




Gas chromatography/mass spectrometry-based metabolite profiling of coffee beans obtained from different altitudes and origins with various postharvest processing

Fitri Amalia¹ · Pingkan Aditiawati² · Yusianto³ · Sastia Prama Putri^{1,2,4}  · Eiichiro Fukusaki^{1,4,5}

Received: 9 November 2020 / Accepted: 2 July 2021 / Published online: 12 July 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Introduction Coffee is a popular beverage because of its pleasant aroma and distinctive flavor. The flavor of coffee results from chemical transformations influenced by various intrinsic and extrinsic factors, including altitude, geographical origin, and postharvest processing. Despite is the importance of grading coffee quality, there is no report on the dominant factor that influences the metabolomic profile of green coffee beans and the correlated metabolites for each factor.

Objective This study investigated the total metabolite profile of coffees from different altitudes and coffees subjected to different postharvest processing.

Method Arabica green coffee beans obtained from different geographical origins and different altitudes (400 and 800 m) and produced by different postharvest processes (dry, honey, and washed process) were used in this study. Coffee samples obtained from altitudes of 400–1600 m above sea level from various origins that were produced by the washed method were used for further study with regard to altitudes. Samples were subjected to gas chromatography/mass spectrometry (GC/MS) analysis and visualized using principal component analysis (PCA) and orthogonal partial least squares (OPLS) regression analysis.

Results The PCA results showed sample separation based on postharvest processing in PC1 and sample separation based on altitude in PC2. A clear separation between samples from different altitudes was observed if the samples were subjected to the same postharvest processing method, and the samples were of the same origin. Based on this result, OPLS analysis was conducted using coffee samples obtained from various altitudes with the same postharvest processing. An OPLS model using altitude as a response variable and 79 metabolites annotated from the GC/MS analysis as an explanatory variable was constructed with good R^2 and Q^2 values.

Conclusion Postharvest processing was found to be the dominant factor affecting coffee metabolite composition; this was followed by geographical origin and altitude. The metabolites glutamic acid and galactinol were associated with the washed and honey process, while glycine, lysine, sorbose, fructose, glyceric acid, and glycolic acid were associated with the dry process. Two metabolites with high variable influence on projection scores in the OPLS model for altitude were inositol and serotonin, which showed positive and negative correlations, respectively. This is the first study to report characteristic coffee metabolites obtained from different altitudes.

Keywords Coffee · Altitude · Postharvest process · Gas chromatography/mass spectrometry · Orthogonal partial least squares

✉ Sastia Prama Putri
sastia_putri@bio.eng.osaka-u.ac.jp

¹ Department of Biotechnology, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

² School of Life Sciences and Technology, Institut Teknologi Bandung, Jl. Ganesha No.10, Lb. Siliwangi, Coblong, Kota Bandung, Jawa Barat 40132, Indonesia

³ Indonesian Coffee and Cocoa Research Institute, Jl. PB. Sudirman 90, Jember, Indonesia

⁴ Industrial Biotechnology Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

⁵ Osaka University Shimadzu Omics Innovation Research Laboratories, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan

1 Introduction

Coffee is a beverage consumed globally owing to its pleasant aroma and flavor (Cordoba et al., 2019). Hence, flavor and aroma are essential aspects to determine coffee quality (Buffo, 2018). The sensorial aspects of coffee are correlated with non-volatile compounds that are derived from a long chain of chemical transformations from seed to cup (Joët & Dussert, 2018; Seninde and Chambers IV, 2020). The sensorial quality of coffee is closely related to the biochemical composition of raw, green coffee beans (Worku et al., 2018). Hence, the biochemical composition of green coffee beans was investigated in previous studies (Joët et al., 2010; Worku et al., 2018). Green coffee beans are non-aromatic (contain green, musty notes) and contain many non-volatile compounds that act as flavor and aroma precursors for numerous volatile compounds generated through the roasting process (Fisk et al., 2012; Joët & Dussert, 2018). Pyridines and pyrroles are examples of volatile compound classes found in roasted coffee and are produced from the reaction between trigonelline and amino acids during the roasting process (Fisk et al., 2012; Tsegay et al., 2019; Seninde and Chambers IV, 2020). Therefore, the biochemical composition of green coffee beans is crucial for identifying the key precursors of coffee flavor and aroma. The biochemical composition of green coffee beans may be influenced by several factors, including species, geographical origin, postharvest processing, roasting, and brewing methods (Joët & Dussert, 2018; Putri et al., 2019; Sunarharum et al., 2014).

Arabica and Robusta are two commercially available species that account for 60% and 40% of the worldwide production, respectively (Clifford et al., 2018). Arabica and Robusta have distinct aromas and flavors, wherein Arabica is usually described as having a pleasant aroma, higher acidity, and is less bitter than Robusta, while Robusta is described as bitter, mild, and with earthy notes (Hoffmann, 2014; Joët & Dussert, 2018). The differences in flavors are due to their biochemical composition with Arabica green coffee beans having notably lower caffeine and malic acid content than Robusta beans (Putri et al., 2019). After roasting, Arabica coffee beans contain more 2-acetylfurans and furfurals than Robusta coffee beans. This enhances the taste properties of balsam and almond in Arabica coffee beans (Seninde and Chambers IV, 2020). Owing to its superior taste, Arabica coffee is widely consumed in the US and European countries and is sold as a specialty coffee grade with 20–60% higher price than that of regular coffee (Sittipod et al., 2019; Tolessa, 2017). A specialty coffee grade is given to coffee with a cup score higher than 80, according to sensory evaluation performed by trained panelists called Q-graders whose evaluation is based on the Specialty Coffee Association of

America (SCAA) standard procedure (SCAA, 2015; Seninde and Chambers IV, 2020).

Arabica coffee is produced by countries located in the tropical belt. Interestingly, the flavor quality of Arabica differs depending on the geographical origin; therefore, high-grade coffee is commonly marketed as single-origin coffees (Purnomo et al., 2019). Previous metabolomics studies have reported the influence of sugars, amino acids, and organic acids on samples based on geographical origin (Putri et al., 2019; Wei et al., 2012a, 2012b). The metabolomics approach has also been used to investigate the influence of soil and climate conditions on the chemical composition of coffee beans (da Silva Taveira et al., 2014). Geographical origin encompasses many factors that result in the uniqueness of each single-origin coffee; among these factors, altitude and place of cultivation are generally believed to affect the flavor of Arabica coffee (Joët et al., 2010). Coffee grown at high altitudes (1000–1200 m above sea level) has good sensory quality with better aroma, acidity, and fruitiness than that grown at lower altitudes featuring earthy and green notes (Avelino et al., 2007; Bertrand et al., 2012; Davis et al., 2012; Schroth et al., 2015). Studies on the correlation between biochemical composition and altitudes have only analyzed targeted metabolites, including sugars, chlorogenic acid, and caffeine (Joët et al., 2010). However, whereas an earlier study reported that caffeine content was unaffected by altitude (Joët et al., 2010), a more recent study reported that caffeine content decreased as altitude increased (Tolessa et al., 2017). Increase (Avelino et al., 2005) or decrease (Tolessa et al., 2017) of chlorogenic acid content with increasing altitude has also been reported. Therefore, future studies covering broader metabolite coverage have been suggested to better clarify the relationship between altitude and coffee chemical composition (Joët et al., 2010).

The final coffee quality is vastly affected by postharvest processing. A series of steps, including the removal of the pericarp, flesh, and mucilage from the fruit, are carried out to obtain dried coffee beans or green coffee beans suitable for roasting (de Melo Pereira et al., 2019; Lee et al., 2015). There are three types of postharvest methods: dry, hybrid, and washed. The dry process or natural process method is performed by drying the harvested coffee fruit under the sun and turning it periodically for 10–25 days. After the drying process, the removal of other coffee fruit parts was conducted to yield coffee beans. The washed process or wet process method is performed by mechanically de-pulping the pericarp and flesh of coffee fruit. Subsequently, mucilage from coffee beans was removed by fermentation for 24–48 h. Afterwards, the coffee beans are then dried under the sun until the water content reaches 11–13%. The hybrid process is a combination of dry and washed processes, in which de-pulping the pericarp and flesh of coffee fruit are first performed, followed by sun-drying of coffee beans with

the mucilage layer intact. The honey process and pulped natural process fall under the classification of a hybrid process (Hoffmann, 2014; Lee et al., 2015). The type of post-harvest process method employed affects the distinct sensory properties of Arabica coffee beans, with the washed process coffee beans having fruity and acidic attributes and less bitterness (Lee et al., 2015). This sensory profile of washed-process coffee beans is developed during fermentation step when microbial activity yields reducing sugars, amino acids, organic acids aldehydes, esters, alcohols, and ketones that contribute to the aroma and flavor of the roasted beans and coffee brew (Seninde and Chambers IV, 2020; Ruta & Farcasanu, 2021). In addition, the selection of the postharvest process method also reflected in the chemical composition of green coffee beans. A chemical composition study of green coffee beans obtained from dry process and washed process methods has shown that the green coffee beans obtained from dry process have a higher content of sugar compounds, caffeic acid, caffeine, trigonelline, and several caffeoylquinic acid isomers, whereas green coffee beans obtained from the washed process are high in citric acid, erythritol, glutamic acid, and aspartic acid (Bytof et al., 2005; De Bruyn et al., 2017; Knopp et al., 2006; Selmar et al., 2002). Based on these studies, the employed postharvest process method was found to influence the chemical composition of green coffee beans and their sensory properties. Therefore, along with geographical origin and altitude, the postharvest process is an important factor influencing coffee quality.

Geographical origin, altitude, and postharvest process were described as important factors for the biochemical composition of green coffee beans and overall coffee quality. These factors were studied separately in previous studies using various analytical instruments (Joët et al., 2010; Bertrand et al., 2012; De Bruyn et al., 2017; Putri et al., 2019; Tsegay et al., 2020). The effect of postharvest processes on coffee quality were studied using metabolite profiling approach by multiple analytical instruments such as high-performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD) for free carbohydrate and sugar alcohols, gas chromatography with flame ionization detection (GC-FID) for short-chain fatty acids and ethanol, ultraperformance liquid chromatography coupled to mass spectrometry (UPLC-MS) for organic acids, and high-performance liquid chromatography coupled to tandem mass spectrometry (HPLC-MS/MS) for amino acids (De Bruyn et al., 2017; Zhang et al., 2019). The altitude effect on coffee quality and biochemical compositions covering sugar and sugar alcohols were studied by HPAEC-PAD, chlorogenic acids using HPLC and fatty acids and volatile compounds using gas chromatography mass spectrometry (Joët et al., 2010; Bertrand et al., 2012; Tsegay et al., 2020). Studies on coffee based on its geographical origin were studied by NMR-based fingerprinting and GC/MS-based fingerprinting.

NMR-based fingerprinting workflow offer non-destructive analysis covering sugar compounds, amino acids, and coffee-related metabolites (Wei et al., 2012a, 2012b). GC/MS-based fingerprinting require extraction and derivatization with more informative and wider-range metabolite coverage that fall into low molecular hydrophilic compounds such as amino acids, organic acids, sugars, and coffee-related compounds (Putri et al., 2019). Since each of method and analytical instrument offer different targeted metabolite coverage, it has its own merits and demerits depending on the purpose of each study. Therefore, in this study, we aimed to identify factors that influence coffee quality the most by analyzing the coffee from various geographical origin, altitude, and postharvest processes simultaneously. Thus, the metabolite correlated with each factor could be determined. Therefore, gas chromatography/mass spectrometry-based metabolic profiling, combined with multivariate analysis, was employed to analyze differences in the metabolite profiles of coffee from different altitudes, geographical origins, and postharvest processes simultaneously. The simultaneous analysis of multiple factors allowed unsupervised analysis of the dominant factors that influenced the green coffee bean metabolite profile, leading to descriptive analysis of the profile. Our study using extraction and derivatization GC/MS analysis workflow to obtain the metabolites belong to low molecular hydrophilic class compounds that might be in complement with previous study focusing on non-polar fractions of coffee beans as well as volatile compounds and other coffee-related metabolites to describe the effects of geographical origin, altitude, and postharvest process on coffee quality (Joët et al., 2010; Bertrand et al., 2012; De Bruyn et al., 2017; Zhang et al., 2019; Putri et al., 2019; Tsegay et al., 2020). The results of this study could be used by coffee industry stakeholders to decide the most crucial factors impacting coffee production.

2 Material and methods

2.1 Chemicals

The chemicals used in this study are described by Jumhawan et al. (2013) and Putri et al. (2019). All chemicals were of analytical grade. Methanol (99.8%), infinity pure pyridine, and ribitol were purchased from FUJIFILM Wako Pure Chemical Industries, Ltd. (Osaka, Japan). Chloroform (99%) was purchased from Kishida Chemical (Osaka, Japan). Methoxyamine hydrochloride was purchased from Sigma Aldrich (Milwaukee, WI, USA). N-Methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) and a standard alkane mixture (C9–C40) were purchased from GL Science Inc. (Tokyo, Japan).

2.2 Green coffee bean sample collection

Green coffee bean samples were collected from the Indonesian Coffee and Cocoa Research Institute (ICCRI) coffee collection which was collected and purchased from producers. The green coffee beans were stored in a vacuum-sealed pack and prevent direct sun exposure during sample transportation. A total of 17 specialty grade samples of Arabica green coffee beans were sourced. The samples were received in green coffee bean form and were obtained from coffee fruit harvested from coffee plantations at various altitudes across Indonesia. These ranged from 400 to 1600 m above sea level. Coffee fruit harvested from trees planted at 400 m and 800 m above sea level were processed by three different process: dry, honey, and washed process method. Since Arabica coffee is usually cultivated in higher altitude (1000 m above sea level), one sample each from lower altitude (400 m and 800 m) were used in this study. Coffee fruit harvested

from trees planted at more than 1000 m above sea level were processed by washed process. Samples were stored at $-30\text{ }^{\circ}\text{C}$ until extraction and analysis. Detailed information on the samples used, including their cup scores, is shown in Table 1.

2.3 Metabolite extraction and derivatization

Metabolite extraction and derivatization have been performed using a published protocol with minor modifications (Jumhawan et al., 2013; Putri et al., 2019). The metabolite extraction of each sample was performed in a technical replicate ($n=5$). Green coffee beans were added to a 50 mL grinding tube (Yasui Kikai Co., Osaka, Japan) along with a metal cone (Yasui Kikai Co., Osaka, Japan). The tube was closed tightly and quenched by dipping in liquid nitrogen. The closed tube was then immediately put into a Multi-beads shocker (Yasui Kikai Co., Osaka, Japan) and

Table 1 List of sample sets

Sample name/code	Coffee processing method	Altitude (m above sea level)	Origin	Cultivar	Total cupping score
Sample set I					
400-DB	Dry process	400	Bajulmati	Andungsari 2 K	83.42
400-HB	Honey process	400	Bajulmati	Andungsari 2 K	82.58
400-WB	Washed process	400	Bajulmati	Andungsari 2 K	81.33
800-DK	Dry process	800	Kalibendo	S795	81.92
800-HK	Honey process	800	Kalibendo	S795	81.50
800-WK	Washed process	800	Kalibendo	S795	82.83
Sample set II					
400-BM	Washed process	400	Bajulmati	Andungsari 2 K	81.33
800-KB	Washed process	800	Kalibendo	S795	82.83
1000-AG	Washed process	1000	Aceh Gayo	Gayo 1	85.19
1200-AS	Washed process	1200	Andungsari	Mix of Andungsari 2 K/ KOMASTI/BP7000	84.06
1400-B	Washed process	1400	Bondowoso	Mix of S795/Cobra/USDA	84.08
1400-M1	Washed process	1400	Manglayang	Kartika	85.08
1600-B	Washed process	1600	Bondowoso	Mix of S795/Cobra/USDA	83.58
1600-M1	Washed process	1600	Manglayang	Mix Variety Java Preanger	84.25
1000-B	Washed process	1000	Bondowoso	Mix of S795/Cobra/USDA	84.92
1200-B	Washed process	1200	Bondowoso	Mix of S795/Cobra/USDA	83.42
1200-F	Washed process	1200	Flores	S795	86.44
1400-M2	Washed process	1400	Manglayang	Kartika	80.08
1600-M2	Washed process	1600	Manglayang	Mix Variety Java Preanger	82.92
Sample set III					
1000-B	Washed process	1000	Bondowoso	Mix of S795/Cobra/USDA	84.92
1200-B	Washed process	1200	Bondowoso	Mix of S795/Cobra/USDA	83.42
1400-B	Washed process	1400	Bondowoso	Mix of S795/Cobra/USDA	84.08
1600-B	Washed process	1600	Bondowoso	Mix of S795/Cobra/USDA	83.58

DB Dry process-Bajulmati, *HB* honey process-Bajulmati, *WB* washed process-Bajulmati, *DK* dry process-Kalibendo, *HK* honey process-Kalibendo, *WK* washed process-Kalibendo, *BM* Bajulmati, *KB* Kalibendo, *AG* aceh gayo, *AS* Andungsari, *B* Bondowoso, *F* Flores, *M* Manglayang

shaken at 2000 rpm for 30 s. This protocol was repeated 2–3 times to produce a fine powder of green coffee beans. Ten milligrams of green coffee powder were transferred to a 2 mL Eppendorf tube and extracted using 1 mL of a mixed solvent (methanol, ultrapure water, and chloroform = 5/2/2 (v/v/v) ratio) containing an internal standard (ribitol 0.2 mg/mL). The samples and blank (mixed solvent only) were vortexed and incubated at 1200 rpm at 37 °C for 30 min. Then, the samples and blank were centrifuged at $16,000 \times g$ for 3 min at 4 °C. The supernatant (900 μ L) was transferred to a 1.5 mL Eppendorf tube and 400 μ L ultrapure water was added before the samples were vortexed and re-centrifuged. Subsequently, 100 μ L of the aqueous layer was transferred to a new tube with a pierced cap. For quality control (QC) purposes, 100 μ L of the aqueous layer obtained from samples were collected in one pool by transferring it into a 15 mL Eppendorf tube. The collection of sample extracts in the QC pool was homogenized and 100 μ L extract was transferred into a 1.5 mL Eppendorf tube with a pierced cap. The sample, QC, and blank extracts were concentrated by vacuum centrifugation for 20 min and lyophilized overnight.

The derivatization process consists of two steps, oximation and trimethylsilylation, as described in the published procedure (Jumhawan et al., 2013; Putri et al., 2019). The dried extract was subjected to oximation by adding 100 μ L methoxyamine hydrochloride (20 mg/mL in pyridine), then vortexed and incubated at 1200 rpm and 30 °C for 90 min. Subsequently, 50 μ L MSTFA was added and the mixture was re-vortexed and incubated at 1200 rpm and 37 °C for 30 min. Sample mixture (100 μ L) was transferred into a GC/MS vial prior to GC/MS analysis.

2.4 GC/MS analysis

The workflow for GC/MS analysis of coffee bean extract has previously been developed and published elsewhere with several modifications (Jumhawan et al., 2013; Putri et al., 2019). GC-Q/MS analysis was performed on a GCM-SQP 2010 Ultra (Shimadzu, Kyoto, Japan) equipped with a 30 m \times 0.25 mm i.d. fused silica capillary column coated with 0.25 μ m InertCap 5MS/NP (GL Science, Inc., Tokyo, Japan) and an AOC-20i/s (Shimadzu, Kyoto, Japan) as an autosampler. System control and data acquisition were conducted using GC/MS solution software (Shimadzu, Kyoto, Japan). The mass spectrometer was tuned and calibrated prior to analysis. The standard alkane mixture (C9–C40) was injected at the beginning of the retention index calculation. The derivatized sample and blank (1 μ L) were injected in random order with QC injections (1 μ L) every four injections. The injections were conducted in split mode, 25:1 (v/v), and at an injection temperature of 270 °C. Helium was used as the carrier gas with a flow rate of 1.12 mL/min and a linear velocity of 39 cm/s. The column and transfer

line temperature settings are identical to those previously published (Jumhawan et al., 2013; Putri et al., 2019). The ion source temperature was 280 °C and ions were generated by electron ionization at 70 eV. The mass spectra were recorded at 20 scans per second over the mass range 85–500 m/z.

2.5 Data processing

The obtained chromatogram data were converted to a netCDF format using the GCMS Solution software package (Shimadzu, Kyoto, Japan). The netCDF files were then converted into ABF format using Reifycs Abf Converter (<https://www.reifycs.com/AbfConverter/>). Peak detection, deconvolution analysis, identification, and alignment were performed using MS-DIAL version 4.18 (Tsugawa et al., 2015; Lai et al., 2018; Tsugawa et al., 2019), which is a free software (http://prime.psc.riken.jp/Metabolomics_Software/MS-DIAL/). The minimum height for peak detection was set as 1000 amplitude with a linear weight moving average smoothing method. Smoothing level began at 5 scans and average peak width of 20 scans. In deconvolution analysis the sigma window value was set as 1, and the EI spectra cut-off was 200 amplitude. Peak identification used the retention index (RI) based on the retention time of the standard alkane mixture (C9–C40) and GL-Science DB library (InertCap 5MS-NP, Kovats RI, 494 records; http://prime.psc.riken.jp/Metabolomics_Software/MS-DIAL/GCMS%20DB_InertCap%205MS-NP_GLscience.msp). The parameters for peak identification and alignment were: RI tolerance 10, RT tolerance 0.5 min, m/z tolerance 0.5 Da, EI similarity cut-off 70%, and identification score was cut-off at 70%. Metabolite identification was conducted by comparing the mass spectra and RI with the GL-Science DB library. After peak detection, the height of the detected peaks was normalized by the height of the internal standard peaks. The data matrix was constructed consisting of the raw height and normalized height of peaks. Filtering was then performed to exclude peaks that were not of biological origin by referring to the blank chromatogram from the data matrix (Putri et al., 2019). Moreover, peaks with a relative standard deviation of more than 30% in QC and with a raw height less than 3000 were also removed. The final data matrix with filtered and assigned peaks was constructed and used for further analysis.

2.6 Multivariate analysis

PCA and OPLS regression analyses were performed using the commercially available SIMCA-P + version 13.0.3 (Umetrics, Umeå, Sweden). The data matrix was subjected to PCA and OPLS with auto-scale and without transformation, to analyze all the obtained metabolites equally (van den Berg et al., 2006). The relative intensity differences between

samples on selected metabolites were evaluated using one-way analysis of variance and post hoc multiple comparison tests using JASP version 0.13.1 (JASP Team, 2020), which is a freely available software (<https://jasp-stats.org>).

2.7 Sensory analysis

The green coffee beans used in this study were subjected to sensory analysis at the Indonesian Coffee and Cocoa Research Institute (ICCRI) facility in Jember, Indonesia. The roasting process of the beans and the sensory analysis of the roasted beans, and coffee brew, was conducted according to a standardized protocol which was published by the Specialty Coffee Association of America (SCAA). The sensory analysis was performed by a certified panelist (Q-grader) on ten parameters which were rated from 0–10, with zero indicating the least strong and ten for the strongest taste. The score from each parameter was accumulated and the total cup score was obtained. The cup score was used to determine coffee bean grade (SCAA, 2015).

3 Result and discussion

3.1 Postharvest processing and geographical origin effects on the metabolome profile of coffee obtained from various altitudes and origins, processed by different postharvest process methods

There were 142 metabolites detected in GC/MS analysis. After data filtering, 79 detected metabolites were selected for multivariate analysis. Seventy peaks consisted of 50 annotated metabolites and 29 unknown metabolites. The annotated metabolites were categorized as amino acids, sugar compounds, sugar alcohol compounds, organic acids, and other class compounds (Table 2). The unknown metabolites found in this study were numbered based on their retention times. The data matrix of normalized peaks was constructed and subjected to PCA for unsupervised data exploration and visualization of information based on sample variance. The result of PCA were shown in Fig. 1.

The PCA score plot (Fig. 1a) and a loading plot (Fig. 1b) of sample set I (Table 1) showed that the coffee samples were clustered based on postharvest process in the first principal component (41%). Samples were clustered into two groups based on postharvest process: dry process method in the positive quadrant and honey and washed process methods in the negative quadrant (Fig. 1a). Furthermore, most metabolites in the PCA loading plot (Fig. 1b) were located in the positive quadrant compared to the negative quadrant, suggesting that green coffee beans produced by dry process method had more metabolite accumulation than

beans produced by honey and washed process methods. This might be due to the existence or absence of the de-pulping step in the coffee processing methods. Unlike dry process, honey and washed process methods require the de-pulping step, which causes injury and leads to the loss of sugar compounds (De Bruyn et al., 2017; Hoffmann, 2014; Selmar et al., 2014).

The metabolites correlated with green coffee beans, which were obtained from coffee fruit after the dry process method, were glycine, lysine, sorbose, fructose, glyceric acid, and glycolic acid (Fig. 1c). In contrast, metabolites correlated with honey and washed processed coffee beans were galactinol and glutamic acid (Fig. 1c). Green coffee beans obtained from the dry process method were reported to have a higher glucose and fructose content (De Bruyn et al., 2017). The current study also corroborated this finding. It was also found that other sugar metabolites, including galactose and sorbitol were more abundant in green beans produced by dry process method compare with other methods (honey process method and washed process method) (Fig. 1c; Fig. S1). In green coffee beans, produced by honey and washed process method, the higher content of galactinol was found (Fig. 1c). Reports mentioning galactinol correlated with coffee processing are scarce. A previous report mentioning galactinol found that it is correlated with coffee beans obtained from a particular area in Brazil and may be a potential metabolite marker (da Silva Taveira et al., 2014). Galactinol was also found to be correlated with coffee beans based on the geographical origin and the fermentation process (Aditiawati et al., 2020; Putri et al., 2019). For the first time this study points out that galactinol is correlated with the postharvest processes, especially honey and washed process methods. Galactinol is a less common sugar alcohol, with proven function is as a substrate for larger soluble oligosaccharides in plant systems, including seeds (Sengupta et al., 2015). The results of our study could be used to determine whether the galactinol content in green coffee beans contributes to the flavor development of coffee beans during the roasting process. As established by several studies on coffee, galactinol could be added to the list of coffee constituents along with other sugar compounds, amino acids, caffeine, and chlorogenic acids. Glycine, lysine and glutamic acid were also found to be related to postharvest processing. In this study, glycine and lysine content was found to be higher in green coffee beans produced by the dry process method (Fig. 1c). Glutamic acid content was found to be higher in green coffee beans produced by honey and washed processed methods (Fig. 1c). These findings found to be consistent with previous studies (Bytof et al., 2005; Selmar et al., 2014). Glycine and lysine are not the dominant amino acids in green Arabica coffee beans and the content of glycine and lysine are relatively low compared to other amino acids (Bytof et al., 2005; Murkovic & Derler, 2006).

Table 2 List of detected metabolites

No	Metabolite name	Class	RT ^a (min)	RI ^b	Quant. <i>m/z</i>	DPSS ^c (%)	RDPSS ^d (%)	MSI Level ^e
1	Alanine	Amino Acid	4.99	1108	116	89.5	93.3	1
2	Asparagine	Amino Acid	10.29	1685	116	95.4	99.2	1
3	Aspartic acid	Amino acid	9.09	1533	232	98.2	97.7	1
4	Glutamic acid	Amino acid	9.88	1631	246	98.2	99.5	1
5	Glycine	Amino acid	7.16	1317	174	98.3	89.5	1
6	Isoleucine	Amino acid	7.03	1302	158	80.7	89.5	1
7	Leucine	Amino acid	6.81	1280	158	93	87.9	1
8	Lysine	Amino acid	12.11	1940	174	92.9	94.8	1
9	Phenylalanine	Amino acid	9.96	1641	218	91.3	95.3	1
10	Proline	Amino acid	7.06	1306	142	77.4	86.2	1
11	Pyroglutamic acid	Amino acid	9.09	1533	156	93.3	93.7	1
12	Serine	Amino acid	7.68	1372	204	97.5	95.7	1
13	Tryptophan	Amino acid	14.03	2248	202	89.8	93.5	1
14	Tyrosine	Amino acid	12.23	1958	218	97.8	97.5	1
15	Valine	Amino acid	6.24	1224	144	95.2	95.1	1
16	Chlorogenic acid	Organic acid	18.59	3173	345	89.1	98	1
17	Fumaric acid	Organic acid	7.47	1350	245	99.1	96	1
18	Glucarate	Organic acid	12.88	2059	333	92	93.7	1
19	Gluconic acid	Organic acid	12.78	2043	333	86.2	95.4	1
20	Glyceric acid	Organic acid	7.4	1342	147	97.6	90.3	1
21	Glycolic acid	Organic acid	4.67	1079	147	96.6	96.3	1
22	Citric acid	Organic acid	11.44	1842	273	94.8	99	1
23	Lactic acid	Organic acid	4.51	1065	147	99.1	98.5	1
24	Malic acid	Organic acid	8.81	1499	147	97.8	99	1
25	Quinic acid	Organic acid	11.79	1892	345	94.8	98.8	1
26	Succinic acid	Organic acid	7.17	1317	147	90.8	93.3	1
27	Fructose	Sugar	11.95	1916	103	97.7	98	1
28	Galactose	Sugar	12.03	1928	205	92.4	98.2	1
29	Galactose + Glucose	Sugar	12.2	1953	205	91.7	97	1
30	Glucose	Sugar	12.07	1934	205	92.6	98.4	1
31	Maltose	Sugar	16.99	2814	204	83.1	79.9	1
32	Melezitose	Sugar	20.67	3585	361	79.2	85.1	1
33	Raffinose	Sugar	20.17	3502	361	92	94.5	1
34	Sorbose	Sugar	11.88	1906	103	96.7	97.7	1
35	Sucrose	Sugar	16.48	2707	361	94.4	97	1
36	Galactinol	Sugar alcohol	18.17	3076	204	98.9	99.3	1
37	Inositol	Sugar alcohol	13.34	2132	217	96.6	99.1	1
38	Mannitol	Sugar alcohol	12.3	1969	205	96.3	98.8	1
39	Meso-erythritol	Sugar alcohol	9.03	1526	217	98	92.7	1
40	Sorbitol	Sugar alcohol	12.35	1976	205	96.5	97.9	1
41	Threitol	Sugar alcohol	8.97	1518	147	94.8	92.5	1
42	Xylitol	Sugar alcohol	10.67	1736	217	98.7	88.7	1
43	2-Aminoethanol	Other	6.76	1276	174	98.9	96.1	1
44	4-Aminobutyric acid	Other	9.16	1542	174	99	99.7	1
45	Caffeine	Other	11.58	1862	194	99.1	99.1	1
46	Glycerol	Other	6.85	1284	147	97.9	97.5	1
47	N-Methylnicotinate	Other	8.94	1514	210	97.5	98.2	1
48	Phosphate	Other	6.83	1283	299	94.3	95.2	1
49	Putrescine	Other	10.82	1756	174	90.1	92.7	1
50	Serotonin	Other	15.37	2490	174	83.6	77.4	1

Table 2 (continued)

No	Metabolite name	Class	RT ^a (min)	RI ^b	Quant. <i>m/z</i>	DPSS ^c (%)	RDPSS ^d (%)	MSI Level ^e
51	Unknown 1	Unknown	4.20	1037	117	N/A	N/A	4
52	Unknown 2	Organic acid	4.31	1047	117	N/A	N/A	3
53	Unknown 3	Unknown	5.83	1184	147	N/A	N/A	4
54	Unknown 4	Unknown	6.65	1264	98	N/A	N/A	4
55	Unknown 5	Unknown	7.71	1375	156	N/A	N/A	4
56	Unknown 6	Unknown	7.99	1404	147	N/A	N/A	4
57	Unknown 7	Unknown	8.21	1430	160	N/A	N/A	4
58	Unknown 8	Unknown	9.13	1537	174	N/A	N/A	4
59	Unknown 9	Unknown	9.28	1556	120	N/A	N/A	4
60	Unknown 10	Organic acid	9.52	1585	129	N/A	N/A	3
61	Unknown 11	Unknown	9.59	1593	210	N/A	N/A	4
62	Unknown 12	Unknown	11.01	1781	217	N/A	N/A	4
63	Unknown 13	Unknown	11.07	1789	217	N/A	N/A	4
64	Unknown 14	Unknown	12.40	1983	147	N/A	N/A	4
65	Unknown 15	Organic acid	12.76	2039	217	N/A	N/A	3
66	Unknown 16	Organic acid	12.90	2061	204	N/A	N/A	3
67	Unknown 17	Organic acid	13.06	2087	333	N/A	N/A	3
68	Unknown 18	Unknown	13.85	2216	204	N/A	N/A	4
69	Unknown 19	Unknown	13.93	2231	204	N/A	N/A	4
70	Unknown 20	Organic acid	17.96	3026	345	N/A	N/A	3
71	Unknown 21	Sugar	18.39	3126	204	N/A	N/A	3
72	Unknown 22	Organic acid	18.49	3148	345	N/A	N/A	3
73	Unknown 23	Sugar	18.71	3200	204	N/A	N/A	3
74	Unknown 24	Unknown	18.72	3204	249	N/A	N/A	4
75	Unknown 25	Unknown	18.80	3222	307	N/A	N/A	4
76	Unknown 26	Organic acid	18.83	3229	249	N/A	N/A	3
77	Unknown 27	Organic acid	18.90	3244	307	N/A	N/A	3
78	Unknown 28	Organic acid	19.17	3305	324	N/A	N/A	3
79	Unknown 29	Unknown	22.04	3776	204	N/A	N/A	4

There are 79 metabolites detected in all samples

^aRetention time in minute(s)

^bRetention indices (RI) are calculated using a standard alkane mixture (C9–C40)

^cDPSS: Dot product similarity score

^dRDPSS: Reverse dot product similarity score

^eMass spectra of metabolite peaks that are compared only with the in-house library are considered Metabolic Standards Initiative (MSI) level 1 identification. Several unknown compounds are considered MSI level 3 based on similarity to known compounds of a chemical class in library and MSI level 4 if there are no similarities with any chemical class compound in library

Although it is not found as dominant amino acids in green coffee beans, the previous study found glycine and lysine content higher in dry processed coffee beans and there was no further discussion related to this finding in the previous study (Bytof et al., 2005). In contrast, glutamic acid, aspartic acid, and alanine were found to be in high concentration in green coffee beans (Bytof et al., 2005; Murkovic & Derler, 2006). This study applied the metabolomics study workflow which used an auto scale for data pretreatment and gave all the obtained metabolites equal importance. Further, it eliminated the importance of concentration (van den Berg

et al., 2006; Sumner et al., 2007). The differences in the statistical analysis between this study and previous studies might be the reason for the limited discussion regarding amino acids, especially glycine and lysine, in relation to the coffee postharvest process method. Therefore, it would be interesting to closely observe the amino acid profile in future coffee studies using different and more advanced analytical and statistical methods. The current study also found higher glyceric acid and glycolic acid content in green coffee beans produced by the dry process method (Fig. 1c). Reports related to glyceric acid and glycolic acid in correlation with

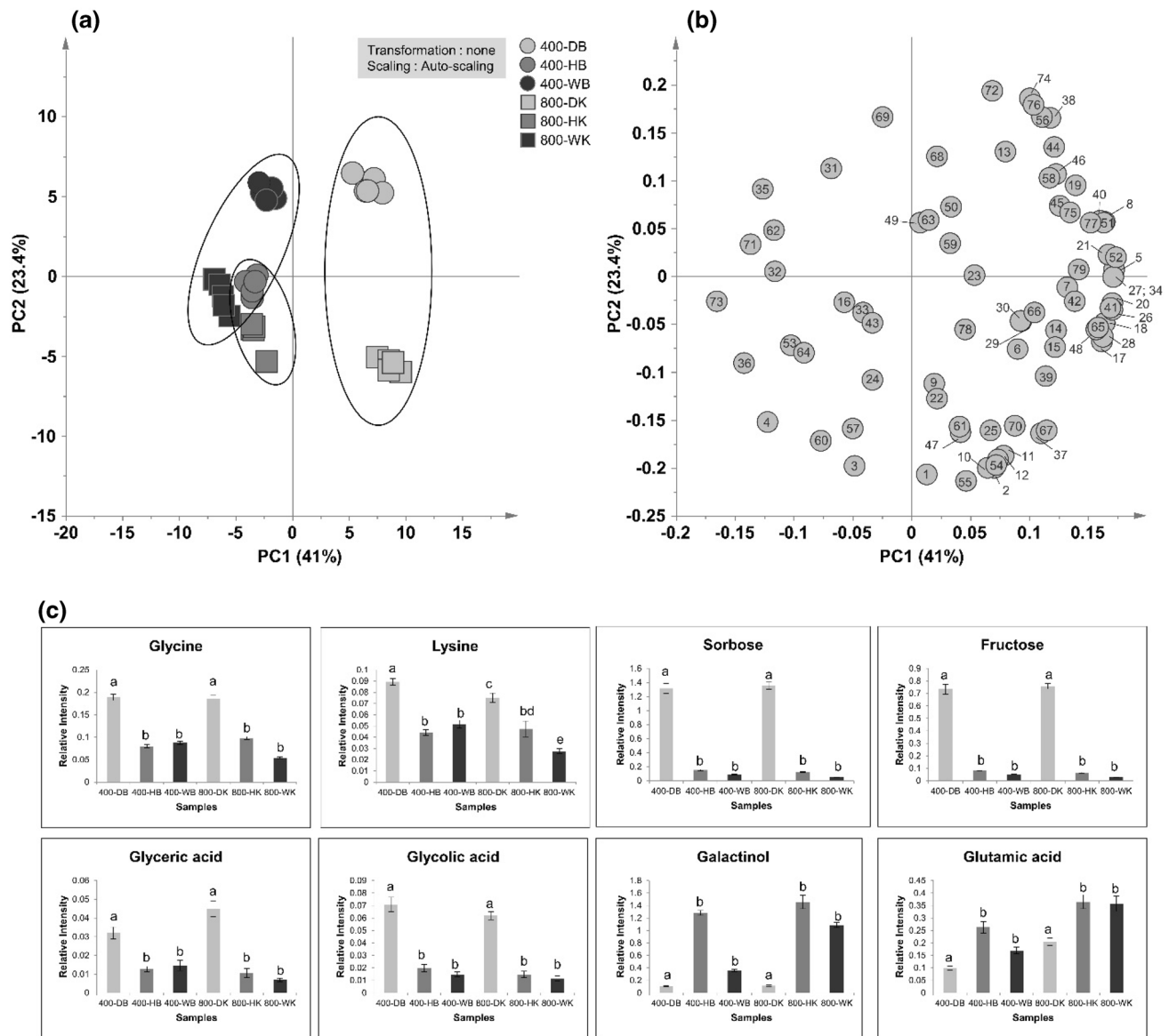


Fig. 1 Principal component analysis (PCA) results of sample clustering based on postharvest processing. Extract of coffee bean samples ($n=5$) were obtained from two different altitudes, 400 m (circles) and 800 m (boxes) above sea level. **a** PCA score plot separates dry processed coffee beans (light grey, right quadrant) from honey processed (dark grey) and washed processed (black) coffee beans. **b** PCA load-

ings plot shows the majority of metabolites correlated with dry processed coffee beans compared to honey and washed processed coffee beans. **c** The relative intensity of correlated metabolites with sample clustering; different letters indicate significant differences according to Tukey's post hoc multiple comparisons test with p values < 0.01 and the error bars indicate standard deviation

coffee processing are limited. Organic acids were found in both green and roasted coffee beans (Ginz et al., 2000). Citric acid, malic acid, chlorogenic acid, and quinic acid were predominantly found in green coffee beans. Meanwhile, formic acid, acetic acid, glycolic acid, and lactic acid were reported to be products of carbohydrate degradation during the roasting process of coffee beans (Clarke & Vitzthum, 2002). However, in a recent study the accumulation of other organic acids, such as lactic acid, gluconic acid, and gluconic acid, has been observed due to fruit fermentation

during the dry process method (De Bruyn et al., 2017). In addition to previous studies, this study found that glyceric acid and glycolic acid accumulation in green coffee beans were produced by the dry process method, assuming that glyceric acid and glycolic acids may also be derived from carbohydrates such as pentose and hexose sugars contained in coffee fruit during the dry processing method (Davidek et al., 2006). The accumulation of organic acids, including glyceric acid and glycolic acid, in green coffee beans produced by the dry process method might increase the total

titratable acidity (TTA) and might influence the final cup of coffee (Barbosa, 2019). The findings of more organic acid class compounds in green coffee beans such as glyceric acid, glycolic acid, lactic acid glucuronic acid, succinic acid, and the known coffee-related organic acids such as citric acid, malic acid, chlorogenic acid, and quinic acid might be useful for future investigation of flavor development in roasting process models.

Since the green coffee bean samples were clustered based on postharvest process, further analysis was conducted using samples obtained from various altitudes and origins, processed using the same postharvest process method (Table 1 sample set II) to eliminate the effect of postharvest process on the sample. However, when using the same metabolome set as the explanatory variable, the samples were not clustered based on altitude in the PCA results (Fig. S2). There might be other factors correlated with the coffee metabolite profile aside from altitude, such as geographical origin. Putri et al. (2019) describe the specialty coffee bean metabolome profile clustering based on geographical origin within Indonesia. Therefore, using the metabolome profile obtained in this study, samples from the same origin and postharvest process were used to eliminate the influence of the latter two in order to observe the correlation between altitude and the coffee metabolome profile. Since this study found a metabolome profile that consists of non-volatile, polar hydrophilic compounds, there is also the possibility that increasing the metabolite coverage, such as including volatile compounds or lipid fractions, might vary the clustering in coffee beans in a future study.

3.2 Correlation between altitude and the coffee metabolome profile obtained from same postharvest process and same geographical origin with various altitudes

Green coffee bean samples obtained from four different altitudes with the same origin and postharvest process method (Table 1 sample set III) were used for further analysis to observe the correlation between altitude and the coffee metabolome profile. The correlation between altitude and the coffee metabolome profile was observed in the PCA analysis (Fig. 2) and showed that the samples were separated based on altitude based on the first principal component (34.4%) (Fig. 2a). Samples were separated from 1000 m above sea level in the right quadrant to 1600 m above sea level on the left quadrant of the PCA. The PCA loading plot (Fig. 2b) showed that metabolites were correlated with sample separation based on altitude. Several sugar compounds, succinic acid, and glycine were correlated with the sample obtained from 1000 m above sea level, while inositol and unknown number 24 were correlated with samples obtained above

1000 m in altitude (Table S2). Sugar compounds such as fructose, sorbose, glucose, galactose + glucose, along with succinic acid and glycine content were decreased as altitude increased (Fig. 2c). In contrast, inositol and unknown number 24 content increased as altitude increased (Fig. 2c).

Previous reports focusing on altitude correlations with the coffee metabolome profile are limited to caffeine, chlorogenic acids, and sugar compounds (Joët et al., 2010; Tolessa et al., 2017; Worku et al., 2018). Hence, studies mentioning glycine and inositol correlating with altitude were scarce. The accumulation of amino acids (including glycine) during seed filling at high temperatures was observed in rice grain and soybean (Carrera et al., 2011; Yamakawa & Hakata, 2010). Low altitude coffee farms have higher temperatures than coffee farms found at higher altitudes (Joët et al., 2010) and might cause the accumulation of glycine in coffee beans in a similar manner to other seeds. The accumulation of glycine in green coffee beans obtained from lower altitude might be reported for the first time through this study. Sugar compound is one of coffee flavor precursor and the accumulation of sugar compounds in coffee beans are different between low altitude and high altitude. Fructose, sorbose, glucose, and galactose + glucose were accumulated in low altitude coffee beans while inositol accumulated in high altitude. The differences of sugar accumulation might lead into different taste characteristics between coffee beans obtained from low altitude and high altitude. Fructose and glucose were degraded during roasting process and could produce aliphatic acids (Ginz et al., 2000). Inositol relatively thermostable and has sweet taste properties (Wei et al., 2012a, 2012b). Therefore, the content of fructose, glucose and inositol in green coffee beans may influences coffee quality of coffee beans based on altitude. Moreover, inositol and its derivatives might have probiotic properties and are beneficial for human health (Caputo et al., 2020). Succinic acid commonly found in green coffee beans along with other organic acid such as citric acid, malic acid, quinic acid, and lactic acid (Evangelista, 2014; Wang et al., 2018). Succinic acid content usually decreased during postharvest processing of coffee, resulting low concentration in final green coffee beans (Cardoso et al., 2021). However, there was little to no discussion about succinic acid in relation with altitude. To our knowledge, this is the first report on the correlation of succinic acid with coffee beans obtained from low altitude. Succinic acid might be contributed to the quality of coffee beans obtained from lower altitude as it might be correlated with the acidity profile of coffee after roasting (Cardoso et al., 2021). Unknown 24 also correlated with high altitude. This compound had lower than 80% of the total similarity score with the library. Hence, further comparison of unknown compounds with the metabolites listed in the library was conducted to gain

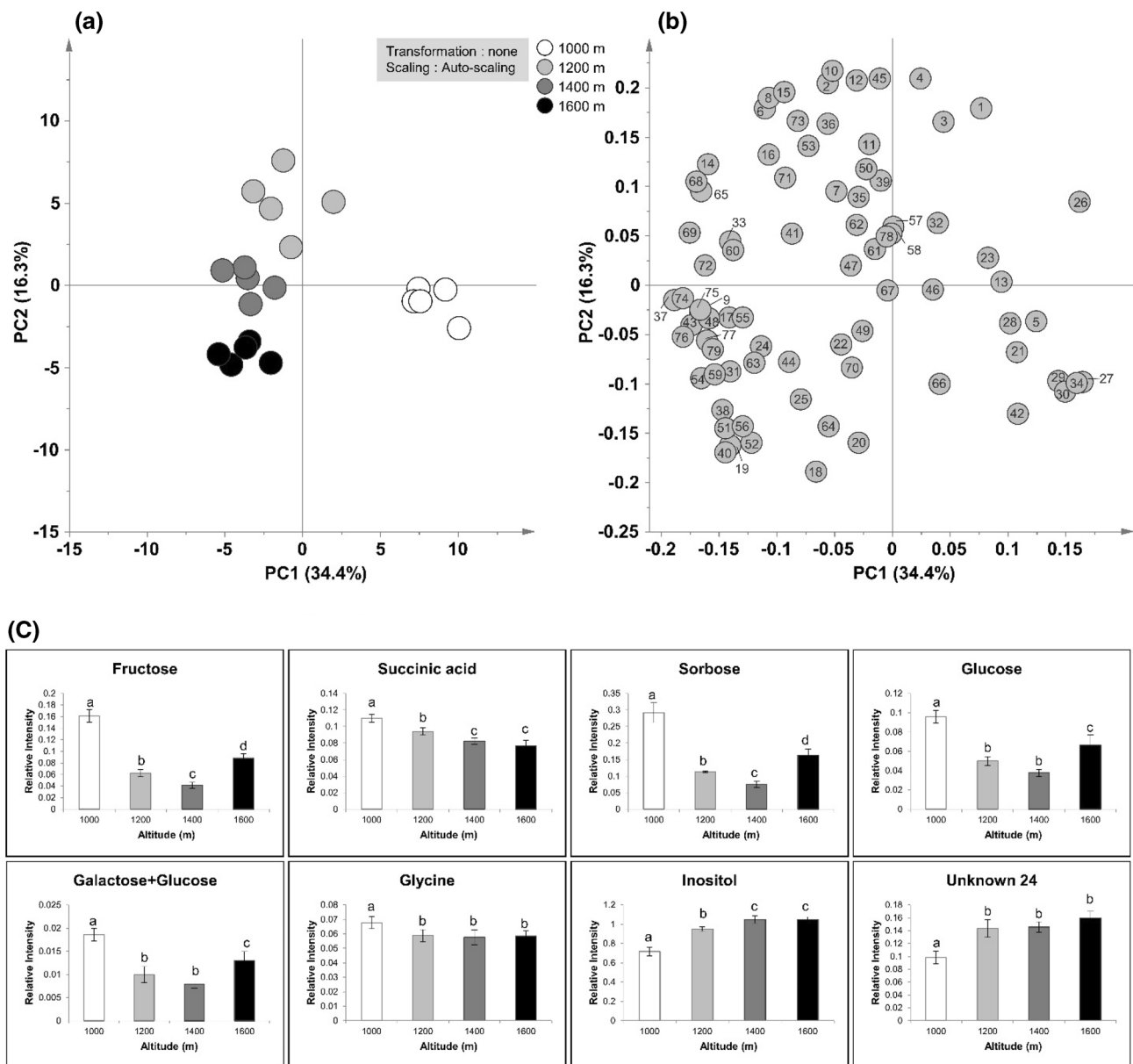


Fig. 2 Principal component analysis (PCA) results of sample clustering based on altitude. Extract of coffee bean samples ($n=5$) were obtained from four altitudes, 1000, 1200, 1400, and 1600 m above sea level, and processed using the washed processing method. **a** PCA score plot separates samples from 1000–1600 m above sea level (right to left quadrant) in the first principal component. **b** PCA load-

ing plot shows the metabolites correlating with sample separation. **c** The relative intensity of correlated metabolites with sample separation; the different letters indicate significant differences according to Tukey's post hoc multiple comparisons test with p values < 0.05 and the error bars indicate standard deviation

more information. Unknown 24 was similar to *p*-coumaric acid and chlorogenic acid, with total similarity scores of 23% and 25.8%, respectively. *p*-coumaric acid belong to the hydroxycinnamic acid and hydroxycinnamate family compounds that are commonly found in coffee (Matei et al., 2019). Hydroxycinnamic acid and hydroxycinnamates are conjugate compounds in chlorogenic acids, the most abundant metabolites in coffee (Matei et al., 2019). Further study is needed to investigate unknown 24 since

there is a possibility it may be related to hydroxycinnamic acid and hydroxycinnamate family compound groups and/or chlorogenic acid in general.

3.3 OPLS regression model for coffee obtained from different altitudes

The OPLS results using sample set II (Table 1) are shown in Fig. 3. The OPLS model of the metabolome profile

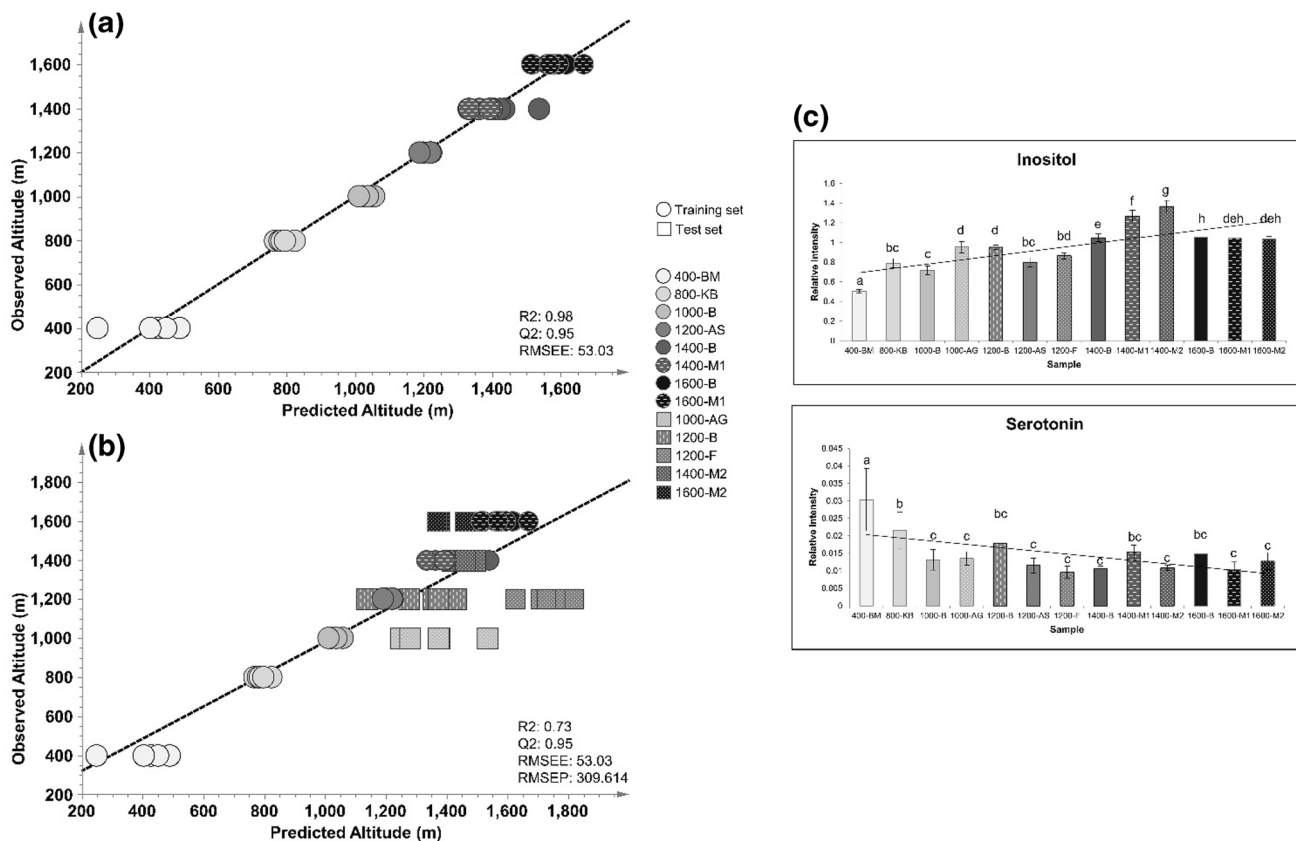


Fig. 3 Orthogonal partial least squares (OPLS) regression model construction based on altitude. **a** The OPLS training set was constructed using extract of six coffee bean samples ($n=5$) obtained from 400–1600 m above sea level with an $R^2=0.96$ and $Q^2=0.95$. **b** The prediction set was added to the OPLS model with good fit and predictive ability ($R^2=0.73$, $Q^2=0.96$, root mean squared error of

estimation (RMSEE)=53.03, and root mean square error of prediction (RMSEP)=309.614). **c** The relative intensity of metabolites with the highest variable influence on projection (VIP) score; the different letters indicate significant differences according to Tukey's post hoc multiple comparisons test with p values < 0.05 and the error bars indicate standard deviation

correlated with altitude (Fig. 3a, b) has an $R^2=0.96$ and a $Q^2=0.95$, suggesting that the model has a good correlation between variables (Alexander et al., 2015; Varmuza & Filzmoser, 2009; Worley & Powers, 2013). Moreover, the model has a low root mean squared error of estimation (RMSEE) and root mean square error of prediction (RMSEP) values, which indicate model accuracy and robustness (Putri & Fukusaki, 2015). The metabolites correlated with altitude were sorted by their variable importance in the projection (VIP) scores (Table S3). Metabolites with a VIP score of more than 1 are considered important (Ikram et al., 2020). This study found 31 metabolites with VIP scores of more than 1, including two metabolites with VIP scores of more than two: inositol and serotonin. These two metabolites had the highest VIP scores among all metabolites (Table S3). Inositol had a positive correlation with altitude, while serotonin had a negative correlation with altitude.

The relative abundance of inositol and serotonin in coffee samples is shown in Fig. 3c. The relative abundance of inositol increased as altitude increased. There was a significant

increase in inositol content in green beans of the sample obtained from 400 m compared to the sample obtained from 800 m above sea level. The increased inositol content continued from the sample obtained from 1000 m to 1600 m above sea level, although not in a linear manner, since there was a variety of inositol content in samples obtained from 1000 m and 1200 m above sea level. The relative abundance of serotonin has a decreasing trend as the altitude increase. The serotonin content was gradually decreasing, and interestingly, after 1000 m above sea level, the difference of serotonin content is no longer significant.

Coffee quality in relation to the altitude coffee is grown at has been widely discussed (Joët et al., 2010; Tolessa et al., 2017; and Worku et al., 2018). Altitude positively correlates with coffee sensory quality (Avelino et al., 2007; Tolessa et al., 2017). However, the compound(s) that describes the correlation between coffee and altitude remains unknown due to limited compound coverage (Joët et al., 2010). In this study, GC/MS-based metabolite profiling and OPLS multivariate analysis methods were employed to investigate the relationship between

altitude and the coffee metabolome profile and identified two metabolites with VIP scores of more than 2. Inositol is positively correlated with increasing altitude, which is consistent with the PCA analysis results (Fig. 2) and may influence coffee quality due to its sweet taste properties (Wei et al., 2012a, 2012b). Historically, the correlation between inositol and altitude was reported to be insignificant by univariate analysis (Joët et al., 2010). However, the multivariate statistical analysis in the current study could conclusively find an underlying correlation between inositol and altitude. In contrast, serotonin was negatively correlated with increasing altitude. Serotonin or 5-hydroxytryptamine, an amine derived from tryptophan or 5-hydroxytryptophan, is predominantly present in most plants, including coffee (Cirilo et al., 2003; Martins & Gloria, 2010; Vasconcelos et al., 2007). In these reports, the role of serotonin in coffee was not mentioned clearly. However, in general, biogenic amines have been reported as food quality and safety markers and have been found in low-quality defective fermented beans with undesirable flavors (Buffo, 2018; Farah, 2012; Oliveira et al., 2005; Ruiz-Capillas & Herrero, 2019). Since the serotonin is a biogenic amine, it was assumed that serotonin might have an undesirable effect on quality in coffee beans obtained from lower altitudes (under 1000 m above sea level). Since the difference in abundance of serotonin was no longer significant at 1000 m above sea level, 1000 m above sea level might be the cut-off level for coffee quality in terms of altitude. This result is in line with previous altitude studies, where the common altitude for the growth of Arabica coffee trees was 1000–1200 m above sea level and coffee fruit and bean harvested from the preferred altitude would have superior quality characteristics (Borém et al., 2019; Davis et al., 2012; Schroth et al., 2015; Tolessa et al., 2017; Worku et al., 2018).

The simultaneous analysis of coffee beans obtained from various geographical origin, altitude and postharvest processes conducted in this study demonstrated that the postharvest process gave the greatest effect in coffee metabolites followed by geographical origin and altitude. The metabolites that showed correlation with each factor described in this study might potentially serve as new coffee quality markers. It is worth noting that our present study focused only on low molecular hydrophilic compounds but there are other compounds such as non-polar compounds, volatiles, or other coffee-related compounds such as trigonelline and phenolic compounds that might be important to explain coffee quality correlated with the factors described in this study. As have been reported previously, ethanal was correlated with fruity taste and was found in high altitude coffee while major fatty acids such as palmitic, stearic, linoleic, and oleic acid and volatile compounds were positively correlated with coffees from high altitude and coffee quality (Betrand et al., 2012; Tsegay et al., 2020). Studies targeting on volatiles were also conducted to describe differences in geographical origins and postharvest process as well as the modifications

within the postharvest process (de Melo Pereira et al., 2019; Marek et al., 2020). Future studies employing multiple analytical instruments and approaches covering a wider range of metabolite may give a more comprehensive outlook on the importance of multiple factors related to coffee origins, processes as well as environmental conditions.

4 Conclusion

A chemical composition study of coffee beans obtained from different altitudes, postharvest processing, and origin using GC/MS-based metabolite profiling focusing on low molecular hydrophilic metabolites was conducted. In this study, we found that the postharvest process had the greatest effect on the coffee metabolome profile, followed by geographical origin and altitude. The influence of altitude on green coffee beans could be found in coffee obtained from the same origin. Furthermore, inositol and serotonin were found to be correlated with altitude based on the OPLS model with good robustness. The current study might increase the validity of the positive correlation between inositol and altitude. Serotonin was reported for the first time as a negatively correlated metabolite with altitude. This information might be beneficial for potential quality profiling of coffee beans obtained from trees grown at low altitudes. Since the serotonin content was lower and no significant difference was observed after an altitude of 1000 m above sea level, this could be used as an altitude cut-off level for Arabica coffee tree plantations. The findings of this study might be useful for further studies related to coffee quality in the future.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11306-021-01817-z>.

Acknowledgements The authors thank Indonesian National Commission for UNESCO and L'Oreal Indonesia for the L'Oreal Award for Women in Science funding received by Sastia P. Putri and to Dr. Agung Wahyu Susilo and late Dr. Sukrisno Widyotomo of Indonesian Coffee and Cocoa Research Institute for their kind support in this study. This study represents a portion of a dissertation submitted by Fitri Amalia to Osaka University in partial fulfillment of requirements for her Ph.D.

Author contribution FA and SPP design the experiments; FA performed the experiments, analyzed the data, and wrote the manuscript; PA conducted sample collection, participated in experiment design; Y provided the sample, conducted the sensory analysis and aided in data interpretation; SPP analyzed the data and wrote the manuscript; EF conceived the study and participated in its design and coordination. All authors have read and agreed to the published version of the manuscript.

Data availability The metabolomics and the metadata reported in this paper was submitted to <https://www.ebi.ac.uk/metabolights/> with study identifier MTBLS1608.

Declarations

Conflict of interest The authors declare no conflict of interest.

Human and animal rights This article does not contain any studies with animal performed by any of the authors. In conducting the cupping testing, we are using the trained and certified assessor from Indonesia Coffee and Cocoa Research Institute (ICCRI).

References

- Aditiawati, P., Astuti, D. I., Kriswantoro, J. A., KhanzaKamarisimaIri-fune, S. M. T., Amalia, F., Fukusaki, E., & Putri, S. P. (2020). GC/MS-based metabolic profiling for the evaluation of solid state fermentation to improve quality of Arabica coffee beans. *Metabolomics*, *16*(5), 1–14. <https://doi.org/10.1007/s11306-020-01678-y>
- Alexander, D. L. J., Tropsha, A., & Winkler, D. A. (2015). Beware of R2: Simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models. *Journal of Chemical Information and Modeling*, *55*(7), 1316–1322. <https://doi.org/10.1021/acs.jcim.5b00206>
- Avelino, J., Barboza, B., Araya, J. C., Fonseca, C., Davrieux, F., Guyot, B., & Cilas, C. (2005). Effects of slope exposure, altitude and yield on coffee quality in two altitude terroirs of Costa Rica, Orosi and Santa María de Dota. *Journal of the Science of Food and Agriculture*, *85*(11), 1869–1876. <https://doi.org/10.1002/jsfa.2188>
- Avelino, J., Barboza, B., Davrieux, F., & Guyot, B. (2007). Shade effects on sensory and chemical characteristics of coffee from very high altitude plantations in Costa Rica. In *Second international symposium on multi-strata agroforestry systems with perennial crops: Making ecosystem services count for farmers, Consumers and the Environment*, September, 1–6.
- 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. <https://doi.org/10.1016/j.foodchem.2012.06.060>
- Borém, F. M., Luz, M. P. S., Sáfiadi, T., Volpato, M. M. L., Alves, H. M. R., Borém, R. A. T., & Maciel, D. A. (2019). Meteorological variables and sensorial quality of coffee in the mantiqueira region of Minas Gerais. *Coffee Science*, *14*(1), 38–47. <https://doi.org/10.25186/cs.v14i1.1522>
- Buffo, R. (2018). Advances in research on coffee flavour compounds. In P. Lashermes (Ed.), *Achieving sustainable cultivation of coffee* (pp. 291–309). Burleigh Dodds Science Publishing. <https://doi.org/10.19103/as.2017.0022.15>
- Bytof, G., Knopp, S. E., Schieberle, P., Teutsch, I., & Selmar, D. (2005). Influence of processing on the generation of γ -aminobutyric acid in green coffee beans. *European Food Research and Technology*, *220*(3–4), 245–250. <https://doi.org/10.1007/s00217-004-1033-z>
- Caputo, M., Bona, E., Leone, I., Samà, M. T., Nuzzo, A., Ferrero, A., Aimaretti, G., Marzullo, P., & Prodám, F. (2020). Inositols and metabolic disorders: From farm to bedside. *Journal of Traditional and Complementary Medicine*, *10*(3), 252–259. <https://doi.org/10.1016/j.jtcme.2020.03.005>
- Cardoso, W. S., Agnoletti, B. Z., de Freitas, R., de Abreu Pinheiro, F., & Pereira, L. L. (2021). Biochemical aspects of coffee fermentation. In L. L. Pereira & T. R. Moreira (Eds.), *Food engineering series: Quality determinants in coffee production* (pp. 166–168). Springer Nature. https://doi.org/10.1007/978-3-030-54437-9_4
- Carrera, C. S., Reynoso, C. M., Funes, G. J., Martínez, M. J., Dardanelli, J., & Resnik, S. L. (2011). Amino acid composition of soybean seeds as affected by climatic variables. *Pesquisa Agropecuária Brasileira*, *46*(12), 1579–1587. <https://doi.org/10.1590/S0100-204X2011001200001>
- Cirilo, M. P. G., Coelho, A. F. S., Araújo, C. M., Gonçalves, F. R. B., Nogueira, F. D., & Glória, M. B. A. (2003). Profile and levels of bioactive amines in green and roasted coffee. *Food Chemistry*, *82*(3), 397–402. [https://doi.org/10.1016/S0308-8146\(02\)00560-5](https://doi.org/10.1016/S0308-8146(02)00560-5)
- Clarke, R. J., & Vitzthum, O. G. (2002). *Coffee: Recent developments*. Blackwell Science Ltd.
- Clifford, M. N., Ludwig, I. A., & Crozier, A. (2018). Chemical composition of coffee beans: an overview. In P. Lashermes (Ed.), *Achieving sustainable cultivation of coffee* (pp. 195–214). Burleigh Dodds Science Publishing. <https://doi.org/10.19103/as.2017.0022.11>
- Cordoba, N., Pataquiva, L., Osorio, C., Moreno, F. L. M., & Ruiz, R. Y. (2019). Effect of grinding, extraction time and type of coffee on the physicochemical and flavour characteristics of cold brew coffee. *Scientific Reports*, *9*(1), 1–12. <https://doi.org/10.1038/s41598-019-44886-w>
- da Silva Taveira, J. H., Borém, F. M., Figueiredo, L. P., Reis, N., Franca, A. S., Harding, S. A., & Tsai, C. J. (2014). Potential markers of coffee genotypes grown in different Brazilian regions: A metabolomics approach. *Food Research International*, *61*, 75–82. <https://doi.org/10.1016/j.foodres.2014.02.048>
- Davídek, T., Devaud, S., Robert, F., & Blank, I. (2006). Sugar fragmentation in the maillard reaction cascade: Isotope labeling studies on the formation of acetic acid by a hydrolytic β -dicarbonyl cleavage mechanism. *Journal of Agricultural and Food Chemistry*, *54*(18), 6667–6676. <https://doi.org/10.1021/jf060667q>
- Davis, A. P., Gole, T. W., Baena, S., & Moat, J. (2012). The impact of climate change on indigenous Arabica coffee (*Coffea arabica*): Predicting future trends and identifying priorities. *PLoS ONE*, *7*(11), 10–14. <https://doi.org/10.1371/journal.pone.0047981>
- de Barbosa, M. S. G., Scholz, M. B. S., Kitzberger, C. S. G., & Benassi, M. T. (2019). Correlation between the composition of green Arabica coffee beans and the sensory quality of coffee brews. *Food Chemistry*, *292*, 275–280. <https://doi.org/10.1016/j.foodchem.2019.04.072>
- De Bruyn, F., Zhang, S. J., Pothakos, V., Torres, J., Lambot, C., Moroni, A. V., Callanan, M., Sybesma, W., Weckx, S., & De Vuyst, L. (2017). Exploring the impacts of postharvest processing on the microbiota and metabolite profiles during green coffee bean production. *Applied and Environmental Microbiology*, *83*(1), 1–16.
- de Melo Pereira, G. V., de Carvalho Neto, D. P., Magalhães Júnior, A. I., Vásquez, Z. S., Medeiros, A. B. P., Vandenberghe, L. P. S., & Soccol, C. R. (2019). Exploring the impacts of postharvest processing on the aroma formation of coffee beans—a review. *Food Chemistry*, *272*(August 2018), 441–452. <https://doi.org/10.1016/j.foodchem.2018.08.061>
- Evangelista, S. R., Cruz Pedrozo Miguel, M. G., Souza Cordeiro, C., Silva, C. F., Marques Pinheiro, A. C., & Schwan, R. F. (2014). Inoculation of starter cultures in a semi-dry coffee (*Coffea arabica*) fermentation process. *Food Microbiology*. <https://doi.org/10.1016/j.fm.2014.05.013>
- Farah, A. (2012). Coffee constituents. In Y. F. Chu (Ed.), *Coffee: Emerging health effects and disease prevention* (pp. 21–50). Wiley-Blackwell. <https://doi.org/10.1002/9781119949893.ch2>
- Fisk, I. D., Kettle, A., Hofmeister, S., Virdie, A., & Kenny, J. S. (2012). Discrimination of roast and ground coffee aroma. *Flavour*, *1*(1), 1–9. <https://doi.org/10.1186/2044-7248-1-14>
- Ginz, M., Balzer, H. H., Bradbury, A. G. W., & Maier, H. G. (2000). Formation of aliphatic acids by carbohydrate degradation during roasting of coffee. *European Food Research and Technology*, *211*(6), 404–410. <https://doi.org/10.1007/s002170000215>
- Hoffmann, J. (2014). *The world atlas of coffee*. Octopus Publishing Group.

- Ikrām, M. M. M., Ridwani, S., Putri, S. P., & Fukusaki, E. (2020). GC-MS based metabolite profiling to monitor ripening-specific metabolites in pineapple (*Ananas comosus*). *Metabolites*, *10*(4), 1–15. <https://doi.org/10.3390/metabo10040134>
- JASP Team. (2020). JASP (version 0.13.1) [Computer Software]. <https://jasp-stats.org/>.
- Joët, T., & Dussert, S. (2018). Environmental and genetic effects on coffee seed biochemical composition and quality. In P. Lashermes (Ed.), *Achieving sustainable cultivation of coffee* (pp. 49–67). Burleigh Dodds Science Publishing. <https://doi.org/10.19103/as.2017.0022.03>
- Joët, T., Laffargue, A., Descroix, F., Doubeau, S., Bertrandkochko, B. A., & Dussert, S. (2010). Influence of environmental factors, wet processing and their interactions on the biochemical composition of green Arabica coffee beans. *Food Chemistry*, *118*(3), 693–701. <https://doi.org/10.1016/j.foodchem.2009.05.048>
- Jumhawan, U., PutriYusiantoMarwani, S. P. E., Bamba, T., & Fukusaki, E. (2013). Selection of discriminant markers for authentication of Asian palm civet coffee (Kopi Luwak): A metabolomics approach. *Journal of Agricultural and Food Chemistry*, *61*(33), 7994–8001. <https://doi.org/10.1021/jf401819s>
- Knopp, S., Bytof, G., & Selmar, D. (2006). Influence of processing on the content of sugars in green Arabica coffee beans. *European Food Research and Technology*, *223*(2), 195–201. <https://doi.org/10.1007/s00217-005-0172-1>
- Lai, Z., Tsugawa, H., Wohlgenuth, G., Mehta, S., Mueller, M., Zheng, Y., Ogiwara, A., Meissen, J., Showalter, M., Takeuchi, K., Kind, T., Beal, P., Arita, M., & Fiehn, O. (2018). Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. *Nature Methods*. <https://doi.org/10.1038/nmeth.4512>
- Lee, L. W., Cheong, M. W., Curran, P., Yu, B., & Liu, S. Q. (2015). Coffee fermentation and flavor—an intricate and delicate relationship. *Food Chemistry*, *185*, 182–191. <https://doi.org/10.1016/j.foodchem.2015.03.124>
- Marek, G., Dobrzański, B., Oniszczuk, T., Combrzyński, M., Cwikła, 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*. <https://doi.org/10.3390/s20072124>
- Martins, A. C. C. L., & Gloria, M. B. A. (2010). Changes on the levels of serotonin precursors—tryptophan and 5-hydroxytryptophan—during roasting of Arabica and Robusta coffee. *Food Chemistry*, *118*(3), 529–533. <https://doi.org/10.1016/j.foodchem.2009.05.016>
- Matei, M. F., Seung-Hun, L., & Kuhnert, N. (2019). CHAPTER 24: Chlorogenic acids. In coffee. In A. Farah (Ed.), *Coffee: Production, quality and chemistry* (pp. 565–583). The Royal Society of Chemistry. <https://doi.org/10.1039/9781782622437-00565>
- Murkovic, M., & Derler, K. (2006). Analysis of amino acids and carbohydrates in green coffee. *Journal of Biochemical and Biophysical Methods*, *69*(1–2), 25–32. <https://doi.org/10.1016/j.jbbm.2006.02.001>
- Oliveira, S. D., Franca, A. S., Glória, M. B. A., & Borges, M. L. A. (2005). The effect of roasting on the presence of bioactive amines in coffees of different qualities. *Food Chemistry*, *90*(1–2), 287–291. <https://doi.org/10.1016/j.foodchem.2004.03.056>
- Purnomo, M., Daulay, P., Utomo, M. R., & Riyanto, S. (2019). Moderating role of connoisseur consumers on sustainable consumption and dynamics capabilities of Indonesian single origin coffee shops. *Sustainability*. <https://doi.org/10.3390/su11051319>
- Putri, S. P., & Fukusaki, E. (2015). *Mass spectrometry-based metabolomics: A practical guide*. CRC Press.
- 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), 1–11. <https://doi.org/10.1007/s11306-019-1591-5>
- Ruiz-Capillas, C., & Herrero, A. M. (2019). Impact of biogenic amines on food quality and safety. *Foods*. <https://doi.org/10.3390/foods8020062>
- Ruta, L. L., & Farcasanu, I. C. (2021). Coffee and yeasts: From flavor to biotechnology. *Fermentation*, *7*(1), 9. <https://doi.org/10.3390/fermentation7010009>
- Schroth, G., Läderach, P., Blackburn Cuero, D. S., Neilson, J., & Bunn, C. (2015). Winner or loser of climate change? A modeling study of current and future climatic suitability of Arabica coffee in Indonesia. *Regional Environmental Change*, *15*(7), 1473–1482. <https://doi.org/10.1007/s10113-014-0713-x>
- Selmar, D., Bytof, G., & Knopp, S. E. (2002). New aspects of coffee processing: The relation between seed germination and coffee quality. In *Proceedings of the international congress of ASIC*, Paris, France, pp. 14–18.
- Selmar, D., Kleinwächter, 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* (pp. 431–476). CRC Press. <https://doi.org/10.1016/B978-0-12-384730-0.00074-4>
- Sengupta, S., Mukherjee, S., Basak, P., & Majumder, A. L. (2015). Significance of galactinol and raffinose family oligosaccharide synthesis in plants. *Frontiers in Plant Science*, *6*(AUG), 1–11. <https://doi.org/10.3389/fpls.2015.00656>
- Seninde, D. R., & Chambers, E. (2020). Coffee flavor: A review. *Beverages*, *6*(3), 1–25. <https://doi.org/10.3390/beverages6030044>
- Sittipod, S., Schwartz, E., Paravisini, L., & Peterson, D. G. (2019). Identification of flavor modulating compounds that positively impact coffee quality. *Food Chemistry*. <https://doi.org/10.1016/j.foodchem.2019.125250>
- Specialty Coffee Association of America. (2015, December 16). *SCAA protocols—Cupping specialty coffee*. Specialty Coffee Association of America. Retrieved June 26, 2020, from <http://www.scaa.org/PDF/resources/cupping-protocols.pdf>.
- Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., Fan, T. W.-M., Fiehn, O., Goodacre, R., Griffin, J. L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A. N., Lindon, J. C., Marriott, P., Nicholls, A. W., & Viant, M. R. (2007). Proposed minimum reporting standards for chemical analysis. *Metabolomics*, *3*(3), 211–221. <https://doi.org/10.1007/s11306-007-0082-2>
- Sunarharum, W. B., Williams, D. J., & Smyth, H. E. (2014). Complexity of coffee flavor: A compositional and sensory perspective. *Food Research International*, *62*, 315–325. <https://doi.org/10.1016/j.foodres.2014.02.030>
- Tolessa, K., D’heer, J., Duchateau, L., & Boeckx, P. (2017). 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. <https://doi.org/10.1002/jsfa.8114>
- Tsegay, G., Redi-Abshiro, M., Chandravanshi, B. S., Ele, E., Mohammed, A. M., & Mamo, H. (2019). Volatile profile of green coffee beans from *Coffea arabica*. Plants grown at different altitudes in Ethiopia. *Bulletin of the Chemical Society of Ethiopia*, *33*(3), 401–413. <https://doi.org/10.4314/bcse.v33i3.2>
- Tsegay, G., Redi-Abshiro, M., Chandravanshi, B. S., Ele, E., Mohammed, A. M., & Mamo, H. (2020). Effect of altitude of coffee plants on the composition of fatty acids of green coffee beans. *BMC Chemistry*, *14*(1), 1–11. <https://doi.org/10.1186/s13065-020-00688-0>
- Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., Kanazawa, M., Vandergeheynst, J., Fiehn, O., & Arita, M. (2015). MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature Methods*. <https://doi.org/10.1038/nmeth.3393>

- Tsugawa, H., Nakabayashi, R., Mori, T., Yamada, Y., Takahashi, M., Rai, A., Sugiyama, R., Yamamoto, H., Nakaya, T., Yamazaki, M., Kooke, R., Bac-Molenaar, J. A., Oztolan-Erol, N., Keurentjes, J. J. B., Arita, M., & Saito, K. (2019). A cheminformatics approach to characterize metabolomes in stable-isotope-labeled organisms. *Nature Methods*. <https://doi.org/10.1038/s41592-019-0358-2>
- van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K., & van der Werf, M. J. (2006). Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics*, 7, 1–15. <https://doi.org/10.1186/1471-2164-7-142>
- Varmuza, K., & Filzmoser, P. (2009). *Introduction to multivariate statistical analysis in chemometrics*. CRC Press.
- Vasconcelos, A. L. S., Franca, A. S., Glória, M. B. A., & Mendonça, J. C. F. (2007). A comparative study of chemical attributes and levels of amines in defective green and roasted coffee beans. *Food Chemistry*, 101(1), 26–32. <https://doi.org/10.1016/j.foodchem.2005.12.049>
- Wang, C., Sun, J., Lassabliere, B., Yu, B., Zhao, F., Zhao, F., Chen, Y., & Liu, S. Q. (2018). Potential of lactic acid bacteria to modulate coffee volatiles and effect of glucose supplementation: Fermentation of green coffee beans and impact of coffee roasting. *Journal of the Science of Food and Agriculture*, 99(1), 409–420. <https://doi.org/10.1002/jsfa.9202>
- Wei, F., Furihata, K., Koda, M., Hu, F., Kato, R., Miyakawa, T., & Tanokura, M. (2012). ¹³C NMR-based metabolomics for the classification of green coffee beans according to variety and origin. *Journal of Agricultural and Food Chemistry*, 60(40), 10118–10125. <https://doi.org/10.1021/jf3033057>
- Wei, F., Furihata, K., Koda, M., Hu, F., Miyakawa, T., & Tanokura, M. (2012). Roasting process of coffee beans as studied by nuclear magnetic resonance: Time course of changes in composition. *Journal of Agricultural and Food Chemistry*, 60(4), 1005–1012. <https://doi.org/10.1021/jf205315r>
- Worku, M., de Meulenaer, B., Duchateau, L., & Boeckx, P. (2018). Effect of altitude on biochemical composition and quality of green arabica coffee beans can be affected by shade and postharvest processing method. *Food Research International*, 105(November), 278–285. <https://doi.org/10.1016/j.foodres.2017.11.016>
- Worley, B., & Powers, R. (2013). Multivariate analysis in metabolomics. *Current Metabolomics*, 1(1), 92–107. <https://doi.org/10.2174/2213235x11301010092>
- Yamakawa, H., & Hakata, M. (2010). Atlas of rice grain filling-related metabolism under high temperature: Joint analysis of metabolome and transcriptome demonstrated inhibition of starch accumulation and induction of amino acid accumulation. *Plant and Cell Physiology*, 51(5), 795–809. <https://doi.org/10.1093/pcp/pcq034>
- Zhang, S. J., De Bruyn, F., Pothakos, V., Torres, J., Falconi, C., Moccand, C., Weckx, S., & De Vuyst, L. (2019). Following coffee production from cherries to cup: Microbiological and metabolomic analysis of wet processing of *Coffea arabica*. *Applied and Environmental Microbiology*, 85(6), 1–22. <https://doi.org/10.1128/AEM.02635-18>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.