SUPPLEMENTARY DATA

Supplementary LC-MS^2 Data Processing

For the first approach, LC-MS^2 vendor datasets output files in .d (generated in this study) and .wiff (Vakilian et al., 2021) extension were converted to .mzXML format using the MSconvert tool from ProteoWizard software version 3.0.24033 (Chambers et al., 2012) and imported to GNPS to perform Classical Molecular Networking (CMN). The molecular network was created using a precursor and fragment ion mass tolerance of 0.02 Da, a Cosine Score of 0.6, a requisite minimum of 4 matching ions. Peaks from blank samples were filtered from the network. Default values were retained for other parameters. From this result, both the feature relative quantification table (also known as bucket table) and the aligned features as an .MGF file were exported. For approach two, we analyzed the .MGF file using MolDiscovery (Cao et al., 2021), DEREPLICATOR+ (Mohimani et al., 2018), and CSI:FingerID *in-silico* tools. The latter tool is integrated within the SIRIUS software, aiming to determine the chemical formula ([N/A] (Ludwig et al., 2020) and molecular structure (Dührkop et al., 2015; [Hoffmann et al., 2022). For MolDiscovery and DEREPLICATOR+, the parameters chosen were as follows: a precursor and fragment ion mass tolerance of 0.02 Da; maximum charge of 1; AllDB as predefined database. For CSI:FingerID (SIRIUS Software ver 5.8.6), the parameters chosen were as follows: adducts [M + H]+, [M + H_2O + H]+, [M - H_2O + H]+, [M + H_3N + H]+, [M + K]+ and [M + Na]+; PubChem and Bio database were scanned; and only compounds with a mass < 850 Da were selected to reduce processing times. To describe the chemical profile of the extracts, only features (nodes) within the molecular network (already blanked) were utilized. Redundant identifications were removed, retaining only one identification per feature using the priority order: GNPS > MolDiscovery > DEREPLICATOR+ > CSI:FingerID. An initial mass-range filter was applied for each annotation tool (except GNPS) according to [PMID: 34140479]) Cao et al.'s (2021) hints. Features under the following ranges were kept: MolDiscovery > 600 Da; DEREPLICATOR+ > 400 Da; CSI:FingerID < 600 Da. Annotations with low mass accuracy (greater than 10 ppm) were filtered out. The putative identified metabolites retained after the filtering steps conformed the spectral library and their chemical class was determined using the ClassyFire (classyfire.wishartlab.com) chemical ontology nomenclature (Feunang et al., 2016). The generated output, considered as an enriched molecular network, was visualized in Cytoscape (version 3.10.1) (Shannon et al., 2003) and colored at the class level. The putative metabolite identifications by GNPS spectral matching and *in-silico* tools are level 2 and 3, respectively, according to the Metabolomics Standards Initiative reporting standards (Sumner et al., 2007). Most of the procedures, including filtering and integration of annotations and chemical classes into the molecular network, were carried out using KNIME (ver. 4.7.0) (Fillbrunn et al., 2017). Data visualization was done using R software version 4.2.2 and RStudio version 2023.12.1+402. For data manipulation, the dplyr library ([N/A] Wickham H, 2023) was used. Pie charts plots were constructed using the Plotly library ([N/A] Plotly Technologies Inc, 2015), while upset plots were created using ggplot ([N/A] Wickham, 2016) and ComplexUpset (Lex et al., 2014) libraries.

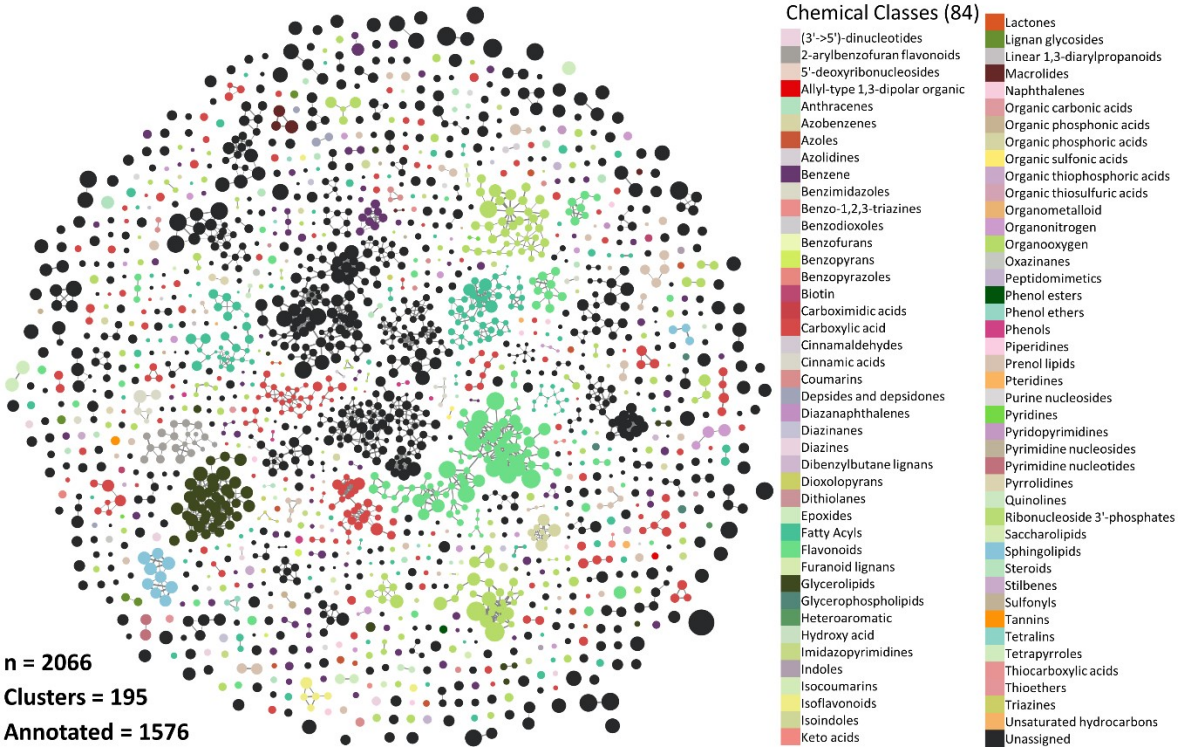

Figure S1. Complete molecular network displaying the metabolite repertoire of *Aloe vera* and *Moringa olefeira*. Each node represents a unique MS^2-containing peak, and the color indicates the associated chemical class (ClassyFire ontology). Nodes within a cluster (comprising at least two connected nodes) indicate MS^2 spectral similarity between them (Cosine score ≥ 0.6). The assignment of the chemical classes to clusters was determined by the predominant class within each respective cluster. The size of each node indicates the proportion of the metabolite mass, (m/z). Singletons are included in this network.