



OPEN ACCESS

EDITED BY

Arbi Gueat,
Northern Border University, Saudi Arabia

REVIEWED BY

Lei Zhang,
Second Military Medical University, China
Walid Elfalleh,
Imam Muhammad ibn Saud Islamic
University, Saudi Arabia
Yang Hui Zhao,
Central South University of Forestry and
Technology, China

*CORRESPONDENCE

Wenjun Li
✉ liwenjunao@ms.xjb.ac.cn

[†]These authors have contributed
equally to this work

RECEIVED 26 November 2025

REVISED 12 March 2026

ACCEPTED 26 March 2026

PUBLISHED 27 April 2026

CITATION

Wariss HM, Yang L, Gu J, Liu D and Li W
(2026) Phylogenomic and metabolomic
approaches to provide insight for
species delimitation of cultivated *Ferula*
species in Xinjiang, China.
Front. Plant Sci. 17:1754844.
doi: 10.3389/fpls.2026.1754844

COPYRIGHT

© 2026 Wariss, Yang, Gu, Liu and Li. This
is an open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Phylogenomic and metabolomic approaches to provide insight for species delimitation of cultivated *Ferula* species in Xinjiang, China

Hafiz Muhammad Wariss^{1,2,3,4†}, Lei Yang^{1,3,5,6†}, Jinxin Gu^{1,2,6},
Danhui Liu^{1,2,3,5} and Wenjun Li^{1,2,3,4,6*}

¹State Key Laboratory of Ecological Safety and Sustainable Development in Arid Lands, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, China, ²Xinjiang Key Laboratory of Biodiversity Conservation and Application in Arid Lands, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, China, ³China-Tajikistan Belt and Road Joint Laboratory on Biodiversity Conservation and Sustainable Use, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, China, ⁴Department of Botany, University of Sargodha, Sargodha, Punjab, Pakistan, ⁵The Specimen Museum of Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, China, ⁶College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China

The genus *Ferula* is an important medicinal group within the Apiaceae family, valued globally for its health and culinary uses. Recently, it has gained significant attention because of its cultivation and use to meet growing demands for food and medicine. However, in the Xinjiang production region, farmers often cultivate multiple *Ferula* species together. In the market, leaves from different species are frequently bundled and sold, leading to inconsistent product quality. This issue greatly reduces the economic value of *Ferula* as a food resource and affects the efficacy and safety of its clinical applications. Correctly identifying species is essential for the sustainable use and conservation of these plants. Traditional methods mainly rely on floral and fruit traits, but the monocarpic or polycarpic nature of these plants makes species identification difficult, especially since flowering or fruiting specimens are hard to obtain in cultivated conditions. In this study, leaf samples from 30 *Ferula* specimens collected from ten cultivated locations in Xinjiang were analyzed using phylogenetic methods and ultraperformance liquid chromatography coupled with tandem mass spectrometry (UPLC-MS/MS) to provide insight for species identification and discrimination. Phylogenomic analysis based on the complete plastome was used to clarify relationships among species. The plastome super-barcoding results showed that cultivated *Ferula* species clustered into three distinct clades: *F. sinkiangensis*, *F. teterrima*, and *F. fukanensis*. Untargeted metabolomic profiling detected 7,169 metabolites, and multivariate analyses (PCA, PLS-DA and HCA) revealed clear chemotaxonomic separation among the three species, with clustering patterns consistent with the phylogenetic tree. Our findings demonstrate that the integration of plastome super-barcoding with metabolomic fingerprinting offers a robust strategy to species discrimination in cultivated *Ferula*. This combined approach provides complementary chemotaxonomic evidence for differentiating species, with direct implications for ensuring appropriate use in distinct culinary, medicinal, and pharmaceutical applications.

KEYWORDS

authentication, *Ferula*, metabolomics, multivariate analysis, phylogenomics

1 Introduction

Plants have long served as a vital source of pharmacologically active compounds that support and maintain human health (Shaw et al., 2025). For decades, industries have heavily relied on naturally harvested resources, such as *Polygonati Rhizoma*, as raw materials (Zhu et al., 2014; Guo et al., 2022). However, the growing demand for these materials has led to the overexploitation of natural populations, causing dramatic declines in their availability (Li et al., 2020; Guo et al., 2022). Conserving wild plant populations is substantially crucial and urgently required in order to avoid their extinction, which is extensively used as traditional medicine (Bakhshipour et al., 2019). To address these deficiencies, commercial cultivation programs have been implemented, such as those for medicinal *Polygonatum* in China since 2000, which have alleviated raw material dearth while generating economic benefits and reducing rural poverty (Shen et al., 2019; Guo et al., 2022). The cultivation of medicinal plants has become increasingly significant in recent years due to their essential role in the pharmaceutical and food industries (Guo et al., 2022). Similarly, the cultivation of *Ferula* species, a valuable medicinal plant, in Xinjiang holds great potential. This initiative could address the increasing demand for its medicinal parts while also creating economic opportunities.

The genus *Ferula* L. (Apiaceae), comprising approximately 180–185 species, is widely distributed from the Mediterranean region to Central Asia (Mohammadhosseini et al., 2019; Salehi et al., 2019; Wariss et al., 2025). Morphologically, the genus is characterized as perennial mono- and polycarpic herbs with thick, tall stems and trisected rosette leaves. Its umbels are arranged in panicles, typically without bracteoles. *Ferula* is recognized as one of the most taxonomically complex genera in the Apiaceae family, due to the similar morphologies of its species (Ajani et al., 2008; Yang et al., 2022). *Ferula* is distinguished by the presence of oleo-gum-resins in its roots and stems, which impart a distinct bitter taste and a pungent odor due to high sulfur-containing compounds and additional phytochemicals (Jiang et al., 2023; Khayat et al., 2023). Historically, *Ferula* has held significant value in traditional practices, serving as a versatile herbal medicine, a flavoring agent, and a seasoning in culinary preparations (Boghrati and Iranshahi, 2019; Wang et al., 2023a, Wang et al., 2023b; Bagheri et al., 2024). The medicinal significance of *Ferula* is rooted in its extensive use for treating various ailments, including skin infections, diarrhea, intestinal parasites, influenza, and malaria (Wang et al., 2020; Sonigra and Meena, 2021). Numerous bioactive metabolites have been isolated from *Ferula* species, such as kellerin, umbelliprenin, galbanic acid, and sinkiangenorin, etc., exhibiting a wide array of biological activities, including antiviral, anti-inflammatory, anticancer, antidiabetic, antibacterial, etc., highlighting the immense therapeutic potential of *Ferula* species (Boghrati and Iranshahi, 2019; Mohammadhosseini et al., 2019).

In China, the genus *Ferula*, known as “a wei shu”, includes 31 species, 25 of which are endemic to the Xinjiang (Jiang et al., 2023; Wang et al., 2023c; Wu et al., 2025). Among this diversity, several *Ferula* species native to Xinjiang serve as regional substitutes for the traditional Chinese medicine (TCM) “a wei” (*Ferula assafoetida* L.) (Wang et al., 2018; Zhang et al., 2019; Li et al., 2020). *Ferula*, has a

documented history dating back to the Tang Dynasty classic “Tang Bencao” (AD 659) and is listed in the Pharmacopoeia of the People’s Republic of China (Zhang et al., 2020a). It is also recognized in the Uyghur Medicine Criteria and traditionally used to treat rheumatoid arthritis and stomach disorders (Yang et al., 2006; Wang et al., 2020; Shoaib et al., 2023; Wang et al., 2023d). *Ferula*, a genus of significant ethnobotanical and economic importance in Xinjiang, has historically been used as a vegetable and a medicinal herb. Over time, cultivation efforts emerged in local communities, and *Ferula* species were introduced from the wild into home gardens for culinary, ornamental, and therapeutic purposes. However, the plant’s rising commercial value has spurred large-scale cultivation by companies and farmers, who often collect seeds indiscriminately from wild populations, local gatherers, or informal networks without knowing or prioritizing species-specific identification. Commercial cultivation intensified due to the plant’s high market value, driven by demand for its leaves and roots in food and traditional medicine, yet this practice introduced significant ambiguity in *Ferula* species identification. The fresh aerial parts, particularly the leaves, are sold at a rate of \$9 per kilogram and are commonly used in vegetable dishes, curries, and traditional local cuisine (Figure 1). Growers and sellers often label products generically as “*Ferula*” or “awei” (its Chinese name), obscuring critical distinctions between species such as *Ferula sinkiangensis* K.M.Shen, *F. ferulaeoides* Korovin, and others, which vary in chemical composition and therapeutic efficacy. This ambiguity has fragmented the understanding of cultivated species and raised concerns about the consistency, safety, and effectiveness of *Ferula*-based products.

Taxonomic delimitation within *Ferula* has traditionally depended on floral and fruit morphology, which are often difficult to access in cultivated or immature specimens. This creates a major challenge for applied contexts, where cultivation and market circulation focus on vegetative parts (e.g., leaves) for culinary or medicinal purposes. Additionally, many *Ferula* species are monocarpic (or polycarpic with delayed, irregular flowering), making reproductive structures unavailable for years or entirely absent in harvested material. Despite this, harvested materials from different species are sold in markets under the broad name “*Ferula*” and are valued for their medicinal properties, leading to inconsistent and highly variable product quality. Therefore, resolving this taxonomic uncertainty is essential for standardizing cultivation practices, conserving wild genetic resources, and ensuring the quality, integrity, and reputation of *Ferula* products in local and international markets.

Species identification has traditionally been a challenging task when relying solely on morphological evidence (Duminil and Di Michele, 2009). However, modern methods that integrate morphological data with genetic analysis and chemotaxonomy, including DNA sequencing and metabolomics, have emerged as effective approaches for molecular authentication and distinguishing medicinal plants (Duan et al., 2012; Gallon et al., 2018; Chervin et al., 2019; Negrin et al., 2019; Zhao et al., 2019; Cheng et al., 2020; Fayek et al., 2021). Advancements in high-throughput sequencing technologies have led to the use of genomic data, such as complete plastid genome (plastome) sequences



FIGURE 1

Cultivation and utilization of *Ferula* species in Xinjiang, China. (A) Field cultivation of different *Ferula* species; (B) Harvesting of aerial parts and young leaves; (C) post-harvest packing and weighing; (D) young plants cultivated for culinary and medicinal purposes; and (E, F) Fresh leaves sold at local bazaar in Urumqi, Xinjiang, China.

(Chong et al., 2022; Shen et al., 2024), which are now recommended as “super-barcodes” to address challenges in species identification and delimitation that standard DNA barcodes cannot resolve (Nock et al., 2011; Dodsworth, 2015). This advancement greatly improves the resolution of species identification and the authentication of traditional medicinal plants (Alberts and Meyer, 2022; Wang et al., 2023e; He et al., 2024). Plant secondary metabolites play a crucial role in distinguishing morphologically similar species (Nguyen et al., 2016). The phytochemical diversity inherent in these metabolites provides a foundation for understanding the evolution and classification of plant species (Shen et al., 2022). Metabolomics, as a widely adopted method, enables comprehensive profiling and comparison of phytochemical compositions across species (Li et al., 2022). This approach, when applied in chemotaxonomy, offers both qualitative and quantitative insights into a vast array of metabolites, significantly enhancing the accuracy of species identification and quality assessment (Masson et al., 2014; Zhang et al., 2023).

In the current work, we hypothesize that an integrated approach, combining phylogenomic analysis based on complete plastomes with untargeted metabolomic profiling, will provide a robust and reliable method for accurately delineating and authenticating the cultivated *Ferula* species in Xinjiang, overcoming the limitations of traditional methods. Therefore, the primary objectives of this study was: (1) to construct a phylogenomic framework for cultivated Xinjiang *Ferula* using complete plastome sequences; (2) to profile the metabolomic diversity of these accessions using UPLC-MS/MS; and (3) to integrate phylogenetic and chemotaxonomic data to establish a definitive method for species identification. The results of this

research will serve as a foundation for the accurate authentication of cultivated *Ferula* species, ensuring the integrity of medicinal materials, with potential applications in medicinal and sustainable utilization of this valuable medicinal resource.

2 Materials and methods

2.1 Plant materials

In this study, field visits were conducted to ten agricultural farms in Xinjiang, China, to collect fresh leaf material of cultivated *Ferula* species. Given that the primary objective of this study was species authentication, the sampling strategy prioritized capturing interspecific diversity. From each farm, three individual plants were sampled, yielding a total of 30 fresh leaf samples collected across the ten sites. Therefore, for each species identified, we collected multiple biological replicates (distinct individuals), ensuring that metabolic variation could be assessed at the species level. Detailed information on the samples and their collection sites is provided in [Supplementary Table 1](#). To minimize the confounding effects of environment and development on metabolomic profiles, strict sampling controls were implemented. All leaf tissue was harvested from plants at an identical phenological stage (pre-flowering) and using the same leaf position (Basal leaf) to minimize developmental variation. For phylogenetic analysis, leaf material from each of the 30 individual plants were collected and immediately preserved in silica gel. For non-targeted metabolomic analysis, leaves were separately collected from the same 30 individual plants,

transported to the laboratory in dry ice, and stored at -80°C until processing. All sampling was conducted in June 2024. Additionally, sample 31st corresponds to a wild *F. fukanensis* specimen obtained from the herbarium. The collection and identification processes complied with institutional, national, and international guidelines (Buck and Hamilton, 2011; Gallagher et al., 2025). Voucher specimens were deposited at the Herbarium of the Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences (XJBI).

2.2 Morphological evaluation

All photographs were captured using a Nikon Z7II camera to record morphological features for morphometric studies. Macro images were obtained using a Laowa 25mm F2.8 2.5-5X macro lens mounted on a sliding track to ensure stability and precision in capturing fine details. The clarity of the photographs was enhanced through the depth of field stacking using Helicon Focus Pro version 6.3. This technique combined multiple images taken at varying focal depths, resulting in high-resolution composite images suitable for detailed morphological studies.

2.3 Phylogenomic studies

2.3.1 Genomic DNA extraction, library preparation, and sequencing

Genomic DNA was extracted from silica-dried leaves using the plant genome extraction kit (DP320) from Tiangen Biochemical Technology (Beijing, China) following the manufacturer's instructions. The quality and quantity of the extracted DNA were assessed using a NanoDrop spectrophotometer and agarose gel electrophoresis. All samples had an A260/A280 ratio between 1.8 and 2.0 and an A260/A230 ratio > 2.0 , confirming high purity and minimal contamination. A total amount of $1.5\text{ }\mu\text{g}$ DNA per sample was used for library preparation. Sequencing libraries were generated using Truseq Nano DNA HT Sample Preparation Kit (Illumina USA) following the manufacturer's recommendations, and index codes were added for sample identification. Briefly, the DNA sample was fragmented by sonication (Covaris M220, 200 cycles per burst, 55 seconds treatment time) to a size of 350 bp, then the DNA fragments were end polished, A-tailed, and ligated with the full-length adapter for Illumina sequencing, followed by further PCR amplification (8 cycles). Finally, PCR products were purified (AMPure XP system) and libraries were analyzed for size distribution by Agilent2100 Bioanalyzer and quantified using real-time PCR. Paired-end sequencing (150 bp) of genomic DNA was conducted on the Illumina NovaSeq 6000 platform with an insert size of approximately 350 bp. Raw data were processed using in-house C scripts for quality control, removing unreliable reads and PCR duplicates (identical paired-end reads) to ensure reliability and reduce bias in subsequent analysis. The complete plastome sequences newly generated in this study have been deposited in GenBank under accession numbers PQ834963 to PQ834992 (a full list is provided in Supplementary Table 1). For subsequent phylogenetic analysis, the newly sequenced plastomes were aligned with publicly available, morphologically verified plastome

sequences of *Ferula* species downloaded from GenBank (e.g., OP324724 for *F. teterrima*, MW411057 for *F. sinkiangensis*), which served as positive controls to validate our assembly and annotation pipeline.

2.3.2 Genome assembly and annotation

The GetOrganelle v. 1.7.7.0 pipeline (<https://github.com/Kinggerm/GetOrganelle>) was used to obtain plastid-like reads by filter paired-end read (Jin et al., 2020) and SPAdes v. 3.10 plugin in the GetOrganelle pipeline was used to assemble reads (Bankevich et al., 2012). The complete circular assembly was extracted and visualized using Bandage v. 0.8.1 (Wick et al., 2015). The genomes were annotated using CpGAVAS (Liu et al., 2012) and PGA (<https://github.com/quxiaojian/PGA>), followed by manual checked and adjustments in Geneious v. 9.1.7 (Kearse et al., 2012) to correct putative start/stop codons and intron positions following comparisons with the reference plastome (*F. transiliensis* (Regel & Herder) Pimenov: ON324040). Finally, the plastome sequences generated in this study were submitted to GenBank for accession numbers (Supplementary Table 1).

2.3.3 Phylogenetic analysis

The 31 plastomes newly sequenced in this study were combined with 20 additional *Ferula* plastomes and two outgroup species (*Cuminum cyminum* L., *Daucus aureus* Desf.), all of which were downloaded from the NCBI database, and employed for phylogenetic analysis. In total, 53 complete plastome sequences were initially aligned using MAFFT v.7.221 with the G-INS-I algorithm in Geneious (Kato and Standley, 2013). The resulting alignment was subsequently trimmed in TrimAl v.1.2 (Capella-Gutierrez et al., 2009) using the 'automated1' parameter to remove ambiguously aligned regions, sequences of uncertain homology, and single-taxon insertions (Capella-Gutierrez et al., 2009; Chen et al., 2016). ModelFinder (Kalyaanamoorthy et al., 2017) was used to select the best-fit model of nucleotide substitution with the Bayesian Information Criterion (BIC) within Phylosuite v.1.2.3 (Zhang et al., 2020b). The Maximum Likelihood (ML) tree was inferred using IQ-TREE v. 2 (Stamatakis, 2014; Drew et al., 2017) with a GTR+I+G model and assessed with 1,000 bootstrap replicates. MrBayes v.3.2.7 (Ronquist et al., 2012) was employed to construct a Bayesian Inference (BI) phylogenetic tree utilizing the optimal DNA substitution model (GTR+F+I+G4). The Markov chain Monte Carlo (MCMC) analysis was run with two independent replicates, each consisting of four chains starting from a random tree. The chains were sampled every 1,000 generations over 20 million generations, continuing until the average standard deviation of split frequencies fell below 0.01. A burn-in of 25% was discarded to account for initial chain non-stationarity. Phylogenetic trees were visualized using FigTree v.1.4.2 (Rambaut, 2014). In addition to plastome analysis, we also extracted and analyzed the nuclear Internal Transcribed Spacer (ITS) and External Transcribed Spacer (ETS) sequences from the 31 sampled taxa. Each marker was aligned separately using MAFFT v.7.221, after which the sequences were concatenated using Phylosuite. Phylogenetic analysis of the concatenated nuclear dataset followed the same methodology as applied to the plastome data.

2.4 Metabolomic analysis

2.4.1 Sample preparation and extraction

Metabolites extraction from *Ferula* samples was performed according to the standard protocol provided by Wuhan MetWare Biotechnology Co., Ltd (Li et al., 2023). A comprehensive quality control (QC) strategy was implemented throughout the analytical process. To monitor system background and cross-contamination, solvent blanks (70% methanol) were injected at the beginning, end, and after every 10 experimental samples. Additionally, a pooled QC sample, prepared from equal aliquots of all experimental extracts, was analyzed after every 10 experimental injections to monitor instrument stability over time. Briefly, plant samples were prepared using vacuum freeze-drying technology in a lyophilizer (Scientz-100F) for 63 hours. The dried samples were ground into a fine powder using a mixer mill (MM 400, Retsch) operating at 30 Hz for 1.5 min. To correct for potential losses during sample processing and fluctuations in instrument response, 2-chlorophenylalanine (98% purity, 1 ppm final concentration; J&K Scientific, CAS: 14091-11-3) was added as an internal standard to the extraction solvent. A total of 50 mg of powdered sample was weighed using an analytical balance (MS105DM) and extracted with 1200 μ L of pre-cooled (-20 °C) 70% methanolic aqueous solution containing the internal standard. The order of sample extraction and LC-MS injection was not randomized, which is a potential source of technical bias. However, the high reproducibility observed in the quality control (QC) samples interspersed throughout the analytical sequence demonstrates exceptional system stability. This indicates that non-biological technical variation was effectively controlled and had minimal impact on the identification of differential metabolites. Samples were vortexed every 30 min for 30 sec, repeated six times, and centrifuged (12,000 rpm, 3 minutes). The resulting supernatants were filtered through a 0.22 μ m microporous membrane and stored in vials at -80 °C before performing the ultra-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) analysis.

2.4.2 Metabolite analysis by UPLC-MS/MS

Leaf sample extracts were analyzed using the UPLC-MS/MS system in both positive and negative ionization modes, following the methods described by Wuhan MetWare Biotechnology Co., Ltd. and as previously referenced (Li et al., 2022; Feng et al., 2024; Li et al., 2024). Both analyses utilized a Waters ACQUITY Premier coupled with an HSS T3 column (1.8 μ m, 2.1 mm $\times$ 100 mm) with mobile phases consisting of 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B). The gradient program was as follows: started at 5% B, increased linearly to 20% B over 2 min, then to 60% B over the next 3 min, and further increased to 99% B in 1 min, holding at 99% B for 1.5 min. The mobile phase was then returned to 5% B within 0.1 min and held for 2.4 min. The chromatographic conditions for both methods were identical, with a column temperature of 40 °C, a flow rate of 0.4 mL min⁻¹, and an injection volume of 4 μ L.

The mass spectrometer was operated using the information-dependent acquisition (IDA) mode with Analyst TF 1.7.1 Software (Sciex, Concord, ON, Canada). The ion source conditions included ion source gas 1 (GAS1) and gas 2 (GAS2) at 50 psi, a curtain gas (CUR) at 25 psi, and a source temperature (TEM) of 550 °C. The ion spray voltage floating (ISVF) was set to +5000 V for positive mode and -4000 V for negative mode, with a declustering potential (DP) of $\pm$ 60 V. The TOF MS scans were conducted over a mass range of 50–1000 Da with an accumulation time of 200 min and dynamic background subtraction enabled. For product ion scans, the mass range was set from 25–1000 Da with an accumulation time of 40 min, and a collision energy of $\pm$ 30 V with a spread of 15 V. The resolution was set to UNIT, and a maximum of 18 candidate ions were monitored per cycle, with isotopes within 4 Da excluded and a mass tolerance of 50 ppm (Li et al., 2024).

2.4.3 Data processing and multivariate analysis

The acquired MS files from the mass spectrometry analysis were exported in mzXML format by ProteoWizard (Holman et al., 2014). The mzXML files were processed using the XCMS R Package for peak extraction, alignment, and retention time correction. XCMS parameters included mass tolerance of $\pm$ 15 ppm, and peak width range of 5–30 seconds. Peaks with a missing rate exceeding 50% in any sample group were excluded, and missing values were imputed by means of the K-nearest neighbors (KNN) (K = 5) method using R package impute 1.56.0. The peak areas were calibrated using the Support Vector Regression (SVR) method. After calibration and filtering, metabolites were identified based on the self-built MetWare database (MWDB) (Chen et al., 2013), integrating public databases, viz., PubChem (Kim et al., 2025), Metlin (Smith et al., 2005), and predictive libraries, and aligned by MetDNA2 (Zhou et al., 2022). The internal standard was used for process monitoring only and was not used for quantitative normalization of other metabolites. Finally, metabolites with an identification score above 0.5 and a QC sample coefficient of variation (CV) value below 0.5 were retained. Data from both positive (ESI⁺) and negative (ESI⁻) ion modes were merged, retaining metabolites with the highest identification scores and the lowest CV values.

To understand the similarities and differences between metabolic components among species, a multivariate analysis was conducted on mass spectrometry data obtained from UPLC-MS/MS full scan mode using the XCMS software. Multivariate analyses, including principal component analysis (PCA), hierarchical cluster analysis (HCA), and partial least squares-discriminant analysis (PLS-DA), were conducted using the statistics function in the R package prcomp (ver. 3.5.1), ComplexHeatmap (v. 1.2.1; 2.71.1009), and MetaboAnalyst R package (v 4.0, R) (www.r-project.org) respectively, using full-scan-MS acquisitions in both positive and negative ion modes. Metabolite data were normalized using unit variance (UV) scaling (Base package 4.1.2) and visualized using heatmaps generated with the R ComplexHeatmap package to gain insight into the distinct metabolic profiles among *Ferula* samples used in this study. Initially, to understand the structure of the metabolomic dataset, unsupervised PCA was performed, with

relationships visualized using score plots of the first two principal components (PC1 vs. PC2). QC samples displayed a well-clustered distribution in principal component analysis (Supplementary Figures 1, 2), confirming analytical system stability and data reliability throughout the acquisition process. The PLS-DA model's performance was assessed using R^2X , R^2Y , and Q^2 metrics. R^2X and R^2Y represent the explained variance of the X and Y matrices, respectively, while Q^2 indicates the model's predictive capability. Values closer to 1 signify a more robust model; $Q^2 > 0.5$ denotes an effective model, and $Q^2 > 0.9$ reflects an outstanding model. It is important to note that no independent external test set was used in this study. Therefore, the PLS-DA model and its derived VIP scores should be interpreted as exploratory tools for highlighting group differences and guiding biomarker selection, rather than as a validated predictive classifier for new samples. Model validity was confirmed through permutation testing (200 iterations). The p-values for R^2Y and Q^2 were calculated based on the frequency of random models outperforming the original model in explained variance (R^2Y) and predictive ability (Q^2). The PLS-DA model demonstrates strong discriminative power ($R^2Y = 0.57$ – 0.59) and predictive accuracy ($Q^2 = 0.608$), validated by permutation testing ($p < 0.005$) in positive ion mode (Supplementary Figures 3, 4). Similarly, in negative ion mode, the PLS-DA model demonstrated strong discriminative power ($R^2Y = 0.59$, $Q^2 = 0.69$) and robustness (permutation test $p < 0.005$ for both Q^2 and R^2Y) (Supplementary Figures 5, 6), revealing that these findings support its use in identifying biologically relevant metabolites driving group differences. Subsequently, supervised PLS-DA was applied to enhance group discrimination among species. Metabolites

contributing significantly to class separation were identified based on variable importance in projection (VIP) scores, with a threshold of $VIP > 1.0$ used to select discriminatory features. Model validity was confirmed via 7-fold cross-validation (CV) and permutation testing ($n = 1,000$ iterations, $p < 0.05$). Metabolites were prioritized using VIP scores ($VIP > 1.0$) derived from the PLS-DA model (biological replicates ≥ 3). Finally, HCA was employed to elucidate the chemotaxonomic relationships among *Ferula* species with results displayed as dendrogram-coupled heatmaps showing sample and metabolite clustering patterns.

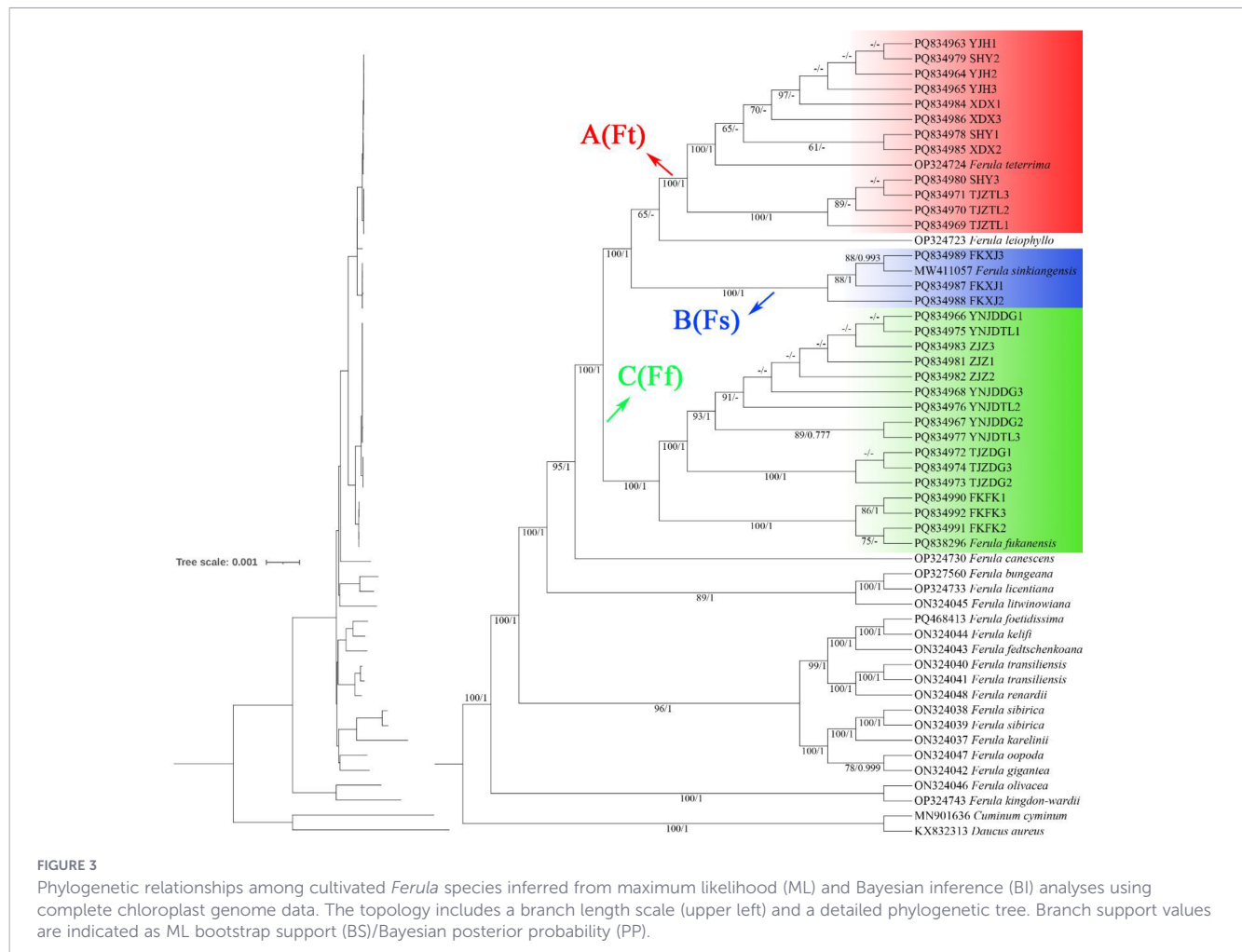
3 Results and discussion

3.1 Morphological study of *Ferula* species

The morphological comparisons of *Ferula* species can't reveal notable variations in leaf morphology differences across the examined samples (Figure 2). Images A–J represent the general habit and detailed leaf morphology of different specimens examined in this study. At the same time, subpanels (A1–J2) in (Figure 2) highlight adaxial and abaxial views of key morphological features including lobation, venation and surface pubescence. The collected sample leaflets are all trilobate and 2–3 times pinnately lobed. Through macro photography observation, it was found that the upper and lower surfaces of A, C, E, F, G, I, J have short, soft hairs, and the leaf margin and veins are very distinct. The leaf underside has significantly more hair than the upper surface. B and D only have short, soft hairs visible along the leaf margin (Figure 2B); H has



FIGURE 2
Leaf morphology of cultivated *Ferula* species. (A) YJH; (B) YNJDG; (C) TJZTL; (D) TJZDG; (E) YNJDTL; (F) SHY; (G) ZJZ; (H) XDX; (I) FKXJ; (J) FKFK; (A1–J1) The adaxial surface of the leaf. (A2–J2) The abaxial surface of the leaf.



dense long, soft hairs on both the upper and lower surfaces of the leaf, with the underside again having significantly more hair than the upper surface (Figure 2). Subsequently, key morphological features such as growth form, leaf arrangement and shape, venation, inflorescence structure, color and shape of flowers and fruits, were primarily used for preliminary identification of *Ferula* species as described in various taxonomic treatments of *Ferula* and Apiaceae (Korovin et al., 1951; Sheh et al., 2005). The traditional classification of *Ferula* species is mainly based on the morphological and anatomical characteristics (Korovin et al., 1951; Safina and Pimenov, 1983). However, in some cases, morphological differences between species are very small, while significant intraspecific variation can occur due to environmental influences, which significantly affects the species identification (Des Roches et al., 2018; Wang et al., 2022). For example, in this study, samples A, C, F, and H were identified as *F. teterrima* Kar. & Kir. based on our phylogenetic analysis (Figure 3). However, sample H exhibited distinct long, soft hairs likely influenced by its age (8 years) and its growth in a highly arid, undisturbed slope environment with minimal human interference. The phenotype of sample H may reflect morphological plasticity in response to local conditions, further reinforcing the need for molecular and metabolomic data for accurate identification. Furthermore, samples B, D, E, and G were all identified as *F. fukanensis* K.M.Shen based on phylogenetic

analyses (Figure 3), but they could not be distinguished from one another based on their pubescence (hairiness) or leaf shape (Figure 2). Cultivated *Ferula* plants are typically harvested in batches, primarily for their young leaves, which makes it impossible to accurately determine the provenance of the germplasm just by examining the young leaves. Morphology serves as a preliminary criterion for plant identification (Kaplan, 2001), while molecular and metabolomic evidence can provide more reliable methods for accurately identifying complex plant species (CBOL Plant Working Group et al., 2009; Yang et al., 2013; Du Preez et al., 2025).

3.2 Phylogenetic analysis of *Ferula* species

In this study, based on complete plastome sequences, 53 sequences from *Ferula* were used to construct the ML and BI trees (Figure 3). The topologies of the ML and BI phylogenetic trees were consistent, and most nodes have strong support values (PP $\geq$ 0.997, BS $\geq$ 99%) therefore, only the ML bootstrap support (BS) and Bayesian posterior probability (PP) values indicated at each node (Figure 3). The ML phylogenetic tree illustrates the evolutionary relationship among *Ferula* species based on whole plastome genomic data. The cultivated *Ferula* cluster within three major clades: clade A (red), clade B (blue), and clade C (green) (Figure 3).

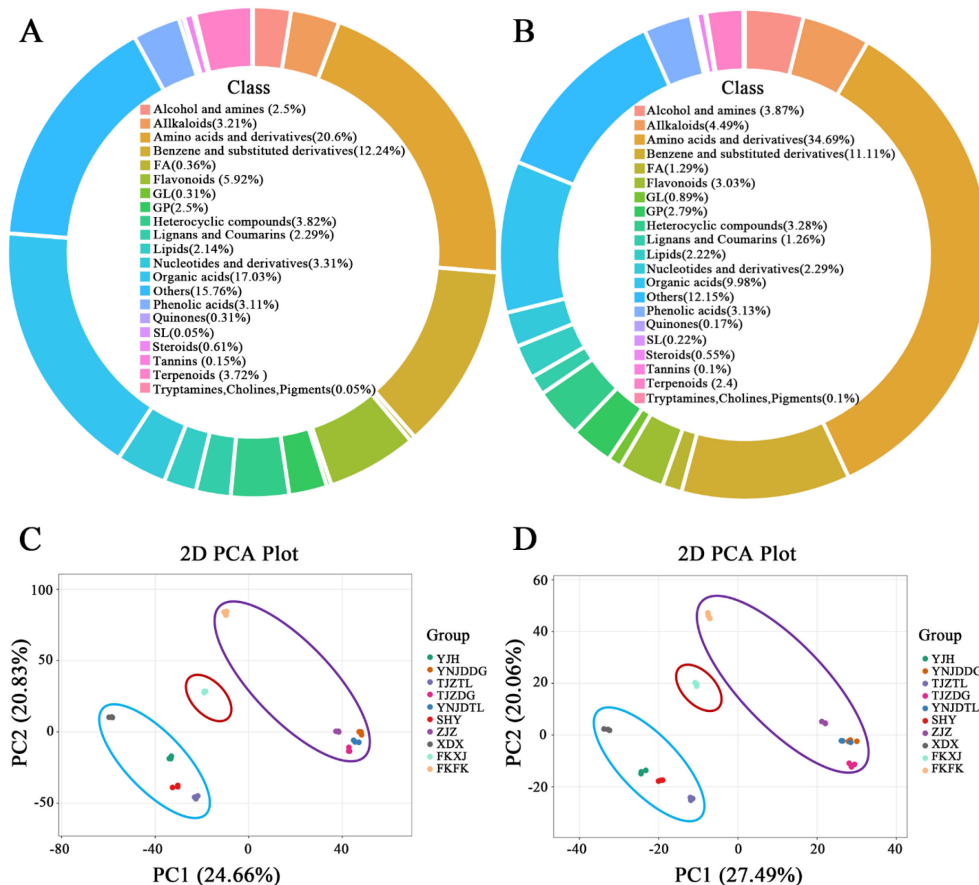


FIGURE 4

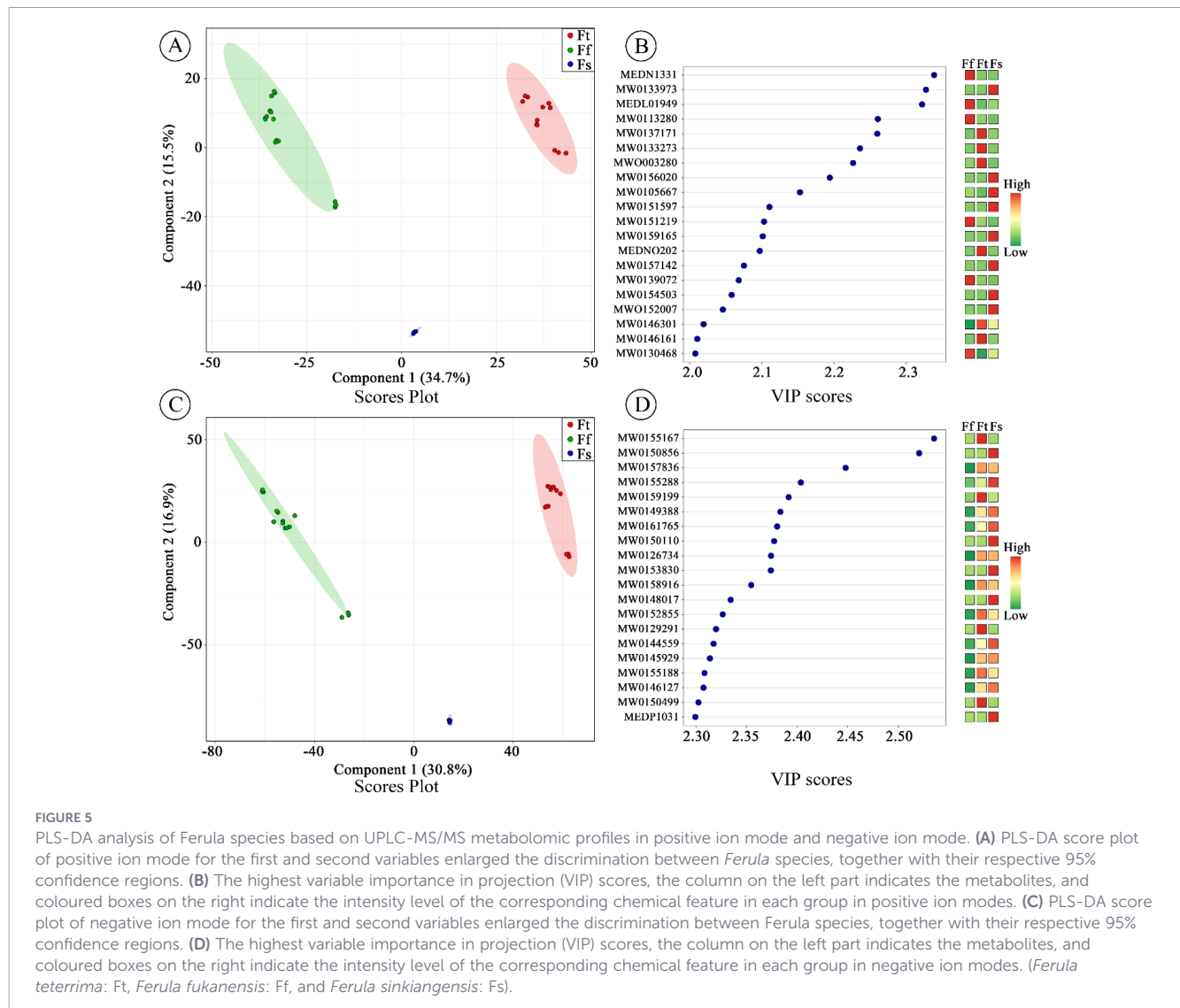
Metabolomic profiles and PCA of cultivated *Ferula* species based on UPLC-MS/MS data in (A) positive ion mode and (B) negative ion mode. (C) PCA scores plot in positive ion mode and (D) PCA scores plot in negative ion mode, illustrating species separation along PC1 and PC2. FA, Fatty Acids; GL, Glycerolipids; GP, Glycerophospholipids; SL, Sphingolipids.

The samples YJH, XDX, TJZTL, and SHY clustered together with *F. teterrima* to form clade A. Clade B consists of FKXJ (FKXJ1, FKXJ2, FKXJ3) with *F. sinkiangensis*, confirming their distinct lineage. The samples YNJDTL, YNJDGD, ZJZ, TJZDG, and FKFK, along with *F. fukanensis*, formed clade C, and the relationships of the three clades were well-resolved with strong support (BS = 100%, PP = 1) (Figure 3). Furthermore, the results of nuclear gene phylogenetic analysis based on ML and BI phylogenies showed that clade B (blue) exhibited a polyphyletic relationship, with lineages embedded within clade A (red). In contrast, clade C (green) retained a strong monophyletic relationship (Supplementary Figures 7, 8).

Multiple accessions of the same species are labeled according to Supplementary Table 1. Chloroplast genomic data supported the identification of cultivated *Ferula* species into three clades corresponding to *Ferula teterrima* (A/Ft), *F. sinkiangensis* (B/Fs), and *F. fukanensis* (C/Ff) in the phylogenetic tree. Phylogenetic tree nodes marked with -/- indicates support less than 50).

Accurately delineating species boundaries is a crucial step in exploiting and conserving plant species (Guo et al., 2022; Zhao et al., 2025). *Ferula* is a medicinal plant, cultivated for the pharmaceutical industry and rural development; however, its taxonomic delimitation remains ambiguous. The monocarpic nature of *Ferula* significantly limits access to reproductive structures, which are critical for traditional taxonomic identification (Pimenov and Leonov, 1993;

Downie et al., 2010). Accurate identification of *Ferula* requires complete, mature specimens including roots, stem bases, basal leaves, inflorescences, flowers, and ripe fruits. Relying solely on young basal leaves, particularly in early growth stages, is highly unreliable and often leads to misidentification (Panahi et al., 2018; Negrin et al., 2019). Furthermore, repeated harvesting of cultivated specimens for leaf material further reduces the availability of diagnostic morphological characters in market samples. These constraints render conventional morphological assessment unreliable for species delimitation in this genus. DNA barcoding enables direct comparison of genetic differences between species (Tan et al., 2020). Since its introduction by Canadian biologist Paul Hebert (Costa et al., 2007; Hebert et al., 2003; Liu et al., 2025a), DNA barcoding has advanced rapidly. For plants, commonly proposed plastid DNA barcodes include: *matK*, *rbcl*, *trnH-psbA*, and ITS (Chase et al., 2005; Kress et al., 2005; Cowan et al., 2006; Xue et al., 2024). The first-generation plant DNA barcode, based on 2–3 plastid fragments and the ribosomal ITS sequence, has significantly advanced plant species identification, biodiversity research and evolutionary studies (Cowan et al., 2006; Dodsworth, 2015). Our nuclear gene phylogeny exhibits topological discordance with the plastome-based phylogeny, particularly with *F. sinkiangensis* being nested within the *F. teterrima* clade (Supplementary Figures 7, 8). Two plausible explanations may account for this discordance. The first is taxonomic: as noted in the

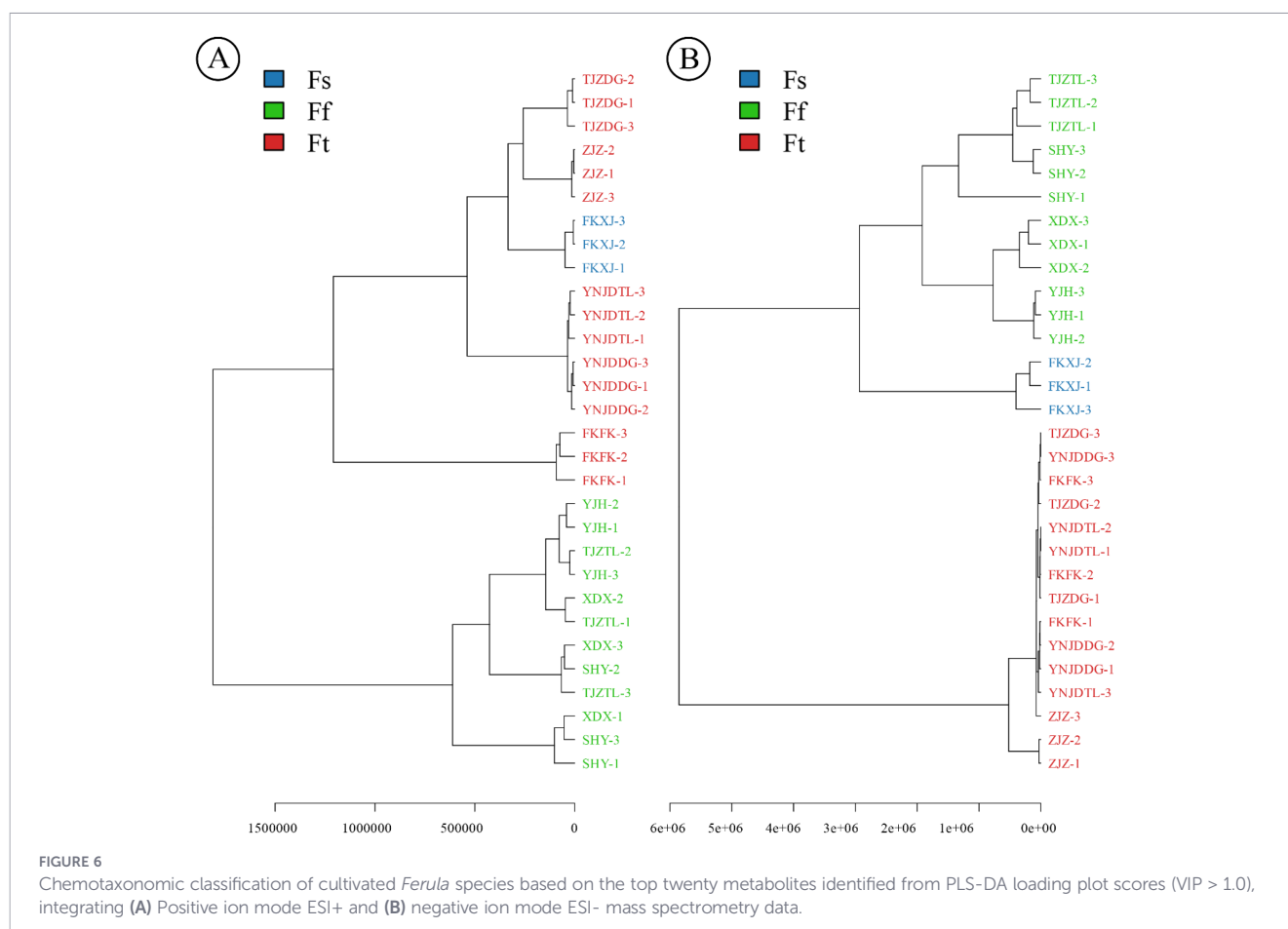


Flora of Xinjiangensis, when *F. sinkiangensis* was originally described, some authors have subsequently treated *F. sinkiangensis* as synonymous with *F. teterrima*. The Flora emphasizes that *F. sinkiangensis* can be distinguished from *F. teterrima* by its multiple vittae in the fruit, a character that has led most recent taxonomic treatments to maintain them as separate species. However, our phylogenetic data alone cannot definitively resolve this long-standing taxonomic ambiguity (Shen, 2011). Second, the chloroplast genome is maternally inherited and may exhibit evolutionary asynchrony with the nuclear genome. Potential biological factors leading to such discordance include incomplete lineage sorting (ILS) and gene introgression (Zhao et al., 2023; Liu et al., 2025b). Given that the sampled *Ferula* individuals had not flowered or exchanged genes prior to collection, contemporary gene flow can be largely excluded, making ILS a more plausible explanation than recent hybridization or chloroplast capture. ILS occurs when ancestral polymorphisms persist through successive speciation events and are randomly sorted into descendant lineages, leading to gene trees that deviate from the species tree. This phenomenon is particularly pronounced in lineages that underwent rapid radiation, such as *Ferula*, where coalescence times may be long relative to speciation intervals (Kurzyńska-Mitynik

et al., 2008; Panahi et al., 2015, 2018). Additionally, the nuclear ITS/ETS fragments analyzed were relatively short and contained limited phylogenetically informative sites. This resulted in low nodal support across much of the *Ferula* phylogeny (Supplementary Figures 7, 8), particularly within taxonomically complex clades, further obscuring relationships among closely related species. The limitations of first-generation barcodes in resolving complex groups like *Ferula* highlight the need for more comprehensive approaches. Therefore, for authenticating cultivated *Ferula* germplasm, we utilized whole plastome genomic data to provide a more robust and reproducible framework (Figure 3). Using the plastome as a super-barcode offers a powerful solution, enabling more effective species identification within such challenging taxa (Liu et al., 2021; Lv et al., 2023; Dewar et al., 2025), even as the full evolutionary history remains partially unresolved.

3.3 Metabolic profiles of *Ferula* species via UPLC-MS/MS

Metabolomics is a powerful tool for identifying and characterizing medicinal plants through the analysis of their



chemical profiles. Therefore, when integrated with phylogenetic data, metabolomic profiling provides powerful complementary evidence for species identification and chemotaxonomic classification. It helps differentiate plant species, validate herbal products, and explore their therapeutic potential by identifying key metabolites. Techniques like LC-MS, GC-MS, and metabolomics analyses offer insights into plant variability, aiding in the authentication and quality control of herbal medicines (Wallace et al., 2020; Wang et al., 2023e; Zhang et al., 2023). The UPLC-MS/MS, an untargeted metabolomics approach, renowned for its superior sensitivity and separation efficiency, was employed to profile metabolites (Lommen et al., 2007). Standardized extraction and dilution protocols were used to ensure unbiased analysis. Data processing steps like background subtraction and peak alignment were followed to improve the accuracy and reliability of metabolite identification. Reproducibility was confirmed by overlaying extracted ion chromatogram (EIC) and total ion chromatograms (TICs) analysis in UPLC-MS/MS in negative mode (Supplementary Figures 9, 10) and positive mode, respectively (Supplementary Figures 11, 12), which showed consistent retention times (RT) and peak intensities. The metabolic analysis was performed on 30 *Ferula* samples from 10 cultivated locations, which were then divided into 10 clusters for additional comparison. In total, 7,169 metabolites (Supplementary Table 2) were detected across both ionization modes, with 5,600 metabolites identified in positive ion mode and 1,569 metabolites in negative ion mode through an

untargeted metabolomic approach using liquid chromatography-mass spectrometry (LC-MS). Among these, the metabolites classes in positive and negative ion modes, respectively, included Alcohol and amines (2.5%, 3.87%), Alkaloids (3.21%, 4.49%), Amino acids and derivatives (20.6%, 34.69%), Flavonoids (5.92%, 3.03%), Organic acids (17.03%, 9.98%), Phenolic acids (3.11%, 3.13%), Terpenoids (3.72%, 2.4%), etc. (Figures 4A, B). The accumulation of secondary metabolites varies across plant species (Wang et al., 2016), reflecting their role in plant fitness and adaptation to environmental conditions (Hoffmann et al., 2017). These specialized metabolites exhibit species- or group-specific profiles, yet their biosynthesis can be influenced by environmental factors such as climate (temperature, light, and water) and soil properties (Negrin et al., 2019). Similarly, dynamic nature of metabolite composition was observed in the metabolic profiles across both ionization modes of *Ferula* species (Figures 4A, B). The metabolomic differences identified among the *Ferula* species likely primarily reflect inherent species-specific biochemical pathways with minimal confounding effects from phenotypic plasticity induced by environmental or developmental variation. This interpretation is supported by our controlled sampling design, wherein all leaf tissues were collected from plants at a uniform developmental stage and grown under the same agricultural conditions. Consequently, the distinct metabolic profiles we report can be more confidently attributed to genetic divergence among the studied species. However, we acknowledge a limitation

of this study: while our sampling strategy controlled for macro-environmental factors, it cannot entirely rule out the influence of micro-environmental heterogeneity (e.g., localized differences in soil rhizosphere or light exposure) or unaccounted-for epigenetic modifications that could also contribute to the chemotypic variation. Future studies employing common garden experiments or controlled environment studies would be necessary to definitively disentangle the genetic and environmental components of the observed metabolomic diversity. These results underscore the capacity of the advanced UPLC-MS/MS method to capture extensive chemical information from studied samples, which will be valuable for classification, and understanding the metabolic diversity of *Ferula* species. The detection of a higher number of metabolites in the positive ionization mode compared to the negative mode suggests a preferential ionization of certain compounds. This trend is consistent with findings from other plant metabolomics studies, where compounds such as Alcohol, amines, and amino acids derivatives ionize more efficiently in positive mode, while other metabolite classes, such as certain organic acids and flavonoids, may produce stronger signals in negative mode (Hu et al., 2020; Zhang et al., 2023).

3.3.1 PCA and PLS-DA

Multivariate analysis, including PCA and PLS-DA was performed using UPLC-MS/MS data obtained through XCMS software to assess the heterogeneity of metabolites among *Ferula* accessions and to classify the studied cultivated species of *Ferula*. The PCA (Figure 4) analysis of 30 *Ferula* samples revealed a clear separation into three distinct clades, which were assigned to *F. fukanensis* (YNJDDG, TJZDG, YNJDTL, ZJZ and FKFK), *F. sinkiangensis* (FKXJ), and *F. teterrima* (YJH, TJZTL, SHY and XDX) based on the phylogenetic tree (Figure 3). This clustering closely aligned with the clades identified in the phylogenetic analysis based on plastome sequences, highlighting the consistency between metabolomics and genetic data results. PCA, an unsupervised pattern recognition method, is commonly employed to reduce the dimensionality of complex numerical datasets while preserving essential variance (Farag et al., 2015; Fayek et al., 2021). It is widely used for the classification, quality assessment, and discrimination of herbal medicinal plants. Recent studies have highlighted its effectiveness in distinguishing botanical raw materials like *Panax notoginseng* and authenticating plant materials from various geographical regions using metabolomics data (Zhu et al., 2014; Zhang et al., 2023). The PCA (Figure 4), based on the UPLC-MS/MS dataset, illustrates the clear discrimination of species, as evidenced by the tight clustering of individual replicates within each sample. In the PCA score plot for positive ion mode (Figure 4C), the first two principal components (PC1 and PC2) together account for 45.49% of the total variance. PC1 accounted for 24.66% of the variance, while PC2 contributed 20.83%. This aligns with the studies (Souard et al., 2018) that multivariate statistical analysis involving multiple groups with multiple detection objects can lead to a lower variance explained by the forward PCs. The *Ferula* accessions were categorized into three distinct clades, each reflecting unique metabolic profiles and

highlighting differences among the studied species. In negative ion mode (Figure 4D), the first two principal components (PCs) accounted for 47.55% of the total variation among the samples in the PCA. The PCA plot showed that species including YNJDDG, TJZDG, YNJDTL, and ZJZ clustered closely together, suggesting a high level of metabolic homogeneity. This pattern aligns with the phylogenetic relationships constructed in this study (Figure 3). Despite this homogeneity, FKFK showed a distinct separation in the PCA (Figures 4C, D). However, phylogenetic analysis confirmed that FKFK belongs to the same species, identified as *F. fukanensis*. In contrast, FKXJ exhibited notable metabolic heterogeneity and showed distinct separation from the other groups (Figure 3), aligning with its phylogenetic position (Figure 3) and supporting its classification as a separate species, *F. sinkiangensis*. Additionally, species including YJH, TJZTL, SHY, and XDX demonstrated close clustering in the PCA plot, indicating minimal variation in their metabolite profiles. These results are consistent with phylogenetic analyses, which identified these species as belonging to *F. teterrima*. The results of PCA support the inferred relationships observed for the *Ferula* species based on the phylogenetic relationship (Figure 3). Similarly, employing UPLC-MS metabolomics has been effective in discriminating the diversity of metabolites at the intragenic level among *Vicia* species, revealing potential chemotaxonomic markers for their discrimination (Fayek et al., 2021). Negrin et al. studied the metabolomics of Holly (*Ilex*) species and related taxa in the Aquifoliaceae family, with results demonstrating that *Ilex* species can be differentiated through metabolite profiling (Negrin et al., 2019).

Plant secondary metabolites frequently exhibit taxon-specific chemical structures and biosynthetic pathways, strengthening their utility in taxonomic classification (Endara et al., 2018; Shen et al., 2024). The PLS-DA score plots (Figure 5), a supervised pattern recognition method, was carried out to distinguish among different cultivated *Ferula* species. PLS-DA reveal distinct clustering patterns among taxonomic groups (*Ferula teterrima*: Ft, *Ferula fukanensis*: Ff, and *Ferula sinkiangensis*: Fs), underscoring significant metabolic differentiation driven by taxonomic affiliation through phylogenetic analysis based on plastome data (Figure 3). PLS-DA displays separation with two components of 34.7% and 15.5% variance, isolating XJ from Ft/Ff (Figure 5A) in positive ion mode, while in negative ion mode with the first two components of 30.8% and 16.9% in score plot, revealed separation of all groups. (Figure 5C). These divergent clustering patterns suggest taxonomic-specific metabolic signatures, consistent with prior observations that secondary metabolite profiles correlate strongly with phylogenetic relationships (Figure 3). Furthermore, to decipher the molecular basis of observed separations in our dataset, we prioritized variables using variable importance in projection (VIP) scores derived from PLS-DA. VIP values quantify the relative contribution of individual metabolites to class discrimination, with higher scores indicating stronger discriminatory power. In this study, the top 20 metabolites demonstrated exceptional discriminative capacity, achieving VIP thresholds of >2.5 (positive ionization mode) and >2.0 (negative mode) (Figures 5B, D). The compound annotated as Phe-Gln-Ser was among the top metabolites contributing to the separation of Ft from other species in this dataset (e.g., MW0155167), 1-

Phenylethanol for Fs samples (e.g., MEDP1031) in positive mode, and 4-Methoxyphenyllactic acid for Ff samples in negative mode (e.g., MEDN1331). These results suggest that the metabolic patterns captured by PLS-DA could serve as a reference for distinguishing among Fs, Ff, and Ft in a controlled, exploratory context. However, as no independent test set was used, these models require validation with external samples before any predictive application. Nonetheless, the identification of species-associated metabolites (e.g., Phe-Gln-Ser for Ft, 1-Phenylethanol for Fs) provides strong chemotaxonomic markers that support the genetic distinctions revealed by plastome phylogenomic.

Furthermore, a dendrogram of metabolomic profiles under positive (A) and negative (B) ion modes (Figure 6) was consistent with the PLS-DA findings, confirming species-specific chemotaxonomic separation among Fs, Ff, and Ft samples. In positive ion mode, Fs forms a separate clade within Ft, while Ff shows clear separation from other samples, while Fs and Ft clustered separately with minimal overlap. In negative ion mode, separation remained evident, though Fs and Ff displayed closer proximity, reflecting partial overlap in metabolite composition. The clustering analyses support the robustness of PLS-DA models, demonstrating that cultivated *Ferula* species possess distinct metabolomic fingerprints that align with taxonomic boundaries consistent with previous chemotaxonomic studies (Chilaka et al., 2025).

Beyond serving as a supportive line of evidence for taxonomy, these metabolomic distinctions have direct medicinal and authentication relevance, markers that segregate species correspond to compound classes (sesquiterpene coumarins, sulfur-containing resins) that underpin many reported bioactivities of *Ferula* (antimicrobial, anti-inflammatory, and enzyme-modulating effects) (Jiang et al., 2023; Wariss et al., 2026), so species-level misidentification could meaningfully alter pharmacological profiles of commercial products. The integration of plastome data with untargeted metabolomics provides a more robust framework for species identification and characterization in cultivated *Ferula*. Chloroplast genomes offer robust genetic resolution, while metabolomic profiles capture the biochemical phenotype. This dual approach enables the correlation of genotype with chemotype, facilitates the discovery of diagnostic markers, and supports the selection of high-value chemotypes. Crucially, it demonstrates how metabolic evidence can corroborate and add a functional dimension to phylogenetic hypotheses. Future research should focus on quantitative targeted assays, MS/MS-based metabolite annotation, and dense geographic and phenological sampling in order to put this into practice. This will lay the groundwork for standardized authentication and quality control in pharmaceutical and industrial applications.

3.3.3 Hierarchical cluster analysis (HCA) of *Ferula* species

The HCA (Supplementary Figure 13) illustrates the metabolite profiles of 30 samples across ten groups, revealing distinct patterns of metabolic variation. To further visualize the chemical differences

among the *Ferula* species, a heat map was generated based on metabolic analysis of UPLC-MS/SM in the negative ion mode, representing the metabolic variation across the 30 samples (Supplementary Figure 13). This approach highlights the species-specific metabolic signatures and provides a comprehensive view of the chemical diversity within the *Ferula* species. The Z-scores of metabolites were used to normalize and visualize the intensity data, with warmer colors (red/orange) indicating higher metabolite levels and cooler colors (green) representing lower levels. The samples clustered both by rows (metabolites) and columns (sample groups), reflecting similarities and differences in metabolic composition across groups.

HCA is an exploratory analysis that classifies variables by clustering samples with similar features, and separating samples with different features (Cheng et al., 2020; Shen et al., 2024). The HCA (Supplementary Figure 13) reveals three major clusters, highlighting distinct patterns of metabolic variation among the samples. The species TJZDG, ZJZ, YNJDDG, and YNJDTL formed a distinct clade, indicating high metabolic homogeneity and supporting their identification as *F. fukanensis*. This clustering pattern aligns with their proximity to the PCA plot (Figure 4) and is consistent with the phylogenetic relationships constructed in this study (Figure 3). In contrast, FKFK and FKXJ with the other four species formed two distinct clusters, highlighting the unique metabolic profile of these species. The HCA further identified a sister clade consisting of FKXJ, YJH, TJZTL, SHY, and XDX, indicating metabolic homogeneity within this subgroup. Notably, FKXJ is a sister species within this clade, revealing its distinct metabolic profile. This clustering pattern supports the recognition of these samples as distinct species identified as *F. teterrima* (YJH, TJZTL, SHY, and XDX) and *F. sinkiangensis* (FKXJ), respectively.

These findings support the coherence between the metabolomic profiles and the PCA and PLS-DA results, emphasizing the distinct chemical signatures and evolutionary relationships among the *Ferula* species. The structure of secondary metabolites and their biosynthetic pathways are often specific and restricted to taxonomically related organisms, making them useful for classification (Endara et al., 2018; García-Pérez et al., 2024). Phytochemical profiles were used to assess the Chemotaxonomic relationship of *Ferula* species reflecting underlying genetic morphology and its interactions with the environmental factors (Zhang et al., 2023; He et al., 2025). Metabolomic fingerprinting contributes to the classification of many crop plant species including date palm fruit (Farag et al., 2014) and *Vicia* species (Fayek et al., 2021). For instance, Metabolomics, combined with multivariate analyses, has proven effective in discriminating between *Camellia sinensis* (Ku et al., 2010), *Ilex* (Negrin et al., 2019) and blueberry samples (Ma et al., 2013).

4 Conclusion

Plant derived drugs have long faced challenges due to limited natural resources. Currently, artificial cultivation is the most

practical approach to ensure a reliable supply of raw materials for natural medicines. However, challenges remain for multi-source medicinal materials like *Ferula*, used for culinary and medicinal purposes. Traditional morphological methods often fall short, especially when only vegetative parts are available. This is the first study to integrate morphological, genetic, and metabolic fingerprinting using a metabolome-based approach (UPLC-MS/MS) unified with chemometric tools such as HCA, PCA, and PLS-DA, to systematically delineate cultivated *Ferula* species in Xinjiang. Phylogenomic analysis based on complete plastome sequences clearly classified the studied accessions into three distinct clades: *F. sinkiangensis*, *F. fukanensis*, and *F. teterrima*. Untargeted UPLC-MS/MS metabolomic profiling further validated this classification, revealing significant interspecific differences in metabolite composition, with clustering patterns that strongly aligned with phylogenetic relationships. Our findings demonstrate that integrating phylogenomics with complementary metabolomic profiling provides a robust framework for species discrimination in cultivated *Ferula*, overcoming the limitations of traditional identification methods. Specifically, the metabolomic data offer chemotaxonomic support for the phylogenetic distinctions, aiding in species differentiation. The established plastome reference dataset also serves as a valuable resource for future germplasm monitoring and trade tracking. Collectively, this study establishes an integrated chemotaxonomic foundation for the standardized authentication and quality control of *Ferula* species and offers a replicable strategy for the classification and sustainable utilization of other medicinal plants.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repository and accession number(s) can be found in the article/[Supplementary Material](#).

Author contributions

HMW: Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft. LY: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. JG: Data curation, Investigation, Writing – review & editing. DL: Writing – review & editing. WL: Conceptualization, Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation, Writing – review & editing.

References

Ajani, Y., Ajani, A., Cordes, J. M., Watson, M. F., and Downie, S. R. (2008). Phylogenetic analysis of nrDNA ITS sequences reveals relationships within five groups of Iranian Apiaceae subfamily Apioideae. *Taxon*. 57, 383–401. doi: 10.2307/25066011

Funding

The author(s) declared that financial support was received for this work and/or its publication. This study was supported by grants from the Natural Science Foundation of Xinjiang Uygur Autonomous Region (2025D01E56), the Western Young Scholars Project of the Chinese Academy of Sciences (2022-XBQNXZ-003), the Tianshan Youth Program of Xinjiang Uygur Autonomous Region (2022TSYCCX0088), and the Capacity-building project of strategic biological resources program of the Chinese Academy of Sciences (KFJ-BRP-017-089).

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2026.1754844/full#supplementary-material>

Alberts, P. S. F., and Meyer, J. J. M. (2022). Integrating chemotaxonomic-based metabolomics data with DNA barcoding for plant identification: A case study on south-east African Erythroxylaceae species. *South. S. Afr. J. Bot.* 146, 174–186. doi: 10.1016/j.sajb.2021.10.005

- Bagheri, S. M., Hakimzadeh, E., and Allahtavakoli, M. (2024). A review on antidiabetic potential of genus *Ferula* (Apiaceae). *Curr. Tradit. Med.* 10, E090323214509. doi: 10.2174/221508380966230309111323
- Bakhsipour, M., Mafakheri, M., Kordrostami, M., Zakir, A., Rahimi, N., Feizi, F., et al. (2019). *In vitro* multiplication, genetic fidelity and phytochemical potentials of *Vaccinium arctostaphylos* L.: An endangered medicinal plant. *Indus. Crops Prod.* 141, 111812. doi: 10.1016/j.indcrop.2019.111812
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Bio.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Boghrati, Z., and Iranshahi, M. (2019). *Ferula* species: A rich source of antimicrobial compounds. *J. Herb. Med.* 16, 100244. doi: 10.1016/j.hermed.2018.10.009
- Buck, M., and Hamilton, C. (2011). The Nagoya Protocol on access to genetic resources and the fair and equitable sharing of benefits arising from their utilization to the Convention on Biological Diversity. *Rev. Eur. Community Int. Environ. Law* 20, 47–61. doi: 10.1111/j.1467-9388.2011.00703.x
- CBOL Plant Working Group, Hollingsworth, P. M., Forrest, L. L., Spouge, J. L., Hajibabaei, M., Ratnasingham, S., van der Bank, M., et al. (2009). A DNA barcode for land plants. *Proc. Natl. Acad. Sci.* 106, 12794–12797. doi: 10.1073/pnas.0905845106
- Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Chase, M. W., Salamin, N., Wilkinson, M., Dunwell, J. M., Kesanakurthi, R. P., Haidar, N., et al. (2005). Land plants and DNA barcodes: short-term and long-term goals. *P. R. Soc. B-Biol. Sci.* 360, 1889–1895. doi: 10.1098/rstb.2005.1720
- Chen, Y. P., Drew, B. T., Li, B., Soltis, D. E., Soltis, P. S., and Xiang, C. L. (2016). Resolving the phylogenetic position of *Ombrocharis* (Lamiaceae), with reference to the molecular phylogeny of tribe Elsholtzieae. *Taxon*. 65, 123–136. doi: 10.12705/651.8
- Chen, W., Gong, L., Guo, Z., Wang, W., Zhang, H., Liu, X., et al. (2013). A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: Application in the study of rice metabolomics. *Mol. Plant* 6, 1769–1780. doi: 10.1093/mp/ss080
- Cheng, Z., Shu, H., Zhang, S., Luo, B., Gu, R., Zhang, R., et al. (2020). From folk taxonomy to species confirmation of *Acorus* (Acoraceae): Evidences based on phylogenetic and metabolomic analyses. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00965
- Chervin, J., Talou, T., Audonnet, M., Dumas, B., Camborde, L., Esquerré-Tugayé, M. T., et al. (2019). Deciphering the phylogeny of violets based on multiplexed genetic and metabolomic approaches. *Phytochem.* 163, 99–110. doi: 10.1016/j.phytochem.2019.04.001
- Chilaka, C. A., Aparicio-Muriana, M. D. M., Petchkongkaew, A., Quinn, B., Birse, N., and Elliott, C. T. (2025). A combined elementomics, metabolomics, and chemometrics approach as tools to identify the geographic origins of black pepper. *Food Chem.* 492, 145420. doi: 10.1016/j.foodchem.2025.145420
- Chong, X., Li, Y., Yan, M., Wang, Y., Li, M., Zhou, Y., et al. (2022). Comparative chloroplast genome analysis of 10 *Ilex* species and the development of species-specific identification markers. *Indus. Crops Prod.* 187, 115408. doi: 10.1016/j.indcrop.2022.115408
- Costa, F. O., de Waard, J. R., Boutillier, J., Ratnasingham, S., Dooh, R. T., Hajibabaei, M., et al. (2007). Biological identifications through DNA barcodes: the case of the Crustacea. *Can. J. Fish. Aquat. Sci.* 64, 272–295. doi: 10.1139/F07-008
- Cowan, R. S., Chase, M. W., Kress, W. J., and Savolainen, V. (2006). 300,000 species to identify: problems, progress, and prospects in DNA barcoding of land plants. *Taxon*. 55, 611–616. doi: 10.2307/25065638
- Des Roches, S., Post, D. M., Turley, N. E., Bailey, J. K., Hendry, A. P., Kinnison, M. T., et al. (2018). The ecological importance of intraspecific variation. *Nat. Ecol. Evol.* 2, 57–64. doi: 10.1038/s41559-017-0402-5
- Dewar, A. E., Belcher, L. J., and West, S. A. (2025). A phylogenetic approach to comparative genomics. *Nat. Rev. Genet.* 26, 395–405. doi: 10.1038/s41576-024-00803-0
- Dodsworth, S. (2015). Genome skimming for next-generation biodiversity analysis. *Trends Plant Sci.* 20, 525–527. doi: 10.1016/j.tplants.2015.06.012
- Downie, S. R., Spalik, K., Katz-Downie, D. S., and Reduron, J.-P. (2010). Major clades within Apiaceae subfamily Apioideae as inferred by phylogenetic analysis of nrDNA ITS sequences. *Plant Divers. Evol.* 128, 111. doi: 10.1127/1869-6155/2010/0128-0005
- Drew, B. T., González-Gallegos, J. G., Xiang, C. L., Kriebel, R., Drummond, C. P., Walker, J. B., et al. (2017). *Salvia* united: The greatest good for the greatest number. *Taxon*. 66, 133–145. doi: 10.12705/661.7
- Duan, L. X., Chen, T. L., Li, M., Chen, M., Zhou, Y. Q., Cui, G. H., et al. (2012). Use of the metabolomics approach to characterize chinese medicinal material Huangqi. *Mol. Plant* 5, 376–386. doi: 10.1093/mp/ssr093
- Duminil, J., and Di Michele, M. (2009). Plant species delimitation: A comparison of morphological and molecular markers. *Plant Biosyst.* 143, 528–542. doi: 10.1080/11263500902722964
- Du Preez, B., Schrire, B. D., Dreyer, L. L., Stirton, C. H., Chimphango, S. B. M., and Muasya, A. M. (2025). Global biogeographic patterns of the genus *Indigofera* (Fabaceae: Indigoferaceae). *Braz. J. Bot.* 48, 19. doi: 10.1007/s40415-024-01045-4
- Endara, M. J., Coley, P. D., Wiggins, N. L., Forrister, D. L., Younkin, G. C., Nicholls, J. A., et al. (2018). Chemocoding as an identification tool where morphological- and DNA-based methods fall short: Inga as a case study. *New Phytol.* 218, 847–858. doi: 10.1111/nph.15020
- Farag, M. A., Mohsen, M., Heinke, R., and Wessjohann, L. A. (2014). Metabolomic fingerprints of 21 date palm fruit varieties from Egypt using UPLC/PDA/ESI-qTOF-MS and GC-MS analyzed by chemometrics. *Food Res. Int.* 64, 218–226. doi: 10.1016/j.foodres.2014.06.021
- Farag, M. A., Sakna, S. T., El-fiky, N. M., Shabana, M. M., and Wessjohann, L. A. (2015). Phytochemical, antioxidant and antidiabetic evaluation of eight *Bauhinia* L. species from Egypt using UHPLC-PDA-qTOF-MS and chemometrics. *Phytochem.* 119, 41–50. doi: 10.1016/j.phytochem.2015.09.004
- Fayek, N. M., Mekky, R. H., Dias, C. N., Kropf, M., Heiss, A. G., Wessjohann, L. A., et al. (2021). UPLC-MS metabolome-based seed classification of 16 *Vicia* species: A prospect for phyto-equivalency and chemotaxonomy of different accessions. *J. Agric. Food. Chem.* 69, 5252–5266. doi: 10.1021/acs.jafc.0c06054
- Feng, Z., Fan, K., Yao, Z., Wang, H., Wu, X., Tang, L., et al. (2024). Integrative metabolome and transcriptome analysis reveals GbKCS and GbMYB involved in the biosynthesis of ginkgolic acids. *Indus. Crops Prod.* 220, 119225. doi: 10.1016/j.indcrop.2024.119225
- Gallagher, R., Roger, E., Packer, J., Slatyer, C., Rowley, J., Cornwell, W., et al. (2025). Incorporating citizen science into IUCN Red List assessments. *Conser. Bio.* 39, e14329. doi: 10.1111/cobi.14329
- Gallon, M. E., Monge, M., Casoti, R., Da Costa, F. B., Semir, J., and Gobbo-Neto, L. (2018). Metabolomic analysis applied to chemosystematics and evolution of megadiverse Brazilian *Vernonieae* (Asteraceae). *Phytochem.* 150, 93–107. doi: 10.1016/j.phytochem.2018.03.007
- García-Pérez, P., Becchi, P. P., Zhang, L., Rocchetti, G., and Lucini, L. (2024). Metabolomics and chemometrics: The next-generation analytical toolkit for the evaluation of food quality and authenticity. *Trends Food Sci. Technol.* 147, 104481. doi: 10.1016/j.tifs.2024.104481
- Guo, X., Shi, N., Xie, P., Zhang, G., Liu, H., and Ji, Y. (2022). Plastome sequencing for accurate and effective authentication of *Polygonatum kingianum* (Asparagaceae). *Indus. Crops Prod.* 184, 115056. doi: 10.1016/j.indcrop.2022.115056
- He, M., Han, X., Qin, X., Bao, J., Li, H., Xie, Q., et al. (2024). Comparative chloroplast genome analyses provide new insights into phylogeny of *Taraxacum* and molecular markers for distinguishing rubber producing dandelions from their weedy relatives in China. *Indus. Crops Prod.* 207, 117712. doi: 10.1016/j.indcrop.2023.117712
- He, G., Man, J., Chen, Y., Zhang, X., Wang, X., An, K., et al. (2025). Identification of *Salvia miltiorrhiza* germplasm resources based on the metabolomics and DNA barcoding. *Front. Pharmacol.* 15. doi: 10.3389/fphar.2024.1518906
- Hebert, P. D. N., Cywinska, A., Ball, S. L., and DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *P. R. Soc. B-Biol. Sci.* 270, 313–321. doi: 10.1098/rspb.2002.2218
- Hoffmann, J. F., Carvalho, I. R., Barbieri, R. L., Rombaldi, C. V., and Chaves, F. C. (2017). *Butia* spp. (Arecaceae) LC-MS-based metabolomics for species and geographical origin discrimination. *J. Agric. Food. Chem.* 65, 523–532. doi: 10.1021/acs.jafc.6b03203
- Holman, J. D., Tabb, D. L., and Mallick, P. (2014). Employing ProteoWizard to convert raw mass spectrometry data. *Curr. Protoc. Bioinf.* 46, 13.24.11–13.24.19. doi: 10.1002/0471250953.bi1324s46
- Hu, Y. Y., Zhang, X. J., Zhang, Z. H., Wang, Z., and Tan, N. H. (2020). Qualitative and quantitative analyses of quinones in multi-origin *Rubia* species by ultra-performance liquid chromatography-tandem mass spectrometry combined with chemometrics. *J. Pharm. Biomed. Anal.* 189, 113471. doi: 10.1016/j.jpba.2020.113471
- Jiang, M., Lan, S., Peng, M., Wang, Z., and Zhuang, L. (2023). The diversity of *Ferula* species and environmental factors on metabolite composition using untargeted metabolomics. *Food Biosci.* 56, 103075. doi: 10.1016/j.fbio.2023.103075
- Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., dePamphilis, C. W., Yi, T. S., et al. (2020). GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Bio.* 21, 241. doi: 10.1186/s13059-020-02154-5
- Kalyanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285
- Kaplan, D. R. (2001). The science of plant morphology: definition, history, and role in modern biology. *Am. J. Bot.* 88, 1711–1741. doi: 10.2307/3558347
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199
- Khayat, M. T., Alharbi, M., Ghazawi, K. F., Mohamed, G. A., and Ibrahim, S. R. M. (2023). *Ferula sinkiangensis* (Chou-AWei, Chinese *Ferula*): Traditional uses, phytoconstituents, biosynthesis, and pharmacological activities. *Plants*. 12, 902. doi: 10.3390/plants12040902

- Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., et al. (2025). PubChem 2025 update. *Nucleic Acids Res.* 53, D1516–d1525. doi: 10.1093/nar/gkai1059
- Korovin, E., Pimenov, M., and Kinkizaeva, G. (1951). Genus ferula L. Flora of the USSR. Ed. Academy of sciences of the USSR. ML62-155.
- Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A., and Janzen, D. H. (2005). Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci.* 102, 8369–8374. doi: 10.1073/pnas.0503123102
- Ku, K. M., Choi, J. N., Kim, J., Kim, J. K., Yoo, L. G., Lee, S. J., et al. (2010). Metabolomics analysis reveals the compositional differences of shade grown tea (*Camellia sinensis* L.). *J. Agric. Food. Chem.* 58, 418–426. doi: 10.1021/jf902929h
- Kurzyna-Młynik, R., Oskolski, A. A., Downie, S. R., Kopacz, R., Wojewódzka, A., and Spalik, K. (2008). Phylogenetic position of the genus *Ferula* (Apiaceae) and its placement in tribe Scandiceae as inferred from nrDNA ITS sequence variation. *Plant Syst. Evol.* 274, 47–66. doi: 10.1007/s00606-008-0022-2
- Li, L., He, L., Su, X., Amu, H., Li, J., and Zhang, Z. (2022). Chemotaxonomy of *Aster* species from the Qinghai-Tibetan Plateau based on metabolomics. *Phytochem. Anal.* 33, 23–39. doi: 10.1002/pca.3058
- Li, Z., Jiang, H., Jiang, X., Zhang, L., and Qin, Y. (2023). Integrated physiological, transcriptomic, and metabolomic analyses reveal that low-nitrogen conditions improve the accumulation of flavonoids in snow *Chrysanthemum*. *Indus. Crops Prod.* 197, 116574. doi: 10.1016/j.indcrop.2023.116574
- Li, H., Lv, Q., Liu, A., Wang, J., Sun, X., Deng, J., et al. (2022). Comparative metabolomics study of Tartary (*Fagopyrum tataricum* (L.) Gaertn) and common (*Fagopyrum esculentum* Moench) buckwheat seeds. *Food Chem.* 371, 131125. doi: 10.1016/j.foodchem.2021.131125
- Li, W., Su, Z., Yang, L., Cao, Q., and Feng, Y. (2020). Genetic diversity of the critically endangered *Ferula sinkiangensis* KM Shen (Apiaceae) and the implications for conservation. *Turk. J. Bot.* 44, 145–152. doi: 10.15666/aer/2306_1175511768
- Li, H., Xu, C., Zhou, S., Huang, S., Wu, Z., Jiangfang, Y., et al. (2024). Integrated metabolomic and transcriptomic analysis of flavonoid biosynthesis in *Ricinus communis* L. *Indus. Crops Prod.* 222, 119684. doi: 10.1016/j.indcrop.2024.119684
- Liu, K., Guo, C., Ren, Y., Bai, Y., Zhang, Z., and Dong, W. (2025a). Phylogenomics and evolution of the *Acer* section *Lithocarpa*. *Scient. Horticul.* 341, 113988. doi: 10.1016/j.scienta.2025.113988
- Liu, D. H., Liu, Q. R., Tojibaev, K. S., Sukhorukov, A. P., Wariss, H. M., Zhao, Y., et al. (2025b). Phylogenomics provides new insight into the phylogeny and diversification of Asian *Lappula* (Boraginaceae). *Mol. Phyl. Evol.* 208, 108361. doi: 10.1016/j.jmpev.2025.108361
- Liu, Z. F., Ma, H., Ci, X. Q., Li, L., Song, Y., Liu, B., et al. (2021). Can plastid genome sequencing be used for species identification in Lauraceae? *Bot. J. Linn. Soc* 197, 1–14. doi: 10.1093/botlinnean/boab018
- Liu, C., Shi, L., Zhu, Y., Chen, H., Zhang, J., Lin, X., et al. (2012). CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genom.* 13, 715. doi: 10.1186/1471-2164-13-715
- Lommen, A., van der Weg, G., van Engelen, M. C., Bor, G., Hoogenboom, L. A. P., and Nielen, M. W. F. (2007). An untargeted metabolomics approach to contaminant analysis: Pinpointing potential unknown compounds. *Anal. Chim. Acta* 584, 43–49. doi: 10.1016/j.aca.2006.11.018
- Lv, S. Y., Ye, X. Y., Li, Z. H., Ma, P. F., and Li, D. Z. (2023). Testing complete plastomes and nuclear ribosomal DNA sequences for species identification in a taxonomically difficult bamboo genus *Fargesia*. *Plant Divers.* 45, 147–155. doi: 10.1016/j.pld.2022.04.002
- Ma, C., Dastmalchi, K., Flores, G., Wu, S.-B., Pedraza-Peñalosa, P., Long, C., et al. (2013). Antioxidant and metabolite profiling of North American and Neotropical blueberries using LC-TOF-MS and multivariate analyses. *J. Agric. Food. Chem.* 61, 3548–3559. doi: 10.1021/jf400515g
- Masson, J., Liberto, E., Brevard, H., Bicchi, C., and Rubiolo, P. (2014). A metabolomic approach to quality determination and authentication of raw plant material in the fragrance field. *Iris rhizomes: A case study. J. Chromatogr. A.* 1368, 143–154. doi: 10.1016/j.chroma.2014.09.076
- Mohammadhosseini, M., Venditti, A., Sarker, S. D., Nahar, L., and Akbarzadeh, A. (2019). The genus *Ferula*: Ethnobotany, phytochemistry and bioactivities - A review. *Indus. Crops Prod.* 129, 350–394. doi: 10.1016/j.indcrop.2018.12.012
- Negrin, A., Long, C., Motley, T. J., and Kennelly, E. J. (2019). LC-MS metabolomics and chemotaxonomy of caffeine-containing holly (*Ilex*) species and related taxa in the Aquifoliaceae. *J. Agric. Food. Chem.* 67, 5687–5699. doi: 10.1021/acs.jafc.8b07168
- Nguyen, H. T., Lee, D. K., Lee, W. J., Lee, G., Yoon, S. J., Shin, B. K., et al. (2016). UPLC-QTOFMS based metabolomics followed by stepwise partial least square-discriminant analysis (PLS-DA) explore the possible relation between the variations in secondary metabolites and the phylogenetic divergences of the genus *Panax*. *J. Chromatogr. B.* 1012-1013, 61–68. doi: 10.1016/j.jchromb.2016.01.002
- Nock, C. J., Waters, D. L. E., Edwards, M. A., Bowen, S. G., Rice, N., Cordeiro, G. M., et al. (2011). Chloroplast genome sequences from total DNA for plant identification. *Plant Biotechnol. J.* 9, 328–333. doi: 10.1111/j.1467-7652.2010.00558.x
- Panahi, M., Banasiak, L., Piwczynski, M., Puchałka, R., Kanani, M. R., Oskolski, A. A., et al. (2018). Taxonomy of the traditional medicinal plant genus *Ferula* (Apiaceae) is confounded by incongruence between nuclear rDNA and plastid DNA. *Bot. J. Linn. Soc* 188, 173–189. doi: 10.1093/botlinnean/boy055
- Panahi, M., Banasiak, L., Piwczynski, M., Puchałka, R., Oskolski, A. A., and Spalik, K. (2015). Phylogenetic relationships among *Dorema*, *Ferula* and *Leutea* (Apiaceae: Scandiceae: Ferulinae) inferred from nrDNA ITS and cpDNA noncoding sequences. *Taxon.* 64, 770–783. doi: 10.12705/644.8
- Pimenov, M. G., and Leonov, M. V. E. (1993). *The genera of the Umbelliferae: a nomenclator*. Kew, UK: Royal Botanic Gardens. doi: 10.2307/4114493
- Rambaut, A. (2014). *FigTree 1.4. 2 software* (UK: Institute of evolutionary biology, Univ. Edinburgh).
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029
- Safina, L., and Pimenov, M. (1983). The carpoanatomical features of the species of the genus *Ferula* of the subgenus *Peucedanoides* (Apiaceae) in connection with the systematics of the genus. *Bot. Zhurn. (Leningrad)*. 68, 730–739. doi: 10.1097/01.ftd.0000179845.53213.39
- Salehi, M., Naghavi, M. R., and Bahmankar, M. (2019). A review of *Ferula* species: Biochemical characteristics, pharmaceutical and industrial applications, and suggestions for biotechnologists. *Indus. Crops Prod.* 139, 111511. doi: 10.1016/j.indcrop.2019.111511
- Shaw, R. E., Farquharson, K. A., Bruford, M. W., Coates, D. J., Elliott, C. P., Mergeay, J., et al. (2025). Global meta-analysis shows action is needed to halt genetic diversity loss. *Nature* 638, 704–710. doi: 10.1038/s41586-024-08458-x
- Sheh, M., Pu, F., Pan, Z., Watson, M. F., Cannon, J., Holmes-Smith, I., et al. (2005). *Apiaceae* (Umbelliferae). *Flora of China* 14, 1–205.
- Shen, G. M. (2011). *Apiaceae (Umbelliferae)* (Urumqi: Xinjiang science & Technology Publishing House).
- Shen, J., Li, P., Chu, H., Li, Y., Meng, X., Li, Z., et al. (2024). Pharmacophylogenetic insights into *Scutellaria strigillosa* Hemsl.: chloroplast genome and untargeted metabolomics, quantitative analysis and antibacterial analysis. *Front. Plant Sci.* 15. doi: 10.3389/fpls.2024.1472204
- Shen, J., Li, P., Wang, Y., Yang, K., Li, Y., Yao, H., et al. (2022). Pharmacophylogenetic study of *Scutellaria baicalensis* and its substitute medicinal species based on the chloroplast genomics, metabolomics, and active ingredient. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.951824
- Shen, Z., Lu, T., Zhang, Z., Cai, C., Yang, J., and Tian, B. (2019). Authentication of traditional Chinese medicinal herb “Gusuibu” by DNA-based molecular methods. *Indus. Crops Prod.* 141, 111756. doi: 10.1016/j.indcrop.2019.111756
- Shoaib, T. H., Abdelmoniem, N., Mukhtar, R. M., Alqhtani, A. T., Alalawi, A. L., Alawaji, R., et al. (2023). Molecular docking and molecular dynamics studies reveal the anticancer potential of medicinal-plant-derived lignans as MDM2-P53 interaction inhibitors. *Molecules* 28, 6665. doi: 10.3390/molecules28186665
- Smith, C. A., Maille, G. O., Want, E. J., Qin, C., Trauger, S. A., Brandon, T. R., et al. (2005). METLIN: A metabolite mass spectral database. *Therap. Drug Monit.* 27, 747–751.
- Sonigra, P., and Meena, M. (2021). Metabolic profile, bioactivities, and variations in the chemical constituents of essential oils of the *Ferula* genus (Apiaceae). *Front. Pharmacol.* 11. doi: 10.3389/fphar.2020.608649
- Souard, F., Delporte, C., Stoffelen, P., Thévenot, E. A., Noret, N., Dauvergne, B., et al. (2018). Metabolomics fingerprint of coffee species determined by untargeted-profiling study using LC-HRMS. *Food Chem.* 245, 603–612. doi: 10.1016/j.foodchem.2017.10.022
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Tan, W., Gao, H., Jiang, W., Zhang, H., Yu, X., Liu, E., et al. (2020). The complete chloroplast genome of *Gleditsia sinensis* and *Gleditsia japonica*: genome organization, comparative analysis, and development of taxon specific DNA mini-barcodes. *Sci. Rep.* 10, 16309. doi: 10.1038/s41598-020-73392-7
- Wallace, E. D., Todd, D. A., Harnly, J. M., Cech, N. B., and Kellogg, J. J. (2020). Identification of adulteration in botanical samples with untargeted metabolomics. *Anal. Bioanal. Chem.* 412, 4273–4286. doi: 10.1007/s00216-020-02678-6
- Wang, J., Huo, X., Wang, H., Dong, A., Zheng, Q., and Si, J. (2023a). Undescribed sesquiterpene coumarins from the aerial parts of *Ferula sinkiangensis* and their anti-inflammatory activities in lipopolysaccharide-stimulated RAW 264.7 macrophages. *Phytochem.* 210, 113664. doi: 10.1016/j.phytochem.2023.113664
- Wang, W., Jiang, Y., Chen, Y., Luo, W., He, D., Wang, Y., et al. (2022). Using intraspecific variation of functional traits and environmental factors to understand the formation of nestedness patterns of a local forest community. *J. Plant Ecol.* 15, 1185–1198. doi: 10.1093/jpe/rtac039
- Wang, J. L., Sang, C. Y., Wang, J., Li, P. L., Chai, T., Naghavi, M. R., et al. (2023b). Sesquiterpene coumarins from *Ferula sinkiangensis* and their anti-pancreatic cancer effects. *Phytochem.* 214, 113824. doi: 10.1016/j.phytochem.2023.113824

- Wang, D., Sun, Y., Liu, Q., Ye, C., Zhao, S., and Zhang, H. (2023c). Ferula sinkiangensis against gastric cancer: a network pharmacology, molecular docking and cell experiment study. *Transl. Cancer Res.* 12, 743–764. doi: 10.21037/tcr-22-2292
- Wang, S., Tu, H., Wan, J., Chen, W., Liu, X., Luo, J., et al. (2016). Spatio-temporal distribution and natural variation of metabolites in citrus fruits. *Food Chem.* 199, 8–17. doi: 10.1016/j.foodchem.2015.11.113
- Wang, X., Wang, Z., Jiang, P., He, Y., Mu, Y., Lv, X., et al. (2018). Bacterial diversity and community structure in the rhizosphere of four Ferula species. *Sci. Rep.* 8, 5345. doi: 10.1038/s41598-018-22802-y
- Wang, J., Wang, H., Zhang, M., Li, X., Zhao, Y., Chen, G., et al. (2020). Sesquiterpene coumarins from Ferula sinkiangensis K.M.Shen and their cytotoxic activities. *Phytochem.* 180, 112531. doi: 10.1016/j.phytochem.2020.112531
- Wang, J., Zheng, Q., Wang, H., Shi, L., Wang, G., Zhao, Y., et al. (2023d). Sesquiterpenes and sesquiterpene derivatives from Ferula: Their chemical structures, biosynthetic pathways, and biological properties. *Antioxidants*. 13. doi: 10.3390/antiox13010007
- Wang, S., Zhou, N., Shi, N., Zhang, G., Liu, H., Guo, X., et al. (2023e). Testing and using complete plastomes for authentication of medicinal Polygonatum species (Asparagaceae). *Indus. Crops Prod.* 197, 116557. doi: 10.1016/j.indcrop.2023.116557
- Wariss, H. M., Liu, T., Zhang, H., Wu, J., Yang, Z., and Li, W. (2025). Genetic diversity and population structure of the endangered medicinal plant Ferula sinkiangensis. *Glob. Ecol. Conser.* 58, e03437. doi: 10.1016/j.gecco.2025.e03437
- Wariss, H. M., Yang, L., Ahmad, S., Yaseen, A., Riaz, M., Tariq, A., et al. (2026). The ethnomedicinal, pharmacological, and phytochemical potential of Ferula sinkiangensis K.M.Shen: An updated and comprehensive review. *J. Ethnopharmacol.* 354, 120444. doi: 10.1016/j.jep.2025.120444
- Wick, R. R., Schultz, M. B., Zobel, J., and Holt, K. E. (2015). Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 31, 3350–3352. doi: 10.1093/bioinformatics/btv383
- Wu, J., Liu, D., Wariss, H. M., Zhang, H., Su, M., Li, W., et al. (2025). Genetic diversity and construction of core collection provides new insight for the conservation of edible Allium galanthum in Xinjiang. *Sci. Hortic.* 341, 113961. doi: 10.1016/j.scienta.2025.113961
- Xue, T. T., Janssens, S. B., Liu, B. B., and Yu, S. X. (2024). Phylogenomic conflict analyses of the plastid and mitochondrial genomes via deep genome skimming highlight their independent evolutionary histories: A case study in the cinquefoil genus Potentilla sensu lato (Potentilleae, Rosaceae). *Mol. Phyl. Evol.* 90, 107956. doi: 10.1016/j.ympev.2023.107956
- Yang, L., Abduraimov, O., Tojibaev, K., Shomurodov, K., Zhang, Y. M., and Li, W. J. (2022). Analysis of complete chloroplast genome sequences and insight into the phylogenetic relationships of Ferula L. *BMC Genom.* 23, 643. doi: 10.1186/s12864-022-08868-z
- Yang, J. R., An, Z., Li, Z. H., Jing, S., and Qina, H. L. (2006). Sesquiterpene coumarins from the roots of Ferula sinkiangensis and Ferula teterrima. *Chem. Pharm. Bull.* 54, 1595–1598. doi: 10.1002/chin.200717175
- Yang, J. B., Tang, M., Li, H. T., Zhang, Z. R., and Li, D. Z. (2013). Complete chloroplast genome of the genus Cymbidium: lights into the species identification, phylogenetic implications and population genetic analyses. *BMC Evol. Bio.* 13, 1–12. doi: 10.21203/rs.3.rs-2284029/v1
- Zhang, D., Gao, F., Jakovic, I., Zou, H., Zhang, J., Li, W. X., et al. (2020b). PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Resour.* 20, 348–355. doi: 10.1111/1755-0998.13096
- Zhang, L., Sun, X., Si, J., Li, G., and Cao, L. (2019). Umbelliprenin isolated from Ferula sinkiangensis inhibits tumor growth and migration through the disturbance of Wnt signaling pathway in gastric cancer. *PloS One* 14, e0207169. doi: 10.1371/journal.pone.0207169
- Zhang, T., Wang, Z., Lv, X., Dang, H., and Zhuang, L. (2020a). Variation of rhizosphere bacterial community diversity in the desert ephemeral plant Ferula sinkiangensis across environmental gradients. *Sci. Rep.* 10, 18442. doi: 10.1038/s41598-020-75591-8
- Zhang, X., Yang, L. E., Hu, Y., Wu, X., Wang, Z., Miao, Y., et al. (2023). Integrating morphology, molecular phylogeny and chemotaxonomy for the most effective authentication in Chinese Rubia with insights into origin and distribution of characteristic Rubiaceae-type cyclopeptides. *Indus. Crops Prod.* 191, 115775. doi: 10.1016/j.indcrop.2022.115775
- Zhao, Y., Zhao, F., Salmaki, Y., Paton, A., Cai, J., Drew, B. T., et al. (2023). Home at last: Molecular phylogenetic analyses support subsuming *Metastachydium* within *Phlomis* (Lamiaceae). *Taxon* 72, 590–606. doi: 10.1002/tax.12935
- Zhao, Y., Yu, D., Kuo, W., Huang, J., Guo, J., Sun, M., et al. (2025). Nuclear phylogenomics provide evidence to clarify key morphological evolution and whole-genome duplication across rosids. *J. Integ. Plant Bio.* 67, 2704–2730. doi: 10.1111/jipb.13972
- Zhao, Y., Zhao, Y.-Y., Du, Y., and Kang, J. S. (2019). Characterization and classification of three common Bambusoideae species in Korea by an HPLC-based analytical platform coupled with multivariate statistical analysis. *Indus. Crops Prod.* 130, 389–397. doi: 10.1016/j.indcrop.2018.12.096
- Zhou, Z., Luo, M., Zhang, H., Yin, Y., Cai, Y., and Zhu, Z. J. (2022). Metabolite annotation from knowns to unknowns through knowledge-guided multi-layer metabolic networking. *Nat. Commun.* 13, 6656. doi: 10.1038/s41467-022-34537-6
- Zhu, J., Fan, X., Cheng, Y., Agarwal, R., Moore, C. M. V., Chen, S. T., et al. (2014). Chemometric analysis for identification of botanical raw materials for pharmaceutical use: A case study using Panax notoginseng. *PloS One* 9, e87462. doi: 10.1371/journal.pone.0087462