

1 Mass spectrometry-based untargeted metabolomics approaches for
2 comprehensive structural annotation of bioactive metabolites from bushy
3 cashew (*Anacardium humile*) fruits

4
5
6
7
8
9
10
11 Gabriel F. dos Santos ^a, Nerilson M. Lima ^{a,*}, Gesiane S. Lima ^a, Jussara V. Roque ^a, Gagan
12 Preet ^b, Ernest Oppong-Danquah ^c, Teresinha J. A. S. Andrade ^d, Marcel Jaspars ^b, Boniek
13 Gontijo Vaz ^{a,*}

14
15
16
17
18
19
20
21 ^a Institute of Chemistry, Federal University of Goiás, Goiania, 59078-970, GO, Brazil

22
23 ^b Marine Biodiscovery Centre, Department of Chemistry, University of Aberdeen, Aberdeen
24 AB24 3UE, Scotland, UK

25
26
27
28 ^c GEOMAR Centre for Marine Biotechnology (GEOMAR-Biotech), Research Unit Marine
29 Natural Product Chemistry, GEOMAR Helmholtz Centre for Ocean Research Kiel, Am Kiel-
30 Kanal 44, 24106 Kiel, Germany

31
32
33
34
35 ^d Instituto Federal de Educação, Ciência e Tecnologia do Maranhão, 65635-468, Presidente
36 Dutra (MA), Brazil

37
38
39
40 *nerilsonmarques@gmail.com (Nerilson M. Lima); boniek@ufg.br (Boniek G. Vaz)

41
42
43
44
45 **Abstract**

46
47
48 *Anacardium humile* (bushy cashew) is a native Brazilian plant with substantial pharmacological
49 potential and noteworthy commercial significance in the food industry. This study introduces
50 an untargeted metabolomics approach based on mass spectrometry for the comprehensive
51 structural annotation of bioactive metabolites. The fruits were collected from three distinct sites
52 and subjected to LC-HRMS/MS analysis. A total of eighty-eight compounds were putatively
53 annotated across various metabolite classes. This unveiled a metabolic profile characterized by

26 notable concentrations of polyphenols, including flavonoids, tannins, phenolic acids, and
27 quinones, while aliphatic acids and terpenes were found in limited quantities. Noteworthy, no
28 significant disparities in the metabolic content were observed among the collection sites. The
29 three principal metabolites (peonidin 3-*O*-glucopyranoside, methylcyanidin and
30 methyldephinidin) underwent assessment for antioxidant activity via molecular docking
31 analysis, subsequently generating structure-based feature pharmacophores. The process of
32 annotation propagation yielded a comprehensive qualitative appraisal of the *A. humile*
33 metabolome. The outcomes obtained offer potential candidates for further exploration of their
34 nutraceutical attributes.

36 Keywords

37 *Anacardium humile*; bushy cashew; food chemistry; mass spectrometry; untargeted
38 metabolomics; molecular networking; molecular docking; pharmacophore.

40 1. Introduction

41 *Anacardium humile* St. Hil (Anacardiaceae) is an indigenous medicinal plant from Brazil,
42 popularly known as bushy cashew, with potentially significant health benefits as food for its
43 nutritional, medicinal, and agricultural importance. Various biological activities have been
44 reported for *A. humile*, such as anti-inflammatory, anticancer, antidiarrheal, and antidiabetic
45 properties (Kubo et al., 1993; Luiz-Ferreira et al., 2008; Urzêda et al., 2013). Additionally, stem
46 extracts have been reported to regulate blood glucose, while leaf and bark infusions are used to
47 treat gastric disorders such as ulcers and gastritis (Luiz-Ferreira et al., 2008; Urzêda et al.,
48 2013).

49 The chemical diversity of *A. humile* comprises a significant level of volatile compounds as
50 well as organic acids such as resorcinolic acid, ascorbic acid, and anacardic acids (Assunção &

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51 Mercadante, 2003; Bicalho et al., 2000). Previous studies have indicated the presence of
52 carotenoids (α -carotene, β -carotene, and β -cryptoxanthin), alkaloids, and polyphenols with
53 significant antioxidant properties. The presence of tannins such as (procyanidin B2), phenolic
54 acids (gallic acid), and flavonoids (catechin, glycosylated quercetin) present in *A. humile* have
55 shown antiglycation properties (Assunção & Mercadante, 2003; Lima Júnior et al., 2021; Luiz-
56 Ferreira et al., 2008; Royo et al., 2015). There is currently little information on the metabolic
57 profiling of bushy cashews or the contribution of individual metabolites to the nutritional value.
58 Therefore, research studies aiming at the rapid and comprehensive acquisition of the metabolic
59 profile of these species are necessary.

60 Metabolic coverage assessment to detect the bioactive constituents and identify the
61 metabolites responsible for nutritional properties requires modern analytical tools. Considering
62 that the process of isolation and structural elucidation of the chemical components of complex
63 matrices is a time-consuming and labour-intensive process, metabolomic approaches using
64 advanced analytical technologies and bioinformatics tools have become a necessary pre-
65 requisite in finding potentially new bioactive molecules. Among the techniques employed in
66 these approaches, mass spectrometry is widely acknowledged as the preferred strategy for
67 exploring food chemistry. It allows the analysing of various chemical species with diverse
68 physicochemical properties, even in minute quantities within complex metabolite mixtures
69 (Domínguez et al., 2020; Jorge et al., 2016).

70 Applications of mass spectrometry-based analytical platforms to profile plant metabolomes of
71 food interest have been implemented in routine analysis in research laboratories (Domínguez
72 et al., 2020; Lu et al., 2018). However, the massive amount of spectral data generated requires
73 robust tools for data exploration and organisation to obtain information on the metabolome of
74 the target species. To address challenges in interpreting these large datasets and translating
75 chemical composition into biological knowledge, sophisticated dereplication tools based on

1 76 spectral fragmentation profiles are used. In addition, *in silico* analysis technologies help
2 77 characterise unknown molecules and develop methods for assessing target metabolites. Data
3
4 78 analysis tools, such as the Global Natural Products Social Molecular Networking (GNPS)
5
6
7 79 platform, have allowed the exploration of plant metabolomes as well as assisting and
8
9
10 80 accelerating the discovery of new bioactive agents (Pilon et al., 2021; Ramabulana et al., 2021).

11
12 81 This work aims to apply untargeted metabolomics approaches to screen for bioactive
13
14 82 metabolites in bushy cashew fruits by high-resolution mass spectrometry (HRMS) combined
15
16
17 83 with *in silico* fragmentation tools. It also utilizes the GNPS platform tools, including classic
18
19 84 Molecular Networking (MN), Feature-Based Molecular Networking (FBMN), Network
20
21
22 85 Annotation Propagation (NAP), Dereplicator+, Suspect library, molDiscovery, MS2LDA, and
23
24 86 MolNetEnhancer to enhance metabolite annotation. Subsequently, based on the collective
25
26
27 87 results, pharmacophore and molecular docking analyses were conducted to generate structure-
28
29 88 based feature pharmacophores for the antioxidant activity of the major metabolites. This study
30
31
32 89 represents the first comprehensive exploration of the molecular diversity of *A. humile*, a typical
33
34 90 fruit found in the Brazilian savanna. Furthermore, it recognises the inherent antioxidant
35
36
37 91 properties of the plant's constituents, thereby enhancing its potential value as both a dietary
38
39 92 resource and a pharmaceutical asset.

40 41 42 93 2. Materials and methods

43 44 45 94 2.1. Chemicals and materials

46
47 95 HPLC-grade acetonitrile and methanol were purchased from Tedia Company (Fair-
48
49
50 96 field, USA). Formic acid and Caffeine-¹³C₃ were purchased from Sigma Aldrich (St. Louis,
51
52
53 97 USA). Progesterone-*d*₉ was purchased from CDN Isotopes (Quebec, Canada). Ultrapure water
54
55 98 was produced using a water purification system (Master System MS2000, Gehaka, São Paulo,
56
57 99 Brazil) with a resistivity of 18.2 MΩcm.

60 100

101 2.2. *Sample preparation*

1
2 102 *Anacardium humile* fruits were collected from three different locations in Goiás State,
3
4
5 103 Brazil (Goiânia (S1), Uruana (S2), and Campinaçu (S3) cities). Bushy cashews were sliced
6
7 104 manually using a sterile knife, frozen at -80 °C, and freeze-dried for 48h. For metabolomics
8
9
10 105 analyses, 10 mg of freeze-dried fruits were extracted with 1 mL of methanol, vortexed for 1
11
12 106 min, and centrifuged (10 min, 15000 rpm at room temperature). This solution was diluted with
13
14 107 methanol (1:5, v:v) and was transferred to injecting vials for LC-MS/MS analysis. Stable
15
16
17 108 isotopes, Caffeine-¹³C₃ and Progesterone-*d*₉ (2.5 µg mL⁻¹) were used as internal standards.

18
19 109

20
21
22 110 2.3. *LC-MS analysis*

23
24 111 LC-MS/MS analyses were performed on an HPLC-UV 1220 Infinity II (Agilent
25
26
27 112 Technologies) coupled with a Q-Exactive hybrid Quadrupole-Orbitrap high-resolution mass
28
29 113 spectrometer (Thermo Scientific) as well as an electrospray ionisation source. The column used
30
31
32 114 in this study was an InfinityLab Poroshell 120 EC-C18 column (4.6 × 100 mm × 2.7 µm
33
34 115 Agilent). All samples were analysed using a gradient elution program. In both ESI positive and
35
36 116 negative modes, the binary mobile phase comprised A (water with 0.1% formic acid) and B
37
38
39 117 (methanol). The gradient elution started at 5% (B) and linearly increased to 100% (B) in 40
40
41 118 min and kept constant for 10 min at 100% (B). The eluent was then restored to the initial
42
43
44 119 conditions in 10 min. The flow rate was set at 0.3 mL min⁻¹. The injection volume was 30 µL,
45
46 120 and the column temperature was 35 °C. The ESI source conditions were set as follows: spray
47
48
49 121 voltage 3.5 kV (in both ionisation modes); the capillary temperature was 250 °C (positive
50
51 122 mode) and 320 °C (negative mode); S-lens RF level 60 V (in both ionisation modes); sheath
52
53
54 123 gas flow rate at 47 L min⁻¹ (positive mode) and 35 L min⁻¹ (negative mode); and aux gas flow
55
56 124 rate at 11 L min⁻¹ (positive mode) and 10 L min⁻¹ (negative mode). In both ESI positive and
57
58 125 negative modes, high-resolution mass spectra were obtained in the Full MS/data dependent -
59
60
61
62
63
64
65

126 MS² (dd-MS²) mode. The mass range in the full MS scanning experiments was m/z 100-1200.

127 The top 5 (TopN, 5, loop count 5) most abundant precursors were sequentially transferred for

128 collision-induced fragmentation acquisition. The collision energy for target analytes was 20,

129 30, and 35 eV. Resolving power was 140,000 and 70,000 for full MS and dd-MS² acquisitions,

130 respectively.

131

132 2.4. Putative compound annotation

133 The files acquired in the Q-Exactive hybrid Quadrupole-Orbitrap mass spectrometer for

134 the methanolic extracts were converted from raw into (.mzML) format using MSConvert

135 software (ProteoWizard, Palo Alto, CA, US) before being processed using MZmine software,

136 version 2.53. This study utilised metadata to organise compound information following the

137 GNPS online workflow (<https://ccms-ucsd.github.io/GNPSDocumentation/>). This platform

138 also curates MS/MS spectral library categorised based on the quality of the spectra as gold

139 (thoroughly characterised structures), silver (a compound in crude extract), and bronze (partial

140 annotation) (Wang et al., 2016). Metabolite annotations were based on searching the

141 experimental spectra against the GNPS spectral library using the tools classic Molecular

142 Networking – MN (Wang et al., 2016), Feature-Based Molecular Networking – FBMN

143 (Nothias et al., 2020), Dereplicator+ (Mohimani et al., 2018), Network Annotation Propagation

144 – NAP (da Silva et al., 2018), molDiscovery (Cao et al., 2021), MS2LDA (Van Der Hoof et

145 al., 2016), MolNetEnhancer (Ernst et al., 2019), and analysis of chromatographic data such as

146 retention time and UV spectra. These tools allow the integration of orthogonal annotation

147 methodologies and tandem mass spectrometry data (MS/MS) to explore and obtain the

148 metabolome of plants used as food.

149

150 2.5. Molecular Docking

151 Molecular docking analysis was performed using Autodock Vina v.1.2.0 (The Scripps
1
2 152 Research Institute, La Jolla, CA, USA) docking software (Trott & Olson, 2010). The receptor
3
4
5 153 site was predicted using LigandScout (Inte: Ligand) Advanced software (Wolber & Langer,
6
7 154 2005) (evaluation license key: 81809629175371877209), which identifies putative binding
8
9
10 155 pockets by creating a grid surface and calculating the buriedness value of each grid point on
11
12 156 the surface. The resulting pocket grid consists of several clusters of grid points, rendered using
13
14 157 an iso surface connecting the grid points. The iso surface represents space that may be suitable
15
16
17 158 for creating a pocket.

19 159 The x-ray crystal structure of Human Peroxiredoxin 5, a Novel Type of Mammalian
20
21
22 160 Peroxiredoxin (PDB: 1HD2) (Declercq et al., 2001), was retrieved from the Protein Data Bank
23
24 161 and utilised to perform docking simulations. Default search parameters were used where the
25
26
27 162 number of binding modes was 10, exhaustiveness was 8, and the maximum energy difference
28
29 163 was 3 kcal/mol.

31 164 LigandScout (Inte: Ligand) Advanced software (Wolber & Langer, 2005) (evaluation
32
33
34 165 license key: 81809629175371877209) was used to generate 3D pharmacophore models.
35
36 166 LigandScout's algorithm calculates and displays chemical interactions between protein–ligand
37
38
39 167 complexes.

41 168

44 169 3. Results and discussion

46 170 An untargeted metabolomics approach was employed to assess the metabolome of *A.*
47
48
49 171 *humile*. The molecular diversity of its fruits and the evaluation of the abundance of ions and
50
51 172 metabolite content were determined using high-resolution mass spectrometry. This was
52
53
54 173 complemented with dereplication tools, which involved molecular structure searches in robust
55
56 174 databases and computer-assisted fragmentation. Furthermore, given the limited data on the
57
58
59 175 metabolic composition of *A. humile* in the existing literature, the scope of the study was

176 extended to identify biomarkers of the genus *Anacardium* besides describing the classes and
177 subclasses of secondary metabolites present in the species. For this purpose, the structural
178 similarity of MS data was assessed by comparing it with the spectral library from the GNPS
179 ecosystem. Furthermore, potential candidates were identified *in silico* using specialised
180 platforms such as Sirius, Dereplicator+, and NAP, utilising MS data obtained in both positive
181 and negative ionisation modes. The detection of metabolites was performed in fruit samples of
182 *A. humile* from three different locations in Brazil.

183 Putative metabolite annotation was performed based on MS/MS fragmentation patterns,
184 calculation of empirical formulas, chromatographic analysis data such as retention time and
185 UV spectra, inspection of candidates and their analogues suggested by the spectral library,
186 which were further prioritised according to chemotaxonomic and chemophenetic data from
187 family Anacardiaceae (Wang et al., 2016). The library matches were assessed for both positive
188 and negative ionization mode data, and the merge networks polarity tool was employed to
189 explore data from both polarities within a unified workflow.

190 *Classic Molecular Networking (MN)*

191 The library matches using the classic molecular networking (MN) yielded 2164 hits with
192 198 unique library compounds in positive ionisation mode and 457 hits with 139 unique library
193 compounds in negative ionisation mode after excluding nodes with repeated hits. In contrast,
194 the merge networks polarity tool generated 2406 hits. Molecular networking-based strategies
195 allowed us to annotate three times more metabolites in positive ionisation mode than negative
196 mode. According to the Metabolomic Standard Initiative-MSI, all annotations were manually
197 inspected, resulting in level 2 or 3 (Sumner et al., 2007). To obtain hits with higher structural
198 similarity and the exact biosynthetic origin within molecular families, the threshold for the
199 cosine score similarity was set to 0.7. The structural annotations were verified in the
200 metabolome of the genus *Anacardium* using databases such as the Dictionary of Natural

201 Products. Initially, the annotations provided by the spectral library and the molecular family
1
2 202 analysis were evaluated, suggesting precursor ions with similar structures or functional groups.
3

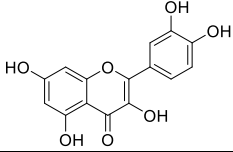
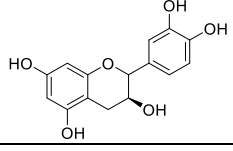
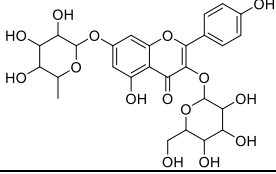
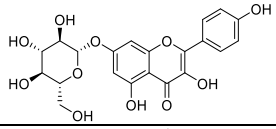
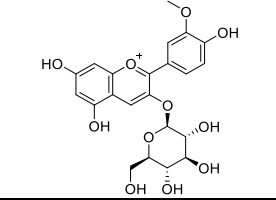
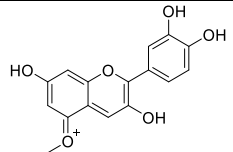
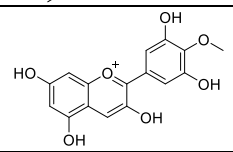
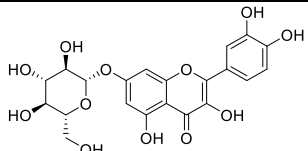
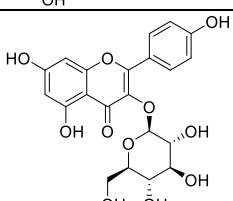
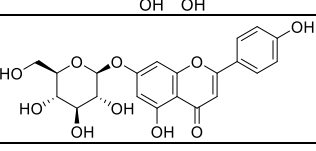
4
5 203 After the dereplication process from the MN analysis, 88 metabolites belonging to different
6
7 204 classes of secondary metabolites in aglycone and heteroside form were putatively annotated,
8
9 205 including terpenes, steroids, flavonoids, tannins, quinones, alkaloids, coumarins, phenolic
10
11 206 acids, and simple phenols. Additionally, many primary metabolites, such as carboxylic acids,
12
13 207 sugars, and amino acids, were annotated in all samples. A high molecular diversity of derivative
14
15 208 compounds has been annotated within these classes (see **Table 1**). In addition, a complex array
16
17 209 of primary metabolites was detected in all samples, such as carbohydrates, amino acids, and
18
19 210 lipids. Based on library matches, polyphenols were the major metabolites in all analyses. *O*-
20
21 211 glycosylated compounds were abundant, and *C*-glycosylated compounds were in the minority.
22
23 212 Alkaloids were detected in low concentrations in the fruits of *A. humile*, which information
24
25 213 corroborates with the data reported for the genus *Anacardium*. Regarding the collection sites,
26
27 214 no substantial variations were observed in metabolite content. All metabolites annotated were
28
29 215 previously described in the family Anacardiaceae.
30
31
32
33

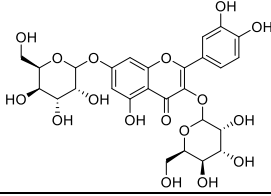
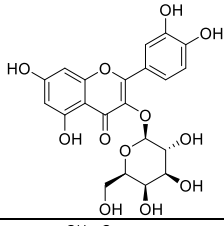
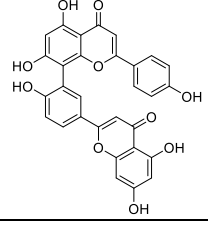
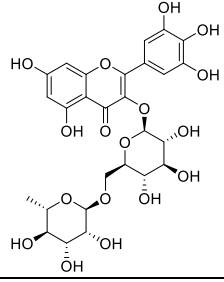
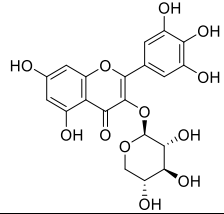
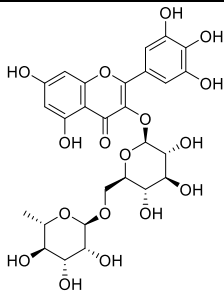
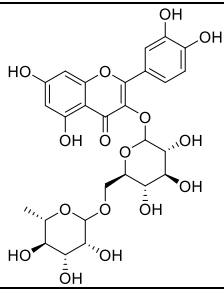
34 216 MS/MS, fragmentation pattern analysis, indicated the characteristic and diagnostic ions of
35
36 217 the annotated metabolites. The glycosylated phenolic compounds found in high concentrations
37
38 218 in the fruits were easily identified by the neutral loss of a sugar moiety, which is determined by
39
40 219 the loss of 162 Da for hexosides, 146 Da for deoxyhexosides and 132 Da for pentosides in *O*-
41
42 220 glycosylated flavonoids (Mannocho-Russo et al., 2020). *C*-glycosylated flavonoids were
43
44 221 characterised by the loss of H₂O and 120 Da (Mannocho-Russo et al., 2020). The compounds
45
46 222 putatively identified as **myricetin 3-*O*-xyloside** and **myricetin 3-rutinoside** at *m/z* 449.073 and
47
48 223 **625.141 [M-H]⁻**, respectively, were identified by the consecutive losses of sugar moieties and
49
50 224 also based on their UV spectra and retention times. Furthermore, the flavonoid aglycone
51
52 225 fragmentation exhibits a pathway characterised by forming fragments originating from retro
53
54
55
56
57
58
59
60
61
62
63
64
65

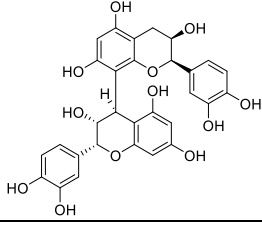
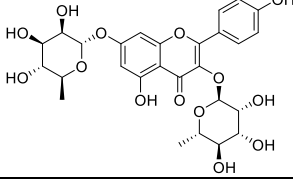
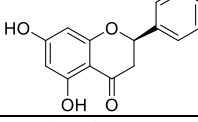
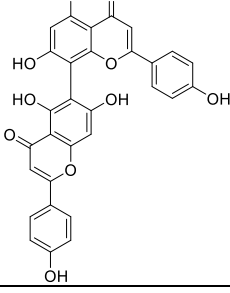
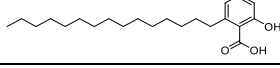
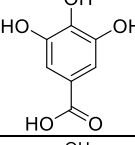
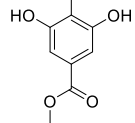
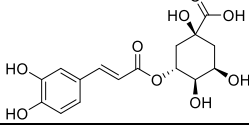
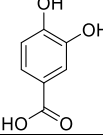
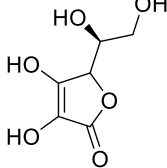
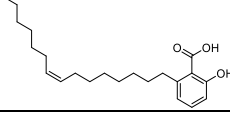
226 Diels-Alder reactions coupled with losses of neutral molecules such as CO₂ and H₂O (Yang et
1 al., 2021). Similarly, the flavonoid monohexosides and dihexosides isoquercitrin, guajavarin,
2 227 al., 2021). Similarly, the flavonoid monohexosides and dihexosides isoquercitrin, guajavarin,
3
4 228 **myricetin 3-*O*-rutinoside, myricetin 3-*O*-galactoside, myricetin 3-*O*-xyloside**, and rutin showed
5
6
7 229 deprotonated molecular ions [M-H]⁻ at *m/z* 463.089, 433.078, 625.141, 479.083, 449.073, and
8
9
10 230 609.146, respectively, previously reported in the literature (**Fig. 1**). Meanwhile, these
11
12 231 compounds were connected with a mass difference of 146 Da, referring to a unit of
13
14 232 deoxyhexoside, 16 Da as a possible extra hydroxyl group, and 30 Da as an extra OCH₂. In total,
15
16
17 233 thirteen *O*-glycosylated flavonoids and four aglycone flavonoids were detected in positive
18
19 234 ionisation mode, and eleven *O*-glycosylated flavonoids and six aglycone flavonoids were
20
21
22 235 detected in negative ionisation mode. The retention time observed in the chromatographic
23
24 236 analysis allowed us to improve the annotation and distinguish the *O*-glycosylated flavonoids.
25
26
27 237 From mass spectrometry fragmentation, it was possible to perform structure-based propagation
28
29 238 and guided detection of metabolites not annotated by the GNPS library since compounds of the
30
31
32 239 same molecular family are structurally related and share molecular substructures that allow the
33
34 240 putative identification of new molecules. Hence, from the difference between the nodes
35
36 241 indicated by *m/z* 609.181 and 463.123 [M+H]⁺, an analogue of the flavone diosmin was
37
38
39 242 annotated, whose mass difference is 146 Da referring to a unit of deoxyhexosides. The detailed
40
41 243 inspection of their MS/MS spectra allowed us to infer the glycosylation type. Through mass
42
43
44 244 spectral similarity networking, accurate annotation of the flavan-3-ols catechin (*m/z* 289.072)
45
46 245 and epigallocatechin (*m/z* 305.067) was achieved in negative ionisation mode [M-H]⁻. Through
47
48
49 246 their retention times, accurate masses, and MS/MS similar fragmentation patterns obtained in
50
51 247 negative mode [M-H]⁻, the methoxylated anthocyanidins 5-methylcyanidin (*m/z* 300.946) and
52
53
54 248 methyldephinidin (*m/z* 316.920) were annotated, which are widely distributed in fruits of the
55
56 249 genus *Anacardium* (de Brito et al., 2007).

58 250

251 **Table 1.** Results of the metabolite annotation in the fruits of *A. humile* through molecular networking
 252 analysis.

Parent mass	Adduct	Molecular formula	Metabolite name	Chemical structure	Chemical class
303.217	[M+H] ⁺	C ₁₅ H ₁₀ O ₇	Quercetin		Flavonoid
289.072	[M-H] ⁻	C ₁₅ H ₁₄ O ₆	Catechin		Flavonoid
595.158	[M+H] ⁺	C ₂₇ H ₃₀ O ₁₅	Kaempferol 3- <i>O</i> -glucoside-7- <i>O</i> -rhamnoside		Flavonoid
449.385	[M+H] ⁺	C ₂₁ H ₂₀ O ₁₁	Kaempferol 7- <i>O</i> -glucoside		Flavonoid
463.123	[M+H] ⁺	C ₂₂ H ₂₃ O ₁₁	Peonidin 3-glucopyranoside		Flavonoid
300.946	[M+H] ⁺	C ₁₆ H ₁₃ O ₆	5-methylcyanidin		Flavonoid
316.920	[M+H] ⁺	C ₁₆ H ₁₃ O ₇	Methyldephinidin		Flavonoid
449.385	[M+H] ⁺	C ₂₁ H ₂₀ O ₁₁	Luteolin 7- <i>O</i> -glucoside		Flavonoid
449.385	[M+H] ⁺	C ₂₁ H ₂₀ O ₁₁	Astragalin		Flavonoid
433.247	[M+H] ⁺	C ₂₁ H ₂₀ O ₁₀	Apigenin glucoside		Flavonoid

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65
627.469	[M+H] ⁺	C ₂₇ H ₃₀ O ₁₇	Quercetin 3,7-di-O-glucoside		Flavonoid																																																											
463.125	[M-H] ⁻	C ₂₁ H ₂₀ O ₁₂	Hyperoside		Flavonoid																																																											
537.168	[M-H] ⁻	C ₃₀ H ₁₈ O ₁₀	Amentoflavone		Flavonoid																																																											
627.156	[M+H] ⁺	C ₂₇ H ₃₀ O ₁₇	Myricetin 3-rutinoside		Flavonoid																																																											
449.073	[M-H] ⁻	C ₂₀ H ₁₈ O ₁₂	Myricetin 3-O-xyloside		Flavonoid																																																											
625.141	[M-H] ⁻	C ₂₇ H ₃₀ O ₁₇	Myricetin 3-rutinoside		Flavonoid																																																											
611.161	[M+H] ⁺	C ₂₇ H ₃₀ O ₁₆	Rutin		Flavonoid																																																											

1						
2						
3	579.145	[M+H] ⁺	C ₃₀ H ₂₆ O ₁₂	Procyanidin B2		Flavonoid
4						
5						
6						
7						
8						
9	579.341	[M+H] ⁺	C ₂₇ H ₃₀ O ₁₄	Kaempferitrin		Flavonoid
10						
11						
12						
13						
14						
15	257.972	[M+H] ⁺	C ₁₅ H ₁₂ O ₄	Pinocembrin		Flavonoid
16						
17						
18						
19						
20						
21						
22	537.168	[M-H] ⁻	C ₃₀ H ₁₈ O ₁₀	Agathisflavone		Flavonoid
23						
24						
25						
26						
27						
28	349.183	[M+H] ⁺	C ₂₂ H ₃₆ O ₃	6-Pentadecylsalicylic Acid		Phenolic acid
29						
30						
31						
32	171.050	[M+H] ⁺	C ₇ H ₆ O ₅	Gallic acid		Phenolic acid
33						
34						
35						
36						
37	185.115	[M+H] ⁺	C ₈ H ₈ O ₅	Methyl gallate		Phenolic acid
38						
39						
40						
41						
42	355.070	[M+H] ⁺	C ₁₆ H ₁₈ O ₉	Chlorogenic acid		Phenolic acid
43						
44						
45						
46						
47	153.019	[M-H] ⁻	C ₇ H ₆ O ₄	Protocatehuic acid		Phenolic acid
48						
49						
50						
51						
52	177.034	[M+H] ⁺	C ₆ H ₈ O ₆	Ascorbic acid		Organic acid
53						
54						
55						
56						
57	345.244	[M-H] ⁻	C ₂₂ H ₃₄ O ₃	Ginkgolic acid		Organic acid
58						
59						
60	253					
61						
62						
63						
64						
65						

254 The systematic investigation of all nodes associated with the flavonoid class allowed us
1
2 255 the annotation propagation of related compounds in the molecular networks from negative
3
4
5 256 ionization mode, such as the galloyl flavonol glycoside derived from quercetin 3-*O*-(6"-
6
7 257 galloyl)-glucopyranoside (m/z 615.100 [M-H]⁻) and the flavanone naringenin (m/z 271.060
8
9
10 258 [M-H]⁻). A detailed assessment of the flavonoid profiling using mass spectral data from two
11
12 259 ionization modes showed other flavonoid structures that were not connected to other nodes
13
14
15 260 from the molecular networking approach, including the flavonoids quercetin (m/z 303.217
16
17 261 [M+H]⁺; empirical formula: C₁₅H₁₀O₇), **kaempferol 7-*O*-glucoside** (m/z 449.385 [M+H]⁺;
18
19 262 C₂₁H₂₀O₁₁), **luteolin 7-*O*-glucoside** (m/z 449.385 [M+H]⁺; C₂₁H₂₀O₁₁), astragalin (m/z 449.385
20
21
22 263 [M+H]⁺; C₂₁H₂₀O₁₁), apigenin glucoside (m/z 433.247 [M+H]⁺; C₂₁H₂₀O₁₀), quercetin 3,7-di-
23
24 264 *O*-glucoside (m/z 627.469 [M+H]⁺; C₂₇H₃₀O₁₇), hyperoside (m/z 463.125 [M-H]⁻; C₂₁H₂₀O₁₂),
25
26
27 265 and the biflavonoid amentoflavone (m/z 537.168 [M-H]⁻; C₃₀H₁₈O₁₀) previously reported in *A.*
28
29 266 *humile* (Lima Júnior et al., 2021; Luiz-Ferreira et al., 2008) and *A. occidentale* (Andarwulan et
30
31
32 267 al., 2012; Salehi et al., 2020; Taiwo et al., 2017). The isomers were separated by their
33
34 268 chromatographic profile, and their precursor ions displayed a different fragmentation pattern.
35
36 269 Applying this same approach, the annotation of low molecular weight metabolites provided
37
38
39 270 some bioactive phenolic acids such as gallic acid (m/z 171.050 [M+H]⁺; C₇H₆O₅), methyl
40
41 271 gallate (m/z 185.115 [M+H]⁺; C₈H₈O₅), protocatechuic acid (m/z 154.980 [M+H]⁺; C₇H₆O₄),
42
43
44 272 and chlorogenic acid (m/z 355.070 [M+H]⁺; C₁₆H₁₈O₉). All these phenolic acids have been
45
46 273 previously reported in the *Anacardium* genus (Andarwulan et al., 2012; Lima Júnior et al.,
47
48
49 274 2021; Luiz-Ferreira et al., 2008). The fragmentation characteristics and biosynthetic knowledge
50
51 275 indicated the oxidation pattern and location of the sugar units.
52

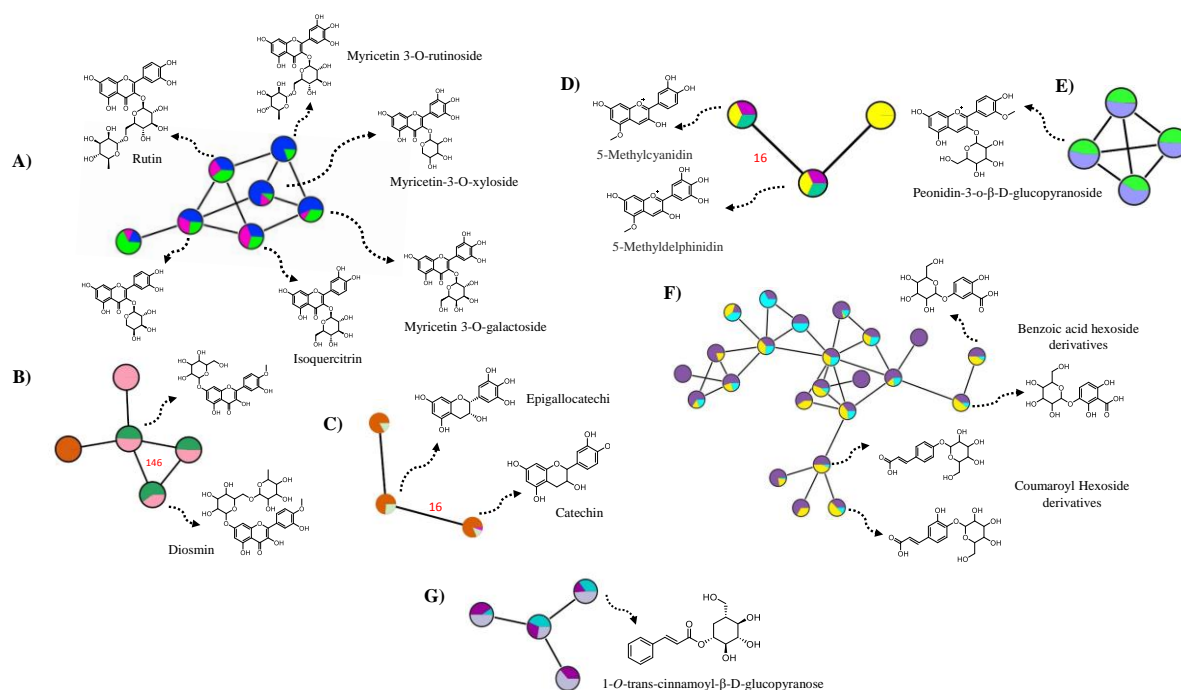
53 276 A more significant number of flavonoids were detected, mainly in the negative ionization
54
55
56 277 mode. A high diversity of flavonoids, mainly *O*-glycosyl flavonols, was confirmed by the large
57
58 278 number of clusters observed in the MN. Combined with detailed taxonomical knowledge,
59
60
61
62
63
64
65

279 allowed for the annotation of metabolites not available in the reference library spectra. MN has
1
2 280 been successfully applied to explore the molecular complexity of *A. humile* fruits and annotate
3
4
5 281 structurally related molecules.
6

7 282 In summary, *O*-glycosylated flavonoids belonging to the subclass of flavonols were
8
9
10 283 dominant primarily based on the structure of myricetin and quercetin and linked mainly to
11
12 284 sugars, glucose and rhamnose. Through the evaluation of the annotations proposed by the MN
13
14
15 285 library and manual inspection of fragmentation spectra, four phenolic acid glycosides were
16
17 286 putatively identified and annotated as benzoic acid hexoside derivatives and coumaroyl
18
19 287 hexoside derivatives (**Fig. 1**). The phenolic acid glycosides classified as cinnamic acid
20
21
22 288 derivatives showed deprotonated molecular ions $[M-H]^-$ at m/z 325.09 (coumaroyl hexoside)
23
24 289 and 341.08 (caffeic acid hexoside); and the benzoic acid hexoside derivatives showed
25
26
27 290 deprotonated molecular ions $[M-H]^-$ at m/z 315.07 (benzoic acid + 2O, O-Hex) and 299.08
28
29 291 (benzoic acid + O, O-Hex) with a mass difference of 16 Da indicating a hydroxyl group as the
30
31
32 292 only structural difference. In addition, its spectrum showed a typical fragment with the
33
34 293 elimination of CO₂. The metabolite annotated as 1-*O*-trans-cinnamoyl-glucofuranose at m/z
35
36 294 309.10 $[M-H]^-$ showed a fragment in MS/MS at m/z 147.04 from neutral loss of the sugar
37
38
39 295 moiety (Glc, 162 Da), previously isolated from cashew apple (*Anacardium occidentale* L.)
40
41 296 (Michodjehoun-Mestres et al., 2009). Its derivative 1-*O*-trans-cinnamoyl-(6-*O*-galloyl)-
42
43
44 297 glucofuranose at m/z 461.110 ($[M-H]^-$) exhibited high spectral similarity and spectral
45
46 298 matching to reference library spectra. The negative ionization mode was more sensitive for
47
48
49 299 detecting these polyphenols.
50

51 300 The presence of other glycosylated polyphenols was investigated in *A. humile* fruits. These
52
53
54 301 analyses found a high abundance of gallotannins (hydrolyzable tannins) formed by polygalloyl
55
56 302 esters of glucose. In this work, two types of gallotannins with degrees of polymerization of 2
57
58 303 (m/z 483.08 $[M-H]^-$) to 3 (m/z 635.089 $[M-H]^-$) galloyl units linked to glucose were putatively
59
60
61
62
63
64
65

304 identified. Their fragmentation patterns were mainly characterized by the loss of 170 Da and
 1 305 152 Da, referring to the loss of a unit of gallic acid and galloyl fission, respectively.



307 **Fig. 1.** The cluster of *O*-glycosylated and aglycone flavonoids is putatively annotated by a molecular
 308 network obtained from MS/MS data of the *A. humile* fruits. The colour inside the nodes indicates the
 309 three different sample locations. (A) S1 = blue, S2 = pink, and S3 = green; (B) S1 = pink, S2 = orange,
 310 and S3 = green; (C) S1 = orange, S2 = nude, and S3 = pink; (D) S1 = yellow, S2 = purple, and S3 =
 311 green; (E) S1 = blue, S2 = green, and S3 = red. The cluster of phenolic acids hexosides is also annotated
 312 by molecular network obtained from MS/MS data of the *A. humile* fruits. (F) S1 = purple, S2 = cyan,
 313 and S3 = yellow; (G) S1 = lavender, S2 = cyan, and S3 = violet.

314
 315 Although gas chromatography-coupled mass spectrometry is the most used technique in
 316 the investigation of chemical markers of *Anacardium*, the analysis by LC-MS/MS with
 317 electrospray ionization enables the detection of the various markers present in the fruits of *A.*
 318 *humile*. The interpretation of MS² data and evaluation of structurally related molecules
 319 annotated by the spectral library and network propagation of *m/z* differences, as well as
 320 retention time analysis, provided the structure of the 16 chemical markers classified as cardols,
 321 anacardic acids, 2-methylcardols, and cardanols with different degrees of unsaturation
 322 (saturated, monounsaturated, di-unsaturated, and tri-unsaturated), confirmed by data reported
 323 in the literature (Gomes Júnior et al., 2018; Kubo et al., 1994). The fragmentation pattern of

324 these metabolites (**Fig. 2**) in negative ionization mode was inspected in detail by observing the
1
2 325 neutral losses, relative abundances, and diagnostic ions based on their network connection. The
3
4
5 326 cluster containing these molecules shows the metabolites of each class with C15-alkyl side-
6
7 327 chain with 0 - 3 double bonds. The chemical marker annotated were **cardols** ((**1**) C_{15:3,5}-
8
9
10 328 [8(Z),ll(Z),14-pentadecatrienyl]resorcinol; (**2**) C_{15:2,5}-[8(Z),ll(Z)-pentadecadienyl]resorcinol;
11
12 329 (**3**) C_{15:1,5}-[8(Z)-pentadecenyl]resorcinol; (**4**) C_{15:0,5}-pentadecylresorcinol), **Anacardic acids**
13
14 330 ((**5**) C_{15:3,6}-[8(Z),11(Z),14-pentadecatrienyl]salicylic acid; (**6**) C_{15:2,6}-[8(Z),ll(Z)-
15
16
17 331 pentadecadienyl]salicylic acid; (**7**) C_{15:2,6}-[8(Z)-pentadecenyl]salicylic acid; (**8**) C_{15:0,6}-
18
19 332 pentadecylsalicylic acid), 2-methylcardols ((**9**) C_{15:3,2}-methyl-5-[8(Z),ll(Z),14-
20
21
22 333 pentadecatrienyl]resorcinol; (**10**) C_{15:2,2}-methyl-5-[8(Z),ll(Z)-pentadecadienyl]resorcinol; (**11**)
23
24 334 C_{15:1,2}-methyl-5-[8(Z)-pentadecenyl]resorcinol; (**12**) C_{15:0,2}-methyl-5-pentadecylresorcinol),
25
26 335 **Cardanols** ((**13**) C_{15:3,3}-[8(Z),ll(Z),14-pentadecatrienyl]phenol; (**14**) C_{15:2,3}-[8(Z),ll(Z)-
27
28
29 336 pentadecadienyl]phenol; (**15**) C_{15:1,3}-[8(Z)-pentadecenyl]phenol; (**16**) C_{15:0,3}-
30
31
32 337 pentadecylphenol) (Gomes Júnior et al., 2018; Kubo et al., 1994). The metabolites were
33
34 338 detected in both positive and negative ionization modes.

35
36 339
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

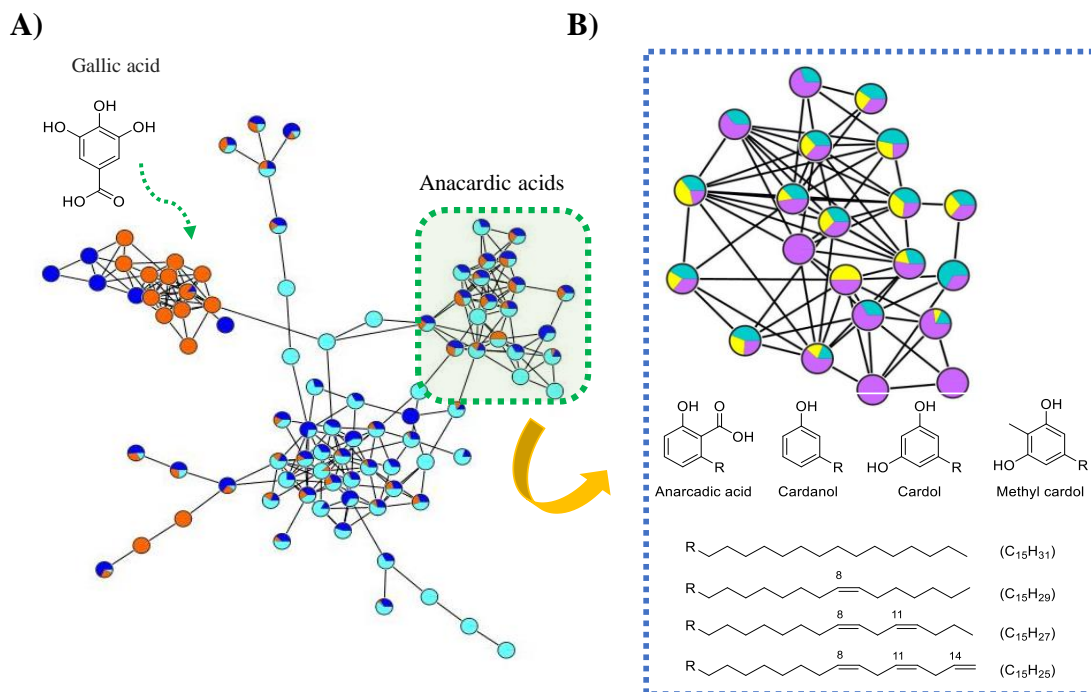


Fig. 2. Cluster of anacardic acids putatively annotated by molecular network obtained from MS/MS data of the *A. humile* fruits. (A) S1 = blue, S2 = orange, and S3 = cyan; (B) S1 = cyan, S2 = yellow, and S3 = purple.

In summary, from the analysis of MN, 91 hits with gold classification were obtained from the GNPS spectral library matching and 80 hits with a cosine score > 0.9 in the positive ionization mode. Using negative ionization mode (ESI(-)), a reduced number of hits was obtained, yielding 38 hits with the gold classification and 21 hits with a cosine score > 0.9. The main classes and subclasses of the metabolites annotated in this analysis are described in **Fig.**

3.

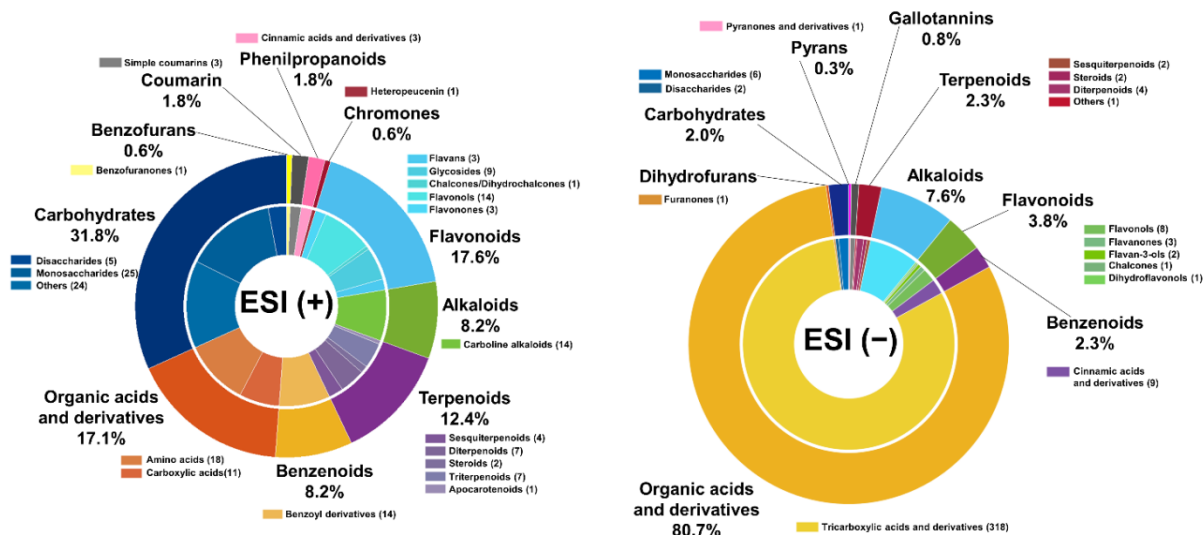


Fig. 3. Chemical classes of the unique library compounds annotated with gold classification and cosine score > 0.9 from both ionization polarities.

Suspect Spectral Library

A new investigation was carried out using the suspect spectral library, a propagated spectral library that provides structural candidates from mass differences and specific modifications of known molecules (Bittremieux et al., 2022). The workflow in positive ionization mode from the suspect spectral library yielded 10806 hits with 569 unique library compounds, 51 with gold classification, and 119 with cosine score > 0.9. The assessment from the library class *in silico* with MZ error < 10 ppm yielded 313 hits, and after inspection of the MS/MS data, metabolites belonging to the class of terpenes, organic acids, and flavonoids were annotated. Only metabolites with many shared peaks, cosine score > 0.9, and common MZ error (ppm) were evaluated. This approach annotated the terpenes β -amyrin acetate and squalene, the organic acid benzoic acid, and the flavonoids epigallocatechin gallate and **myricetin 3-O-hexoside**, previously annotated by the classic MN. In addition, the carbohydrates (maltotriose, **glucopyranose**, **2-phenylethyl 6-O-xylopyranoside**, and galactotriose) were annotated. These were not detected in the classic MN, thus evidencing this tool's efficiency in annotation propagation. On the other hand, in the negative ionization mode, less than 10% of ESI (+) structural candidates were generated. This workflow provided 182 unique library

372 compounds, 37 with a gold classification and 24 with a cosine score > 0.9. Evaluating the same
1
2 373 characteristics described for the ESI (+), only two organic acids and two flavonoids were
3
4
5 374 annotated.

8 375 *Merge Polarity Networks*

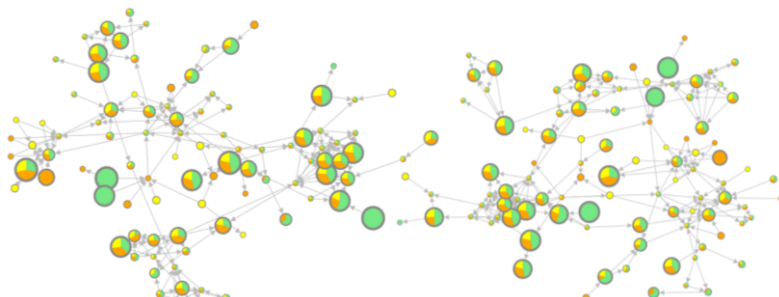
10 376 Notwithstanding the extensive search against the GNPS spectral library and the use of the
11
12
13 377 annotation propagation strategy to explore structural relationships, many ions remained
14
15 378 unannotated. This further motivated the application of more sophisticated dereplication tools.
16
17
18 379 The Merge Polarity Networks function was applied, combining data from both ionization
19
20 380 modes, and yielding more accurate annotations by calculating similarity scores between
21
22
23 381 structures (Tanimoto). This assessment showed 2406 pairs and yielded four relatively abundant
24
25 382 features (m/z 214.09; 163.06; 147.08; 348.07 [M+H]⁺) with Tanimoto values equal to 1 (perfect
26
27 383 agreement), and eight structures were compatible in both ionization modes but were not similar
28
29
30 384 (Tanimoto = -1). This allowed the annotation of kaurenoic acid glucopyranosyl derivative at
31
32
33 385 m/z 645.29 [M+H]⁺, 1,6-anhydro-glucose at m/z 163.06 [M+H]⁺ and the flavonoid 4''-O-
34
35 386 acetylmyricitrin at m/z 627.16 [M+H]⁺.

38 387 *Feature-Based Molecular Networking (FBMN)*

41 388 Although classic MN is very useful for metabolic annotation through spectral library
42
43 389 matching, propagating annotations of unannotated nodes and reducing the size of datasets in
44
45
46 390 untargeted metabolomics, it has some limitations. It does not use MS1 data or provide the
47
48 391 resolution of isomers with different retention times in the chromatographic analysis or coeluting
49
50
51 392 metabolites. Besides, MN does not allow the accurate inference of the relative amounts of the
52
53 393 metabolites (abundance of the ions) between samples. However, these challenges can be solved
54
55 394 by employing Feature-Based Molecular Networking (FBMN), which provides greater
56
57
58 395 resolving power than MN by using MS1 information to verify retention time, peak intensity,

396 ion mobility and isotope patterns (Nothias et al., 2020). Therefore, FBMN was employed to
397 discriminate isomeric/isobaric compounds and infer the ion abundance of some annotated
398 substances and their analogues using chromatographic peak area and peak intensity through
399 feature detection and alignment analysis. FBMN analysis detected 860 features in ESI (+)-MS
400 mode and 340 in ESI (-)-MS. From the relative intensities of the ions, it was possible to infer
401 that the antioxidant compounds gallic acid, ascorbic acid, and anthocyanins were the most
402 abundant metabolites found in *A. humile* (Lima Júnior et al., 2021). In addition, the FBMN
403 provided a highly reliable annotation of polyphenols found in the species (**Fig. 4**).

(A)



(B)

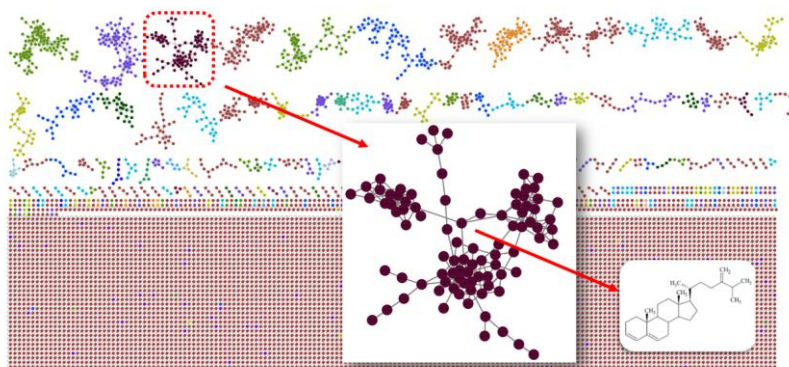


Fig. 4. (A) Full MS/MS-based Feature-Based Molecular Networking (FBMN) in negative ionization mode of the fruit extracts from *A. humile*. Flavonoids are highlighted in the largest cluster. S1 = green, S2 = yellow, and S3 = orange. The node size represents the ions' intensity ratios, and the nodes' colour represents the three locations where *A. humile* fruits were collected. The graphical presentation of the nodes (pie charts) represents the relative abundances of the ions between samples from the different sites. This allowed us to infer the different concentrations of metabolites between the samples and verify the best ionization polarity for the metabolites. (B) Chemical diversity of the secondary metabolites from *A. humile* fruits using the MolNetEnhancer tool acquired in ESI positive mode at the class level.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

415 Other workflows were performed using FBMN data, such as Qemistree software, which
416 predicts molecular profiles from tandem mass spectrometry data using structure source CSI:
417 FingerID and Sirius (Tripathi et al., 2021). This workflow generated 76 hits, and the spectral
418 library revealed the presence of carboxylic acids and derivatives (5 hits), phenol esters (1 hit),
419 keto acids and derivatives (1), organonitrogen compounds (2), steroids and steroid derivatives
420 (1), benzene and substituted derivatives (3), indoles and derivatives (1), and benzofurans (1).

421 *Network Annotation Propagation (NAP)*

422 MS/MS-based metabolomics experiments were analyzed in detail using *in silico*
423 fragmentation tools and complementary methodology such as NAP to improve the reliability
424 of the metabolic annotation suggested by the spectral library to annotate unmatched precursor
425 ions and obtain a highly reliable metabolic profile. The results obtained from the NAP
426 workflow in positive ionization mode were putatively assigned as epigallocatechin, maltose,
427 sucrose, **peonidin 3-O-glucoopyranoside** and corroborated the annotation made by the library.
428 Of the metabolites not annotated by the library (N/A), NAP workflow yielded 64 candidates
429 with consensus with Metfrag (Fusion ID), being the majority belonging to carboxylic acids,
430 glycosylated aldehyde (helicin resulting from the oxidation of the benzylic hydroxy group of
431 salicin to the corresponding aldehyde), chromenes, lactones, quinones (2,6-
432 dimethoxyquinone), triterpenes, and triterpenoid saponin. The MetFrag yielded several
433 candidates belonging to the classes of phenolic acid (benzoic acid derivatives and phenolic
434 compounds with catechol groups, vanillic acid) and terpenoids (diterpene). From metabolites
435 not annotated by the library (N/A), NAP generated 43 candidates with consensus with Metfrag
436 (Fusion ID). Most metabolites were classified as carboxylic acids and derivatives (linear and
437 aromatic acids and lactones), simple phenols, quinones, glycosylated aromatic substances,
438 terpenes, flavonoids (aglycones and heterosides), and phenylpropanoids (phenolic acids). This

439 tool improved the structural annotation based on the GNPS library and provided reliable
1
2 440 candidates from *in silico* spectra matching.
3
4

6 441 *Dereplicator+*

7

8 442 High-confidence identification methods from *in silico* database search were employed in
9
10 443 this study, including Dereplicator+. This high-throughput identification tool improves the
11
12 444 annotation of peptides and small molecules and has been successfully applied in metabolomic
13
14 445 approaches to natural products (Mohimani et al., 2018). The positive ionization mode resulted
15
16 446 in 47 unique compounds, including phenolic acid, chromenes, and terpenes. Terpenes and their
17
18 447 derivatives were the most annotated class of natural products (over 50%). After inspection of
19
20 448 the MS/MS spectra, the tricyclic triterpenoid achilleol B, the triterpenoid betulinaldehyde, the
21
22 449 steroid ergosta-3,5,24(28)-triene, and a chromanol derivative were annotated. In the negative
23
24 450 ionization mode, 15 unique library compounds belonging to flavonoids, tannins, quinones, and
25
26 451 terpenes classes were obtained.
27
28
29
30
31
32

33 452 *MolDiscovery*

34
35

36 453 MolDiscovery was employed to accelerate chemical structural annotation and enhance
37
38 454 the efficiency and accuracy of metabolic identification. This probabilistic model uses a
39
40 455 fragmentation algorithm to generate MS/MS fragmentations and match compounds with their
41
42 456 MS/MS spectra (Cao et al., 2021). As recommended, the analysis of unique metabolites
43
44 457 generated by the workflow with high molecular weight was conducted, as the tool demonstrates
45
46 458 greater accuracy for masses exceeding 600 Da, attributed to the frequently elevated count of
47
48 459 fragment ions. Furthermore, the fragmentation of small molecules considers the type of
49
50 460 fragmented bonds and other factors such as moiety (Cao et al., 2021). MolDiscovery exhibited
51
52 461 2,226 unique metabolites in (+)-ESI mode with high mass ranges >600 Da and a cut-off score
53
54 462 of 15 (159 hits). A larger score implies a greater probability of correct annotation (Cao et al.,
55
56
57
58
59
60
61
62
63
64
65

463 2021). The assessment led to the annotation of a 4-hydroxybenzoic acid derivative (3-
464 decaprenyl-4-hydroxybenzoic acid), the terpene kaurenoic acid glucopyranosyl derivative of
465 m/z 645.29 (Score: 29.35), the flavonoid 4''-*O*-acetylmyricitrin at m/z 627.16 (Score: 26.30),
466 and the steroid periplocoside M of m/z 605.37 (Score: 24.66). The workflow yielded 1205
467 unique metabolites in negative ionization mode, with 185 hits with mass >600 Da. Data
468 inspection was performed only for metabolites of high mass (>400 Da). For large molecules
469 with high mass ranges >600 Da and cut-off score above 15, 82 hits were detected. Hits with the
470 highest score (above 20) were taken into consideration, leading to the putative identification of
471 four flavonoids, encompassing a biflavanone, a coumaroyl isoflavone, a
472 tetrahydroxyflavanone, and a flavone glucopyranoside. The carbohydrate cellotetraose and the
473 terpene grifolinone B 16-deoxo (score 31.28) were identified with high scores. Tannins were
474 the most annotated chemical class, mainly hydrolyzable tannins such as coriariin J (score
475 38.18), ellagitannin puniacortein A (score 32.05), heterophylliin A (score 28.51), and the high
476 molecular weight phillyraeoidin A (m/z 1877.33). This tool proved a reliable and efficient
477 spectral annotation method for high-mass natural products such as tannins and high molecular
478 weight flavonoids such as **tricin 7-O-rutinoside** of m/z 637 (Score: 24.02) annotated in this
479 study.

480 **MS2LDA**

481 To refine the annotation of the observed metabolites, the unsupervised substructure
482 discovery MS2LDA (Van Der Hooft et al., 2016) was explored to discover particular chemical
483 substructures (motifs) obtained from information on fragment peaks and neutral losses from
484 MS² data. The MS2LDA workflow generated 967 MS2LDA motifs for ESI(+) and 1297
485 MS2LDA motifs for ESI(-), which were analyzed in detail to find specific structural features
486 and assist in the structural identification of molecules. The Mass2Motif indicated the presence
487 of hydrolyzable tannins by the neutral losses of 152 and 170 Da (galloyl groups) and neutral

488 losses of 44 (free carboxyl) and 18 (water) from glycosylated derivatives, corroborating the
1
2 489 high abundance of gallotannins and structurally close polyphenols in *A. humile* fruits. In
3
4
5 490 addition, many neutral losses of 162 Da and 146 Da revealed the presence of sugar residue such
6
7 491 as hexose and pentose, which can be attributed to the high abundance of *O*-glycosylated
8
9
10 492 flavonoids found in all samples. The results obtained from the MS2LDA workflow improved
11
12 493 the confidence level of the annotations, expanded the metabolite annotation, and characterized
13
14
15 494 metabolites not yet reported in the species through the analysis of functional groups and core
16
17 495 structures provided by the information from Mass2Motif.
18
19

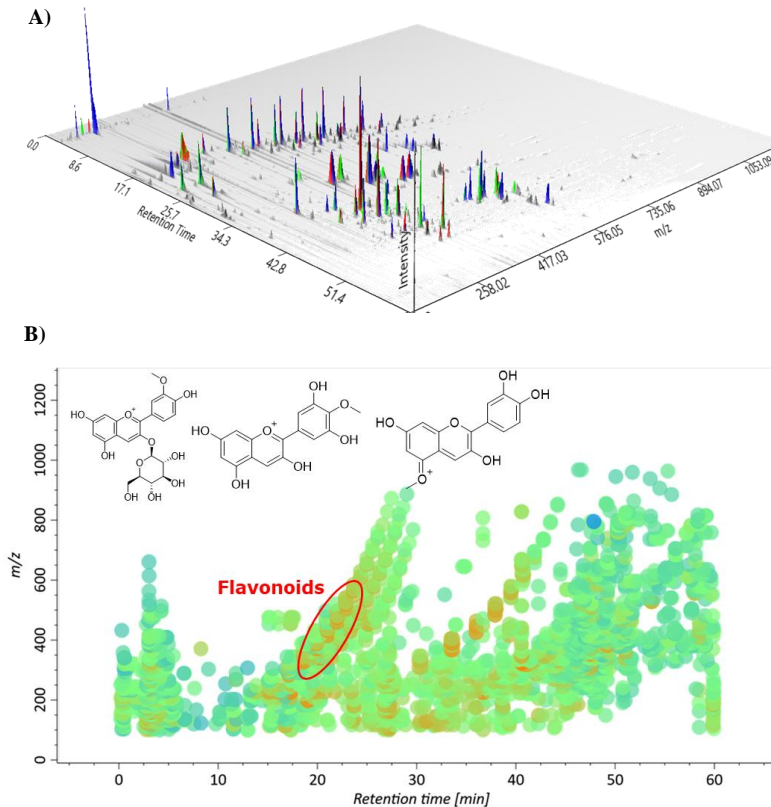
20 496 *MolNetEnhancer*

23 497 An integrated and comprehensive overview of the *A. humile* metabolome was obtained by
24
25 498 combining result outputs from computational metabolomics strategies to explore unannotated
26
27
28 499 features further. Data obtained from other computational metabolomics tools were combined
29
30 500 with the data mining tool MolNetEnhancer, which was used for putative chemical
31
32
33 501 classification, and the clusters were analyzed at superclasses, NP class and direct parent levels.
34
35 502 Chemical structural information from MN (Wang et al., 2016), NAP (da Silva et al., 2018), and
36
37
38 503 Dereplicator+ (Mohimani et al., 2018) were used for the MolNetEnhancer workflow (Ernst et
39
40 504 al., 2019). The most observed metabolite classes at the superclass level were benzenoids and
41
42
43 505 organooxygen compounds in both ionization modes. Analyzing other classification levels, it
44
45 506 was observed that MolNetEnhancer metabolite annotation showed a predominance of
46
47
48 507 secondary metabolites classified as flavonoids (35%), followed by phenolic acids derivatives
49
50 508 (18%), tannins (10%), terpenes (7%), coumarins (4%), and chromenos (2%). This tool
51
52 509 increased the metabolic coverage and allowed the detection of other secondary metabolites not
53
54
55 510 recovered by MN, including steroids (**Fig. 4**).

57 511 There was no significant difference in the metabolic profiling concerning collection sites,
58
59
60 512 as represented in **Fig. 5**. Based on metabolomic profiling obtained of the extracts of *A humile*
61
62
63
64
65

513 fruits from three collection sites, the flavonoids were the major metabolites in all analyses.
514 Considering that flavonoids have remarkable free radical scavenging properties, the flavonoids
515 **peonidin 3-O-glucoopyranoside**, methylcyanidin and methyldephinidin were selected for
516 determination of antioxidant activity using molecular docking.

517



518
519 **Fig. 5.** 3D visualization (A) and spatial ion distributions showing the annotated major metabolites (B)
520 from chromatograms of the extracts of *A. humile* fruits from three collection sites.

522 *Determination of antioxidant activity using molecular docking*

523 As **peonidin 3-O-glucoopyranoside**, methylcyanidin and methyldephinidin are the major
524 metabolites in *A. humile* based on the above metabolomics studies. A molecular docking study
525 was undertaken to assess the antioxidant activity of these compounds. The docking study was
526 carried out on the crystal structure of Human Peroxiredoxin 5 (PRDX5, PDB: 1HD2) (**Declercq**
527 **et al., 2001**), a novel type of *Mammalian peroxiredoxin* that has antioxidative and
528 cytoprotective functions during oxidative stress. The peroxiredoxins define an emerging family

529 of peroxidases able to reduce hydrogen peroxide and alkyl hydroperoxides with the use of
 1
 2 530 reducing equivalents derived from thiol-containing donor molecules such as thioredoxin,
 3
 4 531 glutathione, trypanothione and AhpF. Peroxiredoxins have been identified in prokaryotes as
 5
 6
 7 532 well as in eukaryotes (Arden et al., 2015). Peroxiredoxin 5 is widely expressed in tissues and
 8
 9
 10 533 located cellularly to mitochondria, peroxisomes, and cytosol and, therefore, implicated in
 11
 12 534 antioxidant protective mechanisms (Knoops et al., 2011).

13
 14 535 The three major compounds previously mentioned were subjected to docking analysis, and
 15
 16
 17 536 the specificities of their interaction with these targets, as shown in Fig. 6, were investigated.
 18
 19 537 The best-docked complexes were obtained based on binding energies and interacting residues.

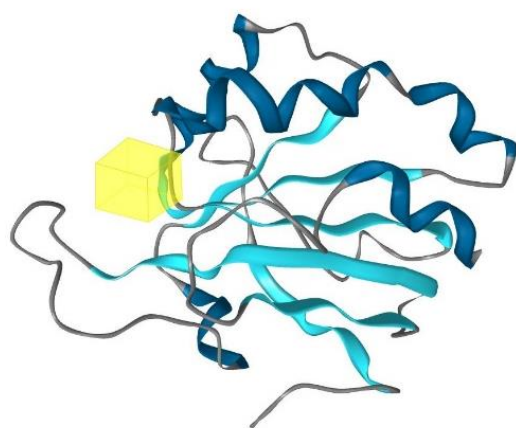
20
 21
 22 538 Docking poses were analyzed and compared to the co-crystallized standard antioxidant
 23
 24 539 nordihydroguaiaretic acid (NDGA). All three major compounds in this molecular docking
 25
 26 540 study docked very well compared to the standard Nordihydroguaiaretic acid (Mala et al., 2020)
 27
 28
 29 541 (Table 2).

30
 31 542

32
 33
 34 543 **Table 2** Results of the metabolite annotation in the fruits of *A. humile*.

Compounds	Docking Score (-) (kcal/mol)
	PDB ID: 1HD2 (Human Peroxiredoxin 5)
Nordihydroguaiaretic acid (Standard)	5.3
Peonidin 3-O-glucoopyranoside	5.02
Methylcyanidin	5.46
Methyldephinidin	6.20

42 544



58 545

60 546 **Fig. 6.** Binding site (yellow colour) of Human Peroxiredoxin 5.

547 *Pharmacophore Evaluation (Structure-based pharmacophore)*

1
2 548 Pharmacophores represent chemical functions. A pharmacophore abstractly describes
3
4
5 549 steric and electronic features required to trigger (or block) biological response (Khedkar et al.,
6
7 550 2007). A pharmacophore model can explain how structurally diverse ligands bind at the
8
9
10 551 receptor site based on common interaction points (Seidel et al., 2019). Structure-based
11
12 552 pharmacophore modelling is a method for pharmacophore development based on the target
13
14
15 553 protein's structural features (Szwabowski et al., 2023). In this method, the possible active site
16
17 554 in protein where the interactions of co-crystallized ligand occur will be analyzed. Due to its
18
19 555 simplicity, this method is computationally very efficient and exceptionally well suited for the
20
21
22 556 virtual screening of a wide range of compound libraries. This method searches for interactions
23
24 557 between ligands and the macromolecule. The three major peonidin 3-O-glucopyranoside,
25
26
27 558 methylcyanidin and methyldephinidin were used for pharmacophore evaluation.

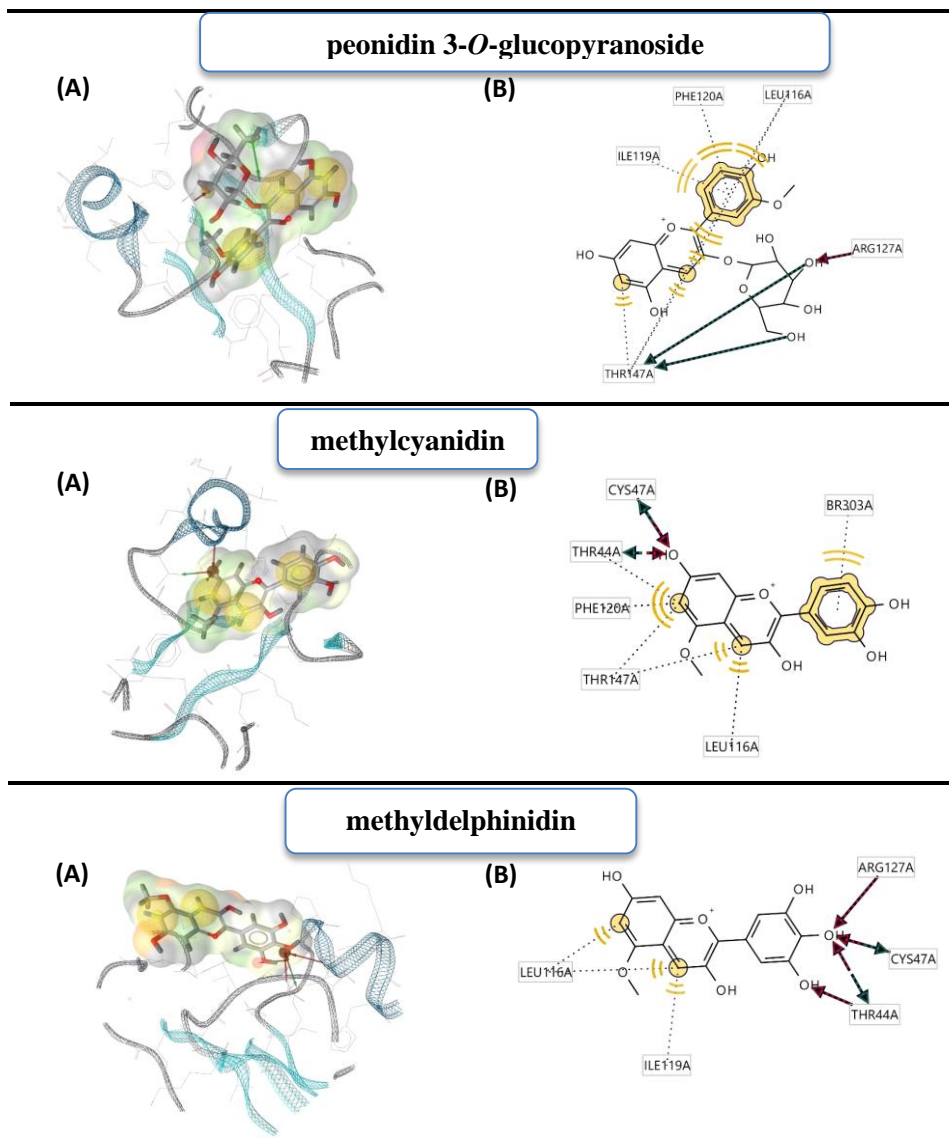
28
29 559 Ligand peonidin 3-O-glucopyranoside interacts with the macromolecule PRDX5, as shown
30
31
32 560 in Fig. 7. It shows hydrophobic effects with amino acids Ile119A and Phe120A. Leu116A and
33
34 561 Thr147A. This pharmacophore also represents a hydrogen bond acceptor (HBA) feature with
35
36 562 nearby amino acid residue Arg127A. Looking at the hydrogen bond donor (HBD) feature, the
37
38
39 563 ligand interacts with Thr147A amino acid residue. Based on the interactions, this structure
40
41 564 provides a pharmacophore with three features hydrophobic effects (H), hydrogen bond acceptor
42
43
44 565 (HBA) and hydrogen bond donor (HBD), with a pharmacophore score of (0.7).

45
46 566 Ligand methylcyanidin interacts with the macromolecule PRDX5 as shown in Fig. 7. It
47
48
49 567 shows hydrophobic effects with BR303A and amino acids Leu116A, Thr147A, Phe120A, and
50
51 568 Thr44A. This pharmacophore has hydrogen bond acceptor (HBA) and hydrogen bond donor
52
53 569 (HBD) features with nearby amino acid residues Cys47A and Thr44A. Based on the
54
55
56 570 interactions, this structure provides a pharmacophore with three features hydrophobic effects

1 571 (H), hydrogen bond acceptor (HBA) and hydrogen bond donor (HBD), with a pharmacophore
2 572 score of (0.89).
3

4
5 573 Ligand methyldephinidin interacts with the macromolecule PRDX5, as shown in **Fig. 7**
6
7 574 (A) and (B). It shows hydrophobic effects with the amino acids Leu116A and Ile119A. This
8
9
10 575 pharmacophore also has both hydrogen bond acceptor (HBA) and hydrogen bond donor (HBD)
11
12 576 features with nearby amino acid residues Cys47A and Thr44A. Arg127A only has hydrogen
13
14
15 577 bond acceptor (HBA) interactions with the ligand. Based on the interactions, this structure
16
17 578 provides a pharmacophore with three features hydrophobic effects (H), hydrogen bond acceptor
18
19 579 (HBA) and hydrogen bond donor (HBD), with a pharmacophore score of (0.87). The above
20
21
22 580 three structure-based pharmacophores further generated a shared feature pharmacophore, as
23
24 581 shown in **Fig. 8**.

25
26
27 582
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65



590 4. Conclusion

1
2
3 591 The application of the integrated metabolomics approach and the combination of structural
4
5 592 annotation and data mining tools provided critical insights into the characterization and
6
7 593 quantification of constituents of the healthy diet, with promising application in food chemistry.
8
9
10 594 This study has revealed a comprehensive metabolic profile and a high content of bioactive
11
12 595 polyphenols in bushy cashew fruits, which hold significant potential from a biotechnological
13
14 596 perspective. Moreover, the state-of-the-art computational tools employed in this study enriched
15
16
17 597 molecular networks, allowing the rapid and effective annotation of a vast array of metabolites
18
19
20 598 not yet reported in *A. humile*. Additionally, molecular docking analysis was performed on
21
22 599 **peonidin 3-O-glucoopyranoside**, methylcyanidin and methyldephinidin to assess their binding
23
24
25 600 to key receptors associated with antioxidant mechanisms. Compared to standard compounds,
26
27 601 these compounds exhibited strong critical potentials related to antioxidant activity. A structure-
28
29 602 based pharmacophore model was proposed to guide future studies, potentially aiding in
30
31 603 selecting and synthesising cost-effective compounds with good bioactivity.
32
33

34 604

37 605 **Declaration of Competing Interest**

38
39 606 The authors declare that they have no known competing financial interests or personal
40
41 607 relationships that could have appeared to influence the work reported in this paper.
42
43

44 608

47 609 **Acknowledgements**

48
49 610 The authors acknowledge financial support from the Coordenação de Aperfeiçoamento de
50
51 611 Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e
52
53 612 Tecnológico (CNPq) for the institutional and financial support.
54
55

56 613

59 614 **References**

615 Andarwulan, N., Kurniasih, D., Apriady, R.A., Rahmat, H., Roto, A.V., Bolling, B.W., 2012.
1
2 616 Polyphenols, carotenoids, and ascorbic acid in underutilized medicinal vegetables. J.
3
4
5 617 Funct. Foods 4, 339–347.
6
7 618 Arden, P., Nelson, K. J., Derek, P., Poole, L.B., Karplus P.A., 2015. Peroxiredoxins: guardians
8
9
10 619 against oxidative stress and modulators of peroxide signaling. Trends Biochem Sci. 40(8),
11
12 620 435–45.
13
14 621 Assunção, R.B., Mercadante, A.Z., 2003. Carotenoids and ascorbic acid from cashew apple
15
16
17 622 (*Anacardium occidentale* L.): variety and geographic effects. Food Chem. 81, 495–502.
18
19 623 Bicalho, B., Pereira, A. S., Aquino Neto, F.R., Pinto, A.C., Rezende, C.M., 2000. Application
20
21
22 624 of high-temperature gas chromatography-mass spectrometry to the investigation of
23
24 625 glycosidically bound components related to cashew apple (*Anacardium occidentale* L. var.
25
26
27 626 nanum) Volatiles. J. Agric. Food Chem. 48, 1167–1174.
28
29 627 Bittremieux, W., Avalon, N.E., Thomas, S.P., Kakhkhorov, S.A., Gauglitz, J.M., Gerwick,
30
31
32 628 W.H., Jarmusch, A.K., Rima, F., 2022. Open Access Repository-Scale Propagated Nearest
33
34 629 Neighbor Suspect Spectral Library for Untargeted Metabolomics. BioRxiv 1–32.
35
36 630 Cao, L., Guler, M., Tagirdzhanov, A., Lee, Y.Y., Gurevich, A., Mohimani, H., 2021.
37
38
39 631 MolDiscovery: learning mass spectrometry fragmentation of small molecules. Nat.
40
41 632 Commun. 12, 1–13.
42
43 633 Da Silva, R.R., Wang, M., Nothias, L.F., van der Hooft, J.J.J., Caraballo-Rodríguez, A.M., Fox,
44
45
46 634 E., Balunas, M.J., Klassen, J.L., Lopes, N.P., Dorrestein, P.C., 2018. Propagating
47
48
49 635 annotations of molecular networks using *in silico* fragmentation. PLoS Comput. Biol. 14,
50
51 636 1–26.
52
53 637 De Brito, E.S., Pessanha de Araújo, M.C., Lin, L.Z., Harnly, J., 2007. Determination of the
54
55
56 638 flavonoid components of cashew apple (*Anacardium occidentale*) by LC-DAD-ESI/MS.
57
58 639 Food Chem. 105, 1112–1118.
59
60
61
62
63
64
65

640 Declercq, J.P., Evrard, C., Clippe, A., Stricht, D.V, Bernard, A., Knoop, B., 2001. Crystal
1 structure of human peroxiredoxin 5, a novel type of mammalian peroxiredoxin at 1.5 Å
2 641 resolution. *J. Mol. Biol.* 311, 751–759.
3
4 642
5
6
7 643 Domínguez, I., Frenich, A.G., Romero-González, R., 2020. Mass spectrometry approaches to
8 ensure food safety. *Anal. Methods* 12, 1148–1162.
9 644
10
11 645 Ernst, M., Kang, K. Bin, Caraballo-Rodríguez, A.M., Nothias, L.F., Wandy, J., Chen, C.,
12 Wang, M., Rogers, S., Medema, M. H., Dorrestein, P. C., Van der Hooft, J. J. J., 2019.
13 646
14 Molnetenhancer: Enhanced molecular networks by integrating metabolome mining and
15 647 annotation tools. *Metabolites* 9, 144.
16
17 648
18
19 649 Gomes Júnior, A.L., Tchekalarova, J.D., Da Conceição Machado, K., Silva Moura, A.K.,
20 Jardim Paz, M.F.C., Ferreira Da Mata, A.M.O., Nogueira, T.R., Islam, M.T., De Sousa
21 650 Rios, M.A., Das Gracias Lopes Cito, A.M., Uddin, S.J., Shilpi, J.A., Das, A.K., Da Silva
22 651 Lopes, L., De Carvalho Melo-Cavalcante, A.A., 2018. Anxiolytic effect of anacardic acids
23 from cashew (*Anacardium occidentale*) nut shell in mice. *IUBMB Life* 70, 420–431.
24 652
25
26 653
27
28 654 Jorge, T. F., Rodrigues, J. A., Caldana, C., Schmidt, R., Dongen, J. T. Van, Thomas-Oates, J.,
29 António, C., 2016. Mass spectrometry- based plant metabolomics: Metabolite responses
30 655 to abiotic stress. *Mass Spectrom. Rev.* 35, 620–649.
31
32 656
33
34 657 Khedkar, S.A., Malde, A.K., Coutinho, E.C., Srivastava, S., 2007. Pharmacophore modeling in
35 drug discovery and development: an overview. *Med Chem.* 3(2), 187-97.
36 658
37
38 659 Knoop, B., Goemaere, J., Van der Eecken, V., Declercq, J.P., 2022. Peroxiredoxin 5: structure,
39 mechanism, and function of the mammalian atypical 2-Cys peroxiredoxin. *Antioxid*
40 660 *Redox Signal.* 15, 817–29.
41 661
42
43 662 Kubo, I., Kinst-Hori, I., Yokokawa, Y., 1994. Tyrosinase inhibitors from *Anacardium*
44 663 *occidentale* fruits. *J. Nat. Prod.* 57, 545–551.
45
46 664
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- 665 *(Anacardium occidentale)* apple juice. J. Agric. Food Chem. 41, 1012–1015.
- 1
- 2 666 Lima Júnior, J.P. de, Franco, R.R., Saraiva, A.L., Moraes, I.B., Espindola, F.S., 2021.
- 3
- 4 667 *Anacardium humile* St. Hil as a novel source of antioxidant, antiglycation and α -amylase
- 5
- 6
- 7 668 inhibitors molecules with potential for management of oxidative stress and diabetes. J.
- 8
- 9 669 Ethnopharmacol. 268, 113667.
- 10
- 11
- 12 670 Lu, H., Zhang, H., Chingin, K., Xiong, J., Fang, X., Chen, H., 2018. Ambient mass
- 13
- 14 671 spectrometry for food science and industry. TrAC, Trends Anal. Chem. 107, 99–115.
- 15
- 16
- 17 672 Luiz-Ferreira, A., Cola-Miranda, M., Barbastefano, V., Hiruma-Lima, C.A., Vilegas, W., Brito,
- 18
- 19 673 A.R.M.S., 2008. Should *Anacardium humile* St. Hil be used as an antiulcer agent? A
- 20
- 21 674 scientific approach to the traditional knowledge. Fitoterapia 79, 207–209.
- 22
- 23
- 24 675 Mala John, G.S., Takeuchi, S., Venkatraman, G., Rayala, S.K. (2002). Nordihydroguaiaretic
- 25
- 26 676 Acid in Therapeutics: Beneficial to Toxicity Profiles and the Search for its Analogs. Curr
- 27
- 28 677 Cancer Drug Targets 20(2), 86–103.
- 29
- 30
- 31 678 Mannocho-Russo, H., Bueno, P.C.P., Bauermeister, A., De Almeida, R.F., Dorrestein, P.C.,
- 32
- 33
- 34 679 Cavalheiro, A.J., Bolzani, V.S., 2020. Can statistical evaluation tools for chromatographic
- 35
- 36 680 method development assist in the natural products workflow? A case study on selected
- 37
- 38 681 species of the plant family Malpighiaceae. J. Nat. Prod. 83, 3239–3249.
- 39
- 40
- 41 682 Michodjehoun-Mestres, L., Amraoui, W., Brillouet, J.M., 2009. Isolation, characterization, and
- 42
- 43 683 determination of 1-*O*-trans-Cinnamoyl- β -D-glucopyranose in the epidermis and flesh of
- 44
- 45 684 developing cashew apple (*Anacardium occidentale* L.) and four of its genotypes. J. Agric.
- 46
- 47 685 Food Chem. 57, 1377–1382.
- 48
- 49
- 50
- 51 686 Mohimani, H., Gurevich, A., Shlemov, A., Mikheenko, A., Korobeynikov, A., Cao, L.,
- 52
- 53 687 Shcherbin, E., Nothias, L. F., Dorrestein, P.C., Pevzner, P.A., 2018. Dereplication of
- 54
- 55 688 microbial metabolites through database search of mass spectra. Nat. Commun. 9, 1–12.
- 56
- 57
- 58 689 Nothias, L.F., Petras, D., Schmid, R., Dührkop, K., Rainer, J., Sarvepalli, A., Protsyuk, I., Ernst,
- 59
- 60
- 61
- 62
- 63
- 64
- 65

690 M., Tsugawa, H., Fleischauer, M., Aicheler, F., Aksenov, A.A., Alka, O., Allard, P.M.,
1
2 691 Barsch, A., Cachet, X., Caraballo-Rodriguez, A.M., Da Silva, R.R., Dang, T., Dorrestein,
3
4 692 P.C., 2020. Feature-based molecular networking in the GNPS analysis environment. Nat.
5
6
7 693 Methods 17, 905–908.
8
9
10 694 Pilon, A.C., Vieira, N.C., Amaral, J.G., Monteiro, A.F., Da Silva, R.R., Spíndola, L.S., Castro-
11
12 695 Gamboa, I., Lopes, N.P., 2021. Molecular networks: An analysis on annotations and
13
14 696 discovery of new assets. Quim. Nova 44, 1168–1179.
15
16
17 697 Ramabulana, A.T., Petras, D., Madala, N.E., Tugizimana, F., 2021. Metabolomics and
18
19 698 molecular networking to characterize the chemical space of four *Momordica* plant species.
20
21 699 Metabolites 11, 763.
22
23
24 700 Royo, V.D.A., Mercadante-Simões, M.O., Ribeiro, L.M., De Oliveira, D.A., Aguiar, M.M. R.,
25
26 701 Costa, E.R., & Ferreira, P.R.B., 2015. Anatomy, histochemistry, and antifungal activity of
27
28 702 *Anacardium humile* (Anacardiaceae) leaf. Microsc. Microanal. 21, 1549–1561.
29
30
31 703 Salehi, B., Gültekin-Özgüven, M., Kirkin, C., Özçelik, B., Morais-Braga, M.F.B., Carneiro, J.
32
33 704 N.P., Bezerra, C.F., Silva, T.G. da, Coutinho, H.D.M., Amina, B., Armstrong, L.,
34
35 705 Selamoglu, Z., Sevindik, M., Yousaf, Z., Sharifi-Rad, J., Muddathir, A.M., Devkota, H.
36
37 706 P., Martorell, M., Jugran, A.K., Martins, N., 2020. Antioxidant, antimicrobial, and
38
39 707 anticancer effects of *Anacardium* plants: An ethnopharmacological perspective. Front.
40
41 708 Endocrinol. 11, 295.
42
43
44 709 Seidel, T., Schuetz, D. A., Garon, A., Langer, T., 2019. The Pharmacophore Concept and Its
45
46 710 Applications in Computer-Aided Drug Design. Prog Chem Org Nat Prod. 110, 99-141.
47
48
49 711 Sumner, L.W., Amberg, A., Barrett, D., Beale, M., Beger, R., Daykin, C., M.Fan, T., Fiehn,
50
51 712 O., Goodacre, R., Griffin, J., Hankemeir, T., Hardy, N., Harnly, J.M., Higashi, R., Kopka,
52
53 713 J., Lane, A., Lindon, J., Marriott, P., Nicholls, A., Reily, M., Thaden, J., Viant, M., 2007.
54
55 714 Proposed minimum reporting standards for chemical analysis. Metabolomics 3, 211–221.
56
57
58
59
60
61
62
63
64
65

715 Szwabowski, G.L., Cole, J.A., Baker, D.L., Parrill, A.L., 2023. Structure-based pharmacophore
1 modeling. Automated random pharmacophore model generation. J Mol Graph Model.
2 716
3
4
5 717 121, 108429.
6
7 718 Taiwo, B.J., Fatokun, A.A., Olubiyi, O.O., Bamigboye-Taiwo, O.T., van Heerden, F.R.,
8
9
10 719 Wright, C.W., 2017. Identification of compounds with cytotoxic activity from the leaf of
11
12 720 the Nigerian medicinal plant, *Anacardium occidentale* L. (Anacardiaceae). Bioorg. Med.
13
14 721 Chem. 25, 2327–2335.
15
16
17 722 Tripathi, A., Vázquez-Baeza, Y., Gauglitz, J.M., Wang, M., Dührkop, K., Nothias-Esposito,
18
19 723 M., Acharya, D.D., Ernst, M., van der Hooft, J.J.J., Zhu, Q., McDonald, D., Brejnrod, A.
20
21 724 D., Gonzalez, A., Handelsman, J., Fleischauer, M., Ludwig, M., Böcker, S., Nothias, L.
22
23
24 725 F., Knight, R., Dorrestein, P.C., 2021. Chemically informed analyses of metabolomics
25
26 726 mass spectrometry data with Qemistree. Nat. Chem. Biol. 17, 146–151.
27
28
29 727 Trott, O., Olson, A.J., 2010. AutoDock Vina: Improving the speed and accuracy of docking
30
31 728 with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem.
32
33 729 31, 455–461.
34
35
36 730 Urzêda, M.A., Marcussi, S., Silva Pereira, L.L., França, S.C., Pereira, A.M.S., Pereira, P.S., Da
37
38
39 731 Silva, S.L., Guimarães, C.L.S., Calderon, L.A., Stábeli, R.G., Soares, A.M., Couto, L.B.,
40
41 732 2013. Evaluation of the hypoglycemic properties of *Anacardium humile* aqueous extract.
42
43 733 Evid.-Based Complement. Alternat. Med. 191080.
44
45
46 734 Van Der Hooft, J.J.J., Wandy, J., Barrett, M.P., Burgess, K.E.V., Rogers, S., 2016. Topic
47
48 735 modeling for untargeted substructure exploration in metabolomics. Proc. Natl. Acad. Sci.
49
50 736 U.S.A. 113, 13738–13743.
51
52
53 737 Wang, M., Carver, J.J., Phelan, V.V., Sanchez, L.M., Garg, N., Peng, Y., Nguyen, D.D.,
54
55
56 738 Watrous, J., Kaponó, C.A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V.,
57
58 739 Meehan, M.J., Liu, W.T., Crüsemann, M., Boudreau, P.D., Esquenazi, E., Sandoval-

740 Calderón, M., Bandeira, N., 2016. Sharing and community curation of mass spectrometry
1
2 741 data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* 34,
3
4
5 742 828–837.
6
7 743 Wolber, G., Langer, T., 2005. LigandScout: 3-D pharmacophores derived from protein-bound
8
9
10 744 ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* 45, 160–169.
11
12 745 Yang, M., Li, J., Zhao, C., Xiao, H., Fang, X., Zheng, J., 2021. LC-Q-TOF-MS/MS detection
13
14 746 of food flavonoids: principle, methodology, and applications. *Crit. Rev. Food Sci. Nutr.*
15
16
17 747 21, 1–21.
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 Mass spectrometry-based untargeted metabolomics approaches for
2 comprehensive structural annotation of bioactive metabolites from bushy
3 cashew (*Anacardium humile*) fruits

4
5 Gabriel F. dos Santos ^a, Nerilson M. Lima ^{a,*}, Gesiane S. Lima ^a, Jussara V. Roque ^a, Gagan
6 Preet ^b, Ernest Oppong-Danquah ^c, Teresinha J. A. S. Andrade ^d, Marcel Jaspars ^b, Boniek
7 Gontijo Vaz ^{a,*}

8
9 ^a Institute of Chemistry, Federal University of Goias, Goiania, 59078-970, GO, Brazil

10 ^b Marine Biodiscovery Centre, Department of Chemistry, University of Aberdeen, Aberdeen
11 AB24 3UE, Scotland, UK

12 ^c GEOMAR Centre for Marine Biotechnology (GEOMAR-Biotech), Research Unit Marine
13 Natural Product Chemistry, GEOMAR Helmholtz Centre for Ocean Research Kiel, Am Kiel-
14 Kanal 44, 24106 Kiel, Germany

15 ^d Instituto Federal de Educação, Ciência e Tecnologia do Maranhão, 65635-468, Presidente
16 Dutra (MA), Brazil

17 *nerilsonmarques@gmail.com (Nerilson M. Lima); boniek@ufg.br (Boniek G. Vaz)

18
19 **Abstract**

20 *Anacardium humile* (bushy cashew) is a native Brazilian plant with substantial pharmacological
21 potential and noteworthy commercial significance in the food industry. This study introduces
22 an untargeted metabolomics approach based on mass spectrometry for the comprehensive
23 structural annotation of bioactive metabolites. The fruits were collected from three distinct sites
24 and subjected to LC-HRMS/MS analysis. A total of eighty-eight compounds were putatively
25 annotated across various metabolite classes. This unveiled a metabolic profile characterized by

26 notable concentrations of polyphenols, including flavonoids, tannins, phenolic acids, and
27 quinones, while aliphatic acids and terpenes were found in limited quantities. Noteworthy, no
28 significant disparities in the metabolic content were observed among the collection sites. The
29 three principal metabolites (peonidin 3-*O*-glucopyranoside, methylcyanidin and
30 methyldephinidin) underwent assessment for antioxidant activity via molecular docking
31 analysis, subsequently generating structure-based feature pharmacophores. The process of
32 annotation propagation yielded a comprehensive qualitative appraisal of the *A. humile*
33 metabolome. The outcomes obtained offer potential candidates for further exploration of their
34 nutraceutical attributes.

36 Keywords

37 *Anacardium humile*; bushy cashew; food chemistry; mass spectrometry; untargeted
38 metabolomics; molecular networking; molecular docking; pharmacophore.

40 1. Introduction

41 *Anacardium humile* St. Hil (Anacardiaceae) is an indigenous medicinal plant from Brazil,
42 popularly known as bushy cashew, with potentially significant health benefits as food for its
43 nutritional, medicinal, and agricultural importance. Various biological activities have been
44 reported for *A. humile*, such as anti-inflammatory, anticancer, antidiarrheal, and antidiabetic
45 properties (Kubo et al., 1993; Luiz-Ferreira et al., 2008; Urzêda et al., 2013). Additionally, stem
46 extracts have been reported to regulate blood glucose, while leaf and bark infusions are used to
47 treat gastric disorders such as ulcers and gastritis (Luiz-Ferreira et al., 2008; Urzêda et al.,
48 2013).

49 The chemical diversity of *A. humile* comprises a significant level of volatile compounds as
50 well as organic acids such as resorcinolic acid, ascorbic acid, and anacardic acids (Assunção &

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51 Mercadante, 2003; Bicalho et al., 2000). Previous studies have indicated the presence of
52 carotenoids (α -carotene, β -carotene, and β -cryptoxanthin), alkaloids, and polyphenols with
53 significant antioxidant properties. The presence of tannins such as (procyanidin B2), phenolic
54 acids (gallic acid), and flavonoids (catechin, glycosylated quercetin) present in *A. humile* have
55 shown antiglycation properties (Assunção & Mercadante, 2003; Lima Júnior et al., 2021; Luiz-
56 Ferreira et al., 2008; Royo et al., 2015). There is currently little information on the metabolic
57 profiling of bushy cashews or the contribution of individual metabolites to the nutritional value.
58 Therefore, research studies aiming at the rapid and comprehensive acquisition of the metabolic
59 profile of these species are necessary.

60 Metabolic coverage assessment to detect the bioactive constituents and identify the
61 metabolites responsible for nutritional properties requires modern analytical tools. Considering
62 that the process of isolation and structural elucidation of the chemical components of complex
63 matrices is a time-consuming and labour-intensive process, metabolomic approaches using
64 advanced analytical technologies and bioinformatics tools have become a necessary pre-
65 requisite in finding potentially new bioactive molecules. Among the techniques employed in
66 these approaches, mass spectrometry is widely acknowledged as the preferred strategy for
67 exploring food chemistry. It allows the analysing of various chemical species with diverse
68 physicochemical properties, even in minute quantities within complex metabolite mixtures
69 (Domínguez et al., 2020; Jorge et al., 2016).

70 Applications of mass spectrometry-based analytical platforms to profile plant metabolomes of
71 food interest have been implemented in routine analysis in research laboratories (Domínguez
72 et al., 2020; Lu et al., 2018). However, the massive amount of spectral data generated requires
73 robust tools for data exploration and organisation to obtain information on the metabolome of
74 the target species. To address challenges in interpreting these large datasets and translating
75 chemical composition into biological knowledge, sophisticated dereplication tools based on

1 76 spectral fragmentation profiles are used. In addition, *in silico* analysis technologies help
2 77 characterise unknown molecules and develop methods for assessing target metabolites. Data
3
4 78 analysis tools, such as the Global Natural Products Social Molecular Networking (GNPS)
5
6
7 79 platform, have allowed the exploration of plant metabolomes as well as assisting and
8
9
10 80 accelerating the discovery of new bioactive agents (Pilon et al., 2021; Ramabulana et al., 2021).

11
12 81 This work aims to apply untargeted metabolomics approaches to screen for bioactive
13
14 82 metabolites in bushy cashew fruits by high-resolution mass spectrometry (HRMS) combined
15
16
17 83 with *in silico* fragmentation tools. It also utilizes the GNPS platform tools, including classic
18
19 84 Molecular Networking (MN), Feature-Based Molecular Networking (FBMN), Network
20
21
22 85 Annotation Propagation (NAP), Dereplicator+, Suspect library, molDiscovery, MS2LDA, and
23
24 86 MolNetEnhancer to enhance metabolite annotation. Subsequently, based on the collective
25
26
27 87 results, pharmacophore and molecular docking analyses were conducted to generate structure-
28
29 88 based feature pharmacophores for the antioxidant activity of the major metabolites. This study
30
31
32 89 represents the first comprehensive exploration of the molecular diversity of *A. humile*, a typical
33
34 90 fruit found in the Brazilian savanna. Furthermore, it recognises the inherent antioxidant
35
36
37 91 properties of the plant's constituents, thereby enhancing its potential value as both a dietary
38
39 92 resource and a pharmaceutical asset.

40 41 42 93 2. Materials and methods

43 44 45 94 2.1. Chemicals and materials

46
47 95 HPLC-grade acetonitrile and methanol were purchased from Tedia Company (Fair-
48
49
50 96 field, USA). Formic acid and Caffeine-¹³C₃ were purchased from Sigma Aldrich (St. Louis,
51
52
53 97 USA). Progesterone-*d*₉ was purchased from CDN Isotopes (Quebec, Canada). Ultrapure water
54
55 98 was produced using a water purification system (Master System MS2000, Gehaka, São Paulo,
56
57 99 Brazil) with a resistivity of 18.2 MΩcm.

60 100

101 2.2. *Sample preparation*

102 *Anacardium humile* fruits were collected from three different locations in Goiás State,
103 Brazil (Goiânia (S1), Uruana (S2), and Campinaçu (S3) cities). Bushy cashews were sliced
104 manually using a sterile knife, frozen at -80 °C, and freeze-dried for 48h. For metabolomics
105 analyses, 10 mg of freeze-dried fruits were extracted with 1 mL of methanol, vortexed for 1
106 min, and centrifuged (10 min, 15000 rpm at room temperature). This solution was diluted with
107 methanol (1:5, v:v) and was transferred to injecting vials for LC-MS/MS analysis. Stable
108 isotopes, Caffeine-¹³C₃ and Progesterone-*d*₉ (2.5 µg mL⁻¹) were used as internal standards.

109

110 2.3. *LC-MS analysis*

111 LC-MS/MS analyses were performed on an HPLC-UV 1220 Infinity II (Agilent
112 Technologies) coupled with a Q-Exactive hybrid Quadrupole-Orbitrap high-resolution mass
113 spectrometer (Thermo Scientific) as well as an electrospray ionisation source. The column used
114 in this study was an InfinityLab Poroshell 120 EC-C18 column (4.6 × 100 mm × 2.7 µm
115 Agilent). All samples were analysed using a gradient elution program. In both ESI positive and
116 negative modes, the binary mobile phase comprised A (water with 0.1% formic acid) and B
117 (methanol). The gradient elution started at 5% (B) and linearly increased to 100% (B) in 40
118 min and kept constant for 10 min at 100% (B). The eluent was then restored to the initial
119 conditions in 10 min. The flow rate was set at 0.3 mL min⁻¹. The injection volume was 30 µL,
120 and the column temperature was 35 °C. The ESI source conditions were set as follows: spray
121 voltage 3.5 kV (in both ionisation modes); the capillary temperature was 250 °C (positive
122 mode) and 320 °C (negative mode); S-lens RF level 60 V (in both ionisation modes); sheath
123 gas flow rate at 47 L min⁻¹ (positive mode) and 35 L min⁻¹ (negative mode); and aux gas flow
124 rate at 11 L min⁻¹ (positive mode) and 10 L min⁻¹ (negative mode). In both ESI positive and
125 negative modes, high-resolution mass spectra were obtained in the Full MS/data dependent -

126 MS² (dd-MS²) mode. The mass range in the full MS scanning experiments was m/z 100-1200.

127 The top 5 (TopN, 5, loop count 5) most abundant precursors were sequentially transferred for

128 collision-induced fragmentation acquisition. The collision energy for target analytes was 20,

129 30, and 35 eV. Resolving power was 140,000 and 70,000 for full MS and dd-MS² acquisitions,

130 respectively.

131

132 2.4. Putative compound annotation

133 The files acquired in the Q-Exactive hybrid Quadrupole-Orbitrap mass spectrometer for

134 the methanolic extracts were converted from raw into (.mzML) format using MSConvert

135 software (ProteoWizard, Palo Alto, CA, US) before being processed using MZmine software,

136 version 2.53. This study utilised metadata to organise compound information following the

137 GNPS online workflow (<https://ccms-ucsd.github.io/GNPSDocumentation/>). This platform

138 also curates MS/MS spectral library categorised based on the quality of the spectra as gold

139 (thoroughly characterised structures), silver (a compound in crude extract), and bronze (partial

140 annotation) (Wang et al., 2016). Metabolite annotations were based on searching the

141 experimental spectra against the GNPS spectral library using the tools classic Molecular

142 Networking – MN (Wang et al., 2016), Feature-Based Molecular Networking – FBMN

143 (Nothias et al., 2020), Dereplicator+ (Mohimani et al., 2018), Network Annotation Propagation

144 – NAP (da Silva et al., 2018), molDiscovery (Cao et al., 2021), MS2LDA (Van Der Hoof et

145 al., 2016), MolNetEnhancer (Ernst et al., 2019), and analysis of chromatographic data such as

146 retention time and UV spectra. These tools allow the integration of orthogonal annotation

147 methodologies and tandem mass spectrometry data (MS/MS) to explore and obtain the

148 metabolome of plants used as food.

149

150 2.5. Molecular Docking

151 Molecular docking analysis was performed using Autodock Vina v.1.2.0 (The Scripps
1
2 152 Research Institute, La Jolla, CA, USA) docking software (Trott & Olson, 2010). The receptor
3
4
5 153 site was predicted using LigandScout (Inte: Ligand) Advanced software (Wolber & Langer,
6
7 154 2005) (evaluation license key: 81809629175371877209), which identifies putative binding
8
9
10 155 pockets by creating a grid surface and calculating the buriedness value of each grid point on
11
12 156 the surface. The resulting pocket grid consists of several clusters of grid points, rendered using
13
14
15 157 an iso surface connecting the grid points. The iso surface represents space that may be suitable
16
17 158 for creating a pocket.

19 159 The x-ray crystal structure of Human Peroxiredoxin 5, a Novel Type of Mammalian
20
21
22 160 Peroxiredoxin (PDB: 1HD2) (Declercq et al., 2001), was retrieved from the Protein Data Bank
23
24 161 and utilised to perform docking simulations. Default search parameters were used where the
25
26
27 162 number of binding modes was 10, exhaustiveness was 8, and the maximum energy difference
28
29 163 was 3 kcal/mol.

31 164 LigandScout (Inte: Ligand) Advanced software (Wolber & Langer, 2005) (evaluation
32
33
34 165 license key: 81809629175371877209) was used to generate 3D pharmacophore models.
35
36
37 166 LigandScout's algorithm calculates and displays chemical interactions between protein–ligand
38
39 167 complexes.

41 168

44 169 3. Results and discussion

46 170 An untargeted metabolomics approach was employed to assess the metabolome of *A.*
47
48
49 171 *humile*. The molecular diversity of its fruits and the evaluation of the abundance of ions and
50
51
52 172 metabolite content were determined using high-resolution mass spectrometry. This was
53
54 173 complemented with dereplication tools, which involved molecular structure searches in robust
55
56
57 174 databases and computer-assisted fragmentation. Furthermore, given the limited data on the
58
59 175 metabolic composition of *A. humile* in the existing literature, the scope of the study was

176 extended to identify biomarkers of the genus *Anacardium* besides describing the classes and
177 subclasses of secondary metabolites present in the species. For this purpose, the structural
178 similarity of MS data was assessed by comparing it with the spectral library from the GNPS
179 ecosystem. Furthermore, potential candidates were identified *in silico* using specialised
180 platforms such as Sirius, Dereplicator+, and NAP, utilising MS data obtained in both positive
181 and negative ionisation modes. The detection of metabolites was performed in fruit samples of
182 *A. humile* from three different locations in Brazil.

183 Putative metabolite annotation was performed based on MS/MS fragmentation patterns,
184 calculation of empirical formulas, chromatographic analysis data such as retention time and
185 UV spectra, inspection of candidates and their analogues suggested by the spectral library,
186 which were further prioritised according to chemotaxonomic and chemophenetic data from
187 family Anacardiaceae (Wang et al., 2016). The library matches were assessed for both positive
188 and negative ionization mode data, and the merge networks polarity tool was employed to
189 explore data from both polarities within a unified workflow.

190 *Classic Molecular Networking (MN)*

191 The library matches using the classic molecular networking (MN) yielded 2164 hits with
192 198 unique library compounds in positive ionisation mode and 457 hits with 139 unique library
193 compounds in negative ionisation mode after excluding nodes with repeated hits. In contrast,
194 the merge networks polarity tool generated 2406 hits. Molecular networking-based strategies
195 allowed us to annotate three times more metabolites in positive ionisation mode than negative
196 mode. According to the Metabolomic Standard Initiative-MSI, all annotations were manually
197 inspected, resulting in level 2 or 3 (Sumner et al., 2007). To obtain hits with higher structural
198 similarity and the exact biosynthetic origin within molecular families, the threshold for the
199 cosine score similarity was set to 0.7. The structural annotations were verified in the
200 metabolome of the genus *Anacardium* using databases such as the Dictionary of Natural

201 Products. Initially, the annotations provided by the spectral library and the molecular family
1
2 202 analysis were evaluated, suggesting precursor ions with similar structures or functional groups.
3

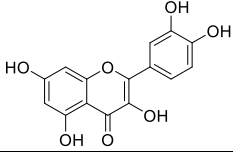
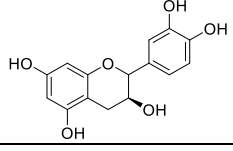
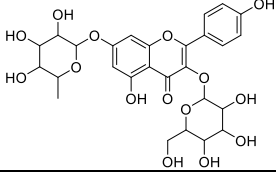
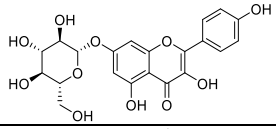
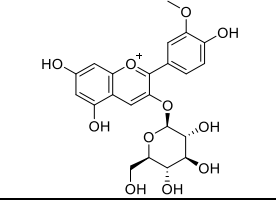
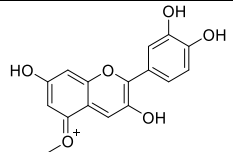
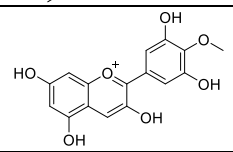
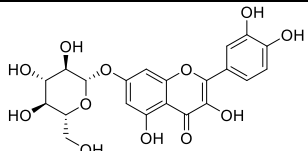
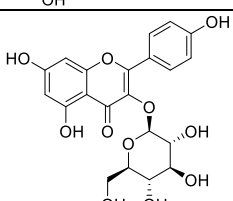
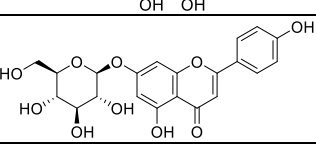
4
5 203 After the dereplication process from the MN analysis, 88 metabolites belonging to different
6
7 204 classes of secondary metabolites in aglycone and heteroside form were putatively annotated,
8
9 205 including terpenes, steroids, flavonoids, tannins, quinones, alkaloids, coumarins, phenolic
10
11 206 acids, and simple phenols. Additionally, many primary metabolites, such as carboxylic acids,
12
13 207 sugars, and amino acids, were annotated in all samples. A high molecular diversity of derivative
14
15 208 compounds has been annotated within these classes (see **Table 1**). In addition, a complex array
16
17 209 of primary metabolites was detected in all samples, such as carbohydrates, amino acids, and
18
19 210 lipids. Based on library matches, polyphenols were the major metabolites in all analyses. *O*-
20
21 211 glycosylated compounds were abundant, and *C*-glycosylated compounds were in the minority.
22
23 212 Alkaloids were detected in low concentrations in the fruits of *A. humile*, which information
24
25 213 corroborates with the data reported for the genus *Anacardium*. Regarding the collection sites,
26
27 214 no substantial variations were observed in metabolite content. All metabolites annotated were
28
29 215 previously described in the family Anacardiaceae.
30
31
32
33
34
35

36 216 MS/MS, fragmentation pattern analysis, indicated the characteristic and diagnostic ions of
37
38 217 the annotated metabolites. The glycosylated phenolic compounds found in high concentrations
39
40 218 in the fruits were easily identified by the neutral loss of a sugar moiety, which is determined by
41
42 219 the loss of 162 Da for hexosides, 146 Da for deoxyhexosides and 132 Da for pentosides in *O*-
43
44 220 glycosylated flavonoids (Mannocho-Russo et al., 2020). *C*-glycosylated flavonoids were
45
46 221 characterised by the loss of H₂O and 120 Da (Mannocho-Russo et al., 2020). The compounds
47
48 222 putatively identified as myricetin 3-*O*-xyloside and myricetin 3-rutinoside at *m/z* 449.073 and
49
50 223 625.141 [M-H]⁻, respectively, were identified by the consecutive losses of sugar moieties and
51
52 224 also based on their UV spectra and retention times. Furthermore, the flavonoid aglycone
53
54 225 fragmentation exhibits a pathway characterised by forming fragments originating from retro
55
56
57
58
59
60
61
62
63
64
65

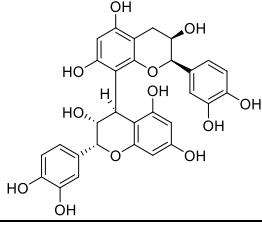
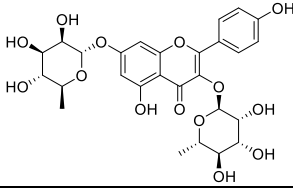
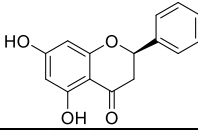
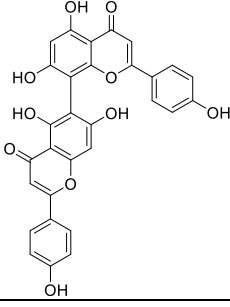
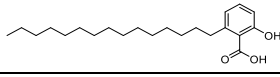
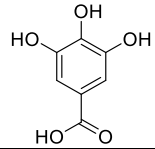
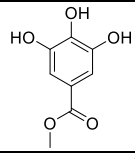
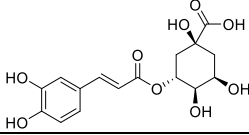
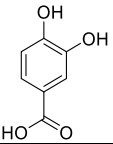
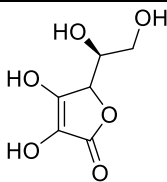
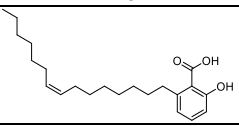
226 Diels-Alder reactions coupled with losses of neutral molecules such as CO₂ and H₂O (Yang et
1 al., 2021). Similarly, the flavonoid monohexosides and dihexosides isoquercitrin, guajavarin,
2 227 al., 2021). Similarly, the flavonoid monohexosides and dihexosides isoquercitrin, guajavarin,
3
4
5 228 myricetin 3-*O*-rutinoside, myricetin 3-*O*-galactoside, myricetin 3-*O*-xyloside, and rutin showed
6
7 229 deprotonated molecular ions [M-H]⁻ at *m/z* 463.089, 433.078, 625.141, 479.083, 449.073, and
8
9
10 230 609.146, respectively, previously reported in the literature (**Fig. 1**). Meanwhile, these
11
12 231 compounds were connected with a mass difference of 146 Da, referring to a unit of
13
14 232 deoxyhexoside, 16 Da as a possible extra hydroxyl group, and 30 Da as an extra OCH₂. In total,
15
16
17 233 thirteen *O*-glycosylated flavonoids and four aglycone flavonoids were detected in positive
18
19 234 ionisation mode, and eleven *O*-glycosylated flavonoids and six aglycone flavonoids were
20
21
22 235 detected in negative ionisation mode. The retention time observed in the chromatographic
23
24 236 analysis allowed us to improve the annotation and distinguish the *O*-glycosylated flavonoids.
25
26
27 237 From mass spectrometry fragmentation, it was possible to perform structure-based propagation
28
29 238 and guided detection of metabolites not annotated by the GNPS library since compounds of the
30
31
32 239 same molecular family are structurally related and share molecular substructures that allow the
33
34 240 putative identification of new molecules. Hence, from the difference between the nodes
35
36 241 indicated by *m/z* 609.181 and 463.123 [M+H]⁺, an analogue of the flavone diosmin was
37
38
39 242 annotated, whose mass difference is 146 Da referring to a unit of deoxyhexosides. The detailed
40
41 243 inspection of their MS/MS spectra allowed us to infer the glycosylation type. Through mass
42
43
44 244 spectral similarity networking, accurate annotation of the flavan-3-ols catechin (*m/z* 289.072)
45
46 245 and epigallocatechin (*m/z* 305.067) was achieved in negative ionisation mode [M-H]⁻. Through
47
48
49 246 their retention times, accurate masses, and MS/MS similar fragmentation patterns obtained in
50
51 247 negative mode [M-H]⁻, the methoxylated anthocyanidins 5-methylcyanidin (*m/z* 300.946) and
52
53
54 248 methyldephinidin (*m/z* 316.920) were annotated, which are widely distributed in fruits of the
55
56 249 genus *Anacardium* (de Brito et al., 2007).

58 250

251 **Table 1.** Results of the metabolite annotation in the fruits of *A. humile* through molecular networking
 252 analysis.

Parent mass	Adduct	Molecular formula	Metabolite name	Chemical structure	Chemical class
303.217	[M+H] ⁺	C ₁₅ H ₁₀ O ₇	Quercetin		Flavonoid
289.072	[M-H] ⁻	C ₁₅ H ₁₄ O ₆	Catechin		Flavonoid
595.158	[M+H] ⁺	C ₂₇ H ₃₀ O ₁₅	Kaempferol 3- <i>O</i> -glucoside-7- <i>O</i> -rhamnoside		Flavonoid
449.385	[M+H] ⁺	C ₂₁ H ₂₀ O ₁₁	Kaempferol 7- <i>O</i> -glucoside		Flavonoid
463.123	[M+H] ⁺	C ₂₂ H ₂₃ O ₁₁	Peonidin 3-glucopyranoside		Flavonoid
300.946	[M+H] ⁺	C ₁₆ H ₁₃ O ₆	5-methylcyanidin		Flavonoid
316.920	[M+H] ⁺	C ₁₆ H ₁₃ O ₇	Methyldephinidin		Flavonoid
449.385	[M+H] ⁺	C ₂₁ H ₂₀ O ₁₁	Luteolin 7- <i>O</i> -glucoside		Flavonoid
449.385	[M+H] ⁺	C ₂₁ H ₂₀ O ₁₁	Astragalin		Flavonoid
433.247	[M+H] ⁺	C ₂₁ H ₂₀ O ₁₀	Apigenin glucoside		Flavonoid

1					
2					
3	627.469	$[M+H]^+$	$C_{27}H_{30}O_{17}$	Quercetin 3,7-di- <i>O</i> -glucoside	Flavonoid
4					
5					
6					
7					
8					
9	463.125	$[M-H]^-$	$C_{21}H_{20}O_{12}$	Hyperoside	Flavonoid
10					
11					
12					
13					
14					
15	537.168	$[M-H]^-$	$C_{30}H_{18}O_{10}$	Amentoflavone	Flavonoid
16					
17					
18					
19					
20					
21					
22					
23					
24					
25	627.156	$[M+H]^+$	$C_{27}H_{30}O_{17}$	Myricetin 3-rutinoside	Flavonoid
26					
27					
28					
29					
30					
31					
32					
33	449.073	$[M-H]^-$	$C_{20}H_{18}O_{12}$	Myricetin 3- <i>O</i> -xyloside	Flavonoid
34					
35					
36					
37					
38					
39					
40					
41					
42	625.141	$[M-H]^-$	$C_{27}H_{30}O_{17}$	Myricetin 3-rutinoside	Flavonoid
43					
44					
45					
46					
47					
48					
49					
50					
51					
52	611.161	$[M+H]^+$	$C_{27}H_{30}O_{16}$	Rutin	Flavonoid
53					
54					
55					
56					
57					
58					
59					
60					
61					
62					
63					
64					
65					

1						
2						
3	579.145	[M+H] ⁺	C ₃₀ H ₂₆ O ₁₂	Procyanidin B2		Flavonoid
4						
5						
6						
7						
8						
9	579.341	[M+H] ⁺	C ₂₇ H ₃₀ O ₁₄	Kaempferitrin		Flavonoid
10						
11						
12						
13						
14						
15	257.972	[M+H] ⁺	C ₁₅ H ₁₂ O ₄	Pinocembrin		Flavonoid
16						
17						
18						
19						
20						
21						
22	537.168	[M-H] ⁻	C ₃₀ H ₁₈ O ₁₀	Agathisflavone		Flavonoid
23						
24						
25						
26						
27						
28	349.183	[M+H] ⁺	C ₂₂ H ₃₆ O ₃	6-Pentadecylsalicylic Acid		Phenolic acid
29						
30						
31						
32	171.050	[M+H] ⁺	C ₇ H ₆ O ₅	Gallic acid		Phenolic acid
33						
34						
35						
36						
37	185.115	[M+H] ⁺	C ₈ H ₈ O ₅	Methyl gallate		Phenolic acid
38						
39						
40						
41						
42	355.070	[M+H] ⁺	C ₁₆ H ₁₈ O ₉	Chlorogenic acid		Phenolic acid
43						
44						
45						
46						
47	153.019	[M-H] ⁻	C ₇ H ₆ O ₄	Protocatehuic acid		Phenolic acid
48						
49						
50						
51						
52	177.034	[M+H] ⁺	C ₆ H ₈ O ₆	Ascorbic acid		Organic acid
53						
54						
55						
56						
57	345.244	[M-H] ⁻	C ₂₂ H ₃₄ O ₃	Ginkgolic acid		Organic acid
58						
59						
60	253					
61						
62						
63						
64						
65						

254 The systematic investigation of all nodes associated with the flavonoid class allowed us
1
2 255 the annotation propagation of related compounds in the molecular networks from negative
3
4
5 256 ionization mode, such as the galloyl flavonol glycoside derived from quercetin 3-*O*-(6"-
6
7 257 galloyl)-glucopyranoside (m/z 615.100 [M-H]⁻) and the flavanone naringenin (m/z 271.060
8
9
10 258 [M-H]⁻). A detailed assessment of the flavonoid profiling using mass spectral data from two
11
12 259 ionization modes showed other flavonoid structures that were not connected to other nodes
13
14
15 260 from the molecular networking approach, including the flavonoids quercetin (m/z 303.217
16
17 261 [M+H]⁺; empirical formula: C₁₅H₁₀O₇), kaempferol 7-*O*-glucoside (m/z 449.385 [M+H]⁺;
18
19 262 C₂₁H₂₀O₁₁), luteolin 7-*O*-glucoside (m/z 449.385 [M+H]⁺; C₂₁H₂₀O₁₁), astragalin (m/z 449.385
20
21 263 [M+H]⁺; C₂₁H₂₀O₁₁), apigenin glucoside (m/z 433.247 [M+H]⁺; C₂₁H₂₀O₁₀), quercetin 3,7-di-
22
23 264 *O*-glucoside (m/z 627.469 [M+H]⁺; C₂₇H₃₀O₁₇), hyperoside (m/z 463.125 [M-H]⁻; C₂₁H₂₀O₁₂),
24
25 265 and the biflavonoid amentoflavone (m/z 537.168 [M-H]⁻; C₃₀H₁₈O₁₀) previously reported in *A.*
26
27 266 *humile* (Lima Júnior et al., 2021; Luiz-Ferreira et al., 2008) and *A. occidentale* (Andarwulan et
28
29 267 al., 2012; Salehi et al., 2020; Taiwo et al., 2017). The isomers were separated by their
30
31 268 chromatographic profile, and their precursor ions displayed a different fragmentation pattern.
32
33
34 269 Applying this same approach, the annotation of low molecular weight metabolites provided
35
36 270 some bioactive phenolic acids such as gallic acid (m/z 171.050 [M+H]⁺; C₇H₆O₅), methyl
37
38 271 gallate (m/z 185.115 [M+H]⁺; C₈H₈O₅), protocatechuic acid (m/z 154.980 [M+H]⁺; C₇H₆O₄),
39
40 272 and chlorogenic acid (m/z 355.070 [M+H]⁺; C₁₆H₁₈O₉). All these phenolic acids have been
41
42 273 previously reported in the *Anacardium* genus (Andarwulan et al., 2012; Lima Júnior et al.,
43
44 274 2021; Luiz-Ferreira et al., 2008). The fragmentation characteristics and biosynthetic knowledge
45
46 275 indicated the oxidation pattern and location of the sugar units.
47
48
49
50
51
52

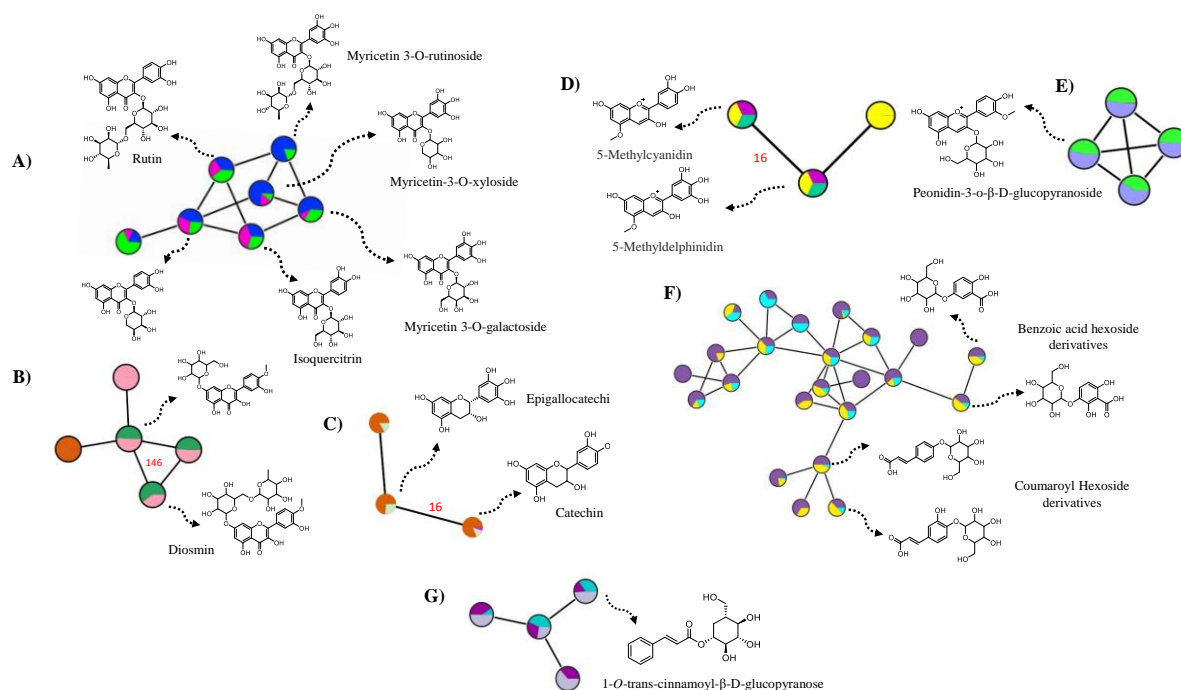
53 276 A more significant number of flavonoids were detected, mainly in the negative ionization
54
55 277 mode. A high diversity of flavonoids, mainly *O*-glycosyl flavonols, was confirmed by the large
56
57 278 number of clusters observed in the MN. Combined with detailed taxonomical knowledge,
58
59
60
61
62
63
64
65

279 allowed for the annotation of metabolites not available in the reference library spectra. MN has
1
2 280 been successfully applied to explore the molecular complexity of *A. humile* fruits and annotate
3
4
5 281 structurally related molecules.
6

7 282 In summary, *O*-glycosylated flavonoids belonging to the subclass of flavonols were
8
9
10 283 dominant primarily based on the structure of myricetin and quercetin and linked mainly to
11
12 284 sugars, glucose and rhamnose. Through the evaluation of the annotations proposed by the MN
13
14
15 285 library and manual inspection of fragmentation spectra, four phenolic acid glycosides were
16
17 286 putatively identified and annotated as benzoic acid hexoside derivatives and coumaroyl
18
19 287 hexoside derivatives (**Fig. 1**). The phenolic acid glycosides classified as cinnamic acid
20
21
22 288 derivatives showed deprotonated molecular ions $[M-H]^-$ at m/z 325.09 (coumaroyl hexoside)
23
24 289 and 341.08 (caffeic acid hexoside); and the benzoic acid hexoside derivatives showed
25
26
27 290 deprotonated molecular ions $[M-H]^-$ at m/z 315.07 (benzoic acid + 2O, O-Hex) and 299.08
28
29 291 (benzoic acid + O, O-Hex) with a mass difference of 16 Da indicating a hydroxyl group as the
30
31
32 292 only structural difference. In addition, its spectrum showed a typical fragment with the
33
34 293 elimination of CO₂. The metabolite annotated as 1-*O*-trans-cinnamoyl-glucofuranose at m/z
35
36 294 309.10 $[M-H]^-$ showed a fragment in MS/MS at m/z 147.04 from neutral loss of the sugar
37
38
39 295 moiety (Glc, 162 Da), previously isolated from cashew apple (*Anacardium occidentale* L.)
40
41 296 (Michodjehoun-Mestres et al., 2009). Its derivative 1-*O*-trans-cinnamoyl-(6-*O*-galloyl)-
42
43
44 297 glucofuranose at m/z 461.110 ($[M-H]^-$) exhibited high spectral similarity and spectral
45
46 298 matching to reference library spectra. The negative ionization mode was more sensitive for
47
48
49 299 detecting these polyphenols.
50

51 300 The presence of other glycosylated polyphenols was investigated in *A. humile* fruits. These
52
53
54 301 analyses found a high abundance of gallotannins (hydrolyzable tannins) formed by polygalloyl
55
56 302 esters of glucose. In this work, two types of gallotannins with degrees of polymerization of 2
57
58 303 (m/z 483.08 $[M-H]^-$) to 3 (m/z 635.089 $[M-H]^-$) galloyl units linked to glucose were putatively
59
60
61
62
63
64
65

304 identified. Their fragmentation patterns were mainly characterized by the loss of 170 Da and
 1 305 152 Da, referring to the loss of a unit of gallic acid and galloyl fission, respectively.



307 **Fig. 1.** The cluster of *O*-glycosylated and aglycone flavonoids is putatively annotated by a molecular
 308 network obtained from MS/MS data of the *A. humile* fruits. The colour inside the nodes indicates the
 309 three different sample locations. (A) S1 = blue, S2 = pink, and S3 = green; (B) S1 = pink, S2 = orange,
 310 and S3 = green; (C) S1 = orange, S2 = nude, and S3 = pink; (D) S1 = yellow, S2 = purple, and S3 =
 311 green; (E) S1 = blue, S2 = green, and S3 = red. The cluster of phenolic acids hexosides is also annotated
 312 by molecular network obtained from MS/MS data of the *A. humile* fruits. (F) S1 = purple, S2 = cyan,
 313 and S3 = yellow; (G) S1 = lavender, S2 = cyan, and S3 = violet.

314
 315 Although gas chromatography-coupled mass spectrometry is the most used technique in
 316 the investigation of chemical markers of *Anacardium*, the analysis by LC-MS/MS with
 317 electrospray ionization enables the detection of the various markers present in the fruits of *A.*
 318 *humile*. The interpretation of MS² data and evaluation of structurally related molecules
 319 annotated by the spectral library and network propagation of *m/z* differences, as well as
 320 retention time analysis, provided the structure of the 16 chemical markers classified as cardols,
 321 anacardic acids, 2-methylcardols, and cardanols with different degrees of unsaturation
 322 (saturated, monounsaturated, di-unsaturated, and tri-unsaturated), confirmed by data reported
 323 in the literature (Gomes Júnior et al., 2018; Kubo et al., 1994). The fragmentation pattern of

324 these metabolites (**Fig. 2**) in negative ionization mode was inspected in detail by observing the
1
2 325 neutral losses, relative abundances, and diagnostic ions based on their network connection. The
3
4
5 326 cluster containing these molecules shows the metabolites of each class with C15-alkyl side-
6
7 327 chain with 0 - 3 double bonds. The chemical marker annotated were **cardols** ((**1**) C_{15:3,5}-
8
9
10 328 [8(Z),ll(Z),14-pentadecatrienyl]resorcinol; (**2**) C_{15:2,5}-[8(Z),ll(Z)-pentadecadienyl]resorcinol;
11
12 329 (**3**) C_{15:1,5}-[8(Z)-pentadecenyl]resorcinol; (**4**) C_{15:0,5}-pentadecylresorcinol), **Anacardic acids**
13
14
15 330 ((**5**) C_{15:3,6}-[8(Z),11(Z),14-pentadecatrienyl]salicylic acid; (**6**) C_{15:2,6}-[8(Z),ll(Z)-
16
17 331 pentadecadienyl]salicylic acid; (**7**) C_{15:2,6}-[8(Z)-pentadecenyl]salicylic acid; (**8**) C_{15:0,6}-
18
19 332 pentadecylsalicylic acid), 2-methylcardols ((**9**) C_{15:3,2}-methyl-5-[8(Z),ll(Z),14-
20
21
22 333 pentadecatrienyl]resorcinol; (**10**) C_{15:2,2}-methyl-5-[8(Z),ll(Z)-pentadecadienyl]resorcinol; (**11**)
23
24 334 C_{15:1,2}-methyl-5-[8(Z)-pentadecenyl]resorcinol; (**12**) C_{15:0,2}-methyl-5-pentadecylresorcinol),
25
26
27 335 **Cardanols** ((**13**) C_{15:3,3}-[8(Z),ll(Z),14-pentadecatrienyl]phenol; (**14**) C_{15:2,3}-[8(Z),ll(Z)-
28
29 336 pentadecadienyl]phenol; (**15**) C_{15:1,3}-[8(Z)-pentadecenyl]phenol; (**16**) C_{15:0,3}-
30
31
32 337 pentadecylphenol) (Gomes Júnior et al., 2018; Kubo et al., 1994). The metabolites were
33
34 338 detected in both positive and negative ionization modes.

35
36 339
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

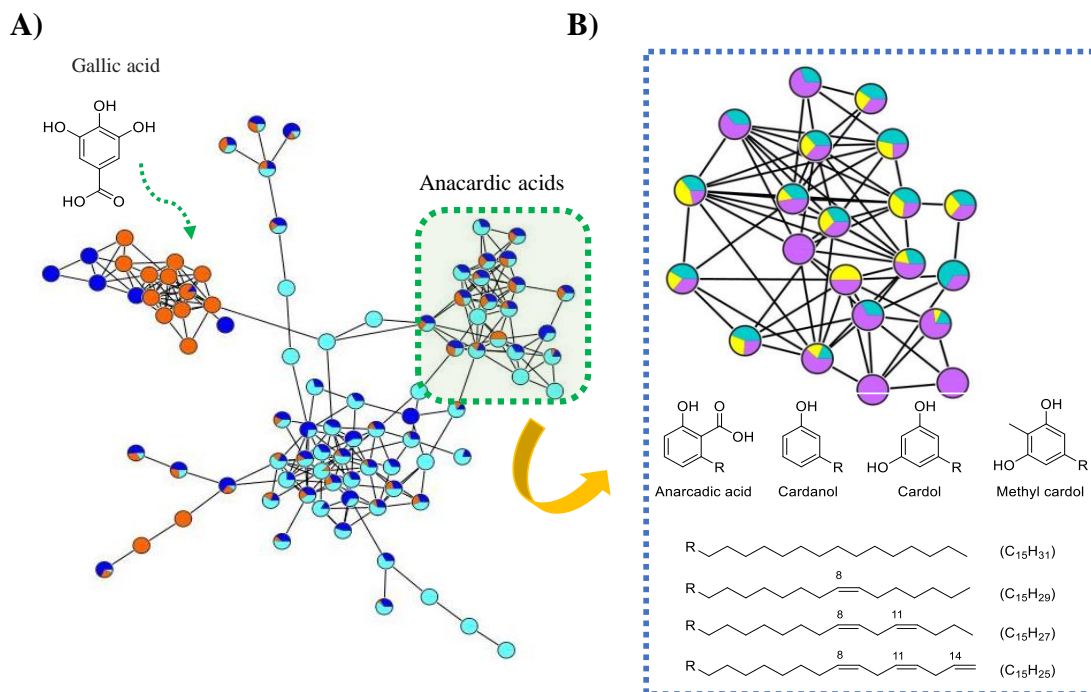


Fig. 2. Cluster of anacardic acids putatively annotated by molecular network obtained from MS/MS data of the *A. humile* fruits. (A) S1 = blue, S2 = orange, and S3 = cyan; (B) S1 = cyan, S2 = yellow, and S3 = purple.

In summary, from the analysis of MN, 91 hits with gold classification were obtained from the GNPS spectral library matching and 80 hits with a cosine score > 0.9 in the positive ionization mode. Using negative ionization mode (ESI(-)), a reduced number of hits was obtained, yielding 38 hits with the gold classification and 21 hits with a cosine score > 0.9. The main classes and subclasses of the metabolites annotated in this analysis are described in Fig.

3.

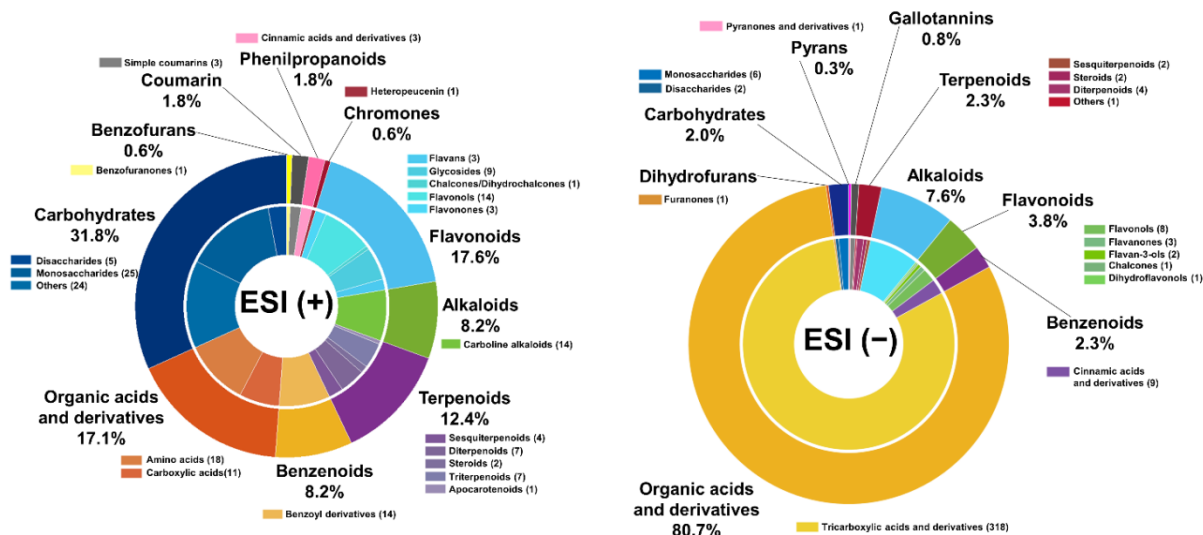


Fig. 3. Chemical classes of the unique library compounds annotated with gold classification and cosine score > 0.9 from both ionization polarities.

Suspect Spectral Library

A new investigation was carried out using the suspect spectral library, a propagated spectral library that provides structural candidates from mass differences and specific modifications of known molecules (Bittremieux et al., 2022). The workflow in positive ionization mode from the suspect spectral library yielded 10806 hits with 569 unique library compounds, 51 with gold classification, and 119 with cosine score > 0.9. The assessment from the library class *in silico* with MZ error < 10 ppm yielded 313 hits, and after inspection of the MS/MS data, metabolites belonging to the class of terpenes, organic acids, and flavonoids were annotated. Only metabolites with many shared peaks, cosine score > 0.9, and common MZ error (ppm) were evaluated. This approach annotated the terpenes β -amyrin acetate and squalene, the organic acid benzoic acid, and the flavonoids epigallocatechin gallate and myricetin 3-*O*-hexoside, previously annotated by the classic MN. In addition, the carbohydrates (maltotriose, glucopyranose, 2-phenylethyl 6-*O*-xylopyranoside, and galactotriose) were annotated. These were not detected in the classic MN, thus evidencing this tool's efficiency in annotation propagation. On the other hand, in the negative ionization mode, less than 10% of ESI (+) structural candidates were generated. This workflow provided 182 unique library

372 compounds, 37 with a gold classification and 24 with a cosine score > 0.9. Evaluating the same
1
2 373 characteristics described for the ESI (+), only two organic acids and two flavonoids were
3
4
5 374 annotated.
6
7

8 375 *Merge Polarity Networks*

9

10 376 Notwithstanding the extensive search against the GNPS spectral library and the use of the
11
12
13 377 annotation propagation strategy to explore structural relationships, many ions remained
14
15 378 unannotated. This further motivated the application of more sophisticated dereplication tools.
16
17
18 379 The Merge Polarity Networks function was applied, combining data from both ionization
19
20
21 380 modes, and yielding more accurate annotations by calculating similarity scores between
22
23 381 structures (Tanimoto). This assessment showed 2406 pairs and yielded four relatively abundant
24
25 382 features (m/z 214.09; 163.06; 147.08; 348.07 [M+H]⁺) with Tanimoto values equal to 1 (perfect
26
27 383 agreement), and eight structures were compatible in both ionization modes but were not similar
28
29
30 384 (Tanimoto = -1). This allowed the annotation of kaurenoic acid glucopyranosyl derivative at
31
32
33 385 m/z 645.29 [M+H]⁺, 1,6-anhydro-glucose at m/z 163.06 [M+H]⁺ and the flavonoid 4''-O-
34
35 386 acetylmyricitrin at m/z 627.16 [M+H]⁺.
36
37

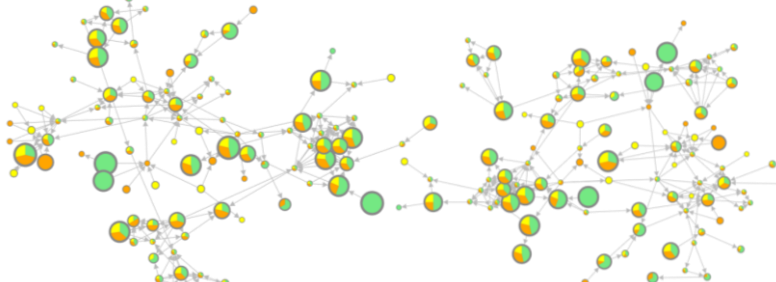
38 387 *Feature-Based Molecular Networking (FBMN)*

39

40
41 388 Although classic MN is very useful for metabolic annotation through spectral library
42
43 389 matching, propagating annotations of unannotated nodes and reducing the size of datasets in
44
45
46 390 untargeted metabolomics, it has some limitations. It does not use MS1 data or provide the
47
48 391 resolution of isomers with different retention times in the chromatographic analysis or coeluting
49
50
51 392 metabolites. Besides, MN does not allow the accurate inference of the relative amounts of the
52
53 393 metabolites (abundance of the ions) between samples. However, these challenges can be solved
54
55 394 by employing Feature-Based Molecular Networking (FBMN), which provides greater
56
57
58 395 resolving power than MN by using MS1 information to verify retention time, peak intensity,
59
60
61
62
63
64
65

396 ion mobility and isotope patterns (Nothias et al., 2020). Therefore, FBMN was employed to
397 discriminate isomeric/isobaric compounds and infer the ion abundance of some annotated
398 substances and their analogues using chromatographic peak area and peak intensity through
399 feature detection and alignment analysis. FBMN analysis detected 860 features in ESI (+)-MS
400 mode and 340 in ESI (-)-MS. From the relative intensities of the ions, it was possible to infer
401 that the antioxidant compounds gallic acid, ascorbic acid, and anthocyanins were the most
402 abundant metabolites found in *A. humile* (Lima Júnior et al., 2021). In addition, the FBMN
403 provided a highly reliable annotation of polyphenols found in the species (**Fig. 4**).

(A)



(B)

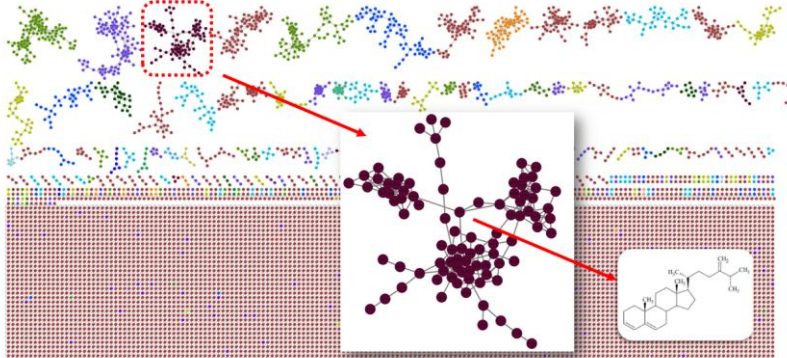


Fig. 4. (A) Full MS/MS-based Feature-Based Molecular Networking (FBMN) in negative ionization mode of the fruit extracts from *A. humile*. Flavonoids are highlighted in the largest cluster. S1 = green, S2 = yellow, and S3 = orange. The node size represents the ions' intensity ratios, and the nodes' colour represents the three locations where *A. humile* fruits were collected. The graphical presentation of the nodes (pie charts) represents the relative abundances of the ions between samples from the different sites. This allowed us to infer the different concentrations of metabolites between the samples and verify the best ionization polarity for the metabolites. (B) Chemical diversity of the secondary metabolites from *A. humile* fruits using the MolNetEnhancer tool acquired in ESI positive mode at the class level.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

415 Other workflows were performed using FBMN data, such as Qemistree software, which
416 predicts molecular profiles from tandem mass spectrometry data using structure source CSI:
417 FingerID and Sirius (Tripathi et al., 2021). This workflow generated 76 hits, and the spectral
418 library revealed the presence of carboxylic acids and derivatives (5 hits), phenol esters (1 hit),
419 keto acids and derivatives (1), organonitrogen compounds (2), steroids and steroid derivatives
420 (1), benzene and substituted derivatives (3), indoles and derivatives (1), and benzofurans (1).

421 *Network Annotation Propagation (NAP)*

422 MS/MS-based metabolomics experiments were analyzed in detail using *in silico*
423 fragmentation tools and complementary methodology such as NAP to improve the reliability
424 of the metabolic annotation suggested by the spectral library to annotate unmatched precursor
425 ions and obtain a highly reliable metabolic profile. The results obtained from the NAP
426 workflow in positive ionization mode were putatively assigned as epigallocatechin, maltose,
427 sucrose, peonidin 3-*O*-glucopyranoside and corroborated the annotation made by the library.
428 Of the metabolites not annotated by the library (N/A), NAP workflow yielded 64 candidates
429 with consensus with Metfrag (Fusion ID), being the majority belonging to carboxylic acids,
430 glycosylated aldehyde (helicin resulting from the oxidation of the benzylic hydroxy group of
431 salicin to the corresponding aldehyde), chromenes, lactones, quinones (2,6-
432 dimethoxyquinone), triterpenes, and triterpenoid saponin. The MetFrag yielded several
433 candidates belonging to the classes of phenolic acid (benzoic acid derivatives and phenolic
434 compounds with catechol groups, vanillic acid) and terpenoids (diterpene). From metabolites
435 not annotated by the library (N/A), NAP generated 43 candidates with consensus with Metfrag
436 (Fusion ID). Most metabolites were classified as carboxylic acids and derivatives (linear and
437 aromatic acids and lactones), simple phenols, quinones, glycosylated aromatic substances,
438 terpenes, flavonoids (aglycones and heterosides), and phenylpropanoids (phenolic acids). This

1
2
3
4
5
6 439 tool improved the structural annotation based on the GNPS library and provided reliable
7
8 440 candidates from *in silico* spectra matching.

9
10
11
12
13
14
15
16 441 *Dereplicator+*

17
18 442 High-confidence identification methods from *in silico* database search were employed in
19
20 443 this study, including *Dereplicator+*. This high-throughput identification tool improves the
21
22 444 annotation of peptides and small molecules and has been successfully applied in metabolomic
23
24 445 approaches to natural products (Mohimani et al., 2018). The positive ionization mode resulted
25
26 446 in 47 unique compounds, including phenolic acid, chromenes, and terpenes. Terpenes and their
27
28 447 derivatives were the most annotated class of natural products (over 50%). After inspection of
29
30 448 the MS/MS spectra, the tricyclic triterpenoid achilleol B, the triterpenoid betulinaldehyde, the
31
32 449 steroid ergosta-3,5,24(28)-triene, and a chromanol derivative were annotated. In the negative
33
34 450 ionization mode, 15 unique library compounds belonging to flavonoids, tannins, quinones, and
35
36 451 terpenes classes were obtained.

37
38
39
40
41
42
43
44 452 *MolDiscovery*

45
46 453 *MolDiscovery* was employed to accelerate chemical structural annotation and enhance
47
48 454 the efficiency and accuracy of metabolic identification. This probabilistic model uses a
49
50 455 fragmentation algorithm to generate MS/MS fragmentations and match compounds with their
51
52 456 MS/MS spectra (Cao et al., 2021). As recommended, the analysis of unique metabolites
53
54 457 generated by the workflow with high molecular weight was conducted, as the tool demonstrates
55
56 458 greater accuracy for masses exceeding 600 Da, attributed to the frequently elevated count of
57
58 459 fragment ions. Furthermore, the fragmentation of small molecules considers the type of
59
60 460 fragmented bonds and other factors such as moiety (Cao et al., 2021). *MolDiscovery* exhibited
61
62 461 2,226 unique metabolites in (+)-ESI mode with high mass ranges >600 Da and a cut-off score
63
64 462 of 15 (159 hits). A larger score implies a greater probability of correct annotation (Cao et al.,
65

2021). The assessment led to the annotation of a 4-hydroxybenzoic acid derivative (3-decaprenyl-4-hydroxybenzoic acid), the terpene kaurenoic acid glucopyranosyl derivative of m/z 645.29 (Score: 29.35), the flavonoid 4''-*O*-acetylmyricitrin at m/z 627.16 (Score: 26.30), and the steroid periplocoside M of m/z 605.37 (Score: 24.66). The workflow yielded 1205 unique metabolites in negative ionization mode, with 185 hits with mass >600 Da. Data inspection was performed only for metabolites of high mass (>400 Da). For large molecules with high mass ranges >600 Da and cut-off score above 15, 82 hits were detected. Hits with the highest score (above 20) were taken into consideration, leading to the putative identification of four flavonoids, encompassing a biflavanone, a coumaroyl isoflavone, a tetrahydroxyflavanone, and a flavone glucopyranoside. The carbohydrate cellotetraose and the terpene grifolinone B 16-deoxo (score 31.28) were identified with high scores. Tannins were the most annotated chemical class, mainly hydrolyzable tannins such as coriariin J (score 38.18), ellagitannin puniacortein A (score 32.05), heterophylliin A (score 28.51), and the high molecular weight phillyraeoidin A (m/z 1877.33). This tool proved a reliable and efficient spectral annotation method for high-mass natural products such as tannins and high molecular weight flavonoids such as tricetin 7-*O*-rutinoside of m/z 637 (Score: 24.02) annotated in this study.

MS2LDA

To refine the annotation of the observed metabolites, the unsupervised substructure discovery MS2LDA (Van Der Hooft et al., 2016) was explored to discover particular chemical substructures (motifs) obtained from information on fragment peaks and neutral losses from MS² data. The MS2LDA workflow generated 967 MS2LDA motifs for ESI(+) and 1297 MS2LDA motifs for ESI(-), which were analyzed in detail to find specific structural features and assist in the structural identification of molecules. The Mass2Motif indicated the presence of hydrolyzable tannins by the neutral losses of 152 and 170 Da (galloyl groups) and neutral

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

488 losses of 44 (free carboxyl) and 18 (water) from glycosylated derivatives, corroborating the
489 high abundance of gallotannins and structurally close polyphenols in *A. humile* fruits. In
490 addition, many neutral losses of 162 Da and 146 Da revealed the presence of sugar residue such
491 as hexose and pentose, which can be attributed to the high abundance of *O*-glycosylated
492 flavonoids found in all samples. The results obtained from the MS2LDA workflow improved
493 the confidence level of the annotations, expanded the metabolite annotation, and characterized
494 metabolites not yet reported in the species through the analysis of functional groups and core
495 structures provided by the information from Mass2Motif.

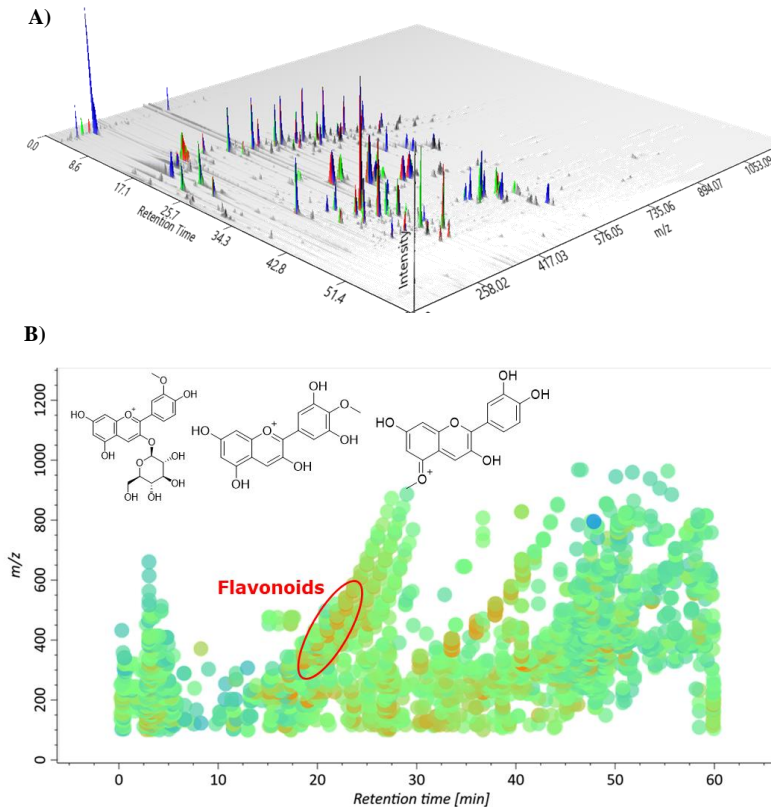
496 *MolNetEnhancer*

497 An integrated and comprehensive overview of the *A. humile* metabolome was obtained by
498 combining result outputs from computational metabolomics strategies to explore unannotated
499 features further. Data obtained from other computational metabolomics tools were combined
500 with the data mining tool MolNetEnhancer, which was used for putative chemical
501 classification, and the clusters were analyzed at superclasses, NP class and direct parent levels.
502 Chemical structural information from MN (Wang et al., 2016), NAP (da Silva et al., 2018), and
503 Dereplicator+ (Mohimani et al., 2018) were used for the MolNetEnhancer workflow (Ernst et
504 al., 2019). The most observed metabolite classes at the superclass level were benzenoids and
505 organooxygen compounds in both ionization modes. Analyzing other classification levels, it
506 was observed that MolNetEnhancer metabolite annotation showed a predominance of
507 secondary metabolites classified as flavonoids (35%), followed by phenolic acids derivatives
508 (18%), tannins (10%), terpenes (7%), coumarins (4%), and chromenos (2%). This tool
509 increased the metabolic coverage and allowed the detection of other secondary metabolites not
510 recovered by MN, including steroids (**Fig. 4**).

511 There was no significant difference in the metabolic profiling concerning collection sites,
512 as represented in **Fig. 5**. Based on metabolomic profiling obtained of the extracts of *A humile*

513 fruits from three collection sites, the flavonoids were the major metabolites in all analyses.
514 Considering that flavonoids have remarkable free radical scavenging properties, the flavonoids
515 peonidin 3-*O*-glucopyranoside, methylcyanidin and methyldelphinidin were selected for
516 determination of antioxidant activity using molecular docking.

517



518
519 **Fig. 5.** 3D visualization (A) and spatial ion distributions showing the annotated major metabolites (B)
520 from chromatograms of the extracts of *A. humile* fruits from three collection sites.

522 *Determination of antioxidant activity using molecular docking*

523 As peonidin 3-*O*-glucopyranoside, methylcyanidin and methyldelphinidin are the major
524 metabolites in *A. humile* based on the above metabolomics studies. A molecular docking study
525 was undertaken to assess the antioxidant activity of these compounds. The docking study was
526 carried out on the crystal structure of Human Peroxiredoxin 5 (PRDX5, PDB: 1HD2) (Declercq
527 et al., 2001), a novel type of *Mammalian peroxiredoxin* that has antioxidative and
528 cytoprotective functions during oxidative stress. The peroxiredoxins define an emerging family

529 of peroxidases able to reduce hydrogen peroxide and alkyl hydroperoxides with the use of
 1
 2 530 reducing equivalents derived from thiol-containing donor molecules such as thioredoxin,
 3
 4 531 glutathione, trypanothione and AhpF. Peroxiredoxins have been identified in prokaryotes as
 5
 6
 7 532 well as in eukaryotes (Arden et al., 2015). Peroxiredoxin 5 is widely expressed in tissues and
 8
 9
 10 533 located cellularly to mitochondria, peroxisomes, and cytosol and, therefore, implicated in
 11
 12 534 antioxidant protective mechanisms (Knoops et al., 2011).

13
 14 535 The three major compounds previously mentioned were subjected to docking analysis, and
 15
 16
 17 536 the specificities of their interaction with these targets, as shown in **Fig. 6**, were investigated.
 18
 19 537 The best-docked complexes were obtained based on binding energies and interacting residues.

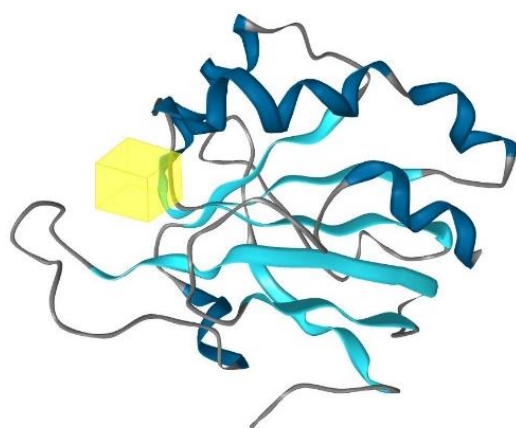
20
 21
 22 538 Docking poses were analyzed and compared to the co-crystallized standard antioxidant
 23
 24 539 nordihydroguaiaretic acid (NDGA). All three major compounds in this molecular docking
 25
 26 540 study docked very well compared to the standard Nordihydroguaiaretic acid (Mala et al., 2020)
 27
 28
 29 541 (**Table 2**).

30
 31 542

32
 33
 34 543 **Table 2** Results of the metabolite annotation in the fruits of *A. humile*.

Compounds	Docking Score (-) (kcal/mol)
	PDB ID: 1HD2 (Human Peroxiredoxin 5)
Nordihydroguaiaretic acid (Standard)	5.3
Peonidin 3- <i>O</i> -glucopyranoside	5.02
Methylcyanidin	5.46
Methyldephinidin	6.20

42 544



58 545

60 546 **Fig. 6.** Binding site (yellow colour) of Human Peroxiredoxin 5.

547 *Pharmacophore Evaluation (Structure-based pharmacophore)*

1
2 548 Pharmacophores represent chemical functions. A pharmacophore abstractly describes
3
4
5 549 steric and electronic features required to trigger (or block) biological response (Khedkar et al.,
6
7 550 2007). A pharmacophore model can explain how structurally diverse ligands bind at the
8
9
10 551 receptor site based on common interaction points (Seidel et al., 2019). Structure-based
11
12 552 pharmacophore modelling is a method for pharmacophore development based on the target
13
14
15 553 protein's structural features (Szwabowski et al., 2023). In this method, the possible active site
16
17 554 in protein where the interactions of co-crystallized ligand occur will be analyzed. Due to its
18
19 555 simplicity, this method is computationally very efficient and exceptionally well suited for the
20
21
22 556 virtual screening of a wide range of compound libraries. This method searches for interactions
23
24 557 between ligands and the macromolecule. The three major peonidin 3-*O*-glucopyranoside,
25
26
27 558 methylcyanidin and methyldephinidin were used for pharmacophore evaluation.

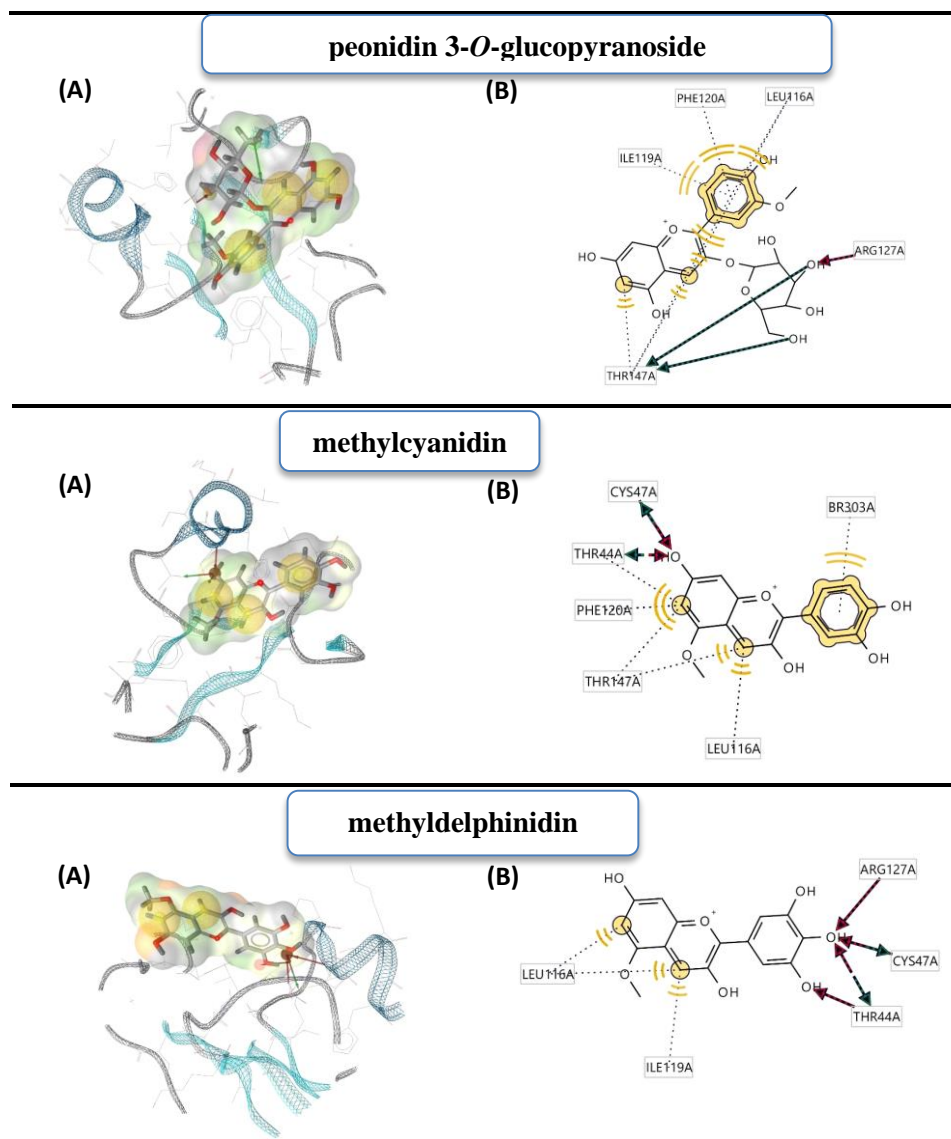
28
29 559 Ligand peonidin 3-*O*-glucopyranoside interacts with the macromolecule PRDX5, as shown
30
31
32 560 in **Fig. 7**. It shows hydrophobic effects with amino acids Ile119A and Phe120A. Leu116A and
33
34 561 Thr147A. This pharmacophore also represents a hydrogen bond acceptor (HBA) feature with
35
36
37 562 nearby amino acid residue Arg127A. Looking at the hydrogen bond donor (HBD) feature, the
38
39 563 ligand interacts with Thr147A amino acid residue. Based on the interactions, this structure
40
41 564 provides a pharmacophore with three features hydrophobic effects (H), hydrogen bond acceptor
42
43
44 565 (HBA) and hydrogen bond donor (HBD), with a pharmacophore score of (0.7).

45
46 566 Ligand methylcyanidin interacts with the macromolecule PRDX5 as shown in **Fig. 7**. It
47
48
49 567 shows hydrophobic effects with BR303A and amino acids Leu116A, Thr147A, Phe120A, and
50
51 568 Thr44A. This pharmacophore has hydrogen bond acceptor (HBA) and hydrogen bond donor
52
53
54 569 (HBD) features with nearby amino acid residues Cys47A and Thr44A. Based on the
55
56 570 interactions, this structure provides a pharmacophore with three features hydrophobic effects

1 571 (H), hydrogen bond acceptor (HBA) and hydrogen bond donor (HBD), with a pharmacophore
2 572 score of (0.89).
3

4
5 573 Ligand methyldephinidin interacts with the macromolecule PRDX5, as shown in **Fig. 7**
6
7 574 (A) and (B). It shows hydrophobic effects with the amino acids Leu116A and Ile119A. This
8
9
10 575 pharmacophore also has both hydrogen bond acceptor (HBA) and hydrogen bond donor (HBD)
11
12 576 features with nearby amino acid residues Cys47A and Thr44A. Arg127A only has hydrogen
13
14
15 577 bond acceptor (HBA) interactions with the ligand. Based on the interactions, this structure
16
17 578 provides a pharmacophore with three features hydrophobic effects (H), hydrogen bond acceptor
18
19 579 (HBA) and hydrogen bond donor (HBD), with a pharmacophore score of (0.87). The above
20
21
22 580 three structure-based pharmacophores further generated a shared feature pharmacophore, as
23
24 581 shown in **Fig. 8**.
25

26
27 582
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65



590 4. Conclusion

1
2
3 591 The application of the integrated metabolomics approach and the combination of structural
4
5 592 annotation and data mining tools provided critical insights into the characterization and
6
7 593 quantification of constituents of the healthy diet, with promising application in food chemistry.
8
9
10 594 This study has revealed a comprehensive metabolic profile and a high content of bioactive
11
12 595 polyphenols in bushy cashew fruits, which hold significant potential from a biotechnological
13
14 596 perspective. Moreover, the state-of-the-art computational tools employed in this study enriched
15
16
17 597 molecular networks, allowing the rapid and effective annotation of a vast array of metabolites
18
19
20 598 not yet reported in *A. humile*. Additionally, molecular docking analysis was performed on
21
22 599 peonidin 3-*O*-glucopyranoside, methylcyanidin and methyldephinidin to assess their binding
23
24
25 600 to key receptors associated with antioxidant mechanisms. Compared to standard compounds,
26
27 601 these compounds exhibited strong critical potentials related to antioxidant activity. A structure-
28
29
30 602 based pharmacophore model was proposed to guide future studies, potentially aiding in
31
32 603 selecting and synthesising cost-effective compounds with good bioactivity.

33
34 604

35 36 37 605 **Declaration of Competing Interest**

38
39 606 The authors declare that they have no known competing financial interests or personal
40
41
42 607 relationships that could have appeared to influence the work reported in this paper.

43
44 608

45 46 47 609 **Acknowledgements**

48
49 610 The authors acknowledge financial support from the Coordenação de Aperfeiçoamento de
50
51 611 Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e
52
53
54 612 Tecnológico (CNPq) for the institutional and financial support.

55
56 613

57 58 59 614 **References**

615 Andarwulan, N., Kurniasih, D., Apriady, R.A., Rahmat, H., Roto, A.V., Bolling, B.W., 2012.
1
2 616 Polyphenols, carotenoids, and ascorbic acid in underutilized medicinal vegetables. J.
3
4
5 617 Funct. Foods 4, 339–347.
6
7 618 Arden, P., Nelson, K. J., Derek, P., Poole, L.B., Karplus P.A., 2015. Peroxiredoxins: guardians
8
9
10 619 against oxidative stress and modulators of peroxide signaling. Trends Biochem Sci. 40(8),
11
12 620 435–45.
13
14 621 Assunção, R.B., Mercadante, A.Z., 2003. Carotenoids and ascorbic acid from cashew apple
15
16 622 (*Anacardium occidentale* L.): variety and geographic effects. Food Chem. 81, 495–502.
17
18
19 623 Bicalho, B., Pereira, A. S., Aquino Neto, F.R., Pinto, A.C., Rezende, C.M., 2000. Application
20
21 624 of high-temperature gas chromatography-mass spectrometry to the investigation of
22
23 625 glycosidically bound components related to cashew apple (*Anacardium occidentale* L. var.
24
25 626 nanum) Volatiles. J. Agric. Food Chem. 48, 1167–1174.
26
27
28
29 627 Bittremieux, W., Avalon, N.E., Thomas, S.P., Kakhkhorov, S.A., Gauglitz, J.M., Gerwick,
30
31 628 W.H., Jarmusch, A.K., Rima, F., 2022. Open Access Repository-Scale Propagated Nearest
32
33 629 Neighbor Suspect Spectral Library for Untargeted Metabolomics. BioRxiv 1–32.
34
35
36 630 Cao, L., Guler, M., Tagirdzhanov, A., Lee, Y.Y., Gurevich, A., Mohimani, H., 2021.
37
38 631 MolDiscovery: learning mass spectrometry fragmentation of small molecules. Nat.
39
40 632 Commun. 12, 1–13.
41
42
43 633 Da Silva, R.R., Wang, M., Nothias, L.F., van der Hooft, J.J.J., Caraballo-Rodríguez, A.M., Fox,
44
45 634 E., Balunas, M.J., Klassen, J.L., Lopes, N.P., Dorrestein, P.C., 2018. Propagating
46
47 635 annotations of molecular networks using *in silico* fragmentation. PLoS Comput. Biol. 14,
48
49 636 1–26.
50
51
52
53 637 De Brito, E.S., Pessanha de Araújo, M.C., Lin, L.Z., Harnly, J., 2007. Determination of the
54
55 638 flavonoid components of cashew apple (*Anacardium occidentale*) by LC-DAD-ESI/MS.
56
57 639 Food Chem. 105, 1112–1118.
58
59
60
61
62
63
64
65

- 640 Declercq, J.P., Evrard, C., Clippe, A., Stricht, D.V, Bernard, A., Knoops, B., 2001. Crystal
1 structure of human peroxiredoxin 5, a novel type of mammalian peroxiredoxin at 1.5 Å
2
3 641
4
5 642 resolution. *J. Mol. Biol.* 311, 751–759.
6
- 7 643 Domínguez, I., Frenich, A.G., Romero-González, R., 2020. Mass spectrometry approaches to
8
9 644 ensure food safety. *Anal. Methods* 12, 1148–1162.
10
11 645 Ernst, M., Kang, K. Bin, Caraballo-Rodríguez, A.M., Nothias, L.F., Wandy, J., Chen, C.,
12
13 646 Wang, M., Rogers, S., Medema, M. H., Dorrestein, P. C., Van der Hooft, J. J. J., 2019.
14
15 647 Molnetenhancer: Enhanced molecular networks by integrating metabolome mining and
16
17 648 annotation tools. *Metabolites* 9, 144.
18
19
20
21 649 Gomes Júnior, A.L., Tchekalarova, J.D., Da Conceição Machado, K., Silva Moura, A.K.,
22
23 650 Jardim Paz, M.F.C., Ferreira Da Mata, A.M.O., Nogueira, T.R., Islam, M.T., De Sousa
24
25 651 Rios, M.A., Das Gracias Lopes Cito, A.M., Uddin, S.J., Shilpi, J.A., Das, A.K., Da Silva
26
27 652 Lopes, L., De Carvalho Melo-Cavalcante, A.A., 2018. Anxiolytic effect of anacardic acids
28
29 653 from cashew (*Anacardium occidentale*) nut shell in mice. *IUBMB Life* 70, 420–431.
30
31
32
33 654 Jorge, T. F., Rodrigues, J. A., Caldana, C., Schmidt, R., Dongen, J. T. Van, Thomas-Oates, J.,
34
35 655 António, C., 2016. Mass spectrometry- based plant metabolomics: Metabolite responses
36
37 656 to abiotic stress. *Mass Spectrom. Rev.* 35, 620–649.
38
39
40
41 657 Khedkar, S.A., Malde, A.K., Coutinho, E.C., Srivastava, S., 2007. Pharmacophore modeling in
42
43 658 drug discovery and development: an overview. *Med Chem.* 3(2), 187-97.
44
45
46 659 Knoops, B., Goemaere, J., Van der Eecken, V., Declercq, J.P., 2022. Peroxiredoxin 5: structure,
47
48 660 mechanism, and function of the mammalian atypical 2-Cys peroxiredoxin. *Antioxid*
49
50 661 *Redox Signal.* 15, 817–29.
51
52
53 662 Kubo, I., Kinst-Hori, I., Yokokawa, Y., 1994. Tyrosinase inhibitors from *Anacardium*
54
55 663 *occidentale* fruits. *J. Nat. Prod.* 57, 545–551.
56
57
58 664 Kubo, I., Ochi, M., Vieira, P.C., Komatsu, S., 1993. Antitumor agents from the cashew
59
60
61
62
63
64
65

665 (*Anacardium occidentale*) apple juice. J. Agric. Food Chem. 41, 1012–1015.

666 Lima Júnior, J.P. de, Franco, R.R., Saraiva, A.L., Moraes, I.B., Espindola, F.S., 2021.

667 *Anacardium humile* St. Hil as a novel source of antioxidant, antiglycation and α -amylase

668 inhibitors molecules with potential for management of oxidative stress and diabetes. J.

669 Ethnopharmacol. 268, 113667.

670 Lu, H., Zhang, H., Chingin, K., Xiong, J., Fang, X., Chen, H., 2018. Ambient mass

671 spectrometry for food science and industry. TrAC, Trends Anal. Chem. 107, 99–115.

672 Luiz-Ferreira, A., Cola-Miranda, M., Barbastefano, V., Hiruma-Lima, C.A., Vilegas, W., Brito,

673 A.R.M.S., 2008. Should *Anacardium humile* St. Hil be used as an antiulcer agent? A

674 scientific approach to the traditional knowledge. Fitoterapia 79, 207–209.

675 Mala John, G.S., Takeuchi, S., Venkatraman, G., Rayala, S.K. (2002). Nordihydroguaiaretic

676 Acid in Therapeutics: Beneficial to Toxicity Profiles and the Search for its Analogs. Curr

677 Cancer Drug Targets 20(2), 86–103.

678 Mannocho-Russo, H., Bueno, P.C.P., Bauermeister, A., De Almeida, R.F., Dorrestein, P.C.,

679 Cavalheiro, A.J., Bolzani, V.S., 2020. Can statistical evaluation tools for chromatographic

680 method development assist in the natural products workflow? A case study on selected

681 species of the plant family Malpighiaceae. J. Nat. Prod. 83, 3239–3249.

682 Michodjehoun-Mestres, L., Amraoui, W., Brillouet, J.M., 2009. Isolation, characterization, and

683 determination of 1-*O*-trans-Cinnamoyl- β -D-glucopyranose in the epidermis and flesh of

684 developing cashew apple (*Anacardium occidentale* L.) and four of its genotypes. J. Agric.

685 Food Chem. 57, 1377–1382.

686 Mohimani, H., Gurevich, A., Shlemov, A., Mikheenko, A., Korobeynikov, A., Cao, L.,

687 Shcherbin, E., Nothias, L. F., Dorrestein, P.C., Pevzner, P.A., 2018. Dereplication of

688 microbial metabolites through database search of mass spectra. Nat. Commun. 9, 1–12.

689 Nothias, L.F., Petras, D., Schmid, R., Dührkop, K., Rainer, J., Sarvepalli, A., Protsyuk, I., Ernst,

690 M., Tsugawa, H., Fleischauer, M., Aicheler, F., Aksenov, A.A., Alka, O., Allard, P.M.,
1
2 691 Barsch, A., Cachet, X., Caraballo-Rodriguez, A.M., Da Silva, R.R., Dang, T., Dorrestein,
3
4
5 692 P.C., 2020. Feature-based molecular networking in the GNPS analysis environment. Nat.
6
7 693 Methods 17, 905–908.
8
9
10 694 Pilon, A.C., Vieira, N.C., Amaral, J.G., Monteiro, A.F., Da Silva, R.R., Spíndola, L.S., Castro-
11
12 695 Gamboa, I., Lopes, N.P., 2021. Molecular networks: An analysis on annotations and
13
14 696 discovery of new assets. Quim. Nova 44, 1168–1179.
15
16
17 697 Ramabulana, A.T., Petras, D., Madala, N.E., Tugizimana, F., 2021. Metabolomics and
18
19 698 molecular networking to characterize the chemical space of four *Momordica* plant species.
20
21 699 Metabolites 11, 763.
22
23
24 700 Royo, V.D.A., Mercadante-Simões, M.O., Ribeiro, L.M., De Oliveira, D.A., Aguiar, M.M. R.,
25
26 701 Costa, E.R., & Ferreira, P.R.B., 2015. Anatomy, histochemistry, and antifungal activity of
27
28 702 *Anacardium humile* (Anacardiaceae) leaf. Microsc. Microanal. 21, 1549–1561.
29
30
31 703 Salehi, B., Gültekin-Özgülven, M., Kirkin, C., Özçelik, B., Morais-Braga, M.F.B., Carneiro, J.
32
33 704 N.P., Bezerra, C.F., Silva, T.G. da, Coutinho, H.D.M., Amina, B., Armstrong, L.,
34
35 705 Selamoglu, Z., Sevindik, M., Yousaf, Z., Sharifi-Rad, J., Muddathir, A.M., Devkota, H.
36
37 706 P., Martorell, M., Jugran, A.K., Martins, N., 2020. Antioxidant, antimicrobial, and
38
39 707 anticancer effects of *Anacardium* plants: An ethnopharmacological perspective. Front.
40
41 708 Endocrinol. 11, 295.
42
43
44 709 Seidel, T., Schuetz, D. A., Garon, A., Langer, T., 2019. The Pharmacophore Concept and Its
45
46 710 Applications in Computer-Aided Drug Design. Prog Chem Org Nat Prod. 110, 99-141.
47
48
49 711 Sumner, L.W., Amberg, A., Barrett, D., Beale, M., Beger, R., Daykin, C., M.Fan, T., Fiehn,
50
51 712 O., Goodacre, R., Griffin, J., Hankemeir, T., Hardy, N., Harnly, J.M., Higashi, R., Kopka,
52
53 713 J., Lane, A., Lindon, J., Marriott, P., Nicholls, A., Reily, M., Thaden, J., Viant, M., 2007.
54
55 714 Proposed minimum reporting standards for chemical analysis. Metabolomics 3, 211–221.
56
57
58
59
60
61
62
63
64
65

715 Szwabowski, G.L., Cole, J.A., Baker, D.L., Parrill, A.L., 2023. Structure-based pharmacophore
1 modeling. Automated random pharmacophore model generation. J Mol Graph Model.
2 716
3
4 717 121, 108429.
5
6
7 718 Taiwo, B.J., Fatokun, A.A., Olubiyi, O.O., Bamigboye-Taiwo, O.T., van Heerden, F.R.,
8
9
10 719 Wright, C.W., 2017. Identification of compounds with cytotoxic activity from the leaf of
11
12 720 the Nigerian medicinal plant, *Anacardium occidentale* L. (Anacardiaceae). Bioorg. Med.
13
14 721 Chem. 25, 2327–2335.
15
16
17 722 Tripathi, A., Vázquez-Baeza, Y., Gauglitz, J.M., Wang, M., Dührkop, K., Nothias-Esposito,
18
19 723 M., Acharya, D.D., Ernst, M., van der Hooft, J.J.J., Zhu, Q., McDonald, D., Brejnrod, A.
20
21 724 D., Gonzalez, A., Handelsman, J., Fleischauer, M., Ludwig, M., Böcker, S., Nothias, L.
22
23
24 725 F., Knight, R., Dorrestein, P.C., 2021. Chemically informed analyses of metabolomics
25
26 726 mass spectrometry data with Qemistree. Nat. Chem. Biol. 17, 146–151.
27
28
29 727 Trott, O., Olson, A.J., 2010. AutoDock Vina: Improving the speed and accuracy of docking
30
31 728 with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem.
32
33 729 31, 455–461.
34
35
36 730 Urzêda, M.A., Marcussi, S., Silva Pereira, L.L., França, S.C., Pereira, A.M.S., Pereira, P.S., Da
37
38
39 731 Silva, S.L., Guimarães, C.L.S., Calderon, L.A., Stábeli, R.G., Soares, A.M., Couto, L.B.,
40
41 732 2013. Evaluation of the hypoglycemic properties of *Anacardium humile* aqueous extract.
42
43 733 Evid.-Based Complement. Alternat. Med. 191080.
44
45
46 734 Van Der Hooft, J.J.J., Wandy, J., Barrett, M.P., Burgess, K.E.V., Rogers, S., 2016. Topic
47
48 735 modeling for untargeted substructure exploration in metabolomics. Proc. Natl. Acad. Sci.
49
50 736 U.S.A. 113, 13738–13743.
51
52
53 737 Wang, M., Carver, J.J., Phelan, V.V., Sanchez, L.M., Garg, N., Peng, Y., Nguyen, D.D.,
54
55
56 738 Watrous, J., Kaponó, C.A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V.,
57
58 739 Meehan, M.J., Liu, W.T., Crüsemann, M., Boudreau, P.D., Esquenazi, E., Sandoval-

740 Calderón, M., Bandeira, N., 2016. Sharing and community curation of mass spectrometry
1
2 741 data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* 34,
3
4
5 742 828–837.
6
7 743 Wolber, G., Langer, T., 2005. LigandScout: 3-D pharmacophores derived from protein-bound
8
9
10 744 ligands and their use as virtual screening filters. *J. Chem. Inf. Model.* 45, 160–169.
11
12 745 Yang, M., Li, J., Zhao, C., Xiao, H., Fang, X., Zheng, J., 2021. LC-Q-TOF-MS/MS detection
13
14 746 of food flavonoids: principle, methodology, and applications. *Crit. Rev. Food Sci. Nutr.*
15
16
17 747 21, 1–21.
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: