikova, R., & Jham, G. N. (1998). Lipid classes, fatty acid composition and triacylglycerol molecular species in crude coffee beans harvested in Brazil. *Food Research International*, *31*(6–7), 479–486. https://doi.org/10.1016/S0963-9969(99)00016-2

Nocairi, H., Mostafa Qannari, E., Vigneau, E., & Bertrand, D. (2005). Discrimination on latent components with respect to patterns. Application to multicollinear data. *Computational Statistics & Data Analysis*, *48*(1), 139–147. https://doi.org/10.1016/j.csda.2003.09.008

Núñez, N., Collado, X., Martínez, C., Saurina, J., & Núñez, O. (2020). Authentication of the Origin, Variety and Roasting Degree of Coffee Samples by Non-Targeted HPLC-UV Fingerprinting and Chemometrics. Application to the Detection and Quantitation of Adulterated Coffee Samples. *Foods*, *9*(3), 378. https://doi.org/10.3390/foods9030378

Núñez, N., Martínez, C., Saurina, J., & Núñez, O. (2021). High-performance liquid chromatography with fluorescence detection fingerprints as chemical descriptors to authenticate the origin, variety and roasting degree of coffee by multivariate chemometric methods. *Journal of the Science of Food and Agriculture*, *101*(1), 65–73. https://doi.org/10.1002/jsfa.10615

Núñez, N., Saurina, J., & Núñez, O. (2021). Non-targeted HPLC-FLD fingerprinting for the detection and quantitation of adulterated coffee samples by chemometrics. *Food Control*, *124*, 107912. https://doi.org/https://doi.org/10.1016/j.foodcont.2021.107912

Obeidat, S. M., Hammoudeh, A. Y., & Alomary, A. A. (2018). Application of FTIR Spectroscopy for Assessment of Green Coffee Beans According to Their Origin. *Journal of Applied Spectroscopy*, *84*(6), 1051–1055. https://doi.org/10.1007/s10812-018-0585-9

Ogle, D. H., Wheeler, P., & Dinno, A. (2021). *FSA: Fisheries Stock Analysis* (R package version 0.8.32).

Okubo, N., & Kurata, Y. (2019a). Nondestructive classification analysis of green coffee beans by using near-infrared spectroscopy. *Foods*, *8*(2). https://doi.org/10.3390/foods8020082

Oliveira, M., Ramos, S., Delerue-Matos, C., & Morais, S. (2015). Espresso beverages of pure origin coffee: Mineral characterization, contribution for mineral intake and geographical discrimination. *Food Chemistry*, *177*, 330–338. https://doi.org/10.1016/j.foodchem.2015.01.061

Ongo, E. A., Montevecchi, G., Antonelli, A., Sberveglieri, V., & Sevilla, F. (2020). Metabolomics fingerprint of Philippine coffee by SPME-GC-MS for geographical and varietal classification. *Food Research International*, *134*(April), 109227. https://doi.org/10.1016/j.foodres.2020.109227

Ongo, E., Falasconi, M., Sberveglieri, G., Antonelli, A., Montevecchi, G., Sberveglieri, V., Concina, I., & Iii, F. S. (2012). Chemometric discrimination of philippine civet coffee using electronic nose and gas chromatography mass spectrometry. *Procedia Engineering*, *47*, 977–980. https://doi.org/10.1016/j.proeng.2012.09.310

Peng, C. yi, Zhang, Y. ling, Song, W., Cai, H. mei, Wang, Y., & Granato, D. (2019). Characterization of Brazilian coffee based on isotope ratio mass spectrometry (δ13C, δ18O, δ2H, and δ15N) and supervised chemometrics. *Food Chemistry*, *297*, 124963. https://doi.org/10.1016/j.foodchem.2019.124963

Pereira, L. L., Cardoso, W. S., Guarçoni, R. C., da Fonseca, A. F. A., Moreira, T. R., & Caten, C. S. ten. (2017). The consistency in the sensory analysis of coffees using Q-graders. *European Food Research and Technology*, *243*(9), 1545–1554. https://doi.org/10.1007/s00217-017-2863-9

Perez, M., Domínguez-López, I., López-Yerena, A., & Vallverdú Queralt, A. (2021). Current strategies to guarantee the authenticity of coffee. *Critical Reviews in Food Science and Nutrition*. https://doi.org/10.1080/10408398.2021.1951651

Putri, S., & Fukusaki, E. (2018). Achieving sustainable cultivation of coffee. In P. Lashermes (Ed.), *Achieving sustainable cultivation of coffee* (First Ed). Burleigh Dodds Science Publishing. London. https://doi.org/10.4324/9781351114363

Putri, S. P., Irifune, T., Yusianto, & Fukusaki, E. (2019). GC/MS based metabolite profiling of Indonesian specialty coffee from different species and geographical origin. *Metabolomics*, *15*(10), 126. https://doi.org/10.1007/s11306-019-1591-5

R Core Team. (2020). R: A language and environment for statistical computing. In *R Foundation for Statistical Computing* (4.0.3).

Risticevic, S., Carasek, E., & Pawliszyn, J. (2008). Headspace solid-phase microextraction–gas chromatographic–time-of-flight mass spectrometric methodology for geographical origin verification of coffee. *Analytica Chimica Acta*, *617*(1–2), 72–84. https://doi.org/10.1016/j.aca.2008.04.009

Robert, J. V., de Gois, J. S., Rocha, R. B., & Luna, A. S. (2022). Direct solid sample analysis using synchronous fluorescence spectroscopy coupled with chemometric tools for the geographical discrimination of coffee samples. *Food Chemistry*, *371*, 131063. https://doi.org/10.1016/j.foodchem.2021.131063

Rohart, F., Gautier, B., Singh, A., & Lê Cao, K.-A. A. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Computational Biology*, *13*(11), e1005752. https://doi.org/10.1371/journal.pcbi.1005752

Romano, R., Santini, A., Le Grottaglie, L., Manzo, N., Visconti, A., & Ritieni, A. (2014). Identification markers based on fatty acid composition to differentiate between roasted Arabica and Canephora (Robusta) coffee varieties in mixtures. *Journal of Food Composition and Analysis*, *35*(1), 1–9. https://doi.org/10.1016/j.jfca.2014.04.001

Rosiana, N. (2020). Dinamika Pola Pemasaran Kopi pada Wilayah Sentra Produksi Utama di Indonesia. *Jurnal Agrosains Dan Teknologi*, *5*(1), 1–10.

Rstudio, T. (2020). RStudio: Integrated Development for R. In *Rstudio Team, PBC, Boston, MA URL http://www.rstudio.com/* (1.3.1093). https://doi.org/10.1145/3132847.3132886

Santos, J. R., & Rodrigues, J. A. (2020). Characterization of volatile carbonyl compounds in defective green coffee beans using a fan assisted extraction process. *Food Control*, *108*, 106879. https://doi.org/10.1016/j.foodcont.2019.106879

Schipilliti, L., Bonaccorsi, I., Buglia, A. G., & Mondello, L. (2019). Comprehensive Isotopic Data Evaluation (CIDE) of Carbon Isotope Ratios for Quality Assessment and Traceability of Coffee. *Food Analytical Methods*, *12*(1), 121–127. https://doi.org/10.1007/s12161-018-1344-5

Scholz, M. B. D. S., Kitzberger, C. S. G., Pereira, L. F. P., Davrieux, F., Pot, D., Charmetant, P., & Leroy, T. (2014). Application of near infrared spectroscopy for green coffee biochemical phenotyping. *Journal of Near Infrared Spectroscopy*, *22*(6), 411–421. https://doi.org/10.1255/jnirs.1134

Selmar, D., Kleinwachter, M., & Bytof, G. (2014). Metabolic responses of coffee beans during processing and their impact on coffee flavor. In R. F. Schwan & G. H. Fleet (Eds.), *Cocoa and Coffee Fermentations* (1st Edition, pp. 432–462). CRC Press. London. https://doi.org/https://doi.org/10.1201/b17536

Sezer, B., Apaydin, H., Bilge, G., & Boyaci, I. H. (2018). Coffee arabica adulteration: Detection of wheat, corn and chickpea. *Food Chemistry*, *264*, 142–148. https://doi.org/10.1016/j.foodchem.2018.05.037

Silva, A. C. R., da Silva, C. C., Garrett, R., & Rezende, C. M. (2020). Comprehensive lipid analysis of green Arabica coffee beans by LC-HRMS/MS. *Food Research International*, *137*, 109727. https://doi.org/10.1016/j.foodres.2020.109727

Speer, K., & Kölling-Speer, I. (2006). The lipid fraction of the coffee bean. *Brazilian Journal of Plant Physiology*, *18*(1), 201–216. https://doi.org/10.1590/S1677-04202006000100014

Suhandy, D., & Yulia, M. (2017). The Use of Partial Least Square Regression and Spectral Data in UV-Visible Region for Quantification of Adulteration in Indonesian Palm Civet Coffee. *International Journal of Food Science*, *2017*, 1–7. https://doi.org/10.1155/2017/6274178

Sun, Z. J., Wells, D. P., Segebade, C., Maschner, H., & Benson, B. (2013a). A provenance study of coffee by photon activation analysis. *Journal of Radioanalytical and Nuclear Chemistry*, *296*(1), 293–299. https://doi.org/10.1007/s10967-012-2021-6

Tarazona, P., Feussner, K., & Feussner, I. (2015). An enhanced plant lipidomics method based on multiplexed liquid chromatography-mass spectrometry reveals additional insights into cold- and drought-induced membrane remodeling. *The Plant Journal*, *84*(3), 621–633. https://doi.org/10.1111/tpj.13013

Thammarat, P., Kulsing, C., Wongravee, K., Leepipatpiboon, N., & Nhujak, T. (2018). Identification of Volatile Compounds and Selection of Discriminant Markers for Elephant Dung Coffee Using Static Headspace Gas Chromatography—Mass Spectrometry and Chemometrics. *Molecules*, *23*(8), 1910. https://doi.org/10.3390/molecules23081910

Thorburn Burns, D., Tweed, L., & Walker, M. J. (2017). Ground Roast Coffee: Review of Analytical Strategies to Estimate Geographic Origin, Species Authenticity and Adulteration by Dilution. *Food Analytical Methods*, *10*(7), 2302–2310. https://doi.org/10.1007/s12161-016-0756-3

Toci, A. T., de Moura Ribeiro, M. V., de Toledo, P. R. A. B., Boralle, N., Pezza, H. R., & Pezza, L. (2018). Fingerprint and authenticity roasted coffees by 1H-NMR: the Brazilian coffee case. *Food Science and Biotechnology*, *27*(1), 19–26. https://doi.org/10.1007/s10068-017-0243-7

Toci, A. T., Farah, A., Pezza, H. R., & Pezza, L. (2016). Coffee Adulteration: More than Two Decades of Research. Critical Reviews in Analytical Chemistry, *46*(2), 83-92 https://doi.org/10.1080/10408347.2014.966185

Toci, A. T., Neto, V. J. M. F., Torres, A. G., & Farah, A. (2013). Changes in triacylglycerols and free fatty acids composition during storage of roasted coffee. *LWT - Food Science and Technology*, *50*(2), 581–590. https://doi.org/10.1016/j.lwt.2012.08.007

Valentin, J. L., & Watling, R. J. (2013). Provenance establishment of coffee using solution ICP-MS and ICP-AES. *Food Chemistry*, *141*(1), 98–104. https://doi.org/10.1016/j.foodchem.2013.02.101

Wang, X., Lim, L.-T., & Fu, Y. (2020). Review of analytical methods to detect adulteration in coffee. *Journal of AOAC International*, *103*(2), 295–305. https://doi.org/10.1093/JAOCINT/QSZ019

Watson, A. D. (2006). Lipidomics: A global approach to lipid analysis in biological systems. *Journal of Lipid Research*, *47*(10), 2101–2111. https://doi.org/10.1194/jlr.R600022-JLR200

Wei, F., Furihata, K., Koda, M., Hu, F., Kato, R., Miyakawa, T., & Tanokura, M. (2012). ^{13}C NMR-based metabolomics for the classification of green coffee beans according to variety and origin. *Journal of Agricultural and Food Chemistry*, *60*(40), 10118–10125. https://doi.org/10.1021/jf3033057

Wilson, B. R., Conley, J. F., Harris, T. M., & Lafone, F. (2012). New terrains of taste: Spatial analysis of price premiums for single origin coffees in Central America. *Applied Geography*, *35*(1–2), 499–507. https://doi.org/10.1016/j.apgeog.2012.10.004

Worku, M., Upadhayay, H. R., Latruwe, K., Taylor, A., Blake, W., Vanhaecke, F., Duchateau, L., & Boeckx, P. (2019). Differentiating the geographical origin of Ethiopian coffee using XRF- and ICP-based multi-element and stable isotope profiling. *Food Chemistry*, *290*, 295–307. https://doi.org/10.1016/j.foodchem.2019.03.135

Yang, N., Liu, C., Liu, X., Degn, T. K., Munchow, M., & Fisk, I. (2016). Determination of volatile marker compounds of common coffee roast defects. *Food Chemistry*, *211*, 206–214. https://doi.org/10.1016/j.foodchem.2016.04.124

Yang, S., Li, C., Mei, Y., Liu, W., Liu, R., Chen, W., Han, D., & Xu, K. (2021). Determination of the Geographical Origin of Coffee Beans Using Terahertz Spectroscopy Combined With Machine Learning Methods. *Frontiers in Nutrition*, *8*, 1–10. https://doi.org/10.3389/fnut.2021.680627

Yener, S., Romano, A., Cappellin, L., Märk, T. D., Del Pulgar, J. S., Gasperi, F., Navarinic, L., Biasioli, F., Sánchez del Pulgar, J., Gasperi, F., Navarini, L., Biasioli, F., Del Pulgar, J. S., Gasperi, F., Navarinic, L., & Biasioli, F. (2014). PTR-ToF-

MS characterisation of roasted coffees (*C. arabica*) from different geographic origins. *Journal of Mass Spectrometry*, *49*(9), 929–935. https://doi.org/10.1002/jms.3455

Zambonin, C. G., Balest, L., De Benedetto, G. E., & Palmisano, F. (2005). Solid-phase microextraction–gas chromatography mass spectrometry and multivariate analysis for the characterization of roasted coffees. *Talanta*, *66*(1), 261–265. https://doi.org/10.1016/j.talanta.2004.11.023

Zhu, M., Long, Y., Ma, Y., Chen, Y., Yu, Q., Xie, J., Li, B., & Tian, J. (2021). Comparison of chemical and fatty acid composition of green coffee bean (Coffea arabica L.) from different geographical origins. *LWT*, *140*, 110802. https://doi.org/10.1016/j.lwt.2020.110802