Journal of Food Composition and Analysis 148 (2025) 108607

Contents lists available at ScienceDirect

Journal of Food Composition and Analysis

journal homepage: www.elsevier.com/locate/jfca

Metabolomic profiling of ten cacao genotypes from Antioquia, Colombia reveals local effects on nutritional composition

Elkin Galeano [a,*], Juan M. Lopez [b,*], Catalina Agudelo [a], Edison Osorio [a], Helena Maruenda [b]

[a] Grupo de Investigación en Sustancias Bioactivas, Facultad de Ciencias Farmacéuticas y Alimentarias, Universidad de Antioquia UdeA, Calle 70 No. 52-21, Medellín, Colombia

[b] Pontificia Universidad Católica del Perú, Departamento de Ciencias - Química, CERMN, Av. Universitaria 1801, Lima 32, Peru

ARTICLE INFO

Keywords:
Theobroma cacao
Metabolomics
Geographic variation
Genetic diversity
Proton NMR spectroscopy
Bioactive metabolites
Antioquia

ABSTRACT

Theobroma cacao beans exhibit considerable genetic and geographic diversity, which significantly influences their chemical composition. These variations affect the nutritional, health-promoting, and organoleptic qualities of cacao products. This study aimed to characterize the metabolic profiles of ten cacao genotypes cultivated in two distinct regions of Antioquia, Colombia, using proton nuclear magnetic resonance (^{1}H NMR) spectroscopy, and to evaluate the impact of geographic origin on key metabolites related to cacao quality. Twenty-one bioactive metabolites, including amino acids, sugars, and secondary metabolites, were identified and quantified in cacao beans from the Urabá and Northeast regions. Multivariate statistical analyses were employed to discern metabolic patterns associated with genotype and cultivation sites. Significant differences in metabolite concentrations were observed between the regions. Beans from Urabá displayed higher levels of amino acids, xanthine derivatives, and antioxidant compounds such as theobromine and epicatechin. Genotypes TSH565, FEAR5, ICS1, and CCN51 showed superior metabolic profiles in Urabá, while FSV41 was notable in the Northeast. Multivariate analyses revealed distinct clustering by geographic origin, highlighting the influence of environmental factors on metabolite expression. The stable, warm, and humid climate of Urabá appears to promote the accumulation of metabolites linked to enhanced protein quality, energy metabolism, and health benefits.

1. Introduction

Theobroma cacao L. (Malvaceae) is a tree endemic to the lower eastern equatorial slopes of the Andes in South America (Motamayor et al., 2002). It is well known worldwide for its valuable seeds, which are used as the starting material in chocolate production.

Historically, chocolate, praised by ancient civilizations such as the Inca, Maya, and Aztecs, was valued not only for its stimulating properties but also for its nutritional benefits. Early uses of cacao involved the preparation of beverages rich in antioxidants and providing essential minerals, such as magnesium, potassium, and iron. These nutrients contributed to the dietary needs of these populations, supporting both physical endurance and cognitive functions (Díaz-Valderrama et al., 2020).

In contemporary Latin America, cacao and its derivatives are a source of nutrition. The raw cacao bean is particularly rich in polyphenols, bioactive compounds renowned for their potent antioxidant properties, which help combat oxidative stress, a key factor linked to chronic conditions such as cardiovascular diseases and metabolic syndrome (Martín & Ramos, 2016). Regular consumption of moderate amounts of dark chocolate with a high cacao content is associated with improved cardiovascular health, enhanced blood circulation, and benefits to neuroplasticity, reducing the risk of neurodegenerative diseases (Camandola et al., 2019; Katz et al., 2011). These effects contribute to an improved overall nutritional status. Recent reviews have also highlighted the anti-obesity and broader health-promoting properties of this valuable Mesoamerican crop (Hernández-Pérez & Paredes-López, 2024).

Moreover, cacao products have evolved beyond traditional beverages to encompass a wide range of edible forms, including cacao nibs,

* Corresponding authors.
E-mail addresses: elkin.galeano@udea.edu.co (E. Galeano), juan.lopez@pucp.edu.pe (J.M. Lopez), catalina.agudelor@udea.edu.co (C. Agudelo), edison.osorio@udea.edu.co (E. Osorio), hmaruen@pucp.edu.pe (H. Maruenda).

https://doi.org/10.1016/j.jfca.2025.108607
Received 14 July 2025; Received in revised form 4 November 2025; Accepted 10 November 2025
Available online 13 November 2025

cacao powder, cocoa, chocolate, and various health supplements. These products are now well integrated into modern diets, valued not only as sources of calories but also for their rich content of essential micronutrients (Zimmermann & Ellinger, 2020).

Antioquia is a department in Colombia (South America) and has been the second most important department in terms of cacao cultivation, with an average of approximately 7154 t year^{-1} (≈11 % of national cacao production). The distribution of the cacao cultivars in Antioquia is heterogeneous. First, the coastal plain area in Urabá (northwest Antioquia) which dedicates 10376 ha in the cultivation of cacao (43 % of a total of 24380 ha reported for Antioquia), produces 3 – 6 kt of cacao beans (40–45 %). Due to its tropical location, this region experiences a warm and humid Af-type climate (Manciu et al., 2023), an average air temperature of approximately 27 °C, and an annual rainfall range of 7–15 mm (Suárez et al., 2015). The soils of this quaternary alluvial plain are deep, loamy-silty fluventic inceptisols with pHs values between 5.3–5.8, organic carbon content of 2.7–3.6 % and, exchangeable aluminum content of 1.3–2.0 cmol kg^{-1} (Causil-Pastrana et al., 2024; León-Moreno et al., 2019). These conditions favor vigorous cacao growth, provided the adequate fertilization protocols for these plants are followed (León-Moreno et al., 2019). Second, the mountainous region in northeast Antioquia, which contributes with 4168 ha (≈17 % of the total amount reported for Antioquia) yielding 1.2–2.5 kt of product (≈15–20 %), due to its altitude, is characterized by a humid submontane climate, 23.5 °C average temperature, and, approximately, 2600–3100 mm annual rainfall. The steep metamorphic slopes become ferralitic acrisols (F-Ar and F-Ar-A complexes) that are significantly more acidic (pH 4.8–5.2), have low effective cation exchange capacity (< 8 cmol kg^{-1}), and high Fe/Al oxide activity, all of which limit nutrient storage and root volume (Mira-Giraldo et al., 2025).

The quality, nutritional value, and organoleptic properties of any plant-derived product are influenced by multiple factors, including the crop's genetic makeup, local climate and ecosystem, soil characteristics, agronomic management practices, and postharvest treatments. Together, these variables shape the metabolome of the final product. In *T. cacao*, the key factors affecting bean quality have been recently reviewed (Putri et al., 2024).

Metabolomics, a state-of-the-art analytical approach that uses high-throughput instrumentation, such as mass spectrometry (MS) or nuclear magnetic resonance spectroscopy (NMR) to characterize the molecular composition of a vast number of samples, has proven useful in plant sciences. Hyphenated MS platforms based on traditional chromatographic instrumentation provide higher sensitivity (in the pico-to-nano Molar range) (Gowda & Djukovic, 2014) and a larger (> 500) set of compounds detected per scan (Roessner et al., 2001; Sawada & Hirai, 2013) compared to less than 100 identifiable compounds by NMR-based metabolic profiling, in a concentration range of high nM to low uM (depending on magnet, probe, # scans, hardware). However, ^{1}H NMR enables the simultaneous detection and quantitation (under fully relaxed conditions) of a wide range of chemically diverse metabolites without the need of physical separation (Simmler et al., 2014).

In cacao, the use of ^{1}H NMR- metabolic profiling has evolved from, the simultaneous detection and quantification of amino acids, polyalcohols, organic acids, sugars, methylxanthines, catechins, and phenolics in fermented cacao beans of different varieties and geographical origin (Caligiani et al., 2010; D'Souza et al., 2017), to the demonstration of how determinant is the fermentation level on the metabolome of the cacao beans, when compared to origin and variety (Caligiani et al., 2014). Furthermore, the use of HR-MAS ^{1}H NMR fingerprints in the chemometric analyses of cacao to trace geographical origin, demonstrated that by incorporating lipid content in the assessment the discriminating capability of the methodology was refined when contrasted to solution ^{1}H NMR profiling, solely (Marseglia et al., 2016). More recently, the power of combining methods in the construction of more robust discrimination methods was exemplified by employing time multi-element isotope-ratio mass spectrometry (IRMS) coupled with ^{1}H NMR fingerprints in the evaluation of origin and variety of cacao beans (Bindereif et al., 2019; Truzzi et al., 2023, 2024). The chemical changes in cacao followed by ^{1}H NMR metabolic profiling has also been employed with success in the 1) quality control evaluation of cacao-derived products, such as, chocolate butter blends; 2) valorization of *T. cacao* by-products, such as pod husk exposed to different dehydration treatments (Villalón-López et al., 2023), and 3) assessment of water-deficit-resistant cacao genotypes in search of better *T. cacao* hybrids (Boutchouang et al., 2024).

The production yields of Urabá (northwest) and northeast Antioquia - which, together, represent more than half of Antioquia's annual cacao bean production (Fedecacao – Fondo Nacional del Cacao, 2025) and their contrasting environmental differences, provide valuable context for the study here proposed: employ high-throughput NMR metabolic profiling, together with multivariate and correlation analyses, to study ten *T. cacao* genotypes - four of which were genetically developed by the Federación Nacional de Cacaoteros (FEDECACAO) in Colombia (Henao Ramírez et al., 2018), an organization that is dedicated to the quality and quality control of cacao, and six universal/commercial varieties-cultivated in two different geographical sites of the Department of Antioquia (Colombia), Urabá and northeast Antioquia. The integrative approach aims to elucidate the combined effects of genetic background and geographic origin on the metabolic composition of cacao beans, providing valuable insights for breeding programs and cultivation strategies focused on enhancing cacao quality and regional differentiation.

2. Materials and methods

2.1. Plant material and reagents

Ten cacao genotypes were evaluated, including six universal/commercial genotypes (EET96, TSH565, CCN51, ICS1, ICS60, and ICS95) and four regional genotypes provided by FEDECACAO (FNC): FSV41, FEAR5, FEC2, and FTA2. These genotypes were grown in two distinct geographic locations in Antioquia-Colombia, Urabá (northwest Antioquia, Turbo and Chigorodó towns) and the Northeast Antioquia region (town of Vegachí). The cacao pods were harvested during the first two weeks of December 2019.

Sampling design: For each genotype in each region, five independent biological samples were collected from different farms, yielding fifty samples per region and one hundred in total. This design was chosen to maximize metabolomic representativeness and to characterize the average behavior of each genotype across the regions.

All reagents, including deuterium oxide (99.98 %), 3-(trimethylsilyl) propionic-2,2,3,3-d_4 acid sodium salt (TMSP-d_4), potassium phosphate monobasic (KH_2PO_4), and potassium phosphate dibasic trihydrate ($K_2HPO_4{\cdot}3H_2O$) were purchased from Merck (Darmstadt, HE, Germany). Methanol (HPLC grade) was purchased from J. T. Baker (Griesheim, DE), and the water used was purified using MilliQ (Merck KGaA-Darmstadt, HE, DE).

2.2. Sample preparation

The collected cacao pods were processed following methodologies described in the literature (Agudelo et al., 2022). In brief, the unfermented dried beans were mechanically triturated (IKA® A 11 basic, IKA Works, Inc., Wilmington, NC, USA and passed through a No. 60 (250 μm) sieve. The sieved material was stored in sealed plastic containers at −20°C until further use.

The extraction process was carried out following the method of Caligiani and colleges (Caligiani et al., 2014), with minor modifications. In brief, approximately 200 mg of the cacao powder was treated with 20 mL of 80:20 v/v water: methanol solution. The suspension was stirred for 10 min at 90 °C using a thermostatic bath set to 95 °C. The suspension, cooled to room temperature, was centrifuged at 3000 x g for

5 min and concentrated to dryness under reduced pressure on a rotary evaporator (Heidolph, Schwabach, Germany) with automatic vacuum control (water bath, 30 °C, condenser, 10 °C, flask rotation, 120–180 rpm, pressure, 37 mbar (±5 mbar)). The dried extract was redissolved in 1.0 mL of potassium phosphate buffer (100 mM, pH 6.02) in D_2O, containing TMSP-d_4 at 5.85 mM. The sample was sonicated for 15 min at room temperature and centrifuged at 13000 x g for 10 min at room temperature. The supernatant was filtered through a 13 mm PTFE 0.45 µm syringe filter, and 550 µL of the filtered solution was transferred to an NMR tube for analysis. Five independent extractions for each cacao sample were performed. This procedure was used for each of the five independent samples per genotype per region.

2.3. NMR analyses

All NMR spectra were recorded at 298 K on a Bruker Avance III HD 500 MHz spectrometer equipped with a TCI He-cooled cryoprobe, autosampler, and automatic tuning and matching, available at CERMN-PUCP (Bruker®, Billerica, MA, U.S.A). The ^{1}H NMR spectra were acquired with a 30° flip angle, 5 Hz field strength water presaturation, 128 scans, 16 dummy scans, 64k data points, a spectral width of 10 kHz, and a total relaxation time of 5.28 s. The flip angle was calibrated for each sample using 360° pulse optimization. The compounds were assigned through ^{1}H–^{1} H COSY, ^{1}H–^{1}H TOCSY, ^{1}H-^{1}H J-resolved, ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC, and ^{1}H Pure-shift (Sapphire-Psyche) experiments (Lopez et al., 2019; Moutzouri et al., 2017), and corroborated against the available NMR database using the Chenomx NMR suite v8.6 software (Chenomx Inc, Edmonton, AB. Canada).

2.4. NMR spectra processing

NMR spectra were postprocessed with Mnova v14.1.0 software (Mestrelab Research, S.L; Santiago de Compostela, Spain). Each experiment was zero-filled to 256 k complex points, Fourier transformed with 0.3 Hz exponential apodization, and phase and baseline corrected.

2.4.1. Untargeted metabolomic profiling

The spectra were aligned and binned with a width of 0.04 ppm across the retained range; bin intensities were the integrated areas. To minimize dilution effects while preserving relative composition, data were normalized to total area (excluding residual water signals, from 4.90 to 4.50 ppm and TMSP-d_4, from −0.05–0.05 ppm) and mean-centered. An initial PCA (Hotelling's T^2, 95 % ellipse) was used to screen for outliers and batch effects; none, required removal. Subsequently, PLS-DA (NIPALS) was performed on the bucket matrix to evaluate genotype and growing location separation. Variable-importance scores (VIP) were computed, and features with VIP $\geq$ 1 were shortlisted. Shortlisted bins were mapped to metabolite resonances and cross-checked against the 1D/2D NMR experiments (see NMR analyses section) for assignment validation.

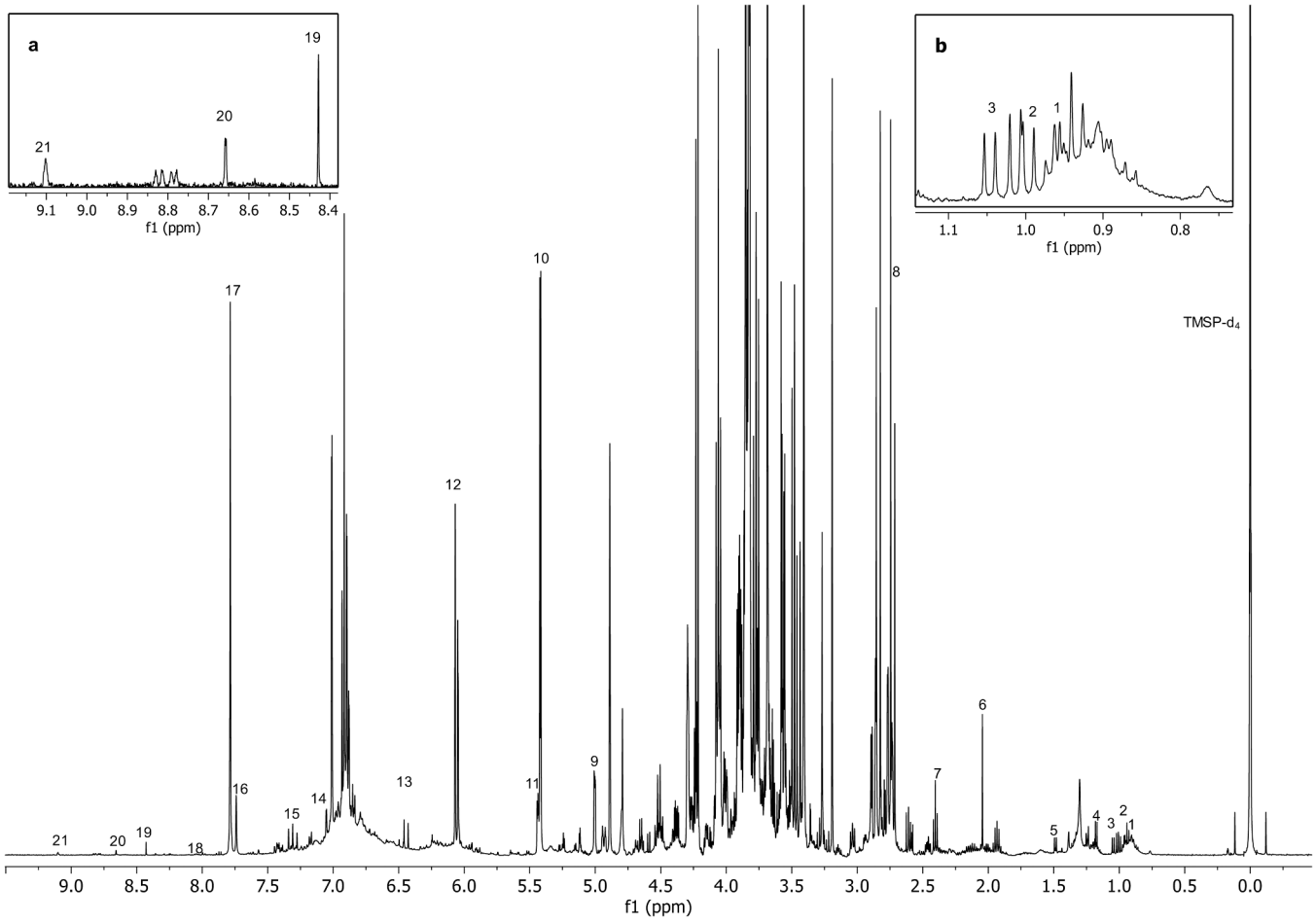

Fig. 1. ^{1}H NMR spectrum of a representative cacao methanol-aqueous extract (sample CCN-N1). Sample prepared with 100 mM potassium phosphate buffer-pH 6.02 in D_2O, containing TMSP-d_4 at 5.85 mM, as an internal standard. The signals used for the quantification of each metabolite are indicated: Ile (1), Leu (2), Val (3), DHT (4), Ala (5), Ach (6), GABA (7), CIT (8), STA (9), SUC (10), G1P (11), EPI (12), CAD (13), Tyr (14), Phe (15), CAF (16), THE (17), Xan (18), FOR (19), IMI (20), and T (21). Top panel A: Zoomed-in spectrum region between 8.4 and 9.2 ppm. Top panel B Zoomed-in spectrum region between 0.8 and 1.1 ppm.

2.4.2. Targeted metabolic profiling

The twenty-one metabolites, which discriminated samples per genotype and cultivation site, were quantified by targeted profiling and line-shape deconvolution in Chenomx NMR Suite. Quantification used diagnostic, minimally overlapped signals (see Fig. 1) with TMSP-d_4 at 5.85 mM as the qNMR internal standard; deconvolution was applied for overlapped signals. The resulting concentration matrix was used to interpret the patterns revealed in the untargeted stage and to re-evaluate group structure.

2.5. Statistical analysis

In each area, fifty cacao samples were collected in each region. For each clone, five samples were taken from different farms to maximize metabolomic representativeness and characterize the average behavior of each species across regions. Ninety-eight cacao extracts were analyzed, and two samples were excluded due to spoilage during analysis. The Mann-Whitney-Wilcoxon test was used to compare the behavior of the genotypes according to their geographic origin. The metabolite content of each genotype grown in both regions was compared using the Kruskal-Wallis test. A p-value < 0.05 denotes statistical significance. The integral tables obtained were analyzed with Statgraphics® v19 Centurion software (Statgraphics Technologies, Inc., The Plains, VA, U.S.A), and box plots were generated with the RStudio: Integrated Development Environment for R. (Boston, U.S.A) software package. Principal Component Analysis (PCA) was performed using The Unscrambler X (Camo Analytics, Oslo, Norway) software. The PLS-DA model was validated by class-balanced 7-fold cross-validation (LV selection by Q^2), CV-ANOVA, and permutation testing (n = 200); R^2Y, Q^2, CV-ANOVA p-value, and cross-validated balanced accuracy are reported in the explanatory text of Fig. 2.

3. Results and discussion

Free amino acids, phenolic compounds, reducing sugars, and various organic acids are key quality indicators of raw cacao beans, as they influence plant development, contribute to the formation of characteristic aromas during fermentation, and are associated with the health-promoting properties of cacao and its chocolate-derived products (Paparella et al., 2025). In this study, several metabolites, including amino acids– alanine (Ala), leucine (Leu), phenylalanine (Phe), tyrosine (Tyr), valine (Val), isoleucine (Ile), and nonproteinogenic γ-aminobutyric acid (GABA); key cacao secondary metabolites– epicatechin (EPI), caffeic acid (CAD), caffeine (CAF), xanthine (XAN), and theobromine (THE); cellular energy metabolites– citrate (CIT), formate (FOR), stachyose (STA), glucose-1-phosphate (G1P), and sucrose (SUC); and other compounds key in plant metabolism, such as acetylcholine (ACh), dihydrothymine (DHT), imidazole (IMI), and trigonelline (T), were identified in the ^{1}H NMR profile of the cacao pod extract (Fig. 1, Table S1), using Chenomx software and 2D NMR and pure-shift spectra (Figure S1-S5). The quantification results for the ten genotypes grown in two geographic regions, Urabá and Northeast, are presented in Table 1.

3.1. Amino acid content

Amino acids are known to participate in the development of the flavor and aroma of cacao (Rohsius et al., 2006). As previously reported, the amino acid content varies among clones and between locations (Rohsius et al., 2006). The results are displayed in Figure S6 (Supplementary Material). In the Urabá region, all the amino acids presented differences among the clones (Table 1). The clone *FEAR5* showed the highest content of all the amino acids, with Val and Ile being the most abundant. Samples from the Northeast region showed lower amounts of all amino acids compared to those from the Urabá region. *ICS60* clones from the Northeast region, on average, exhibited lower concentrations of amino acids (Ile, Ala, Leu, Phe, and Val). Conversely, *FSV41* showed higher concentrations of amino acids (Ile, Ala, Leu, and Val) than the other clones from the Northeast region. In contrast, FEAR5 was the clone from the Urabá regions with the highest amino acid concentration (all of them). These elevated amino acid levels, especially in beans from Urabá, are not only important for flavor development but also increase nutritional value by providing essential amino acids that support protein synthesis in the human diet (Fang et al., 2020).

3.2. Sugar content

Differences in sugar accumulation among *T. cacao* genotypes have been reported and the results here corroborate these findings (Rangel-Fajardo et al., 2011). The sugars identified and quantified were STA, SUC, and G1P. Among these, STA presented the most significant difference between clones (Table 1) with concentrations oscillating between 13.4 and 62.4 mM for the Northeast region and between 14.2 and 76.0 mM for the Urabá region.

As shown in Figure S7, Urabá was the region with the highest sugar

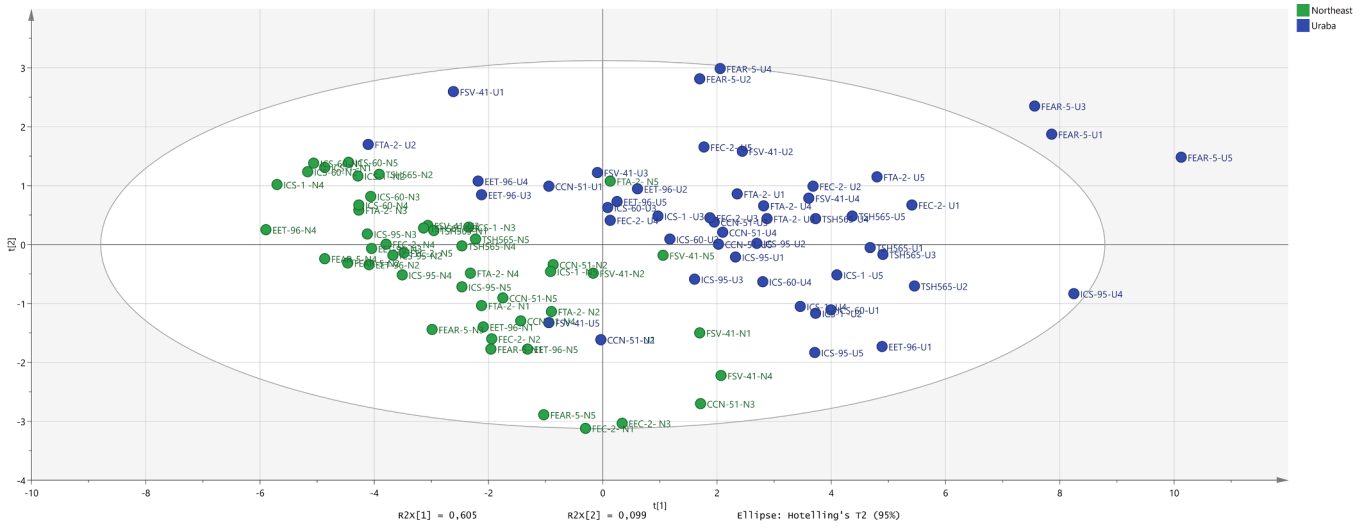

Fig. 2. PLS-DA was performed with the full 1H NMR spectra of the ten *T. cacao* bean genotypes grown at two different locations (ninety-eight extracts). Blue and green dots indicate samples from the Urabá and Northeast regions of Antioquia, Colombia, respectively. The model discriminates against geographical origin with 69.3 % accuracy, and ANOVA cross-validation yielded a p-value of 1.06e-16.

Table 1
Concentrations (mM) of discriminant metabolites identified in the aqueous extract of *T. cacao* bean genotypes.

	Acetylcholine (ACh) Concentración (mM)	Caffeic acid (CAD)	Caffeine (CAF)	Citrate (CIT)	Dihydrothymine (DHT)	Epicatechin (EPI)	Formate (FOR)	Gamma-Aminobutyric acid (GAB)	Glucose-1-phosphate (G1P)	Imidazole (IMI)	Isoleucine (Ile)
CCN51 Urabá *	0.04 ± 0.01	0.94 ± 0.21	0.51 ± 0.07	6.64 ± 1.12	0.21 ± 0.02	1.90 ± 0.58	0.22 ± 0.11	1.01 ± 0.26	1.79 ± 0.31	0.14 ± 0.05	0.12 ± 0.03
CCN51 Northeast *	0.04 ± 0.003	1.14 ± 0.32	0.55 ± 0.16	4.98 ± 0.45	0.21 ± 0.03	2.67 ± 0.65	0.13 ± 0.02	0.57 ± 0.08	1.93 ± 0.39	0.12 ± 0.05	0.09 ± 0.02
p-value	0.876391	0.270101	0.587155	0.0149992	0.724964	0.0865861	0.125498	0.00693916	0.547916	0.428591	0.0438873
FEC2 Urabá	0.05 ± 0.01	0.82 ± 0.18	0.63 ± 0.14	9.81 ± 1.89	0.23 ± 0.02	1.94 ± 0.57	0.51 ± 0.17	0.96 ± 0.22	1.69 ± 0.29	0.18 ± 0.05	0.15 ± 0.04
FEC2 Northeast	0.01 ± 0.001	0.70 ± 0.29	0.73 ± 0.26	6.38 ± 1.77	0.25 ± 0.06	1.46 ± 0.60	0.07 ± 0.01	0.54 ± 0.14	1.76 ± 0.53	0.05 ± 0.04	0.06 ± 0.01
p-value	0.000514	0.454388	0.500107	0.018275	0.547129	0.234608	0.003993	0.007305	0.808769	0.002409	0.000937
EET96 Urabá	0.04 ± 0.01	0.74 ± 0.28	0.87 ± 0.30	6.37 ± 1.64	0.26 ± 0.12	2.16 ± 0.70	0.31 ± 0.10	0.76 ± 0.22	1.20 ± 0.32	0.09 ± 0.04	0.09 ± 0.03
EET96 Northeast	0.03 ± 0.01	0.59 ± 0.25	0.52 ± 0.20	3.40 ± 1.32	0.19 ± 0.02	1.06 ± 0.56	0.14 ± 0.04	0.18 ± 0.07	1.26 ± 0.50	0.08 ± 0.03	0.06 ± 0.02
p-value	0.107232	0.414810	0.063597	0.013549	0.227634	0.024996	0.007349	0.003269	0.825886	0.563396	0.118044
FEAR5 Urabá	0.06 ± 0.01	0.82 ± 0.34	0.60 ± 0.25	12.80 ± 2.88	0.24 ± 0.04	1.82 ± 0.71	0.56 ± 0.04	1.51 ± 0.38	1.59 ± 0.57	0.17 ± 0.05	0.31 ± 0.07
FEAR5 Northeast	0.01 ± 0.003	0.19 ± 0.07	0.61 ± 0.27	8.51 ± 2.10	0.21 ± 0.02	0.35 ± 0.18	0.11 ± 0.01	0.23 ± 0.05	1.71 ± 0.51	0.08 ± 0.03	0.09 ± 0.03
p-value	0.000518	0.013171	0.933895	0.027482	0.157760	0.001955	0.000004	0.001510	0.724702	0.008796	0.000292
FSV41 Urabá	0.05 ± 0.01	0.86 ± 0.25	0.99 ± 0.53	8.37 ± 2.35	0.24 ± 0.03	1.42 ± 0.72	0.32 ± 0.17	0.81 ± 0.23	1.24 ± 0.58	0.12 ± 0.06	0.09 ± 0.03
FSV41 Northeast	0.03 ± 0.01	1.20 ± 0.23	1.31 ± 0.46	8.53 ± 1.78	0.25 ± 0.03	1.56 ± 0.58	0.24 ± 0.10	0.55 ± 0.13	1.30 ± 0.33	0.10 ± 0.06	0.11 ± 0.02
p-value	0.064064	0.051808	0.334845	0.908010	0.643905	0.748920	0.389674	0.059124	0.834028	0.761489	0.240169
FTA2 Urabá	0.04 ± 0.01	0.93 ± 0.29	0.38 ± 0.14	11.82 ± 3.70	0.23 ± 0.03	1.55 ± 0.62	0.48 ± 0.26	1.02 ± 0.37	1.48 ± 0.48	0.09 ± 0.04	0.13 ± 0.04
FTA2 Northeast	0.02 ± 0.01	0.77 ± 0.23	0.68 ± 0.20	5.57 ± 1.74	0.19 ± 0.02	1.16 ± 0.37	0.27 ± 0.05	0.53 ± 0.16	1.43 ± 0.48	0.11 ± 0.03	0.06 ± 0.02
p-value	0.071983	0.358164	0.022934	0.009078	0.051111	0.266827	0.141069	0.025986	0.867018	0.436877	0.017780
ICS95 Urabá	0.06 ± 0.03	1.59 ± 0.47	1.66 ± 0.45	11.98 ± 1.25	0.22 ± 0.03	3.02 ± 1.28	0.76 ± 0.52	0.72 ± 0.29	1.53 ± 0.30	0.15 ± 0.03	0.13 ± 0.02
ICS95 Northeast	0.03 ± 0.005	0.78 ± 0.14	0.60 ± 0.12	5.30 ± 0.33	0.18 ± 0.01	0.99 ± 0.14	0.12 ± 0.04	0.25 ± 0.05	1.19 ± 0.12	0.07 ± 0.02	0.06 ± 0.01
p-value	0.111570	0.013149	0.002583	0.000018	0.026499	0.023171	0.302321	0.019023	0.075907	0.002174	0.000652
ICS1 Urabá	0.05 ± 0.02	1.40 ± 0.53	1.38 ± 0.54	9.50 ± 2.67	0.20 ± 0.01	2.93 ± 1.16	0.31 ± 0.14	0.96 ± 0.25	1.72 ± 0.11	0.13 ± 0.03	0.14 ± 0.01
ICS1 Northeast	0.01 ± 0.040	0.39 ± 0.19	0.49 ± 0.16	6.83 ± 2.43	0.18 ± 0.02	0.68 ± 0.05	0.13 ± 0.04	0.41 ± 0.15	0.76 ± 0.41	0.12 ± 0.07	0.05 ± 0.02
p-value	0.022209	0.004986	0.042851	0.162375	0.145240	0.026518	0.024356	0.004792	0.002694	0.770959	0.000157
ICS60 Urabá	0.03 ± 0.01	1.23 ± 0.35	1.09 ± 0.31	7.90 ± 2.81	0.21 ± 0.02	2.20 ± 0.69	0.28 ± 0.05	0.95 ± 0.16	1.62 ± 0.26	0.13 ± 0.03	0.14 ± 0.01
ICS60 Northeast	0.02 ± 0.004	0.35 ± 0.09	0.42 ± 0.08	4.71 ± 0.50	0.17 ± 0.01	1.03 ± 0.19	0.11 ± 0.03	0.13 ± 0.02	0.58 ± 0.09	0.05 ± 0.01	0.03 ± 0.01
p-value	0.040805	0.012888	0.020293	0.106129	0.004226	0.004226	0.004226	0.001807	0.000066	0.020403	0.000000
TSH565 Urabá	0.05 ± 0.001	1.17 ± 0.07	0.74 ± 0.05	10.20 ± 0.68	0.29 ± 0.03	3.39 ± 0.52	0.30 ± 0.04	1.54 ± 0.08	1.67 ± 0.20	0.23 ± 0.02	0.15 ± 0.01
TSH565 Northeast	0.02 ± 0.00	0.56 ± 0.10	0.44 ± 0.06	5.37 ± 0.33	0.20 ± 0.02	1.30 ± 0.25	0.22 ± 0.05	0.41 ± 0.04	1.08 ± 0.18	0.08 ± 0.01	0.06 ± 0.01
p-value	3.6e−10	0.000002	0.000025	0.000001	0.000596	0.000041	0.039387	1.77e−9	0.001283	3.37e−8	0.000001

	L-Alanine (Ala)	Leucine (Leu)	Phenylalanine (Phe)	Stachyose (STA)	Sucrose (SUC)	Theobromine (THE)	Trigonelline (TRI)	Tyrosine (Tyr)	Valine (Val)	Xanthine (XAN)
Concentración (mM)										
CCN51 Urabá *	0.29 ± 0.12	0.15 ± 0.03	0.31 ± 0.03	46.97 ± 4.47	1.58 ± 0.21	109.15 ± 18.22	0.06 ± 0.02	1.21 ± 0.32	0.11 ± 0.02	0.03 ± 0.01
CCN51 Northeast *	0.07 ± 0.01	0.10 ± 0.02	0.26 ± 0.10	42.22 ± 5.81	1.78 ± 0.14	104.87 ± 31.76	0.05 ± 0.01	1.04 ± 0.10	0.08 ± 0.02	0.02 ± 0.01
p-value	0.016910	0.010831	0.297517	0.184778	0.129077	0.800550	0.282120	0.306497	0.011906	0.454331
FEC2 Urabá	0.53 ± 0.11	0.19 ± 0.05	0.27 ± 0.05	56.21 ± 11.91	1.37 ± 0.25	95.10 ± 22.46	0.06 ± 0.01	1.35 ± 0.30	0.13 ± 0.02	0.04 ± 0.01
FEC2 Northeast	0.15 ± 0.03	0.08 ± 0.04	0.22 ± 0.08	33.83 ± 9.47	1.69 ± 0.64	106.57 ± 37.91	0.05 ± 0.01	0.69 ± 0.15	0.05 ± 0.01	0.03 ± 0.03
p-value	0.001018	0.003359	0.269696	0.011050	0.332931	0.576583	0.785853	0.002308	0.000199	0.704414
EET96 Urabá	0.32 ± 0.08	0.14 ± 0.06	0.28 ± 0.09	35.00 ± 9.65	1.31 ± 0.32	115.68 ± 39.67	0.05 ± 0.01	1.01 ± 0.31	0.09 ± 0.02	0.05 ± 0.02
EET96 Northeast	0.06 ± 0.02	0.07 ± 0.02	0.12 ± 0.06	26.53 ± 9.02	1.55 ± 0.28	78.51 ± 32.40	0.03 ± 0.01	0.33 ± 0.10	0.06 ± 0.02	0.02 ± 0.02
p-value	0.001526	0.054825	0.009922	0.189290	0.242483	0.143320	0.057853	0.001538	0.034261	0.066788
FEAR5 Urabá	0.79 ± 0.17	0.28 ± 0.08	0.40 ± 0.10	57.55 ± 14.85	1.66 ± 0.43	88.84 ± 36.84	0.05 ± 0.02	1.97 ± 0.52	0.29 ± 0.07	0.03 ± 0.02

(*continued on next page*)

Table 1 (*continued*)

FEAR5 Northeast	0.12 ± 0.03	0.09 ± 0.03	0.11 ± 0.05	33.73 ± 8.45	1.73 ± 0.36	58.61 ± 27.24	0.03 ± 0.01	0.35 ± 0.09	0.09 ± 0.03	0.01 ± 0.01
p-value	0.000795	0.001367	0.000393	0.014293	0.771950	0.178429	0.087726	0.001925	0.000382	0.099418
FSV41 Urabá	0.46 ± 0.18	0.10 ± 0.02	0.31 ± 0.10	35.53 ± 13.45	1.20 ± 0.46	79.06 ± 33.99	0.06 ± 0.02	1.13 ± 0.27	0.08 ± 0.02	0.03 ± 0.02
FSV41 Northeast	0.32 ± 0.07	0.12 ± 0.02	0.24 ± 0.06	50.85 ± 10.46	1.27 ± 0.33	98.68 ± 30.51	0.07 ± 0.02	0.78 ± 0.20	0.10 ± 0.02	0.03 ± 0.02
p-value	0.150150	0.240561	0.206399	0.079132	0.779568	0.365084	0.755703	0.051615	0.112212	0.689328
FtA2 Urabá	0.41 ± 0.13	0.15 ± 0.04	0.29 ± 0.10	41.70 ± 11.24	1.52 ± 0.41	98.16 ± 32.98	0.07 ± 0.03	1.39 ± 0.49	0.11 ± 0.04	0.05 ± 0.02
FtA2 Northeast	0.22 ± 0.14	0.07 ± 0.02	0.23 ± 0.09	33.25 ± 6.35	1.42 ± 0.27	75.82 ± 23.84	0.04 ± 0.01	0.69 ± 0.09	0.06 ± 0.01	0.03 ± 0.01
p-value	0.050201	0.002945	0.332080	0.181690	0.646396	0.254600	0.085072	0.030933	0.017454	0.152222
ICS95 Urabá	0.39 ± 0.05	0.15 ± 0.03	0.41 ± 0.06	48.41 ± 9.29	1.58 ± 0.29	162.49 ± 44.85	0.08 ± 0.02	1.01 ± 0.39	0.12 ± 0.02	0.07 ± 0.03
ICS95 Northeast	0.06 ± 0.01	0.05 ± 0.01	0.13 ± 0.01	25.43 ± 5.19	1.30 ± 0.13	59.96 ± 11.63	0.04 ± 0.01	0.42 ± 0.03	0.05 ± 0.01	0.02 ± 0.01
p-value	0.000066	0.000161	0.000221	0.003187	0.124379	0.003160	0.015291	0.026416	0.000284	0.012484
ICS1 Urabá	0.42 ± 0.04	0.14 ± 0.02	0.37 ± 0.05	50.74 ± 5.44	1.70 ± 0.10	144.79 ± 58.68	0.08 ± 0.02	1.25 ± 0.32	0.12 ± 0.02	0.06 ± 0.03
ICS1 Northeast	0.12 ± 0.04	0.06 ± 0.02	0.14 ± 0.07	23.41 ± 9.85	0.99 ± 0.26	54.45 ± 24.42	0.04 ± 0.02	0.54 ± 0.19	0.05 ± 0.02	0.02 ± 0.01
p-value	0.000011	0.001278	0.001024	0.001677	0.001257	0.015929	0.009414	0.004236	0.000369	0.070078
ICS60 Urabá	0.39 ± 0.04	0.14 ± 0.04	0.34 ± 0.08	49.16 ± 9.54	1.63 ± 0.23	117.31 ± 39.08	0.07 ± 0.03	1.26 ± 0.20	0.12 ± 0.20	0.05 ± 0.01
ICS60 Northeast	0.04 ± 0.004	0.05 ± 0.01	0.11 ± 0.02	36.32 ± 4.62	0.54 ± 0.05	45.08 ± 9.63	0.03 ± 0.01	0.38 ± 0.26	0.03 ± 0.003	0.01 ± 0.003
p-value	0.000458	0.000458	0.008594	0.031769	0.001907	0.004883	0.021798	0.000836	0.001022	0.009766
TSH565 Urabá	0.42 ± 0.02	0.14 ± 0.03	0.47 ± 0.04	61.42 ± 7.16	1.82 ± 0.15	123.65 ± 8.48	0.07 ± 0.01	1.99 ± 0.09	0.13 ± 0.01	0.06 ± 0.01
TSH565 Northeast	0.19 ± 0.02	0.07 ± 0.01	0.16 ± 0.03	31.19 ± 3.77	1.20 ± 0.22	59.75 ± 6.57	0.03 ± 0.01	0.61 ± 0.09	0.05 ± 0.01	0.02 ± 0.005
p-value	4.36e−8	0.000664	0.000001	0.000032	0.000893	0.000001	0.000782	9.31e−9	0.000002	0.000179

Comparative concentration (mM) of the 21 metabolites identified in the aqueous extract of the beans of ten *T. cacao* genotypes grown at two different sites in Colombia: Urabá and Northeast. Results are the mean ± sd of five replicates measured by quantitative ^{1}H NMR spectroscopy. *Results are the mean ± sd of four replicates

content. There was no difference in the contents of G1P and SUC between the genotypes grown at this location, contrary to what was observed for STA; in particular, *TSH565* was among the groups containing the highest STA content. In Northeast, the *CCN51* genotype stands out for having the highest concentration of G1P, as does FSV*41* for its high STA content and *CCN51* and *FEAR5* for their high levels of SUC. On the other hand, ICS60 contained the lowest concentrations of GP1 and SUC, yet it also exhibited the greatest regional disparity, with GP1 and SUC concentrations. The concentrations of these metabolites are three times higher in the ICS60 clones from Urabá than in those cultivated in the Northeast region. The observed differences in sugar content have nutritional relevance because sugars represent a primary energy source. Variations in SUC and STA levels can affect the caloric and glycemic profiles of cacao-derived products, which is important for consumers concerned with energy balance and metabolic health (Darand et al., 2021).

3.3. Acetylcholine content

All *T. cacao* clones cultivated in the Urabá region exhibited higher ACh concentrations than their counterparts grown in the Northeast region, suggesting that Urabá is a location that would stimulate the greatest resistance to environmental stress and more favorable growth and development properties for *T. cacao* plants (Wessler et al., 2001).

Since this metabolite affects seed germination and early growth in several crops (e.g., bean and pea seedlings), ACh may participate in controlling the transport of reserve substances from cotyledons to rapidly developing parts of seedlings (Tretyn & Kendrick, 1991). It has also been demonstrated that ACh can stimulate plant growth (Tretyn & Kendrick, 1991). However, the mechanism of ACh activity depends on its interaction with different growth regulators. This difference seems to be related to the light environment and pH range under which the plant is grown (Tretyn & Kendrick, 1991).

Considering all these interrelationships, the results in Urabá for all different clones are worth mentioning. Individually, *TSH565* from Urabá (0.05 ± 0.001 mM) and *CCN51* from the Northeast (0.04 ± 0.003 mM) were the most notable, containing the highest amount of ACh (Table 1, Figure S8). In addition to its agronomic role, the presence of ACh may also offer neuroactive benefits, as it is a key neurotransmitter involved in cognitive function. The presence of this substance in cacao extracts confers neuroprotective properties and adds a functional food dimension to cacao-derived products (Gallo & Gámiz, 2023).

3.4. GABA

T. cacao is considered one of the best sources of GABA and its concentration is known to vary according to variety, growing conditions, and postharvest processes (de Araujo et al., 2021). In this study, concerning unfermented beans, significant differences (*p-value* < 0.05) were found in the concentration of GABA in all the *T. cacao* clones cultivated at both locations, except for FSV41, (Table 1). Again, Urabá stands out for producing cacao beans with higher amounts of GABA than those of Northeast: *TSH565* (1.54 ± 0.08), *FEAR5* (1.51 ± 0.38 mM), and *ICS60* (0.95 ± 0.16 mM). In turn, the clones with the lowest GABA values were *CCN51* (0.57 ± 0.08 mM), *FEC2* (0.54 ± 0.14 mM) and EET96 (0.18 ± 0.07 mM), all from the Northeast region (Figure S9).

The role of GABA in plants is not entirely clear. GABA may be involved in the regulation of plant growth and development, as well as in its ability to combat pathogens and environmental stress (Guo et al., 2023). Other studies have shown that GABA can protect cacao plants against damage caused by cold, drought, and fungal diseases (Kaspal et al., 2021). In addition, in another context, foods such as GABA-enriched dark chocolate could improve blood pressure- regulating effects. From an economic point of view, this type of enriched formulated product with high added value could diversify the functional food market (Koh et al., 2023). Hence, the results obtained here pinpoint

Urabá as a location that stimulates better GABA contents and growing conditions for *T. cacao* plants. Given GABA's role in reducing stress and its potential to lower blood pressure, the higher GABA levels in Urabá beans may increase the functional value of cacao, which aligns with current trends toward nutritionally fortified products (Koh et al., 2023).

3.5. Secondary metabolites

According to the statistical analysis (Table 1), CAF was the most variable metabolite quantified between clones grown in Urabá. In contrast, THE and EPI contents did not show significant differences between the clones at this site. The clones at the Urabá region with the highest contents of the secondary metabolites analyzed here were: *TSH565* (0.74 ± 0.05 mM of CAF, 3.39 ± 0.52 mM of EPI, 123.65 ± 8.48 mM of THE), *ICS95* (1.66 ± 0.45 mM CAF, 3.02 ± 1.28 mM EPI, 162.49 ± 44.85 mM THE) and *FSV41* (0.99 ± 0.53 mM CAF, 1.42 ± 0.72 mM EPI, 79.06 ± 33.99 mM THE) (Figure S10). On the other hand, in the Northeast region, the *T. cacao* clone with the highest EPI level was *CCN51* (2.67 ± 0.65 mM). Moreover, the Urabá region generally exhibited higher THE concentrations across most clones compared to the Northeast region. This difference was statistically significant (*p-value*: 0.0018), with Urabá having a significantly higher average THE concentration (mean concentration: 113.42 mM) than the Northeast region (mean concentration: 74.23 mM). Clones *ICS95* from Urabá had the highest THE concentration among all the clones and regions (162.49 mM), whereas the clone *ICS60* in Northeast had the lowest THE concentration (45.08 mM).

Studies on *T. cacao* have suggested that CAF and THE are overexpressed as a defense response mechanism. Aneja & Gianfagna reported that pathogen attack increased CAF levels in *T. cacao* stems by up to eightfold compared with levels in healthy stems (Aneja & Gianfagna, 2001). Additionally, it has been demonstrated that the contents of CAF and THE in cacao beans are associated with plant growth and better yields of fruit production (Senanayake & Wijesekera, 1971). Environmental factors such as water availability, season, temperature variation, and light have been reported to significantly affect the composition of methylxanthines in plants (Ahmed et al., 2019). These factors, in addition to their genetic counterparts, may further explain the variability reported here for these methylxanthines. Secondary metabolites such as CAF, THE, and EPI not only contribute to the characteristic bitter and stimulating taste of cacao but also provide antioxidant and potential cardioprotective benefits, enhancing the overall nutritional profile of cacao-based products (Goya et al., 2022).

3.6. Chemometric analysis

To better understand the chemical variations of the *T. cacao* bean samples, a principal component analysis (PCA) was performed with the twenty-one metabolites previously identified as the most variable importance projection by an exploratory PLS-DA (VIP value > 1). The results obtained do not show a clear separation between the two cultivation sites, the Northeast (N) and Urabá (U) regions, for all the clones studied (Fig. 2). However, the PCA biplot, revealed that geographical location impacts the content of certain metabolites in four clones (*FEAR5, TSH565, ICS95*, and *ICS1*) from the Urabá region, whereas in the case of the genotypes from the Northeast region, no characteristic variabilities among the measured metabolites were observed (Figure S11). The separation of the samples may be attributed to differences in the hydrophobic amino acid and secondary metabolite contents: *FEAR5*, Ile, Ala, Leu, and Val; for *TSH565*, Phe, STA, and ACh; and for clones *ICS95* and *ICS1*, their methylxanthines, CAD and EPI. Samples from the Northeast region predominantly occupied quadrants II and III, which correspond to lower metabolite concentrations.

Furthermore, PCA was performed on the samples from each geographical site, as shown in Figure S12 for Urabá, and Figure S13, for the Northeast regions. The results indicate that the metabolic variability among genotypes is more marked for the Urabá region than for Northeast region, which may be related to inherent differences associated with this cultivation site. However, in the PCA biplot obtained with data from the Northeast region (Figure S13), a clear separation between the two genotypes was observed: in the case of *CCN51,* the presence of ACh, Tyr, and CAD was the determinant, whereas for *FSV41*, the behavior was associated mainly with Val, Ile, Leu, CAF, DHT, and STA. The results suggested that most of the *T. cacao* beans from this geographical region shared similar metabolic profiles despite their genetic makeup.

To discern patterns of synchronized behavior, pairwise Spearman correlation analyses were performed on the quantified metabolites. The resulting correlation coefficients were visualized in a heatmap (Fig. 3).

Spearman correlation analysis of metabolites from cacao beans collected in the Urabá region revealed several significant associations, suggesting potential metabolic or biosynthetic linkages. Notably, Tyr showed a strong positive correlation with GABA (r = 0.94). Although these compounds arise from different metabolic pathways—Tyr is derived from Phe, while GABA, a nonproteinogenic amino acid, is synthesized from glutamate—this correlation may reflect coordinated regulation or shared physiological roles. De Araujo et al. reported that varying crop conditions, such as soil quality and regional factors, can influence the biochemical profile of cacao, particularly affecting free amino acids, including Leu, Tyr, Phe, and GABA (de Araujo et al., 2021).

Similarly, Ile and Val, both branched-chain amino acids (BCAAs) with similar structures and related biosynthetic pathways, were strongly correlated (r = 0.92). This shared metabolic origin likely explains the observed association in their concentrations within cacao beans (de Araujo et al., 2021).

Among secondary metabolites, CAD and THE were highly correlated (r = 0.86). Theobromine also correlated strongly with EPI (r = 0.85), and CAD correlated with EPI (r = 0.85). These relationships highlight the interconnected biosynthesis of polyphenols and methylxanthines, a phenomenon previously documented (Gallego et al., 2019).

Spearman correlation analysis of metabolites from cacao beans collected in the Northeast region revealed similar patterns to those observed in Urabá. Strong positive correlations were found between Tyr and GABA (r = 0.88), Ile and Val (r = 0.98), and between polyphenols and methylxanthines: CAD with EPI (r = 0.82), and THE with CAF (r = 0.81). The persistence of these correlations across regions suggests robust biochemical relationships among these metabolic pathways.

Additionally, a positive correlation was identified between the energy-related metabolites SUC and G1P (r = 0.87). Importantly, SUC concentration serves as a positive discriminating factor in the Cacao Quality Index (CQI), which was proposed by Araujo et al. to systematize the chemical composition of cacao beans with their sensory quality. This index helps identify key compounds that act as discriminants of cacao bean quality (Araujo et al., 2014).

Furthermore, T exhibited positive correlations with THE (r = 0.80), CAD (r = 0.83), and Phe (r = 0.83), suggesting its potential involvement in CAF biosynthesis. Trigonelline, a niacin derivative, and THE, a methylxanthine alkaloid, both play significant roles in cacao bean biochemistry (Genovese & Barros, 2019). Previous research has shown that the concentrations of T and THE are linked due to shared biosynthetic pathways and their roles in plant responses to environmental stress. For example, the production of methylxanthines such as THE involves methylation of xanthosine, a process influenced by metabolites including T and CAD (Gallego et al., 2019).

This study also revealed that metabolic expression varied among cacao clones even when cultivated at the same geographic site. Such information is crucial for crop renewal decisions and for selecting raw materials for industrial use. For instance, cacao beans from Urabá contained higher levels of hydrophobic amino acids and sugars than those from the Northeast region, suggesting these samples may undergo superior fermentation and potentially yield fine-flavored cacao products (Granvogl et al., 2006). Conversely, the lower metabolic variability observed in Northeast cacao beans likely reflects more homogeneous

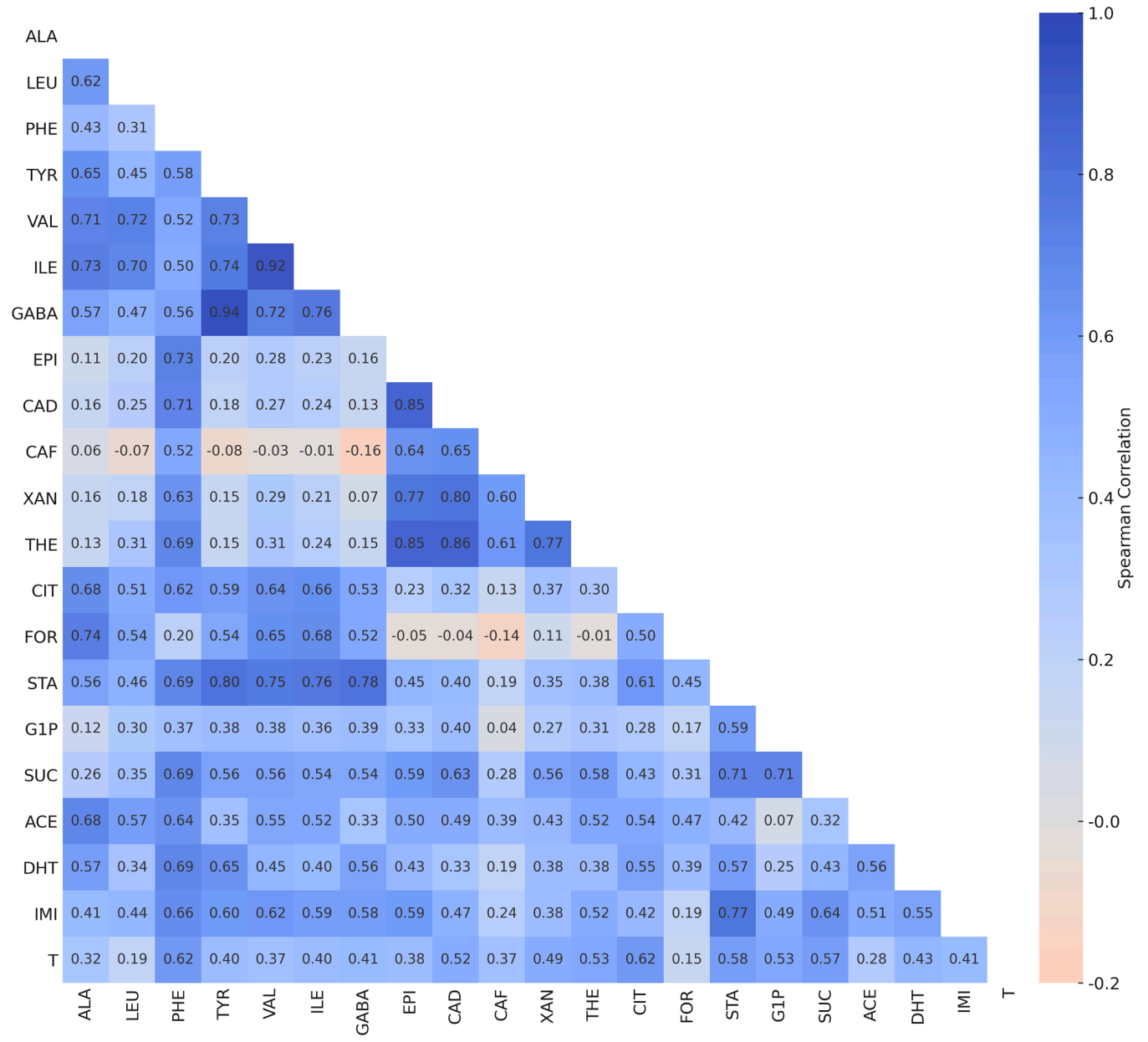

Fig. 3. Spearman correlation heatmaps of metabolite contents in extracts from A) Urabá region, and B) Northeast region. Deep blue is indicative of strong positive correlation, light colors represent no or very weak correlation, and dark red is associated with strong negative (inverse) correlation. No strong negative correlations were detected between metabolites in the regions studied.

agricultural practices or environmental conditions at that site.

Statistical analyses identified clones TSH565, FEAR5, and ICS1 grown in Urabá as producing cacao beans with the highest average levels of all measured metabolites. Within Urabá, FEAR5 exhibited relatively high levels of non-polar amino acids, while ICS1 and ICS95 had elevated concentrations of xanthine derivatives. In the Northeast region, FSV41 showed higher levels of non-polar amino acids, and CCN51 and FEC2 presented increased concentrations of non-polar aromatic amino acids and alkaloid derivatives.

Both universal clones (TSH565 and CCN51) and regional clones (FSV41 and FEAR5) demonstrated higher amounts of metabolites associated with plant growth compared to other genotypes evaluated.

From a nutritional standpoint, the enhanced profiles of amino acids, sugars, and bioactive compounds in cacao beans from the Urabá region indicate superior nutritional and functional qualities. These profiles may improve protein quality, energy balance, and provide additional health benefits, including antioxidant, anti-inflammatory, and neuroprotective effects when consumed.

Overall, the metabolomic profiling data generated by NMR in this study can inform improved cultivation practices and genotype selection. The lower metabolic variability among clones cultivated in Northeast Antioquia suggests more uniform growing conditions, whereas the greater diversity in Urabá highlights opportunities for targeted genotype propagation and large-scale cultivation tailored to environmental conditions.

4. Conclusion

This study applied state-of-the art quantitative ^{1}H NMR metabolic profiling to capture, for the first time, the metabolic differences among

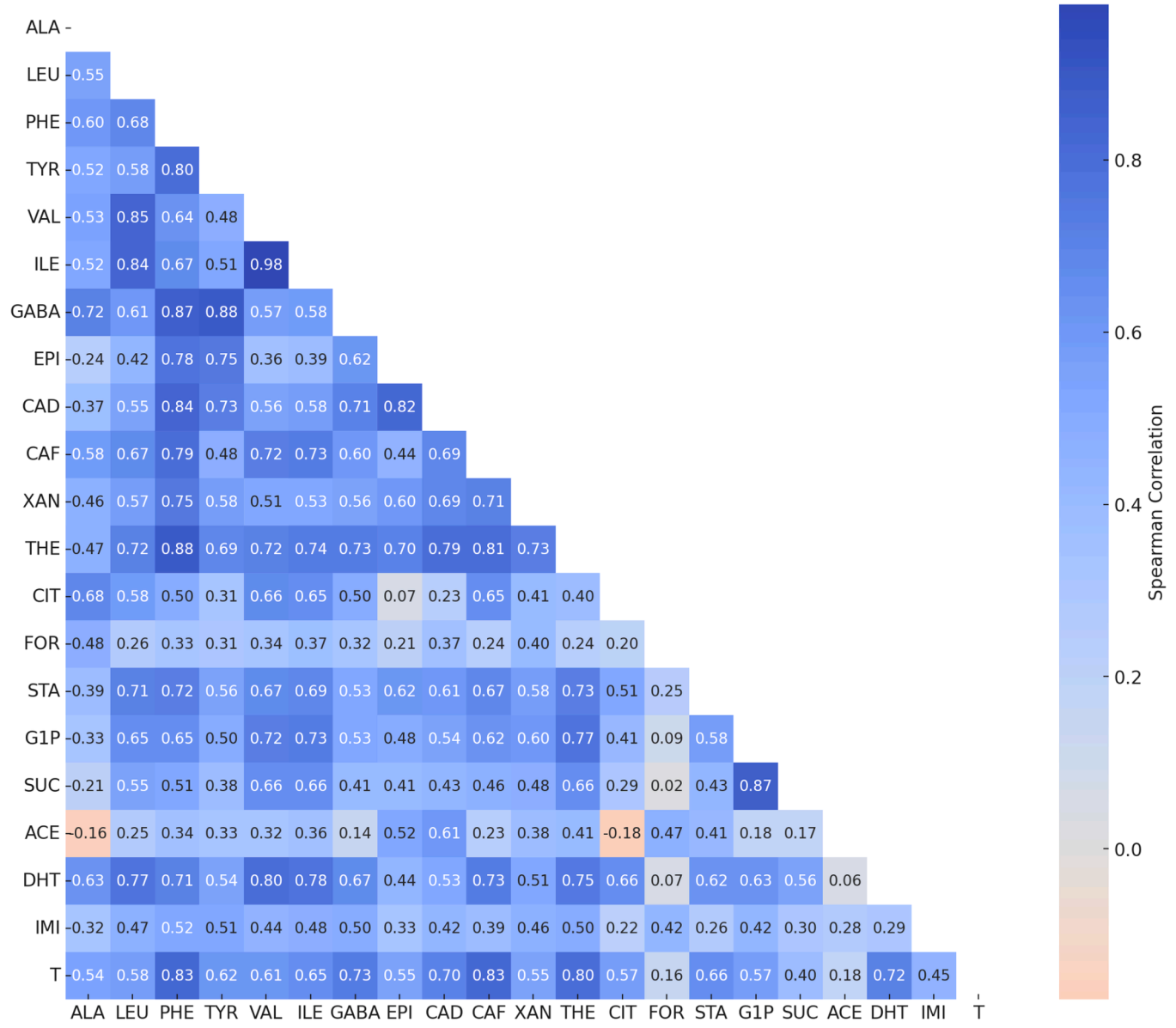

Fig. 3. (*continued*).

ten cacao genotypes grown in five different localities (n = 5 per genotype) across two distinct geographic regions of Antioquia, Colombia.

Our results highlight that cacao beans grown in the Urabá region generally exhibit higher concentrations of nutritionally and functionally important metabolites, such as non-polar amino acids, xanthine derivatives, and antioxidant compounds, compared to those from the Northeast region. Notably, genotypes TSH565, FEAR5, ICS1, and CCN51 demonstrated superior metabolic profiles associated with enhanced plant growth and potential health benefits.

In general, in the Northeast region, the performance of all genotypes showed no characteristic variabilities among the measured key metabolites, demonstrating the determinant role that plays the ecosystem on the metabolome of the cacao beans, over genotype.

NMR for metabolic profiling, as implemented in the current study, has proven once again to be a very powerful technique, as the results underscore the importance of clone selection tailored to specific environmental conditions to optimize the quality of *T. cacao* beans – a socio-economical important commodity for Colombia and other Latin American countries – in regions in which geographical and ecological diversity abound.

In particular, and in the immediate future, our results can guide breeding programs and cultivation strategies aimed at improving yield, nutritional value, and functional properties of cacao beans, in benefit of local producers and consumers.

Declarations

None

CRediT authorship contribution statement

Elkin Galeano: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Catalina Agudelo:** Writing – review & editing, Resources, Methodology. **Juan M. Lopez:** Writing – review & editing, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Helena Maruenda:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition. **Edison Osorio:** Writing – review & editing, Resources, Project

administration, Methodology, Funding acquisition.

Ethics approval and consent to participate

Not applicable.

Compliance with guidelines and applicable legislation

The study complied with relevant institutional, national, and international guidelines and legislation.

Declaration of Competing Interest

The authors certify that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article. In particular:

No financial relationships—such as employment, consultancies, honoraria, stock ownership, paid expert testimony, patents, or research grants—are in place or pending that might be perceived as a source of bias.

No non-financial relationships, including personal, professional, ideological, or academic rivalries—exist that could reasonably be construed as a potential conflict.

No third-party influence occurred: any funding bodies acknowledged in the manuscript had no role in the study design, data collection, analysis, interpretation, manuscript preparation, or the decision to submit for publication.

All authors have full control of the primary data, agree to allow the journal to review these data if requested, and have approved the final version of the manuscript.

Acknowledgments

Elkin Galeano gratefully acknowledges the support of the Alianza Pacífico Plataforma de Movilidad Estudiantil y Académica, the Programa Nacional de Becas y Crédito - Ministerio de Educación (Lima, Perú), and the Departamento de Ciencias at the Pontificia Universidad Católica del Perú for facilitating his stay at CERMN-PUCP, where he conducted the NMR sample preparation and analyses. This study was supported by the Dirección de Gestión de la Investigación-VRI-PUCP (CAP-2022-PI0880).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jfca.2025.108607.

Data availability

The data presented in this study are available in the manuscript and the Supplementary material. Raw data is available upon request.

References

Agudelo, C., Acevedo, S., Carrillo-Hormaza, L., Galeano, E., Osorio, E., 2022. Chemometric Classification of Colombian cacao crops: effects of different genotypes and origins in different years of harvest on levels of Flavonoid and Methylxanthine metabolites in raw Cacao beans. Molecules 27 (7), 2068. https://doi.org/10.3390/molecules27072068.

Ahmed, S., Griffin, T.S., Kraner, D., Schaffner, M.K., Sharma, D., Hazel, M., Leitch, A.R., Orians, C.M., Han, W., Stepp, J.R., Robbat, A., Matyas, C., Long, C., Xue, D., Houser, R.F., Cash, S.B., 2019. Environmental factors variably impact tea secondary metabolites in the context of climate change. Front. Plant Sci. 10, 939. https://doi.org/10.3389/fpls.2019.00939.

Aneja, M., Gianfagna, T., 2001. Induction and accumulation of caffeine in young, actively growing leaves of cocoa (Theobroma cacao L.) by wounding or infection with Crinipellis perniciosa. Physiol. Mol. Plant Pathol. 59 (1), 13–16. https://doi.org/10.1006/pmpp.2001.0337.

Araujo, Q.R., Fernandes, C.A.F., Ribeiro, D.O., Efraim, P., Steinmacher, D., Lieberei, R., Bastide, P., Araujo, T.G., 2014. Cocoa quality index – a proposal. Food Control 46, 49–54. https://doi.org/10.1016/j.foodcont.2014.05.003.

Bindereif, S.G., Brauer, F., Schubert, J.M., Schwarzinger, S., Gebauer, G., 2019. Complementary use of ^{1}H NMR and multi-element IRMS in association with chemometrics enables effective origin analysis of cocoa beans (Theobroma cacao L.). Food Chem. 299, 125105. https://doi.org/10.1016/j.foodchem.2019.125105.

Boutchouang, R.P., Fliniaux, O., Eyamo, J.V.E., Djabou, A.S.M., Fontaine, J.X., Molinié, R., Mesnard, F., Niemenak, N., 2024. Metabolome profiling of cacao (Theobroma cacao L.) callus under drought stress conditions induced by polyethylene glycol (PEG) as osmoticant. Phytochem. Anal. 35 (4), 708–722. https://doi.org/10.1002/pca.3323.

Caligiani, A., Acquotti, D., Cirlini, M., Palla, G., 2010. ^{1}H NMR study of fermented cocoa (Theobroma cacao L.) beans. J. Agric. Food Chem. 58 (23). https://doi.org/10.1021/jf102985w.

Caligiani, A., Palla, L., Acquotti, D., Marseglia, A., Palla, G., 2014. Application of ^{1}H NMR for the characterisation of cocoa beans of different geographical origins and fermentation levels. Food Chem. 157, 94–99. https://doi.org/10.1016/j.foodchem.2014.01.116.

Camandola, S., Plick, N., Mattson, M.P., 2019. Impact of coffee and cacao purine metabolites on neuroplasticity and neurodegenerative disease. Neurochem. Res. 44 (1), 214–227. https://doi.org/10.1007/S11064-018-2492-0.

Causil-Pastrana, E.J., Bernal, M., Delgado-Bejarano, L., Paternina, R.J.E., 2024. Physicochemical properties of soils in morphological characteristics of banana (Musa spp.) in four farms in Urabá-Antioquia. Temas Agrar. 29 (2), 200–210. https://doi.org/10.21897/R19EMB78.

Darand, M., Hajizadeh Oghaz, M., Hadi, A., Atefi, M., Amani, R., 2021. The effect of cocoa/dark chocolate consumption on lipid profile, glycemia, and blood pressure in diabetic patients: A meta-analysis of observational studies. Phytother. Res. 35 (10), 5487–5501. https://doi.org/10.1002/PTR.7183.

de Araujo, Q.R., de Abreu Loureiro, G.A.H., Póvoas, C.E.M., Steinmacher, D., de Almeida, S.S., Ahnert, D., Baligar, V.C., 2021. Effect of different edaphic crop conditions on the free amino acid profile of PH-16 Dry cacao beans. Agronomy 11 (8), 1637. https://doi.org/10.3390/agronomy11081637.

Díaz-Valderrama, J.R., Leiva-Espinoza, S.T., Aime, M.C., 2020. The history of cacao and Its diseases in the Americas. Phytopathology 110 (10), 1604–1619. https://doi.org/10.1094/PHYTO-05-20-0178-RVW.

D'Souza, R.N., Grimbs, S., Behrends, B., Bernaert, H., Ullrich, M.S., Kuhnert, N., 2017. Origin-based polyphenolic fingerprinting of Theobroma cacao in unfermented and fermented beans. Food Res. Int. 99. https://doi.org/10.1016/j.foodres.2017.06.007.

Fang, Y., Li, R., Chu, Z., Zhu, K., Gu, F., Zhang, Y., 2020. Chemical and flavor profile changes of cocoa beans (Theobroma cacao L.) during primary fermentation. Food Sci. Nutr. 8 (8), 4121. https://doi.org/10.1002/FSN3.1701.

Fedecacao – Fondo Nacional del Cacao. (2025, March 25). Reporte producción de cacao en Colombia, 2024. Boletin Fedecacao. ⟨https://www.fedecacao.com.co/post/santander-antioquia-y-arauca-reportaron-la-mayor-producci%C3%B3n-de-cacao-en-2024-sin-embargo-el-contr⟩.

Gallego, A.M., Rojas, L.F., Rodriguez, H.A., Mora, C., Atehortúa, L., Urrea, A.I., Guiltinan, M.J., Maximova, S.N., Gaquerel, E., Zuluaga, M., Pabón-Mora, N., 2019. Metabolomic profile of cacao cell suspensions growing in blue light/dark conditions with potential in food biotechnology. Plant Cell Tissue Organ Culture (PCTOC 139 (2), 275–294. https://doi.org/10.1007/s11240-019-01679-3.

Gallo, M., Gámiz, F., 2023. Choline: an essential nutrient for human health. Page 2900, 15 Nutrients 2023 15 (13), 2900. https://doi.org/10.3390/NU15132900.

Genovese, M.I., Barros, H.R. de M., 2019. Theobroma cacao and Theobroma grandiflorum: Bioactive Compounds and Associated Health Benefits. In: Mérillon, J. M., Ramawat, K.G. (Eds.), Reference Series in Phytochemistry. Springer, Cham, pp. 1049–1070. https://doi.org/10.1007/978-3-319-78030-6_15.

Gowda, N., Djukovic, D., 2014. Overview of mass spectrometry-based metabolomics: opportunities and challenges. Methods Mol. Biol. (Clifton N. J.) 1198, 3. https://doi.org/10.1007/978-1-4939-1258-2_1.

Goya, L., Kongor, J.E., de Pascual-Teresa, S., 2022. From cocoa to chocolate: effect of processing on flavanols and methylxanthines and their mechanisms of action. Int. J. Mol. Sci. 23 (22), 14365. https://doi.org/10.3390/ijms232214365.

Granvogl, M., Bugan, S., Schieberle, P., 2006. Formation of amines and aldehydes from parent amino acids during thermal processing of cocoa and model systems: new insights into pathways of the Strecker reaction. J. Agric. Food Chem. 54 (5), 1730–1739. https://doi.org/10.1021/jf0525939.

Guo, Z., Gong, J., Luo, S., Zuo, Y., Shen, Y., 2023. Role of Gamma-Aminobutyric Acid in Plant Defense Response. Metabolites 13 (6), 741. https://doi.org/10.3390/metabo13060741.

Henao Ramírez, A.M., de la Hoz Vasquez, T., Ospina Osorio, T.M., Garcés, L.A., Urrea Trujillo, A.I., 2018. Evaluation of the potential of regeneration of different Colombian and commercial genotypes of cocoa (Theobroma cacao L.) via somatic embryogenesis. Sci. Hortic. 229, 148–156. https://doi.org/10.1016/j.scienta.2017.10.040.

Hernández-Pérez, T., Paredes-López, O., 2024. Selected Mesoamerican Crops – Anti-Obesity Potential and Health Promotion. A Review. Plant Foods Hum. Nutr. 79 (3), 563–570. https://doi.org/10.1007/S11130-024-01211-9/FIGURES/4.

Kaspal, M., Kanapaddalagamage, M.H., Ramesh, S.A., 2021. Emerging Roles of γ Aminobutyric Acid (GABA) Gated Channels in Plant Stress Tolerance. Plants 10 (10), 2178. https://doi.org/10.3390/plants10102178.

Katz, D.L., Doughty, K., Ali, A., 2011. Cocoa and chocolate in human health and disease. Antioxid. Redox Signal. 15 (10), 2779–2811. https://doi.org/10.1089/ARS.2010.3697.

Koh, W.Y., Lim, X.X., Teoh, E.S.W., Kobun, R., Rasti, B., 2023. The Effects of Gamma-Aminobuytric Acid (GABA) Enrichment on Nutritional, Physical, Shelf-Life, and Sensorial Properties of Dark Chocolate. Foods 12 (1), 213. https://doi.org/10.3390/foods12010213.

León-Moreno, C.E., Rojas-Molina, J., Castilla-Campos, C.E., León-Moreno, C.E., Rojas-Molina, J., Castilla-Campos, C.E., 2019. Physicochemical characteristics of cacao (Theobroma cacao L.) soils in Colombia: Are they adequate to improve productivity? Agron. ía Colomb. 37 (1), 28–38. https://doi.org/10.15446/AGRON.COLOMB.V37N1.70545.

Lopez, J.M., Cabrera, R., Maruenda, H., 2019. Ultra-Clean Pure Shift ^{1}H NMR applied to metabolomics profiling. Sci. Rep. 9 (1), 1–8. https://doi.org/10.1038/s41598-019-43374-5.

Manciu, A., Rammig, A., Krause, A., Quesada, B.R., 2023. Impacts of land cover changes and global warming on climate in Colombia during ENSO events. Clim. Dyn. 61 (1–2), 111–129. https://doi.org/10.1007/S00382-022-06545-1/FIGURES/6.

Marseglia, A., Acquotti, D., Consonni, R., Cagliani, L.R., Palla, G., Caligiani, A., 2016. HR MAS ^{1}H NMR and chemometrics as useful tool to assess the geographical origin of cocoa beans – Comparison with HR ^{1}H NMR. Food Res. Int. 85, 273–281. https://doi.org/10.1016/J.FOODRES.2016.05.001.

Martín, M.A., Ramos, S., 2016. Cocoa polyphenols in oxidative stress: Potential health implications. J. Funct. Foods 27, 570–588. https://doi.org/10.1016/J.JFF.2016.10.008.

Mira-Giraldo, J.M., Cerón-Muñoz, M.F., Medina-Sierra, M., Mira-Giraldo, J.M., Cerón-Muñoz, M.F., Medina-Sierra, M., 2025. Textura del suelo en sistemas productivos de cacao y ganadería en el departamento de Antioquia, Colombia. Rev. De. Invest. Vet. Del. Perú 36 (1), 26349. https://doi.org/10.15381/RIVEP.V36I1.26349.

Motamayor, J.C., Risterucci, A.M., Lopez, P.A., Ortiz, C.F., Moreno, A., Lanaud, C., 2002. Cacao domestication I: The origin of the cacao cultivated by the Mayas. Heredity 89 (5), 380–386. https://doi.org/10.1038/sj.hdy.6800156.

Moutzouri, P., Chen, Y., Foroozandeh, M., Kiraly, P., Phillips, A.R., Coombes, S.R., Nilsson, M., Morris, G.A., 2017. Ultraclean pure shift NMR. Chem. Commun. 53 (73), 10188–10191. https://doi.org/10.1039/C7CC04423B.

Paparella, A., Schirone, M., López, C.C., 2025. The Health Impact of Cocoa from Cultivation to the Formation of Biogenic Amines: An Updated Review. *Page 255*, *14* Foods 2025 14 (2), 255. https://doi.org/10.3390/FOODS14020255.

Putri, D.N., De Steur, H., Juvinal, J.G., Gellynck, X., Schouteten, J.J., 2024. Sensory attributes of fine flavor cocoa beans and chocolate: A systematic literature review. J. Food Sci. 89 (4), 1917–1943. https://doi.org/10.1111/1750-3841.17006.

Rangel-Fajardo, M.A., Delgado-Alvarado, A., Córdova-Tellez, L., Zavaleta-Mancera, H.A., Villegas-Monter, A., López-Andrade, A.P., Vidales-Fernandez, I., 2011. Accumulation of sugars in cacao (Theobroma cacao L.) seeds of three genetic origins and its relationship to desiccation tolerance. Seed Sci. Technol. 39 (1), 1–11. https://doi.org/10.15258/sst.2011.39.1.01.

Roessner, U., Luedemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L., Fernie, A.R., 2001. Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. Plant Cell 13 (1), 11–29. https://doi.org/10.1105/TPC.13.1.11.

Rohsius, C., Matissek, R., Lieberei, R., 2006. Free amino acid amounts in raw cocoas from different origins. Eur. Food Res. Technol. 222 (3–4), 432–438. https://doi.org/10.1007/s00217-005-0130-y.

Sawada, Y., Hirai, Y., 2013. Integrated LC-MS/MS system for plant metabolomics. Comput. Struct. Biotechnol. J. 4 (5), e201301011. https://doi.org/10.5936/csbj.201301011.

Senanayake, U.M., Wijesekera, R.O.B., 1971. Theobromine and caffeine content of the cocoa bean during its growth. J. Sci. Food Agric. 22 (5), 262–263. https://doi.org/10.1002/jsfa.2740220512.

Simmler, C., Napolitano, J.G., McAlpine, J.B., Chen, S.-N., Pauli, G.F., 2014. Universal quantitative NMR analysis of complex natural samples. Curr. Opin. Biotechnol. 25, 51–59. https://doi.org/10.1016/j.copbio.2013.08.004.

Tretyn, A., Kendrick, R.E., 1991. Acetylcholine in plants: Presence, metabolism and mechanism of action. Bot. Rev. 57 (1), 33–73. https://doi.org/10.1007/BF02858764.

Truzzi, E., Marchetti, L., Fratagnoli, A., Rossi, M.C., Bertelli, D., 2023. Novel application of ^{1}H NMR spectroscopy coupled with chemometrics for the authentication of dark chocolate. Food Chem. 404. https://doi.org/10.1016/j.foodchem.2022.134522.

Truzzi, E., Rossi, M.C., Bertelli, D., 2024. ^{1}H NMR targeted and untargeted approaches for the authentication of milk chocolates. Food Control 160, 110367. https://doi.org/10.1016/j.foodcont.2024.110367.

Villalón-López, D.N., Hernández-Ortega, M., Ortiz-Moreno, A., Ceballos, G., Zepeda-Vallejo, G., Valadez-Carmona, L., 2023. ^{1}H NMR-based metabolomics of cacao pod husk: profile comparison of three drying treatments. Explor. Foods Food 1 (3). https://doi.org/10.37349/eff.2023.00009.

Wessler, I., Kilbinger, H., Bittinger, F., Kirkpatrick, C.J., 2001. The Non-neuronal Cholinergic System. Jpn. J. Pharmacol. 85 (1), 2–10. https://doi.org/10.1254/jjp.85.2.

Zimmermann, B.F., Ellinger, S., 2020. Cocoa, chocolate, and human health. Page 698, 12 Nutrients 2020 12 (3), 698. https://doi.org/10.3390/NU12030698.

Glossary

^{1}H NMR: Proton nuclear magnetic resonance spectroscopy
ACh: Acetylcholine
Ala: Alanine
BCAAs: Branched-chain amino acids
CAD: Caffeic acid
CAF: Caffeine
CIT: Citrate
DHT: Dihydrothymine
EPI: Epicatechin
FEDECACAO: Federación Nacional de Cacaoteros
FOR: Formate
G1P: Glucose-1-phosphate
GABA: γ-Aminobutyric acid
IMI: Imidazole
Ile: Isoleucine
Leu: Leucine
masl: Meters above sea level
PCA: Principal component analysis
Phe: Phenylalanine
PLS-DA: Partial least squares discriminant analysis
qNMR: Quantitative nuclear magnetic resonance
STA: Stachyose
SUC: Sucrose
T: Trigonelline
THE: Theobromine
Tyr: Tyrosine
Val: Valine
XAN: Xanthine