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

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

7 a Institute of Chemistry, Federal University of Goias, Goiania, 59078-970, GO, Brazil
8 b Marine Biodiscovery Centre, Department of Chemistry, University of Aberdeen, Aberdeen
9 AB24 3UE, Scotland, UK
10 c GEOMAR Centre for Marine Biotechnology (GEOMAR-Biotech), Research Unit Marine
11 Natural Product Chemistry, GEOMAR Helmholtz Centre for Ocean Research Kiel, Am Kiel-
12 Kanal 44, 24106 Kiel, Germany
13 d Instituto Federal de Educação, Ciência e Tecnologia do Maranhão, 65635-468, Presidente
14 Dutra (MA), Brazil
15 *nerilsonmarques@gmail.com (Nerilson M. Lima); boniek@ufg.br (Boniek G. Vaz)

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

Keywords

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

1. Introduction

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

The chemical diversity of *A. humile* comprises a significant level of volatile compounds as well as organic acids such as resorcinolic acid, ascorbic acid, and anacardic acids (Assunção &
Mercadante, 2003; Bicalho et al., 2000). Previous studies have indicated the presence of
carotenoids (α-carotene, β-carotene, and β-cryptoxanthin), alkaloids, and polyphenols with
significant antioxidant properties. The presence of tannins such as (procyanidin B2), phenolic
acids (gallic acid), and flavonoids (catechin, glycosylated quercetin) present in A. humile have
shown antiglycation properties (Assunção & Mercadante, 2003; Lima Júnior et al., 2021; Luiz-
Ferreira et al., 2008; Royo et al., 2015). There is currently little information on the metabolic
profiling of bushy cashews or the contribution of individual metabolites to the nutritional value.
Therefore, research studies aiming at the rapid and comprehensive acquisition of the metabolic
profile of these species are necessary.

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

Applications of mass spectrometry-based analytical platforms to profile plant metabolomes of
food interest have been implemented in routine analysis in research laboratories (Domínguez
et al., 2020; Lu et al., 2018). However, the massive amount of spectral data generated requires
robust tools for data exploration and organisation to obtain information on the metabolome of
the target species. To address challenges in interpreting these large datasets and translating
chemical composition into biological knowledge, sophisticated dereplication tools based on
spectral fragmentation profiles are used. In addition, *in silico* analysis technologies help characterise unknown molecules and develop methods for assessing target metabolites. Data analysis tools, such as the Global Natural Products Social Molecular Networking (GNPS) platform, have allowed the exploration of plant metabolomes as well as assisting and accelerating the discovery of new bioactive agents (Pilon et al., 2021; Ramabulana et al., 2021).

This work aims to apply untargeted metabolomics approaches to screen for bioactive metabolites in bushy cashew fruits by high-resolution mass spectrometry (HRMS) combined with *in silico* fragmentation tools. It also utilizes the GNPS platform tools, including classic Molecular Networking (MN), Feature-Based Molecular Networking (FBMN), Network Annotation Propagation (NAP), Dereplicator+, Suspect library, molDiscovery, MS2LDA, and MolNetEnhancer to enhance metabolite annotation. Subsequently, based on the collective results, pharmacophore and molecular docking analyses were conducted to generate structure-based feature pharmacophores for the antioxidant activity of the major metabolites. This study represents the first comprehensive exploration of the molecular diversity of *A. humile*, a typical fruit found in the Brazilian savanna. Furthermore, it recognises the inherent antioxidant properties of the plant’s constituents, thereby enhancing its potential value as both a dietary resource and a pharmaceutical asset.

2. Materials and methods

2.1. Chemicals and materials

HPLC-grade acetonitrile and methanol were purchased from Tedia Company (Fairfield, USA). Formic acid and Caffeine-$^{13}$C$_3$ were purchased from Sigma Aldrich (St. Louis, USA). Progesterone-$d_9$ was purchased from CDN Isotopes (Quebec, Canada). Ultrapure water was produced using a water purification system (Master System MS2000, Gehaka, São Paulo, Brazil) with a resistivity of 18.2 MΩcm.
2.2. Sample preparation

Anacardium humile fruits were collected from three different locations in Goiás State, Brazil (Goiânia (S1), Uruana (S2), and Campinaçu (S3) cities). Bushy cashews were sliced manually using a sterile knife, frozen at -80 °C, and freeze-dried for 48h. For metabolomics analyses, 10 mg of freeze-dried fruits were extracted with 1 mL of methanol, vortexed for 1 min, and centrifuged (10 min, 15000 rpm at room temperature). This solution was diluted with methanol (1:5, v:v) and was transferred to injecting vials for LC-MS/MS analysis. Stable isotopes, Caffeine-\textsuperscript{13}C\textsubscript{3} and Progesterone-\textit{d}\textsubscript{9} (2.5 µg mL\textsuperscript{-1}) were used as internal standards.

2.3. LC-MS analysis

LC-MS/MS analyses were performed on an HPLC-UV 1220 Infinity II (Agilent Technologies) coupled with a Q-Exactive hybrid Quadrupole-Orbitrap high-resolution mass spectrometer (Thermo Scientific) as well as an electrospray ionisation source. The column used in this study was an InfinityLab Poroshell 120 EC-C18 column (4.6 × 100 mm × 2.7 µm Agilent). All samples were analysed using a gradient elution program. In both ESI positive and negative modes, the binary mobile phase comprised A (water with 0.1% formic acid) and B (methanol). The gradient elution started at 5% (B) and linearly increased to 100% (B) in 40 min and kept constant for 10 min at 100% (B). The eluent was then restored to the initial conditions in 10 min. The flow rate was set at 0.3 mL min\textsuperscript{-1}. The injection volume was 30 µL, and the column temperature was 35 °C. The ESI source conditions were set as follows: spray voltage 3.5 kV (in both ionisation modes); the capillary temperature was 250 °C (positive mode) and 320 °C (negative mode); S-lens RF level 60 V (in both ionisation modes); sheath gas flow rate at 47 L min\textsuperscript{-1} (positive mode) and 35 L min\textsuperscript{-1} (negative mode); and aux gas flow rate at 11 L min\textsuperscript{-1} (positive mode) and 10 L min\textsuperscript{-1} (negative mode). In both ESI positive and negative modes, high-resolution mass spectra were obtained in the Full MS/data dependent -
MS<sup>2</sup> (dd-MS<sup>2</sup>) mode. The mass range in the full MS scanning experiments was m/z 100-1200. The top 5 (TopN, 5, loop count 5) most abundant precursors were sequentially transferred for collision-induced fragmentation acquisition. The collision energy for target analytes was 20, 30, and 35 eV. Resolving power was 140,000 and 70,000 for full MS and dd-MS<sup>2</sup> acquisitions, respectively.

2.4. Putative compound annotation

The files acquired in the Q-Exactive hybrid Quadrupole-Orbitrap mass spectrometer for the methanolic extracts were converted from raw into (.mzML) format using MSConvert software (ProteoWizard, Palo Alto, CA, US) before being processed using MZmine software, version 2.53. This study utilised metadata to organise compound information following the GNPS online workflow (https://ccms-ucsd.github.io/GNPSDocumentation/). This platform also curates MS/MS spectral library categorised based on the quality of the spectra as gold (thoroughly characterised structures), silver (a compound in crude extract), and bronze (partial annotation) (Wang et al., 2016). Metabolite annotations were based on searching the experimental spectra against the GNPS spectral library using the tools classic Molecular Networking – MN (Wang et al., 2016), Feature-Based Molecular Networking – FBMN (Nothias et al., 2020), Dereplicator+ (Mohimani et al., 2018), Network Annotation Propagation – NAP (da Silva et al., 2018), molDiscovery (Cao et al., 2021), MS2LDA (Van Der Hooft et al., 2016), MolNetEnhancer (Ernst et al., 2019), and analysis of chromatographic data such as retention time and UV spectra. These tools allow the integration of orthogonal annotation methodologies and tandem mass spectrometry data (MS/MS) to explore and obtain the metabolome of plants used as food.

2.5. Molecular Docking
Molecular docking analysis was performed using Autodock Vina v.1.2.0 (The Scripps Research Institute, La Jolla, CA, USA) docking software (Trott & Olson, 2010). The receptor site was predicted using LigandScout (Inte: Ligand) Advanced software (Wolber & Langer, 2005) (evaluation license key: 81809629175371877209), which identifies putative binding pockets by creating a grid surface and calculating the buriedness value of each grid point on the surface. The resulting pocket grid consists of several clusters of grid points, rendered using an iso surface connecting the grid points. The iso surface represents space that may be suitable for creating a pocket.

The x-ray crystal structure of Human Peroxiredoxin 5, a Novel Type of Mammalian Peroxiredoxin (PDB: 1HD2) (Declercq et al., 2001), was retrieved from the Protein Data Bank and utilised to perform docking simulations. Default search parameters were used where the number of binding modes was 10, exhaustiveness was 8, and the maximum energy difference was 3 kcal/mol.

LigandScout (Inte: Ligand) Advanced software (Wolber & Langer, 2005) (evaluation license key: 81809629175371877209) was used to generate 3D pharmacophore models. LigandScout's algorithm calculates and displays chemical interactions between protein–ligand complexes.

3. Results and discussion

An untargeted metabolomics approach was employed to assess the metabolome of A. humile. The molecular diversity of its fruits and the evaluation of the abundance of ions and metabolite content were determined using high-resolution mass spectrometry. This was complemented with dereplication tools, which involved molecular structure searches in robust databases and computer-assisted fragmentation. Furthermore, given the limited data on the metabolic composition of A. humile in the existing literature, the scope of the study was
extended to identify biomarkers of the genus *Anacardium* besides describing the classes and subclasses of secondary metabolites present in the species. For this purpose, the structural similarity of MS data was assessed by comparing it with the spectral library from the GNPS ecosystem. Furthermore, potential candidates were identified *in silico* using specialised platforms such as Sirius, Dereplicator+, and NAP, utilising MS data obtained in both positive and negative ionisation modes. The detection of metabolites was performed in fruit samples of *A. humile* from three different locations in Brazil.

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

*Classic Molecular Networking (MN)*

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

After the dereplication process from the MN analysis, 88 metabolites belonging to different classes of secondary metabolites in aglycone and heteroside form were putatively annotated, including terpenes, steroids, flavonoids, tannins, quinones, alkaloids, coumarins, phenolic acids, and simple phenols. Additionally, many primary metabolites, such as carboxylic acids, sugars, and amino acids, were annotated in all samples. A high molecular diversity of derivative compounds has been annotated within these classes (see Table 1). In addition, a complex array of primary metabolites was detected in all samples, such as carbohydrates, amino acids, and lipids. Based on library matches, polyphenols were the major metabolites in all analyses. O-glycosylated compounds were abundant, and C-glycosylated compounds were in the minority. Alkaloids were detected in low concentrations in the fruits of A. humile, which information corroborates with the data reported for the genus Anacardium. Regarding the collection sites, no substantial variations were observed in metabolite content. All metabolites annotated were previously described in the family Anacardiaceae.

MS/MS, fragmentation pattern analysis, indicated the characteristic and diagnostic ions of the annotated metabolites. The glycosylated phenolic compounds found in high concentrations in the fruits were easily identified by the neutral loss of a sugar moiety, which is determined by the loss of 162 Da for hexosides, 146 Da for deoxyhexosides and 132 Da for pentosides in O-glycosylated flavonoids (Mannochio-Russo et al., 2020). C-glycosylated flavonoids were characterised by the loss of H2O and 120 Da (Mannochio-Russo et al., 2020). The compounds putatively identified as myricetin 3-O-xyloside and myricetin 3-rutinoside at m/z 449.073 and 625.141 [M−H]−, respectively, were identified by the consecutive losses of sugar moieties and also based on their UV spectra and retention times. Furthermore, the flavonoid aglycone fragmentation exhibits a pathway characterised by forming fragments originating from retro
Diels-Alder reactions coupled with losses of neutral molecules such as CO\textsubscript{2} and H\textsubscript{2}O (Yang et al., 2021). Similarly, the flavonoid monohexosides and dihexosides isoquercitrin, guajavarin, myricetin 3-\textit{O}-rutinoside, myricetin 3-\textit{O}-galactoside, myricetin 3-\textit{O}-xyloside, and rutin showed deprotonated molecular ions [M−H]⁻ at m/z 463.089, 433.078, 625.141, 479.083, 449.073, and 609.146, respectively, previously reported in the literature (Fig. 1). Meanwhile, these compounds were connected with a mass difference of 146 Da, referring to a unit of deoxyhexoside, 16 Da as a possible extra hydroxyl group, and 30 Da as an extra OCH\textsubscript{2}. In total, thirteen \textit{O}-glycosylated flavonoids and four aglycone flavonoids were detected in positive ionisation mode, and eleven \textit{O}-glycosylated flavonoids and six aglycone flavonoids were detected in negative ionisation mode. The retention time observed in the chromatographic analysis allowed us to improve the annotation and distinguish the \textit{O}-glycosylated flavonoids. From mass spectrometry fragmentation, it was possible to perform structure-based propagation and guided detection of metabolites not annotated by the GNPS library since compounds of the same molecular family are structurally related and share molecular substructures that allow the putative identification of new molecules. Hence, from the difference between the nodes indicated by m/z 609.181 and 463.123 [M+H]⁺, an analogue of the flavone Diosmin was annotated, whose mass difference is 146 Da referring to a unit of deoxyhexosides. The detailed inspection of their MS/MS spectra allowed us to infer the glycosylation type. Through mass spectral similarity networking, accurate annotation of the flavan-3-ols catechin (m/z 289.072) and epigallocatechin (m/z 305.067) was achieved in negative ionisation mode [M−H]⁻. Through their retention times, accurate masses, and MS/MS similar fragmentation patterns obtained in negative mode [M−H]⁻, the methoxylated anthocyanidins 5-methylcyanidin (m/z 300.946) and methyldelphinidin (m/z 316.920) were annotated, which are widely distributed in fruits of the genus \textit{Anacardium} (de Brito et al., 2007).
Table 1. Results of the metabolite annotation in the fruits of *A. humile* through molecular networking analysis.

<table>
<thead>
<tr>
<th>Parent mass</th>
<th>Adduct</th>
<th>Molecular formula</th>
<th>Metabolite name</th>
<th>Chemical structure</th>
<th>Chemical class</th>
</tr>
</thead>
<tbody>
<tr>
<td>303.217</td>
<td>[M+H]⁺</td>
<td>C₁₅H₁₀O₇</td>
<td>Quercetin</td>
<td><img src="image" alt="Quercetin" /></td>
<td>Flavonoid</td>
</tr>
<tr>
<td>289.072</td>
<td>[M−H]⁻</td>
<td>C₁₅H₁₄O₆</td>
<td>Catechin</td>
<td><img src="image" alt="Catechin" /></td>
<td>Flavonoid</td>
</tr>
<tr>
<td>595.158</td>
<td>[M+H]⁺</td>
<td>C₂₇H₃₀O₁₅</td>
<td>Kaempferol 3-O-glucoside-7-O-rhamnoside</td>
<td><img src="image" alt="Kaempferol" /></td>
<td>Flavonoid</td>
</tr>
<tr>
<td>449.385</td>
<td>[M+H]⁺</td>
<td>C₂₁H₂₀O₁₁</td>
<td>Kaempferol 7-O-glucoside</td>
<td><img src="image" alt="Kaempferol" /></td>
<td>Flavonoid</td>
</tr>
<tr>
<td>463.123</td>
<td>[M+H]⁺</td>
<td>C₂₃H₂₃O₁₁</td>
<td>Peonidin 3-glucopyranoside</td>
<td><img src="image" alt="Peonidin" /></td>
<td>Flavonoid</td>
</tr>
<tr>
<td>300.946</td>
<td>[M+H]⁺</td>
<td>C₁₆H₁₃O₆</td>
<td>5-methylcyanidin</td>
<td><img src="image" alt="5-methylcyanidin" /></td>
<td>Flavonoid</td>
</tr>
<tr>
<td>316.920</td>
<td>[M+H]⁺</td>
<td>C₁₆H₁₃O₇</td>
<td>Methyldelphinid</td>
<td><img src="image" alt="Methyldelphinid" /></td>
<td>Flavonoid</td>
</tr>
<tr>
<td>449.385</td>
<td>[M+H]⁺</td>
<td>C₂₁H₂₀O₁₁</td>
<td>Luteolin 7-O-glucoside</td>
<td><img src="image" alt="Luteolin" /></td>
<td>Flavonoid</td>
</tr>
<tr>
<td>449.385</td>
<td>[M+H]⁺</td>
<td>C₂₁H₂₀O₁₁</td>
<td>Astragalin</td>
<td><img src="image" alt="Astragalin" /></td>
<td>Flavonoid</td>
</tr>
<tr>
<td>433.247</td>
<td>[M+H]⁺</td>
<td>C₂₁H₂₀O₁₀</td>
<td>Apigenin glucoside</td>
<td><img src="image" alt="Apigenin glucoside" /></td>
<td>Flavonoid</td>
</tr>
<tr>
<td>Mass (M+H)</td>
<td>Formula</td>
<td>Molecule Name</td>
<td>Functional Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>---------</td>
<td>---------------</td>
<td>------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>627.469</td>
<td>C_{27}H_{30}O_{17}</td>
<td>Quercetin 3,7-di-O-glucoside</td>
<td>Flavonoid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>463.125</td>
<td>C_{21}H_{20}O_{12}</td>
<td>Hyperoside</td>
<td>Flavonoid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>537.168</td>
<td>C_{30}H_{18}O_{10}</td>
<td>Amentoflavone</td>
<td>Flavonoid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>627.156</td>
<td>C_{27}H_{30}O_{17}</td>
<td>Myricetin 3-rutinoside</td>
<td>Flavonoid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>449.073</td>
<td>C_{20}H_{18}O_{12}</td>
<td>Myricetin 3-O-xyloside</td>
<td>Flavonoid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>625.141</td>
<td>C_{27}H_{30}O_{17}</td>
<td>Myricetin 3-rutinoside</td>
<td>Flavonoid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>611.161</td>
<td>C_{27}H_{30}O_{16}</td>
<td>Rutin</td>
<td>Flavonoid</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
579.145 [M+H]$^+$ $\text{C}_{30}\text{H}_{26}\text{O}_{12}$ Procyanidin B2 Flavonoid

579.341 [M+H]$^+$ $\text{C}_{27}\text{H}_{30}\text{O}_{14}$ Kaempferitrin Flavonoid

257.972 [M+H]$^+$ $\text{C}_{15}\text{H}_{12}\text{O}_{4}$ Pinocembrin Flavonoid

537.168 [M−H]$^-$ $\text{C}_{30}\text{H}_{18}\text{O}_{10}$ Agathisflavone Flavonoid

349.183 [M+H]$^+$ $\text{C}_{22}\text{H}_{36}\text{O}_{3}$ 6-Pentadecysalicylic Acid Phenolic acid

171.050 [M+H]$^+$ $\text{C}_{7}\text{H}_{6}\text{O}_{5}$ Gallic acid Phenolic acid

185.115 [M+H]$^+$ $\text{C}_{9}\text{H}_{8}\text{O}_{5}$ Methyl gallate Phenolic acid

355.070 [M+H]$^+$ $\text{C}_{16}\text{H}_{18}\text{O}_{9}$ Chlorogenic acid Phenolic acid

153.019 [M−H]$^-$ $\text{C}_{7}\text{H}_{6}\text{O}_{4}$ Protocatechuic acid Phenolic acid

177.034 [M+H]$^+$ $\text{C}_{6}\text{H}_{6}\text{O}_{6}$ Ascorbic acid Organic acid

345.244 [M−H]$^-$ $\text{C}_{22}\text{H}_{33}\text{O}_{3}$ Ginkgolic acid Organic acid
The systematic investigation of all nodes associated with the flavonoid class allowed us to annotate propagation of related compounds in the molecular networks from negative ionization mode, such as the galloyl flavonol glycoside derived from quercetin 3-\(O\)-(6\(^\prime\)-galloyl)-glucopyranoside (\(m/z\) 615.100 [M−H]\(^-\)) and the flavanone naringenin (\(m/z\) 271.060 [M−H]\(^-\)). A detailed assessment of the flavonoid profiling using mass spectral data from two ionization modes showed other flavonoid structures that were not connected to other nodes from the molecular networking approach, including the flavonoids quercetin (\(m/z\) 303.217 [M+H]\(^+\); empirical formula: \(C_{15}H_{10}O_7\)), kaempferol 7-\(O\)-glucoside (\(m/z\) 449.385 [M+H]\(^+\); \(C_{21}H_{20}O_{11}\)), luteolin 7-\(O\)-glucoside (\(m/z\) 449.385 [M+H]\(^+\); \(C_{21}H_{20}O_{11}\)), astragalin (\(m/z\) 449.385 [M+H]\(^+\); \(C_{21}H_{20}O_{11}\)), apigenin glucoside (\(m/z\) 433.247 [M+H]\(^+\); \(C_{21}H_{20}O_{10}\)), quercetin 3,7-di-\(O\)-glucoside (\(m/z\) 627.469 [M+H]\(^+\); \(C_{27}H_{30}O_{17}\)), hyperoside (\(m/z\) 463.125 [M-H]; \(C_{21}H_{20}O_{12}\)), and the biflavonoid amentoflavone (\(m/z\) 537.168 [M-H]; \(C_{30}H_{18}O_{10}\)) previously reported in \(A.\) \(humile\) (Lima Júnior et al., 2021; Luiz-Ferreira et al., 2008) and \(A.\) \(occidentale\) (Andarwulan et al., 2012; Salehi et al., 2020; Taiwo et al., 2017). The isomers were separated by their chromatographic profile, and their precursor ions displayed a different fragmentation pattern.

Applying this same approach, the annotation of low molecular weight metabolites provided some bioactive phenolic acids such as gallic acid (\(m/z\) 171.050 [M+H]\(^+\); \(C_7H_6O_3\)), methyl gallate (\(m/z\) 185.115 [M+H]\(^+\); \(C_8H_8O_3\)), protocatechuic acid (\(m/z\) 154.980 [M+H]\(^+\); \(C_7H_6O_4\)), and chlorogenic acid (\(m/z\) 355.070 [M+H]\(^+\); \(C_{16}H_{18}O_9\)). All these phenolic acids have been previously reported in the \(Anacardium\) genus (Andarwulan et al., 2012; Lima Júnior et al., 2021; Luiz-Ferreira et al., 2008). The fragmentation characteristics and biosynthetic knowledge indicated the oxidation pattern and location of the sugar units.

A more significant number of flavonoids were detected, mainly in the negative ionization mode. A high diversity of flavonoids, mainly \(O\)-glycosyl flavonols, was confirmed by the large number of clusters observed in the MN. Combined with detailed taxonomical knowledge,
allowed for the annotation of metabolites not available in the reference library spectra. MN has been successfully applied to explore the molecular complexity of *A. humile* fruits and annotate structurally related molecules.

In summary, *O*-glycosylated flavonoids belonging to the subclass of flavonols were dominant primarily based on the structure of myricetin and quercetin and linked mainly to sugars, glucose and rhamnose. Through the evaluation of the annotations proposed by the MN library and manual inspection of fragmentation spectra, four phenolic acid glycosides were putatively identified and annotated as benzoic acid hexoside derivatives and coumaroyl hexoside derivatives (Fig. 1). The phenolic acid glycosides classified as cinnamic acid derivatives showed deprotonated molecular ions [M−H]− at m/z 325.09 (coumaroyl hexoside) and 341.08 (caffeic acid hexoside); and the benzoic acid hexoside derivatives showed deprotonated molecular ions [M−H]− at m/z 315.07 (benzoic acid + 2O, O-Hex) and 299.08 (benzoic acid + O, O-Hex) with a mass difference of 16 Da indicating a hydroxyl group as the only structural difference. In addition, its spectrum showed a typical fragment with the elimination of CO₂. The metabolite annotated as 1-*O*-trans-cinnamoyl-glucopyranose at m/z 309.10 [M−H]− showed a fragment in MS/MS at m/z 147.04 from neutral loss of the sugar moiety (Glc, 162 Da), previously isolated from cashew apple (*Anacardium occidentale* L.) (Michodjehoun-Mestres et al., 2009). Its derivative 1-*O*-trans-cinnamoyl-(6-*O*-galloyl)-glucopyranose at m/z 461.110 ([M−H]−) exhibited high spectral similarity and spectral matching to reference library spectra. The negative ionization mode was more sensitive for detecting these polyphenols.

The presence of other glycosylated polyphenols was investigated in *A. humile* fruits. These analyses found a high abundance of gallotannins (hydrolyzable tannins) formed by polygalloyl esters of glucose. In this work, two types of gallotannins with degrees of polymerization of 2 (m/z 483.08 [M−H]−) to 3 (m/z 635.089 [M−H]−) galloyl units linked to glucose were putatively
identified. Their fragmentation patterns were mainly characterized by the loss of 170 Da and 152 Da, referring to the loss of a unit of gallic acid and galloyl fission, respectively.

![Diagram showing flavonoids and phenolic acids](image)

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

Although gas chromatography-coupled mass spectrometry is the most used technique in the investigation of chemical markers of *Anacardium*, the analysis by LC-MS/MS with electrospray ionization enables the detection of the various markers present in the fruits of *A. humile*. The interpretation of MS² data and evaluation of structurally related molecules annotated by the spectral library and network propagation of *m/z* differences, as well as retention time analysis, provided the structure of the 16 chemical markers classified as cardols, anacardic acids, 2-methylcardols, and cardanols with different degrees of unsaturation (saturated, monounsaturated, di-unsaturated, and tri-unsaturated), confirmed by data reported in the literature (Gomes Júnior et al., 2018; Kubo et al., 1994). The fragmentation pattern of
these metabolites (Fig. 2) in negative ionization mode was inspected in detail by observing the neutral losses, relative abundances, and diagnostic ions based on their network connection. The cluster containing these molecules shows the metabolites of each class with C15-alkyl side-chain with 0 - 3 double bonds. The chemical marker annotated were cardols ((1) C15:3,5-[8(Z),ll(Z),l4-pentadecatrienyl]resorcinol; (2) C15:2,5-[8(Z),ll(Z)-pentadecadienyl]resorcinol; (3) C15:1,5-[8(Z)-pentadecenyl]resorcinol; (4) C15:0,5-pentadecylresorcinol), Anacardic acids ((5) C15:3,6-[8(Z),ll(Z),l4-pentadecatrienyl]salicylic acid; (6) C15:2,6-[8(Z),ll(Z)-pentadecadienyl]salicylic acid; (7) C15:2,6-[8(Z)-pentadecenyl]salicylic acid; (8) C15:0,6-pentadecylsalicylic acid), 2-methylcardols ((9) C15:3,2-methyl-5-[8(Z),ll(Z)-pentadecatrienyl]resorcinol; (10) C15:2,2-methyl-5-[8(Z),ll(Z)-pentadecadienyl]resorcinol; (11) C15:1,2-methyl-5-[8(Z)-pentadecenyl]resorcinol; (12) C15:0,2-methyl-5-pentadecylresorcinol), Cardanols ((13) C15:3,3-[8(Z),ll(Z),l4-pentadecatrienyl]phenol; (14) C15:2,3-[8(Z),ll(Z)-pentadecadienyl]phenol; (15) C15:1,3-[8(Z)-pentadecenyl]phenol; (16) C15:0,3-pentadecylphenol) (Gomes Júnior et al., 2018; Kubo et al., 1994). The metabolites were detected in both positive and negative ionization modes.
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
compounds, 37 with a gold classification and 24 with a cosine score > 0.9. Evaluating the same characteristics described for the ESI (+), only two organic acids and two flavonoids were annotated.

**Merge Polarity Networks**

Notwithstanding the extensive search against the GNPS spectral library and the use of the annotation propagation strategy to explore structural relationships, many ions remained unannotated. This further motivated the application of more sophisticated dereplication tools. The Merge Polarity Networks function was applied, combining data from both ionization modes, and yielding more accurate annotations by calculating similarity scores between structures (Tanimoto). This assessment showed 2406 pairs and yielded four relatively abundant features ($m/z$ 214.09; 163.06; 147.08; 348.07 [M+H]$^+$) with Tanimoto values equal to 1 (perfect agreement), and eight structures were compatible in both ionization modes but were not similar (Tanimoto = -1). This allowed the annotation of kaurenoic acid glucopyranosyl derivative at $m/z$ 645.29 [M+H]$^+$, 1,6-anhydro-glucose at $m/z$ 163.06 [M+H]$^+$ and the flavonoid 4''-O-acetylmyricitrin at $m/z$ 627.16 [M+H]$^+$.

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

Although classic MN is very useful for metabolic annotation through spectral library matching, propagating annotations of unannotated nodes and reducing the size of datasets in untargeted metabolomics, it has some limitations. It does not use MS1 data or provide the resolution of isomers with different retention times in the chromatographic analysis or coeluting metabolites. Besides, MN does not allow the accurate inference of the relative amounts of the metabolites (abundance of the ions) between samples. However, these challenges can be solved by employing Feature-Based Molecular Networking (FBMN), which provides greater resolving power than MN by using MS1 information to verify retention time, peak intensity,
ion mobility and isotope patterns (Nothias et al., 2020). Therefore, FBMN was employed to discriminate isomeric/isobaric compounds and infer the ion abundance of some annotated substances and their analogues using chromatographic peak area and peak intensity through feature detection and alignment analysis. FBMN analysis detected 860 features in ESI (+)-MS mode and 340 in ESI (-)-MS. From the relative intensities of the ions, it was possible to infer that the antioxidant compounds gallic acid, ascorbic acid, and anthocyanins were the most abundant metabolites found in A. humile (Lima Júnior et al., 2021). In addition, the FBMN provided a highly reliable annotation of polyphenols found in the species (Fig. 4).

![Fig. 4](image-url)

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

**Network Annotation Propagation (NAP)**

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

**Dereplicator+**

High-confidence identification methods from in silico database search were employed in this study, including Dereplicator+. This high-throughput identification tool improves the annotation of peptides and small molecules and has been successfully applied in metabolomic approaches to natural products (Mohimani et al., 2018). The positive ionization mode resulted in 47 unique compounds, including phenolic acid, chromenes, and terpenes. Terpenes and their derivatives were the most annotated class of natural products (over 50%). After inspection of the MS/MS spectra, the tricyclic triterpenoid achilleol B, the triterpenoid betulinaldehyde, the steroid ergosta-3,5,24(28)-triene, and a chromanol derivative were annotated. In the negative ionization mode, 15 unique library compounds belonging to flavonoids, tannins, quinones, and terpenes classes were obtained.

**MolDiscovery**

MolDiscovery was employed to accelerate chemical structural annotation and enhance the efficiency and accuracy of metabolic identification. This probabilistic model uses a fragmentation algorithm to generate MS/MS fragmentations and match compounds with their MS/MS spectra (Cao et al., 2021). As recommended, the analysis of unique metabolites generated by the workflow with high molecular weight was conducted, as the tool demonstrates greater accuracy for masses exceeding 600 Da, attributed to the frequently elevated count of fragment ions. Furthermore, the fragmentation of small molecules considers the type of fragmented bonds and other factors such as moiety (Cao et al., 2021). MolDiscovery exhibited 2.226 unique metabolites in (+)-ESI mode with high mass ranges >600 Da and a cut-off score of 15 (159 hits). A larger score implies a greater probability of correct annotation (Cao et al.,
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-acetylmicrictin 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 punicacortein 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 tricin 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...
losses of 44 (free carboxyl) and 18 (water) from glycosylated derivatives, corroborating the high abundance of gallotannins and structurally close polyphenols in *A. humile* fruits. In addition, many neutral losses of 162 Da and 146 Da revealed the presence of sugar residue such as hexose and pentose, which can be attributed to the high abundance of *O*-glycosylated flavonoids found in all samples. The results obtained from the MS2LDA workflow improved the confidence level of the annotations, expanded the metabolite annotation, and characterized metabolites not yet reported in the species through the analysis of functional groups and core structures provided by the information from Mass2Motif.

**MolNetEnhancer**

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

There was no significant difference in the metabolic profiling concerning collection sites, as represented in Fig. 5. Based on metabolomic profiling obtained of the extracts of *A humile*...
fruits from three collection sites, the flavonoids were the major metabolites in all analyses. Considering that flavonoids have remarkable free radical scavenging properties, the flavonoids peonidin 3-O-glucopyranoside, methylcyanidin and methyldelphinidin were selected for determination of antioxidant activity using molecular docking.

![3D visualization](image1)

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

* Determination of antioxidant activity using molecular docking
  
  As peonidin 3-O-glucopyranoside, methylcyanidin and methyldelphinidin are the major metabolites in *A. humile* based on the above metabolomics studies. A molecular docking study was undertaken to assess the antioxidant activity of these compounds. The docking study was carried out on the crystal structure of Human Peroxiredoxin 5 (PRDX5, PDB: 1HD2) (Declercq et al., 2001), a novel type of *Mammalian peroxiredoxin* that has antioxidative and cytoprotective functions during oxidative stress. The peroxiredoxins define an emerging family
of peroxidases able to reduce hydrogen peroxide and alkyl hydroperoxides with the use of reducing equivalents derived from thiol-containing donor molecules such as thioredoxin, glutathione, trypanothione and AhpF. Peroxiredoxins have been identified in prokaryotes as well as in eukaryotes (Arden et al., 2015). Peroxiredoxin 5 is widely expressed in tissues and located cellullarly to mitochondria, peroxisomes, and cytosol and, therefore, implicated in antioxidant protective mechanisms (Knoops et al., 2011).

The three major compounds previously mentioned were subjected to docking analysis, and the specificities of their interaction with these targets, as shown in Fig. 6, were investigated. The best-docked complexes were obtained based on binding energies and interacting residues. Docking poses were analyzed and compared to the co-crystallized standard antioxidant nordihydroguaiaretic acid (NDGA). All three major compounds in this molecular docking study docked very well compared to the standard Nordihydroguaiaretic acid (Mala et al., 2020) (Table 2).

### Table 2 Results of the metabolite annotation in the fruits of *A. humile*.

<table>
<thead>
<tr>
<th>Compounds</th>
<th>Docking Score (-) (kcal/mol)</th>
<th>PDB ID: 1HD2 (Human Peroxiredoxin 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nordihydroguaiaretic acid (Standard)</td>
<td>5.3</td>
<td></td>
</tr>
<tr>
<td>Peonidin 3-O-glucopyranoside</td>
<td>5.02</td>
<td></td>
</tr>
<tr>
<td>Methylcyanadin</td>
<td>5.46</td>
<td></td>
</tr>
<tr>
<td>Methyldelphinidin</td>
<td>6.20</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 6. Binding site (yellow colour) of Human Peroxiredoxin 5.
Pharmacophore Evaluation (Structure-based pharmacophore)

Pharmacophores represent chemical functions. A pharmacophore abstractly describes steric and electronic features required to trigger (or block) biological response (Khedkar et al., 2007). A pharmacophore model can explain how structurally diverse ligands bind at the receptor site based on common interaction points (Seidel et al., 2019). Structure-based pharmacophore modelling is a method for pharmacophore development based on the target protein's structural features (Szwabowski et al., 2023). In this method, the possible active site in protein where the interactions of co-crystallized ligand occur will be analyzed. Due to its simplicity, this method is computationally very efficient and exceptionally well suited for the virtual screening of a wide range of compound libraries. This method searches for interactions between ligands and the macromolecule. The three major peonidin 3-O-glucopyranoside, methylcyanidin and methyldelphinidin were used for pharmacophore evaluation.

Ligand peonidin 3-O-glucopyranoside interacts with the macromolecule PRDX5, as shown in Fig. 7. It shows hydrophobic effects with amino acids Ile119A and Phe120A. Leu116A and Thr147A. This pharmacophore also represents a hydrogen bond acceptor (HBA) feature with nearby amino acid residue Arg127A. Looking at the hydrogen bond donor (HBD) feature, the ligand interacts with Thr147A amino acid residue. Based on the interactions, this structure provides a pharmacophore with three features hydrophobic effects (H), hydrogen bond acceptor (HBA) and hydrogen bond donor (HBD), with a pharmacophore score of (0.7).

Ligand methylcyanidin interacts with the macromolecule PRDX5 as shown in Fig. 7. It shows hydrophobic effects with BR303A and amino acids Leu116A, Thr147A, Phe120A, and Thr44A. This pharmacophore has hydrogen bond acceptor (HBA) and hydrogen bond donor (HBD) features with nearby amino acid residues Cys47A and Thr44A. Based on the interactions, this structure provides a pharmacophore with three features hydrophobic effects...
(H), hydrogen bond acceptor (HBA) and hydrogen bond donor (HBD), with a pharmacophore score of (0.89).

Ligand methyldelphinidin interacts with the macromolecule PRDX5, as shown in Fig. 7 (A) and (B). It shows hydrophobic effects with the amino acids Leu116A and Ile119A. This pharmacophore also has both hydrogen bond acceptor (HBA) and hydrogen bond donor (HBD) features with nearby amino acid residues Cys47A and Thr44A. Arg127A only has hydrogen bond acceptor (HBA) interactions with the ligand. Based on the interactions, this structure provides a pharmacophore with three features hydrophobic effects (H), hydrogen bond acceptor (HBA) and hydrogen bond donor (HBD), with a pharmacophore score of (0.87). The above three structure-based pharmacophores further generated a shared feature pharmacophore, as shown in Fig. 8.
Fig. 7. Ligands interaction with the macromolecule. (A) 3D view of pharmacophore presents in the macromolecule binding site. (B) 2D view of pharmacophore. Pharmacophore features: Hydrophobic interactions (H), Yellow; Hydrogen bond acceptor (HBA), Red; Hydrogen bond donor (HBD), Green.

Fig. 8. Generated Shared feature pharmacophore.
4. Conclusion

The application of the integrated metabolomics approach and the combination of structural annotation and data mining tools provided critical insights into the characterization and quantification of constituents of the healthy diet, with promising application in food chemistry. This study has revealed a comprehensive metabolic profile and a high content of bioactive polyphenols in bushy cashew fruits, which hold significant potential from a biotechnological perspective. Moreover, the state-of-the-art computational tools employed in this study enriched molecular networks, allowing the rapid and effective annotation of a vast array of metabolites not yet reported in *A. humile*. Additionally, molecular docking analysis was performed on peonidin 3-*O*-glucopyranoside, methylcyanidin and methyldelphinidin to assess their binding to key receptors associated with antioxidant mechanisms. Compared to standard compounds, these compounds exhibited strong critical potentials related to antioxidant activity. A structure-based pharmacophore model was proposed to guide future studies, potentially aiding in selecting and synthesising cost-effective compounds with good bioactivity.

Declaration of Competing Interest

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.

Acknowledgements

The authors acknowledge financial support from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the institutional and financial support.

References


Kubo, I., Ochi, M., Vieira, P.C., Komatsu, S., 1993. Antitumor agents from the cashew


Nothias, L.F., Petras, D., Schmid, R., Dührkop, K., Rainer, J., Sarvepalli, A., Protsyuk, I., Ernst,
M., Tsugawa, H., Fleischauer, M., Aicheler, F., Aksenov, A.A., Alka, O., Allard, P.M.,
Barsch, A., Cachet, X., Caraballo-Rodriguez, A.M., Da Silva, R.R., Dang, T., Dorrestein,

Methods 17, 905–908.


Ramabulana, A.T., Petras, D., Madala, N.E., Tugizimana, F., 2021. Metabolomics and
molecular networking to characterize the chemical space of four *Momordica* plant species.

Metabolites 11, 763.

Royo, V.D.A., Mercadante-Simões, M.O., Ribeiro, L.M., De Oliveira, D.A., Aguiar, M.M. R.,
Costa, E.R., & Ferreira, P.R.B., 2015. Anatomy, histochemistry, and antifungal activity of

Salehi, B., Gültekin-Özgüven, M., Kirkin, C., Özçelik, B., Morais-Braga, M.F.B., Carneiro, J.
Selamoglu, Z., Sevindik, M., Yousaf, Z., Sharifi-Rad, J., Muddathir, A.M., Devkota, H.
anticancer effects of *Anacardium* plants: An ethnopharmacological perspective. Front.
Endocrinol. 11, 295.


Sumner, L.W., Amberg, A., Barrett, D., Beale, M., Beger, R., Daykin, C., M.Fan, T., Fiehn,
O., Goodacre, R., Griffin, J., Hankemeir, T., Hardy, N., Harnly, J.M., Higashi, R., Kopka,
Proposed minimum reporting standards for chemical analysis. Metabolomics 3, 211–221.


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

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

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

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

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

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

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

Abstract

*Anacardium humile* (bushy cashew) is a native Brazilian plant with substantial pharmacological potential and noteworthy commercial significance in the food industry. This study introduces an untargeted metabolomics approach based on mass spectrometry for the comprehensive structural annotation of bioactive metabolites. The fruits were collected from three distinct sites and subjected to LC-HRMS/MS analysis. A total of eighty-eight compounds were putatively annotated across various metabolite classes. This unveiled a metabolic profile characterized by
notable concentrations of polyphenols, including flavonoids, tannins, phenolic acids, and quinones, while aliphatic acids and terpenes were found in limited quantities. Noteworthy, no significant disparities in the metabolic content were observed among the collection sites. The three principal metabolites (peonidin 3-O-glucopyranoside, methylcyanidin and methyldelphinidin) underwent assessment for antioxidant activity via molecular docking analysis, subsequently generating structure-based feature pharmacophores. The process of annotation propagation yielded a comprehensive qualitative appraisal of the *A. humile* metabolome. The outcomes obtained offer potential candidates for further exploration of their nutraceutical attributes.

### Keywords

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

### 1. Introduction

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

The chemical diversity of *A. humile* comprises a significant level of volatile compounds as well as organic acids such as resorcinolic acid, ascorbic acid, and anacardic acids (Assunção &
Mercadante, 2003; Bicalho et al., 2000). Previous studies have indicated the presence of carotenoids (α-carotene, β-carotene, and β-cryptoxanthin), alkaloids, and polyphenols with significant antioxidant properties. The presence of tannins such as (procyanidin B2), phenolic acids (gallic acid), and flavonoids (catechin, glycosylated quercetin) present in A. humile have shown antiglycation properties (Assunção & Mercadante, 2003; Lima Júnior et al., 2021; Luiz-Ferreira et al., 2008; Royo et al., 2015). There is currently little information on the metabolic profiling of bushy cashews or the contribution of individual metabolites to the nutritional value. Therefore, research studies aiming at the rapid and comprehensive acquisition of the metabolic profile of these species are necessary.

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

Applications of mass spectrometry-based analytical platforms to profile plant metabolomes of food interest have been implemented in routine analysis in research laboratories (Domínguez et al., 2020; Lu et al., 2018). However, the massive amount of spectral data generated requires robust tools for data exploration and organisation to obtain information on the metabolome of the target species. To address challenges in interpreting these large datasets and translating chemical composition into biological knowledge, sophisticated dereplication tools based on
spectral fragmentation profiles are used. In addition, in silico analysis technologies help characterise unknown molecules and develop methods for assessing target metabolites. Data analysis tools, such as the Global Natural Products Social Molecular Networking (GNPS) platform, have allowed the exploration of plant metabolomes as well as assisting and accelerating the discovery of new bioactive agents (Pilon et al., 2021; Ramabulana et al., 2021).

This work aims to apply untargeted metabolomics approaches to screen for bioactive metabolites in bushy cashew fruits by high-resolution mass spectrometry (HRMS) combined with in silico fragmentation tools. It also utilizes the GNPS platform tools, including classic Molecular Networking (MN), Feature-Based Molecular Networking (FBMN), Network Annotation Propagation (NAP), Dereplicator+, Suspect library, molDiscovery, MS2LDA, and MolNetEnhancer to enhance metabolite annotation. Subsequently, based on the collective results, pharmacophore and molecular docking analyses were conducted to generate structure-based feature pharmacophores for the antioxidant activity of the major metabolites. This study represents the first comprehensive exploration of the molecular diversity of A. humile, a typical fruit found in the Brazilian savanna. Furthermore, it recognises the inherent antioxidant properties of the plant’s constituents, thereby enhancing its potential value as both a dietary resource and a pharmaceutical asset.

2. Materials and methods

2.1. Chemicals and materials

HPLC-grade acetonitrile and methanol were purchased from Tedia Company (Fairfield, USA). Formic acid and Caffeine-\(^{13}\)C\(_3\) were purchased from Sigma Aldrich (St. Louis, USA). Progesterone-\(^{d9}\) was purchased from CDN Isotopes (Quebec, Canada). Ultrapure water was produced using a water purification system (Master System MS2000, Gehaka, São Paulo, Brazil) with a resistivity of 18.2 M\(\Omega\)cm.
2.2. Sample preparation

*Anacardium humile* fruits were collected from three different locations in Goiás State, Brazil (Goiânia (S1), Uruana (S2), and Campinaçu (S3) cities). Bushy cashews were sliced manually using a sterile knife, frozen at -80 °C, and freeze-dried for 48h. For metabolomics analyses, 10 mg of freeze-dried fruits were extracted with 1 mL of methanol, vortexed for 1 min, and centrifuged (10 min, 15000 rpm at room temperature). This solution was diluted with methanol (1:5, v:v) and was transferred to injecting vials for LC-MS/MS analysis. Stable isotopes, Caffeine-\(^{13}\)C\(_3\) and Progesterone-\(d_9\) (2.5 \(\mu\)g mL\(^{-1}\)) were used as internal standards.

2.3. LC-MS analysis

LC-MS/MS analyses were performed on an HPLC-UV 1220 Infinity II (Agilent Technologies) coupled with a Q-Exactive hybrid Quadrupole-Orbitrap high-resolution mass spectrometer (Thermo Scientific) as well as an electrospray ionisation source. The column used in this study was an InfinityLab Poroshell 120 EC-C18 column (4.6 × 100 mm × 2.7 \(\mu\)m Agilent). All samples were analysed using a gradient elution program. In both ESI positive and negative modes, the binary mobile phase comprised A (water with 0.1% formic acid) and B (methanol). The gradient elution started at 5% (B) and linearly increased to 100% (B) in 40 min and kept constant for 10 min at 100% (B). The eluent was then restored to the initial conditions in 10 min. The flow rate was set at 0.3 mL min\(^{-1}\). The injection volume was 30 \(\mu\)L, and the column temperature was 35 °C. The ESI source conditions were set as follows: spray voltage 3.5 kV (in both ionisation modes); the capillary temperature was 250 °C (positive mode) and 320 °C (negative mode); S-lens RF level 60 V (in both ionisation modes); sheath gas flow rate at 47 L min\(^{-1}\) (positive mode) and 35 L min\(^{-1}\) (negative mode); and aux gas flow rate at 11 L min\(^{-1}\) (positive mode) and 10 L min\(^{-1}\) (negative mode). In both ESI positive and negative modes, high-resolution mass spectra were obtained in the Full MS/data dependent -
MS² (dd-MS²) mode. The mass range in the full MS scanning experiments was m/z 100-1200.

The top 5 (TopN, 5, loop count 5) most abundant precursors were sequentially transferred for collision-induced fragmentation acquisition. The collision energy for target analytes was 20, 30, and 35 eV. Resolving power was 140,000 and 70,000 for full MS and dd-MS² acquisitions, respectively.

2.4. Putative compound annotation

The files acquired in the Q-Exactive hybrid Quadrupole-Orbitrap mass spectrometer for the methanolic extracts were converted from raw into (.mzML) format using MSConvert software (ProteoWizard, Palo Alto, CA, US) before being processed using MZmine software, version 2.53. This study utilised metadata to organise compound information following the GNPS online workflow (https://ccms-ucsd.github.io/GNPSDocumentation/). This platform also curates MS/MS spectral library categorised based on the quality of the spectra as gold (thoroughly characterised structures), silver (a compound in crude extract), and bronze (partial annotation) (Wang et al., 2016). Metabolite annotations were based on searching the experimental spectra against the GNPS spectral library using the tools classic Molecular Networking – MN (Wang et al., 2016), Feature-Based Molecular Networking – FBMN (Nothias et al., 2020), Dereplicator+ (Mohimani et al., 2018), Network Annotation Propagation – NAP (da Silva et al., 2018), molDiscovery (Cao et al., 2021), MS2LDA (Van Der Hooft et al., 2016), MolNetEnhancer (Ernst et al., 2019), and analysis of chromatographic data such as retention time and UV spectra. These tools allow the integration of orthogonal annotation methodologies and tandem mass spectrometry data (MS/MS) to explore and obtain the metabolome of plants used as food.

2.5. Molecular Docking
Molecular docking analysis was performed using Autodock Vina v.1.2.0 (The Scripps Research Institute, La Jolla, CA, USA) docking software (Trott & Olson, 2010). The receptor site was predicted using LigandScout (Inte: Ligand) Advanced software (Wolber & Langer, 2005) (evaluation license key: 81809629175371877209), which identifies putative binding pockets by creating a grid surface and calculating the buriedness value of each grid point on the surface. The resulting pocket grid consists of several clusters of grid points, rendered using an iso surface connecting the grid points. The iso surface represents space that may be suitable for creating a pocket.

The x-ray crystal structure of Human Peroxiredoxin 5, a Novel Type of Mammalian Peroxiredoxin (PDB: 1HD2) (Declercq et al., 2001), was retrieved from the Protein Data Bank and utilised to perform docking simulations. Default search parameters were used where the number of binding modes was 10, exhaustiveness was 8, and the maximum energy difference was 3 kcal/mol.

LigandScout (Inte: Ligand) Advanced software (Wolber & Langer, 2005) (evaluation license key: 81809629175371877209) was used to generate 3D pharmacophore models. LigandScout’s algorithm calculates and displays chemical interactions between protein–ligand complexes.

3. Results and discussion

An untargeted metabolomics approach was employed to assess the metabolome of A. humile. The molecular diversity of its fruits and the evaluation of the abundance of ions and metabolite content were determined using high-resolution mass spectrometry. This was complemented with dereplication tools, which involved molecular structure searches in robust databases and computer-assisted fragmentation. Furthermore, given the limited data on the metabolic composition of A. humile in the existing literature, the scope of the study was
extended to identify biomarkers of the genus *Anacardium* besides describing the classes and subclasses of secondary metabolites present in the species. For this purpose, the structural similarity of MS data was assessed by comparing it with the spectral library from the GNPS ecosystem. Furthermore, potential candidates were identified *in silico* using specialised platforms such as Sirius, Dereplicator+, and NAP, utilising MS data obtained in both positive and negative ionisation modes. The detection of metabolites was performed in fruit samples of *A. humile* from three different locations in Brazil.

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

**Classic Molecular Networking (MN)**

The library matches using the classic molecular networking (MN) yielded 2164 hits with 198 unique library compounds in positive ionisation mode and 457 hits with 139 unique library compounds in negative ionisation mode after excluding nodes with repeated hits. In contrast, the merge networks polarity tool generated 2406 hits. Molecular networking-based strategies allowed us to annotate three times more metabolites in positive ionisation mode than negative mode. According to the Metabolomic Standard Initiative-MSI, all annotations were manually inspected, resulting in level 2 or 3 (Sumner et al., 2007). To obtain hits with higher structural similarity and the exact biosynthetic origin within molecular families, the threshold for the cosine score similarity was set to 0.7. The structural annotations were verified in the metabolome of the genus *Anacardium* using databases such as the Dictionary of Natural
Products. Initially, the annotations provided by the spectral library and the molecular family analysis were evaluated, suggesting precursor ions with similar structures or functional groups. After the dereplication process from the MN analysis, 88 metabolites belonging to different classes of secondary metabolites in aglycone and heteroside form were putatively annotated, including terpenes, steroids, flavonoids, tannins, quinones, alkaloids, coumarins, phenolic acids, and simple phenols. Additionally, many primary metabolites, such as carboxylic acids, sugars, and amino acids, were annotated in all samples. A high molecular diversity of derivative compounds has been annotated within these classes (see Table 1). In addition, a complex array of primary metabolites was detected in all samples, such as carbohydrates, amino acids, and lipids. Based on library matches, polyphenols were the major metabolites in all analyses. O-glycosylated compounds were abundant, and C-glycosylated compounds were in the minority. Alkaloids were detected in low concentrations in the fruits of *A. humile*, which information corroborates with the data reported for the genus *Anacardium*. Regarding the collection sites, no substantial variations were observed in metabolite content. All metabolites annotated were previously described in the family Anacardiaceae.

MS/MS, fragmentation pattern analysis, indicated the characteristic and diagnostic ions of the annotated metabolites. The glycosylated phenolic compounds found in high concentrations in the fruits were easily identified by the neutral loss of a sugar moiety, which is determined by the loss of 162 Da for hexosides, 146 Da for deoxyhexosides and 132 Da for pentosides in O-glycosylated flavonoids (Mannochio-Russo et al., 2020). C-glycosylated flavonoids were characterised by the loss of H2O and 120 Da (Mannochio-Russo et al., 2020). The compounds putatively identified as myricetin 3-O-xylloside and myricetin 3-rutinoside at m/z 449.073 and 625.141 [M–H]−, respectively, were identified by the consecutive losses of sugar moieties and also based on their UV spectra and retention times. Furthermore, the flavonoid aglycone fragmentation exhibits a pathway characterised by forming fragments originating from retro
Diels-Alder reactions coupled with losses of neutral molecules such as CO$_2$ and H$_2$O (Yang et al., 2021). Similarly, the flavonoid monohexosides and dihexosides isoquercitrin, guajavarin, myricetin 3-O-rutinoside, myricetin 3-O-galactoside, myricetin 3-O-xyloside, and rutin showed deprotonated molecular ions [M−H]$^-$ at m/z 463.089, 433.078, 625.141, 479.083, 449.073, and 609.146, respectively, previously reported in the literature (Fig. 1). Meanwhile, these compounds were connected with a mass difference of 146 Da, referring to a unit of deoxyhexoside, 16 Da as a possible extra hydroxyl group, and 30 Da as an extra OCH$_2$. In total, thirteen O-glycosylated flavonoids and four aglycone flavonoids were detected in positive ionisation mode, and eleven O-glycosylated flavonoids and six aglycone flavonoids were detected in negative ionisation mode. The retention time observed in the chromatographic analysis allowed us to improve the annotation and distinguish the O-glycosylated flavonoids. From mass spectrometry fragmentation, it was possible to perform structure-based propagation and guided detection of metabolites not annotated by the GNPS library since compounds of the same molecular family are structurally related and share molecular substructures that allow the putative identification of new molecules. Hence, from the difference between the nodes indicated by m/z 609.181 and 463.123 [M+H]$^+$, an analogue of the flavone diosmin was annotated, whose mass difference is 146 Da referring to a unit of deoxyhexosides. The detailed inspection of their MS/MS spectra allowed us to infer the glycosylation type. Through mass spectral similarity networking, accurate annotation of the flavan-3-ols catechin (m/z 289.072) and epigallocatechin (m/z 305.067) was achieved in negative ionisation mode [M−H]$^-$. Through their retention times, accurate masses, and MS/MS similar fragmentation patterns obtained in negative mode [M−H]$^-$, the methoxylated anthocyanidins 5-methylcyanidin (m/z 300.946) and methyldelphinidin (m/z 316.920) were annotated, which are widely distributed in fruits of the genus *Anacardium* (de Brito et al., 2007).
Table 1. Results of the metabolite annotation in the fruits of *A. humile* through molecular networking analysis.

<table>
<thead>
<tr>
<th>Parent mass</th>
<th>Adduct</th>
<th>Molecular formula</th>
<th>Metabolite name</th>
<th>Chemical structure</th>
<th>Chemical class</th>
</tr>
</thead>
<tbody>
<tr>
<td>303.217</td>
<td>[M+H]^+</td>
<td>C_{15}H_{10}O_7</td>
<td>Quercetin</td>
<td><img src="image" alt="Quercetin structure" /></td>
<td>Flavonoid</td>
</tr>
<tr>
<td>289.072</td>
<td>[M-H]^−</td>
<td>C_{13}H_{14}O_6</td>
<td>Catechin</td>
<td><img src="image" alt="Catechin structure" /></td>
<td>Flavonoid</td>
</tr>
<tr>
<td>595.158</td>
<td>[M+H]^+</td>
<td>C_{27}H_{30}O_{15}</td>
<td>Kaempferol 3-O-glucoside-7-O-rhamnoside</td>
<td><img src="image" alt="Kaempferol structure" /></td>
<td>Flavonoid</td>
</tr>
<tr>
<td>449.385</td>
<td>[M+H]^+</td>
<td>C_{21}H_{20}O_{11}</td>
<td>Kaempferol 7-O-glucoside</td>
<td><img src="image" alt="Kaempferol structure" /></td>
<td>Flavonoid</td>
</tr>
<tr>
<td>463.123</td>
<td>[M+H]^+</td>
<td>C_{23}H_{21}O_{11}</td>
<td>Peonidin 3-glucopyranoside</td>
<td><img src="image" alt="Peonidin structure" /></td>
<td>Flavonoid</td>
</tr>
<tr>
<td>300.946</td>
<td>[M+H]^+</td>
<td>C_{16}H_{13}O_6</td>
<td>5-methylcyanidin</td>
<td><img src="image" alt="5-methylcyanidin structure" /></td>
<td>Flavonoid</td>
</tr>
<tr>
<td>316.920</td>
<td>[M+H]^+</td>
<td>C_{16}H_{13}O_7</td>
<td>Methyldelphinidin</td>
<td><img src="image" alt="Methyldelphinidin structure" /></td>
<td>Flavonoid</td>
</tr>
<tr>
<td>449.385</td>
<td>[M+H]^+</td>
<td>C_{21}H_{20}O_{11}</td>
<td>Luteolin 7-O-glucoside</td>
<td><img src="image" alt="Luteolin structure" /></td>
<td>Flavonoid</td>
</tr>
<tr>
<td>449.385</td>
<td>[M+H]^+</td>
<td>C_{21}H_{20}O_{11}</td>
<td>Astragaline</td>
<td><img src="image" alt="Astragaline structure" /></td>
<td>Flavonoid</td>
</tr>
<tr>
<td>433.247</td>
<td>[M+H]^+</td>
<td>C_{21}H_{20}O_{10}</td>
<td>Apigenin glucoside</td>
<td><img src="image" alt="Apigenin structure" /></td>
<td>Flavonoid</td>
</tr>
</tbody>
</table>
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
579.145 [M+H]$^+$ C$_{30}$H$_{26}$O$_{12}$ Procyanidin B2 Flavonoid

579.341 [M+H]$^+$ C$_{27}$H$_{30}$O$_{14}$ Kaempferitin Flavonoid

257.972 [M+H]$^+$ C$_{15}$H$_{12}$O$_{4}$ Pinocembrin Flavonoid

537.168 [M−H]$^-$ C$_{30}$H$_{18}$O$_{10}$ Agathisflavone Flavonoid

349.183 [M+H]$^+$ C$_{22}$H$_{36}$O$_{3}$ 6-Pentadecylsalicylic Acid Phenolic acid

171.050 [M+H]$^+$ C$_{7}$H$_{6}$O$_{5}$ Gallic acid Phenolic acid

185.115 [M+H]$^+$ C$_{8}$H$_{8}$O$_{5}$ Methyl gallate Phenolic acid

355.070 [M+H]$^+$ C$_{16}$H$_{18}$O$_{9}$ Chlorogenic acid Phenolic acid

153.019 [M−H]$^-$ C$_{7}$H$_{6}$O$_{4}$ Protocatechuic acid Phenolic acid

177.034 [M+H]$^+$ C$_{6}$H$_{8}$O$_{6}$ Ascorbic acid Organic acid

345.244 [M−H]$^-$ C$_{22}$H$_{33}$O$_{3}$ Ginkgolic acid Organic acid
The systematic investigation of all nodes associated with the flavonoid class allowed us to annotate propagation of related compounds in the molecular networks from negative ionization mode, such as the galloyl flavonol glycoside derived from quercetin 3-\textit{O}-\textit{(6"-galloyl)}-glucopyranoside (\textit{m/z} 615.100 \text{[M−H]}^−) and the flavanone naringenin (\textit{m/z} 271.060 \text{[M−H]}^−). A detailed assessment of the flavonoid profiling using mass spectral data from two ionization modes showed other flavonoid structures that were not connected to other nodes from the molecular networking approach, including the flavonoids quercetin (\textit{m/z} 303.217 \text{[M+H]}^+; empirical formula: C_{15}H_{10}O_{7}), kaempferol 7-\textit{O}-glucoside (\textit{m/z} 449.385 \text{[M+H]}^+; C_{21}H_{20}O_{11}), luteolin 7-\textit{O}-glucoside (\textit{m/z} 449.385 \text{[M+H]}^+; C_{21}H_{20}O_{11}), astragalin (\textit{m/z} 449.385 \text{[M+H]}^+; C_{21}H_{20}O_{11}), apigenin glucoside (\textit{m/z} 433.247 \text{[M+H]}^+; C_{21}H_{20}O_{10}), quercetin 3,7-di-\textit{O}-glucoside (\textit{m/z} 627.469 \text{[M+H]}^+; C_{27}H_{30}O_{17}), hyperoside (\textit{m/z} 463.125 \text{[M−H]}^−; C_{21}H_{20}O_{12}), and the biflavonoid amentoflavone (\textit{m/z} 537.168 \text{[M−H]}^−; C_{30}H_{18}O_{10}) previously reported in \textit{A. humile} (Lima Júnior et al., 2021; Luiz-Ferreira et al., 2008) and \textit{A. occidentale} (Andarwulan et al., 2012; Salehi et al., 2020; Taiwo et al., 2017). The isomers were separated by their chromatographic profile, and their precursor ions displayed a different fragmentation pattern.

Applying this same approach, the annotation of low molecular weight metabolites provided some bioactive phenolic acids such as gallic acid (\textit{m/z} 171.050 \text{[M+H]}^+; C_{7}H_{6}O_{3}), methyl gallate (\textit{m/z} 185.115 \text{[M+H]}^+; C_{8}H_{8}O_{3}), protocatechuic acid (\textit{m/z} 154.980 \text{[M+H]}^+; C_{7}H_{6}O_{4}), and chlorogenic acid (\textit{m/z} 355.070 \text{[M+H]}^+; C_{16}H_{18}O_{9}). All these phenolic acids have been previously reported in the \textit{Anacardium} genus (Andarwulan et al., 2012; Lima Júnior et al., 2021; Luiz-Ferreira et al., 2008). The fragmentation characteristics and biosynthetic knowledge indicated the oxidation pattern and location of the sugar units.

A more significant number of flavonoids were detected, mainly in the negative ionization mode. A high diversity of flavonoids, mainly \textit{O}-glycosyl flavonols, was confirmed by the large number of clusters observed in the MN. Combined with detailed taxonomical knowledge,
allowed for the annotation of metabolites not available in the reference library spectra. MN has been successfully applied to explore the molecular complexity of *A. humile* fruits and annotate structurally related molecules.

In summary, *O*-glycosylated flavonoids belonging to the subclass of flavonols were dominant primarily based on the structure of myricetin and quercetin and linked mainly to sugars, glucose and rhamnose. Through the evaluation of the annotations proposed by the MN library and manual inspection of fragmentation spectra, four phenolic acid glycosides were putatively identified and annotated as benzoic acid hexoside derivatives and coumaroyl hexoside derivatives (Fig. 1). The phenolic acid glycosides classified as cinnamic acid derivatives showed deprotonated molecular ions [M−H]$^-$ at m/z 325.09 (coumaroyl hexoside) and 341.08 (caffeic acid hexoside); and the benzoic acid hexoside derivatives showed deprotonated molecular ions [M−H]$^-$ at m/z 315.07 (benzoic acid + 2O, O-Hex) and 299.08 (benzoic acid + O, O-Hex) with a mass difference of 16 Da indicating a hydroxyl group as the only structural difference. In addition, its spectrum showed a typical fragment with the elimination of CO$_2$. The metabolite annotated as 1-*O*-trans-cinnamoyl-glucopyranose at m/z 309.10 [M−H]$^-$ showed a fragment in MS/MS at m/z 147.04 from neutral loss of the sugar moiety (Glc, 162 Da), previously isolated from cashew apple (*Anacardium occidentale* L.) (Michodjehoun-Mestres et al., 2009). Its derivative 1-*O*-trans-cinnamoyl-(6-*O*-galloyl)-glucopyranose at m/z 461.110 ([M−H]$^-$) exhibited high spectral similarity and spectral matching to reference library spectra. The negative ionization mode was more sensitive for detecting these polyphenols.

The presence of other glycosylated polyphenols was investigated in *A. humile* fruits. These analyses found a high abundance of gallotannins (hydrolyzable tannins) formed by polygalloyl esters of glucose. In this work, two types of gallotannins with degrees of polymerization of 2 (m/z 483.08 [M−H]$^-$) to 3 (m/z 635.089 [M−H]$^-$) galloyl units linked to glucose were putatively
identified. Their fragmentation patterns were mainly characterized by the loss of 170 Da and 152 Da, referring to the loss of a unit of gallic acid and galloyl fission, respectively.

![Diagram](image)

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

Although gas chromatography-coupled mass spectrometry is the most used technique in the investigation of chemical markers of *Anacardium*, the analysis by LC-MS/MS with electrospray ionization enables the detection of the various markers present in the fruits of *A. humile*. The interpretation of MS² data and evaluation of structurally related molecules annotated by the spectral library and network propagation of *m/z* differences, as well as retention time analysis, provided the structure of the 16 chemical markers classified as cardols, anacardic acids, 2-methylcardols, and cardanol with different degrees of unsaturation (saturated, monounsaturated, di-unsaturated, and tri-unsaturated), confirmed by data reported in the literature (Gomes Júnior et al., 2018; Kubo et al., 1994). The fragmentation pattern of
these metabolites (Fig. 2) in negative ionization mode was inspected in detail by observing the neutral losses, relative abundances, and diagnostic ions based on their network connection. The cluster containing these molecules shows the metabolites of each class with C15-alkyl side-chain with 0 - 3 double bonds. The chemical marker annotated were cardols ((1) C_{15:3},5-\{8(Z),ll(Z),l4-pentadecatrienyl\}resorcinol; (2) C_{15:2},5-\{8(Z),ll(Z)-pentadecadienyl\}resorcinol; (3) C_{15:1},5-\{8(Z)-pentadecenyl\}resorcinol; (4) C_{15:0},5-pentadecylresorcinol), Anacardic acids ((5) C_{15:3},6-\{8(Z),l1(Z),l4-pentadecatrienyl\}salicylic acid; (6) C_{15:2},6-\{8(Z),ll(Z)-pentadecadienyl\}salicylic acid; (7) C_{15:2},6-\{8(Z)-pentadecenyl\}salicylic acid; (8) C_{15:0},6-pentadecylsalicylic acid), 2-methylcardols ((9) C_{15:3},2-methyl-5-\{8(Z),ll(Z)-pentadecatrienyl\}resorcinol; (10) C_{15:2},2-methyl-5-\{8(Z),ll(Z)-pentadecadienyl\}resorcinol; (11) C_{15:1},2-methyl-5-\{8(Z)-pentadecenyl\}resorcinol; (12) C_{15:0},2-methyl-5-pentadecylresorcinol), Cardanols ((13) C_{15:3},3-\{8(Z),ll(Z),l4-pentadecatrienyl\}phenol; (14) C_{15:2},3-\{8(Z),ll(Z)-pentadecadienyl\}phenol; (15) C_{15:1},3-\{8(Z)-pentadecenyl\}phenol; (16) C_{15:0},3-pentadecylphenol) (Gomes Júnior et al., 2018; Kubo et al., 1994). The metabolites were detected in both positive and negative ionization modes.
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...
compounds, 37 with a gold classification and 24 with a cosine score > 0.9. Evaluating the same
characteristics described for the ESI (+), only two organic acids and two flavonoids were
annotated.

Merge Polarity Networks

Notwithstanding the extensive search against the GNPS spectral library and the use of the
annotation propagation strategy to explore structural relationships, many ions remained
unannotated. This further motivated the application of more sophisticated dereplication tools.
The Merge Polarity Networks function was applied, combining data from both ionization
modes, and yielding more accurate annotations by calculating similarity scores between
structures (Tanimoto). This assessment showed 2406 pairs and yielded four relatively abundant
features (m/z 214.09; 163.06; 147.08; 348.07 [M+H]+) with Tanimoto values equal to 1 (perfect
agreement), and eight structures were compatible in both ionization modes but were not similar
(Tanimoto = -1). This allowed the annotation of kaurenoic acid glucopyranosyl derivative at
m/z 645.29 [M+H]+, 1,6-anhydro-glucose at m/z 163.06 [M+H]+ and the flavonoid 4''-O-
acetylmyricitrin at m/z 627.16 [M+H]+.

Feature-Based Molecular Networking (FBMN)

Although classic MN is very useful for metabolic annotation through spectral library
matching, propagating annotations of unannotated nodes and reducing the size of datasets in
untargeted metabolomics, it has some limitations. It does not use MS1 data or provide the
resolution of isomers with different retention times in the chromatographic analysis or coeluting
metabolites. Besides, MN does not allow the accurate inference of the relative amounts of the
metabolites (abundance of the ions) between samples. However, these challenges can be solved
by employing Feature-Based Molecular Networking (FBMN), which provides greater
resolving power than MN by using MS1 information to verify retention time, peak intensity,
ion mobility and isotope patterns (Nothias et al., 2020). Therefore, FBMN was employed to

discriminate isomeric/isobaric compounds and infer the ion abundance of some annotated

substances and their analogues using chromatographic peak area and peak intensity through

feature detection and alignment analysis. FBMN analysis detected 860 features in ESI (+)-MS

mode and 340 in ESI (-)-MS. From the relative intensities of the ions, it was possible to infer

that the antioxidant compounds gallic acid, ascorbic acid, and anthocyanins were the most

abundant metabolites found in *A. humile* (Lima Júnior et al., 2021). In addition, the FBMN

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

**Network Annotation Propagation (NAP)**

MS/MS-based metabolomics experiments were analyzed in detail using *in silico* fragmentation tools and complementary methodology such as NAP to improve the reliability of the metabolic annotation suggested by the spectral library to annotate unmatched precursor ions and obtain a highly reliable metabolic profile. The results obtained from the NAP workflow in positive ionization mode were putatively assigned as epigallocatechin, maltose, sucrose, peonidin 3-O-glucopyranoside and corroborated the annotation made by the library.

Of the metabolites not annotated by the library (N/A), NAP workflow yielded 64 candidates with consensus with Metfrag (Fusion ID), being the majority belonging to carboxylic acids, glycosylated aldehyde (helcin resulting from the oxidation of the benzylic hydroxy group of salicin to the corresponding aldehyde), chromenes, lactones, quinones (2,6-dimethoxyquinone), triterpenes, and triterpenoid saponin. The MetFrag yielded several candidates belonging to the classes of phenolic acid (benzoic acid derivatives and phenolic compounds with catechol groups, vanillic acid) and terpenoids (diterpene). From metabolites not annotated by the library (N/A), NAP generated 43 candidates with consensus with Metfrag (Fusion ID). Most metabolites were classified as carboxylic acids and derivatives (linear and aromatic acids and lactones), simple phenols, quinones, glycosylated aromatic substances, terpenes, flavonoids (aglycones and heterosides), and phenylpropanoids (phenolic acids).
tool improved the structural annotation based on the GNPS library and provided reliable candidates from *in silico* spectra matching.

*Dereplicator*+

High-confidence identification methods from *in silico* database search were employed in this study, including Dereplicator+. This high-throughput identification tool improves the annotation of peptides and small molecules and has been successfully applied in metabolomic approaches to natural products (Mohimani et al., 2018). The positive ionization mode resulted in 47 unique compounds, including phenolic acid, chromenes, and terpenes. Terpenes and their derivatives were the most annotated class of natural products (over 50%). After inspection of the MS/MS spectra, the tricyclic triterpenoid achilleol B, the triterpenoid betulinaldehyde, the steroid ergosta-3,5,24(28)-triene, and a chromanol derivative were annotated. In the negative ionization mode, 15 unique library compounds belonging to flavonoids, tannins, quinones, and terpenes classes were obtained.

*MolDiscovery*

MolDiscovery was employed to accelerate chemical structural annotation and enhance the efficiency and accuracy of metabolic identification. This probabilistic model uses a fragmentation algorithm to generate MS/MS fragmentations and match compounds with their MS/MS spectra (Cao et al., 2021). As recommended, the analysis of unique metabolites generated by the workflow with high molecular weight was conducted, as the tool demonstrates greater accuracy for masses exceeding 600 Da, attributed to the frequently elevated count of fragment ions. Furthermore, the fragmentation of small molecules considers the type of fragmented bonds and other factors such as moiety (Cao et al., 2021). MolDiscovery exhibited 2.226 unique metabolites in (+)-ESI mode with high mass ranges >600 Da and a cut-off score of 15 (159 hits). A larger score implies a greater probability of correct annotation (Cao et al.,...
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 punicacortein 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 tricin 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
losses of 44 (free carboxyl) and 18 (water) from glycosylated derivatives, corroborating the high abundance of gallotannins and structurally close polyphenols in *A. humile* fruits. In addition, many neutral losses of 162 Da and 146 Da revealed the presence of sugar residue such as hexose and pentose, which can be attributed to the high abundance of O-glycosylated flavonoids found in all samples. The results obtained from the MS2LDA workflow improved the confidence level of the annotations, expanded the metabolite annotation, and characterized metabolites not yet reported in the species through the analysis of functional groups and core structures provided by the information from Mass2Motif.

*MolNetEnhancer*

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

There was no significant difference in the metabolic profiling concerning collection sites, as represented in Fig. 5. Based on metabolomic profiling obtained of the extracts of *A. humile*
fruits from three collection sites, the flavonoids were the major metabolites in all analyses. Considering that flavonoids have remarkable free radical scavenging properties, the flavonoids peonidin 3-O-glucopyranoside, methylcyanidin and methyldelphinidin were selected for determination of antioxidant activity using molecular docking.

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

**Determination of antioxidant activity using molecular docking**

As peonidin 3-O-glucopyranoside, methylcyanidin and methyldelphinidin are the major metabolites in *A. humile* based on the above metabolomics studies. A molecular docking study was undertaken to assess the antioxidant activity of these compounds. The docking study was carried out on the crystal structure of Human Peroxiredoxin 5 (PRDX5, PDB: 1HD2) (Declercq et al., 2001), a novel type of *Mammalian peroxiredoxin* that has antioxidative and cytoprotective functions during oxidative stress. The peroxiredoxins define an emerging family
of peroxidases able to reduce hydrogen peroxide and alkyl hydroperoxides with the use of reducing equivalents derived from thiol-containing donor molecules such as thioredoxin, glutathione, trypaonothione and AhpF. Peroxiredoxins have been identified in prokaryotes as well as in eukaryotes (Arden et al., 2015). Peroxiredoxin 5 is widely expressed in tissues and located cellurally to mitochondria, peroxisomes, and cytosol and, therefore, implicated in antioxidant protective mechanisms (Knoops et al., 2011).

The three major compounds previously mentioned were subjected to docking analysis, and the specificities of their interaction with these targets, as shown in Fig. 6, were investigated. The best-docked complexes were obtained based on binding energies and interacting residues. Docking poses were analyzed and compared to the co-crystallized standard antioxidant nordihydroguaiaretic acid (NDGA). All three major compounds in this molecular docking study docked very well compared to the standard Nordihydroguaiaretic acid (Mala et al., 2020) (Table 2).

### Table 2 Results of the metabolite annotation in the fruits of A. humile.

<table>
<thead>
<tr>
<th>Compounds</th>
<th>Docking Score (-) (kcal/mol)</th>
<th>PDB ID: 1HD2 (Human Peroxiredoxin 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nordihydroguaiaretic acid (Standard)</td>
<td>5.3</td>
<td></td>
</tr>
<tr>
<td>Peonidin 3-O-glucopyranoside</td>
<td>5.02</td>
<td></td>
</tr>
<tr>
<td>Methylcyanidin</td>
<td>5.46</td>
<td></td>
</tr>
<tr>
<td>Methyldelphinidin</td>
<td>6.20</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 6. Binding site (yellow colour) of Human Peroxiredoxin 5.
Pharmacophore Evaluation (Structure-based pharmacophore)

Pharmacophores represent chemical functions. A pharmacophore abstractly describes steric and electronic features required to trigger (or block) biological response (Khedkar et al., 2007). A pharmacophore model can explain how structurally diverse ligands bind at the receptor site based on common interaction points (Seidel et al., 2019). Structure-based pharmacophore modelling is a method for pharmacophore development based on the target protein's structural features (Szwabowski et al., 2023). In this method, the possible active site in protein where the interactions of co-crystallized ligand occur will be analyzed. Due to its simplicity, this method is computationally very efficient and exceptionally well suited for the virtual screening of a wide range of compound libraries. This method searches for interactions between ligands and the macromolecule. The three major peonidin 3-\textit{O}-glucopyranoside, methylcyanidin and methyldelphinidin were used for pharmacophore evaluation.

Ligand peonidin 3-\textit{O}-glucopyranoside interacts with the macromolecule PRDX5, as shown in Fig. 7. It shows hydrophobic effects with amino acids Ile119A and Phe120A. Leu116A and Thr147A. This pharmacophore also represents a hydrogen bond acceptor (HBA) feature with nearby amino acid residue Arg127A. Looking at the hydrogen bond donor (HBD) feature, the ligand interacts with Thr147A amino acid residue. Based on the interactions, this structure provides a pharmacophore with three features hydrophobic effects (H), hydrogen bond acceptor (HBA) and hydrogen bond donor (HBD), with a pharmacophore score of (0.7).

Ligand methylcyanidin interacts with the macromolecule PRDX5 as shown in Fig. 7. It shows hydrophobic effects with BR303A and amino acids Leu116A, Thr147A, Phe120A, and Thr44A. This pharmacophore has hydrogen bond acceptor (HBA) and hydrogen bond donor (HBD) features with nearby amino acid residues Cys47A and Thr44A. Based on the interactions, this structure provides a pharmacophore with three features hydrophobic effects...
(H), hydrogen bond acceptor (HBA) and hydrogen bond donor (HBD), with a pharmacophore score of (0.89).

Ligand methyldelphinidin interacts with the macromolecule PRDX5, as shown in Fig. 7 (A) and (B). It shows hydrophobic effects with the amino acids Leu116A and Ile119A. This pharmacophore also has both hydrogen bond acceptor (HBA) and hydrogen bond donor (HBD) features with nearby amino acid residues Cys47A and Thr44A. Arg127A only has hydrogen bond acceptor (HBA) interactions with the ligand. Based on the interactions, this structure provides a pharmacophore with three features hydrophobic effects (H), hydrogen bond acceptor (HBA) and hydrogen bond donor (HBD), with a pharmacophore score of (0.87). The above three structure-based pharmacophores further generated a shared feature pharmacophore, as shown in Fig. 8.
Fig. 7. Ligands interaction with the macromolecule. (A) 3D view of pharmacophore presents in the macromolecule binding site. (B) 2D view of pharmacophore. Pharmacophore features: Hydrophobic interactions (H), Yellow; Hydrogen bond acceptor (HBA), Red; Hydrogen bond donor (HBD), Green.

Fig. 8. Generated Shared feature pharmacophore.
4. Conclusion

The application of the integrated metabolomics approach and the combination of structural annotation and data mining tools provided critical insights into the characterization and quantification of constituents of the healthy diet, with promising application in food chemistry. This study has revealed a comprehensive metabolic profile and a high content of bioactive polyphenols in bushy cashew fruits, which hold significant potential from a biotechnological perspective. Moreover, the state-of-the-art computational tools employed in this study enriched molecular networks, allowing the rapid and effective annotation of a vast array of metabolites not yet reported in *A. humile*. Additionally, molecular docking analysis was performed on peonidin 3-*O*-glucopyranoside, methylcyanidin and methyldelphinidin to assess their binding to key receptors associated with antioxidant mechanisms. Compared to standard compounds, these compounds exhibited strong critical potentials related to antioxidant activity. A structure-based pharmacophore model was proposed to guide future studies, potentially aiding in selecting and synthesising cost-effective compounds with good bioactivity.

Declaration of Competing Interest

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.

Acknowledgements

The authors acknowledge financial support from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the institutional and financial support.

References


Kubo, I., Ochi, M., Vieira, P.C., Komatsu, S., 1993. Antitumor agents from the cashew


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


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: