

UPLC-QTOF-MS^E-Based Metabolic Profile to Screening Candidates of Biomarkers of Dwarf-Cashew Clones Resistant and Susceptible to Anthracnose (*Colletotrichum gloeosporioides* (Penz.) Penz. & Sacc.)

Priscilla C. C. Alves,^{#,a} Jhonyson A. C. Guedes,^{id #,b,c} Luiz A. L. Serrano,^c
Marlon V. V. Martins,^c Debora B. de Sousa,^b Gisele S. Silva,^{id c} Paulo R. V. Ribeiro,^c
Dávila S. Zampieri^b and Guilherme J. Zocolo^{id *,c}

^aDepartamento de Química Analítica e Físico-Química, Universidade Federal do Ceará,
Av. Humberto Monte, s/n, Pici, 60455-760 Fortaleza-CE, Brazil

^bDepartamento de Química Orgânica e Inorgânica, Universidade Federal do Ceará,
Av. Humberto Monte, s/n, Pici, 60455-760 Fortaleza-CE, Brazil

^cEmbrapa Agroindústria Tropical, Rua Dra. Sara Mesquita, 2270, Pici, 60020-181 Fortaleza-CE, Brazil

Investigating specialized plant metabolites, traditionally referred to as ‘secondary metabolites’ present in leaf extracts of cashew trees (*Anacardium occidentale*) resistant and susceptible to anthracnose disease was carried out using metabolomics combined with chemometric tools. We used clones of dwarf-cashew with the following variations of characteristics: resistant and healthy (CCP 76, BRS 226, BRS 189), susceptible and healthy (BRS 265), and another clone also susceptible but affected by the disease (BRS 265). The UPLC-QTOF-MS^E (ultra performance liquid chromatography coupled with quadrupole time-of-flight mass spectrometry) analysis of the leaves of *Anacardium* clones allowed us to annotate a total of 39 metabolites. The multiple orthogonal partial discriminant analysis of the least-squares (OPLS-DA) allowed twelve metabolites to be potential biomarkers of differentiation among the clones studied. Namely, the triene-(17:3)-anacardic acid found on CCP 76 and BRS 189 clones, respectively, is the main candidate for biomarker of resistance. While catechin, B-type procyanidin isomers, and procyanidin dimer monogallate identified in BRS 265 are significant potential biomarkers of susceptibility.

Keywords: *Anacardium occidentale*, *Colletotrichum gloeosporioides*, chemometric analysis, dwarf-cashew clone, clone resistance, chromatographic analysis

Introduction

The cashew tree (*Anacardium occidentale* L., Anacardiaceae family) is an important economic plant native to Brazil.¹ In 2020, 4.2 million tons of cashew nuts were produced all over the world, with the most prominent producers being the African continent (58.4%), Asia (37.9%), and Brazil (3.7%).²

Anthracnose (*Colletotrichum gloeosporioides* (Penz.) Penz. & Sacc.) is currently the main disease of cashew in the producing area of West Africa and recent areas of Central America. In 2012, the fungus *Colletotrichum* spp. was recognized as one of the 10 most critical pathogenic fungi in the world due to its scientific importance and

relevance in terms of the economic losses it causes in the orchards it affects.³ In Brazil, anthracnose is among the main phytopathologies that affect the cashew crop, compromising the productivity of orchards.⁴ The fungus attacks both young and adult plants, causing damage to leaves, cashew nuts, peduncles, and inflorescences, an act that leads to significant production losses.^{5,6} The months of August and September are crucial for the aggravation of the disease in Brazilian orchards due to sporadic rains, a fact that favors the reproduction and dispersion of the pathogen in the field.⁴

Cashew cultivation occurs in predominantly tropical regions (between latitudes 30° N and 30° S), where high temperatures and high humidity favor the high incidence of anthracnose. Moreover, because of the introduction of improved cashew clones in extensive areas of monocultures, phytosanitary issues became more significant for the cashew production system, mainly in the coastal regions

*e-mail: guilherme.zocolo@embrapa.br

[#]These authors contributed equally to this work.

Dedicated to Prof Carol H. Collins.

Editor handled this article: Andréa R. Chaves (Associate)

of Northeast Brazil. Epidemics have become more frequent in the cultivated field and successful strategies for control and prevent these infectious diseases have created new challenges in the face of aggressive pathogens and host vulnerability.⁷

In Brazil, there are numerous species of *Colletotrichum* associated with anthracnose in the fruit of economic relevance, such as cashew,⁸ grape,⁹ mango,¹⁰ guavas, and papaya.¹¹ Furthermore, anthracnose has been associated with a loss of productivity in different species of plants. In 2000, anthracnose disease in the cashew tree was responsible for a 40% loss in cashew tree yield in Brazil.⁸

Recently, the genotypic strategy has been explored to improve the cashew crop in Mozambique¹² and Brazil.¹³⁻¹⁵ In the case of Brazil, anthracnose was considered the main disease of the cashew until the last decade, so the national breeding programs were directed to obtain clones resistant to this disease. Resistant dwarf-cashew clones were introduced in the northeast region since the 1980s.¹³ In general, the anthracnose resistance factor was reported as a result of high genetic variability in dwarf-cashew clones. Successes were obtained with several clones, but one of the most productive, BRS 265, has been shown susceptible.⁴

Although there are records of fungicides for the control of anthracnose,¹⁶ such as copper oxychlorate, copper hydroxide, captafol, benomyl, ditianon, anilazine, bitertanol, among others,⁴ alternative management strategies should be sought. The use of resistant genotypes is one of the adequate strategies in managing diseases. The advantages of the genetic management strategy include obtaining pest and disease-resistant cashew trees, thus reducing the use of agrochemicals. In this way, more economical and healthy harvests are obtained for the consumption of the peduncle. There is also an increase in productivity throughout the year, even in long periods of drought.¹³ Another important factor is that low cashew trees facilitate the harvesting of the fruit; therefore, allowing better use of the pseudo fruit. In addition, they corroborate to increase the number of plants *per* area and also in the size of the cashew nut.¹⁷

In the context of the evaluation of clones resistant to anthracnose, metabolomics is a powerful tool. Metabolic fingerprint has been widely applied in evaluating food and crop quality.¹⁸ Moreover, identifying genetic modification's influence on plant metabolism provides valuable insights into genetic improvement.¹⁹

Thus, considering the potential for resistance against pathogens of dwarf-cashew clones, together with the high cashew economic perspective as a commercial commodity, it is an emerging need in agribusiness to develop innovative, simple, and inexpensive tools to assist in the selection of

cashew clones that effective increase in annual productivity and quality of cashew cultivation. Thus, we performed a comparative analysis of the metabolites present in the leaves of different health and disease clones of *A. occidentale* using the UPLC-QTOF-MS^E (ultra performance liquid chromatography coupled with quadrupole time-of-flight mass spectrometry) technique, combined with multivariate data analysis models (principal component analysis (PCA), orthogonal partial least squares discriminatory analysis (OPLS-DA), and S-Plot) to find possible differences in the chemical profile associated with susceptibility and resistance to anthracnose disease.

Experimental

Plant material

Comparative studies of the metabolomics profiles were conducted from leave extracts of resistant early *A. occidentale* clones, encoded as CCP 76 (C1), BRS 226 (C2), BRS 189 (C3), and a susceptible anthracnose clone, encoded as BRS 265 (C4_SH for the clone susceptible and healthy) and BRS 265 (C4_SD for the clone susceptible and diseased). The leaves were collected at the experimental field of Embrapa, located in the municipality of Pacajus (Ceará state, Brazil, latitude: 04°10'21" S, longitude: 38°27'38" W) on 2016 July 08 between 9 a.m. and 10 a.m. The orchard with cashew trees was five years old, and no application of pesticides was carried out during the analysis year to avoid external interference. The access to Genetic Resources was registered in the Genetic Heritage Management Council (Conselho de Gestão do Patrimônio Genético-CGEN) under the code AF91C72.

The orchard from which the samples were collected consisted of 16 rows with 30 dwarf-cashew in each line. The 16 rows of the orchard were divided into four blocks (randomized block design), with each row consisting of one clone of dwarf-cashew of the Embrapa: C1, C2, C3 and C4. In each row of the orchard, six plants were sampled, with two healthy leaves being removed from each plant quadrant, totaling eight leaves *per* plant. Since 6 plants were sampled *per* row of the orchard and 8 leaves were collected from each plant, we have a total of 48 leaves *per* clone sample in each block (6 plants × 8 leaves). Finally, considering the 4 blocks, we have a total of 192 leaves collected *per* clone.

The leaves were removed from branches without inflorescence and corresponded to the first pair of mature leaves from the apex of the branch. Additionally, samples of leaves attacked by the anthracnose, naturally infected by the pathogen *Colletotrichum goe孢子oides*, were removed

from the susceptible clone C4 (BRS 265). BRS 265 clone was the only one to present the disease within the group analyzed. The collected leaves were submitted to the paralysis of their metabolism through the process known as quenching, in which the leaves are immersed in liquid nitrogen and crushed in a porcelain crucible. The crushed material was then oven-dried and dried at 40 °C for 72 h.

Chemicals

The solvents used were purchased from LiChrosolv[®] of the Sigma-Aldrich Chemical Company (St. Louis, MO, USA). High purity Milli-Q water (Billerica, MA, USA) was used for all methods. The (+)-catechin (Lot BCBF0735V) and the Procyanidin B2 standards (Lot 13021848) were acquired from Sigma (St. Louis, USA) and Extrasynthese (Lyon, France), respectively.

Preparation of extracts

The extraction of the organic compounds of *A. occidentale* was performed using an adapted methodology, previously reported.²⁰⁻²²

Initially, 50 mg of the cashew leaves were weighed separately and transferred to test tubes. Next, 4 mL of hexane were added to each test tube and the mixture was homogenized in a vortex system for 1 min. The non-polar extraction was performed in an ultrasonic bath at a fixed power of 135 W for 20 min. Subsequently, the polar metabolites were extracted from the same samples extracted initially by hexane using 4 mL of ethanol/water solution (7:3) under conditions similar to the previous procedure. Thus, the obtained mixture was again homogenized in a vortex system and taken to the ultrasonic bath. The test tubes were subjected to centrifugation at 3000 rpm for 10 min to remove the suspended plant material. Lastly, 1 mL of the polar fraction was filtered (0.22 µm poly(tetrafluoroethylene) (PTFE) filter), collected, and stored in vials for UPLC-QTOF-MS^E. The extraction procedure was performed on twelve biological replicates for each of the five-set of samples (C1, C2, C3, C4_SH, and C4_SD). In addition, blank extraction was performed in quintuplicate, where solvents were added to a test tube with subsequent execution of the extraction methodology described above, n = 65 extractions.

Analysis by UPLC-QTOF-MS^E

The sample solutions were analyzed in Waters Acquity equipment of ultra performance liquid chromatography (UPLC, Waters, Milford, MA, USA) coupled to a

Quadrupole/TOF system (QTOF, Waters, Milford, MA, USA). The UPLC analysis conditions include using Waters Acquity UPLC BEH column (150 mm × 2.1 mm, 1.7 µm), at a fixed temperature of 40 °C. An exploratory gradient using water (A) and acetonitrile (B) (both with 0.1% formic acid) as mobile phases varying from 2 to 95% B (0-15 min), flow rate of 0.4 mL min⁻¹ and injection volume of 5 µL was the method adopted for the analysis of the metabolic profiles of the extracts of *A. occidentale*. The chemical profiles were determined by coupling the Waters ACQUITY UPLC system to a QTOF mass spectrometer (Waters, Milford, MA, USA) with the electrospray ionization interface (ESI) in negative ionization mode. The ESI⁻ modes were acquired in the range of 110-1200 Da in MS and 50-1200 in MS/MS, with a fixed source temperature of 120 °C and a desolvation temperature of 350 °C. A desolvation gas flow of 500 L h⁻¹ for the ESI⁻ mode. The capillary voltage was 3 kV. Leucine enkephalin was used as a lock mass. The spectrometer operated with MS/MS centroid programming using a tension ramp from 20 to 40 V. The instrument was controlled by the MassLynx 4.1 software program (Waters Corporation, USA).

Chemometric data analysis

UPLC-QTOF-MS^E data were analyzed using the MarkerLynx XS v4.1 software²³ program (Waters Corporation) to identify potential discriminatory biomarkers from leaves of *A. occidentale*. For data collection, the method's parameters were defined at a retention time interval of 0.70-17.0 min, a mass range of 110-1200 Da, a mass tolerance of 0.05 Da, and a noise elimination level set at 5. A list was generated with the identification of the peaks detected using the pairs retention time (t_R)-mass data (m/z). An arbitrary identification was assigned to each of these pairs (t_R)-(m/z) based on their elution order from the UPLC system. Ion identification was based on t_R and m/z values compared to previously published data and analytical standards. The ion intensities for each detected peak were normalized against the sum of the peak intensities within that sample using the MarkerLynx XS v4.1 software program. Ions from different samples were identical when t_R and m/z values were matched. The raw data were submitted to PCA using the Pareto model. OPLS-DA was used to validate the PCA model and identify the differential metabolites. S-Plot selected the biomarkers. Comparative analysis of the specialized metabolites present in *Anacardium* leaves, using the UPLC-QTOF-MS^E data is combined with multivariate data analysis models, PCA, and OPLS-DA, variability in the variable importance in projection (VIP) and S-Plot. The significant differential

retention time and exact mass pairs in the S-Plots were selected and imported back into MarkerLynx XS for compound identification, after filtering using analysis of variance (ANOVA) p value ≤ 0.05 and VIP > 5 .

Results and Discussion

Chromatographic analysis

The liquid-liquid microextraction method (EtOH:H₂O, 7:3 (v/v)) was employed to obtain leaf extracts from anthracnose-resistant and susceptible dwarf-cashew clones, both healthy and unhealthy. In total, five sets of dwarf-cashew clone samples were investigated. These included four healthy lineages named C1, C2, C3, C4_SH, and a single diseased clone called C4_SD. Subsequently, the chemical profiles of the cashew leaves were examined using UPLC-QTOF-MS^E in negative mode (ESI⁻). The UPLC system offers several advantages, including improved resolution, increased sensitivity, faster analysis times, reduced solvent consumption, and broader applicability. The chromatographic separation was carried out in a way that an exploratory gradient proved to be the best option given the high complexity of the sample, which showed a rich profile in tannins and other difficult-to-separate classes, such as catechins and anacardic acids (Figure 1).

The polar and hydrophilic nature of sugars results in weaker interactions with the hydrophobic stationary phase, leading to faster elution times. The addition of formic acid as an organic modifier aid in the protonation of sugars, contributing to their separation and resolution. Organic acids exhibit stronger interactions with the stationary phase than sugars, leading to longer retention times. Hydrolysable tannins, catechins, and flavonoids display various polarities and hydrophobic interactions with the C18 stationary phase, affecting their elution behavior. Figure 1 illustrates the structural complexity of various compounds, showing that the elution order can be influenced by these factors. In the study, it was observed that higher molecular weight compounds, such as galloylated flavonoids, elute later than their non-galloylated counterparts due to the increased hydrophobic interactions caused by the presence of galloyl groups. Catechins are a type of flavanol, which are known for their hydroxyl groups and hydrophilic nature. The presence of multiple hydroxyl groups in catechins leads to strong hydrogen bonding with polar solvents in the mobile phase, resulting in relatively faster elution times. However, as the structural complexity and hydrophobicity of these compounds increase, they tend to interact more strongly with the hydrophobic stationary phase, leading to longer retention times. Flavonoids, on the other hand, can be glycosylated, meaning they have sugar moieties attached to their structures.

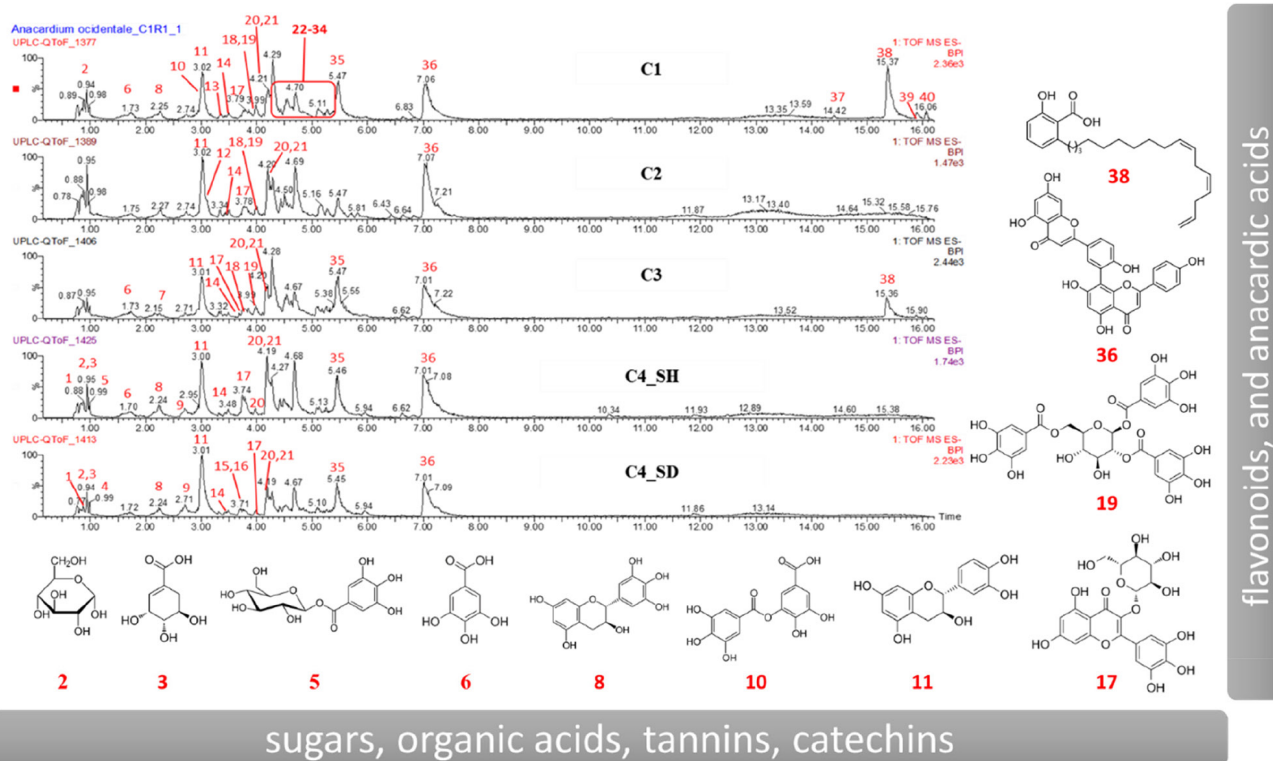


Figure 1. Representative chromatograms of metabolic profiles derived from leaf extracts, as well as some chemical structures of dwarf-cashew clones (*A. occidentale*), are presented in the negative ionization mode (ESI⁻). This includes both healthy plants (C1, C2, C3, C4_SH) and the unhealthy plant (C4_SD).

These sugar groups contribute to increased polarity and hydrophilic properties. In general, glycosylated flavonoids elute earlier than their non-glycosylated counterparts due to their stronger interactions with the polar solvents in the mobile phase. Additionally, the presence of glycosidic groups can enhance the chromatographic separation by providing further selectivity based on the specific sugar moieties and the positions at which they are linked to the flavonoid core.

When considering the interrelationship between catechins, flavonoids, and glycosylated flavonoids in chromatographic separation, it is essential to understand that the elution order is influenced by a combination of factors, including molecular weight, structural complexity, and the presence of functional groups.

Finally, the hydrophobic interactions between anacardic acids and the C18 stationary phase are enhanced by the presence of formic acid as an organic modifier, which improves peak shape, reduces peak broadening, and

promotes the ionization equilibrium and separation of these compounds. It is important to consider the influence of the aliphatic side chain length and varying number of unsaturations in anacardic acid molecules, such as (15:3), (17:3), (15:1), and (17:2). The aliphatic side chain length and the number of unsaturations in anacardic acid molecules play a significant role in determining their hydrophobic interactions with the stationary phase. Compounds with longer aliphatic side chains exhibit stronger hydrophobic interactions, leading to longer retention times. On the other hand, the presence of unsaturations in the side chains can affect the overall hydrophobicity of the anacardic acids. Molecules with a higher number of unsaturations generally exhibit weaker hydrophobic interactions due to increased polarity, resulting in shorter retention times compared to their saturated counterparts, this explains the elution order presented in Table 1, where the retention time (t_R) of anacardic acid (15:3) < anacardic acid (17:3) < anacardic

Table 1. Annotated metabolites from the leaves extracts of dwarf-cashew clones (*A. occidentale*): healthy plants (C1, C2, C3, C4_SH) and the unhealthy plant (C4_SD)

Compound	t_R^a / min	[M – H] ⁻ calculated	[M – H] ⁻ observed	(MS/MS)	Molecular formula	Error / ppm	Peak annotation	CCP76 (C1) ^b	BRS226 (C2) ^b	BRS189 (C3) ^b	BRS265 C4_SH ^c	BRS265 C4_SD ^d	Reference
1	0.87	377.0873	377.0836	191.0555; 267.0721; 165.0398	C ₁₈ H ₁₈ O ₉	-9.8	unknown	-	-	-	+	+	
2	0.93	341.1084	341.1083	179.0550; 161.0450; 143.0354	C ₁₂ H ₂₂ O ₁₁	-0.3	α-D-glucopyranoside	+	-	-	+	+	24
3	0.96	173.0450	173.0433	165.0397; 111.0480	C ₇ H ₁₀ O ₅	-9.8	shikimic acid	-	+	-	+	+	25
4	0.98	191.0192	191.0181	173.0407; 111.0100	C ₆ H ₈ O ₇	-5.8	citric acid	-	-	-	-	+	22
5	0.99	331.0665	331.0660	169.0141; 125.0222	C ₁₃ H ₁₆ O ₁₀	-1.5	galloyl-hexose	-	-	-	+	-	26
6	1.71	169.0137	169.0141	125.0244	C ₇ H ₆ O ₅	2.4	gallic acid	+	-	+	+	-	27
7	2.22	325.0560	325.0558	169.0133; 125.0234	C ₁₄ H ₁₄ O ₉	-0.6	galloyl shikimic acid	-	-	+	-	-	28
8	2.25	305.0661	305.0672	219.0645; 179.0369; 165.0190; 139.0390; 137.0242	C ₁₅ H ₁₄ O ₇	3.6	galocatechin or epigallocatechin	+	-	-	+	+	29
9	2.74	577.1371	577.1366	451.1066; 425.0876; 407.0826; 289.0735	C ₃₀ H ₂₆ O ₁₂	3.5	B-type procyanidin dimer	-	-	-	+	+	30
10	2.96	321.0247	321.0270	169.0126; 125.0238	C ₁₄ H ₁₀ O ₉	7.2	digallic acid	+	-	-	-	-	26
11	3.01	289.0712	289.0705	245.0827; 205.0511	C ₁₅ H ₁₄ O ₆	-2.4	catechin	+	+	+	+	+	analytical standard
12	3.09	453.1033	453.1048	313.0577; 179.0357; 169.0161; 125.0243	C ₂₀ H ₂₂ O ₁₂	3.3	hydroxy-methoxyphenyl-O-(O-galloyl)-hexose	+	+	+	-	-	26

Table 1. Annotated metabolites from the leaves extracts of dwarf-cashew clones (*A. occidentale*): healthy plants (C1, C2, C3, C4_SH) and the unhealthy plant (C4_SD) (cont.)

Compound	t_R^a / min	[M – H] ⁻ calculated	[M – H] ⁻ observed	(MS/MS)	Molecular formula	Error / ppm	Peak annotation	CCP76 (C1) ^b	BRS226 (C2) ^b	BRS189 (C3) ^b	BRS265 C4_SH ^c	BRS265 C4_SD ^d	Reference
13	3.35	635.0884	635.0923	465.0686; 313.0607; 169.0130; 125.0231	C ₂₇ H ₂₄ O ₁₈	6.1	trigalloyl glucose	+	–	–	–	–	31,32
14	3.51	457.0771	457.0772	169.0122; 125.0241	C ₂₂ H ₁₈ O ₁₁	0.2	epigallocatechin gallate	+	+	+	+	+	31,32
15	3.69	577.1346	577.1325	451.1097; 425.0824; 407.0746; 289.0709	C ₃₀ H ₂₆ O ₁₂	–3.6	B-type procyanidin dimer	–	–	–	+	+	29,30
16	3.74	729.1456	729.1520	407.0810; 289.0728	C ₃₇ H ₃₀ O ₁₆	8.8	procyanidin dimer monogallate	–	–	–	+	+	29
17	3.79	479.0826	479.0833	317.0307; 316.0222	C ₃₀ H ₂₆ O ₁₂	1.5	myricetin-3- <i>O</i> -glucoside	+	+	+	+	+	26
18	3.85	631.0935	631.0975	479.0862; 317.0317	C ₂₈ H ₂₄ O ₁₇	6.3	myricetin galloyl hexoside	+	+	+	–	–	22,28
19	3.98	787.0994	787.1033	635.0886; 617.0876; 465.0759; 169.0122; 125.0252	C ₃₄ H ₂₈ O ₂₂	5.0	tetra- <i>O</i> -galloyl-glucoside	+	+	+	+	–	27
20	4.17	463.0877	463.0880	317.0286; 316.0216	C ₂₁ H ₂₀ O ₁₂	–0.6	myricetin-3- <i>O</i> -rhamnoside	+	+	+	+	+	33
21	4.18	463.0877	463.0880	301.0257	C ₂₁ H ₂₀ O ₁₂	–0.6	quercetin-3- <i>O</i> -galactoside	+	+	+	+	+	27,33
22	4.23	463.0877	463.0858	301.0317	C ₂₁ H ₂₀ O ₁₂	–4.1	quercetin-3- <i>O</i> -glucoside	+	+	–	+	+	27,33
23	4.25	615.0986	615.0997	463.0883; 301.0328	C ₂₈ H ₂₄ O ₁₆	1.8	quercetin galloyl hexoside	+	+	+	+	+	28
24	4.28	939.1104	939.1106	770.1005; 769.0918; 617.0865; 169.0112; 125.0230	C ₄₁ H ₃₂ O ₂₆	0.2	penta- <i>O</i> -galloyl-glucoside	+	+	+	+	+	27
25	4.43	433.0771	433.0774	301.0350	C ₂₀ H ₁₈ O ₁₁	0.7	quercetin-3- <i>O</i> -xyloside	+	+	–	+	+	34
26	4.49	447.0927	447.0915	285.0402; 284.0315	C ₂₁ H ₂₀ O ₁₁	–2.7	kaempferol-3- <i>O</i> -glucoside	–	–	–	+	+	33,34
27	4.49	433.0771	433.0771	301.0350	C ₂₀ H ₁₈ O ₁₁	0.0	quercetin-3- <i>O</i> -arabinopyranoside	–	+	–	+	+	33,34
28	4.54	433.0771	433.0771	301.0361	C ₂₀ H ₁₈ O ₁₁	0.0	quercetin-3- <i>O</i> -arabinofuranoside	–	+	–	+	–	33
29	4.55	447.0927	447.0920	285.0386; 284.0287	C ₂₁ H ₂₀ O ₁₁	–1.6	kaempferol-3- <i>O</i> -glucoside dimer	+	+	+	+	+	35
30	4.57	1091.1213	1091.1285	939.1167; 769.0932; 617.0867; 169.0123	C ₄₈ H ₃₆ O ₃₀	6.6	hexagalloyl hexoside	+	+	+	+	+	22,26
31	4.59	599.1037	599.1074	285.0398; 169.0123; 125.0252	C ₂₈ H ₂₄ O ₁₅	6.2	cyanidin-3- <i>O</i> -(2'' galloyl)-galactoside	–	+	–	+	–	26
32	4.69	447.0927	447.0909	301.0343	C ₂₁ H ₂₀ O ₁₁	–4.0	quercetin-3- <i>O</i> -rhamnoside	+	+	+	+	+	26,36
33	4.75	585.0880	585.0923	301.0341; 169.0125	C ₂₇ H ₂₂ O ₁₅	7.3	quercetin galloyl pentoside isomer	+	+	+	+	+	28

Table 1. Annotated metabolites from the leaves extracts of dwarf-cashew clones (*A. occidentale*): healthy plants (C1, C2, C3, C4_SH) and the unhealthy plant (C4_SD) (cont.)

Compound	t _R ^a / min	[M – H] ⁻ calculated	[M – H] ⁻ observed	(MS/MS)	Molecular formula	Error / ppm	Peak annotation	CCP76 (C1) ^b	BRS226 (C2) ^b	BRS189 (C3) ^b	BRS265 C4_SH ^c	BRS265 C4_SD ^d	Reference
34	5.27	585.0880	585.0912	301.0350; 169.0152	C ₂₇ H ₂₂ O ₁₅	5.5	quercetin galloyl pentoside isomer	–	+	–	–	–	28
35	5.44	349.0560	349.0562	198.0495; 197.0426; 169.0128; 124.0166	C ₁₆ H ₁₄ O ₉	0.6	ethyl 2,4-dihydroxy-3-(3,4,5-trihydroxybenzoyl)	+	+	+	+	+	27
36	7.02	537.0822	537.0809	417.0616; 375.0501	C ₃₀ H ₁₈ O ₁₀	–2.4	amentoflavone or agatisflavone	+	+	+	+	+	26
37	14.41	341.2117	341.3125	297.2204; 119.0514; 106.0428	C ₂₂ H ₃₀ O ₃	5.3	anacardic acid (15:3)	+	–	–	–	–	28
38	15.38	369.2430	369.2406	325.2503	C ₂₄ H ₃₄ O ₃	–6.5	anacardic acid (17:3)	+	–	+	–	–	37
39	15.90	345.2430	345.2428	301.2550	C ₂₂ H ₃₄ O ₃	–0.6	anacardic acid (15:1)	+	–	–	–	–	28
40	16.06	371.2586	371.2582	327.2686	C ₂₄ H ₃₆ O ₃	–1.1	anacardic acid (17:2)	+	–	–	–	–	28

^aRetention time; ^bC1, C2, C3 (resistant and healthy); ^cC4_SH (susceptible and healthy); ^dC4_SD (susceptible and unhealthy).

acid (15:1) < anacardic acid (17:2), following a pattern consistent with the literature.²⁸

Annotation of metabolites

To accurately annotate the peaks identified in the chromatogram, only reference data related to the Anacardiaceae family, *Anacardium* genus, and *A. occidentale* species were considered. Specifically, all tribes within the *Anacardium* genus were taken into account: Anacardiaceae, Rhoaceae, Spondiadeae, Semecarpeae, and Dobineae.³⁸

In addition, to reduce the possibility of error in structural determination, we used extensive bibliographic research in different databases (SciFinder, ChemSpider and PubChem). The research of chemotaxonomic reference in comprehensive databases has been extensively explored in the literature,^{22,39–41} allowing us to increase the accuracy of annotation of metabolites in several plants.

Figure 1 shows the chromatograms for the five sets of samples analyzed in the negative mode (ESI⁻) by UPLC-QTOF-MS^E. Chromatograms were evaluated along with mass spectra, with the detection of forty distinct metabolites, of which thirty-nine metabolites were annotated (Table 1).

Among the compound class, many anacardic acids, flavonoid, and procyanidin were found. Epigallocatechin gallate (**14**), myricetin-3-*O*-glucoside (**17**), tetra-*O*-galloyl-glucoside (**19**), myricetin-3-*O*-rhamnoside (**20**), quercetin-3-*O*-galactoside (**21**), penta-*O*-galloyl-glucoside (**24**), kaempferol-3-*O*-glucoside dimer (**29**), hexagalloyl hexoside (**30**), quercetin-3-*O*-rhamnoside (**32**) and amentoflavone (**36**) were observed in all dwarf-cashew

clones. Additionally, digallic acid (**10**) was observed only in the healthy clone (C1), whereas citric acid (**4**) was observed only in the diseased clone (C4_SD).

In general, the chemical profile of healthy and diseased lineages of C4_SH and C4_SD, respectively, is very alike. However, some relevant differences noted are due to the presence or absence of quercetin and procyanidin compounds. It is observed clearly that compounds **23** and **25** are present only in the diseased clone, while compounds **5**, **6**, **19**, and **28** are present only in the healthy clone. It was noted that no anacardic acid was detected in healthy (C4_SH) and diseased (C4_SD) susceptible lineages.

The compound quercetin-3-*O*-arabinofuranoside (**28**) is prominent in the healthy clone, whereas B-type procyanidin dimer (**15**) and procyanidin dimer monogallate (**16**) are found in the unhealthy clone (C4_SD). Additionally, gallic acid (**6**) and citric acid (**4**) were found in the healthy and unhealthy clones (C4_SH and C4_SD), respectively. The descriptions of the identification of the leading chemical compounds that differentiate the profile of the analyzed dwarf-cashew clones are presented below.

Quercetin and procyanidin compound

The compounds **21**, **22**, **23**, **25**, **27**, **28**, **32**, **33** and **34** were identified as quercetin derivatives. Quercetin-3-*O*-galactoside (**21**) and quercetin-3-*O*-glucoside (**22**) present as precursor ions at *m/z* 463.0880 and 463.0858, respectively. Both fragments encountered are compatible with the loss of a hexose [M – H – 162]⁻ from the peak at *m/z* 301.0257 for the first and *m/z* 301.0317 for the second compound.^{27,33}

The glycosylated quercetin **23**, **33**, and **34** have also been annotated with the respective precursor ions at m/z 615.0997, 585.0923, 585.0912. The first molecule (**23**) is a quercetin galloyl hexoside, and the last two (**33** and **34**) are possibly quercetin galloyl pentoside isomers, respectively.

For these three metabolites, the typical presence of the quercetin skeleton (m/z 301.0328) was identified. Besides that, for compound **23**, it was observed the loss of mass of a galloyl group $[M - H - 152]^-$. Additionally, the identification of compounds **33** and **34** was possible due to the presence of two fragments at m/z 301.0350 and 169.0152, which correspond to quercetin and gallic acid, respectively.²⁸

The compounds **25**, **27**, and **28** were detected for common precursor ions at m/z 433.077. It was observed that these quercetins presented the same patterns of fragmentation and low mass error from 0.0 to 0.7 ppm. In this way, the attempt to identify isomers of quercetins was made based on the retention time, as reported in the literature.^{33,34} Due to a limitation of the MS technique, isomers cannot be unambiguously identified based only on MS data.⁴² Hence, the quercetin aglycone anion was identified by the presence of the fragment of quercetin. They were tentatively identified as quercetin-3-*O*-xyloside (m/z 433.0774), quercetin-3-*O*-arabinopyranoside (m/z 433.0771) and quercetin-3-*O*-arabinofuranoside (m/z 433.0771), respectively.

The quercetin rhamnoside (**32**) was verified as an isoflavone at 447.0909 Da, since the fragmentation pattern was the sequential loss of quercetin residue (m/z 301.0343) in accordance with the previous report.²⁶

Procyanidin compound

Procyanidins were also identified in this study; however, only in leaf extracts of dwarf-cashew C4_SH and C4_SD, the clone susceptible to anthracnose. The presence of precursor ions at m/z 577.1366 and 577.1325 led to the identification of compounds **9** and **15** as B-type procyanidin dimers, respectively.^{29,30} Additionally, The MS/MS spectrum on precursor ion at m/z 577 gave product ions characteristic of procyanidins $[M - H - 126]^-$ at m/z 451, a peak $[M - H - 152]^-$ at m/z 425, $[M - H - 170]^-$ and another peak at m/z 407.⁴³

The prominent ion at m/z 729.1520 (**16**) was possible to identify as procyanidin dimer monogallate due to the presence of essential fragments at m/z 407.0810 and 289.0728.^{29,44} To corroborate the identification of procyanidin dimer type B, the data obtained in UPLC-QTOF-MS^E were compared to the retention time

and chromatographic profile obtained from analyzes of procyanidin B2 and catechin standards. The chromatogram of procyanidin B2 and catechin has a prominent peak at m/z 577 and 289 (t_R 3.25 and 2.32 min, respectively).

It was observed that the retention time and mass fragmentation profile of the identified procyanidins (**9**, **15**, and **16**) in the extracts of cashew leaves resemble those of the standard used. Therefore, the identification of procyanidin derivatives successfully was confirmed based on patterns of procyanidin B2 and catechin.

Additionally, Figure 2 shows the mechanism of the fragmentation pattern of B-type procyanidin dimer (**16**) based on data reported in the literature that corroborate our study.⁴⁵⁻⁴⁸

Initially, it is proposed that the precursor ion detected as procyanidin dimer monogallate (**16**, m/z 729.1520) may undergo fragmentation leading to the formation of B-type procyanidin dimer (m/z 577.1325) and release of compound 3,4,5-trihydroxybenzoic acid ($C_7H_6O_3$) in its neutral or ionic form. Although precursor ion at m/z 577 was not identified from MS/MS spectra of **16**, it is commonly reported in the literature.^{29,44}

Thenceforth, the fragments illustrated are formed from various mechanisms such as retro-Diels-Alder (RDA) reactions, quinone methide reaction (QM), and heterocyclic ring B fragmentation (HRF).

The fragmentation pattern of procyanidin dimers **9** and **15** at m/z 577 gave product ions at m/z 289 through a QM, produced by cleavage in interflavanoid C-C linkage, which is a bond that interconnects the flavanol moieties.^{45,49}

In addition, the procyanidins **9** and **15** also presented fragments at m/z 451 $[M - H - 126]^-$ generated by the heterocyclic fission of the B ring and 1,3,5-trihydroxybenzene output; m/z 425 $[M - H - 125]^-$ is formed by RDA fragmentation of one unit of catechin. Lastly, the loss of group B in m/z 407.0746 $[M - H - 18]^-$ refers to the elimination of a water molecule of the F ring.

Finally, all procyanidins identified in this study (**9**, **15**, and **16**) exhibit the same fragment at m/z 289, which is typically formed via the QM mechanism. The precursor ion m/z 289 $[M - H]^-$ detected is evidence that the compounds **15** and **16** have in its scaffold catechin dimer.

Gallotannins

Compounds **7**, **12**, **13**, **18**, and **19** were found in dwarf-cashew clones resistant to anthracnose. Compounds **24** and **30** were already found in all clones targeted for investigation.

The compound **7**, present only in clone C3, was identified as galloyl shikimic acid due to the presence of the

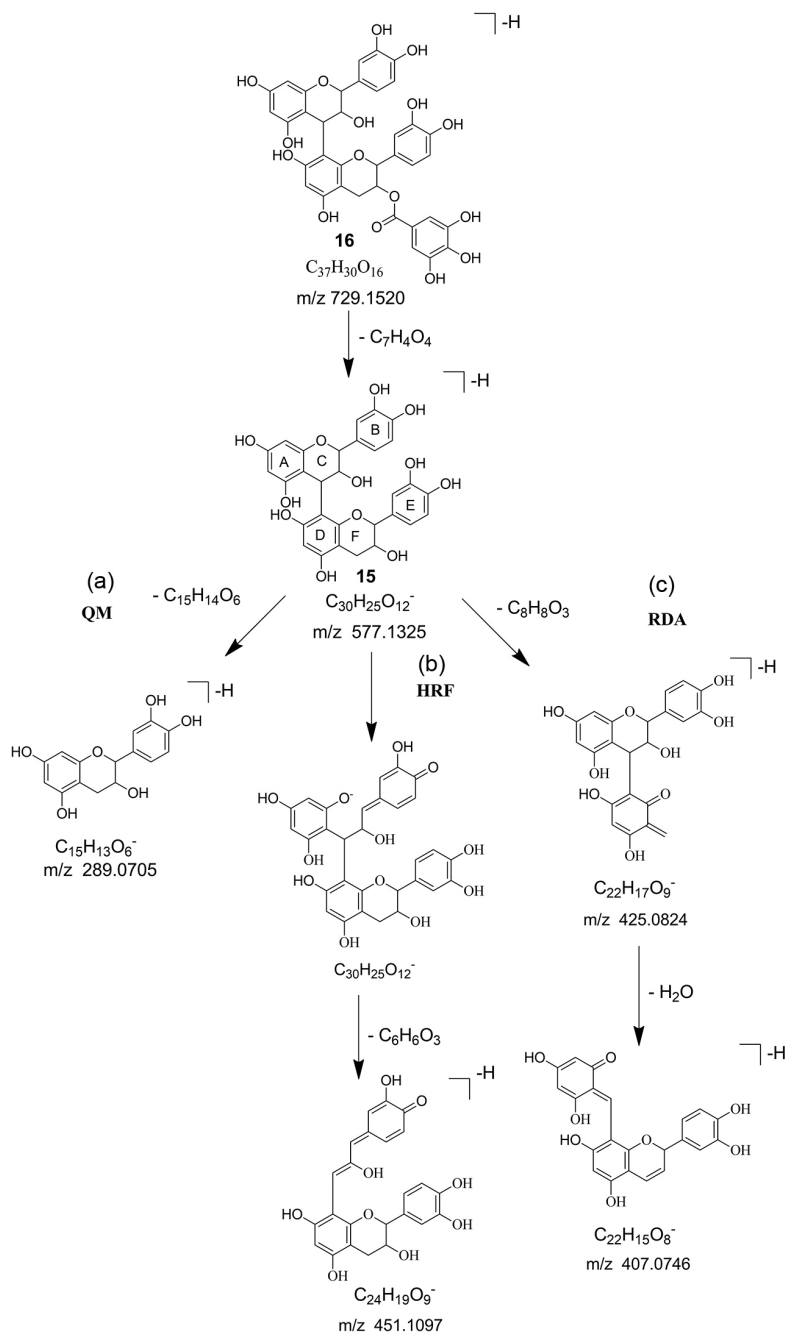


Figure 2. Proposed fragmentation pathways of B-type procyanidin dimer from procyanidin dimer monogallate: (a) quinone methide (QM) reaction, (b) heterocyclic ring fusion (HRF), and (c) retro-Diels-Alder (RDA) reactions.⁴⁵⁻⁴⁸

ion m/z 325.0558 $[M - H]^-$ and its fragments with values at m/z 169.0133 and 125.0234 given the loss of one molecule of CO_2 $[M - H - 156 - 44]^-$.²⁸

Hydroxy-methoxyphenyl-*O*-(*O*-galloyl)-hexose was attributed to compound **12**, which was present in clone C2. This identification was made possible by the presence of the precursor ion in m/z 453.1048. Through the analysis of the MS/MS spectrum, it was possible to identify fragments at m/z 313.0577, 179.0357, 169.0161, and 125.0243 as reported in the literature.²⁶

The precursor ion at m/z 635.0923 was identified as trigalloyl glucose (**13**). The presence of ions at m/z 465.0686, 313.0607, 169.0130, and 125.0231 in the MS/MS spectrum allied to the data reported in the literature confirm the identity of the compound.^{31,32}

Myricetin galloyl hexoside (**18**) was identified in clones C1, C2 and C3 and, through analysis of the obtained spectra, it was noted the presence of precursor ion at m/z 631.0975. In addition, it was identified two fragments mass at m/z 479.0862 and 317.0317 formed due to the

successive losses of galloyl $[M - H - 152]^-$ and hexose moiety $[M - H - 152 - 162]^-$, respectively.^{22,28}

Compounds **24** (m/z 939.1106) and **30** (m/z 1091.1285) were identified as penta-*O*-galloyl-glucoside and hexagalloyl hexoside, respectively. Hence, the two compounds showed similar ions fragments in m/z 769 and 617, corresponding to the loss of a neutral gallic acid molecule and a fraction of galloyl.^{22,26,27}

Anacardic acid

Four metabolites derived from anacardic acids with different side-chain tips were identified. They have fifteen or seventeen carbon atoms with varying degrees of unsaturation (monoene, diene, or triene). They elute at the end of the chromatographic run due to their lipophilic side-chain of alk(en)yl.²⁸ Compounds were determined according to the length of their side chain and the number of double bonds in the side chain.

The metabolite **37** was identified as anacardic acid (15:3) due to the presence of m/z at 341.3125. The MS/MS spectrum showed peaks in m/z 297.2204, 119.0514, and 106.0428. The fragments were produced by the loss of a CO_2 molecule from the phenolic carboxyl group by the fragmentation at the allyl position of the non-saturated anacardic acid and the removal of a phenyl group, respectively.

Compounds **38**, **39**, and **40** were identified according to their precursor ions, m/z 369.2406, 345.2428, and 371.2582, as anacardic acids 17:3, 15:1, and 17:2, respectively. Their spectra in MS/MS showed fragment ions $[M - H - 44]^-$ in common, indicating a pattern CO_2 loss of the phenolic carboxyl group.^{28,37}

Comparatively, it is noted that the resistant clone C1 showed all the anacardic acids identified in this study. On the other hand, a single anacardic acid (17:3) has also been identified in the resistant healthy clone C3. However, none of the anacardic acid was identified in the susceptible clone (C4), either healthy or diseased.

Multivariate analysis

Principal component analysis (PCA)

PCA represents an unsupervised linear mixture model that attempts to explain the variation within a set of data by a smaller number of components correlated.⁵⁰ Hence, data obtained from a PCA model indicate clustering trends within the data matrix.⁵¹⁻⁵³ Thus, the evaluation of the extent of differences or similarities of the dwarf-cashew sample sets was designed PCA.

The data processing by the PCA method revealed a

separation of the major components by a specific group. The generated model presented discrimination of the samples equal to 36.22% (R^2X [1] = 0.2257 and R^2X [2] = 0.1365, where the variable R^2X is called the explained variation, component [1] and [2]).

We observed the separation of the five groups of samples analyzed with the PCA, represented by C1, C2, C3, C4_SH, and C4_SD. However, by following PC1 (Figure 3), it is possible to verify a clear distinction between non-susceptible (C1, C2, and C3) and susceptible clones (C4_SH, and C4_SD). Since the set of samples that represent the non-susceptible clones (C1, C2, and C3) are in PC1 positive and the susceptible clones are in PC1 negative.

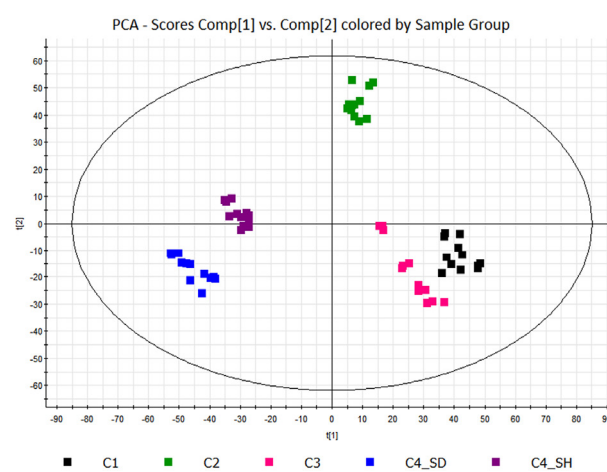


Figure 3. Score plot of the principal component analysis for *A. occidentale* clones: C1, C2, C3, C4_SH, and C4_SD.

It is important to note that the clusters between clone analyzes are promoted by the similarities and differences in the metabolic profiles of the different clones. Thus, considering that through PC1 it is possible to visualize and distinguish two large groups of samples that represent the non-susceptible and susceptible clones, we can infer that there are metabolic differences between the susceptible and non-susceptible clones.

Since, through the PCA, we verified that there are differences between the analyzed clones, we can now use appropriate statistical tools to determine the metabolites that contribute to conferring the clone's resistance to anthracnose to determine the candidates for biomarkers associated with resistance.

OPLS-DA analysis of leaf extracts of *A. occidentale*

To find the main candidates for biomarkers that may be associated with the diverse behaviors of the target plants, the extracts obtained from the collected samples were compared using an OPLS-DA model, with scores charts, and graphs of dispersion (S-Plot). The OPLS-DA

loadings plots, such as the S-Plot, enable the extraction of such statistically significant variables and identification of shared/unique structures in the samples.⁵¹⁻⁵⁴

To verify the accuracy and reliability of the OPLS-DA model, two parameters are used: the R²Y variable called explained variance, which provides a measure of model fit for the original data, plus the Q² variable, said predicted variance, and provides a measure of internal consistency between the original and predictive data of the cross-validation. The models with R²Y and Q² parameters close to 1 are considered excellent, although values above 0.5 are accepted when the components of the samples are highly complex.^{52,55}

The data were modeled by OPLS-DA (Figure 4) so that the groups of samples in which the clones represent them were compared in pairs, where the non-susceptible clones (C1, C2, and C3) were compared with the susceptible and diseased clones (C4_SD). Finally, it was also compared the two susceptible clones, diseased (C4_SD) with healthy (C4_SH). In this way, it is possible to infer that the sets of diseased and healthy samples are different. The values for both model quality parameters were satisfactory ((Figure 4a) R²Y = 0.99 and Q² = 0.98; (Figure 4b) R²Y = 0.99 and Q² = 0.99; (Figure 4c) R²Y = 0.99 and Q² = 0.98; (Figure 4d) R²Y = 0.99 and Q² = 0.97), suggesting that there is a statistically significant difference between the metabolic compositions of the analyzed samples.

The scatter plot (S-Plot) shows the integration and classification of variables with higher correlation and variance between the groups, showing the most relevant metabolites in the study. Each point refers to an ion in this graph, containing information about its retention times (t_R) and mass/charge ratio (m/z).^{52,56}

Figure 4 shows the S-Plot to analyze the variables responsible for separating the groups. The discriminant ions are highlighted in red; these ions were selected based on the variable importance in projection (VIP). Thus, ions with VIP > 5 were selected as the discriminating metabolites among the sample sets.

In the negative axis of the scatter plots S-Plot, Figures 4a, 4b, and 4c are the ions responsible for the discrimination of the resistant clones, whereas in the positive axis are those of the susceptible and diseased clone (C4_SD). In the comparison between plants of clone on Figure 4d, the candidates for biomarkers linked to the healthy plant (C4_SH) are on the negative axis, and the positive axis those related to the diseased plant (C4_SD). The identification of the ions was made through their retention time-mass/charge ratio (t_R - m/z), relating them to the previously identified compounds, Table 1.

Candidates for resistance biomarkers

The chemical compounds associated with resistance were determined through multivariate analysis, where non-susceptible clones were compared with a susceptible and diseased clone. Thus, using the S-Plot graphs and VIP > 5, it was possible to infer the candidates for biomarkers linked to resistance. In this context, Table 2 lists the potential biomarkers and the VIP values correlated with resistant, susceptible, and diseased clones, which were inferred from the S-Plot (Figure 4).

Through Table 2, we can infer that the metabolites hydroxy-methoxyphenyl-*O*-(*O*-galloyl)-hexose, myricetin galloyl hexoside, penta-*O*-galloyl-glucoside, anacardic acids (17:3; 17:2; 15:1), shikimic acid, quercetin galloyl pentoside isomer and tetra-*O*-galloyl-glucoside may be associated with resistance to anthracnose. Since, according to OPLS-DA and S-Plot graphs, these are the metabolites responsible for distinguishing the groups of samples from the non-susceptible clones (C1, C2, and C3) of the susceptible and diseased clone (C4_SD), Figure 4.

In order to corroborate with the observations obtained from the S-Plot graphs, we can check the graphs of the average distribution of the biomarkers, Figures 5a, 5b, 5c and 5d. Thus, we can observe more clearly that the metabolites mentioned above (also contained in Table 2) are more pronounced in resistant clones when compared with non-resistant clones to anthracnose.

The metabolites tetra-*O*-galloyl-glucoside and penta-*O*-galloyl-glucoside are polyphenols, more specifically hydrolyzable tannin that belongs to the group of gallotannins, which have already been found in plants of different genus and species. Tannin compounds are considered part of the plant's natural defense system against environmental stressors.^{57,58} The literature also reports the antifungal action of polyphenols against anthracnose.⁵⁹ Thus, reinforcing the biomarker as a metabolite possibly associated with the clones' resistance to anthracnose.

Besides, several biological effects of this metabolite have been reported, including anticancer, anti-adipogenic, anti-microbial, anti-diabetic, anti-inflammatory, and anti-oxidative activities. In addition, there are also other reports of the bioactivity of this metabolite, such as antiviral activity against hepatitis B virus (HBV), hepatitis C virus (HCV), human immunodeficiency virus (HIV), and herpes simplex virus (HSV).^{57,58,60,61}

Several bioactivities have already been attributed to them about anacardic acids, which have been identified as metabolites responsible for devising the distinction between resistant and non-resistant clones. Studies with these metabolites revealed potential use against

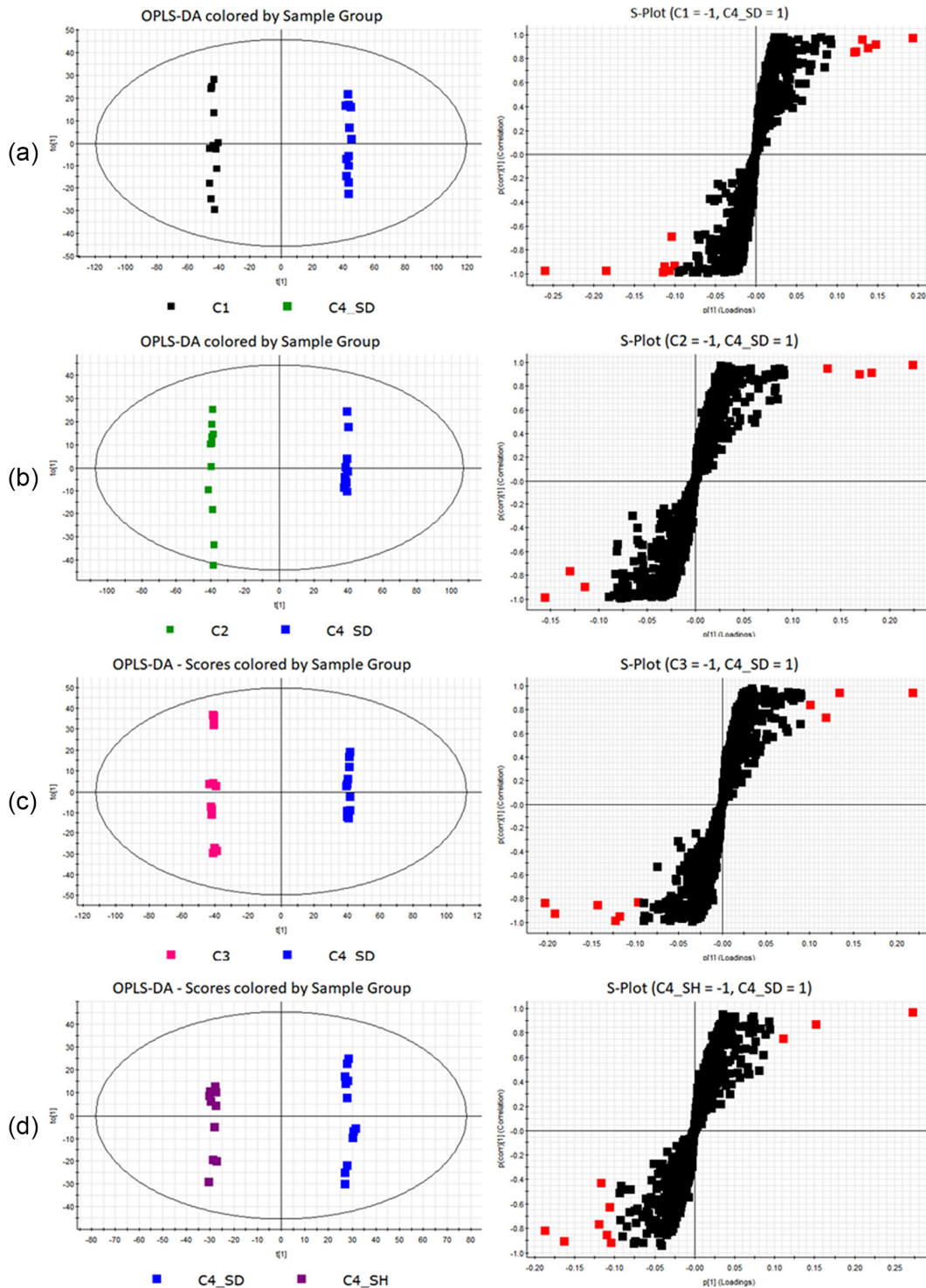


Figure 4. Graphic of scores (left) and S-Plot from OPLS-DA (right) for leaf extracts of *A. occidentale*: (a) C1 vs. C4_SD (susceptible and diseased); (b) C2 vs. C4_SD (susceptible and diseased); (c) C3 vs. C4_SD (susceptible and diseased) and (d) C4_SH (susceptible and healthy) vs. C4_SD (susceptible and diseased).

important bacteria that cause dental caries (*Streptococcus mutans*), acne (*Propionibacterium aches*), ulcers (*Helicobacter pylori*), and other infectious conditions (*Staphylococcus aureus*).⁶² Regarding the antifungal activity of anacardic acids, the literature lacks more

detailed studies; however, a preliminary investigation indicated that these metabolites do not have great efficacy against fungal growth. However, these metabolites may play an important role in inhibiting spore germination. In this way, anacardic acid-producing plants may be

more resistant to the development of diseases caused by fungal pathogens.⁶² Thus, this evidence reported in the literature and the information obtained through

chemometric analysis can corroborate to strengthen the hypothesis of the beneficial relationship of anacardic acids with the leaves of *Anacardium occidentale* clones.

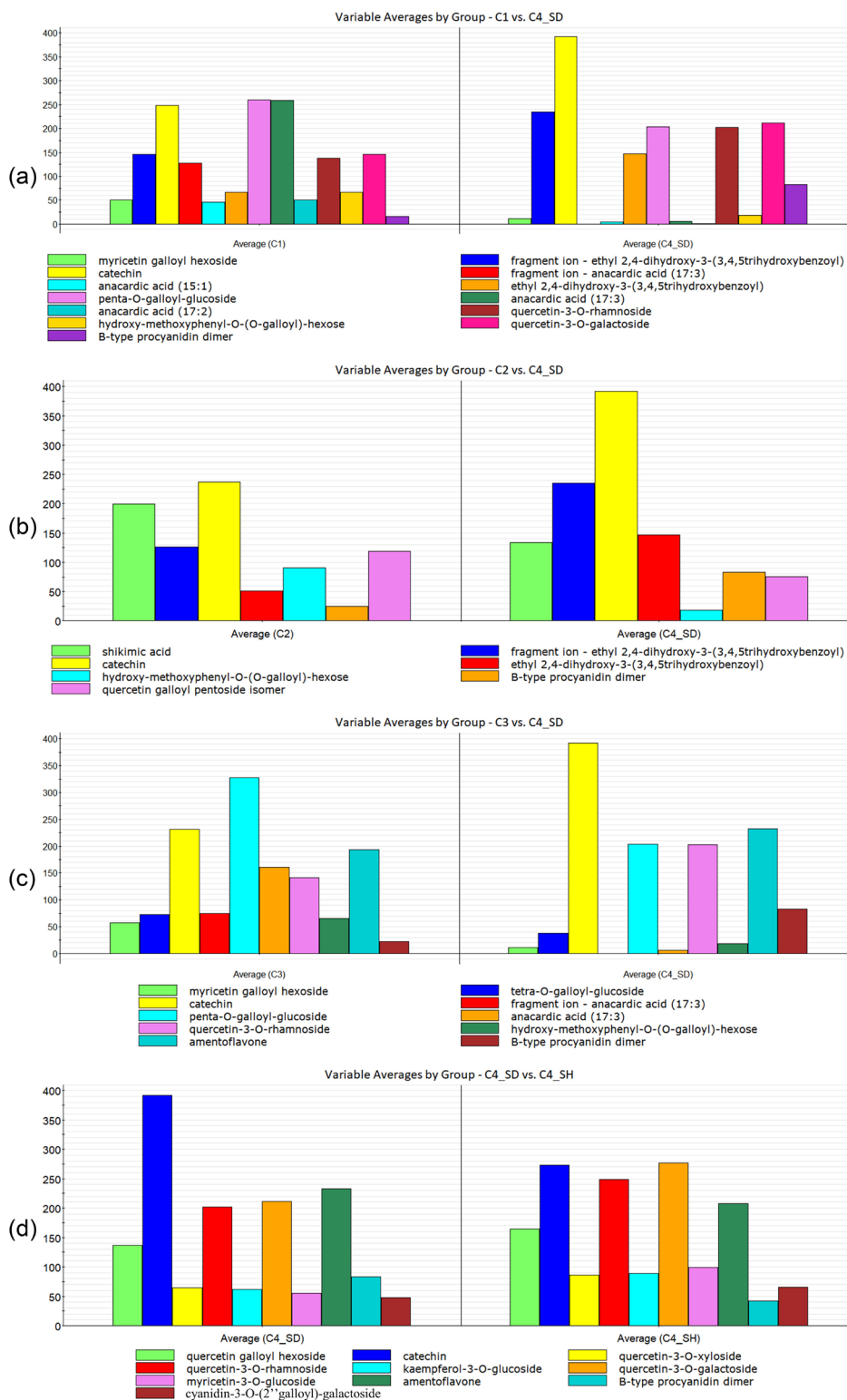


Figure 5. Mean variation of the discriminant metabolites from the leaf extracts of dwarf-cashew clones: (a) C1 vs. C4_SD (susceptible and diseased); (b) C2 vs. C4_SD (susceptible and diseased); (c) C3 vs. C4_SD (susceptible and diseased) and (d) C4_SD (susceptible and diseased) vs. C4_SH (susceptible and healthy).

Table 2. Potential biomarkers, along with the VIP values, pointed through the OPLS-DA and S-Plot

Comparison	t_R^a / min	Mass (m/z)	Compound associated with the plant		VIP ^b
			Healthy	Susceptible and diseased	
C1 vs. C4_SD	3.09	453.1048	hydroxy-methoxyphenyl- <i>O</i> -(<i>O</i> -galloyl)-hexose	–	6.03143
	3.85	631.0975	myricetin galloyl hexoside	–	5.29549
	4.28	939.1106	penta- <i>O</i> -galloyl-glucoside	–	5.45557
	15.38	325.2503	fragment ion - anacardic acid (17:3)	–	9.69652
	15.38	369.2406	anacardic acid (17:3)	–	13.6958
	15.90	345.2428	anacardic acid (15:1)	–	9.69652
	16.06	371.2582	anacardic acid (17:2)	–	5.93149
	2.74	577.1366	–	B-type procyanidin dimer	6.99247
	3.01	289.0705	–	catechin	10.3065
	4.18	463.0880	–	quercetin-3- <i>O</i> -galactoside	6.53039
	4.69	447.0909	–	quercetin-3- <i>O</i> -rhamnoside	6.43365
	5.44	197.0426	–	fragment ion - ethyl 2,4-dihydroxy-3-(3,4,5trihydroxybenzoyl)	7.87872
	5.44	349.0562	–	ethyl 2,4-dihydroxy-3-(3,4,5trihydroxybenzoyl)	7.36085
	C2 vs. C4_SD	0.96	173.0433	shikimic acid	–
3.09		453.1048	hydroxy-methoxyphenyl- <i>O</i> -(<i>O</i> -galloyl)-hexose	–	8.18934
4.75		585.0923	quercetin galloyl pentoside isomer	–	6.0198
2.74		577.1366	–	B-type procyanidin dimer	7.20447
3.01		289.0705	–	catechin	11.8627
5.44		197.0424	–	fragment ion - ethyl 2,4-dihydroxy-3-(3,4,5trihydroxybenzoyl)	9.64248
5.44		349.0562	–	ethyl 2,4-dihydroxy-3-(3,4,5trihydroxybenzoyl)	8.97728
C3 vs. C4_SD	3.09	453.1048	hydroxy-methoxyphenyl- <i>O</i> -(<i>O</i> -galloyl)-hexose	–	6.4372
	3.85	631.0975	myricetin galloyl hexoside	–	6.17815
	3.98	787.1033	tetra- <i>O</i> -galloyl-glucoside	–	5.05398
	4.28	939.1106	penta- <i>O</i> -galloyl-glucoside	–	10.0749
	15.38	325.2503	fragment ion - anacardic acid (17:3)	–	7.5472
	15.38	369.2406	anacardic acid (17:3)	–	10.7223
	2.74	577.1366	–	B-type procyanidin dimer	7.11072
	3.01	–	–	catechin	11.5700
	4.69	–	–	quercetin-3- <i>O</i> -rhamnoside	6.23384
	7.02	–	–	amentoflavone	5.36072
C4_SH vs. C4_SD	3.79	–	myricetin-3- <i>O</i> -glucoside	–	8.60503
	4.69	–	quercetin-3- <i>O</i> -rhamnoside	–	6.05042
	4.49	–	kaempferol-3- <i>O</i> -glucoside	–	5.57408
	4.18	–	quercetin-3- <i>O</i> -galactoside	–	9.8749
	4.25	–	quercetin galloyl hexoside	–	6.33076
	4.43	–	quercetin-3- <i>O</i> -xyloside	–	5.79357
	4.59	–	cyanidin-3- <i>O</i> -(2'' galloyl)-galactoside	–	5.46027
	2.74	–	–	B-type procyanidin dimer	8.01379
	3.01	–	–	catechin	14.3827
	7.02	–	–	amentoflavone	5.86459

^aRetention time; ^bvariable importance in projection (VIP).

Considering that the chemometric tools indicated three different anacardic acids (17:3; 17:2; 15:1) as metabolites associated with clones resistant to anthracnose disease, which is caused by fungi of the genus *Colletotrichum*.

Candidates for biomarkers of susceptibility

The leaves extract from the susceptible and diseased clone (C4_SD) was compared with the leaves extract

from the three resistant clones and the susceptible and healthy clone (Figure 4). It was possible to infer that the metabolites B-type procyanidin dimer, catechin, quercetin-3-*O*-galactoside, quercetin-3-*O*-rhamnoside, ethyl 2,4-dihydroxy-3-(3,4,5trihydroxybenzoyl) and amentoflavone are associated with diseased and susceptible clones. This finding is supported by the S-Plot (Figure 4) and corroborated by the average distribution graphs of the biomarkers shown in Figure 5. Furthermore, in Table 2, we can observe all the markers together with the VIP values.

Evaluating Table 2, we can observe and infer important information about the extracts of the clones. Initially, we can verify that the ethyl 2,4-dihydroxy-3-(3,4,5trihydroxybenzoyl) metabolite presents itself as a marker of diseased clones. However, this behavior is not observed when evaluating the comparison between healthy and disease susceptible clones. Thus, this metabolite may be more evident in susceptible clones and not directly a marker of diseased clones.

Similarly, the metabolites quercetin-3-*O*-galactoside and quercetin-3-*O*-rhamnoside also appear to be characteristic of diseased clones. However, when we evaluated the comparison of susceptible clones (healthy and diseased), we verified that these compounds appear as markers related to susceptible and healthy clones. Thus, these metabolites should not be characteristic specifically of susceptible and diseased clones, but of susceptible clones (healthy and diseased), is even more pronounced in susceptible and healthy clones than in susceptible and diseased clones (Figure 5 and Table 2).

In the case of the B-type procyanidin dimer and catechin metabolites, both were identified as susceptible and diseased clones' markers in all comparisons, including the comparisons between healthy and diseased susceptible clones. Thus, there is a great possibility that the diseased plants are intensifying the biosynthesis of these flavonoids to act on the plant defense system, aiming to combat the anthracnose disease to which the diseased clones are affected. It is important to emphasize that these metabolites were also identified by healthy clones (Table 1). However, the tools used in the chemometric analyzes indicate that these metabolites are at higher levels of concentration in the diseased clones (Figure 5).

Conclusions

The extraction followed by the analyses by UPLC-QTOF-MS^E allowed the annotation of 39 metabolites in the four clones studied. Among the metabolites noted, anacardic acids, flavonoids, procyanidins, and others stand out. In general, it is possible to observe similarities and

differences between the chemical profiles of the clones. These differences, as well as the similarities, were evaluated and explored through multivariate analysis. Through the PCA it was possible to observe the grouping and the clear distinction between the groups of samples.

In addition, it was possible, through the analysis of OPLS-DA, S-Plot and VIP scores, to determine the characteristic biomarkers of the different sets of samples. Thus, it was possible to infer that the metabolites hydroxy-methoxyphenyl-*O*-(*O*-galloyl)-hexose, myricetin galloyl hexoside, penta-*O*-galloyl-glucoside, anacardic acids (17:3; 17:2; 15:1), shikimic acid, quercetin galloyl pentoside isomer and tetra-*O*-galloyl-glucoside are supposedly associated with the resistance of the clones against anthracnose. On the other hand, it was also possible to infer that the metabolites catechin and B-type procyanidin dimer are associated with the diseased clones. This fact may indicate that the clones may intensify the biosynthesis of these metabolites to act in the defense mechanism of the plant organism against anthracnose disease.

Acknowledgments

The authors gratefully acknowledge financial support from the CNPq, National Council for Scientific and Technological Development (303791/2016-0), INCT BioNat, National Institute of Science and Technology (grant No. 465637/2014-0). We would also like to thank Embrapa (SEG 03.14.01.012.00.00). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Author Contributions

Priscilla C. C. Alves was responsible for conceptualization, data curation, formal analysis, investigation, visualization, writing original draft; Jhonyson A. C. Guedes for conceptualization, data curation, investigation, software, validation, visualization, writing original draft, writing-review and editing; Luiz A. L. Serrano for conceptualization, data curation, visualization, writing-review and editing; Marlon V. V. Martins for conceptualization, data curation, visualization, writing-review and editing; Debora B. de Sousa for conceptualization, investigation, validation, visualization, writing original draft; Gisele S. Silva for conceptualization, investigation, validation, visualization, writing original draft; Paulo R. V. Ribeiro for conceptualization, data curation, investigation, software, validation, visualization; Dávila S. Zampieri for conceptualization, validation, visualization, writing-review and editing; Guilherme J. Zocolo for conceptualization, funding acquisition, investigation, project administration, resources, validation, visualization, writing original draft, writing-review and editing.

References

1. Brito, E. S.; Silva, E. O.; Rodrigues, S.; In *Exotic Fruits Reference Guide*; Rodrigues, S.; Silva, E. O.; Brito, E. S., eds.; Elsevier: London, 2018.
2. Food and Agriculture Organization of the United Nations; *Crops and Livestock Products*, <https://www.fao.org/faostat/en/#data/QCL/visualize>, accessed in June 2023.
3. Dean, R.; Van Kan, J. A. L.; Pretorius, Z. A.; Hammond-Kosack, K. E.; Di Pietro, A.; Spanu, P. D.; Rudd, J. J.; Dickman, M.; Kahmann, R.; Ellis, J.; Foster, G. D.; *Mol. Plant Pathol.* **2012**, *13*, 414. [Crossref]
4. Freire, F. C. O.; Cardoso, J. E.; dos Santos, A. A.; Viana, F. M. P.; *Crop Prot.* **2002**, *21*, 489. [Crossref]
5. Uaciquete, A.; Korsten, L.; Van der Waals, J. E.; *Crop Prot.* **2013**, *49*, 66. [Crossref]
6. Veloso, J. S.; Lima, W. G.; Reis, A.; Doyle, V. P.; Michereff, S. J.; Câmara, M. P. S.; *Plant Pathol.* **2021**, *70*, 167. [Crossref]
7. Barros, L. M.; Cavalcanti, J. J. V.; de Paiva, J. R.; Crisóstomo, J. R.; Corrêa, M. P. F.; Lima, A. C.; *Pesq. Agropec. Bras.* **2000**, *35*, 2197. [Crossref]
8. Wonni, I.; Sereme, D.; Ouedraogo, I.; Kassankagno, A. I.; Dao, I.; Ouedraogo, L.; Nacro, S.; *Adv. Plants Agric. Res.* **2017**, *6*, 78. [Crossref]
9. Guginski-Piva, C. A.; Bogo, A.; Gomes, B. R.; Menon, J. K.; Nodari, R. O.; Welter, L. J.; *J. Plant Dis. Prot.* **2018**, *125*, 405. [Crossref]
10. Lima, N. B.; Lima, W. G.; Tovar-Pedraza, J. M.; Michereff, S. J.; Câmara, M. P. S.; *Eur. J. Plant Pathol.* **2015**, *141*, 679. [Crossref]
11. Oliveira, P. D. L.; de Oliveira, K. Á. R.; Vieira, W. A. S.; Câmara, M. P. S.; de Souza, E. L.; *Int. J. Food Microbiol.* **2018**, *266*, 87. [Crossref]
12. Uaciquete, A.; Korsten, L.; Van der Waals, J. E.; *Crop Prot.* **2013**, *50*, 6. [Crossref]
13. de Alencar, P. G.; de Espindola, G. M.; Carneiro, E. L. N. C.; *Land Use Policy* **2018**, *79*, 1. [Crossref]
14. Cipriano, A. K. A. L.; Gondim, D. M. F.; Vasconcelos, I. M.; Martins, J. A. M.; Moura, A. A.; Moreno, F. B.; Monteiro-Moreira, A. C. O.; Melo, J. G. M.; Cardoso, J. E.; Paiva, A. L. S.; Oliveira, J. T. A.; *J. Proteomics* **2015**, *113*, 90. [Crossref]
15. Pinto, O. R. O.; Cardoso, J. E.; Maia, A. H. N.; Pinto, C. M.; Lima, J. S.; Viana, F. M. P.; Martins, M. V. V.; *Crop Prot.* **2018**, *112*, 282. [Crossref]
16. Ministério da Agricultura Pecuária e Abastecimento; *Agrofit*, https://agrofit.agricultura.gov.br/agrofit_cons/principal_agrofit_cons, accessed in June 2023.
17. Muniz, C. R.; Freire, F. C. O.; Viana, F. M. P.; Cardoso, J. E.; Cooke, P.; Wood, D.; Guedes, M. I. F.; *Micron* **2011**, *42*, 419. [Crossref]
18. Krastanov, A.; *Biotechnol. Biotechnol. Equip.* **2010**, *24*, 1537. [Crossref]
19. Wen, W.; Li, K.; Alseikh, S.; Omranian, N.; Zhao, L.; Zhou, Y.; Xiao, Y.; Jin, M.; Yang, N.; Liu, H.; Florian, A.; Li, W.; Pan, Q.; Nikoloski, Z.; Yan, J.; Fernie, A. R.; *Plant Cell* **2015**, *27*, 1839. [Crossref]
20. Chagas-Paula, D. A.; Zhang, T.; da Costa, F. B.; Edrada-Ebel, R. A.; *Metabolites* **2015**, *5*, 404. [Crossref]
21. Nehme, C. J.; de Moraes, P. L. R.; Tininis, A. G.; Cavalheiro, A. J.; *Biochem. Syst. Ecol.* **2008**, *36*, 602. [Crossref]
22. Luz, L. R.; Porto, D. D.; Castro, C. B.; Silva, M. F. S.; Alves Filho, E. G.; Canuto, K. M.; de Brito, E. S.; Becker, H.; Pessoa, C. Ó.; Zocolo, G. J.; *J. Chromatogr. B* **2018**, *1099*, 97. [Crossref]
23. *MarkerLynx XS*, v4.1 SCN719; Waters Corporation, Milford, USA, 2009.
24. Kim, J.; Choi, J. N.; Kang, D.; Son, G. H.; Kim, Y.-S.; Choi, H.-K.; Kwon, D. Y.; Lee, C. H.; *Biosci. Biotechnol. Biochem.* **2011**, *75*, 732. [Crossref]
25. Avula, B.; Wang, Y.-H.; Smillie, T. J.; Khan, I. A.; *Chromatographia* **2009**, *69*, 307. [Crossref]
26. Abu-Reidah, I. M.; Ali-shtayeh, M. S.; Jamous, R. M.; Arráez-Román, D.; Segura-Carretero, A.; *Food Chem.* **2015**, *166*, 179. [Crossref]
27. Dorta, E.; González, M.; Lobo, M. G.; Sánchez-Moreno, C.; de Ancos, B.; *Food Res. Int.* **2014**, *57*, 51. [Crossref]
28. Erşan, S.; Güçlü Üstündağ, Ö.; Carle, R.; Schweiggert, R. M.; *J. Agric. Food Chem.* **2016**, *64*, 5334. [Crossref]
29. Sobeh, M.; Mahmoud, M. F.; Hasan, R. A.; Abdelfattah, M. A. O.; Sabry, O. M.; Ghareeb, M. A.; El-Shazly, A. M.; Wink, M.; *Sci. Rep.* **2018**, *8*, 9343. [Crossref]
30. Fongang, A. L. M.; Nguemfo, E. L.; Nangue, Y. D.; Zanguue, C. B.; Fouokeng, Y.; Azebaze, A. G. B.; Llorent-Martínez, E. J.; Córdova, M. L. F.; Dongmo, A. B.; Vierling, W.; *J. Ethnopharmacol.* **2017**, *203*, 11. [Crossref]
31. Sandhu, A. K.; Gu, L.; *J. Agric. Food Chem.* **2010**, *58*, 4681. [Crossref]
32. Al Sayed, E.; Martiskainen, O.; Sinkkonen, J.; Pihlaja, K.; Ayoub, N.; Singab, A.-E. N.; El-Azizi, M.; *Nat. Prod. Commun.* **2010**, *5*, 1934578X1000500. [Crossref]
33. Shukri, M. A. M.; Alan, C.; *J. Trop. Agric. Food Sci.* **2010**, *38*, 221. [Link] accessed in June 2023
34. Schieber, A.; Berardini, N.; Carle, R.; *J. Agric. Food Chem.* **2003**, *51*, 5006. [Crossref]
35. Heras, R. M.-L.; Quifer-Rada, P.; Andrés, A.; Lamuela-Raventós, R.; *J. Funct. Foods* **2016**, *23*, 370. [Crossref]
36. Regazzoni, L.; Arlandini, E.; Garzon, D.; Alfredo, N.; Beretta, G.; Maffei, R.; *J. Pharm. Biomed. Anal.* **2013**, *72*, 202. [Crossref]
37. da Silva, D. P. B.; Florentino, I. F.; da Silva Moreira, L. K.; Brito, A. F.; Carvalho, V. V.; Rodrigues, M. F.; Vasconcelos, G. A.; Vaz, B. G.; Pereira-Junior, M. A.; Fernandes, K. F.; Costa, E. A.; *J. Ethnopharmacol.* **2018**, *213*, 395. [Crossref]

38. Schulze-Kaysers, N.; Feuereisen, M. M.; Schieber, A.; *RSC Adv.* **2015**, *5*, 73301. [Crossref]
39. de Figueirêdo, M. C. B.; Potting, J.; Serrano, L. A. L.; Bezerra, M. A.; Barros, V. S.; Gondim, R. S.; Nemecek, T.; *J. Clean. Prod.* **2016**, *112*, 131. [Crossref]
40. Chagas-Paula, D. A.; Oliveira, T.; Zhang, T.; Edrada-Ebel, R.; Da Costa, F.; *Planta Med.* **2015**, *81*, 450. [Crossref]
41. Savaris, C. R.; Ghiraldi, D. M. B.; Rovigatti Chiavelli, L. U.; de Oliveira Santin, S. M.; Milaneze-Gutierrez, M. A.; Kischkel, B.; Negri, M.; Scariot, D. B.; Garcia, F. P.; Nakamura, C. V.; Pomini, A. M.; *Biochem. Syst. Ecol.* **2018**, *80*, 11. [Crossref]
42. dos Santos, A. E.; Kuster, R. M.; Yamamoto, K. A.; Salles, T. S.; Campos, R.; de Meneses, M. D. F.; Soares, M. R.; Ferreira, D.; *Parasit. Vectors* **2014**, *7*, 130. [Crossref]
43. Liu, H.; Zou, T.; Gao, J.-M.; Gu, L.; *Food Chem.* **2013**, *141*, 488. [Crossref]
44. Rockenbach, I. I.; Jungfer, E.; Ritter, C.; Santiago-Schübel, B.; Thiele, B.; Fett, R.; Galensa, R.; *Food Res. Int.* **2012**, *48*, 848. [Crossref]
45. Rue, E. A.; Rush, M. D.; van Breemen, R. B.; *Phytochem. Rev.* **2018**, *17*, 1. [Crossref]
46. Said, R. B.; Hamed, A. I.; Mahalel, U. A.; Al-Ayed, A. S.; Kowalczyk, M.; Moldoch, J.; Oleszek, W.; Stochmal, A.; *Int. J. Mol. Sci.* **2017**, *18*, 512. [Crossref]
47. Yan, T.; Hu, G.-S.; Wang, A.-H.; Hong, Y.; Jia, J.-M.; *Nat. Prod. Res.* **2014**, *28*, 1834. [Crossref]
48. Friedrich, W.; Eberhardt, A.; Galensa, R.; *Eur. Food Res. Technol.* **2000**, *211*, 56. [Crossref]
49. Karonen, M.; Loponen, J.; Ossipov, V.; Pihlaja, K.; *Anal. Chim. Acta* **2004**, *522*, 105. [Crossref]
50. Bartel, J.; Krumsiek, J.; Theis, F. J.; *Comput. Struct. Biotechnol. J.* **2013**, *4*, [Crossref]
51. Wiklund, S.; Johansson, E.; Sjöström, L.; Mellerowicz, E. J.; Edlund, U.; Shockcor, J. P.; Gottfries, J.; Moritz, T.; Trygg, J.; *Anal. Chem.* **2008**, *80*, 115. [Crossref]
52. Guedes, J. A. C.; Santiago, Y. G.; dos Reis Luz, L.; Silva, M. F. S.; Ramires, C. M. C.; de Lima, M. A. C.; de Oliveira, V. R.; do Ó Pessoa, C.; Canuto, K. M.; de Brito, E. S.; Lima, M. F.; Alves, R. E.; Zampieri, D.; do Nascimento, R. F.; Zocolo, G. J.; *Phytochem. Lett.* **2020**, *40*, 26. [Crossref]
53. Matos, T. K.; Guedes, J. A. C.; Alves Filho, E. G.; Luz, L. R.; Lopes, G. S.; do Nascimento, R. F.; de Sousa, J. A.; Canuto, K. M.; de Brito, E. S.; Dias-Pini, N. S.; Zocolo, G. J.; *J. Braz. Chem. Soc.* **2021**, *32*, 1617. [Crossref]
54. Madala, N. E.; Piater, L. A.; Steenkamp, P. A.; Dubery, I. A.; *SpringerPlus* **2014**, *3*, 254. [Crossref]
55. Dendena, B.; Corsi, S.; *Agron. Sustainable Dev.* **2014**, *34*, 753. [Crossref]
56. Guedes, J. A. C.; Alves Filho, E. G.; Silva, M. F. S.; Rodrigues, T. H. S.; Ramires, C. M. C.; Lima, M. A. C.; Silva, G. S.; Pessoa, C. Ó.; Canuto, K. M.; Brito, E. S.; Alves, R. E.; Nascimento, R. F.; Zocolo, G. J.; *J. Braz. Chem. Soc.* **2020**, *31*, 331. [Crossref]
57. Cao, Y.; Himmeldirk, K. B.; Qian, Y.; Ren, Y.; Malki, A.; Chen, X.; *J. Nat. Med.* **2014**, *68*, 465. [Crossref]
58. Torres-León, C.; Ventura-Sobrevilla, J.; Serna-Cock, L.; Ascacio-Valdés, J. A.; Contreras-Esquível, J.; Aguilar, C. N.; *J. Funct. Foods* **2017**, *37*, 176. [Crossref]
59. Gómez-Maldonado, D.; Lobato-Calleros, C.; Aguirre-Mandujano, E.; Leyva-Mir, S. G.; Robles-Yerena, L.; Vernon-Carter, E. J.; *LWT - Food Sci. Technol.* **2020**, *126*, 109337. [Crossref]
60. Bae, S.; Kim, S. Y.; Do, M. H.; Lee, C. H.; Song, Y.-J.; *Antiviral Res.* **2017**, *144*, 266. [Crossref]
61. Zhang, J.; Li, L.; Kim, S.-H.; Hagerman, A. E.; Lü, J.; *Pharm. Res.* **2009**, *26*, 2066. [Crossref]
62. Schultz, D. J.; Wickramasinghe, N. S.; Klinge, C. M.; In *Recent Advances in Phytochemistry*; Romeo, J. T., ed.; Elsevier: Oxford, 2006.

Submitted: March 31, 2023

Published online: June 20, 2023

