

# IMPACT OF THE HARVESTING YEAR ON THE POSSIBLE AUTHENTICATION OF GEOGRAPHICAL ORIGIN OF GREEN COFFEA ARABICA USING PROFILE OF VOLATILES

Alžbeta Demianová\*, Alica Bobková, Lukáš Jurčaga, Marek Bobko, Ľubomír Belej, Katarína Poláková, Miroslav Žiak

#### Address(es):

Slovak University of Agriculture in Nitra, Faculty of Biotechnology and Food Sciences, Institute of Food Sciences, Trieda Andreja Hlinku 2, 949 76 Nitra, Slovakia.

\*Corresponding author: xdemianova@uniag.sk

ABSTRACT

https://doi.org/10.55251/jmbfs.5407

#### ARTICLE INFO

Received 12. 11. 2021 Revised 23. 9. 2022 Accepted 3. 11. 2022 Published 1. 12. 2022

Regular article

Authentication of coffee is highly required. This study aimed to understand the relative abundance of volatiles of green coffee harvested during two years. Using volatiles (GC-MS) and Linear Discriminant Analysis, we focused on the geographical origin identification. We analyzed samples of green *Coffea arabica* from Africa, Central America, and South America, harvested in 2018 and the same samples harvested in 2019. A total of 215 different volatiles were detected. Based on their chemical structure and the functional chemical group, they were divided into categories: furan derivates, aldehydes, ketones, alcohols, organic acids, hydrocarbons (alkanes, alkenes, alkynes, aromatic hydrocarbons), terpenoids, heterocyclic compounds, nitriles, amines. Green Arabica contained mostly organic acids and esters, aldehydes, hydrocarbons, and alcohols. We observed significant differences in aromatic hydrocarbons and furan derivates by comparing the volatiles profiles of African coffee beans collected in 2019 and 2019. The profile of Central American samples (both years) was homogenous; thus, no significant differences were observed. The aroma profile of South American coffees harvested in 2018 was explained by organic acids and esters, alkenes applying LDA. More than 94% of the variability between Africa, Central, and South America coffees harvested in 2018 was explained by organic acids and esters, alkenes, alkenes, alkenes, alkenes, and esters, alkenes to aldehydes and ketones, organic acids and esters, alkenes, terpenoids, and aldehydes. These appear to be useful for geographical authentication regardless of the year of harvesting.

Keywords: green coffee, authentication, harvesting year, gas chromatography, linear discriminant analysis

## INTRODUCTION

Averagely 165 million 60 kg bags of coffee are produced every year, meaning that coffee is one of the most sought beverages, especially in the western world, and the second most traded commodity in the world market. Coffee quality is defined by chemical composition and organoleptic properties. The final quality of green beans and the content of the chemical substances can be affected by various reasons, e.g., coffee species, altitude, soil composition, variations of temperature, thus mainly by the geographical region of the growing area (Kathurima et al., 2000; Musanto et al., 2011; Fassio et al., 2017). Based on these facts, several scientific studies have focused on the effect of environmental factors, such as growing altitude, shade, and harvest period, on coffee plants' quality, biochemical and chemical composition (Toleso et al., 2016).

Coffee beans, as a plant material, may contain almost a thousand different chemical compounds and isomers. As it was reported, green coffee has a complex composition. The main compounds are polysaccharides, monosaccharides, lipids, sterols, fatty acids, phenolic acids, polyphenols, alkaloids, proteins, free amino acids, vitamins, and minerals. Moreover, great attention is paid to substances with an antioxidant effect (e.g., chlorogenic acids and their isomers, hydroxycinnamic acids, caffeine, caffeic acid, phenolic compounds, and others). These can scavenge free radicals and thus protect the body from oxidative stress (Yılmaz et al., 2014; Semen et al., 2017). Undoubtedly, the vast popularity of coffee strongly correlates with its sensory properties, primarily characteristic of aroma and flavor. Certain volatiles, named as the key components, are responsible for the final sensory properties. The main components of green coffee were identified as isoamylalcohol (10.4%), hexanal (10.4%) and hexocosane (8.2%), while furfurylalcohol (13.6%), furfurylacetate (10.7%), and 5-methyl furfural (9.27%) were identified as the main components of the roasted coffee (Poyraz et al., 2016). Used HS-SPME-GC-MS, was able to quantify different volatile compounds like alcohols, aldehydes, ketones, pyrazines, pyridines, and furans in green and roasted coffee beans, which belong to different chemical classes. However, given applied temperature and time, the concentration of volatile compounds in roasted coffee can undergo dramatic changes depending on the thermal profile applied during the roasting process and strongly depends on the green coffee bean composition (**Caporaso et al., 2018**). Given world-wide consumption and common illegal practice in whole food industry, it was necessary to create approaches and analysis suitable for the authentication of coffee (**Toci et al., 2015**). Specific analytical techniques are used to identify possible adulteration of coffee. However, recently technique of fingerprinting or profiling, which is represented by the targeted identification of the statistical approach, may show a vital conclusion regarding the authentication process (**Suoto et al., 2015**).

The main aim of our research was to study the profile of volatile compounds in green beans of *Coffea arabica* from various geographical regions from Africa, Central America, and South America repeatedly for 2 harvesting years. And based on the obtained profiles develop a statistical model using machine learning approaches, especially Linear Discriminant Analysis (LDA), and determine the impact of harvesting year on the possibility of authenticating the geographical origin of green *C.arabica*.

# MATERIAL AND METHODS

#### Material

Samples of green *Coffea arabica* were purchased from Barzzuz (Banská Bystrica, Slovakia). The company focuses on importing and processing coffee from various well-known producers. Eighteen samples originating from Africa, Central America, and South America (the origin of samples was proved by the certificated) were analyzed. The first group of samples (group A) were harvested in 2019, rest for a year, and were analyzed in 2021. The second group (group B) was harvested in the same regions in 2018, rest for a year, and then were analyzed in 2020. A detailed description of samples is shown in Table 1, and the geographical range is shown in Figure 1.

Table 1 Description of analyzed samples of green Coffea arabica

ID	Continent	Country	Variety	Processing	Region	Altitude (mamsl)
1	Africa	Burundi	В	W	Kayanza	1700-1800
2	Africa	Rwanda	K,B	W	Gasetsa	1500-1700
3	Africa	Ethiopia	Н	W	Yigacheffe	1700-2000
4	South America	Columbia	C1,B,T	W	Multiple	1200-2000
5	South America	Brazil	В	PN	São Sebastião da Grama	1150-1350
6	South America	Peru	B,C1	W	SHB Monte Verde Amazona	1400-1800
7	Central America	Costarica	C1	W	San Rafael Tarrazu	1200-2000
8	Central America	Guatemala	В	W	Antigua Guatemala, Sacatepequez	1450-1800
9	Central America	El Salvador	В	W	Apaneca	1250
Mater Maniatan D	Doughou C1 Cotom	- C Calaba II	Haidaan V Van	T Tomin W/ (mar	DN	

Note: Variety: B – Bourbon, C1 – Caturra, G – Geisha, H – Heirloom, K – Kent, T – Typica, W (washed, wet processed), PN - pulped natural; mamsl-meters above mean sea level



Figure 1 Geographical range of analyzed samples

### Determination of volatile compounds using gas chromatography (GS-MS)

Twenty grams of homogenized green coffee beans were put into 40 mL glass vials with septum Archon caps ptfe/sil. The samples was warmed up to 35 °C for 15 minutes in Metaltermoblock Liebisch Labortechnik. Self-sorption was performed with Fiber: Carboxen®/PDMS (CAR/ PDMS) 2 cm, time 30 minutes at 35 °C, followed by GC-MS analysis. Volatiles were determined according to method previously described in Sádecká et al., (2014) with modifications. The Agilent Technologies 6890 gas chromatograph (Agilent Technology, Palo Alto, USA) equipped with an Agilent Technologies 5973 selective inertial detector (MSD). Volatiles were separated using a J&W 122-7333 DB-WAXetr 30 m x 0.25 mm x 0.5 µm capillary column. The carrier gas was helium. The injector temperature was set to 250 °C. The oven temperature was programmed to be isothermal at 50 °C for 1 minute, then heated to 250 °C at a rate of 5 °C.min<sup>-1</sup>. Input parameters were as follows: splitless mode, initial temperature: 250 °C, pressure: 88.9 kPa, flow rate: 20.0 mL.min<sup>-1</sup>, cleaning time: 1.00 min, total flow rate: 24.6 mL.min<sup>-1</sup>. Electron ionization (EI) was set to 70 eV. The transfer line temperature and ion source temperature were 280 °C. The mass spectrometer collected data in full scan mode. Identification was performed by comparing the target compounds' mass spectra and chromatography data with reference materials in the NIST 14 library.

#### **Statistical Analysis**

For the summarizing and describing of our results descriptive statistic was used. To discover any possible significant differences within selected geographical groups of green coffee beans, ANOVA Duncan test and REGWQ was used. This statistical analysis was performed using Microsoft Office Excel 365 for iOS. Linear Discriminant Analysis was used to create a model that would be able to determine the geographical origin of single bean green *Coffea arabica* (XL Stat for iOS, Addinsoft), and Tanagra (Lyon 2 University Berges du Rhone Campus).

#### **RESULTS AND DISCUSSION**

#### Profile of volatiles in green coffee harvested in 2018

Obtained volatiles were divided into several categories based on their chemical composition and functional group. Selected categories were furan derivates, aldehydes, alcohols, organic acids and their esters, terpenoids, hydrocarbons (separately divided into alkanes, alkenes, alkynes, and aromatic hydrocarbons), heterocyclic compounds, nitriles, ketones, amines.

African samples showed that in 2018 the most abundant compounds were alcohols, organic acids and their esters, aldehydes, terpenoids, and aromatic hydrocarbons. These observations are in accordance with **Tsegay et al.**, (2019), who reported that the most abundant groups in green coffee are alcohols and organic acids. Among organic acids, we observed mainly acetic acid; hexanoic acid; 2-butenoic acid; 3-methyl-; pentanoic acid, and ethyl acetate. These reached the highest concentrations in the group. The most abundant alcohols were ethanol, 1-penthanol; 1-hexanol; maltol; and 2,3-butanediol. Given that furan is considered a potential carcinogen and strict legislation is applied, furan derivates were evaluated separately. Our samples from Africa showed a concentration range from 3.26 - 12.19%. This by far highest value was detected in a sample from Burundi, where furfural; 2- methyl furan; 3-methyl furan, and furan 2-methanol were abundant the most.

On the other hand, furan derivates concentration of African samples harvested in 2019 showed significantly different values (p-value < 0.05) ranging from 3.26 to 5.61%, however, the highest concentration within the current year was measured again in the Burundi sample. The concentration of aromatic hydrocarbons was significantly higher in African coffees in 2019. In 2018, GC-MS detected mainly benzene, 1,3-dimethyl-; toluene; and o-xylene, however, only in minor concentrations. On the other hand, in 2019, this group reached 19.81%, containing mainly methylbenzene. Average concentrations of each chemical group are shown in Figure 2.



Figure 2 Average values of each chemical group of volatiles regarding year of harvesting of green *Coffe arabica* from Africa. Note: B-year 2018, A-year 2019; significantly different values (P-value < 0.05) are marked with upper index\*

J Microbiol Biotech Food Sci / Demianová et al. 2022 : 12 (3) e5407

Within South America green coffee, we observed that none of the selected chemical groups showed a significant difference (p-value > 0.05) in observed harvesting years (Figure 3). **Procida et al. 2020** focused their research on the volatiles profiling of green Arabica coffee from various continents, including South America. They observed that 4-methyl-2,3-dihydrofuran, n-hexanol, limonene and nonanal, appear involved in the characterization of the geographical origin of the analyzed samples. Apart from 4-methyl-2,3-dihydrofuran, these were also identified in our samples. However, a direct comparison cannot be made given mentioned study did not focus on the robustness of these substances regarding the geographical origin during two years of harvesting.

The last observed group was Central America. Comparably to African samples, ANOVA - REGWQ shows significant differences (p-value < 0.05) in average values of aromatic hydrocarbons. In 2018, this group had the highest concentrations in Costa Rica (17.67%), mainly methylbenzene (11.19%). The lowest concentration of aromatic hydrocarbons was detected in Guatemala

(0.30%). In 2019 Guatemala, showed relatively similar values (0.71%), but Costa Rica contained only 0.37%, and methylbenzene was measured in minor concentrations. Another significant difference was in alkanes. Samples from 2018 contained mainly pentane, 1-chloro-; butane, 1-chloro-, and undecane. Whereas samples from 2019 contained undecane, 5-ethyl-; decane; and cyclopropane, 1,1-dimethyl. Average values of the group are shown in Figure 5.

Importantly, all three selected geographical groups showed the same pattern in volatiles profile regardless the year of harvesting. The highest concentration range was measured within alcohols, organic acids and esters, and other groups reached relatively similar concentrations, which indicates that the profile of volatiles may be stable enough in order to be used further in combination with an advanced statistical approach to develop a statistical model suitable for geographical origin identification of green *Coffea arabica*.



Figure 4 Average values of each chemical group of volatiles regarding year of harvesting of green *Coffea arabica* from South America Note: B – year 2018, A – year 2019



Figure 5 Average values of each chemical group of volatiles regarding year of harvesting of green Arabica coffee from Central America Note: B – year 2018, A – year 2019; significantly different values (p-value < 0.05) are highlighted red

Even though the detection of volatiles is considered routine, geographical origin authentication of green coffee based on fingerprinting and advanced statistical approach is still limited, given that many are focused on roasted coffee (**Caporaso** et al., 2018). Ding et al., (2018) and Bobková et al., (2021) previously used Linear Discriminant Analysis (LDA) to distinguish the geographical origin of *Cynomorium songaricum* and *Camelia sinensis* from various countries. That indicates its possible use for other commodities in the food sector. LDA is a dimensionality reduction technique generally used to classify patterns between classes. LDA is a classifier used to find a linear combination of features, separating two or more data classes. The succeeding combination can be used as a linear classifier. This approach tries to maximize the between-class variance and minimize the within-class variance (Forina et al., 2009). Firstly, by applying LDA on the profile of volatiles detected in samples harvested in 2018, we observed that Bartlett's test reached significant values (p-value < 0.05). The test represents the significance of the model and how well each function separates cases into groups and confirm that the within-class covariance matrices are different. Moreover, LDA proved that only two factors are needed to explain 100% of the variability between three geographical groups. Variables representation shows which initial variables (groups of volatiles) are correlated with these factors. From Figure 6 is obvious that 94.22% of the variability is explained by alcohols, organic acids, ketones, aldehydes, and alkenes. The rest of the variability. Furthermore, LDA performed in Tanagra accordingly evaluated the most significant input parameters to be organic acids and esters (p-value = 0.07), alkenes (p-value = 0.0144), and ketones (p-value = 0.00025).





These data confirm that the groups are discriminated, and continents are well separated, especially African samples are clearly separated. However, on the LDA map (Figure 6) we see that Central and South American centroids tends to each other, which can be explained by the close geographical locality and similarity in climatic, environmental conditions, and altitude of growing areas.

To observe the harvesting year's impact on the possibility of identifying the geographical origin of green coffee, we added samples harvested in 2019 to the LDA model. In this case, Rao's approximation and Bartlett's test reached significance (p-value < 0.05). This indicated that the model could distinguish selected continents on a significant level. As in 2018, the most significant parameter among all were ketones (p-value < 0.05). Nevertheless, in 2019, organic acids, alcohols, and alkenes parameters did not reach the same level. These previously significant parameters were substituted by aldehydes (p-value = 0.045) and terpenoids (p-value = 0.048). This indicated that year of harvesting; its climate conditions, respectively, may affect the volatile profile of green *Coffea arabica*. Bertrand et al. (2012) also confirmed that climate change, which generally involves a substantial increase in average temperatures in mountainous tropical regions, could impact the quality of coffee aroma, thus volatiles.

To quantify the model's functionality. Given the number of samples, the leave-oneout cross-validation was used. This process is repeated several times, always using different sample. As an output, the Confusion matrix for training samples, Crossvalidation of prior and posterior classification, and membership probabilities are calculated. The latter is a matrix that predicts the classification accuracy.

In our study, the confusion matrix for training samples reached 100% accuracy of origin identification regardless of the harvesting year. According to the confusion matrix, LDA could identify all training samples in all three continents correctly. However, the Cross-validation of prior and posterior classification, and membership probabilities contained two miss-classified samples from Africa identified as South American coffee, one sample from South America identified as South American coffee. These results showed that the predict overall accuracy of the model was lowered to 77.78% (Table 2).

from $\ to$	Africa	Central America	South America	Total	% correct
Africa	4	0	2	6	66.66
Central America	0	5	1	6	83.33
South America	1	0	5	6	83.33
Total	5	5	8	18	77.77

Results indicates that not all input parameters are needed for the identification of geographical origin regardless the year of harvesting. Given this justified argument we re-run LDA using only organic acids and ester, aldehydes, ketones, alkenes, and terpenoids. In this case Rao's approximation and Bartlett's test verify the significance (p-value < 0.05). The Confusion matrix for training samples reached 100% accuracy of origin identification regardless of the harvesting year. According to the confusion matrix, LDA could correctly identify all training samples in all three continents. But the Confusion matrix for the cross-validation results contained one miss-classified samples from Africa identified as Central American coffee, These results showed that the predict overall accuracy of the model was lowered to 88.89% (Table 3).

from \ to	Africa	Central America	South America	Total	% correct
Africa	5	1	0	6	83.33
Central America	0	5	1	6	83.33
South America	0	0	6	6	100.00
Total	5	6	7	18	88.89

### CONCLUSION

Given coffee's popularity, geographical authentication of green beans is very required in the coffee industry. Our research showed that GC-MS analysis of volatiles provides sufficient information suitable for further statistical approaches. We observed significant differences in aromatic hydrocarbons and furan derivates by comparing the volatiles profiles of African coffee beans collected in 2018 and 2019. The profile of Central American coffee beans from two years was homogenous. Thus, no significant differences were observed. The aroma profile of South American coffees showed significant differences in aromatic hydrocarbons and alkanes (p-value < 0.05).

To create a model suitable for geographical origin identification, LDA was used. Analyzed samples from 2018 suggest that the most significant parameters among volatiles are organic acids and esters, alkenes, and ketones. According to LDA, these parameters explained most of the variability between African, South American, and Central American coffee. Coffee is a biological material; thus, the impact of harvesting years must be considered. Therefore, samples from the same growing areas harvested in 2019 were added to LDA model. The confusion matrix showed that training samples were correctly identified among all geographical groups regardless of the year's harvest. However, leave-one-out testing showed several misclassified samples that lowered the model's overall accuracy. Furthermore, based on LDA calculations we could reduce input parameters and determined organic acids and esters, alkenes, terpenoids, aldehydes, and ketones to be most useful for geographical authentication regardless the year of harvesting. Results showed that harvesting year might affect the volatiles profile of coffee. More importantly, samples from both years suggest that the group of ketones were the most significant parameters in geographical origin identification regardless of the year. Therefore, a more detailed investigation of ketones in green coffee is needed, given that group may contain individual markers suitable for origin identification that can simplify whole identification process.

Acknowledgments: This research has been supported by The Ministry of Education, Science, Research and Sport of the Slovak Republic, grant VEGA 1/0734/20

### REFERENCES

Bertrand, B., Boulanger, R., Dussert, S., Ribeyre, F., Berthiot, L., Descroix, F., & Joët, T. (2012). Climatic factors directly impact the volatile organic compound fingerprint in green Arabica coffee bean as well as coffee beverage quality. *Food Chemistry*, 135(4), 2575–2583. <u>http://doi:10.1016/j.foodchem.2012.06.060</u>

Bobková, A., Demianová, A., Belej, E., Harangozo, E., Bobko, M., Jurčaga, L., Poláková, K., Božiková, M., Bilčík, M., Árvay, J. (2021). Detection of Changes in Total Antioxidant Capacity, the Content of Polyphenols, Caffeine, and Heavy Metals of Teas in Relation to Their Origin and Fermentation. *Foods*, 10(8), 1821. http://doi:10.3390/foods10081821

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. http://doi:10.1016/j.foodres.2018.03.077

Ding, J., Gu, C., Huang, L., & Tan, R. (2018). Discrimination and Geographical Origin Prediction of Cynomorium songaricum Rupr. from Different Growing Areas in China by an Electronic Tongue. *Journal of Analytical Methods in Chemistry*, 2018, 1–6. http://doi:10.1155/2018/5894082

Fassio, L. O., Malta, M. R., Carvalho, G. R., Liska, G. R., Lima, P. M., Nadaleti, D. H. S., Pimenta, C. J. (2017). Fatty Acids Profile of Coffea arabica L. Resistant to Leaf Rust Grown in Two Environments of Minas Gerais, Brazil. *Journal of Agricultural Science*, 9(12), 88. <u>http://doi:10.5539/jas.v9n12p88</u>

Forina, M., Casale, M., & Oliveri, P. (2009). Application of Chemometrics to Food Chemistry. Comprehensive Chemometrics, 75–128. <u>http://doi:10.1016/b978-044452701-1.00124-1</u>

Kathurima, C., Gichimu, B.M., Kenji, G.M., Muhoho, S.M., Boulanger, R. (2000). Evaluation of beverage quality and green bean physical characteristics of selected Arabica coffee genotypes in Kenya. *African Journal of Food Science*, 3, 365–371. Mussatto, S. I., Machado, E. M. S., Martins, S., & Teixeira, J. A. (2011). Production, Composition, and Application of Coffee and Its Industrial Residues. *Food and Bioprocess Technology*, 4(5), 661–672. <u>http://doi:10.1007/s11947-011-0565-z</u>

Poyraz, E. İ., Öztürk, N., Kıyan, H. T., & Demirci, B. (2016). Volatile compounds of Coffea arabica L. green and roasted beans. *Anadolu University Journal Of Science And Technology –C Life Sciences and Biotechnology*, 5(1). http://doi:10.18036/btdc.13390

Procida, G., Lagazio, C., Cateni, F., Zacchigna, M., & Cichelli, A. (2020). Characterization of Arabica and Robusta volatile coffees composition by reverse carrier gas headspace gas chromatography-mass spectrometry based on a statistical approach. *Food Science and Biotechnology*, 29(10), 1319–1330. http://doi:10.1007/s10068-020-00779-7

Sádecká, J., Kolek, E., Pangallo, D., Valík, L., & Kuchta, T. (2014). Principal volatile odorants and dynamics of their formation during the production of May Bryndza cheese. *Food Chemistry*. 150, 301–306. https://doi.org/10.1016/j.foodchem.2013.10.163

Semen, S., Mercan, S., Yayla, M., & Açıkkol, M. (2017). Elemental composition of green coffee and its contribution to dietary intake. *Food Chemistry*, 215, 92–100. <u>http://doi:10.1016/j.foodchem.2016.07.176</u>

Souto, U. T. de C. P., Barbosa, M. F., Dantas, H. V., de Pontes, A. S., Lyra, W. da S., Diniz, P. H. G. D., de Aráujo, U.C.M., da Silva, E. C. (2014). Screening for Coffee Adulteration Using Digital Images and SPA-LDA. *Food Analytical Methods*, 8(6), 1515–1521. <u>http://doi:10.1007/s12161-014-0020-7</u>

Toci, A.T., Farah, A., Pezza, H.R., Pezza, L. (2015). Coffee Adulteration: More than Two Decades of Research. *Analytical Chemistry*. 46(2),83–92. http://dx.doi.org/10.1080/10408347.2014.966185

Tolessa, K., D'heer, J., Duchateau, L., & Boeckx, P. (2016). Influence of growing altitude, shade and harvest period on quality and biochemical composition of Ethiopian specialty coffee. *Journal of the Science of Food and Agriculture*, 97(9), 2849–2857. http://doi:10.1002/jsfa.8114

Tsegay, G., Redi-Abshiro, M., S. Chandravanshi, B., Ele, E., M. Mohammed, A., & Mamo, H. (2019). Volatile profile of green coffee beans from Coffea arabica L. plants grown at different altitudes in Ethiopia. *Bulletin of the Chemical Society of Ethiopia*, 33(3), 401. <u>http://doi:10.4314/bcse.v33i3.2</u>

Yilmaz, P. K.; Hacibekiroğlu, I.; Kolak, U. (2014). Effect of roasting on antioxidant and anticholinesterase capacities of coffee. *Journal of Food and Nutrition Research*, 53(3), pp. 232-239.