Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Computational metabolomics reveals flavonoid architectures underlying species-specific chemical identity in medicinal Plectranthus

Akhona Myoli
University of Johannesburg
Ntakadzeni Edwin Madala
University of Venda
Farhahna Allie
University of Johannesburg
Justin J.J. van der Hooft
Wageningen University & Research
Fidele Tugizimana
ftugizimana@uj.ac.za

University of Johannesburg

Research Article

Keywords: Computational metabolomics, GNPS, MS2LDA 2.0, MS2Query, natural products, flavonoids, Plectranthus

Posted Date: July 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10436578/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Metabolomics on October 25th, 2026. See the published version at https://doi.org/10.1007/s11306-026-02546-x.

Abstract

Plant metabolomes encode biochemical information underlying species diversity, ecological adaptation, and pharmacological potential. Within the genus *Plectranthus*, diterpenoids are widely regarded as both the primary bioactive constituents and species-specific chemical signatures; however, the extent to which other bioactive metabolite classes contribute to species-specific chemical signatures remains poorly understood. Here, we applied a computational metabolomics approach to characterize and compare the leaf metabolomes of *Plectranthus hadiensis* and *Plectranthus ambiguus*, two medicinally important species used in South African ethnopharmacology. Untargeted LC–MS/MS profiling combined with feature-based molecular networking followed by comparative metabolomics and GNPS library matching, MS2Query, and MS2LDA 2.0-enabled structure and substructure-level annotations revealed extensive metabolic diversity beyond diterpenoids. While the prenol lipid chemical class, consisting of diterpenoids, was confirmed as a major discriminatory chemical class, flavonoids emerged as key drivers of species-level chemical differentiation. Distinct structural patterns were observed, with glycosylated flavones predominating in *P. hadiensis*, whereas methylated flavones and glycosylated flavonols were enriched in *P. ambiguus*. These compositional differences reflect underlying variation in flavonoid biosynthetic modifications, suggesting divergent metabolic regulation between species. Therefore, computational metabolomics revealed that flavonoid structural diversity constitutes a previously underappreciated layer of chemical specialization in *Plectranthus* plants and contributes substantially to species-specific chemical identity beyond the emphasized diterpenoids. By moving beyond targeted phytochemical screening, this work establishes a scalable framework for resolving species-specific metabolic architectures in medicinal plants. These findings have implications for chemotaxonomy, natural product discovery, and the rational exploration of plant-derived bioactives, contributing to a deeper understanding of the molecular basis of ethnomedicinal efficacy.

1. Introduction

Throughout history, plants have played and continue to play a key role in the discovery and development of pharmaceutical drugs and multifunctional cosmeceuticals (Nicolai et al., 2020; Chaachouay and Zidane, 2024). Among these plants, the genus *Plectranthus* has been recognized for its bioactive compounds, which have garnered interest in its pharmacological potential (Guo, 2017; Newman and Cragg, 2020). Also known as spurflowers, the genus comes from the family Lamiaceae and consists of approximately 300 species that are distributed through the tropical and warm regions of the world including Africa, India and Australia (Arumugam et al., 2016; Marquesl et al., 2012). The worldwide distribution of *Plectranthus* species translates to an assortment of ethnomedicinal uses that vary culturally from region to region. In South Africa, *P. ambiguus* (known as iBoza elincane in isiZulu) and *P. hadiensis* (known as iLozane in isiXhosa) grow freely in the southern KwaZulu-Natal (KZN) and the northern parts of the Eastern Cape (EC) provinces. Phenotypically, the small leaves of *P. ambiguus* are easily distinguishable from the large leaves of *P. hadiensis* – this is one of the morphological differences between the two plant species (Van Wyk and Smith 2001; Van Jaarsveld, 2006; Hao et al., 2022).

Furthermore, both *P. ambiguus* and *P. hadiensis* are part of the most traditionally utilized *Plectranthus* species in South Africa (Rice et al., 2011). These species have been widely used in ethnomedicine for the treatment of various ailments, including respiratory conditions, skin infections, fever and inflammation (Matisa et al., 2019). This longstanding ethnomedicinal relevance of the two species highlights the need for investigating their chemical composition in greater detail. Previous studies have identified that *Plectranthus* species, including *P. ambiguus* and *P. hadiensis* are particularly rich in specialized metabolites such as diterpenoids (Abdel-Mogib et al., 2002). For example, modified abietane-type diterpenoids are frequently highlighted as the major bioactive constituents contributing to the pharmacological properties of these plants. Additionally, recently, other classes of compounds such as phenolic acids and flavonoids have also been associated with their biological activity (Śliwiński et al., 2020; Lambrechts and Lall, 2021; Barbosa et al., 2023).

However, prior phytochemical studies have largely focused on the classical screening methods targeting these well-known abietane diterpenoids (Kubínová et al., 2014; Śliwiński et al., 2020; Sitarek et al., 2020; Ntungwe et al., 2021; Nicolus et al, 2023). While these studies have provided valuable insights, they offer only a partial view of the overall chemical repertoire of these species. Based on the available literature, as far as we know, no comprehensive metabolomic comparison has been conducted between *P. ambiguus* and *P. hadiensis*. Such knowledge gaps on the metabolomic landscape of these species restricts our understanding of their phytochemical complexity and potentially overlooks other important bioactive compounds. Moreover, the exclusive focus on diterpenoids constrains our ability to identify shared or species-specific metabolic fingerprints that may explain differences or similarities in their traditional uses and biological activities.

Thus, reported herein is a comprehensive metabolomics-based approach to investigate and compare the complex metabolite profiles of *P. hadiensis* and *P. ambiguus*. This untargeted, computational metabolomics work integrates feature finding with mzmine 4, comparative metabolomics and annotation tools available within Feature-based Molecular Networking (FBMN), as well as MS2Query and MS2LDA 2.0 (Heuckeroth et al., 2024, Nothias et al., 2020, de Jonge et al., 2023; van der Hooft et al., 2016) to visualize and uncover hidden layers of chemical diversity that may underpin species-level metabolic differentiation. As such, the work enables a holistic exploration of the chemical diversity and chemical signatures of these two traditionally important South African medicinal plants. This comparative analysis contributes to bridging indigenous knowledge with modern analytical science, expanding the current understanding of *Plectranthus* phytochemistry beyond classical diterpenoid profiles. Thus, through the application of an integrated computational metabolomics pipeline, this study, provides new insights into the chemical differentiation and potential pharmacological significance of *P. ambiguus* and *P. hadiensis*.

2. Experimental procedure

2.1 Metabolite extraction and sample preparation

Plectranthus ambiguus and *Plectranthus hadiensis* plants (60 days old) were purchased from Random Harvest Nursery in Krugersdorp, South Africa. The leaves of the plants were harvested, air-dried, and crushed into fine powder. For metabolite extraction, 200 mg of plant material was accurately measured and dissolved in 10 mL of 80% methanol (Chemlab, UK). The samples were then placed in a digital rotisserie tube rotator (Thermo Fisher, Johannesburg, South Africa) at 70 rpm and spun overnight. The crude extracts were centrifuged afterwards at 1330 rpm at 4°C using a benchtop fixed-angle centrifuge (Thermo Fisher, Johannesburg, South Africa). The supernatants, after centrifugation, were filtered using 0.22 µm nylon filters into glass vials and stored at 4°C until further analysis. For this study, five independent biological replicates were used.

2.2 Data acquisition – LC-MS/MS analysis of crude methanolic leaf extracts

Samples (methanol extracts) were analyzed on a liquid chromatography–quadrupole time-of-flight mass spectrometry (LC-qToF-MS) system (LCMS-9030, Shimadzu Corporation, Kyoto, Japan). Chromatographic separation was achieved by injecting 3 µL sample onto a Shim-pack C18 column (100 mm × 2.1 mm, 2.7 µm) (Shimadzu Corporation, Kyoto, Japan) maintained at 55°C. The reverse-phase chromatographic separation was performed using a binary mobile phase and a gradient elution program. The solvent system consisted of solvent A (0.1% formic acid (HPLC grade, Merck, Darmstadt, Germany) in Milli-Q water) and solvent B (consisting of methanol (HPLC grade, Romil SpS, Cambridge, UK) with 0.1% formic acid), with a flow rate of 0.4 mL/min. The gradient elution was performed as follows, B referring to organic composition (*i.e*., solvent B): 5% B for 3 min, 5–40% B over 3–5 min, 40–95% B over 5–12 min, then 95% B for from 12 min to 18 min. The gradient was changed back to initial 5% B at 18 min and kept to 20 min. The column was re-equilibrated for 3 min. The column effluents from chromatographic separation were further analysed with the high-definition mass spectrometer, equipped with electrospray ionization (ESI). Following method optimization, mass spectral data were acquired in negative ionization mode based on optimal ion efficiency. The mass spectrometer was operated under the following conditions: 4.0 kV interface voltage, with interface temperature of 300°C; 3 L/min flow rate for nebulization and dry gas; DL temperature of 250°C, and 400°C for heat block; detector was operated at 1.8 kV voltage. For monitoring the accuracy of acquired mass-to-charge ratio (*m/z*), sodium iodide (NaI) was used as a calibration solution. Both full scan (MS^1) and tandem MS/MS (MS^2) spectral data were acquired over an *m/z* range of 100–1200 Da. MS/MS experiments were conducted using a data-dependent acquisition (DDA) method with an intensity threshold of 5000 counts. Ion fragmentation was achieved via Collision-Induced Dissociation (CID) method using argon as a collision gas at a collision energy of 30 eV with a spread of 5 eV.

2.4 Peak detection and alignment

The acquired raw data files in the mzML format (ESI negative centroid data) were processed using mzmine 4.8.30. For MS1 mass detection, the factor of the lowest signal parameter was selected with the corresponding noise factor set at 8.0, and for MS2, it was set at 1.1. Thereafter, the parameters for the chromatogram builder were set as follows: the minimum consecutive scans were set to 5, with the

minimum intensity for consecutive scans and minimum absolute height set at 6.0E3 and 7.0E3, respectively. To resolve the detected peak features, the chromatographic threshold was set at 85% with the minimum search range retention time (RT) set at 0.04. Sample alignment was then performed with the weight for *m/z* and RT set at 3 and 1, respectively, to generate an aligned feature list. Gap filling of the aligned feature list was then performed with the intensity threshold set at 20%. The resulting gap-filled and aligned feature list (peak area matrix) was composed of 3531 features (detected peaks). The full set of mzmine processing parameters are provided as a batch file in Zenodo (Doi: 10.5281/zenodo.20112661).

2.5 Statistical analysis and identification of discriminant markers

The feature quantification table, exported from MZmine, was then used to generate ecological diversity metrics and multivariate chemometric models. Using each feature as a unique entry, ecological diversity metrics (Shannon and Simpson indices) were calculated to determine metabolite diversity (equations in Section 1 of the **Supplementary file**) following normalization of the feature quantification table (peak area matrix) using the total sum scaling. The Shannon and Simpson diversity indices were computed using the vegan package in R. To compute the multivariate chemometric models, the feature quantification table was imported into soft independent modelling of class analogy (SIMCA) version 18.0 software (Sartorius, South Africa). Principal Component Analysis (PCA) and Orthogonal Partial Least Square-Discriminant Analysis (OPLS-DA) models were computed for data exploration, group classification, and discriminant analysis.

2.6 Computational strategies for metabolite annotation

Post MZmine data processing, the GNPS export files, both the spectral .mgf file and feature quantification table, were exported into the newly released GNPS2 ecosystem (Wang et al., 2016) and FBMN was then computed. The parameters used for computing FBMN for the spectral dataset included precursor ion mass tolerance of 0.25 Da with fragment ion mass tolerance of 0.15 Da. For spectral similarity, a cosine score cut-off was set at 0.65, with a minimum of 5 matched fragment ions. The default GNPS2 spectral libraries (all reference spectra in the LC library irrespective of the ionization mode) were used for metabolite annotation. To visualize and analyze the computed network, the Cytoscape network visualization software (version 3.8.2) (Shannon et al., 2003; Smoot et al., 2011) was used. All putative node annotations and some unmatched/unidentified nodes, visualized in Cytoscape, were verified using their empirical formulae calculated from accurate mass and fragmentation patterns. In addition to the manual inspection of annotations, PubChem (https://pubchem.ncbi.nlm.nih.gov/) was also used to verify the exact spectral library matches obtained from GNPS, MS2Query and MS2LDA by comparing the fragmentation patterns of each annotation to its corresponding PubChem entry.

In addition to annotations sourced from GNPS spectral library matching, MS2Query (version 1.5.4) was also carried out for analogue and exact metabolite. MS2Query is a tool that enables high-throughput MS2 library searching, supporting the simultaneous detection of structural analogues and exact spectral

matches. The tool provides reliable predictions for both analogues and exact library matches using MS2 information (de Jonge et al., 2023). The same mass spectral file (.mgf file) used for FBMN was used as an input file for MS2Query. Subsequently, the MS2Query run was executed via the command line in Python 3.9, following the script provided on its official GitHub page (https://github.com/iomega/ms2query) and using the negative ESI mode MS2Query library obtained at (https://zenodo.org/record/7104184). The results were then overlaid on the FBMN using the feature IDs as keys.

To further enhance and increase confidence in the putative metabolite annotations, substructure (Mass2Motifs) annotations were carried out using the MS2LDA 2.0 (MS2 Latent Dirichlet allocation version 2.0) (van der Hooft et al., 2016). The revised scalable workflow of MS2LDA 2.0 enables substructure exploration using MS2 information with new innovative features such as the Spec2Vec-powered automated Mass2Motif Annotation Guidance (MAG) system and queryable and matchable MotifDB MotifSets, while simultaneously leveraging Python-based metabolomics frameworks such as matchms to handle the mass spectrometry data (Huber et al., 2021; Ortega et al., 2025). For the ESI negative dataset acquired in this study, MS2LDA 2.0 was implemented using a Python script provided on its official Github page (https://github.com/vdhooftcompmet/MS2LDA/blob/main/README_CONDA.md) and the model for negative ionization mode was downloaded at https://zenodo.org/records/17521464. The MS2LDA settings included setting the number of discovered motifs to 399 with the number of iterations set at 1000. For the downstream analysis, the results were then uploaded, visualized and further analyzed using the MS2LDAViz 2.0 application (https://ms2lda.org/viz/) where only the negative Motifset LDBs, namely LDB_NEG_MotifDB_01, LDB_NEG_MotifDB_02 and Motifset Rhamnaceae Plant MotifDB were selected as reference Motif Data Bases (MotifDBs). Annotated Mass2Motifs that matched with a similarity score ≥ 0.3 to experimental Mass2Motifs were considered relevant to follow-up on to check if their annotations could be transferred. Within Mass2Motifs, mass spectra with both a probability and overlap score ≥ 0.2 were considered associated to the Mass2Motif and accounting to its degree.

Following the compilation of verified putative annotations from GNPS2 and MS2Query, metabolite class annotations were automatically assigned using ClassyFire (http://classyfire.wishartlab.com/), a chemical class ontology and classification tool integrated into MS2Query and GNPS2 where neighbouring nodes were assigned class annotations based on the ClassyFire-classified structures identified through library matching or analogue search. Data visualization was also done using circlize version 0.4.17 (Gu et al., 2014) in R. All metabolite annotations were carried out at level 2 and 3 of the Metabolomics Standards Initiative (MSI) (Sumner et al., 2007; Schymanski et al., 2014), and all validated annotations (manual and computational) are listed in the **Supplementary Table 1**.

3. Results and Discussion

Mass spectral data were acquired in negative ESI mode from methanolic leaf extracts of *P. hadiensis* and *P. ambiguus*, with five biological replicates analyzed per plant species. After data pre-processing, a

comparative metabolomics analysis of the generated spectral data revealed 1421 shared metabolite features out of a total of 3531 between *P. hadiensis* and *P. ambiguus*, indicating a conserved metabolic inventory between the two plant species. Nonetheless, *P. hadiensis* exhibited 1472 unique metabolite features, suggesting a higher metabolite diversity and a higher degree of divergence from the metabolite profile of *P. ambiguus*, which in turn displayed 638 unique features (Fig. 1A). To further explore the degree of divergence in the metabolite profiles of the two species, metabolite diversity was assessed using ecological metrics, namely the Shannon and Simpson diversity indices.

Although these indices are traditionally applied in community ecology studies to quantify biodiversity, they are also useful in measuring chemodiversity across different plant species. In this study, metabolite diversity analysis provides a statistical measure of chemical variability within the metabolomes of the studied plant species. Thus, in this context, using the untargeted metabolomics dataset, individual metabolite features are used as input and treated as unique entries (i.e., "biological species") with their relative abundances corresponding to species abundance (Peters et al., 2021; Manslodo et al.,2022).

Therefore, the Shannon index as an output, is highly influenced by the presence of rare metabolites and reflects both the richness and evenness of metabolite distribution. In contrast, the Simpson index places greater emphasis on the dominance of highly abundant metabolites (Morris et al., 2014; Santini et al., 2017; Chaura et al., 2026). In this study, Fig. 1B**(i)** shows the Shannon index values where *P. ambiguus* exhibited a lower value, suggesting less metabolite diversity defined by the reduced richness and evenness of metabolite distribution. Conversely, *P. hadiensis* showed the highest diversity, indicating a broader and more evenly distributed metabolite profile. Similarly, the Simpson index values as shown by Fig. 1B**(ii)**, further pointed to *P. hadiensis* as having a more diverse metabolome, which could be used in explaining the species' environmental adaptability and bioactivity potential.

The significance of these two diversity indices was evaluated using the Wilcoxon rank-sum tests, where both the Shannon and the Simpson indices yielded a p-value of 4.11 x 10^{-5}, respectively, demonstrating a statistically significant difference in metabolic complexity and diversity between the two species. Following the initial exploratory metabolite diversity analysis, which revealed a heterogeneous chemical complexity between two plant species, the FBMN approach enabled the visualization of the chemical space of the species' leaf extracts by organizing features into molecular families based on MS/MS spectral similarity. FBMN revealed structural relationships among the detected metabolites (shared and unique features), where each metabolite is represented as a node (**Figure S1**). The nodes in the molecular network were annotated using GNPS spectral library matching, complemented with additional *in-silico* annotation tools such as MS2Query and MS2LDA 2.0. Notably, 43% of the total verified annotations were correctly identified as exact matches (*m/z* difference ≤ 0.003) by MS2Query with prediction scores ≥ 0.5. The reliable *in-silico* compound predictions not only expanded the annotations but also increased confidence in the annotations obtained through GNPS spectral library matching and manual annotations.

These structural annotations uncovered a broad spectrum of metabolite classes. These discovered chemical classes (based on the compound classifications enabled by ClassyFire in MS2Query and GNPS2) included cinnamic acids, flavonoids, prenol lipids, benzodioxanes, glycerophospholipids, coumarins, and tannins (Fig. 1C). While several metabolite features were common to both species, distinct clusters of nodes were uniquely associated with either *P. hadiensis* or *P. ambiguus*, suggesting species-specific chemical signatures as illustrated earlier by Fig. 1A. Although a substantial number (1421 metabolite features) of metabolites are shared between the two species, their species-specific chemical signatures, observed in some of the molecular families, reflect underlying metabolic variation and may explain, in part, their differing traditional uses and pharmacological profiles. However, network-based visualization alone does not quantify the statistical significance or relative contribution of individual features to species differentiation. To further interrogate and validate the observed species-specific chemical signatures, comparative metabolomics analysis with unsupervised and supervised chemometric models was conducted. This chemometric approach complements FBMN by providing a quantitative assessment of the observed differential metabolic profiles across the two species, facilitating the identification of key discriminating features that define the species-specific metabolic differences.

3.1 Chemometric analysis identifies flavonoids as key drivers of species-specific chemical differentiation

Both unsupervised and supervised statistical approaches, PCA and OPLS-DA were computed as described in section 2.4. The PCA scores plot of the Pareto-scaled dataset (**Figure S2**) showed distinct sample grouping where *P. ambiguus* samples were observed to group away from *P. hadiesnsis* samples, highlighting the underlying metabolite differences between the two plants. This observation was further explored by applying a supervised approach to determine the discriminating metabolite features between the two species. The OPLS-DA method enables the identification of metabolite features that are significantly contributing to the separation of two well-defined sample groups (Trygg et al., 2007; Tugizimana et al., 2013). The generated OPLS-DA model, as a binary classifier, separating multivariate relationships into predictive (species differentiation) and orthogonal (unrelated to species differentiation) variation (Fig. 2A), was validated by assessing the robustness, predictive ability, reliability and significance of the models. The computed and validated OPLS-DA model ($p \leq 0.05$) (Fig. 2A) was a perfect classifier and statistically reliable, with very good predictive capability: no signs of possible overfitting, as indicated by cross-validation; and none of the permutated models (**Figure S2B**) performed better than the original model in separating classes, *P. hadiensis* and *P. ambiguus*.

Some of these validations included a 7-fold cross-validation, with R^2 and Q^2 metrics, the analysis of variance testing of cross-validated predictive residuals (CV-ANOVA, *p*-value ≤ 0.05 as a cut-off), the receiver operator characteristic (ROC) curves and response permutation tests (**Figure S2B-C**) (Eriksson et al., 2014; Tugizimana et al., 2016; Westerhuis et al., 2008). Furthermore, the selection of discriminating metabolite features (*P. hadiensis* vs. *P. ambiguus*) was carried out by evaluating the OPLS-DA loading S-

plot (Fig. 2B-D). To avoid overinterpretation of the model and reduce variable selection bias, only features that were statistically significant in contributing to species separation were retained. Hence, variables that combined both high covariation and correlation (as examined on S-plot) were statistically relevant as potential discriminant metabolite features (Tugizimana et al., 2013; Wiklund et al., 2008). However, it is important to note that the S-plot is susceptible to data matrix changes due to correlation sensitivity and dependency on data structure.

Therefore, in this study, additional tests and tools such as the variable importance in projection (VIP) plots, jackknife confidence intervals (used to estimate standard errors in a nonparametric way as an estimate of bias) (**Figure S2D**) were used to further confirm the statistical significance and discriminability of the S-plot-derived potential discriminant markers (Galindo-Prieto et al., 2015; Gika et al., 2014; Tugizimana et al., 2016). Furthermore, from the descriptive statistics, some of the top relevant markers discriminating *P. ambiguus* from *P. hadiensis* are highlighted by Fig. 2C and include *m/z* 313.0706, *m/z* 447.0871, *m/z* 271.0602, *m/z* 625.3220, *m/z* 299.0551 and *m/z* 304.9130. On the other hand, *m/z* 389.1958, *m/z* 589.1187, *m/z* 461.0715, *m/z* 351.1804 and *m/z* 445.0766 (Fig. 2D) were the most relevant markers responsible for discriminating *P. hadiensis* from *P. ambiguus*. The statistically significant markers for both plant species were observed to belong to the prenol lipid and flavonoid chemical classes with some features (351.1804 and *m/z* 389.1958) mapped on MFs that had no chemical class identity on the FBMN (Fig. 2E). Since the two species are predominantly known for their prenol lipid (*e.g.*, diterpenoids) content, the flavonoid chemical class was further investigated for its contribution to the species-specific patterns observed in the chemical landscape of *P. ambiguus* and *P.hadiensis*.

3.2 Flavonoid structural diversification shapes species-specific chemical identity

The combination of chemometric models and mass spectral annotation tools (namely FBMN, MS2Query, and MS2LDA 2.0) pointed to flavonoids as one of the major chemical classes responsible for the metabolite diversity as depicted by the differences in the distinct metabolite profiles of *P. ambiguus* and *P. hadiesis* (Fig. 2). The discriminating features were further evaluated by considering VIP scores, where only metabolite features with VIP scores > 2 were retained as discriminating markers driving the metabolite differences. The criterion for the VIP value was set at > 2 (instead of the commonly used criterion of ≥ 1) to reduce and narrow down a smaller subset of metabolites that are the most relevant features in discriminating the two species. For instance, from the metabolite features that are positively correlated to *P. hadiensis*, *m/z* 445.0766 on the S-plot with a high VIP score of 7.45, was putatively identified through FBMN as baicalin. After baicalin, sorbifolin (*m/z* 299.0551) was observed to have the second highest VIP score of 6.62 among the metabolite features that are positively correlated to *P. ambiguus*, indicating the flavonoid content as a key discriminatory chemical class between *P. hadiesnis* and *P. ambiguus*.

When evaluating the baicalin molecular family (MF), we observed that additional discriminant markers (VIP score > 2) highly abundant in *P. hadiensis* were also present in the same MF. These markers included luteolin-3-O-glucuronide, unknown *m/z* 605.1135 and *m/z* 589.1187 (Fig. 3A). To further investigate the unknown discriminant markers in the baicalin MF, MS2LDA 2.0 was applied to assess the structural architecture and relationship between the known and unknown metabolites in the cluster. In the MS2LDA analysis, a total of 399 Mass2Motifs were detected across the dataset, with 18 Mass2Motifs detected in the baicalin MF (**Figure S3A**). From the 18 main motifs in the MF, motif_14 and motif_187 were found in more than one node (**Figure S3A**). When using MS2LDA and assessing the fragmentation spectra of *m/z* 605.1135, the structure of which remains unidentified, it was observed that *m/z* 605.1135 shared a common fragment (motif_14 = 285.03 Da) with luteolin-3-O-glucuronide and luteolin-3-O-hexoside (VIP score < 2) (Fig. 3A**(i)**).

Notably, in all three connected and structurally related parent ions, motif_14 (285.03 Da) as a shared substructure was found to be associated with luteolin. This was supported by the presence of a luteolin diagnostic fragment of 285.03 Da in the fragmentation pattern of *m/z* 605.1135, luteolin-3-O-glucuronide, and luteolin-3-O-hexoside, respectively. In luteolin-3-O-glucuronide, motif_14 further fragments to produce low-intensity luteolin diagnostic fragments of *m/z* 175.03 and *m/z* 133.02 (Fabre and Rustan, 2001), reducing the probability of motif_14 being any other isomeric flavonoid aglycone; however, further validation experiments are required to fully verify the aglycone identity. Moreover, a glucuronide sugar moiety (neutral loss of 176 Da) specifically indicated that *m/z* 605.1135 is an analogue of luteolin-3-O-glucuronide (Fig. 3A**(i)**). The neighbouring node, unknown *m/z* 589.1187, was shown to be comprised of motif_187 (445 Da). Leveraging on the new Mass2Motif Annotation Guidance (MAG) in MS2LDA 2.0 that uses Spec2Vec (Huber et al., 2021) to find potential matches for parent ions in a large mass spectral library, a metabolite feature whose *m/z* value equals that of motif_187 was recommended to be apigenin-7-O-glucuronide, an annotation that was a spectral hit in GNPS2 and MS2query, respectively. Within the dataset, motif_187 was found to be common in 68 different metabolite features eluting at different times. The pseudospectra (optimized fragmentation peaks) of motif_187 reflected apigenin structural similarities as confirmed by the apigenin diagnostic fragments of *m/z* 269.04 and *m/z* 117.02 with glycosylation shown by the glucuronide sugar moiety (Fig. 3A**(ii)**).

Thus *m/z* 589.1187 was postulated to be a glycosylated analogue of apigenin-7-O-glucuronide with an additional rhamnosyl sugar moiety. The presence and abundance of these glycosylated flavonoids in *P. hadiensis* may have functional significance in the context of medicinal plant applications. For instance, glycosylation plays a vital role in the medicinal properties of flavonoids as it increases the reactivity, solubility and stability of the corresponding bioactive aglycones, thus improving bioavailability and bioactivity of flavonoids (Alseekh et al., 2020). Notwithstanding the predominance of glycosylated flavones in *P. hadiensis*, simple hydroxylated flavone aglycones, namely apigenin (*m/z* 269.0445) and luteolin (*m/z* 285.0393), were also identified as highly abundant discriminating features in *P. hadiensis* (**Figure S4**). Conversely, methylated flavones were observed to be highly abundant discriminating features in *P. ambiguus* and included a molecular family containing sorbifolin (VIP score > 2) (Fig. 3B).

Amongst the 16 Mass2Motifs detected in the sorbifolin MF, motif_4 was the most common substructure (**Figure S3B**). The Spec2Vec matching of MAG suggested the chemical annotation of motif_4 to be 5,7-dihydroxy-3-(4-hydroxyphenyl)-6-methoxychromen-4-one, an isomer of sorbifolin. The probability and overlap score for motif_4 detected in the mass spectra of sorbifolin (*m/z* 299.0551) was observed to be 0.326 and 0.559 respectively. Guided by this and the generated pseudospectra of motif_4, comprised of fragments 284.03 Da, 283.04 Da, 255 Da, 227.03 Da, 212.03 Da, 186.03 Da, 136.99 Da and 117.03Da, we annotated motif_4 as being related to a sorbifolin (sub)structure. This was further supported by the pseudospectra of motif_4, which corresponded and overlapped with the fragmentation pattern of sorbifolin as deposited in the PubChem database (CID:3084390). Therefore, this further increased the confidence in the GNPS-generated annotation of the neighbouring node (*m/z* 315.0499) as 6-methoxyluteolin, since both metabolites are structural analogues of each other with varying hydroxylation and methylation patterns. Although methylated flavonoids generally occur in lower levels than their glycosylated counterparts in plants, it has been suggested that differences in polarity or the accumulation of methylated hydrophobic flavones in medicinal plants, as seen in *P. ambiguus*, could point to specific biological roles (Berim and Gang, 2016). Interestingly, when assessing the glycosylated and methylated discriminant flavonoids in Fig. 3A-B, it was observed that they share a flavone backbone (C6-C3-C6 form) with similar hydroxylation modifications (Fig. 4).

However, the overall chemical architecture of the discriminant flavonoids (VIP score > 2) is distinguished by glycosylation and methylation modifications, respectively, a trend that also carries over to the neighbouring nodes in each MF (Fig. 4A and Fig. 4B). These flavone structural differences are reflected at the species level by the prevalence of flavone glycosylation in *P. hadiensis* (Fig. 4A), whereas flavone methylation was more prominent in *P. ambiguus* (Fig. 4B-C). Therefore, both glycosylation and methylation modifications can be postulated as possible chemical signatures that differentiate the leaf tissues of the two plant species.

Upon further investigations into the flavonoid chemical class, it was observed that discriminant flavonols were also one of the flavonoid subclasses, differentiating the leaf metabolite profiles of the two species. Notably, flavonols (Fig. 5A) and flavanones (Fig. 5B) were observed to be predominantly present in *P. ambiguus* than in *P. hadiensis*. The core structure of flavonols differs from that of the flavones due to the presence of a hydroxyl group on the C-ring at the C-3 position (Todorova et al., 2013; Stachelska et al., 2025; Przybylski et al., 2025). As mentioned earlier, the chemical features and biological activities of flavonoids are influenced by the nature of the substituents attached to the basic structures. Free hydroxyl groups are more reactive than glycosylated and methylated hydroxyl groups, and this is vital for various flavonoid biological activities, such as antioxidant activities.

Thus, structure and activity studies have shown that the presence of the hydroxyl group on the C-ring of flavonols, which often undergoes modifications, is one of the substituents that is frequently associated with the antioxidant activities of flavonoids (Halbwirth, 2010; Verdan et al., 2011). al., 2013; Stachelska et al., 2025; Przybylski et al., 2025). Consequently, the C-ring hydroxyl group of flavanols can undergo glycosylation where a sugar moiety is attached to C-3. In this study, glycosylation on the two flavanol

aglycones resulted in the formation of two quercetin-3-O-hexoside metabolites (potentially a glucoside and a galactoside) and kaempferol-3-O-hexoside, exact metabolite matches as identified through both FBMN and MS2query (Fig. 5A).

Both annotations were further confirmed by the Mass2Motifs discovered through MS2LDA 2.0 as shown by the duplicate MF highlighting all discovered Mass2Motifs (Fig. 6B). From the 14 substructures discovered in the MF, one of the Mass2Motifs in kaempferol-3-O-hexoside, was motif_390, initially detected as an unknown (Fig. 6B). This Mass2Motif (motif_390), was later identified as a kaempferol-related substructure based on its pseudospectra, which included kaempferol diagnostic fragments such as 284.03 Da, 225.02 Da and 151.0 Da (Fig. 6A**(i)**) (Fabre et al., 2001).

The metabolite features annotated with quercetin-3-O-hexoside were associated with motif_295, and the MAG Spec2Vec matching functionality returned the quercetin-3-O-hexoside annotation as a Mass2Motif annotation of motif_295 (Fig. 6A**(ii)**), which increased confidence in the GNPS spectral library and MS2Query-generated annotations. As discussed, glycosylation increases the hydrophilicity of flavonoid aglycones, minimizes toxic side effects, and enhances the specific targeting ability of flavonoids (Hu et al., 2025). Amongst the glycosyl moieties, an additional attachment is seen in the form of acylation (highlighted in blue on Fig. 5A) and is illustrated by the putative annotation of quercetin-3-O-malonylhexoside and quercetin-3-(6”-acetylhexoside) (Fig. 5A). The acylation of the sugar moieties is crucial for enhancing molecular stability, improving water solubility, and increasing resistance to enzymatic degradation of natural products (Urakawa et al., 2024). Various plant species carry out flavonoid malonylation, a type of acylation reaction, in which malonyltransferases transfer a malonyl group from malonyl-CoA with regioselectivity to a hydroxyl group of a flavonoid sugar residue, a reaction common in anthocyanins and flavanols (Kogawo et al., 2007; Kachlicki et al., 2008). Additionally, acetyltransferases facilitate the transfer of acetyl groups from various donors to flavonoid glycosyl moieties (Wang et al., 2022). In addition to influencing the physico-chemical properties of flavonoids, acetylation of sugar residues has been identified as contributing to the cytoprotective effects and anti-inflammatory effects in *in vivo* studies (Urakawa et al., 2024).

In addition to the predominance of modified flavanols, it was observed that within the flavanone subclass, naringenin (VIP score = 3.89) was also one of the significant discriminating features in *P. ambiguus* plants (Fig. 5B and **Figure S5A(i)**). In the naringenin MF, unknown metabolite features (*m/z* 303.0863 and *m/z* 331.0932) were observed to be modified derivatives of naringenin, based on their mass spectral fragmentation patterns that suggested methylation (**Figure S5A(ii)** and **Figure S5A(iii)**). The duplicate MF highlighted 9 Mass2Motifs (**Figure S5B**) where the parent ion masses, namely *m/z* 331.0932 (**Figure S5A (ii)**) and *m/z* 303.0863 (**Figure S5A(iii)**) both were shown to have a naringenin-related substructure (motif_258), confirmed by the presence of the naringenin diagnostic fragment of 271 Da and 119 Da. Both parent ions showed an initial neutral loss of 32 Da, suggesting the loss of a methoxy group.

This observation was further supported by the MS2query analogue annotation of *m/z* 331.0932 as naringenin 5,7-dimethyl ether (precursor analogue *m/z* 299 in negative ESI), suggesting possible methylation on the unknown metabolite features. Naringenin is well known for its biological activities, which range from antioxidant to anti-diabetic properties (Shilpa et al., 2023; Pansare et al., 2025). However, naringenin as an aglycone is known to have poor water solubility, which affects its overall bioavailability and bioactivity (Arafah et al., 2020; Uçar and Göktaş, 2023). The presence of methylated derivatives (*m/z* 303.0863 and 331.0932) of naringenin could potentially enhance the bioavailability of naringenin and thus positively benefit the overall medicinal properties of *P. ambiguus*

The elucidation and comparative study of the metabolite profiles of *P. hadiensis* and *P. ambiguus* has highlighted flavonoids and their respective chemical modifications as species-specific chemical signatures that differentiate the two medicinal plant species. This type of chemo-characterization, driven by the integration of computational metabolomics tools, holds potential in understanding the environmental adaptation strategies and ethnopharmacological uses of these two species. Considering the lack of comprehensive data on the metabolite composition *P. ambiguus* and *P. hadiensis* metabolomes, the information generated in this study lays the foundation for generating insights into finding key metabolites that contribute to each species' unique medicinal properties beyond their well-known diterpenoid content.

4. Conclusion

An in-depth exploration of the metabolite composition of medicinally important *P. hadiensis* and *P. ambiguus* plants was achieved using computational metabolomics strategies. Metabolite annotation facilitated by the integration of mass spectral annotation tools such as feature-based molecular networking, MS2Query, and MS2LDA 2.0 revealed a diverse range of chemical classes that span the metabolic atlas of both plant species. The combination of comparative chemometric models and mass spectral annotation tools pointed to the flavonoid chemical class as one of the key chemical classes that may underpin the species-specific metabolite profiles of *P. hadiensis* and *P. ambiguus* leaf extracts. The findings from this study have provided new insights into the contribution of the differential flavonoid chemical architecture in the chemotaxonomic classification of *Plectranthus* species, where *P. hadiensis* was observed to be rich in glycosylated flavones, while *P. ambiguus* was characterized by the abundance of methylated flavones and glycosylated flavanols. Such observation brings to the forefront the need to explore the ecological relevance of metabolite diversification and its impact on the unique ethnomedicinal applications and biological activities elicited by each species. However, it is important to note that these observations may be influenced by plant-growing environmental conditions and/or unique metabolic biosynthetic pathways. Nonetheless, the current study highlights the potential of *P. hadiensis* and *P. ambiguus* plants as treasure troves of natural products with therapeutic relevance, promoting further exploration of the two plant species under various environmental and experimental conditions.

Declarations

Data and code availability section

The mzmine 4.8.30 batch file, including all data preprocessing settings, together with the MS2LDA analysis config file specifying the analysis settings and the results file (.json) are all made available at Zenodo (Doi: 10.5281/zenodo.20112661). The spectral data presented in this study is made available on the supplementary files as a GNPS job link. The molecular networking files (feature table and mgf files) and raw (mzML) files can be found in MassIVE (MSV000101781).

Acknowledgements

University of Venda, Biochemistry Department are gratefully thanked for access to the SHIMADZU LCMS-9030. The South African National Research Fund (NRF) is highly thanked for the bursary support to A.M.

Author contributions

A.M. performed experiments, data analysis, data curation, wrote the original draft. M.C. supervised the study, performed data curation and analysis. N.E.M. supervised the LCMS analyses, reviewed and edited the manuscript text. F.A. supervision, reviewed and edited the manuscript text. J.J.J.vd.H. supervised the data analysis and curation, reviewed and edited the manuscript text. F.T. conceptualized the study, supervised the study, data analysis and curation, reviewed and edited the manuscript text, and attended to the project administration. All authors read and approved the manuscript.

Funding statement – N/A

Competing Interests

J.J.J.vdH is member of the Scientific Advisory Board of NAICONS Srl., Milano, Italy and consults for Corteva Agriscience, Indianapolis, IN, USA.

References

1. Abdel-Mogib, M., Albar, H. A., and Batterjee, S. M. (2002). Chemistry of the Genus *Plectranthus*. *Molecules 7*(2), 271-301. https://doi.org/10.3390/70200271
2. Alseekh, S., de Souza, L.P., Benina, M., and Fernie, A.R. (2020). The style and substance of plant flavonoid decoration; towards defining both structure and function. Phytochemistry, 174, 1-15. https://doi.org/10.1016/j.phytochem.2020.112347
3. Arumugam, G., Swamy, M. K., and Sinniah, U. R. (2016). *Plectranthus amboinicus* (Lour.) Spreng: Botanical, Phytochemical, Pharmacological and Nutritional Significance. *Molecules 21*(4), 1-26. https://doi.org/10.3390/molecules21040369
4. Arafah, A., Rehman, M.U., Mir, T.M., Wali, A.F, Ali, R., Qamar, W., Khan, R., Ahmad, A., Aga, S.S., Alqahtani, S. and Almatroudi, N.M. (2020). Multi-Therapeutic Potential of Naringenin (4',5,7-

Trihydroxyflavonone): Experimental Evidence and Mechanisms. *Plants (Basel)*, 9(*12*), 1-22. https://doi.org/10.3390/plants9121784

5. Barbosa, M. de O., Wilairatana, P., Leite, G. M. de L., Delmondes, G. de A., Silva, L. Y. S. da, Júnior, S. C. A., Dantas, L. B. R., Bezerra, D. S., Beltrão, I. C. S. L. de, Dias, D. de Q., Ribeiro-Filho, J., Felipe, C. F. B., Coutinho, H. D. M., de Menezes, I. R. A., and Kerntopf Mendonça, M. R. (2023). *Plectranthus* Species with Anti-Inflammatory and Analgesic Potential: A Systematic Review on Ethnobotanical and Pharmacological Findings. *Molecules 28*(15), 1-21. https://doi.org/10.3390/molecules28155653
6. Berim, A., and Gang, D.R. (2016). Methylated flavones: occurrence, importance and biosynthesis. *Phytochemistry reviews*, *15*, 363-390. https://doi.org/10.1007/s11101-015-9426-0
7. Chaachouay, N. and Zidane, L. (2024). Plant-Derived Natural Products: A Source for Drug Discovery and Development. *Drugs and Drug candidates*, *3*(1), 184-207. https://www.mdpi.com/2813-2998/3/1/11
8. de Jonge, N. F., Louwen, J. J. R., Chekmeneva, E., Camuzeaux, S., Vermeir, F. J., Jansen, R. S., Huber, F., and van der Hooft, J. J. J. (2023). MS2Query: reliable and scalable MS2 mass spectra-based analogue search. *Nature Communications*, *14*(1). https://doi.org/10.1038/s41467-023-37446-4
9. Eriksson, L., Trygg, J., and Wold, S. (2014). A chemometrics toolbox based on projections and latent variables. *Journal of Chemometrics*, *28*(5), 332–346. https://doi.org/10.1002/cem.2581
10. Fabre, N., Rustan, I., de Hoffmann, E. and Quetin-Leclercq, J. (2001). Determination of flavone, flavonol, and flavanone aglycones by negative ion liquid chromatography electrospray ion trap mass spectrometry. *Journal of the American Society for Mass Spectrometry*, 12, 707–715. https://doi.org/10.1016/S1044-0305(01)00226-4
11. Kachlicki, P., Einhorn, J., Muth, D., Kerhoas, L., and Stobiecki, M. (2008). Evaluation of glycosylation and malonylation patterns in flavonoid glycosides during LC/MS/MS metabolite profiling. *Journal of Mass Spectrometry*, *43*(5), 572–586. https://doi.org/10.1002/jms.1344
12. Kachlicki, P., Piasecka, A., Stobiecki, M., and Marczak, L. (2016). Structural Characterization of Flavonoids Glycoconjucates and Their Derivatives with Mass Spectrometric Techniques. Molecules, *21*(14945), 1-21. https://doi.org/10.3390/molecules21111494
13. Kogawa, K., Kazuma, K., Kato, N., Noda, N., and Suzuki, M. (2007). Biosynthesis of malonylated flavonoid glycosides on the basis of malonyltransferase activity in the petals of *Cliroria ternatea*, *164*, 886-894. https://doi.org/10.1016/j.jplph.2006.05.006
14. Koirala, N., Thuan, N.H., Ghimire, G.P., van Thang, D., and Sohng, J.K. (2016). Methylation of flavonoids: Chemical structures, bioactivities, progress and perspectives for biotechnological production. *Enzyme and Microbial Technology*, *86*, 103-116. https://doi.org/10.1016/j.enzmictec.2016.02.003
15. Galindo-Prieto, B., Eriksson, L., and Trygg, J. (2015). Variable influence on projection (VIP) for OPLS models and its applicability in multivariate time series analysis. *Chemometrics and Intelligent Laboratory Systems*, *146*, 297–304. https://doi.org/10.1016/j.chemolab.2015.05.001

16. Gika, H. G., Theodoridis, G. A., Plumb, R. S., and Wilson, I. D. (2014). Current practice of liquid chromatography-mass spectrometry in metabolomics and metabonomics. *Journal of Pharmaceutical and Biomedical Analysis*, *87*, 12–25. https://doi.org/10.1016/j.jpba.2013.06.032
17. Gu, Z., Gu, L., Eils, R., Schlesner, M. and Brors, B. (2014). Circlize implements and enhances circular visualization in R. *Bioinformatics*, *30*, 1-2. https://doi.org/10.1093/bioinformatics/btu393
18. Guo, Z. (2017). The modification of natural products for medical use. *Acta Pharmaceutica Sinica B*, *7*(2), 119–136. https://doi.org/10.1016/j.apsb.2016.06.003
19. Halbwirth, H. (2010). The creation and physiological relevance of divergent hydroxylation patterns in the flavonoid pathway. *International Journal of Molecular Sciences*, *11*(2), 595–621. https://doi.org/10.3390/ijms11020595
20. Hao, D. C., He, C. N., Spjut, R. W., and Xiao, P. G. (2022). Editorial: Plant-derived natural compounds in drug discovery: The prism perspective between plant phylogeny, chemical composition, and medicinal efficacy. *Frontiers in Plant Science*, *13*, 1-5. Frontiers Media S.A. https://doi.org/10.3389/fpls.2022.1042695
21. Heuckeroth, S., Damiani, T., Smirnov, A., Mokshyna, O., Brungs, C., Korf, A., Smith, J.D., Stincone, P., Dreolin, N., Nothias, L.F., Hyötyläinen, T., Orešič, M., Karst, U., Dorrestein, P.C., Petras, D., Du, X., van der Hooft J.J.J., Schmid, R. and Pluskal, T. (2024). Reproducible mass spectrometry data processing and compound annotation in MZmine 3. *Nature Protocols*, *19*, 2597–2641. https://doi.org/10.1038/s41596-024-00996-y
22. Hu, L., Luo, Y., Yang, J., and Cheng, C. (2025). Botanical Flavonoids: Efficacy, Absorption, Metabolism and Advanced Pharmaceutical Technology for Improving Bioavailability. *Molecules*, *30*(5), 1-35. https://doi.org/10.3390/molecules30051184
23. Huber, F., Ridder, L., Verhoeven, S., Spaaks, J.H., Diblen, F., Rogers, S. and van der Hooft, J.J.J. (2021). Spec2Vec: Improved mass spectral similarity scoring through learning of structural relationships, *PLos Computational Biology*, 17(2), 1-18. https://doi.org/10.1371/journal.pcbi.1008724
24. Jiang, N., Doseff, A.I., and Grotewold, E. (2016). Flavones: From Biosynthesis to Health Benefits. Plants, *5*(27), 1-25. https://doi.org/10.3390/plants5020027
25. Lambrechts, I. A., and Lall, N. (2021). Traditional usage and biological activity of *Plectranthus madagascariensis* and its varieties: A review. *Journal of Ethnopharmacology*, *269*, 1-13. https://doi.org/10.1016/j.jep.2020.113663
26. Leonte D, Ungureanu D, and Zaharia V. (2023). Flavones and Related Compounds: Synthesis and Biological Activity. *Molecules*, *28*(18), 1-52. https://doi.org/10.3390/molecules28186528
27. Marquesl, A. M., Cristina, M., Limas, H. P., Foe, H. A. C., Passinho3, H. S., Auxiliadora, M., and Kaplanl, C. (2012). Chemical profile analysis of three different species of *Plectranthus*. *Emirates Journal of Food and Agriculture*, *24*(2), 1-5. http://ejfa.info/PLANTSCIENCE
28. Newman, D. J., and Cragg, G. M. (2020). Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *Journal of Natural Products*, *83*(3), 770–803. https://doi.org/10.1021/acs.jnatprod.9b01285

29. Nicolai M, Mota J, Fernandes AS, Pereira F, Pereira P, P Reis C, Robles Velasco MV, Baby AR., Rosado C and Rijo P. (2020) Assessment of the Potential Skin Application of *Plectranthus ecklonii* Benth. *Pharmaceuticals (Basel)*, *13*(6), 1-10. https://doi.org/10.3390/ph13060120
30. Ortega, L.R.T., Dietrich, J., Wandy, J., Mol, H., and van der Hooft, J.J.J. (2025). Large-scale discovery and annotation of hidden substructure patterns in mass spectrometry profiles. bioRxiv, 1-38. https://www.biorxiv.org/content/10.1101/2025.06.19.659491v1
31. Pansare, K., Ahire, Y. and Bairagi, V. (2025). Naringenin: A Potential Flavonoid Phytochemical For Diabetes Management. *Biosciences Biotechnology Research Asia*, 22(*3*), 1-11. http://dx.doi.org/10.13005/bbra/3411
32. Przybylski, T., Czerniel, J., Dobrosielski, J., and Stawny, M. (2025). Flavonol Technology: From the Compounds' Chemistry to Clinical Research. Molecules, *30*, 1-36. https://doi.org/10.3390/molecules30153113
33. Rice, L. J., Brits, G. J., Potgieter, C. J., and Van Staden, J. (2011). *Plectranthus*: A plant for the future? *South African Journal of Botany*, *77*(4), 947–959. https://doi.org/10.1016/j.sajb.2011.07.001
34. Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J. (2014). Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environmental Science Technology*, 48(*4*), 1-2. https://doi.org/10.1021/es5002105
35. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–2504. https://doi.org/10.1101/gr.1239303
36. Shilpa, V.S., Shams, R., Dash, K.K., Pandey, V.K., Dar, A.H., Ayaz Mukarram, S., Harsányi, E. and Kovács B. (2023). Phytochemical Properties, Extraction, and Pharmacological Benefits of Naringin: A Review. *Molecules*, 28(*15*), 1-18. https://doi.org/10.3390/molecules28155623
37. Sysak, S., Wicher, B., Kucinska, M., Kobylka, P., Mlynarczyk, D.T., Lesyl, R., Tykarska, E., Murias, M., Goslinski, T. and Szczolko, W. (2025). Synthesis, physicochemical characterization and biological activity of novel pyrrole flavones. *Scientific reports*, *15*, 1-16. https://doi.org/10.1038/s41598-025-91772-9
38. Śliwiński, T., Sitarek, P., Skała, E., Isca, V. M. S., Synowiec, E., Kowalczyk, T., Bijak, M., and Rijo, P. (2020). Diterpenoids from *Plectranthus* spp. As potential chemotherapeutic agents via apoptosis. *Pharmaceuticals*, *13*(6), 1–16. https://doi.org/10.3390/ph13060123
39. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L., and Ideker, T. (2011). Cytoscape 2.8: New features for data integration and network visualization. *Bioinformatics*, *27*(3), 431–432. https://doi.org/10.1093/bioinformatics/btq675
40. Stachelska, M.A., Karpinski, P., and Kruszewski, B. (2025). A Comprehensive Review of Biological Properties of Flavonoids and Their Role in the Prevention of Metabolic, Cancer and Neurodegenerative Diseases. *Applied sciences*, *15*, 1-40. https://doi.org/10.3390/app151910840
41. Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., Fan, T. W. M., Fiehn, O., Goodacre, R., Griffin, J. L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A. N.,

Lindon, J. C., Marriott, P., Nicholls, A. W., … Viant, M. R. (2007). Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*, *3*(3), 211–221. https://doi.org/10.1007/s11306-007-0082-2

42. Todorova, T.Z., Traykov, M.G., Tadjer, A.V., and Velkov, Zh.A. (2013). Structure of flavones and flavonols. Part I: Role of substituents on the planarity of the system. Computational and Theoretical Chemistry, *1017*, 85-90. https://doi.org/10.1016/j.comptc.2013.05.005
43. Trygg, J., Holmes, E., and Lundstedt, T. (2007). Chemometrics in metabonomics. *Journal of Proteome Research*, 6(2), 469–479. https://doi.org/10.1021/pr060594q
44. Tugizimana, F., Piater, L., and Dubery, I. (2013). Plant metabolomics: A new frontier in phytochemical analysis. *South African Journal of Science*, *109*(5–6), 1-11. https://doi.org/10.1590/sajs.2013/20120005
45. Tugizimana, F., Steenkamp, P. A., Piater, L. A., and Dubery, I. A. (2016). A conversation on data mining strategies in LC-MS untargeted metabolomics: Pre-processing and pre-treatment steps. *Metabolites*, *6*(4), 1-18. https://doi.org/10.3390/metabo6040040
46. Uçar, K. and Göktaş, Z. (2023). Biological activities of naringenin: A narrative review based on in vitro and in vivo studies. *Nutrition Research*, 119,43-55. https://doi.org/10.1016/j.nutres.2023.08.006
47. Urakawa, D., Shioiridani, Y., Igata, S., Hou, D. X., and Sakao, K. (2024). Comparative Analysis of Acetylated Flavonoids' Chemopreventive Effects in Different Cancer Cell Lines. *International Journal of Molecular Sciences*, *25*(14), 1-28. https://doi.org/10.3390/ijms25147689
48. Verdan, A. M., Wang, H. C., García, C. R., Henry, W. P., and Brumaghim, J. L. (2011). Iron binding of 3-hydroxychromone, 5-hydroxychromone, and sulfonated morin: Implications for the antioxidant activity of flavonols with competing metal binding sites. *Journal of Inorganic Biochemistry*, *105*(10), 1314–1322. https://doi.org/10.1016/j.jinorgbio.2011.07.006 b
49. Wang, L., Chen, K., Zhang, M., Ye, M., and Qiao, X. (2022). Catalytic function, mechanism, and application of plant acyltransferases. *Critical Reviews in Biotechnology*, *42*(1), 125–144. https://doi.org/10.1080/07388551.2021.1931015b
50. Wen, L., Jiang, Y., Yang, J., Zhao, Y., Tian, M., and Yang, B. (2017). Structure, bioactivity, and synthesis of methylated flavonoids. Annals of the New York Academy of Science, *1398*, 1-10. https://doi.org/10.1111/nyas.13350
51. Westerhuis, J. A., Hoefsloot, H. C. J., Smit, S., Vis, D. J., Smilde, A. K., Velzen, E. J. J., Duijnhoven, J. P. M., and Dorsten, F. A. (2008). Assessment of PLSDA cross validation. *Metabolomics*, *4*(1), 81–89. https://doi.org/10.1007/s11306-007-0099-6
52. Wiklund, S., Johansson, E., Sjöström, L., Mellerowicz, E. J., Edlund, U., Shockcor, J. P., Gottfries, J., Moritz, T., and Trygg, J. (2008). Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. *Analytical Chemistry*, *80*(1), 115–122. https://doi.org/10.1021/ac0713510

53. Xie, L., Deng, Z., Zhang, J., Dong, H., Wasng, W., Xing, B., and Liu, X. (2022). Comparison of Flavonoid O-Glycoside, C-Glycoside and Their Aglycones on Antioxidant Capacity and Metabolism during In Vitro Digestion and In Vivo. *Foods*, *11*(882), 1-17. https://doi.org/10.3390/foods11060882

Figures

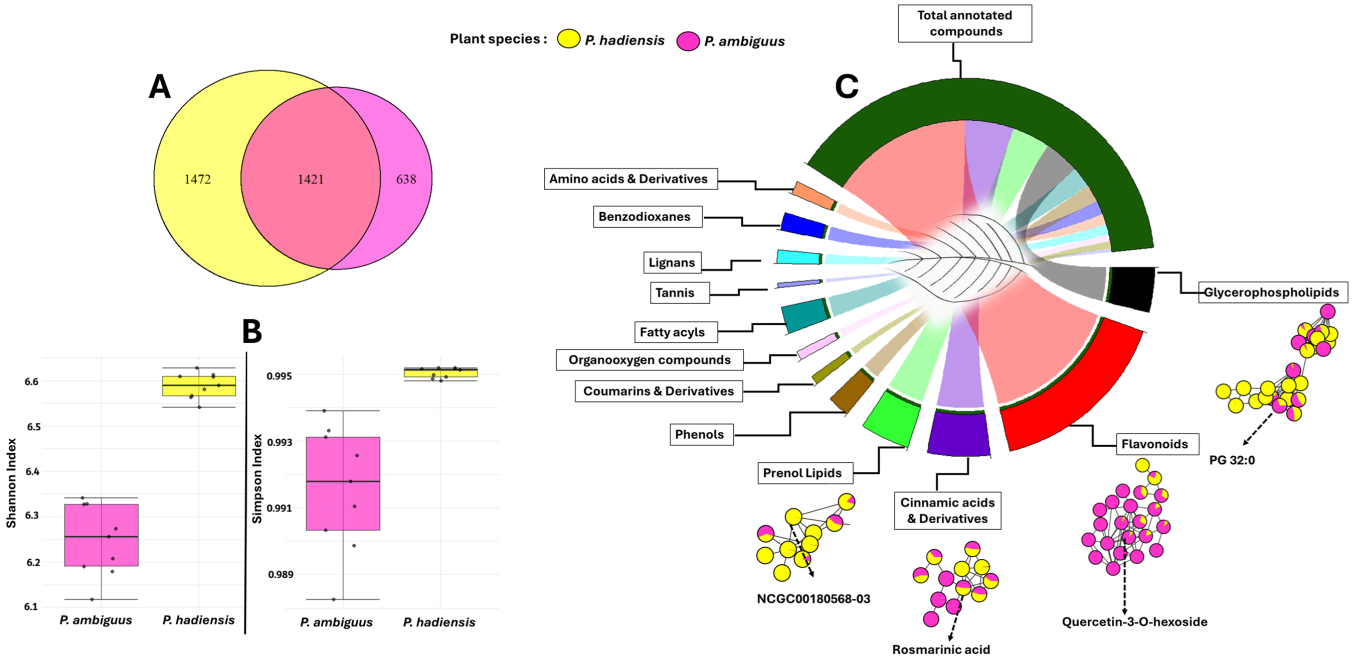


Figure 1

Chemical complexity and diversity of metabolite profiles in *P. hadiensis* and *P. ambiguus*. (**A**) Overview of the detected metabolite features, highlighting shared and unique features between both species. (**B**) Diversity indices reflecting metabolite diversity: (**i**) Shannon index and (**ii**) Simpson index. (**C**) Total number of annotations belonging to the **glycerophospholipids**, **cinnamic acids**, **prenol lipids**and **flavonoids** chemical classes as highlighted by each molecular family in the FBMN. The highlighted molecular families are comprised of metabolites shown as interconnected nodes. Some of the metabolites in each molecular family were identified through GNPS spectral library matching MS2query, MS2LDA 2.0 and manual annotation. The pie charts represent the relative distribution of each metabolite across *P. hadiensis* and *P. ambiguus*

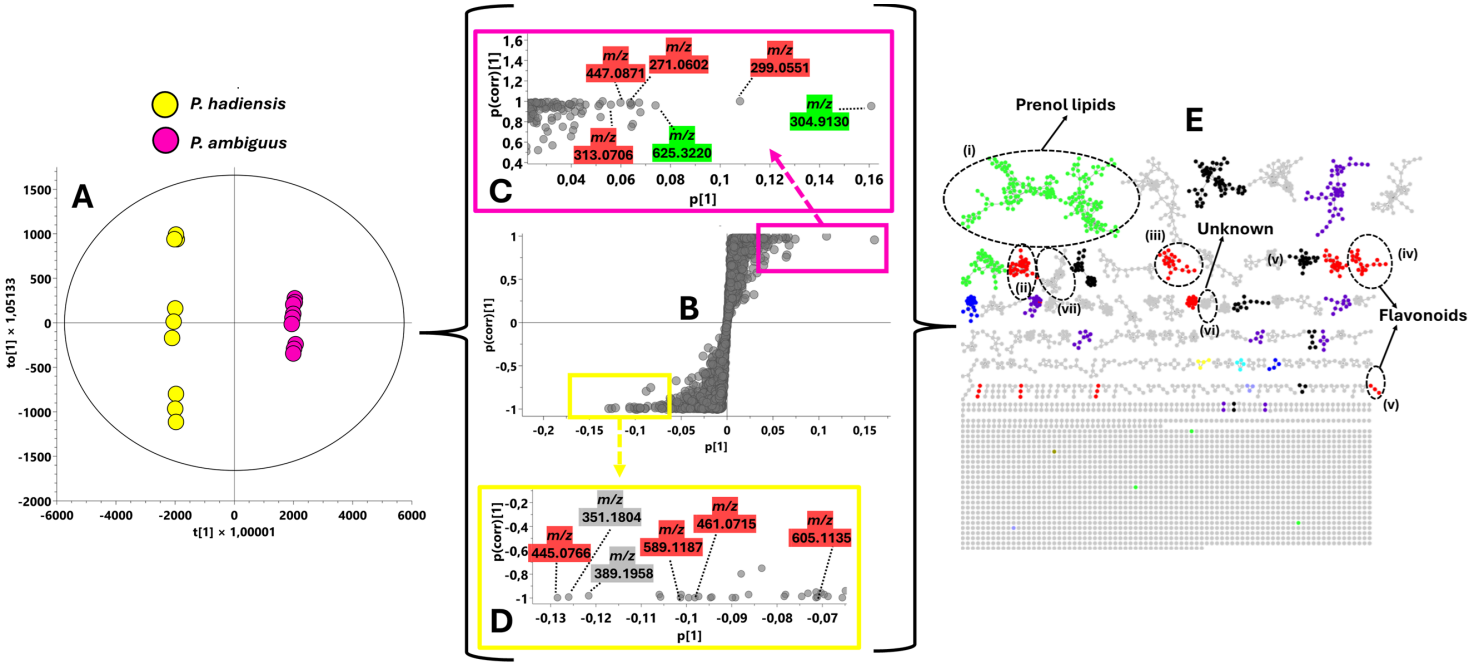


Figure 2

Comparative metabolomics differentiate the metabolomes of *Plectranthus* species. OPLS-DA scores plot (**A**) of *P. hadiensis* and *P. ambiguus* methanolic leaf extracts showing species-specific sample grouping. (**B-D**) S-plot generated from the OPLS-DA model, highlighting the discriminating metabolite features (box on the upper right (**C**) and bottom left (**D**)) responsible for the observed species-specific sample grouping.(**E**) Some of the statistically significant discriminating metabolite features highlighted in **C** and **D** were mapped on the **FBMN of *P. hadiensis* and *P. ambiguus*** and were revealed to belong to the prenol lipid (**i**) and flavonoid (**ii,iii, iv** and **v**) chemical classes, with some mapped on an unknown MF (**vi** and **vii)**

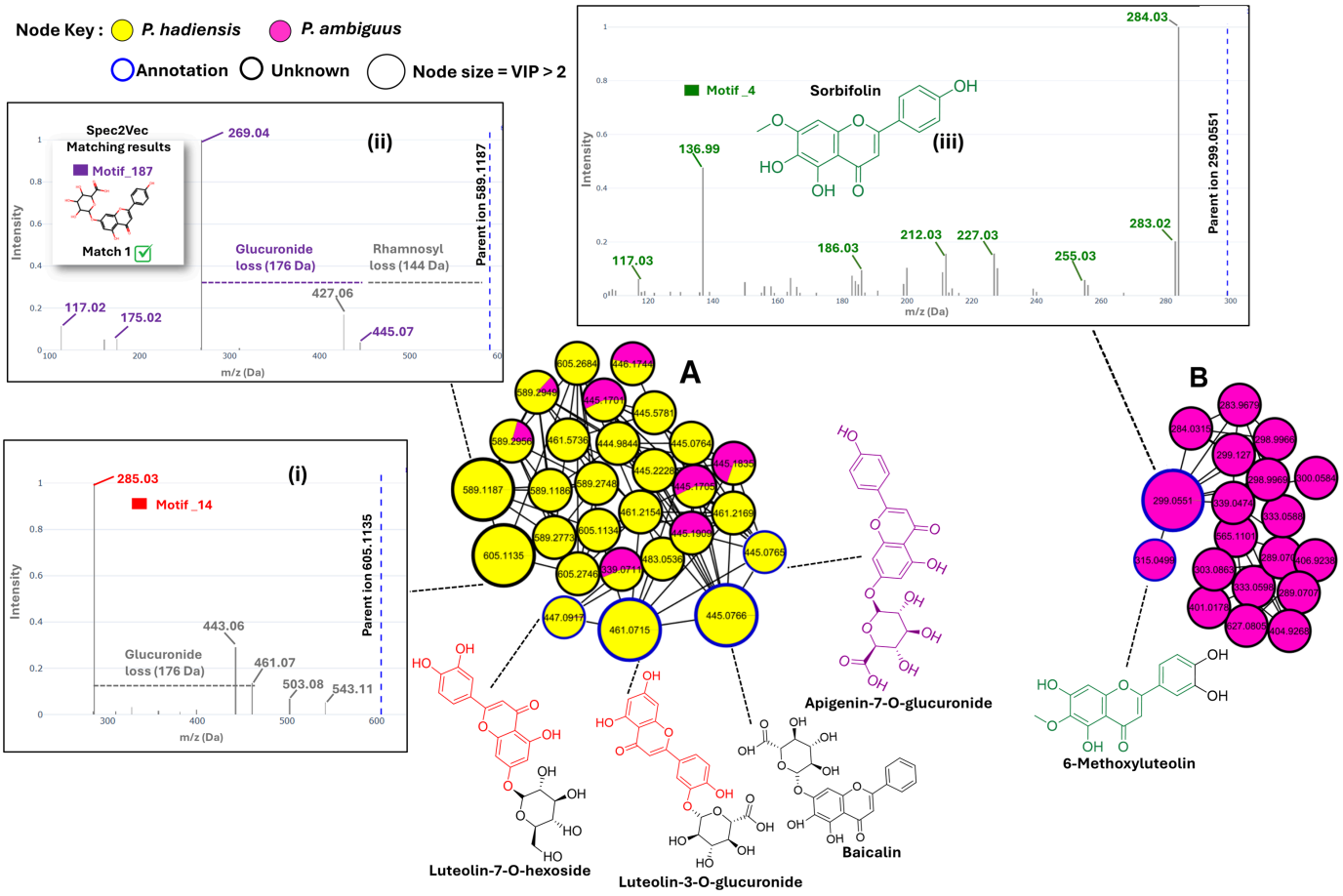


Figure 3

MS2LDA 2.0 substructure discovery on the computed FBMN. (A) An MF of glycosylated flavones with the discovered Mass2Motifs, including **(i)** motif_14 (Luteolin related) and **(ii)** motif_187 (Apigenin-7-O-glucuronide related) annotated through Spec2Vec matching by Mass2Motif Annotation Guidance. (**B)**An MF of methylated flavones with (**iii**) motif_4 (Sorbifolin related). The node keys placed on the top left describe both **A** and **B**molecular families

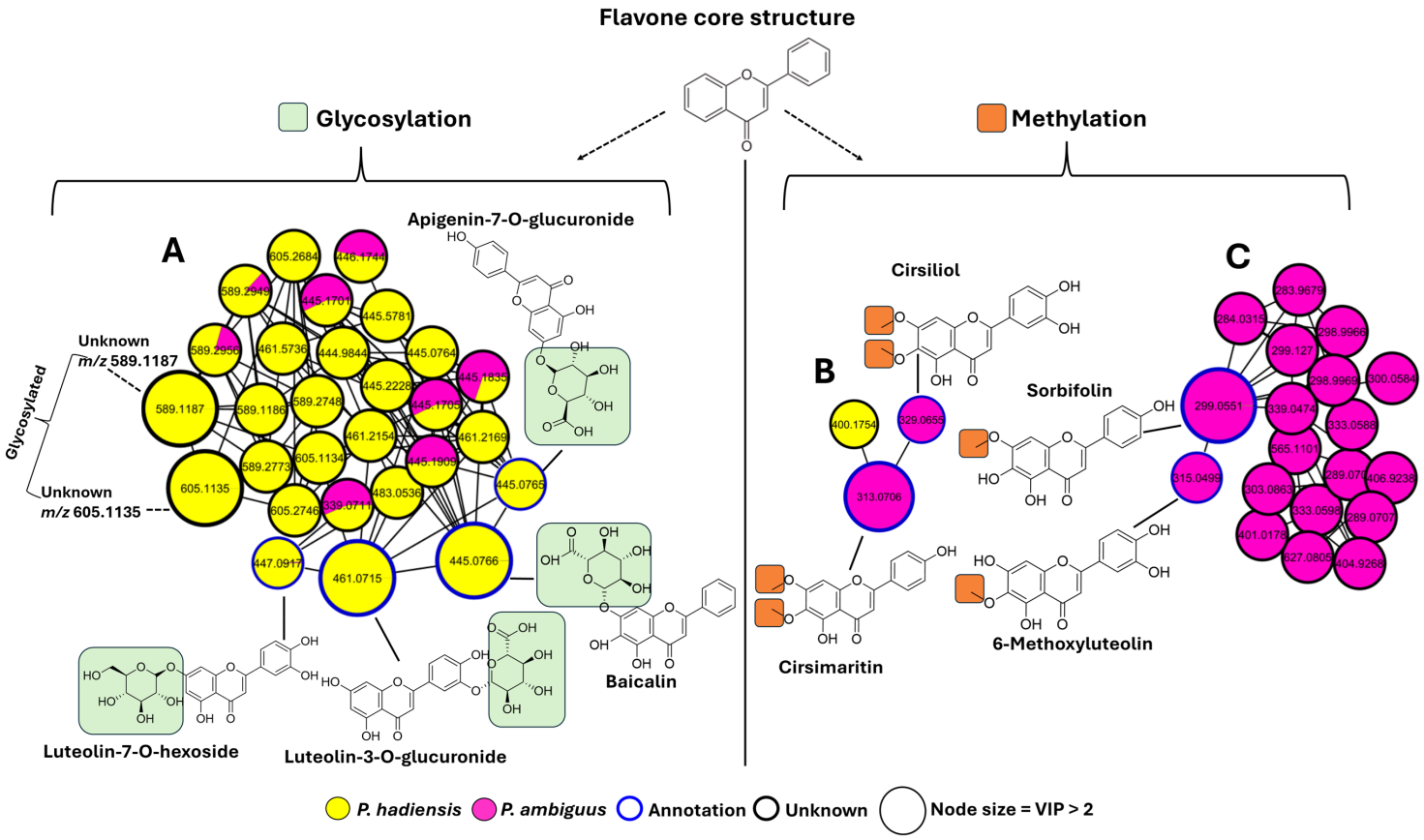


Figure 4

Flavonemolecular families (MFs) from the generated FBMN (**Figure S1**) with glycosylation and methylation modifications in *P. hadiensis* and *P. ambiguus*. **(A)** An MF of glycosylated flavones consisting of luteolin 3-O-glucuronide (*m/z* 605.1135), baicalin (*m/z* 445.0766), *m/z*589.1187 and *m/z* 605.1135 as discriminating metabolite features (VIP score>2) while (**B-C)** are MFs of methylated flavones consisting of cirsimaritin (m/z 313.0706) sorbifolin (*m/z* 299.0551) as discriminating metabolite features (VIP score >2). Glycosylation and methylation are highlighted with the green and orange boxes, respectively. The pie charts indicate the relative abundance of metabolite features across the two plant species.

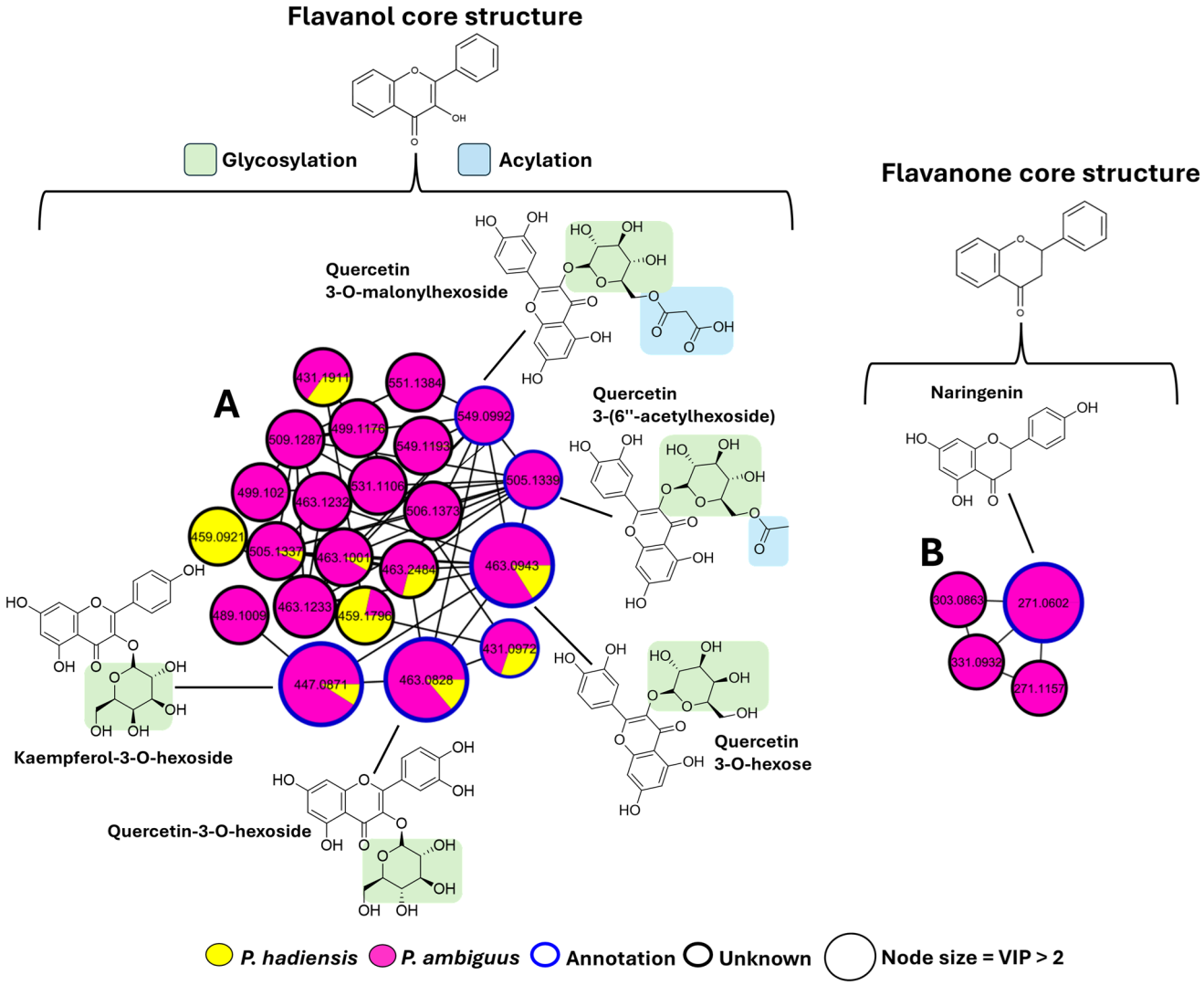


Figure 5

FBMN showing **(A)**flavanols and **(B)** flavonones *P. hadiensis* and *P. ambiguus*, with the sugar moieties highlighted by the green boxes in addition to the acetyl and malonyl conjugates highlighted by blue boxes, respectively. **(A)**Glycosylated flavonols consisting of kaempferol-3-O-hexoside (*m/z*447.0671) and quercetin-3-O-hexoside (*m/z* 463.0828) as discriminating metabolite features. **(B)** A flavonone MF consisting of naringenin (*m/z*271.0602) as a discriminating metabolite feature. The pie charts on both **(A)**and **(B)** indicate the relative abundance of flavonols and flavonones in *P. ambiguus* and *P. hadiensis*.

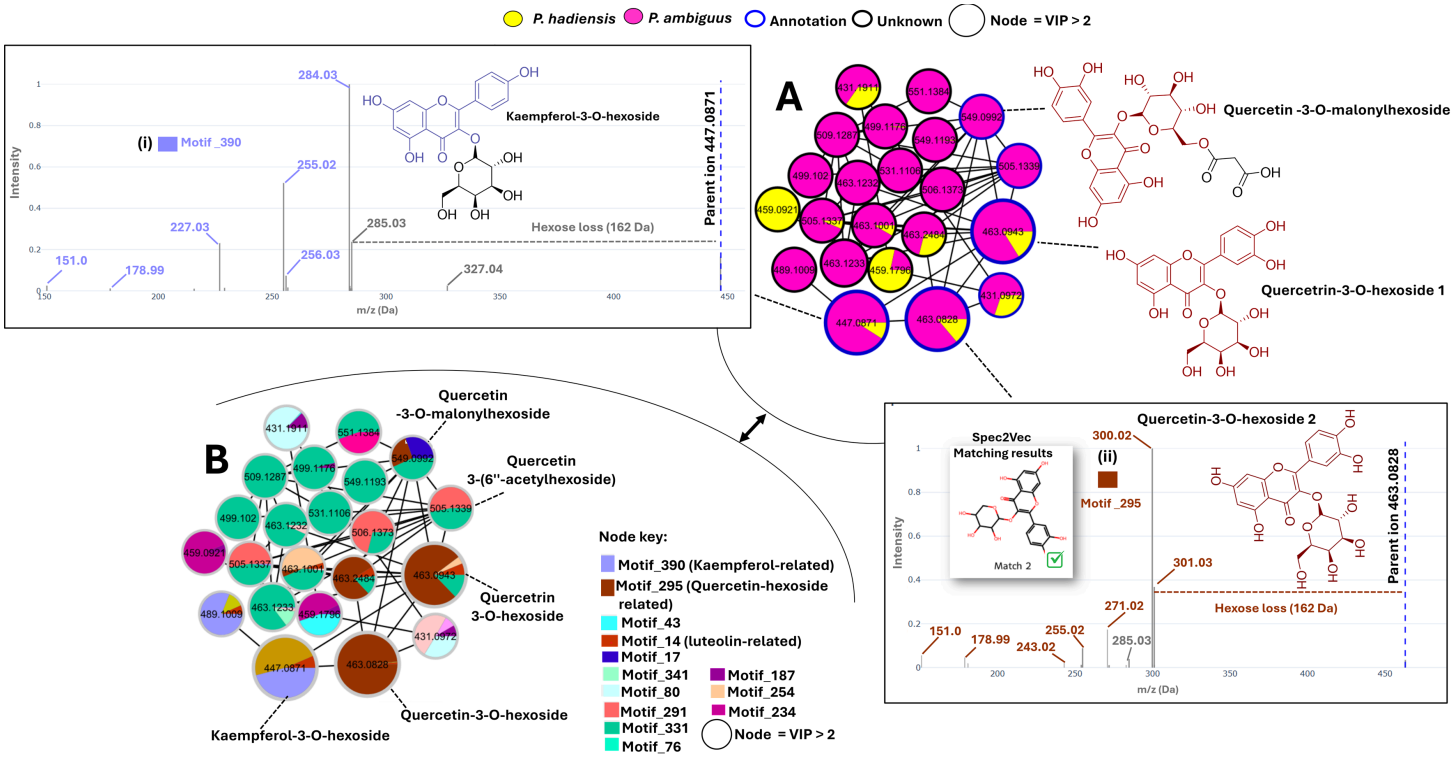


Figure 6

MS2LDA 2.0 substructure discovery on the computed FBMN. (A) An MF of glycosylated flavanols with duplicate MF (**B**) showing 16 discovered Mass2Motifs, including) **(i)** motif_390 (Kaempferol related) and **(ii)**motif_295 (quercetin-3-O-hexoside related). The node key at the top describes MF (**A**), where the pie charts indicate the relative abundance of metabolite features across the two plants species while the pie charts in duplicate MF (**B**) indicate the discovered Mass2Motifs listed at the bottom.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFiles21072026.pdf