Full Paper

Journal of the Brazilian Chemical Society

https://dx.doi.org/10.21577/0103-5053.20260068

J. Braz. Chem. Soc. **2026**, *37-Spe1*, e-20260068, 1-14
©2026 Sociedade Brasileira de Química

Volatile and Non-Volatile Metabolites from "Cacau-jacaré" (*Theobroma mariae*), an Amazonian Unconventional Food Plant

Silvio O. Assunção Jr., [a] ***Eldrinei G. Peres,*** [a,b] ***Fernanda A. S. Rocha,*** [a,c] ***Emilly J. S. P. Lima,*** [a] ***Moysés B. Araújo Jr.,*** [a,d] ***Felipe M. A. Silva,*** [e] ***Hector H. F. Koolen*** [*,a] ***and Giovana A. Bataglion*** [*,b]

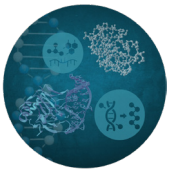
[a]*Grupo de Pesquisas em Metabolômica e Espectrometria de Massas, Escola Superior de Ciências da Saúde, Universidade do Estado do Amazonas, 690065-130 Manaus-AM, Brazil*

[b]*Departamento de Química, Universidade Federal do Amazonas, 69067-005 Manaus-AM, Brazil*

[c]*Faculdade de Engenharia de Alimentos, Universidade Estadual de Campinas, 13083-862 Campinas-SP, Brazil*

[d]*Instituto de Ciências Exatas e Tecnologia, Universidade Federal do Amazonas, 69103-128 Itacoatiara-AM, Brazil*

[e]*Coordenação de Tecnologia e Inovação, Instituto Nacional de Pesquisas da Amazônia, 69080-971 Manaus-AM, Brazil*



The Amazon hosts countless plant species, many still unknown to science. Among catalogued species, hundreds are edible, yet most are rarely consumed or locally restricted. Known in Brazil as PANCs ("*Plantas Alimentícias Não Convencionais*"), these unconventional food plants, these species can enhance food security, especially for vulnerable populations. Beyond nutrition, studies reveal their nutraceutical potential due to specialized metabolites such as polyphenols, terpenes, and alkaloids, which enable applications in supplements, cosmetics, and herbal formulations. In this context, *Theobroma mariae* (Mart.) K. Schum., or "cacau-jacaré", is a little-known fruit-bearing plant with promising applicability. This study annotated its volatile and non-volatile compounds in pulp and seeds using headspace solid-phase microextraction gas chromatography coupled with mass spectrometry (HS-SPME/GC-MS) and ultra-high-performance liquid chromatography-high resolution tandem mass spectrometry (UHPLC-HRMS/MS), respectively. A total of 72 volatile and 32 non-volatile compounds were annotated. Fatty acids predominated in the pulp volatilome, whereas theacrine was the main volatile compound in the seeds. Phenolic compounds dominated the non-volatile fraction, notably flavonoid-*C*-glycosides, *O*-glycosides, and *O*-glucuronides. The findings highlight *T. mariae* as a valuable yet underutilized Amazonian fruit, whose chemical profile resembles that of common cacao and offers potential for novel applications in food, nutraceutical, and cosmetic formulations.

Keywords: cocoa, flavonoid glucoronides, theacrine, unconventional food plants, volatilome

Introduction

The Amazon is a biome of extraordinary biodiversity and immense economic potential, particularly in the food sector, as it harbors more than 15,000 plant species.[1] Many of these are fruit-bearing and edible, such as buriti (*Mauritia flexuosa*), camu-camu (*Myrciaria dubia*), araçá-boi (*Eugenia stipitata*), and açaí (*Euterpe oleracea*), which are locally produced and consumed. Among them, açaí has gained worldwide recognition as a superfruit due to its high levels of anthocyanins and proanthocyanidins, with antioxidant properties, as well as its notable nutritional value.[2] Conversely, many other Amazonian fruits are consumed only by local communities and remain commercially underexplored and largely unknown to the general public. Such plants are referred to as unconventional food plants (PANCs),[3] as they are part of the diet of specific Amazonian groups but are unfamiliar to most of the population. Studies on the chemical composition of PANCs help promote these

*e-mail: giovanabataglion@ufam.edu.br; hkoolen@uea.edu.br
Editor handled this article: Paulo Wender P. Gomes (Guest)
This publication is part of the special issue "Omics Sciences"

This is an open-access article distributed under the terms of the Creative Commons Attribution License.

species, enabling broader use and paving the way for future commercialization – an important step toward generating income for Amazonian communities.

Some PANCs belong to the genus *Theobroma*, which comprises about 40 species native to tropical rainforests of South and Central America.[4] Species of this genus are valued for their fruits, which have culinary and medicinal uses and form part of the Amazonian cultural heritage. Among them, cacao (*Theobroma cacao*) stands out as the source of chocolate.[4] Other species, such as cupuaçu (*Theobroma grandiflorum*) and macambo (*Theobroma bicolor*), are less known globally but are commercially used in the region: their pulps are used to make ice creams, sweets, jams, and juices, while their seeds yield chocolate-like products known as "cupulate" and "bacalate", respectively.[5,6] Although these are the best-known and most cultivated species of the genus, several so-called wild cacao varieties remain underexplored despite their nutritional and commercial potential.[7]

One of the wild cacao species is *Theobroma mariae* (commonly known as "cacau-jacaré") (Figure 1), which occurs in Brazil across the states of Acre, Amazonas, Pará, Rondônia, and Roraima.[8] The species was initially included in the genus *Theobroma* by Schumann,[9] but due to morphological distinctions, it was later reclassified into the genus *Herrania* by Schultes,[10] being renamed *Herrania mariae*. However, recent phylogenetic and molecular genetic analyses confirmed that the species indeed belongs to the genus *Theobroma*, thus reinstating its scientific name *T. mariae*.[4] This species grows as a shrub or a slender, unbranched small tree, commonly found in the understory of pristine Amazonian forests. The oblong and ovoid fruits of *Theobroma mariae* (Figures 1c and 1d) can reach up to 12 cm in length and exhibit well-defined longitudinal ribs. Internally, they contain a sweet white pulp appreciated for its mild flavor.[11]

Despite its close relationship with *T. cacao*, *T. mariae*'s fruits remain underutilized. The pulp is used in preparations such as jams, and flour made from its seeds is used in cakes, leading to the species being regarded as a survival fruit.[3] Analyses of the lipid composition of its seeds revealed high levels of the saturated fatty acid arachidic acid, as well as significant amounts of unsaturated fatty acids, particularly linoleic acid.[12,13] Furthermore, previous studies have reported the presence of alkaloids – especially methylxanthines such as theobromine, theacrine, and caffeine, in concentrations similar to those found in *T. cacao*.[14,15] More recent research has shown that the edible flowers of *T. mariae* contain phenolic compounds such as hyperoside, guaijaverin, astragalin, juglanin, and kaempferol as major constituents.[16]

Despite being an edible fruit, the chemical knowledge regarding this species remains quite limited, particularly with respect to its pulp, especially when considering the potential demonstrated by other recently studied species, such as *Theobroma sylvestre*.[17] Given the consumption of *T. mariae* fruits by local Amazonian communities and the lack of knowledge regarding their chemical composition, this study aimed to investigate the volatile and non-volatile metabolite profiles present in the pulp and seeds of this species. The objective was to elucidate its chemical composition and encourage future studies exploring possible correlations between the identified compounds and their nutraceutical potential.

Experimental

Reagents and standards

The reagents used were as follows: high-performance liquid chromatography (HPLC)-grade methanol (Qhemis, Brazil); liquid chromatography mass spectrometry (LC-MS)-grade

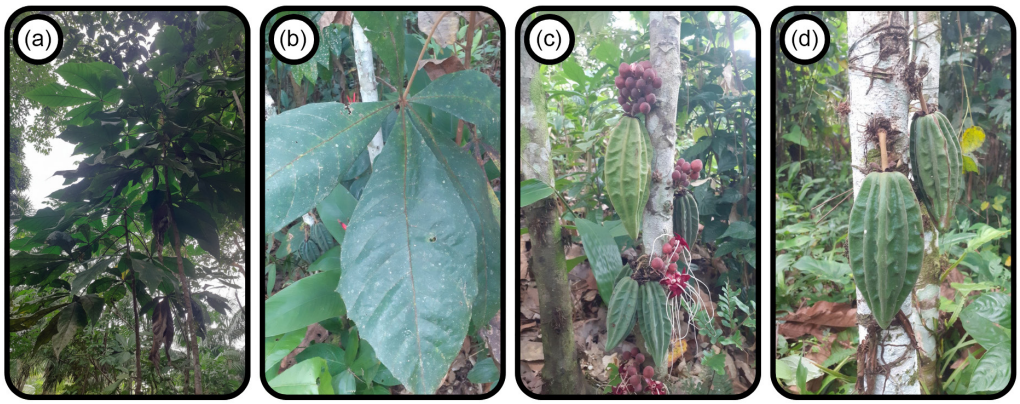


Figure 1. Morphological characteristics of *Theobroma mariae*. (a) Tree showing its slender, unbranched structure. (b) Leaves with a characteristic palmatisect arrangement of leaflets. (c) Inflorescences arranged in fascicles directly on the stem, exhibiting reddish to purplish coloration and developing fruits. (d) Ellipsoid fruits displaying prominent external wing-like projections.

methanol (CHROMASOLV™, Germany); LC-MS-grade water (Merck KGaA, Germany); LC-MS-grade formic acid (Merck KGaA, Germany); sodium chloride (Dinâmica, Brazil); ammonium formate (Tedia, USA); *n*-C_7-*n*-C_{40} saturated alkane standard mixture at 1000 μg mL^{-1} (Supelco, USA); and deionized water from a Purelab® Quest system (ELGA LabWater, United Kingdom).

Specimen collection site

Mature medium- and large-sized fruits (≥ 350 g) of *T. mariae* were collected from three different trees, with more than one fruit sampled *per* tree, in Vila Nogueira, located in the municipality of Alvarães, Amazonas state, Brazil (3°17'57.5" S, 64°47'01" W), in May 2022 during the dry season. The botanical material was identified by Deisy Saraiva, and a voucher specimen (No. 12,690) was deposited at the Herbarium of the Federal University of Amazonas (UFAM), Brazil. Access registration in the National System for the Management of Genetic Heritage and Associated Traditional Knowledge (SisGen) was obtained under No. A24C508.

HS-SPME/GC-MS analysis of volatile organic compounds

The composition of volatile organic compounds (VOCs) was investigated by gas chromatography coupled to mass spectrometry (GC-MS) using headspace solid-phase microextraction (HS-SPME). A manual procedure was used to separate the fruit components from three distinct trees individually. Each fruit was longitudinally opened to expose the seeds surrounded by pulp, and the peel was removed. The mucilaginous pulp adhering to each seed was carefully removed using a sterilized pair of scissors. Subsequently, the seeds were washed with LC-MS-grade water, dried with paper towels, and ground while still fresh, whereas the separated pulp was used directly. Samples of seed and pulp (2 g each, *per* replicate) were weighed into 10 mL SPME vials, followed by the addition of 2 mL of a NaCl solution (100 mg mL^{-1}). The samples were pre-incubated statically for 30 min at 40 °C before extraction. VOCs were extracted using a mixed divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS) fiber (Sigma-Aldrich, USA) for 60 min at 40 °C. After extraction, the fiber was desorbed in the injector, operated in splitless mode, for 5 min at 270 °C.

Data were acquired using a GC-MS QP2010 Plus system (Shimadzu, Japan). Chromatographic separation was carried out on an HP-5MS capillary column (Hewlett-Packard, USA; 30 m × 0.25 mm internal diameter (i.d.), 0.25 μm film thickness) with helium (99.999% purity) as the carrier gas at a flow rate of 1.0 mL min^{-1}. The oven temperature program was as follows: initial temperature of 40 °C (held for 1 min), increased at 3 °C min^{-1} to 200 °C (held for 1 min), then at 5 °C min^{-1} to 300 °C (held for 10 min). The MS parameters were set as follows: ion source temperature of 200 °C, interface temperature of 280 °C, electron ionization at 70 eV, and mass scan range from *m/z* 50 to 550.

All analyses were performed in triplicate and the relative proportion of each compound was estimated by dividing its average peak area by the total chromatographic area and expressing the result as a relative percentage. Compounds were identified based on spectral minimum similarity threshold of 80 with the NIST 2020 library installed at the GCMS solutions software (v. 4.53, Shimadzu, 2021), elution order and calculated retention indices (RI). A mixture of linear alkanes (*n*-C_7-*n*-C_{40}) was analyzed together with the samples to calculate RI values according to the Dool and Kratz equation.[18]

Extraction of non-volatile compounds

The crushed pulp and seeds (100 g *per* replicate) were separately processed in a glass blender with methanol (500 mL) and left to stand for 24 h for extraction. After this period, each extract was diluted to 5 L of water and subjected to ultrasonic treatment for 20 min at 30 °C. The resulting solutions were then centrifuged and vacuum-filtered through 2.7 μm glass microfiber filters (Whatman, UK). Finally, the filtrates were subjected to solid-phase extraction (SPE) using 1 g polymeric reversed-phase Strata-X cartridges (Phenomenex, USA), previously activated with methanol (6 mL) and conditioned with water (6 mL). Sample cleanup was performed with 5 mL of water/methanol (97:3, v/v), followed by elution of the target compounds with 5 mL of methanol. The eluates were then dried under a gentle stream of nitrogen and stored at −20 °C until analysis.

UHPLC-HRMS/MS analysis

The dried extracts were dissolved in methanol (0.9 mg mL^{-1}), centrifuged, and filtered through a 0.22 μm nylon membrane (Allcrom, Brazil). Ultra-high-performance liquid chromatography-high resolution tandem mass spectrometry (UHPLC-HRMS/MS) analyses were performed using a Vanquish Flex UHPLC system (Thermo Scientific, Germany) coupled to an Orbitrap Exploris 240 mass spectrometer (Thermo Scientific, Germany) equipped with a heated electrospray ionization (HESI) source operating in both positive and negative ionization

modes. The mass analyzer operated over an *m/z* range of 100-1000 with a resolution of 120,000 at *m/z* 200 for full-scan acquisition and 15,000 for MS/MS.

Chromatographic separations were carried out on a Hypersil GOLD column (50 × 3 mm, 1.9 µm; Thermo Scientific, Germany) at 40 °C. The mobile phases consisted of 0.1 formic acid in water (A) and methanol (B), both containing 0.1% formic acid for positive ionization mode and 0.1 mM ammonium formate for negative ionization mode, at a flow rate of 0.4 mL min^{-1}. The elution gradient was as follows: 10% B; 0.5-6 min, 10-90% B; 6-8 min, 100% B; 8-11 min, 100-10% B. The injection volume was 2 µL.

The ion source parameters were as follows: ion voltage ± 3.0 kV; sheath gas 35 arbitrary units; auxiliary gas 7 arbitrary units; sweep gas 2 arbitrary units; ion transfer tube temperature 275 °C; vaporizer temperature 320 °C; and RF lens 70. Data-dependent acquisition (DDA) mode was used, targeting the five most intense precursor ions, with an isolation window of 1.5 *m/z*. Normalized collision energies (NCE) of 30, 50, and 70 were applied. For MS/MS data acquisition, the following filters were applied: minimum intensity threshold of 5,000, and dynamic exclusion after one acquisition (10 s duration, 5 ppm mass tolerance, with isotope exclusion).

Compound annotation was manually performed by curation of high-resolution product-ion spectra trought the software Freestyle (v. 1.8, Thermo Scientific, 2021) and by comparison with literature and reference data.

Results and Discussion

Volatilome annotation

Using the headspace solid-phase microextraction gas chromatography coupled with mass spectrometry (HS-SPME/GC-MS) technique, the aroma composition of the pulp and seeds of *T. mariae* ("cacau-jacaré") was characterized. In total, 72 different VOCs were annotated (Table 1), of which 42 were present in the pulp, 38 in the seeds and 22 in both of them. These annotations were classified at level 2, as they relied on spectral similarity with the NIST library and retention index comparisons.

In the pulp, the main classes detected were fatty acids, alcohols, aldehydes, and esters, while alkaloids, aldehydes and terpenes were the dominant classes in the seeds.

Fatty acids (68.79%) were predominant in the pulp, with octadecanoic acid (23.56%) (**59**) and arachidic acid (17.85%) (**65**) being the most abundant VOCs. In the seeds, the same compound were observed (Figure 2), though in lower proportions (16.18% of **65** and 8.70% of **59**). The predominance of long-chain fatty acids (C_{16}-C_{20}) (Figure 2b) differs from the fatty acid profile reported for *T. grandiflorum* pulp, where the major component is hexadecanoic acid,[19] along with short- and medium-chain fatty acids such as butanoic acid and octanoic acid,[20] both of which were not detected in *T. mariae* (Table 1, Figure 2b). In contrast, *T. bicolor* pulps exhibit a wider range of fatty acids, spanning medium- to long-chain compounds, with decanoic and hexadecanoic acids as the main constituents,[21] highlighting the unique composition of *T. mariae* among *Theobroma* species.

Furthermore, aldehydes were the second most abundant class in *T. mariae* pulp (15.32%) and seeds (20.94%), with the aliphatic aldehydes (2*E*)-octenal (2.98%) (**5**) and dodecanal (1.71%) (**29**) being the most representative in the pulp and the aromatic derivatives benzeneacetaldehyde (**4**, 8.53%) and benzaldehyde (**2**, 4.47%) in the seeds (Table 1). The aldehyde profile of *T. mariae* pulp was quite diverse, with alkyl chains ranging from C_7 to C_{19} (with dominance of medium-chain aldehydes, C_8-C_{11}), much broader than those reported for *T. cacao* (restricted to C_8-C_{10})[22] and *T. bicolor* (only compound **7**)[21] (Figure 2a). These aldehydes are associated with citrus and waxy notes, as well as subtle floral and sweet aromas. Conversely, long-chain aliphatic aldehydes (C_{12}-C_{19}) were also present, imparting stronger waxy characteristics.[23,24]

Furthermore, esters were detected in the pulps (5.20%) and seeds (2.39%) were also detected, with methyl (3*Z*,9*Z*,12*Z*)-octadecatrienoate (1.52%) (**69**) and ethyl linoleate (1.16%) (**58**) being the most abundant in pulps and seeds, respectively, conferring herbal and soft notes to this fruit. Interestingly, the ester composition exhibited high variability, including fatty acid esters of C_8, C_{14}, C_{16}, and C_{18} (saturated and/or unsaturated), aromatic esters, salicylates, and jasmonates (Table 1, Figure 2d). Although diverse in *T. mariae*, this compound class was underrepresented compared to the pulps of *T. bicolor* (> 30% ethyl acetate)[21] and *T. cacao* (> 25% 2-pentyl acetate),[22] where esters are the major VOCs.

Additionally, alcohols accounted for only 4.30% of the total VOCs in the pulp (phenylethyl alcohol, 0.96%, **8**), while traces of **8** (0.63%) were found in the seeds. Compound **8** has also been reported in fresh pulps of *T. cacao* and *T. grandiflorum*.[19,22] The distribution of aliphatic alcohols in the pulp of *T. mariae* spans a wide range of alkyl chain lengths (C_8-C_{18}) (Table 1, Figure 2c), contrasting with those previously observed in *T. cacao* (C_5-C_{11})[21] and *T. bicolor* (C_4-C_8).[21,22]

Moreover, a relatively high abundance of theacrine (**60**) (1.70%), a purine alkaloid, was detected in the pulp. This

Table 1. Volatile organic compounds annotated in seeds and pulp samples of *T. mariae* by HS-SPME/GC-MS

No.	Compound[a]	t_R / min[b]	Class	$RI_{exp.}$	$RI_{theo.}$	Relative area / %		Spectral similarity / %
						Seed	Pulp	
1	(2*E*)-Heptenal	10.05	aldehyde	960	958	nd	0.23 ± 0.05	85
2	Benzaldehyde	10.19	aldehyde	964	966	4.47 ± 0.17	nd	93
3	2-Ethyl-1-hexanol	12.55	alcohol	1031	1031	nd	0.48 ± 0.06	88
4	Benzeneacetaldehyde	13.10	aldehyde	1048	1048	8.53 ± 0.87	0.45 ± 0.07	96[c]
5	(2*E*)-Octenal	13.60	aldehyde	1062	1062	nd	2.98 ± 0.13	88
6	Linalool	14.94	monoterpene	1101	1101	nd	2.10 ± 0.29	96
7	Nonanal	15.10	aldehyde	1107	1107	0.81 ± 0.20	1.41 ± 0.12	96[c]
8	Phenylethyl alcohol	15.41	alcohol	1117	1117	0.63 ± 0.19	0.96 ± 0.10	95[c]
9	(2*E*,6*Z*)-Nonadienal	16.66	aldehyde	1157	1156	nd	0.49 ± 0.04	86
10	(2*E*)-Nonenal	16.88	aldehyde	1164	1164	0.20 ± 0.05	0.36 ± 0.03	92[c]
11	(2*E*,6*Z*)-Nonadien-1-ol	17.01	alcohol	1168	1169	nd	0.21 ± 0.02	88
12	(2*E*)-Nonen-1-ol	17.10	alcohol	1171	1171	nd	0.14 ± 0.01	83
13	Menthol	17.46	monoterpene	1182	1182	nd	0.32 ± 0.03	92
14	Methyl salicylate	17.92	ester	1197	1197	nd	0.14 ± 0.01	85
15	α-Terpineol	18.00	monoterpene	1200	1199	nd	0.51 ± 0.07	89
16	Decanal	18.25	aldehyde	1209	1209	0.38 ± 0.08	1.04 ± 0.12	95.5[c]
17	5-Hydroxymethylfurfural	18.94	aldehyde	1232	1233	2.56 ± 0.23	nd	93
18	Nerol	19.57	monoterpene	1254	1254	nd	0.22 ± 0.02	84
19	Phenethyl acetate	19.72	ester	1259	1255	nd	0.13 ± 0.02	86
20	(2*E*)-Decenal	19.94	aldehyde	1266	1265	nd	0.67 ± 0.07	93
21	Nonanoic acid	20.09	fatty acid	1272	1271	nd	0.98 ± 0.07	93
22	α-Ethylidene-benzeneacetaldehyde	20.15	aldehyde	1274	1274	0.32 ± 0.06	nd	92
23	1-Decanol	20.19	alcohol	1275	1274	nd	0.32 ± 0.05	84
24	(2*E*,4*Z*)-Decadienal	20.86	aldehyde	1298	1298	0.45 ± 0.02	nd	82
25	Undecanal	21.19	aldehyde	1310	1310	0.29 ± 0.10	0.60 ± 0.07	95[c]
26	(2*E*,4*E*)-Decadienal	21.52	aldehyde	1323	1321	0.59 ± 0.16	0.65 ± 0.08	96[c]
27	(2*E*)-Undecenal	22.74	aldehyde	1368	1368	0.50 ± 0.16	1.11 ± 0.11	90[c]
28	Vanillin	23.67	aldehyde	1402	1400	0.59 ± 0.19	nd	94
29	Dodecanal	23.91	aldehyde	1412	1412	0.45 ± 0.07	1.71 ± 0.24	95.5
30	Hexanophenone	25.29	ketone	1467	–[b]	0.17 ± 0.06	nd	89
31	β-Acoradiene	25.43	sesquiterpene	1472	1475	0.29 ± 0.07	nd	90
32	1-Dodecanol	25.56	alcohol	1477	1476	nd	0.91 ± 0.11	95
33	Tridecanal	26.48	aldehyde	1515	1513	0.29 ± 0.02	0.61 ± 0.06	90[d]
34	1-Tridecanol	28.02	alcohol	1579	1575	nd	0.17 ± 0.02	85
35	Tetradecanal	28.88	aldehyde	1616	1615	0.28 ± 0.09	0.74 ± 0.08	95.5[c]
36	Methyl dihydroepijasmonate	29.77	ester	1655	1654	nd	0.20 ± 0.03	87
37	1-Tetradecanol	30.64	alcohol	1694	1678	nd	0.30 ± 0.02	90
38	Pentadecanal	31.17	aldehyde	1718	1717	0.23 ± 0.07	0.43 ± 0.12	94.5[c]
39	Octyl caprylate	32.48	ester	1779	1779	nd	0.43 ± 0.06	86
40	2-Ethylhexyl salicylate	33.12	ester	1810	1816	nd	0.31 ± 0.03	89
41	Hexadecanal	33.36	aldehyde	1821	1822	nd	0.50 ± 0.05	94
42	Isopropyl myristate	33.44	ester	1825	1824	nd	0.49 ± 0.07	87
43	Caffeine	34.03	alkaloid	1853	1850	0.15 ± 0.04	nd	86
44	1-Hexadecanol	34.65	alcohol	1884	1883	nd	0.41 ± 0.06	89
45	Homomenthyl salicylate	34.85	ester	1894	1903	nd	0.91 ± 0.07	89

Table 1. Volatile organic compounds annotated in seeds and pulp samples of *T. mariae* by HS-SPME/GC-MS (cont.)

No.	Compound[a]	t_R / min[b]	Class	$RI_{exp.}$	$RI_{theo.}$	Relative area / %		Spectral similarity / %
						Seed	Pulp	
46	5,7-Dimethyl undecane	34.96	hydrocarbon	1899	–	0.07 ± 0.01	nd	87
47	2-Heptadecanone	35.05	ketone	1904	1901	nd	0.25 ± 0.02	84
48	Heptadecanal	35.42	aldehyde	1923	1920	nd	0.53 ± 0.06	95
49	Methyl isohexadecanoate	35.51	ester	1927	–	nd	0.07 ± 0.01	83
50	Hexadecanoic acid	36.25	fatty acid	1965	1965	8.21 ± 1.75	6.30 ± 0.62	94[c]
51	Octadecanal	37.39	aldehyde	2025	2021	nd	0.47 ± 0.04	85
52	1-Octadecanol	38.57	alcohol	2088	2084	nd	0.38 ± 0.07	90
53	2-Nonadecanone	38.90	ketone	2106	2101	0.71 ± 0.19	0.21 ± 0.02	88.5[c]
54	Nonadecanal	39.28	aldehyde	2127	2127	nd	0.34 ± 0.04	86
55	(9*Z*,12*Z*)-Octadecadienoic acid	39.45	fatty acid	2139	2140	8.28 ± 1.94	4.47 ± 0.33	95[c]
56	Oleic acid	39.57	fatty acid	2147	2147	4.95 ± 0.26	8.88 ± 0.23	93[c]
57	(9*E*)-Octadecenoic acid	39.65	fatty acid	2149	–	nd	6.75 ± 0.11	90
58	Ethyl linoleate	39.89	ester	2161	2162	1.16 ± 0.10	nd	92
59	Octadecanoic acid	40.04	fatty acid	2172	2172	16.18 ± 0.90	23.56 ± 0.14	93[c]
60	Theacrine	41.42	alkaloid	2250	–	17.27 ± 3.67	1.70 ± 0.30	94.5[c]
61	Allyl stearate	41.90	ester	2278	–	0.75 ± 0.07	1.00 ± 0.08	87.5[c]
62	γ-Stearolactone	42.64	lactone	2322	–	0.34 ± 0.03	0.63 ± 0.07	83[c]
63	δ-Octadecalactone	43.19	lactone	2356	–	0.27 ± 0.02	nd	87
64	Meadowlactone	43.20	lactone	2357	–	nd	0.46 ± 0.04	84
65	Arachidic acid	43.43	fatty acid	2370	2365	8.70 ± 0.47	17.85 ± 0.20	91.5[c]
66	Hexadecanamide	43.82	amide	2394	–	0.60 ± 0.03	nd	81
67	Glycerol 1-oleate	44.30	monoacylglycerol	2424	–	0.48 ± 0.06	nd	87
68	(8*Z*,11*Z*,14*Z*)-Eicosa-8,11,14-triene	45.08	hydrocarbon	2474	–	1.71 ± 0.16	nd	86
69	Methyl (3*Z*,9*Z*,12*Z*)-octadecatrienoate	45.09	ester	2475	–	nd	1.52 ± 0.08	86
70	β-Sitosterol	48.25	triterpenoid	2643	–	4.92 ± 1.01	nd	85
71	Squalene	50.07	triterpene	2714	–	2.10 ± 0.76	nd	95
72	γ-Tocopherol	53.28	meroterpenoid	2957	–	1.11 ± 0.24	nd	93

[a]Annotation based on spectral similarities and/or retention indexes; [b]Theorical retention index not available; [c]Average among samples; nd = not detected; t_R: retention time; $RI_{exp.}$: experimental retention index; $RI_{theo.}$: theoretical retention index; HS-SPME/GC-MS: headspace solid-phase microextraction gas chromatography coupled with mass spectrometry.

odorless compound had not been previously reported among the volatile constituents of fresh *Theobroma* pulps.[19-22] Its presence may contribute to the mild bitterness of the pulp and potentially to its stimulant properties.[25] On the other hand, compound **60** was identified as the principal VOC detected in the seed samples (17.27%). This purine alkaloid, structurally similar to caffeine (**43**, 0.15% in seeds), is known for its psychostimulant activity, modulating adenosinergic and dopaminergic pathways and acting as a stimulant.[25] The high concentration of methylxanthines, such as theobromine and caffeine, in the seeds is known to contribute to the characteristic bitterness of cocoa-based products. This finding is consistent with previous studies reporting high theacrine concentrations in *T. mariae* seeds.[15] Conversely, our results contrast with those for *T. cacao* and *T. grandiflorum* seeds, which contain higher levels of caffeine and theobromine,[26] indicating that the distinct sensory properties of products derived from *T. mariae* beans may be associated with the presence of theacrine. Interestingly, the accumulation of theacrine, one of the final molecules in the biosynthesis of methylxanthine alkaloids, indicates that the individuals selected for sampling preferentially convert caffeine through N9 methylation and C8 oxidation. This pattern contrasts with previous studies reporting the accumulation of theobromine in *T. mariae* seeds. The higher expression of genes associated with these biosynthetic steps may be influenced by environmental factors such as seasonality, soil type, and the presence or absence of diseases and predators. Low relative abundance of caffeine was observed, while theophylline and theobromine were not detected in *T. mariae* seeds in this study.

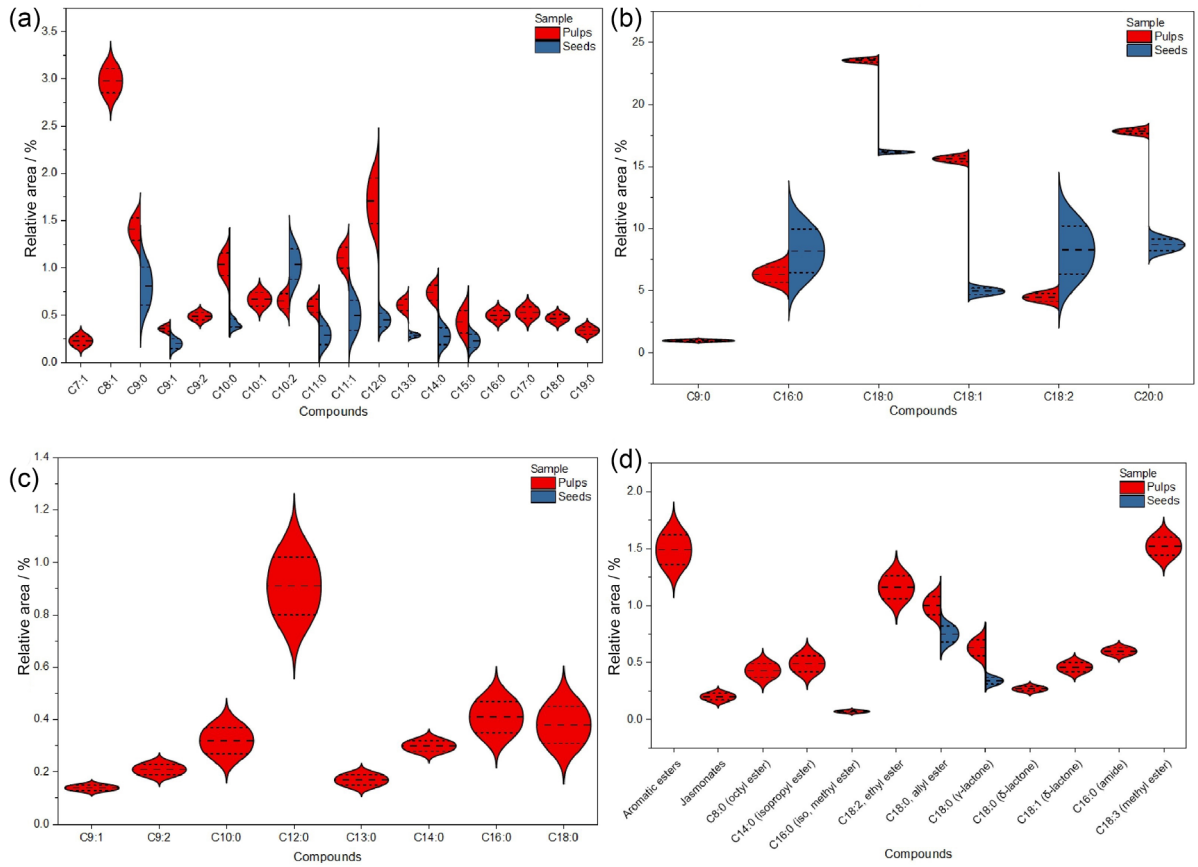


Figure 2. Half violin plots showing the distribution of compounds from different chemical classes in the pulp (red) and seeds (blue) of *T. mariae* samples, where the central dashed line in each violin plot represents the mean relative abundance of a given compound and the dashed lines at the upper and lower extremes indicate the positive and negative standard deviation, respectively; (a) distribution of aldehydes, (b) distribution of fatty acids, (c) distribution of alcohols, and (d) distribution of esters.

Finally, terpenes and terpenoids accounted for 3.15% of the total VOCs in the pulp and 8.42% in the seeds. Mainly composed of the monoterpenes linalool (**6**) (2.10%), α-terpineol (**15**) (0.51%), menthol (**13**) (0.32%), and nerol (**18**) (0.22%), which contribute citrus and floral aromas, with subtle woody notes in the pulp. The terpenoid compounds β-sitosterol (**70**) (4.92%), squalene (**71**) (2.10%), β-acoradiene (**31**) (0.29%), and γ-tocopherol (**72**) (1.11%) were observed exclusively in the seeds. The monoterpenes have been previously reported in other *Theobroma* species, including *T. cacao*, *T. grandiflorum*, and *T. bicolor*,[19-22] whereas the triterpenes **70** and **71** are reported here for the first time in the *Theobroma genus*.

UHPLC-HRMS/MS analysis

The chemical composition analysis of *Theobroma mariae* seeds and pulp allowed the annotation of 32 compounds, including amino acids, benzoic acids, hydroxycinnamic acids, methylxanthines, and flavonoids. Among the flavonoids, aglycones, *O*-glycosylated, *C*-glycosylated, and glucuronide forms were annotated. Annotation was performed manually by determining chemical formulas using high-resolution mass spectrometry and by manual curation of product ion spectra in comparison with literature data, allowing compound identification at level 2. The annotated compounds were numbered according to their elution order (Table 2).

Amino acids

Compound **1** (0.68 min) showed an *m/z* of 175.1186 $[M + H]^+$ (5.14 ppm), with MS/MS fragments at *m/z* 158, 130, 116, 70, and 60, corresponding to losses of H_3N (−17 Da), CH_3NO (−45 Da), CH_5N_3 (−59 Da), $C_2H_7N_3O_2$ (−105 Da), and $C_5H_9NO_2$ (−115 Da), respectively. These fragmentation patterns supported its annotation as the amino acid arginine.[27] In addition, other amino acids were annotated in *T. mariae* samples following similar fragmentation logic, including lysine (**2**, 0.71 min, *m/z* 147.1126 $[M + H]^+$, 4.75 ppm),[28] proline (**3**, 0.71 min, *m/z* 116.0703 $[M + H]^+$, 6.89 ppm),[28] aspartic acid (**4**, 0.72 min, *m/z* 132.0302 $[M - H]^-$, 3.79 ppm),[28] glutamic acid (**5**, 0.73 min, *m/z* 146.0458 $[M - H]^-$, 3.42 ppm),[28]

leucine or isoleucine (**6**, 1.15 min, *m/z* 132.1016 $[M + H]^+$, 6.81 ppm),[28] and phenylalanine (**7**, 1.68 min, *m/z* 166.0860 $[M + H]^+$, 4.82 ppm).[28]

These amino acids were found in greater diversity in *T. mariae* seeds than in the pulp. Interestingly, compounds of this class (**1**, **6**, and **7**) have also been reported in cocoa honey produced from the mucilage surrounding cocoa beans,[29] but in contrast to that study, these amino acids were exclusively detected in the seed samples of *T. mariae* in the present work.

Table 2. Non-volatile compounds annotated in pulp and seed samples by UHPLC-HRMS/MS in positive ($[M + H]^+$ and negative ($[M - H]^-$) ionization modes

Compound	t_R / min	Annotation	*m/z* (error) positive	MS/MS *m/z*	*m/z* (error) negative	MS/MS *m/z*	Tissue	References
1	0.68	Arginine	175.1186 (5.14)	158, 130, 116, 70, 60	nd	–	seeds	27
2	0.71	Lysine	147.1126 (4.75)	130, 84	nd	–	seeds	28
3	0.71	Proline	116.0703 (6.89)	70	nd	–	seeds	28
4	0.72	Aspartic acid	nd	–	132.0302 (3.79)	115, 88	seeds	28
5	0.73	Glutamic acid	nd	–	146.0458 (3.42)	128, 102	seeds	28
6	1.15	Leucine/isoleucine	132.1015 (6.81)	86, 69	nd	–	seeds	28
7	1.68	Phenylalanine	166.0860 (4.82)	120, 103	nd	–	seeds	28
8	2.68	4-Hydroxybenzoic acid	139.0387 (5.75)	111, 93, 65	nd	–	seeds	30
9	2.84	(+)-Catechin / (–)-epicatechin	291.0857 (4.12)	139, 123	nd	–	seeds	31
10	3.07	Theacrine	225.0976 (5.33)	210, 168, 153	nd	–	pulp / seed	32
11	3.76	Vanillin	153.0544 (5.23)	125, 111, 93, 65	151.0400 (3.31)	136	seeds	33,34
12	3.90	Hypolaetin 8-*O*-glucuronide-glucoside	nd	–	639.1191 (0.94)	463, 301	seeds	35
13	4.05	Vanillic acid	nd	–	167.0347 (1.8)	152, 123, 108	pulp / seed	36
14	4.12	Sinapinic acid	nd	–	223.0608 (0.90)	179, 137	pulp	37
15	4.14	Apigenin-6-*C*-arabinosyl-8-*C*-glucoside (isochaftoside) or apigenin-8-*C*-glucoside-6-*C*-arabinosyl (chaftoside)	nd	–	563.1395 (1.06)	443, 383, 353, 269	seeds	38
16	4.21	Hesperetin	303.0857 (3.96)	151, 123	nd	–	seeds	39
17	4.26	Caffeic acid	nd	–	179.0347 (1.67)	161, 135	pulp	40
18	4.31	Luteolin-7-*O*-glucoside (cyranoside)	449.1071 (2.89)	287	447.0919 (1.79)	285	seeds	41
19	4.32	Apigenin-8-*C*-glucoside (vitexin)	433.1123 (2.77)	313, 283, 165	431.0970 (1.85)	269	seeds	42
20	4.35	Kaempferol-3-*O*-rutinoside	nd	–	593.1501 (0.84)	411, 285	seeds	43
21	4.67	Baicalein-7-*O*-glucoronide (baicalin)	nd	–	445.0755 (3.59)	269	seeds	44
22	4.68	Hypolaetin-8-*O*-glucuronide-3''-sulfate (theograndin II)	559.0374 (3.58)	479, 303	557.0228 (1.61)	477, 301, 255	seeds	35,45
23	4.75	Apigenin-7-*O*-malonyl-glucoside	519.1138 (0.19)	271	nd	–	seeds	46,47
24	4.83	Quercetin-3-*O*-glucuronide (miquelianin)	479.0811 (3.13)	303	nd	–	seeds	48
25	4.84	Quercetin-3-*O*-glucoside (isoquercetrin)	465.1021 (2.58)	303	nd	–	seeds	49
26	4.96	Hypolaetin-3''-methyl ether-8-*O*-glucuronide	493.0970 (2.43)	439, 317, 302	nd	–	seeds	35
27	5.01	Azelaic acid	nd	–	187.0972 (1.06)	125, 97	seeds / pulp	50
28	5.01	Isoscutellarein-8-*O*-glucuronide-3''-sulfate (theograndin I)	543.0427 (3.31)	463, 287	541.0279 (1.66)	461, 285, 255	seeds	35,45

Table 2. Non-volatile compounds annotated in pulp and seed samples by UHPLC-HRMS/MS in positive ($[M + H]^+$ and negative ($[M - H]^-$) ionization modes (cont.)

Compound	t_R / min	Annotation	*m/z* (error) positive	MS/MS *m/z*	*m/z* (error) negative	MS/MS *m/z*	Tissue	References
29	5.02	Kaempferol-3-*O*-glucoside (astragalin)	449.1071 (2.89)	287	447.0920 (1.56)	285	seeds	49
30	5.06	Hypolaetin-3'-methyl ether-8-*O*-glucuronide-3"-*O*-sulfate	573.0537 (2.26)	493, 317, 302	571.0389 (0.87)	491, 315, 300, 255	seeds / pulp	35
31	5.11	Kaempferol-3-*O*-glucuronide	463.0866 (2.16)	287	461.0714 (1.30)	285	seeds / pulp	51
32	6.25	Isoscutellarein-8-*O*-glucuronide	nd	-	461.0712 (1.73)	285	seeds	35,52

nd = not detected on sample; t_R: retention time; MS/MS: tandem mass spectrometry; HS-SPME/GC-MS: headspace solid-phase microextraction gas chromatography coupled with mass spectrometry.

Alkaloids

Similarly to the VOC profile, compound **10** (3.07 min) was observed in both pulp and seed samples. It showed *m/z* 225.0976 $[M + H]^+$ (5.33 ppm) with major fragments at *m/z* 210, 168, and 153, due to successive losses of CH_3 (–15 Da), CH_3 + CNO (–57 Da) and $2CH_3$ + CNO (–72 Da), consistent with theacrine.[30] This compound was recently reported in the pulps and seeds of *T. sylvestre*[17] and in previous studies in seeds of *T. mariae* and eight other *Theobroma* species.[15]

Phenolic acids e derivatives

Compound **8** (2.68 min) was annotated as 4-hydroxybenzoic acid (*m/z* 139.0387 $[M + H]^+$, 5.75 ppm) based on its characteristic fragments at *m/z* 111, 93, and 65, corresponding to losses of CO (–28 Da), CH_2O_2 (–46 Da), and $C_2H_2O_3$ (–74 Da), respectively.[30] In addition to these benzenoids, the aldehyde vanillin (**11**, 3.76 min) was identified with *m/z* 153.0544 $[M + H]^+$ (5.23 ppm), displaying typical losses of CO (–28 Da), C_2H_2O (–42 Da), $C_2H_4O_2$ (–60 Da) and $C_3H_4O_3$ (–88 Da), and with *m/z* 151.0401 $[M–H]^-$ (3.92 ppm) showing a CH_3 loss (–15 Da), consistent with literature data.[33,34] The acidic derivative of compound **11**, vanillic acid (**13**, 4.05 min), presented *m/z* 167.0347 $[M - H]^-$ (1.8 ppm) and fragments at *m/z* 152, 123, and 108, corresponding to CH_3 (–15 Da), CO_2 (–44 Da), and $CH_3 + CO_2$ (–59 Da) losses, respectively, in agreement with previously reported data.[36]

Benzoic acid derivatives were observed in both *T. mariae* samples, though this class was more diverse in the pulp. Compounds of this class have been previously reported in *Theobroma* species, including compound **8**, identified as one of the major constituents of *T. cacao* bean shells.[53] Compound **11** has been previously reported only in fermented seeds of *T. grandiflorum*,[54] while glycosylated isomers of compound **13** have recently been described in the pulps of *T. sylvestre*,[17] and the same compound reported here was previously described in the pulps of *T. grandiflorum*.[55]

Moreover, hydroxycinnamic acids were annotated in our analyses, with higher relative abundance observed in the pulp. Additionally, compound **14** (4.12 min, *m/z* 223.0608 $[M - H]^-$, 0.90 ppm) exhibited fragment ions at *m/z* 179, 137, and 97, consistent with the structure of sinapinic acid.[37] This compound has not been previously reported in *Theobroma* fruits, having been isolated only from *T. cacao* leaves.[56]

Compound **17** (4.26 min) showed *m/z* 179.0347 $[M - H]^-$ (1.67 ppm) and produced a fragment ion at *m/z* 135, corresponding to a loss of CO_2 (–44 Da), consistent with the fragmentation pattern of caffeic acid.[40] Compound **17** has been previously reported in *T. grandiflorum* pulp[54] and in wild species such as *Theobroma subincanum* and *Theobroma asperum*.[57]

In addition, the saturated-chain dicarboxylic acid azelaic acid (**27**, 5.01 min, *m/z* 187.0972 $[M - H]^-$, 1.06 ppm) was annotated based on fragment ions at *m/z* 125 and 97, arising from losses of $CO_2 + H_2O$ (–62 Da) and $H_2O + C_3H_4O_2$ (–90 Da), respectively.[58] The presence of this compound suggests potential applications of *T. mariae* fruits beyond food purposes, as substances like **27**, along with polyphenols, are highly desirable in cosmetic formulations.[59,60]

Flavonoid aglycones and their *O*- and *C*-glycosides

In addition to small phenolic acids, *T. mariae* samples were dominated by various flavonoid derivatives, including aglycones, *O*- and *C*-glycosylated forms, which were

exclusively detected in the seeds. Among them, one of the possible flavan-3-ol epimers (catechin/epicatechin) (**9**, 2.84 min) showed *m/z* 291.0857 $[M + H]^+$ (4.12 ppm), with product ions at *m/z* 139 and 123 corresponding to the losses of $C_9H_8O_3$ (−152 Da) and $C_8H_8O_4$ (−168 Da), consistent with literature data.[31]

Furthermore, the di-*C*-glycosylated flavone **15** (4.14 min, *m/z* 563.1395 $[M - H]^-$, 1.06 ppm) exhibited fragmentation typical of *C*-glycosylated derivatives, with losses of $C_4H_8O_4$ (−120 Da), hexose + H_2O (−180 Da), hexose + H_2O + CH_2O (−210 Da), hexose + $C_2H_4O_3$ (−238 Da), hexose + $C_3H_4O_4$ (−297 Da) and hexose + pentose (−294 Da). These patterns supported its annotation as apigenin-6-*C*-arabinosyl-8-*C*-glucoside (isoschaftoside) or its structural isomer apigenin-6-*C*-glucoside-8-*C*-arabinosyl.[38]

The flavanone aglycone hesperetin (**16**, 4.21 min) presented *m/z* 303.0857 $[M + H]^+$ (3.96 ppm), producing fragment ions at *m/z* 151 and 123, corresponding to the losses of $C_8H_8O_3$ (−152 Da) and $C_9H_8O_4$ (−180 Da), respectively.[39] Compound **18** (4.31 min) was annotated as the flavone luteolin-7-*O*-glucoside (cyranoside) (*m/z* 449.1071 $[M + H]^+$, 2.89 ppm), due to the characteristic loss of a hexose unit (−162 Da), generating a fragment ion corresponding to the aglycone (*m/z* 287).[41]

Apigenin-8-*C*-glucoside (vitexin) (**19**, 4.32 min) was annotated based on the deduced formula from *m/z* 433.1123 $[M + H]^+$ (2.77 ppm) and fragment ions at *m/z* 313, 283, and 165.[42] Compound **20** (4.35 min) exhibited *m/z* 593.1501 $[M - H]^-$, consistent with the molecular formula of kaempferol-3-*O*-rutinoside, whose annotation was supported by its characteristic fragmentation showing the aglycone-related ion.[43]

Additionally, other flavonoids were annotated, such as apigenin-7-*O*-malonyl-glucoside (**23**), eluting at 4.75 min with *m/z* 519.1138 $[M + H]^+$ (0.19 ppm), which exhibited a neutral loss of 248 Da corresponding to a malonyl-glucose moiety, forming the fragment ion of the aglycone (*m/z* 271).[46,47] Compound **25** (4.84 min, *m/z* 465.1021 $[M + H]^+$, 2.58 ppm) was annotated as quercetin-3-*O*-glucoside (isoquercitrin) based on its fragmentation pattern showing a characteristic hexose loss that generated an ion at *m/z* 303.[49] Finally, kaempferol-3-*O*-glucoside (astragalin) (**29**, 5.02 min, *m/z* 449.1071 $[M + H]^+$, 2.89 ppm; *m/z* 447.0919 $[M - H]^-$, 1.79 ppm) was annotated according to its fragmentation profile in agreement with literature data[61] (Table 2).

These flavonoids have been previously reported in studies of the *Theobroma* genus. For instance, compound **9**, annotated as catechin, was earlier identified in *T. mariae* seeds. More recent studies have also reported this compound in the pulp of *T. grandiflorum*[55] and in both fresh and roasted *T. cacao* seeds.[61] The alternative flavan-3-ol epimer of compound **9**, epicatechin, has been found in *T. cacao* seeds,[62] *T. grandiflorum* husks,[54] and *T. sylvestre* fruits.[17] Compound **15** was previously described in *T. cacao* pods and veins,[63] while compound **19** was also reported in *T. cacao* pods.[64] In contrast, compound **23** has not been previously reported in any *Theobroma* species. Nevertheless, the aglycone moiety of compounds **15**, **19**, and **23** has been identified in *T. cacao* seeds.[62]

Compound **16** (hesperetin) is commonly reported in *Citrus* species; however, it had not been described in previous studies on *Theobroma*, and was here annotated in the seed samples. Compound **18**, detected in the seeds of this study, had been previously detected in *T. cacao* seeds,[62] with its aglycone also described in *T. grandiflorum* pulp[40] and *T. cacao* seeds.[43] Compound **25**, present in the seed samples, has been previously reported in *T. cacao* husks and veins,[63] while compound **20**, annotated in *T. mariae* seeds, was also found in *T. cacao* seeds.[43] Lastly, compound **29**, also detected in the seed samples, was recently reported as a major component in *T. sylvestre*.[17]

Flavonoid glucoronides

A distinctive feature of *Theobroma* fruits is the presence of glucuronide derivatives, which have been reported in nearly all species analyzed through metabolomic approaches. Interestingly, these compounds are recognized as human metabolites derived from the consumption of flavonoid-rich foods. Due to the low concentration of active flavonoids in the bloodstream, glucuronide metabolites are of great interest for understanding how these compounds exert their effects in tissues.

Compound **12** (3.90 min) exhibited *m/z* 639.1191 $[M - H]^-$ (0.94 ppm), with MS/MS spectra dominated by product ions at *m/z* 463 and 301, corresponding to losses of glucuronide (−176 Da) and glucose (−162 Da), respectively, consistent with the structure of hypolaetin-8-*O*-glucuronide glucoside. Baicalein-7-*O*-glucuronide (**21**, 4.67 min, *m/z* 445.0755 $[M - H]^-$, 3.59 ppm) was annotated based on the neutral loss of 176 Da, in agreement with literature data.[35] Compound **22** (4.68 min, *m/z* 559.0374 $[M + H]^+$, 3.58 ppm) displayed a fragment ion at *m/z* 303 from the loss of a sulfoglucuronide group (−256 Da). The same compound, in negative mode (*m/z* 557.0228 $[M - H]^-$, 1.61 ppm), showed fragment ions at *m/z* 477 and 301, corresponding to competitive losses of sulfate (−80 Da) and sulfoglucuronide (−256 Da), along with *m/z* 255 (sulfoglucuronide$^-$).[35,45] These features supported its annotation as hypolaetin-8-*O*-glucuronide-3''-sulfate, also known as theograndin II.[45]

Additionally, compounds **24** (4.83 min, *m/z* 479.0811 $[M + H]^+$, 3.13 ppm), **26** (4.96 min, *m/z* 493.0970 $[M + H]^+$, 2.43 ppm), and **30** (5.06 min, *m/z* 573.0537 $[M + H]^+$, 2.26 ppm; 5.14 min, *m/z* 571.0389 $[M - H]^-$, 0.87 ppm) were annotated as quercetin-3-*O*-glucuronide (miquelianin), hypolaetin-3"-methylether-8-*O*-glucuronide, and hypolaetin-3'-methylether-8-*O*-glucuronide-3"-*O*-sulfate, respectively, based on characteristic glucuronide losses and MS/MS fragmentation patterns.[35]

Moreover, compound **28** (5.01 min, *m/z* 543.0427 $[M + H]^+$, 3.31 ppm; *m/z* 541.0279 $[M - H]^-$, 1.66 ppm) exhibited the same fragmentation pattern as compound **22** but with a fragment ion corresponding to the aglycone isoscutellarein (*m/z* 285 and 287), consistent with the loss of a sulfoglucuronide moiety (–256 Da). Thus, compound **28** was annotated as isoscutellarein-8-*O*-glucuronide-3"-sulfate (theograndin I).[35,45] Finally, compounds **31** (5.11 min, *m/z* 463.0866 $[M + H]^+$, 2.16 ppm; *m/z* 461.0714 $[M - H]^-$, 1.30 ppm) and **32** (6.25 min, *m/z* 461.0712 $[M - H]^-$, 1.73 ppm) were annotated following the same logic as compounds **24** and **26**, being identified as kaempferol-3-*O*-glucuronide and isoscutellarein-8-*O*-glucuronide, respectively.[35,51]

Glucuronidated flavonoids were predominantly detected in the seeds of *T. mariae* including compounds, **22** and **28**. These compounds were first isolated from *T. grandiflorum* seeds[45] and later observed in both seeds[35] and pulp[51] of this species, which was previously the only *Theobroma* species known to contain these metabolites. Thus, this represents only the second report of such compounds within the genus. Similarly, compound **32**, which corresponds to the desulfated form of compound **28**, was also found in both pulp and seeds of *T. grandiflorum*,[35,52] but in this study it was detected exclusively in seeds. Compounds **12** and **26**, sharing the same aglycone as compound **22** but with distinct substituents, had also been reported in *T. grandiflorum* pulp and seeds.[35,52] Compound **24**, found in the seed samples, was likewise reported in *T. grandiflorum* seeds.[45] Compound **21**, annotated in *T. mariae* seeds, had previously been reported in *T. grandiflorum* pulp.[55] Finally, compound **31** was the only member of this class detected in both pulp and seed samples in the present study, and was also previously reported *T. cacao* pods and veins.[63]

Functional properties

Flavonoids, the most well-represented compound class in this study, offer multiple nutraceutical benefits.

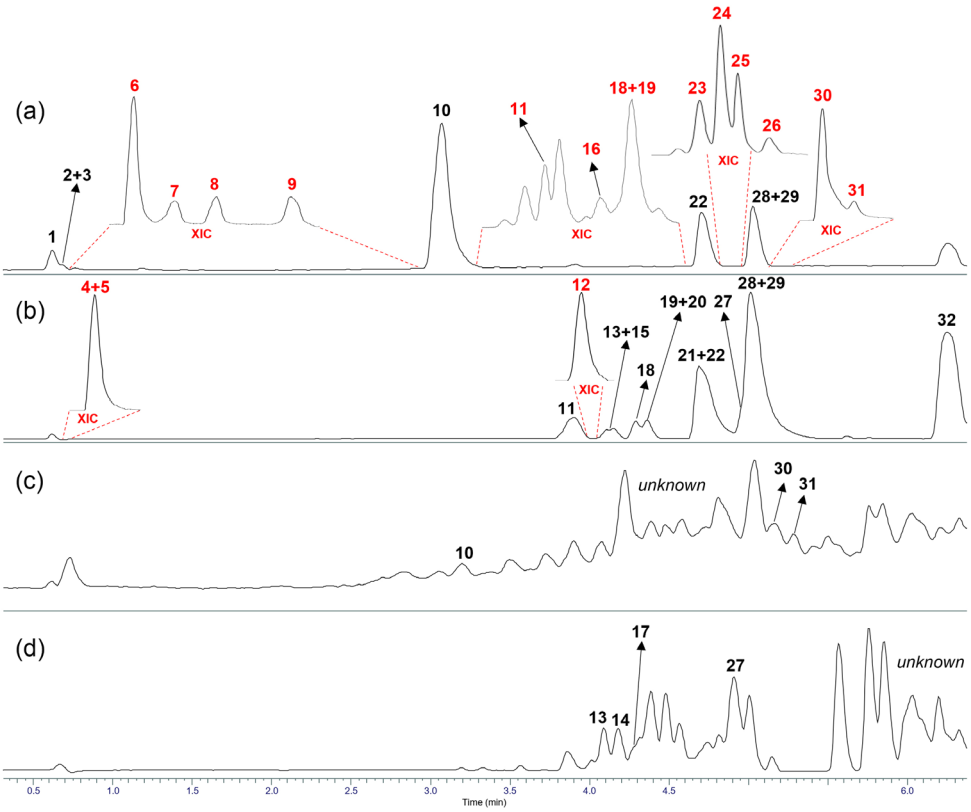


Figure 3. Total ion chromatograms (TIC) obtained from UHPLC-HRMS/MS analysis of seeds and pulps of *T. mariae*. (a) TIC and extracted ion chromatograms (XIC, red numbered peaks) obtained in positive mode for the seeds. (b) TIC and XICs obtained in negative mode for the seeds. (c) TIC obtained in positive mode for the pulp. (d) TIC obtained in negative mode for the pulp.

The consumption of these compounds is associated with cellular protection, metabolic regulation, and preventive effects against age-related diseases. Apigenin and its derivatives, such as compound **20**, exhibit antioxidant and anti-inflammatory properties.[65] Apigenin also protects neurons against oxidative stress, and exerts antidiabetic and antidepressant effects.[66] Another derivative, compound **15**, protects neural cells by inhibiting phosphorylation within signaling pathways.[67] Compound **16** protects neurons from oxidative stress and inflammation-induced damage and prevents the release of neurotoxic substances in neurodegenerative disease models.[68] Compound **9** exerts cardioprotective effects by preventing low-density lipoprotein (LDL) oxidation. In contrast, compound **32** helps reduce total cholesterol, LDL, and triglyceride levels, lowers blood pressure, and improves glucose metabolism by reducing glycemia and enhancing insulin sensitivity.[69] The glycosylated flavonoid **30** improves intestinal structure and reduces pro-inflammatory cytokines.[70] Compound **20** also acts as a hydrogen donor to reactive oxygen species (ROS) and regulates cholesterol metabolism.[71]

In addition to glycosylated flavonoids, glucuronidated ones were also abundant. Although these compounds are typically described as mammalian metabolites, they are present in plant physiology but are relatively rare compared to other conjugates. Their main characteristic is high solubility, which facilitates transport, enhances bioavailability, and promotes excretion.[72] Moreover, the position of glucuronidation directly influences antioxidant potential, with substitutions at positions 3 and 7 showing the highest activity.[73] Thus, studies suggest that glucuronidation not only enables excretion but may also help direct molecules to their specific sites of action.[72] Ongoing research on such flavonoids has demonstrated their therapeutic potential; for instance, compound **21** helps regulate pancreatic function and hypertension.[74] Compounds **22** and **29** exhibit antioxidant activity, with compound **22** showing greater potency than compound **29**.[45] Finally, compound **24** has been reported to promote neurogenesis by stimulating receptor-mediated proliferation and maturation.[75]

In addition to nutraceutical properties, the annotated compounds also exhibit cosmetic applications. Compound **27** stands out in dermatology for its anti-comedogenic action and ability to modulate melanogenesis, making it an effective agent in the treatment of acne and hyperpigmentation.[31,60] Compound **11** functions as a fragrance fixative and also as a stimulator for epithelial cell regeneration, strengthening the skin barrier against pathogens.[76] Compound **17** enhances UV absorption and reduces erythema and wrinkles through its antioxidant and anti-inflammatory properties.[77] Within the flavonoid group, the aglycone of compounds **24** and **25** inhibits collagen degradation and accelerates wound healing, while compound **21**, when integrated into bio-vesicles, optimizes fibroblast repair and combats severe oxidative damage, acting synergistically to preserve the extracellular matrix.[78]

Conclusions

The characterization of *Theobroma mariae* ("cacau-jacaré"), a little-known Amazonian fruit, was achieved through the combination of HS-SPME/GC-MS and UHPLC-HRMS/MS techniques. A total of 78 VOCs and 32 non-volatile compounds were annotated, including several recognized antioxidants. The aromatic profile of its seeds – analyzed for the first time in this species – showed similarity to those of the *T. cacao* cultivars Criollo, Forastero, and Trinitario. The presence of bioactive compounds such as phenolic acids and flavonoids, with antioxidant and therapeutic properties, highlights the potential of this fruit for applications in the food and pharmaceutical industries. These findings expand current knowledge and appreciation of a rare species from the Brazilian Amazon rainforest, paving the way for new commercial uses and the assessment of its economic potential as a natural resource of the Amazon.

Supplementary Information

Supplementary information is available free of charge at http://jbcs.sbq.org.br as PDF file.

Data Availability Statement

The data used in this study are publicly available in the MassIVE repository under the accession code MSV000100491. The dataset can be accessed directly via ftp://MSV000100491@massive-ftp.ucsd.edu, using the password 7355kkgw66HX2VHz for download and consultation.

Acknowledgments

The authors are grateful to the Brazilian agencies: CNPq (315040/2021-1, 443823/2024-3, 407108/2023-8), FAPEAM, FINEP, CAPES (finance code 001).

The base images of the graphical abstract were generated using Microsoft Copilot and later edited in Canva. The text was revised using OpenAI Grammar to ensure grammatical accuracy.

Author Contributions

Conceptualization: H. H. F. K.; data curation: S. O. A. J., E. G. P., F. A. S. R., E. J. S. P. L., M. B. A. J.; formal analysis: S. O. A. J.,

M. B. A. J., G. A. B.; investigation: S. O. A. J., E. G. P., F. A. S. R.; methodology: F. A. S. R., G. A. B., H. H. F. K.; writing the original draft: SOAJ; reviewing, and editing: H. H. F. K.; resources: E. G. P., G. A. B., H. H. F. K.; supervision: M. B. A. J., G. A. B, H. H. F. K.; project funding and administration: H. H. F. K.

References

1. Miranda, C. T. C. D. S.; Soares, S. D.; De Oliveira, W. Q.; Lima, A. D. S.; Neri Numa, I. A.; Pastore, G. M.; *Foods* **2024**, *13*, 2925. [Crossref]
2. Silveira, J. T. D.; Rosa, A. P. C. D.; Morais, M. G. D.; Victoria, F. N.; Costa, J. A. V.; *Food Res. Int.* **2023**, *173*, 113304. [Crossref]
3. Kinupp, V. F.; In *Plantas alimentícias não convencionais (PANC) no Brasil: guia de identificação, aspectos nutricionais e receitas ilustradas*, 1st ed.; Kinupp, V. F.; Lorenzi, H. V.; eds.; Plantarum: São Paulo, BR, 2014.
4. Colli-Silva, M.; Richardson, J. E.; Bossa-Castro, A. M.; Pirani, J. R.; *Brittonia* **2024**, *76*, 53. [Crossref]
5. da Silva, C. V. A.; Salimo, Z. M.; de Souza, T. A.; Reyes, D. E.; Bassicheto, M. C.; de Medeiros, L. S.; Sartim, M. A.; de Carvalho, J. C.; Gonçalves, J. F. C.; Monteiro, W. M.; Tavares, J. F.; De Melo, G. C.; da Silva, F. M. A.; Bataglion, G. A.; Koolen, H. H. F.; *Food Res. Int.* **2024**, *192*, 114729. [Crossref]
6. Tineo, D.; Calderon, M. S.; Maicelo, J. L.; Oliva, M.; Huamán-Pilco, Á. F.; Ananco, O.; Bustamante, D. E.; *Mitochondrial DNA Part B* **2024**, *9*, 227. [Crossref]
7. da Rosa, J. S. D.; Oliveira Moreira, P. I.; Carvalho, A. V.; Freitas-Silva, O.; *Processes* **2024**, *12*, 1353. [Crossref]
8. Flora e Funga do Brasil. [Link] acessed in October 2025
9. Schumann, K. M. In *Flora Brasiliensis*, vol. 12, 1st ed.; Martius C. F. P.; Eichler A. W.; Urban I., eds.; Frid. Fleisch: Munich, DE, 1886. [Crossref]
10. Schultes, R. E.; *J. Arnold Arbor.* **1958**, *34*, 217. [Crossref]
11. Secco, R. S.; do Rosário, A. S.; Praia, T. S.; *Pesquisas Botânicas* **2016**, *69*, 8. [Link] acessed in October 2025
12. Carpenter, D. R.; Hammerstone, J. F.; Romanczyk, L. J.; Aitken, W. M.; *J. Am. Oil Chem. Soc.* **1994**, *71*, 845. [Crossref]
13. Gilabert-Escrivá, M. V.; Gonçalves, L. A. G.; Silva, C. R. S.; Figueira, A.; *J. Sci. Food Agric.* **2002**, *82*, 1425. [Crossref]
14. Marx, F.; Maia, J. G. S.; *Eur. Food Res. Technol.* **1991**, *193*, 460. [Crossref]
15. Hammerstone Jr., J. F.; Romanczyk Jr., L. J.; Aitken, W. M.; *Phytochemistry* **1994**, *35*, 1237. [Crossref]
16. Silva, L. Y. S.; Cavalcante, D. N.; Oliveira, E. S. C.; Salvador, A. M.; Pereira, Z. C.; Consentini, J. C. C.; Furlaneto, G.; Campelo, P. H.; Sanches, E. A.; Azevedo, L.; Bezerra, J. D. A.; *Plants* **2025**, *14*, 377. [Crossref]
17. Fonseca-Júnior, E. Q. D.; de Queiroz Rocha, D.; Souza Coelho Oliveira, E.; Oliveira da Silva, S.; Amaral Souza, F. D. C. D.; Furlaneto, G.; da Silva Cruz, L.; Azevedo, L.; Frota Corrêa, R.; Aparecido Sanches, E.; Campelo, P. H.; de Araújo Bezerra, J.; *Food Biosci.* **2025**, *73*, 107673. [Crossref]
18. Van Den Dool, H.; Kratz, P. D.; *J. Chromatogr.* **1963**, *11*, 463. [Crossref]
19. Boulanger, R.; Crouzet, J.; *Flavour Fragrance J.* **2000**, *15*, 251. [Crossref]
20. Quijano, C. E.; Pino, J. A.; *Food Chem.* **2007**, *104*, 1123. [Crossref]
21. Quijano, C. E.; Pino, J. A.; *J. Essent. Oil Res.* **2009**, *21*, 211. [Crossref]
22. Pino, J. A.; Ceballos, L.; Quijano, C. E.; *J. Essent. Oil Res.* **2010**, *22*, 113. [Crossref]
23. Kanter, J.-P.; Honold, P. J.; Luh, D.; Heiles, S.; Spengler, B.; Fraatz, M. A.; Zorn, H.; Hammer, A. K.; *J. Agric. Food Chem.* **2023**, *71*, 8112. [Crossref]
24. Kanter, J.-P.; Honold, P. J.; Lüke, D.; Heiles, S.; Spengler, B.; Fraatz, M. A.; Harms, C.; Ley, J. P.; Zorn, H.; Hammer, A. K.; *Appl. Microbiol. Biotechnol.* **2022**, *106*, 6095. [Crossref]
25. Dillon, H.; Steven, M.; *J. Exercise Nutr.* **2021**, *4*, 1. [Crossref]
26. Jean-Marie, E.; Jiang, W.; Bereau, D.; Robinson, J.-C.; *Foods* **2022**, *11*, 3966. [Crossref]
27. Vishwanathan, K.; Tackett, R. L.; Stewart, J. T.; Bartlett, M. G.; *J. Chromatogr. B: Biomed Sci. Appl.* **2000**, *748*, 157. [Crossref]
28. Le, A.; Ng, A.; Kwan, T.; Cusmano-Ozog, K.; Cowan, T. M.; *J. Chromatogr. B* **2014**, *944*, 166. [Crossref]
29. Rocha, G. H. A. M.; de Almeida, M. C.; da Silva, L. L. P.; Flores, I. S.; Castiglioni, G. L.; de Oliveira, T. F.; Pereira, J.; *Food Res. Int.* **2025**, *202*, 115694. [Crossref]
30. Amanpour, A.; Kelebek, H.; Selli, S.; *J. Mass Spectrom.* **2019**, *54*, 227. [Crossref]
31. Gardiman, M.; De Rosso, M.; De Marchi, F.; Flamini, R.; *Metabolomics* **2023**, *19*, 25. [Crossref]
32. Wang, Y.-H.; Mondal, G.; Butawan, M.; Bloomer, R. J.; Yates, C. R.; *J. Chromatogr. B* **2020**, *1155*, 122278. [Crossref]
33. Betts, T. A.; Palkendo, J. A.; *J. Chem. Educ.* **2018**, *95*, 1035. [Crossref]
34. Gao, H.; Yin, C.; Li, C.; Li, Y.; Shi, D.; Fan, X.; Yao, F.; Wu, W.; Li, J.; *Curr. Res. Food Sci.* **2023**, *6*, 100519. [Crossref]
35. Yang, H.; Protiva, P.; Cui, B.; Ma, C.; Baggett, S.; Hequet, V.; Mori, S.; Weinstein, I. B.; Kennelly, E. J.; *J. Nat. Prod.* **2003**, *66*, 1501. [Crossref]
36. Martínez-Inda, B.; Jiménez-Moreno, N.; Esparza, I.; Ancín-Azpilicueta, C.; *Antioxidants* **2025**, *14*, 42. [Crossref]
37. Griffiths, L. A.; *Biochemical J.* **1958**, *70*, 120. [Crossref]
38. Brito, A.; Ramirez, J.; Areche, C.; Sepúlveda, B.; Simirgiotis, M.; *Molecules* **2014**, *19*, 17400. [Crossref]
39. Kakouri, E.; Hatziagapiou, K.; Kanakis, C.; Nikola, O.; Lambrou, G. I.; Trigas, P.; Kanaka-Gantenbein, C.; Tarantilis, P. A.; *Appl. Sci.* **2023**, *13*, 4784. [Crossref]

40. Coyago-Cruz, E.; Salazar, I.; Guachamin, A.; Alomoto, M.; Cerna, M.; Mendez, G.; Heredia-Moya, J.; Vera, E.; *Antioxidants* **2025**, *14*, 299. [Crossref]
41. Abdelbaset, S.; Ayoub, I. M.; Mohamed, O. G.; Tripathi, A.; Eldahshan, O. A.; El-Kersh, D. M.; *BMC Complementary Med. Ther.* **2024**, *24*, 1. [Crossref]
42. Sánchez-Rabaneda, F.; Jáuregui, O.; Casals, I.; Andrés-Lacueva, C.; Izquierdo-Pulido, M.; Lamuela-Raventós, R. M.; *J. Mass Spectrom.* **2003**, *38*, 35. [Crossref]
43. Barbosa-Carvalho, A. P.; Texeira Costa Leitão, D. D. S.; Campos De Siqueira, F.; Zerlotti Mercadante, A.; Campos Chisté, R.; *J. Food Sci.* **2021**, *86*, 4045. [Crossref]
44. Aissat, A. K.; Chaher-Bazizi, N.; Richard, T.; Kilani-Atmani, D.; Pedrot, E.; Renouf, E.; Atmani, D.; Valls Fonayet, J.; *J. Food Compos. Anal.* **2022**, *106*, 104286. [Crossref]
45. Chen, G.; Netzel, M. E.; Mantilla, S. M. O.; Phan, A. D. T.; Netzel, G.; Sivakumar, D.; Sultanbawa, Y.; *Molecules* **2023**, *28*, 1608. [Crossref]
46. Slimestad, R.; Johny, A.; Thomsen, M. G.; Karlsen, C. R.; Rosnes, J. T.; *Molecules* **2022**, *27*, 7335. [Crossref]
47. Masala, V.; Jokić, S.; Aladić, K.; Molnar, M.; Tuberoso, C. I. G.; *Molecules* **2024**, *29*, 2600. [Crossref]
48. Andrade-Pinheiro, J. C.; Sobral de Souza, C. E.; Ribeiro, D. A.; Silva, A. D. A.; da Silva, V. B.; dos Santos, A. T. L.; Juno Alencar Fonseca, V.; de Macêdo, D. G.; da Cruz, R. P.; Almeida-Bezerra, J. W.; Machado, A. J. T.; de Freitas, T. S.; de Brito, E. S.; Ribeiro, P. R. V.; da Costa, J. G. M.; Coutinho, H. D. M.; Kowalska, G.; Rowiński, R.; Kowalski, R.; Morais-Braga, M. F. B.; *Plants* **2023**, *12*, 415. [Crossref]
49. Hermund, D. B.; Larsen, L. K.; Trangbæk, S. R.; Madsen, Q. R. T.; Sørensen, A. M.; Kaya, J.; Jacobsen, C.; *J. Am. Oil Chem. Soc.* **2025**, *102*, 35. [Crossref]
50. Lu, L.; Deng, Y.; Li, J.; Feng, X.; Zou, H.; *Int. J. Mol. Sci.* **2025**, *26*, 5853. [Crossref]
51. Barros, H. R. D. M.; García-Villalba, R.; Tomás-Barberán, F. A.; Genovese, M. I.; *J. Funct. Foods* **2016**, *22*, 477. [Crossref]
52. Zhang, B.; Dong, Y.; Yu, N.; Sun, Y.; Xing, Y.; Yang, F.; Yu, X.; Sun, W.; Sun, J.; Li, X.; Xiu, Z.; *J. Funct. Foods* **2019**, *54*, 53. [Crossref]
53. Andrade, J. K. S.; Barros, R. G. C.; Pereira, U. C.; Gualberto, N. C.; de Oliveira, C. S.; Shanmugam, S.; Narain, N.; *Food Chem.* **2022**, *373*, 131494. [Crossref]
54. Rodríguez, M. S.; Martínez León, A. J.; Porras, L. S.; Giraldo, I. A.; Rojas, E.; Lavao, F. E.; Martínez, K.; *Plants* **2025**, *14*, 2632. [Crossref]
55. Yang, X.; Shi, J.; Li, H.; Zhang, K.; Li, J.; Song, Q.; *Anal. Bioanal. Chem.* **2023**, *415*, 6511. [Crossref]
56. Ahmad, W.; Ali, A.; Mohsin, A.; Ji, X.; Aziz, M.; Wang, L.; Zhao, L.; *Food Meas. Charact.* **2024**, *18*, 1302. [Crossref]
57. Qin, S.; Wang, W.; Li, D.; Pu, Y.; Cui, T.; Wang, W.; *J. Pharm. Biomed. Anal.* **2023**, *234*, 115509. [Crossref]
58. King, S.; Campbell, J.; Rowe, R.; Daly, M.; Moncrieff, G.; Maybury, C.; *J. Cosmet. Dermatol.* **2023**, *22*, 2650. [Crossref]
59. Pena, L. S. C.; Santos, J. R.; Caetité, A. R. M.; *Res. Soc. Dev.* **2022**, *11*, e58711730594. [Crossref]
60. Aryal, B.; Adhikari, B.; Aryal, N.; Bhattarai, B. R.; Khadayat, K.; Parajuli, N.; *BioMed Res. Int.* **2021**, *2021*, 1. [Crossref]
61. Apaydın, D.; *Food Meas. Charact.* **2024**, *18*, 5297. [Crossref]
62. Vargas-Arana, G.; Merino-Zegarra, C.; Tang, M.; Pertino, M. W.; Simirgiotis, M. J.; *Antioxidants* **2022**, *11*, 595. [Crossref]
63. Abdul Karim, A.; Azlan, A.; Ismail, A.; Hashim, P.; Abd Gani, S. S.; Zainudin, B. H.; Abdullah, N. A.; *BMC Complementary Altern. Med.* **2014**, *14*, 381. [Crossref]
64. Pugliese, A. G.; Tomas-Barberan, F. A.; Truchado, P.; Genovese, M. I.; *J. Agric. Food Chem.* **2013**, *61*, 2720. [Crossref]
65. Salehi, B.; Venditti, A.; Sharifi-Rad, M.; Kręgiel, D.; Sharifi-Rad, J.; Durazzo, A.; Lucarini, M.; Santini, A.; Souto, E. B.; Novellino, E.; Antolak, H.; Azzini, E.; Setzer, W. N.; Martins, N.; *Int. J. Mol. Sci.* **2019**, *20*, 1305. [Crossref]
66. Guan, S.; Sun, L.; Wang, X.; Huang, X.; Luo, T.; *Evidence-Based Complementary Altern. Med.* **2022**, *2022*, 1. [Crossref]
67. Evans, J. A.; Mendonca, P.; Soliman, K. F. A.; *Nutrients* **2022**, *14*, 2228. [Crossref]
68. Shin, J. H.; Shin, S. H.; *J. Microbiol. Biotechnol.* **2024**, *34*, 2425. [Crossref]
69. Tang, E.; Lin, H.; Yang, Y.; Xu, J.; Lin, B.; Yang, Y.; Huang, Z.; Wu, X.; *Front. Nutr.* **2024**, *11*, 1481203. [Crossref]
70. Cen, C.; Li, J.; Zhou, P.; Fisher, D.; Hien, N. T. T.; Musabaev, E.; Pronyuk, K.; Zhao, L.; *Front. Pharmacol.* **2025**, *16*, 1648614. [Crossref]
71. Docampo, M.; Olubu, A.; Wang, X.; Pasinetti, G.; Dixon, R. A.; *J. Agric. Food Chem.* **2017**, *65*, 7607. [Crossref]
72. Day, A. J.; Bao, Y.; Morgan, M. R. A.; Williamson, G.; *Free Radical Biol. Med.* **2000**, *29*, 1234. [Crossref]
73. Qiu, X.; Huang, R.; Xie, J.; Luo, S.; Cheng, X.; Cui, J.; Hu, D.; *Biology* **2025**, *14*, 637. [Crossref]
74. Shahbaz, M.; Imran, M.; Momal, U.; Naeem, H.; Alsagaby, S. A.; Al Abdulmonem, W.; El-Ghorab, A. H.; Ghoneim, M. M.; Hussain, M.; Shaker, M. E.; Abdelgawad, M. A.; Mujtaba, A.; Jbawi, E. A.; *Food Agric. Immunol.* **2023**, *34*, 2265690. [Crossref]
75. Tazon, A. W.; Awwad, F.; Meddeb-Mouelhi, F.; Desgagné-Penix, I.; *BioChem* **2024**, *4*, 323. [Crossref]
76. Wang, Z.; Liu, T.; Lu, B.; Zhang, J.; Zhan, J.; Liu, X.; Wang, M.; Zhang, J.; *J. Mol. Struct.* **2025**, *1348*, 143558. [Crossref]
77. Zaborowski, M. K.; Dlugosz, A.; Blaszak, B.; Szulc, J.; Leis, K.; *Molecules* **2024**, *29*, 3206. [Crossref]
78. Chen, L.; Wang, F.; Hu, X.; Li, N.; Gao, Y.; Xue, F.; Xie, L.; Xie, M.; *J. Cosmet. Dermatol.* **2025**, *24*, e70024. [Crossref]

Submitted: November 2, 2025
Final version online: April 16, 2026