Theor. Exp. Plant Physiol. (2025) 37:27
https://doi.org/10.1007/s40626-025-00373-4

Non-targeted metabolomic analysis of field-grown *Coffea arabica* cultivars reveals distinct leaf metabolic signatures

Jéfyne Campos Carréra · Leonor Guerra-Guimarães · John Charles D'Auria · Luana de Jesus Sartori · Carla Pinheiro · Vânia Aparecida Silva · Margarete Lordelo Volpato · Gladyston Rodrigues Carvalho · Fabio Akira Mori

Received: 24 October 2024 / Accepted: 25 April 2025

Abstract This study conducted a non-targeted metabolomic analysis of five *Coffea arabica* L. cultivars grown in the field experimental areas of the Cerrado Mineiro (Minas Gerais State, Brazil) to identify their metabolic fingerprints. The five cultivars selected for this study were chosen based on their specific genetic backgrounds and traits, including disease resistance, productivity, and cup quality. A total of 463 metabolic features were detected in the overall *C. arabica* metabolome, with the major metabolic classes comprising sugars, amino acids, lipids, phenylpropanoids, and phenolic compounds. Among these, 41 metabolites were identified as key discriminators among the five cultivars. Partial least squares discriminant analysis (PLS-DA) revealed distinct metabolic profiles, highlighting ferulic acid, theobromine, octopamine, rosmarinic acid, and gibberellin as key metabolites. These findings emphasize the importance of phenolic compounds and alkaloids in cultivar discrimination. The most relevant metabolic markers associated with environmental stress tolerance

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40626-025-00373-4.

J. C. Carréra · L. de Jesus Sartori · F. A. Mori
UFLA - Universidade Federal de Lavras, Lavras, Minas Gerais 37200-000, Brazil

L. Guerra-Guimarães (✉)
CIFC - Centro de Investigação das Ferrugens do Cafeeiro, Jardim Botânico da Ajuda, Instituto Superior de Agronomia, Universidade de Lisboa, Calçada da Ajuda, 1300-011 Lisbon, Portugal
e-mail: leonorguimaraes@isa.ulisboa.pt

L. Guerra-Guimarães
LEAF - Linking Landscape, Environment, Agriculture and Food Research Centre and TERRA Associated Laboratory, Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, 1349-017 Lisbon, Portugal

J. C. D'Auria
Research Group Metabolic Diversity, Department of Molecular Genetics, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK Gatersleben), OT Gatersleben, Corrensstraße 3, 06466 Seeland, Germany

C. Pinheiro
UCIBIO Applied Molecular Biosciences Unit, Department of Life Sciences, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal

C. Pinheiro
Associate Laboratory i4HB Institute for Health and Bioeconomy, NOVA School of Science and Technology, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal

V. A. Silva · M. L. Volpato · G. R. Carvalho
EPAMIG – Empresa de Pesquisa Agropecuária de Minas Gerais, Lavras, Minas Gerais 37200-000, Brazil

Published online: 27 May 2025

suggest their potential as biochemical indicators for selecting resilient cultivars, thereby contributing to coffee breeding programs. Notably, this study is the first documented characterization of the leaf metabolome of field-grown *C. arabica* cultivars, with Catiguá MG2 emerging as the most distinct. Our findings demonstrate the efficacy of metabolomic fingerprinting via non-targeted metabolomic as a powerful tool for differentiating coffee cultivars and for precision breeding strategies.

Keywords GC–TOF–MS · Phenolics · Alkaloids · Cerrado Mineiro (Brazil)

1 Introduction

Brazil is currently the world's leading producer and exporter of coffee, accounting for over one-third of global coffee production (International Coffee Organization 2024). Within Brazil, Minas Gerais state is the primary region for cultivating *Coffea arabica* L., a position largely attributed to breeding research initiatives with developed numerous cultivars since the 1970s (Caixeta et al. 2022). The establishment of breeding programs, coupled with government incentives, has enabled the expansion of coffee cultivation into the Cerrado Mineiro, a region situated in northwest Minas Gerais renowned for its high-quality coffee (Nagay 1999; Caixeta et al. 2022). Field characterization has a crucial role in identifying cultivars with superior yield potential, excellent cup quality, and optimal adaptation to the region's specific edaphoclimatic conditions (De Carvalho et al. 2022). Additionally, metabolic profiling provides a biochemical perspective on cultivar performance, linking field traits to the underlying metabolic processes that drive plant growth, tolerance to stressful conditions, and quality conditions.

Metabolomics enables the assessment of chemical compound variations in plants, stemming from species differences and environmental changes (Shen et al. 2023). Specifically, non-targeted metabolomics enables the identification of metabolic signatures by providing a comprehensive snapshot of the metabolome, encompassing both primary and specialized metabolites (Shen et al. 2023). By examining the specialized metabolism, distinct metabolic fingerprints can be identified within a given species, providing valuable insights into metabolic differentiation and adaptation mechanisms (Dussarrat et al. 2023).

The identification of metabolic signatures is crucial for understanding the biochemical differences among *C. arabica* cultivars. These signatures, shaped by genetic, physiological, and environmental factors (Kitashova et al. 2023), offer valuable tools for cultivar differentiation. They are characterized by distinct metabolic profiles, which are influenced by plant genetics, cultivation practices, and stress responses (Martins et al. 2014; Mihai et al. 2024). Since the leaf metabolome provides insights into key plant processes – such as health, stress responses, and development -, its evaluation is essential for understanding factors that impact coffee bean quality and yield (Da Silva et al. 2021; Chekol et al. 2024). In the context of coffee cultivars, metabolic signatures can help identify variations in key biochemical pathways, such as those governing alkaloid production and polyphenol biosynthesis (Cangeloni et al. 2022; Castro-Moretti et al. 2023). In addition, studies on coffee leaf metabolism have identified diverse classes of metabolites, including chlorogenic acids, benzophenones, flavonoids, xanthones, and alkaloids, that are critical to plant physiology and defense mechanisms (Souard et al. 2017; Montis et al. 2022).

By linking metabolic profiles to specific cultivars, a deeper understanding of the biochemical foundations of phenotypic diversity can be achieved. This knowledge is fundamental for breeding programs, precision agriculture, and the development of targeted strategies, aimed at enhancing both plant resilience and the sensory attributes of coffee products (Martins et al. 2014; Guimarães et al. 2021; Chekol et al. 2024).

While coffee metabolomics has traditionally focused on species differences within the Rubiaceae family, particularly the coffee bean's role in beverage quality (Amalia et al. 2023; Castro-Moretti et al. 2023; Zhai et al. 2024), recent studies have begun to explore how plant tissues, including leaves, influence overall plant health and productivity (Castro-Moretti et al. 2023). As a key indicator of plant stress responses, nutrient status, and growth, the leaf metabolome provides valuable insights that indirectly affect coffee bean quality (Sardans et al. 2011; Bollen et al. 2024). To address this gap, this study applied non-targeted polar metabolomic analysis to the leaves of five field-grown *C. arabica* cultivars. By characterizing

the unique metabolome of each cultivar and identifying potential metabolic signatures relevant to differentiation, this study aims to shed light on how leaf metabolic profiles serve as predictors of plant performance, contributing to optimized coffee production, plant health and enhanced beverage quality.

2 Material and methods

2.1 Plant material and field experiments

Seven-year-old *C. arabica* plants used in this study were grown since 2016 in an Empresa de Pesquisa Agropecuária de Minas Gerais (EPAMIG) experimental germplasm field located in the municipality of Patrocínio, MG, Brazil (latitude: 18° 57' S, longitude: 46° 54' W), which is part of the coffee-producing region known as the Cerrado Mineiro. The selection of five cultivars used in the present study was based on their specific traits, resistance to *Hemileia vastatrix* (coffee rust) and *Meloidogyne exigua* (nematode), as well as their productivity and cup quality (Table 1). The study utilized rainfed plants (in dry season, august 2022). The plants were arranged in a 0.7 m spacing in the same crop row (plot). The soil conditions in terms of macro and micronutrient availability and soil texture were standardized across the plot (Supplementary Table 1). A pair of fully expanded leaves from the middle third of five individual plants per cultivar were collected. The biological replicates were obtained from pooled individuals (Sumner et al. 2007) and the biological sampling was n = 5 for each coffee cultivar. Subsequently, the leaf midribs were excised and discarded, and the remaining material was preserved in liquid nitrogen. Leaf collection occurred between 07:00 and 09:00 h considering the sun-exposed side of the plant (Supplementary Fig. 1).

2.2 Leaf extract

The leaf samples were ground with liquid nitrogen, followed by lyophilization, and subsequently stored in vacuum-sealed conical centrifuge tubes (Supplementary Fig. 1). The extraction procedure was carried out with adaptations based on the method proposed by Salem et al. (2016, 2020), designed for the extraction of polar metabolites, lipids, starch and proteins from a single sample of leaf tissue. This involved transferring 30 mg of lyophilized material into microtubes containing 1.5 mL of the extraction solvent mixture, Methyl-tert-butyl-ether: methanol (MTBE:MeOH, 3:1, v:v), along with internal standards U-^{13}C sorbitol and L-Alanine-d4 (50 µL/100 mL).

The microtubes were vortexed and incubated on an orbital shaker (40 rpm) for 45 min at 4 °C followed

Table 1 Classification of the five *Coffea arabica* cultivars studied: genetic origin, resistance to coffee leaf rust (CLR) and nematode, cup quality and coffee production

Cultivar	Genetic origin(1)	CLR resistance	Nematode resistance	Cup quality	Coffee production
Catuaí Vermelho IAC 144	Caturra amarelo IAC 476-11 × Mundo Novo IAC 374-19	Susceptible(2,3)	Susceptible(2)	Good(2)	High(2)
Catiguá MG2	Catuaí amarelo IAC 86 × Híbrido de Timor UFV 440-10	Resistant(2,3)	Susceptible(2,4)	Excellent(5)	High(4)
Catiguá MG3	Catuaí amarelo IAC 86 × Híbrido de Timor UFV 440-10	Resistant(2,3)	Resistant(2,4)	Good(4)	High(2,4)
Paraíso 2	Catuaí amarelo IAC 30 × Híbrido de Timor UFV 445-46	Resistant(3)	Susceptible(2)	Excellent(5)	High(2)
Sarchimor MG 8840	Villa Sarchi CIFC 971/10 × Híbrido de Timor CIFC 832/2	Resistant(2)	Susceptible(2)	Regular(2)	High(2)

(1) IAC—Instituto Agronômico de Campinas (Brazil); MG—Minas Gerais (Brazil); CIFC—Centro de Investigação das Ferrugens do Cafeeiro (Portugal); UFV—Universidade Federal de Viçosa (Brazil). (2) De Carvalho et al. (2022); (3) Oliveira et al. (2021); (4) Fazuoli et al. (2007); (5) Empresa de Pesquisa Agropecuária de Minas Gerais (2019)

by a 15 min sonication step in a water bath at 4 °C. The microtubes were then centrifuged (10,000×*g*) for 10 min at 4 °C. From the soluble fractions, aliquots of 0.5 mL were transferred to new microtubes and dried in the speed vacuum for 45 min for the metabolomic analyses.

2.3 Metabolite profiling

Dried metabolic extracts in sealed, crimped vials were derivatized using 10 µL of a 20 mg/mL solution of methoxyamine hydrochloride in pyridine and 10 µL of *N*-trimethylsilyl-*N*-methyl trifluoroacetamide (MSTFA) following the protocol of Erban et al. (2007) and injected using a Gerstel MPS2-XL (Gerstel, Muhlheim/Ruhr, Germany) autosampler. A LECO Pegasus BT time-of-flight mass spectrometer (LECO, St. Joseph, MI, USA) hyphenated with an Agilent 8890 gas chromatograph and helium as the carrier gas at 1.0 mL/min flow and linear velocity as flow control mode. The front inlet temperature was set at 230 °C, and samples were run in splitless mode. The capillary column used was an Agilent DB-35MS (30 m×0.25 mm×0.25 µm). The temperature program started at 85 °C in isothermal mode for 2 min. The isothermal step was followed by a 15 °C/min ramp to 360 °C. Fatty acid methyl esters (FAME MIX: C8-C30) were used in the determination of the retention time index (Kind et al. 2009). The ion source was set to 250 °C. The recorded mass range is m/z = 50–600 at 30 scans/s. A solvent delay of 3 min was used. Filament bias current was − 70 V, and detector voltage ranged from approximately 1900–2000 V depending on detector age. The instrument was tuned and validated according to EPA tune compliance.

Metabolic features were identified by employing the LECO ChromaTOF software in conjunction with the GOLM Metabolome Database. Peak intensities were determined using the TargetSearch package (Cuadros-Inostroza et al. 2009) within the R software (R Core Team 2023). For normalisation, the concentration of the internal standard L-Alanine-d_4 was utilized. All metabolic class identifications were manually validated based on GOLM (Hummel et al. 2013), KEGG (Kanehisa et al. 2023), and HMDB (Wishart et al. 2022) metabolite databases. After annotating metabolic classes, 123 metabolic features were excluded from subsequent analysis. These features included reagents (44), unassigned or unknown features (10), and those not clearly identified as plant metabolites (69).

2.4 Data analysis

Metabolic classes were analyzed using Microsoft Office Excel (Microsoft Corp., Redmond, WA, USA) and R software (R Core Team 2023). Univariate and multivariate analyses were conducted using MetaboAnalyst 6.0 (Pang et al. 2024). Metabolic features with constant or single value across samples were excluded from the statistical analysis. Additionally, features with more than 50% missing values were removed, and for the remaining features, missing values were imputed with 1/5 of the minimum positive value for each variable. All data were log-transformed (base 10), and differences among the five groups (cultivars) were evaluated using a one-way analysis of variance (ANOVA). The significance level was set at $p < 0.05$. Significant metabolic features were further analyzed using partial least squares discriminant analysis (PLS-DA) and hierarchical cluster heatmap to cluster the cultivar samples based on auto-scaled variables. The cross-validation method was a fivefold cross-validation (CV), with a maximum of five components considered for selection and evaluation based on three distinct accuracy metrics (Supplementary Fig. 2). A variable importance plot (VIP) was generated to identify the most significant variables for distinguishing the five groups (cultivars) in the PLS-DA model (VIP > 1).

3 Results

Non-targeted metabolomic analysis identified 463 metabolic features (MFs) in the leaves of *C. arabica*, consistently present in all the cultivars examined (Supplementary Fig. 3). As multiple MFs can correspond to the same metabolite, duplicates were then removed, resulting in the identification of 398 unique metabolites (Supplementary Table 2). Metabolites were categorized by their chemical and functional metabolic characteristics using databases, such as, GOLM and KEGG pathways, enabling the identification of key classes in the *C. arabica* metabolome. Major metabolic classes identified were sugars (21%), amino acids (20%), lipids (12%) phenylpropanoids

and phenolics (11%), and other organic acids (11%) as shown in Fig. 1. The compositional breakdown provides valuable insights into the metabolic diversity and complexity of *C. arabica* leaves, highlighting the roles of primary and secondary metabolites in coffee plant physiology.

Among the 463 metabolic features (MFs) identified, only 41 MFs were statistically significant in distinguishing the five cultivars assessed. Amino acids, sugars, alkaloids, and organic acids collectively accounted for 68% of these significant MFs (Table 2). In the cross-validation of the PLS-DA model, the accuracy exceeded 80% up to the third component (Supplementary Fig. 2). The PLS-DA analysis distinguished Catiguá cultivars (MG2 and MG3) from the others (Catuaí, Paraíso 2, and Sarchimor), explaining 44.7% of the total variance in the first component (Fig. 2). The second component accounted for 10.1% of the variance, effectively separating the Sarchimor cultivar from the others.

A hierarchical cluster heatmap analysis allowed visualisation of the relative values of the 41 metabolites that significantly change between cultivars (Fig. 3). Based on the variable importance plot (VIP), 15 out of the 41 significant metabolites were identified as the most relevant variables for distinguishing the five cultivars (highlighted in bold in the heatmap, Fig. 3). These significant 15 variables remained mostly consistent across the PLS-DA components (Supplementary Fig. 4). Five key metabolites emerged as the most important for the model: ferulic acid, theobromine, and octopamine, followed by rosmarinic acid and gibberellin A4 (VIP scores > 1.5) (Fig. 3 and Supplementary Fig. 4, B to F). The Catiguá MG2 cultivar exhibited elevated concentrations of four key metabolites, with the exception of octopamine, which was present in higher concentrations in the Paraíso 2 and Catuaí cultivars (Fig. 3 and Supplementary Fig. 4, B to F, VIP Scores and heatmap).

4 Discussion

In field-based plant metabolomics studies, maintaining the viability of plant material from collection to analysis remainss a major challenge. The metabolic extraction protocol outlined by Salem et al. (2020), adapted in this study for freeze-dried leaves, successfully facilitated the non-targeted metabolomics analysis of *C. arabica* cultivars. This approach allowed the identification of a substantial number of metabolic features, highlighting the most significant metabolic classes, such as sugars, amino acids, phenols, and lipids, in *C. arabica* cultivars grown in Cerrado

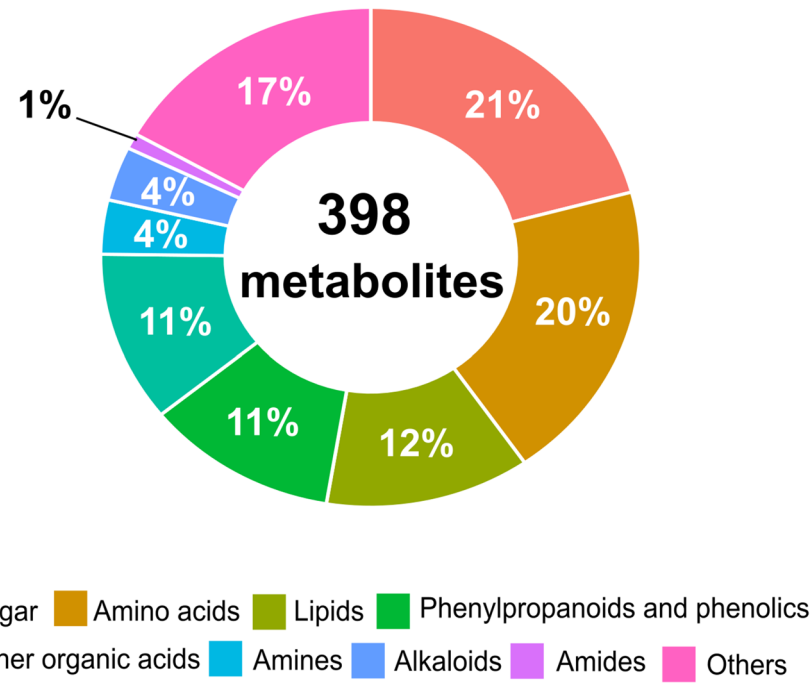


Fig. 1 Distribution of 398 unique metabolites detected in *C. arabica* leaves, categorized into various metabolic classes

Table 2 Classification of the statistically significant metabolic features distinguishing the five *C. arabica* cultivars

Metabolic feature[(a)]	p-value[(b)]	RI interval[(c)]	Linked metabolite	Metabolic class
1-Pyrroline-2-carboxylate (1TMS)	< 0.01	374,286–374,526	1-Pyrroline-2-carboxylate	Amino acids
Butanoic acid, 4-amino-3-hydroxy- (3TMS)	< 0.001	425,294–426,562	4-Amino-3-hydroxybutanoate	Amino acids
Cystathionine (4TMS)	< 0.001	753,045–754956	L-Cystathionine	Amino acids
Tryptophan (2TMS)_1	< 0.001	790,065–790537	Tryptophan	Amino acids
Tryptophan (2TMS)_2	< 0.01	814,376–814,681	Tryptophan	Amino acids
Glutamic acid, N-acetyl- (3TMS)	< 0.01	579,209–581005	N-Acetyl-L-glutamate	Amino acids
Hippuric acid (1TMS)	< 0.01	692,005–693045	Hippuric acid	Amino acids
Leucine (2TMS)	< 0.01	307,233–307462	L-Leucine	Amino acids
Orotic acid, 4,5-dihydro- (3TMS)	< 0.001	615,528–616,572	L-Dihydroorotate	Amino acids
Tyrosine, 3-nitro- (3TMS)	< 0.01	795,791–796,300	3-Nitrotyrosine	Amino acids
4-Hydroxyphenyl-beta-glucopyranoside (5TMS)	< 0.001	871,143–871,361	Arbutin	Sugar
Arabitol (5TMS)	< 0.001	501,789–503019	D-Arabitol	Sugar
beta-D-Fructofuranosyl-(2,1)-beta-D-Fructofuranose (1MEOX) (8TMS) MP	< 0.01	855,615–856,221	Inulobiose	Sugar
Gentiobiose (1MEOX) (8TMS) MP	< 0.001	925,546–926,040	Gentiobiose	Sugar
Gluconic acid-1,5-lactone (4TMS)	< 0.001	624,583–625,875	D-Glucono-1,5-lactone	Sugar
Digitoxose (1MEOX) (3TMS) MP	< 0.01	479,899–480,784	D-Digitoxose	Sugar
Glyceric acid (3TMS)	< 0.01	346,482–346,736	L-Glycerate	Sugar
Digitoxose (1MEOX) (3TMS) BP	< 0.01	479,892–481,076	D-Digitoxose	Sugar
Glycerol (3TMS)	< 0.01	295,565–295,805	Glycerol	Sugar
2-Piperidinecarboxylic acid (2TMS)	< 0.01	374,241–374,353	Pipecolic acid	Alkaloids
Theobromine (1TMS)	< 0.001	785,361–787,089	Theobromine	Alkaloids
Theophylline (1TMS)	< 0.001	737,150–737800	Theophylline	Alkaloids
Thiazole, 4-methyl-5-hydroxyethyl- (1TMS)	< 0.001	426,506–426715	4-Methyl-5-(2'-hydroxyethyl)-thiazole	Alkaloids
Tryptamine, 1-methyl- (2TMS)	< 0.001	785,205–786414	N-Methyltryptamine	Alkaloids
Ferulic acid, trans- (2TMS)	< 0.001	754,284–755,047	Ferulic acid	Phenylpropanoids or phenolic compounds
Hydrocaffeic acid (3TMS)	< 0.001	666,766–667,192	3,4-Dihydroxyphenylpropanoate	Phenylpropanoids or phenolic compounds
Rosmarinic acid (5TMS)	< 0.01	1,133,997–1,135,298	Rosmarinic acid	Phenylpropanoids or phenolic compounds
Aconitic acid, cis- (3TMS)	< 0.01	581,718–583,537	cis-Aconitate	Other organic acids
Hippuric acid, 2-hydroxy- (3TMS)	< 0.01	685,475–686,300	Salicyluric acid	Other organic acids
Malic acid (3TMS)	< 0.01	442,458–443,060	L-Malate	Other organic acids
Shikimic acid (4TMS)	< 0.001	584,778–586,715	Shikimate	Other organic acids
Phenethylamine (2TMS)	< 0.001	489,914–490,638	Phenethylamine	Amines

Table 2 (continued)

Metabolic feature[a]	p-value[b]	RI interval[c]	Linked metabolite	Metabolic class
Dopamine (4TMS)	< 0.01	692,779–693,045	Dopamine	Amines
Octopamine (4TMS)	< 0.001	666,161–666,481	Octopamine	Amines
Ergosterol (1TMS)	< 0.01	1,052,029–1053269	Ergosterol	Lipids
Gibberellin A4 (2TMS)	< 0.001	943,755–944,850	Gibberellin A4	Lipids
Ascorbic acid (4TMS)	< 0.01	650,395–651185	L-Ascorbic acid	Others
Dehydroascorbic acid dimer (2MEOX) MP	< 0.01	624,221–625,314	Dehydroascorbic acid	Others
Galactonic acid-1,4-lactone (4TMS)	< 0.001	617,495–618,546	D-Galactono-1,4-lactone	Others
Kynurenine (3TMS)	< 0.01	779,498–780,164	L-Kynurenine	Others
Orotic acid (3TMS)	< 0.01	584,880–586808	Orotic acid	Others

[a]Products of the derivatization process, annotated according to the GOLM specific library

[b]ANOVA and Tukey HSD test ($p < 0.05$)

[c]RI (retention index) interval of the metabolic features found in the samples

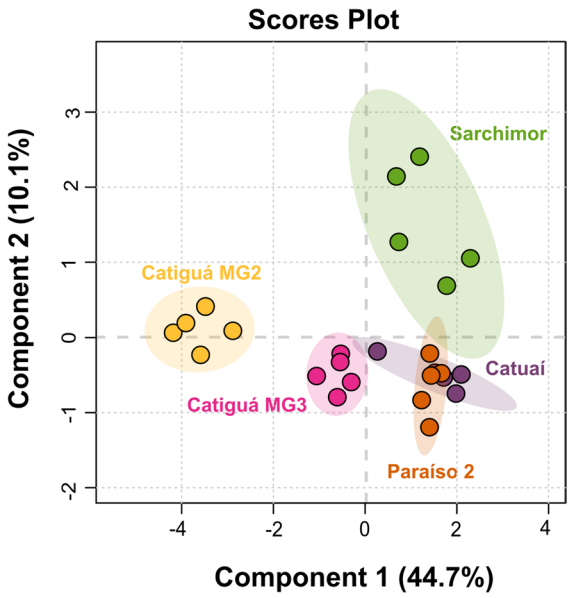
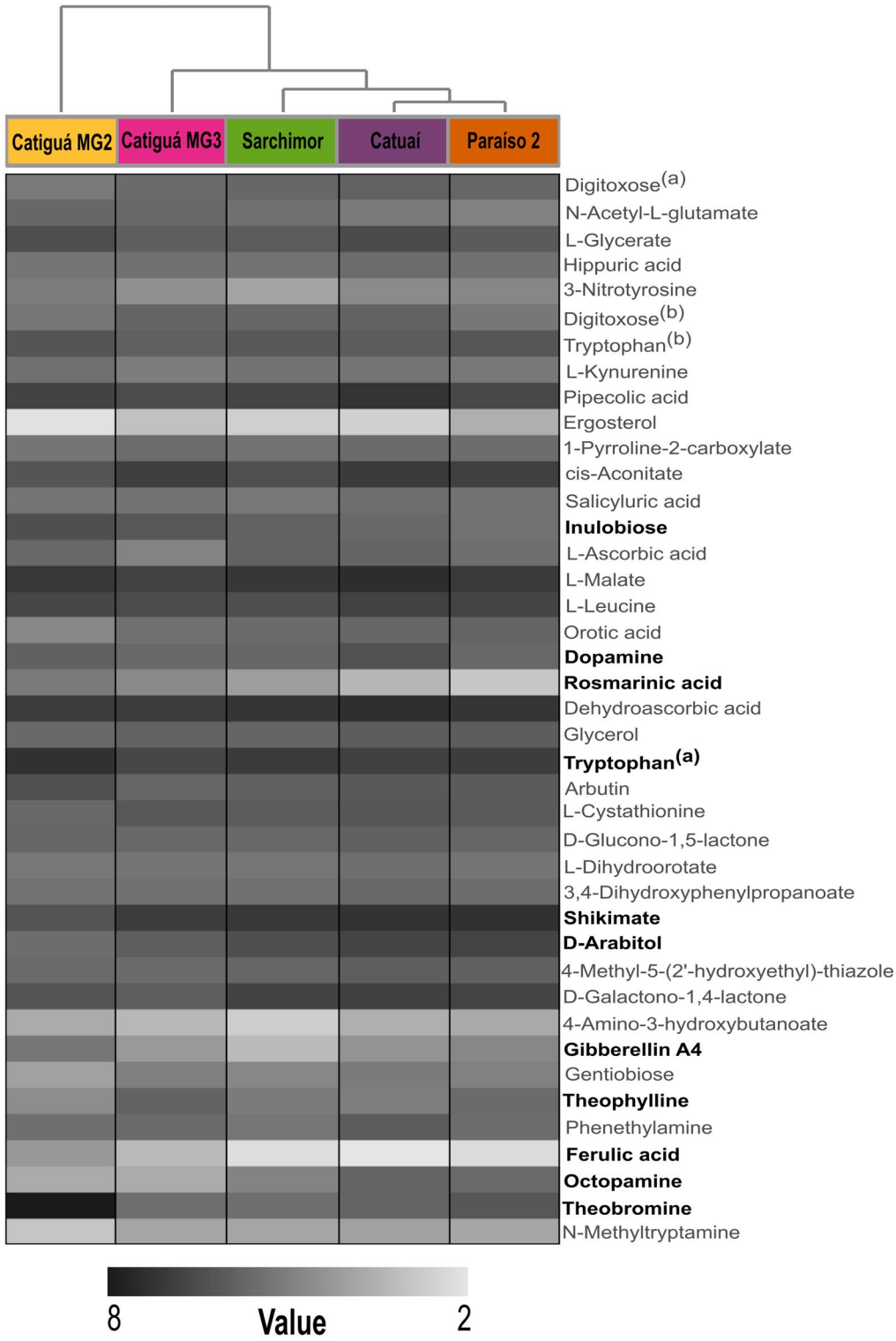


Fig. 2 Leaf non-targeted metabolomic analysis of the five *C. arabica* cultivars. A: Partial Least Squares Discriminant Analysis (PLS-DA) performed for the significant metabolites that change between cultivars, according to the ANOVA and Tukey HSD test ($p < 0.05$)

Mineiro, one of Brazil's most important coffee-producing regions. Although the lipid concentration identified was lower than typically reported for coffee plants (Bastian et al. 2021), this finding aligns with the method employed, which is designed to extract polar metabolites (Salem et al. 2020). Additionally, several of the metabolites identified in the leaves of the five *C. arabica* cultivars have already been described in coffee leaves by other studies (Chen 2018; Cangeloni et al. 2022; Montis et al. 2022; Castro-Moretti et al. 2023), providing consistency and validation of the findings.

To distinguish the five *C. arabica* cultivars, sugars and amino acids accounted for nearly 50% of the most significant metabolites, but other important compound classes, such as phenylpropanoids, alkaloids, and lipids, also played key roles. Notably, two phenolic compounds, ferulic acid and rosmarinic acid, were crucial for differentiating the cultivars. Ferulic acid, synthesized via the shikimate pathway (Manivel and Chen 2021), plays essential roles in structural and physiological processes. Its contrasting accumulation patterns, along with shikimate, suggest a bottleneck in its conversion in the Catuaí, Paraíso 2, and Sarchimor cultivars. Ferulic acid contributes to cell wall stability (Kumar and Pruthi 2014), potentially enhancing tolerance to environmental stresses. This compound is a precursor of chlorogenic acids (CGA) (Marques and Farah 2009), a major class of phenolic acids known to accumulate in young coffee leaves (Monteiro et al. 2020). Previous studies have shown its presence in leaves (Chen 2018) and fruits (Kumar and Pruthi 2014), but it has not been detected in stems (Carréra et al. 2024), highlighting its tissue-specific roles. Beyond its structural functions, ferulic acid may also exhibit antioxidant properties against biotic stressors (Ahlawat et al. 2024).


Catiguá MG2
Catiguá MG3
Sarchimor
Catuaí
Paraíso 2
Digitoxose(a)
N-Acetyl-L-glutamate
L-Glycerate
Hippuric acid
3-Nitrotyrosine
Digitoxose(b)
Tryptophan(b)
L-Kynurenine
Pipecolic acid
Ergosterol
1-Pyrroline-2-carboxylate
cis-Aconitate
Salicyluric acid
Inulobiose
L-Ascorbic acid
L-Malate
L-Leucine
Orotic acid
Dopamine
Rosmarinic acid
Dehydroascorbic acid
Glycerol
Tryptophan(a)
Arbutin
L-Cystathionine
D-Glucono-1,5-lactone
L-Dihydroorotate
3,4-Dihydroxyphenylpropanoate
Shikimate
D-Arabitol
4-Methyl-5-(2'-hydroxyethyl)-thiazole
D-Galactono-1,4-lactone
4-Amino-3-hydroxybutanoate
Gibberellin A4
Gentiobiose
Theophylline
Phenethylamine
Ferulic acid
Octopamine
Theobromine
N-Methyltryptamine
8
Value
2

◂**Fig. 3** Heatmap analysis of the metabolites that significantly change between cultivars, according to the ANOVA and Tukey HSD test ($p < 0.05$). The color scale represents metabolic feature relative value, shown in a gradient from black (higher value) to white (lower value). Metabolic features highlighted in bold represent key discriminants for cultivar differentiation based on the PLS-DA model. (Color figure online)

Similarly, rosmarinic acid, another phenylpropanoid, was consistently detected in the cultivars studied here, as well as in *C. arabica* Obatã IAC 1669-20 (Castro-Moretti et al. 2023). However, it has been reported as absent in *C. arabica* Mundo Novo 376/4 and Iapar-59 (Da Silva et al. 2021). These results suggest that rosmarinic acid could serve as a cultivar-specific biomarker. Additionally, its antioxidant and anti-stress properties further support its role in enhancing cultivar resilience.

Alkaloids, a well-known and functionally significant group of coffee metabolites, accounted for 12% of the significant features identified. Among these, theobromine, a methylxanthine involved in caffeine biosynthesis (Jin et al. 2016; Montis et al. 2024), was particularly noteworthy. A recent study demonstrated a biochemical connection between leaf and bean metabolites, demonstrating that caffeine concentrations in young leaves (Monteiro et al. 2020) were comparable to those found in green beans (Zheng et al. 2004). This relationship underscores the importance of leaf metabolomics not only for understanding plant physiological responses but also for providing insights into traits associated with coffee bean quality. Theobromine concentrations were significantly higher in the Catiguá MG2 cultivar compared to the others, reflecting its unique metabolic profile (Fig. 3A). Dopamine, while classified as a biogenic amine, exhibits alkaloid-like properties, and was also identified as one of the significant metabolites. Derived from tyrosine, through the action of polyphenol oxidases and DOPA decarboxylase, dopamine plays a crucial role in plant growth, development, and in mitigating stress responses to both biotic and abiotic factors (Cherubino Ribeiro et al. 2023; Liu et al. 2020). Similarly, octopamine, another tyrosine-derived amine (Lee et al. 2009; Kanehisa et al. 2023), also plays a key role in distinguishing the five cultivars (Fig. 3C). Together, these findings underscore the role of alkaloids and biogenic amines in not only distinguishing cultivars but also enhancing stress resilience and potentially influencing cup quality.

Beyond alkaloids, metabolites derived from the shikimate pathway, such as tryptophan, also emerged as key distinguishing features. Tryptophan, which accumulated more in the Catiguá MG2 cultivar, is synthesized after anthranilate conversion (Desmet et al. 2021) and plays a pivotal role in stress response pathways. Along with dopamine and shikimate (Desmet et al. 2021; Cherubino Ribeiro et al. 2023), tryptophan highlighted the critical role of alkaloids in distinguishing the coffee cultivars. Gibberellin A4, a bioactive gibberellin and growth-promoting hormone, also varied significantly among cultivars. This compound is known to enhance drought stress resilience (Shohat et al. 2021), exhibit antioxidant activity (Nani et al. 2022) and, when combined with gibberellin A7 (GA4 + 7), mitigates the severe reduction of photosynthetic pigments, proteolysis, and lipid peroxidation (Ritonga et al. 2023). These findings are particularly relevant given that leaf samples were collected during the dry season, suggesting a link between gibberellin A4 concentrations and the varying drought tolerances of the studied cultivars. Such differences provide valuable insights into how specific metabolites contribute to cultivar adaptability under challenging environmental conditions.

The identification of these metabolic signatures in coffee leaves not only provides a biochemical basis for distinguishing *C. arabica* cultivars but also highlights their broader ecological and agronomic significance. Metabolites like theobromine, rosmarinic acid, and ferulic acid are closely tied to physiological and stress-adaptive traits, suggesting that targeted metabolomics could play a vital role in addressing challenges posed by climate change and increasing pest pressures. For instance, phenolic compounds could enhance pathogen resistance, while methylxanthines and gibberellins could support drought tolerance, ensuring the resilience of coffee plants in diverse growing conditions.

Future research should focus on targeted metabolomics and enzyme activity assays to better understand the regulatory mechanisms underlying these metabolic pathways. By studying how metabolites like alkaloids, phenylpropanoids, and gibberellins respond to specific stressors—such as drought, pests, or temperature fluctuations—researchers could identify key markers for breeding resilient cultivars. Multi-omics approaches integrating metabolomics with transcriptomics and proteomics under controlled

stress conditions would provide a deeper understanding of the dynamic interplay between genotype, environment, and metabolism. For example, exploring caffeine metabolism or the biosynthesis of phenolic compounds in stressed plants could yield actionable insights for breeding and agricultural optimization.

Finally, the potential applications of leaf metabolomics extend beyond differentiation to breeding programs aimed at improving both agronomic and sensory attributes of coffee (Gamboa-Becerra et al. 2019). Metabolites like theobromine, ferulic acid, and gibberellin A4 not only contribute to stress tolerance but also indirectly influence traits associated with flavor and cup quality. Breeders could leverage these biomarkers to select cultivars that combine superior resilience with desirable sensory profiles. For instance, Catiguá MG2's distinct metabolic profile—marked by elevated concentrations of theobromine, rosmarinic acid, and gibberellin A4—reflects its adaptability to drought and its superior flavor characteristics, as confirmed in previous studies (Reichel et al. 2023).

By establishing the metabolic underpinnings of *C. arabica* cultivar differentiation, this study demonstrates how leaf metabolomics can support coffee breeding programs, improve agricultural practices, and ensure the future of high-quality coffee production. Moving forward, combining metabolomics with agronomic and ecological data will be essential for developing strategies to enhance coffee sustainability and productivity in the face of evolving industry demands and environmental challenges.

Funding Open access funding provided by FCT|FCCN (b-on). Financial support was provided by the Brazilian agencies: National Counsel of Technological and Scientific Development- CNPq, the Brazilian Consortium Coffee Research and Development, the National Institute of Coffee Science and Technology (INCT Café/CNPq), Coordination for the Improvement of Higher Education Personnel (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior — CAPES, funding code: 001), and the Foundation for Research Support of the State of Minas Gerais (FAPEMIG). Financial support by the Access to Research Infrastructures, Horizon2020 Programme of the European Union (EPPN2020 Grant Agreement 731013). Financial support by Portuguese funds from FCT - Fundação para a Ciência e a Tecnologia, I.P., in the scope of the project UIDP/04378/2020 (DOI: 10.54499/UIDP/04378/2020) and UIDB/04378/2020 (DOI: 10.54499/UIDB/04378/2020) of the Research Unit on Applied Molecular Biosciences - UCIBIO and the project LA/P/0140/2020 (DOI: 10.54499/LA/P/0140/2020) of the Associate Laboratory Institute for Health and Bioeconomy - i4HB. This research was also supported by LEAF – Linking Landscape, Environment, Agriculture and Food Research Center under the projects UIDB/04129/2020 (DOI: 10.54499/UIDB/04129/2020) and UIDP/04129/2020 (DOI: 10.54499/UIDP/04129/2020); and Associate Laboratory TERRA LA/P/0092/2020 (DOI: 10.54499/LA/P/0092/2020) all funded by FCT I.P., Portugal.

Data availability Data available on request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Ahlawat YK, Singh M, Manorama K, Nita L, Zaid A, Zulfiqar F (2024) Plant phenolics: neglected secondary metabolites in plant stress tolerance. Braz J Bot 47:703–721. https://doi.org/10.1007/s40415-023-00949-x

Amalia F, Irifune T, Takegami T, Yusianto SU, Putri SP, Fukusaki E (2023) Identification of potential quality markers in Indonesia's Arabica specialty coffee using GC/MS-based metabolomics approach. Metabolomics 19:1–11. https://doi.org/10.1007/s11306-023-02051-5

Bastian F, Hutabarat OS, Dirpan A, Nainu F, Harapan H, Emran TB, Simal-Gandara J (2021) From plantation to cup: changes in bioactive compounds during coffee processing. Foods. https://doi.org/10.3390/foods10112827

Bollen R, Rojo-Poveda O, Verleysen L, Ndezu R, Tshimi EA, Mavar H, Ruttink T, Honnay O, Stoffelen P, Stévigny C, Souard F, Delporte C (2024) Metabolite profiles of green leaves and coffee beans as predictors of coffee sensory quality in Robusta (*Coffea canephora*) germplasm from the Democratic Republic of the Congo. Appl Food Res 4(2):100560. https://doi.org/10.1016/j.afres.2024.100560

Caixeta ET, Resende MDV, Alkimim ER, Sousa TV, de Oliveira ACB, Pereira AA, Alves RS (2022) Aceleração do melhoramento do cafeeiro via seleção genômica: agilidade e eficácia no lançamento de novas cultivares, Documentos 17. Embrapa Café, Brasília

Cangeloni L, Bonechi C, Leone G, Consumi M, Andreassi M, Magnani A, Rossi C, Tamasi G (2022) Characterization of extracts of coffee leaves (*Coffea arabica* L.) by spectroscopic and chromatographic/spectrometric techniques. Foods. https://doi.org/10.3390/foods11162495

Carréra JC, De Souza RR, Gama Batista AC, Campolina GA, Da Silva Júnior FG, Gavilanes ML, Guimarães RJ, Cardoso M das G, Mori FA (2024) Using underutilized residues of coffee to obtain valuable dietary and antioxidant bioactive compounds. J Sci Food d Agric 104(5):2660–2668. https://doi.org/10.1002/jsfa.13151

Castro-Moretti FR, Cocuron J-C, Castillo-Gonzalez H, Escudero-Leyva E, Chaverri P, Guerreiro-Filho O, Slot JC, Alonso AP (2023) A metabolomic platform to identify and quantify polyphenols in coffee and related species using liquid chromatography-mass spectrometry. Front Plant Sci. https://doi.org/10.3389/fpls.2022.1057645

Chekol H, Warkineh B, Shimber T, Mierek-Adamska A, Dabrowska GB, Degu A (2024) Drought stress responses in Arabica coffee genotypes: physiological and metabolic insights. Plants 13:828. https://doi.org/10.3390/plants13060828

Chen X (2018) A review on coffee leaves: phytochemicals, bioactivities, and applications. Crit Rev Food Sci Nutr 59(6):1008–1025. https://doi.org/10.1080/10408398.2018.1546667

CherubinoRibeiro TH, de Oliveira RR, das Neves TT, Santiago WD, Mansur BL, Saczk AA, Vilela de Resende ML, Chalfun-Junior A (2023) Metabolic pathway reconstruction indicates the presence of important medicinal compounds in *Coffea* such as L-DOPA. Int J Mol Sci. https://doi.org/10.3390/ijms241512466

Cuadros-Inostroza Á, Caldana C, Redestig H, Kusano M, Lisec J, Peña-Cortés H, Willmitzer L, Hannah MA (2009) TargetSearch—a Bioconductor package for the efficient preprocessing of GC-MS metabolite profiling data. BMC Bioinform 10:1–12. https://doi.org/10.1186/1471-2105-10-428

Da Silva JAG, de Resende MLV, Ribeiro IS, Lima AR, Albuquerque LRM, Monteiro ACA, Pereira MHB, dos Santos Botelho DM (2021) Chemical composition, production of secondary metabolites and antioxidant activity in coffee cultivars susceptible and partially resistant to bacterial Halo blight. Plants. https://doi.org/10.3390/plants10091915

De Carvalho CHS, Bartelega L, Sera GH, Matiello JB, Santinato SRAF, Hotz AL (2022) Catálogo de cultivares de café arábica. Documentos 16. Embrapa Café, Brasília

Desmet S, Morreel K, Dauwe R (2021) Origin and function of structural diversity in the plant specialized metabolome. Plants. https://doi.org/10.3390/plants10112393

Dussarrat T, Schweiger R, Ziaja D, Nguyen TT, Krause L, Jakobs R, Eilers EJ, Müller C (2023) Influences of chemotype and parental genotype on metabolic fingerprints of tansy plants uncovered by predictive metabolomics. Sci Rep. https://doi.org/10.1038/s41598-023-38790-7

Empresa de Pesquisa Agropecuária de Minas Gerais – EPAMIG (2019) Cultivares de café – Programa de Melhoramento Genético EPAMIG, EMBRAPA Café, UFV e UFLA. Epamig, Minas Gerais

Erban A, Schauer N, Fernie AR, Kopka J (2007) Nonsupervised construction and application of mass spectral and retention time index libraries from time-of-flight gas chromatography-mass spectrometry metabolite profiles. In: Weckwerth W (ed) Metabolomics—methods in Molecular Biology™, v. 358. Humana Press, pp 19–38. https://doi.org/10.1007/978-1-59745-244-1_2

Fazuoli LC, de Carvalho HS, Carvalho GR et al (2007) CULTIVARES DE CAFÉ ARÁBICA (*Coffea arabica* L.). In: Carvalho CHS (ed) Cultivares de café. 2007. Embrapa, Brasília, pp 125–198

Gamboa-Becerra R, Hernández-Hernández MC, González-Ríos Ó, Suárez-Quiroz ML, Gálvez-Ponce E, Ordaz-Ortiz JJ, Winkler R (2019) Metabolomic markers for the early selection of *Coffea canephora* plants with desirable cup quality traits. Metabolites 9:214. https://doi.org/10.3390/metabo9100214

Guimarães PS, Schenk JCM, Cintra LC, Giachetto PF, Silvarolla MB, Padilha L, Maluf MP (2021) Large-scale prospection of genes on caffeine-free Coffea arabica plants—discovery of novel markers associated with development and secondary metabolism. Plant Gene 27:100314. https://doi.org/10.1016/j.plgene.2021.100314

Hummel J, Strehmel N, Bölling C, Schmidt S, Walther D, Kopka J (2013) Mass spectral search and analysis using the golm metabolome database. In: Weckwerth W, Kahl G (eds) The handbook of plant metabolomics. Wiley-VCH Verlag GmbH & Co. KGaA, pp 321–343. https://doi.org/10.1002/9783527669882.ch18

International Coffee Organization (2024) Trade statistic tables, production, coffee trade status. http://www.ico.org/trade_statistics.asp. Accessed 20 June 2024

Jin JQ, Yao MZ, Ma CL, Ma JQ, Chen L (2016) Natural allelic variations of TCS1 play a crucial role in caffeine biosynthesis of tea plants and their related species. Plant Physiol Biochem 100:18–26. https://doi.org/10.1016/j.plaphy.2015.12.020

Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M (2023) KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res 51(D1):D587–D592

Kind T, Wohlgemuth G, Lee DY, Lu Y, Palazoglu M, Shahbaz S, Fiehn O (2009) FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. Anal Chem 81(24):10038–10048. https://doi.org/10.1021/ac9019522

Kitashova A, Brodsky V, Chaturvedi P, Pierides I, Ghatak A, Weckwerth W, Nägele T (2023) Quantifying the impact of dynamic plant-environment interactions on metabolic regulation. J Plant Physiol 290:154116. https://doi.org/10.1016/j.jplph.2023.154116

Kumar N, Pruthi V (2014) Potential applications of ferulic acid from natural sources. Biotechnol Rep 4:86–93. https://doi.org/10.1016/j.btre.2014.09.002

Lee K, Kang K, Park M, Park S, Back K (2009) Enhanced octopamine synthesis through the ectopic expression of tyrosine decarboxylase in rice plants. Plant Sci 176:46–50. https://doi.org/10.1016/j.plantsci.2008.09.006

Liu Q, Gao T, Liu W, Liu Y, Zhao Y, Liu Y, Li W, Ding K, Ma F, Li C (2020) Functions of dopamine in plants: a

review. Plant Signal Behav. https://doi.org/10.1080/15592324.2020.1827782

Manivel P, Chen X (2021) Chlorogenic, caffeic, and ferulic acids and their derivatives in foods. In: Xiao J, Sarker SD, Asakawa Y (eds) Handbook of dietary phytochemicals. Springer, Singapore, pp 1033-1063. https://doi.org/10.1007/978-981-15-4148-3_22

Marques V, Farah A (2009) Chlorogenic acids and related compounds in medicinal plants and infusions. Food Chem 113(4):1370–1376. https://doi.org/10.1016/j.foodchem.2008.08.086

Martins SCV, Araújo WL, Tohge T, Fernie AR, DaMatta FM (2014) In high-light-acclimated coffee plants the metabolic machinery is adjusted to avoid oxidative stress rather than to benefit from extra light enhancement in photosynthetic yield. PLoS ONE 9(4):e94862. https://doi.org/10.1371/journal.pone.0094862

Mihai RA, Ortiz-Pillajo DC, Iturralde-Proaño KM, Vinueza-Pullotasig MY, Sisa-Tolagasí LA, Villares-Ledesma ML, Melo-Heras EJ, Cubi-Insuaste NS, Catana RD (2024) Comprehensive assessment of coffee varieties (*Coffea arabica* L.; *Coffea canephora* L.) from Coastal, Andean, and Amazonian Regions of Ecuador; a holistic evaluation of metabolism, antioxidant capacity and sensory attributes. Horticulturae 10:200. https://doi.org/10.3390/horticulturae10030200

Monteiro Â, Colomban S, Azinheira HG, Guerra-Guimarães L, Silva M do C, Navarini L, Resmini M (2020) Dietary antioxidants in coffee leaves: impact of botanical origin and maturity on chlorogenic acids and xanthones. Antioxidants 9(1):6. https://doi.org/10.3390/antiox9010006

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

Montis A, Delporte C, Noda Y, Stoffelen P, Stévigny C, Hermans C, Van Antwerpen P, Florence S (2024) Targeted metabolomics and transcript profiling of methyltransferases in three coffee species. Plant Sci. https://doi.org/10.1016/j.plantsci.2024.112117

Nagay JHC (1999) Café do Brasil: dois séculos de história. Formação Econômica 3:17–23

Nani BD, Rosalen PL, Lazarini JG, De Araújo LP, Dos Reis MS, Breseghello I, Cunha TM, De Alencar SM, Da Silveira NJ, Franchin M (2022) A study on the anti-NF-κB, anti-candida, and antioxidant activities of two natural plant hormones: gibberellin A4 and A7. Pharmaceutics 14(7):1347. https://doi.org/10.3390/pharmaceutics14071347

Oliveira ACB, Pereira AA, Caixeta ET, Resende MDV, Ribeiro MF (2021) Cultivares de café resistentes à ferrugem: alternativa viável para a cafeicultura das Matas de Minas. Embrapa Café, Brasília

Pang Z, Lu Y, Zhou G, Hui F, Xu L, Viau C, Spigelman AF, MacDonald PE, Wishart DS, Li S, Xia J (2024) MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation. Nucleic Acids Res 52(W1):W398–W406. https://doi.org/10.1093/nar/gkae253

RCoreTeam (2023) R: a language and environment for statistical computing. version 4.3.1. R Foundation for statistical computing, Vienna, Austria. https://www.R-project.org/. Accessed 27 Oct 2023

Reichel T, De Resende MLV, Nadaleti DHS, Santos FO, Botelho CE (2023) Potential of rust-resistant arabica coffee cultivars for specialty coffee production. Biosci J. https://doi.org/10.14393/BJ-v39n0a2023-66

Ritonga FN, Zhou D, Zhang Y, Song R, Li C, Li J, Gao J (2023) The roles of gibberellins in regulating leaf development. Plants 12(6):1243. https://doi.org/10.3390/plants12061243

Salem MA, Jüppner J, Bajdzienko K, Giavalisco P (2016) Protocol: a fast, comprehensive and reproducible one-step extraction method for the rapid preparation of polar and semi-polar metabolites, lipids, proteins, starch and cell wall polymers from a single sample. Plant Methods 12:45. https://doi.org/10.1186/s13007-016-0146-2

Salem MA, Perez de Souza L, Serag A, Fernie AR, Farag MA, Ezzat SM, Alseekh S (2020) Metabolomics in the context of plant natural products research: from sample preparation to metabolite analysis. Metabolites. https://doi.org/10.3390/metabo10010037

Sardans J, Peñuelas J, Rivas-Ubach A (2011) Ecological metabolomics: overview of current developments and future challenges. Chemoecology 21:191–225. https://doi.org/10.1007/s00049-011-0083-5

Shen S, Zhan C, Yang C, Fernie AR, Luo J (2023) Metabolomics-centered mining of plant metabolic diversity and function: past decade and future perspectives. Mol Plant 16:43–63. https://doi.org/10.1016/j.molp.2022.09.007

Shohat H, Cheriker H, Kilambi HV, Illouz Eliaz N, Blum S, Amsellem Z, Tarkowská D, Aharoni A, Eshed Y, Weiss D (2021) Inhibition of gibberellin accumulation by water deficiency promotes fast and long-term 'drought avoidance' responses in tomato. New Phytol 232:1985–1998. https://doi.org/10.1111/nph.17709

Souard F, Delporte C, Stoffelen P, Thévenot EA, Noret N, Dauvergne B, Kauffmann J-M, Van Antwerpen P, Stévigny C (2017) Metabolomics fingerprint of coffee species determined by untargeted-profiling study using LC-HRMS. Food Chem 245:603–612. https://doi.org/10.1016/j.foodchem.2017.10.022

Sumner LW, Amberg A, Barrett D et al (2007) Proposed minimum reporting standards for chemical analysis. Metabolomics 3:211–221. https://doi.org/10.1007/s11306-007-0082-2

Wishart DS, Guo A, Oler E et al (2022) HMDB 5.0: the human metabolome database for 2022. Nucleic Acids Res 50(D1):D622–D631. https://doi.org/10.1093/nar/gkab1062

Zhai H, Dong W, Fu X, Li G, Hu F (2024) Integration of widely targeted metabolomics and the e-tongue reveals the chemical variation and taste quality of Yunnan Arabica coffee prepared using different primary processing methods. Food Chem. https://doi.org/10.1016/j.fochx.2024.101286

Zheng Q, Nagai C, Ashihara H (2004) Pyridine nucleotide cycle and trigonelline (N-methylnicotinic acid) synthesis in developing leaves and fruits of *Coffea arabica*. Physiol Plant 122(4):404–411. https://doi.org/10.1111/j.1399-3054.2004.00422.x