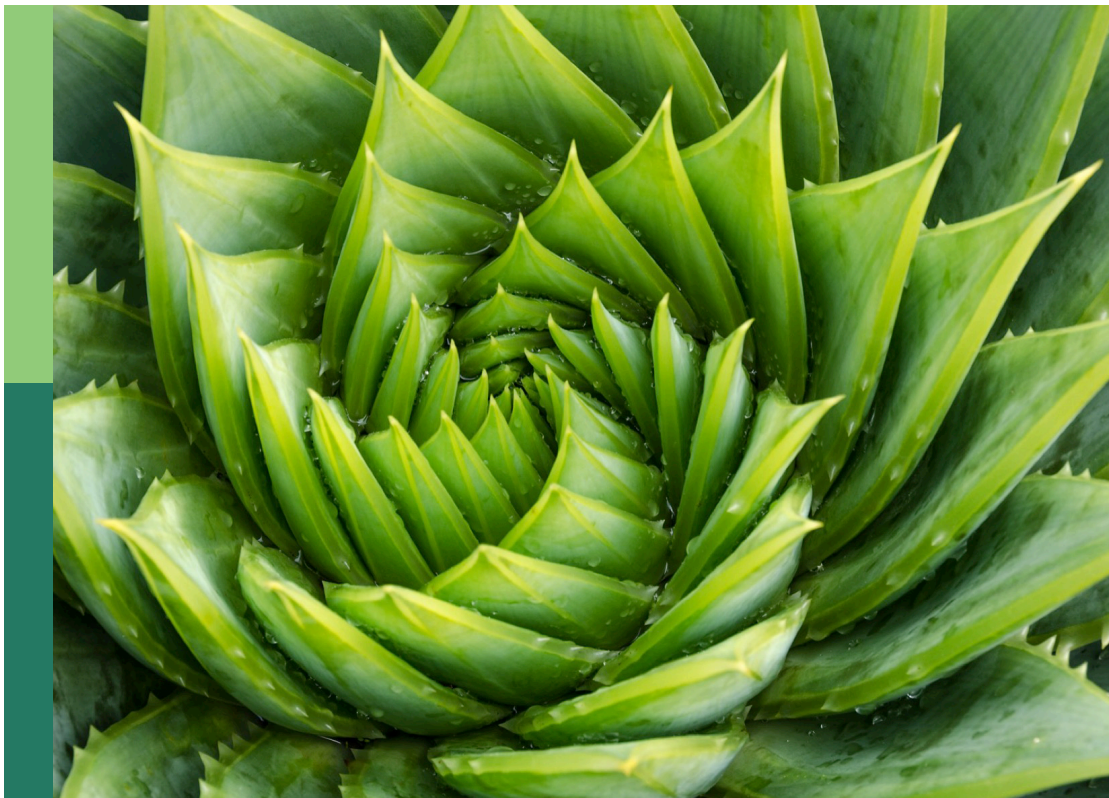
frontiers | Research Topics

Plant-derived natural compounds in drug discovery: The prism perspective between plant phylogeny, chemical composition, and medicinal efficacy

Edited by
Chunnian He, Da-Cheng Hao, Richard Spjut and Peigen Xiao

Published in
Frontiers in Plant Science
Frontiers in Pharmacology



ISSN 1664-8714
ISBN 978-2-83250-566-3
DOI 10.3389/978-2-83250-566-3

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Plant-derived natural compounds in drug discovery: The prism perspective between plant phylogeny, chemical composition, and medicinal efficacy

Topic editors

Chunnian He — Chinese Academy of Medical Sciences and Peking Union Medical College, China

Da-Cheng Hao — Dalian Jiaotong University, China

Richard Spjut — World Botanical Associates, Inc. CA, United States

Peigen Xiao — Chinese Academy of Medical Sciences and Peking Union Medical College, China

Citation

He, C., Hao, D.-C., Spjut, R., Xiao, P., eds. (2022). *Plant-derived natural compounds in drug discovery: The prism perspective between plant phylogeny, chemical composition, and medicinal efficacy*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-83250-566-3

Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 07 October 2022
DOI 10.3389/fpls.2022.1042695

OPEN ACCESS

EDITED AND REVIEWED BY
Massuo Jorge Kato,
University of São Paulo, Brazil

*CORRESPONDENCE
Chun-Nian He
cnhe@implad.ac.cn
Da-Cheng Hao
hao@djtu.edu.cn

SPECIALTY SECTION
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

RECEIVED 12 September 2022
ACCEPTED 27 September 2022
PUBLISHED 07 October 2022

CITATION
Hao D-C, He C-N, Spjut RW and Xiao P-G (2022) Editorial: Plant-derived natural compounds in drug discovery: The prism perspective between plant phylogeny, chemical composition, and medicinal efficacy. *Front. Plant Sci.* 13:1042695. doi: 10.3389/fpls.2022.1042695

Editorial: Plant-derived natural compounds in drug discovery: The prism perspective between plant phylogeny, chemical composition, and medicinal efficacy

Da-Cheng Hao[1,2]*, Chun-Nian He[3]*, Richard W. Spjut[4] and Pei-Gen Xiao[3]

[1]School of Environment and Chemical Engineering, Biotechnology Institute, Dalian Jiaotong University, Dalian, China, [2]Institute of Molecular Plant Science, University of Edinburgh, Edinburgh, United Kingdom, [3]Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [4]World Botanical Associates, Bakersfield, CA, United States

KEYWORDS

pharmacophylogeny, plant phylogeny/phylogenomics, chemical composition, biological activity, multi-omics

Editorial on the Research Topic

Plant-derived natural compounds in drug discovery: The prism perspective between plant phylogeny, chemical composition, and medicinal efficacy

Plant based natural medicine research has evolved from a long history going back to the beginning of human civilization to the present in building on innovative research method systems of pharmacophylogeny and pharmacophylogenomics (Hao and Xiao, 2020). "Pharmacophylogeny" was conceived by Professor Xiao Pei-Gen four decades ago as a result of long-term studies of Chinese researchers, especially since the 1950s. The objective is to disentangle the intricate relationships and connectivity between medicinal plant phylogeny, phytochemical profiles and bioactivities/therapeutic utilities (Figure 1), so as to benefit pioneering plant-based drug research and development (R&D). The historical research in this field while continuing with the status quo, pharmacophylogeny, has become increasingly familiar to more and more researchers. "Pharmacophylogenomics" is proposed to reflect the mounting range of applications of omics based pharmacophylogeny in phytomedicine research. Pharmacophylogeny/pharmacophylogenomics is a multidisciplinary integration, involving molecular phylogeny/phylogenomics, plant morphology, chemotaxonomy, phytochemistry, molecular biology and omics, ethnopharmacology/pharmacology, and the like; for example, Spjut (1985) had

recognized the genus as the lowest chemotaxonomic level of diversity to suggest a phytogeographical approach in searching novel antitumor compounds. In the advent of phylogeny, this could become a phylogeographic approach. Pharmacophylogeny suggests that healing plants of the related taxonomic groups are more likely to possess the analogous chemical profiles/efficacies (Figure 1), which is a law explored from long-term herbal medicine practice, and is used in practice after being confirmed and perfected by scientific research. It played an active role in bioprospecting domestic resources to replace imported medicines. Currently, it is very useful to expand medicinal plant resources (Cui et al.), along with authentication/quality regulation of herbal medicines, and predicting the chemicals or bioactive constituents of herbals and identification/quantification of chemicals (Figure 1). In the coming years, pharmacophylogeny and pharmacophylogenomics could be more powerful in mining original natural medicines, refining ethnopharmacology understandings, therefore advancing the workable protection and application of old/natural remedial resources.

The goal of this Research Topic is to gain novel mechanism insights into the phylogeny/evolution, phytometabolites and pharmacological effects of selected medicinal genera/families. Where possible, we propose that such explorations should be conducted within the context of pharmacophylogeny and/or pharmacophylogenomics. To this end, this Research Topic presents three original articles that reconstructed the phylogenetic trees of *Scutellaria* (Shen et al.), tribe Polygonateae (Wang et al.) and *Arnebia* (Sun et al.) respectively based on the whole chloroplast (cp) genome sequences. The Lamiaceae genus *Scutellaria* has more than 360 species, among which over 70 species have a long history of medicinal use. The molecular phylogeny and metabolomic information are essential to understand the medicinal value of this genus and to develop alternative medicinal resources. The complete cp genomes of 17 *Scutellaria* species, which were used to reconstruct the phylogenetic tree (Shen et al.), disagrees with the traditional morphological grouping. Convincingly, *S. baicalensis* (*Radix Scutellariae* in traditional Chinese medicine (TCM)) is most closely related to *S. viscidula*, followed by *S. hypericifolia*, *S. amoena* and *S. likiangensis*; correspondingly, the metabolomic analyses and phytometabolite content determination reveal the overall similarity of phytometabolite profiles between *S. baicalensis* and its four substitute species. The cp genome sequencing also suggests petA-

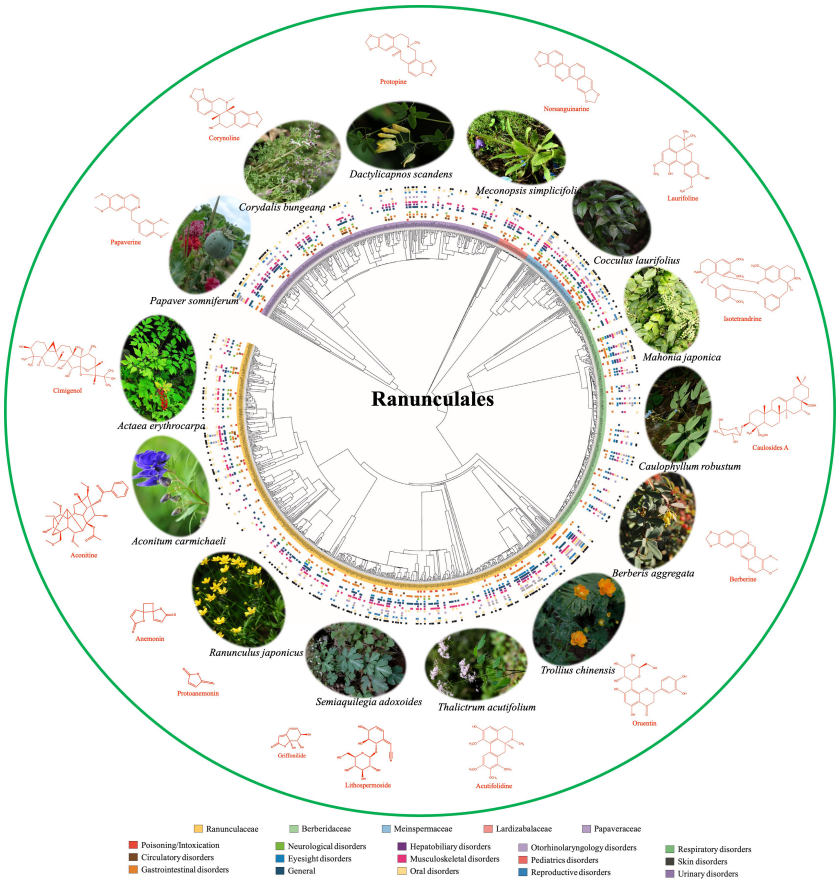


FIGURE 1
An illustrated overview of pharmacophylogeny, taking the order Ranunculales as an example.

psbL as a potential barcoding marker for distinguishing *S. baicalensis* and its substitutes, and the interspecific chemodiversity could lead to development of novel clinical utility. A set of cp genome sequences of Asparagaceae tribe Polygonateae (Wang et al.) is essential for unearthing the links between phylogenomics and chemotaxonomy of medicinal taxa therein. The complete cp genomes of 26 Polygonateae species were *de novo* assembled and characterized; the cp genome-based phylogeny suggests the monophyly of *Polygonatum*, an important TCM genus, and also suggests that *Heteropolygonatum* may be its sister group, except that *Disporopsis*, *Maianthemum* and *Disporum* may have diverged earlier. On the contrary, the phyllotaxy of *Polygonatum* is not stable at the intraspecies level, which cannot be taxonomically used as the unique morphology marker.

Arnebiae Radix is an old oriental medication with diverse activities. The genome skimming methods were utilized to obtain cp genomes of five *Arnebia* species (Sun et al.), by which the phylogenetic relationship of five *Arnebia* species was completely resolved. The origin plants of *Arnebiae Radix A. guttata* and *A. euchroma* were of high genetic diversity, and had three and two subclades respectively. The cp genome is an useful genetic resource for phylogeny and evolution studies at both species and subspecies/population levels. The genetic resources provided in these studies will aid the conservation and exploitation of various medicinal taxonomic groups. Notwithstanding, hybridization has led to incomplete lineage sorting and polyploidy during long-term evolution of taxonomic groups; cp phylogeny does not equal to the exact species one (Lu et al., 2022). In RNA-Seq based phylotranscriptomics, nuclear orthologous genes are concatenated to reconstruct the phylogenetic tree (Hao et al., 2021), which, along with genome skimming (Sun et al.) and Hyb-Seq (Lu et al., 2022), represents the promising complementary approach in pharmacophylogenomics.

According to pharmacophylogeny, taxa in sister phylogeny groups have closely related genetic features (Hao and Xiao, 2020). Thus, they are more likely to possess analogous biosynthesis pathways and their chemical arsenals could be more similar (Liu et al., 2021). The various tiers of chemical similarity result in the global resemblance of bioactivity or pharmacological efficacy (Hao et al., 2022a; 2022b Liu et al.). Pharmacophylogeny successfully guides the development of novel curative taxa (Hao and Xiao, 2020), while circumventing the limitations of old-style approaches and enabling the targeted studies. In contrast, the random discovery (accidental discovery) relies too much on luck; new molecular entities obtained by combinatorial chemistry, high throughput screening or computer aided drug design very often fail due to improper pharmacokinetic and pharmacodynamic properties (Hao et al., 2022c). A complete knowledgebase of chemodiversity/phytometabolism of systematic groups of interest is an essential base for pharmacophylogeny. This Research Topic presents a few original articles that carefully investigate the specialized metabolism and chemodiversity of *Chrysanthemum* (Hao et al.), *Oxytropis* (Jia et al.), *Broussonetia* (Jiao et al.), *Schisandra* (Liu et al.), *Bletilla* (Liu et al.) and *Verbena* (Peng et al.). More than 9,000 flavonoids are distributed in 245 families of seed plant (Zhang et al., 2020). In *Broussonetia papyrifera* leaves, flavonoids accumulate gradually in leaf development to until its maturity (Jiao et al.), and female plants have greater flavonoid leaf content than in male plant, despite their analogous composition. The 192 identified flavonoids include flavonols, flavones, flavan-3-ols, flavonoid carbonoside, anthocyanins, among others, and their biosynthesis follows the well-known flavonoid biosynthetic pathways. Some differentially expressed genes and metabolites along the flavonoid biosynthetic pathway were quantified by transcriptome sequencing and metabolomic analyses respectively.

The phenolic acids are distributed in 167 families of seed plant (Zhang et al., 2020). Natural phenolic compounds are commonly distributed in food and TCM (Hao et al.). In a suspension culture of *Bletilla striata* (Liu et al.), the content of Dactylorhin A peaked at the earliest, i.e., nine days post inoculation, followed by p-hydroxybenzylalcohol, Militarine and Coelonin. Based on full-length transcriptome data, multiple unigenes involved in the biosynthesis of HBA, militarine, dactylorhin A and coelonin were identified. In the transcriptome analysis of *Verbena officinalis*, 206, 229 and 115 unigenes were identified in the biosynthetic pathways of iridoid glycoside, phenylethanol glycoside and flavonoid respectively (Peng et al.). The contents of these phytometabolites were highest in leaves, followed by stem and root, which is consistent with inter-tissue variations of biosynthetic gene expression levels.

The above studies not only contribute a large number of unigenes for the ortholog extraction and phylotranscriptomic inference, but also greatly enrich the phytometabolite database at the genus/species level. In light of the expanding knowledge of plant chemodiversity, it is possible to study the phylogenetic distribution of diverse types of compounds, e.g., less studied pyrone glycoside (Shen et al.), at the subfamily, tribe, genus or subgenus levels, so as to bioprospecting medicinal taxa more precisely.

A triple helix systems perspective, i.e., an essentially holistic view, would be established if we qualitatively and quantitatively investigate the association between molecular phylogeny and phytochemical data (Zhang et al., 2020), between phylogeny and bioactivity/efficacy (Hao et al., 2022a; 2022b), and between phytometabolite and bioactivity (Song et al.; Sun et al.). Based on extensive ethnomedicinal uses, the thorough phytochemistry and bioactivity characterizations contribute to the safe administration of herbals and discovery of lead compounds (Souza et al., 2018). In total, 167 compounds were identified from essential oils (EOs) of six Lamiaceae herbs by GC-MS analysis (Sun et al.). All EOs had promising anti-inflammatory activity in rats with adjuvant arthritis, and *Perilla frutescens* EO was the best. They alleviated the joint swelling in the rat models; the lymphocyte infiltration and cartilage damage were significantly inhibited. The TNF-α, IL-1, and IL-6 were reduced in rat articular tissues by EOs. Six EOs had

substantial toxicity on tumor cell lines; they also had the significant antioxidant activity, and *Salvia japonica* EO was the strongest. Berberine, baicalin and saponin from *Coptis chinensis*, *Scutellaria baicalensis* and *Vincetoxicum atratum* respectively were combined to produce the formula BBS (Song et al.), and effects of BBS on tumor growth and protein expression were as strong as those of formula HHB, a combination of the above three herbs. Although these bottom-up and intermediate approaches cannot replace the top-down approach, i.e., treatment investigations of clinical patients, they are very important in confirming or falsifying tentative conclusions obtained from the phylogenetic signals in phytometabolites or bioactivity/efficacy across species phylogenetic tree. The bioactivity studies provide more reference data for pharmacophylogeny research and drug discovery from ethnomedicinal plants.

In summary, the articles of this Research Topic illustrate the newest knowledge and prospects of pharmacophylogeny, pharmacophylogenomics and related concepts, and their mounting usefulness in phytomedicine R&D. The deeper investigations of genomics of curative plants, metabolomics, and ethnopharmacology-based bioactivity will enable the workable protection and consumption of botanical possessions. Of note, species relationships may be cryptic due to genetic and/or epigenetic factors, or species of the same taxonomic group may be morphologically and biochemically distinct, making it possible for closely related taxa to have diametrically opposite clinical effects. During long-term evolution, the phytochemical diversification parallels the expansion of biological complexity. Biologically active compounds are generally specialized metabolites, and the distribution of these compounds on phylogenetic trees of different levels is very worthy of deeper study. Many medicinally important phytometabolites, e.g., alkaloids, terpenoids and phenylpropanoids, can be further subdivided structurally, and the phylogenetic signals of different subtypes/subclasses could be distinct. For compounds with similar structures, not only the resemblances and variances of their bioactivities should be explored, but also those of their pharmacokinetic and pharmacodynamic properties should be clarified one by one (Hao et al., 2022c). Therefore, there is a great deal of work to be conducted on various topics of pharmacophylogeny. During the past two decades, various omics techniques have continued to evolve towards elucidating the cryptic connections between plant phylogeny, phytometabolites and their biosynthesis, as well as their bioactivities, which promisingly provide lead entities to develop novel plant-derived drugs.

Author contributions

D-CH and C-NH prepared the draft and finalized the manuscript. P-GX and RS edited and corrected the draft. All authors contributed to the article and approved the submitted version.

Funding

The authors are grateful to the Journal for presenting this Editorial article. D-CH is supported by China Scholarship Council (202108210156). This work is also funded by the Innovation Team and Talents Cultivation Program of the National Administration of Traditional Chinese Medicine (ZYYCXTDD-202005) and CAMS Innovation Fund for Medical Sciences (CIFMS; 2021-I2M-1–071).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Hao, D. C., Li, P., Xiao, P. G., and He, C. N. (2021). Dissection of full-length transcriptome and metabolome of dichocarpum (Ranunculaceae): implications in evolution of specialized metabolism of ranunculales medicinal plants. *PeerJ* 9, e12428. doi: 10.7717/peerj.12428

Hao, D. C., Lyu, H. Y., Wang, F., and Xiao, P. G. (2022b). Evaluating potentials of species rich taxonomic groups in cosmetics and dermatology: Clustering and dispersion of skin efficacy of asteraceae and ranunculales plants on the species phylogenetic tree. *Curr. Pharm. Biotech.* doi: 10.2174/1389201023666220324123926

Hao, D. C., Wang, F., and Xiao, P. G. (2022c). Impact of drug metabolism/pharmacokinetics and their relevance upon traditional medicine-based anti-COVID-19 drug research. *Curr. Drug Metab.* 23 (5), 374–393. doi: 10.2174/1389200223666220418110133

Hao, D. C., and Xiao, P. G. (2020). Pharmaceutical resource discovery from traditional medicinal plants: Pharmacophylogeny and pharmacophylogenomics. *Chin. Herb. Med.* 12 (2), 104–117. doi: 10.1016/j.chmed.2020.03.002

Hao, D. C., Zhang, Y. L., He, C. N., and Xiao, P. G. (2022a). Distribution of therapeutic efficacy of ranunculales plants used by ethnic minorities on the phylogenetic tree of Chinese species. *Evid. Based Comple. Alter. Med.* 2022, 9027727. doi: 10.1155/2022/9027727

Liu, H. B., Ma, P., Xu, L. J., and Xiao, P. G. (2021). Informatics and big data: A new stage of pharmacophylogeny in medicinal plants. *Mod. Chin. Med.* 23 (9), 1506–1511. doi: 10.13313/j.issn.1673-4890.20210901002

Lu, W. X., Hu, X. Y., Wang, Z. Z., and Rao, G. Y. (2022). Hyb-seq provides new insights into the phylogeny and evolution of the chrysanthemum zawadskii species complex in China. *Cladistics*. doi: 10.1111/cla.12514

Souza, E. N. F., Williamson, E. M., and Hawkins, J. A. (2018). Which plants used in ethnomedicine are characterized? phylogenetic patterns in traditional use related to research effort. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00834

Spjut, R. W. (1985). Limitations of a random screen: Search for new anticancer drugs in higher plants. *Econo. Bot.* 39, 266–288. doi: 10.1007/BF02858796

Zhang, Y. Z., Deng, T., Sun, L., Landis, J. B., Moore, M. J., Wang, H., et al. (2020). Phylogenetic patterns suggest frequent multiple origins of secondary metabolites across the seed-plant 'tree of life'. *Natl. Sci. Rev.* 8 (4), nwaa105. doi: 10.1093/nsr/nwaa105

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 20 May 2022
doi: 10.3389/fpls.2022.900030

Integrative Metabolome and Transcriptome Analysis of Flavonoid Biosynthesis Genes in *Broussonetia papyrifera* Leaves From the Perspective of Sex Differentiation

Peng Jiao[1], *Li Chaoyang*[2], *Zhai Wenhan*[3], *Dai Jingyi*[1], *Zhao Yunlin*[1]* and *Xu Zhenggang*[1,3]*

[1] Hunan Research Center of Engineering Technology for Utilization of Environmental and Resources Plant, Central South University of Forestry and Technology, Changsha, China, [2] Central South Inventory and Planning Institute of National Forestry and Grassland Administration, Changsha, China, [3] College of Forestry, Northwest A&F University, Yangling, China

OPEN ACCESS

Edited by:
Da-Cheng Hao,
Dalian Jiaotong University, China

Reviewed by:
Feng Xu,
Yangtze University, China
Zou Xiaoxing,
Fujian Agriculture and Forestry University, China
Chongde Sun,
Zhejiang University, China

***Correspondence:**
Zhao Yunlin
zyl8291290@163.com
Xu Zhenggang
xuzhenggang@nwafu.edu.cn

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: 19 March 2022
Accepted: 15 April 2022
Published: 20 May 2022

Citation:
Jiao P, Chaoyang L, Wenhan Z, Jingyi D, Yunlin Z and Zhenggang X (2022) Integrative Metabolome and Transcriptome Analysis of Flavonoid Biosynthesis Genes in Broussonetia papyrifera Leaves From the Perspective of Sex Differentiation. Front. Plant Sci. 13:900030. doi: 10.3389/fpls.2022.900030

Flavonoids are important secondary metabolites involved in plant development and environmental responses. Sex differences in flavonoids are common in plants. *Broussonetia papyrifera* is a dioecious plant that is rich in flavonoids. However, few studies have been done on its molecular mechanism, especially sex differences. In the present study, we performed an integrated transcriptomics and metabolomics analysis of the sex differences in the accumulation of flavonoids in *B. papyrifera* leaves at different developmental stages. In general, flavonoids accumulated gradually with developmental time, and the content in female plants was higher than that in male plants. The composition of flavonoids in female and male plants was similar, and 16 kinds of flavonoids accumulated after flowering. Correspondingly, a significant enrichment of differentially expressed genes and metabolites was observed in the flavonoid biosynthesis pathway. WGCNA and qRT-PCR analyses identified several key genes regulating the accumulation of flavonoids, such as those encoding *CHS*, *CHI* and *DFR*. In addition, 8 TFs were found to regulate flavonoid biosynthesis by promoting the expression of multiple structural genes. These findings provide insight into flavonoid biosynthesis in *B. papyrifera* associated molecular regulation.

Keywords: ***Broussonetia papyrifera*, flavonoids, transcription factors, sex differentiation, WGCNA**

INTRODUCTION

Dioecious plants are an important part of terrestrial ecosystems and play a positive role in protecting species diversity and maintaining ecosystem stability. From an evolutionary perspective, dioeciousness increases genetic diversity, allowing plants to exhibit a variety of morphologies and traits (Al-Dossary et al., 2021). On the other hand, environmental stress or human selection can further contribute to differences in plant traits between sexes. Actually, sex differences in metabolites are common in plants. In *Rhamnus alpinus*, male plants have higher concentrations of defense-related compounds (anthraquinones) in the leaves than female plants, thus indicating increased resistance to biotic stress in male plants (Banuelos and Obeso, 2004). In *Populus cathayana*, under normal conditions, the sucrose concentration of male leaves

was higher than that of female leaves. Nitrogen or potassium deficiency significantly enhanced leaf sucrose in females, but had less effect on males (Yang et al., 2015; Wu et al., 2021). Until now, most reports have focused on the physiological basis of dioecious plants, but few studies have been done on molecular biology, which is a crucial question. Because changes in physiological traits may be due to a genetic basis or in combination with other factors such as allocation. It is necessary to distinguish temporary changes in stress from stable genetic features.

Flavonoids, as a class of polyphenolic compounds with a C6–C3–C6 double aromatic ring, are important secondary metabolites in plants development. Flavonoids can improve the adaptability of plants to the various terrestrial environments, and participate in plant ecological defense, such as promoting plant growth (Stafford, 1991), resisting diseases (Agati et al., 2012), resisting the damage of ultraviolet radiation (Harborne and Williams, 2000), scavenging oxygen free radicals (Agati et al., 2009, 2011; Lillo et al., 2010; Tattini et al., 2010), driving away foragers, and preventing the invasion of harmful microorganisms (Falcone Ferreyra et al., 2012). Therefore, flavonoids may be a secondary metabolite in response to the environment. In addition to being beneficial to plants themselves, flavonoids have also been developed and utilized for their medicinal effects. More and more evidences suggest that flavonoids and related compounds can provide beneficial effects on cancer and chronic diseases, including cardiovascular disease, type II diabetes and non-alcoholic fatty liver disease (NAFLD), due to their immunomodulatory, anti-inflammatory and antioxidant properties (González-Gallego et al., 2010; Ying et al., 2013; Pisonero-Vaquero et al., 2014). Sex differentiation of flavonoid content has recently been emphasized because it not only reflects the survival strategies of different sexual plants, but also affects the development and utilization. For instances, the content of flavonoids in female *Hippophae rhamnoides* leaves is slightly lower than that of male plants (Lu et al., 2018). Under limited phosphorus conditions, the synthesis of flavonoids and condensed tannins in leaves, as well as the synthesis of phenolic acids in stems and roots in females *P. tremula*, were greater than in males (Randriamanana et al., 2014). In *Ginkgo biloba*, more phenolics were extracted from leaves of male trees than from female trees (Koczka et al., 2015).

Broussonetia papyrifera (paper mulberry) is a dioecious plant belonging to the Moraceae family. Once an important source of forage, the distribution of *B. papyrifera* has expanded with human migration (Chang et al., 2015). It is widely distributed in Asia and Pacific countries, and is an important economic plant in most areas (Zheng et al., 2002). In recent year, due to its easily digestible crude fiber and high protein content, *B. papyrifera* is still used as forage to address the shortage of feedstuff. At the same time, the tree has the characteristics of strong germination, fast growth, resistance to salt stress, heavy metals and air pollution, and has been used as a pioneer plant in polluted areas (Li et al., 2012; Zhang et al., 2013; Zhao et al., 2014). Compared with other woody plants, *B. papyrifera* is rich in flavonoids, polyphenols, and fructose (Feng et al., 2019), which may be the reason why *B. papyrifera* has been used as a traditional Chinese medicine for a long history (Feng et al., 2008). For instance, prenylflavone derivatives from *B. papyrifera* have an inhibitory effect on cancer cells (Guo et al., 2013), and polyphenols from *B. papyrifera* can inhibit coronavirus protease (Park et al., 2017). Broussochalcone A, papyriflavonol A, 3′-(3-methylbut-2-enyl)-3′,4′,7-trihydroxyflavane, and broussoflavan A from *B. papyrifera* have been proved to be potent the main protease (Mpro) inhibitors. There are more effective than the two repurposed drugs (lopinavir and darunavir) and may serve as promising anti-COVID-19 drugs (Ghosh et al., 2021). *B. papyrifera* leaves and roots are rich in flavonoids. The high flavonoid content may result from the significant expansion of gene family in the flavonoid synthesis pathway (Peng et al., 2019). On the other hand, previous studies have shown that there are sex differences in the accumulation of flavonoids in *B. papyrifera* (Zhao, 2011), but the molecular mechanism is still unclear. Understanding synthesis mechanism of flavonoids in *B. papyrifera* leaves of different sexes will not only help us to understand the survival strategies of *B. papyrifera* in the wild, but also help us to select superior varieties.

At present, metabolomic and transcriptomic analysis have been widely used to study the response mechanisms of metabolite accumulation under different conditions (Zhang et al., 2021). In our study, we focus on the molecular mechanisms of flavonoid biosynthesis in *B. papyrifera* leaves, especially their sex differences. Our work aims to: (I) elucidate the accumulation of flavonoids in *B. papyrifera* leaves of different sexes, and (II) screen the key genes for the synthesis of *B. papyrifera* flavonoids. The results of this work may provide insight into flavonoid biosynthesis in *B. papyrifera* associated molecular regulation.

Abbreviations: ANR, Anthocyanin reductase; AP2/ERF, APETALA2/ethylene-responsive element binding factors; CHS, Chalcone synthase; CHI, Chalcone isomerase; DAFs Differentially accumulated flavonoids; DFR, Dihydroflavonol reductase; DEGs, Differentially expressed genes; F3H, Flavanone 3-hydroxylase; HSFs, Heat shock factors; MYB, v-myb avian myeloblastosis viral oncogene homolog; NFYA-HAP2, Nuclear transcription factor Y; qRT-PCR, Quantitative reverse transcription-polymerase chain reaction; TFs, Transcription factors; UPLC, Ultra performance liquid chromatography; WGCNA, Weighted gene co-expression network analysis.

MATERIALS AND METHODS

Plant Materials

Broussonetia papyrifera leaves were collected from April 2019 to October 2019 on the campus of Central South University of Forestry and Technology (28°6′25.48″ N, 112°59′37.68″ E), China. Three adult female (F) and three adult male (M) *B. papyrifera* were selected as experimental samples. They were divided into three groups, each consisting of a male and a female, and they were grown together. The distance between each group was about 200 meters. The six samples were the same age and well-grown. Subsequently, three typical phenological periods of *B. papyrifera* were selected for sampling. The first was 30 days after flowering (the flowering stage, F30/M30). At this point, the gender could be clearly distinguished, and the leaves were fully expanded. The second was 120 days after flowering (the fruiting stage, F120/M120). It was the peak fruiting period of *B papyrifera*,

with green and red fruits accounting for about half. The third was 180 days after flowering (the defoliation stages, F180/M180). At this point, the leaves began to fall (**Supplementary Figure S1**). Every treatment was repeated three times. These leaves were immediately frozen in liquid nitrogen and stored at −80°C for metabolite extraction, transcriptome sequencing, and qRT-PCR analysis.

Total Flavonoids Analysis

The content of total flavonoids was determined by the aluminum nitrate chromogenic method (Yang, 2018). In brief, 1.00 g freeze-dried leaf powder mixed with 5 ml 70% (v/v) ethanol, and then extracted at 50°C for 1.5 h. The supernatant was collected by centrifugation at 8,000 r/min for 10 min at room temperature, and then diluted to 25 ml with 70% (v/v) ethanol. The content of total flavonoids was measured by spectrophotometer UV-1200 (Mapada, China) at 510 nm using rutin as the standard substance.

Flavonoid Identification and Quantification

100 mg of vacuum freeze-dried leaf samples were dissolved in 1.0 ml 70% (v/v) methanol. The supernatant was collected by centrifugation at 10,000 g for 10 min. The test solution was obtained by filtration with a microporous membrane (0.22 μm Pore Size). Flavonoids were analyzed by Metware Biotechnology Co., Ltd. (Wuhan, China) using an LC-ESI-MS/MS System (UPLC, Shim-Pack UFLC SHIMADZU CBM30A System, www.shimadzu.com.cn/; MS, Applied Biosystems 4000 Q TRAP, www.appliedbiosystems.com.cn/). Metabolite quantification was performed using a scheduled multiple reaction monitoring (MRM) method, which has been previously described (Chen et al., 2013). Setting $|\text{Log}_2\ (\text{Fold Change})| \geq 2$ and $|\text{Log}_2\ (\text{Fold Change})| \leq 0.5$, and $P < 0.05$ as thresholds for DAFs.

RNA-seq Analysis

The total RNA was extracted from frozen leaf samples using an EASYspin Plus Plant RNA Kit by Biomarker Technologies Co., Ltd. (Beijing, China). A total amount of 1 μg RNA per sample was used as input material for the RNA sample preparations. Sequencing libraries were generated using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) following manufacturer's recommendations and transcriptome sequencing were performed by Biomarker Technologies Co., Ltd. (Beijing, China) using an Illumina Hiseq 2000 platform. Clean data were obtained by removing reads containing adapter, reads containing ploy-N and low quality reads from raw data. Transcriptome assembly was accomplished using Trinity (Grabherr et al., 2011) by default. The assembled transcript was mapped to the clean reads to obtain the mapped reads. Gene expression levels were estimated by RNA-seq by Expectation-Maximization (RSEM) (Li and Dewey, 2011), and fragments per kilobase of transcript per million mapped reads (FPKM) was used for gene/transcript level quantification. Differential expression analysis of two samples was performed using the EBSeq R package. *P* value was adjusted using *q* value. *q* value < 0.005 and $|\log_2\ (\text{Fold Change})| > 1$ was set as the threshold for significantly differential expression. Gene functions were annotated based on the following databases: NR (NCBI non-redundant protein sequences), Pfam (Protein family), KOG/COG/eggNOG (Clusters of Orthologous Groups of proteins), Swiss-Prot (A manually annotated and reviewed protein sequence database), KEGG (Kyoto Encyclopedia of Genes and Genomes), GO (Gene Ontology). We used KOBAS (Chen et al., 2011) software to test the statistical enrichment of differential expression genes in the KEGG pathways.

Co-expression Network Analysis for Construction of Modules

To identify key regulatory genes in the flavonoid biosynthesis pathway, WGCNA was performed using the WGCNA R package (v 1.70). The adjacency matrix between different genes was constructed with a soft power of 26 and a dynamic tree cut procedure (merge cut height = 0.25, min module size = 30) was used to screen similar modules in the hierarchical tree. The module eigengene was defined as the first principal component of a given module, and then used to represent the expression profile of module genes in each sample. The Pearson correlations between the eigengenes of each module and the abundance of flavonoids were performed using OriginPro 2016. Genes associated with flavonoid biosynthesis pathways were identified by KEGG annotation (ko00941, ko00942, ko00943 and ko00944) (Guo et al., 2020). Genes related to flavonoid biosynthesis and TFs in flavonoid biosynthesis highly correlated modules were constructed co-expression networks with edge weights >0.3, and visualized the candidate target genes with Cytoscape software (Version 3.9.0) (Ding et al., 2021). Genes with high degrees in the network were selected as hub genes.

qRT-PCR Analysis

Total RNA was extracted from leaf samples according to the instructions of Biofit kit (Tsingke, China). The strand cDNA was synthesized using Goldenstar RT6 cDNA synthesis kit (Tsingke, China). The expression levels of representative unigenes and selected key genes in the flavonoid biosynthesis pathway were analyzed by qRT-PCR using FQD-96A (Bioer Technology, China). The 20 μl reaction mixture contained 10 μl SYBR® Green Real-time PCR Master Mix (CWBIO), 1 μl cDNA template, 0.8 μM of each forward and reverse primer, and 7.4 μl ddH_2O. The following cycling parameters were applied for amplification: 95°C for 1 min followed by 40 cycles of 95°C for 15 s, 60°C for 15 s, 72°C for 30 s, then followed by 95°C for 5 s, 60°C for 1 min, 0.11°C / s to 95°C, 50°C for 30 s for plate reading. The primers list is shown in **Supplementary Table S1**. The actin gene was selected as an internal reference (Xu et al., 2019). Three replicates were performed for each sample. Quantitative data were analyzed using the $2^{-\Delta\Delta CT}$ method.

Statistical Analysis

Differences between samples were determined by one-way analysis of variance (ANOVA) and significant differences were calculated by the least significant difference (LSD) test at $P < 0.05$.

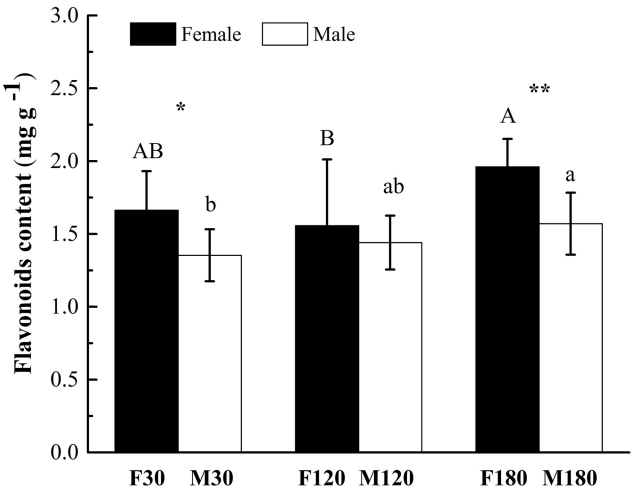


FIGURE 1 | Total flavonoid contents in developing *Broussonetia papyrifera* leaves. Asterisks indicate statistically significant differences between female and male plants: no symbol $P > 0.05$; $^{*}P < 0.05$; $^{**}P < 0.01$. Capital letters indicate statistically significant differences ($P < 0.05$) among the three development stages of female plants. Lowercase letters indicate statistically significant differences ($P < 0.05$) among the three development stages of male plants.

RESULTS

Total Flavonoids in *Broussonetia papyrifera* Leaves

The content of flavonoids in leaves of female *B. papyrifera* was higher than that of male at three developmental stages. There were significant differences between flowering and defoliation stages ($P < 0.05$). Accumulation of flavonoids increased with leaf development. F180 and M180 had the highest content at 1.961 ± 0.191 mg g^{-1} and 1.571 ± 0.213 mg g^{-1}, respectively (**Figure 1**).

Identification of Flavonoids in *Broussonetia papyrifera* Leaves

A total of 192 kinds of flavonoids substances were detected in *B. papyrifera* leaves, including 85 flavones, 46 flavonols, 23 flavonoid carbonoside, 10 flavan-3-ols, 9 anthocyanins, 6 dihydroflavone, 7 dihydroflavonol, 4 isoflavones and 2 chalcones (**Supplementary Table S2**). 170 kinds of flavonoids were identified in F30 or M30, and 188 kinds of flavonoids were identified in F120, M120, F180 or M180 (**Figure 2A**). 163 kinds of flavonoids were shared among 6 leaf samples. Besides, 16 kinds of flavonoids were found in the fruiting and defoliation stage but not in flowering stage (**Figure 2B**). Comparing the sex differences of each flavonoid, there were 50 DAFs in flowering stage, 49 DAFs in fruiting stage and 72 DAFs in defoliation stage. Among this DAFs, the content of most flavonoids in female *B. papyrifera* was higher than that in male (**Supplementary Figure S2**). Cluster analysis of 192 kinds of flavonoids showed F30 and M30 were located in the same branch, and F120, M120, F180 and M180 were located in the same branch. At the same time, these flavonoids were divided into 3 clade, clade I had high content substances in the F30 and M30, clade II high content substances in the M30 and F120, and clade III high content substances in the F180 and M120 (**Figure 3A**).

According to KEGG annotation results, 27 substances were identified in the flavonoids biosynthesis pathway. Remarkably, most substances were down-regulated in male samples compared with female samples (**Supplementary Table S3**). Cluster analysis of 27 flavonoids showed that F30, M30 and M180 were located in the same branch, while F120, M120 and F180 were located in the same branch. The number of flavonoids with high content in the former was lower than that in the latter (**Figure 3B**).

Transcriptome Analysis of *Broussonetia papyrifera* Leaves of Different Sexes

The throughput and sequencing quality were high enough to warrant further analysis. Each library received 22244561–22999380 clean reads and 17941528–21223800 mapped reads. The mapped ratio was in the range of 83.79–85.73%. Q30 percentages and GC percentages were 93.08–94.45% and 46.70–47.37%, respectively (**Supplementary Table S4**). During the development of *B. papyrifera* leaves, a total of 5,672 DEGs were obtained, including 1,475 at flowering stage, 4,391 at fruiting stage and 1,030 at defoliation stage (**Supplementary Figure S3**). Comparing the DEGs with the common databases, 4,887 differentially expressed genes were annotated, accounting for 86.16% (**Supplementary Table S5**).

The KEGG enrichment analysis revealed that most genes were mainly enriched in "Ribosome," "Plant-pathogen interaction," and "Carbon metabolism" (**Figure 4**). Of note, there were 6 DEGs for F30 vs. M30, 11 DEGs for F120 vs. M120, and 11 DEGs for F180 vs. M180 in the flavonoid biosynthesis pathway. According to KEGG annotation results, 20 genes were identified in the flavonoids biosynthesis pathway (**Supplementary Table S6**). Among them, 14 genes were only highly expressed in the fruiting and defoliation stages, accounting for 70% (**Figure 5**; **Supplementary Table S6**). Among them, *CHI*, *CHS2*, *CHS3*, *CHS4*, *CHS5*, *CHS6*, *CHS7*, and *DFR* were highly expressed mainly in female plants. *CHS1*, *HCT5* and *HCT4* were highly expressed in male plants. *ANR*, *CCOMT2* and *CYP73A* were highly expressed in male plants at fruiting stage, and highly expressed in female plants at defoliation stage (**Figure 5**). Notably, the expression of *F3H* was different in the three stages, the male plants were higher than the female plants in the flowering and fruiting stages, and the male plants were lower than the female plants in defoliation stage (**Figure 5**). Therefore, some structural genes may regulate the sex differentiation of flavonoid synthesis in *B. papyrifera* leaves, especially in the fruiting and defoliation stages.

Expression of Flavonoid Synthesis-Related Genes in Different Sexes

The profiles of transcriptional data and metabolic data during leaf development of *B. papyrifera* were exhibited in the flavonoid synthesis pathway (**Figure 6**). A total of 13 kinds of flavonoids were regulated by 12 genes in the pathway. Among these genes, the differences in gene

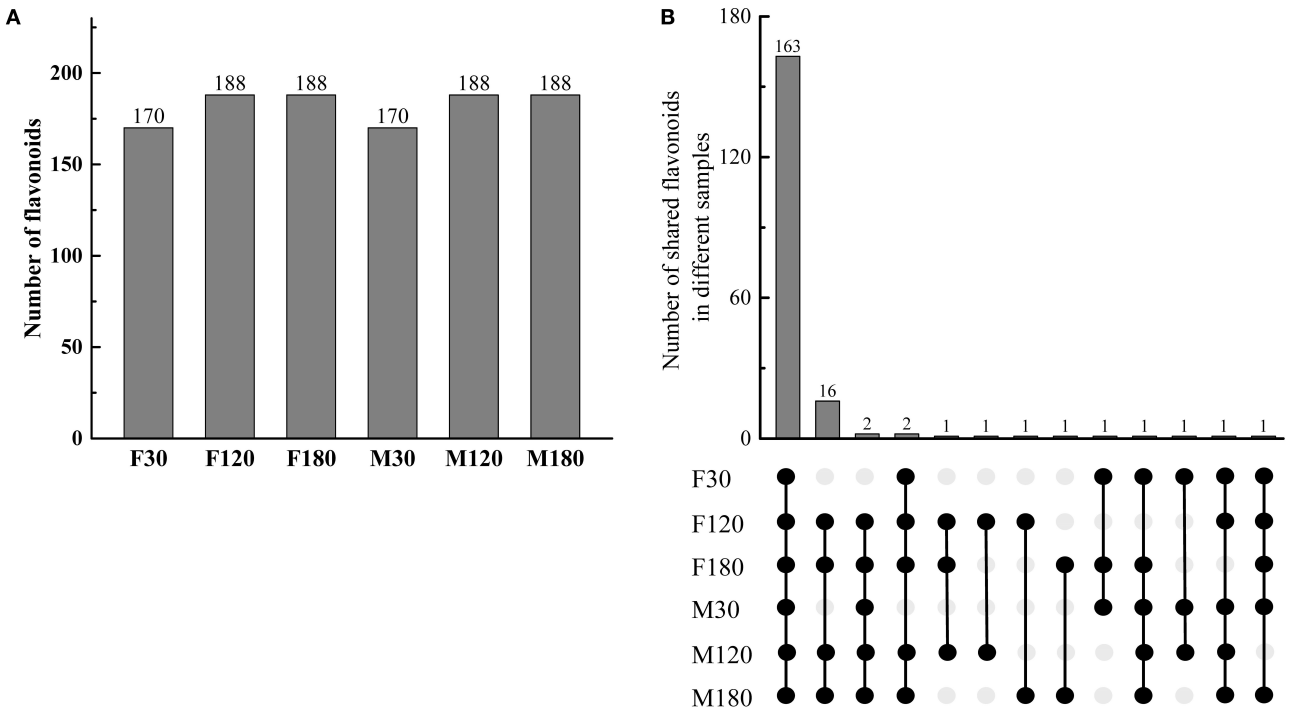


FIGURE 2 | Gender differences in the types and numbers of flavonoid metabolites in *Broussonetia papyrifera* leaves. **(A)** Numbers of flavonoids in different samples. **(B)** Numbers of shared flavonoids in different samples. The black solid circle indicated the presence of compounds in samples, and the gray solid circle indicated the absence of compounds in samples.

expression of *F3H*, *CHS1* and *CYP75B1* had little effect on the content of flavonoids. High expression of *CHS2*, *CHS3*, *CHS4*, *CHS5*, *CHS6*, *CHS7*, *DFR*, *ANR* and *CHI* promoted the accumulation of flavonoids such as phlorizin, dihydrokaempferol, garbanzol, epigallocatechin, epiafzelechin, catechin, afzelechin, etc.

Co-expression Analysis to Identify Differential Genes Related to Flavonoid Synthesis

A gene cluster tree was constructed according to the correlation of expression levels between genes, and one branch of the tree corresponds to gene clusters whose expression levels were highly correlated (**Figure 7A**). The average adjacency coefficient of each gene was 196 (**Supplementary Figure S4**). 16 common expression modules were obtained (**Figure 7B**). Except for the gray module consisting of non-clustered genes, the turquoise module contained the largest number of genes (1,688). The midnightblue module contained the smallest number of genes (36). Correlation coefficients between modules and flavonoids showed that red and turquoise modules had the highest correlation with flavonoids synthesis (**Supplementary Figure S5**). The red module was significantly correlated with mws1173 (garbanzol) and pme0371 (naringenin 7-O-glucoside), while turquoise was correlated with mws0463 (hesperetin), mws1068 (kaempferol), mws4057 (7,4'-dihydroxyflavone) and pme3285 (afzelechin) (**Supplementary Figure S5**).

KEGG enrichment analysis showed that in the red module, flavonoid biosynthesis, cysteine and methionine metabolism, and biosynthesis of amino acids were the top 3 in metabolism pathways (**Supplementary Figure S6**). Eight genes related to flavonoid synthesis, namely *CHS2*, *CHS3*, *CH4*, *CH5*, *CHS6*, *CHS7*, *ANR*, and *DFR*, were enriched in the flavonoid synthesis pathway (**Supplementary Table S6**). In the turquoise module, carbon metabolism, biosynthesis of amino acids, oxidative phosphorylation, glycolysis/gluconeogenesis, and purine metabolism were the top 5 in metabolism pathways (**Supplementary Figure S6**). In addition, a gene related to flavonoid synthesis, *CHI*, was enriched in the flavonoid synthesis pathway (**Supplementary Table S6**).

Using the FPKM values of genes to draw heat maps (**Figure 8**), the results showed that genes in the red module were highly expressed in F180 and significantly higher than in other samples, while genes in the turquoise module were highly expressed in F120. Although the red and turquoise modules were highly correlated with flavonoid synthesis, the genes in them exhibited distinct expression patterns and might perform distinct functions.

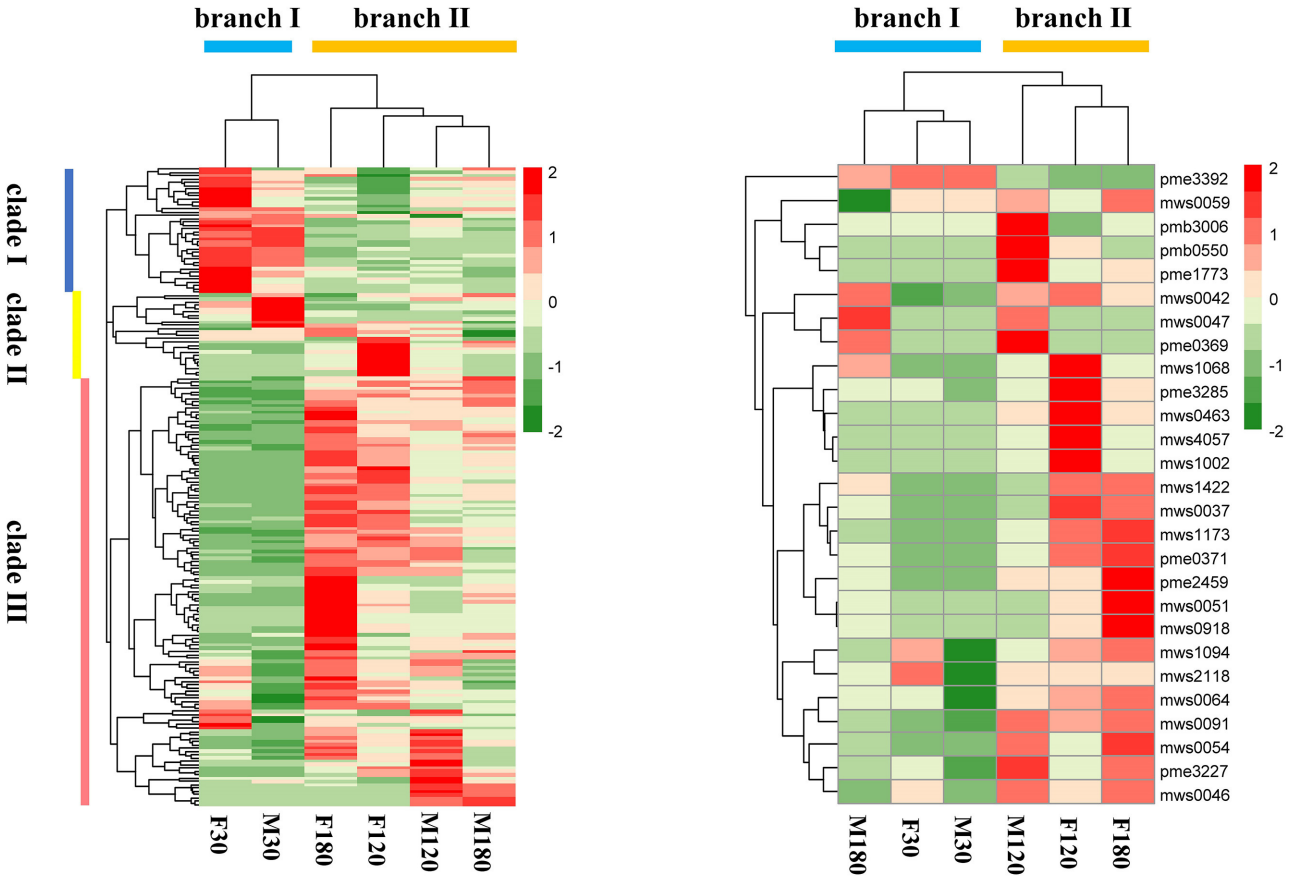


FIGURE 3 | Cluster analysis of flavonoids in *Broussonetia papyrifera* leaves. **(A)** 192 flavonoids found in samples. **(B)** 27 flavonoids identified in the flavonoids biosynthesis pathway.

Identification of Key Genes for Flavonoids Synthesis

Key genes for flavonoids synthesis were screened using the gene expression levels of the major modules (**Table 1**). In the red module, *CHS2*, *CH3*, *CHS4*, *CHS5*, *CHS6*, *CHS7*, *ANR* and *DFR* were highly expressed at defoliation stage. In the turquoise module, *CHI* were highly expressed at fruiting stage and defoliation stage. *CHS2*, *CH7*, *DFR* and *CHI* had FPKM > 100 in our data, these genes may be key genes for flavonoid accumulation of *B. papyrifera* leaves.

Identification of Transcription Factors for Flavonoids Synthesis

TFs were screened using the co-expression network of the major modules (**Supplementary Table S7**; **Figure 9**). The results showed that there were eight structural genes related to flavonoid synthesis, which were *CHS2*, *CHS3*, *CH4*, *CH5*, *CHS6*, *CHS7*, *DFR*, and *CHI*. There were eight TFs related to flavonoid synthesis. Among them, HSFs1, HSFs2, MYB1 were strongly associated with *CHS* and *DFR*. AP2, NFYA-HAP2, WRKY33a, WRKY33b and MYB2 were strongly associated with *CHI*.

qRT-PCR Analysis

To verify the stability of transcriptome, qRT-PCR verification on 10 representative unigenes in the flavonoid biosynthesis pathway was performed (**Supplementary Table S1**; **Supplementary Figure S7**). The correlation coefficient between qRT-PCR results and RNA-seq results reached 0.82. Therefore, our transcriptome sequencing data were generally accurate.

On the other hand, qRT-PCR verification on four key genes for flavonoid accumulation of *B. papyrifera* leaves was performed (**Figure 10**). The results showed that *CHS2*, *CHS7*, *DFR* and *CHI* were highly expressed in female plants at defoliation stage. By comparing the differences between female plants and male plants at different stages, *CHS2* gene had the largest difference, reaching 35.77 times.

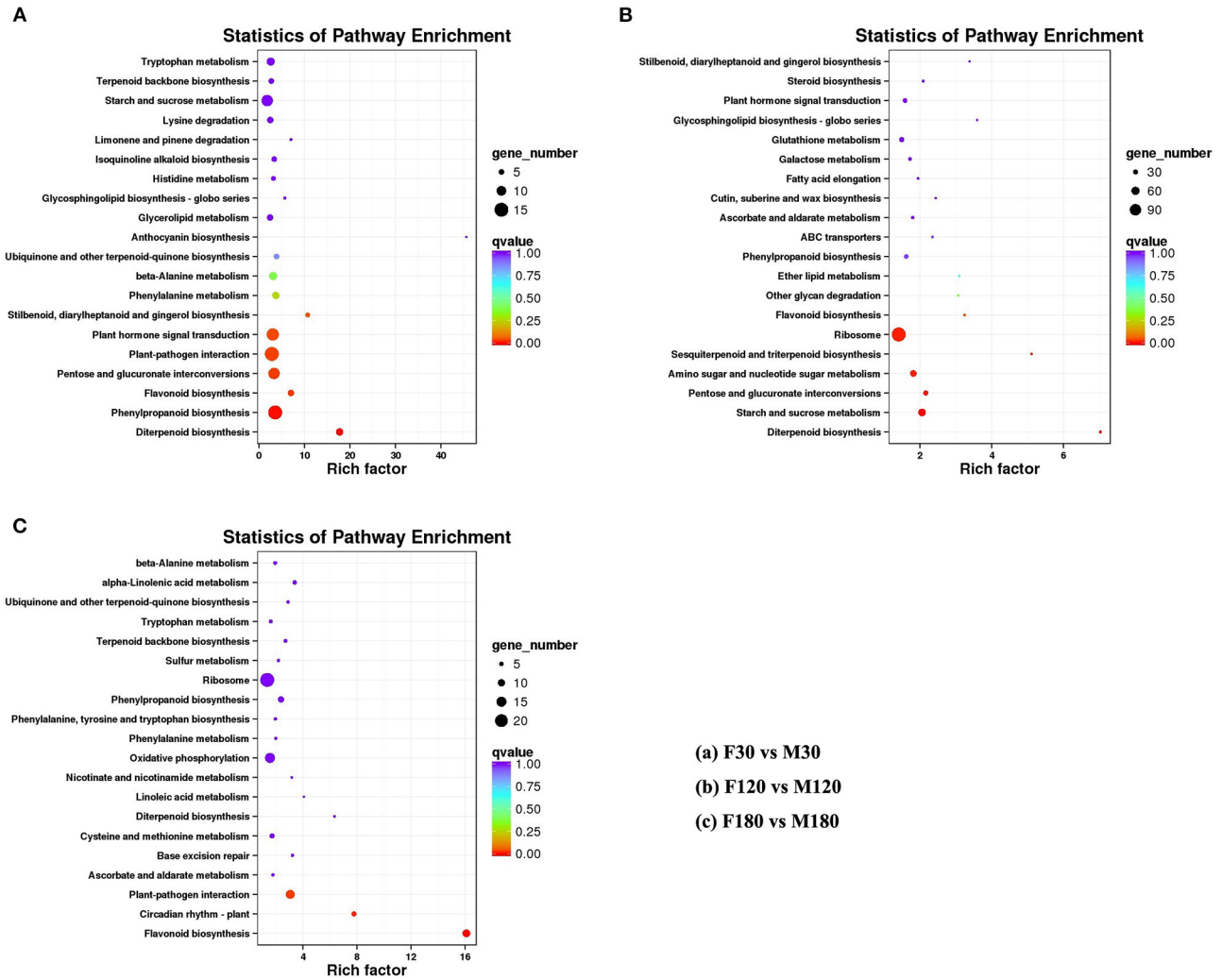


FIGURE 4 | Significant pathways of metabolites in developing *Broussonetia papyrifera* leaves by KEGG functional enrichment. **(A)** F30 vs. M30. **(B)** F120 vs. M120. **(C)** F180 vs. M180.

DISCUSSION

In the present study, we investigated sex differences in flavonoid accumulation of *B. papyrifera* leaves. The results showed that flavonoids gradually accumulated with developmental time, and the content in female plants was higher than that in male plants (**Figure 1**; **Supplementary Figure S2**). Moreover, the composition of flavonoids in female and male plants was very similar, and 16 kinds of flavonoids accumulated after flowering (**Figure 2**). Xu et al. (2009) studied the seasonal changes of total flavonoids in different parts of *B. papyrifera*, and the results showed that flavonoids in *B. papyrifera* leaves increased with the developmental stage, which was consistent with our research. Generally speaking, the different selection pressures faced by female and male plants lead to different nutritional requirements. Female plants have a higher investment in reproduction and show stronger compensatory mechanisms (Delph and Herlihy, 2012). Studies have found that in resource-rich environments, the sex ratio tends to favor female plants (Ward et al., 2002), while in resource-poor environments, the sex ratio tends to favor male plants (Dawson and Ehleringer, 1993; Chen and Li, 2014; Li et al., 2017). Therefore, we hypothesize that the different survival strategies of male and female plants of *B. papyrifera* leaves lead to differences in flavonoids.

Combined with the results of WGCNA and qRT-PCR analysis, *CHS2*, *CHS7*, *CHI* and *DFR* may the key genes for flavonoid accumulation of *B. papyrifera* leaves, and the expression of these genes was affected by sex differences and developmental stages (**Figures 5**, **9**, **10**; **Table 1**). *CHS* directs the phenylpropane metabolic pathway to the synthesis of flavonoids. In *B. papyrifera* leaves, *CHS* enzymes code for proteins exhibiting polyketide synthase activity that accept either p-dihydrocoumaroyl-CoA or p-coumaroyl-CoA as starter

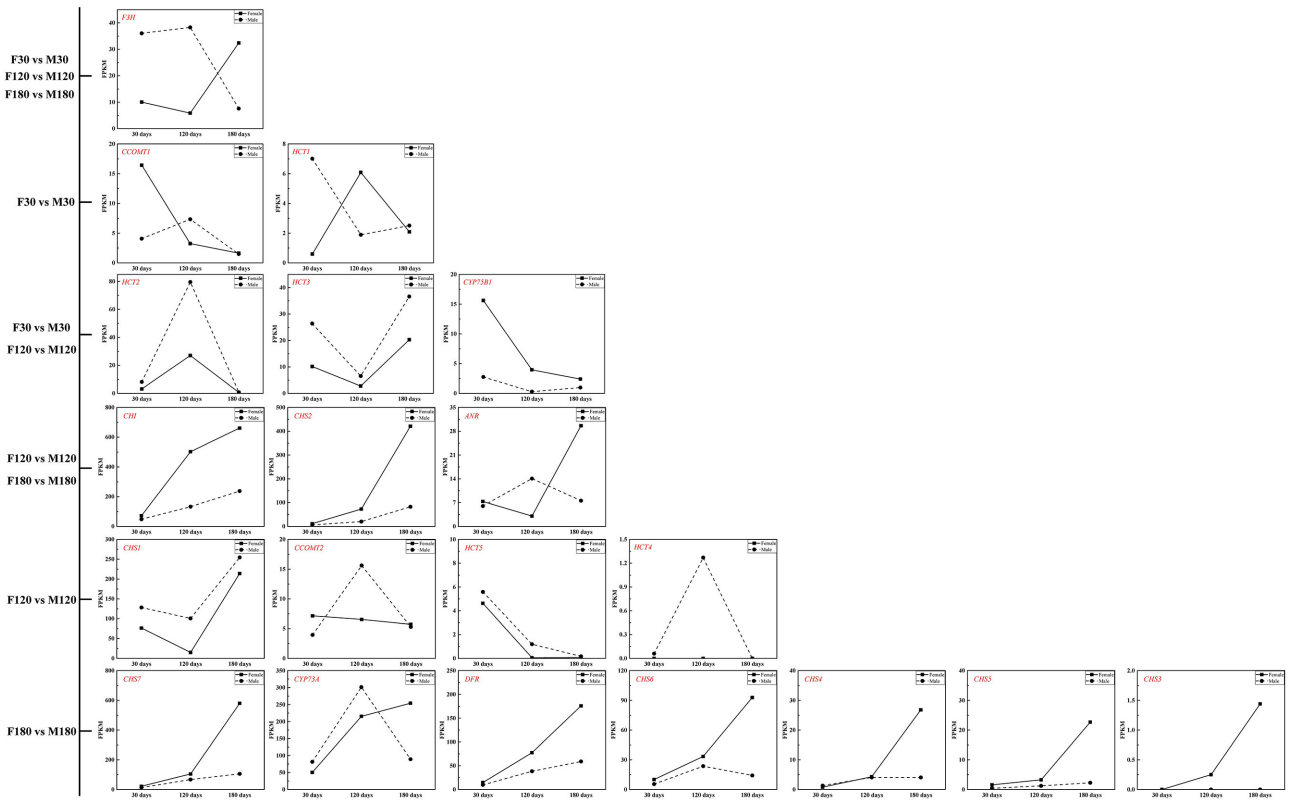


FIGURE 5 | Differentially expressed genes identified in the flavonoids biosynthesis pathway.

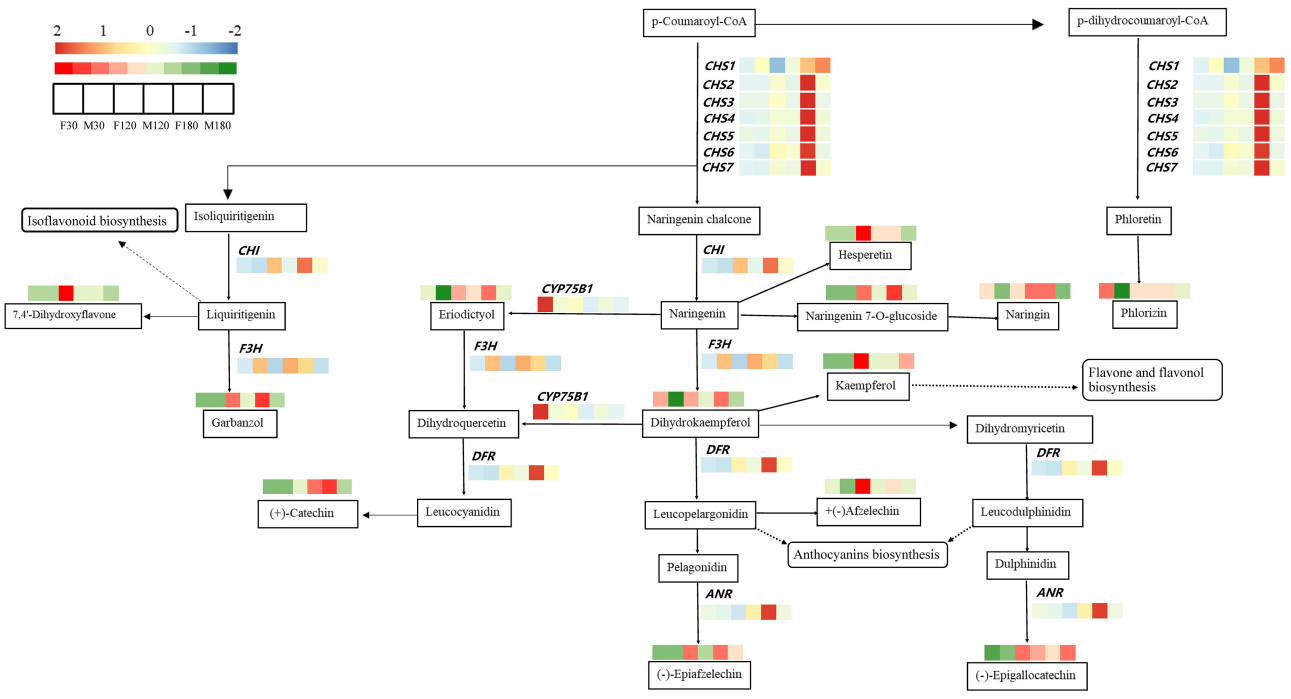


FIGURE 6 | The transcriptional profile of structural genes and metabolites in the flavonoids biosynthesis pathway of *Broussonetia papyrifera* leaves.

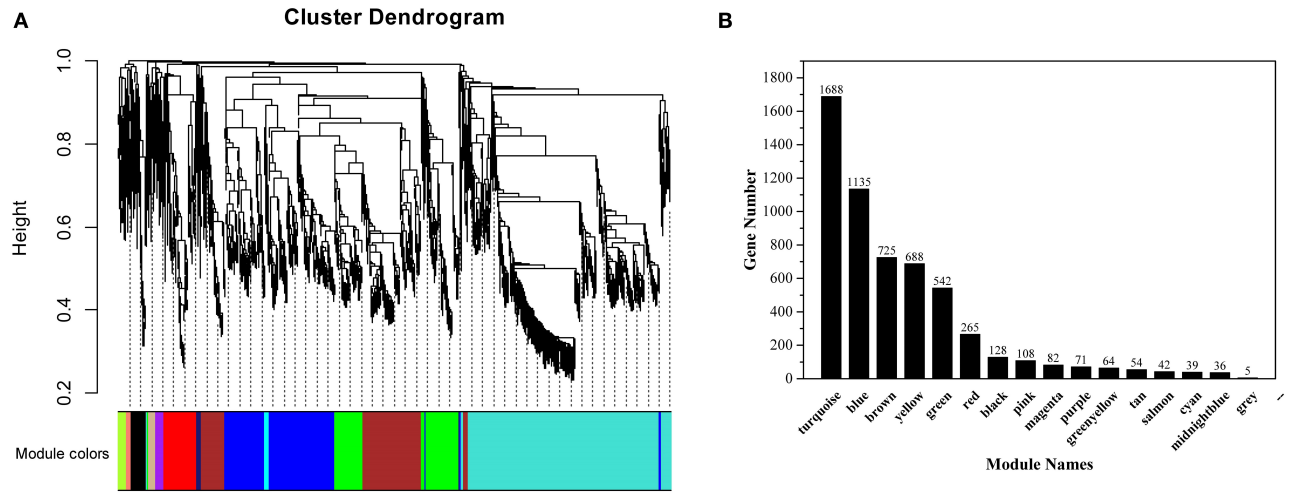


FIGURE 7 | Co-expression network analysis of RNA-seq and physiological trait data. **(A)** Clustering dendrogram of differentially expressed genes and modules identified by the weighted gene co-expression network analysis. **(B)** The number of genes in each module.

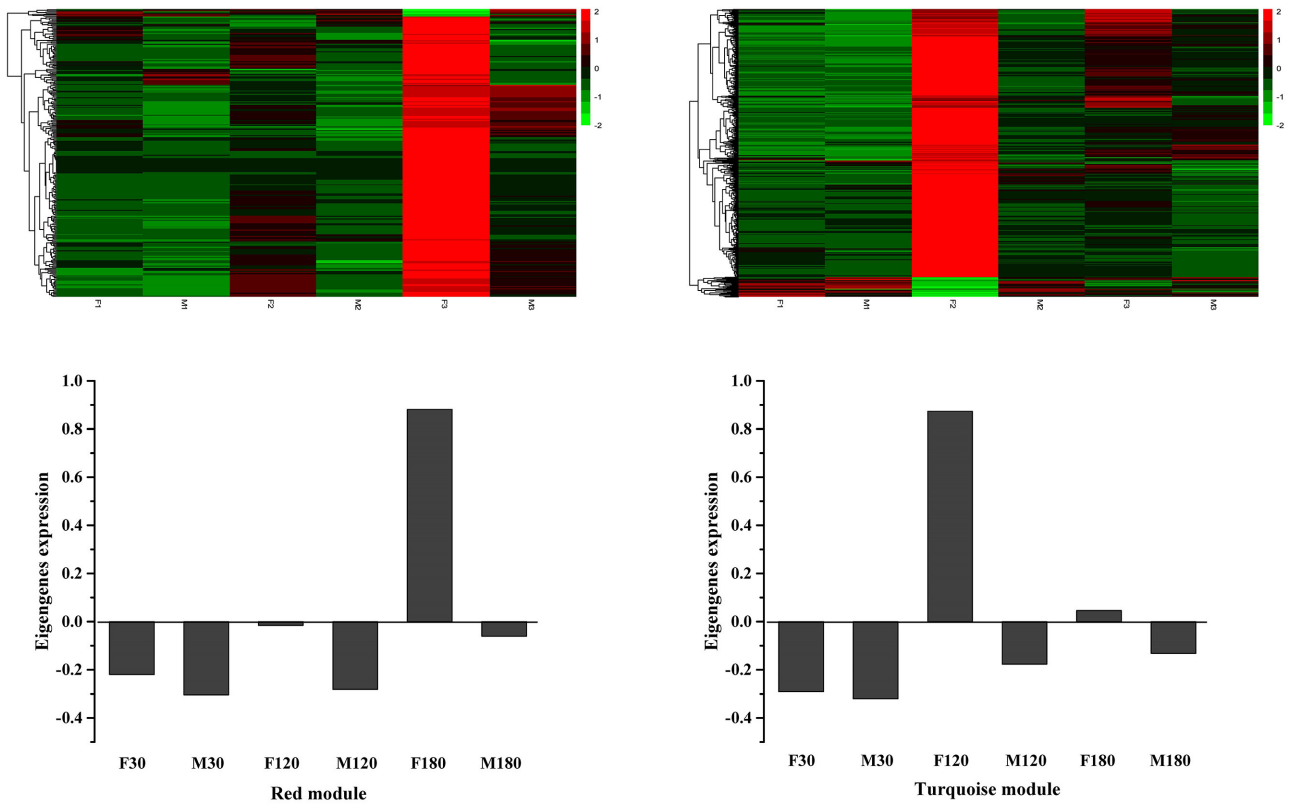


FIGURE 8 | Eigengenes expression profiles of red and turquoise modules.

CoA substrates, leading to production of phloretin, naringenin chalcone, and isoliquiritigenin (**Figure 6**), which supports the findings in apple (Yahyaa et al., 2017). *CHI* enzymes (EC 5.5.1.6) accept either naringenin chalcone or isoliquiritigenin as starter, leading to production of naringenin and liquiritigenin (**Figure 6**). *DFR* enzymes (EC 1.1.1.219) catalyz the NADPH-mediated and stereospecific reduction of dihydroflavonols to leucoanthocyanidins, which are the precursors for the formation

TABLE 1 | Identification of flavonoid-related genes in the developing *Broussonetia papyrifera* leaves.

Gene name	FPKM values[a]						Module
	F30	M30	F120	M120	F180	M180	
ANR	7.34	6.01	3.05	14.08	29.63	7.57	Red
CHS2	11.39	6.56	73.33	20.51	**421.1**	82.76	Red
CHS3	0	0	0.25	0	1.44	0	Red
CHS4	0.79	1.31	4.3	4.07	26.79	4.06	Red
CHS5	1.57	0.42	3.25	1.23	22.67	2.28	Red
CHS6	10.13	5.48	33.37	23.55	92.84	14.24	Red
CHS7	21.69	12.84	**104.69**	67.29	**580.22**	**105.64**	Red
DFR	14.67	9.69	77.47	38.38	**175.87**	59	Red
CHI	71.88	48.25	**502.77**	**132.94**	**661.16**	**237.24**	Turquoise

[a]*Expressed unigenes with FPKM values >100 were highlighted with bold font.*

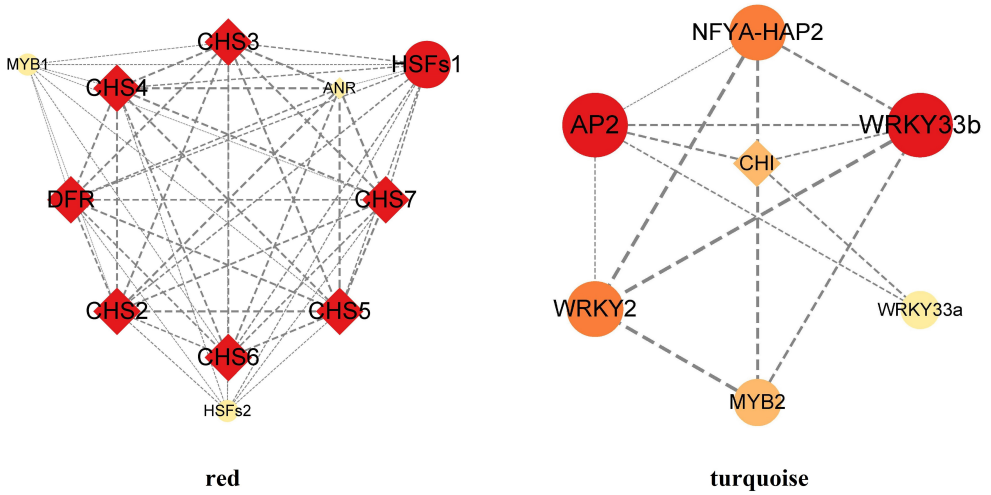


FIGURE 9 | Co-expression network related to flavonoids biosynthesis.

of anthocyanins, catechins, and proanthocyanidins (Fischer et al., 2003). Specifically, it accepts either dihydrokaempferol, dihydroquercetin or dihdydromyricerin as starter, leading to production of leucopelargonidin, leucocyanidin and leucodelphinidin (**Figure 6**). However, these results are based on statistical method, and the function of these genes needs to be further verified. So far, little was known about the flavonoid-regulated genes in *B. papyrifera* leaves. Relevant molecular studies have only demonstrated that four key enzyme families (*CHS*, *F3'H*, *I2'H*, and *DFR*) in the flavonoid synthesis pathway exhibited significant expansion (Peng et al., 2014; Feng et al., 2019). The high flavonoid abundance and strong disease resistance of *B. papyrifera* may originate from these expanded gene families.

CHS2, *CHS7*, *CHI* and *DFR* were highly expressed in female *B. papyrifera* (**Figures 5**, **10**). Most of them are upstream genes in flavonoids biosynthesis pathway. Correlation coefficients between modules and flavonoids showed that *CHS2*, *CHS7* and *DFR* was significantly correlated with garbanzol and naringenin 7-O-glucoside. *CHI* was significantly correlated with hesperetin, kaempferol, 7,4'-dihydroxyflavone and afzelechin (**Supplementary Figure S5**). Therefore, we speculate that the key to affecting the accumulation of flavonoids in *B. papyrifera* leaves of different sexes is the expression of upstream genes in flavonoids biosynthesis pathway, and we can increase the accumulation of flavonoids by enhancing the expression of upstream genes.

TFs that may interact with flavonoid genes were screened from two key modules closely related to flavonoids, and 8 reliable TFs were found (**Figure 9**). Among them, HSFs and MYB TFs were strongly correlated with *CHS*. *DFR*. AP2, NFYA-HAP2, WRKY and MYB TFs were strongly correlated with *CHI*.

The HSFs family mediates the activation of genes responsive to diverse stresses. The N-terminal region of the HSF protein sequence contains a highly conserved DNA-binding domain, which can accurately identify and bind the heat shock cis-elements (HSE) of the downstream target gene promoter (Wang et al., 2020). Our results suggest that HSFs1 and

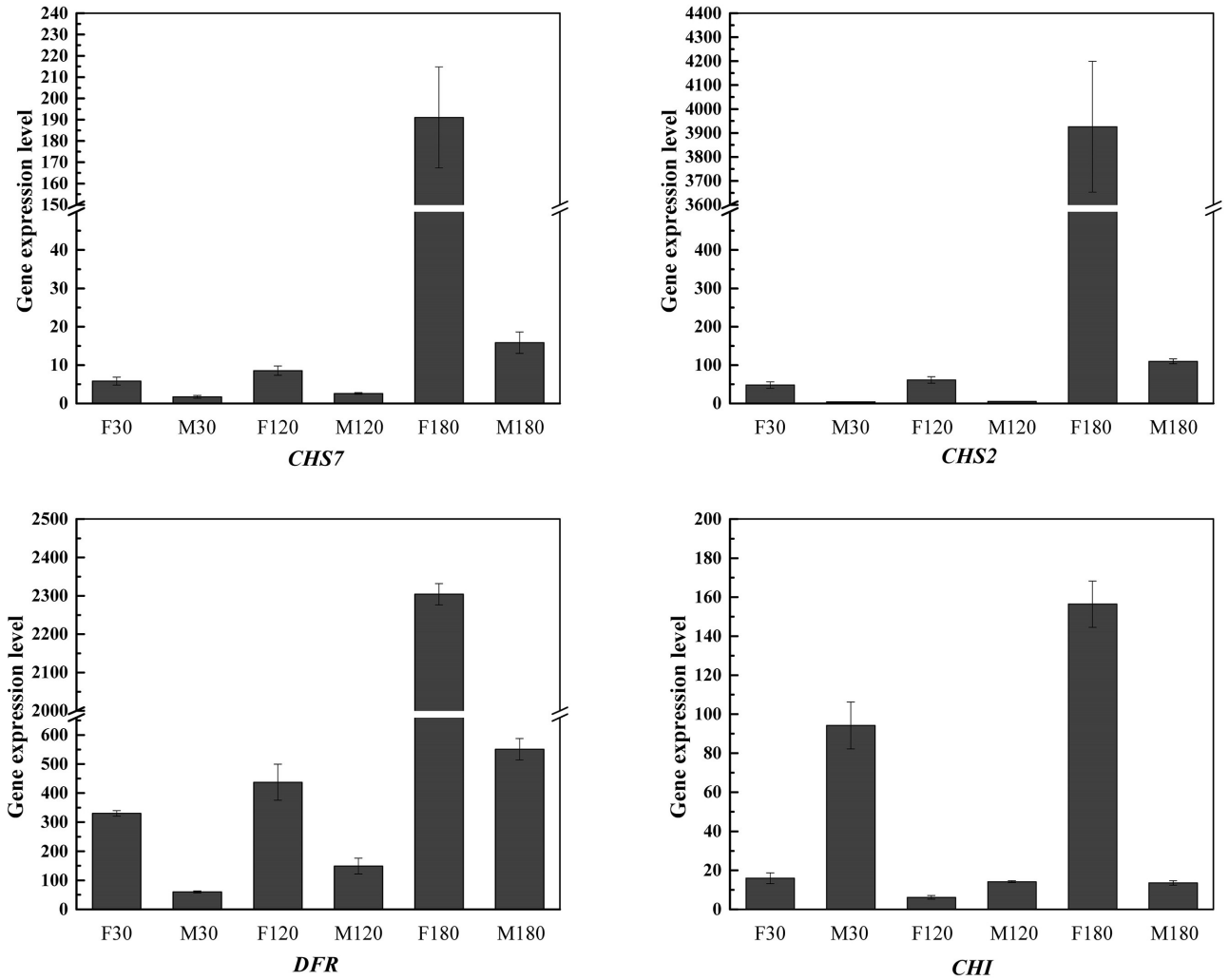


FIGURE 10 | qRT-PCR analysis of four structural genes associated with flavonoids biosynthesis.

HSFs2 genes in *B. papyrifera* may promote the expression of genes encoding related enzymes, such as *CHS* and *DFR* (**Figure 9**). The results are consistent with studies on other species. In soybean, HSFB2b can bind to the HSE cis-elements and promotes expression of *GmCHS* and *GmFLS* (Bian et al., 2020). In *Malus domestica*, MdHSFA8a can regulate the expression of *MdDFR*, *MdFLS*, and *MdANS* (Wang et al., 2020).

The MYB family is one of the largest families of regulatory proteins, and plays important roles in many secondary metabolic pathways (Zhang et al., 2018b). We found that the MYB1 gene was strongly associated with *CHS* and *DFR*, while the MYB2 gene was strongly associated with *CHI* (**Figure 9**). These genes may be involved in the flavonoid synthesis. In recent years, many MYB genes associated with the flavonoid biosynthesis have been isolated and characterized. Overexpression of GbMYBFL under the control of the *CaMV35S* promoter promoted the accumulation of flavonoids and anthocyanin in *G. biloba* (Zhang et al., 2018b). MYB6, a major regulator of flavonoid synthesis in *P. tomentosa*, up-regulated the expression of flavonoid biosynthetic genes (Wang et al., 2019).

The AP2/ERF TFs can directly or indirectly participate in multiple processes of plant development such as seed development, morphogenesis of organs such as flowers and fruits by responding to the regulation of ethylene, cytokinin and auxin; in addition to primary metabolism. AP2/ERF also has a significant effect on plant secondary metabolism, especially in regulating the synthesis of main medicinal active components (such as artemisinin, paclitaxel and lignin) (Zhang et al., 2018a). In citrus, the *CitCHIL*1 gene is transcriptionally activated by three AP2/ERF TFs, and significantly enhanced the accumulation of flavonoids

(Zhao et al., 2021), and our results supported this point (**Figure 9**).

The WRKY family is involved in plant growth, development, abiotic stress and regulation of secondary metabolites (Dai et al., 2021). Our results suggested that WRKY33a and WRKY33b may regular the transcription of *CHI*, which is involved in the production of naringenin and liquiritigenin in *B. papyrifera*. In addition, the WRKY family is also involved in regulating other pathways of flavonoid synthesis. In apple calli, MdWRKY11 up-regulated the expression of *F3H*, *FLS*, *DFR*, *ANS*, and *UFGT* and increased the accumulation of flavonoids and anthocyanin (Wang et al., 2018). In grape, overexpression of VvWRKY26 possibly promoted the accumulation of proanthocyanidin biosynthesis (Amato et al., 2017).

Although there are few reports on the regulation of flavonoid synthesis by NFYA-HAP2 TFs, the NFYA-HAP2 gene may be involved in the stress resistance *B. papyrifera*. For example, NFYA-HAP2 showed significant responses to heat shock treatment and *F. oxysporum* infection in *Brassica oleracea* (Kim et al., 2015).

At present, based on the statistical method, we have screened out the key genes and TFs that may regulate the synthesis of *B. papyrifera* flavonoids. To provide more detailed information about the function of the genes and TFs in *B. papyrifera*, it is necessary to use biological techniques such as yeast two-hybrid and yeast one-hybrid methods to verify their transcriptional activities and interactions with the target genes promoter in the future study.

CONCLUSION

In this study, we presented the integrative metabolome and transcriptome analysis of flavonoid biosynthesis genes in *B. papyrifera* leaves from the perspective of sex differentiation. A total of 192 flavonoids substances were detected in *B. papyrifera* leaves. The accumulation law of flavonoids was that they accumulated gradually with the development time, and the content in female plants was higher than that in male plants. The composition of flavonoids in female and male plants was very similar, and 16 kinds of flavonoids accumulated after flowering. Four key structural genes involved in flavonoids accumulation during leaves development were identified. Most of them are upstream genes in the flavonoids biosynthesis pathway, and overexpression of these genes increased the accumulation of flavonoids, such as garbanzol, naringenin 7-O-glucoside, hesperetin, kaempferol, 7,4'-dihydroxyflavone and afzelechin. Furthermore, 8 reliable TFs (two HSFs, two MYBP, three AP2/ERF, one NFYA-HAP2, and two WRKY) were found. Based on statistical methods and previous findings, these TFs may interact with flavonoid genes *CHI*, *CHS* and *DFR* in *B. papyrifera*. This study will be useful for further research on flavonoid biosynthesis in *B. papyrifera* associated molecular regulation.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. The RNA-Seq data presented in the study are deposited in the NCBI's Short Read Archive (SRA) database (http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi) under accession number PRJNA819010.

AUTHOR CONTRIBUTIONS

PJ: conceptualization, writing—original draft, and visualization. LC: investigation. ZY: supervision and funding acquisition. XZ: methodology and writing review and editing. ZW and DJ: data administration. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Key Projects of National Forestry and Grassland Bureau (201801), Postgraduate Scientific Research Innovation Project of Hunan Province (CX20200711), China Postdoctoral Science Foundation (2020M683592), and Natural Science Foundation of Shaanxi Province (2022JQ-202).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.900030/full#supplementary-material

Supplementary Figure S1 | Phenotypes of *Broussonetia papyrifera* leaves of different sexes at three developmental stages.

Supplementary Figure S2 | Differentially accumulated flavonoids in leaf samples.

Supplementary Figure S3 | Differentially expressed genes in leaf samples.

Supplementary Figure S4 | Analysis of network topology for various soft-thresholding powers.

Supplementary Figure S5 | Module-trait correlations. Each cell contains the corresponding correlation and *p* value. Red is positive correlation. Green is negative correlation.

Supplementary Figure S6 | KEGG functional analysis of red and turquoise modules in metabolism pathways.

Supplementary Figure S7 | Pearson correlation analysis between qRT-PCR results and RNA-seq results.

Supplementary Table S1 | The primers list of genes for qRT-PCR analysis.

Supplementary Table S2 | Types and relative contents of flavonoids in *Broussonetia papyrifera* leaves.

Supplementary Table S3 | KEGG analysis of differential metabolites.

Supplementary Table S4 | Transcriptome sequencing quality analysis.

Supplementary Table S5 | Annotation of differentially expressed genes.

Supplementary Table S6 | KEGG analysis of structural genes in the flavonoid biosynthesis pathway.

Supplementary Table S7 | Candidate co-express genes related to flavonoid biosynthesis.

REFERENCES

Agati, G., Azzarello, E., Pollastri, S., and Tattini, M. (2012). Flavonoids as antioxidants in plants: location and functional significance. *Plant Sci.* 196, 67–76. doi: 10.1016/j.plantsci.2012.07.014

Agati, G., Biricolti, S., Guidi, L., Ferrini, F., Fini, A., and Tattini, M. (2011). The biosynthesis of flavonoids is enhanced similarly by UV radiation and root zone salinity in *L. vulgare* leaves. *J. Plant Physiol.* 168, 204–212. doi: 10.1016/j.jplph.2010.07.016

Agati, G., Stefano, G., Biricolti, S., and Tattini, M. (2009). Mesophyll distribution of 'antioxidant' flavonoid glycosides in Ligustrum vulgare leaves under contrasting sunlight irradiance. *Ann. Bot.* 104, 853–861. doi: 10.1093/aob/mcp177

Al-Dossary, O., Alsubaie, B., Kharabian-Masouleh, A., Al-Mssallem, I., Furtado, A., and Henry, R. J. (2021). The jojoba genome reveals wide divergence of the sex chromosomes in a dioecious plant. *Plant J.* 108, 1283–1294. doi: 10.1111/tpj.15509

Amato, A., Cavallini, E., Zenoni, S., Finezzo, L., Begheldo, M., Ruperti, B., et al. (2017). A grapevine TTG2-like WRKY transcription factor is involved in regulating vacuolar transport and flavonoid biosynthesis. *Front. Plant Sci.* 7, 1979. doi: 10.3389/fpls.2016.01979

Banuelos, M. J., and Obeso, J. R. (2004). Resource allocation in the dioecious shrub *Rhamnus alpinus*: the hidden costs of reproduction. *Evol. Ecol. Res.* 6, 397–413.

Bian, X.-H., Li, W., Niu, C.-F., Wei, W., Hu, Y., Han, J.-Q., et al. (2020). A class B heat shock factor selected for during soybean domestication contributes to salt tolerance by promoting flavonoid biosynthesis. *New Phytol.* 225, 268–283. doi: 10.1111/nph.16104

Chang, C.-S., Liu, H.-L., Moncada, X., Seelenfreund, A., Seelenfreund, D., and Chung, K.-F. (2015). A holistic picture of Austronesian migrations revealed by phylogeography of Pacific paper mulberry. *Proc. Nat. Acad. Sci.* 112, 13537–13542. doi: 10.1073/pnas.1503205112

Chen, J., and Li, C. (2014). Sex-specific responses to environmental stresses and sexual competition of dioecious plants. *Chin. J. Appl. Environ. Biol.* 2014, 743–750. doi: 10.3724/SP.J.1145.2014.06008

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/sst080

Chen, X., Mao, X., Huang, J., Yang, D., Wu, J., Dong, S., et al. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucl. Acids Res.* 39(suppl_2), w316–w322. doi: 10.1093/nar/gkr483

Dai, X., Liu, J., and Li, S. (2021). Bioinformatics analysis of WRKY gene family in *Yulania biondii*. *J. Jiangsu For. Sci. Technol.* 48, 1–9. doi: 10.3969/j.issn.1001?7380.2021.06.001

Dawson, T. E., and Ehleringer, J. R. (1993). Gender-specific physiology, carbon isotope discrimination, and habitat distribution in boxelder, *Acer negundo*. *Ecology* 74, 798–815. doi: 10.2307/1940807

Delph, L. F., and Herlihy, C. R. (2012). Sexual, fecundity, and viability selection on flower size and number in a sexually dimorphic plant. *Evol. Int. J. Organ. Evol.* 66, 1154–1166. doi: 10.1111/j.1558-5646.2011.01510.x

Ding, T., Zhang, R., Zhang, H., Zhou, Z., Liu, C., Wu, M., et al. (2021). Identification of gene co-expression networks and key genes regulating flavonoid accumulation in apple (Malus × domestica) fruit skin. *Plant Sci.* 304, 110747. doi: 10.1016/j.plantsci.2020.110747

Falcone Ferreyra, M. L., Rius, S., and Casati, P. (2012). Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Front. Plant Sci.* 3, 222. doi: 10.3389/fpls.2012.00222

Feng, J., Dong, P., Li, R., Li, C., Xie, X., and Shi, Q. (2019). Effects of wood fiber properties on mold resistance of wood polypropylene composites. *Int. Biodeterior. Biodegr.* 140, 152–159. doi: 10.1016/j.ibiod.2019.04.005

Feng, W., Li, H., and Zheng, X. (2008). Chemical constituents from the leaves of *Broussonetia papyrifera*. *Acta Pharm. Sin.* 43, 173–180. doi: 10.16438/j.0513-4870.2008.02.005

Fischer, T. C., Halbwirth, H., Meisel, B., Stich, K., and Forkmann, G. (2003). Molecular cloning, substrate specificity of the functionally expressed dihydroflavonol 4-reductases from *Malus domestica* and *Pyrus communis* cultivars and the consequences for flavonoid metabolism. *Arch. Biochem. Biophys.* 412, 223–230. doi: 10.1016/S0003-9861(03)00013-4

Ghosh, R., Chakraborty, A., Biswas, A., and Chowdhuri, S. (2021). Identification of polyphenols from *Broussonetia papyrifera* as SARS CoV-2 main protease inhibitors using in silico docking and molecular dynamics simulation approaches. *J. Biomol. Struct. Dyn.* 39, 6747–6760. doi: 10.1080/07391102.2020.1802347

González-Gallego, J., García-Mediavilla, M. V., Sánchez-Campos, S., and Tuñón, M. J. (2010). Fruit polyphenols, immunity and inflammation. *Br. J. Nutr.* 104, S15–S27. doi: 10.1017/S0007114510003910

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Guo, F., Feng, L., Huang, C., Ding, H., Zhang, X., Wang, Z., et al. (2013). Prenylflavone derivatives from Broussonetia papyrifera, inhibit the growth of breast cancer cells *in vitro* and *in vivo*. *Phytochem. Lett.* 6, 331–336. doi: 10.1016/j.phytol.2013.03.017

Guo, Y., Gao, C., Wang, M., Fu, F.-,f., El-Kassaby, Y. A., Wang, T., et al. (2020). Metabolome and transcriptome analyses reveal flavonoids biosynthesis differences in *Ginkgo biloba* associated with environmental conditions. *Ind. Crops Prod.* 158, 112963. doi: 10.1016/j.indcrop.2020.112963

Harborne, J. B., and Williams, C. A. (2000). Advances in flavonoid research since 1992. *Phytochemistry* 55, 481–504. doi: 10.1016/S0031-9422(00)00235-1

Kim, Y.-W., Jung, H.-J., Park, J.-I., Hur, Y., and Nou, I.-S. (2015). Response of NBS encoding resistance genes linked to both heat and fungal stress in *Brassica oleracea*. *Plant Physiol. Biochem.* 86, 130–136. doi: 10.1016/j.plaphy.2014.11.009

Koczka, N., Móczár, Z., Stefanovits-Bányai, É., and Ombódi, A. (2015). Differences in antioxidant properties of ginkgo leaves collected from male and female trees. *Acta Pharm.* 65, 99–104. doi: 10.1515/acph-2015-0001

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinform.* 12, 1–16. doi: 10.1186/1471-2105-12-323

Li, H., Li, T., Tu, B., Kou, Y., and Li, X. (2017). Host species shapes the co-occurrence patterns rather than diversity of stomach bacterial communities in pikas. *Appl. Microbiol. Biotechnol.* 101, 5519–5529. doi: 10.1007/s00253-017-8254-0

Li, M., Yan, L., and Li, H. (2012). Improvement of paper mulberry tolerance to abiotic stresses by ectopic expression of tall fescue FaDREB1. *Tree Physiol.* 32, 104–113. doi: 10.1093/treephys/tpr124

Lillo, C., Lea, U. S., and Ruoff, P. (2010). Nutrient depletion as a key factor for manipulating gene expression and product formation in different branches of the flavonoid pathway. *Plant Cell Environ.* 31, 587–601. doi: 10.1111/j.1365-3040.2007.01748.x

Lu, W., Jianhua, T., Jucheng, W., Peng, S., Yicai, H., and Xuebin, G. (2018). Comparative analysis on total flavonoids content of *Hippophae rhamnoides* linn. leaf in Shanxi. *South African J. Bot.* 47, 33. doi: 10.1016/j.sajb.2012.11.005

Park, J.-Y., Yuk, H. J., Ryu, H. W., Lim, S. H., Kim, K. S., Park, K. H., et al. (2017). Evaluation of polyphenols from *Broussonetia papyrifera* as coronavirus protease inhibitors. *J. Enzyme Inhib. Med. Chem.* 32, 504–512. doi: 10.1080/14756366.2016.1265519

Peng, X., Liu, H., Chen, P., Tang, F., Hu, Y., Wang, F., et al. (2019). A chromosome-scale genome assembly of paper mulberry (*Broussonetia papyrifera*) provides new insights into its forage and papermaking usage. *Mol. Plant* 12, 661–677. doi: 10.1016/j.molp.2019.01.021

Peng, X., Teng, L., Wang, X., Wang, Y., and Shen, S. (2014). De novo assembly of expressed transcripts and global transcriptomic analysis from seedlings of the paper mulberry (*Broussonetia kazinoki* x *Broussonetia papyifera*). *PLoS ONE* 9, e108613. doi: 10.1371/journal.pone.0097487

Pisonero-Vaquero, S., Garcia-Mediavilla, M. V., Jorquera, F., Majano, P. L., Benet, M., Jover, R., et al. (2014). Modulation of PI3K-LXRα-dependent lipogenesis mediated by oxidative/nitrosative stress contributes to inhibition of HCV replication by quercetin. *Lab. Investig.* 94, 262–274. doi: 10.1038/labinvest.2013.156

Randriamanana, T. R., Nybakken, L., Lavola, A., Aphalo, P. J., Nissinen, K., and Julkunen-Tiitto, R. (2014). Sex-related differences in growth and carbon allocation to defence in *Populus tremula* as explained by current plant defence theories. *Tree Physiol.* 34, 471–487. doi: 10.1093/treephys/tpu034

Stafford, H. A. (1991). Flavonoid evolution: an enzymic approach. *Plant Physiol.* 96, 680–685. doi: 10.1104/pp.96.3.680

Tattini, M., Galardi, C., Pinelli, P., Massai, R., and Agati, G. (2010). Differential accumulation of flavonoids and hydroxycinnamates in leaves of *Ligustrum vulgare* under excess light and drought stress. *New Phytol.* 163, 547–561. doi: 10.1111/j.1469-8137.2004.01126.x

Wang, L., Lu, W., Ran, L., Dou, L., Yao, S., Hu, J., et al. (2019). R2R3-MYB transcription factor MYB 6 promotes anthocyanin and proanthocyanidin biosynthesis but inhibits secondary cell wall formation in *Populus tomentosa*. *Plant J.* 99, 733–751. doi: 10.1111/tpj.14364

Wang, N., Liu, W., Yu, L., Guo, Z., Chen, Z., Jiang, S., et al. (2020). HEAT SHOCK FACTOR A8a modulates flavonoid synthesis and drought tolerance. *Plant Physiol.* 184, 1273–1290. doi: 10.1104/pp.20.01106

Wang, N., Liu, W., Zhang, T., Jiang, S., Xu, H., Wang, Y., et al. (2018). Transcriptomic analysis of red-fleshed apples reveals the novel role of MdWRKY11 in flavonoid and anthocyanin biosynthesis. *J. Agric. Food Chem.* 66, 7076–7086. doi: 10.1021/acs.jafc.8b01273

Ward, J., Dawson, T., and Ehleringer, J. (2002). Responses of *Acer negundo* genders to interannual differences in water availability determined from carbon isotope ratios of tree ring cellulose. *Tree Physiol.* 22, 339–346. doi: 10.1093/treephys/22.5.339

Wu, X., Liu, J., Meng, Q., Fang, S., Kang, J., and Guo, Q. (2021). Differences in carbon and nitrogen metabolism between male and female *Populus cathayana* in response to deficient nitrogen. *Tree Physiol.* 41, 119–133. doi: 10.1093/treephys/tpaa108

Xu, Y., Guo, Z., Tan, L., Bu, X., and Long, L. (2009). Seasonal variation of total flavonoids contents for different parts of *Broussonetia papyrifera*. *Res. Pract. Chin. Med.* 23, 16–17. doi: 10.13728/j.1673-6427.2009.04.029

Xu, Z., Dong, M., Peng, X., Ku, W., Zhao, Y., and Yang, G. (2019). New insight into the molecular basis of cadmium stress responses of wild paper mulberry plant by transcriptome analysis. *Ecotoxicol. Environ. Saf.* 171, 301–312. doi: 10.1016/j.ecoenv.2018.12.084

Yahyaa, M., Ali, S., Davidovich-Rikanati, R., Ibdah, M., Shachtier, A., Eyal, Y., et al. (2017). Characterization of three chalcone synthase-like genes from apple (*Malus* x domestica Borkh.). *Phytochemistry* 140, 125–133. doi: 10.1016/j.phytochem.2017.04.022

Yang, Y. (2018). *Study on Extraction, Purification and Antioxidant Activity of Total Flavonoids from Apocynum venetum*. Master, South China University of Technology.

Yang, Y., Jiang, H., Wang, M., Korpelainen, H., and Li, C. (2015). Male poplars have a stronger ability to balance growth and carbohydrate accumulation than do females in response to a short-term potassium deficiency. *Physiol. Plant.* 155, 400–413. doi: 10.1111/ppl.12325

Ying, H.-Z., Liu, Y.-H., Yu, B., Wang, Z.-Y., Zang, J.-N., and Yu, C.-H. (2013). Dietary quercetin ameliorates nonalcoholic steatohepatitis induced by a high-fat diet in gerbils. *Food Chem. Toxicol.* 52, 53–60. doi: 10.1016/j.fct.2012.10.030

Zhang, L., Zhang, Z., Fang, S., Liu, Y., and Shang, X. (2021). Integrative analysis of metabolome and transcriptome reveals molecular regulatory mechanism of flavonoid biosynthesis in *Cyclocarya paliurus* under salt stress. *Ind. Crops Prod.* 170, 113823. doi: 10.1016/j.indcrop.2021.113823

Zhang, M., Fang, Y., Ji, Y., Jiang, Z., and Wang, L. (2013). Effects of salt stress on ion content, antioxidant enzymes and protein profile in different tissues of *Broussonetia papyrifera*. *South Afric. J. Bot.* 85, 1–9. doi: 10.1016/j.sajb.2012.11.005

Zhang, Q., Chen, J., Li, L., Zhao, M., Zhang, M., and Wang, Y. (2018a). Research progress on plant AP2/ERF transcription factor family. *Biotechnol. Bull.* 34, 1–7. doi: 10.13560/j.cnki.biotech.bull.1985.2017-1142

Zhang, W., Xu, F., Cheng, S., and Liao, Y. (2018b). Characterization and functional analysis of a MYB gene (GbMYBFL) related to flavonoid accumulation in *Ginkgo biloba*. *Genes Genom.* 40, 49–61. doi: 10.1007/s13258-017-0609-5

Zhao, C., Liu, X., Gong, Q., Cao, J., Shen, W., Yin, X., et al. (2021). Three AP2/ERF family members modulate flavonoid synthesis by regulating type IV chalcone isomerase in citrus. *Plant Biotechnol. J.* 19, 671–688. doi: 10.1111/pbi.13494

Zhao, H. (2011). *Quality Evaluation of Broussonetia papyrifera and the Intra-Specific Variation Under Heavy Metal Stress*. Master, Fujian University of Traditional Chinese Medicine.

Zhao, X., Liu, J., Xia, X., Chu, J., Wei, Y., Shi, S., et al. (2014). The evaluation of heavy metal accumulation and application of a comprehensive bio-concentration index for woody species on contaminated sites in Hunan, China. *Environ. Sci. Pollut. Res.* 21, 5076–5085. doi: 10.1007/s11356-013-2393-3

Zheng, H., Huang, B., Qin, L., and Zhang, Q. (2002). Biological character and resources distribution *Broussonetia*. *Chin. Wild Plant Resour.* 21, 11–13. doi: 10.3969/j.issn.1006-9690.2002.06.004

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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.

frontiers | Frontiers in Pharmacology

ORIGINAL RESEARCH
published: 26 May 2022
doi: 10.3389/fphar.2022.907826

A Novel Chinese Herbal and Corresponding Chemical Formula for Cancer Treatment by Targeting Tumor Maintenance, Progression, and Metastasis

Ying-Chyi Song[1]*, *Der-Yen Lee*[1] and *Pei-Yen Yeh*[2]*

[1]Graduate Institute of Integrated Medicine, College of Chinese Medicine, China Medical University, Taichung, Taiwan, [2]TCM division, Jin-Mi company, Taipei, Taiwan

We characterized a so-called "heirloom recipe" Chinese herbal formula (temporarily named Formula X) that contains five Chinese medical botanical drugs, Huang-Lian (*Coptis chinensis* Franch. [Ranunculaceae]), Huang-Qin (*Scutellaria baicalensis* Georgi [Lamiaceae]), Bai-Wei (*Vincetoxicum atratum* (Bunge) C. Morren and Decne. [Apocynaceae]), E-Zhu (*Curcuma aromatica* Salisb. [Zingiberaceae]) and Bai-Zhu (*Atractylodes macrocephala* Koidz. [Asteraceae]). Formula X inhibited the growth of various cancer cells and decreased the expression levels of a panel of proteins, including CD133, Myc, PD-L1, and Slug, in cancer cells. We further found that the inhibition of growth and protein expression were exerted by Huang-Lian, Huang-Qin, and Bai-Wei (formula HHB), which exhibited the same biological effects as those of Formula X. Furthermore, we selected three active chemicals, berberine, baicalin, and saponin from Huang-Lian, Huang-Qin, and Bai-Wei, respectively, to produce a chemical formulation (formula BBS), which exhibited similar effects on cell growth and protein expression as those induced by formula HHB. Both the formulae HHB and BBS suppressed tumor growth in an animal study. Moreover, they decreased the protein levels of Myc and PD-L1 in tumor cells *in vivo*. In summary, we established a novel Chinese herbal formula and a chemical formula that targeted three important processes, tumor maintenance (tumor stem cells), progression, and metastasis, and that influenced the response of tumors to host immunosuppression, for the potentially effective treatment of cancer patients.

Keywords: cancer, botanical drug, Huang-Lian (*Coptis chinensis* Franch.), Huang-Qin (*Scutellaria baicalensis* Georgi), Bai-Wei (*Vincetoxicum atratum* (Bunge))

OPEN ACCESS

Edited by:
Richard Spjut,
World Botanical Associates, Inc., United States

Reviewed by:
Yuyang Zhou,
Soochow University, China
Viqar Syed,
Uniformed Services University of the Health Sciences, United States

***Correspondence:**
Ying-Chyi Song
songyingchyi@gmail.com
Pei-Yen Yeh
yehpeiy@yahoo.com.tw

Specialty section:
This article was submitted to Ethnopharmacology, a section of the journal Frontiers in Pharmacology

Received: *30 March 2022*
Accepted: *16 May 2022*
Published: *26 May 2022*

Citation:
Song Y-C, Lee D-Y and Yeh P-Y (2022) A Novel Chinese Herbal and Corresponding Chemical Formula for Cancer Treatment by Targeting Tumor Maintenance, Progression, and Metastasis. Front. Pharmacol. 13:907826. doi: 10.3389/fphar.2022.907826

INTRODUCTION

Traditional Chinese herbal medicine is widespread in Asia, particularly in Taiwan, China, Japan and Korea. It was developed from a philosophical basis different from that of modern Western medicine. A large number of traditional Chinese medicines (TCM) have still not been fully investigated using modern scientific knowledge. How TCM should be used is controversial at present (Corson and Crews, 2007; Xu and Xia, 2019), and the main concern might be that there is no well-established

system to identify and characterize the basis of TCM, such as two "Chi", "Ying," and "Yang", in the human body. Furthermore, unlike pure chemical drugs, whose history, development, and production can be traced and whose biological functions at the cellular and molecular levels have been characterized, Chinese herbal formulae are usually composed of several Chinese botanical drugs; thus, it is difficult to clearly identify the underlying mechanism of their biological effects. However, TCM has been the major medical remedy in Asia for thousands of years and is used in many aspects at present. TCM should be taken as a huge repository of potentially new medical materials and concepts and deserves in-depth exploration.

The fight against cancer has been long. Although we have continuously found new strategies for cancer treatment, their efficacy is restricted, or they have failed due to the genomic instability of cancer cells. Various abnormally mechanisms are developed in cancer cells in response to therapeutic treatment, which eventually cause drug resistance of cancer cells. Multitarget or combinatory molecular targeted therapy might be a more effective cancer treatment (Vasan et al., 2019). Along this line of thinking, Chinese herbal formula, which usually contains several components and potentially affect multiple targets, might be good candidates for cancer treatment. Although investigating how Chinese herbal formulae should be used or fit into modern medicine requires more effort, studying these formulae can at least serve as a starting point for the discovery of potential cancer therapies.

An interesting concept in the theory of TCM is that there are two opposite elements in the human body, Yin-Yang and cold-heat. Among them, the terms heat and cold might not merely represent measurable body temperature and are used to describe two counteracting elements in the human body. Although this concept has not been well elucidated and accepted by modern medical science, these two opposite elements are thought to form a Yin-Yang or cold-heat buffer system to maintain normal physiological conditions. Any intrinsic or extrinsic factors that alter the balance of the buffer systems might cause human disease. In particular, "overheat" inside the body might induce inflammation, and inflammation might increase the level of "heat". This finding could be interpreted as the heat being like a cellular factor, the level of which is auto-loop increased. Numerous studies have indicated that inflammation is highly associated with the initiation and progression of tumors (Crusz and Balkwill, 2015; Greten and Grivennikov, 2019). Therefore, it is believed in TCM that releasing the "overheat" is expected to effectively prevent or treat tumors.

The botanical drugs Huang-Lian (*Coptis chinensis* Franch. [Ranunculaceae]), Huang-Qin (*Scutellaria baicalensis* Georgi [Lamiaceae]), and Bai-Wei (*Vincetoxicum atratum* (Bunge) C. Morren and Decne. [Apocynaceae]) belong to a family that is classified as having abilities to remove heat and toxins in TCM. A rapidly accumulating number of studies indicates that this family of botanical drugs exerts anticancer effects either alone or with other tumor therapeutic agents. Multiple molecules that these botanical drugs target have been identified to be involved in various oncogenic processes, including cell growth, metastasis, drug resistance and in the response of tumor cells to host immunosuppression (Zhang et al., 2017; Luo et al., 2019).

In addition to the formulae recorded in classic Chinese medical books, many formulae are kept as heirloom recipes and are commonly used in the Chinese community. Although there may be no comprehensively solid evidence to support the function of such formulae, some of them have been used for a long time and are probably helpful for human health. It might be a loss that those formulae will eventually be distorted or lost over time if we view TCM negatively without even performing further characterization. Collecting and testing such formulae in detail is necessary and important for the reasonable usage of Chinese herbal formulae.

We utilized a formula (temporarily named Formula X) comprising five medical botanical drugs, Huang-Lian (HL; *Coptis chinensis* Franch. [Ranunculaceae]), Huang-Qin (HQ; *Scutellaria baicalensis* Georgi [Lamiaceae]), Bai-Wei (BW; *Vincetoxicum atratum* (Bunge) C. Morren and Decne. [Apocynaceae]), E-Zhu (EZ; *Curcuma aromatica* Salisb. [Zingiberaceae]), and Bai-Zhu (BZ; *Atractylodes macrocephala* Koidz. [Asteraceae]). This formula was originally used for anti-fever and anti-inflammation. We tried to identify whether it has an anticancer effect because four components of this formula (HL, HQ, BW, and EZ) have been reported to exert anticancer activity (Ikemoto et al., 2000; Jeon et al., 2011; Wang et al., 2017; Zhang et al., 2017; Gong et al., 2019; Luo et al., 2019; Zhang et al., 2019). In this study, we found that the water extract of Formula X had wide anticancer activity and suppressed the expression of various oncogenic genes, including Her2, estrogen receptor (ER), CD133, Myc, PD-L1 and Slug. We further deconstructed Formula X and then characterized every single component and the various combinations among them, we found that the extract from three components, HL, HQ, and BW together, exhibited the same inhibitory effect on cell growth and gene expression as that from Formula X. By searching previous findings about HL, HQ and BW, three chemicals, berberine, baicalin and saponin, which were used to formulate a new chemical formula (formula BBS), were selected for further studies (Katoh, 2011; Bretones et al., 2015; Stine et al., 2015; Zhang et al., 2017; Zhang et al., 2019). Formula BBS showed the same effect on cell growth and gene expression as that of the herbal formulae. Both the formulae HHB and BBS suppressed tumor growth and the protein expression of Myc and PD-L1 in tumor cells in an animal study.

MATERIALS AND METHODS

Cell Culture and Chemicals

The culture medium (Corning Mediatech, Inc.) and cancer cells used in this study was DMEM for MCF7 (breast cancer) and Huh7 (hepatoma) cells, RPMI1640 for T47D (breast cancer), HCC 1954 (breast cancer), AGS (gastric cancer), LoVo (colon cancer), Mia-Paca2 (pancreatic cancer) and U2OS (osteosarcoma) cells, and DMEM/F12 for MDA-MB231 (breast cancer) cells. The medium was supplemented with 10% fetal bovine serum (FBS). The cells were incubated in a humidified incubator with 5% CO_2 at 37°C. The drugs

berberine, baicalin and saponin were purchased from Sigma. They were prepared with concentrations of 10 mg/ml as a stock solution and stored in a -20 °C refrigerator.

Preparation of Botanical Drug Extract

We obtained botanical drugs from commercial Chinese medical herbal stores as ordinary people would. The botanical drugs were imported from China, and they have been regulated and routinely checked by Taiwan's food and drug administration. The botanical drugs were directly dried materials not subjected to other manufacturing processes. Ten Gram of a single dried botanical drug was added to 100 ml of water (concentration indicated as 100 mg/ml), and then sterilized by high-pressure saturated steam at 121°C for 20 min with an autoclave sterilizer. The supernatants were stored at −20°C. To prepare a formula containing multiple botanical drugs, an equal amount of each herbal extract was mixed. The concentrations indicated in the figures mean the concentration of every single botanical drug in the herbal formula. The identities of the botanical drugs were determined by LC-ESI-MS (detailed method and results are shown in the supplemental method and **Supplementary Figure S1**) in Lab. The chemical structure of berberine, baicalin and C21 steroidal saponin, the major constituent in the extract from Huang-Lian, Huang-Qin, and Bai-Wei respectively, was derived from the previous studies (Liang et al., 2007; Zhou et al., 2016; Luo et al., 2019) and shown in the **Supplementary Figure S2**. The samples of each botanical drug were stored for further identification and can be supplied upon request. To maintain the consistency of the biological function across different lots, we further performed a functional assay based on a cytotoxicity assay and Western blotting of specific proteins in each preparation.

Evaluation of Cell Number

The effect of the botanical drug extract on cell growth was determined by a tetrazolium-based semiautomated colorimetric assay (MTT assay). Cells were seeded at a density of 3,000 cells/well in a 96-well plate. After overnight culture, the cells were treated with various combinations of challenging agents for 4 days as indicated in the figures. Cell numbers were evaluated by the MTT assay with an ELISA reader at OD_{540}.

Western Blotting Analysis

The change in the protein expression profile was determined by Western blotting analysis. After administration of various treatments as indicated in the figures, the cells were lysed with RIPA solution containing a cocktail of protease inhibitors. Aliquots of lysates were subjected to Western blotting analysis. Anti-CD133 and anti-estrogen receptor antibodies were purchased from Santa Cruz Biotechnology. Antibodies against Her2, EGFR, Myc, Slug and PD-L1 were obtained from Cell Signaling Technology. The images were developed with a chemiluminescence reagent.

Animal Study

To establish the syngeneic mouse model of pancreatic tumors, Pan18 cells (GFP-LUC-tagged) derived from pancreatic tumors in EKP (elastase-CreER; LSL-KrasG12D; p53+/−) mice, kindly gifted from Dr. Yuh-Pyng Sher (China Medical University, Taichung, Taiwan), were used.

Female C57BL/6 mice, 6 weeks of age, were obtained from the National Laboratory Animal Breeding and Research Center (Taipei, Taiwan). All animals were housed at the Animal Center of China Medical University (CMU) and maintained in accordance with the institutional animal care protocol. All animal studies were approved by the animal committee of CMU (2017–128). To assess the therapeutic effect of the herbal or chemical formula, tumors were inoculated with 1×10^5 Pan18

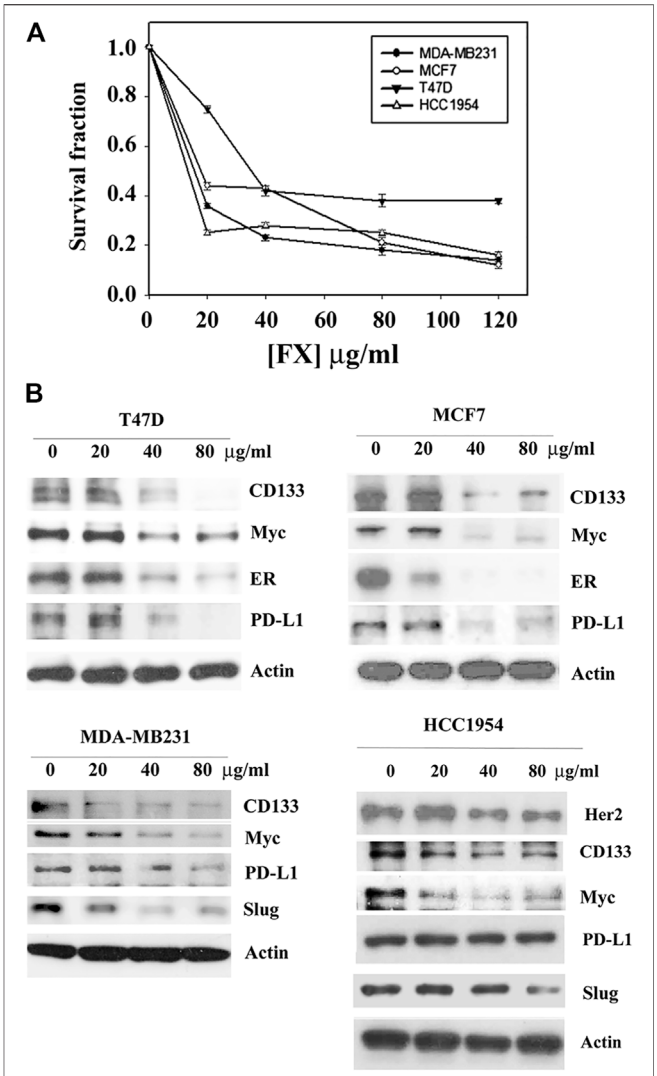


FIGURE 1 | Formula X inhibits cell growth and suppresses oncogenic protein expression in different types of breast cancer cells. **(A)** The indicated cells were treated with a serial dose of Formula X (FX) for 4 days. The level of surviving cells was determined by MTT assay. All data were obtained from at least two independent experiments. The survival fraction showed the fold change relative to the levels of the untreated control. The data are presented as the mean ± SD. **(B)** The cells were treated with various doses of Formula X for 24 h. The lysates were prepared, and the effect of Formula X on the selected protein expression as indicated in the figure was characterized by Western blotting analysis. Actin staining served as a loading control.

cells into the abdominal region of the mouse. After 7 days or 31 days, the mice were orally administered (p.o.) drug a concentration of 50 mg/kg three times a week (once every 2-3 days). The tumor volumes were monitored every 2-3 days after tumor implantation. Tumor sizes were calculated according to the following formula: (length × width^2)/2.

The lysates of tumor cells were prepared from sacrificed mice and subjected to Western blotting to detect the protein levels of Myc and PD-L1.

Statistical Test

Results were expressed as the mean ± SD/SEM. Differences between groups were assessed by Student's t-test. p values less than 0.05 were considered significant.

RESULTS

Formula X Inhibits Cell Growth and Suppresses Oncogenic Protein Expression in Breast Cancer Cells

The effect of Formula X on cancer cell growth was first determined in breast cancer cells. Four breast cancer cell lines that represent different types of breast cancer were selected for the study. T47D and MCF7 are estrogen receptor (ER)-positive luminal cancer cells, MDA-MB231 is a triple-negative basal cancer cell line, and HCC1954 is a Her2-overexpressing breast cancer cell line. Formula X inhibited the growth of all 4 cell lines in a dose-dependent manner (**Figure 1A**). Next, we examined the impact of Formula X on the levels of several oncogenic proteins. The Myc protein is an important oncogenic protein that is widely involved in cancer cell growth and progression (Bretones et al., 2015; Stine et al., 2015). The Slug protein is related to cancer cell metastatic ability and might cooperate with Myc to play a critical role in the status of tumor stem cells (Katoh, 2011; Zhou et al., 2019). CD133 is a marker of tumor stem cells (Jang et al., 2017). The PD-L1 protein helps cancer cells escape immunosuppression (Bellmunt et al., 2017). Taken together, we tried to identify the potential effect of Formula X on three important cancer processes, the maintenance, progression and metastasis of cancer, and on the response of cancer cells to the host immune system. The changes in these protein levels under Formula X treatment were determined by Western blotting analysis. As shown in **Figure 1B**, Formula X decreased the levels of these proteins in a dose-dependent manner in all four breast cancer cell lines, except the level of PD-L1 protein in HCC1954 cells was not affected. In addition, Formula X also suppressed the expression of ER in T47D and MCF7 cells and Her2 in HCC1954 cells. There was no or very little expression of Slug in T47D or MCF7 cells.

Formula X Inhibits the Growth of a Wide Spectrum of Cancer Cells

We next expanded the study by determining the effect of Formula X on various tissue-original cancer cell lines, including AGS (gastric cancer), LoVo (colon cancer), Huh7 (hepatoma), U2OS (osteosarcoma), and Mia-Paca2 (pancreatic cancer) cells. Formula X inhibited the growth of all cancer cells tested (**Figure 2A**). Consistent with the results from Formula X-treated breast cancer cell lines, Formula X decreased the level of selected proteins in all cancer cells tested, except the Slug protein level of U2OS cells was not decreased by Formula X (**Figure 2B**). Because epidermal growth factor receptor (EGFR) is a critical target for colon cancer therapy (Tamas et al., 2015), the change in EGFR expression under Formula X treatment was also examined. Formula X decreased the protein level of EGFR in LoVo cells (**Figure 2B**). The staining of Slug in Huh7 and LoVo cells and the staining of PD-L1 in U2OS cells were very weak, which might be due to their low expression in these cells (**Figure 2B**), and therefore, it is hard to determine the changes of those proteins under Formula X treatment.

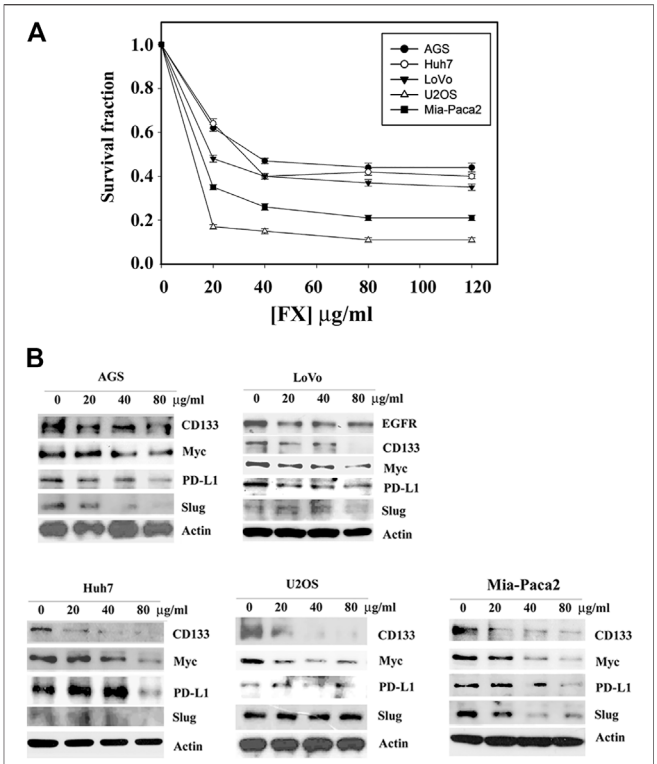


FIGURE 2 | Formula X inhibits cell growth and suppresses oncogenic protein expression in a wide spectrum of cancer cells. **(A)** The indicated cells were treated with a serial dose of Formula X (FX) for 4 days. The level of surviving cells was determined by MTT assay. All data were obtained from at least two independent experiments. The survival fraction showed the fold change relative to the levels of the untreated control. The data are presented as the mean ± SD. **(B)** The cells were treated with various doses of Formula X for 24 h. The lysates were prepared, and the effect of Formula X on the selected protein expression as indicated in the figure was characterized by Western blotting analysis. Actin staining served as a loading control.

Inducibility of Drug Resistance to Formula X in Cancer Cells

Cancer therapy often fails due to the development of drug resistance. Experimentally, the drug resistance of cancer

cells can be induced by short-term treatment of cancer cells with a nonlethal dosage of anticancer drugs (Yeh et al., 2002). We then tried to identify whether drug resistance to Formula X would be induced in cancer cells. The cancer cells indicated in **Figure 3A** were treated with 40 μg/ml Formula X for 24 h and then released for another 24 h. The sensitivity of these cells (-XR cells) to Formula X was evaluated. As shown in **Figure 3A**, no change in sensitivity was observed in Huh7-XR, U2OS-XR or HCC1954-XR cancer cells; a slightly more sensitive response was observed in AGS-XR and T47D-XR cancer cells; and slightly more resistance was observed in MDA-MB231-XR cancer cells. In contrast, a dramatic increase in resistance was induced in Mia-Paca2-XR cancer cells.

Western blotting analysis showed that Formula X decreased the protein levels of CD133 and PD-L1 but increased the protein levels of Myc and Slug in Mia-Paca2-XR cells (**Figure 3B**). In contrast, Formula X still suppressed the protein levels of Myc and Slug in T47D and MDA-MB231-XR cells (**Figure 3B**). In addition, the expression levels of PD-L1 in T47D-XR and MDA-MB231-XR cells were not suppressed by Formula X, suggesting that PD-L1 might not be involved in the sensitivity of these cancer cells to Formula X.

Deconstruction of Formula X

To further identify the function of each component of Formula X, the sensitivity of T47D and MDA-MB231 cells to each single drug of Formula X was determined. As shown

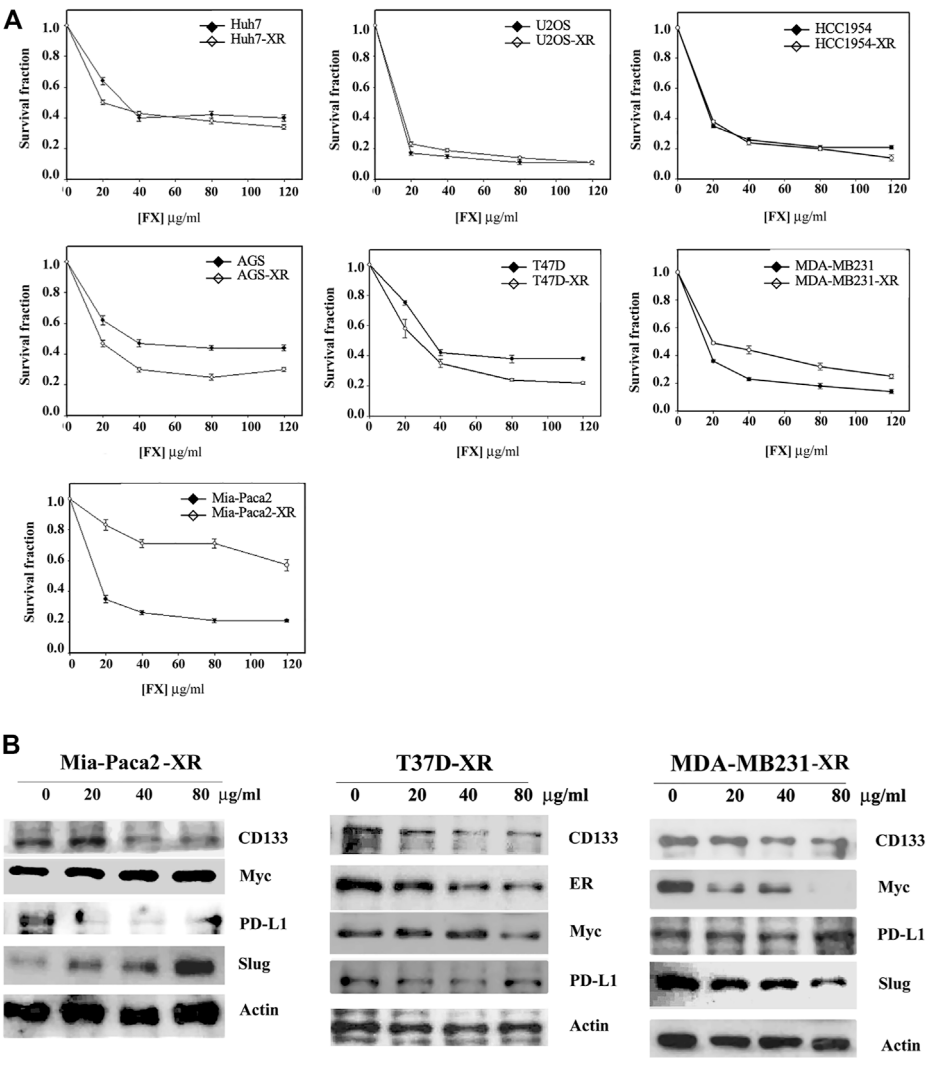


FIGURE 3 | Determination of inducible drug resistance by Formula X treatment. The cells were treated with 40 μg/ml Formula X (FX) for 24 h and then released for another 24 h (-XR cells). Subsequently, the XR cells were treated with a serial dose of Formula X for 4 days. The level of surviving cells was determined by MTT assay. For direct comparison, the curve of untreated cells was redrawn from **Figures 1A**, **2A**. The data are shown as the mean ± SD. **(B)** The XR cells manipulated from Mia-Paca2, T47D and MDA-MB231 cells were treated with a serial dose of Formula X. Western blotting analysis showed the change in protein expression in these cells.

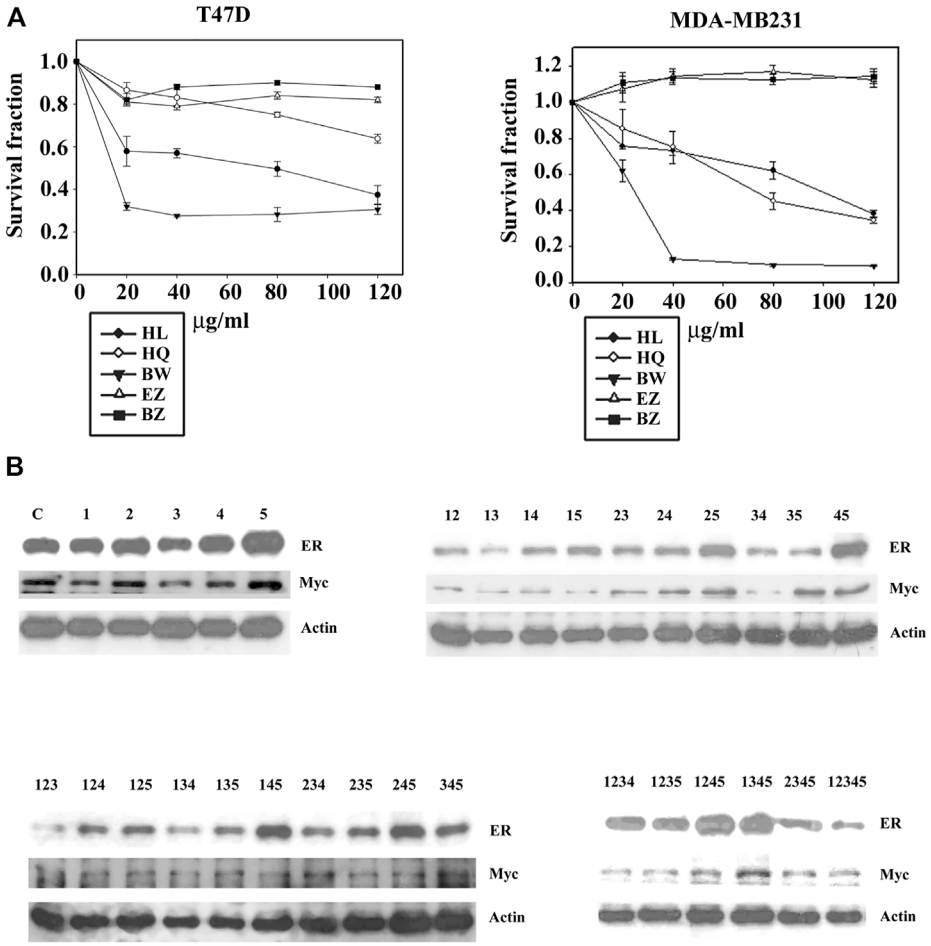


FIGURE 4 | Deconstruction of Formula X. **(A)** T47D and MDA-MB231 cells were treated with various doses of each single component of Formula X (FX) for 4 days. The level of surviving cells was determined by MTT assay. The survival fraction showed the fold change relative to the levels observed in the untreated control. The data are presented as the mean ± SD. **(B)** T47D cells were treated with 40 μg/ml of each single or all possible combinations among the five components of Formula X for 24 h. Western blotting analysis was used to characterize the effect of each treatment on the level of the indicated protein. 1) HL, 2) HQ, 3) BW, 4) EZ, 5) BZ.

TABLE 1 | T47D cells were seeded at a density of 2000 cells/well in 96-well plates. The cells were treated with a single drugs as indicated or various combinations of drugs at 40 μg/ml for 4 days. The number of survival fractions is shown as the fold change relative to levels observed in the untreated control. The results are presented as the mean ± SD. 1) HL, 2) HQ, 3) BW, 4) EZ, 5) BZ.

Treatment	1 (HL)	2 (HQ)	3 (BW)	4 (EZ)	5 (BZ)					
Survival fraction	0.57 ± 0.03	0.83 ± 0.02	0.39 ± 0.04	1.17 ± 0.01	1.03 ± 0.02					
Treatment	1 + 2	1 + 3	1 + 4	1 + 5	2 + 3	2 + 4	2 + 5	3 + 4	3 + 5	4 + 5
Survival fraction	0.62 ± 0.03	0.37 ± 0.01	0.83 ± 0.07	0.78 ± 0.02	0.34 ± 0.04	1.00 ± 0.01	0.88 ± 0.01	0.54 ± 0.07	0.49 ± 0.04	0.90 ± 0.04
Treatment	1 + 2 + 3	1 + 2 + 4	1 + 2 + 5	1 + 3 + 4	1 + 3 + 5	1 + 4 + 5	2 + 3 + 4	2 + 3 + 5	2 + 4 + 5	3 + 4 +5
Survival fraction	0.37 ± 0.04	0.89 ± 0.03	0.82 ± 0.01	0.42 ± 0.02	0.49 ± 0.02	0.82 ± 0.02	0.50 ± 0.03	0.49 ± 0.04	1.02 ± 0.09	0.52 ± 0.05
Treatment	1 + 2 + 3 + 4	1 + 2 + 3 + 5	1 + 2 + 4 + 5	1 + 3 + 4 + 5	2 + 3 + 4 + 5					
Survival fraction	0.46 ± 0.09	0.33 ± 0.01	0.97 ± 0.01	0.43 ± 0.01	0.51 ± 0.08					
Treatment	1 + 2 + 3 + 4 +5									
Survival fraction	0.40 ± 0.02									

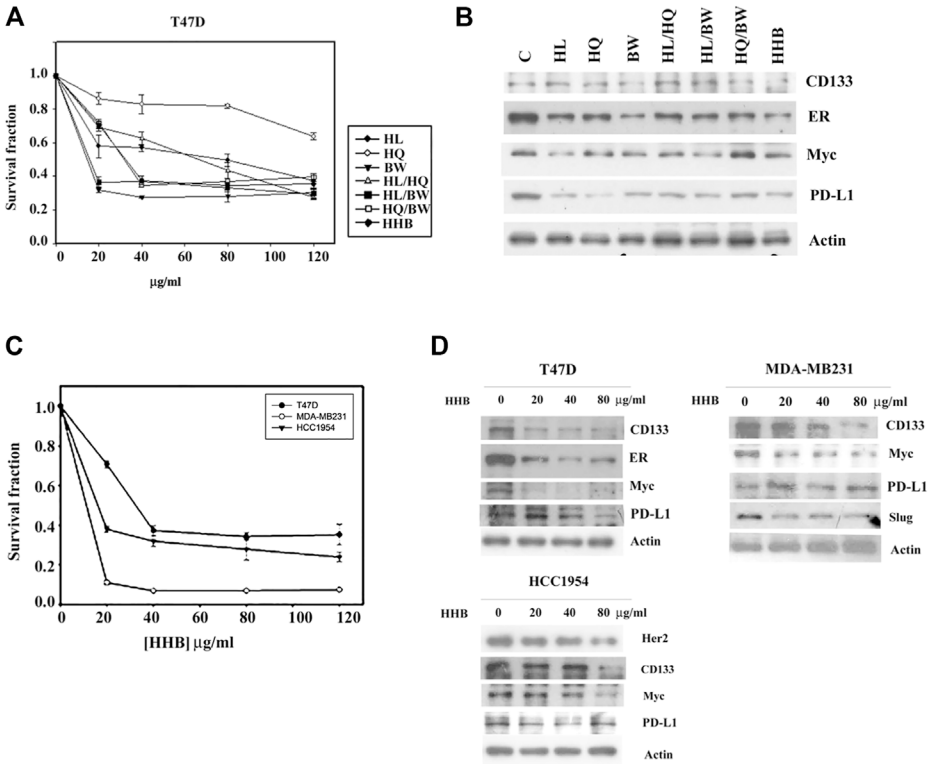


FIGURE 5 | The biological effect of the formula HHB. **(A)** T47D cells were treated with various doses of HL, HQ and BW or all possible combinations for 4 days. The level of surviving cells was determined by MTT assay. Formula HHB is a combination treatment with HL, HQ and BW. **(B)** T47D cells were treated with 40 μg/ml of each single or all possible combinations of the indicated agents for 24 h. Western blotting analysis was performed to characterize the change in the protein expression profile. Three breast cancer cell lines, T47D, MDA-MB231 and HCC 1954, were treated with various doses of the formula HHB. The effect of the formula HHB on cell growth was determined by MTT assay **(C)**, and the effect on protein expression was characterized by Western blotting analysis **(D)**. The survival fraction showed the fold change relative to the levels observed in the untreated control. The data are presented as the mean ± SD.

in **Figure 4A**, BW had the highest inhibitory effect on both cell lines. HL and HQ showed moderate inhibitory effects. However, EZ and BZ had no inhibitory effect and even slightly increased the growth of MDA-MB231 cells.

We then decided to fully deconstruct Formula X by determining the level of growth inhibition induced by each single component and all possible combinations among them in T47D cells. The data are shown in **Table 1**. The results showed that HL and BW were the major components exerting inhibition of growth. HQ exhibited a smaller inhibitory activity; in contrast, both EZ and BZ had no effect on T47D cell growth. All the combinations containing HL and/or BW showed a similar level of growth inhibition; for example, the combinations of HL/BW, HQ/BW, HL/HQ/BW and HL/HQ/BW/BZ exhibited comparable inhibitory effects to that of the whole Formula X (the survival fractions were 0.37, 0.34, 0.37, 0.33, and 0.40, respectively).

The effect of all kinds of combinations of components on protein expression was characterized on the basis of the change in ER and Myc in T47D cells to assess their potential function in cancer cell growth. The results showed that the combinations containing HL and/or BW could effectively suppress ER and Myc expression; in contrast, BZ alone or combined with other components increased protein expression and particularly the level of ER expression (**Figure 4B**). Of all ten combinations of the three components, all combinations containing BZ increased ER expression. The highest level of ER was observed in HL/EZ/BZ- and HQ/EZ/BZ-treated T47D cells, and the lowest level of ER was observed in HL/HQ/BW-treated cells.

We then removed EZ and BZ from Formula X. The biological effect of the new formula (HHB) was characterized using T47D cells as an assay system. As shown in **Figure 5A**, both single components and all possible combinations among these three components suppressed cell growth. Western blotting analysis showed that all treatments decreased the levels of proteins, including CD133, ER, Myc and PD-L1, to different degrees, and HHB suppressed the levels of all proteins to their lowest levels (**Figure 5B**). To further confirm this result, three breast cancer cell lines, T47D, MDA-MB231 and HCC 1954, were treated with a dose course of HHB. The formula HHB suppressed the growth of all three cancer cell lines in a dose-dependent manner (**Figure 5C**). MDA-MA231 cells were the most sensitive to the formula HHB. Western blotting analysis showed that the formula HHB decreased the expression of selected proteins in all three breast cancer cell lines, except the expression of PD-L1 was not affected by the formula HHB in MDA-MB231 cells (**Figure 5D**).

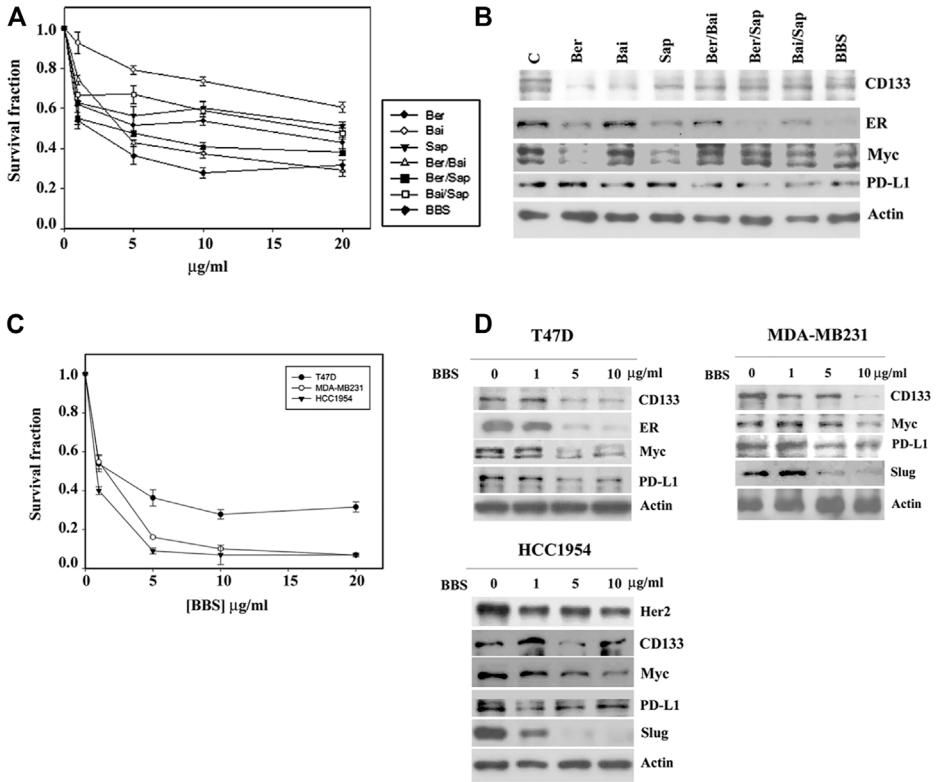


FIGURE 6 | The biological effect of the BBS formula. **(A)** T47D cells were treated with various doses of berberine (Ber), baicalin (Bai) and saponin (Sap) or all possible combinations for 4 days. The level of surviving cells was determined by MTT assay. Formula BBS is a combination treatment with berberine (Ber), baicalin (Bai), and saponin (Sap). **(B)** T47D cells were treated with 5 μg/ml of each single or all possible combinations of the indicated agents for 24 h. Western blotting analysis was performed to characterize the change in the protein expression profile. Three breast cancer cell lines, T47D, MDA-MB231 and HCC 1954, were treated with various doses of the BBS formula. The effect of the BBS formula on cell growth was determined by MTT assay **(C)**, and protein expression was characterized by Western blotting analysis **(D)**. The survival fraction showed the fold change relative to the levels observed in the untreated control. The data are presented as the mean ± SD.

Establishment of a Chemical Formula

The quality control of medical botanical drugs is always an important issue because the botanical drug content might be different with changes in environmental conditions. Physicochemical methods might be used to identify the botanical drug but might not be able to fully determine the biological and/or medical effect of the botanical drug. To set up a functional assay system to control the quality of botanical drugs, in this study, we used a cytotoxic assay and Western blotting of Myc to assess that the consistency of every preparation had been maintained. On the other hand, much effort has been made to identify the principle effective compound in medical botanical drugs and to discover the many related effective chemical compounds of specific medical botanical drugs (Tan et al., 2011; Cai et al., 2018; Luo et al., 2019; Wang et al., 2019). We searched the previous findings about HL, HQ and BW (Liang et al., 2007; Zheng et al., 2007; Xu et al., 2016; Mortazavi et al., 2020; Orzechowska et al., 2020) and selected three chemicals, berberine, baicalin and saponin, to form a chemical formula (formula BBS) for further studies. Among them, C21 steroidal saponin was found to be a major component in BW; however, it was commercially unavailable. A mixture of saponins (S4521) obtained from Sigma was used, which might be expected to have similar biological characteristics to those of C21 steroidal saponins. Consistent with the effect of the formula HHB, both a single agent and different combinations among three chemicals inhibited the growth of T47D cells, and the three chemicals together (formula BBS) achieved maximal growth inhibition (**Figure 6A**). Western blotting analysis showed that all treatments exerted different effects on the protein levels of CD133, ER, Myc and PD-L1, while the formula BBS suppressed the expression levels of all four proteins (**Figure 6B**).

Furthermore, a serial concentration of the formula BBS was used to treat three breast cancer cell lines, T47D, MDA-MB231 and HCC 1954. The formula BBS exhibited a dose-dependent effect of growth inhibition in all three cancer cell lines. MDA-MB231 and HCC1954 cells showed similar sensitivities to the BBS formula, and T47D cells exhibited a higher resistance to the BBS formula (**Figure 6C**). Western blotting analysis indicated that the BBS formula suppressed protein expression in these three cancer cell lines in a pattern similar to that of the HHB formula (**Figure 6D** and compared to **Figure 5D**). Our results suggested that the BBS formula could at least partly replace the HHB formula on the basis of its effects on breast cancer cells.

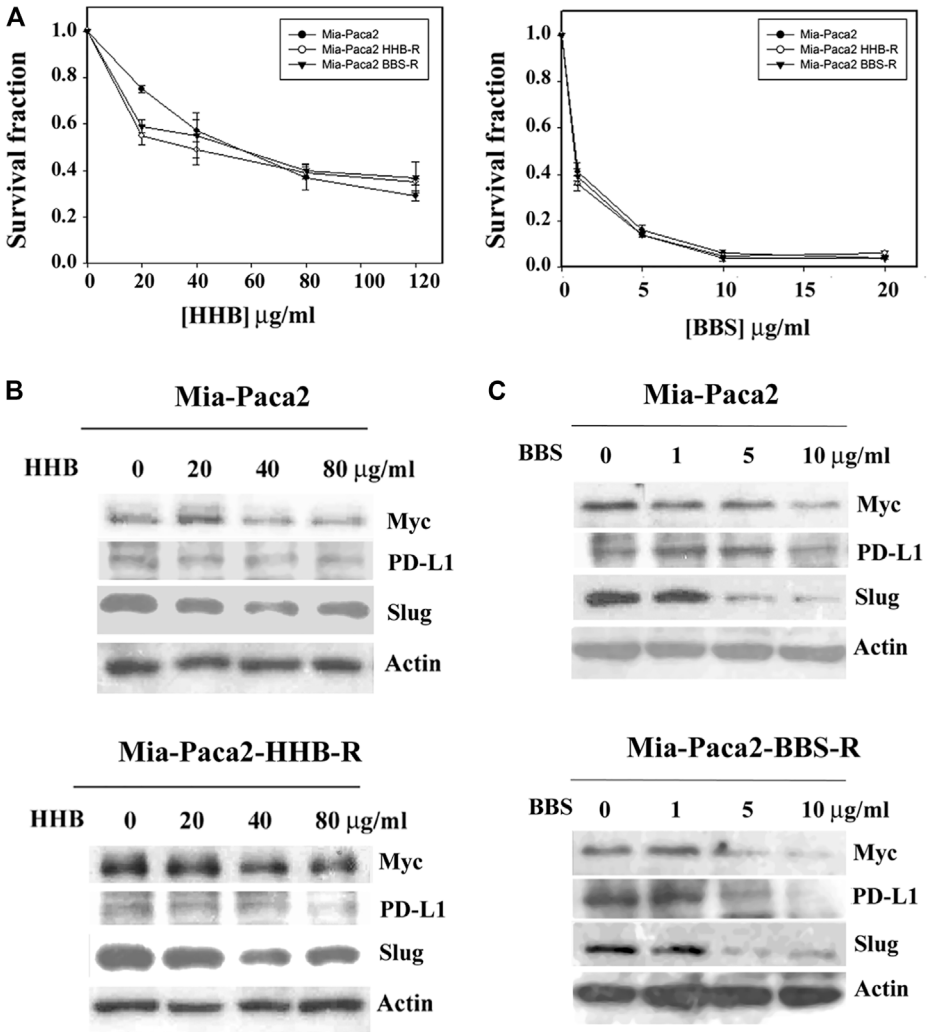


FIGURE 7 | Determination of inducible drug resistance by formula HHB and BBS treatment. **(A)** Mia-Paca2 cells were treated with 40 μg/ml formula HHB or 5 μg/ml formula BBS for 24 h and then released for another 24 h (Mia-Paca2 HHB-R cells and Mia-Paca2 BBS-R cells, respectively). Subsequently, the HHB-R/BBS-R cells were treated with a serial dose of HHB/BBS for 4 days. The level of surviving cells was determined by MTT assay. The survival fraction shows the fold change relative to the levels observed in the untreated control. The data are presented as the mean ± SD. The cells shown in the figure were treated with a serial dose of the formula HHB **(B)** or the formula BBS **(C)** for 24 h. The changes in Myc, PD-L1 and Slug expression under the indicated treatments were characterized by Western blotting analysis.

Formulae HHB and BBS do Not Induce Drug Resistance in Mia-Paca2 Cells

A previous experiment showed that Formula X induced drug resistance in Mia-Paca2 cells (**Figure 3A**). Whether the formulae HHB and BBS also induce drug resistance in Mia-Paca2 cells was determined. Mia-Paca2 cells were treated with 40 μg/ml formula HHB or 5 μg/ml formula BBS for 24 h and then released for another 24 h (Mia-Paca2-HHB-R and Mia-Paca2-BBS-R cells, respectively). The effect of the formula HHB or BBS on growth inhibition in these manipulated cells was evaluated. As shown in **Figure 7A**, neither HHB nor BBS induced drug resistance in Mia-Paca2 cells. Consistently, Myc, PD-L1 and Slug expression was suppressed in untreated and both HHB-R and BBS-R Mia-Paca2 cells (**Figures 7B,C**).

Both Formulae HHB and BBS Exhibited Inhibitory Effects on Tumor Growth and Myc and PD-L1 Expression in Animal Studies

Since these formulae decreased the protein level of PD-L1 in cancer cells, a mouse pancreatic cancer cell, Pan18, implanted mouse was used as an immunocompetent model system to assay drug inhibitory effects on tumor growth in animals. As shown in **Figures 8A,B**, while Formula X did not suppress tumor growth, formulae HHB and BBS decreased the rate of tumor growth. To evaluate the potential therapeutic effect of formula BBS, a group of Pan18-implanted mice was treated with formula BBS starting at 31 days after implantation. The formula BBS decreased the growth rate of tumors, suggesting that the formula BBS is a potentially effective therapeutic formula (**Figure 8B**; BBS-T).

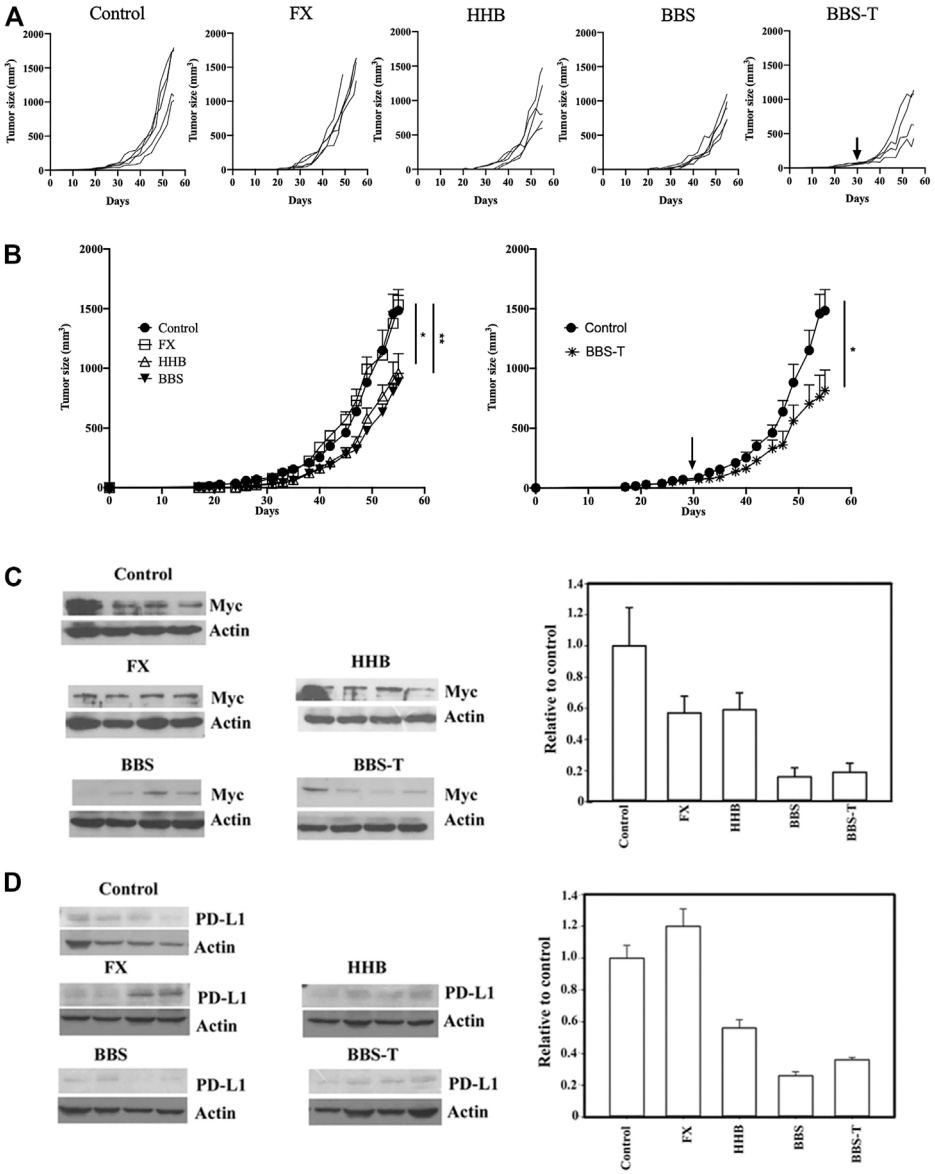


FIGURE 8 | The antitumor effect of formula FX/HHB/BBS treatment. **(A)** Pan18 tumor-bearing mice (4-5 animals/group) were orally administered ddH_2O (control), Formula X (FX), HHB, or BBS on Day 7 or orally administered formula BBS on Day 31 (to detect potential therapeutic effects; BBS-T). Each curve represents each individual mouse. **(B)** The results were pooled and are expressed as the means ± SEMs. The tumor diameters are shown (mm^3). $^*p < 0.05$, $^{**}p < 0.01$. The lysates from the tumors of each sacrificed mouse were subjected to Western blotting to detect the protein levels of Myc **(C)** and PD-L1 **(D)**. No particular region of the image was solely manipulated. The density of each band was determined by an internet available program (ImageJ) and normalized to the density of the actin band. The data are expressed as the fold change relative to the levels observed in the control ±SD (right panel of Figures C and D).

To identify whether these formulae could affect the expression of protein in the tumor-bearing mice as was observed in cultured cells, the levels of Myc and PD-L1 were detected in all tumors from the scarified mice. As shown in **Figure 8C**, all treatments decreased the expression of Myc to different degrees. The protein level of Myc was reduced to half of the level of the control group in Formula X- and HHB-treated groups and reduced to below 20% of the level of the control group in the formula BBS-treated group. Formula X did not suppress, or even slightly increase, the protein level of PD-L1, whereas all other treatments decreased the protein level of PD-L1 to a similar degree as they did for Myc expression (**Figure 8D**).

DISCUSSION

In this study, we deconstructed and characterized each component of the Chinese herbal Formula X. Interestingly, we found that two components (EZ and BZ) of Formula X exhibited effects opposite to those of the other three components, while they did not

influence the biological effects of Formula X in cultured cells. According to the description regarding the functions of EZ and BZ in TCM (the information can be obtained from the websites en.wikibooks.org/wiki/Traditional Chinese Medicine (in English), and cht.a-hospital.com (in Chinese)), both have functions in the stomach and pancreatic meridians, which means they may protect and enhance the function of the digestive system, or in other words, they may influence nutrient uptake. This finding might represent the philosophy of TCM. The formula used in TCM is usually based on the role, so called "jun, chen, zuo and shi" (Wu et al., 2014), in which some components are principle effective components and the others are helpers to increase the effect or decrease the side effects induced by major components or to enhance the speed of drug delivery in the body.

However, the results of the animal study indicated that Formula X did not inhibit tumor growth in animals. In contrast, both formulae HHB and BBS reduced tumor growth in animals. This result suggested that EZ and BZ might reverse the inhibitory effects of HL, HQ and BW on tumor growth in animals, although Formula X showed a similar effect of growth inhibition as those of formulae HHB and BBS in the culture-cell study. This finding raises the possibility that the metabolic products of EZ and BZ in animals might cause some uncharacterized processes to occur that reverses the effect of the other three components of Formula X. Furthermore, Western blotting analysis demonstrated that Formula X did not decrease the expression of PD-L1 but still reduced the protein level of Myc. This result suggested that immuno-suppression of tumor growth might play a critical role and that EZ and BZ could block the immune response of host animals by restoring the level of PD-L1 in tumor cells. This phenomenon probably could not be observed when using immunodeficient mice as the study system. In addition, whether the finding of no effect of Formula X on tumor growth in animals is specific to a particular type of tumor cell or is a common phenomenon remains to be identified.

Berberine has been reported to inhibit cancer cell growth and metastasis, and exhibit anti-inflammatory properties (Wang et al., 2020). Similarly, baicalin (Verma et al., 2021) and saponin (Elekofehinti et al., 2021) are also reported to be involved in multiple processes of cancer cells, including cell growth and metastasis. The underlying mechanism is not well-identified how a single molecule such as berberine, baicalin, or saponin could affect such many biological processes. In this study, we showed all three formulae affected the expression levels of a panel of proteins, including growth factor and steroid receptors, oncogenic proteins, tumor stem cell-related proteins, and immune response-related proteins, in various cancer cells. It is reasonable that these formulae might target multiple molecules and signaling pathways, but the possibility that they target a "master" molecule in the signal transduction network could not be ruled out. For example, the Myc protein was identified to regulate the expression of PD-L1 and CD133 (Gopisetty et al., 2013; Casey et al., 2016).

The primary purpose of this study is to characterize a Chinese herbal formula through deconstruction and reconstitution of the formula. Our results transformed the original formula (Formula X) into a new formula (Formula HHB), and rationally combined three chemicals to establish a novel chemical formula (Formula BBS), which might be more suitable for modern medicine to potentially treat cancer. In this study, the changes in protein expression under treatment might be served as molecular markers to determine the effect of the individual or combined component. However, the underlying mechanism of both herbal and chemical formulae has not been characterized. To identify the direct target or the signaling transduction pathway is important and necessary for further understanding and developing both herbal and chemical formulae. We indeed try to identify and characterize the underlying mechanism of the effect of both Formula HHB and Formula BBS in a single type of cancer by modern biochemical and molecular techniques and expect to deeper elucidate the function of both formulae.

We observed that the formulae HHB and BBS suppressed Myc expression and did not induce drug resistance in pretreated Mia-Paca2 cells. This result indicated that EZ and BZ might be involved in the induction of drug resistance by increasing Myc protein expression when compared with the result from Formula X pretreated Mia-Paca2 cells (Mia-Paca2-XR).

Western blot analysis showed that every single component of formula HHB reduced level of selected protein to a different degree. The inhibitory effect of HL and HQ on PD-L1 expression, BW on CD133 and ER expression, and HL on Myc expression was higher than that of other compounds (**Figure 5B**). This result might represent the potential advantage benefit of traditional Chinese herbal medicine, in which the combined different botanical drugs cooperatively affect multiple targets. Similar results were also observed in the chemical formula BBS (**Figure 6B**). Our results suggested that deconstruction and reconstitution of a Chinese herbal formula could further understand the possible underlying mechanism of its modern biological function. Importantly, we can establish a new chemical formula with a rational combination of different chemical drugs.

Comparing the doses of the formulae HHB and BBS used, 40 μg/ml formula HHB and 5 μg/ml formula BBS were needed to reduce T47D cell growth to 40% of the control level. The levels of the three chemicals (berberine, baicalin and saponin) in the formula BBS were generally below 10% of the content of the corresponding botanical drugs (HL, HQ and BW) in the formula HHB, and the efficiency of extraction should not be 100%; therefore, the amount of these three chemicals in the extract from 40 μg/ml formula HHB was theoretically below 5 μg/ml. This suggests that the inhibitory effect of the formula HHB on tumor growth might not be totally caused by these three chemicals. In an animal study, both formulae HHB and BBS suppressed tumor growth to a similar degree, while the formula BBS decreased Myc and PD-L1 levels to a much lower level than the formula HHB did. Taken together, these findings suggest that some uncharacterized components in the extract of the formula HHB exerted effects of growth inhibition through PD-L1- and Myc-independent pathways.

Furthermore, we demonstrated that these herbal formulae decreased the protein levels of PD-L1 and several other important oncogenic proteins. This raises the possibility that they might be combined with conventional cancer therapies, such as immuno- or chemotherapy. In particular, they suppressed the protein levels of Her2, estrogen receptor and Myc in different types of breast

cancer cells, suggesting that they might be broadly effective in breast cancer, although this possibility remains to be further characterized.

Formula X was originally used to treat fever, and no record indicates that it has anticancer activity. Our study indeed suggests that Formula X might not be suitable for cancer treatment, partly because it increases PD-L1 expression and consequently reduces the host immunosuppression of tumor growth. However, through thoughtful characterization, we transformed it into an effective novel herbal formula. We believe this approach provides a method to renew and understand traditional Chinese medicine.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

All mice were housed in a specific pathogen-free room at the Animal Center, China Medical University, Taichung, Taiwan and maintained in accordance with institutional animal care guidelines. The animal protocol was approved by the Animal Care and Use Committee of China Medical University, Taichung, Taiwan (CMUIACUC-2017-128).

AUTHOR CONTRIBUTIONS

Y-CS performed the animal study and wrote the manuscript. D-YL performed LC-ESI-MS analysis for the identity of botanical drugs. P-YY performed the tumor cell growth assay and Western blotting analysis and wrote the manuscript.

FUNDING

This study was supported by Jin-Mi company to P-YY and the grant (MOST 105-2320-B-039-005, CMU110-MF-75) from Ministry of Science and Technology, Taiwan and China Medical University, Taiwan to Y-CS.

ACKNOWLEDGMENTS

We thank Dr. YS Lu (Department of Oncology, National Taiwan University Hospital) for helping with the cell growth and protein expression assays.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2022.907826/full#supplementary-material

REFERENCES

Bellmunt, J., Powles, T., and Vogelzang, N. J. (2017). A Review on the Evolution of PD-1/pd-L1 Immunotherapy for Bladder Cancer: The Future Is Now. *Cancer Treat. Rev.* 54, 58–67. doi:10.1016/j.ctrv.2017.01.007

Bretones, G., Delgado, M. D., and León, J. (2015). Myc and Cell Cycle Control. *Biochim. Biophys. Acta* 1849 (5), 506–516. doi:10.1016/j.bbagrm.2014.03.013

Cai, F. F., Zhou, W. J., Wu, R., and Su, S. B. (2018). Systems Biology Approaches in the Study of Chinese Herbal Formulae. *Chin. Med.* 13, 65. doi:10.1186/s13020-018-0221-x

Casey, S. C., Tong, L., Li, Y., Do, R., Walz, S., Fitzgerald, K. N., et al. (2016). Erratum for the Report "MYC Regulates the Antitumor Immune Response through CD47 and PD-L1" by S. C. Casey, L. Tong, Y. Li, R. Do, S. Walz, K. N. Fitzgerald, A. M. Gouw, V. Baylot, I. Gütgemann, M. Eilers, D. W. Felsher. *Science* 352 (6282), 227–231. doi:10.1126/science.aac993510.1126/science.aaf7984

Corson, T. W., and Crews, C. M. (2007). Molecular Understanding and Modern Application of Traditional Medicines: Triumphs and Trials. *Cell* 130 (5), 769–774. doi:10.1016/j.cell.2007.08.021

Crusz, S. M., and Balkwill, F. R. (2015). Inflammation and Cancer: Advances and New Agents. *Nat. Rev. Clin. Oncol.* 12 (10), 584–596. doi:10.1038/nrclinonc.2015.105

Elekofehinti, O. O., Iwaloye, O., Olawale, F., and Ariyo, E. O. (2021). Saponins in Cancer Treatment: Current Progress and Future Prospects. *Pathophysiology* 28 (2), 250–272. doi:10.3390/pathophysiology28020017

Gong, B., Kao, Y., Zhang, C., Zhao, H., Sun, F., and Gong, Z. (2019). Exploring the Pharmacological Mechanism of the Herb Pair "HuangLian-GanJiang" against Colorectal Cancer Based on Network Pharmacology. *Evid. Based Complement. Altern. Med.* 2019, 2735050. doi:10.1155/2019/2735050

Gopisetty, G., Xu, J., Sampath, D., Colman, H., and Puduvalli, V. K. (2013). Epigenetic Regulation of CD133/PROM1 Expression in Glioma Stem Cells by Sp1/myc and Promoter Methylation. *Oncogene* 32 (26), 3119–3129. doi:10.1038/onc.2012.331

Greten, F. R., and Grivennikov, S. I. (2019). Inflammation and Cancer: Triggers, Mechanisms, and Consequences. *Immunity* 51 (1), 27–41. doi:10.1016/j.immuni.2019.06.025

Ikemoto, S., Sugimura, K., Yoshida, N., Yasumoto, R., Wada, S., Yamamoto, K., et al. (2000). Antitumor Effects of Scutellariae Radix and its Components Baicalein, Baicalin, and Wogonin on Bladder Cancer Cell Lines. *Urology* 55 (6), 951–955. doi:10.1016/s0090-4295(00)00467-2

Jang, J. W., Song, Y., Kim, S. H., Kim, J., and Seo, H. R. (2017). Potential Mechanisms of CD133 in Cancer Stem Cells. *Life Sci.* 184, 25–29. doi:10.1016/j.lfs.2017.07.008

Jeon, J., Park, K. A., Lee, H., Shin, S., Zhang, T., Won, M., et al. (2011). Water Extract of Cynanchi Atrati Radix Regulates Inflammation and Apoptotic Cell Death through Suppression of IKK-Mediated NF-Kb Signaling. *J. Ethnopharmacol.* 137 (1), 626–634. doi:10.1016/j.jep.2011.06.022

Katoh, M. (2011). Network of WNT and Other Regulatory Signaling Cascades in Pluripotent Stem Cells and Cancer Stem Cells. *Curr. Pharm. Biotechnol.* 12 (2), 160–170. doi:10.2174/138920111794295710

Liang, M., Zheng, Z., Yuan, Y., Kong, L., Shen, Y., Liu, R., et al. (2007). Identification and Quantification of C21 Steroidal Saponins from Radix Cynanchi Atrati by High-Performance Liquid Chromatography with Evaporative Light Scattering Detection and Electrospray Mass Spectrometric Detection. *Phytochem. Anal.* 18 (5), 428–435. doi:10.1002/pca.998

Luo, H., Vong, C. T., Chen, H., Gao, Y., Lyu, P., Qiu, L., et al. (2019). Naturally Occurring Anti-cancer Compounds: Shining from Chinese Herbal Medicine. *Chin. Med.* 14, 48. doi:10.1186/s13020-019-0270-9

Mortazavi, H., Nikfar, B., Esmaeili, S. A., Rafieenia, F., Saburi, E., Chaichian, S., et al. (2020). Potential Cytotoxic and Anti-metastatic Effects of Berberine on Gynaecological Cancers with Drug-Associated Resistance. *Eur. J. Med. Chem.* 187, 111951. doi:10.1016/j.ejmech.2019.111951

Orzechowska, B. U., Wróbel, G., Turlej, E., Jatczak, B., Sochocka, M., and Chaber, R. (2020). Antitumor Effect of Baicalin from the Scutellaria Baicalensis Radix Extract in B-Acute Lymphoblastic Leukemia with Different Chromosomal Rearrangements. *Int. Immunopharmacol.* 79, 106114. doi:10.1016/j.intimp.2019.106114

Stine, Z. E., Walton, Z. E., Altman, B. J., Hsieh, A. L., and Dang, C. V. (2015). MYC, Metabolism, and Cancer. *Cancer Discov.* 5 (10), 1024–1039. doi:10.1158/2159-8290.CD-15-0507

Tamas, K., Walenkamp, A. M., de Vries, E. G., van Vugt, M. A., Beets-Tan, R. G., van Etten, B., et al. (2015). Rectal and Colon Cancer: Not Just a Different Anatomic Site. *Cancer Treat. Rev.* 41 (8), 671–679. doi:10.1016/j.ctrv.2015.06.007

Tan, W., Lu, J., Huang, M., Li, Y., Chen, M., Wu, G., et al. (2011). Anti-cancer Natural Products Isolated from Chinese Medicinal Herbs. *Chin. Med.* 6 (1), 27. doi:10.1186/1749-8546-6-27

Vasan, N., Baselga, J., and Hyman, D. M. (2019). A View on Drug Resistance in Cancer. *Nature* 575 (7782), 299–309. doi:10.1038/s41586-019-1730-1

Verma, E., Kumar, A., Devi Daimary, U., Parama, D., Girisa, S., Sethi, G., et al. (2021). Potential of Baicalein in the Prevention and Treatment of Cancer: A Scientometric Analyses Based Review. *J. Funct. Foods* 86, 104660. doi:10.1016/j.jff.2021.104660

Wang, H., Li, H., Chen, F., Luo, J., Gu, J., Wang, H., et al. (2017). Baicalin Extracted from Huangqin (Radix Scutellariae Baicalensis) Induces Apoptosis in Gastric Cancer Cells by Regulating B Cell Lymphoma (Bcl-2)/Bcl-2-Associated X Protein and Activating Caspase-3 and Caspase-9. *J. Tradit. Chin. Med.* 37 (2), 229–235. doi:10.1016/s0254-6272(17)30049-3

Wang, J., Wu, M. Y., Tan, J. Q., Li, M., and Lu, J. H. (2019). High Content Screening for Drug Discovery from Traditional Chinese Medicine. *Chin. Med.* 14, 5. doi:10.1186/s13020-019-0228-y

Wang, Y., Liu, Y., Du, X., Ma, H., and Yao, J. (2020). The Anti-cancer Mechanisms of Berberine: A Review. *Cancer Manag. Res.* 12, 695–702. doi:10.2147/CMAR.S242329

Wu, L., Wang, Y., Li, Z., Zhang, B., Cheng, Y., and Fan, X. (2014). Identifying Roles of "Jun-Chen-Zuo-Shi" Component Herbs of QiShenYiQi Formula in Treating Acute Myocardial Ischemia by Network Pharmacology. *Chin. Med.* 9, 24. doi:10.1186/1749-8546-9-24

Xu, J., and Xia, Z. (2019). Traditional Chinese Medicine (TCM) – Does its Contemporary Business Booming and Globalization Really Reconfirm its Medical Efficacy & Safety? *Med. Drug Discov.* 1, 100003. doi:10.1016/j.medidd.2019.100003

Xu, X. H., Li, T., Fong, C. M., Chen, X., Chen, X. J., Wang, Y. T., et al. (2016). Saponins from Chinese Medicines as Anticancer Agents. *Molecules* 21 (10), 1326. doi:10.3390/molecules21101326

Yeh, P. Y., Chuang, S. E., Yeh, K. H., Song, Y. C., Ea, C. K., and Cheng, A. L. (2002). Increase of the Resistance of Human Cervical Carcinoma Cells to Cisplatin by Inhibition of the MEK to ERK Signaling Pathway Partly via Enhancement of Anticancer Drug-Induced NF Kappa B Activation. *Biochem. Pharmacol.* 63 (8), 1423–1430. doi:10.1016/s0006-2952(02)00908-5

Zhang, X. Q., Yao, C., Bian, W. H., Chen, X., Xue, J. X., Zhu, Z. Y., et al. (2019). Effects of Astragaloside IV on Treatment of Breast Cancer Cells Execute Possibly through Regulation of Nrf2 via PI3K/AKT/mTOR Signaling Pathway. *Food Sci. Nutr.* 7 (11), 3403–3413. doi:10.1002/fsn3.1154

Zhang, Y., Liang, Y., and He, C. (2017). Anticancer Activities and Mechanisms of Heat-Clearing and Detoxicating Traditional Chinese Herbal Medicine. *Chin. Med.* 12, 20. doi:10.1186/s13020-017-0140-2

Zheng, Z., Zhang, W., Kong, L., Liang, M., Li, H., Lin, M., et al. (2007). Rapid Identification of C21 Steroidal Saponins in Cynanchum Versicolor Bunge by Electrospray Ionization Multi-Stage Tandem Mass Spectrometry and Liquid Chromatography/tandem Mass Spectrometry. *Rapid Commun. Mass Spectrom.* 21 (3), 279–285. doi:10.1002/rcm.2829

Zhou, W., Gross, K. M., and Kuperwasser, C. (2019). Molecular Regulation of Snai2 in Development and Disease. *J. Cell Sci.* 132 (23), jcs235127. doi:10.1242/jcs.235127

Zhou, Y., Yang, Z. Y., and Tang, R. C. (2016). Bioactive and UV Protective Silk Materials Containing Baicalin - the Multifunctional Plant Extract from Scutellaria Baicalensis Georgi. *Mater Sci. Eng. C Mater Biol. Appl.* 67, 336–344. doi:10.1016/j.msec.2016.05.063

Conflict of Interest: Author P-YY was employed by Jin-Mi company

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 09 June 2022
doi: 10.3389/fpls.2022.920826

Phylogenomics and Genetic Diversity of *Arnebiae Radix* and Its Allies (*Arnebia*, Boraginaceae) in China

Jiahui Sun[1†], Sheng Wang[1†], Yiheng Wang[1], Ruishan Wang[1], Kangjia Liu[2], Enze Li[2], Ping Qiao[1], Linyuan Shi[1], Wenpan Dong[2], Luqi Huang[1]* and Lanping Guo[1]**

[1] State Key Laboratory Breeding Base of Dao-di Herbs, National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China, [2] Laboratory of Systematic Evolution and Biogeography of Woody Plants, School of Ecology and Nature Conservation, Beijing Forestry University, Beijing, China

OPEN ACCESS

Edited by:
Da-Cheng Hao,
Dalian Jiaotong University, China

Reviewed by:
Zhangjian Shan,
Institute of Botany (CAS), China
Jie Shen,
Weifang Medical University, China

***Correspondence:**
Wenpan Dong
wpdong@bjfu.edu.cn
Luqi Huang
huangluqi01@126.com
Lanping Guo
glp01@126.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: *15 April 2022*
Accepted: *11 May 2022*
Published: *09 June 2022*

Citation:
Sun J, Wang S, Wang Y, Wang R, Liu K, Li E, Qiao P, Shi L, Dong W, Huang L and Guo L (2022) Phylogenomics and Genetic Diversity of Arnebiae Radix and Its Allies (Arnebia, Boraginaceae) in China. Front. Plant Sci. 13:920826. doi: 10.3389/fpls.2022.920826

Arnebiae Radix is a traditional medicine with pleiotropic properties that has been used for several 100 years. There are five species of *Arnebia* in China, and the two species *Arnebia euchroma* and *Arnebia guttata* are the source plants of *Arnebiae Radix* according to the Chinese Pharmacopoeia. Molecular markers that permit species identification and facilitate studies of the genetic diversity and divergence of the wild populations of these two source plants have not yet been developed. Here, we sequenced the chloroplast genomes of 56 samples of five *Arnebia* species using genome skimming methods. The *Arnebia* chloroplast genomes exhibited quadripartite structures with lengths from 149,539 and 152,040 bp. Three variable markers (*rps16-trnQ*, *ndhF-rpl32*, and *ycf1b*) were identified, and these markers exhibited more variable sites than universal chloroplast markers. The phylogenetic relationships among the five *Arnebia* species were completely resolved using the whole chloroplast genome sequences. *Arnebia* arose during the Oligocene and diversified in the middle Miocene; this coincided with two geological events during the late Oligocene and early Miocene: warming and the progressive uplift of Tianshan and the Himalayas. Our analyses revealed that *A. euchroma* and *A. guttata* have high levels of genetic diversity and comprise two and three subclades, respectively. The two clades of *A. euchroma* exhibited significant genetic differences and diverged at 10.18 Ma in the middle Miocene. Three clades of *A. guttata* diverged in the Pleistocene. The results provided new insight into evolutionary history of *Arnebia* species and promoted the conservation and exploitation of *A. euchroma* and *A. guttata.*

Keywords: *Arnebiae Radix*, *Arnebia euchroma*, *Arnebia guttata*, chloroplast genome, genetic diversity, phylogeography, phylogenomics

INTRODUCTION

Arnebiae Radix is a traditional Persian, Unani, Ayurvedic, and Chinese medicine with pleiotropic properties that has been used for several 100 years (Zhu, 1982; Kumar et al., 2021). *Arnebiae Radix* is an ingredient in 122 Chinese patent medicines and 195 Chinese medicine prescriptions. Shikonin, shikonofurans, and red naphthoquinones are the main effective chemical constituents,

and they have been widely used for the treatment of infections, inflammation, and bleeding for their anti-inflammatory, anti-fungal, and anti-angiogenic activities (Gupta et al., 2014; Feng et al., 2020; Kumar et al., 2021). Shikonin and its derivatives exhibit anti-cancer and anti-tumorigenic activities and have the potential to be used for the development of anti-tumor drugs (Feng et al., 2020; Liao et al., 2020; Kumar et al., 2021). *Arnebiae Radix* has also been used in the printing, dyeing, and cosmetics industries (Ma et al., 2021). Although *Arnebiae Radix* is widely used for its varied medicinal effects, the evolutionary history of its source plants remains unclear.

Arnebiae Radix is the dried root of *Arnebia euchroma* (Royle ex Benth.) I. M. Johnst. and *Arnebia guttata* Bunge according to the Chinese Pharmacopoeia (2020 version). *Arnebia* Forsskål is a genus in the family Boraginaceae and the tribe Lithospermeae. *Arnebia* comprises ca. 25 species and are distributed from North Africa to Central Asia and the Himalayas (Johnston, 1954). Only six *Arnebia* species to date have been reported from China (Zhu, 1982), including *A. euchroma*, *A. guttata*, *A. decumbens* (Ventenat) Cosson & Kralik, *A. tschimganica* (B. Fedtschenko) G. L. Chu, *A. szechenyi* Kanitz, and *A. fimbriata* Maximowicz. *A. guttata* (Inner Mongolia *Arnebiae Radix*) mainly occurs in Xizang, Xinjiang, western Gansu, Ningxia, and Inner Mongolia in China. *A. euchroma*, commonly known as Xinjiang *Arnebiae Radix*, is an endangered herb (Kala, 2000; Gupta et al., 2014) that grows naturally on the slopes and dry patches in cold desert temperate zones of the western Himalayas (Xinjiang, western Xizang) (Lal et al., 2020). *A. szechenyi* inhabits sunny rocky slopes and sand dune edges; most populations of this species occur in areas surrounding the Tengger Desert in Northwest China (Fu et al., 2021). *A. decumbens* is an annual plant, and its native range extends from North Africa to Mongolia, including the Arabian Peninsula; it is also native to the Canary Islands. In China, its range is restricted to northern Xinjiang. *A. fimbriata* has a densely gray–white hirsute and is distributed in western Gansu, Ningxia, Qinghai, and Mongolia. *A. tschimganica* is an endangered herb in China (Qin et al., 2017). Recent studies have shown that this species belongs to the monotypic *Ulugbekia* [*Ulugbekia tschimganica* (B. Fedtsch.) Zakirov] (Weigend et al., 2009).

Owing to its various medicinal properties, *Arnebia* species have been overexploited, and the populations of some species have declined so much. There are two major outstanding problems regarding the exploitation and conservation of *Arnebiae Radix* and its source plants that need to be addressed. First, the sources of the medicinal materials of *Arnebiae Radix* in the market are complex according to one field study (Liu et al., 2020). Identification of source species is difficult based on morphological features; there is thus an urgent need to develop more efficient methods for species identification, such as DNA barcoding (Liu et al., 2020; Xu et al., 2021). Second, a sound understanding of patterns of genetic diversity and genetic divergence among wild plant populations is important for plant seeding and the conservation of threatened species (Funk et al., 2012). However, few studies have evaluated the population genetics of the source plants of *Arnebiae Radix*. Improved genetic markers need to be developed to facilitate the identification of the source species of *Arnebiae Radix* and clarify the population genetics of *A. euchroma* and *A. guttata*.

Chloroplasts are key plant organelles that are involved in photosynthesis and important biological processes, such as fatty acid and amino acid synthesis (Sugiura, 1992). The chloroplast genome of higher plants is a double-stranded and circular DNA molecule with a typical quadripartite structure (Dong et al., 2013, 2021a; Villanueva-Corrales et al., 2021), including a pair of inverted repeats (IRs) as well as a large single copy (LSC) and a small single copy (SSC) region. The chloroplast genome encodes approximately 80 protein-coding, 30 transfer RNA (tRNA), and four ribosomal RNA (rRNA) genes, and the structure of the chloroplast genome, including the gene content and gene order, is conserved. An increasing number of chloroplast genomes are being sequenced due to the development of sequencing technology and genome assembly methods. The nucleotide substitution rate of the chloroplast genome is moderate compared with that of the nuclear genome; it is also uniparentally inherited and exhibits low rates of recombination (Schwarz et al., 2017; Dong et al., 2020). Chloroplast genome sequences are thus effective for inferring phylogenetic relationships at various levels of divergence (Dong et al., 2018, 2022b), characterizing population structure (Qiao et al., 2020; Xiao et al., 2021), identifying species (Ślipiko et al., 2020; Dong et al., 2021b), and elucidating patterns of genome evolution and molecular evolution (Dong et al., 2020). The chloroplast genomes of *Arnebia* have been compared with those of *Lithospermum* (Park et al., 2020); the full chloroplast genome sequences of *A. guttata* and *A. euchroma* have been published, and molecular markers have been developed at the genus level through comparative analysis. However, interspecific and intraspecific variation among *Arnebia* species has not yet been clarified. Multiple samples of species and genotypes would aid the development of markers for the identification of *Arnebiae Radix* species as well as for assessments of the population structure of *A. euchroma* and *A. guttata*.

Here, the chloroplast genomes of 56 *Arnebiae Radix* samples and its allies (including five *Arnebia* species) were sequenced and assembled using genome skimming methods. Specifically, we aimed to (i) elucidate variation in the chloroplast genome among *Arnebia* species in China, (ii) evaluate the divergence times among *Arnebia* species, (iii) identify markers that could be used to discriminate between different *Arnebia* species, and (iv) evaluate the genetic structure of and genetic divergence between *A. euchroma* and *A. guttata*. The chloroplast genome resources presented in this study will aid the conservation and exploitation of *Arnebia* species.

MATERIALS AND METHODS

Plant Material

Fifty-six samples covering five Arnebia species were included in this study. A total of 50 samples were collected from field in China. DNA of four samples from Russia, and two from Mongolia were acquired from the DNA bank of

China in Institute of Botany, Chinese Academy of Sciences. The samples included seven genotypes of *A. decumbens*, 15 genotypes of *A. euchroma*, six genotypes of *A. fimbriata*, 17 genotypes of *A. guttata*, and 11 genotypes of *A. szechenyi*. Samples were derived from various localities in Asia and were representative of the geographical distributions of the five *Arnebia* species. Sample information is provided in **Supplementary Table 1**.

Genome Sequencing and Chloroplast Genome Assembly

Total DNA was extracted using the modified CTAB method (Li et al., 2013). DNA quality was measured using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, United States). A total of 500 ng of DNA was used for sequencing library preparation. Total DNA was sheared to 350-bp fragments using an ultrasonicator. Illumina paired-end DNA library construction and paired-end whole-genome shotgun resequencing (150 bp) on an Illumina HiSeq X-ten platform at Novogene (Tianjin, China) were performed; each sample yielded approximately 5 Gb of data.

Quality control of the raw data was conducted using Trimmomatic 0.36 (Bolger et al., 2014) with the following parameters: LEADING, 20; TRAILING, 20; SLIDING WINDOW, 4:15; MIN LEN, 36; and AVG QUAL, 20. GetOrganelle (Jin et al., 2020) with k-mer lengths of 85, 95, and 105 was used to assemble the whole chloroplast genomes. If GetOrganelle was unable to successfully assemble the whole chloroplast genome, we followed the methods of Dong et al. (2022a) to assemble the chloroplast genome sequences. Genomes were annotated using Plann (Huang and Cronk, 2015), and the published genome of *A. guttata* (GenBank Accession number: MT975391) was used as the reference sequence. Annotated whole chloroplast genome sequences were submitted to GenBank. The *Arnebia* chloroplast genomes maps were depicted using Chloroplot (Zheng et al., 2020).

Development of Molecular Markers

All the chloroplast genome sequences were aligned using MAFFT 7 (Katoh and Standley, 2013) and adjusted manually using Se-Al 2.0 (Rambaut, 1996); for example, alignment errors associated with polymeric repeat structures and small inversions were corrected to avoid overestimation of sequence divergence. mVISTA and nucleotide diversity (π) were used to analyze interspecific variation in *Arnebia* chloroplast genomes (Frazer et al., 2004). Mutational hotspots were identified by sliding window methods; the window size was set to 800 bp, and the step size was set to 100 bp.

Nucleotide diversity, variable, and parsimony-informative sites were used to assess marker variability for hypervariable markers (mutational hotspot regions). The three universal chloroplast DNA markers, *rbcL*, *matK*, and *trnH-psbA*, were used in this analysis to compare variation between hypervariable markers and universal chloroplast DNA markers. The number of variable and parsimony-informative sites and nucleotide diversity (π) were calculated using MEGA 7.0 (Kumar et al., 2016) and DnaSP v6 (Rozas et al., 2017).

Phylogenetic Analyses

We used all the coding genes to reconstruct the phylogenetic relationships among *Arnebia* and other Lithospermeae species at the species level. The dataset included 12 *Arnebia* samples and samples from eight other Boraginaceae species. Coding genes were extracted using Geneious Prime v2020.0.5 based on annotation of the chloroplast genomes. Nucleotide and amnio acid sequences were used for phylogenetic analyses. The whole chloroplast genome dataset of all *Arnebia* samples was used to infer the phylogenetic relationships among all genotypes.

Maximum likelihood (ML) and Bayesian inference (BI) were used to infer phylogenetic relationships. ML analyses were run using RAxML-NG (Kozlov et al., 2019) with 500 bootstrap replicates. ModelFinder (Kalyaanamoorthy et al., 2017) was used to select the best-fit model of nucleotide substitution with the Bayesian information criterion. The BI tree was generated in MrBayes v3.2 (Ronquist et al., 2012). The BI analysis was run with two independent chains and priors for 20 million generations, with sampling every 1,000 generations. The stationary phase was examined using Tracer 1.6 (Rambaut et al., 2014), and the first 25% of the sampled trees was discarded. The remaining trees were used to generate a majority-rule consensus tree to estimate posterior probabilities.

Time Calibration of the Phylogeny

The chloroplast gene data were used to estimate the divergence times among *Arnebia* species. Two priors from the findings of (Chacón et al., 2019) were used for these analyses. The crown age of Lithospermeae was constrained to 42.5 Mya [95% highest posterior density (HPD): 35.3–51.5 Mya]. The crown age of *Onosma* and *Lithospermum* was 32.1 Mya. The two priors were placed under a normal distribution with a standard deviation of 1.

We used the whole chloroplast genome data of 56 genotypes of the five *Arnebia* species to infer divergence times at the intraspecific level. Three priors from the above results were used for this analysis: (i) the crown age of *Arnebia* (the root of the tree); (ii) the crown age of *A. guttata* and *A. decumbens*; and (3) the crown age of *A. euchroma* and *A. szechenyi*/*A. fimbriata*. All three priors were placed under a normal distribution with a standard deviation of 1.

The divergence time analyses were performed in BEAST 2 (Bouckaert et al., 2014). The prior tree Yule model and GTR model were selected with the uncorrelated lognormal distribution relaxed molecular clock model, and the Markov Chain Monte Carlo (MCMC) tool was run for 500,000,000 generations with sampling every 50,000 generations. Tracer 1.6 (Rambaut et al., 2014) was used to evaluate convergence and to ensure a sufficient and effective sample size for all parameters surpassing 200. A maximum credibility tree was generated with mean heights in TreeAnnotator, with the first 10% of the trees discarded as burn-in.

Genetic Diversity and Population Differentiation of *Arnebia euchroma* and *Arnebia guttata*

Number of variable sites, nucleotide diversity (π), and number of haplotypes were calculated using MEGA 7.0 and DnaSP v6 within the five species to characterize intraspecific variation in genetic diversity. Filtered intraspecific SNPs were used to analyze population structure with an admixture model-based clustering method implemented in STRUCTURE v. 2.3.4 (Falush et al., 2007). Allele frequencies were assumed to be correlated among populations (K = 1 to K = 10). The most likely number of clusters was determined based on both LnP(D) and Δk. Principal component analysis (PCA) was also used to assess genetic structure. PCA was conducted using Plink (Purcell et al., 2007), and graphs were built using the ggbiplot package in R.

RESULTS

Structure, Gene Content, and Sequences in Five *Arnebia* Species Chloroplast Genomes

The whole chloroplast genomes of the five *Arnebia* species had the quadripartite structure typical of most angiosperm species (**Figure 1**). The length of the 56 chloroplast genomes ranged between 149,539 and 152,040 bp (**Table 1** and **Supplementary Table 2**). The LSC (between 80,462 and 82,946 bp) and SSC (between 17,143 and 17,336 bp) were separated by two IR regions (IRa and IRb, between 25,753 and 26,053 bp). The overall GC content was approximately 37.5–37.8%, and the GC content was slightly higher in the IR (43.0%) regions than in the LSC (35.6%) and SSC (31.3%) regions. The *Arnebia* chloroplast genome encoded 112 unique genes, including 78 protein-coding genes, 30 tRNA genes, and four rRNA genes. Eighteen genes had introns; 16 genes had a single intron, and two genes (*ycf3* and *clpP*) had two introns. There were three intron-containing genes (*ndhB*, *trnI-GAU*, and *trnA-UGC*) in the IR regions. The *matK* gene was located in *trnK*-UUU, which is the largest intron in the chloroplast genome, and the *rps12* gene was *trans*-spliced, with two copies of the 3′ end in the IR region and the 5′ end in the LSC region. The IR/SC boundaries slightly differed among the five *Arnebia* species. For example, the *trnH* gene was located in the LSC region, 12 to 42 bp from the IRa/LSC boundary.

Genome Variation and Molecular Marker Identification

The mVISTA results revealed high synteny among the chloroplast genomes of the five *Arnebia* species, which indicates that the chloroplast genome is highly conserved in this genus (**Supplementary Figure 1**). The 65 whole chloroplast genomes had an aligned length of 157,292 bp (**Table 2**), including 4,921 variable sites (3.13%) and 3,996 parsimony-informative sites (2.54%). The overall nucleotide diversity (π) of the chloroplast genome was 0.00634. The nucleotide diversity was lowest for the IR region (0.00177) and highest for the SSC region (0.01234). The SSC region also had the highest proportion of variable sites and parsimony-informative sites. These results indicate that the SSC region had the highest mutation rate and sequence divergence and that the IR region was more conserved compared with other regions (**Table 2** and **Supplementary Figure 1**).

The sliding window method was used to identify mutational hotspots in the whole chloroplast genome using a window size of 800 bp (**Figure 2**). The π values ranged from 0 to 0.02439. Nucleotide diversity was lowest in the IR region. Three peaks were identified in the 65 whole chloroplast genomes, with π values $>$0.02. The three markers contained the two intergenic regions *rps16-trnQ* and *ndhF-rpl32* and the coding region *ycf1b*. *rps16-trnQ* was located in the LSC region, and *ndhF-rps32* and *ycf1b* were located in the SSC region.

We compared the number of variable sites among the three hypervariable markers and the three universal DNA barcodes (*rbcL*, *matK*, and *trnH-psbA*). Marker information is shown in **Table 3**. The length of the three hypervariable markers was 971 bp (*rps16-trnQ*), 1,640 bp (*ndhF-rpl32*), and 1,843 bp (*ycf1b*). *Ycf1b* had the greatest number of variable sites (153 sites), followed by *ndhF-rpl32* (125 sites) and *rps16-trnQ* (72 sites). The number of variable sites for *rbcL*, *matK*, and *trnH-psbA* was 67, 73, and 40, respectively. *Ycf1b* had the highest π value and showed more sequence divergence among the six markers, followed by *trnH-psbA* and *rps16-trnQ*. *RbcL* and *matK* were less variable according to the π values.

Plastid Phylogenomics of *Arnebia*

All 78 protein-coding genes of the chloroplast genome were used to infer the phylogenetic relationships among *Arnebia* species and their allies through BI and ML (**Figure 3**). The nucleotide data matrix contained 67,557 sites, including 4,712 variable sites and 1,675 parsimony-informative sites. The amino acid data matrix

TABLE 1 | Chloroplast genome features of the five *Arnebia* species.

Species	Nucleotide length (bp)				GC content (%)	Number of genes			
	LSC	IR	SSC	Total		Protein	tRNA	rRNA	Total
Arnebia decumbens	80,462–80,826	25,943–25,945	17,188–17,203	149,539–149,919	37.8	78	30	4	112
Arnebia euchroma	81,086–81,708	25,918–25,968	17,256–17,336	150,224–150,900	37.6	78	30	4	112
Arnebia fimbriata	82,846–82,946	25,896–25,899	17,289–17,298	151,927–152,040	37.5	78	30	4	112
Arnebia guttata	80,737–81,265	25,967–26,053	17,143–17,213	149,840–150,410	37.7	78	30	4	112
Arnebia szechenyi	82,064–82,569	25,753–25,940	17,156–17,302	150,867–151,701	37.6-37.7	78	30	4	112

TABLE 2 | Analyses of the variable sites in chloroplast genomes of *Arnebia*.

Regions	Aligned length (bp)	Variable sites		Information sites		Nucleotide diversity	Number of haplotypes
		Numbers	%	Numbers	%		
LSC	87,252	3,518	4.03	2,861	3.28	0.00843	49
SSC	17,664	964	5.46	791	4.48	0.01234	48
IR	26,179	214	0.82	168	0.64	0.00177	36
Whole chloroplast genome	157,202	4,921	3.13	3,996	2.54	0.00654	49

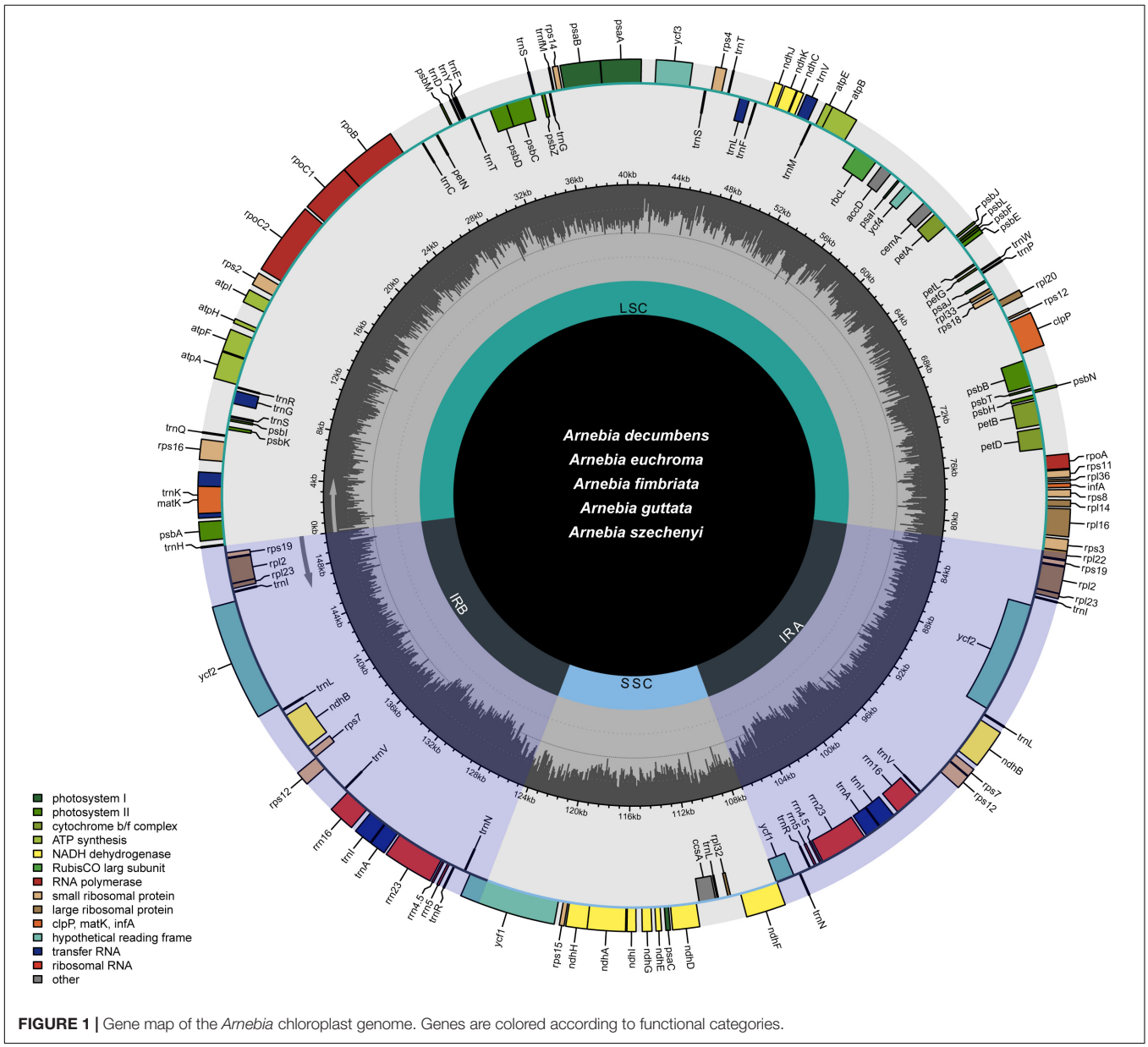


FIGURE 1 | Gene map of the *Arnebia* chloroplast genome. Genes are colored according to functional categories.

contained 22,519 sites, including 1,834 variable sites and 626 parsimony-informative sites.

The topologies of the ML and BI tree were consistent, and most nodes had strong support values. *Alkanna tinctoria* was the first group to diverge from Lithospermeae. The nucleotide dataset indicated that *Onosma* was the second group to diverge from [name]; however, the sequence of divergence inferred from the amino acid dataset differed from that inferred by the nucleotide

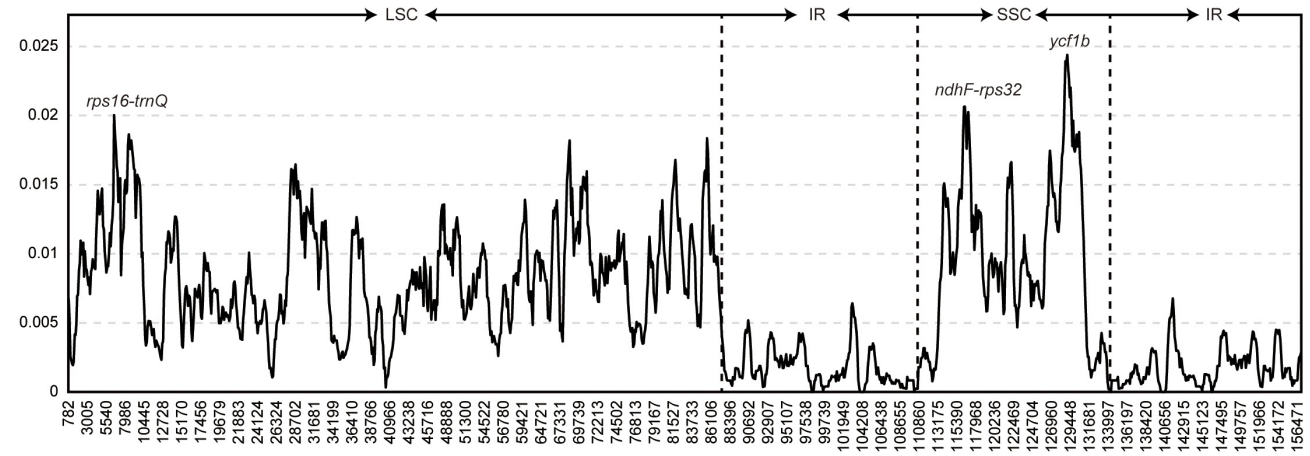


FIGURE 2 | Hypervariable regions in the *Arnebia* chloroplast genome. Window length: 800 bp; step size: 100 bp. Three regions with the highest π values were marked. *x*-axis: position of the midpoint of a window; *y*-axis: nucleotide diversity of each window.

TABLE 3 | Three hypervariable regions and the universal markers of chloroplast genomes of *Arnebia*.

Barcode	Number of sequences	Length (bp)	Variable sites		Information sites		Nucleotide diversity (π)
			Numbers	%	Numbers	%	
rps16-trnQ	56	971	72	7.42	60	6.18	0.02002
ndhF-rpl32	56	1,640	125	7.62	107	6.52	0.01852
ycf1b	56	1,843	153	8.30	130	7.05	0.02087
trnH-psbA	56	472	67	14.19	54	11.44	0.02019
matK	56	1,536	73	4.75	57	3.71	0.00936
rbcL	56	1,434	40	2.79	34	2.37	0.00668

dataset. Both datasets indicated that *Buglossoides* was sister to *Lithospermum* and formed a clade. *Arnebia*, *Buglossoides*, and *Lithospermum* formed a monophyletic group with low support values (BS = 59/PP = 0.72) based on the nucleotide dataset; the amnio acid dataset indicated that *Arnebia* was sister to *Onosma* (BS = 73/PP = 1). The inferred phylogenetic relationships among the five *Arnebia* species predicted by both datasets were similar (BS = 100/PP = 1). *A. decumbens* and *A. guttata* formed a clade with strong support values. *A. fimbriata* and *A. szechenyi* formed a clade that was sister to *A. euchroma* with moderate support values (BS = 62/PP = 0.87 or BS = 62/PP = 0.99). All the samples for each species formed a monophyletic group with high support values based on the whole chloroplast genome sequences.

Divergence Time Estimation

Divergence time estimates showed that the stem and crown nodes of *Arnebia* were 28.84 Ma (95% HPD: 20.05–35.69 Ma) in the middle Oligocene and 20.89 Ma (95% HPD: 12.82–29.46 Ma) in the early Miocene, respectively (**Figure 4**). The divergence time between *A. guttata* and *A. decumbens* was 12.75 Ma (95% HPD: 5.37–20.99 Ma) in the middle Miocene. The three species *A. fimbriata*, *A. szechenyi*, and *A. euchroma* diverged at 18.41 Ma, and the divergence time between *A. szechenyi* and *A. euchroma* was 14.54 Ma. The genotype-level divergence times were estimated using all samples (**Figure 5**). The crown ages of *A. szechenyi*, *A. euchroma*, *A. fimbriata*, *A. guttata*, and *A. decumbens* were 5.26, 3.48, 10.18, 2.54, and 1.68 Ma, respectively.

Intraspecific Diversity and Genetic Structure of *Arnebia euchroma*

The 15 aligned chloroplast genomes of *A. euchroma* were 152,175 bp in length. A total of 1,167 mutation sites were identified, including 339 singleton and 827 parsimony-informative sites (**Table 4**). We also identified 265 indels in the *A. euchroma* chloroplast genome. Phylogenetic analysis was performed using ML and BI and whole chloroplast genomes (**Figure 6A**). These samples were clearly divided into two clades (AE-I and AE-II). Population structure was analyzed using *K* values ranging from 1 to 10; the populations were clearly divided into two clades with $K = 2$ (**Figure 6B**). The PCA results revealed two major groups (**Figure 6C**). The AE-I clade included four samples located in southern Tibet (Jilong County and Zada County) (**Figure 6D**). The two clades exhibited significant genetic differences and diverged at 10.18 Ma in the middle Miocene (**Figure 5**).

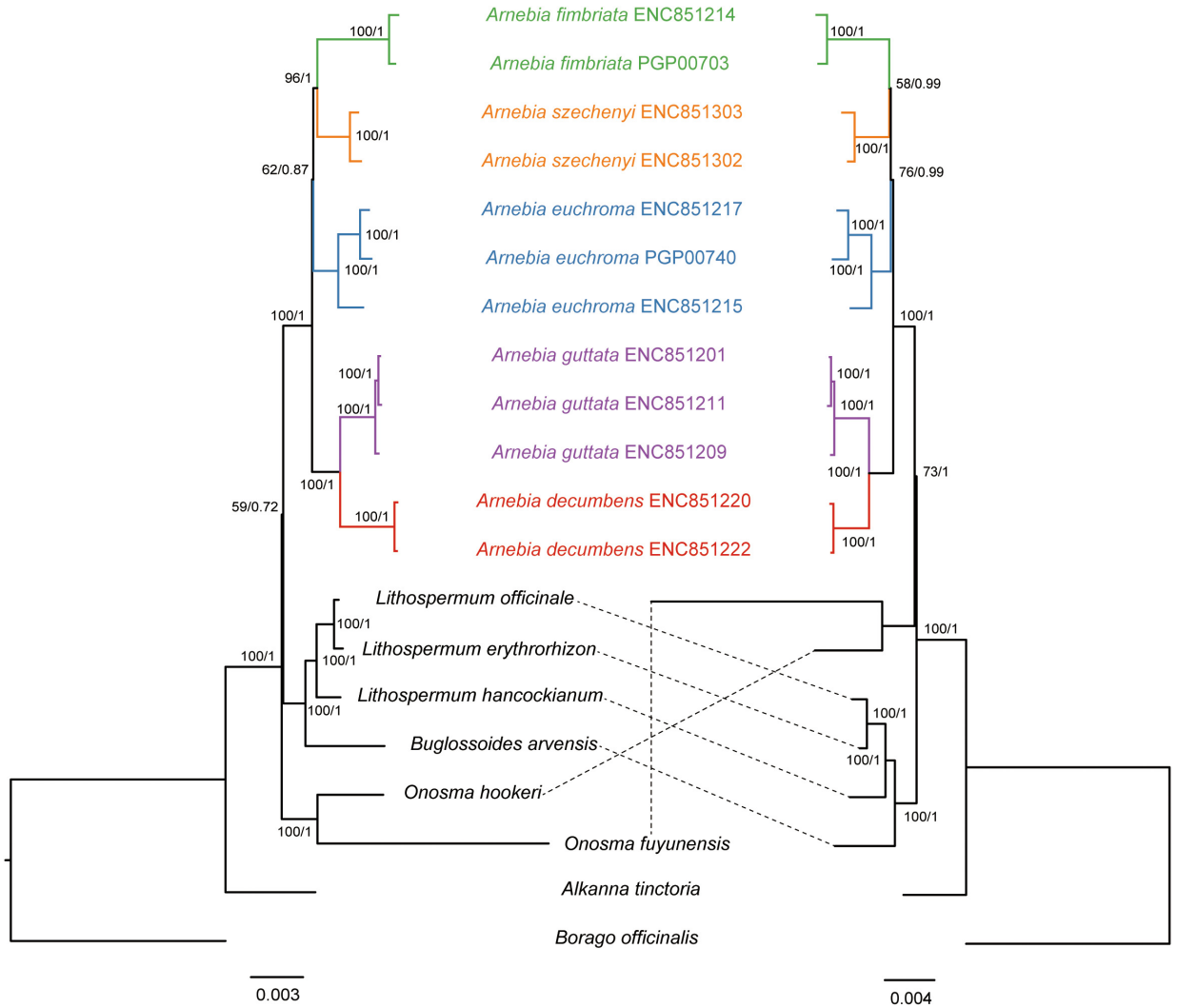


FIGURE 3 | Phylogenetic trees of Lithospermeae based on all protein-coding genes. The nucleotide dataset **(left)** and amino acid dataset **(right)**. Maximum likelihood (ML) bootstrap support values/Bayesian posterior probabilities are shown at each node.

Intraspecific Diversity and Genetic Structure of *Arnebia guttata*

The aligned 17 chloroplast genomes of *A. guttata* were 150,929 bp in length. A total of 313 mutation sites were screened, including 77 singleton and 236 parsimony-informative sites. The nucleotide diversity and number of haplotypes were 0.00058 and 14 for the 17 *A. guttata* chloroplast genome sequences, respectively. The sequence divergence among the *A. guttata* chloroplast genome sequences was lower than that among the *A. euchroma* chloroplast genome sequences (**Table 4**). The samples were clearly divided into three clades (AG-I, AG-II, and AG-III) according to the phylogenetic relationships (**Figure 7A**), STRUCTURE analysis (**Figure 7B**), and PCA (**Figure 7C**). The AG-I clade contained eight samples from Mongolia, Russia, Gansu Province, and northwestern Xinjiang, China. The AG-II clade included two samples from Tacheng, Xinjiang. The AG-III clade contained seven samples from northwestern Tibet (**Figure 7D**). The divergence times among the three clades ranged from 1.2 to 2.54 Ma in the Pleistocene (**Figure 5**).

DISCUSSION

Inter- and Intraspecific Variation in the Chloroplast Genomes of *Arnebia* Species

The aim of this study was to explore patterns of chloroplast genome divergence in *Arnebia* species in China. The chloroplast genomes of the five *Arnebia* species were conserved, as little variation was observed in their size, GC content, and gene order among species. The organization of *Arnebia* genomes was similar to that of other members of the family Boraginaceae and exhibited a typical quadripartite structure. The mVISTA

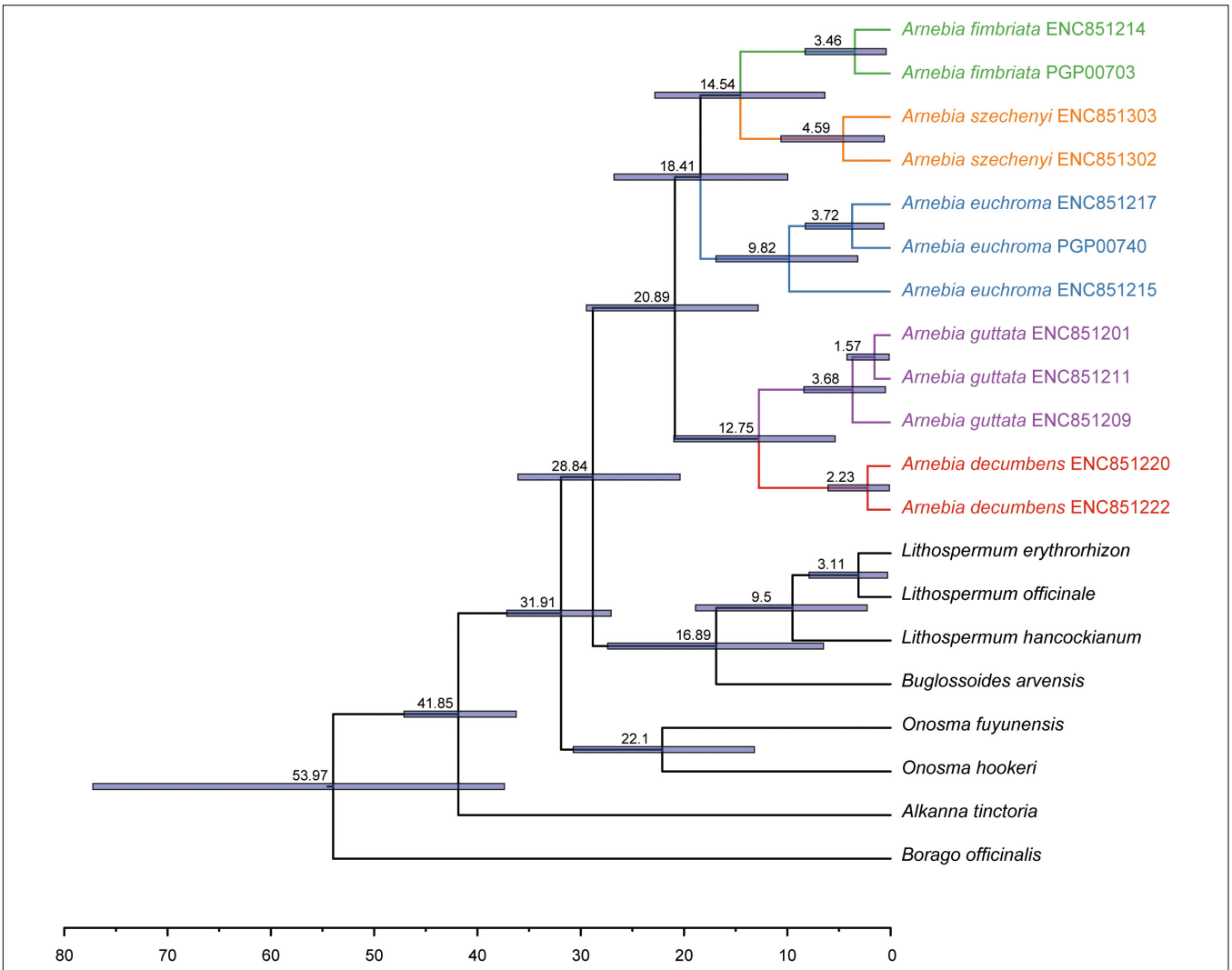


FIGURE 4 | Divergence times of Lithospermeae obtained from BEAST analysis based on all coding protein-coding genes with two priors. The mean divergence time of the nodes is shown next to the nodes, and the blue bars correspond to the 95% highest posterior density (HPD).

TABLE 4 | Intraspecific chloroplast genome sequence divergence of five *Arnebia* species.

Species	Number of samples	Alignment length (bp)	Number of variable sites			Nucleotide polymorphism	
			Polymorphic	Singleton	Parsimony informative	Nucleot ide diversity	Haplotypes
Arnebia decumbens	7	150,133	157	86	70	0.00041	7
Arnebia euchroma	15	152,175	1,167	339	827	0.00241	14
Arnebia fimbriata	6	152,386	304	135	168	0.00089	6
Arnebia guttata	17	150,929	313	77	236	0.00058	14
Arnebia szechenyi	11	152,438	709	445	259	0.00133	10

results and DnaSP results revealed that the LSC and SSC regions were more variable than the IR regions, and the non-coding regions were more variable than the coding regions. Mutation events were not random but clustered into hotspots in the chloroplast genome (**Figure 2** and **Supplementary Figure 1**). Mutational hotspot regions were more variable than universal chloroplast markers (*rbcL*, *matK*, and *trnH-psbA*) or commonly used markers, such as *ndhF* and *trnL-F*. Three variable regions (*rps16-trnQ*, *ndhF-rpl32*, and *ycf1b*) were identified in the *Arnebia* chloroplast genome. *rps16-trnQ* was located in the LSC

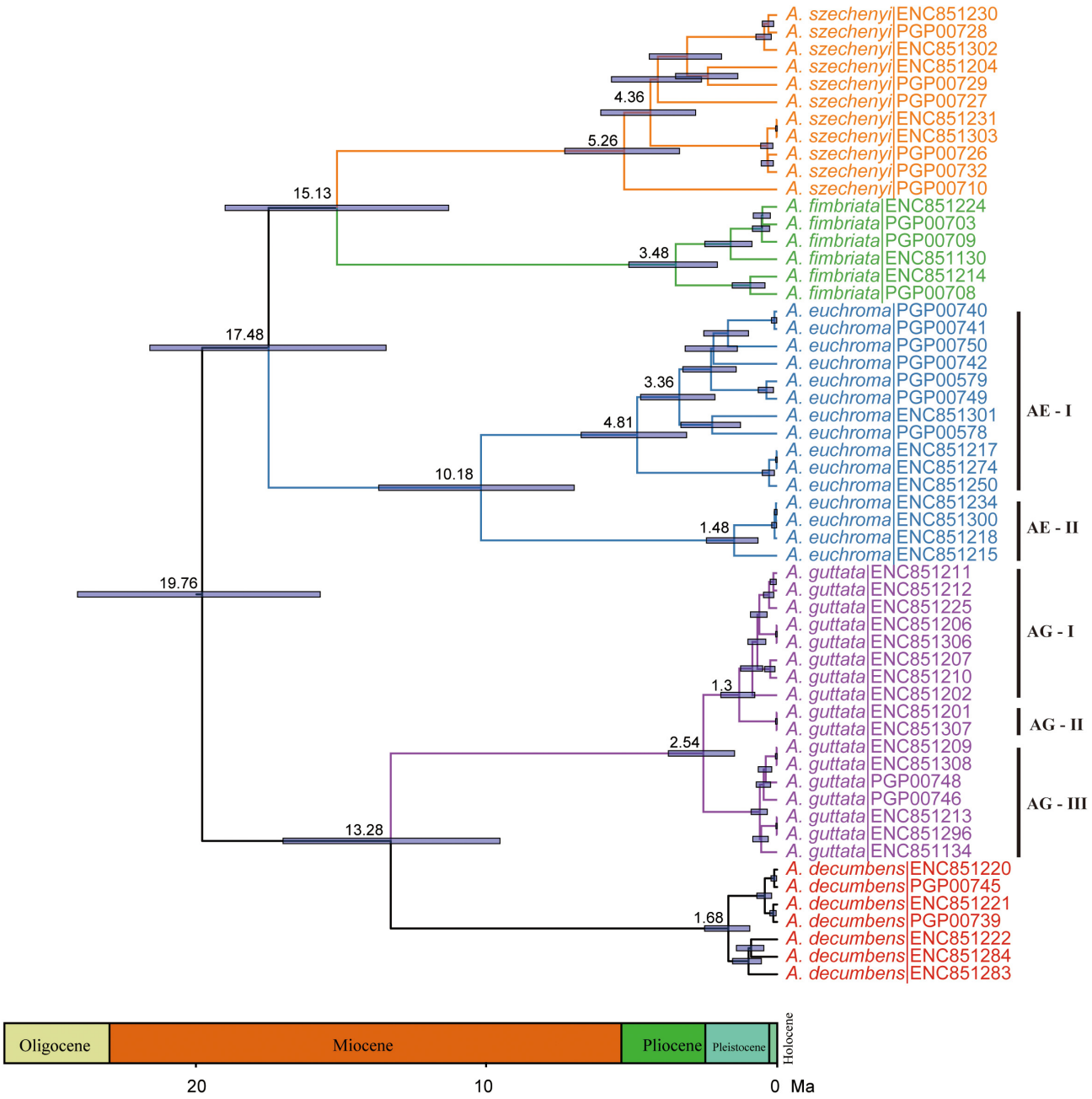


FIGURE 5 | Divergence times of *Arnebia* obtained from BEAST analysis based on the 65 whole chloroplast genomes. The mean divergence time of the nodes is shown next to the nodes, and the blue bars correspond to the 95% HPD.

region, and this marker has long been used for the reconstruction of plant phylogenies and species identification (Smedmark et al., 2008). The *rps16-trnQ* marker contains the *rps16* intron and the intergenic spacer *rps16-trnQ* (Shaw et al., 2005; Dong et al., 2012). *rps16-trnQ* shows greater variation in the chloroplast genome. However, it contains a larger inversion in papilionoid species. The intergenic spacer *ndhF-rpl32* is located in the SSC region. This marker was identified by Shaw et al. (2007) and Dong et al. (2012), and both of these studies suggested that this marker could be used to conduct phylogenetic studies at the species or subspecies level because of its higher level of variability compared with other chloroplast markers. The *ycf1* gene is the second longest gene in the chloroplast genome, and this gene spans the IR and SSC regions. Two variable regions in this gene, *ycf1*a and *ycf1*b, have been identified (Dong et al., 2012, 2015). In the *Arnebia* chloroplast genome, *ycf1*b was the most variable marker (**Figure 2** and **Table 3**). *ycf1* has also been shown to vary substantially among different varieties (Magdy et al., 2019; Xiao et al., 2021).

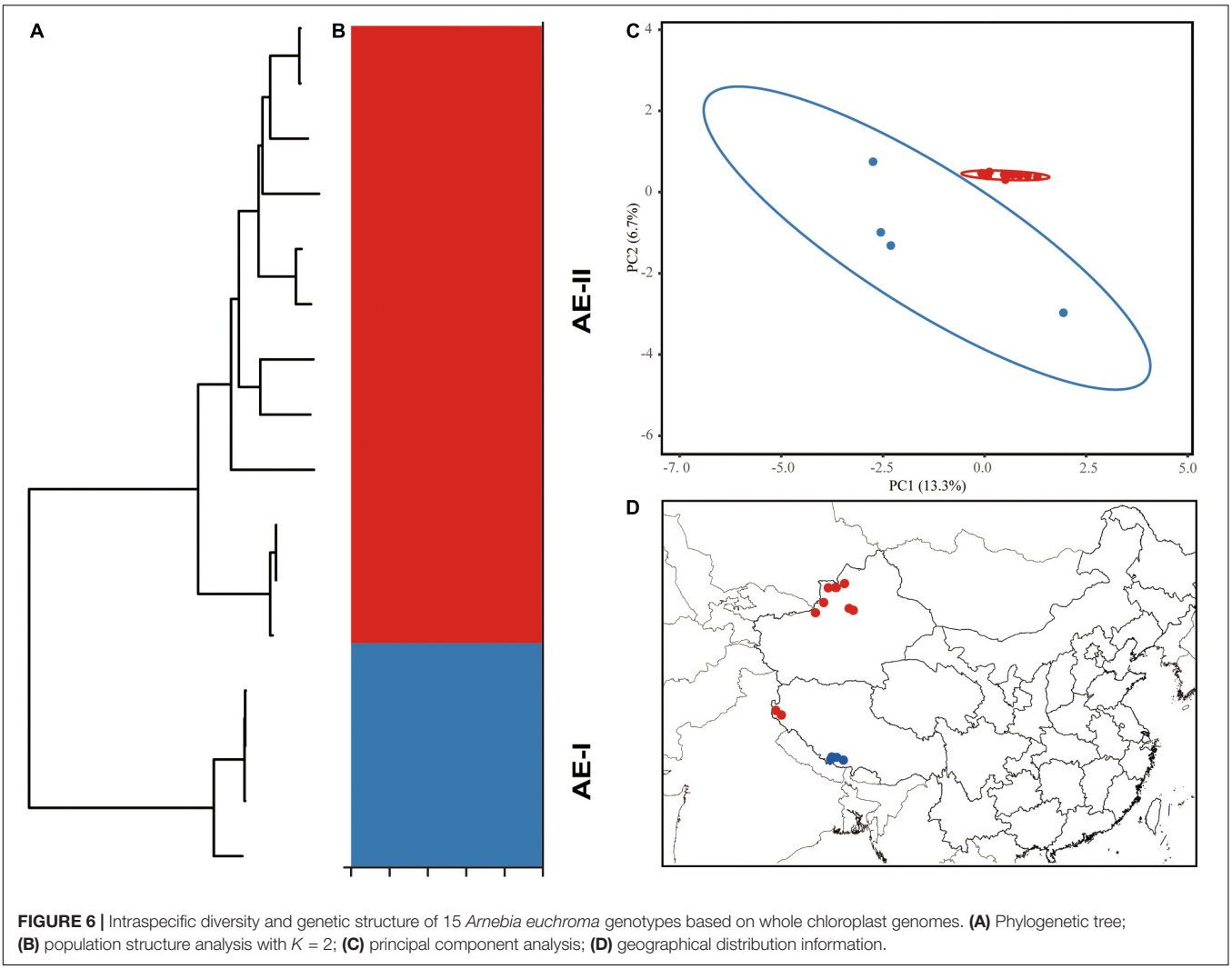


FIGURE 6 | Intraspecific diversity and genetic structure of 15 *Arnebia euchroma* genotypes based on whole chloroplast genomes. **(A)** Phylogenetic tree; **(B)** population structure analysis with $K = 2$; **(C)** principal component analysis; **(D)** geographical distribution information.

Chloroplast genome markers have long been used in plant population genetic studies, but these markers generally have few polymorphic sites. We identified many intraspecific mutations in *Arnebia* species (**Table 4**), and *A. euchroma* and *A. szechenyi* had a higher number of variable sites in the chloroplast genome. The polymorphic sites generate a sufficient number of haplotypes, and most of the sampled population exhibited private haplotypes. There were more variable sites in the chloroplast genomes of *Arnebia* species compared with other woody plants. Variation in the substitution rate among taxa has been examined by various studies (Smith and Donoghue, 2008; Amanda et al., 2017; Schwarz et al., 2017; Choi et al., 2021). Generation time variation is the generally accepted explanation for substitution rate variation; specifically, nucleotide substitution rates are thought to be negatively correlated with generation time (Smith and Donoghue, 2008; Amanda et al., 2017).

The genetic information contained in the chloroplast genome at the inter- and intraspecific levels can improve our understanding of plant evolution and population genetics. The rich genetic variation of *Arnebia* species can also enhance the breeding of herbal varieties.

Phylogenetic Relationships and Divergence Times Among *Arnebia* Species

Arnebia species are annual or perennial plants. Johnston (1954) suggested that the genus comprises two sections: Sect. *Euarnebia*, which includes one annual species *A. tetrastigma* Forsskål, and sect. *Strobilia*, which is further subdivided into three subgroups according to its life cycle (annual/perennial) and the presence of a pubescent annulus at the base of the corolla tube. Molecular data have indicated that *Arnebia s.l.* (including *Macrotomia*) is not monophyletic (Coppi et al., 2015); however, the molecular markers used in these studies do not provide sufficient resolution for resolving species-level relationships in this group (Weigend et al., 2009; Coppi et al., 2015; Chacón et al., 2019). In this study, we used whole chloroplast genomes to infer the phylogenetic relationships among *Arnebia* species in China, and

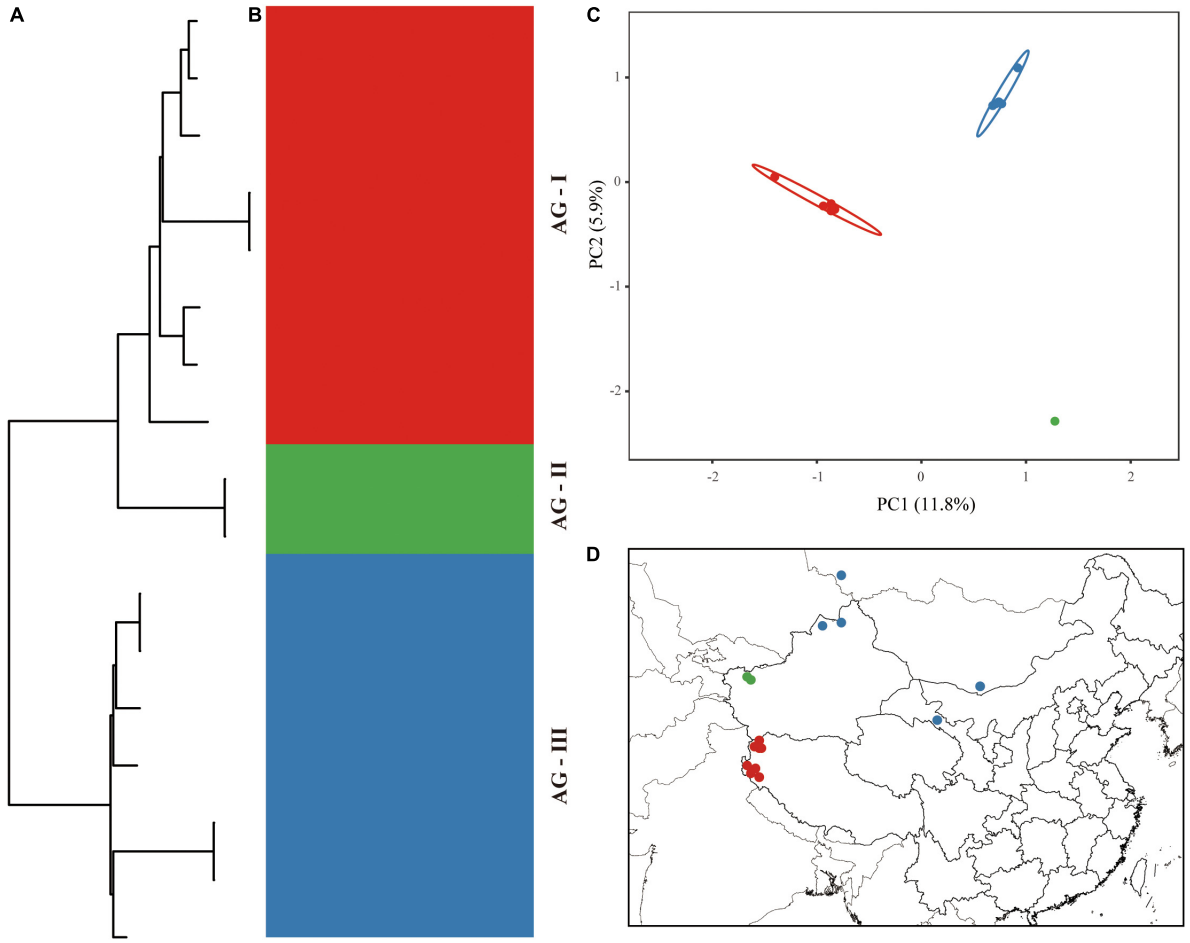


FIGURE 7 | Intraspecific diversity and genetic structure of 17 *A. guttata* genotypes based on whole chloroplast genomes. **(A)** Phylogenetic tree; **(B)** population structure analysis with $K = 3$; **(C)** principal component analysis; **(D)** geographical distribution information.

the whole chloroplast genomes provided sufficient information for resolving phylogenetic relationships at the species and subspecies level. The chloroplast genome tree did not group species by life cycle, and this finding was inconsistent with the internal transcribed spacer tree, which indicates that *Arnebia s.s.* formed a monophyletic group subdivided into two lineages corresponding to the groups of annual and perennial species (Coppi et al., 2015).

We estimated the divergence time of *Arnebia* based on the chloroplast genome at the species and subspecies levels. *Arnebia* originated from the Oligocene and diversified in the middle Miocene (**Figure 4**; Chacón et al., 2019). Late Oligocene warming and the progressive uplift of Tianshan and the Himalayas in the early Miocene (Favre et al., 2015) might be the main factors underlying the origin and diversification of *Arnebia*.

Genetic Divergence of *Arnebia euchroma* and *Arnebia guttata* in China

Arnebia euchroma and *A. guttata* are the two source plants of *Arnebiae Radix*. Wild populations of these species have declined due to overharvesting, habitat destruction, and fragmentation (Samant et al., 1998; Qin et al., 2017). Both species are difficult to grow *via* conventional agricultural practices (Kumar et al., 2021), and this has resulted in the increased use of wild resources. Studies of genetic diversity are essential for plant seeding, agricultural practices, as well as the conservation of endangered species.

Our analyses based on the whole chloroplast genome revealed a considerably high level of genetic variation within *A. euchroma* and *A. guttata* (**Table 4**). Most of the samples had specific genotypes. *A. euchroma* was divided into two clades that showed significant genetic divergence (**Figure 6**). Divergence time estimation indicated that the two clades diverged the in late Miocene (10.18 Ma), and intraspecific divergence occurred earlier compared with other plants. We carefully examined the morphological characteristics of the populations from southern Xizang (Jilong County and Zada County) (AE-I clade), and none of them were misidentified. The AE-I clade might contain cryptic species, and more samples and field studies need to be performed in the future to explore this possibility. *A. guttata* was the source of Inner Mongolia *Arnebiae Radix*. Chinese populations

comprised three subclades (**Figure 7**), which were consistent with the geographical distributions of this species. The divergence time of this species was estimated to be 2.54 Ma in the Pleistocene.

Several studies have shown that some herbaceous plants occurring in drylands show high levels of genetic variation (Meng and Zhang, 2011; Zhang et al., 2017; Fu et al., 2021). Ecological conditions in the drylands and the long evolutionary histories of desert plants are probably associated with the high levels of genetic variation (Meng and Zhang, 2011). The drylands and the mountains likely result in discrete patches of vegetation, which reduces gene flow and genetic similarity among populations (Maestre et al., 2012).

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the GenBank, accession number ON529903-ON529958.

AUTHOR CONTRIBUTIONS

WD, LG, and LH conceived and designed the work. JS, SW, RW, KL, PQ, and LS collected the samples. JS, EL, YW, SW, RW, KL, PQ, and LS performed the experiments and analyzed the data. WD, JS, and SW wrote the manuscript. LG and LH revised the manuscript. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (82173934), the Fundamental Research Funds for the Central Public Welfare Research Institutes (ZZXT201901), CACMS Innovation Fund (CI2021A03904) and Key Project at Central Government level: the ability establishment of sustainable use for valuable Chinese medicine resources (2060302).

ACKNOWLEDGMENTS

We thank the DNA Bank of China in Institute of Botany, Chinese Academy of Sciences for providing materials.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.920826/full#supplementary-material

Supplementary Figure 1 | Comparison of the five *Arnebia* species chloroplast genomes using mVISTA.

REFERENCES

Amanda, R., Li, Z., van de Peer, Y., and Ingvarsson, P. K. (2017). Contrasting rates of molecular evolution and patterns of selection among gymnosperms and flowering plants. *Mol. Biol. Evol.* 34, 1363–1377. doi: 10.1093/molbev/msx069

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C. H., Xie, D., et al. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comp. Biol.* 10:e1003537.

Chacón, J., Luebert, F., Selvi, F., Cecchi, L., and Weigend, M. (2019). Phylogeny and historical biogeography of Lithospermeae (Boraginaceae): Disentangling the possible causes of Miocene diversifications. *Mol. Phylogenet. Evol.* 141:106626. doi: 10.1016/j.ympev.2019.106626

Choi, K., Weng, M.-L., Ruhlman, T. A., and Jansen, R. K. (2021). Extensive variation in nucleotide substitution rate and gene/intron loss in mitochondrial genomes of *Pelargonium*. *Mol. Phylogenet. Evol.* 155:106986. doi: 10.1016/j.ympev.2020.106986

Coppi, A., Cecchi, L., Nocentini, D., and Selvi, F. (2015). Arnebia purpurea: a new member of formerly monotypic genus Huynhia (Boraginaceae-Lithospermeae). *Phytotaxa* 204, 123–136.

Dong, W., Li, E., Liu, Y., Xu, C., Wang, Y., Liu, K., et al. (2022a). Phylogenomic approaches untangle early divergences and complex diversifications of the olive plant family. *BMC Biol.* 20:92. doi: 10.1186/s12915-022-01297-0

Dong, W., Liu, J., Yu, J., Wang, L., and Zhou, S. (2012). Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS One* 7:e35071. doi: 10.1371/journal.pone.0035071

Dong, W., Liu, Y., Li, E., Xu, C., Sun, J., Li, W., et al. (2022b). Phylogenomics and biogeography of Catalpa (Bignoniaceae) reveal incomplete lineage sorting and three dispersal events. *Mol. Phylogenet. Evol.* 166:107330. doi: 10.1016/j.ympev.2021.107330

Dong, W., Liu, Y., Xu, C., Gao, Y., Yuan, Q., Suo, Z., et al. (2021a). Chloroplast phylogenomic insights into the evolution of *Distylium* (Hamamelidaceae). *BMC Genom.* 22:293. doi: 10.1186/s12864-021-07590-6

Dong, W., Sun, J., Liu, Y., Xu, C., Wang, Y., Suo, Z., et al. (2021b). Phylogenomic relationships and species identification of the olive genus *Olea* (Oleaceae). *J. Syst. Evol.* 2021:802. doi: 10.1111/jse.12802

Dong, W., Xu, C., Cheng, T., Lin, K., and Zhou, S. (2013). Sequencing angiosperm plastid genomes made easy: a complete set of universal primers and a case study on the phylogeny of Saxifragales. *Genome Biol. Evol.* 5, 989–997. doi: 10.1093/gbe/evt063

Dong, W., Xu, C., Li, C., Sun, J., Zuo, Y., Shi, S., et al. (2015). ycf1, the most promising plastid DNA barcode of land plants. *Sci. Rep.* 5:8348. doi: 10.1038/srep08348

Dong, W., Xu, C., Wen, J., and Zhou, S. (2020). Evolutionary directions of single nucleotide substitutions and structural mutations in the chloroplast genomes of the family Calycanthaceae. *BMC Evol. Biol.* 20:96. doi: 10.1186/s12862-020-01661-0

Dong, W., Xu, C., Wu, P., Cheng, T., Yu, J., Zhou, S., et al. (2018). Resolving the systematic positions of enigmatic taxa: manipulating the chloroplast genome data of Saxifragales. *Mol. Phylogenet. Evol.* 126, 321–330. doi: 10.1016/j.ympev.2018.04.033

Falush, D., Stephens, M., and Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* 7, 574–578. doi: 10.1111/j.1471-8286.2007.01758.x

Favre, A., Päckert, M., Pauls, S. U., Jähnig, S. C., Uhl, D., Michalak, I., et al. (2015). The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biol. Rev.* 90, 236–253. doi: 10.1111/brv.12107

Feng, J., Yu, P., Zhou, Q., Tian, Z., Sun, M., Li, X., et al. (2020). An integrated data filtering and identification strategy for rapid profiling of chemical constituents, with Arnebiae Radix as an example. *J. Chromatograph.* 1629:461496. doi: 10.1016/j.chroma.2020.461496

Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 32, W273–W279. doi: 10.1093/nar/gkh458

Fu, M.-J., Wu, H.-Y., Jia, D.-R., and Tian, B. (2021). Evolutionary history of a desert perennial Arnebia szechenyi (Boraginaceae): Intraspecific divergence, regional expansion and asymmetric gene flow. *Plant Divers.* 43, 462–471. doi: 10.1016/j.pld.2021.04.002

Funk, W. C., Mckay, J. K., Hohenlohe, P. A., and Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends Ecol. Evol.* 27, 489–496. doi: 10.1016/j.tree.2012.05.012

Gupta, K., Garg, S., Singh, J., and Kumar, M. (2014). Enhanced production of napthoquinone metabolite (shikonin) from cell suspension culture of Arnebia sp. and its up-scaling through bioreactor. *3 Biotech* 4, 263–273. doi: 10.1007/s13205-013-0149-x

Huang, D. I., and Cronk, Q. C. B. (2015). Plann: A command-line application for annotating plastome sequences. *Appl. Plant Sci.* 3:1500026. doi: 10.3732/apps.1500026

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 Biol.* 21:241. doi: 10.1186/s13059-020-02154-5

Johnston, I. M. (1954). Studies in the Boraginaceae XXVI. Further revaluations of the genera of the Lithospermeae. *J. Arnold Arboretum* 35, 1–81.

Kala, C. P. (2000). Status and conservation of rare and endangered medicinal plants in the Indian trans-Himalaya. *Biol. Conserv.* 93, 371–379.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285

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

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35, 4453–4455. doi: 10.1093/bioinformatics/btz305

Kumar, A., Shashni, S., Kumar, P., Pant, D., Singh, A., and Verma, R. K. (2021). Phytochemical constituents, distributions and traditional usages of Arnebia euchroma: a review. *J. Ethnopharmacol.* 271:113896. doi: 10.1016/j.jep.2021.113896

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Lal, M., Samant, S., Kumar, R., Sharma, L., Paul, S., Dutt, S., et al. (2020). Population ecology and niche modelling of endangered Arnebia euchroma in Himachal Pradesh, India-An approach for conservation. *Med. Plants-Int. J. Phytomed. Rel. Indust.* 12, 90–104.

Li, J., Wang, S., Jing, Y., Wang, L., and Zhou, S. (2013). A modified CTAB protocol for plant DNA extraction. *Chin. Bull. Bot.* 48, 72–78.

Liao, M., Yan, P., Liu, X., Du, Z., Jia, S., Aybek, R., et al. (2020). Spectrum-effect relationship for anti-tumor activity of shikonins and shikonofurans in medicinal Zicao by UHPLC-MS/MS and chemometric approaches. *J. Chromatograph. B* 1136:121924. doi: 10.1016/j.jchromb.2019.121924

Liu, J., Zhao, C., Guo, L., Ma, S., Zheng, J., and Wang, J. (2020). The plant original identification of Inner Mongolia Arnebia radix based on morphology and DNA barcoding. *Res. Squ.* 2020:6478.

Ma, S.-J., Geng, Y., Ma, L., and Zhu, L. (2021). Advances in Studies on Medicinal Arnebiae Radix. *Mod. Chin. Med.* 23, 177–184. doi: 10.1016/j.fitote.2019.03.005

Maestre, F. T., Salguero-Gómez, R., and Quero, J. L. (2012). It is getting hotter in here: determining and projecting the impacts of global environmental change on drylands. *Philos. Trans. R. Soc. B: Biol. Sci.* 367, 3062–3075. doi: 10.1098/rstb.2011.0323

Magdy, M., Ou, L., Yu, H., Chen, R., Zhou, Y., Hassan, H., et al. (2019). Pan-plastome approach empowers the assessment of genetic variation in cultivated Capsicum species. *Horticult. Res.* 6:108. doi: 10.1038/s41438-019-0191-x

Meng, H.-H., and Zhang, M.-L. (2011). Phylogeography of Lagochilus ilicifolius (Lamiaceae) in relation to Quaternary climatic oscillation and aridification in northern China. *Biochem. Syst. Ecol.* 39, 787–796.

Park, I., Yang, S., Song, J.-H., and Moon, B. C. (2020). Dissection for floral micromorphology and plastid genome of valuable medicinal borages arnebia and lithospermum (Boraginaceae). *Front. Plant Sci.* 11:606463. doi: 10.3389/fpls.2020.606463

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Human Genet.* 81, 559–575. doi: 10.1086/519795

Qiao, J., Zhang, X., Chen, B., Huang, F., Xu, K., Huang, Q., et al. (2020). Comparison of the cytoplastic genomes by resequencing: insights into the genetic diversity and the phylogeny of the agriculturally important genus Brassica. *BMC Genom.* 21:480. doi: 10.1186/s12864-020-06889-0

Qin, H., Yang, Y., Dong, S., He, Q., Jia, Y., Zhao, L., et al. (2017). Threatened species list of China's higher plants. *Biodiv. Sci.* 25:696. doi: 10.1186/s13002-020-00420-1

Rambaut, A. (1996). *Se-Al: sequence alignment editor. version 2.0.*

Rambaut, A., Suchard, M., Xie, D., and Drummond, A. (2014). *Tracer v1. 6.*

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Hohna, 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

Rozas, J., Ferrer-Mata, A., Sanchez-Delbarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248

Samant, S. S., Dhar, U., and Rawal, R. S. (1998). Biodiversity status of a protected area in West Himalaya: askot wildlife sanctuary. *Int. J. Sust. Dev. World Ecol.* 5, 194–203.

Schwarz, E. N., Ruhlman, T. A., Weng, M.-L., Khiyami, M. A., Sabir, J. S. M., Hajarah, N. H., et al. (2017). Plastome-wide nucleotide substitution rates reveal accelerated rates in Papilionoideae and correlations with genome features across legume subfamilies. *J. Mol. Evol.* 84, 187–203. doi: 10.1007/s00239-017-9792-x

Shaw, J., Lickey, E. B., Beck, J. T., Farmer, S. B., Liu, W. S., Miller, J., et al. (2005). The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* 92, 142–166. doi: 10.3732/ajb.92.1.142

Shaw, J., Lickey, E. B., Schilling, E. E., and Small, R. L. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am. J. Bot.* 94, 275–288. doi: 10.3732/ajb.94.3.275

Ślipiko, M., Myszczyński, K., Buczkowska, K., Bączkiewicz, A., Szczecińska, M., and Sawicki, J. (2020). Molecular delimitation of European leafy liverworts of the genus *Calypogeia* based on plastid super-barcodes. *BMC Plant Biol.* 20:243. doi: 10.1186/s12870-020-02435-y

Smedmark, J. E. E., Rydin, C., Razafimandimbison, S. G., Khan, S. A., Liede-Schumann, S., and Bremer, B. (2008). A phylogeny of Urophylleae (Rubiaceae) based on rps16 intron data. *Taxon* 57, 24–32.

Smith, S. A., and Donoghue, M. J. (2008). Rates of molecular evolution are linked to life history in flowering plants. *Science* 322, 86–89. doi: 10.1126/science.1163197

Sugiura, M. (1992). The chloroplast genome. *Plant Mol. Biol.* 19, 149–168.

Villanueva-Corrales, S., García-Botero, C., Garcés-Cardona, F., Ramírez-Ríos, V., Villanueva-Mejía, D. F., and Álvarez, J. C. (2021). The complete chloroplast genome of plukenetia volubilis provides insights into the organelle inheritance. *Front. Plant Sci.* 12:60. doi: 10.3389/fpls.2021.667060

Weigend, M., Gottschling, M., Selvi, F., and Hilger, H. H. (2009). Marbleseeds are gromwells – Systematics and evolution of Lithospermum and allies (Boraginaceae tribe Lithospermeae) based on molecular and morphological data. *Mol. Phylogenet. Evol.* 52, 755–768. doi: 10.1016/j.ympev.2009.05.013

Xiao, S., Xu, P., Deng, Y., Dai, X., Zhao, L., Heider, B., et al. (2021). Comparative analysis of chloroplast genomes of cultivars and wild species of sweetpotato (Ipomoea batatas [L.] Lam). *BMC Genom.* 22:262.

Xu, H., Li, P., Ren, G., Wang, Y., Jiang, D., and Liu, C. (2021). Authentication of three source spices of arnebiae radix using DNA Barcoding and HPLC. *Front. Pharmacol.* 12:14. doi: 10.3389/fphar.2021.677014

Zhang, Y., Yu, Q., Zhang, Q., Hu, X., Hu, J., and Fan, B. (2017). Regional-scale differentiation and phylogeography of a desert plant Allium mongolicum (Liliaceae) inferred from chloroplast DNA sequence variation. *Plant Syst. Evol.* 303, 451–466.

Zheng, S., Poczai, P., Hyvönen, J., Tang, J., and Amiryousefi, A. (2020). Chloroplot: an online program for the versatile plotting of organelle genomes. *Front. Genet.* 11:124. doi: 10.3389/fgene.2020.576124

Zhu, G.-L. (1982). A study on the taxonomy and distribution of Lithospermum and Arnebia in China. *J. Syst. Evol.* 20, 323–328.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 17 June 2022
doi: 10.3389/fpls.2022.875404

Genome-Wide Identification of Genes Related to Biosynthesis of Phenolic Acid Derivatives in *Bletilla striata* at Different Suspension Culture Stages

Houbo Liu[1,2†], *Ceyin Huang*[1†], *Qingqing Li*[1], *Mufei Wang*[1], *Shiji Xiao*[3], *Junhua Shi*[4], *Yihuai He*[4], *Weie Wen*[1*], *Lin Li*[1*] *and Delin Xu*[1*]

[1] Department of Cell Biology, Zunyi Medical University, Zunyi, China, [2] Department of Dermatology, Chengdu Second People's Hospital, Chengdu, China, [3] School of Pharmacy Chemistry, Zunyi Medical University, Zunyi, China, [4] Affiliated Hospital of Zunyi Medical University, Zunyi, China

OPEN ACCESS

Edited by:
Da-Cheng Hao,
Dalian Jiaotong University, China

Reviewed by:
Liubov Skrypnik,
Immanuel Kant Baltic Federal University, Russia
Pei Li,
Chinese Academy of Medical Sciences and Peking Union Medical College, China
Pavel Feduraev,
Immanuel Kant Baltic Federal University, Russia

***Correspondence:**
Delin Xu
xudelin2000@163.com
orcid.org/0000-0003-3695-2997
Lin Li
lilinzmc2015@163.com
Weie Wen
wenweie@live.com

[†]These authors share first authorship

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: 04 March 2022
Accepted: 25 April 2022
Published: 17 June 2022

Citation:
Liu H, Huang C, Li Q, Wang M, Xiao S, Shi J, He Y, Wen W, Li L and Xu D (2022) Genome-Wide Identification of Genes Related to Biosynthesis of Phenolic Acid Derivatives in Bletilla striata at Different Suspension Culture Stages. Front. Plant Sci. 13:875404. doi: 10.3389/fpls.2022.875404

To screen the genes regulating the biosynthesis of phenolic acid derivatives from the genome of *Bletilla striata*, we designed a suspension culture system to sample the cells for the following experiments. The contents of four phenolic acid derivatives were determined by high-performance liquid chromatography, and several full-length transcriptome sequencings of RNA samples at 10 time points were performed for bioinformatics analysis. The correlation analysis was used to identify and verify the key DEGs involved in the biosynthesis of the four phenolic acid derivatives. The results showed that the contents of p-hydroxybenzylalcohol (HBA), Dactylorhin A, Militarine, and Coelonin peaked at 33 days postinoculation (Dpi), 18 Dpi, 39 Dpi, and 39 Dpi of the culture system, respectively. Based on transcriptome data, 80 DEGs involved in the biosynthesis of phenolic acid derivatives were obtained. The KEGG pathway enrichment analysis classified them mostly into five metabolic pathways: phenylpropane biosynthesis, starch and sucrose metabolic, cyanoamino acid metabolism, gluconeogenesis and glycolysis, and phenylalanine metabolism. qPCR analysis revealed that the relative gene expression levels were consistent with the overall trend of transcriptome sequencing results. Among them, 14, 18, 23, and 41 unigenes were found to be involved in the synthesis of HBA, Dactylorhin A, Coelonin, and Militarine, respectively. These unigenes laid a solid foundation for elucidating the biosynthesis mechanism of phenolic acid derivatives in suspension cells of *B. striata*.

Keywords: *Bletilla striata*, phenolic acid derivatives, suspension culture system, biosynthesis, regulate genes

INTRODUCTION

Bletilla striata is a perennial herb of *Orchidaceae* andalso a traditional and precious Chinese herbal medicine with a high concentration of medicinal metabolites. For thousands of years, it has been prescribed in traditional Chinese medicine to treat hematemesis and promote wound healing (Zhao et al., 2021). A recent report detailed that *B. striata* has therapeutic effects on gastrointestinal bleeding, hemoptysis, traumatic bleeding, and postpartum hemorrhage (Zhang et al., 2022). Phenolic acid derivatives are a class of compounds with a polyhydroxyphenol structure, which are important secondary metabolites for supporting most pharmacological activities, such

as antioxidant, anti-tumor, and anti-inflammatory, of *B. striata* (Kassim et al., 2010; Anantharaju et al., 2016; Nigro et al., 2017; Milovanović et al., 2021; Yuan et al., 2021). Among the metabolites, p-hydroxybenzylalcohol (HBA), Dactylorhin A, Militarine, and Coelonin were the most abundant active components in *B. striata*. In our previous study, we found that the content of these secondary metabolites could reach 0.793 mg/g, 7.792 mg/g, 9.447 mg/g, and 0.345 mg/g of dried suspension cells, respectively (Pan, 2019). A number of studies further demonstrated that HBA has anti-inflammatory, anti-oxidation, anti-convulsion, anti-depression, and ameliorating effects on memory impairment, as well as protective benefits against brain damage caused by cerebral ischemia and sedative and hypnotic effects (Luo et al., 2017; Ding et al., 2019; Zhang D. et al., 2020). In recent years, many kinds of research have further confirmed the anti-cancer and antioxidant properties of dihydrophenanthrene compounds (Boudjada et al., 2019; Jiang et al., 2019). In conclusion, HBA, Dactylorhin A, Militarine, and Coelonin are natural chemicals with the potential for drug development.

However, a variety of phenolic acid derivatives have been identified in species such as *B. striata*, *Flickingeria fimbriata,* (Wu et al., 2017) and *Dendrobium scabrilingue* (Sarakulwattana et al., 2020), but their biosynthetic mechanisms and the synergistic regulation mechanisms of multiple genes are still unclear. Therefore, it is important to find the genes that regulate the biosynthesis of secondary metabolites through genetic improvement way or engineering bacteria.

Former studies found that the biosynthesis pathways of phenolic acids mainly include the shikimic acid metabolic pathway, the phenylpropanoid metabolic pathway, and the flavonoid metabolic pathway (Marchiosi et al., 2020). However, the biosynthetic pathways of phenolic acid derivatives have not been clearly explained. By analyzing the metabolic pathways of key time nodes related to HBA, Dactylorhin A, Militarine, and Coelonin, we see that they may provide a theoretical basis for exploring the synthetic methods for phenolic acid derivatives.

In this study, aided by the full-length transcriptome sequencing of PacBio Sequel and Illumina short-read sequencing technology (Zhou et al., 2021), we first profiled gene expression in different growth stages of the lag phase, exponential phase, deceleration phase, and decline phase of the suspension culture system of *B. striata*. Then, we screened out the genes involved in the biosynthesis of phenolic acid derivatives. Finally, we verified the expression patterns of selected genes in the suspension system. The obtained conclusions may provide a theoretical foundation for the genetic breeding of *B. striata* as well as an experimental basis for exploring the molecular biosynthesis mechanism of phenolic acid derivatives.

MATERIALS AND METHODS

Material

The capsules were harvested from the *B. striata* Germplasm Garden of Zunyi Medical University, Xinpu District, Zunyi City, Guizhou Province, China (27°42' N, 107°01' E). After disinfection (Li et al., 2020), matured seeds obtained from capsules were induced for the following suspension culture. The information on the compound standards used for measuring the contents of phenolic acid derivatives is shown in **Supplementary Table S3**. In addition, the compound structures are shown in **Supplementary Figure S1**.

Methods

Culture of Suspension Cells

Suspension culture was carried out for the efficient biosynthesis of phenolic acid by selecting *B. striata* with relatively uniform growth and good induction. Based on the contents of various compounds, we used MS (Murashige & Skoog Medium) as the base medium, with the addition of 1.0 mg/L of 6-BA, 3.0 mg/L of 2,4-D, and 30.0 g/L of sucrose. The pH was adjusted to 5.8–6.0. After 45 days of induction and 30 days of subculture, the content of secondary metabolites of suspension-cultured cells was measured every 3 days for 45 days postinoculation (Dpi). In addition, the content of phenolic acid derivatives was calculated according to the standard curve (Pan, 2019). The suspension-cultured cells were stored at 3 Dpi, 9 Dpi, 18 Dpi, 27 Dpi, 30 Dpi, 33 Dpi, 36 Dpi, 39 Dpi, and 42 Dpi for measuring the chemical content and RNA-Seq sequencing as described below.

Content Determination of Phenolic Acid Derivatives

Starting from the day of the directional suspension culture, the contents of phenolic acid derivatives were measured every 3 days, and the sampling was repeated three times. After being drained and dried, the cell clusters were crushed with a mortar, and 0.2 g was accurately weighed and placed in a round-bottom flask. Then, the cells were heated and refilled with 70% methanol-water for 2 h, transferred to a 5-ml volumetric flask filled with 70% methanol-water to maintain a constant volume, shaken well, and filtered using a 0.45 μm filter membrane for subsequent experiments. The content of phenolic acid derivatives in cells was determined by HPLC (Waters e2695).

The chromatographic conditions and gradient elution conditions were as follows: Chromatographic conditions: column: Universil C18 column (250 mm×4.6 mm, 5 μm, Kromat Corporation); other conditions were listed in **Supplementary Table S4**.

Transcriptome Data Assembly and Sequencing

After grinding the selected callus materials of 10 time points with liquid nitrogen, total RNA was extracted by Trizol and stored at −80 °C for subsequent sequencing on PacBio Sequel and Illumina Hiseq 2500 platforms, which were entrusted to us by Novogene Co., Ltd. To obtain clean reads, high-throughput sequencing from a large amount of raw data was obtained, from which the adaptor sequence, primer sequence, and low-quality reads were removed. The LoRDEC (Salmela and Rivals, 2014) software was used to correct the PacBio Sequel data with the Illumina Hiseq 2500 data. The CD-HIT (Fu et al., 2012) software was used to de-redundant the corrected consensus sequence, and the obtained transcript set was applied as the reference sequence (ref) for subsequent analysis. Then, the clean reads of each sample obtained by Illumina sequencing were compared to the reference to determine the transcripts detected in each sample.

TABLE 1 | The information of qPCR primers.

Gene	Primer name	Primer	Primer length (nt)	Product length (bp)	Tm (°C)
NA1	NA1-F	ACGGAGGCGGAGGCAACTAC	20	119	63.12
	NA1-R	GCAGCAGCCGTGGTAGCAAC	20		63.04
NA2	NA2-F	GCAGCGGCGGAATCTCTTCG	20	82	62.49
	NA2-R	TGCTGTGGAGCGGGATGACC	20		63.51
NA3	NA3-F	CTGGTGTGCGAATGTTGCTTCATC	24	148	59.24
	NA3-R	GCTTTCACGTATCAGGCGTTTGC	23		59.61
NA4	NA4-F	GCAATGGGCTACAGGACCAGAAG	23	80	60.74
	NA4-R	TCCGCCGCCTCCGTGATAAC	20		63.34
C4H1	C4H1-F	GCGGCGGTGACTCTTGCTATC	21	138	61.64
	C4H1-R	GGTGGCTAGATTGCGGTGATTGAG	24		60.6
PAL1	PAL1-F	CCAGACTCGCCATTGCTGCTATC	23	149	60.88
	PAL1-R	GGCTATCTCTGCTCCCTTGAAACC	24		60.05
PAL2	PAL2-F	GCTCCTTGCAAGAGTCGATTT	21	117	55.40
	PAL2-R	CAGGCAGAGTCCCTCCATTA	20		56.50
BGLU1	BGLU1-F	CACCATCCCATAGCCCGTTTCTAC	24	139	60.04
	BGLU1-R	TCAACTCCCGTAGCAAAGCCATTC	24		59.76
BGLU2	BGLU2-F	GAAGGACGGTAGCACAGCAGATG	23	145	60.53
	BGLU2-R	CACCGCTCCTCTCCCTCTTGG	21		63.22
BGLU3	BGLU3-F	CGGCTCTCAGGGATGGCTCAG	21	140	63.51
	BGLU3-R	CTTTGGCACTCTCTCACAGGCTTC	24		60.48
actin	actin-F	AATCCCAAGGCAAACAGA	18	-	51.00
	actin-R	CACCATCACCAGAATCCAG	19		53.00

Screening of DEGs

Using the RSEM (Li and Dewey, 2011) software to carry out the statistical analysis on the comparison results of Bowtie2 and perform FPKM conversion to determine the expression levels of the genes detected in the samples. The resulting q-values were adjusted by Benjamini and Hochberg's method. DEGs among different samples were screened based on $|\log_2(\text{FoldChange})| \geq 2$ and a q-value < 0.05.

GO Enrichment and KEGG Pathway Analysis of DEGs

We used GO (http://www.geneontology.org/) and KEGG databases (http://www.genome.jp/kegg/) to perform functional enrichment analysis on the selected DEGs. It was considered that the GO function or KEGG pathway was significantly enriched (q-value < 0.05), and the detection was annotated to DEGs in different GO functional groups or metabolic pathways.

Screen the key Genes Involved in Regulating the Synthesis of Phenolic Acid Derivatives

Using the K-means cluster analysis method, the DEGs with similar expressions were grouped into one category, and the differential gene expression trend was analyzed. Using the R package to analyze the correlation between DEGs and phenolic acid derivatives of HBA, Dactylorhin A, Militarine, and Coelonin at 10 time points (Spearman correlation coefficient), the screening criteria were q-value < 0.05 and $R > 0.05$. Subsequently, according to the accumulation and change curves of the four compounds on different days, DEGs at different time points were analyzed by the KEGG pathway and GO function analysis. The DEGs with significant upregulation or downregulation were screened. The main metabolic pathways involved in DEGs were determined. Finally, combined with correlation analysis, the key genes involved in the synthesis of phenolic acid derivatives were screened.

QPCR Verification of DEGs

The detected DEGs closely related to each derivative were randomly selected for verification via qPCR. Approximately 1 μg total RNA was used as the template for synthesizing the first cDNA strand by reverse transcription. The SYBR PrimeScriptTM RT-PCR Kit (TaKaRa) fluorescence quantitative kit was used to carry out qPCR reactions on an AGS 4800 real-time qPCR instrument (Hangzhou Anyang Technology Co., Ltd., Hangzhou, China). Primer Premier software was used to design specific primers for qPCR reactions of different randomly selected genes (**Table 1**). The approaches for cDNA reverse transcription and the qPCR reaction system are detailed in **Supplementary Tables S5**, **S6**. Besides, the cDNA reverse transcription reaction conditions were 42 °C for 60 min and 95 °C for 3 min. In addition, qPCR reaction conditions were enzyme activation at 95°C for 5 min; denaturation at 95°C for 5 s, and extension at 60°C for 30 s, 45 cycles.

Data Statistical Analysis

Each experiment was conducted with three biological replicates and three technical replicates. The $2^{-\Delta\Delta Ct}$ method was used to analyze gene expression levels in qPCR arrays. The statistical analysis of the data was

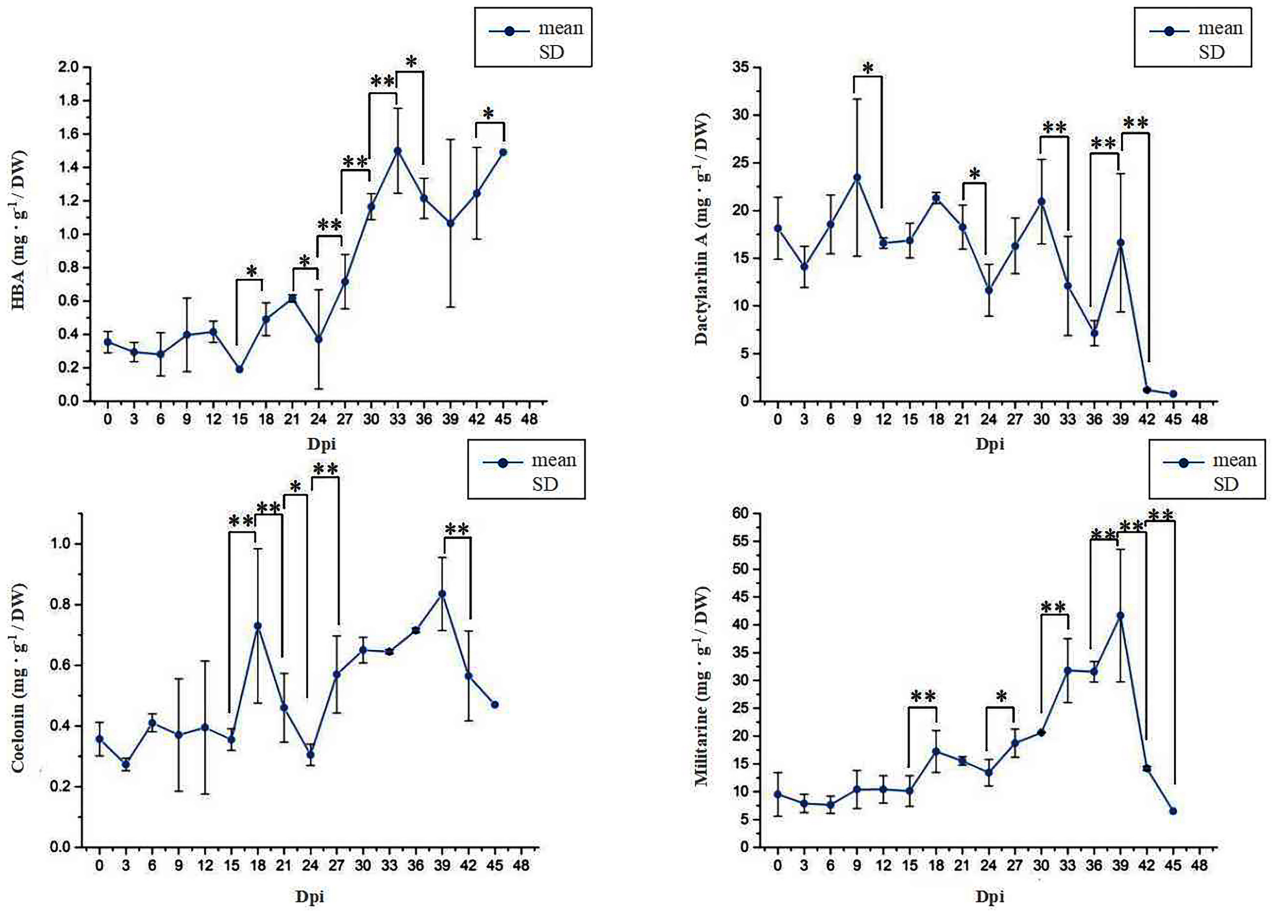


FIGURE 1 | The content changes of four phenolic acid derivatives. * Represents the significant difference between the two groups compared with the previous group at a level of $P < 0.05$. ** represents the significant difference at a level of $P < 0.01$.

run in SPSS, and the statistical methods of the one-way ANOVA and multiple comparisons were applied to the harvested data.

RESULTS

Determination of the Content of Phenolic Acid Derivatives in Suspension Cells of *B. striata*

The content of the four metabolites varied significantly in different growth stages of suspension culture cells (**Figure 1**). With the growth of culture, the accumulation of HBA in the suspension culture cells increased first and then decreased before finally reaching the maximum peak at 33 Dpi. The accumulation of Dactylorhin A gradually reduced as the number of days increased and then decreased sharply after 30 Dpi. The accumulation of Militarine peaked at 39 Dpi and then reduced. The accumulation of Coelonin first increased and then decreased with the increase of time, reaching its peak at 39 Dpi. According to the changing trend of the content of different compounds, we selected 18 Dpi, 27 Dpi, and 33 Dpi for HBA; 18 Dpi, 36 Dpi, and 42 Dpi for Dactylorhin A; 3 Dpi, 30 Dpi, 33 Dpi, and 39 Dpi for Militarine; and 3 Dpi, 30 Dpi, 33 Dpi, and 39 Dpi for Coelonin for subsequent experimental analysis.

Transcriptome Sequencing Analysis of Suspension Cells of *B. striata*

By performing full-length transcriptome sequencing on suspension cells of *B. striata* at different time points, a total of 50.02 Gb of raw data were obtained, which were deposited in the SRA database under the accession number SRR18045794. A total of 902,688 CCS (Circular Consensus Sequence) readings were obtained, including 719,981 reads with 5'- primers, 785,110 reads with 3'- primers, and 781,562 reads with Poly A tail. There were 620,018 full-length (FL) reads. The number of full-length non-chimeric reads (FLNC) was 472,211, and the average length of FLNC was 2,423 bp. The FLNC sequence of the same transcript was converted using the ICE algorithm Clustering, which was performed to eliminate redundancy, and a total of 246,933 identical transcripts were obtained. Using CD-HIT to perform de-redundancy analysis on high-quality transcripts, a

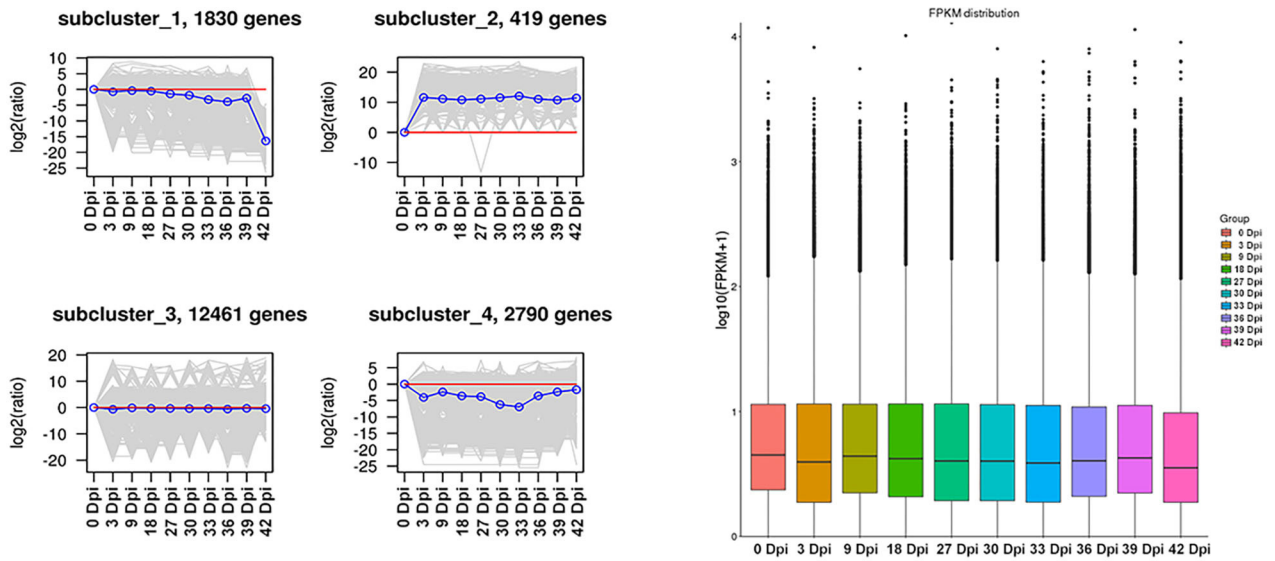


FIGURE 2 | Distribution of DEGs at different time points. **(A)** Expression trend of DEGs by K-means. **(B)** FPKM box plots of different samples.

total of 100,276 transcripts were obtained, which were used as the final transcript sequence for further analysis (Li et al., 2020).

Analysis of DEGs at Various Time Points

Distribution of DEGs

In order to explore the expression trend of DEGs during the development of suspension cells of *B. striata*, we conducted a K-means analysis on DEGs, which resulted in four gene expression trend graphs (**Figure 2A**). Although gene cluster 3 contained the largest number of genes, the changes in expression levels in each period were not obvious. The gene expression levels of gene cluster 2 were upregulated throughout the development of suspension cells of *B. striata*, whereas gene clusters 1 and 4 were reduced.

The FPKM value was used to quantify gene expression, and it can be seen from the results that there is little difference in gene expression levels among different samples (**Figure 2B**). This showed that, after FPKM quantification, the analysis results of DEGs between each sample are accurate and reliable. Subsequently, DESeq2 software was used to analyze the significant difference in the DEGs between different samples, and pairwise comparisons of suspension cells at eight time points were made. It was found that the number of DEGs gradually increased during the growth and breeding of suspension cells. A total of 2,052 DEGs were selected from the 3 Dpi vs. 30 Dpi comparison, which was significantly higher than the other comparisons. In the following comparison groups (3 Dpi vs. 30 Dpi, 18 Dpi vs. 36 Dpi, 27 Dpi vs. 33 Dpi, 36 Dpi vs. 42 Dpi, 30 Dpi vs. 33 Dpi, 39 Dpi vs. 42 Dpi), the number of upregulated genes was significantly higher than that of downregulated genes. In the other three comparison groups (18 Dpi vs. 27 Dpi, 30 Dpi vs. 39 Dpi, 33 Dpi vs. 39 Dpi), the number of upregulated genes was significantly lower than that of downregulated genes. These results indicated that DEGs were closely involved in the synthesis of phenolic acid derivatives in the suspension culture cells of *B. striata*, reflecting the obvious spatio-temporal differences in the synthesis of secondary metabolites.

We analyzed some comparisons of different time points. We found that there were 56 (18 Dpi vs. 27 Dpi vs. 33 Dpi), 96 (18 Dpi vs. 36 Dpi vs. 42 Dpi), 12 (3 Dpi vs. 30 Dpi vs. 39 Dpi vs. 42 Dpi), and 38 (3 Dpi vs. 30 Dpi vs. 33 Dpi vs. 39 Dpi) DEGs in different time groups, respectively (**Figure 3**), indicating that these DEGs may play an important role in the growth and development of *B. striata*.

The results of the Pearson correlation analysis showed that the biological duplication between each group was highly correlated. During the development of suspension cells, the correlation values (R^2) of the following groups were higher than the other comparison groups: 3 Dpi vs. 9 Dpi ($R^2 = 0.861$), 3 Dpi vs. 18 Dpi ($R^2 = 0.845$), 3 Dpi vs. 27 Dpi ($R^2 = 0.0.848$), 9 Dpi vs. 18 Dpi ($R^2 = 0.854$), 9 Dpi vs. 27 Dpi ($R^2 = 0.0.828$), 18 Dpi vs. 27 Dpi ($R^2 = 0.822$), 18 Dpi vs. 30 Dpi ($R^2 = 0.816$). In the process of growth and development, $R^2 > 0.8$ indicated that the differences in the expression of related genes between the different periods were small (**Figure 4**).

DEGs Screening Involved Biosynthesis of Phenolic Acid Derivatives in Suspension Cells

Correlation Analysis Between DEGs and Four Main Secondary Metabolites at Each Time Point

We detected correlations between the four phenolic acid derivatives and 17,500 DEGs. Among them, HBA, Dactylorhin

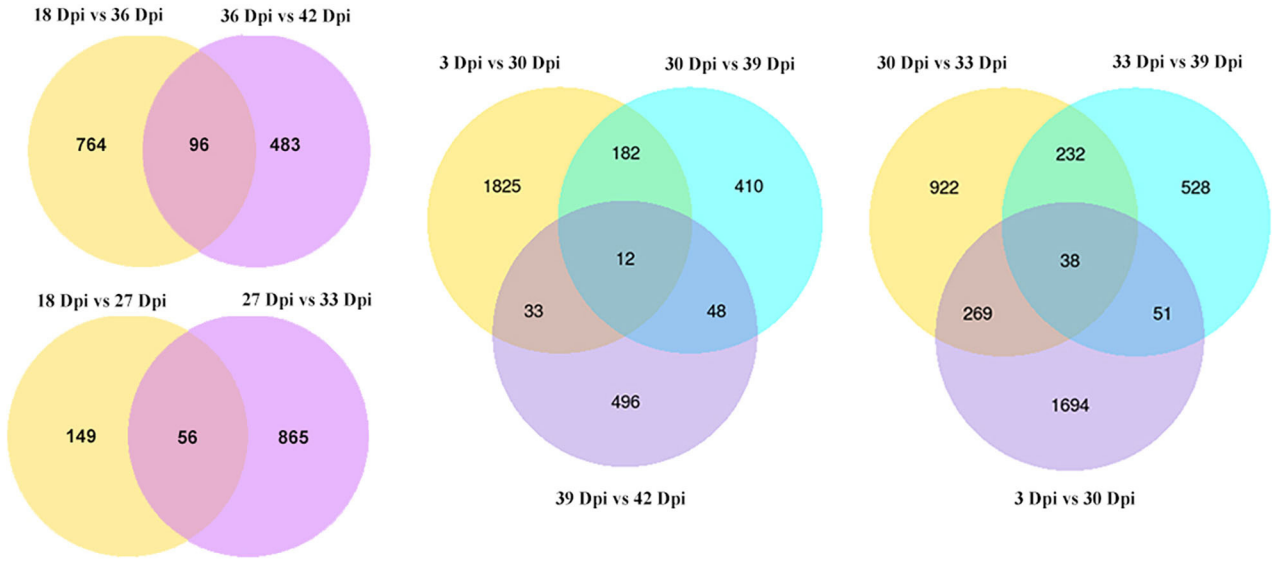


FIGURE 3 | Number of DEGs between different combinations.

A, Militarine, and Coelonin were significantly correlated with 5,134, 3,567, 1,687, and 599 DEGs, respectively ($P < 0.05$) (**Supplementary Table S1**).

DEGs Analysis of HBA

Based on the change in HBA content, we found that it increased gradually from 0 Dpi to 27 Dpi, and then the HBA content began to increase exponentially, where it stabilized at 33 Dpi. Three time points (18 Dpi, 27 Dpi, and 33 Dpi) with significant differences in HBA content were selected for differential expression analysis. At the 18 Dpi vs. 27 Dpi time point, the number of upregulated genes was 78, and the number of downregulated genes was 127. At the 27 Dpi vs. 33 Dpi time point, the number of upregulated genes was 712, while the number of downregulated genes was 209.

The functional clustering arrays of GO and KEGG on the DEGs imply a mechanism for explaining how HBA was biosynthesized. GO function analysis found that, at the 18 Dpi vs. 27 Dpi time point, DEGs were mainly enriched in the binding process, cellular process, metabolic process, and catalytic activity. At the 27 Dpi vs. 33 Dpi time point, DEGs were mainly enriched in the hydrolase activity and nucleoside phosphatase activity. A KEGG pathway analysis of DEGs at three time points found that a total of 39 DEGs were significantly enriched in four pathways (q-value < 0.05), namely Circadian rhythm-plant, Cyanoamino acid metabolism, starch and sucrose metabolism, and Phenylpropanoid biosynthesis pathways. At the 18 Dpi vs. 27 Dpi time point, there were 6 DEGs involved in the circadian rhythm of plants (ko04712). At the 27 Dpi vs. 33 Dpi time point, 30 DEGs were significantly enriched in the metabolic pathways of starch and sucrose (ko00500); 18 DEGs were significantly enriched in the cyanoamino acid metabolism pathway (ko00940); and 16 DEGs were enriched in the cyanogen amino acid metabolism pathway (ko00460) (**Supplementary Table S2**). We selected the pathways of phenylpropanoid biosynthesis and cyanoamino acid metabolism, which are closely involved in the synthesis of HBA. There were 11 upregulated genes with significant differences in the phenylpropane metabolic pathways, encoding β-glucosidase (BGLU) and peroxidase (PRX); there were seven downregulated genes encoding phenylalanine ammonia-lyase (PAL) and BGLU,. In the cyanoamino acid metabolism pathway, the up and downregulated genes with bigger differences were involved in encoding BGLU. Compared with the results of correlation analysis, 18 unigenes (8 enzymes) were finally screened out as being potentially involved in HBA synthesis (**Table 2**).

DEGs Analysis of Dactylorhin A

Three time points—18 Dpi, 36 Dpi, and 42 Dpi—were detected to have significant differences with Dactylorhin A content, and these time points were selected for further differential expression analysis. Pairwise comparisons of 18 Dpi vs. 36 Dpi and 36 Dpi vs. 42 Dpi found 860 and 579 DEGs, respectively, with a total of 96 DEGs appearing to overlap. Among them, at the 18 Dpi vs. 36 Dpi time point, the number of upregulated genes was 554, and the number of downregulated genes was 306. At the 36 Dpi vs. 42 Dpi time point, the number of upregulated genes was 323, and the number of downregulated genes was 256.

The functional clustering arrays of GO and KEGG on the DEGs imply a mechanism for explaining how Dactylorhin A was biosynthesized. GO function analysis found that, at the 18 Dpi vs. 36 Dpi time point, 602 DEGs were annotated into the GO function database, with the majority of them enriched in hydrolase activity, O-acyltransferase activity, and carbohydrate metabolic process, and cysteine-type peptidase activity. In addition, at the 36 Dpi vs. 42 Dpi time point, 363 DEGs were mainly annotated in binding carboxylic acid

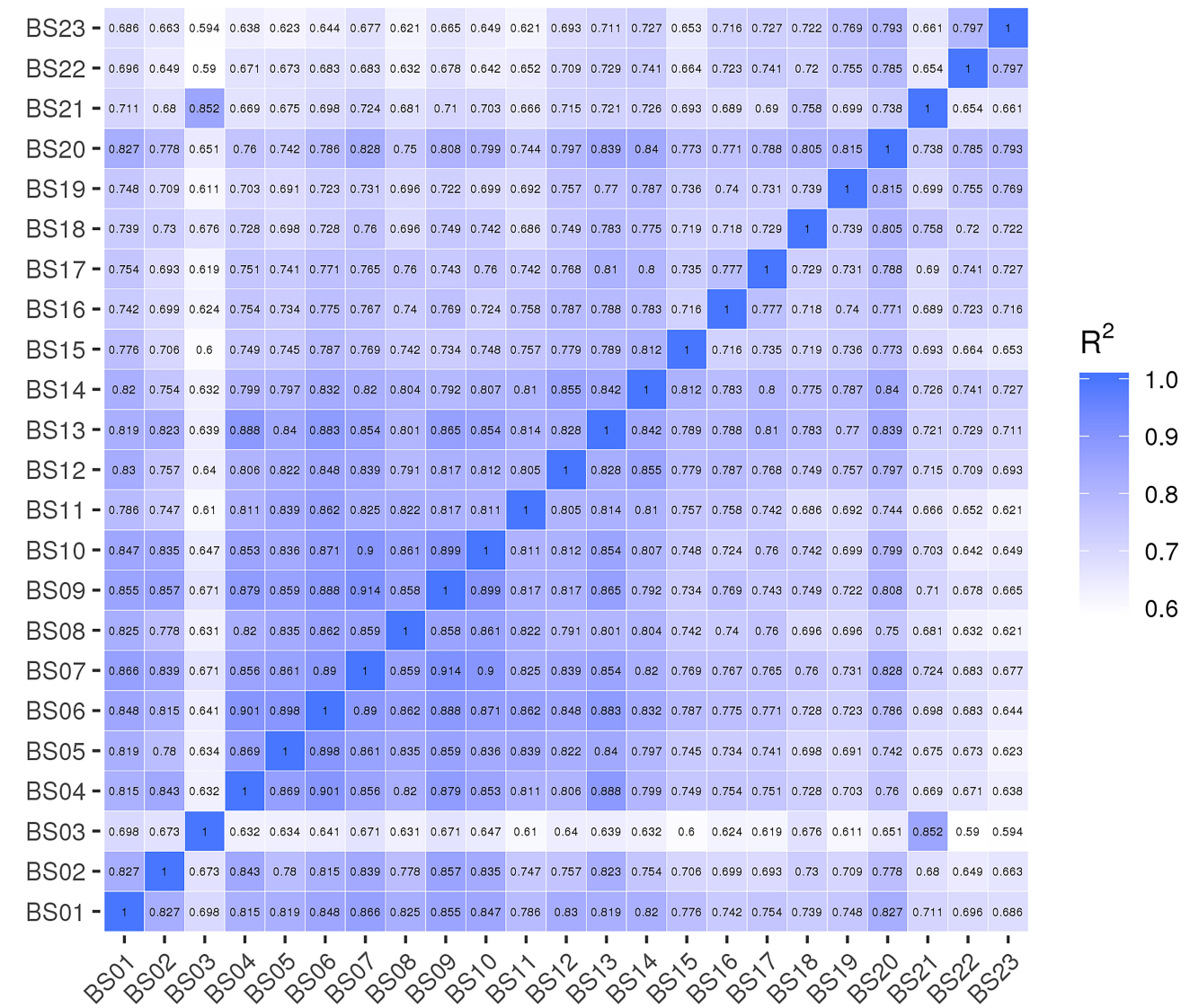


FIGURE 4 | Correlation analysis among samples.

metabolic process and oxidoreductase activity. While the KEGG pathway analysis found that, at the 18 Dpi vs. 36 Dpi time point, a total of 51 DEGs were significantly enriched in four pathways (q-value < 0.05), namely cyanoamino acid metabolism, starch and sucrose metabolism, phenylpropanoid biosynthesis, and alanine, aspartate, and glutamate metabolism pathways. In addition, the DEGs at the 36 Dpi vs. 42 Dpi were mainly annotated for the TCA cycle and starch and sucrose metabolism, among others (**Supplementary Table S2**). Selecting the metabolic pathways of starch and sucrose, which were closely related to the synthesis of Dactylorhin A, we found that there were 13 upregulated genes with significant differences, encoding UGP, BGLU, SPS, TPS, scrK, and 13 downregulated genes, encoding INV, SUS, TPS, BAM, and BGLU. Combined with the above analysis, 23 unigenes (related to 11 enzymes) were screened out as being involved in the synthesis of Dactylorhin A (**Table 2**).

DEGs Analysis of Coelonin

Based on the change in the content of Coelonin, we found that its content increased slowly from 3 Dpi to 30 Dpi and to 33 Dpi, remained relatively stable, and then increased exponentially before reaching its peak at 39 Dpi. Therefore, a pairwise comparison of the DEGs of 3 Dpi, 30 Dpi, 33 Dpi, and 39 Dpi was made (3 Dpi vs. 30 Dpi, 30 Dpi vs. 33 Dpi, 33 Dpi vs. 39 Dpi); there were 2,052, 1,461, and 849 DEGs, respectively. Among them, in the 3 Dpi vs. 30 Dpi time point, the number of

TABLE 2 | Genes involved in the biosynthesis of secondary metabolites in suspension cells of *Bletilla striata*.

Secondary metabolites	Metabolic pathway	Entry ID	Key enzymes	Gene No.
HBA	Phenylpropanoid biosynthesis	K10775	Phenylalanine ammonia-lyase (PAL)	1
		K01188	Beta-glucosidase (BGLU)	11
		K05349		
	Starch and sucrose metabolism	K01179	Endoglucanase (E3.2.1.4, EG)	1
		K01194	Alpha,alpha-trehalase (TREH)	1
		K13648	alpha-1,4-galacturonosyltransferase (GAUT)	1
		K16055	Trehalose 6-phosphate synthase/phosphatase (TPS)	1
	Circadian rhythm - plant	K16240	Protein suppressor of PHYA-105 1 (SPA1)	1
		K12120	Phytochrome A (PHYA)	1
Dactylorhin A	Starch and sucrose metabolism	K01193	Beta-fructofuranosidase (INV)	1
		K16055	Trehalose 6-phosphate synthase (TPS)	2
		K13648	Beta-amylase (E3.2.1.2, BAM)	2
		K00847	Fructokinase (E2.7.1.4, scrK)	2
		K01184	Polygalacturonase (E3.2.1.15, PG)	1
	Phenylpropanoid biosynthesis	K01904	4-coumarate-CoA ligase (4CL)	1
		K01188	Beta-glucosidase (BGLU)	6
		K00430	Peroxidase (E1.11.1.7, POD)	1
		K00083	Cinnamyl-alcohol dehydrogenase (CAD)	1
	Alanine, aspartate and glutamate metabolism	K01953	Asparagine synthase (AsnB)	5
		K00261	Glutamate dehydrogenase (GLUD1-2)	1
Coelonin	Phenylpropanoid biosynthesis	K10775	Phenylalanine ammonia-lyase (PAL)	3
		K01904	4-coumarate-CoA ligase (4CL)	1
		K00430	Peroxidase (E1.11.1.7, POD)	1
	Starch and sucrose metabolism	K00695	Sucrose synthase (SUS)	1
		K13648	Alpha-1,4-galacturonosyltransferase (GAUT)	1
	Glycolysis/ Gluconeogenesis	K13951	Alcohol dehydrogenase 1/7 (ADH1/7)	3
		K01568	Pyruvate decarboxylase (PDC)	2
		K00850	6-phosphofructokinase 1 (PFK)	1
		K01623	Fructose-bisphosphate aldolase (ALDO)	1
Militarine	Phenylpropanoid biosynthesis	K10775	Phenylalanine ammonia-lyase (PAL)	5
		K00487	Trans-cinnamate 4-monooxygenase (CYP73A)	1
		K01904	4-coumarate-CoA ligase (4CL)	2
		K05349	Beta-glucosidase (BGLU)	2
	Starch and sucrose metabolism	K01087	Trehalose 6-phosphate phosphatase (otsB)	3
		K16055	Trehalose 6-phosphate synthase (TPS)	2
		K01193	Beta-fructofuranosidase (INV)	1
		K00847	Fructokinase (E2.7.1.4, scrK)	1
	Glycolysis/ Gluconeogenesis	K00850	6-phosphofructokinase 1 (PFK)	1
		K00134	Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	2
		K08679	UDP-glucuronate 4-epimerase (GAE)	1
		K01689	Enolase (ENO)	1
		K01623	Fructose-bisphosphate aldolase (ALDO)	3
		K00873	Pyruvate kinase (PK)	2
		K00627	Dihydrolipoamide acetyltransferase (DLAT)	1
		K01568	Pyruvate decarboxylase (PDC)	6
		K18857	Alcohol dehydrogenase 1 (ADH1)	5
		K13951	Alcohol dehydrogenase 1/7 (ADH1/7)	2

upregulated genes was 1,162, and the number of downregulated genes was 890. At the 30 Dpi vs. 33 Dpi time point, the number of upregulated genes was 902, and the number of downregulated genes was 559. At the 33 Dpi vs. 39 Dpi time point, the number of upregulated genes was 316, and the number of downregulated genes was 533.

The functional clustering arrays of GO and KEGG on the DEGs imply a mechanism for explaining how Coelonin was biosynthesized. Through the comparative analysis of the GO database, the unigenes annotated by the GO can be divided into three categories: biological process, molecular function, and cellular component. At the 3 Dpi vs. 30 Dpi time point, DEGs were mainly involved in the metabolic process, cellular process, and organic substance metabolic process in biological processes; hydrolase activity, catalytic activity, and binding in molecular function. In the suspension cells of the 30 Dpi vs. 33 Dpi and 33 Dpi vs. 39 Dpi, DEGs were significantly enriched in the plant-type cell wall organization, plant-type cell wall organization, cell wall organization, and external encapsulating structure organization in biological processes; significantly enriched molecular functions were the structural constituents of the cell wall, such as structural molecule activity and others. The cell component that was significantly enriched was the histone deacetylase complex. Among them, the unigenes that are more abundant in biological processes may be related to the pathway of Coelonin synthesis concerned in this study.

KEGG enrichment results showed that, in comparison with 3 Dpi, 53 DEGs were involved in the metabolic pathways of starch and sucrose metabolism (ko00500) in 30 Dpi; 40 DEGs were enriched in the phenylpropanoid biosynthesis pathway (ko00940). At the 30 Dpi vs. 33 Dpi time point, there were 40 DEGs enriched in starch and sucrose metabolic pathways (ko00500). At the 33 Dpi vs. 39 Dpi time point, DEGs were enriched in the cyanoamino acid metabolism pathway (ko00460) (**Supplementary Table S2**).

We selected the phenylpropanoid biosynthesis, phenylalanine metabolism, and cyanoamino acid metabolism pathways, which were closely related to the synthesis of Coelonin. And according to the correlation analysis between gene expression and Coelonin content, the differential genes related to Coelonin metabolism were found. The results showed that the key regulatory genes mainly encoded PAL, 4CL, POD, SUS, GAUT, ADH, PDC, PFK, and ALDO. Finally, 14 unigenes (9 enzymes) were identified as potential Coelonin synthesis candidates (**Table 2**).

DEGs Analysis of Militarine

According to the change in the content of Militarine, it was found that from 3 Dpi to 30 Dpi, its content increased slowly and then increased exponentially before reaching its peak at 39 Dpi and then dropping sharply. Pairwise comparison of the DEGs of 3 Dpi, 30 Dpi, 39 Dpi, and 42 Dpi (3 Dpi vs. 30 Dpi, 30 Dpi vs. 39 Dpi, 39 Dpi vs. 42 Dpi) showed that there were 2052, 652, and 589 DEGs, respectively, and 1825, 496, and 410 DEGs appeared in only one DEG set, respectively; 33, 48, and 182 DEGs appeared in two DEG sets, respectively, and a total of 12 DEGs appeared in 3 DEG sets. Among them, at the 30 Dpi vs. 39 Dpi time point, the number of upregulated genes was 296, and the number of downregulated genes was 356; at the 39 Dpi vs. 42 Dpi time point, the number of upregulated genes was 387, and the number of downregulated genes was 202.

According to GO analysis, 1,344 DEGs were annotated into the GO database at 3 Dpi vs. 30 Dpi time point, with the majority of them enriched in hydrolase activity, asparagine synthase (glutamine-hydrolyzing) activity, catalytic activity, single-organism metabolic process, and carbohydrate metabolic process, among others. At the 30 Dpi vs. 39 Dpi time point, 476 DEGs were significantly enriched in cysteine-type peptidase activity, covalent chromatin modification, and histone modification. The KEGG pathway analysis of DEGs at three time points found that a total of 136 DEGs were significantly enriched in five pathways at 3 Dpi vs. 30 Dpi time point, namely starch and sucrose metabolism, phenylpropanoid biosynthesis, glycolysis/gluconeogenesis, photosynthesis, and phenylalanine metabolism pathways; at the 30 Dpi vs. 39 Dpi time point, a total of 30 DEGs were significantly enriched in Spliceosome; at the 39 Dpi vs. 42 Dpi time point, DEGs were mainly enriched in plant hormone signal transduction and phenylpropanoid biosynthesis (**Supplementary Table S2**). Based on the above analysis, 41 unigenes (18 key enzymes) of the main regulatory genes were screened out, which may be involved in the synthesis of Militarine, as shown in **Table 2**.

Expression Analysis of Related Genes Regulating the Synthesis of Phenolic Acid Derivatives

In order to verify the accuracy of gene expression in the transcriptome sequencing results, 10 DEGs were randomly selected from samples of four periods and verified by qPCR, in which the *β-actin* was applied as the internal reference gene. The results showed that the relative gene expression levels were consistent with the transcriptome sequencing results, indicating that the data obtained by transcriptome sequencing were accurate and reliable (**Figure 5**).

DISCUSSION

The secondary metabolites were produced by plants during their growth and development, as well as to defend against environmental stresses. In addition, the synthesis and accumulation of these metabolites can be significantly improved through suspension cell culture (Yang et al., 2018; Isah, 2019; Jiang et al., 2020). The secondary metabolites in suspension cells provide the material basis for their growth and reproduction. We analyzed the DEGs in different growth stages and found that many genes involved in the biosynthesis of secondary metabolites expressed significant differences in different growth stages. Moreover, the content of secondary metabolites also changed significantly, indicating that, with the change in suspension cell morphology and continuous growth and development, the synthesis of secondary metabolites of suspension cells is regulated by a variety of genes.

As time goes on, cultured suspension cells synthesize and accumulate secondary metabolites through different metabolic processes, thereby improving their ability to synthesize and accumulate HBA, Dactylorhin A, Militarine, and Coelonin in suspension cells of *B. striata* (Pan, 2019). This study found that the contents of HBA at 18 Dpi, 27 Dpi, and 33 Dpi, Dactylorhin A at 18 Dpi, 36 Dpi, and 42 Dpi, Coelonin at 3 Dpi, 30 Dpi, 33 Dpi, and 39 Dpi, and Militarine at 3 Dpi, 30 Dpi 39 Dpi,

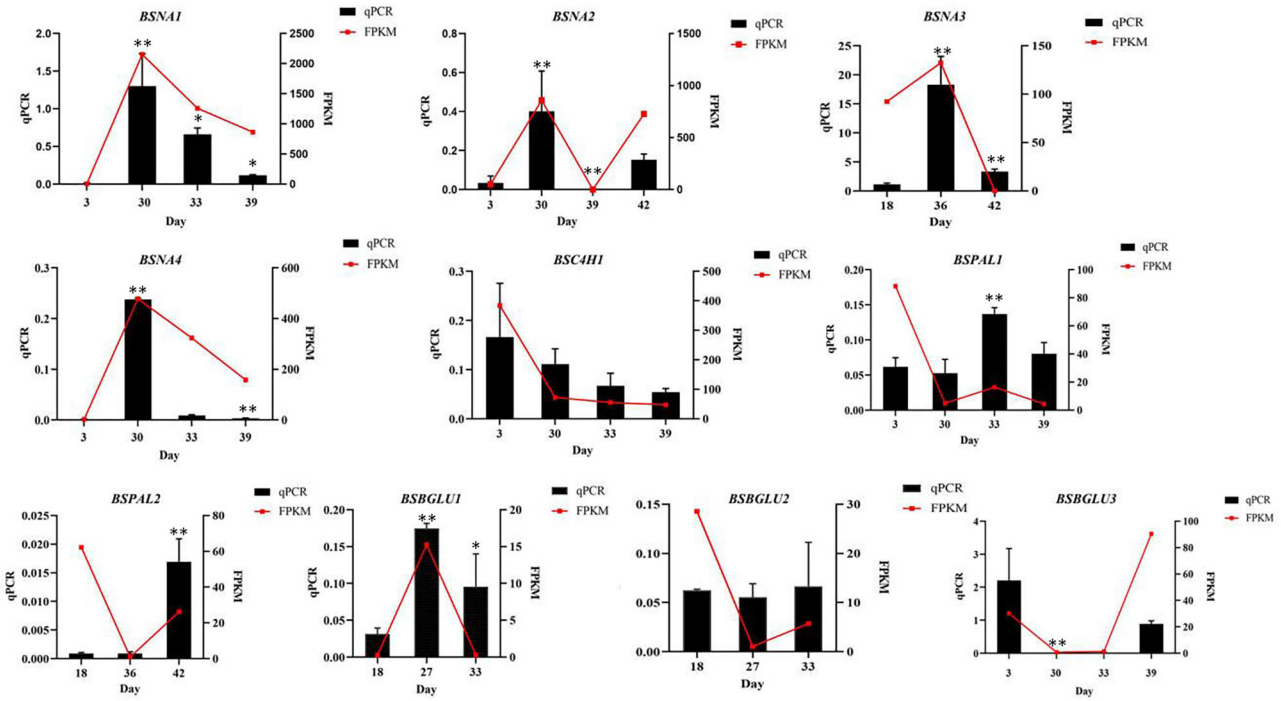


FIGURE 5 | Comparison of expression of 10 genes obtained by qPCR analysis and by RNA-seq (FPKM values). * Represents the significant difference between the two groups compared with the previous group at a level of $P < 0.05$. ** represents the significant difference at a level of $P < 0.01$.

and 42 Dpi were significantly different by HPLC, respectively. It was found that, compared with the high-efficiency culture lag zone, the content of the four metabolites in the exponential phase increased significantly; the content of Dactylorhin A, Militarine, and Coelonin decreased significantly in the declining phase, while HBA slowly increased. According to the chemical structures of these four compounds (**Supplementary Figure 1**), the main structures of Dactylorhin A and Militarine were the same, which implies that they may have connections in the biosynthesis process. The increase in Militarine content resulted in a reduction of Dactylorhin A synthesis. HBA was simple in structure, which could be an integral part of forming Militarine and Coelonin. Therefore, HBA contributes to the increase of Militarine and Coelonin.

Genes involved in various metabolic processes of cultured suspension cells are critical in the process of the synthesis of secondary metabolites. The metabolism of phenolic acid compounds was a relatively complicated process. The shikimic acid pathway, the phenylpropanoid metabolic pathway, and the flavonoid metabolic pathway were the main pathways for the synthesis of phenolic acid compounds (Saxe et al., 2021). Among them, phenylalanine and tyrosine are direct precursors of phenolic acid biosynthesis, which are converted from glucose into a multistep metabolism via the shikimic acid pathway and then into the phenylpropanoid metabolic pathway (Feduraev et al., 2020; El-Shora et al., 2021). Additionally, phosphoenolpyruvate and erythrose-4-phosphoric acid were the main precursor substance for the synthesis of phenolic acid compounds (Deng et al., 2020). Phenylalanine was deaminated and transformed into trans-cinnamic acid by phenylalanine ammonia-lyase (PAL) (Trivedi et al., 2022). Subsequently, phenolic acids such as trans-cinnamic acid, p-coumarate, and ferulic acid are formed under the action of cinnamate 4-hydroxylase (C4H), 4-coumaratecoa-ligase (4CL), O-methyltransferase (OMT), and other enzymes catalyzed further into caffeic acid and chlorogenic acid (Wohl and Petersen, 2020). HBA was one of the main components of phenolic compounds and the main precursor of aromatic compounds (Zhang Z. L. et al., 2020). Former studies have confirmed that Dactylorhin A and Militarine were glycoside compounds isolated from plants, which were produced by the loss of water from the hydroxyl amino thiol group of monosaccharides or oligosaccharides and the hemiacetal hydroxyl group of another molecule. With the growth and development of suspension-cultured cells, the expression levels of genes involved in the metabolism of starch and sucrose and the biosynthetic pathway of phenylpropanoid differed significantly. The differences in the expression of these genes may be the main reason for the significant differences in the content of glycosides in different periods. BGLU is a group of heterogeneous hydrolases that have the ability to hydrolyze and synthesize glycosidic bonds and have great application value in various fields (Santos et al., 2018; Sun et al., 2020). The enhanced expression of genes involved in BGLU may promote the synthesis of carbohydrates and provide energy for the growth and development of suspension-cultured cells. At the same time, BGLU was regulated by a large

number of genes at different time points and was involved in the synthesis of phenolic acids, which indicated that BGLU may be involved in the synthesis of phenolic acids and other substances and maintaining cell morphology during the development of *B. striata*. Coelonin was a dihydrophenanthrene compound. The methoxy and hydroxyl groups were the main substituents on the benzene ring of the dihydrophenanthrene compound. From the point of view of molecular structure, phenolic acid and its derivatives can be obtained after the oxidation of dihydrophenanthrene compounds.

In this study, we detected several interesting co-changes between gene expression and chemical content, which may infer the genes' function. The expression of i1_LQ_BSbp1_C76772/f1p0 increased steadily from 0 to 8 Dpi but decreased dramatically later, which was positively correlated with the content changes of Dactylorhin A. The expression of i3_LQ_BSbp1_C942 /f1p19 was continuously downregulated from 0 to 36 Dpi but upregulated from 36 Dpi to 39 Dpi and then downregulated again, which was positively correlated with the content change of Dactylorhin A. The expression of i3_LQ_BSbp1_C5868/f1p2 was continuously downregulated at the stage of 0–18 Dpi but then upregulated, which showed a positively significant correlation with the content of Militarine. Based on these results, we predict that the TPS plays a very important role in plant metabolic regulation. In addition, we also speculated that these genes are closely related to the biosynthesis of Dactylorhin A and Militarine. Similar to this analysis and speculation, we built the relationship between genes i2_HQ_BSbp1_C5228/f3p1 and Dactylorhin A and gene i2_LQ_BSbp1_C117598/f1p12 and Militarine. In addition, the INV may be key to the synthesis of Dactylorhin A and Militarine. Therefore, during the growth of *B. striata* suspension culture cells, with the passage of time, the expression of key enzyme genes regulating the synthetic biological pathway is different in different periods, and through interaction, the content of phenolic acid derivatives is different in different periods.

In different growth periods, the metabolic pathways involved in the synthesis of these four secondary metabolites of DEGs mainly include phenylpropanoid biosynthesis, phenylalanine metabolism, cyanoamino acid metabolism, and starch and sucrose metabolism. The results of the enrichment of metabolic pathways indicated that carbohydrate metabolism was needed as a basis during development, and hormones and some secondary metabolites were required to co-regulate. According to the KEGG pathway, GO enrichment analysis, and correlation analysis of DEGs, the final screening of enzymes involved in the synthesis of HBA, Dactylorhin A, Militarine, and Coelonin were 8, 11, 18, and 9, respectively, and the related unigenes were 18, 23, 41, and 14, which led to the speculation that these genes which were screened out may be the key regulators of phenolic acid derivatives. Therefore, the differential expression of key enzyme genes in the biosynthetic pathway at different periods may ultimately lead to differences in the content of phenolic acid derivatives at different periods.

At present, four phenolic acid derivatives of secondary metabolites have been isolated and identified, but the metabolic mechanism of their synthesis and accumulation has not been studied. It was found that the content of Dactylorhin A, Militarine, and other compounds in *B. striata* was higher than the cultivated ones (Wang et al., 2014). In this study, gene expression profiles of suspension cells at different growth stages were obtained through the transcriptome method, and a total of 17,500 DEGs were screened. The preliminary screening showed that the genes of PAL, 4CL, BGLU, INV, TPS, BAM, and PDC may regulate the key enzyme of phenolic acid derivatives. Through the gene identification and regulatory mechanism exploring of four main secondary metabolites, the key regulated genes in this study could be overexpressed or silenced in the future to further verify their functions in the biosynthesis of four phenolic acid derivatives in *B. striata*. Thus, the metabolic mechanism of synthesis and accumulation of the four secondary metabolites will be thoroughly explained. This provided data for an in-depth study of the synthesis and regulation mechanisms of phenolic acid derivatives in suspension cells of *B. striata*.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: National Center for Biotechnology Information (NCBI) BioProject database under accession numbers PRJNA807395 and PRJNA807396.

AUTHOR CONTRIBUTIONS

DX conceived, supervised, wrote, and reviewed the manuscript. HL, CH, WW, and LL originally wrote and reviewed the draft. HL, CH, QL, MW, SX, JS, and YH conducted the experiments and carried out the analysis. DX founded and administrated the project. All authors read and approved the final version.

FUNDING

This research was financially supported by the National Natural Science Foundation of China (31560079, 31960074), the Science and Technology Department Foundation of Guizhou Province (No. [2017]5733-050, [2019]-027, [2019]5657), the Special Joint Bidding Project of Zunyi Sci & Tech Bureau and Zunyi Medical University (ZSKHHZ-2020-91), the Honghuagang Sci & Tech Project (ZHKHNZT [2020]04), and the Program for Excellent Young Talents of Zunyi Medical University (15zy-004).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.875404/full#supplementary-material

Supplementary Figure 1 | Structure of four phenolic acid derivatives.

Supplementary Table S1 | Correlation between different phenolic acid derivatives.

Supplementary Table S2 | Annotation of DEGs.

Supplementary Table S3 | Main experimental reagent.

Supplementary Table S4 | Parameter values for HPLC assay.

Supplementary Table S5 | cDNA synthesis.

Supplementary Table S6 | qPCR reaction system.

REFERENCES

Anantharaju, P. G., Gowda, P. C., Vimalambike, M. G., and Madhunapantula, S. V. (2016). An overview on the role of dietary phenolics for the treatment of cancers. *Nutr. J.* 15, 99–104. doi: 10.1186/s12937-016-0217-2

Boudjada, A., Touil, A., Bensouici, C., Bendif, H., and Rhouati, S. (2019). Dioscorea communis phenanthrene and dihydrophenanthrene derivatives from with anticholinesterase, and antioxidant activities. *Nat. Product Res.* 33, 3278–3282. doi: 10.1080/14786419.2018.1468328

Deng, N., Zheng, B., Li, T., and Liu, R. H. (2020). Assessment of the phenolic profiles, hypoglycemic activity, and molecular mechanism of different highland barley (Hordeum vulgare L.) varieties. *Int. J. Molec. Sci.* 21 (4):1175–1194. doi: 10.3390/ijms21041175

Ding, Y. F., Bao, X. M., Lao, L. F., Ling, Y. X., Wang, Q. W., and Xu, S. J. (2019). p-Hydroxybenzyl Alcohol prevents memory deficits by increasing neurotrophic factors and decreasing inflammatory factors in a mice model of alzheimer's disease. *J. Alzheimer's Dis.* 67, 1007–1019. doi: 10.3233/JAD-180910

El-Shora, H. M., Massoud, G. F., and Gad, D. (2021). Activation of metabolic pathways associated with phenolic biosynthesis in garden cress leaves under lead stress. *Res. J. Pharmac. Biol. Chem. Sci.* 11, 128–140. doi: 10.33887/rjpbcs/2020.11.6.16

Feduraev, P., Skrypnik, L., Riabova, A., Pungin, A., Tokupova, E., Maslennikov, P., et al. (2020). Phenylalanine and tyrosine as exogenous precursors of wheat (Triticum aestivum L.) secondary metabolism through PAL-associated pathways. *Plants* 9, 476–494. doi: 10.3390/plants9040476

Fu, L. M., Niu, B. F., and Zhu, Z. W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28, 3150. doi: 10.1093/bioinformatics/bts565

Isah, T. (2019). Stress and defense responses in plant secondary metabolites production. *Biol. Res.* 52, 39. doi: 10.1186/s40659-019-0246-3

Jiang, D. Q., Wang, H. Y., Kang, C. Z., Jiang, J. Y., Du, Y. X., Zhang, Y., et al. (2020). Influence and mechanism of stress combination on medicinal plants secondary metabolism. *China J. Chin. Mater. Medica* 45, 2009–2016. doi: 10.19540/j.cnki.cjcmm.20200302.106

Jiang, F. S., Li, M. Y., Wang, H. Y., Ding, B., Zhang, C. C., Ding, Z. S., et al. (2019). Coelonin, an anti-inflammation active component of Bletilla striata and its potential mechanism. *Int. J. Molec. Sci.* 20, 4422–4439. doi: 10.3390/ijms20184422

Kassim, M., Achoui, M., Mustafa, M. R., Mohd, M. A., and Yusoff, K. M. (2010). Ellagic acid, phenolic acids, and flavonoids in Malaysian honey extracts demonstrate in vitro anti-inflammatory activity. *Nutr. Res.* 30, 650–659. doi: 10.1016/j.nutres.2010.08.008

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* 12, 323. doi: 10.1186/1471-2105-12-323

Li, L., Liu, H. B., Wen, W. E., Huang, C. Y., Li, X. M., Xiao, S. J., et al. (2020). Full transcriptome analysis of callus suspension culture system of bletilla striata. *Front. Genet.* 11, 995–1006. doi: 10.3389/fgene.2020.00995

Luo, L., Kim, S. W., Lee, H. K., Kim, I. D., Lee, H., and Lee, J. K. (2017). Anti-oxidative effects of 4-hydroxybenzyl alcohol in astrocytes confer protective effects in autocrine and paracrine manners. *PLoS ONE.* 12, e0177322. doi: 10.1371/journal.pone.0177322

Marchiosi, R., Santos, W. D. D., Constantin, R. P., Lima, R. B. D., Soares, A. R., Finger-Teixeira, A., et al. (2020). Biosynthesis and metabolic actions of simple phenolic acids in plants. *Phytochem. Rev.* 19, 865–906. doi: 10.1007/s11101-020-09689-2

Milovanović, V., Petrović, Z., Novaković, S., Bogdanović, G. A., Simijonović, D., Mladenović, M., et al. (2021). Pyrazole derivatives of medically relevant phenolic acids: insight into antioxidative and anti-LOX activity. *Med. Chem.* 17, 807–819. doi: 10.2174/1573406416666200602152643

Nigro, D., Laddomada, B., Mita, G., Blanco, E., Colasuonnod, P., Simeone, R., et al. (2017). Genome-wide association mapping of phenolic acids in tetraploid wheats. *J. Cereal Sci.* 75, 25–34. doi: 10.1016/j.jcs.2017.01.022

Pan, Y. C. (2019). *Establishment of Bletilla Striata Liquid Suspension Culture System and Determination of Secondary Metabolites.* Zunyi: Zunyi Medical University. p. 1–118.

Salmela, L., and Rivals, E. (2014). LoRDEC: accurate and efficient long read error correction. *Bioinformatics* 30, 3506–3514. doi: 10.1093/bioinformatics/btu538

Santos, S. J. C., Parras, M. L., Sibeli, C., and John, W. R. (2018). Glucose tolerant and glucose stimulated β-glucosidases - A review. *Bioresour. Technol.* 267, 704–713. doi: 10.1016/j.biortech.2018.07.137

Sarakulwattana, C., Mekboonsonglarp, W., Likhitwitayawuid, K., Rojsitthisak, P., and Sritularak, B. (2020). New bisbibenzyl and phenanthrene derivatives from Dendrobium scabrilingue and their α-glucosidase inhibitory activity. *Nat. Prod. Res.* 34, 1694–1701. doi: 10.1080/14786419.2018.1527839

Saxe, H. J., Horibe, T., Balan, B., Butterfield, T. S., Feinberg, N. G., Zabaneh, C. M., et al. (2021). Two UGT84A family glycosyltransferases regulate phenol, flavonoid, and tannin metabolism in juglans regia (English Walnut). *Front. Plant Sci.* 12, 626483. doi: 10.3389/fpls.2021.626483

Sun, J., Ma, Y. S., Xu, X. R., Liu, Z. C., and Zou, L. (2020). Molecular cloning and bioinformatics analyses of a GH3 beta-glucosidase from Auricularia heimuer. *Biotechnol. Biotechnol. Equip.* 34, 850–859. doi: 10.1080/13102818.2020.1807407

Trivedi, V. D., Chappell, T. C., Krishna, N. B., Shetty, A., Sigamani, G. G., Mohan, K., et al. (2022). In-depth sequence-function characterization reveals multiple paths to enhance phenylalanine ammonia-lyase (PAL) activity. *ACS Catal.* 12, 2381–2396. doi: 10.1021/acscatal.1c05508

Wang, A. M., Yan, Y., Lan, B., Liao, S. G., Wang, Y. L., Li, Y. J., et al. (2014). Simultaneous determination of nine chemical markers of bletillae rhizoma by ultra performance liquid chromatography. *China J. Chin. Mater. Medica.* 39, 2051–2055. doi: 10.4268/cjcmm20141121

Wohl, J., and Petersen, M. (2020). Phenolic metabolism in the hornwort Anthoceros agrestis: 4-coumarate CoA ligase and 4-hydroxybenzoate CoA ligase. *Plant Cell Rep.* 39, 1129–1141. doi: 10.1007/s00299-020-02552-w

Wu, Y. P., Liu, W. J., Zhong, W. J., Chen, Y. J., Chen, D. N., He, F., et al. (2017). Phenolic compounds from the stems of Flickingeria fimbriata. *Nat. Prod. Res.* 31, 1518–1522. doi: 10.1080/14786419.2017.1278599

Yang, L., Wen, K. S., Ruan, X., Zhao, Y. X., Wei, F., and Wang, Q. (2018). Response of plant secondary metabolites to environmental factors. *Molecules* 23, 762. doi: 10.3390/molecules23040762

Yuan, Y., Xiang, J., Zheng, B., Sun, J. J., Luo, D. L., Li, P. Y., et al. (2021). Diversity of phenolics including hydroxycinnamic acid amide derivatives, phenolic acids contribute to antioxidant properties of Proso millet. *Food Sci. Technol.* 154, 112611. doi: 10.1016/j.lwt.2021.112611

Zhang, D., Zhu, H., Bian, X., Zhao, Y., Zang, P., Gao, Y. G., et al. (2020). The antidepressant effect of 4-hydroxybenzyl alcohol 2-naphthoate through monoaminergic, GABAergic system and BDNF signaling pathway. *Nat. Product Res.* 34, 2328–2331. doi: 10.1080/14786419.2018.1531862

Zhang, Y. H., Kou, Y. G., Zheng, T., and Chen, L. (2022). Research Progress on hemostatic mechanism and clinical application of Bletilla striata. *China Pharm.* 25, 321–323. doi: 10.19962/j.cnki.issn1008-049X.2022.02.023

Zhang, Z. L., Gao, Y. G., Zang, P., Gu, P. P., Zhao, Y., He, Z. M., et al. (2020). Research progress on mechanism of gastrodin and p-hydroxybenzyl alcohol on central nervous system. *China J. Chin. Mater. Medica.* 45, 312–320. doi: 10.19540/j.cnki.cjcmm.20190730.401

Zhao, F. F., Li, J., Xu, G. B., Wang, Y. L., Xi, X. L., and Liao, S. G. (2021). Study on the hemostatic active compounds from the non-polysaccharide fraction of Bletilla striata. *Nat. Product Res. Develop.* 33, 426–432.

Zhou, R. Z., Wang, Y. H., Cui, Z. Y., Li, S. C., Wang, Y. B., Zheng, G. W., et al. (2021). Genomic sequencing analysis of Panax notoginseng based on PacBio's third-generation sequencing technology. *Lishizhen Med. Mater. Medica Res.* 32, 719–723. doi: 10.3969/j.issn.1008-0805.2021.03.60

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 24 June 2022
doi: 10.3389/fpls.2022.882189

Comparative Analysis of Chloroplast Genome and New Insights Into Phylogenetic Relationships of *Polygonatum* and Tribe Polygonateae

Jing Wang[1,2], Jun Qian[1], Yuan Jiang[1], Xiaochen Chen[2], Baojiang Zheng[2], Shilin Chen[3], Fajian Yang[4], Zhichao Xu[1,2] and Baozhong Duan[1]**

[1] College of Pharmaceutical Science, Dali University, Dali, China, [2] Heilongjiang Key Laboratory of Plant Bioactive Substance Biosynthesis and Utilization, College of Life Science, Northeast Forestry University, Harbin, China, [3] Key Laboratory of Beijing for Identification and Safety Evaluation of Chinese Medicine, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China, [4] Baoshan College of Traditional Chinese Medicine, Baoshan, China

Members of *Polygonatum* are perennial herbs that have been widely used in traditional Chinese medicine to invigorate Qi, moisten the lung, and benefit the kidney and spleen among patients. However, the phylogenetic relationships and intrageneric taxonomy within *Polygonatum* have long been controversial because of the complexity of their morphological variations and lack of high-resolution molecular markers. The chloroplast (cp) genome is an optimal model for deciphering phylogenetic relationships in related families. In the present study, the complete cp genome of 26 species of Trib. Polygonateae were *de novo* assembled and characterized; all species exhibited a conserved quadripartite structure, that is, two inverted repeats (IR) containing most of the ribosomal RNA genes, and two unique regions, large single sequence (LSC) and small single sequence (SSC). A total of 8 highly variable regions (*rps16-trnQ-UUG, trnS-GCU-trnG-UCC*, *rpl32-trnL-UAG*, *matK-rps16*, *petA-psbJ, trnT-UGU-trnL-UAA*, *accD-psaI,* and *trnC-GCA-petN*) that might be useful as potential molecular markers for identifying *Polygonatum* species were identified. The molecular clock analysis results showed that the divergence time of *Polygonatum* might occur at ∼14.71 Ma, and the verticillate leaf might be the ancestral state of this genus. Moreover, phylogenetic analysis based on 88 cp genomes strongly supported the monophyly of *Polygonatum*. The phylogenetic analysis also suggested that *Heteropolygonatum* may be the sister group of the *Polygonatum*, but the *Disporopsis, Maianthemum,* and *Disporum* may have diverged earlier. This study provides valuable information for further species identification, evolution, and phylogenetic research of *Polygonatum*.

Keywords: ***Polygonatum*, chloroplast genome, phylogenetics, divergence time, Trib. Polygonateae**

OPEN ACCESS

Edited by:
Da-Cheng Hao,
Dalian Jiaotong University, China

Reviewed by:
Aaron Floden,
Missouri Botanical Garden, United States
Bingyue Lu,
Honghe University, China

****Correspondence:***
Zhichao Xu
zcxu@nefu.edu.cn
Baozhong Duan
bzduan@126.com

Specialty section:
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

Received: *23 February 2022*
Accepted: *03 June 2022*
Published: *24 June 2022*

Citation:
Wang J, Qian J, Jiang Y, Chen X, Zheng B, Chen S, Yang F, Xu Z and Duan B (2022) Comparative Analysis of Chloroplast Genome and New Insights Into Phylogenetic Relationships of Polygonatum and Tribe Polygonateae. Front. Plant Sci. 13:882189. doi: 10.3389/fpls.2022.882189

INTRODUCTION

Polygonatum Mill (1754) is an essential medicinal and edible species widely distributed in warm–temperate zones of the Northern Hemisphere and Northeastern Asia (Floden, 2015). There are approximately 70 species recognized worldwide (Floden and Schilling, 2018), with 39 present in China, 20 of which are endemic to the region (Chen and Tamura, 2000). The underground rhizomes of *Polygonatum* have crucial medicinal value in moistening lungs, relieving thirst, replenishing the spleen, and increasing immunity (Jiao et al., 2018a). Among them, four species

[*Polygonatum odoratum* (Mill.) Druce, *Polygonatum sibiricum* Red., *Polygonatum cyrtonema* Hua, and *Polygonatum kingianum* Coll. et Hemsl] were listed in the Chinese Pharmacopoeia (Chinese Pharmacopoeia Commission, 2020). Modern studies have demonstrated that some *Polygonatum* species were rich in nutrients and functional components and were regarded as a new enormous potential miscellaneous grain (Si and Zhu, 2021). The previous survey has revealed that *Polygonati rhizoma* is often contaminated with several common adulterants in herbal markets, such as *Polygonatum cirrhifolium*, *Polygonatum humile*, *Polygonatum stenophyllum*, *Polygonatum filipes*, and *Polygonatum verticillatum* (Yang et al., 2015; Jiao et al., 2018b; Wang Y. et al., 2019; Wang Z. W. et al., 2019). Because the morphology of these species is similar, changeable, and indistinguishable, it seriously affects the safety and effectiveness of clinical drug use (Ali et al., 2021).

In addition, the phylogenetic position of *Polygonatum* has been controversial for many years. Some previous taxonomy places the genus within Convallariaceae, Ruscaceae, and Asparagaceae based on morphological and molecular phylogenies (Tang, 1978; Rudall et al., 2000; Kim et al., 2012). Concerning the intraspecific relationship of *Polygonatum*, according to the leaf order, Baker and Esq (1875) divided it into three groups: section (sect.) *Alternifolia* (=sect. *Polygonatum*), sect. *Verticillata,* and sect. *Oppositifolia*, but Tang (1978) considered that the classification mentioned above might be inappropriate for identification. He divided the genus into eight series (ser.): including ser. *Alternifolia*, ser. *Altelobata*, ser. *Bracteata*, ser. *Punctata*, ser. *Kingiana*, ser. *Hookeriana*, ser. *Verticillata*, and ser. *Oppositifolia*. Tamura (1993) considered that *Polygonatum* could be divided into sect. *Polygonatum* and sect. *Verticillata* according to stamen morphology, chromosome number, karyotype, and filament. Based on the *rpl16* gene and *trnK* gene, Wu et al. (2000) found that the opposite leaves and verticillate leaves of *Polygonatum* were polyphyletic. More recent molecular phylogenies based on one to several genes have suggested that the *Polygonatum* could be divided into three groups (Meng et al., 2014; Floden and Schilling, 2018), recommended that the *petA-psbJ* plastid gene region is combined with the nuclear ribosomal *ITS* for *Polygonatum* identification. However, the phylogenetic position of some species [i.e., *P. franchetii* Hua, *P. alternicirrhosum* Hand.-Mzt., *P. verticillatum* (L.) All., *P. punctatum* Royle ex Kunth] is in dispute (Meng et al., 2014; Floden, 2017; Floden and Schilling, 2018; Jiao, 2018; Zhao et al., 2019; Xia et al., 2022). Moreover, the current taxonomy of Tribe (Trib.) Polygonateae has long been debated, which has been divided into three to eight genera. For example, the genus was divided into eight genera: *Polygonatum, Disporopsis, Smilacina, Maianthemum, Disporum, Clintonia*, and *Streptopus* (Tang, 1978). Whereas Tamura et al. (1997) proposed three genera: *Polygonatum, Disporopsis*, and *Heteropolygonatum*. Thus, the boundaries and relationships of Trib. Polygonateae remain problematic.

Despite these potential issues, using the chloroplast (cp) genome for phylogenetic estimates generally shows promise for resolving deep relationships among the plant lineages (Nie et al., 2020). Compared with the traditional DNA fragments, the cp genome was relatively conserved and slightly varied (Liang et al., 2020). The method has recently been applied to many research fields, such as taxonomic revision, systematic evolution, and species identification (Chen et al., 2019; Henriquez et al., 2020). Floden and Schilling (2018) and Xia et al. (2022) used cp genome data to reconstruct the phylogeny of *Polygonatum*, and the results supported the three groups and their sister relationship with *Heteropolygonatum*. Although these studies resolved the phylogenetic relationships of some species of *Polygonatum*, the phylogenetic relationships among the genera of Trib. Polygonateae and some species of *Polygonatum* were still unclear. In addition, the reliability of some analyses still needs to be further clarified due to the limited number of samples in the previous study. Given this, it is necessary to provide further support for the intra-generic relationships, divergence times, and genomic characteristics of Trib. *Polygonateae* based on a larger sample size. In the present study, we *de novo* assembled and annotated the cp genome of 26 species, including 23 species of Asparagaceae (18 species of *Polygonatum*, four species of *Disporopsis,* and one species of *Maianthemum*), and three species of *Disporum* (Colchicaceae). Besides, comparative analysis and phylogenetic evolution of the cp genome were conducted. The present results provide a basis for species identification, phylogenetic studies, resource development, and utilization of *Polygonatum* medicinal plants.

MATERIALS AND METHODS

Plant Material and DNA Sequencing

The fresh and healthy leaves of *Polygonatum, Disporopsis, Maianthemum*, and *Disporum* were collected in the field or Germplasm Resource Garden (China), and then the leaf tissue was frozen fresh at –20°C. Numbers after taxa names refer to the locality, and sample information is shown in **Supplementary Figure 1** and **Supplementary Table 1**. The specimens were identified following the taxonomic key and external morphology diagnosis proposed by related literature (Tang, 1978). The voucher specimens have been deposited at the herbarium of Dali University. Total genomic DNA was extracted from tissue samples using the Plant Genomic DNA kit (Tiangen, Beijing, China). The extracted DNA was quantified on a Nanodrop 2000 spectrophotometer (Nanodrop Technologies, Thermo Scientific, United States), and all PCR products were tested for the presence of amplified products on agarose gels. The library of each sample was prepared using 30 μl of high-quality (>100 ng/) genomic DNA. All libraries were sequenced on the Illumina NovaSeq system (Illumina, San Diego, CA, United States).

Genome Assembly and Annotation

The pair-end reads were trimmed for adapter and low-quality reads (Phred score < 30) using NGS QC Toolkit v.2.3.3 software. The cp genomes of *P. sibiricum* (NC029485), *Disporopsis fuscopicta* (MW248136), *Maianthemum bicolor* (NC035970), and *Disporum cantoniense* (MW759302) were downloaded from the National Center of Biotechnology Information (NCBI). The genome above was then used as the reference sequence.

TABLE 1 | Summary of cp genome features.

Species	Total length (bp)	GC content (%)	AT content (%)	LSC length (bp)	SSC length (bp)	IR length (bp)	Gene number	Protein-coding gene number	rRNA gene number	tRNA gene number	GenBank accession
P. kingianum	155,802	37.7	62.3	84,625	18,525	26,326	133	85	8	38	MZ029091
P. cirrhifolium	156,021	37.6	62.4	84,618	18,573	26,415	133	87	8	38	MZ029092
P. sibiricum 1	155,512	37.7	62.3	84,533	18,417	26,281	134	88	8	38	MZ029093
P. cyrtonema	155,512	37.7	62.3	84,462	18,292	26,379	132	86	8	38	MZ029094
P. alternicirrhosum	155,806	37.7	62.3	84,588	18,524	26,347	131	85	8	38	OL405009
P. filipes	155,472	37.7	62.3	84,422	18,292	26,379	133	87	8	38	OL405010
P. franchetii	155,228	37.7	62.3	84,164	18,418	26,323	133	87	8	38	OL405011
P. hookeri	155,953	37.6	62.4	84,573	18,550	26,415	133	87	8	38	OL405012
P. humile	155,185	37.7	62.3	84,102	18,455	26,314	133	87	8	38	OL405013
P. hunanense	155,456	37.7	62.3	84,286	18,426	26,372	133	88	8	38	OL405014
P. involucratum	155,372	37.7	62.3	84,282	18,450	26,320	133	87	8	38	OL405015
P. odoratum	154,576	37.8	62.2	83,493	18,459	26,312	133	87	8	38	OL405016
P. prattii	155,887	37.6	62.4	84,503	18,554	26,415	133	87	8	38	OL405017
P. mengtzense	155,590	37.7	62.3	84,539	18,427	26,312	131	85	8	38	OL587680
P. stewartianum	155,847	37.7	62.3	84,629	18,524	26,347	131	85	8	38	OL405018
P. uncinatum	155,681	37.7	62.3	84,596	18,529	26,278	133	87	8	38	OL405019
P. sibiricum 2	155,514	37.7	62.3	84,536	18,416	26,281	133	87	8	38	OL405024
P. zanlanscianense	155,827	37.6	62.4	84,463	18,534	26,415	133	87	8	38	OL405020
P. stenophyllum	155,961	37.7	62.3	84,609	18,561	26,395	132	86	8	38	OL405025
Disporopsis aspersa	156,110	37.7	62.3	85,055	18,525	26,265	131	85	8	38	OL405021
Disporopsis longifolia	156,008	37.7	62.3	85,048	18,480	26,280	131	85	8	38	OL405022
Disporopsis fuscopicta	155,934	37.7	62.3	84,923	18,527	26,242	131	85	8	38	OL405023
Disporopsis pernyi	156,072	37.7	62.3	85,017	18,525	26,265	131	85	8	38	OL587681
Disporum megalanthum	156,583	37.6	62.4	84,973	18,020	26,795	127	81	8	38	OL405026
Disporum uniflorum	156,588	37.6	62.4	84,977	18,023	26,794	127	81	8	38	OL405027
Disporum cantoniense	156,562	37.6	62.4	84,926	18,002	26,815	127	81	8	38	OL587682
M. fuscum	156,711	37.6	62.4	85,218	18,447	26,523	131	85	8	38	OL405028

The cp genome was assembled using GetOrganelle v.1.6.4, exploiting Bowtie2 v.2.4.4, SPAdes v.3.13.0, and Blast v.2.5.0 as dependencies (get_organelle_from_reads.py -1 R1.fq -2 R2.fq -o cp_output -R 15 -k 21,45,65,85,105 -F embplant) (Jin et al., 2020). All clean reads were mapped to the database, and then the mapping data were extracted based on similarity and coverage. Subsequently, the assembled contigs were visualized and removed redundant sequence by Bandage v.0.8 to generate the complete circular cp genome (Wick et al., 2015). Finally, the reads were remapped to assembled cp genome by Bowtie2, and Jellyfish v.2.2.3 was then used to determine the reverse repeat region boundary. After assembly, circular cp genomes were annotated using online tools CpGAVAS2 and GeSeq based on the reference cp genome (Michael et al., 2017; Shi et al., 2019). The Apollo was used to correct the start codons, stop codons, and intron/exon boundaries (Lewis et al., 2002). Annotated cp genome sequences were submitted to the GenBank database of the NCBI to obtain specific accession numbers (**Table 1**). Fully annotated cp genome circle diagrams were drawn by OrganellarGenomeDRAW (OGDRAW) online (Lohse et al., 2007).

Genome Structure and Comparisons Analysis

The GC content was analyzed using Geneious v.9.0. Four types of the dispersed repeat sequence, including forward (F), complementary (C), palindromic (P), and reverse (R), were detected using the REPuter program[1] (Stefan et al., 2001).

[1]https://bibiserv.cebitec.uni-bielefeld.de/reputer/

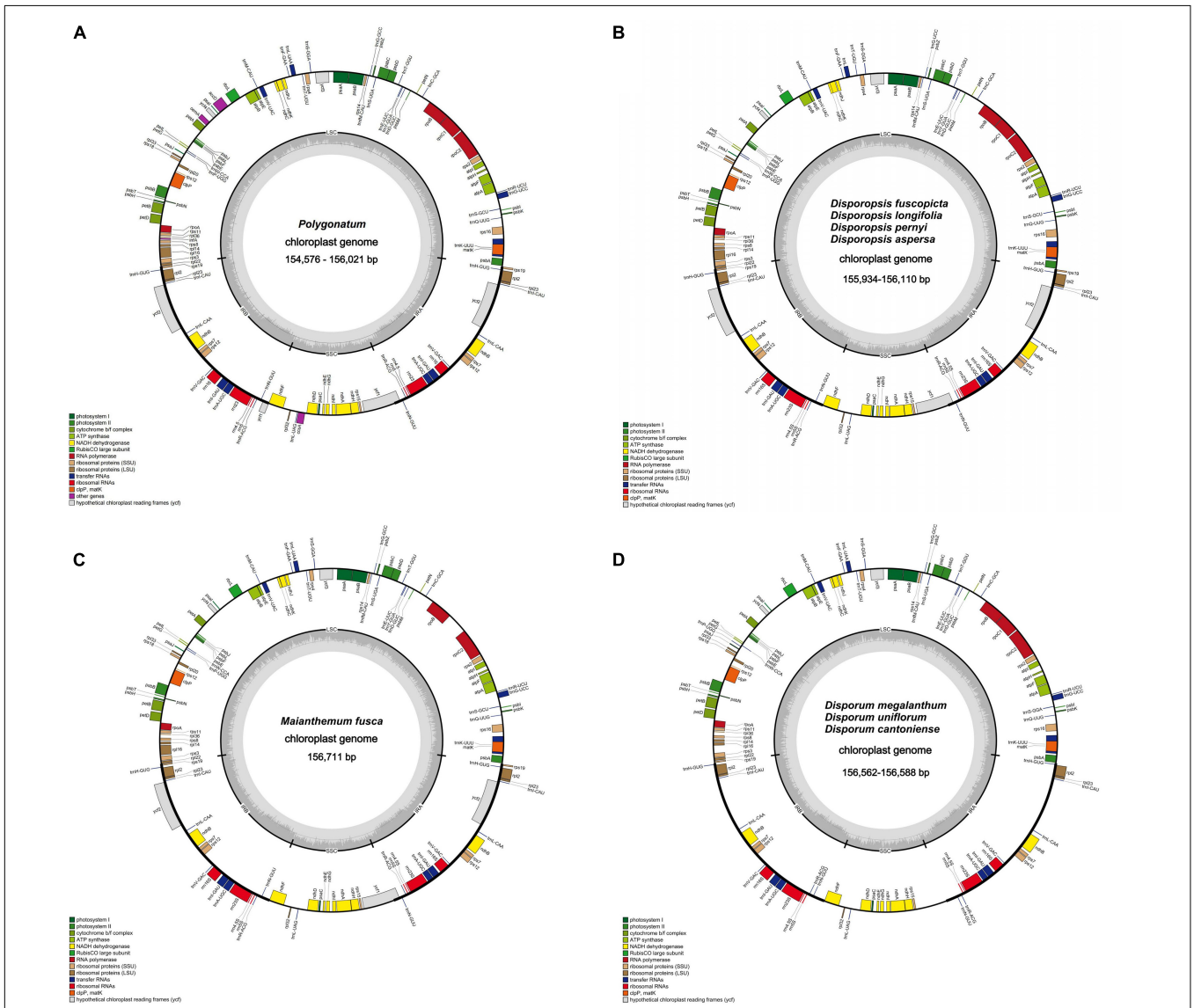


FIGURE 1 | Chloroplast genome map: **(A)** *Polygonatum*, **(B)** *Disporopsis*, **(C)** *Maianthemum*, **(D)** *Disporum*. Genes inside and outside the circle are transcribed in a clockwise and counter-clockwise direction.

Tandem repeats were detected using Phobos v.3.3.12[2] (Mayer et al., 2010) with default parameter values. The cp genome of *P. sibiricum* (MZ029093) was selected as a reference for coordinate positions, and indels and SNPs were counted within the non-overlapping 150 bp window for 18 *Polygonatum* plastomes (**Supplementary Table 7a**; Liu et al., 2020). The region IRb was removed for the analyses of repeats to avoid over-representing the repeats following Abdullah et al. (2019). Spearman's Rho correlations were calculated based on substitutions, Indels, and oligonucleotide repeats using Minitab v. 18 (Akoglu, 2018).

The criteria for repeat determination include a minimum repeat size of 20 bp with the similarity between repeat pairs of 90% by putting edit value 3. Furthermore, the simple sequence repeats (SSRs) were analyzed using MISA software[3] with the parameters of "10" for mono-, "5" for di-, "4" for tri-, and "3" for tetra- and penta- nucleotide motifs (Beier et al., 2017). The cp genomes were compared with mVISTA under the Shuffle-LAGAN mode. The cp genome junctions were visualized and compared using IRscope[4] online (Amiryousefi et al., 2018). The cp genomes were aligned using the MAFFT (Katoh and Standley, 2013). Additionally, the nucleotide variability across the cp genome sequences was analyzed using DnaSP v.6.12.03, with a window length of 600 sites and a step size of 200 sites.

Phylogenetic Analyses and Ancestral Character State Reconstruction

Phylogenetic reconstruction included 27 *de novo* assembled sequences (**Table 1**), and 61 cp genomes downloaded from NCBI (**Supplementary Table 2**). At the same time, two species, *Dioscorea esculenta* (NC052854) and *Dioscorea schimperian* (NC039855), were used as outgroups. A total of 88 sequences were aligned using MAFFT with default parameters and trimmed using trimAl v.1.4 with option automated. Neighbor-Joining (NJ) analyses were performed using the MEGA X, applied with 1,000 bootstrap replicates at each branch node (Sudhir et al., 2018). The alignment was also evaluated using bootstrap analysis on 1,000 in a maximum likelihood (ML) by IQ-tree (Nguyen et al., 2015), with parameters: iqtree -s input -m MFP -b 1000 -nt AUTO -o NC052854, NC039855, best-fit nucleotide substitution model.

The leaf arrangement was selected to analyze the phyllotaxy evolution of *Polygonatum*. The phyllotaxy information was obtained from taxonomic literature and the Flora of China (Chen and Tamura, 2000; Jiao, 2018). The states of phyllotaxy were coded: alternate (A), verticillate (B), opposite (C), and the crowd (D). In the case of some species with more than two-character states, we coded the character state based on their dominant status. For example, the leaf arrangement for *Polygonatum prattii* was coded as alternate because it usually has alternate leaves, although there is an opposite leaf arrangement or three-verticillate leaves occasionally. For *Polygonatum hunanense*, its leaves were mainly verticillate, sometimes with a few alternate or opposite leaves, thus coded as verticillate in the analyses. Statistical Dispersal-Vicariance Analysis (S-DIVA) and Multistate Reconstruction BayesTraits (MRBT) method were conducted in RASP software to infer ancestral character states (Yu et al., 2015).

Divergence Time Estimation

The divergence times of *Polygonatum* were calculated using the Markov chain Monte Carlo (MCMC) tree program of PAML (Puttick, 2019). IQ-tree was used to estimate the best tree topology of the data set. According to the previous study (Chen et al., 2013; Eguchi and Tamura, 2016; Wang et al., 2016; Xia et al., 2022), we used four calibration points to restrict each node: (F1) 115.9–137.4 Ma for the root node, (F2) 58.3–76.6 Ma for Asparagaceae stem age, (F3) 56.4–72.7 Ma for Asparagaceae crown age, (F4) 14.34–27.54 Ma for *Polygonatum* and *Heteropolygonatum*. The clock model uses the independent rate model (IRM), which follows a lognormal distribution. Nucleotide substitution selects the HKY model with alpha for gamma rates at sites set to 0.5. The birth–death process is used to generate uniform node age priors in the tree, using the default parameter ($\lambda = 1$, $\mu = 1$, $s = 0.1$). The posterior probabilities of parameters were calculated using MCMC samples. The first 10% trees were discarded as burn-in and then sampled every 10 iterations until 20,000 samples were gathered.

RESULTS

Sequencing, Assembly, and Annotation

The raw data of 27 individuals were filtered to remove adapters and low-quality reads; 3–5 Gb data were obtained for each species in this study. After assembly and splicing, the complete cp genomes of the circular tetrad structure were obtained (**Figure 1**). The annotated result suggested that the cp genome length of *P. odoratum* (154,576 bp) was the smallest, and the cp genome length of *M. fuscum* (156,711 bp) was the largest among the 27 individuals. The length of the LSC region ranged from 83,493 bp (*P. odoratum*) to 85,218 bp (*M. fuscum*). The length of the SSC region ranged from 18,002 bp (*D. cantoniense*) to 18,573 bp (*P. cirrhifolium*), and the length of the IRa and IRb regions ranged from 26,242 bp (*D. fuscopicta*) to 26,815 bp (*D. cantoniense*). The GC content of the cp genomes ranged from 37.6 to 37.8% and varied among the different regions of the cp genomes. In addition, the number of genes and introns were highly conserved (**Table 1**), and the same suite of rRNA genes and tRNA genes was found in all taxa. All genomes have 85–88 protein-coding genes, except for *D. cantoniense, D. megalanthum,* and *D. uniflorum*, with 83 protein-coding genes (lacking *rps16* and *rpl32* genes) (**Supplementary Table 4**). It is worth noting that 19 genes were repeated in the *Polygonatum*, which were involved in photosynthesis and self-replication (**Supplementary Table 3**).

Repeat Analysis

Repetitive sequences in the cp genome play a critical role in genome evolution and rearrangements. Oligonucleotide repeats analysis of four types of repeats in the cp genome, including Forward (F), Reverse (R), Palindromic (P), and Complementary (C), was performed by REPuter. The number of repeat types

[2]http://www.rub.de/ecoevo/cm/cm_phobos.htm

[3]http://pgrc.ipk-gatersleben.de/misa/

[4]https://irscope.shinyapps.io/irapp/

varied among the 26 cp genomes and presented random permutations, but most repeat sequences existed in 20–29 bp (**Supplementary Figure 2**). The abundance of F and P repeats was higher than that of R and C repeats (**Supplementary Figure 2**). The minimum number of repeats was found in *Disporum uniflorum* (60), whereas the maximum was found in *P. cirrhifolium*, *P. sibiricum*, and *P. zanlanscianense* (86). Complete details have been listed in **Supplementary Tables 5, 6**. Moreover, Spearman's Rho correlation coefficients were obtained between tandem repeats, indels, and SNPs (**Supplementary Table 7b**). All these correlations showed a significant value (tandem repeats and indels: $p < 0.001$, tandem repeats and SNPs: $p < 0.001$, indels and SNPs: $p < 0.001$). The average correlation values between tandem repeats and indels, indels and SNPs, and tandem repeats in 26 *Polygonatum* species were 0.469, 0.351, and 0.267, respectively. Furthermore, we identified 67 (*P. franchetii*)–87 (*M. fuscum*) SSRs per cp genome consisting of mono- to hexa-nucleotide repeating units (**Supplementary Table 8**). Most of the SSRs were located in the intergenic areas. More than half of these SSRs (52.94–66.23%) were mononucleotide A/T motif, followed by dinucleotide (18.29–26.47%) with a predominant motif of AT/TA, tetranucleotide repeats (10.81–13.89%) with a predominant motif of AAAT/ATTT, AATC/ATTG, trinucleotide (2.94–6.94%), pentanucleotide (2.44–4.35%) with a predominant motif of AAACG/CGTTT, and hexanucleotides (0–1.35%) were absent in the cp genome of *M. fuscum* (**Supplementary Figure 3**).

Inverted Repeats Regions Contraction and Expansion

The contraction and expansion of IR regions revealed variation in LSC/IR/SSC regions (**Figure 2**). The *rpl22* gene was present in the LSC region, and *rpl2* and t*rnH* existed entirely in the IRb region. The *rps19* gene was present in the junction of the IRa/IRb/LSC region in four genera (*Polygonatum, Heteropolygonatum, Disporopsis,* and *Maianthemum*), but in *Disporum,* the *rps19* gene was absent in IRa/LSC region and existed completely in the IRb region. Notably, *rps19* was started in IRa regions and integrated into the LSC by 60 base pairs in *P. cyrtonema*, whereas in all other species of *Polygonatum,* the *rps19* gene exists completely in the LSC region. Additionally, the *ndhF* gene was observed at the junction of IRb/SSC and integrated into the SSC varied from 21 to 34 bp. Another truncated copy of the *ycf1* gene was observed in all species at the IRa/SSC junction, which starts in IRa regions and integrates into the SSC. In addition, *rpl2* and *trnH* existed in the IRa, and *psbA* existed in the LSC. It is worth noting that the *rps19*

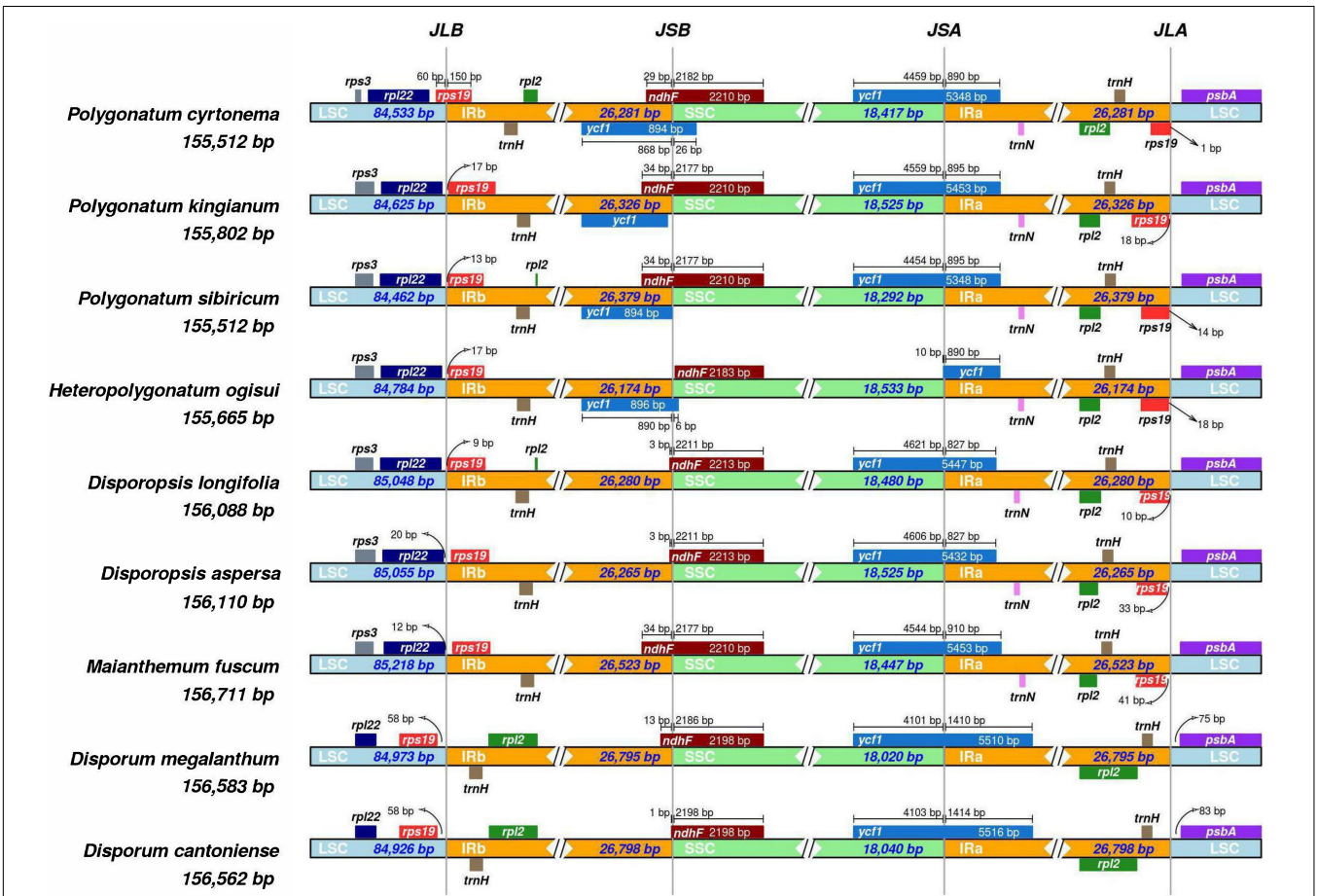


FIGURE 2 | Comparisons of the borders of LSC, SSC, and IRa/b regions among the *Polygonatum*, *Heteropolygonatum*, *Disporopsis*, *Maianthemum*, and *Disporum*.

and *trnN* genes existed entirely in the IRa region of four genera, whereas the two genes were missing in the *Disporum* at the junction of IRa and LSC (**Supplementary Figure 4**). Moreover, the genome alignment analysis showed that the cp genomes among the 26 species were relatively conserved, and no inversions, translocations, and genomic rearrangements were detected (**Supplementary Figures 5, 6**).

Comparative Chloroplast Genomic Analysis

Comparison of overall sequence variation showed that the cp genome within *Polygonatum* is highly conserved. The IR regions had lower sequence divergence than LSC and SSC regions. In addition, the coding region was more conserved than the non-coding regions. Furthermore, except for the more remarkable mutation in *ndhA, ycf1*, and *ycf2* genes, most of the protein-coding genes of *Polygonatum* were pretty conserved. The highest divergence in intergenic regions was found in the *rps16-trnQ-UUG, trnS-GCU-trnG-UCC, trnT-UGU-trnL-UAA, ndhC-trnV-UAC, rpl32-trnL-UAG, trnV-GAC-rps7,* and *accD-psaI*. The most divergent in the coding region were the *ycf1* and *ycf2* open reading frames (**Supplementary Figure 7**). Moreover, the sliding window analysis demonstrated that the seven regions had greater nucleotide diversity values (>0.01), including *matK-rps16, trnS-GCU-trnG-UCC, rpl32-trnL-UAG, trnC-GCA-petN, petA-psbJ, ccsA,* and *ycf1* (**Figure 3**). The polymorphism loci of these variability regions are listed in **Supplementary Table 9**. Among these regions, nucleotide diversity values of *rps16-trnQ-UUG, trnS-GCU-trnG-UCC, rpl32-trnL-UAG,* and *trnC-GCA-petN* were greater than 0.01, and the ycf1 gene was the lowest (0.00047). The insertions/deletions (indels) diversity of *trnS-GCU-trnG-UCC* and *matK-rps16* were 7.076 and 5.181, respectively, with no indel events detected in the *ccsA* gene. Furthermore, the cp genome of *Polygonatum, Heteropolygonatum, Disporopsis*, and *Maianthemum* is similar with an average similarity of 99% but different from that of *Disporum* with an average similarity of 85% based on the global comparison (**Supplementary Figure 8**).

Phylogenetic Analysis

The ML and NJ phylogenetic trees were inferred using 87 species, with the *Dioscorea* as the outgroup. The consensus trees obtained from the inference analyses were resolved, and most nodes were supported with maximum support (100% bootstrap support, **Figure 4** and **Supplementary Figure 9**). The core Asparagaceae includes the subfamily of Scilloideae, Nolinoddeae, and Agavoideae, which form a monophyletic group (group I). Scilloideae and Agavoideae were sister taxa within the three subfamilies, and Nolinoddeae was a sister group to the clade of Scilloideae + Agavoideae. In addition, the results showed that most species of Trib. Polygonateae were placed in the crown of the phylogenetic tree, including *Polygonatum*, *Heteropolygonatu*m, and *Disporopsis*, and supported the monophyly of three genera. However, within this clade, the *Maianthemum* and Ophiopogoneae were sisters to a clade formed by *Disporopsis*, *Heteropolygonatum,* and *Polygonatum*, while *Disporum* was polyphyletic across three separate clades and distantly related to Trib. Polygonateae. In addition, the *Polygonatum* is further divided into sect. *Sibirica*, sect. *Polygonatum*, and sect. *Verticillata*. And the sister relationship was between sect. *Sibirica* and sect. *Polygonatum,* whereas sect. *Verticillata* was placed as sister to sect. *Polygonatum* + sect. *Sibirica* with high support (100% B/S). It is worth mentioning that sect. *Sibirica* only includes a species of *P. sibiricum.* The NJ and ML analyses produced trees with similar topologies, although some poorly supported groups were sensitive to changes in the mode of inference. The position of several species was unresolved, including *P. hunanense* and *P. kingianum*, which varied among trees recovered using distinct phylogenetic inference methods.

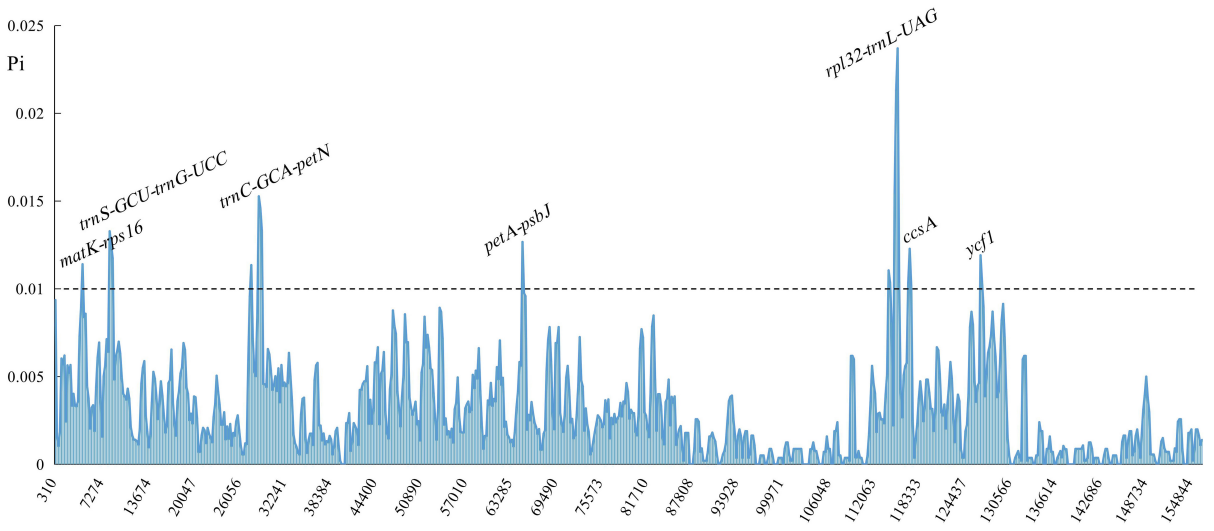


FIGURE 3 | Sliding window analysis of *Polygonatum* cp genome. The X-axis represents the midpoint of the window; the Y-axis represents nucleotide diversity values.

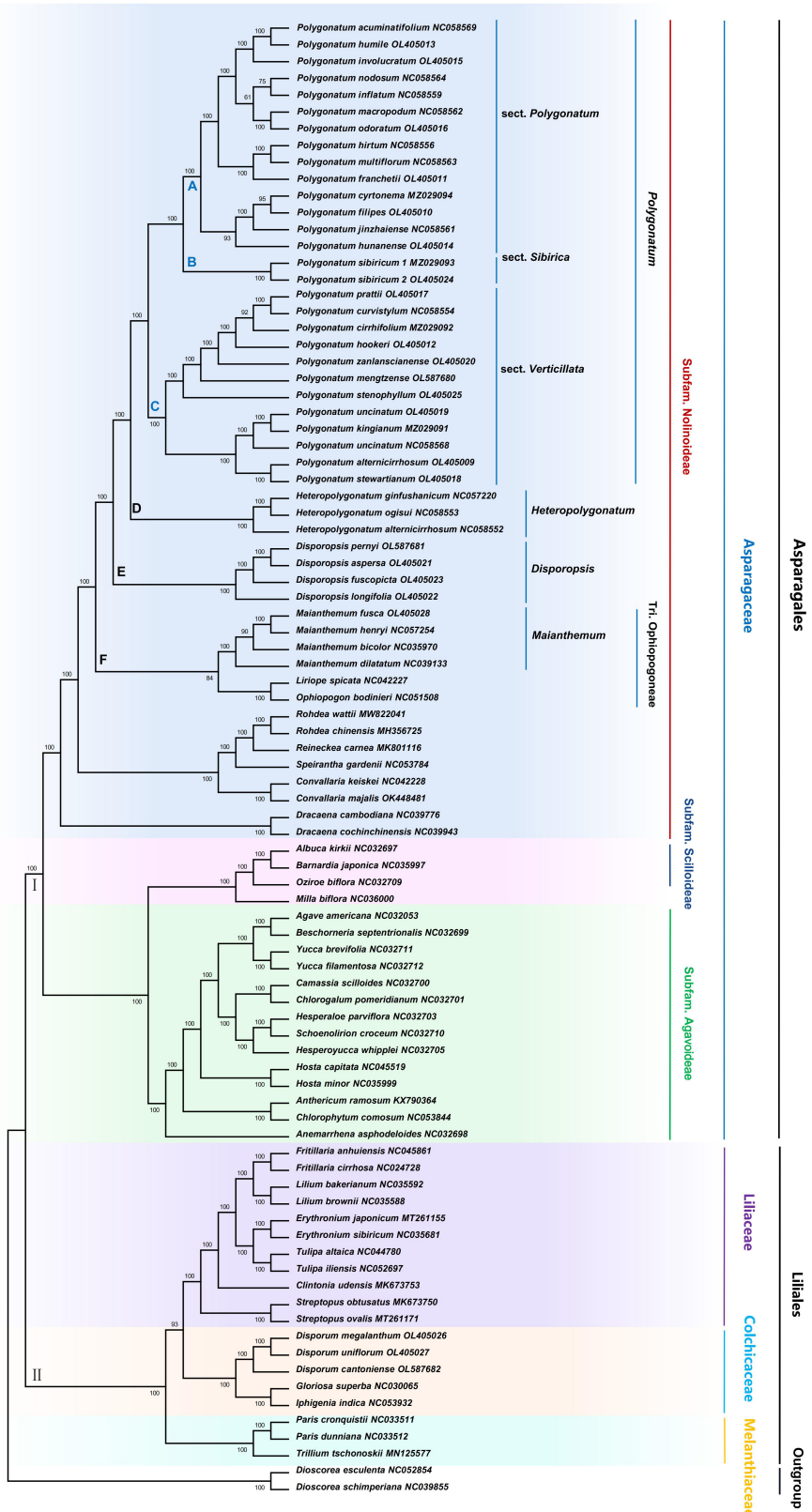


FIGURE 4 | Maximum likelihood phylogenetic tree based on complete cp genome. *Dioscorea esculenta* and *D. schimperiana* were used as outgroups. Numbers at nodes are bootstrap support values.

Divergence Time Estimation

Results of divergence time for the node of the 95% highest posterior density (HPD) intervals are shown in **Supplementary Table 10**. A complete chronogram is demonstrated in **Figure 5**. The extant genera of the *Polygonatum* and *Disporopsis* have shared a common ancestor at the beginning of the Eocene (41.68 Ma, 31.97–57.29, 95% HPD), while the split between *Polygonatum* and *Heteropolygonatum* is estimated to occur at 16.56 Ma (HPD = 13.57–20.56 Ma, 95%), and sect. *Verticillata*, sect. *Polygonatum* and sect. *Sibirica* might share a common ancestor at 14.71 Ma (1.32–18.57 Ma, 95% HPD), and the divergence times between sect. *Polygonatum* and sect. *Sibirica* was formed at approximately 11.80 Ma. Sampled specimens of *Maianthemum* and *Ophiopogon* were estimated to have originated in 52.22 Ma. Moreover, the divergence times of the *Disporum* occurred at 128.56 Ma, having shared a common ancestor with the Asparagaceae.

Reconstruction of Leaf Morphological Character

In the classification of *Polygonatum*, the arrangement of leaves was usually concerned. Phyllotaxy is one of the main characteristics of *Polygonatum* taxonomy, character transformation of leaf order is essential to understanding the evolution of *Polygonatum*. The phyllotaxy was used to reconstruct ancestral traits of *Polygonatum* and its relative species. As illustrated in **Figure 6**, the S-DIVA results showed that the verticillate leaves arrangement was the most likely ancestral state of *Polygonatum*, which was consistent with the MRBT method (B: $p = 0.95$). In addition, sect. *Polygonatum* is marked by phyllotaxy with an alternate leaf, except for *P. hunanense*. In its sister clade, sect. *Sibirica* is mostly a verticillate leaf arrangement. Sect. *Verticillata* includes species that appear to have a combination of the opposite, alternate, and verticillate leaves. In addition, alternate and verticillate leaves evolved more than once. Notably, the *Heteropolygonatum* and *Disporopsis* showed alternative phyllotaxy.

DISCUSSION

Chloroplast Genome Structure and Comparative Analysis

In the present study, we *de novo* assembled the cp genome of 26 species of Trib. Polygonateae and performed comparative analyses. The cp genome of 26 species exhibited a quadripartite structure with two IR regions separated by the LSC and SSC regions. Higher GC content was observed in the IR region compared with the LSC and SSC regions, consistent with previous reports (Liang et al., 2021). In addition, our results found that the total length, GC content, and gene composition of the cp genome were almost identical in all species. Previous studies have found that the angiosperms possessed a highly conserved nature in the cp genome at the genus level (Yu et al., 2019; Shahzadi et al., 2020), but we found that the *rps16* and *rpl32* genes lost in *Disporum,* and this variation may be specific to *Disporum*.

Inverted repeat contraction and expansion could cause gene duplication, the origination of pseudogenes, and length variation in the cp genome, which were considered critical evolutionary phenomena. In the present study, *rps19* is present in the IR region in four genera (*Polygonatum, Heteropolygonatum, Disporopsis,* and *Maianthemum*), except for *Disporum*. Previous studies of monocotyledons' cp genome revealed that the *rps19* gene existed in the IR region (Ahmed et al., 2012; Nock et al., 2014; Henriquez et al., 2020). Whereas the *de novo* assembled genome of *Disporum* contrast with previous studies, revealing integration of *rps19* into the LSC region. Moreover, the *ycf1* gene was duplicated in the IRa and IRb regions in *Polygonatum* and *Heteropolygonatum*, while in the other three genera (*Disporopsis, Maianthemum,* and *Disporum*), this gene only exists in the junction of the SSC/IRa region, which is consistent with a previous study (Wang et al., 2011).

At the genus level, weak-to-strong correlations among tandem repeats, SNPs, and indels have been observed in the *Polygonatum*. We found a weak correlation between tandem repeats and SNPs, a moderate correlation between indels and SNPs, and a strong correlation between tandem repeats and indels in *Polygonatum* plastomes. A recent study has confirmed that the plastomes exhibited strong associations between tandem repeats, indels, and substitutions in Araceae and Malvaceae (Abdullah et al., 2020, 2021). Our results also supported prior findings that tandem repeats play an important role in generating the indels and SNPs. These results have practical implications in selecting appropriate loci for comparative analyses.

Phylogenetic and Taxonomic Resolution

The phylogeny and classification of *Polygonatum* have long been debated (Feng et al., 2020). This study used 88 cp genome, including most of the basal monocot family Asparagaceae, to construct the phylogenetic tree. Results of the NJ and ML phylogenies analysis confirmed the position of *Polygonatum* within the Asparagaceae, which were congruent and largely concordant with recent phylogenomic studies (Zhao et al., 2019; Xia et al., 2022). There is strong support for the monophyly of many major clades of Asparagaceae, including *Polygonatum*, *Heteropolygonatum*, *Disporopsis, Maianthemum*, and *Rohdea*. In a previous study, the *Polygonatum* was subdivided into two sections: sect. *Polygonatum* and sect. *Verticillata* based on the *trnK* (Tamura et al., 1997), but in our phylogeny, the *Polygonatum* was recovered as monophyletic in NJ and ML analyses, which were divided into three sections: sect. *Sibirica*, sect. *Polygonatum*, and sect. *Verticillata*, and can be strongly supported as a sister relationship between (1) *Polygonatum* and *Heteropolygonatum*, and (2) sect. *Sibirica* and sect. *Polygonatum,* and (3) sect. *Verticillata* and sect. *Polygonatum* + sect. *Sibirica*, respectively, which is consistent with previous studies (Meng et al., 2014; Floden, 2017; Zhao et al., 2019; Xia et al., 2022).

Moreover, we found several interesting implications of phylogeny in this study. First, both NJ and ML analyses provided strong evidence for the monophyly of *P. sibiricum* in sect. *Sibirica*. These results were supported by the findings of other

FIGURE 5 | Divergence time estimation based on cp genome sequences. The divergence times are exhibited on each node, whereas the blue bars represent the 95% highest posterior density interval for each node age.

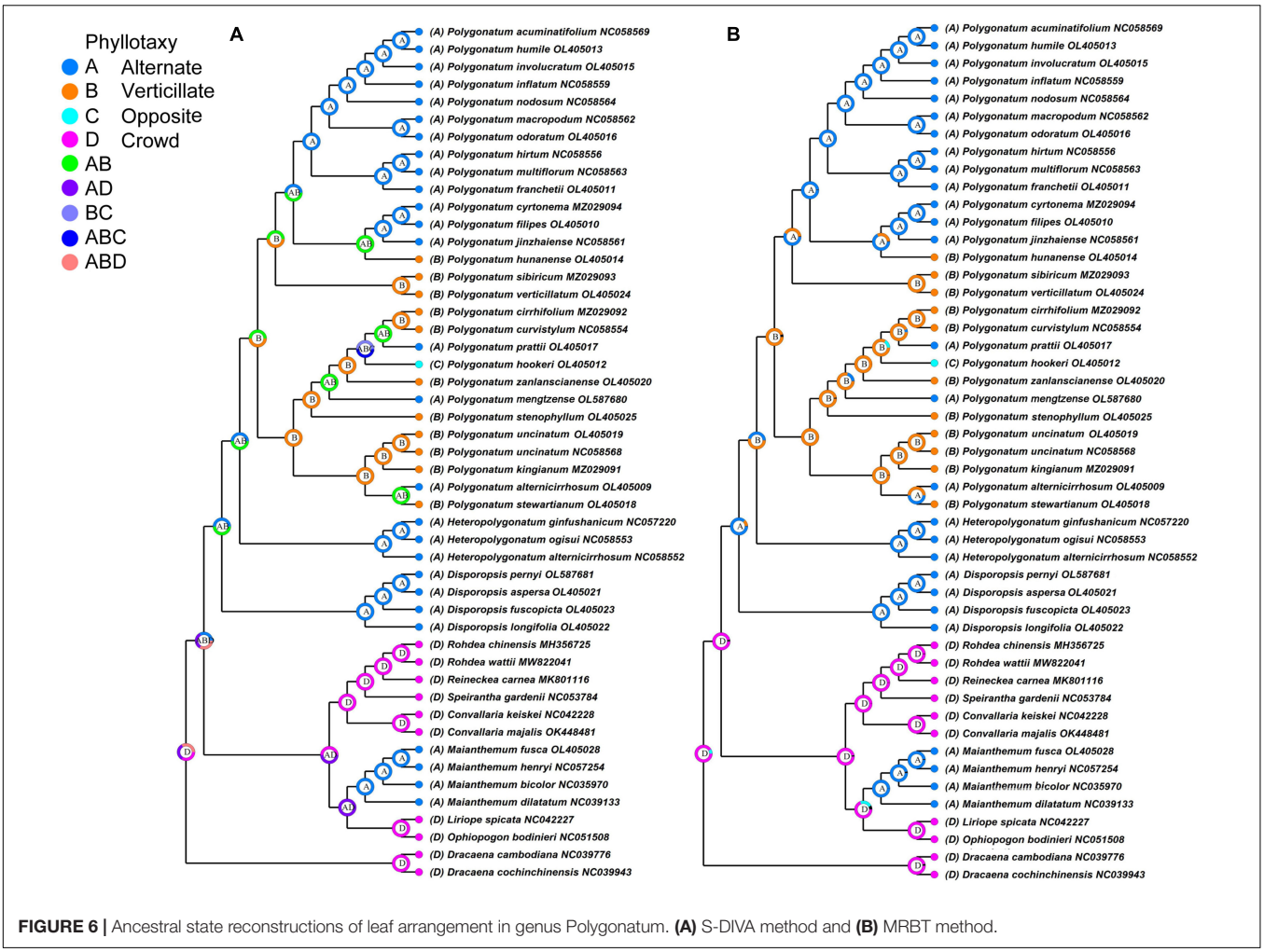


FIGURE 6 | Ancestral state reconstructions of leaf arrangement in genus Polygonatum. **(A)** S-DIVA method and **(B)** MRBT method.

researchers based on *matK*, *trnK*, *rbcL*, *psbA-trnH*, and cp genome (Meng et al., 2014; Floden, 2017; Zhao et al., 2019; Xia et al., 2022). In contrast, the cp genome tree placed *P. sibiricum* nested within sect. *Verticillata* or sect. *Sibirica* in a previous study (Floden and Schilling, 2018). Therefore, additional phylogeny analysis using an enormous collection of *Polygonatum* should be performed in the future. Second, previous phylogenetic studies of *Polygonatum* based on the cp genome have shown that the *Heteropolygonatum alternicirrhosum* belongs to the *Heteropolygonatum* and strongly supported the monophyly of *Heteropolygonatum* (Xia et al., 2022). However, in our study, the ML and NJ tree which strongly supported *P. alternicirrhosum* was deeply nested within the *Polygonatum* and supported the monophyly of *Polygonatum* and *Heteropolygonatum*. Our analyses suggest that *Polygonatum* may be the sister to the *Heteropolygonatum*. In addition, another previous study also supports that the species belongs to *Polygonatum* (Chen and Tamura, 2000). Therefore, our results suggested that *P. alternicirrhosum* (OL405009) and *H. alternicirrhosum* (NC058552) should be considered as two species, in contrast with other recent estimations, mainly to the influential Floden's (2014) study, which suggested that *P. alternicirrhosum* should be transferred to *Heteropolygonatum* based of morphological, cytological comparisons and molecular data. In addition, Zhao et al. (2019) indicated that the *P. alternicirrhosum* should be recovered within the *Polygonatum*. Another example is the *Polygonatum mengtzense*, which was considered most closely related to the *P. punctatum* (Floden, 2014). In this study, maximum likelihood analysis reveals that *P. mengtzense* was deeply nested within sect. *Verticillata*, which was supported by molecular and chromosome evidence (Floden, 2014; Xia et al., 2022). Finally, there is a significant divergence in the classification of Trib. Polygonateae. In the previous study, Kim et al. (2010) placed *Disporopsis* as a sister to *Polygonatum* + *Maianthemumin* in a Bayesian analysis. In the present study, the phylogenetic analyses corroborated the monophyly of *Polygonatum, Heteropolygonatum, Disporopsis, Maianthemum,* and *Disporum*. The *Disporum* is sister to Glorioseae and nested within the family Colchicaceae (more distantly related species), consistent with previous genetic studies (Tamura et al., 2013).

It is interesting to note that the *Maianthemum* was traditionally included with Trib. Polygonateae based on morphology (Chen and Tamura, 2000), and other research based

on multiple plastid markers (*atpB*, *ndhF*, *rbcL*, *matK*, *psbA-trnH*, *trnC-petN*, *atpB-rbcL*, and *rps16*) also supports its placement in Trib. Polygonateae (Chen et al., 2013; Meng et al., 2014; Zhao et al., 2019). However, ML analysis in our study reveals that the *Maianthemum* was deeply nested within Trib. Ophiopogoneae rather than Trib. Polygonateae, which agrees with former studies based on multiple loci (e.g., *petA-psbJ*, *ETS*, *ITS*, and *rps10*; or *trnL-F*, *rps16*, *rpl16*, *psbA-trnH, rbcL, trnK, trnC-petN,* and *ITS*) (Floden, 2017; Floden and Schilling, 2018; Meng et al., 2021). Our results suggested that the Trib. Polygonateae should include only three genera (*Disporopsis, Heteropolygonatum*, and *Polygonatum*). Therefore, the results revealed that the convergent evolution of some traits may have misled previous relationships. Further phylogenetic analysis is needed within the *Maianthemum*.

Diversification History and Leaf Arrangements

Results of divergence time estimates suggest that the elevated diversification rates of *Polygonatum* occurred from approximately 15–0.1 Ma during the late Miocene and early Pliocene. The two main lineages, sect. *Verticillata* and sect. *Sibirica* + sect. *Polygonatum* seem to have radiated since the mid-Miocene (sect. *Verticillata*: 11.52 Ma; sect. *Sibirica* + sect. *Polygonatum*: 11.80 Ma; **Figure 5**; **Supplementary Table 10**). Notably, the diversification rates of *Polygonatum* slowly increased during this period, attributed to uplifts of the Qinghai–Tibetan Plateau in the early Miocene (Xue et al., 2021). In addition to the aforementioned tectonic rearrangements and mountain formation in East Asia, the global climatic fluctuations and aridification that occurred in the Mid-Miocene Climatic Optimum (MMCO, 15–17 Ma; Zachos et al., 2001) also accelerated the diversification rates of *Polygonatum*. Global warming occurred at approximately 15 Ma (MMCO), followed by a gradual decrease in temperature (Zachos et al., 2001). These climatic changes might have influenced the plant diversification and promoted radiation of *Polygonatum* species.

Variation in phyllotaxy morphology represents an important character source for species delimitation. The phyllotaxy diversity (alternate, opposite, and verticillate) caused some confusion in classifying the *Polygonatum*. Our study showed the evolutionary trend of *Polygonatum* from verticillate leaves to alternate leaves, and this suggests that verticillate leaf is the ancestral state and agrees with the previous molecular studies of *Polygonatum* (Xia et al., 2022). It is noteworthy that the phyllotaxy of *Polygonatum* is an unstable character even for the same species. Therefore, phyllotaxy cannot be used as the unique taxonomic feature for classifying *Polygonatum*.

CONCLUSION

In this study, the complete cp genome of 26 species of Trib. Polygonateae was *de novo* assembled from Illumina reads. In all of our analyses, these cp genomes were generally conserved and exhibited similar gene content and genomic structure. A total of 8 highly variable loci were identified across the *Polygonatum* cp genome, which could serve as potential markers for phylogenetic and population genetics studies. The monophyly of *Polygonatum* was confirmed, and phylogenetic analysis indicated that the genus consists of three sections (sect. *Sibirica*, sect. *Polygonatum*, and sect. *Verticillata*). Meanwhile, the phylogenetic analysis suggested that *Heteropolygonatum* may be the sister group of the *Polygonatum*, but the *Disporopsis*, *Maianthemum*, and *Disporum* may have diverged earlier. In conclusion, our results enhanced the genomic information for *Polygonatum* and provided valuable insight into the phylogenetic relationships among the genera involved in Trib. Polygonateae. The results also contribute to the bioprospecting and conservation of the *Polygonatum*.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

JW, ZX, SC, and BD participated in the conception and design of the research. JW, JQ, FY, YJ, and BZ collected and identified the species. JW, JQ, and XC were responsible for analyzing and processing data. JW wrote the manuscript. ZX and BD revised the manuscript. All authors agreed to the submitted version of the manuscript.

FUNDING

This work was supported by the Yunnan Academician Expert Workstation (202205AF150026 and 202105AF150053), the National Natural Science Foundation of China (Grant No. 31860080), the key technology projects in Yunnan Province of China (202002AA100007), the National Key R&D Program of China (Grant No. 2019YFC1711100), Heilongjiang Touyan Innovation Team Program (Tree Genetics and Breeding Innovation Team), and the Yunnan Xingdian Talent Support Plan (YNWR-QNBJ-2020251).

ACKNOWLEDGMENTS

We would like to thank Zhongyu Zhou, Jie Liu, and Qingshu Yang for their assistance in obtaining specimens for this study. We thank Yaping Wang at Jinan University for her assistance with language polishing. We are also thankful to Northeast Forestry University and the China Academy of Chinese Medical Sciences for technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.882189/full#supplementary-material

REFERENCES

Abdullah, Henrique, C. L., Croat, T. B., Poczai, P., and Ahmed, I. (2020). Mutational dynamics of Aroid chloroplast genomes II. *Front. Genet.* 11:610838. doi: 10.3389/fgene.2020.610838

Abdullah, Mehmood, F., Shahzadi, I., Ali, Z., Islam, M., Naeem, M., et al. (2021). Correlations among oligonucleotide repeats, nucleotide substitutions, and insertion-deletion mutations in chloroplast genomes of plant family Malvaceae. *J. Syst. Evol.* 59, 388–402. doi: 10.1111/jse.12585

Abdullah, Shahzadi I., Mehmood F., Ali Z., Malik M. S., Waseem S., et al. (2019). Comparative analyses of chloroplast genomes among three *Firmiana* species: identification of mutational hotspots and phylogenetic relationship with other species of Malvaceae. *Plant Genet.* 19:100199. doi: 10.1016/j.plgene.2019.100199

Ahmed, I., Bigg, P. J., Matthews, P. J., Collins, L. J., Hendy, M. D., Lockhart, P. J., et al. (2012). Mutational dynamics of Aroid chloroplast genomes. *Genom. Biol. Evol.* 4, 1316–1323. doi: 10.1093/gbe/evs110

Akoglu, H. (2018). User's guide to correlation coefficients. *Turk. J. Emerg. Med.* 3, 91–93. doi: 10.1016/j.tjem.2018.08.001

Ali, M., Liu, Y. J., Xia, Q. P., Bahadur, S., and Shuaib, M. (2021). Pollen micromorphology of eastern Chinese *Polygonatum* and its role in taxonomy by using scanning electron microscopy. *Microsc. Res. Techniq.* 84, 1451–1461. doi: 10.1002/jemt.23701

Amiryousefi, A., Hyvönen, J., and Poczai, P. (2018). IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34, 3030–3031. doi: 10.1093/bioinformatics/bty220

Baker, J. G., and Esq, F. L. S. (1875). Revision of the genera and species of Asparagaceae. *Bot. J. Linn. Soc.* 80, 547–632. doi: 10.1111/j.1095-8339.1875.tb00349.x

Beier, S., Thie, T., Münch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33, 2583–2585. doi: 10.1093/bioinformatics/btx198

Chen, Q., Wu, X. B., and Zhang, D. Q. (2019). Phylogenetic analysis of *Fritillaria cirrhosa* D. Don and its closely related species based on complete chloroplast genomes. *Peer J.* 7:e7480. doi: 10.7717/peerj.7480

Chen, S., Kim, D. K., Chase, M. W., and Kim, J. H. (2013). Networks in a large-scale phylogenetic analysis: reconstructing evolutionary history of Asparagales (Lilianae) based on four plastid genes. *PLoS One* 8:e59472. doi: 10.1371/journal.pone.0059472

Chen, X. Q., and Tamura, M. N. (2000). *Flora of China. Science Press*. Beijing: Beijing and Missouri Botanical Garden Press, St. Louis.

Chinese Pharmacopoeia Commission (2020). *Pharmacopoeia of the People's Republic of China*. Beijing: China medicine science and technology press.

Eguchi, S., and Tamura, M. N. (2016). Evolutionary timescale of monocots determined by the fossilized birth-death model using a large number of fossil records. *Evolution* 70, 1136–1144. doi: 10.1111/evo.12911

Feng, T. H., Jia, Q. J., Meng, X., Chen, X. L., and Liang, Z. S. (2020). Evaluation of genetic diversity and construction of DNA fingerprinting in *Polygonatum* Mill. based on EST-SSR and SRAP molecular markers. *3 Biotech* 10:322. doi: 10.1007/s13205-020-02316-z

Floden, A. J. (2014). New names in *Heteropolygonatum* (Asparagaceae). *Phytotaxa* 188, 218–226. doi: 10.11646/phytotaxa.188.4.4

Floden, A. J. (2015). A new *Polygonatum* (Asparagaceae) endemic to the Truong Son of southern Vietnam. *Phytotaxa* 197, 125–131. doi: 10.11646/phytotaxa.197.2.5

Floden, A. J. (2017). *Molecular phylogenetic Studies of the Genera of Tribe Polygonateae (Asparagaceae: Nolinoideae): Disporopsis, Heteropolygonatum, and Polygonatum*. PhD diss., University of Tennessee. Knoxville: University of Tennessee

Floden, A., and Schilling, E. E. (2018). Using phylogenomics to reconstruct phylogenetic relationships within tribe Polygonateae (Asparagaceae), with a special focus on *Polygonatum*. *Mol. Phylogenet. Evol.* 129, 202–213. doi: 10.1016/j.ympev.2018.08.017

Henriquez, C. L., Abdullah, Ahmed, I., Carlsen, M. M., and Mckain, M. R. (2020). Evolutionary dynamics of chloroplast genomes in subfamily Aroideae (Araceae). *Genomics* 112, 2349–2360. doi: 10.1016/j.ygeno.2020.01.006

Jiao, J. (2018). *A Study on Germplasm Resources of Polygonati Rhizoma*. Doctoral thesis, Northwest A&F University, Xianyang, 1–274.

Jiao, J., Huang, W. L., Bai, Z. Q., Liu, F., Ma, C. D., Liang, Z., et al. (2018a). DNA barcoding for the efficient and accurate identification of medicinal Polygonati Rhizoma in China. *PLoS One* 13:e0201015. doi: 10.1371/journal.pone.0201015

Jiao, J., Jia, X. R., Liu, P., Zhang, Q. M., Liu, F., Pengzhou, Xi, et al. (2018b). Species identification of Polygonati Rhizoma in China by both morphological and molecular marker methods. *C. R. Biol.* 341, 102–110. doi: 10.1016/j.crvi.2017.10.004

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. *Genom. Biol.* 21, 241. doi: 10.1186/s13059-020-02154-5

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

Kim, D. K., Kim, J. S., and Kim, J. H. (2012). The phylogenetic relationships of Asparagales in Korea based on five plastid DNA regions. *J. Plant Biol.* 55, 325–341. doi: 10.1007/s12374-011-0016-4

Kim, J. H., Kim, D. K., Forest, F., Fay, M. F., and Chase, M. W. (2010). Molecular phylogenetics of Ruscaceae sensu lato and related families (Asparagales) based on plastid and nuclear DNA sequences. *Ann. Bot.* 106, 775–790. doi: 10.1093/aob/mcq167

Lewis, S. E., Searle, S., Harris, N., Gibson, M., and Clamp, M. E. (2002). Apollo: a sequence annotation editor. *Genome Biol.* 3, 1-82.14. doi: 10.1186/gb-2002-3-12-research0082

Liang, C. L., Wang, L., Ma, W. S., and Xu, J. (2021). A comparative study of complete chloroplast genome for the genus *Salvia*. *J. Plant Biochem. Biot.* 30, 117–125. doi: 10.1007/s13562-020-00575-8

Liang, H., Zhang, Y., Deng, J. B., Gao, G., and Yang, R. W. (2020). The complete chloroplast genome sequences of 14 *Curcuma* species: insights into genome evolution and phylogenetic relationships within Zingiberales. *Front. Genet.* 11:802. doi: 10.3389/fgene.2020.0080

Liu, Q., Li, X. Y., Li, M. Z., Xu, W. K., Schwarzacher, T., Heslop-Harrison, J. S., et al. (2020). Comparative chloroplast genome analyses of *Avena*: insights into evolutionary dynamics and phylogeny. *BMC Plant Biol.* 20:406. doi: 10.1186/s12870-020-02621-y

Lohse, M., Drechsel, O., and Bock, R. (2007). OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* 52, 267–274. doi: 10.1007/s00294-007-0161-y

Mayer, C., Leese, F., and Tollrian, R. (2010). Genome-wide analysis of tandem repeats in *Daphnia pulex* - a comparative approach. *BMC Genom.* 11:277. doi: 10.1186/1471-2164-11-277

Meng, R., Meng, Y., Yang, Y. P., and Nie, Z. L. (2021). Phylogeny and biogeography of *Maianthemum* (Asparagaceae: Nolinoideae) revisited with emphasis on its divergence pattern in SW China. *Plant Divers.* 43, 93–101. doi: 10.1016/j.pld.2021.02.001

Meng, Y., Nie, Z. L., Deng, T., Wen, J., and Yang, Y. P. (2014). Phylogenetics and evolution of phyllotaxy in the Solomon's seal genus *Polygonatum* (Asparagaceae: Polygonateae). *Bot. J. Linn. Soc.* 176, 435–451. doi: 10.1111/boj.12218

Michael, T., Pascal, L., Tommaso, P., Elena, S. U. E., Axel, F., Bock, R., et al. (2017). GeSeq-versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 47, W6–W11. doi: 10.1093/nar/gkx391

Nguyen, L., Schmidt, H. A., Haeseler, A. V., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Nie, L. P., Cui, Y. X., Chen, X. L., Xu, Z. C., Sun, W., Song, J., et al. (2020). Complete chloroplast genome sequence of the medicinal plant *Arctium lappa* (Burdock). *Genome* 63, 53–60. doi: 10.1139/gen-2019-0070

Nock, C. J., Baten, A., and King, G. J. (2014). Complete chloroplast genome of *Macadamia integrifolia* confirms the position of the Gondwanan early-diverging eudicot family Proteaceae. *BMC Genom.* 9:S13. doi: 10.1186/1471-2164-15-S9-S13

Puttick, M. N. (2019). MCMCtreeR: functions to prepare MCMCtree analyses and visualize posterior ages on trees. *Bioinformatics* 35, 5321–5322. doi: 10.1093/bioinformatics/btz554

Rudall, P. J., Rudall, J. G., and Chase, M. W. (2000). Systematics of Ruscaceae/Convallariaceae: a combined morphological and molecular investigation. *Bot. J. Linn. Soc.* 134, 73–92. doi: 10.1006/bojl.2000.0365

Shahzadi, I., Abdullah, Mehmood, F., Ali, Z., Ahmed, I., Mirza, B., et al. (2020). Chloroplast genome sequences of *Artemisia maritima* and *Artemisia absinthium*: comparative analyses, mutational hotspots in genus *Artemisia* and phylogeny in family Asteraceae. *Genomics* 112, 1454–1463. doi: 10.1016/j.ygeno.2019.08.016

Shi, L. C., Chen, H. M., Jiang, M., Wang, L. Q., Wu, X., Huang, L., et al. (2019). CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* 47, W65–W73. doi: 10.1093/nar/gkz345

Si, J. P., and Zhu, A. G. (2021). Polygonati Rhizoma-a new high quality crop with great potential and not occupying farmland. *Sci. Sin. Vitae* 51, 1477–1484.

Stefan, K., Choudhuri, J. V., Enno, O., Chris, S., Jens, S., Giegerich, R., et al. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 29, 4633–4642. doi: 10.1093/nar/29.22.4633

Sudhir, K., Glen, S., Li, M., Christina, K., and Koichiro, T. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Tamura, M. N. (1993). Biosystematic studies on the genus *Polygonatum* (Liliaceae) III. Morphology of staminal filaments and karyology of eleven *Eurasian* species. *Bot. Jahrb. Syst.* 115, 1–26.

Tamura, M. N., Schwarzbach, A. E., Kruse, S., and Reski, R. (1997). Biosystematic studies on the genus *Polygonatum* (Convallariaceae) IV. molecular phylogenetic analysis based on restriction site mapping of the chloroplast gene *trnK*. *Feddes Rep.* 108, 159–168. doi: 10.1002/fedr.19971080306

Tamura, M. N., Yamada, M., Fuse, S., and Hotta, M. (2013). Molecular phylogeny and axonomy of the genus *Disporum* (Colchicaceae). *Acta Phytotaxon. Geobot.* 64, 137–147. doi: 10.18942/apg.KJ00008918923

Tang, Y. C. (1978). "*Polygonatum* Mill," in *Flora Reipublicae Popularis Sinicae*, eds F. T. Wang and T. Tang (Beijing China: Science Press).

Wang, J. J., Yang, Y. P., Hang, S., Wen, J., Tao, D., Nie, Z. L., et al. (2016). The biogeographic South-North divide of *Polygonatum* (Asparagaceae Tribe Polygonateae) within Eastern Asia and its recent dispersals in the Northern Hemisphere. *PLoS One.* 11:e166134. doi: 10.1371/journal.pone.0166134

Wang, W. Q., Messing, J., and Badger, J. H. (2011). High-throughput sequencing of three *Lemnoideae* (duckweeds) chloroplast genomes from total DNA. *PLoS One.* 6:e24670. doi: 10.1371/journal.pone.0024670

Wang, Y., Liu, X. Q., Su, H., Yin, S. K., Han, C. X., Hao, D., et al. (2019). The regulatory mechanism of chilling-induced dormancy transition from endo-dormancy to non-dormancy in *Polygonatum kingianum* Coll. et Hemsl rhizome bud. *Plant Mol. Biol.* 99, 205–217. doi: 10.1007/s11103-018-0812-z

Wang, Z. W., Du, F. Q., Zhang, L., Wang, S. M., and Wei, S. H. (2019). Advances in classification and identification of the plants of Polygonati Rhizoma and its adulterants. *North Hortic.* 24, 130–136. doi: 10.11937/bfyy.20191881

Wick, R. R., Schultz, M. B., Justin, Z., and Holt, K. E. (2015). Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* 31, 3350–3352. doi: 10.1093/bioinformatics/btv383

Wu, S. A., Lu, H. L., Yang, J., Rao, G. Y., You, R. L., and Zhong, Y. (2000). Molecular systematic studies on the tribe Polygonateae (s.l.) in China based on RFLPs data of PCR-amplified chloroplast DNA fragments. *Acta Phytotaxon. Sin.* 38, 97–110.

Xia, M. Q., Liu, Y., Liu, J. J., Chen, D. H., Shi, Y., Chen, Z., et al. (2022). Out of the Himalaya-Hengduan Mountains: phylogenomics, biogeography and diversification of *Polygonatum* Mill. (Asparagaceae) in the Northern Hemisphere. *Mol. Phylogenet. Evol.* 169:107431. doi: 10.1016/j.ympev.2022.107431

Xue, T., Gadagkar, S. R., Albright, T. P., Yang, X., Li, J., Xia, C., et al. (2021). Prioritizing conservation of biodiversity in an alpine region: distribution pattern and conservation status of seed plants in the Qinghai-Tibetan plateau. *Glob. Ecol. Conserv.* 32:e01885. doi: 10.1016/j.gecco.2021.e01885

Yang, P., Zhou, H., Xin, T. Y., Ma, S. J., Duan, B. Z., Yao, H., et al. (2015). Identification study of DNA Barcode sequences in the medicinal plants of *Polygonatum*. *World Chin. Med.* 10, 1173–1176.

Yu, X. Y., Zuo, L. H., Lu, D. D., Lu, B., Yang, M. S., Wang, J., et al. (2019). Comparative analysis of chloroplast genomes of five *Robinia* species: genome comparative and evolution analysis. *Gene* 689, 141–151. doi: 10.1016/j.gene.2018.12.023

Yu, Y., Harris, A. J., Blair, C., and He, X. (2015). RASP (Reconstruct Ancestral State in Phylogenies)*: a* tool for historical biogeography. *Mol. Phylogenet. Evol.* 87, 46–49. doi: 10.1016/j.ympev.2015.03.008

Zachos, J., Pagani, M., Sloan, L., Thomas, E., and Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292, 686–693. doi: 10.1126/science.1059412

Zhao, L. H., Zhou, S. D., and He, X. J. (2019). A phylogenetic study of Chinese *Polygonatum* (Polygonateae, Asparagaceae). *Nord. J. Bot.* 37:e02019. doi: 10.1111/njb.02019

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 01 August 2022
DOI 10.3389/fpls.2022.919294

OPEN ACCESS

EDITED BY
Chunnian He,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

REVIEWED BY
Minhui Li,
Baotou Medical College, China
Jiushi Liu,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

*CORRESPONDENCE
Yaping Kou
kouyaping@caas.cn
Lanping Guo
glp01@126.com
Lanyue Zhang
zhanglanyue@gdut.edu.cn

SPECIALTY SECTION
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

RECEIVED 13 April 2022
ACCEPTED 04 July 2022
PUBLISHED 01 August 2022

CITATION
Sun J, Sun P, Kang C, Zhang L, Guo L and Kou Y (2022) Chemical composition and biological activities of essential oils from six lamiaceae folk medicinal plants.
Front. Plant Sci. 13:919294.
doi: 10.3389/fpls.2022.919294

Chemical composition and biological activities of essential oils from six lamiaceae folk medicinal plants

Jiahui Sun[1], Peipei Sun[2], Chuanzhi Kang[1], Lanyue Zhang[3]*, Lanping Guo[1]* and Yaping Kou[4,5]*

[1]National Resources Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, China, [2]Institute of Food Science and Technology, Chinese Academy of Agriculture Sciences, Beijing, China, [3]Guangdong Provincial Key Laboratory of Plant Resources Biorefinery, School of Biomedical and Pharmaceutical Sciences, Guangdong University of Technology, Guangzhou, China, [4]National Center of China for Flowers Improvement, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China, [5]Key Laboratory of Biology and Genetic Improvement of Flower Crops (North China), Ministry of Agriculture and Rural Affairs, Beijing, China

Essential oils have attracted wide attention in recent years due to their extensive applications in natural functional ingredients, pharmaceutical preparations, biomedical products, and the cosmetics industry. In this study, the chemical compositions and biological activities of essential oils extracted from six Lamiaceae herbs, including *Pogostemon cablin* (Blanco) Benth. (PCEO), *Perilla frutescens* (L.) Britton (PFEO), *Salvia japonica* Thunb. (SJEO), *Rosmarinus officinalis* L. (ROEO), *Lavandula angustifolia* Mill. (LAEO), and *Agastache rugosa* (Fisch. & C. A. Mey.) Kuntze (AREO), were determined and analyzed. A total of 167 components were identified from the six essential oils by GC-MS analysis, with 35, 24, 47, 46, 54, and 37 components in PCEO, PFEO, SJEO, ROEO, LAEO, and AREO, respectively. Hierarchical cluster analysis of chemical compositions showed that the composition of the six essential oils was significantly different in content, and they were clearly divided into six classes. However, all of these six essential oils exhibited promising anti-inflammatory activity by inhibiting the expression of interleukin-1, interleukin-6, tumor necrosis factor-α, and cyclooxygenase-2 in rats with adjuvant arthritis, among which PFEO had the best performance. In addition, the six essential oils displayed significant cytotoxicity on B16 (IC_{50} = 86.91–228.91 μg/mL) and LNCaP cell lines (IC_{50} = 116.4–189.63 μg/mL). Meanwhile, all of them presented satisfactory antioxidant activity (IC_{50} = 4.88–13.89 μg/mL) compared with Trolox C (IC_{50} = 13.83 μg/mL), and SJEO (IC_{50} = 7.93 μg/mL) served as an optimal candidate natural antioxidant by DPPH assay. Taken together, these results indicate that the six Lamiaceae essential oils manifest excellent and diverse biological activities, enabling them to be used as perfect natural functional ingredients in antioxidant, antitumor, or anti-arthritic drugs. This study provides more references for pharmaphylogeny research and drug discovery from folk medicinal plants.

KEYWORDS

Lamiaceae herb, essential oil, chemical composition, folk medicinal plants, biological activities

Introduction

The Lamiaceae family contains 236 genera and about 7,173 species, almost cosmopolitan, except for the coldest regions of high latitude. Seven diversity centers were recognized: (1) Mediterranean and SW Central Asia; (2) Africa south of the Sahel and Madagascar; (3) China; (4) Australia; (5) South America; (6) Northern America and Mexico; (7) Indomalesian region (SE Asia). Based on their distribution, these species-rich genera fall generally into two groups: those mostly tropical in origin, including Vitex (Viticoideae), Clerodendrum (Ajugoideae), Callicarpa (Incertae sedis), Ocimum, Hyptis (Nepetoideae), and those which probably had a temperate origin, but often extend into the montane tropics, such as Ajuga, (Ajugoideae), Scutellaria (Scutellarioideae), Stachys (Lamioideae), Salvia, Clinopodium, Mentha, (Nepetoideae). In addition, Teucrium (Ajugoideae) has a distinct distribution in the southern hemisphere, but is most species-rich in the north, spreading down to the mountains of tropical Africa. The Lamiaceae plants contain many aromatic and medicinal plants that are widely used in traditional and modern medicine, food, and cosmetics industries (Vukovic et al., 2009; Nieto, 2017; Borges et al., 2019). In traditional and modern medicine, some Lamiaceae species, such as *Perilla frutescens* (L.) Britton, *Pogostemon cablin* (Blanco) Benth., *Rosmarinus officinalis* L., *Lavandula angustifolia* Mill., and *Agastache rugosa* (Fisch. & C. A. Mey.) Kuntze, are pervasively applied to dispel fever, expel superficial evils, eliminate stasis, induce diuresis, promote blood circulation, and reduce edema (Guo et al., 2016; Luo et al., 2019). *R. officinalis*. has been used as an analgesic, antispasmodic, and antidepressant to cure intercostal neuralgia, headaches, migraine, insomnia emotional upset, and depression in folk medicine (Ghasemzadeh Rahbardar and Hosseinzadeh, 2020). For a long time, the biological activities of extracts from these plant species have been studied, such as the antitumor, antioxidant, antimicrobial, and anti-inflammatory activities (Nieto, 2017; Guo et al., 2019; Karpinski, 2020).

Essential oil, an important category of plant extracts, has a multidirectional action mode and a variety of biological activities (Wojtunik-Kulesza et al., 2019). Essential oils can disrupt the cell and cell membrane via a permeabilization process. The lipophilic compounds of essential oils can pass through the cell wall; damage polysaccharides, fatty acids, and phospholipids; change the permeability for H^+ and K^+ cations to affect cellular pH; and damage organelles and disintegrate mitochondrial membrane (Karpinski, 2020). What is more, essential oils inhibit the biosynthesis of fungal DNA, RNA, proteins, and so on. They are widely applied in cosmetic additives, natural functional food, pharmaceutical preparations, and biomedical products (Nieto, 2017; Zhang et al., 2017b, 2020a). Specifically, essential oils from medicinal plants have attracted increasing attention in recent years for their multifaceted biological activities and diverse chemical compositions (Santos and Rao, 2000; El-Sayed et al., 2014; Xue, 2016). Lamiaceae family plants rich in essential oils have significant values in natural medicine, pharmacology, cosmetology, and aromatherapy. Some Lamiaceae species that are used in traditional medicine have been employed in the characterization of their essential oils, such as bioactivities and phytochemical composition. For example, essential oils of *P. cablin* (PCEO), *P. frutescens* (PFEO), *R. officinalis* (ROEO), and *L. angustifolia* (LAEO) are mainly composed of patchouli alcohol, linalool, α-terpineol, β-pinene, DL-menthol, and isobornyl acetate, which show strong anti-inflammatory activity by inhibiting the expression of interleukin 6 (IL-6), cyclooxygenase 2 (COX-2), tumor necrosis factor α (TNF-α), and nuclear factor-kappa B (NF-kB) in tetradecanoylphorbol acetate (TPA)-induced inflammation models (Luo et al., 2019; Zhang et al., 2020b). Moreover, these essential oils also demonstrate high antioxidant, antibacterial, and antifungal activities (Luo et al., 2019). The essential oil of *A. rugosa* (AREO), mainly composed of methyleugenol, estragole, and eugenol, exhibits strong pesticide activity against *Meloidogyne incognita*, with a LC_{50} value of 47.3 μg/mL (Li et al., 2013). ROEO also shows suppression of the lipopolysaccharide (LPS)-induced inflammation by inhibiting the expression of COX-2 and inducible nitric oxide synthase (iNOS) and blocking the production of TNF-α (Yu et al., 2013). Lavender essential oils have been used cosmetically and therapeutically for centuries, and their biological activities have been extensively studied (Cavanagh and Wilkinson, 2002). Borges et al. (2019) indicated that ROEO possesses strong anti-inflammatory activity and can be used as a remedy for inflammatory diseases (Borges et al., 2019). Though the essential oil extracted from *Salvia japonica* Thunb. (SJEO) has been widely used, little is known about its biological activity. Herb essential oils exert their diverse biological activities by acting on various pathways using different chemical components. The composition and bioactivity of essential oils extracted from Lamiaceae plants have been analyzed in many studies; however, their anti-inflammatory, antioxidant, antitumor, and anti-arthritis activities need to be systematically explored from multiple aspects (Nikolić et al., 2014; Vyry Wouatsa et al., 2014; Park et al., 2016; Mamadalieva et al., 2017, 2019; Mouahid et al., 2017; Borges et al., 2019; Karpinski, 2020).

The chemical composition of plant essential oil is influenced by numerous factors, such as the growing environment, harvest time, and plant organ used for essential oil extraction. Therefore, it is necessary to determine the phytochemical composition of the essential oil before carrying out further studies on their

Abbreviations: AREO, the essential oil of *Agastache rugosa* (Fisch. & C. A. Mey.) Kuntze; LAEO, the essential oil of *Lavandula angustifolia* Mill.; PCEO, the essential oil of *Pogostemon cablin* (Blanco) Benth.; PFEO, the essential oil of *Perilla frutescens* (L.) Britton; ROEO, the essential oil of *Rosmarinus officinalis* L.; SJEO, the essential oil of *Salvia japonica* Thunb.

bioactivities. The current study aimed to elucidate the biological activity through the determination of the chemical composition of the essential oils extracted from six Lamiaceae plants with gas chromatography-mass spectrometry (GC-MS). The diversity of chemical components was analyzed by hierarchical cluster analysis in six essential oils from six Lamiaceae folk medicinal plants. Diverse biological activities, including anti-inflammatory, antitumor, and antioxidant activities, were evaluated through corresponding models. The model of complete Freund's adjuvant (CFA)-induced rheumatoid arthritis was used to evaluate the anti-inflammatory activity and related mechanisms of the six essential oils, and the LNCaP and B16 cell lines were used to estimate the antitumor activity. The 1,1-diphenyl-2-picrylhydrazyl (DPPH) free radical method was carried out to determine free radical scavenging capacity of these essential oils. Our study provides additional data to support the use of essential oils from Lamiaceae plants as a drug treatment.

Materials and methods

Plant materials and chemicals

The local name, storage locations, and collection dates of *A. rugosa*, *L. angustifolia*, *P. frutescens*, *P. cablin*, *R. officinalis*, and *S. japonica* samples are shown in Table 1. The leaves of *A. rugosa*, *P. cablin*, *P. frutescens*, *R. officinalis* were used for essential oil extraction, while the aerial parts of *L. angustifolia* and *S. japonica* were used for essential oils extractions. All plant samples were confirmed by Professor Nian Liu of Zhongkai University of Agriculture and Engineering (Guangzhou, China). All chemicals used in this study were purchased from Aladdin Reagent Inc. (Shanghai, China).

Essential oil extraction

The steam distillation method was used to extract essential oils from the six plant samples as previously described (Zhang et al., 2020b). The plants were cleaned, ground, passed through a 0.45-mm sieve, and distilled with a steam distillation device for 3.5 h (Zhang et al., 2017b). The isolated essential oils were dried as previously described in a former research and stored in individual brown glass bottles at 4 °C until use (Xiang et al., 2017).

GC-MS analysis

The phytochemical compositions of PCEO, PFEO, SJEO, ROEO, LAEO, and AREO were identified according to previous methods (Xiang et al., 2017; Zhang et al., 2017b) using a GC-MS system with a DB-5MS capillary column (0.25 mm × 30 m, i.d. 0.25 μm) (Agilent, Santa Clara, CA, USA). The carrier gas was Helium (He) at a flow rate of 1 mL/min. The initial temperature was set as 40°C for 1 min, then increased to 280°C by 3°C/min, and held at 280°C for 5 min. The split ratio was set as 100:1. For MS conditions, the ionization conditions were as follows: pressure of 50 kPa, electron energy of 70 eV, and ion source temperature of 200°C.

Each component of essential oils was determined based on the retention index (RI), which was calculated using a series of n-alkanes (C_6-C_{40}) (Xue et al., 2016). Additionally, the mass spectrum of each compound was searched against the NIST Standard Reference Database (NIST Chemistry WebBook, 2014, over 40,000 compounds in this database) and databases published elsewhere (Zhang et al., 2017a). The total ionization chromatography (TIC) was obtained and used for determining the contents of each compound (Supplementary Figure S1).

Animals

The animal experiments were carried out following ethical guidelines of the Laboratory Animal Center of Sun Yat-sen University. Male rats (6–8 weeks old, 210 ± 30 g body weight) were purchased from Sun Yat-sen University and raised under normal conditions (25 ± 2°C, 12/12 h light/dark cycle). Food and water were fed as required during the experiment.

Experimental treatments of adjuvant arthritis

According to our previous study, the rats were acclimatized for 7 days and then divided into 10 groups, with 10 rats in each group: (a) normal control (Con) group, (b) model (CFA) group, (c) negative control (NC) group, (d) positive control (PC) group, (e) PCEO group, (f) PFEO group, (g) ROEO group, (h) SJEO group, (i) LAEO group, and (j) AREO group (Zhang et al., 2020a).

Rats in the Con. group were not given any treatments. In the model group, CFA (0.1 mL) was subcutaneously injected into each rat after routine sterilization from Day 8 to Day 21 to induce arthritis. Then these rats were raised under normal conditions until use. Tween 80 was given to the rats in the NC group, while ibuprofen (100 mg/kg, dissolved in Tween 80) was given to those in the PC group from Day 8 to Day 20 (Khayyal et al., 2005). Rats in PCEO, PFEO, ROEO, SJEO, LAEO, and AREO groups were treated with corresponding essential oils (100 mg/kg, dissolved in Tween 80) from Day 8 till the end of the experiment. The arthritis score was recorded from Day 8 to Day 20 for all groups. The 5-point method was used to assess and grade the severity of the swelling, erythema, or stiffness in the paw: 0 = no signs of illness; 1 = mild swelling and erythema in the ankle/wrist; 2 = swelling and erythema in the

TABLE 1 Latin name, local name, voucher specimen number, and collection time of six Lamiaceae plants.

Latin name	Local name	Voucher number	Collection time	Storage location
Pogostemon cablin (Blanco) Benth.	Guanghuoxiang	2018-100C	2018.10	Institute of Natural Medicine & Green Chemical, School of Chemical Engineering and Light Industry, Guangdong University of Technology
Salvia japonica Thunb.	Shuweicao	2018-102C	2018.08	
Perilla frutescens (L.) Britton	Zisuye	2018-106C	2018.07	
Rosmarinus officinalis L.	Midiexiang	2018-103C	2018.09	
Lavandula angustifolia Mill.	Xunyicao	2018-101C	2018.12	
Agastache rugosa (Fisch. & C. A. Mey.) Kuntze	Huoxiang	2018-105C	2018.11	

ankle/wrist; 3 = severe swelling and erythema in the ankle/wrist; and 4 = severe illness in the paw or front leg. Both the hind feet were graded, and the total score was not allowed to exceed 8 (Funk et al., 2010). The rats were then sacrificed and their ankle joints were sampled and stored in 4% (v/v) paraformaldehyde for subsequent histological analysis.

Histological analysis and immunohistochemical staining

As described previously (Zhang et al., 2020a), the ankle joints were paraffin-embedded and the sections were observed using light microscopy (Olympus IX71, OLYMPUS, Japan). The IL-1, IL-6, COX-2, and TNF-α antibodies, all diluted at a 1:200 ratio, were used for immunohistochemical staining. The number of positive cells was counted with ImageJ using photos taken with a fluorescence microscope (NIH, Bethesda, MD, USA).

For immunohistochemical analysis, the sections were incubated overnight with IL-1β (dilution 1:200), IL-6 (dilution 1:200), COX-2 (dilution 1:200), and TNF-α (dilution 1:200) antibodies at 4°C and then treated with a secondary antibody and alkaline phosphatase-labeled streptavidin (1:200) at 25°C for up to 1 h. Sections were developed with 3,30-diaminobenzidine (DAB) solution. Image analysis software (Image-Pro Plus) was used to count the number of positive cells, and these were observed using a fluorescence microscope (NIH, Bethesda, MD) (Zhang et al., 2020b).

Determination of antitumor activity

Anittumor activity of essential oil was investigated via human prostate cancer cell model LNCap and mouse B16 melanoma cell lines *in vitro*. In PCEO, PFEO, ROEO, SJEO, LAEO, and AREO groups, the cytotoxicity of essential oils on LNCaP and B16 cells treated with essential oils was assessed using MTT [3 - (4,5 - dimethyl - 2 - thiazolyl) - 2,5 - diphenyltetrazolium bromide] values described in Zhang et al. (2020a). The B16 and LNCaP cells were cultured in DMEM and RPMI1640, respectively, supplemented with 10% FBS, 2 mM glutamine, 100 mg/mL streptomycin, and 100 U/mL penicillin. The cells were maintained in a humidified 5% CO_2 incubator at 37 °C and were subcultured every 3–4 days to maintain logarithmic growth and allowed to grow for 24 h before the treatments were applied. The cells were then treated with different concentrations of the essential oil (Table 3), and then the absorbance at 570 nm was read on a microplate reader. The IC_{50} value of MTT assays is defined as essential oils concentration resulting in a 50% reduction of absorbance (Kanipandian and Thirumurugan, 2014). Hydroquinone and Paclitaxel were used as positive controls for B16 and LNCaP cells, respectively (Zhang et al., 2017c; Rodboon et al., 2020).

Determination of antioxidant activity

The antioxidant activity of the six essential oils was evaluated by DPPH free radical scavenging capacity (Xiang et al., 2017; Zhang et al., 2020a). DPPH solution (67 μg/mL) was mixed with each of the essential oils at various concentrations and incubated at 25°C for 30 min in the dark. The absorbance was then read at 517 nm. The scavenging percentage was calculated as follows:

$$\text{Scavenging percentage}(\%) = [1 - (A_{sample} - -A_{\text{blank}})/A_{control}] \times 100$$

Statistical analysis

One-way analysis of variance (ANOVA) was used to test statistical significance, and the result was considered significant at $P \leq 0.05$. Data were presented as mean ± standard

TABLE 2 Chemical compositions of six Lamiaceae plants.

No	Compounds [i]	RI2 [ii]	RI3 [iii]	Ref	Cas	Relative content (%) [iv]					
						PCEO	PFEO	SJEO	ROEO	LAEO	AREO
1	Menthol	1181		b	015356-70-4	0.22	-	-	-	-	-
2	Terpenol	1200	1010	c	000098-55-5	0.28	-	-	3.36	-	-
3	Cinnamaldehyde	1283	1277		014371-10-9	1.41	-	-	0.99	0.18	2.05
4	p-Anethole	1291	1289	d	000104-46-1	0.49	-	-	3.73	0.48	0.92
5	p-Allylguaiacol/Eugenol	1370	1373	b	000097-53-0	5.64	-	-	-	-	8.29
6	Copaene	1394	1394	b	003856-25-5	0.68	0.18	-	-	0.09	1.34
7	β-Patchoulene	1405	1406	b	000514-51-2	0.19	-	-	-	-	0.40
8	β-Elemene	1407	1382	a	000515-13-9	0.20	-	0.24	-	-	0.33
9	β-Caryophyllene	1443	1441	b	000087-44-5	1.17	7.76	1.04	0.68	0.81	2.12
10	α-Guaiaene	1457	1455		003691-12-1	1.04	-	-	-	-	1.62
11	Seychellene	1472	1375	c	020085-93-2	0.86	-	-	-	-	1.34
12	Humulene	1478	1364	a	006753-98-6	0.22	0.92	0.87	-	0.13	-
13	α-Patchoulene	1485	1288	c	000560-32-7	0.47	-	-	-	-	0.76
14	α-Elemene	1488	1348	c	005951-67-7	0.12	-	-	-	-	-
15	γ-Patchoulene	1492	1424	a	000508-55-4	0.14	-	-	-	-	-
16	α-Selinene	1511	1587	c	000473-13-2	0.14	-	0.25	-	-	-
17	α-Gurgujene	1519	1519		000489-40-7	0.75	-	-	-	-	1.09
18	δ-Guaijene	1527	1525		003691-11-0	2.39	-	0.16	-	-	3.48
19	δ-Cadinene, (+)-	1541	1547		000483-76-1	0.35	0.28	-	0.13	-	0.60
20	Calamenene A	1544	1543		000483-77-2	0.15	-	-	-	-	-
21	Caryophyllene oxide	1613	1613	b	001139-30-6	1.22	-	0.21	0.72	4.31	1.94
22	Epicurzerenone	1629	1623		020085-85-2	0.16	-	-	0.14	-	-
23	α-Humulene epoxide II	1640	1615		019888-34-7	0.18	0.16	0.12	0.53	-	0.23
24	Viridiflorol	1654	1620		000552-02-3	1.29	-	-	0.91	-	1.02
25	Viridflorene	1672	1656		021747-46-6	0.40	-	-	-	-	-
26	Pogostole	1688	1655	b	021698-41-9	3.96	-	-	-	-	3.94
27	Patchouli alcohol	1707	1587	a	005986-55-0	43.04	-	0.12	0.26	0.20	45.70
28	Rotundone	1736	1722		018374-76-0	1.59	-	-	-	-	1.53
29	Pogostone	1743	1641	a	023800-56-8	14.35	-	-	-	-	11.92
30	Longifolenealdehyde	1755	1668		019890-84-7	0.12	-	-	-	-	-
31	Cycloisosativene	1766	1530		022469-52-9	0.31	-	-	-	-	-
32	Perhydrofarnesyl acetone	1850	1848		000502-69-2	0.12	-	-	-	0.23	0.13
33	Corymbolone	1922	1899		097094-19-4	0.31	-	-	-	-	0.20
34	Palmitic acid	1962	1961		000057-10-3	0.16	-	-	-	-	-
35	Linoelaidic acid	2142			000506-21-8	0.16	-	-	-	-	-
36	L-α-Pinene	939	945	e	007785-26-4	-	0.50	5.97	-	-	-
37	Sabinene	979	977		003387-41-5	-	0.23	0.96	-	-	-
38	(-)-beta-Pinene	985	1010	e	018172-67-3	-	0.15	-	-	-	-
39	D-Limonened	1035	1100	e	005989-27-5	-	0.69	4.13	-	-	-
40	cis-Linalool oxide	1079	1078		005989-33-3	-	0.24	-	-	5.69	-
41	Linalool	1111	1230	e	000078-70-6	-	67.65	8.89	1.56	29.84	-
42	1,2-Dihydrolinalool	1137	1125	b	018479-51-1	-	0.93	-	-	-	-
43	3,7-Octadiene-2,6-diol, 2,6-dimethyl-, (3E)-	1195	1191		051276-34-7	-	0.15	-	-	-	-
44	Elsholtzione	1208	1454		000488-05-1	-	0.22	-	-	-	-

(Continued)

TABLE 2 Continued

No	Compounds [i]	RI2 [ii]	RI3 [iii]	Ref	Cas	Relative content (%) [iv]					
						PCEO	PFEO	SJEO	ROEO	LAEO	AREO
45	trans-Shisool	1282	1248		022451-48-5	-	0.24	-	-	-	-
46	perrilaldehyde	1290	1632	e	002111-75-3	-	1.45	-	-	-	-
47	α-trans-Bergamoptene	1499	1490		013474-59-4	-	1.60	-	-	-	-
48	β-Copaene	1504	1460	b	018252-44-3	-	0.32	-	-	-	-
49	β-Maaliene	1518	1418		000489-29-2	-	0.16	-	-	-	-
50	Myristicin	1536	1382	g	000607-91-0	-	0.29	-	-	-	-
51	Elemicin	1562	1560	b	000487-11-6	-	0.15	-	-	-	-
52	Nerolidol	1572	1564	f	000142-50-7	-	0.27	-	-	-	-
53	β-Asarone	1687	1641	b	005273-86-9	-	0.57	-	-	-	-
54	Diisobutyl phthalate	1881	1873		000084-69-5	-	0.90	-	-	-	-
55	Isocitronellene	919	917		085006-04-8	-	-	0.21	-	-	-
56	α-Thujene	930	931	b	002867-05-2	-	-	0.21	-	-	-
57	Camphene	956	957	b	000079-92-5	-	-	0.83	-	-	-
58	β-Pinene	984	1023	k	000127-91-3	-	-	1.01	-	-	-
59	β-Myrcene	992	988	b	000123-35-3	-	-	0.61	-	-	-
60	o-Cymene	1031	1029	b	000527-84-4	-	-	1.31	-	-	-
61	Eucalyptol	1040	1039	b	000470-82-6	-	-	6.15	6.05	0.40	0.11
62	α-Terpinolene	1095	10.94	b	000586-62-9	-	-	0.16	-	-	-
63	(+)-2-Bornanone	1160	1144	k	000464-49-3	-	-	9.54	15.10	-	-
64	Benzyl acetate	1172	1168	b	000140-11-4	-	-	5.73	-	-	-
65	Isononyl acetate	1177	1180		058430-94-7	-	-	0.42	-	-	-
66	Borneol	1180	1178	b	000507-70-0	-	-	3.20	14.58	1.84	-
67	Terpinen-4-ol	1189	1188		000562-74-3	-	-	0.84	-	1.03	-
68	L-α-Terpineol	1201	1514	k	010482-56-1	-	-	2.33	4.90	4.28	-
69	γ-Terpineol	1207	1217	b	000586-81-2	-	-	0.32	-	11.11	-
70	Linalyl acetate	1259	1262	b	000115-95-7	-	-	3.26	-	0.12	-
71	1,4-Dimethyl-4-ethenyl-cyclohexene	1290	949		001743-61-9	-	-	1.22	-	-	-
72	Acetic acid, bornyl ester	1298			092618-89-8	-	-	1.47	-	-	-
73	Safrole	1302	1288	b	000094-59-7	-	-	2.65	-	-	-
74	γ-Terpinene	1309	1269	b	000099-85-4	-	-	0.23	-	-	-
75	δ-terpinyl acetate	1327	1316		093836-50-1	-	-	0.22	-	-	-
76	Triacetin	1353	1344	h	000102-76-1	-	-	6.84	-	-	-
77	Terpenyl acetate	1362	1367	j	000080-26-2	-	-	10.52	-	-	-
78	α-Terpinene	1364	1243	k	000099-86-5	-	-	2.57	-	-	-
79	α-Cubebene	1394	1382	b	017699-14-8	-	-	0.28	-	-	-
80	Methyl eugenol	1410	1410		000093-15-2	-	-	0.13	-	-	-
81	Longifolene	1433	1427	b	000475-20-7	-	-	0.67	-	-	-
82	Santalene	1436	1431		000512-61-8	-	-	0.50	-	0.17	-
83	Citroviol	1439	1413		000128-51-8	-	-	3.79	-	-	-
84	Coumarin	1461	1458	b	000091-64-5	-	-	0.55	-	0.34	-
85	Germacrene D	1505	1508	b	023986-74-5	-	-	0.26	-	-	-
86	β-Eudesmene	1512	1507		017066-67-0	-	-	0.26	-	-	-
87	γ-Cadinene	1536	1534	b	039029-41-9	-	-	0.06	-	-	-
88	trans-β-Nerolidol	1573	1567		040716-66-3	-	-	0.88	-	-	-

(Continued)

TABLE 2 Continued

No	Compounds [i]	RI2 [ii]	RI3 [iii]	Ref	Cas	Relative content (%) [iv]					
						PCEO	PFEO	SJEO	ROEO	LAEO	AREO
89	(-)-Spathulenol	1605	1619		077171-55-2	-	-	0.25	-	-	-
90	α-Curcumene	1555	1499	h	000644-30-4	-	-	-	-	0.10	-
91	trans-Sabinenhydrate	1106	1106		017699-16-0	-	-	-	0.11	-	-
92	Fenchol	1124	1125		001632-73-1	-	-	-	0.24	-	-
93	cis-p-Menth-2-ene-1-ol	1131	1129		029803-82-5	-	-	-	0.46	-	-
94	Chrysanthenone	1135	1126		000473-06-3	-	-	-	0.14	-	-
95	Camphene hydrate	1163	1150		000465-31-6	-	-	-	0.25	-	-
96	dl-Isopulegol	1168	1167	b	050373-36-9	-	-	-	0.20	-	-
97	L-4-terpineneol	1189	1182		020126-76-5	-	-	-	2.77	-	-
98	Trimethylphenylsilane	1196			000768-32-1	-	-	-	0.26	-	-
99	Estragole	1207	1201		000140-67-0	-	-	-	0.18	-	-
100	Myrtenol	1211	1201	b	000515-00-4	-	-	-	0.54	-	-
101	S-cis-Sabinol	1217	1179	b	003310-02-9	-	-	-	0.25	-	-
102	Levo verbenone	1228	1204	i	001196-01-6	-	-	-	15.29	-	-
103	Citronellol	1231	1230		000106-22-9	-	-	-	0.78	-	-
104	Pulegone	1252	1250	b	000089-82-7	-	-	-	0.19	-	-
105	Thujone	1253	1117		000546-80-5	-	-	-	0.65	-	-
106	Geraniol	1258	1255		000106-24-1	-	-	-	0.90	-	-
107	3-Carvomenthenone	1268	1268		000089-81-6	-	-	-	0.43	-	-
108	Biosol	1288	1332		003228-02-2	-	-	-	0.68	-	-
109	Carvacrol	1306	1306		000499-75-2	-	-	-	0.63	-	-
110	Piperitenone	1358	1348		000491-09-8	-	-	-	0.51	-	-
111	Chavibetol	1372	1362	l	000501-19-9	-	-	-	17.72	1.32	-
112	Safranal	1401	1596		000116-26-7	-	-	-	0.30	0.44	-
113	cis-β-Farnesene	1462	1434		028973-97-9	-	-	-	0.11	-	-
114	Aceteugenol	1534	1525		000093-28-7	-	-	-	0.17	-	-
115	Caryophylla-4(12),8(13)-dien-5-beta-ol	1664	1644		019431-80-2	-	-	-	0.26	-	-
116	9,9-Dimethyl-9-silafluorene	1700			013688-68-1	-	-	-	0.17	-	-
117	Isopimara-9(11),15-diene	1954	1920		039702-28-8	-	-	-	0.11	-	-
118	cis-Biformene	2045	2004		005957-33-5	-	-	-	0.15	-	-
119	Dehydroabietan	2095	2057		019407-28-4	-	-	-	0.11	-	-
120	Phytol	2119	2122		000150-86-7	-	-	-	0.11	-	-
121	Osthole	2166	2174		000484-12-8	-	-	-	0.45	0.43	-
122	1-Octen-3-ol	978	972		003391-86-4	-	-	-	-	0.25	-
123	Lavender lactone	1045	1046		001073-11-6	-	-	-	-	0.15	-
124	(E)-Linalool furanoxide	1095	1094		034995-77-2	-	-	-	-	4.10	-
125	Hotrienol	1110	1108		029957-43-5	-	-	-	-	0.88	-
126	(-)-Alcanfor	1158	1146		000464-48-2	-	-	-	-	0.56	-
127	Nerol oxide	1160	1164		001786-08-9	-	-	-	-	0.43	-
128	(+/-)-Lavandulol	1172	1170		058461-27-1	-	-	-	-	3.38	-
129	α-Phellandren-8-ol	1177	1181		001686-20-0	-	-	-	-	0.35	-
130	linalool oxide (III)	1181	1199		039028-58-5	-	-	-	-	0.38	-

(Continued)

TABLE 2 Continued

No	Compounds [i]	RI2 [ii]	RI3 [iii]	Ref	Cas	Relative content (%) [iv]					
						PCEO	PFEO	SJEO	ROEO	LAEO	AREO
131	Hexyl butyrate	1192	1195		002639-63-6	-	-	-	-	0.19	-
132	Cryptone	1199	1184		000500-02-7	-	-	-	-	1.11	-
133	Carveol II	1228	1231		001197-06-4	-	-	-	-	0.21	-
134	p-Cumenol	1231	1247		000099-89-8	-	-	-	-	0.10	-
135	Vernol	1234	1233		000106-25-2	-	-	-	-	0.85	-
136	p-Cumic aldehyde	1253	1249		000122-03-2	-	-	-	-	0.47	-
137	D-Carvone	1255	1225		002244-16-8	-	-	-	-	0.26	-
138	Lavandulyl propionate	1293	1375		059550-34-4	-	-	-	-	6.42	-
139	Bornyl acetate	1297	1286		000076-49-3	-	-	-	-	0.29	-
140	p-Cymen-7-ol	1300	1302		000536-60-7	-	-	-	-	0.43	-
141	Nerol acetate	1366	1347		000141-12-8	-	-	-	-	1.09	-
142	Geranyl acetate	1385	1386		000105-87-3	-	-	-	-	1.78	-
143	2-Caren-4-ol	1492	1181		006617-35-2	-	-	-	-	0.22	-
144	α-Muurolene	1518			031983-22-9	-	-	-	-	0.11	-
145	Cadina-3,9-diene	1541	1529		000523-47-7	-	-	-	-	0.12	-
146	cis-3-Hexenyl benzoate	1584	1550		025152-85-6	-	-	-	-	0.26	-
147	α-Bisabolol	1702	1701		000515-69-5	-	-	-	-	0.12	-
148	cis-14-nor-Muurol-5-en-4-one	1721	1696		063180-33-6	-	-	-	-	0.16	-
149	Benzyl Benzoate	1787	1789		000120-51-4	-	-	-	-	0.11	-
150	Nerolidol	2040	2030		007212-44-4	-	-	-	-	0.31	-
151	1-Nonadecene	2280	1960		018435-45-5	-	-	-	-	0.18	-
152	Eicosane	2490			000112-95-8	-	-	-	-	0.29	-
153	Pentacosane	2499			000629-99-2	-	-	-	-	0.10	-
154	Menthol	1181	1171		001490-04-6	-	-	-	-	-	0.29
155	Guaia-6,9-diene	1489	1450		036577-33-0	-	-	-	-	-	0.21
156	Patchoulene	1492	1485		001405-16-9	-	-	-	-	-	0.22
157	γ-Muurolene	1495	1483		030021-74-0	-	-	-	-	-	0.18
158	Alloaromadendrene	1511	1495		025246-27-9	-	-	-	-	-	0.27
159	cis-Calamenene	1544	1531		072937-55-4	-	-	-	-	-	0.27
160	Cashmeran	1592	1503		033704-61-9	-	-	-	-	-	1.25
161	3-Ethylphenol	1598			000620-17-7	-	-	-	-	-	1.84
162	α-Isonootkatol	1610			1380573-94-3	-	-	-	-	-	0.92
163	Neointermedeol	1664	1656		005945-72-2	-	-	-	-	-	0.39
164	Eremophilene	1669			010219-75-7	-	-	-	-	-	0.23
165	γ-Gurjunene	1673	1664		022567-17-5	-	-	-	-	-	0.16
166	(E)-2-Hexenal	1750			006728-26-3	-	-	-	-	-	1.21
167	Isoeremophilene	1765	1721		004630-07-3	-	-	-	-	-	0.43
	Total/%					84.29	86.01	92.47	98.76	89.21	98.92

Retention index (RI) and relative content of identified compounds in six essential oils.

[a] Chen et al. (2017); [b] Luo et al. (2019); [c] Zhang et al. (2019); [d] Vieira et al. (2019); [e] Li (2010); [f] Xue et al. (2016); [g] Gao and Zhang (2019); [h] Zhang et al. (2017a); [j] Mouahid et al. (2017); [k] Li et al. (2016); [l] Vyry Wouatsa et al. (2014).

[i] Compound listed in the order of elution from methyl silicone capillary column.

[ii] Retention index relative to n-alkanes (C_6-C_{40}) on the same methyl silicone capillary column.

[iii] Literature indices.

[iv] PFEO, the essential oils of Perilla frutescens (L.) Britton; PCEO, the essential oils of Pogostemon cablin (Blanco) Benth.; SJEO, the essential oils of Salvia japonica Thunb.; ROEO, the essential oils of Rosmarinus officinalis L.; LAEO, the essential oils of Lavandula angustifolia Mill.; AREO, the essential oils of Agastache rugosa (Fisch. & C. A. Mey.) Kuntze.

deviation (SD). Based on the content of each component, hierarchical cluster analysis (HCA) was performed for chemical compositions in the six essential oils using pheatmap package (version 1.0.12).

Results and discussion

Phytochemical compositions of the six essential oils

A total of 167 components were identified from six essential oils (Table 2) using GC-MS. Based on the species and quantity of compositions in each essential oil, the relationship of the six Lamiaceae folk medicinal plants was analyzed via hierarchical cluster. The results showed that the six essential oils were clearly divided into six classes, and the six essential oils have their unique principal components (Figure 1). The proportions of corymbolone, pogostone, patchouli alcohol, pogostole, rotundone, menthol, α-patchoulene, seychellene, α-guaiaene, α-gurgujene, and δ-guaijene showed similarity in PCEO and AREO. The proportions of L-α-terpineol, osthole, and safranal showed similarity in LAEO and ROEO. The proportions of nerolidol and linalool showed similarity in LAEO and PFEO. The proportion of terpinen-4-ol showed similarity in LAEO and SJEO. The proportions of (+)-2-bornanone and eucalyptol showed similarity in SJEO and ROEO. The proportion of humulene showed similarity in SJEO and PFEO. The proportion of α-selinene showed similarity in SJEO and PCEO. The proportion of epicurzerenone showed similarity trend in PCEO and ROEO (Figure 1). β-Caryophyllene was present in all six essential oils, while patchouli alcohol and caryophyllene oxide were found in all essential oils except PFEO. In summary, PCEO has the latest relationship with AREO, followed by PFEO, SJEO, ROEO, and LAEO. This relationship was also partly proved by the biological activities of six essential oils in the following results.

To further investigate the difference of the six essential oils, the major compounds of essential oils were compared in the analysis. The total relative contents of PCEO, PFEO, ROEO, SJEO, LAEO, and AREO were 84.29, 86.01, 98.76, 92.47, 89.21, and 98.92%, respectively. The dominant components of PCEO were patchouli alcohol (43.04%), pogostone (14.35%), and p-allylguaiacol/eugenol (5.638%). The major components of PFEO were linalool (67.65%) and β-caryophyllene (7.7564%). The main components of SJEO were terpenyl acetate (10.52%), (+)-2-bornanone (9.54%), camphor (9.54%), linalool (8.89%), L-α-pinene (5.97%), triacetin (6.84%), eucalyptol (6.15%), plastolin I (5.73%), and D-limonened (4.13%). The principal components of ROEO were chavibetol (17.72%), verbenone (15.29%), camphor (15.10%), borneol (14.58%), eucalyptol (6.05%), and α-terpineol (4.90%). The key components of LAEO were linalool (29.84%), γ-terpineol (11.11%), caryophyllene oxide (4.31%), *cis*-linalool oxide (5.69%), L-α-terpineol (4.28%), (E)-linalool furanoxide (4.10%), and lavandulyl propionate (6.42%). The primary components of AREO were patchouli alcohol (45.70%), pogostone (11.92%), and eugenol (8.29%). Linalool was found in LAEO, PFEO, ROEO, and SJEO, exhibiting the highest relative content in PFEO (67.66%), followed by LAEO (29.84%), SJEO (8.89%), and ROEO (1.56%). ROEO and AREO shared many common components, including patchouli alcohol, pogostone, and eugenol. In POEO and AREO, the relative contents of patchouli alcohol, pogostone, and eugenol were 43.04, 14.35, and 5.64%; and 45.70, 11.92, and 8.29%, respectively (Table 2). Meanwhile, the chemical structures of 15 components are shown in Figure 2 to provide more reference for later researchers, including patchouli alcohol, eugenol,β-caryophyllene, pogostone, L-α-pinene, *cis*-linalool oxide, linalool, eucalyptol, (+)-2-bornanone, benzyl acetate, (+)-borneol, L-α-terpineol, γ-terpineol, triacetin, terpenyl acetate, levo verbenone, chavibetol, (E)-linalool furanoxide, and lavandulyl propionate. All 15 components were picked up according to their content, the and content was more than 4 %.

The chemical composition of essential oil is influenced by various factors. Some components have been previously reported in the essential oils of Lamiaceae species. For example, PFEO is proposed to mainly contain linalool (46.55%) and 2-hexanoylfuran (30.79%); ROEO is composed of α-pinene (45.35%) and D-limonene (18.42%); PCEO mainly contains patchouli alcohol (28.27%), α-bulnesene (18.29%), and α-guaiene (14.53%); LAEO mainly includes isononyl acetate (22.52%), α-pinene (11.47%), and benzyl acetone (10.93%); and SJEO mainly contains *o*-cymene (41.20%), (Z, E)-α-farnesene (10.82%), and γ-muurolene (9.89%) (Luo et al., 2019).

TABLE 3 Antitumor activity of six Lamiaceae plants essential oils.

Essential oil	IC_{50} (μg/ml) [a]	
	B16	LNCaP
PCEOs[b]	$109.32^{ACFc} \pm 9.51$ [d]	$138.24^{ACDE} \pm 13.74$
PFEOs	$144.36^{BE} \pm 7.75$	$127.90^{ACDE} \pm 18.76$
ROEOs	$93.96^{CF} \pm 8.65$	$189.63^{B} \pm 11.93$
SJEOs	$228.91^{D} \pm 13.89$	$129.40^{CDE} \pm 15.78$
LAEOs	$152.69^{E} \pm 23.23$	$116.41^{DE} \pm 14.50$
AREOs	$86.91^{F} \pm 6.29$	$126.2^{E} \pm 17.79$
Hydroquinone	495.5^{G}	
Paclitaxel		$0.04^{F} \pm 0.01$

[a] IC_{50} values of EOs were calculated by MTT assay.
[b] PCEOs, Pogostemon cablin (Blanco) Benth. essential oils, PFEOs, Perilla frutescens (L.) Britton essential oils, SJEOs, Salvia japonica Thunb. essential oils, ROEOs, Rosmarinus officinalis L. essential oils, LAEOs, Lavandula angustifolia Mill. essential oils, AREOs, Agastache rugosa (Fisch. & C. A. Mey.) Kuntze. essential oils.
[c] values followed by different capitals are significantly different ($P \leq 0.05$) in One-way ANOVA. ABCDEFG values followed by different capitals are significant different ($p \leq 0.01$) in one-way ANOVA.
[d] Data are means ± SD of 3 replications.

10.3389/fpls.2022.919294

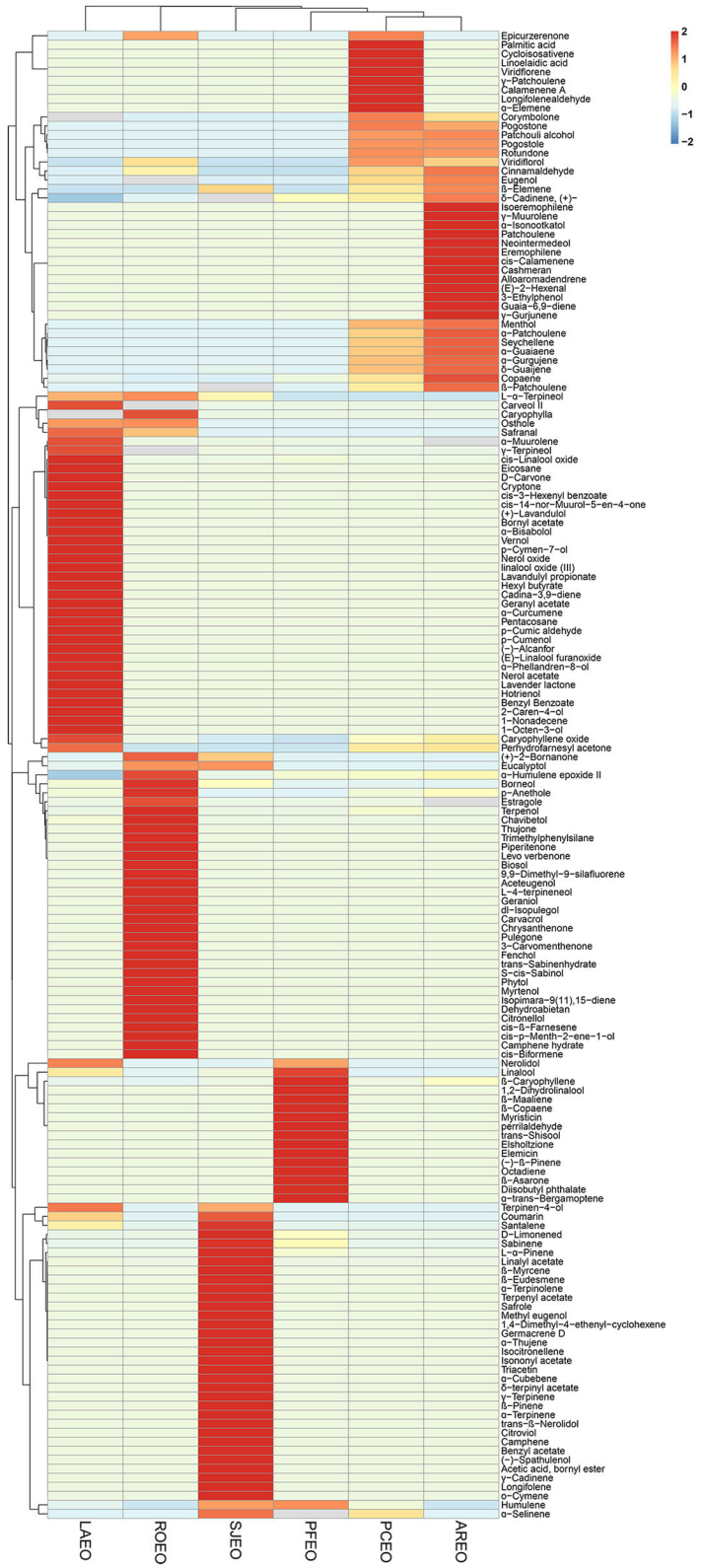


FIGURE 1
Hierarchical cluster analysis for the chemical compositions in six essential oils.

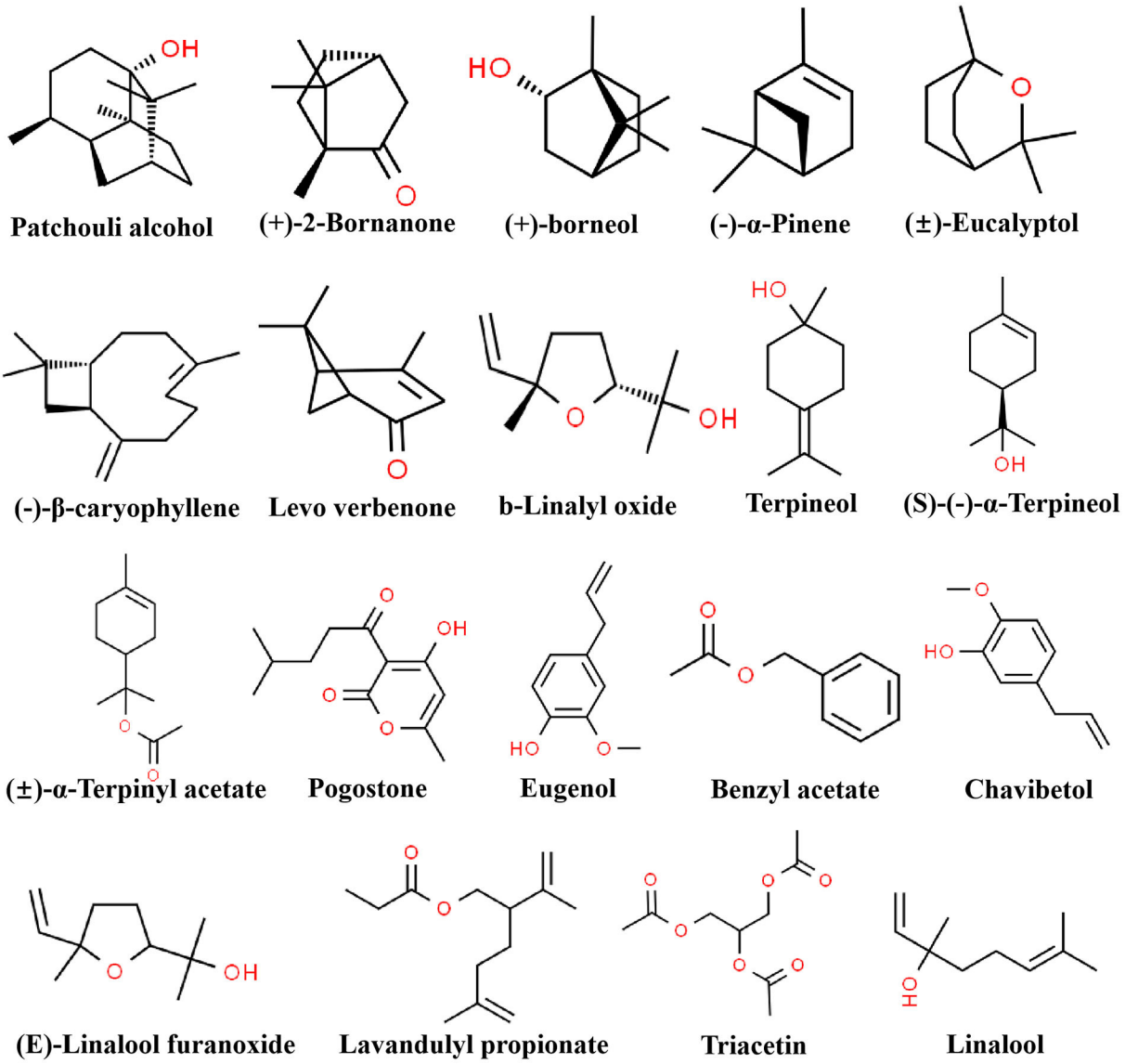


FIGURE 2
The chemical structure of 15 components in six essential oils. The relative content of each component was more than 4%.

Our results indicated that the six Lamiaceae essential oils have diverse chemical compositions, and they could serve as good sources of eugenol, patchouli alcohol, linalool, eucalyptol, β-caryophyllene, terpenyl acetate, chavibetol, camphor, γ-terpineol, borneol, and α-pinene. Some of these essential oils have been reported to demonstrate promising antioxidant, anti-nociceptive, anti-cardiotoxicity, anti-cancer, and anti-inflammatory activities (Sharma et al., 1994; Santos and Rao, 2000; Khan et al., 2014; Vyry Wouatsa et al., 2014; Fidyt et al., 2016; Mamadalieva et al., 2017; Nieto, 2017; Lian et al., 2018; Liu et al., 2018; de Souza et al., 2019; Luo et al., 2019; Oner et al., 2019).

Anti-arthritis activity of the six essential oils

As shown in Figure 3, all the six Lamiaceae essential oils displayed inhibitory effects on adjuvant arthritis in rats. Compared to the model group, PCEO, PFEO, ROEO, SJEO, LAEO, and AREO at a concentration of 100 mg/kg exhibited inhibitory effects on arthritis, among which PFEO manifested the highest anti-inflammatory activity, while PCEO showed the lowest. This result was consistent with that obtained from the PC group (ibuprofen treatment), which is effective for alleviating joint swelling in the rat models of arthritis.

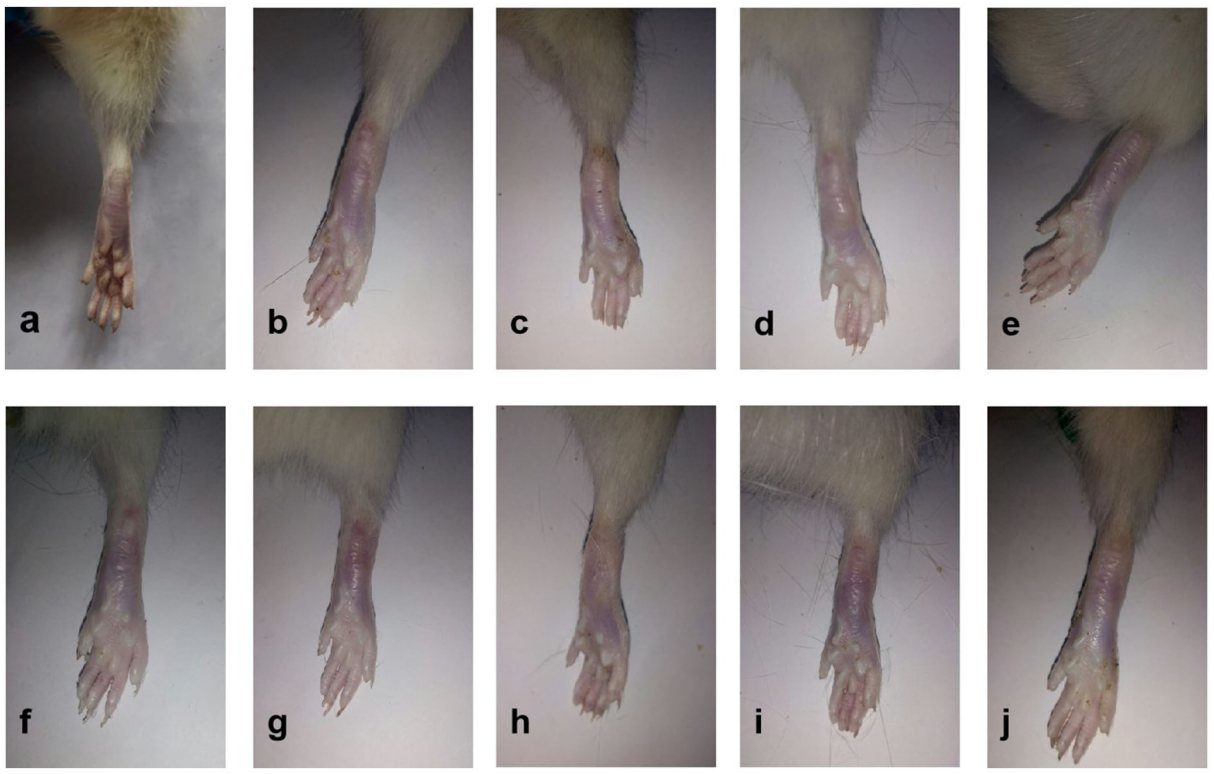


FIGURE 3
The appearance of rats with CFA-induced adjuvant arthritis. **(a)** Normal control (Con) group; **(b)** model (CFA) group; **(c)** negative control (NC) group; **(d)** positive control (PC) group; **(e)** PCEO group; **(f)** PFEO group; **(g)** ROEO group; **(h)** SJEO group; **(i)** LAEO group; and **(j)** AREO group.

The ankle joints of rats in the model group were significantly swollen compared with those of the control (Figure 3). In addition, the arthritis score of the positive control group (ibuprofen treatment) was significantly lower than that of the model group (Figure 4), which implied the success of CFA-induced arthritis. The arthritis scores of six essential oil treatment groups were lower than that of the model group, with PL displaying the lowest score, which indicated that PFEO might possess the strongest anti-arthritis capacity (Figure 4).

To obtain further insight into the anti-arthritis effect of six essential oils, histological and immunohistochemical characterizations were conducted using articular tissues. Severe cartilage damage, capillary hyperplasia, synovial proliferation, and lymphocyte infiltration were observed in the model group, while lymphocyte infiltration and cartilage damage were significantly inhibited in the PC group (ibuprofen treatment) (Figure 5). PCEO, PFEO, ROEO, SJEO, LAEO, and AREO exhibited similar effects relative to ibuprofen, which may largely decrease the damage.

Complete Freund's adjuvant can induce numerous inflammatory responses of cytokines, including COX-2, iNOS, IL-1, and IL-6 (Zhang et al., 2020c). To further understand the anti-inflammatory mechanism of these Lamiaceae essential oils, the spatial-temporal expression profiles of inflammatory cytokines in rat articular tissues were investigated (Figure 6). In the model group, CFA treatment greatly induced the expression of COX-2, TNF-α, IL-1, and IL-6 in rat articular tissues, while the essential oil treatments notably reduced that of TNF-α, IL-1, and IL-6 compared to the Con group. Nevertheless, COX-2 expression was slightly decreased after ibuprofen treatment. Caryophyllene was a common component shared by the six essential oils, which was reported to inhibit the expression of TNF, IL-1β, and COX-2 in APP/PS1 mice (Alzheimer-like phenotype) through CB2 receptor activation and the PPARγ pathway (Cheng et al., 2014). In our study, PFEO exhibited the greatest anti-inflammatory capacity inhibiting adjuvant arthritis and largely reduced the expression of inflammatory cytokines including TNF-α, IL-1, and IL-6. Linalool, a dominant component in PFEO, exhibited notably anti-inflammatory potential by reducing the expression of IL-1β and TNF-α in BV2 microglia cell lines (Li et al., 2015). Our results are consistent with those of previous studies, indicating that linalool may be a potential anti-inflammatory compound in those essential oils. In all, our results demonstrated that essential oils of Lamiaceae

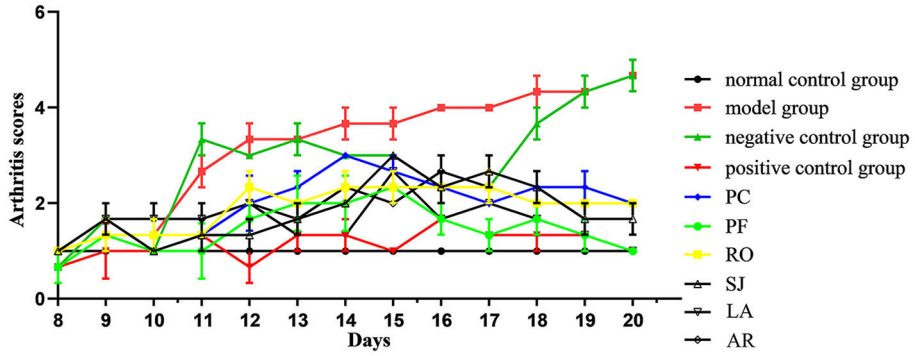


FIGURE 4
Arthritis scores of rats with CFA-induced adjuvant arthritis. The scores of normal control group, model (CFA) group, negative control (NC) group, positive control group, PCEO group (PC), PFEO group (PF), ROEO group (RO), SJEO group (SJ), LAEO group (LA), AREO group (AR) are shown with different color lines.

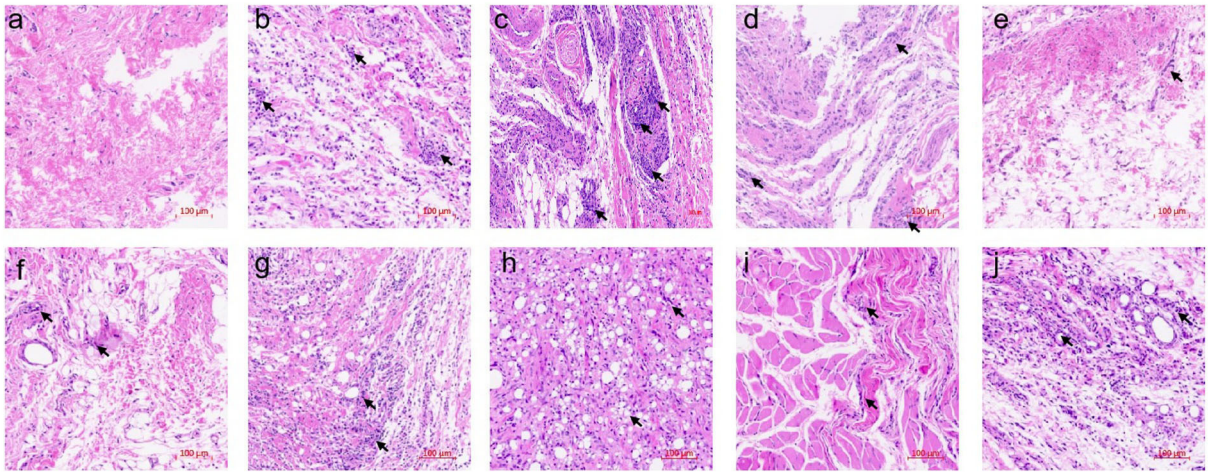


FIGURE 5
Histological sections of rat articular tissues showing severe cartilage damage, capillary hyperplasia, synovial proliferation, and lymphocyte infiltration (200× magnification). **(a)** Normal control (Con) group; **(b)** model (CFA) group; **(c)** negative control (NC) group; **(d)** positive control (PC) group; **(e)** PCEO group; **(f)** PFEO group; **(g)** ROEO group; **(h)** SJEO group; **(i)** LAEO group; **(j)** AREO group. Arrows indicate the lesion sites.

herbs play a key role in alleviating inflammation by inhibiting inflammatory cytokine expression.

Antitumor activity of the six essential oils

The antitumor activity was evaluated by examining *in vitro* inhibitory effects of these essential oils on LNCaP and B16 cells, and the results are shown in Table 3. The IC_{50} values of the six essential oils on LNCaP cells were between 116.41 and 189.63 μg/mL; LAEO (116.41 μg/mL) showed the strongest inhibitory effect, followed by AREO (126.20 μg/mL), PFEO (127.90 μg/mL), SJEO (129.40 μg/mL), PCEO (138.24 μg/mL), and ROEO (189.63 μg/mL). The IC_{50} values of these essential oils on B16 cells ranged from 86.91 to 228.91 μg/mL; AREO (86.91 μg/mL) exhibited the highest inhibitory effect, followed by ROEO (93.96 μg/mL), PCEO (109.32 μg/mL), PFEO (144.36 μg/mL), LAEO (152.69 μg/mL), and SJEO (189.63 μg/mL).

Previous studies have shown that patchouli alcohol, linalool, caryophyllene, borneol, and camphor achieve anticancer effects by inhibiting the expression of inflammatory factors (Santos and Rao, 2000; de Lima et al., 2014; Fidyt et al., 2016; Lian et al., 2018; Oner et al., 2019). Our results showed that the

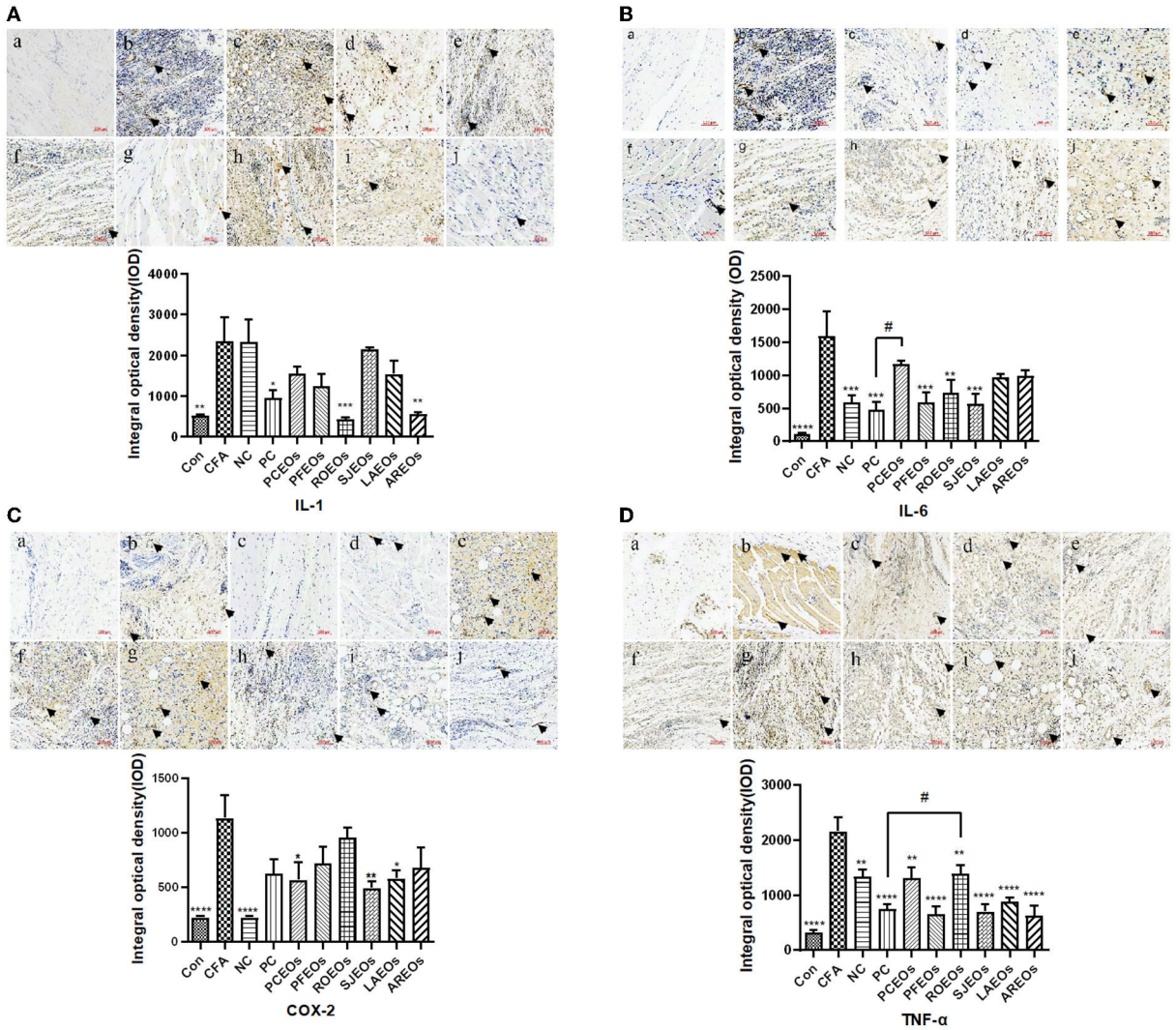


FIGURE 6
Immunohistochemical staining of rat articular tissues for COX-2 **(A)** IL-1 **(B)**, IL-6 **(C)**, and TNF-α **(D)** (200 × magnification). (a) normal control (Con) group, (b) model (CFA) group, (c) negative control (NC) group, (d) positive control (PC) group, (e) PCEO group, (f) PFEO group, (g) ROEO group, (h) SJEO group, (i) LAEO group, and (j) AREO group. Statistical significance relative to the model (CFA) group is indicated, $^{*}p \leq 0.05, ^{**}p \leq 0.01, ^{***}p \leq 0.001, ^{****}p \leq 0.0001$, significantly different from the model (CFA) group. Arrows indicate the lesion sites. Statistical significance relative to the positive control (PC) group is indicated, $^{\#}p \leq 0.05$.

dominant components of six essential oils had different degrees of anticancer effects on B16 and LNCaP cells, consistent with previous studies. PCEO and AREO had better performance in inhibiting B16 and LNCaP cells. Patchouli alcohol, a dominant component of PCEO (43.04%) and AREO (45.70%), maybe a key anticancer ingredient in these two essential oils. Linalool was a dominant component in LAEO (29.84%) and PFEO (66.65%) and may be a key component that plays an important role in their anticancer activity. Borneol (14.59%) and camphor (15.10%) were the major components of ROEO, which were likely the key antitumor component in our *in vitro* cell experiment. Similarly, linalool (8.89%) and camphor (9.54%) may be principal components participating in the anticancer activity of SJEO.

Antioxidant activity of the six essential oils

The DPPH method, which is stable, simple, and fast, was employed to assess free radical-scavenging activity of the six essential oils (Luo et al., 2019; Zhang et al., 2020a,b). In this study, our results showed that the IC_{50} values of the six essential oils were between 7.93 and

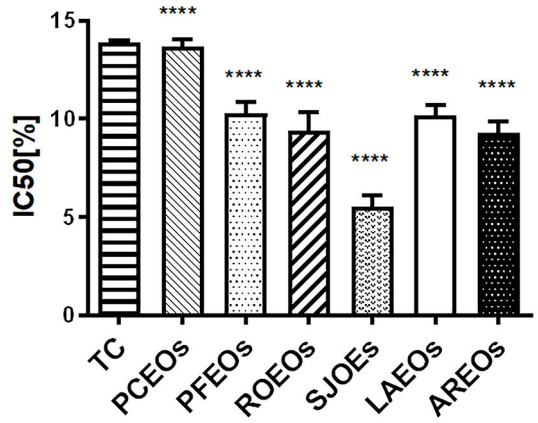


FIGURE 7
Measurement of DPPH free radical scavenging capacity of the six Lamiaceae essential oils using Trolox C (TC) as the positive control. Statistical significance relative to the Trolox C (TC) group is indicated, $^{****}p \leq 0.0001$, significantly different from the positive control group.

13.83 μg/mL (Figure 7). SJEO (4.88 μg/mL) showed the highest free radical scavenging capacity, while PCEO had minimal capacity (13.89 μg/mL). The antioxidant activity of AREO (8.79 μg/mL), ROEO (8.98 μg/mL), LAEO (9.80 μg/mL), and PFEO (10.87 μg/mL) was superior to that of Trolox C (13.80 μg/mL).

The antioxidant activity of some Lamiaceae essential oils had been assessed using the DPPH method in the previous study (Spiridon et al., 2011; Luo et al., 2019). Luo et al. (2019) demonstrated that LAEO (IC_{50} = 1.07%) possesses higher antioxidant activity than PCEO (IC_{50} = 1.60%), ROEO (IC_{50}, 2.80%), and PFEO (IC_{50}, 18.77%). Spiridon et al. (2011) showed that LAEO (IC_{50} = 96.67 μg/mL) can effectively scavenge free radicals. In this study, SJEO (IC_{50} = 4.88 μg/mL) showed the highest activity on scavenging free radicals, followed by AREO, ROEO, LAEO, and PFEO.

The antioxidant activity of essential oils may be largely associated with their major components, among which eucalyptol, α-pinene, and linalool have been reported to demonstrate antioxidant activity (Khan et al., 2014; Oner et al., 2019; Zhang et al., 2020c). For example, α-pinene, with an IC_{50} of 310 μg/mL, was reported to possess antioxidant activity in DPPH assays (Bouzenna et al., 2017). In this study, terpenyl acetate (10.52%), camphor (9.54%), α-pinene (5.97%), linalool (5.97%), and benzyl acetate (5.73%) were the principal components of SJEO, which may account for its high antioxidant activity. Chavibetol (17.72%), camphor (15.10%), verbenone (15.29%), borneol (14.58%), and eucalyptol (6.04%) were the dominant components of ROEO, which may lead to high antioxidant activity. Taken together, our results indicate that some essential oils, as well as their major components, may serve as potent natural antioxidants.

Conclusion

The demand for essential oils is growing due to their potential applications in pharmacology and industry. The phytochemical composition of essential oils takes charge of their bio-activities and their pharmaceutical effects. However, the chemical composition of plant essential oils is affected by many factors, such as a growing environment and identification methods. In this study, we used modified methods to identify different chemical component profiles of essential oils extracted from six folk medicinal plants. A total of 167 components were identified and analyzed by GC-MS, among which the dominant components were patchouli alcohol, pogostone, linalool, chavibetol, β-caryophyllene, terpenyl acetate, camphor, chavibetol, verbenone, borneol, and γ-terpineol (Table 2). HCA of chemical compositions was first analyzed in six essential oils, which provides more reference for pharmaphylogeny research (Figure 1). Meanwhile, the chemical structure of 15 major compounds was exhibited to provide more reference for later research. Our results were remarkably different from the former reported and provide a new insight into the phytochemical composition of the six essential oils.

The six Lamiaceae essential oils exhibited diverse anti-inflammatory activities on CFA-induced adjuvant arthritis in rats. PFEO, with a high linalool concentration (67.65%), showed higher anti-inflammatory activity relative to the rest of the essential oils and ibuprofen. Anti-inflammation was achieved by inhibiting the expression of COX-2, IL-1, IL-6, and TNF-α. All six essential oils also demonstrated different DPPH radical scavenging capacities except PCEO. SJEO, with high concentrations of terpenyl acetate (10.52%) and camphor (9.54%), showed the highest antioxidant capacity. The six essential oils also exhibited significantly different antitumor activities on LNCaP and B16 cells. AREO, with a high proportion of patchouli alcohol (45.70%), showed the highest antitumor capacity by inhibiting B16 cells with the lowest concentration of 86.91 μg/mL. LAEO, high in linalool (29.84%), showed promising antitumor capacity by inhibiting LNCaP cells at the lowest dosage of 116.5 μg/mL. Collectively, these six Lamiaceae essential oils, possessing varied chemical compositions and biological activities, exhibit potential for serving as bio-functional additives in biomedical products, such as anti-inflammatory and antitumor drugs.

Five of the six medicinal plants belong to the subfamily Nepetoideae (Dumort.) Burnett, except for *P. cablin*, which belongs to the subfamily Lamioideae Harley, indicating that species with similar bio-activities and pharmaceutical

effects are clustered in their phylogenetic relationships. More specifically, based on our study, PCEO showed the lowest effects on anti-inflammatory and antioxidant activities, which is consistent with the phylogeny (Supplementary Figure S2). In the views of pharmaphylogeny, species that are closely related are not only similar in physiological characteristics, but at the same time, it is also reflected in the similarity of phytochemical components. Our results showed that the principal components of the six essential oils were greatly distinct. Perhaps because the genes expressing these chemical compositions have suffered different selection pressures during evolution, or they have been of independent origin and have undergone different evolutionary pathways. However, we can still find chemical compositions that match the phylogenetic tree. For example, *L. angustifolia* and *P. frutescens* are sister groups, and the proportion of nerolidol and linalool showed similarity in LAEO and PFEO; *R. officinalis* and *S. japonica* are closely related, and the content of (+)-2-bornanone and eucalyptol showed similarity in SJEO and ROEO.

Although these six Lamiaceae plants are widely used and cultivated in China, only *A. rugosa*, *P. frutescens* and *S. japonica* are native to China. *L. angustifolia* and *R. officinalis* are native to Mediterranean, and *P. cablin* is distributed around the equator in Southeast Asia. This research provided more references for pharmaphylogeny and drug discovery from folk medicinal plants, and more studies need to be done for further exploring the drugs' function.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding authors.

Ethics statement

The animal study was reviewed and approved by the animal experiments were carried out following Ethical Guidelines of the Laboratory Animal Center of Sun Yat-sen University.

Author contributions

YK: data curation, writing–original draft, and writing–review and editing. LG: conceptualization and supervision. JS: data curation and software. PS: software and visualization. CK: data curation. LZ: conceptualization and methodology. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the National Key Research and Development Program of China (No. 2017YFC1700701), CACMS Innovation Fund (CI2021A03909), and the Key Laboratory of Biology and Genetic Improvement of Horticultural Crops, the Ministry of Agriculture and Rural Affairs, China.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.919294/full#supplementary-material

References

Borges, R. S., Ortiz, B. L. S., Pereira, A. C. M., Keita, H., and Carvalho, J. C. T. (2019). *Rosmarinus officinalis* essential oil: A review of its phytochemistry, anti-inflammatory activity, and mechanisms of action involved. *J. Ethnopharmacol.* 229, 29–45. doi: 10.1016/j.jep.2018.09.038

Bouzenna, H., Hfaiedh, N., Giroux-Metges, M. A., Elfeki, A., and Talarmin, H. (2017). Potential protective effects of alpha-pinene against cytotoxicity caused by aspirin in the IEC-6 cells. *Biomed. Pharmacother.* 93, 961–968. doi: 10.1016/j.biopha.2017.06.031

Cavanagh, H. M., and Wilkinson, J. M. (2002). Biological activities of *lavender* essential oil. *Phytother. Res.* 16, 301–308. doi: 10.1002/ptr.1103

Chen, H., Zhang, L., Zhang, S., and Luo, Y. (2017). GC-MS Analysis of Volatile Oil Constiuents in Herbal Pair Pogostemonis Herba-Eupatorii Herba and the Single Herbs. *Tradi. Chine. Drug Rese. Clinil Pharm.* 28, 781–785. (Chinese).

Cheng, Y., Dong, Z., and Liu, S. (2014). beta-Caryophyllene ameliorates the Alzheimer-like phenotype in APP/PS1 Mice through CB2 receptor activation and the PPARgamma pathway. *Pharmacology* 94, 1–12. doi: 10.1159/000362689

de Lima, V. T., Vieira, M. C., Kassuya, C. A., Cardoso, C. A., Alves, J. M., Foglio, M. A., et al. (2014). Chemical composition and free radical-scavenging, anticancer and anti-inflammatory activities of the essential oil from Ocimum kilimandscharicum. *Phytomedicine* 21, 1298–1302. doi: 10.1016/j.phymed.2014.07.004

de Souza, W. F. M., Mariano, X. M., Isnard, J. L., de Souza, G. S., de Souza Gomes, A. L., de Carvalho, R. J. T., et al. (2019). Evaluation of the volatile composition, toxicological and antioxidant potentials of the essential oils and teas of commercial Chilean boldo samples. *Food Res. Int.* 124, 27–33. doi: 10.1016/j.foodres.2018.12.059

El-Sayed, R. M., Moustafa, Y. M., and El-Azab, M. F. (2014). Evening primrose oil and celecoxib inhibited pathological angiogenesis, inflammation, and oxidative stress in adjuvant-induced arthritis: novel role of angiopoietin-1. *Inflammopharmacology* 22, 305–317. doi: 10.1007/s10787-014-0200-5

Fidyt, K., Fiedorowicz, A., Strzadala, L., and Szumny, A. (2016). beta-caryophyllene and beta-caryophyllene oxide-natural compounds of anticancer and analgesic properties. *Cancer Med.* 5, 3007–3017. doi: 10.1002/cam4.816

Funk, J. L., Frye, J. B., Oyarzo, J. N., Zhang, H., and Timmermann, B. N. (2010). Anti-arthritic effects and toxicity of the essential oils of turmeric (Curcuma longa L.). *J. Agric. Food Chem.* 58, 842–849. doi: 10.1021/jf9027206

Gao, Y., and Zhang, Z. X. (2019). Extraction and component analysis of essential oil from perilla leaf. *Guangzhou Chem. Indus* 47, 112–114. (Chinese).

Ghasemzadeh Rahbardar, M., and Hosseinzadeh, H. (2020). Therapeutic effects of rosemary (Rosmarinus officinalis L.) and its active constituents on nervous system disorders. *Iran J. Basic Med. Sci.* 23, 1100–1112. doi: 10.22038/ijbms.2020.45269.10541

Guo, S., Geng, Z., Zhang, W., Liang, J., Wang, C., Deng, Z., et al. (2016). The chemical composition of essential oils from cinnamomum camphora and their insecticidal activity against the stored product pests. *Int. J. Mol. Sci.* 17, 1836. doi: 10.3390/ijms17111836

Guo, S. S., Wang, Y., Pang, X., Geng, Z. F., Cao, J. Q., and Du, S. S. (2019). Seven herbs against the stored product insect: Toxicity evidence and the active sesquiterpenes from Atractylodes lancea. *Ecotoxicol. Environ. Saf.* 169, 807–813. doi: 10.1016/j.ecoenv.2018.11.095

Kanipandian, N., and Thirumurugan, R. (2014). A feasible approach to phyto-mediated synthesis of silver nanoparticles using industrial crop Gossypium hirsutum (cotton) extract as stabilizing agent and assessment of its in vitro biomedical potential. *Ind. Crop. Prod.* 55, 1–10. doi: 10.1016/j.indcrop.2014.01.042

Karpinski, T. M. (2020). Essential oils of lamiaceae family plants as antifungals. *Biomolecules* 10, 103. doi: 10.3390/biom10010103

Khan, A., Vaibhav, K., Javed, H., Tabassum, R., Ahmed, M. E., Khan, M. M., et al. (2014). 1,8-cineole (eucalyptol) mitigates inflammation in amyloid Beta toxicated PC12 cells: relevance to Alzheimer's disease. *Neurochem. Res.* 39, 344–352. doi: 10.1007/s11064-013-1231-9

Khayyal, M. T., El-Ghazaly, M. A., Abdallah, D. M., Okpanyi, S. N., Kelber, O., and Weiser, D. (2005). Mechanisms involved in the anti-inflammatory effect of a Standized Willow Bark Extract. *Arzneimittelforschung* 55, 11. doi: 10.1055/s-0031-1296917

Li, H. Q., Liu, Q. Z., Liu, Z. L., Du, S. S., and Deng, Z. W. (2013). Chemical composition and nematicidal activity of essential oil of *Agastache rugosa* against Meloidogyne incognita. *Molecules* 18, 4170–4180. doi: 10.3390/molecules18044170

Li, W. S. (2010). Qualitative analysis of volatile oil from Perilla leaf by GC-MS. *Henan Sci. Tech*. 14,168–169. (Chinese).

Li, Y., Duan, Y. Q., Wang, H. X., and Xia, J. J. (2016). Analysis of aroma components in sage oil by GC combined with retention index. *Cereals Oils* 29, 65–68. (Chinese).

Li, Y., Lv, O., Zhou, F., Li, Q., Wu, Z., and Zheng, Y. (2015). Linalool Inhibits LPS-Induced Inflammation in BV2 Microglia Cells by Activating Nrf2. *Neurochem. Res.* 40, 1520–1525. doi: 10.1007/s11064-015-1629-7

Lian, D. W., Xu, Y. F., Ren, W. K., Fu, L. J., Chen, F. J., Tang, L. Y., et al. (2018). Unraveling the novel protective effect of patchouli alcohol against helicobacter pylori-induced gastritis: insights into the molecular mechanism *in vitro* and *in vivo*. *Front. Pharmacol.* 9, 1347. doi: 10.3389/fphar.2018.01347

Liu, W. J., Yin, Y. B., Sun, J. Y., Feng, S., Ma, J. K., Fu, X. Y., et al. (2018). Natural borneol is a novel chemosensitizer that enhances temozolomide-induced anticancer efficiency against human glioma by triggering mitochondrial dysfunction and reactive oxide species-mediated oxidative damage. *Onco. Targets Ther.* 11, 5429–5439. doi: 10.2147/OTT.S174498

Luo, W., Du, Z., Zheng, Y., Liang, X., Huang, G., Zhang, Q., et al. (2019). Phytochemical composition and bioactivities of essential oils from six Lamiaceae species. *Ind. Crop. Prod.* 133, 357–364. doi: 10.1016/j.indcrop.2019.03.025

Mamadalieva, N. Z., Sharopov, F., Satyal, P., Azimova, S. S., and Wink, M. (2017). Composition of the essential oils of three Uzbek Scutellaria species (Lamiaceae) and their antioxidant activities. *Nat. Prod. Res.* 31, 1172–1176. doi: 10.1080/14786419.2016.1222383

Mamadalieva, N. Z., Youssef, F. S., Ashour, M. L., Sasmakov, S. A., Tiezzi, A., and Azimova, S. S. (2019). Chemical composition, antimicrobial and antioxidant activities of the essential oils of three Uzbek Lamiaceae species. *Nat. Prod. Res.* 33, 2394–2397. doi: 10.1080/14786419.2018.1443088

Mouahid, A., Dufour, C., and Badens, E. (2017). Supercritical CO_2 extraction from endemic Corsican plants; comparison of oil composition and extraction yield with hydrodistillation method. *J. CO_2 Util.* 20, 263–273. doi: 10.1016/j.jcou.2017.06.003

Nieto, G. (2017). Biological activities of three essential oils of the lamiaceae family. *Medicines (Basel)* 4, 63. doi: 10.3390/medicines4030063

Nikolić, M., Jovanović, K. K., Marković, T., Marković, D., Gligorijević, N., Radulović, S., et al. (2014). Chemical composition, antimicrobial, and cytotoxic properties of five Lamiaceae essential oils. *Ind. Crop. Prod.* 61, 225–232. doi: 10.1016/j.indcrop.2014.07.011

Oner, Z., Altinoz, E., Elbe, H., and Ekinci, N. (2019). The protective and therapeutic effects of linalool against doxorubicin-induced cardiotoxicity in Wistar albino rats. *Hum. Exp. Toxicol.* 38, 803–813. doi: 10.1177/0960327119842634

Park, C. G., Jang, M., Yoon, K. A., and Kim, J. (2016). Insecticidal and acetylcholinesterase inhibitory activities of Lamiaceae plant essential oils and their major components against Drosophila suzukii (Diptera: Drosophilidae). *Ind. Crop. Prod.* 89, 507–513. doi: 10.1016/j.indcrop.2016.06.008

Rodboon, T., Okada, S., and Suwannalert, P. (2020). Germinated 2020 riceberry rice enhanced protocatechuic acid and vanillic acid to suppress melanogenesis through cellular oxidant-related tyrosinase activity in B16 cells. *Antioxidants (Basel)*. 9, 247. doi: 10.3390/antiox9030247

Santos, F. A., and Rao, V. S.. (2000). Antiinflammatory and antinociceptive effects of 1,8-cineole a terpenoid oxide present in many plant essential oils. *Phytother Research* 14, 240–4. doi: 10.1002/1099-1573(200006)14:4<240::aid-ptr573>3.0.co;2-x

Sharma, J. N., Srivastava, K. C., and Gan, E. K. (1994). Suppressive effects of eugenol and ginger oil on arthritic rats. *Pharmacology.*, 49, 5. doi: 10.1159/000139248

Spiridon, I., Colceru, S., Anghel, N., Teaca, C. A., Bodirlau, R., and Armatu, A. (2011). Antioxidant capacity and total phenolic contents of oregano (Origanum vulgare), lavender (*Lavandula angustifolia*) and lemon balm (*Melissa officinalis*) from Romania. *Nat. Prod. Res.* 25, 1657–1661. doi: 10.1080/14786419.2010.521502

Vieira, J. N., Gonçalves, C. L., Villarreal, J., Goncalves, V., Lund, R., Freitag, R., et al. (2019). Chemical composition of essential oils from the apiaceae family, cytotoxicity, and their anti-fungal activity in vitro against candida species from oral cavity. *Braz. J. Biol.* 79, 432–437. doi: 10.1590/1519-6984.182206

Vukovic, N., Sukdolak, S., Solujic, S., and Niciforovic, N. (2009). Antimicrobial activity of the essential oil obtained from roots and chemical composition of the volatile constituents from the roots, stems, and leaves of Ballota nigra from Serbia. *J. Med. Food* 12, 435–441. doi: 10.1089/jmf.2008.0164

Vyry Wouatsa, N. A., Misra, L., and Venkatesh Kumar, R. (2014). Antibacterial activity of essential oils of edible spices, Ocimum canum and Xylopia aethiopica. *J. Food Sci.* 79, M972–M977. doi: 10.1111/1750-3841.12457

Wojtunik-Kulesza, K. A., Kasprzak, K., and Oniszczuk, T., Oniszczuk, A. (2019). Natural monoterpenes: Much more thanonly a scent. *Chem. Biodiv.* 16, e19004. doi: 10.1002/cbdv.201900434

Xiang, H., Zhang, L., Yang, Z., Chen, F., Zheng, X., and Liu, X. (2017). Chemical compositions, antioxidative, antimicrobial, anti-inflammatory and antitumor activities of Curcuma aromatica Salisb. essential oils. *Ind. Crop. Prod.* 108, 6–16. doi: 10.1016/j.indcrop.2017.05.058

Xue, H. J., Wei, J. N., Magalhães, S., Zhang, B., Song, K. Q., Liu, J., et al. (2016). Contact pheromones of 2 sympatric beetles are modified by the host plant and affect mate choice. *Behav. Ecol.* 27, 895–902. doi: 10.1093/beheco/arv238

Xue, S. (2016). Research of the composition and antioxidant activity of essential oil from Folium Perillaes extracted by different methods. *Sci. Tech. Food. Indus.,* 37, 67–74. (Chinese).

Yu, M. H., Choi, J. H., Chae, I. G., Im, H. G., Yang, S. A., More, K., et al. (2013). Suppression of LPS-induced inflammatory activities by *Rosmarinus officinalis* L. *Food Chem.* 136, 047–1054. doi: 10.1016/j.foodchem.2012.08.085

Zhang, L., Liang, X., Ou, Z., Ye, M., Shi, Y., Chen, Y., et al. (2020a). Screening of chemical composition, anti-arthritis, antitumor and antioxidant capacities of essential oils from four Zingiberaceae herbs. *Ind. Crop. Prod.* 149, 112342. doi: 10.1016/j.indcrop.2020.112342

Zhang, L., Liang, X., Wang, B., Lin, Z., Ye, M., Ma, R., et al. (2020b). Six herbs essential oils suppressing inflammatory responses via inhibiting COX-2/TNF-α/IL-6/NF-κB activation. *Microchem. J.* 156, 104769. doi: 10.1016/j.microc.2020.104769

Zhang, L., Pan, C., Ou, Z., Liang, X., Shi, Y., Chi, L., et al. (2020c). Chemical profiling and bioactivity of essential oils from Alpinia officinarum Hance from ten localities in China. *Ind. Crop. Prod.* 153, 112583. doi: 10.1016/j.indcrop.2020.112583

Zhang, L., Yang, Z., Chen, D., Huang, Z., Li, Y., Lan, X., et al. (2017c). Variation on composition and bioactivity of essential oils of four common *Curcuma* Herbs. *Chem Biodivers.* 14, e1700280. doi: 10.1002/cbdv.201700280

Zhang, L., Yang, Z., Wei, J., Su, P., Chen, D., Pan, W., et al. (2017a). Contrastive analysis of chemical composition of essential oil from twelve Curcuma species distributed in China. *Ind. Crop. Prod.* 108, 17–25. doi: 10.1016/j.indcrop.2017.06.005

Zhang, L., Yang, Z., Wei, J., Su, P., Pan, W., Zheng, X., et al. (2017b). Essential oil composition and bioactivity variation in wild-growing populations of Curcuma phaeocaulis Valeton collected from China. *Ind. Crop. Prod.* 103, 274–282. doi: 10.1016/j.indcrop.2017.04.019

Zhang, R. Z., Liu, D. P., Xiao, C. R., Tu, L. F., and Luo, Y. M. (2019). GC-MS analysis of volatile oil constituents in herbal pair pogostemonis herba-perillae folium and the single herbs. *Tradi. Chine. Drug Rese. Clinil Pharm.* 37, 67–74. (Chinese).

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 04 August 2022
DOI 10.3389/fpls.2022.955075

OPEN ACCESS

EDITED BY
Chunnian He,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

REVIEWED BY
Yi Wang,
Zhejiang University, China
Yanbing Shen,
Tianjin University of Science and Technology, China
Zhubo Dai,
Tianjin Institute of Industrial Biotechnology (CAS), China

*CORRESPONDENCE
Chunhua Wang
pharmwch@126.com
Xiaohui Yan
yanxh@tjutcm.edu.cn
Zheng Li
lizheng@tjutcm.edu.cn

SPECIALTY SECTION
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

RECEIVED 28 May 2022
ACCEPTED 27 June 2022
PUBLISHED 04 August 2022

CITATION
Peng S, Li F, Yu K, Zhou F, Yu H, Liu H, Guo J, Li G, Wang C, Yan X and Li Z (2022) Integrating transcriptome and chemical analyses to reveal the anti-Alzheimer's disease components in *Verbena officinalis* Linn.
Front. Plant Sci. 13:955075.
doi: 10.3389/fpls.2022.955075

Integrating transcriptome and chemical analyses to reveal the anti-Alzheimer's disease components in *Verbena officinalis* Linn

Shuhuan Peng[1,2], Fangyi Li[2], Kuo Yu[2], Fengshu Zhou[2], Heshui Yu[2], Hui Liu[1], Jialiang Guo[1], Guoqiang Li[3], Chunhua Wang[1,2]*, Xiaohui Yan[2]* and Zheng Li[2]*

[1]School of Medicine, Foshan University, Foshan, China, [2]College of Pharmaceutical Engineering of Traditional Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin, China, [3]School of Food Science and Engineering, Foshan University, Foshan, China

Verbena officinalis Linn. is a kind of traditional Chinese medicine, which has a long history of application and shows good effects on neuroprotection. Therefore, we consider that *V. officinalis* may be a potential drug for treating Alzheimer's disease (AD). First, ultra-performance liquid chromatography-mass spectrometry (UPLC-MS) pointed out that the main chemical components in *V. officinalis* were iridoid glycosides, phenylethanoid glycosides, and flavonoids. These compounds were used for molecular docking and the results showed that these compounds had good anti-AD activity. To explore the biosynthetic pathway of anti-AD components in *V. officinalis*, UPLC and ultraviolet (UV) spectrophotometry were used for contents determination and the result was leaf > stem > root. At the same time, 92,867 unigenes were annotated in *V. officinalis* transcriptome; 206, 229, 115 related unigenes were, respectively, annotated in iridoid glycoside, phenylethanoid glycoside, and flavonoid pathway, of which 61, 73, and 35 were differential expression genes. The components had relatively high expression in leaves, which was consistent with the quantitative results. In addition, the tissue distribution particularity of verbenalin may be related to the branching of pathways. Meanwhile transcription factors VoWRKY6 and VoWRKY7 may be involved in the regulation of iridoid glycoside biosynthesis. Further, VoWRKY3, VoWRKY9, and VoWRKY12 may be related to flavonoid biosynthesis. The above research is helpful to explore the biosynthetic pathway of anti-AD components and the regulation mechanism of active components and to further explore the anti-AD effect of *V. officinalis*.

KEYWORDS

Alzheimer's disease, *Verbena officinalis*, transcriptome, content determination, tissue expression difference, biosynthetic pathway

Introduction

Alzheimer's disease (AD) is a common neurodegenerative disease, which mostly occurs among elderly people. Its clinical symptoms are mainly expressed as hypomnesia, thinking slow, and language and learning disability. As humans get older, the risk of disease increases. Nowadays, with the trend of global aging, the threat of AD to human beings has increased. At present, the pathogenesis of AD is not clear. Researchers have put forward many related hypotheses, including amyloid hypothesis, tau hyperphosphorylation hypothesis, neuroinflammation hypothesis, mitochondrial dysfunction hypothesis, cholinergic hypothesis, and so on (Holmes, 2013; Lane et al., 2018; DeTure and Dickson, 2019). The drugs for clinical treatment of AD include acetylcholinesterase (AChE) inhibition galantamine, lisdimene, tacrine, donepezil, and N-methyl-D-aspartate receptor antagonist memantine (Ghezzi et al., 2013). Although these drugs can effectively improve the symptoms of AD, they cannot fundamentally treat AD.

Verbena officinalis Linn. (Verbenaceae), is a popular herb in folk medicine for the treatment of skin disease, edema, and dysmenorrhea. Pharmacological studies have shown that it has many pharmacological effects such as anti-tumor, antibacterial, antiviral, anti-early pregnancy, anti-inflakangmatory, and antioxidantion (Kubica et al., 2020). Notably, *V. officinalis* also shows a good neuroprotective effect. Lai et al. (2006) treated neurons *in vitro* with aqueous extract, which could reduce both destructions of neurites and neuronal apoptosis. The plant aqueous extract may play a role by inhibiting the activation of caspase-2 and caspase-3 by nerve cells. Tan and Wang (2011) prepared AD mice by giving D-galactose. After administration of *V. officinalis* decoction, they found that the number of first dark avoidance errors after 24 h significantly reduced, indicating that *V. officinalis* decoction improved its directional learning and memory ability. Outside, *V. officinalis*' good anti-inflammatory effects may also be a potential factor in the treatment of AD (Calvo et al., 1998; Speroni et al., 2007; Grawish et al., 2016).

There are many chemical components in *V. officinalis*, and the material basis of its anti-AD activity is not clear. Because the main components of the plant are iridoid glycosides, flavonoids, and phenylethanoid glycosides, they are considered active ingredients for anti-AD. Iridoid glycosides are characteristic components of this plant and a large number of studies have shown that the biosynthetic pathway of iridoid glycosides is the same as that of other terpenoids. Geranyl diphosphate is produced through the mevalonate (MVA) or 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway, and iridoid glycosides are generated through the downstream synthesis pathway (Krithika et al., 2015; Ahmad et al., 2018). Many studies have used RNA sequencing (RNA-Seq) technology to explore the biosynthetic pathway of medicinal plants which is rich in iridoid glycoside such as *Lamium barbatum* (Li et al., 2010), *Gardenia jasminoides* (Pan et al., 2021), and *Rehmannia glutinosa* (Sun et al., 2012). The flavonoid biosynthetic pathway is relatively clear, mainly involving the phenylpropane pathway, which has been clarified in *Arabidopsis thaliana* and other plants (Saito et al., 2013; Lou et al., 2021). In addition, the flavonoid biosynthetic pathway is also associated with secondary metabolites such as flavonol, flavanone, dihydroflavonol, and anthocyanin (Liu et al., 2021). Acteoside is a typical phenylethanoid glycoside compound. In the biosynthesis of acteoside, caffeic acid is produced by the phenylalanine pathway and 3-hydroxyloside is generated by the dopamine pathway or tyramine pathway. Bai et al. (2014) isolated a glycosyltransferase from *Rhodiola rosea*, UGT73B6, which can catalyze tyrosol to generate salidroside. And salidroside is a precursor of 3-hydroxyloside biosynthesis in *Escherichia coli*. However, the cross pathway between downstream acyl transfer and glycosyl modification is not clear and needs to be further explored.

To explore the anti-AD active compounds of *V. officinalis*, we first analyzed the chemical components by ultra-performance liquid chromatography-mass spectrometry (UPLC-MS) to investigate the potential therapeutic effects and possible mechanisms for AD. Iridoid glycosides, flavonoids, and phenylethanoid glycosides were docked with AD target protein by molecular docking technology. Then, we quantitatively analyzed the active components to determine the tissue difference in component content. Meanwhile, the RNA-Seq technology was used for *V. officinalis*. With the help of gene expression difference analysis and quantitative analysis of tissue differences, we speculated the enzymes related to the biosynthesis of active components and the biosynthetic pathway involving anti-AD active components, to provide clues for the further study of the anti-AD activity of *V. officinalis*.

Materials and methods

Preparation of plant materials

We herborized *V. officinalis* seedlings (7 May 2021) from Wangshan, Huangtan, Wencheng, Wenzhou, Zhejiang, China (27°44'10.23" N, 119°58'2.12" E). They were cultured at the Tianjin University of Traditional Chinese Medicine, in July 2021 and identified by Dr. Wang Chunhua. Plant samples are stored in the Pharmaceutical Engineering College of Tianjin University of Traditional Chinese Medicine. We selected normal growth and disease-free plants, and washed the tissue with pure water, sterile water, and RNA-free water. Quickly cut the plant leaf, stem, and root tissues into small pieces (three biological replicates for each tissue, from three plants cultured under the same conditions, a total of nine samples). Then, quickly froze them with liquid nitrogen, and transferred them to the refrigerator at −80°C for standby (**Figure 1**). After fresh plants were picked, the leaf, stem, root, and total plant tissue dried naturally, and we ground them into powder for sample.

FIGURE 1
Plant and tissues of *Verbena officinalis*. **(A)** Whole *V. officinalis*. plant. **(B)** Leaves. **(C)** Stems. **(D)** Roots.

Ultra-performance liquid chromatography-mass spectrometry

The Waters ACQUITY UPLC^{TM} System (Waters, Milford, MA, United States) combined with Q Exactive^{TM} Plus (Q Exactive Plus–Orbitrap MS., Thermo Fisher Scientific., Waltham, MA, United States) and Xcalibur 4.1 software (Thermo Fisher Scientific., Waltham, MA, United States). Acetonitrile and methanol used in the experiment are chromatographic pure reagents (Thermo Fisher Scientific., Waltham, MA, United States) and pure water from Mill-Q system (Millipore., Illkirch-Graffenstaden, France). The chromatographic column is Waters Acquity UPLC® BEH C_{18} column (100 mm × 2.1 mm, 1.7 μm, Waters, Milford, MA, United States), column temperature was 25°C, the mobile phases were acetonitrile(A) and water (B) with a flow rate of 0.2 ml/min, and the injection volume was 1 μl. The gradient program was 0–1 min, 5% A; 1–35 min, 5–95% A; 35–40 min, 95%A. Q Exactive MS source parameters include heated electrospray spray ionization source (HESI), positive

and negative mode detection, 100-1,500 Da in full MS scan mode, a 3.5 kV spray voltage, a 320°C capillary temperature, a 350°C auxiliary gas heater temperature, with a sheath gas flow rate of 40 arb, and the auxiliary gas flow rate at 15 arb. The resolution of the full MS was 70,000 and the resolution of the dd-MS2 was 17,500.

The total plant powders (0.1 g) were added to 5 ml of 80% (v/v) methanol for ultrasonic extraction (power 250 W, frequency 40 kHz), and the extraction time was 2 h. The solution was filtered after cooling, and 80% (v/v) methanol was added to the solution a total 5 ml. The test solution was centrifuged at 14,000 r/min at 4°C for 15 min, and the supernatant was taken for 0.22 μm filter membrane for UPLC-MS.

Molecule docking

The 3D structure of small molecule was downloaded from PubChem.[1] NLRP3 (6NPY), BACE1 (6UWP), AChE (6CQZ), and GSK-3β (1J1C) were downloaded from the PDB database.[2] PyMol 2.5.0 software was used for the water removal and ligand removal of target proteins, and the hydrogenation was completed with Autodock Tools 1.5.6. Finally, Autodock vina software was used for the docking of small molecules and target proteins.

Analysis of tissue contents difference

The Agilent Series 1290 UPLC (Agilent Technologies., Santa Clara, CA, United States) was used to determine the contents of verbenalin and acteoside. Standard references are verbenalin (Saizhiwei Technology Co., Tianjin, China, content ≥ 98%, batch number: wkq21100922) and acteoside (Saizhiwei Technology Co., Tianjin, China, content ≥ 98%, batch number: wkq22031108). Acetonitrile is chromatographic pure (Thermo Fisher, Waltham, MA, United States), pure water is taken from Mill-Q (Millipore., Illkirch-Graffenstaden, France), and methanol is analytically pure (Concord Technology., Tianjin, China). The chromatographic column is Waters Acquity UPLC® BEH C_{18} column (100 mm × 2.1 mm, 1.7 μm, Waters, Milford, MA, United States), column temperature was 30°C, the mobile phases were acetonitrile (A) and water (B) with a flow rate of 0.2 ml/min, detective wave was 240 nm and 334 nm, and the injection volume was 1 μl. The gradient program was 0–1 min, 5% A; 1–3 min, 5–23% A; 3–13 min, 23–95% A. The standard curve equation with 240 nm of verbenalin was $Y = 7517.9X + 44.166$, $R^2 = 0.9993$, the linear range was 0.032625 ∼ 0.58 mg/ml. And the standard curve equation with 334 nm of acteoside was $Y = 3814.3X - 23.406$, $R^2 = 0.9996$, the linear range was 0.08775 ∼ 1.56 mg/ml.

The sample powders (0.1 g) were added to 10 ml of 80% (v/v) methanol for ultrasonic extraction (power 250 W, frequency 40 kHz), and the extraction time was 2 h. The solution was filtered after cooling, and 80% (v/v) methanol was added to solution a total 10 ml. The test solution was centrifuged at 10,000 r/min for 10 min, and the supernatant was taken for 0.22 μm filter membrane for UPLC analysis. Each sample was measured three times and the experiment was repeated three times for parallel to calculate the content of chemical components in each tissue.

Otherwise, we used Agilent Cary 8454 UV-Vis (Agilent Technologies., Santa Clara, CA, United States) to determine the contents of total iridoid glycosides, phenylethanoid glycosides, and flavonoids in *V. officinalis*. verbenalin (Saizhiwei Technology Co., Tianjin, China., content ≥ 98%., batch number: wkq21100922), as the main component of *V. officinalis* was used as the quantitative analysis of total iridoid glycosides. The standard curve equation with 240 nm of total iridoid glycosides was $Y = 26.489X + 0.0071$, $R^2 = 0.9995$, the linear range was 4.716 ∼ 29.868 μg/ml. For the quantitative determination of phenylethanoside, acteoside (Saizhiwei Technology Co., Tianjin, China., content ≥ 98%., batch number: wkq22031108) was used as standard. At 334 nm, the standard curve equation was $Y = 26.354X - 0.0047$, $R^2 = 0.9999$, the linear range was 4.496 ∼ 29.224 μg/ml. For the content determination of total flavonoids, apigenin (Saizhiwei Technology Co., Tianjin, China., content ≥ 98%., batch number: wkq21072811) was selected as the standard control, colored with 1% (v/v) triethylamine, and analyzed quantitatively at 383 nm. The standard curve equation was $Y = 121.5X - 0.002$, $R^2 = 0.9998$, the linear range was 0.944 ∼ 6.136 μg/ml. The sample powders (0.1 g) were added to 20 ml of 80% (v/v) methanol for ultrasonic extraction (power 250 W, frequency 40 kHz), and the extraction time was 2 h. The solution was filtered after cooling, and 80% (v/v) methanol was added to solution a total 100 ml. Each sample was measured three times and the experiment was repeated three times in parallel to calculate the content of chemical components in each tissue.

RNA extraction, cDNA library construction, and sequencing

The leaf, stem, and root tissues of *V. officinalis* (3 copies each for biological replicate) were used to extract the total RNA from the cryopreserved plant tissues by using the Plant Total RNA Purification Kit (Tiangen biotech., Guangzhou, China) according to the instructions. The RNA quality was evaluated by using Agilent 2100 Bioanalyzer (Agilent Technologies., Santa Clara, CA, United States), and the degradation degree of RNA was detected by agarose gel electrophoresis. Based on the structural characteristics of most eukaryotic mRNAs

1 https://pubchem.ncbi.nlm.nih.gov/

2 https://www.rcsb.org/

with PolyA tail, the mRNA with PolyA tail was enriched by oligo (dT) magnetic beads, fragmented and reverse transcribed by ultrasound, and then repaired at the end. A total of 9 cDNA libraries were obtained by PCR amplification with linker, index, and sequencing primers. Each cDNA library was sequenced on Illumina Novaseq 6000 platform with paired end. The sequencing was done by Novogene (Novogene Co., Ltd., Beijing, China).

De novo assembly and unigene annotation

The original data were transformed into sequence data through base identification and presented in the fastq format. FastQC was used to ensure the quality and reliability of data analysis, the reads with adapter, reads containing, and low-quality reads were moved to obtain clean data. The clean reads data were assembled *de novo* into transcripts using Trinity (Grabherr et al., 2011). Based on Trinity splicing, Corset (Davidson and Oshlack, 2014) software was used for hierarchical clustering. After removing redundancy, take the longest transcript in each transcript cluster as a non-repetitive sequence gene.

In order to obtain the functional information of genes, the program blast2go (Götz et al., 2008) was used to compare unigenes with NR (NCBI non-redundant protein sequences), NT (NCBI nucleoside sequences), Pfam (Protein family), KOG/COG (Clusters of Orthologous Groups of proteins/euKaryotic Ortholog Groups) and Swiss prot (A manually annotated and reviewed protein sequence database), KEGG (Kyoto Encyclopedia of Genes and Genomes), and GO (Gene Ontology) database.

Differential expression analysis

Taking the transcriptome by Trinity as the reference sequence, Bowtie2 program (Langmead and Salzberg, 2012) was used to map the clean reads into the transcript assembly, and RSEM software (Li and Dewey, 2011) for the FPKM value (expected number of fragments per kilobase of transcript sequence per million base pairs sequence) to measure the quantitative analysis of gene expression level. Differential expression genes (DEGs) were screened (screening standard was $|\log_2(\text{FoldChange})| > 1$ and $p < 0.005$) by DESeq2 (Love et al., 2014).

We performed GO function enrichment and KEGG pathway enrichment analysis on the DEGs. The biosynthetic pathways of iridoid glycosides, phenylethanoid glycosides, and flavonoids in references, combined with the results of database function annotation, and relevant genes were identified. The value of FPKM was used to measure the level of expression. It is used to predict the biosynthetic pathway of related compounds. ChemDraw 20.0 and Origin 2021 were used to draw the heat map of the biosynthesis pathway and gene expression, respectively.

The iTAK software (Zheng et al., 2016) was used to predict plant transcription factors. In addition, we downloaded the sequence of WRKY transcription factors (TFs) in various plants from NCBI, and used MAGA7 (Kumar et al., 2016) software to analyze the developmental tree of WRKY in *A. thaliana*, *Dendrobium officinale*, *Catharanthus roseus*, *Artemisia annua*, *Oryza sativa*, *R. glutinosa,* and *V. officinalis*. Sequence alignment was performed with MUSCLE (Edgar, 2004). The sequence was cut with TBools (Chen C. et al., 2020) and the phylogenetic tree was constructed with Neighbor-Joining with 1,000 bootstrap replicates. The iTOL[3] tools were used to beautify the phylogenetic tree.

Results

Ultra-performance liquid chromatography-mass spectrometry

In this study, UPLC-MS was used to analyze and identify the chemical components of *V. officinalis* extract under positive and negative ion modes, respectively. There were responses in both positive and negative ion modes. A total of 16 compounds were identified, including 8 compounds in the positive ion mode and 8 compounds in the negative ion mode (**Supplementary Figure 1**). We scanned the medicinal materials in the positive and negative modes (**Table 1**). The results showed that the main chemical components of *V. officinalis* were iridoid glycosides, phenylethanoid glycosides, and flavonoids, among which iridoid glycosides mainly included verbenalin, hastatoside, gentiopicroside, aucubin, 3,4-dihydroverbenalin, and swertiamarine. Phenylethanoid glycosides include acetoside, isoacteoside, 2′-acetylacetoside, jionoside D, and cistanoside F. Flavonoids mainly include acacetin-7-*O*-rutinoside, apigenin-7-*O*-glucoside, glucosyl-6-pedalitin, 4'-hydroxyl wogonoside, and acacetin.

The UPLC-MS results showed that the main compounds in *V. officinalis* were iridoid glycosides, phenylethanoid glycosides, and flavonoids. Literature research shows that iridoid glycosides are characteristic components of *V. officinalis*, and verbenalin has a high content (Council of Europe, 2020). As a representative of phenylethanoid glycosides, acetoside also has a high content in *V. officinalis* (Deepak and Handa, 2000). The above two components were identified in UPLC-MS, and flavonoids were also identified in the experiment. Therefore, we consider that the above three compounds are the main anti-AD active

3 https://itol.embl.de/

TABLE 1 Result of UPLC-MS in *Verbena officinalis*.

Type	Compound	Molecular weight	Measured mass (*m/z*)	Ion mode	Retention time	MS/MS fragment
Iridoid glycosides	Aucubin	346.3	345.10	$[M-H]^-$	1.11	183.09, 179.06, 167.11, 149.09
	Swertiamarine	374.3	373.11	$[M-H]^-$	1.21	419.12, 179.06, 149.04, 141.02
	Hastatoside	404.4	427.12	$[M + Na]^+$	6.44	243.09, 225.08, 207.06, 193.05
	Gentiopicroside	356.3	357.12	$[M + H]^+$	6.77	195.07, 177.05
	Verbenalin	388.4	389.14	$[M + H]^+$	7.03	357.12, 227.09, 195.06
	3,4-Dihydroverbenalin	390.4	391.16	$[M + H]^+$	7.87	229.11, 197.08
Phenylethanol glycosides	2'-Acetylacetoside	666.2	665.21	$[M-H]^-$	1.08	503.16, 179.05
	Acetoside	624.6	623.20	$[M-H]^-$	9.34	461.20, 315.11, 161.02
	Cistanoside F	488.4	487.14	$[M-H]^-$	10.04	179.03, 161.02, 135.04
	Jionoside D	638.6	637.21	$[M-H]^-$	10.53	461.07, 193.05, 175.07, 135.04
	Isoacteoside	624.6	623.20	$[M-H]^-$	10.99	461.17, 161.02
Flavonoids	Apigenin-7-*O*-glucoside	432.3	433.10	$[M + H]^+$	7.56	445.08, 269.05
	Glucosyl-6-pedalitin	478.1	477.09	$[M-H]^-$	8.15	300.06, 271.91, 166.02
	4'-Hydroxyl wogonoside	462.1	463.28	$[M + H]^+$	22.49	477.27, 301.20, 269.21
	Acacetin	284.0	283.04	$[M-H]^-$	24.14	241.03, 159.03, 153.09, 135.04, 115.07, 111.08
	Acacetin-7-*O*-rutinoside	592.5	593.28	$[M + H]^+$	25.32	447.32, 285.20, 270.21, 242.28

components of *V. officinalis*. These components are the main direction to study the anti-AD effect of *V. officinalis*.

Molecular docking

To investigate the potential of chemical components for the treatment of AD, we chose target proteins for docking with compounds in *V. officinalis*. Chronic inflammation is suspected to be associated with the occurrence of AD. The activation of NLRP3 leads to the release of cytokines, which plays a key role in the pathogenesis of AD (Shao et al., 2015; Chen D. B. et al., 2020). The deposition of Aβ produces neurotoxicity, which maybe leads to the occurrence of AD. Beta-site APP Cleaving Enzyme1 (BACE1) is the rate-limiting enzyme for Aβ, and its activity reduction can prevent Aβ deposition (Moussa-Pacha et al., 2020). Elenbecesta was used as a control to measure the effect of the compounds on BACE1 (Das et al., 2021). Acetylcholinesterase (AChE) can degrade acetylcholine. When the AChE activity increases, the concentration of acetylcholine decreases, which may lead to the loss of acetylcholine at cholinergic synapses, and lead to cognitive and memory impairment. Donepezil is a clinically applied ChE inhibitor and is used as the reference drug in this experiment (Ambure et al., 2014). Hyperphosphorylation of tau protein will destroy the structure and physiological function of neurons. The expression level of GSK-3β is related to the phosphorylation of tau protein. In the experiment, TWS119 was used as a positive control drug (Eskandarzadeh et al., 2021). The NLRP3, BACE1, AChE, and GSK-3β are closely related to Alzheimer's disease. **Supplementary Figure 2** shows the relationship between these target proteins and AD (Loera-Valencia et al., 2019).

Verbenalin, hastatoside, aucubin, swertiamarine, acetoside, isoacteoside, jionoside D, cistanoside F, acacetin-7-*O*-rutinoside, apigenin-7-*O*-glucoside, glucosyl-6-pedalitin, and acacetin were used to molecular docking (**Table 2**). Verbenalin (**Figure 2A**), swertiamarin, and jionoside D have a good binding effect with NLRP3, which may play a role in the treatment of AD by reducing inflammation. Phenylethanoid glycosides including acetoside (**Figure 2B**) and isoacteoside have good docking effects with NLRP3 and BACE1, which may reduce the inflammatory response and inhibit Aβ deposition. Apigenin-7-*O*-glucoside (**Figure 2C**) and glucosyl-6-pedalitin were higher than positive drugs in molecular docking with NLRP3, BACE1, and AChE. Acacetin-7-*O*-rutinoside (**Figure 2D**) has a good binding effect on all four protein targets, which is better than that of apigenin-7-*O*-glucoside, glucosyl-6-pedalitin. Acacetin also has multi-target effects.

De novo assembly and unigene annotation

There were the sequencing data of 9 cDNA libraries (3 biological replicates in each tissue) and obtained 62.55 GB clean data. The clean data of each sample reached more than 6 GB. The overall sequencing error rate of each sample was 0.03%

TABLE 2 Docking results of *Verbena officinalis*. compounds with AD target protein molecules.

Type	Compounds	NLRP3 Binding affinity (kcal/mol)	BACE1 Binding affinity (kcal/mol)	AChE Binding affinity (kcal/mol)	GSK-3β Binding affinity (kcal/mol)
Positive control	Itanapracedyang	−7.9	–	–	–
	Elenbecesta	–	−8.2	–	–
	Donepezil	–	–	−9.8	–
	TWS119	–	–	–	−9.8
Iridoid glycosides	Verbenalin	**−8.4**	−7.9	−7.9	−8.8
	Aucubin	−7.5	−7.1	−7.4	−7.1
	Swertiamarine	**−8.1**	−7.4	−7.4	−8.4
Phenylethanol Glycosides	Acetoside	**−8.8**	**−8.8**	−8.5	−8.0
	Isoacteoside	**−9.0**	**−8.5**	−9.1	−9.4
	Jionoside D	**−8.9**	−8.2	−8.7	−8.7
	Cistanoside F	−7.7	−7.3	−7.3	−7.6
Flavonoids	Acacetin-7-*O*-tinoside	**−10.6**	**−9.5**	**−9.6**	**−10.2**
	Apigenin-7-*O*-glucoside	**−9.8**	**−9.1**	**−10.0**	−9.1
	Glucosyl-6-pedalitin	**−9.6**	**−8.4**	**9.9**	−9.3
	Acacetin	**−8.3**	**−8.3**	9.3	−8.5

The words in red are those with a binding energy greater than the positive drug.

(the error rate of all libraries was 0.03%), Q20 base ratio more than 97.21%, Q30 base ratio more than 92.3%, indicating that the data were qualified and meet the analysis requirements (**Supplementary Table 1**).

High-quality sequences were spliced by Trinity to obtain 2,45,315 transcripts, with an average length of 1,208 bp. After removing redundant and clustering analysis, 92,867 unigenes were obtained, with an average length of 10,44bp (**Supplementary Table 2** and **Figure 3A**).

Unigenes were compared to 7 databases for gene function annotation. Unigenes annotated in NR for 57.11% of the total base number, NT 40.71%, Pfam 47.43%, go 47.43%, Ko 24.76%, Swissprot 50.83%, and KOG 21.51% (**Figure 3B**). In addition, 71.68% of unigenes in *V. officinalis* were annotated in at least one database and 7.78% of genes were annotated in seven databases.

In the comparison of the NR database, it was found that the species with a high matching degree with *V. officinalis* similar sequence are *Sesamum indicum* (accounting for 25.3%), followed by *Handroanthus impetiginosus* (accounting for 16.6%), *Erythranthe guttata* (accounting for 7.7%), *Quercus suber* (accounting for 4.7%), and the others less than 2.5% (**Figure 4A**).

The function prediction and classification statistics of unigenes can be carried out through the KOG database. A total of 19,977 unigenes were annotated in the KOG database and were classified according to 26 groups (**Figure 4B** and **Supplementary Table 3**). The five highest numbers are Post-translational modification (2,720 unigenes), General function prediction only (2,514 unigenes), Translation, ribosomal structure and biogenesis (2,013 unigenes), Signal translation mechanisms (1,940 unigenes), and Internal trafficking and secret and vesicular transport (1,508 unigenes).

There were 4 KEGG metabolic pathways in *V. officinalis* transcriptome, including 34 branches and 302 metabolic pathways (**Figure 4C** and **Supplementary Table 4**). Among them, the top five pathways are Ribosome (1,043 unigenes), Carbon metabolism (832 unigenes), Spliteosome (792 unigenes), Protein processing in endoplastic reticulum (722 unigenes), and Biosynthesis of amino acids (647 unigenes).

GO functions are divided into biological process, cellular component, and molecular function. In the annotation results, unigenes of the three types of functions are involved, and 43 branches are annotated in the second level (**Figure 4D** and **Supplementary Table 5**). In the biological process category, it mainly involves cellular process (25,366 unigenes), metallic process (23,237 unigenes), and biological regulation (8,832 unigenes). In the cell group classification, it mainly includes cellular analytical entity (18,034 unigenes), intracellular (10,686 unigenes), and protein-containing complex (7,873 unigenes). Among the molecular functional categories, they are mainly concentrated in binding (24,152 unigenes), catalytic activity (19,166 unigenes), and transporter activity (3,289 unigenes).

Differential expression analysis

To explore the metabolic differences of *V. officinalis* in different tissues, the samples were analyzed by differential expression analysis. A total of 92,867 unigenes were annotated in *V. officinalis* transcriptome, including 15,715 DEGs (**Figure 5A**).

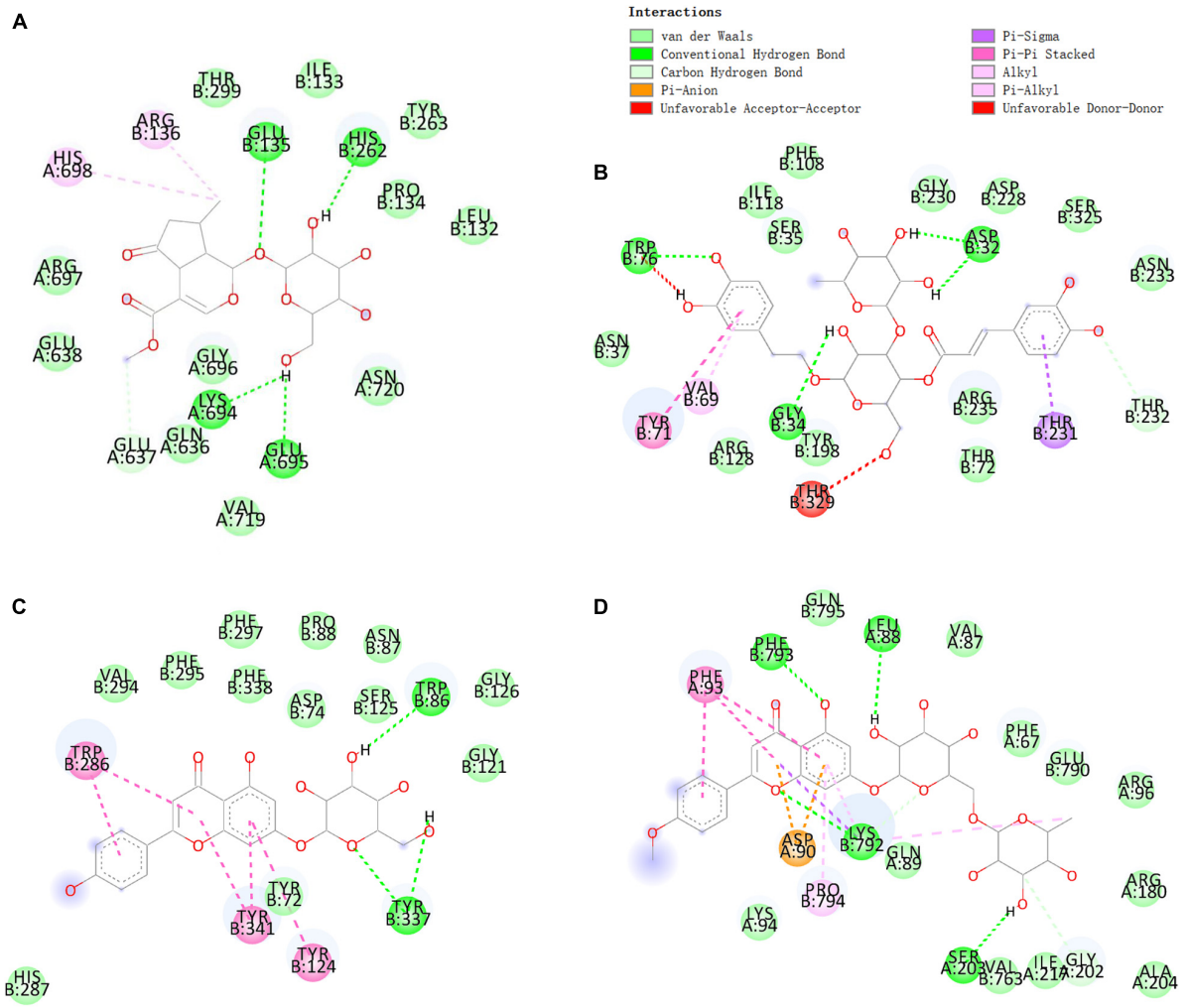


FIGURE 2
Molecular docking of some compounds in *Verbena officinalis*. with AD targets. **(A)** verbenalin docking with NLRP3. **(B)** Acetoside docking with BACE1. **(C)** Apigenin-7-*O*-glucoside with AChE. **(D)** Acacetin-7-*O*-rutinoside docking with GSK-3β. The figure on the upper left explains that different colors represent different types of compounds that bind with amino acids.

Under the condition of $p < 0.05$ and $|\log2foldchange| > 1$, the differential genes between the tissues were screened.

There are 12,939 DEGs in the leaf vs. root group (Figure 5B and Supplementary Table 6), with the largest number, and the up-regulated genes account for 47.0%. Therefore, it is speculated that there may be great differences in contents of secondary metabolites in leaf vs. root. Leaf vs. stem is the least, with 2,822 DEGs and 45.5% up-regulated genes (Figure 5C and Supplementary Table 7). Root vs. stem-group with 8,563 DEGs (Figure 5D and Supplementary Table 8), and the up-regulated genes accounted for 54.1%. The results suggest that there is little difference between leaves and stems, while more difference between leaves and roots. There may be some differences in the types and contents of anti-AD active components in the three tissues. Through differential expression analysis, the differential expression of some genes can also be retrieved. According to the distribution characteristics of tissues with the same active ingredients and gene expression, we consider that there is a certain correlation between metabolites and genes. The results showed that there were many differentially expressed genes in the leaves and roots of *V. officinalis*, suggesting that the active components in the leaves were significantly different from those in the roots.

A total of 37,906 unigenes were expressed in the root, stem, and leaf. And 42,313 unigenes were expressed in only one tissue (Supplementary Figure 3). In the comparison of differences between groups, there were 427 DEGs in the three groups of leaf vs. stem, leaf vs. root, and root vs. stem (Figure 6A). To explore the functions and possible regulatory mechanisms of these DEGs, we classified the differential genes and mapped them to the database. The results showed that there were 2,812 DEGs in the leaf vs.

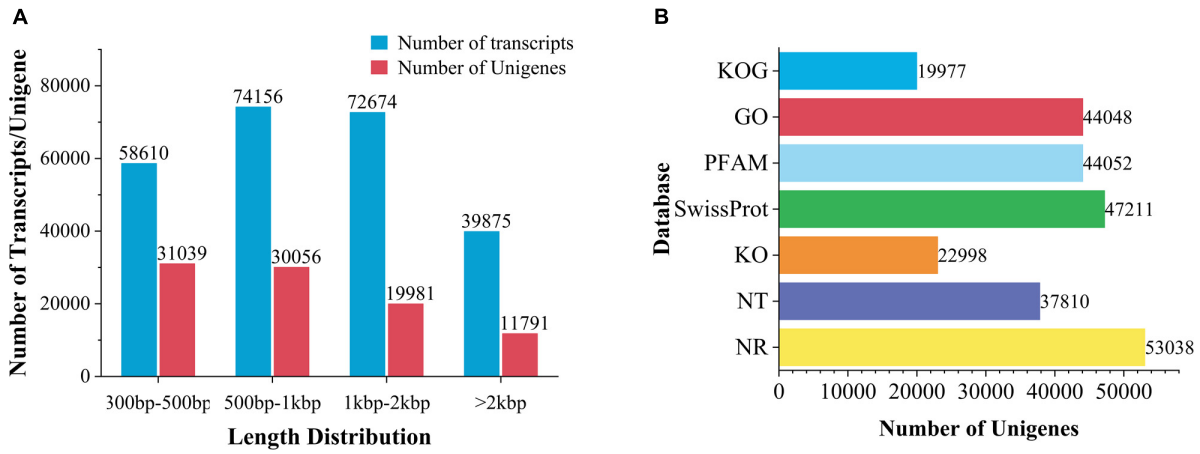


FIGURE 3
Transcription splicing and database annotation. **(A)** Transcripts and Unigenes length distribution. **(B)** Number of gene annotations in the database.

root group, which were mapped in 120 KEGG pathways (Figure 6B). The top five pathways were plant pathway interaction (137), Plant hormone signal transmission (114), start and cross metabolism (108), phynopropanoid biosynthesis (94), and Glyoxylate and dicarboxylate metabolism (68). In the KEGG database, 519 DEGs of leaf vs. stem were mapped in 96 pathways (Figure 6C), mainly concentrated in the pathways such as Phylpropanoid biosynthesis (50), Start and sucrose metabolism (33), Glycoxylate and dicarboxylate metabolism (24), Plant hormone signal transmission (24), and Glycolysis/gluconeogenesis (23). The 2,010 DEGs of root vs. stem-group are mapped in 120 pathways (Figure 6D) and mainly reflected in Plant pathway interaction (86), Phenotropic biosynthesis (74), Plant hormone signal transmission (72), Glyoxylate and dicarboxylate metabolism (48), and Pyruvate metabolism (48).

Content determination results

The accumulation of iridoid glycosides, phenylethanoid glycosides, and flavonoids in different tissues of *V. officinalis* was studied by content determination. We determined the contents of verbenalin and acteoside by UPLC. From the result, we can know that the contents of the leaf, stem, and root are 4.19 ± 0.09, 4.16 ± 0.03, and 11.33 ± 1.09, respectively (Figure 7A). However, there are some differences in the content. The content of leaf and stem is similar, and roots are significantly up-regulated. The contents of acteoside in leaf, stem, and root were 50.94 ± 1.11, 9.92 ± 0.08, and 9.08 ± 0.97 mg/g, respectively (Figure 7B). The contents of the stem and root were similar and the leaf was significantly up-regulated, which was the same as our understanding that the upper part was an effective component, but verbenalin was not so.

Therefore, the contents of total iridoid glycosides, total phenylethanoid glycosides, and total flavonoids in *V. officinalis* were determined by ultraviolet spectrophotometry. The contents of total iridoid glycosides in the leaves, stems and roots were 17.60, 7.86, and 7.00% (Figure 7C), phenylethanoid glycosides were 10.50, 3.84, and 3.06% (Figure 7D), and flavonoids were 4.65, 1.02, and 0.82% respectively (Figure 7E). The common feature is leaf > stem > root. The results suggest that the genes related to the biosynthetic pathway of the above compounds in *V. officinalis* are highly expressed in the leaf.

Transcriptome analysis of compounds

The biosynthesis pathway of iridoids mainly involves Terpenoid backbone biosynthesis, Sesquiterpenoid and triterpenoid biosynthesis, Indole alkaloid biosynthesis, Isoquinoline alkaloid biosynthesis, Tropane, piperidine, and pyridine alkaloid biosynthesis. The main pathways involved in the biosynthesis of phenylethanoid glycosides are Tyrosine metabolism, Phenylalanine metabolism, Tryptophan metabolism, and Phenylpropanoid biosynthesis. The main pathways involved in flavonoids biosynthesis are Phenylalanine metabolism, Phenylpropanoid biosynthesis, Flavone and flavonol biosynthesis, and Flavonoid biosynthesis. Annotations of the related pathway genes are shown in Table 3.

Biosynthesis of iridoid glycosides

There are differences in the expression of genes related to iridoid glycoside synthesis in the root, stem, and leaf of *V. officinalis*. Iridoid glycosides are characteristic compounds in *V. officinalis*. They are special monoterpenoids. Their biosynthesis pathway is the same as that of other terpenoids (Figure 8A). Through transcriptome analysis, 206 unigenes

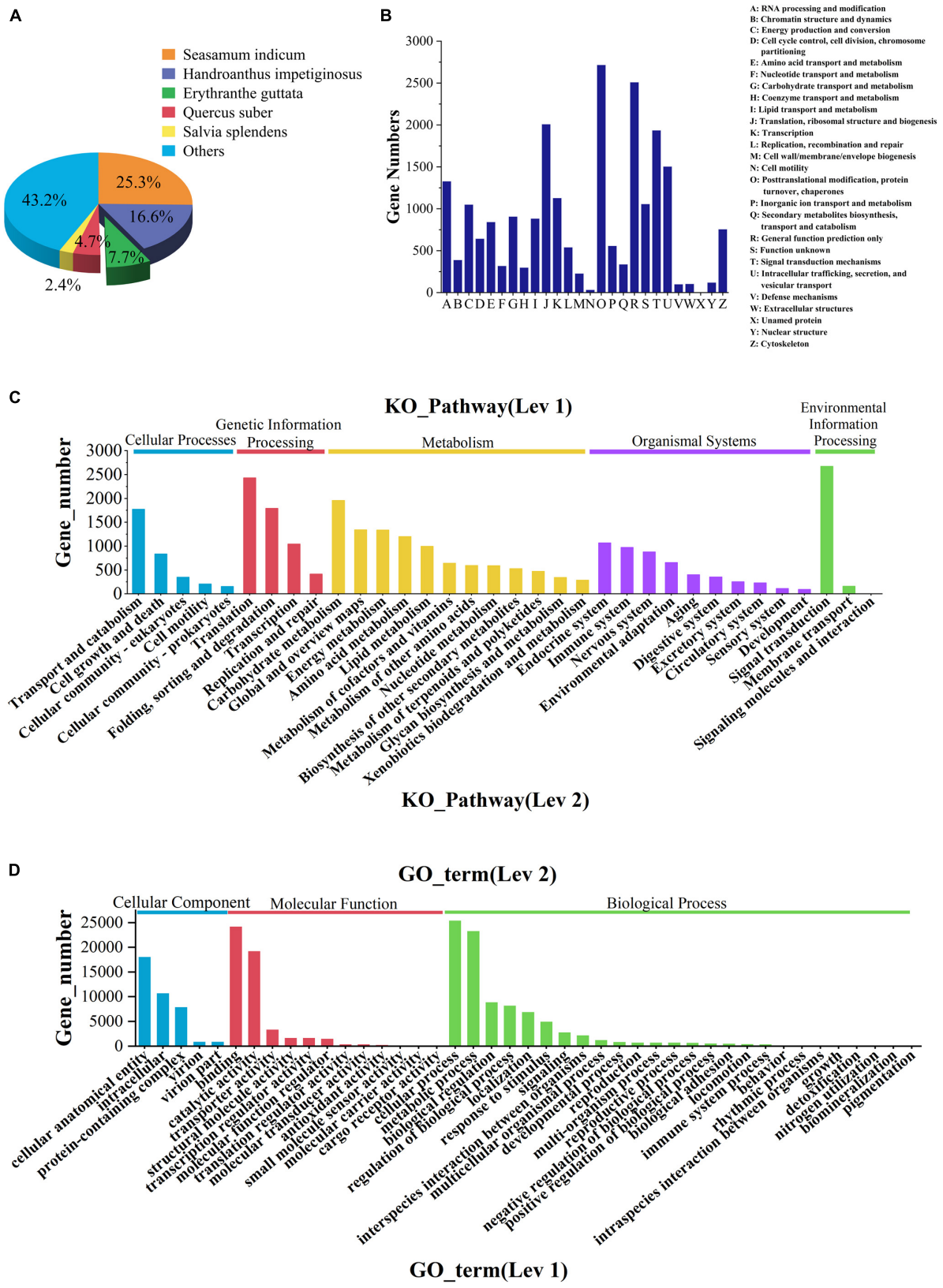


FIGURE 4
Database annotation classification. **(A)** Species distribution of transcriptomic unigenes against NR database. **(B)** KOG classification. **(C)** KEGG pathway classification. **(D)** GO classification.

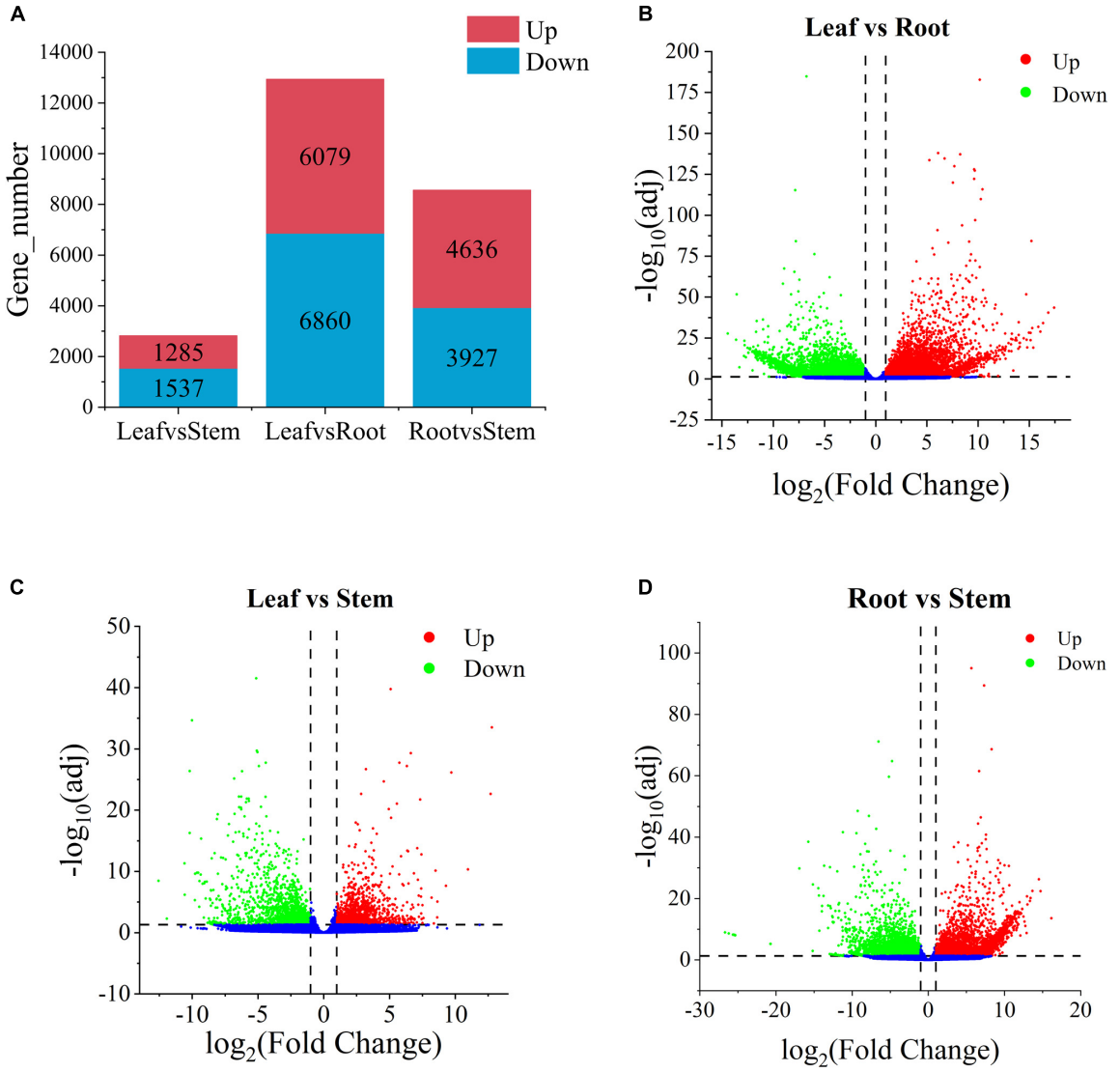


FIGURE 5
Histogram of the number of differential genes in each tissue of and the volcanoplots of differential genes in *Verbena officinalis*. **(A)** Histogram of differential genes. **(B)** Volcanoplot of Leaf vs. Root. **(C)** Volcanoplot of Leaf vs. Stem. **(D)** Volcanoplot of Root vs. Stem DEGs. The screening criteria for DEGs are $p < 0.05$ and $|\log_2\text{FoldChange}| > 1$ by DESeq2.

involving the above pathways were annotated, including 61 DEGs. The heat map was drawn based on the FPKM value of the differential genes (**Figure 8B** and **Supplementary Table 9**).

Modern studies have shown that the important precursors are IPP and DMAPP. These two substances have two recognized production pathways. One is to take acetyl CoA as the material (Tholl, 2015; Bergman et al., 2019), through acetyl CoA acetyltransferase (ACAT), the MVA pathway of HMG CoA synthase (HMGS), HMG CoA reductase (HMGCR), MVA kinase (MK), phosphomevalonate kinase (PMK), and diphosphomevalonate decarboxylase decarboxylase (MVD). Through database annotation and sequence comparison, we annotated 48 MVA pathway genes in *V. officinalis* transcriptome. There were *ACAT* (cluster-10847.0), *HMGCR* (cluster-10611.1) and MVD (cluster-7516.0) were DEGs and all of them up-regulated in roots. The results suggest that the biosynthesis pathway of terpenoids in *V. officinalis* root may be through the MVA pathway.

The other is the MEP pathway (Volke et al., 2019). The materials are pyruvate and G3P. The enzymes include 1-deoxy-D-xylulose-5-phosphate synthase (DXS), 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR), 2-C-methyl-D-erythrol 4-phosphate cydylyl transfer (CMS),

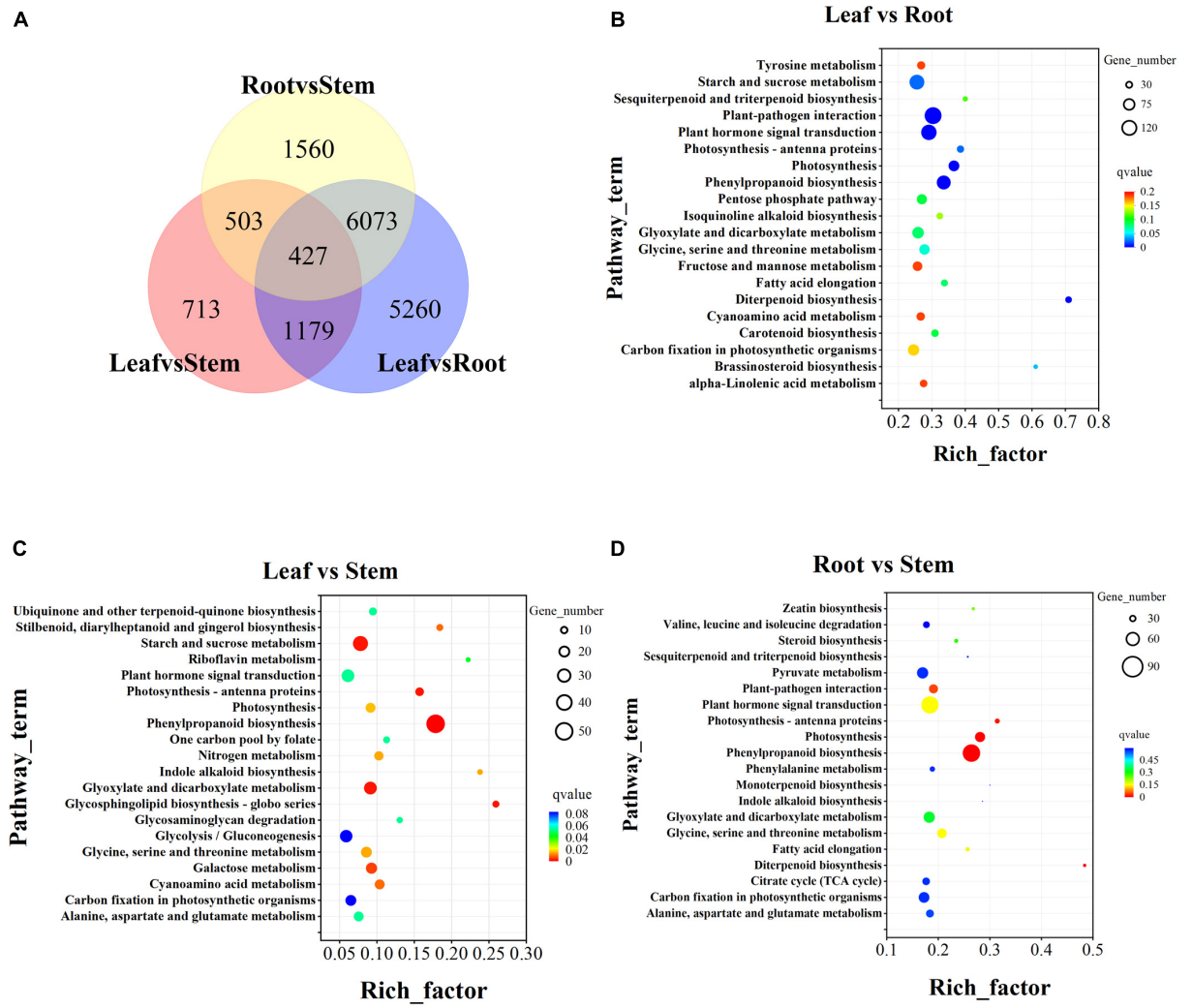


FIGURE 6
Transcriptome difference analysis. **(A)** Differential venn diagram of three tissues and differential venn diagram between groups. **(B)** Top 20 of KEGG pathway enrichment of Leaf vs. Root DEGs. **(C)** Top 20 of KEGG pathway enrichment of Leaf vs. Stem DEGs. **(D)** Top 20 of KEGG pathway enrichment of Root vs. Stem DEGs. The screening criteria for DEGs are $p < 0.05$ and $|\log_2 \text{FoldChange}| > 1$ by DESeq2.

4-diphosphodyl-2-C-methyl-d-phosphate kinase (CMK), 2-C-methyl-d-erythritol 2,4-cyclodiphosphate synthase (MCS), (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (HDS), and 4-hydroxy-3-methylbut-2-en-1-yldiphosphate reductase (HDR). In recent years, more and more studies tend to believe that the precursors IPP and DMAPP of terpenoid synthesis are mostly provided by the MEP pathway (Ma et al., 2017; Zhu et al., 2022). In this study, 39 single genes were annotated in the MEP pathway, including 14 DEGs. When analyzing the differential expression of genes, we found that *DXS*, *DXR*, *MCT*, *MCS*, *HDS*, and *HDR* in the MEP pathway are highly expressed in the leaf, suggesting that the synthesis of terpenoids in the leaf is high and realized through MEP pathway.

In the downstream pathway, enzyme genes are also highly expressed in leaves, such as geranyl diphosphate synthase (GPPS), geranyl diphosphate diphosphate (GES), geraniol-10-hydroxylase (G10H), 10-hydroxygeranioxido reduction (10-HGO), iridoid synthesis (IS), and so on. Isomerization between IPP and DMAPP is realized by isopentenyl diphosphate delta isomerase (IDI). *GPPS* is the key enzyme in the formation of GPP from IPP and DMAPP. Simkin et al. identified the enzyme *CrGES* that catalyzes geranyl diphosphate to produce geraniol from *C. roseus* for the first time (Simkin et al., 2013). It should be noted that only two *GES* genes were annotated in the transcriptome, of which cluster-193.45926 had a low expression in all three tissues (FPKM < 1), and another *GES* (cluster-193.25638) had a high expression in leaves. Colu et al. identified *G10H* as *CYP76B6* in *C. roseus*. This enzyme can catalyze geraniol to produce 10-hydroxygeraniol (Collu et al., 2001). Then, 10-hydroxygeraniol is catalyzed by *10-HGO* to

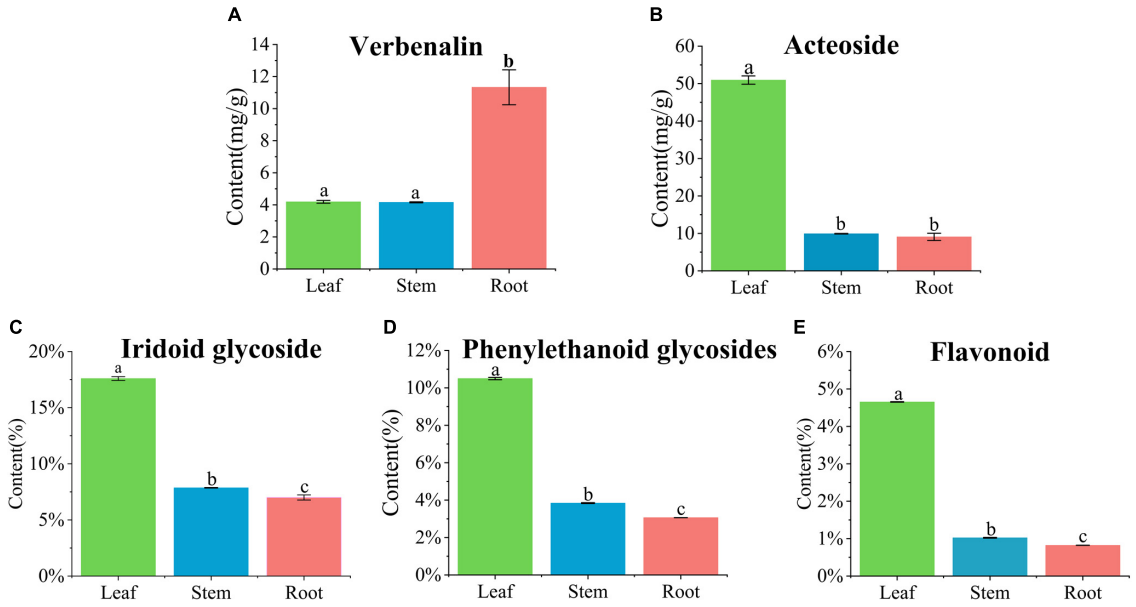


FIGURE 7
Quantification in the leaf, stem, and root of *Verbena officinalis*. **(A)** Verbenalin. **(B)** Acteoside. **(C)** Total iridoid glycosides. **(D)** Total phenylethanol glycosides. **(E)** Total flavonoids. Values are expressed as the means ± standard errors of three independent samples. Significant differences ($p < 0.05$) were analyzed using Origin2021 and indicated by lowercase letters a, b, and c in the leaf, stem, and root.

TABLE 3 Pathways involved in biosynthesis of iridoid glycosides, flavonoids and phenylethanol glycosides.

Pathway name	KO ID	Input number	DEGs number	Leaf vs. Stem	Leaf vs. Root	Stem vs. Root
Tyrosine metabolism	ko00350	127	39	5	34	19
Phenylalanine metabolism	ko00360	122	61	9	26	23
Tryptophan metabolism	ko00380	104	21	3	12	13
Terpenoid backbone biosynthesis	ko00900	153	30	9	28	17
Indole alkaloid biosynthesis	ko00901	30	15	5	9	6
Monoterpenoid biosynthesis	ko00902	22	9	–	8	6
Sesquiterpenoid and triterpenoid biosynthesis	ko00909	37	17	2	14	9
Phenylpropanoid biosynthesis	ko00940	282	127	50	94	74
Flavone and flavonol biosynthesis	ko00944	4	–	–	–	–
Flavonoid biosynthesis	ko00941	60	16	4	12	10
Isoquinoline alkaloid biosynthesis	ko00950	68	24	–	22	12
Tropane, piperidine and pyridine alkaloid biosynthesis	ko00960	53	14	–	13	10

generate 10-oxogeranial. There are three genes annotated in which the expression of cluster-193.33255 is low. The other two genes *10-HGO1* (cluster-193.30177) and *10-HGO2* (cluster-193.30279) are differential genes, which are highly expressed in leaf. Therefore, *GES*, *10-HGO1* and *10-HGO2*. Then *IS*, iridoid oxidation (IO), 7-deoxyloganic acid glucosyl transferase (7-DLGT), a series of key enzymes catalyze the formation of 7-deoxyglycic acid, which may then be hydroxylated, methylated, and oxidized to produce iridoid glycosides (Sun et al., 2012; Krithika et al., 2015; Ji et al., 2017; Ye et al., 2019).

The genes *IO6* (cluster-193.30787), *HDS2* (cluster-193.31836), and *HDR3* (cluster-193.30998) have the highest FPKM values in this pathway, and they are significantly up-regulated in leaf. Generally speaking, the number of genes with high expression in roots is relatively small, while the expression of related genes in stems is generally low, suggesting that the content of related iridoids glycoside is high in the leaf of plants. Thus, *GES*, *10-HGO1,* and *10-HGO2* may be the key enzymes in this process.

Iridoid glycosides are characteristic compounds in *V. officinalis*, mainly including verbenalin, hastatoside, gentiopicroside, aucubin, *etc*., in *V. officinalis* tissue. The total iridoid glycosides in the leaf were significantly increased, but the content of verbenalin in the root was the highest. Five

10.3389/fpls.2022.955075

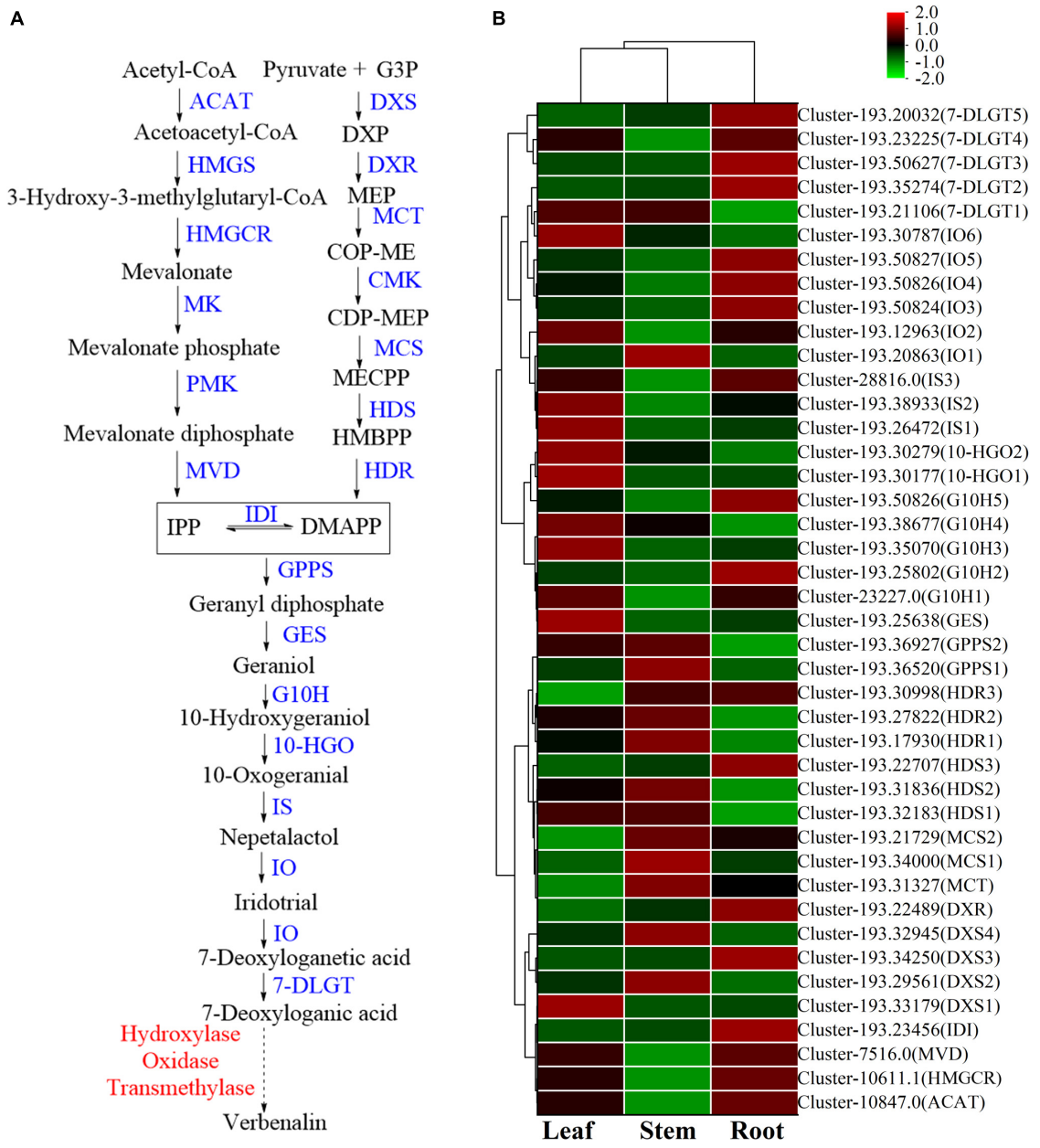


FIGURE 8
The iridoid glycosides biosynthesis pathway and the heatmap of corresponding genes in *Verbena officinalis*. **(A)** Predicted biosynthetic pathways of iridoid glycosides. Compounds are in black, enzyme names are in blue, and the red have not been reported. The solid arrows are the processes that have been reported, and the dashed lines have not been reported. **(B)** Heatmap of iridoid glycosides pathway DEGs.

7-DLGT DEGs were identified in *V. officinalis* transcriptome, of which cluster-193.35274, cluster-193.50627, cluster-193.23225, and cluster-193.20032 were significantly up-regulated in roots. Meanwhile, the *IO* candidate genes cluster-193.12963, cluster-193.50824, cluster-193.50826, and cluster-193.50827 also showed significant up-regulation in the root. Therefore, the higher expression of these genes in roots may be related to the tissue distribution of verbenalin. The difference in tissue distribution between verbenalin and other iridoid glycosides may be due to the branching of verbenalin in the biosynthetic pathway, and the highly expressed genes *7-DLGT* and *IO* in the root may play a role in the biosynthesis of verbenalin.

Biosynthesis of phenylethanoid glycosides

The content of phenylethanoid glycosides in *V. officinalis* is high. Acteoside is a typical representative of phenylethanoid glycosides. Its structure is composed of caffeic acid hydroxytyrosol, glucose, and C3-modified rhamnose. The

biosynthesis of acteoside (Saimaru and Orihara, 2010) mainly includes the phenylalanine metabolic pathway, dopamine pathway/tyramine pathway, and the glycosylation pathway (**Figure 9A**). In the *V. officinalis* transcriptome database, the KEGG pathways mainly include phenylpropanoid biosynthesis and tyrosine metabolism. In the above synthetic pathways, 229 related genes were annotated, including 73 DEGs (**Figures 9B,C** and **Supplementary Table 10**). The results of the UPLC-MS analysis showed that the phenylethanoid glycosides in *V. officinalis* mainly included acteoside, isoacteoside, and 2'-acetylacetoside.

On the one hand, the phenylalanine metabolic pathway (Zhang et al., 2019; Yang et al., 2021) takes phenylalanine as raw material and generates cafeyl CoA through phenylalanine ammonia lyase (PAL), *trans* cinnate 4-hydroxylase (C4H), 4-coumarate CoA ligase (4CL), shikimate o-hydroxy cinnamoyl transferase (HCT), coumarate-3-hydroxylase (C3H), and caffeoyl shikimate esterase (CSE). In *V. officinalis* transcriptome, the *PAL* (10), *C4H* (10), *4CL* (44), *HCT* (22), *C3H* (3), and *CSE* (18) were all annotated. Most of the candidate genes of *4CL*, *HCT,* and *CSE* were significantly up-regulated in leaves. On the other hand, in the dopamine pathway/tyramine pathway, tyrosine is catalyzed by polyphenol oxidase (PPO), tyrosine decarboxylase (TyDC), primary amine oxidase (AOC3), and aldehyde dehydrogenase (ALDH) to produce 3.4-dihydroxyphenylthanol. It is connected with a molecule of glucose to produce 3-hydroxysalidroside. Meanwhile, tyrosine can also be catalyzed by *TyDC*, *AOC3*, *ALDH,* and *UGT* to produce salidroside. Then 3-hydroxysalidroside is formed by hydroxylation. All the above genes are annotated in the transcriptome, *PPO*, *TyDC,* and *UGT* were significantly up-regulated in the leaf. So far, the synthetic precursors of acteoside, cafeyl CoA, and 3-hydroxysalidroside have been generated, and then the final product is generated through glycosyl modification and cross-binding reaction, but the downstream process is not clear and needs to be further explored (Alipieva et al., 2014; Zhou et al., 2020).

By referring to the existing research and KEGG pathway, the biosynthetic pathway of acteoside in *V. officinalis* was speculated, and the genes of the phenylalanine metabolic pathway and dopamine pathway/tyramine pathway were all annotated. Therefore, this biosynthetic pathway is feasible in *V. officinalis*. In addition, the genes with the highest expression in this pathway are *CSE* (cluster-193.31738) and *PPO* (cluster-193.31298), which were significantly up-regulated in the leaf, which was consistent with the content determination results.

Biosynthesis of flavonoids

Flavonoids are common compounds in medicinal plants and one of the main chemical components in *V. officinalis*. Its biosynthetic pathway is relatively clear (**Figure 10A**). KEGG pathway involved in flavonoid biosynthesis pathway and mainly includes Phylpropanoid biosynthesis, Flavone and flavonol biosynthesis, Flavonoid biosynthesis, *etc*. In the above synthetic pathways, 115 related genes were annotated, including 35 DEGs (**Figure 10B** and **Supplementary Table 11**), mainly concentrated in *4CL* (14), *FLS* (9). *V. officinalis* contains many flavonoids.

As the raw material for the synthesis of flavonoids, L-phenylalanine through *PAL*, *C4H*, *4CL*, chalcone synthase (CHS), chalcone isomerase (CHI). A series of enzymes catalyze the formation of flavonoid synthetic skeleton 2*S*-naringeni. In addition, 7 *CHS* and 17 *CHI* genes were annotated, and *CHI* included 4 DEGs. Then 2*S*-naringeni modified by flavone synthase (FNSI), flavanone-3-hydroxylase/naringenin 3-dioxygenase (F3H), Flavonol Synthase (FLS), flavanoid 3',5'-hydroxylase (F3'5'H) and other enzymes to form myricetin, dihydromyricetin, tricetin and other flavonoids (Zhao et al., 2020; Liu et al., 2021; Lou et al., 2021). All the above genes were annotated in *V. officinalis* transcriptome, including *FNSI* (2), *F3H* (3), *FLS* (14), *F3'5'H* (8), and *FNSI* (1). In addition, *FLS* (9) and *F3'5'H* (2) were DEGs.

The results of the UPLC-MS analysis show that it mainly includes acacetin-7-*O*-rutinoside, apigenin-7-*O*-glucoside, *etc*. In the determination of total flavonoids, it was the highest in the leaf. Through KEGG pathway analysis, we mapped the flavonoid biosynthesis pathway. The number of related genes with high expression in the root is relatively small, while more in the leaf. The candidate genes of *4CL*, *CHI*, *CHS*, *F3'5'H*, *FLS,* and other enzymes are significantly up-regulated in the leaf. Among them, *F3'5'H* has 8 candidate genes were annotated, and only *F3'5'H1* (cluster-193.34318) was significantly up-regulated in leaf, which was consistent with the content determination results. At the same time, it also suggested that these genes may be the rate-limiting enzyme of flavonoids in *V. officinalis*.

Transcription factors analysis

Transcription factors (TFs) play a vital role in plant growth. They affect the production and accumulation of plant secondary metabolites by promoting or inhibiting genes expression. A total of 3,717 genes were annotated in *V. officinalis* transcriptome, belonging to 90 categories of TFs (**Supplementary Table 12**). Among them, the largest number of genes was C2H2 (403, accounting for 12.71%), followed by bHLH (basic helix loop helix, 158, accounting for 4.98%), AP2/ERF ERF (apetala2/ethylene-responsive factor, 152, accounting for 4.79%), and C3H (cys3his zinc finger domain-containing protein, 131, accounting for 4.13%).

There are 111 VoWRKYs were annotated in *V. officinalis* transcriptome, accounting for 3.50% of the total transcription factor genes. To explore the role of WRKYs in the growth and development of *V. officinalis*, 16 gene sequences with $|\log_2 \text{FoldChange}| > 4$ in *V. officinalis* WRKY were screened (figure, the expression statistics of genes with

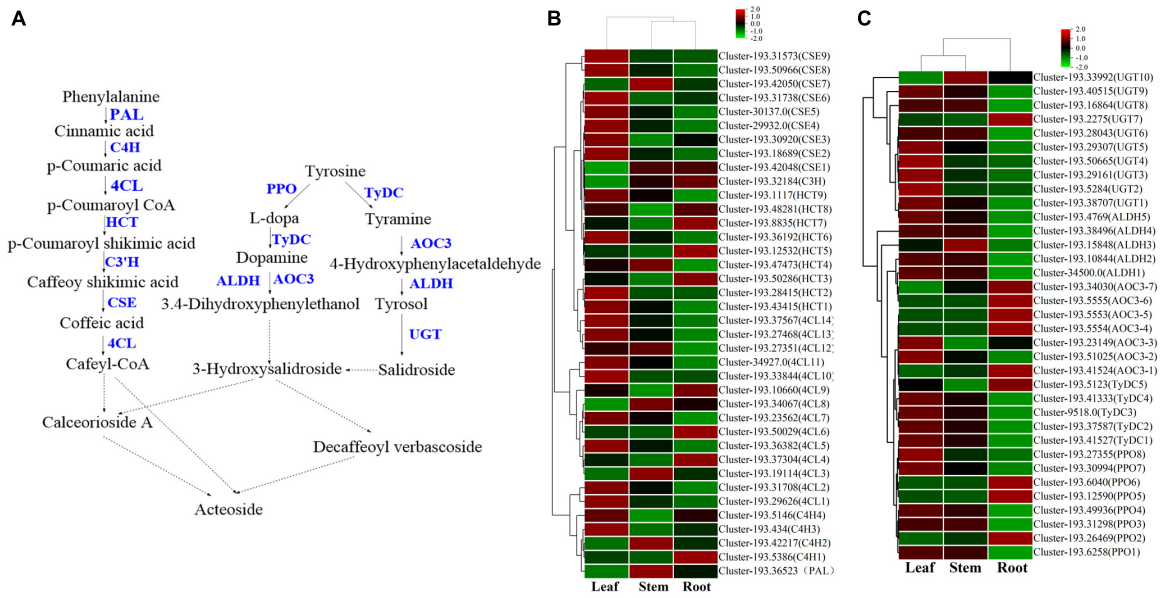


FIGURE 9
The phenylethanoid glycosides biosynthesis pathway and the heatmap of corresponding genes in *Verbena officinalis*. **(A)** Predicted biosynthetic pathways of phenylethanoid glycosides. Compounds are in black, enzyme names are in blue. The solid arrows are the processes that have been reported, and the dashed lines have not been reported. **(B)** Heatmap of phenylalanine metabolic pathway DEGs. **(C)** Heatmap of dopamine/tyramine pathway DEGs.

$|\log_2 \text{FoldChange}| > 4$ in *V. officinalis* WRKY). After removing redundant genes, 12 genes (**Supplementary Table 13**) were selected to build a phylogenetic tree with AtWRKYs, NtWRKY, DoWRKYs, CrWRKY, OsWRKY, and AaWRKYs (**Figure 11**). According to the results of phylogenetic analysis, VoWRKY6 and DoWRKY3 are in the same branch, and their gene sequences are similar. Therefore, VoWRKY6 may have similar functions to DoWRKY3, affect TPS, and regulate the biosynthesis of terpenoids in *V. officinalis*. VoWRKY7 is close to AaWRKY1 1, which may also play a role in regulating terpene biosynthesis. OsWRKY13 and AtWRKY23 are related to flavonoid biosynthesis. Therefore, VoWRKY3, VoWRKY9, and VoWRKY12, which are close to them, may also play a role in regulating the flavonoid biosynthesis of *V. officinalis*.

Discussion

Verbena officinalis is a medicinal plant of the genus *Verbena*. In addition, its relatives, *Verbena bonariensis* L., *Verbena brasiliensis* Vell, *Verbena hybrid* Voss, *Verbena tenera* Sprang, *Verbena hastata* L., *Verbena litoralis* Kunth, *etc*. also belong to the genus of *Verbena*, and most of which have good medicinal values and are used as folk medicine. To explore the genetic relationship between *Verbena* plants, Yuan et al. used the pentatricopeptide repeat gene family to analyze the phylogeny of *Verbena* plants. Hence, the phylogenetic framework was constructed, and it provided an important basis for the phylogenetic relationship of *Verbena* plants (Yuan et al., 2010). Ruzicka et al. constructed the phylogenetic tree of *Verbena* by comparing the internal transcribed spacer sequences and molecular markers of *Verbena*. Under the two conditions, the difference between *V. officinalis* and *V. hastata* was the smallest (Ruzicka et al., 2009).

Based on the genetic relationship and genetic similarity, *Verbena* plants may have similar secondary metabolites and pharmacological effects (Gong et al., 2022). De Lima et al. analyzed *V. littoralis* by UHPLC-ESI-HRMS and found that phenylethanoid glycosides, iridoid glycosides, and triterpenoids were identified in plants (De Lima et al., 2018). Soares et al. (2020) analyzed the main chemical components of *V. minutiflora* by HPLC-DAD. The results showed that iridoid glycosides, phenylethanoid glycosides, and flavonoids all had high contents in plants. el-Hela et al. (2000) isolated the chemical constituents of *V. bipinnatifida*, and successfully identified two iridoid glycosides and two phenylethanoid glycosides. A large number of literature studies have shown that iridoid glycosides and phenylethanoid glycosides are common in *Verbena* plants, and verbenalin, hastatoside, acteoside, and isacteoside are high content in these plants. This also suggests that these plants may have similar pharmacological activities.

At present, there are few studies on the *Verbena* plants. Only *V. officinalis* and *V. bonariensis* have some chemical constituents and pharmacological effects research. In previous reports, nuclear magnetic resonance (Shu et al., 2014; Luo et al., 2021), HPLC-DAD (Liu et al., 2012), HPLC-MS

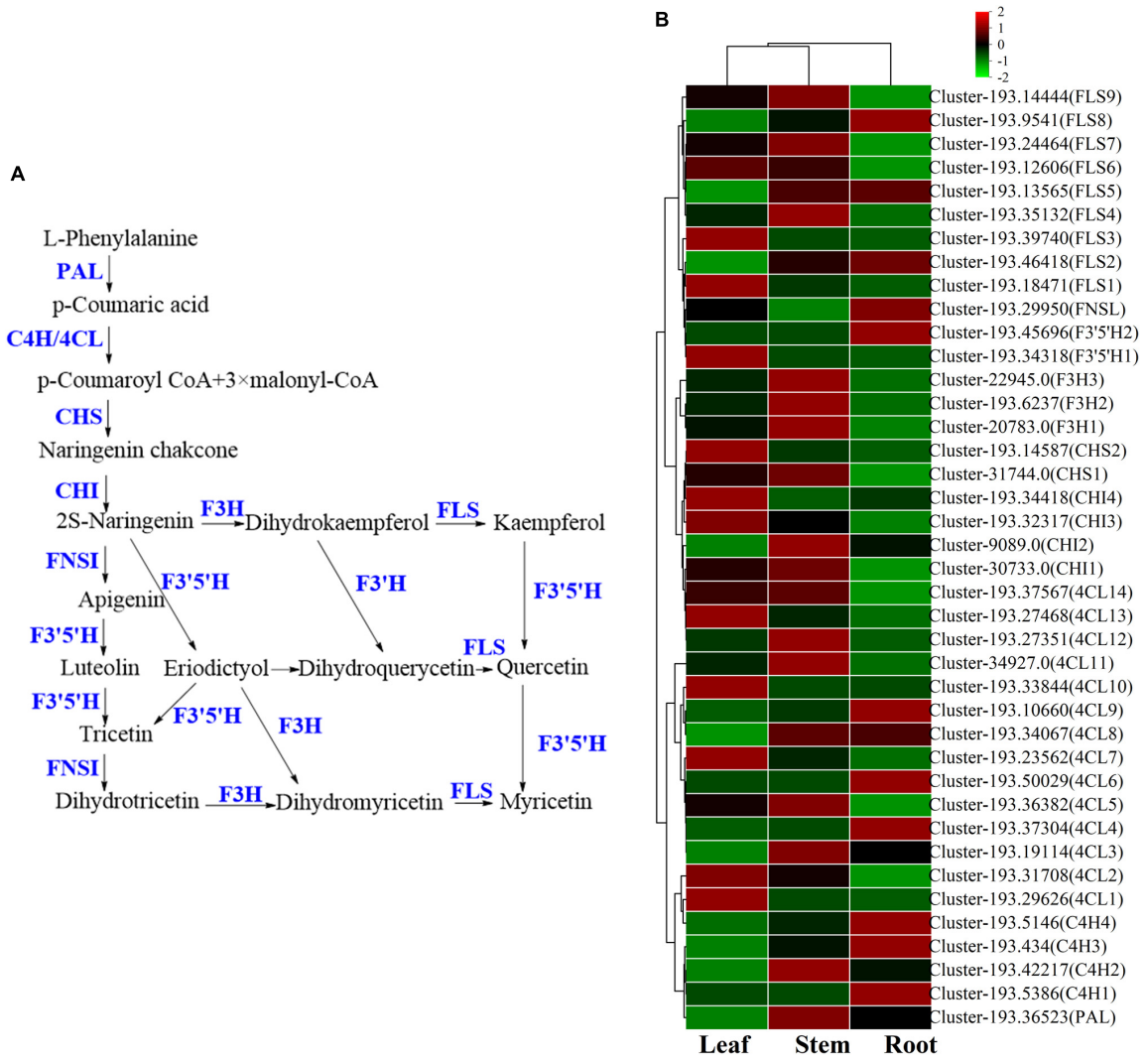


FIGURE 10
The flavonoids biosynthesis pathway and the heatmap of corresponding genes in *Verbena officinalis*. **(A)** Predicted biosynthetic pathways of flavonoids. Compounds are in black, while enzyme names are in blue. **(B)** Heatmap of flavonoids pathway DEGs.

(El-Hela et al., 2010) and other technologies were used to study the chemical components and content of *V. officinalis*. In this study, we used UPLC-MS technology to identify the chemical components in *V. officinalis* extract. We not only used the chromatographic separation technology, but also combined the high selectivity and strong structural identification ability of mass spectrometry. A total of 16 compounds in *V. officinalis* extract were detected by this method, and the structure of each compound was further confirmed by multi-stage full scan mass spectrometry. Iridoid glycosides, phenylethanoid glycosides, and flavonoids, as the main components of *V. officinalis*, are considered the main components for anti-AD effect. Therefore, we preliminarily investigated the activity of the compounds with the help of molecular docking technology. The iridoid glycosides, phenylethanoid glycosides, and flavonoids of *V. officinalis* showed good activity in the anti-AD molecular docking experiment. Iridoid glycosides showed good binding to inflammation-related proteins NLRP3, and phenylethanoid glycosides showed good binding to NLRP3 and BACE1. Flavonoids have a good binding effect with target proteins NLRP3, BACE1, AChE, and GSK-3β. This result suggested that the compounds in *V. officinalis* had a good potential for anti-AD activity.

To further explore the active components against AD, we divided *V. officinalis* tissue into three ones: leaf, stem, and root. The contents of verbenalin and acteoside were determined by UPLC. The results showed significant differences in the three tissues. However, the content of verbenalin is the highest in the root, which is contrary to the fact that the upper part is used as medicine. Therefore, we determined the tissue content

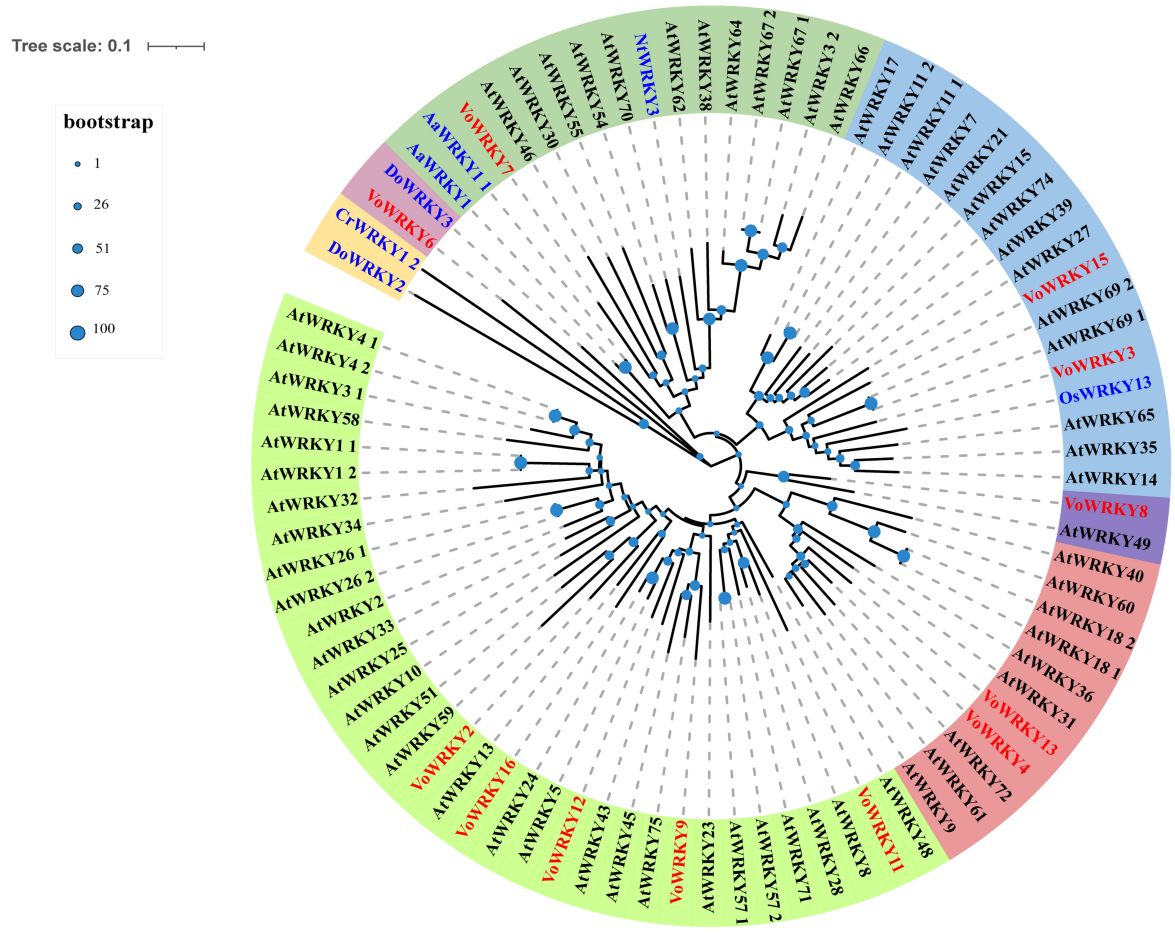


FIGURE 11
Phylogenetic tree of WRKY transcription factors in *Verbena officinalis*. Red font is *Verbena officinalis* TFs, black font is *Arabidopsis thaliana* transcription factor, and blue font is transcription factor of other species. The size of the blue node represents the bootstrap value.

differences of total iridoid glycosides, total phenylethanoid glycosides, and total flavonoids by a UV spectrophotometer. The three components have the same tissue distribution characteristics. The content is leaf > stem > root.

Meanwhile, we obtained the transcriptome date by RNA-Seq and annotated the gene functions through database comparison. In addition, through KEGG pathway enrichment and literature retrieval, we speculated and mapped the biosynthetic pathways of iridoid glycosides, phenylethanoid glycosides, and flavonoids. The related genes and DEGs of pathways were identified, and the key rate-limiting enzymes of iridoid glycosides may be *GES*, *10-HGO1* and *10-HGO2*. In addition, the tissue distribution of verbenalin is special, which may be related to the branching of the biosynthetic pathway in the downstream approach. The accumulation of metabolites in different tissues is not completely consistent with the level of gene expression, indicating that the genes related to substance synthesis and decomposition are differentially controlled or regulated. The speculation and gene identification of the phenylethanoid glycoside biosynthesis pathway have been predicted, but the search for key rate-limiting enzymes needs further experiments. *F3'5'H1* was speculated to be the key gene of flavonoid biosynthesis because of its unique up-regulation in candidate genes. TFs analysis also led us to speculate that the WRKY related to the biosynthesis of iridoid glycosides in transcriptome may be VoWRKY6 and VoWRKY7, while VoWRKY3, VoWRKY9, and VoWRKY12 may be involved in regulating the biosynthesis of flavonoids in *V. officinalis*.

However, our experiment has some shortcomings. There are many key enzymes involved in the biosynthetic pathways, and we have not verified their functions. In addition, the molecular docking technology verified that the anti-AD activity components of the compounds remained at the virtual level. Therefore, our next step is to verify the function of key enzymes involved in the pathway. At the same time, we also plan to improve the exploration of anti-AD activity to the pharmacological level, and establish an AD model

through model organism *Caenorhabditis elegans* to verify the activity of compounds.

Previously, there was no report on the analysis of *V. officinalis* transcriptome. Therefore, we creatively analyzed the biosynthesis pathway of iridoid glycosides, phenylethanoid glycosides, and flavonoids by using RNA-Seq technology, and combinedthe results with content determination to explore the biosynthetic pathway of anti-AD ingredients. First, UPLC-MS helped us to identify the main chemical constituents of *V. officinalis*. After confirming the existence of chemical components, we investigated the activities of these compounds by molecular docking technology and directly explored the anti-AD activity of the compounds in *V. officinalis*. This method of combining chemical analysis and transcriptome analysis is novel and has great application potential in the future. Through this technique, we investigated the possible regulatory mechanism of the active components in *V. officinalis*. Some compounds in *V. officinalis* show predominant multi-target anti-AD activity, so they can further target the pathogenesis. Natural products have significant advantages in finding new anti-AD agents and provide clues for the prevention and treatment of AD. Meanwhile, because *V. officinalis* has similar active components and gene sequences with *Verbena* plants, it is suggested that other *Verbena* plants may have similar pharmacological activities with *V. officinalis*. At the same time, the study of anti-AD active components in *V. officinalis* will provide a reference for the research and development of other *Verbena* plants.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA842648.

Author contributions

CW, XY, and ZL conceived and designed the project. SP, KY, and FL performed the experiments, processed the experimental data and analyzed it. SP wrote the manuscript. HY helped JG draw the figures in the article. CW, GL, and HL checked and corrected the manuscript. All authors read and approved the final manuscript.

Funding

This work was financially supported by the National Key R&D Program of China, Synthetic Biology Research (No. 2019YFA0905300), the National Natural Science Foundation of China (Nos. 82074276, 81872832, and 82173781), Key Field Projects of Guangdong Universities (Intelligent Manufacturing; 2020ZDZX2057), and Innovation Team and Talents Cultivation Program of National Administration of Traditional Chinese Medicine (No. ZYYCXTD-D-202002).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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.2022.955075/full#supplementary-material

References

Ahmad, B., Banerjee, A., Tiwari, H., Jana, S., Bose, S., and Chakrabarti, S. (2018). Structural and functional characterization of the Vindoline biosynthesis pathway enzymes of *Catharanthus roseus*. *J. Mol. Model.* 24:53. doi: 10.1007/s00894-018-3590-2

Alipieva, K., Korkina, L., Orhan, I. E., and Georgiev, M. I. (2014). Verbascoside-a review of its occurrence, (bio)synthesis and pharmacological significance. *Biotechnol. Adv.* 32, 1065–1076. doi: 10.1016/j.biotechadv.2014.07.001

Ambure, P., Kar, S., and Roy, K. (2014). Pharmacophore mapping-based virtual screening followed by molecular docking studies in search of potential acetylcholinesterase inhibitors as anti-Alzheimer's agents. *Biosystems* 116, 10–20. doi: 10.1016/j.biosystems.2013.12.002

Bai, Y., Bi, H., Zhuang, Y., Liu, C., Cai, T., Liu, X., et al. (2014). Production of salidroside in metabolically engineered *Escherichia coli*. *Sci Rep*. 4:6640. doi: 10.1038/srep06640

Bergman, M. E., Davis, B., and Phillips, M. A. (2019). Medically useful plant terpenoids: biosynthesis, occurrence, and mechanism of action. *Molecules* 24:3961. doi: 10.3390/molecules24213961

Calvo, M. I., Vilalta, N., San Julián, A., and Fernández, M. (1998). Anti-inflammatory activity of leaf extract of *Verbena officinalis* L. *Phytomedicine* 5, 465–467. doi: 10.1016/S0944-7113(98)80043-3

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Chen, D. B., Gao, H. W., Peng, C., Pei, S. Q., Dai, A. R., Yu, X. T., et al. (2020). Quinones as preventive agents in Alzheimer's diseases: focus on NLRP3 inflammasomes. *J. Pharm. Pharmacol.* 72, 1481–1490. doi: 10.1111/jphp.13332

Collu, G., Unver, N., Peltenburg-Looman, A. M., Van der Heijden, R., Verpoorte, R., and Memelink, J. (2001). Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS Lett.* 508, 215–220. doi: 10.1016/s0014-5793(01)03045-9

Council of Europe (2020). *European Pharmacopoeia*, 10th Edn. Strasburg: European Directorate for the Quality Control of Medicines.

Das, B., Singh, N., Yao, A. Y., Zhou, J., He, W., Hu, X., et al. (2021). BACE1 controls synaptic function through modulating release of synaptic vesicles. *Mol. Psychiat.* 26, 6394–6410. doi: 10.1038/s41380-021-01166-2

Davidson, N. M., and Oshlack, A. (2014). Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biol.* 15:410. doi: 10.1186/s13059-014-0410-6

De Lima, R., Guex, C. G., Da Silva, A., Lhamas, C. L., Dos Santos Moreira, K. L., Casoti, R., et al. (2018). Acute and subacute toxicity and chemical constituents of the hydroethanolic extract of *Verbena litoralis* Kunth. *J. Ethnopharmacol.* 224, 76–84. doi: 10.1016/j.jep.2018.05.012

Deepak, M., and Handa, S. S. (2000). Antiinflammatory activity and chemical composition of extracts of *Verbena officinalis*. *Phytother. Res.* 14, 463–465. doi: 10.1002/1099-1573(200009)14:6<463::aid-ptr611<3.0.co;2-g

DeTure, M. A., and Dickson, D. W. (2019). The neuropathological diagnosis of Alzheimer's disease. *Mol. Neurodegener.* 14, 32. doi: 10.1186/s13024-019-0333-5

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

El-Hela, A. A., Al-Amier, H. A., and Ibrahim, T. A. (2010). Comparative study of the flavonoids of some *Verbena* species cultivated in Egypt by using high-performance liquid chromatography coupled with ultraviolet spectroscopy and atmospheric pressure chemical ionization mass spectrometry. *J. Chromatogr. A.* 1217, 6388–6393. doi: 10.1016/j.chroma.2010.08.025

el-Hela, A. A., Sowiński, P., and Krauze-Baranowska, M. (2000). Iridoids and phenylethanoids of *Verbena bipinnatifida* Nutt. *Acta Pol. Pharm.* 57, 65–68.

Eskandarzadeh, M., Kordestani-Moghadam, P., Pourmand, S., Khalili Fard, J., Almassian, B., and Gharaghani, S. (2021). Inhibition of GSK-3β by iridoid glycosides of snowberry (*Symphoricarpos albus*) effective in the treatment of alzheimer's disease using computational drug design methods. *Front. Chem.* 9:709932. doi: 10.3389/fchem.2021.709932

Ghezzi, L., Scarpini, E., and Galimberti, D. (2013). Disease-modifying drugs in Alzheimer's disease. *Drug Des. Dev. Ther.* 7, 1471–1478. doi: 10.2147/DDDT.S41431

Gong, X., Yang, M., He, C. N., Bi, Y. Q., Zhang, C. H., Li, M. H., et al. (2022). Plant Pharmacophylogeny: Review and Future Directions. *Chin. J. Integr. Med.* 28, 567–574. doi: 10.1007/s11655-020-3270-9

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., et al. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36, 3420–3435. doi: 10.1093/nar/gkn176

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883

Grawish, M. E., Anees, M. M., Elsabaa, H. M., Abdel-Raziq, M. S., and Zedan, W. (2016). Short-term effects of *Verbena officinalis* Linn decoction on patients suffering from chronic generalized gingivitis: double-blind randomized controlled multicenter clinical trial. *Quint. Int.* 47, 491–498. doi: 10.3290/j.qi.a35521

Holmes, C. (2013). Review: systemic inflammation and Alzheimer's disease. *Neuropathol. Appl. Neurobiol.* 39, 51–68. doi: 10.1111/j.1365-2990.2012.01307.x

Ji, A., Jia, J., Xu, Z., Li, Y., Bi, W., Ren, F., et al. (2017). Transcriptome-Guided mining of genes involved in crocin biosynthesis. *Front. Plant Sci.* 8:518. doi: 10.3389/fpls.2017.00518

Krithika, R., Srivastava, P. L., Rani, B., Kolet, S. P., Chopade, M., Soniya, M., et al. (2015). Characterization of 10-hydroxygeraniol dehydrogenase from *Catharanthus roseus* reveals cascaded enzymatic activity in iridoid biosynthesis. *Sci. Rep.* 5:8258. doi: 10.1038/srep08258

Kubica, P., Szopa, A., Dominiak, J., Luczkiewicz, M., and Ekiert, H. (2020). *Verbena officinalis* (Common Vervain) - A review on the investigations of this medicinally important plant species. *Planta. Med.* 86, 1241–1257. doi: 10.1055/a-1232-5758

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Lai, S. W., Yu, M. S., Yuen, W. H., and Chang, R. C. (2006). Novel neuroprotective effects of the aqueous extracts from *Verbena officinalis* Linn. *Neuropharmacology* 50, 641–650. doi: 10.1016/j.neuropharm.2005.11.009

Lane, C. A., Hardy, J., and Schott, J. M. (2018). Alzheimer's disease. *Eur. J. Neurol.* 25, 59–70. doi: 10.1111/ene.13439

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with 2. *Nat. Methods.* 9, 357–359. doi: 10.1038/nmeth.1923

Li, B., and Dewey, C. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* 12:323. doi: 10.1186/1471-2105-12-323

Li, H., Yang, S. Q., Wang, H., Tian, J., and Gao, W. Y. (2010). Biosynthesis of the iridoid glucoside, lamalbid, in *Lamium barbatum*. *Phytochemistry* 71, 1690–1694. doi: 10.1016/j.phytochem.2010.06.019

Liu, W., Feng, Y., Yu, S., Fan, Z., Li, X., Li, J., et al. (2021). The flavonoid biosynthesis network in plants. *Int. J. Mol. Sci.* 22:12824. doi: 10.3390/ijms222312824

Liu, Z., Xu, Z., Zhou, H., Cao, G., Cong, X. D., Zhang, Y., et al. (2012). Simultaneous determination of four bioactive compounds in *Verbena officinalis* L. by using high-performance liquid chromatography. *Pharmacogn. Mag.* 8, 162–165. doi: 10.4103/0973-1296.96575

Loera-Valencia, R., Cedazo-Minguez, A., Kenigsberg, P. A., Page, G., Duarte, A. I., Giusti, P., et al. (2019). Current and emerging avenues for Alzheimer's disease drug targets. *J. Intern. Med.* 286, 398–437. doi: 10.1111/joim.12959

Lou, H., Hu, L., Lu, H., Wei, T., and Chen, Q. (2021). Metabolic engineering of microbial cell factories for biosynthesis of flavonoids: A Review. *Molecules* 26:4522. doi: 10.3390/molecules26154522

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8

Luo, D., Chen, N. H., Wang, W. Z., Zhang, J. H., Li, C. J., Zhuo, X. F., et al. (2021). Structurally diverse matrine-based alkaloids with anti-inflammatory effects from *Sophora alopecuroides*. *Chin. J. Chem.* 39, 3339–3346. doi: 10.1002/cjoc.202100526

Ma, D., Li, G., Zhu, Y., and Xie, D. Y. (2017). Overexpression and suppression of *Artemisia annua* 4-hydroxy-3-methylbut-2-enyl diphosphate reductase 1 gene (AaHDR1) differentially regulate artemisinin and terpenoid biosynthesis. *Front. Plant Sci.* 8:77. doi: 10.3389/fpls.2017.00077

Moussa-Pacha, N. M., Abdin, S. M., Omar, H. A., Alniss, H., and Al-Tel, T. H. (2020). BACE1 inhibitors: current status and future directions in treating Alzheimer's disease. *Med. Res. Rev.* 40, 339–384. doi: 10.1002/med.21622

Pan, Y., Zhao, X., Wang, Y., Tan, J., and Chen, D. X. (2021). Metabolomics integrated with transcriptomics reveals the distribution of iridoid and crocin metabolic flux in *Gardenia jasminoides* Ellis. *PLoS One* 16:e0256802. doi: 10.1371/journal.pone.0256802

Ruzicka, J., Lukas, B., Merza, L., Göhler, I., Abel, G., Popp, M., et al. (2009). Identification of *Verbena officinalis* based on ITS sequence analysis and RAPD-derived molecular markers. *Planta. Med.* 75, 1271–1276. doi: 10.1055/s-0029-1185535

Saimaru, H., and Orihara, Y. (2010). Biosynthesis of acteoside in cultured cells of *Olea europaea*. *J. Nat. Med.* 64, 139–145. doi: 10.1007/s11418-009-0383-z

Saito, K., Yonekura-Sakakibara, K., Nakabayashi, R., Higashi, Y., Yamazaki, M., Tohge, T., et al. (2013). The flavonoid biosynthetic pathway in Arabidopsis: structural and genetic diversity. *Plant Physiol. Biochem.* 72, 21–34. doi: 10.1016/j.plaphy.2013.02.001

Shao, B. Z., Xu, Z. Q., Han, B. Z., Su, D. F., and Liu, C. (2015). NLRP3 inflammasome and its inhibitors: a review. *Front. Pharmacol.* 6:262. doi: 10.3389/fphar.2015.00262

Shu, J., Chou, G., and Wang, Z. (2014). Two new iridoids from *Verbena officinalis* L. *Molecules* 19, 10473–10479. doi: 10.3390/molecules190710473

Simkin, A. J., Miettinen, K., Claudel, P., Burlat, V., Guirimand, G., Courdavault, V., et al. (2013). Characterization of the plastidial geraniol synthase from Madagascar periwinkle which initiates the monoterpenoid branch of the alkaloid

pathway in internal phloem associated parenchyma. *Phytochemistry* 85, 36–43. doi: 10.1016/j.phytochem.2012.09.014

Soares, K., Pianoski, K. E., Finger, D., Machado, C. S., QuinÁia, S. P., and Torres, Y. R. (2020). Levels of phenylpropanoids and iridoids in extracts and infusions of *Verbena minutiflora*. *An. Acad. Bras. Cienc*. 92:e20181116. doi: 10.1590/0001-3765202020181116

Speroni, E., Cervellati, R., Costa, S., Guerra, M. C., Utan, A., Govoni, P., et al. (2007). Effects of differential extraction of *Verbena officinalis* on rat models of inflammation, cicatrization and gastric damage. *Planta. Med*. 73, 227–235. doi: 10.1055/s-2007-967116

Sun, P., Song, S., Zhou, L., Zhang, B., Qi, J., and Li, X. (2012). Transcriptome analysis reveals putative genes involved in iridoid biosynthesis in *Rehmannia glutinosa*. *Int. J. Mol. Sci*. 13, 13748–13763. doi: 10.3390/ijms131013748

Tan, W. B., and Wang, Z. F. (2011). Effect of *Verbena officinalis* decoction on learning and memory in Alzheimer's Disease Mice. *Chin. J. Ethnomed. Ethnopharm*. 20, 36–37. doi: 10.3969/j.issn.1007-8517.2011.20.028

Tholl, D. (2015). Biosynthesis and biological functions of terpenoids in plants. *Adv. Biochem. Eng. Biotechnol*. 148, 63–106. doi: 10.1007/10_2014_295

Volke, D. C., Rohwer, J., Fischer, R., and Jennewein, S. (2019). Investigation of the methylerythritol 4-phosphate pathway for microbial terpenoid production through metabolic control analysis. *Microb. Cell Fact*. 18:192. doi: 10.1186/s12934-019-1235-5

Yang, Y., Wu, Y., Zhuang, Y., and Liu, T. (2021). Discovery of glycosyltransferases involved in the biosynthesis of Ligupurpuroside B. *Org. Lett*. 23, 7851–7854. doi: 10.1021/acs.orglett.1c02873

Ye, P., Liang, S. C., Wang, X. M., Duan, L. X., Jiang, F. Y. Y., Yang, J. F., et al. (2019). Transcriptome analysis and targeted metabolic profiling for pathway elucidation and identification of a geraniol synthase involved in iridoid biosynthesis from *Gardenia jasminoides*. *Ind. Crop. Prod*. 132, 48–58. doi: 10.1016/j.indcrop.2019.02.002

Yuan, Y. W., Liu, C., Marx, H. E., and Olmstead, R. G. (2010). An empirical demonstration of using pentatricopeptide repeat (PPR) genes as plant phylogenetic tools: phylogeny of Verbenaceae and the *Verbena* complex. *Mol. Phylogenet. Evol*. 54, 23–35. doi: 10.1016/j.ympev.2009.08.029

Zhang, X., Li, C., Wang, L., Fei, Y., and Qin, W. (2019). Analysis of *Centranthera Grandiflora* benth transcriptome explores genes of catalpol, acteoside and azafrin biosynthesis. *Int. J. Mol. Sci*. 20:6034. doi: 10.3390/ijms20236034

Zhao, C., Wang, F., Lian, Y., Xiao, H., and Zheng, J. (2020). Biosynthesis of citrus flavonoids and their health effects. *Crit. Rev. Food Sci. Nutr*. 60, 566–583. doi: 10.1080/10408398.2018.1544885

Zheng, Y., Jiao, C., Sun, H., Rosli, H. G., Pombo, M. A., Zhang, P., et al. (2016). iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol. Plant* 9, 1667–1670. doi: 10.1016/j.molp.2016.09.014

Zhou, Y., Zhu, J., Shao, L., and Guo, M. (2020). Current advances in acteoside biosynthesis pathway elucidation and biosynthesis. *Fitoterapia* 142:104495. doi: 10.1016/j.fitote.2020.104495

Zhu, P., Chen, Y., Wu, F., Meng, M., and Ji, K. (2022). Expression and promoter analysis of MEP pathway enzyme-encoding genes in *Pinus massoniana* Lamb. *PeerJ* 10:e13266. doi: 10.7717/peerj.13266

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 11 August 2022
DOI 10.3389/fpls.2022.973197

OPEN ACCESS

EDITED BY
Wanchai De-Eknamkul,
Chulalongkorn University,
Thailand

REVIEWED BY
Xiwen Li,
China Academy of Chinese Medical Sciences,
China
Wenpan Dong,
Beijing Forestry University,
China

*CORRESPONDENCE
Da-Cheng Hao
hao@djtu.edu.cn
Lijia Xu
ljxu@implad.ac.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

RECEIVED 19 June 2022
ACCEPTED 18 July 2022
PUBLISHED 11 August 2022

CITATION
Hao D-C, Song Y, Xiao P, Zhong Y, Wu P and Xu L (2022) The genus *Chrysanthemum*: Phylogeny, biodiversity, phytometabolites, and chemodiversity. *Front. Plant Sci.* 13:973197. doi: 10.3389/fpls.2022.973197

The genus *Chrysanthemum*: Phylogeny, biodiversity, phytometabolites, and chemodiversity

Da-Cheng Hao[1,2*†], Yanjun Song[3†], Peigen Xiao[3,4], Yi Zhong[3], Peiling Wu[3] and Lijia Xu[3,4*]

[1]School of Environment and Chemical Engineering, Biotechnology Institute, Dalian Jiaotong University, Dalian, China, [2]Institute of Molecular Plant Science, University of Edinburgh, Edinburgh, United Kingdom, [3]Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [4]Key Laboratory of Bioactive Substances and Resources Utilization of Chinese Herbal Medicine, Ministry of Education, Beijing, China

The ecologically and economically important genus *Chrysanthemum* contains around 40 species and many hybrids and cultivars. The dried capitulum of *Chrysanthemum morifolium* (CM) Ramat. Tzvel, i.e., *Flos Chrysanthemi*, is frequently used in traditional Chinese medicine (TCM) and folk medicine for at least 2,200years. It has also been a popular tea beverage for about 2,000years since Han Dynasty in China. However, the origin of different cultivars of CM and the phylogenetic relationship between *Chrysanthemum* and related Asteraceae genera are still elusive, and there is a lack of comprehensive review about the association between biodiversity and chemodiversity of *Chrysanthemum*. This article aims to provide a synthetic summary of the phylogeny, biodiversity, phytometabolites and chemodiversity of *Chrysanthemum* and related taxonomic groups, focusing on CM and its wild relatives. Based on extensive literature review and in light of the medicinal value of chrysanthemum, we give some suggestions for its relationship with some genera/species and future applications. Mining chemodiversity from biodiversity of *Chrysanthemum* containing subtribe Artemisiinae, as well as mining therapeutic efficacy and other utilities from chemodiversity/biodiversity, is closely related with sustainable conservation and utilization of Artemisiinae resources. There were eight main cultivars of *Flos Chrysanthemi*, i.e., Hangju, Boju, Gongju, Chuju, Huaiju, Jiju, Chuanju and Qiju, which differ in geographical origins and processing methods. Different CM cultivars originated from various hybridizations between multiple wild species. They mainly contained volatile oils, triterpenes, flavonoids, phenolic acids, polysaccharides, amino acids and other phytometabolites, which have the activities of antimicrobial, anti-viral, antioxidant, anti-aging, anticancer, anti-inflammatory, and closely related taxonomic groups could also be useful as food, medicine and tea. Despite some progresses, the genetic/chemical relationships among varieties, species and relevant genera have yet to be clarified; therefore, the roles of pharmacophylogeny and omics technology are highlighted.

KEYWORDS

***Chrysanthemum*, *Chrysanthemum morifolium*, phylogenetic relationship, phytochemistry, chemodiversity, pharmacophylogeny**

Introduction

Anthemideae is a taxonomically controversial Asteraceae (Compositae) tribe, in which phylogenetic relationships are still not settled (Criado Ruiz et al., 2021). In Anthemideae, Artemisiinae and Matricariinae are the two most widely distributed subtribes (Watson et al., 2002; Shen et al., 2021). Artemisiinae (Figure 1) includes 1/3 species of tribe Anthemideae, consisting of more than 600 species of 18–19 genera, which are mainly in the northern hemisphere, especially in Central Asia and East Asia. *Chrysanthemum*, a small Artemisiinae genus, is native to East Asia and northeastern Europe. It contains around 40 species and many hybrids and cultivars, most of which originate from East Asia, and China is the center of diversity (Liu et al., 2012). Innumerable horticultural varieties and cultivars of *Chrysanthemum* exist. The *Chrysanthemum* plants are perennial herbs. Their leaves are undivided or palmately/pinnately divided once or twice. The capitulum is heteromorphic, solitary at the top of the stem, a few or more of which are arranged into corymbose or compound corymbose inflorescences at the top of stem and branch. The marginal flowers are female, tongue shaped, 1-layer (multi-layer in cultivated varieties), and the central disk flowers are bisexual and tubular. The involucre is shallowly discoid, rarely campanulate. The involucral bracts are 4–5 layers, with white, brown, black brown or brownish black edges. The receptacle is protuberant, hemispherical or conical, without stipules. The tongue shaped flowers are yellow, white or red, and the tongue can be as short as 1.5 mm, and as long as 2.5 cm or longer. All tubular flowers are yellow, with five teeth at the top. The style branches are linear, apically truncate. The anther base is obtuse, and apical appendages are lanceolate ovate or oblong. All achenes are isomorphic, nearly cylindrical and narrowed to the lower part, with 5–8 longitudinal veins and no coronal crown hair.

As the most remarkable taxon of *Chrysanthemum*, *Chrysanthemum morifolium* Ramat. Tzvel (CM, known as "Ju Hua" in Chinese) originated in China as a medicinal, food and ornamental plant. Its history began about 3,000 years ago, since Xia and Shang Dynasties (Zhang and Dai, 2009). CM was introduced to Japan as a famous spice during Tang Dynasty (AD 710 to 784), and later entered Europe and then America in the 17th century (Li, 1993). Due to its ornamental, medical and nutritional value, CM appreciation had created an unique chrysanthemum culture in East Asia. The varieties of CM have increased from 300 in Qing Dynasty to over 1,000 nowadays (Zhang, 2008; Mao, 2020). The CM's medical and nutritional values were recognized around 2,300 years ago (Chen, 2005; Zhang, 2008); in the Spring and Autumn Period CM firstly appeared in "*Li Sao*," written by the great poet Yuan Qu, as an edible vegetable. Later it was recorded as a medicine in "*Shen Nong's Classic of the Materia Medica*" in Han Dynasty. Presently, CM is one of the most commonly consumed food, herbal medicine and tea beverage in China and adjacent countries (Lai et al., 2007; Jiang et al., 2013). CM is extensively used by healthcare providers to treat conditions such as dizziness, photophobia with lacrimation, conjunctivitis, headache with fever, red eyes, swollen poison and boils, among others (Chinese Pharmacopoeia Commission, 2020).

Modern studies found that *Chrysanthemum* and related genera contain significant amounts of volatile oils, flavonoids and hydroxycinnamoyl-quinic acids (Beninger et al., 2004; Jiang et al., 2004; Chen et al., 2007; Clifford et al., 2007; Lai et al., 2007; Wang et al., 2008a,b; Duan et al., 2022), and have extensive biological activities (Figure 1), e.g., anti-inflammation, antioxidation, vasodilation, protecting cardiovascular system, anticancer, inhibiting aldose reductase and anti-mutagenic (Zhang et al., 2000; Rajic et al., 2001; Peng et al., 2006; Shao et al., 2020). Given the long use history and important role of *Chrysanthemum* taxa, and in order to better preserve and utilize *Chrysanthemum* resources, a comprehensive review is necessary to synthesize information from existing studies of their phylogeny, biodiversity, chemodiversity and chemotaxonomy. In this study, the up-to-date information of *Chrysanthemum* were collected from major databases including NCBI PubMed, Google Scholar, Web of Science, SciFinder, Wiley online, Elsevier ScienceDirect and China National Knowledge Infrastructure (CNKI), using the keywords such as "Chrysanthemum," "juhua," "phylogenetic," "biosynthesis," the respective phytometabolite name, the respective genus/species name, and the like. Additionally, materia medica books and patents have made their contribution to the summary of botany, traditional uses and cultural significance of *C. morifolium*. The phylogenetic kinship, origin of genuine varieties, and the chemical ingredients and their bioactive effects of *Chrysanthemum* taxa were elaborated from the perspective of pharmacophylogeny, so as to facilitate a holistic understanding of *Chrysanthemum* and related genera and promote the sustainable conservation and utilization of relevant natural resources.

Abbreviations: Acn, Acacetin; AFLP, Amplified fragment length polymorphism; AG, Acetylglucoside; ANS, Anthocyanidin synthase; Api, Apigenin; CDDP, Conserved DNA-derived polymorphism; CHI, Chalcone isomerase; CHS, Chalcone synthase; CIF, Chrysanthemi Indici Flos; CM, Chrysanthemum× morifolium; cp, chloroplast; CYP, Cytochrome p450; DCQA, Dicaffeoylquinic acid; DFR, Dihydroflavanol 4-reductase; Dio, Diosmetin; EST, Expressed sequence tag; F3′H, Flavonoid 3′-hydroxylase; F3′5′H, Flavonoid 3′5′-hydroxylase; FISH, Fluorescence *in situ* hybridization; GT, Glycosyltransferase; IGS, Intergenic spacer; ISSR, Inter-simple sequence repeat; Lut, Luteolin; MG, Malonylglucoside; MW, Molecular weight; nr ITS, Nuclear ribosomal internal transcribed spacer; PCA, Principal component analysis; QTP, Qinghai-Tibetan Plateau; RAPD, Random amplified polymorphic DNA; RFLP, Restriction fragment length polymorphism; SCAR, Sequence-characterized amplified region; SNP, Single nucleotide polymorphism; SSR, Simple sequence repeat; TCM, Traditional Chinese medicine; TPS, Terpene synthase; TS, Terpenoid synthase; UGT, UDP-glucuronosyl and -glucosyltransferase; UPGMA, Unweighted pair group method with arithmetic means; WGD, Whole genome duplication.

Biodiversity and phylogenetic relationship

Intergeneric relationship within Asteraceae tribe Anthemideae

The family Asteraceae, also called Compositae, consists of over 32,000 known species of flowering plants in over 1,900 genera

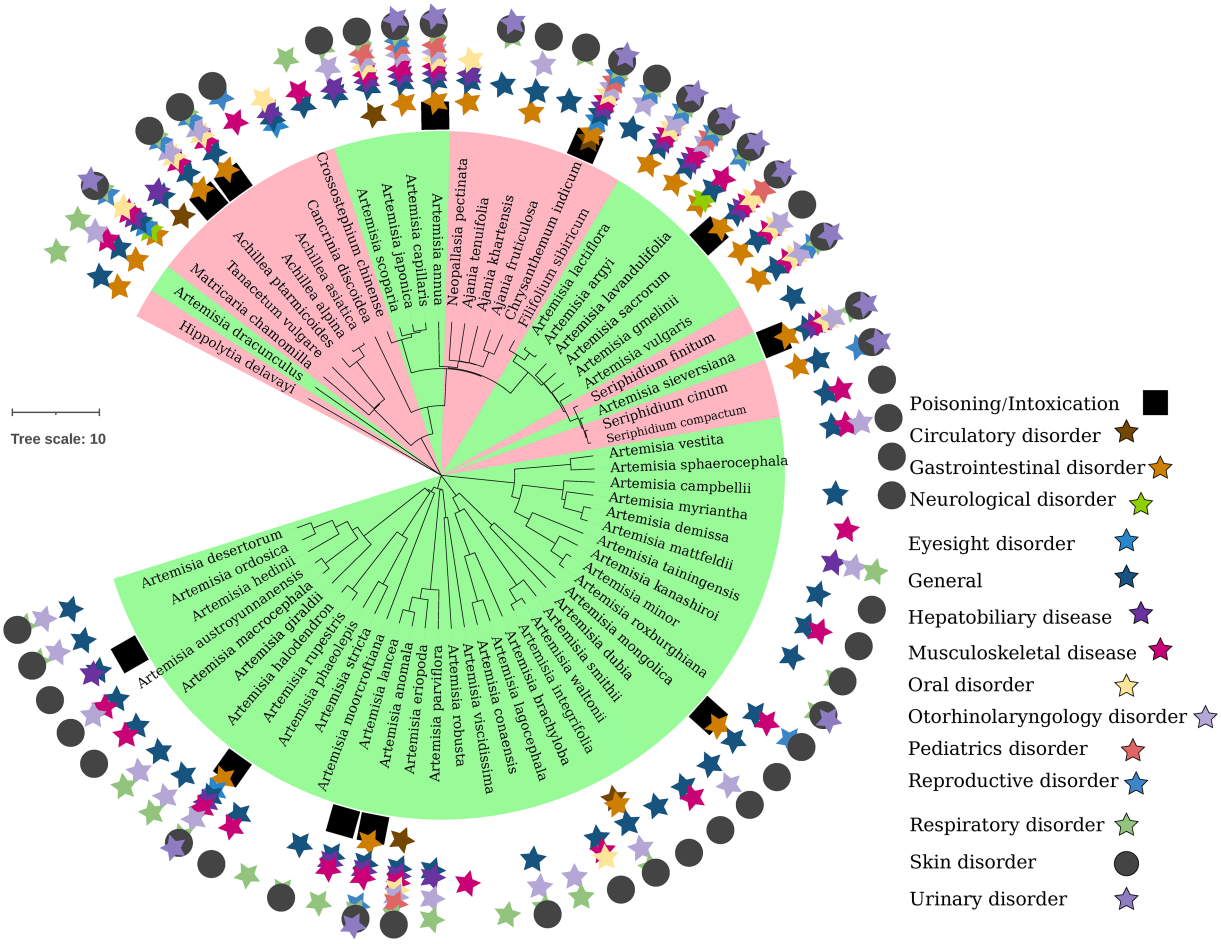


FIGURE 1
Phylogenetic relationship of Artemisiinae species, including *Chrysanthemum* and its wild relatives. The distribution of therapeutic efficacy of ethnomedicinal species on the phylogenetic tree of Chinese taxa is shown. In the inner circle, *Artemisia* species are in green area, and non-*Artemisia* species are in red area. From the inside to the outside of the outer circle are poisoning, circulatory diseases, gastrointestinal diseases, nervous system diseases, eye diseases, other/general diseases, hepatobiliary diseases, musculoskeletal diseases, oral diseases, ear/nose/throat diseases, pediatric diseases, reproductive system diseases, respiratory diseases, skin diseases, and urinary diseases, indicated by stars of different colors, square or circle.

within the order Asterales (Mandel et al., 2019). There are more than 200 genera and more than 2,000 Asteraceae species in China, which are distributed all over the country. Anthemideae is a tribe of the subfamily Asteroideae, plants of which are distributed worldwide with concentrations in central Asia, Mediterranean Basin, and southern Africa. Most species of plant known as chamomile belong to genera of this tribe. There are about 1,800 species classified in 111 Anthemideae genera (Oberprieler et al., 2007). This tribe is divided into 14 subtribes, and the genus *Chrysanthemum* belongs to the subtribe Artemisiinae (Figure 1).

Artemisiinae has 19 genera, 15 of which are distributed in East Asia, and there are eight endemic genera. This subtribe is the most important Anthemideae subtribe in East Asia, including more than 80% of the Anthemideae genera. Based on the extensive collection of East Asian Anthemideae resources, the systematic evolution was studied by means of pollen morphology, molecular systematics, molecular markers and so on (Zhao, 2007). The intergeneric crossing between cultivated *C.* × *grandiflorum* and *Ajania pacifica* was conducted, as well as self-crosses and backcrosses of hybrid F1. The genetic performance and cytological behavior of traits in different hybrid combinations and generations were probed to elucidate the relationship between parental genomes and their roles in different generations of hybrids. The East Asian Anthemideae taxa are mainly diploid and tetraploid, and many monotypic genera native to East and Central Asia, e.g., *Opisthopappus* Shih and *Crossostephium* Less., etc., are diploid. It is speculated that the ancestor of the tribe should be diploidy. Within the genus *Ajania* Polj., the chromosome ploidy differs between species and intraspecific populations.

East Asian Anthemideae can be divided into three groups according to the outer wall decoration of pollen (Zhao, 2007). The first group has obvious thorn-like decoration on the surface, the thorns are larger, longer than 1.5 μm, the base is expanded, and there are hole-like perforations from middle of the thorn to the

thorn base. This group includes genera *Glebionis* and *Argyranthemum* of the subtribe Chrysanthemidae, *Hippolytia*, *Opisthopappus*, *Pyrethrum* and *Tanacetum* of subtribe Tanacetinae, *Leucanthemum* and *Nipponanthemum* of subtribe Leucanthemidae, *Matricaria* of subtribe Matricariinae, *Achillea* of Achilleinae, *Chrysanthemum*, *Ajania* and *Brachanthemum* of subtribe Artemisiinae. The second group has degenerate small spines or inconspicuous spines, basically less than 0.5 μm. This group includes genera *Artemisia*, *Seriphidium*, *Crossostephium*, *Elachanthemum*, *Neopallasia*, *Ajaniopsis*, *Filifolium* and *Kaschgaria*. The third group has shorter surface thorn-like ornamentation, which is, however, very obvious, and its length is between the first and second groups, 1.0–1.5 μm; it is conical, the base does not descend, and the apex is pointed. This group includes genus *Stilpnotepis*, *A. salicifolia* and *A. variifolia* of *Ajania*. Based on the outer wall decoration of pollen, *Phaeostigma* is proposed to be a separate genus from *Ajania*.

To better understand the intergeneric relationships and taxonomic position of small Asian genera of Anthemideae, the sequences of nuclear ribosomal internal transcribed spacer (nr ITS) and chloroplast (cp) trnL-F intergenic spacer (IGS) were used to infer the phylogenetic relationship of 48 Anthemideae taxa (Zhao et al., 2010). The trnL-F sequence was of poor resolving power, but some deletions and insertions were useful in interspecific and intergeneric circumscriptions. Both ITS and ITS/IGS phylogenies suggest two major clades. The clade A is subtribe Artemisiinae, consisting of two groups; one group includes *Chrysanthemum*, *Arctanthemum*, *Ajania*, *Opisthopappus* and *Elachanthemum*, while the other includes *Artemisia*, *Crossostephium*, *Neopallasia* and *Sphaeromeria*. Within the former group, *Chrysanthemum*, *Arctanthemum* and *Ajania* were closely related to each other, and the delimitation of genera *Chrysanthemum* and *Ajania* was ambiguous. The successful hybridization between *C.* × *grandiflorum* and *A. pacifica* and the normal meiotic chromosome behavior of the hybrid also indicate the close relatedness between two genera (Zhao, 2007). *Phaeostigma* was not in *Chrysanthemum* group, which was confirmed by the 6-bp insertion in trnL-F. *Brachanthemum* was excluded, and *Elachanthemum* was included in this group. *Opisthopappus* of subtribe Tanacetinae was proposed to be in the subtribe Artemisiinae, *Chrysanthemum* group (Hong et al., 2015; Figure 2). The molecular phylogeny facilitates the inference of evolution of pollen and capitulum characters. Disentangling old hybridization events may be compromised by the set of evolutionary processes accumulated subsequently (Criado Ruiz et al., 2021), particularly in areas with past climatic instability. The high-throughput sequencing data will facilitate understanding the role and impact of reticulate evolution in the phylogenetic puzzle.

In an amplified fragment length polymorphism (AFLP) analysis of Artemisiinae, 1705 polymorphic bands were obtained (Zhao, 2007). According to the similarity coefficient, both within the whole subtribe, its related groups, and within *Chrysanthemum* and *Ajania*, the high genetic diversity was revealed, indicating obvious differentiation between genera and species. In the cluster diagram, *Elachanthemum* and *Neopallasia* are at the base of the whole subtribe, while *Brachanthemum* is at the base of *Chrysanthemum-Ajania* branch. The population of *Ajania* is always at the base of *Chrysanthemum* population, and the latter may have evolved from different species or populations of *Ajania* in parallel. The close relatedness between *Opisthopappus* and *Chrysanthemum-Ajania* was also confirmed by AFLP results (Zhao et al., 2010). Based on the results of cytology, pollen morphology, molecular sequence analysis and AFLP, it is believed that the East Asian Anthemideae could originate in the Eurasian continent, and the ancestral species of Eurasian continent spread eastward and gradually evolved throughout Central Asia and East Asia. Some genera distributed in Central Asia, e.g., *Kaschgaria*, may be the relatively primitive groups of Artemisiinae.

The geography and ecology play a key role in driving species diversification of Anthemideae and in shaping genotype/phenotype of species. The multi-phased orogenesis of Qinghai-Tibetan Plateau (QTP) and global climate changes over late-Miocene has profoundly influenced the environments and evolution of Anthemideae plants in QTP and adjacent regions (Shen et al., 2021). The DNA sequences of seven low-copy nuclear genes and nrITS were combined to reconstruct a time-calibrated phylogeny of subtribe Artemisiinae (Figure 2). In the monophyletic *Chrysanthemum* group, *Chrysanthemum* and *Ajania* were well resolved, agreeing with capitulum morphology and suggesting the low resolving power of cp and ITS markers (Zhao et al., 2010; Hong et al., 2015; Tyagi et al., 2020; Masuda et al., 2022). Within *Chrysanthemum*, the *C. indicum* complex and *C. zawadskii* complex diverged, which temporally coincide with the late Cenozoic uplift of Northern QTP and associated climatic heterogeneity between eastern and central Asia. The origin of *Chrysanthemum* group might be in Central Asia, then *Chrysanthemum* migrated eastward, in contrast to the *in situ* diversification of *Ajania* (Figure 3). The *C. indicum* complex and *C. zawadskii* complex have distinct distributions in East Asia, i.e., the former in more southern and the latter in more northern regions. The distribution patterns are related with the niche differentiation of different lineages and environmental heterogenization.

Through distant hybridization and aggregating germplasms, the wild Anthemideae genetic resources with excellent resistance, e.g., cold resistance, drought resistance, waterlogging resistance, etc., can be used in modern chrysanthemum breeding (Sun et al., 2010a; Hong, 2014) to realize the innovation of germplasm resources. While breeding new varieties, some basic data for exploring the genetic relationship between *Chrysanthemum* and related genera can be obtained (Wu, 2014), and the phylogenetic positions of some ambiguous taxa can also be clarified. With conventional hybridization methods or embryo rescue conditions, distant hybrids were obtained between *Chrysanthemum* and genera such as *Ajania*, *Crossostephium* (subtribe Artemisiinae), *Opisthopappus*, *Tanacetum*, *Pyrethrum* (subtribe Tanacetinae), *Argyranthemum*,

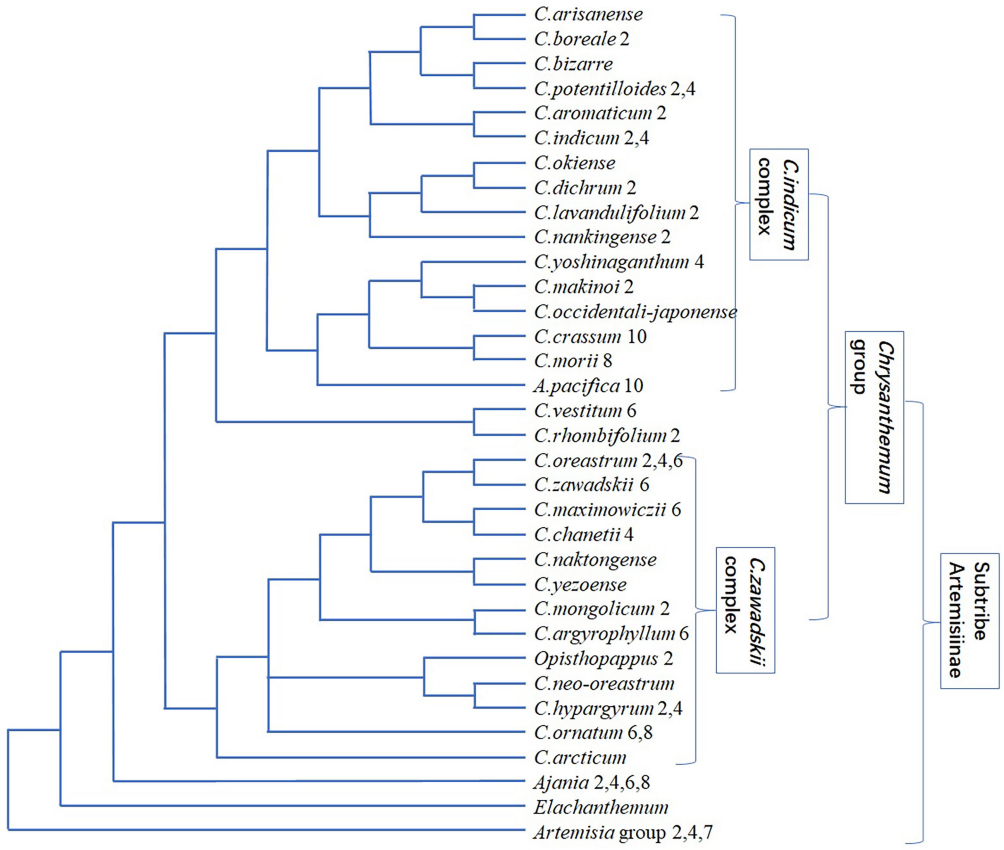


FIGURE 2
Phylogenetic relationship of *Chrysanthemum* taxa. The number after the taxon name represents the ploidy level. This schematic is compiled mainly based on Shen et al., 2021, as the phylogenetic topology inferred from low-copy nuclear genes and nrITS could be more convincing than those inferred from cp sequences (Liu et al., 2012; Lu et al., 2022a). It should be noted that it is challenging to distinguish taxa from each other within *C. indicum* complex and *C. zawadskii* complex, respectively.

Glebionis (subtribe Chrysanthemidae), *Leucanthemella* (subtribe Leucanthemidae; Zhao et al., 2008; Sun et al., 2010a; Shao, 2018). *Chrysanthemum*, *Ajania* and *Opisthopappus* are closely related (Figures 1, 2), and their hybridization is easier to be successful. *Brachanthemum* and *Hippolytia* are close to the clade of *Chrysanthemum* group, and their hybridization with *Chrysanthemum* (e.g., CM, *C. chanetii* and *C. naktongense*×*C. argyrophyllum*) can be successful (Hong, 2014; Xie, 2016). The hybrid without emasculation between *Hippolytia* and *Nipponanthemum* was obtained (Hong et al., 2015); on the ITS tree, both genera are closer to *Brachanthemum* than to *Chrysanthemum*. *Chrysanthemum* sensu lato includes around 38 genera, of which 27 are distributed in China (Shao, 2018). There are abundant wild resources and excellent stress resistance in *Chrysanthemum* sensu lato, which is of great significance to improve cultivated chrysanthemum breeding.

The chromosomal ploidy changes are abundant in *Chrysanthemum* (Figure 2), from diploid to decaploid, and there are also ploidy differences in different populations within the species. *C. crassum* (2n = 10x =90), *C. indicum* (4x = 36) and *C. nakingense* (2x = 18), as the female parent, respectively, were crossed with *Ajania myriantha* (2x = 18), *Crossostephium Chinense* (2x = 18), *Opisthopappus taihangensis* (2x = 18) and *Tanacetum vulgare* (2x = 18), respectively, to explore the genetic relationship between *Chrysanthemum* and its Anthemideae relatives (Tang, 2009). The F1 progeny of *C. indicum*×*Crossostephium Chinense* grew healthily and bloomed luxuriantly; the F1 progeny of *C. indicum*×*O. taihangensis* grew well, but there was no flower bud differentiation. The F1 progeny of *C. indicum*×*T. vulgare* did not grow normally in the field. It is speculated that the genetic relationship between *Chrysanthemum* and related genera from close to far is *Ajania* (Wu, 2007), *Crossostephium*, *Opisthopappus* and *Tanacetum*. In genomic two-color fluorescence *in situ* hybridization (FISH), the chromosomes of *Crossostephium* and *Tanacetum* were not observed in F1 progenies of CM cultivar "Tianzhuiyulu (falling jade dew)" × (*C. indicum*×*Crossosstephium Chinense*) and "Tianzhuiyulu" × (*C. nakingense*×*T. vulgare*), indicating that the chromosomes of two related genera may have been excluded. However, in F1 progeny of "Tianzhuiyulu" × (*C. crassum*×*Crossostephium Chinense*) the genomic chromosomes of three parents were observed, which realized the innovation of CM germplasm and enable full use of

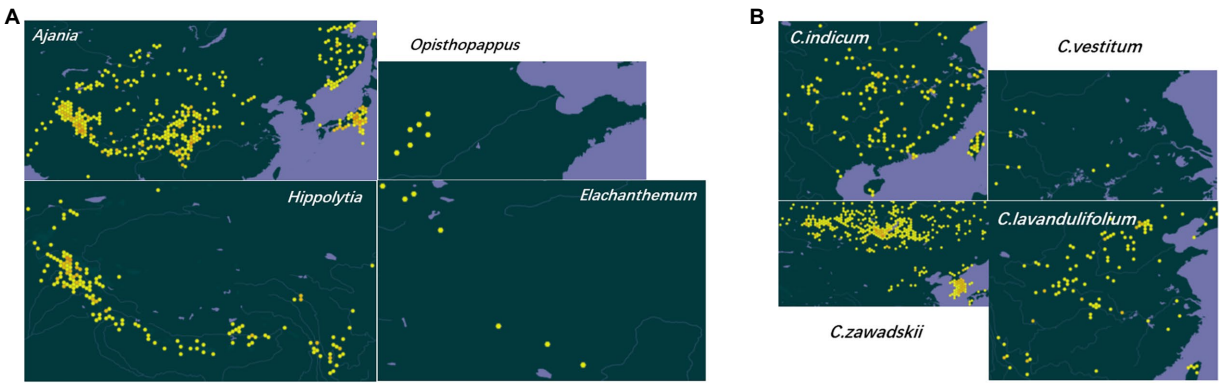


FIGURE 3
Geographical distribution of representative taxa of Asteraceae subtribe Artemisiinae. **(A)** Genus *Ajania* is produced in the vast areas of China except the southeast, Mongolia, North Korea, northern Afghanistan and Russia; *Opisthopappus* is endemic to Taihang Mountains, China; *Elachanthemum* is distributed in northern China, northwestern China and Mongolia; *Hippolytia* is produced in central Asia and the Himalayas. **(B)** *C. indicum* is produced in Northeast China, North China, Central China, South China, Southwest China, India, Japan, North Korea and Russia; *C. vestitum* is produced in western Henan, western Hubei and western Anhui, China; *C. zawadskii* is produced in Heilongjiang, Jilin, Liaoning, Hebei, Shanxi, Inner Mongolia, Shaanxi, Gansu and Anhui of China, Mongolia, Russia and Europe; *C. lavandulifolium* is produced in Jilin, Liaoning, Hebei, Shandong, Shanxi, Shaanxi, Gansu, Qinghai, Xinjiang (eastern), Jiangxi, Jiangsu, Zhejiang, Sichuan, Hubei and Yunnan of China, Korea and Japan. The geographic distribution data are retrieved from Global Biodiversity Information Facility (https://www.gbif.org/).

the excellent germplasm of *Chrysanthemum* related genera to improve CM varieties.

Intrageneric relationship of *Chrysanthemum* and origin of CM

The genus *Chrysanthemum*, associated with polyploidy and hybridization, undergoes rapid speciation and has about 40 species, most of which are distributed in East Asia (Liu et al., 2012). Many taxa are narrowly distributed and grow in specific habitats. The *Chrysanthemum* taxa vary greatly in morphology and ploidy (Figure 2), and the interspecific relationships are not fully understood. The genome size values of 15 species fall into three groups (Luo et al., 2017), which positively correlate with three ploidy levels (2×, 4×, 6×), and there was a genome downsizing after polyploidization in *Chrysanthemum*. Two major phylogenetic clades were inferred based on ITS and trnL-F sequences, i.e., *C. indicum* complex and *C. zawadskii* + *Ajania* group. The genome size and 1Cx values (DNA content of one non-replicated monoploid genome with chromosome number x) in the former group were significantly lower than those in the latter group, even though both have same ploidy level, and they were also correlated with latitude and morphology. The 1Cx values of diploid *Chrysanthemum* species were similar to those of *C. indicum* (Mt. Tianzhu) and *C. indicum* (Fujian) indicate that some populations of tetraploid *C. indicum* originated by autoploidy (Li et al., 2013). The genome size of *C. chanetii* is significantly larger than that of other tetraploids, while *C. vestitum* is smaller than other hexaploids, but the relationship between genome size and evolutionary time cannot be determined due to limited sampling. The divergence between *Chrysanthemum* and *Ajania* may be relatively recent, presumably temporally similar to the divergence of *C. zawadskii*. Hybridization and gene flow between taxa occurred frequently during the evolutionary history of *Chrysanthemum* and *Ajania*, causing considerable incongruency of cp and ITS trees (Li et al., 2014; Lu et al., 2022a). The gene flow had occurred between *C. indicum*, *C. nankingense* and *C. lavandulifolium* (Yang et al., 2006; Liu et al., 2012), and between *Cheiracanthium mongolicum* and *C. chanetii* (Luo et al., 2017), which could be involved in the origin of polyploids of *C. indicum* and *C. zawadskii* complex, respectively.

The ITS and cp rpl16 sequences were combined to resolve the interspecific relationship of *Chrysanthemum* (Wu, 2007). *C. indicum* and *C. nankingense* (once regarded as a variety of the former), *C. dichrum* and *C. lavandulifolium*, *C. japonicum* and *C. yoshinaganthum*, *C. crassum* and *C. rhombifolium* are closely related. The meiotic behavior of several *Chrysanthemum* taxa and their hybrids suggested that three diploid species are closely related to each other (Cui, 2004), among which *C. dichrum* and *C. lavandulifolium* are the closest (Figure 2), followed by *C. lavandulifolium* and *C. nankingense*, and *C. nankingense* and *C. dichrum* are the furthest. *C. japonense* and *C. vestitum*, *C. lavandulifolium* and *C. boreale*, *C. zawadskii* and *C. weyrichii* are also closely related. However, they are still independent species and should not be combined into the single species. In neighbor-joining analysis of DNA sequences, *C. japonicum*, *C. arcticum*, *C. japonense*, *C. vestitum* and their respective subspecies/variety got together in pairs, showing a close kinship. The small genetic distance between *C. zawadskii* and *C. zawadskii* var. *latilobum* shows their close genetic relationship. However, in the morphological cluster analysis, these species and their variety/subspecies were not well aggregated, indicating the inconsistency between the phenotypic evolution and genetic evolution of

Chrysanthemum, with large differentiation of morphological traits and small genetic differentiation of DNA sequences.

In the clustering analysis based on 24 morphological characters and one biological character, Chinese *C. indicum*, *C. lavandulifolium*, *C. nankingense* and *C. dichrum* clustered together (Wu, 2007), and most *Chrysanthemum* taxa originating in Japan clustered together, indicating that there is an obvious geographical isolation between Chinese and Japanese *Chrysanthemum* plants. The Japanese wild species evolutionarily appeared later than Chinese ones. In PCR-RFLP, three CM varieties "Yinxing," "Jinlingwanxia" and "Shenma" had the same enzyme digestion pattern as that of wild *Chrysanthemum* in China, indicating their close genetic relationship. These CM cultivars have close genetic relationship with *C. indicum*, *C. vestitum* and *C. nankingense*, but they have the closest relationship with *C. chanetii*, and are far from *C. ornatum* and *C. japonense*, although wild species native to Japan have also played a positive role in enriching modern CM varieties. According to the morphological characteristics and molecular phylogeny, the wild *Chrysanthemum* species were classified into four groups, the indicum group, makinoi group, zawadskii group, and *Ajania* group (Nakano et al., 2021).

The multiple differentiation and hybridization/polyploidization cycles characterize the reticulate evolution of *C. indicum* complex (Yang et al., 2006) and *C. zawadskii* complex (Lu et al., 2022a), causing the difficulties in systematic classification. The DNA sequences of single-copy nuclear CDS (chrysanthemyl diphosphate synthase) gene and seven cpDNA loci were used to infer the phylogenetic relationship of 32 *Chrysanthemum* taxa and 11 species of the allied genera (Liu et al., 2012). It was found that the affinity between *Chrysanthemum* and *Ajania* is very close, and the resolving power of CDS and cpDNA markers were not enough. When eight nuclear sequences were used (Shen et al., 2021), *Chrysanthemum* and *Ajania* were well separated with only a few exceptions (Figure 2), which agrees with traditional taxonomy mainly based on capitulum morphology. In *Chrysanthemum*, two Chinese endemic species with white ray flowers, *C. rhombifolium* and *C. vestitum*, are sister to *C. indicum* complex with yellow/white ray flowers (Figure 2). The *C. indicum* complex has a and b branches; species of 'b' have island distributions, and species of 'a' are mostly distributed in mainland China. The *C. zawadskii* complex (Figure 2) has white to purple ray flowers with continuous color variation, and *Opisthopappus* falls into this clade. A Hyb-Seq based phylogenetic study found that *C. zawadskii*, *C. naktongense*, *C. chanetii* and *C. maximowiczii* intermingled in a clade with highly complicated phylogenetic relationships (Lu et al., 2022a), and the leaf morphology and phylogenetic analyses were difficult to distinguish them from each other, arguing that they have not yet differentiated into independent species and supporting the taxonomic treatments of *C. zawadskii* complex.

When nuclear CDS was used to infer the phylogenetic relationship, *Phaeostigma* is more closely related to *Chrysanthemum* + *Ajania* than other genera (Liu et al., 2012). According to pollen morphology and DNA data, *A. purpurea* is a member of *Phaeostigma*. The species differentiation in *Chrysanthemum* could be correlated with geographic and environmental factors. The *C. zawadskii* complex is distributed in northern China and *C. indicum* complex in southern China (Figure 3). Many polyploid species, e.g., *C. argyrophyllum*, could originate from divergent ancestors via allopolyploidization. The geographic/ecological conditions, hybridization and polyploidy play important roles in the divergence and speciation of *Chrysanthemum*. The comparative transcriptome analysis of CM and diploid *C. boreale* revealed whole-genome duplication (WGD) and gene selection patterns in these taxa (Won et al., 2017); the transcriptomes of *C. rhombifolium* and *C. dichrum* were also preliminarily characterized (Liu et al., 2021a; Zhang et al., 2021a), but the phylotranscriptomic analysis covering more *Chrysanthemum* species has not been reported.

In order to clarify the origin of CM, the investigation and collection of Chinese medicinal *Chrysanthemum* varieties with less germplasm erosion, ancient ornamental big CM (Daju in Chinese) varieties, *Chrysanthemum* sensu stricto and related genera were highlighted (Zhou, 2009); these plants were used as the materials of various experiments for the origin of CM by using different technical methods. Natural hybridization and cultivation play an important role in the origin of CM. In the distant hybridization of four years, 47 combinations were carried out, and 3,225 seeds, 675 hybrid seedlings and 30 flowering variant seedlings were obtained. *C. vestitum* and *C. indicum*, *C. vestitum* and *C. zawadskii*, *C. vestitum* and *C. lavandulifolium* were easy to cross successfully, and seeds were obtained. In the same distribution area of *C. vestitum* and *C. indicum*, the flowering period of the former is slightly earlier than that of latter, and they meet in flowering period (Wang, 2020). The F1 generation had great variation, and the inflorescence, flower color and leaf type were close to modern CM. The F2 progeny showed greater variation after natural pollination, which was very similar to ornamental small CM (Xiaoju). Hybridization is the main approach of origin of CM. *C. vestitum*, *C. indicum*, *C. zawadskii*, *C. lavandulifolium* and *C. dichrum* are closely related and play a role in the origin of CM to varying degrees. There are more frequent natural crosses between *C. indicum* and *C. vestitum*, possibly due to the weak reproductive isolation between tetraploid and hexaploid (Wang, 2020). The phylogenetic analysis of nuclear LFY sequences suggested that different cultivars had different ancestors (Ma et al., 2016). *C. indicum*, *C. zawadskii* and *C. nankingense* might be the direct ancestors of most CM cultivars examined. *C vestitum* and *C. lavandulifolium* might be the ancestor of some cultivars. Another line of support is from simple sequence repeat (SSR) markers mined from expressed sequence tags (ESTs; Fan et al., 2019) of CM. Transferability of EST-SSR primers among *Chrysanthemum* species was identified in *C. indicum* (96–100%), *C. nankingense* (100%), *C. chanetii* (96%), *C. zawadskii* (96%), and *C. indicum* var. *aromaticum* (92%).

The cluster analysis of ISSR (inter-simple sequence repeat)-PCR showed that the genetic relationship between *C. vestitum* and Daju/medicinal CM is the closest (Zhou, 2009), followed by *C. vestitum* of Yichang, Hubei Province (*C. vestitum var. vestitum*, Wang, 2020), *C. zawadskii* and *C. nankingense*. The main evolution mode of CM could be: Wild *Chrysanthemum* → medicinal CM (economic original chrysanthemum) → ornamental CM. Based on the research results on the origin of CM for half a century, the macro-morphology is combined with cytology, palynology, isozyme, DNA molecular markers, which are further combined with artificial hybridization, field resource investigation, introduction and domestication, and historical analysis, and it is concluded that the cultivated CM (original CM beginning in the era of poet Tao Yuanming, Jin Dynasty) originated in China; it is a hybrid cultigen complex, and is mainly produced by natural interspecific hybridization of some wild species in the middle reaches of Yangtze River (Hubei, Anhui and Henan). After long-term artificial repeated selection of some special variation types, it is carefully cultivated. The main parents are *C. vestitum* and *C. indicum*. Later, *C. zawadskii*, *C. lavandulifolium*, *C. indicum var. aromaticum*, *C. nankingense* and *C. dichrum* participate in the evolution of CM through pollen to varying degrees; the germplasm introgression enhanced the diversification of CM genes, so today's ever-changing CM is formed. The recurrent hybridizations between several wild progenitor species contributed to the evolutionary novelty of CM (Song et al., 2018), and cultivation/selection also play a significant role in the formation of CM.

Many wild *Chrysanthemum* species may be involved in the complex formation process of CM. The ISSR-based UPGMA (unweighted pair group method with arithmetic means) clustering suggested that the *Chrysanthemum* species evolved from low ploidy to high ploidy (Liu, 2010); it can generally be inferred that the flat petal is the basic petal shape of ornamental CM. *C. nankingense* and CM had the closest genetic relationship, *C. chanetii*, *C. japonicum*, *C. japonicum var. wakasaense* were also closely related to CM, and *C. indicum var. aromaticum* was farthest related to other samples. The random amplified polymorphic DNA (RAPD) analysis showed that the genetic diversity of wild species was more abundant than that of CM. The cultivated Xiaoju, rather than Daju, was genetically closer to the wild *Chrysanthemum* species, and the petal shape and flower diameter can be used in the classification of CM. AFLP markers were used to detect the relationships among 12 wild accessions and 62 groundcover cultivars (Chen et al., 2013). The genetic variation is abundant in chrysanthemums. The 74 accessions were classified by UPGMA; the genetic relationship was the most relevant factor in AFLP-marker clustering, and petal type was also informative. AFLP technology could be very efficient for discriminating *Chrysanthemum* species and related genera and reconstructing their genetic relatedness.

Intraspecific relationship of CM and origin of cultivars

Several genus names have been used for CM, such as *Matricaria, Anthemis, Pyrethrum, Chrysanthemum* and *Dendranthema*, reflecting its ambiguous phylogenetic position. The species names of *grandiflora* vs. *morifolium* were contentious. The International Code of Botanical Nomenclature changed the defining species of genus to *C. indicum* (Figure 1), and restored the florist's chrysanthemum to the genus *Chrysanthemum* (Poe and Swofford, 1999). CM is a perennial herb with 50–140 cm height, has erect and slightly purplish red stems (FRPS, 1983); the base of the plant is woody, with much branching covered by densely white pubescent. The leaves of CM have a shape of ovate to lanceolate, with a color of dark green above and greenish beneath covered by densely white short hair on both sides. The flower heads are on the branches separately or aggregated with each other, with 2.5–15 cm in diameter; the involucral bracts of CM have multiple layers and ligulate flowers have diverse colors. The tubular flowers are yellow, blossoming in September to October in the northern hemisphere.

CM has been cultivated for more than 1,600 years and is one of the top ten traditional famous flowers in China. It has rich variation, strong adaptability, wide distribution and complex genetic background, which has brought great difficulties to the investigation and protection of germplasm resources, identification and classification, genetic diversity analysis, and the like. The genetic diversity of 53 CM germplasm samples with different flower diameter, petal type and flower color type was studied using conserved DNA-derived polymorphism (CDDP) markers (Li, 2014) to provide a basis for the classification and identification of chrysanthemum varieties. In the cluster analysis of CDDP amplification results, 53 chrysanthemum materials were divided into six groups. Group 1 includes 30 Daju varieties and two Xiaoju varieties, group 2 contains 16 Xiaoju varieties, and group 3 includes two Xiaoju varieties. *C. nakingense*, *C. indicum* and *C. vestitum* were in groups 4, 5 and 6, respectively. The CM varieties were basically clustered according to flower diameter, which has no direct correlation with petal type and flower color. The phylogenetic analysis based on large sets of single nucleotide polymorphisms (SNPs) revealed that the Xiaoju types and potted/ground chrysanthemum, instead of Daju, are more closely related to the wild progenitor species (Chong et al., 2016).

The CM varieties are classified in terms of the parameters such as flowering stage, flower diameter, flower color, appearance of petals, cultivation form, leaf shape, among others. Both medicinal CM, i.e., *Flos Chrysanthemi* in TCM, and ornamental CM can be classified according to these external characteristics. Before Song Dynasty, the artificially cultivated CM was for food/medicine and viewing, and the wild relatives were also used as medicine. With the increased use of medical varieties, the cultivation flourished in Qing Dynasty, when CM's genuine production areas were formed (Figure 4). The changes of CM producing areas in ancient times are shown in Supplementary Figure S1. With the

passage of time, changes of cultivation sites and processing methods, eight famous CM categories are gradually formed (Figure 4); *Chinese Pharmacopoeia* contains five Daodi medicinal materials (geoherbs) of CM: Hangju, Boju, Gongju, Chuju, Huaiju (Chinese Pharmacopoeia Commission, 2020). There are also local varieties such as Qiju, Jinsi Huangju, Wuyuan Huangju and so on (Liu, 2020). Gongju and Hangju could be closely related (Wang et al., 1999, 2001), Huaiju, Jiju, Boju and Qiju are more closely related, while Chuju is an independent variety. Interestingly, among 26 antifungal terpenoids of 13 CM cultivars, longifolenaldehyde and β-selinene were identified only in Chuju (Xue et al., 2019). At present, Hang Baiju (Hangzhou white CM) and Gongju in the south of Yangtze River are mainly used for tea, while Huaiju, Boju, Chuju, Qiju and Jiju in the north of Yangtze River are mainly used for medicinal purposes.

Hangju, including Hang Huangju (Hangzhou yellow CM) and Hang Baiju, has been cultivated for about 400 years. Huangju was mainly used as medicine while Baiju was consumed as tea. In the 1950s and 1960s in Futianhe Town, Macheng, Hubei Province, Hangju (Big Baiju) was introduced from Tongxiang, Zhejiang Province (Chang et al., 2019). The large-scale planting began in the 1980s. The local CM is produced in Futianhe Town as the core production area, so they are also called "Macheng Fubaiju" or "Fubaiju." The cultivars of Fubaiju mainly include Baiju and Jinju (golden CM), which can also be divided into two major varieties: "local early-flowering" and "local late-flowering" (Xiong, 2014); they were mainly used for tea. Surprisingly, the ISSR cluster analysis showed that the genetic distance between Fubaiju and Hang Baiju was relatively far (Xiong, 2014), and it was closer to heterophyllous chrysanthemum and Chuju in Chuzhou, Anhui Province.

Boju has been cultivated for at least 240 years, but its processing method of sulfur-fumigation arouses controversy. Twenty-one medicinal CM cultivars were heteroploidy (Wang et al., 2012), mostly hexaploid, and some were tetraploid (Huangyaoju) and pentaploid (Dahuangju). The karyotype of Da Boju, Xiao Boju, Chuju, "Xenogeneic Dabaiju" and "Little Baiju" is 2A, and the other 16 types are 2B. The chromosomes of 21 CM cultivars showed polymorphism, and the chromosome length, centromere position and satellite varied among cultivars. In a ISSR analysis (Lyu et al., 2008), both Da Boju and Xiao Boju were closer to different cultivars of Hangju. A cultivar-specific sequence-characterized amplified region (SCAR) marker was developed to detect Da Boju (Cai et al., 2022), which is morphologically similar to ornamental Xiaoju, and could be artificially evolved from Xiaoju.

Chuju, Hangju and Boju have deep notch correctitude-leaves (Li et al., 2010); unlike other cultivars, Chuju and Boju have no auricle. In a UPGMA analysis based on the DNA fingerprint patterns of SSR markers (Feng et al., 2016), Chuju was closer to Boju, followed by Gongju and Huaiju. In a ISSR analysis (Lyu et al., 2008), Chuju was closer to Jiju than to other medicinal CM. The uniqueness of Chuju was also validated by the quality evaluation and identification studies based on monoterpenoids and sesquiterpenoids (Wang et al., 2008b). Gongju is introduced for over 100 years mainly as tea beverage. Gongju grows in an environment of high altitude and high humidity during harvest, and the main processing method was drying by charcoal fire. In a ISSR analysis (Lyu et al., 2008), Gongju was closer to Hangju, agreeing with morphological and metabolomic investigations (Wang et al., 1999, 2001; Nie et al., 2019).

Huaiju is considered as the ancestor of medicinal CM because of its ancient records in medicinal uses. In a RAPD analysis (Dai et al., 2017), Huaiju was not clustered with Hangju, Gongju and *C. indicum* (*Chrysanthemi Indici Flos* (CIF) in TCM; Song et al., 2020), indicating the rich genetic diversity in chrysanthemum germplasm resources. Huaiju samples were clustered into a large group, indicating its high genetic purity. In a ISSR analysis (Lyu et al., 2008), *C. indicum*, *C. nankingense* and a hybrid between *C. indicum* and Gongju were grouped together. The 19 medicinal CM accessions were divided into two groups according to the origin. Most accessions originating in the north of Yangtze River had relatively close genetic relationships, while most accessions cultivated in the south also had relatively close relationships. The resolving power of ISSR is generally higher than that of RAPD (Xu et al., 2006; Luo et al., 2013), facilitating the clarification of inter-cultivar relationship of CM. Medicinal CM germplasm resources are indeed distinct at the molecular level; the differences among medicinal CM cultivation types are related to environmental factors (Huang et al., 2020), but to a greater extent are determined by their genetic factors. The quantitative analysis of morphological variation (Shao et al., 2011) and isozyme analysis (Ding et al., 2008) also help much in elucidating the inter-cultivar relationship, variety identification and conservation at different phenotypic levels.

Phytometabolites and chemodiversity of *Chrysanthemum*

Plant pharmacophylogeny is an interdisciplinary subject that integrates plant systematics, phytochemistry, TCM, pharmacology and other multidisciplinary knowledge systems (Hao and Xiao, 2017, 2020), and plays an important role in guiding the conservation, development and utilization of medicinal plant resources. The core idea of plant pharmacophylogeny is that species in adjacent phylogenetic groups have relatively close genetic characteristics; the similar gene sequences lead to relatively close biosynthetic pathways of various phytometabolites in closely related taxa, manifesting a high degree of similarity in the chemical composition, i.e., (1) specific secondary metabolites are more likely to be distributed in multiple species that are genetically close and (2) natural products in closely related taxa have higher similarity in molecular skeleton composition and the relationship of derivation (Liu et al., 2021b). The similarity of the above two tiers of chemicals is manifested in the overall similarity of biological activity or therapeutic effect in clinical application. In practice, the concept of plant pharmacophylogeny has effectively guided the in-depth development and utilization of

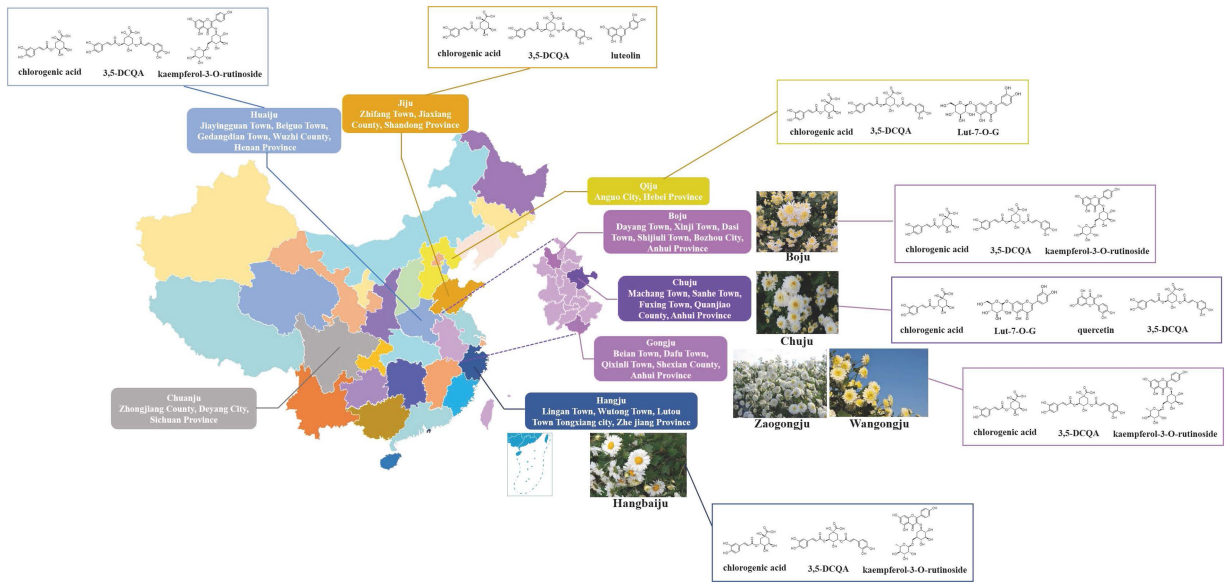


FIGURE 4
Geographical distribution, chemical and morphological characteristics of representative CM cultivars. The medicinal quality (Q) markers of each cultivar are shown: Boju, Huaiju, Gongju and Hangju: chlorogenic acid, 3,5-DCQA, and kaempferol-3-O-rutinoside (Lu et al., 2022b); Chuju: chlorogenic acid, Lut-7-O-G, quercetin, 3,5-DCQA (Yang et al., 2018b); Qiju: chlorogenic acid, 3,5-DCQA, and Lut-7-O-G (Peng et al., 2019); Jiju: chlorogenic acid, luteolin and 3,5-DCQA (Kang et al., 2022). The Q-marker of Chuanju of Zhongjiang, Sichuan Province is not reported. Morphology of medicinal CM: Boju: Inverted conical or cylindrical shape, sometimes slightly flattened and fan-shaped, 1.5–3cm in diameter, discrete (Chinese Pharmacopoeia Commission, 2020). Involucral bracts dish-shaped; involucral bracts 3–4 layers, ovate or elliptic, herbaceous, yellow-green or brown-green, pubescent outside, margin membranous. Chuju: Irregular spherical or oblate spherical, diameter 1.5–2.5cm. Ligulate flowers are white, irregularly twisted, involute, with shriveled edges, sometimes with light brown glandular dots; tubular flowers are mostly hidden. Gongju: Oblate spherical or irregular spherical, 1.5–2.5cm in diameter. Ligulate flowers white or off-white, obliquely ascending, upper part reflexed, margin slightly involute and shriveled, usually without glandular dots; tubular flowers few, exposed. Hangju: It is dish-shaped or oblate spherical, with a diameter of 2.5–4cm, and is often connected in several pieces. Ligulate flowers white or yellow, spreading or slightly folded, adhering to each other, usually without glandular dots; tubular flowers numerous, exposed. Huaiju: Irregular spherical or oblate spherical, diameter 1.5–2.5cm. Ligulate flowers are the most, white or yellow, irregularly twisted, involute, with shriveled edges, and sometimes glandular dots can be seen; most of the tubular flowers are hidden.

medicinal plant resources (Hao and Xiao, 2017, 2020), especially in the development of new medicinal plant species, so as to avoid the blindness in traditional research methods and guarantee the targeted research. Fully understanding the phytometabolites and chemodiversity of *Chrysanthemum* and related taxonomic groups is an indispensable basis for pharmacophylogenetic research.

Volatile oils

Volatile oils were one of the most important bioactive components in Chrysanthemum species, which are mainly composed of hydrocarbons, terpenoids, alcohols, esters, ketones, aldehydes, among others (Ryu et al., 2019; Wang et al., 2021a). The herbivory-induced emission of volatile terpenes in Chrysanthemum is an indirect defense against pest by attracting natural enemies (Xu et al., 2021). The candidate terpene synthase (TPS) genes were identified by comparing the transcriptomes of healthy and pest-infested CM plants. Totally 46 terpenoids were identified in flower heads of 44 Chrysanthemum species/varieties and 43 were identified in the emission (Zhang et al., 2021b). The CM flowers had less terpenoids than their wild relatives, and displayed lower emission rates. The differences were principally determined by seven monoterpenes (camphor, endo-borneol, bornyl acetate, sabinene 1,8-cineole, filifolone, β-myrcene; Figure 5) and five sesquiterpenes (germacrene D, β-ylangene, (E)-β-farnesene, β-copaene, β-caryophyllene). At least 183 monoterpenoids have been reported in Chrysanthemum (Jiang et al., 2021b), e.g., the abundant borneol, camphor, *β*-pinene, *α*-thujone, and verbenone, etc. (Xie et al., 2012). CDS catalyzes the condensation of two molecules of dimethylallyl diphosphate to chrysanthemyl diphosphate, an irregular monoterpene (Liu et al., 2012). Irregular monoterpenes are mainly found in plants of the tribe Anthemideae. The monoterpene glycosides were identified from flowers of edible *C. "Kamiohno"* (Kurimoto et al., 2021).

In *C. indicum*, 17 monoterpenes and 27 sesquiterpenes were identified (Zhou et al., 2021). The recombinant CiTPS1 and 2 produced α-pinene, CiTPS3 was responsible for the production of sesquiterpenoids β-farnesene, petasitene, and α-bisabolene, while CiTPS4 contributed to the production of four monoterpenoids and three sesquiterpenoids. These functionally redundant genes were derived from WGD and segmental duplication of Chrysanthemum (Song et al., 2018; Nakano et al., 2021). Among the expanded C. nankingense gene families, the genes involved in

the biosynthesis of monoterpenoids, sesquiterpenoids, triterpenoids are highly enriched (Song et al., 2018). The terpenoid synthase (TS) and cytochrome p450 (CYP) families are among the most highly enriched functional categories, and 219 TS genes (7 squalene synthases, 158 TPSs, and 54 triterpene cyclases) and 708 CYP genes were identified in C. nankingense genome, suggesting a significant expansion of TPS genes, explicitly in the TPS-a/−b subfamilies. Unlike Chrysanthemum, much less TS and CYP genes were identified in the other three Asteraceae genomes, i.e., artichoke, sunflower, and lettuce.

At least 207 sesquiterpenoids have been identified from Chrysanthemum (Chen et al., 2019; Jiang et al., 2021a,b), including 26 germacrane-type, 26 eudesmane-type, 64 guaianolide-type, 4 bisabolane-type, and 15 other-type sesquiterpenoids (Figure 5), with anti-inflammatory, antibacterial, anti-tumor, insecticidal, and anti-viral activities, etc. CM, *C. indicum*, *C. lavandulifolium*, *C. zawadskii*, and *C. ornatum* are rich in sesquiterpenes, especially the former two. Angeloylcumambrin B, cumambrin A and handelin were reported from more than one species. Germacrane-type sesquiterpenes are monocyclic, composing of a 10-member carbon ring, a methyl group at C-4 and C-10, and an isopropyl group at C-7. Eudesmane-type sesquiterpenes are bicyclic, consisting of two six-member carbon rings with methyl groups at C-4 and C-10, and an isopropyl group at C-7. Guaianolide-type sesquiterpenes have ternary rings consisting of a five-member ring, a seven-member ring, a five-member γ-lactone ring, and methyl groups at C-4 and C-10. They are more abundant than other types of sesquiterpene. The cyclization of farnesyl diphosphate to germacrene A could be the first committed step in sesquiterpene biosynthesis (Jiang et al., 2021b), and oxidations of (+)-germacrene A determine where additional cyclization occurs to generate guaianolides or eudesmanes. In GC/MS analysis of CM, *β*-humulene (*β*-caryophyllene) showed the highest contents, accounting for 16.3% of the total 58 detected volatiles (Sun et al., 2010b); ledene oxide-(I) was also abundant, amounting to 9.0% of total volatiles.

The constituent and yield of CM essential oils varied a lot depending on the harvest time, origin, processing methods, etc. (Xue et al., 2007; Yang et al., 2007; Wu et al., 2015). The flowers were blue-green in the early bloom stage and contained the most abundant volatile oil (Xue et al., 2007). There is correlation between floral volatile components and antioxidant activity of different CM cultivars (Yang et al., 2017a), therefore it is important to quantify contents of volatile oils in the respective cultivar. Chuju, Gongju and Qiju had 2.0–4.0 ml/kg volatile oil, Huangyaoju (a kind of Gongju), Boju and Huaiju had more (4.5 ml/kg), and Jiju had the highest volatile oil content of >10.0 ml/kg, whereas two kinds of Hangju had the lowest content of <1.0 ml/kg (Xu et al., 2005).

Based on the volatile chemical profile of flower tea, Hangju, Huangju, chamomile (*Matricaria chamomilla*), and a new CM "Xiaokuixiang" were well distinguished from each other (Wang et al., 2021a). The main volatile components in chamomile were the ester ethyl 2-methylbutanoate and alcohol 1,8-cineole monomer, but the content of ketones in chamomile was relatively low; ethyl 2-methylbutanoate and ketone 6-methyl-5-hepten-2-one were more abundant in "Xiaokuixiang." The terpenoids 6-methyl-5-hepten-2-one and 1-menthol monomer were salient in Hangju and Huangju, but very few esters were detected in both, agreeing with their close phylogenetic relationship.

Triterpenes and other terpenoids

More than fifty triterpenes were identified from *Chrysanthemum*, including tetracyclic and pentacyclic triterpenes (Figure 5). The former includes stigmastanes, lanostanes, dammaranes, cycloartanes (more than others), tirucallanes. The latter includes lupanes, taraxeranes (more than others), oleananes, and ursanes. Squalene, camphor, DL-α-tocopherol were among the top 10 volatile compounds of 12 γ-irradiated mutant cultivars of CM (Ryu et al., 2019), which are useful for classification and identification of chrysanthemum mutant cultivars. Though, the triterpene biosynthesis genes of *Chrysanthemum* and related taxonomic groups have not been reported, and the bioactivity of *Chrysanthemum* triterpenes is less studied. One diterpene (C20) was found in roots of 12 CM cultivars (Zhang et al., 2020).

Flavonoids

More than 78 CM flavonoids (Figure 6) are identified, and flavones are the most, followed by flavonols and flavonones. They fall into 12 categories (Jo et al., 2020; Mao, 2020), i.e., apigenin (Api) and derivatives (mainly glycosides), diosmetin (Dio) and derivatives, acacetin (Acn) and derivatives, kaempferol and derivatives, quercetin and derivatives, luteolin (Lut) and derivatives, hesperetin and derivatives, eriodictyol and their glycosides, isorhamnetin, baicalin, eupatilin, and anthocyanins. Among them, apigenin and derivatives are the most, mainly Api-7-O-6-AG (acetylglucoside), Api-7-O-G and the like, which is followed by luteolin and its derivatives, e.g., Lut-7-O-G (glucoside), Lut-7-O-MG (malonylglucoside) and so on. These are also abundant in *C. indicum* and *C. nankingense* (Zou et al., 2022). In *C. pacificum*, luteolin conjugates were mostly enriched in flowers (Farag et al., 2015), and non-flowering aboveground parts were rich in quercetin and methoxylated flavone conjugates. Root sample had the lowest contents of all flavones and dicaffeoylquinic acids (DCQAs).

Among 477 metabolites identified in *C. mongolicum* and *C. rhombifolium* (Duan et al., 2022), 69 are listed in the Chinese Pharmacopoeia (2015) and YaTCM database (Li et al., 2018). The flavonoids, e.g., linarin (11.64%), chrysanthin (5.03%), farnesin (3.03%) and genkwanin (2.75%), are abundant in *C. mongolicum*. The contents of genkwanin, trigonelline, Dio, narcissoside, 3,4-dihydroxyphenylacetic acid, linarin, N′,N′-p-coumarin, C-hexosyl-tricetin O-pentoside, chrysoeriol, Acn and

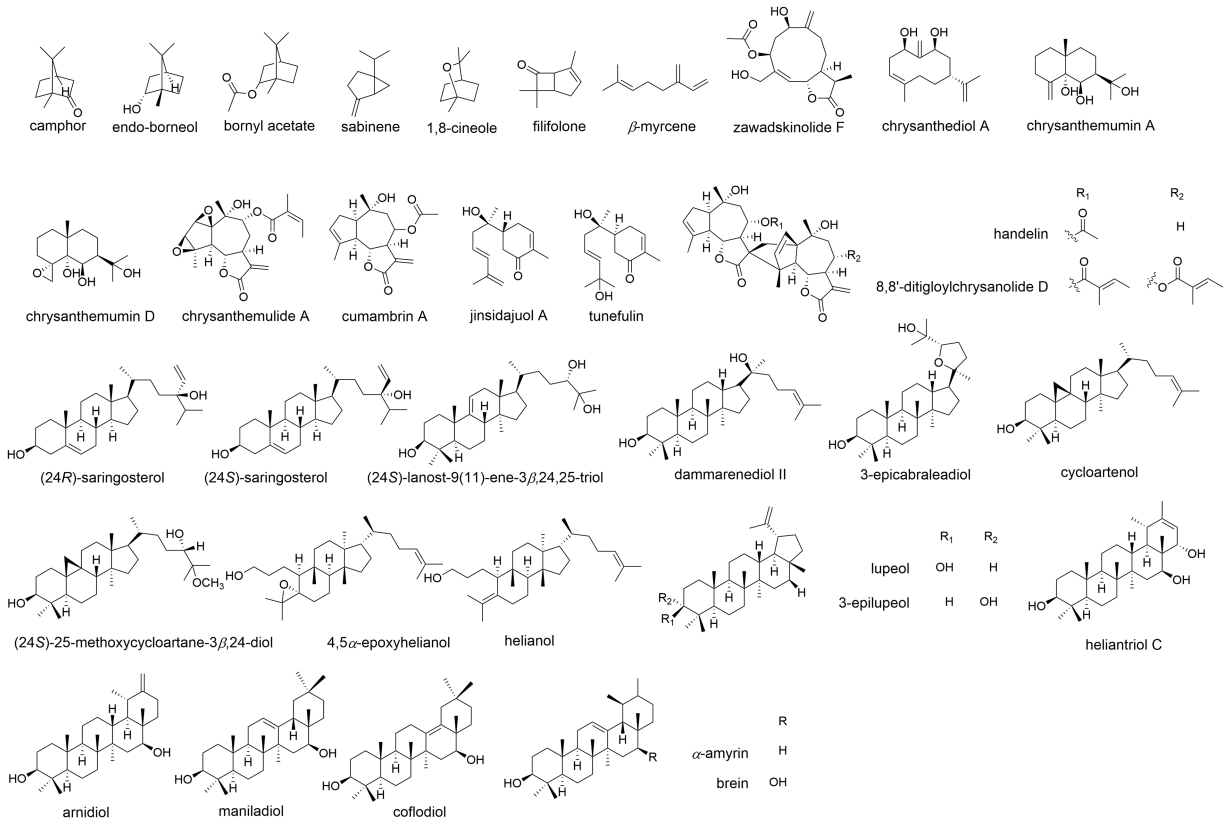


FIGURE 5
The molecular structure of representative terpenoid components of *Chrysanthemum*. A, Monoterpene: camphor, endo-borneol, bornyl acetate, sabinene,1,8-cineole, filifolone, β-myrcene. B, Sesquiterpene: germacrane-type: zawadskinolide F (anti-inflammatory, *C. zawadskii*), chrysanthediol A (anti-viral, CM); eudesmane-type: chrysanthemumin A (anti-viral, *C. indicum*), chrysanthemumin D (anti-viral, *C. indicum*); guaianolide-type: chrysanthemulide A (anti-tumor, anti-inflammatory, *C. indicum*), cumambrin A (treating osteoporosis, *C. ornatum, C. indicum, C. zawadskii*); bisabolene-type: jinsidajuol A (CM), tunefulin (*C. indicum*); others: handelin (anti-aging, *C. ornatum, C. indicum*), 8,8′-ditigloylchrysanolide D (anti-tumor, *C. indicum*). C, Triterpene: tetracyclic: stigmastanes: (24*R*)-saringosterol, (24*S*)-saringosterol (CM); lanostane: (24*S*)-lanost-9(11)-ene-3*β*,24,25-triol (CM); dammaranes: dammarenediol II (CM), 3-epicabraleadiol (anti-viral, CM); cycloartanes: cycloartenol (various bioactivities, CM), (24*S*)-25-methoxycycloartane-3*β*,24-diol (anti-inflammatory, CM); tirucallanes: 4,5*α*-epoxyhelianol (antitubercular, CM), helianol (anti-inflammatory, CM). pentacyclic: lupanes: lupeol (anti-inflammatory, *C. indicum*, CM), 3-epilupeol (antitubercular, CM); taraxeranes: heliantriol C (anti-tumor, CM), arnidiol (anti-tumor, CM); oleananes: maniladiol (antitubercular, CM), coflodiol (anti-tumor, CM); ursanes: *α*-amyrin, brein (CM).

kaempferol-3-O-gentiobioside were significantly different between *C. mongolicum* and *C. rhombifolium* (Duan et al., 2022), which could be biomarkers and inspire drug/food development.

Lut-7-G and quercitrin were the top two abundant compounds of CM, accounting for 85.7% of total flavonoids (Sun et al., 2010b). More Acn 7-O-galactoside resulted in lighter colored flowers and less Acn and kaempferol was associated with yellow flowers (Chen et al., 2021). The principal component analysis (PCA) of metabolic profile data separated 10 CM cultivars according to flower color rather than mutation rates (Sawada et al., 2019). Hot-H_2O extraction of CM tea showed that most flavonoids and CQAs dissolved out at 30 min (Chen et al., 2021), with 20.977 and 8.958 mg/g gross weight (Learn how to make good CM tea).

The contents of taxifolin (dihydroquercetin) were clearly associated with the variation in mutation rates among 10 CM cultivars (Sawada et al., 2019). Taxifolin is a precursor of both cyanidin derivatives (pigments) and quercetin derivatives (colorless co-pigments modifying flower color). Higher accumulation of taxifolin could easily result in higher accumulation of cyanidin and/or quercetin derivatives by mutagenesis, causing changes in flower color. The anthocyanins cyanidin 3-G and cyanidin 3-(3′-malonoyl) glucoside were identified in 23 CM cultivars (Park et al., 2015), among which "Magic," "Angel" and "Relance" had high amounts of anthocyanins and showed a wide range of red and purple colors in their petals. Dihydroflavanol 4-reductase (DFR) and anthocyanidin synthase (ANS) convert dihydrokaempferol to pelargonidin, a pink anthocyanidin. In CM, the B-ring of dihydrokaempferol is further hydroxylated to cyanidin by flavonoid 3′-hydroxylase (F3′H; Mekapogu et al., 2020). Due to the absence of flavonoid 3′5′-hydroxylase (F3′5′H), there is no delphinidin-based anthocyanin accumulated in CM petals. The accumulation of cyanidin and pelargonidin imparts pink to red-purple, orange to red, respectively. To date, the anthocyanin biosynthesis genes of

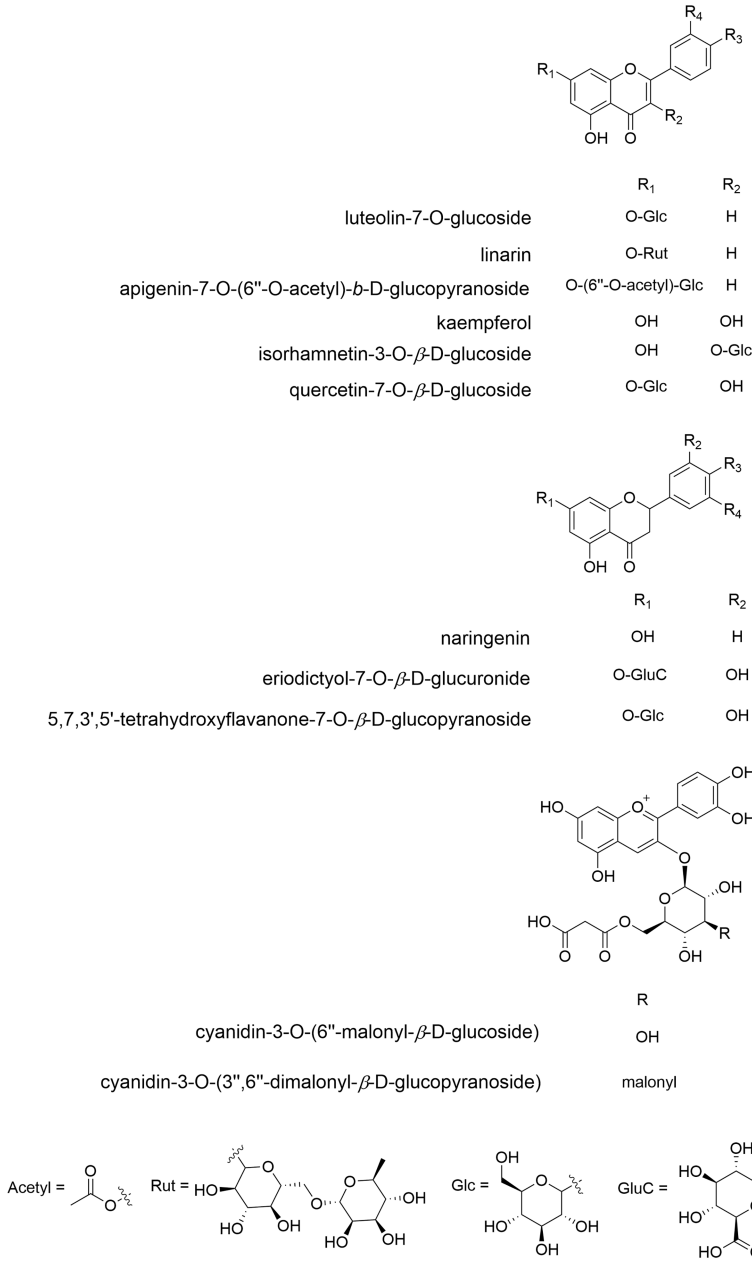


FIGURE 6
The molecular structure of representative flavonoid components of Chrysanthemum. Flavones: Lut-7-O-G (anti-inflammatory, antioxidant, relieving asthma, xanthine oxidase inhibitor, CM, *C. indicum*), linarin (antioxidant, anti-inflammatory, preventing acute lung injury, promoting osteogenic differentiation, inhibiting acetylcholinesterase activity, CM, *C. indicum*, C. zawadskii), Api-7-O-(6″-O-acetyl)-β-D-glucopyranoside (antioxidant, CM). Flavonols: kaempferol (antioxidant, anti-inflammatory, anticancer, CM, *C. indicum*), isorhamnetin 3-O-β-D-G (anti-inflammatory, CM, *C. indicum*), quercetin 7-O-β-D-G (various bioactivities, CM, *C. indicum*). C, flavonones: naringenin (treating depression, CM), eriodictyol-7-O-β-D-glucuronide (immunoregulation, C. zawadskii, *C. indicum*), 5,7,3″,5″-tetrahydroxyflavanone-7-O-β-D-glucopyranoside (*C. indicum*). D, anthocyanins: cyanidin-3-O-(6″-O-malonyl)glucoside (*C. grandiflorum*), cyanidin-3-O-(3″,6″-di-O-dimalonyl-β-glucopyranoside)(*C. grandiflorum*).

wild *Chrysanthemum* taxa have not been reported; it is unknown whether F3′5′H is present in these taxa and how the variation of biosynthesis genes is involved in the contents of anthocyanins as well as spectrum of flower color.

The total flavonoid content of *Chrysanthemum* sensu lato varied between 4 and 10% (Shao, 2018), and its change was not obvious after the end of flowering. The content of linarin in CIF was significantly different among different habitats (Wei et al., 2021), where climatic factors, especially average annual precipitation, annual average sunshine hours, and annual average temperature, pointedly impacted on the linarin content. Some abundant flavonoids with content differences among CM cultivars

can be used as indicators for quantitative identification (Mao, 2020), e.g., Lut-7-O-G, Api-7-O-G, Lut-7-O-β-G, Api-7-O-6-AG. The components with high content, large dispersion among cultivars, and large contribution to classification can be used as evaluation indicators for the quality of CM and cultivar identification, e.g., Api-7-O-6-AG, 3,5-DCQA, 4,5-DCQA, 3,4-DCQA, 3 -CQA. The content of some phytometabolites is relatively low, and their differences between CM cultivars are large, e.g., Api, Dio, 5-CQA, 4-CQA, Acn; some components were not detected in some CM cultivars, e.g., Api, Acn. These are not suitable as indicators for *Chrysanthemum* quality evaluation. The CM cultivar "XiaoYangJu" (a kind of Hangju) had the highest total flavone, total flavonol and total CQA, which is very useful in investigating the biosynthesis pathway of active ingredients and breeding cultivars with the highest specialized metabolite yield.

Whether the flavonoid/anthocyanin biosynthesis is conserved across *Chrysanthemum* species is actually unknown (Mekapogu et al., 2020). In general, the biosynthesis of flavonoids begins with the condensation of 4-coumaroyl-CoA and malonyl-CoA by chalcone synthase (CHS) and chalcone isomerase (CHI) to yield naringenin (Hao et al., 2015). In CM, the flavonoid 3-hydroxylase and F3′H are responsible for the synthesis of dihydrokaempferol and dihydroquercetin as the essential precursors for the corresponding flavonols. Naringenin could also be transformed into flavones via flavone synthase (FS) 1. Flavonols and flavones, via glycosyltransferase (GT), methyltransferase, hydroxyl transferase, and acyltransferase, are subject to diverse structural modifications to generate various flavonoids. In CM, the common feature of flavonol modification was 3-O-glycosylation (Chen et al., 2021), while flavones were involved in 7-O-glycosylation and 7-O-6′-acylation; the metabolic intermediates are then shunted into seven sub-pathways. There were significant positive correlations between flavonoids and corresponding flavonoid glycosides, and there were meaningfully positive correlations between Api and Aca, Api and Lut, Lut and Dio, indicating coordinated sub-pathways in CM, and possibly in its wild relatives. However, few data of other Artemisiinae species are available. The content of total flavonoids in different tissues of *O. taihangensis* was higher than that in corresponding tissues of *C. indicum* (Liu, 2013), but biosynthetic pathways have not been well investigated in these species. Seven *C. indicum* samples from different habitats were distinct in their flavonoid profile (Zou et al., 2022), which was also significantly different from *C. nankingense*. So far, a comprehensive metabolomic similarity analysis of as many *Chrysanthemum* taxa as possible is still lacking, and the role of genetic vs. climatic factors in metabolic differences of populations/species is elusive.

The enrichment of flavonoid biosynthesis genes in the *C. nankingense* genome was revealed by whole genome sequencing (Song et al., 2018). Two key genes, CHS and CHI, are significantly expanded in *C. nankingense* genome (17 CHS genes and eight CHI genes), which showed the differential expression among tissues, and there were significantly high expressions in flowers, suggesting the spatial and temporal regulation of flavonoid biosynthesis. Numerous flavonoid UDP-glucuronosyl and -glucosyltransferase (UGT) gene copies were also identified in *C. nankingense* genome, which is consistent with the phytochemical findings.

Phenolic acids

Common phenolic acids of CM have long been known, e.g., chlorogenic acid, 4-CQA, 3,4-DCQA, 3,5-DCQA, 3,5-DC-epi-QA, 1,3-DC-epi-QA and ethyl caffeate (Xie et al., 2012; Yang et al., 2018a; Lu et al., 2022b). A rare 8-oxa-bicyclo[3.2.1]oct-3-en-2-one ring was reported (Yang et al., 2017b), which is formed through a [5 + 2] cycloaddition of CQA with a D-glucose derivative (Figure 7). *C. indicum* is rich in cryptochlorogenic acid and 3-O-p-coumaroyl quinic acid (Zou et al., 2022).

CQA and flavonoids showed a certain correlation with the stress tolerance of CM plants (Huang, 2010). Importantly, the medicinal quality evaluation indexes of *Chrysanthemum* can be identified from these two categories (Mao, 2020; Chen et al., 2021; Kang et al., 2022). Among the common phytometabolites shared by different medicinal cultivars, compounds with high content and great differences among cultivars, e.g., 4,5-DCQA, flavonoids Acn and Lut, are suitable for screening the index components of quality evaluation, which also have high contribution to CM classification. The content of 4,5- DCQA is high in most cultivars, so it is suitable to be used as a quantitative index component for *Chrysanthemum* quality evaluation. The content of 3-CQA in all samples was high (Mao, 2020), and the difference between cultivars was small. It is suitable to be used as a qualitative index component for *Chrysanthemum* quality evaluation. The low-abundance phytometabolites of *Chrysanthemum* cannot represent its chemical characteristics, e.g., 1-CQA, anthocyanins and carotenoids, which are not suitable for the quality evaluation. The anti-hepatotoxic 1,5-DCQA was at high levels in *C. pacificum* flowers and aerial parts reaching 3,145 and 1,390 μg/g, respectively (Farag et al., 2015), suggesting an alternative natural resource of medicinal compounds.

The correlation analysis between the pharmacologically active CM ingredients showed that there was a very significant positive correlation between chlorogenic acid, isochlorogenic acid B, isochlorogenic acid A and isochlorogenic acid C (Liu, 2020), and chlorogenic acid, isochlorogenic acid B/C and tilianin was significantly positively correlated. There was a very significant positive correlation between isochlorogenic acid C and tilianin, while cynaroside and Lut-7-O-β-D-glucuronic acid had a very significant positive correlation. There was a significant positive association between Dio-7-O-G and linarin, and a very significant positive correlation between the former and tilianin. These interesting results imply some commonly shared biosynthesis pathway of phenolics and flavonoids, as well as coordinated regulation of metabolic networks in *Chrysanthemum*. The phylogenetically related taxa not only have analogous chemical

constituents, their biosynthesis pathways and regulatory components, as well as metabolic networks and signal transduction networks, might also be similar.

The anti-inflammatory components of CM could be 3,5-DCQA, 4,5-DCQA, Lut-7-O-β-G, 1,3-DCQA, 3-CQA, and the like (Peng et al., 2019; Mao, 2020). Some of them were induced by jasmonic acid against western flower thrips (Chen et al., 2019). These compounds can be used as indicator components for quality evaluation of CM, which are also useful in studying the chemical relatedness between CM cultivars (Chen et al., 2021), as well as CM and wild relatives. According to the evaluation results of CM grades, 50 kinds of CM on the market, including Gongju, Hangju, Taiju (made from CM bud), etc., were divided into three grades (Mao, 2020), which was positively correlated with the content of index components. The lower the grade, the lower the index component content. This grading provided a reference for the establishment of TCM quality grade evaluation system. Interestingly, in the abundance of CQA, flavonoids and carotenoids, Boju was largely equivalent to Gongju and many cultivars of Hangju (Chen et al., 2021), although the former is phylogenetically distinct from the latter two (Lyu et al., 2008).

Taiju and full bloom CM of the same cultivar were relatively close in the PCA diagram based on index components (Mao, 2020), indicating that there is a great similarity between their chemical components. Among 13 cultivars, Gongju was the most unique. Gongju, Boju and the characteristic cultivar #1 were quite different from others. The clustering analysis based on contents of 13 components largely agreed with PCA results. Most Hangju samples were clustered into one group, and Huaiju, Chuju and some Hangju samples were clustered together, indicating a certain similarity between them. The characteristic cultivars had certain similarities with Boju, suggesting the kinship between them; alternatively, these might be partially explained by similar cultivation conditions, and/or similar terroir factors (Suo et al., 2011; Hao and Liu, 2021).

The cluster analysis based on chlorogenic acid, Lut, 3,5-DCQA, etc., showed that different CM cultivars were basically clustered by cultivar (Xiong, 2014), and the subgroup composed of Fubaiju cultivars "local early-flowering" and "local late-flowering," derived from Hang Baiju, clustered with Hangju and Gongju. The chemical cluster of Hangju and Gongju was also confirmed by other independent studies (Liu et al., 2019), which is basically consistent with the traditional classification and molecular marker-based classification of CM (Lyu et al., 2008; Mao, 2020).

The content of chlorogenic acid varied greatly among *Chrysanthemum* sensu lato plants, the highest was 0.975 g/100 g (Shao, 2018), and the lowest was only 0.086 g/100 g. After flowering, the content of chlorogenic acid in the plant will increase to a certain extent. The contents of chlorogenic acid, rutinoside, quercetin, Lut, Api in the leaves of *O. taihangensis* were higher than those in flowers, stems and *C. indicum* (Liu, 2013). The contents of chlorogenic acid, rutinoside and quercetin were higher in the stems and leaves, Lut and Api were less abundant in stems, and they were not detected in leaves. In different harvesting periods, the contents of five phenolics in different tissues of *O. taihangensis* were higher than those in corresponding tissues of *C. indicum*. Interestingly, *O. taihangensis* is often locally used as *C. indicum* in folk medicine, and the people in the producing area report that its aroma is strong, and its tea and medicinal value is far superior to that of *C. indicum*.

Oligosaccharide, polysaccharide

The sugar components of CM contribute to its dual value of food/medicine; the expressions of most glycolysis-, GT-related genes were induced by ultraviolet-B radiation (Yang et al., 2018b). The health-promoting raffinose and 1-kestose were abundant in *C. dichrum* leaves, flower buds, and blooming flowers (Liu et al., 2021a). The immunomodulatory and anti-inflammatory JFP1-1-2, a homogenous polysaccharide of non-medicinal parts of CM, is composed of mannose, glucose and galactose in a molar ratio of 4.53:3.06:1.00 (Tao, 2017); another bioactive polysaccharide JFP1-2-2 is composed of mannose, galactose, glucose, xylose and galacturonic acid in 3.0:2.3:1.2:1.0:1.0.

The content of polysaccharide prepared by membrane separation method of six CM cultivars of Kaifeng, Henan Province varied between 47.8 and 64.3% (Zhao, 2015), while those obtained by alcohol precipitation were 18.9 and 25.3%. The 295 g boiling water-extracted crude polysaccharide CMJA (yield 7.8%) and 96 g alkali-extracted polysaccharide JHB (2.5%) were obtained from 3.8 kg of dried Huaiju (Zheng, 2015). CMJA contained eight homogeneous polysaccharide fractions: CMJA0S1 (110 mg, yield 0.4%), CMJA0S2 (2.2%), CMJA1a S2 (2.3%), CMJA1a S3 (0.4%), CMJA1b S2 (0.5%), CMJA1b S3 (2.0%), CMJA2S2 (2.0%) and CMJA2S3 (2.2%). JHB contained a polysaccharide JHB0S2 (532 mg, yield 1.9%). CMJA0S1 is an arabinogalactan with molecular weight (MW) of 4.7×10^4 Da. The MW of CMJA0S2 is 6.5×10^3 Da; its main chain is a galactomannan glucan composed of 1,4-β-Glcp, 1,4-β-Galp, 1,4-β-Manp, with the branched chains composed of 1,6-β-Galp, T-α-Glcp, T-α-Araf, 1,5-α-Araf, etc. CMJA1a S2, CMJA1a S3, CMJA1b S2, CMJA1b S3 and CMJA2S2 are a series of RG-I type pectin polysaccharides with different MWs, and their galacturonic acid content and branched-chain glycosyl composition varied. 1,2-α-Rhap and 1,4-α-Gal Ap constitute the main chain, and at the O-4 position of rhamnose, there are branched chains composed of arabinose and galactose. CMJA2S3 is a HG-type pectin polysaccharide of 1.1×10^4 Da and composed of galacturonic acid. JHB0S2 is a xyloglucan of 1.6×10^4 Da, with 1,4-β-Glcp constituting the main chain and a branch at the O-6 position, connecting xylose residues.

Mineral elements

Mineral elements absorbed by *Chrysanthemum* plants play an important role in regulating the processes of primary and

10.3389/fpls.2022.973197

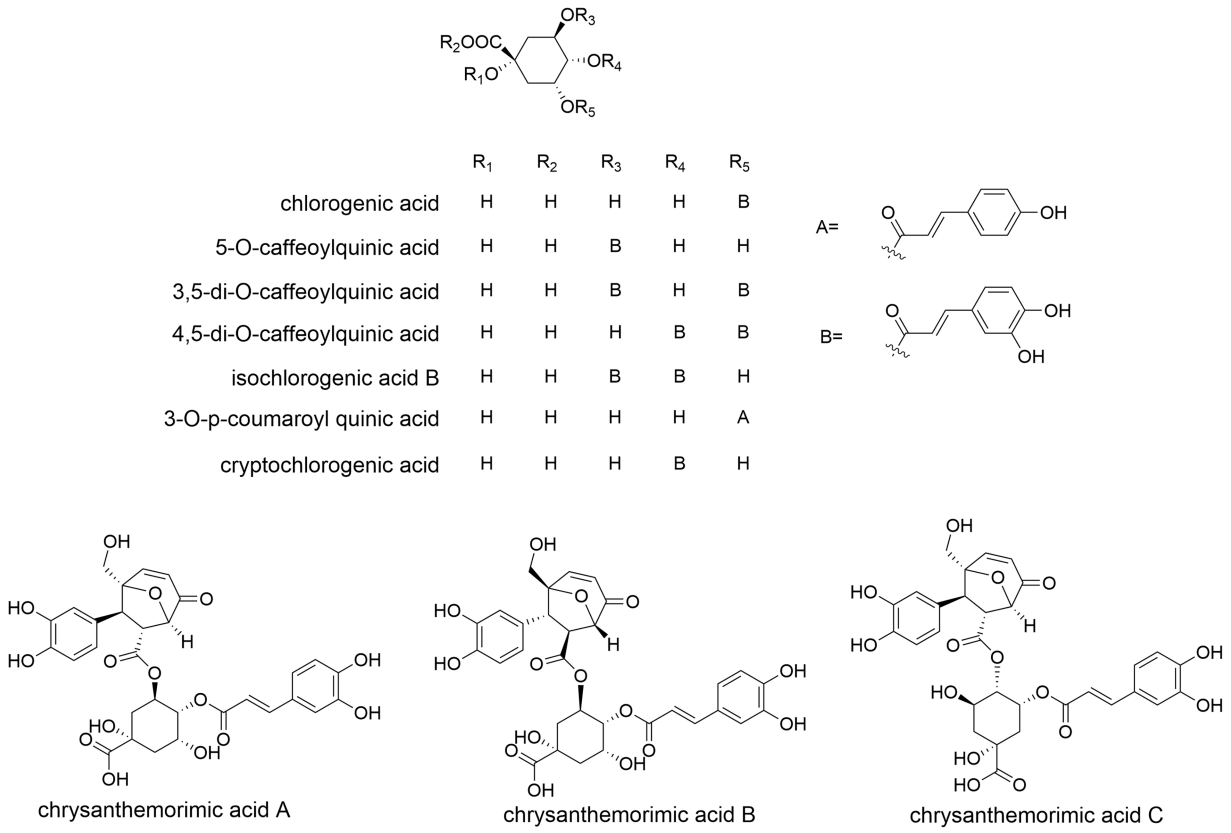


FIGURE 7
The molecular structure of representative phenolic components of *Chrysanthemum*. chlorogenic acid (=3-CQA), 5-CQA, 3,5-DCQA, 4,5-DCQA, isochlorogenic acid B, 3-O-p-coumaroyl quinic acid, cryptochlorogenic acid, chrysanthemorimic acid A/B/C.

secondary metabolism (Liu, 2020). They are the constituent factors of active ingredients in traditional medicine, and have a great impact on the formation and accumulation of bioactive metabolites in TCM plants. CM contained mineral elements such as calcium, magnesium, phosphorus, sulfur, potassium, and indispensable trace elements like copper, iron, zinc, cobalt, manganese, strontium, selenium. The total beneficial mineral elements (K, Ca, Mg, Fe, Na) contained in six CM cultivars of Kaifeng were between 32,008.7 and 40,183.1 mg/kg (Zhao, 2015), while the hazardous As, Pb, Hg and Cd were between 0.01–0.04 and 0.63–1.53 mg/kg. The CM flowers and leaves are rich in N, P, K, Ca, Mg, and Fe (Yan et al., 2021), among which K element had the largest variation, and N, Ca, Fe, Mg, and Zn had a large variation range. The absorption and accumulation of various elements in leaves of different germplasm resources are distinct. There is a strong positive correlation between Ca and Mg/Mn/Cd. The PCA found that CM cultivars were separated based on Mn, Cr, Cu, P, K. From the perspective of mineral elements, Hangju-Fuhuangju, Hangju-Xiaoyangju Late-ripening, Hangju-Sheyangju Late-ripening, Hangju-Dayanghua, Hangju-Subeiju were of good quality. The PCA score map divided the 23 CM samples into four groups, and the cluster analysis heat map divided them into five categories. The samples of the same cultivar clustered together well, indicating that the differences in the mineral element content of CM germplasm resources are closely related to genetic factors.

There was a significant positive correlation between CM isochlorogenic acid A/tilianin and N/P element contents (Liu, 2020; Liu et al., 2021c), isochlorogenic acid B and N element were significantly positively correlated, whereas chlorogenic acid, isochlorogenic acid A, C and cymaroside were significantly negatively correlated with heavy metal Pb. Cynaroside had a very significant negative relationship with Mn; Dio-7-O-G had a very significant positive correlation with Mn and Mg, and had a significant positive correlation with Ca and Pb. There was a significant positive correlation between linarin and Cd/Cr heavy metals. These imply the potential roles of phenolics and flavonoids in *Chrysanthemum* stress/defense response, as well as possible approaches of regulating phytometabolite yields.

Nutrition constituents

Seventeen amino acids were identified in CM, and contents of Lys, Phe and Leu were relatively high. The contents of seven essential amino acids, i.e., Thr, Val, Met, Ile, Leu, Phe and Lys, in

the manufactured goods were 2.3–6 times those in the fresh flowers (Wang et al., 2016). The total amino acids of six CM cultivars of Kaifeng varied between 10.5 and 11.9% (Zhao, 2015), and the essential amino acids were between 2.4 and 3.0%. Asparagine and glutamine were abundant in *C. dichrum* leaves, flower buds, and blooming flowers (Liu et al., 2021a).

Among 28 nucleobases, nucleosides, nucleotides and amino acids of CM of nine geographical origins, eight crucial quality markers, i.e., 2′-deoxyadenosine, guanosine, adenosine 3′,5′-cyclic phosphate (cAMP), guanosine 3′,5′-cyclic monophosphate (cGMP), arginine, proline, glutamate and tryptophan, were identified (Chang et al., 2021), which could be used to discriminate geographical origin of CM.

The crude fat content of six CM cultivars of Kaifeng was between 3.6 and 7.1% (Zhao, 2015). Twenty chemical constituents were detected after methylation of fatty acids, including five saturated fatty acids; unsaturated fatty acids accounted for 69.9% of the total fatty acids. The contents of fatty acids, e.g., capric acid (C10.0), pentadecanoic acid (C15.0), α-linolenic acid (C18.3 N3), eicosanoic acid (C20.0), docosanoic acid (C22.0) and lignoceric acid (C24.0), flavonoids and CQAs were increased in CM by UV-B radiation (Yang et al., 2018b). The contents of moisture, protein, fat, ash and carbohydrate in *C. mongolicum* tea were 8.2, 13.7, 4.1, 6.1 and 67.7%, respectively (Xiao, 2019); all essential amino acids were detected, and the content of amino acids was 11.3 g/100 g. Twenty-two kinds of fatty acids were detected, with abundant linoleic acid (22.8%) and α-linolenic acid (7.4%).

The carotenoids, e.g., zeaxanthin, carotene, cryptoxanthin, lutein, etc., were identified in different CM cultivars (Park et al., 2015; Zhou et al., 2019b; Mao, 2020; Chen et al., 2021). The average total content of carotenoid was highest in the 10% open, followed by 70 and 100% bloom stages at 0.50, 0.43, and 0.36 mg/g FW, respectively. The CM cultivars with higher carotenoid contents displayed yellow or green petal colors (Park et al., 2015). The expression of carotenoid biosynthetic genes in the petals of different CM cultivars at mid-development showed no differences (Kishimoto and Ohmiya, 2006). The antioxidant vitamin C contents of six CM cultivars of Kaifeng varied between 21.7 and 62.7 mg/100 g (Zhao, 2015). CM is beneficial to human health and is expected to be widely developed and applied in plant food, tea making, health care products, and such like. The nutrient composition measurement results were graded (Zhao, 2015), the nutritional index values of each CM cultivar were accumulated to obtain the comprehensive nutritional evaluation index, so as to compare various CM cultivars and select cultivars that are prioritized for human consumption.

When compared with *C. indicum*, quinones, flavonoids (e.g., naringenin 7-O-G, quercetin 4'-O-G), steroids, lipids and carbohydrates were more abundant in *C. nankingense* (Zou et al., 2022), which is usually eaten as a vegetable in Nanjing, China, and its phenylpropanoid, terpene, indole derivatives and alkaloid contents were relatively low. Among identified 477 metabolites, 72 showed significant differences between *C. mongolicum* and *C. rhombifolium* (Duan et al., 2022), mainly flavonoids, organic acids and nucleotides. The metabolomic techniques should be applied in more *Chrysanthemum* species to obtain a holistic and comparative view of their nutritional value.

Other constituents

Coumarin, umbelliferone, lignan, tonghaosu, pyrethrin, jasmolin, cinerin, etc., were also found in CM (Yang et al., 2019; Zhou et al., 2019b; Mao, 2020; Figure 8). Pyrethrins, a natural insecticide, are biosynthesized by the Anthemideae plants such as *Tanacetum cinerariifolium* (Matsuda, 2012) and CM (Mao, 2020). Their monoterpenoid acid and rethrolone-type oxylipin alcohol moieties are biosynthesized via the 2-C-methyl-D: -erythritol 4-phosphate (MEP) and oxylipin pathways, respectively (Lybrand et al., 2020). The organic acids, e.g., azelaic acid, 4-guanidinobutyric acid, are abundant in *C. indicum* (Zou et al., 2022). The anti-diabetic polyacetylene glycosides from flowers of edible *C. "Kamiohno"* are also reported (Kurimoto et al., 2021).

Quality control methods and standards

Five kinds of medicinal CM are listed in the Chinese Pharmacopoeia (2020), including Boju, Hangju, Chuju, Huaiju, Gongju. The diversity of species, origins, cultivation, processing and harvesting methods directly affects the medicinal components and effects. However, the traditional HPLC method of measuring the contents of chlorogenic acid, Lut-7-O-G, 3,5-DCQA is not enough to evaluate the quality of medicinal CM comprehensively. The concept of quality marker (Q-marker) provides a new idea for quality control of TCM. Q-marker is based on the perspective of biological activity, and various methods are utilized to find the chemical components that best reflect biological effects of TCM (Liu et al., 2018). Multiple methods, especially multi-ingredients quantification, chromatographic fingerprint and/or their combination (Nie et al., 2019; Peng et al., 2019), have been applied extensively in the quality control of CM. In recent years, quality verification methods based on the activity-integrated fingerprints have received wide attention. For instance, an antioxidant activity based method combining the online HPLC-DPPH, ESI-MS, and NIR spectra analysis identified 16 antioxidants of five CM cultivars (Zhang et al., 2022). Phenolic acids play a more important role in antioxidant activity, and chlorogenic acid, Lut-7-O-G, 3,5-DCQA, 4,5-DCQA were observed as the main contributors to the overall antioxidant capacity. The chromatographic fingerprint and *in vitro* antioxidant activity assay were combined to show that chlorogenic acid, 3,5-DCQA, 1, 4,5-O-DCQA and kaempferol-3-O-rutinoside could be Q-markers of Hangbaiju, Gongju, Huaiju, Taiju and Boju (Lu et al., 2022b). In the future, the strategy based on activity-integrated fingerprints has great potential in quality control of traditional medicine,

food and other fields, as it can fully reflect the pharmacological information of active ingredients.

Pharmacological properties

The plants of *Chrysanthemum* and related taxonomic groups are traditionally used as ethnomedicine (Figure 1; Jia and Zhang, 2016). *Chrysanthemum* plants have a wide range of pharmacological activities (Figure 9; Supplementary Table S1). In recent years, the main studied taxa include CM (Khan et al., 2020), *C. indicum* (Tian et al., 2020), *C. boreale* (Kim et al., 2022) and *C. zawadskii* (Kim et al., 2019), among others. The most reported activity of *Chrysanthemum* is its anti-inflammatory and immunomodulatory effects (Xue et al., 2021; Kim et al., 2022), which are consistent with traditional efficacy; it also shows great potential for improving chronic metabolic diseases, neurodegenerative diseases, etc. As the main active ingredients, flavonoids, phenolic acids and terpenoids may be the principal material basis for these effects. For example, total flavonoids of CM cultivar Bianliang ziyu prevented hepatotoxicity by inhibiting oxidative stress and apoptosis via the activation of Nrf2 signaling (Tian et al., 2019); caffeoylquinic acids, chlorogenic acid, gallocatechin, Lut-7-OG, Acn-7-O-rutinoside, and anthocyanins of *Chrysanthemum* showed strong antioxidant activities (Supplementary Table S1.). Acn-7-O-rutinoside, Lut-7-OG, and chlorogenic acid also showed the anti-inflammatory activity (Zhang et al., 2019). Flavonoids (including linarin, Dio-7-G, tilianin, etc.) and phenolic acids (including isochlorogenic acid C, isochlorogenic acid A, 1,3-DCQA, etc.) of *Chrysanthemum* improved the inflammatory bowel disease of zebrafish by regulating the expressions of IL-1β, IL-8 and MMP9 (Li et al., 2022). Eriodictyol-7-O-β-d-glucuronide and 5,7-dihydroxy-4-chromene of *C. zawadskii* var. *latilobum* had antiallergic effects in FcεRI-mediated human basophilic KU812F cells (Lee and Shim, 2020). The activities of *Chrysanthemum* flavonoids in metabolic regulation are also salient. For instance, Lut and luteoloside improved blood

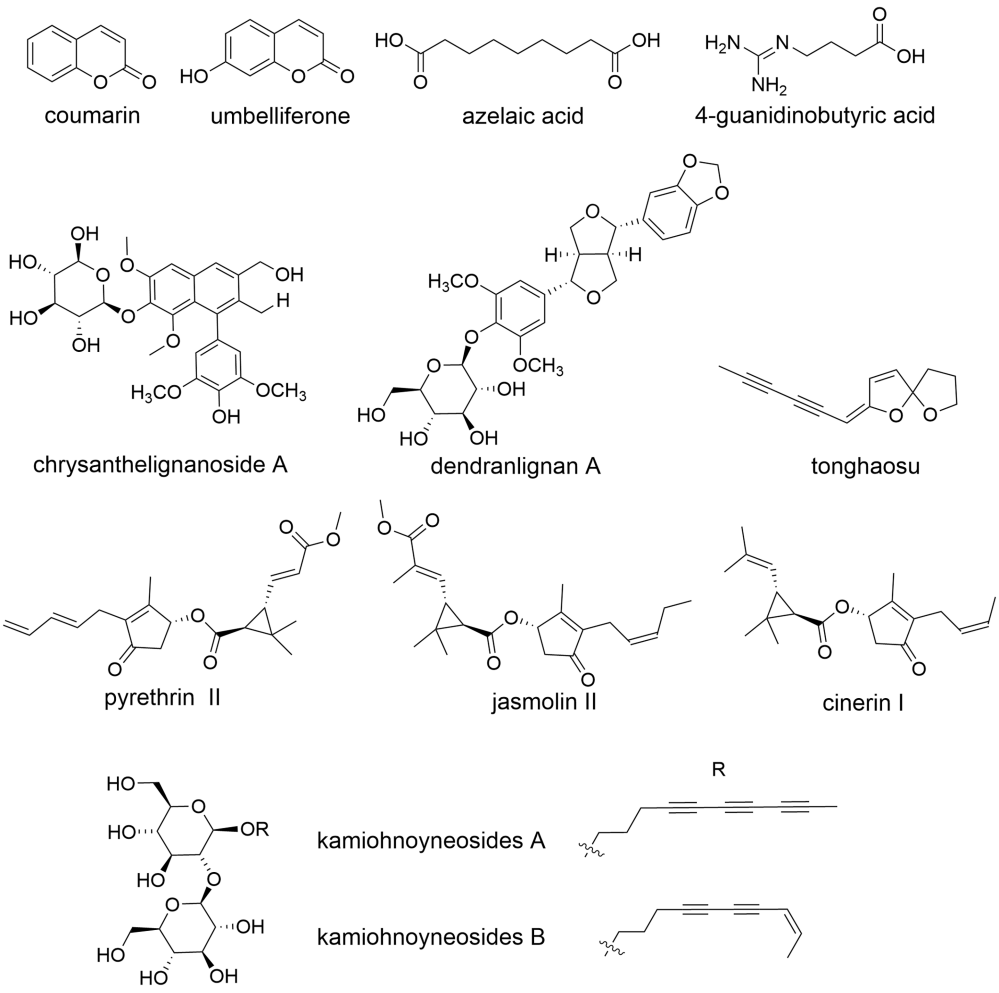


FIGURE 8
Examples of other phytometabolites of *Chrysanthemum*. Coumarin, umbelliferone, azelaic acid (*C. indicum* Zou et al., 2022), 4-guanidinobutyric acid, chrysanthelignanoside A (neuroprotection, CM), dendranlignan A (anti-inflammatory, CM), tonghaosu, pyrethrin II, jasmolin II, cinerin I, kamiohnoyneosides A and B (anti-diabetic, CM).

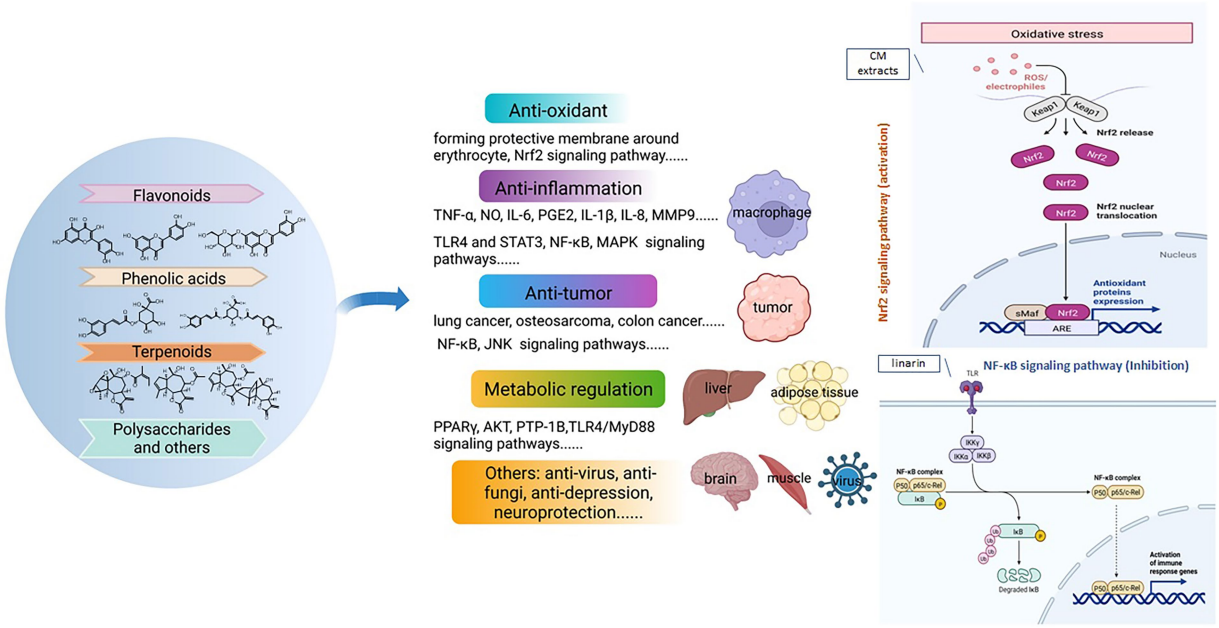


FIGURE 9
Illustration of pharmacological mechanisms exerted by *Chrysanthemum* extracts and compounds. Various bioactive compounds produced in *Chrysanthemum* plants are shown in left panel; their diverse pharmacological activities and possible mechanisms of action are exemplified in the central part; two representative signaling pathways regulated by *Chrysanthemum* components are shown on the right part. ARE: Antioxidant response element; NF-κB: Nuclear factor kappa-B; Nrf2: NF-E2-related factor 2; TLR4: Toll-like receptor 4.

lipids and hepatic steatosis in hyperlipidemia rats by regulating antioxidant levels and lipid metabolism (Sun et al., 2021); naringenin and naringenin-7-O-G inhibited the intracellular lipid accumulation by the activation of PPARγ and phosphorylation of the PI3K/Akt pathway (Nishina et al., 2019); Lut, Acn, and buddleoside controlled the postprandial glucose concentration by inhibiting the activity of α-amylase (Li et al., 2019). Other activities of *Chrysanthemum* flavonoids are also intriguing. Linarin and scutellarein displayed anticancer activity (Jung et al., 2019; Li et al., 2020); buddleoside reduced blood pressure in spontaneously hypertensive rats by inhibiting the vascular TLR4/MyD88 pathway and improving vascular endothelial function (Wang et al., 2021c); Acn-7-O-β-D-rutinoside prevented dexamethasone-evoked muscle atrophy via the Akt/mTOR pathway and decreasing the mitochondrial respiration (Lee et al., 2021).

When compared with flavonoids and phenolics, the antioxidant activity and metabolic regulation of *Chrysanthemum* terpenoids are less reported (Supplementary Table S1.), but their anti-inflammatory, anticancer and other properties are remarkable. For example, chrysanthemulide A (sesquiterpenoid) of *C. indicum* induced apoptosis of osteosarcoma by upregulating death receptor 5 via JNK-mediated autophagosome accumulation (Zhuo et al., 2019); it also showed the anti-inflammatory activity via suppressing the LPS-induced NF-κB pathway and down-regulating MAPK activation (Xue et al., 2018); cumambrin A (sesquiterpenoid) suppressed the osteoclast formation, bone resorption, and RANKL-induced signaling pathways in the treatment of osteoporosis (Zhou et al., 2019a). The recently discovered sesquiterpenes are subject to the in vitro activity screening (Jiang et al., 2021a), whose efficacies and corresponding mechanisms of action warrant further explorations.

The activities of *Chrysanthemum* polysaccharides have also attracted attention (Wang et al., 2021b), e.g., antioxidant, anti-inflammatory and anti-viral (Supplementary Table S1), and other compounds and extracts exhibited diversified bioactivities. The antimicrobial, anti-inflammatory, anti-hypertension and antioxidant effects (Supplementary Table S1) of CM have been utilized by ancient doctors and herbalists to help maintain body balance and relieve much of the sufferings. The other tantalizing effects of CM components, e.g., anti-aging and anti-diabetic, are being revealed by contemporary polypharmacology investigations, which facilitate drug repurposing and benefit more people. The phylogenetically related *Chrysanthemum* species are more likely to possess similar chemical profiles, despite the impact of environmental factors. The most typical examples are CM and wild relatives. The medicinal compounds and therapeutic efficacy of CM are closer to *C. indicum* than to *C. zawadskii*, and are more dissimilar to those of *C. boreale*. Significant or subtle differences in chemical composition between closely related species could be useful in tailoring personalized therapeutic protocol for each patient.

Conclusions and prospects

This article summarized the phylogeny, biodiversity, phytochemistry, and chemodiversity of *Chrysanthemum*, especially

the cultivated hybrid CM. As a unique Chinese culture, the chrysanthemum culture has witnessed the development of history and had a profound influence on human health. The ethnomedicinal experiences, phytochemistry and bioactivity studies are the prerequisite of product research and development; to date more than 120 CM compounds have been isolated and identified, and these monomeric compounds and crude extracts are screened for pharmacological activities *in vivo* and in vitro. The experimental studies validated the traditional medicinal uses of CM, but the responsible chemicals have not been thoroughly determined. Thus, there is a need of bioassay-guided identification of the bioactive components. What's more, the relationship between traditional uses and recent pharmacological findings is not always clear, and it is imperative to investigate the biochemical and physiological mechanisms of components of CM and evolutionarily related taxa, especially their cardiovascular protection, anti-tumor and antioxidant activities. Efforts should also be made to determine the modes of action, bioavailability, pharmacokinetics and physiological pathways of specific functional compounds in *Chrysanthemum* and related taxonomic groups. Moreover, clinical studies, e.g., randomized controlled trial, should be encouraged to identify any side effects and possible interactions between *Chrysanthemum* herbal medicine and other natural medicines/synthetic drugs. Further safety verification and clinical trials should be carried out to expand the application scope of CM and better integrate it into medicinal practice. With multiple chemical ingredients, CM may exert its beneficial effects by gently interacting with different cell signaling pathways and networks, which achieve the same therapeutic efficacy as that of mono-ingredient agents, whereas CM doses are much lower than those of single compounds. In order to better mine the medicinal potential and edible value of different *Chrysanthemum* cultivars and Daodi medicinal materials, the intricate link between phylogenetic relationship, chemical profile, ethnomedicinal knowledge and pharmacological activities should be scrutinized within the pharmacophylogenetic framework (Hao and Xiao, 2017, 2020).

The emerging pharmacophylogenomics (Hao and Xiao, 2017, 2020) calls for much more genomic data of *Chrysanthemum* and related taxonomic groups. The rapid development of reduced-representation genome sequencing enables the cost-effective sequencing of Asteraceae plant (Mérot, 2020). It is of low cost and strong sequencing performance. In the future, it can be used in *Chrysanthemum* phylogeny, plant identification and genome assisted breeding. The cp genome with maternal genetic characteristics has the advantages of multi-copy and conservative structure. The combination of cp genome and high-throughput sequencing technology has become an effective means of plant genetic resources research, which facilitates the study of *Chrysanthemum* classification and genetic resources (Tyagi et al., 2020; Masuda et al., 2022). The accumulation of whole genome sequencing data of *Chrysanthemum* (Song et al., 2018; Nakano et al., 2021; van Lieshout et al., 2022) and relevant taxonomic groups will enable the comprehensive phylogenomic analyses to reveal the hybridization/polyploidization events leading to the speciation and formation of CM. The assembly and annotation of more high-quality *Chrysanthemum* genomes will help to elucidate the genus evolution and its contributions to gene abundance/function.

The genome editing enables the induction of mutations in a targeted genomic region and has recently played a substantial role in Asteraceae functional genomics and biotechnology (Bernard et al., 2019; Park et al., 2019). As precise mutations can be generated in the targeted sequence, it is considered to be more effective than conventional mutation breeding. This breakthrough technology has been adopted for various Asteraceae crops such as chicory and lettuce (Bernard et al., 2019; Park et al., 2019), and could be used in *Chrysanthemum* for desired phenotypes. Unlike transgenic technology, genome editing does not require the transgene to be integrated with the genome, and the integrated transgene could be segregated in the progeny.

At present, there are very few reports on the use of metabolomics in *Chrysanthemum* taxa. The metabolomics technology can be used to comprehensively characterize the chemical components of *Chrysanthemum* species and varieties, obtain quantitative information of each phytometabolite, and quickly screen the characteristic chemical, so as to provide effective data for the quality evaluation and control of *Chrysanthemum* samples, and contribute new ideas and references for the elucidation of medicinal material basis of *Chrysanthemum*. The evaluation and control provide effective data, inspire new ideas and contribute references for the elucidation of the medicinal material basis of chrysanthemum. The metabolic research is conducive to the realization of the integrity and traceability of chrysanthemum quality control. The metabolic makeup of fresh flowers and processed products may vary greatly; how to achieve the transferability of chrysanthemum quality in multiple links such as harvesting, crude processing, fine processing, and production of patent medicines still needs to be studied from multiple hierarchies. However, the relative quantification using the peak area as an index cannot reflect the exact content of compounds in chrysanthemum, therefore it is necessary to establish the quantitative analysis method for the differential components of different species/varieties of *Chrysanthemum*. The compounds detected by UPLC-Q-TOF-MS in plant metabolomics are more comprehensive (Hao et al., 2021). To find the quality control indicators of chrysanthemum from numerous compounds, it is obligatory to combine content determination and pharmacokinetics/pharmacodynamics for comprehensive evaluation, which warrant deeper studies in the pharmacologically active constituents of chrysanthemum.

The current highly intensive industrial systems of agricultural and horticultural production are counter-sustainable, as the energy consumption is particularly intensive for cultivation, and the excessive use of nitrogen fertilizer leads to the worrying emission of greenhouse gas, N_2O (Hao et al., 2021; Wang et al., 2022). The genetic modification (GM) is transforming the prospects of sustainable crop protection, the specialized metabolite (e.g., pyrethrins) biosynthesis genes can be integrated to the target *Chrysanthemum* plants to enhance the stress response and disease

resistance, so as to realize the full potential of GM and offer greener crop protection.

This article applies the systematic and holistic characteristics of plant pharmacophylogeny to the regular arrangement of biodiversity and chemodiversity explorations on *Chrysanthemum*. On the basis of scientific and technological innovation, and with the integration of holistic view of traditional medicine, we have deepened our understanding of the following points: The phylogenetically close taxa are more likely to have similar metabolic profile, and the similar metabolic makeup could result in analogous pharmacokinetic behavior and clinical efficacy (Hao and Xiao, 2020). The studies of pharmacophylogeny are conducive to the sustainable conservation and rational utilization of *Chrysanthemum* resources, as well as the inheritance and innovation of traditional medicine. The medicinal properties of plants can be predicted by virtue of phylogenetic methods (Hao et al., 2022a,b), which has been utilized to explore the regularity of bioactivities of Asteraceae and Ranunculales plants against diseases of multiple human organs. The mapping of different types of *Chrysanthemum* (and related taxonomic groups) compounds onto the species phylogenetic tree could also be beneficial to mining novel sources of economically important chemicals. Disentangling recent speciation events and distribution of compounds/therapeutic effects usually requires a reliable phylogenetic framework, therefore research in phylogeny/evolution of *Chrysanthemum* and related taxonomic groups cannot be overemphasized. Besides the intensive studies of CM, the evaluation of other *Chrysanthemum* species should also keep pace with the times and develop scientifically. The ability to correlate the chrysanthemum metabolome with the medicinal efficacy in the context of pharmacophylogeny will help to develop chrysanthemum cultivars with better curative effect and more commercial value in the near future.

Author contributions

D-CH and PX: conceptualization and writing—review and editing. D-CH, YS, and YZ: methodology and analysis. PX and LX: resources and supervision. D-CH, YS, and PW: data curation and visualization. D-CH and YS: writing—original draft preparation. All authors contributed to the article and approved the submitted version.

Funding

This work is supported by the Scientific Research Funds Project of Liaoning Education Department (JDL2019012), China Scholarship Council (202108210156), CAMS Innovation Fund for Medical Sciences (CIFMS 2021-I2M-1-032), and Hainan Academician Innovation Platform Scientific Research Project and National Science & Technology Fundamental Resources Investigation Program of China (2018FY100700).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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.2022.973197/full#supplementary-material

References

Beninger, C. W., Abou-Zaid, M. M., Kistner, A. L., Hallett, R. H., Iqbal, M. J., Grodzinski, B., et al. (2004). A flavanone and two phenolic acids from *Chrysanthemum morifolium* with phytotoxic and insect growth regulating activity. *J. Chem. Ecol.* 30, 589–606. doi: 10.1023/B:JOEC.0000018631.67394.e5

Bernard, G., Gagneul, D., Alves Dos Santos, H., Etienne, A., Hilbert, J. L., and Rambaud, C. (2019). Efficient genome editing using CRISPR/Cas9 technology in chicory. *Int. J. Mol. Sci.* 20:1155. doi: 10.3390/ijms20051155

Cai, Y., Gao, Y., Zhang, Z., Liu, H., Wang, Y., Ma, Y., et al. (2022). Development and application of a cultivar-specific sequence-characterized amplified region (SCAR) marker for the detection of *Chrysanthemum morifolium* Ramat. *'Daboju'*. *Plants.* 11:604. doi: 10.3390/plants11050604

Chang, X. W., Wei, D. D., Chen, D. J., Yan, H., Sun, X. D., Zhu, W. B., et al. (2019). Historical origin and development of medicinal and tea *Chrysanthemum morifolium* resource. *Modern Chin. Med.* 21, 116–123. doi: 10.13313/j.issn.1673-4890.20181224005

Chang, X., Zhang, Z., Yan, H., Su, S., Wei, D., Guo, S., et al. (2021). Discovery of quality markers of nucleobases, nucleosides, nucleotides and amino acids for Chrysanthemi Flos from different geographical origins using UPLC-MS/MS combined with multivariate statistical analysis. *Front. Chem.* 9:689254. doi: 10.3389/fchem.2021.689254

Chen, J. Y. (2005). Contributions of Chinese Chrysanthemum to the world in the past and future. *J. Chin. Landscape Architect.* 9, 73–75.

Chen, G., Kim, H. K., Klinkhamer, P. G., and Escobar-Bravo, R. (2019). Site-dependent induction of jasmonic acid-associated chemical defenses against western flower thrips in Chrysanthemum. *Planta* 251:8. doi: 10.1007/s00425-019-03292-2

Chen, T., Li, L. P., Lu, X. Y., Jiang, H. D., and Zeng, S. (2007). Absorption and excretion of luteolin and apigenin in rats after oral administration of *Chrysanthemum morifolium* extract. *J. Agri. Food Chem.* 55, 273–277. doi: 10.1021/jf062088r

Chen, S., Liu, J., Dong, G., Zhang, X., Liu, Y., Sun, W., et al. (2021). Flavonoids and caffeoylquinic acids in *Chrysanthemum morifolium* Ramat flowers: A potentially rich source of bioactive compounds. *Food Chem.* 344:128733. doi: 10.1016/j.foodchem.2020.128733

Chen, X. J., Sun, M., Liang, J. G., Xue, H., and Zhang, Q. X. (2013). Genetic diversity of species of Chrysanthemum and related genera and groundcover cultivars assessed by amplified fragment length polymorphic markers. *HortScience* 48, 539–546. doi: 10.21273/HORTSCI.48.5.539

Chinese Pharmacopoeia Commission. (2015). *Pharmacopoeia of the People's Republic of China*. Beijing: China Medical Science Press, 1, 310–311.

Chinese Pharmacopoeia Commission. (2020). *Pharmacopoeia of the People's Republic of China*. Beijing: China Medical Science Press, 1, 323–324.

Chong, X., Zhang, F., Wu, Y., Yang, X., Zhao, N., Wang, H., et al. (2016). A SNP-enabled assessment of genetic diversity, evolutionary relationships and the identification of candidate genes in Chrysanthemum. *Genome Biol. Evol.* 8, 3661–3671. doi: 10.1093/gbe/evw270

Clifford, M. N., Wu, W., Kirkpatrick, J., and Kuhnert, N. (2007). Profiling the chlorogenic acids and other caffeic acid derivatives of herbal Chrysanthemum by LC-MSn. *J. Agri. Food Chem.* 55, 929–936. doi: 10.1021/jf062314x

Criado Ruiz, D., Villa Machío, I., Herrero Nieto, A., and Nieto Feliner, G. (2021). Hybridization and cryptic speciation in the Iberian endemic plant genus Phalacrocarpum (Asteraceae-Anthemideae). *Mol. Phylogenet. Evol.* 156:107024. doi: 10.1016/j.ympev.2020.107024

Cui, N. X. (2004). Meiosis behavior and genetic relationship of several Chrysanthemum plants and their hybrids. MSc thesis. [Nanjing(Jiangsu)]: Nanjing Agricultural University, 1–61.

Dai, Z., Li, S. X., Chen, S. Q., and Zhu, Y. (2017). Quality evaluation of Huaiju germplasm resources. *Chin. J. Expt. Trad. Med. Formulae.* 23, 48–54. doi: 10.13422/j.cnki.syfjx.2017060048

Ding, L., Chen, F. D., Teng, N. J., Fang, W. M., and Zhao, J. Y. (2008). Isozyme analysis of wild and different purpose Chrysanthemum. *J. Nanjing Agri. Univ.* 2008, 37–42.

Duan, X. X., Zhang, W. J., Li, J. J., Xu, H., Hu, J., Zhao, L., et al. (2022). Comparative metabolomics analysis revealed biomarkers and distinct flavonoid biosynthesis regulation in Chrysanthemum mongolicum and C. rhombifolium. *Phytochem. Anal.* 33, 373–385. doi: 10.1002/pca.3095

Fan, M., Gao, Y. K., Gao, Y. H., Wu, Z., Liu, H., and Zhang, Q. X. (2019). Characterization and development of EST-SSR markers from transcriptome sequences of chrysanthemum (Chrysanthemum× morifolium Ramat.). *Hort Science* 54, 772–778. doi: 10.21273/HORTSCI13694-18

Farag, N. F., Farag, M. A., Abdelrahman, E. H., Azzam, S. M., and El-Kashoury, E. S. A. (2015). Metabolites profiling of Chrysanthemum pacificum Nakai parts using UPLC-PDA-MS coupled to chemometrics. *Nat. Prod. Res.* 29, 1342–1349. doi: 10.1080/14786419.2015.1025396

Feng, S. G., He, R. F., Lu, J., Jiang, M., Shen, X., Jiang, Y., et al. (2016). Development of SSR markers and assessment of genetic diversity in medicinal *Chrysanthemum morifolium* cultivars. *Front. Genet.* 7:113. doi: 10.3389/fgene.2016.00113

Flora Reipublicae Popularis Sinicae. (1983). Beijing: Science Press, 76, 35–39.

Hao, D. C., Chen, S. L., Osbourn, A., Kontogianni, V. G., Liu, L. W., and Jordán, M. J. (2015). Temporal transcriptome changes induced by methyl jasmonate in *Salvia sclarea*. *Gene* 558, 41–53. doi: 10.1016/j.gene.2014.12.043

Hao, D. C., Li, P., Xiao, P. G., and He, C. N. (2021). Dissection of full-length transcriptome and metabolome of Dichocarpum (Ranunculaceae): implications in evolution of specialized metabolism of Ranunculales medicinal plants. *PeerJ.* 9:e12428. doi: 10.7717/peerj.12428

Hao, D. C., and Liu, C. X. (2021). Back to beginning: Searching for Rosetta Stone of enhancing herbal medicine quality. *Chin. Herb. Med.* 13, 299–300. doi: 10.1016/j.chmed.2022.03.005

Hao, D. C., Lyu, H. Y., Wang, F., and Xiao, P. G. (2022a). Evaluating potentials of species rich taxonomic groups in cosmetics and dermatology: Clustering and dispersion of skin efficacy of Asteraceae and Ranunculales plants on the species phylogenetic tree. *Curr. Pharm. Biotech.* 23:926. doi: 10.2174/1389201023666220324123926

Hao, D. C., and Xiao, P. G. (2017). *An Introduction of Plant Pharmacophylogeny*. Beijing: Chemical Industry Press.

Hao, D. C., and Xiao, P. G. (2020). Pharmaceutical resource discovery from traditional medicinal plants: Pharmacophylogeny and pharmacophylogenomics. *Chin. Herb. Med.* 12, 104–117. doi: 10.1016/j.chmed.2020.03.002

Hao, D. C., Zhang, Y. L., He, C. N., and Xiao, P. G. (2022b). Distribution of therapeutic efficacy of Ranunculales plants used by ethnic minorities on the phylogenetic tree of Chinese species. *Evid. Based Complement. Altern. Med.* 2022, 9027727. doi: 10.1155/2022/9027727

Hong, G. (2014). Preliminary study on intergeneric hybridization of Chrysanthemum and its related genera (VIII). MSc thesis. [Beijing]: Beijing Forestry University, 1–56.

Hong, G., Wu, X. B., Liu, Y. C., Xie, F., Liu, Z. H., Liu, W. C., et al. (2015). Intergeneric hybridization between Hippolytia kaschgarica (Krascheninnikov) Poljakov and *Nipponanthemum nipponicum* (Franch. ex Maxim.) Kitam. *Genet. Resour. Crop. Evol.* 62, 255–263. doi: 10.1007/s10722-014-0150-1

Huang, Y. (2010). Preliminary study on cross breeding of medicinal Chrysanthemum. MSc thesis. [Nanjing(Jiangsu)]: Nanjing Agricultural University, 1–87.

Huang, Z., Liu, Z. Y., Wang, S. L., Xue, Y. Q., Xue, J. Q., and Zhang, X. X. (2020). Resources investigation of Chinese medicinal *Chrysanthemum morifolium* varieties and its industry status analysis. *J. Chin. Medi. Materia.* 43, 1325–1329. doi: 10.13863/j.issn1001-4454.2020.06.007

Jia, M.R., and Zhang, Y. (2016). *Dictionary of Chinese Ethnic Medicine*. Beijing: China Medical Science Press.

Jiang, S., Wang, M., Jiang, Z., Zafar, S., Xie, Q., Yang, Y., et al. (2021a). Chemistry and pharmacological activity of sesquiterpenoids from the Chrysanthemum genus. *Molecules* 26:3038. doi: 10.3390/molecules26103038

Jiang, H., Xia, Q., Xu, W., and Zheng, M. (2004). *Chrysanthemum morifolium* attenuated the reduction of contraction of isolated rat heart and cardiomyocytes induced by ischemia/reperfusion. *Pharmazie* 59, 565–567.

Jiang, B. P., Xu, L. J., Wang, Q. L., and Xiao, P. G. (2013). *Chrysanthemum morifolium*--- Ethnopharmacological use, phytochemistry and pharmacological activity. *Modern Chin. Med.* 15, 523–530. doi: 10.13313/j.issn.1673-4890.2013.06.016

Jiang, Y. F., Zhang, W. B., Chen, X. L., Wang, W., Köllner, T. G., Chen, S., et al. (2021b). Diversity and biosynthesis of volatile terpenoid secondary metabolites in the Chrysanthemum genus. *Criti. Rev. Plant Sci.* 40, 422–445. doi: 10.1080/07352689.2021.1969504

Jo, Y. D., Ryu, J., Kim, Y. S., Kang, K. Y., Hong, M. J., Choi, H. I., et al. (2020). Dramatic increase in content of diverse flavonoids accompanied with down-regulation of F-box genes in a Chrysanthemum (Chrysanthemum × morifolium (Ramat.) Hemsl.) mutant cultivar producing dark-purple ray florets. *Genes.* 11:865. doi: 10.3390/genes11080865

Jung, C. H., Han, A. R., Chung, H. J., Ha, I. H., and Um, H. D. (2019). Linarin inhibits radiation-induced cancer invasion by downregulating MMP-9 expression via the suppression of NF-κB activation in human non-small-cell lung cancer A549. *Nat. Prod. Res.* 33, 3582–3586. doi: 10.1080/14786419.2018.1484460

Kang, S. H., Nie, J., Chen, K. L., and Deng, H. Y. (2022). Comparison of morphological characteristics and quality evaluation of *Chrysanthemum morifolium*. *J. Chin. Medi. Mate.* 2022, 49–57. doi: 10.13863/j.issn1001-4454.2022.01.009

Khan, I. A., Xu, W., Wang, D., Yun, A., Khan, A., Zongshuai, Z., et al. (2020). Antioxidant potential of *chrysanthemum morifolium* flower extract on lipid and protein oxidation in goat meat patties during refrigerated storage. *J. Food Sci.* 85, 618–627. doi: 10.1111/1750-3841.15036

Kim, K. Y., Oh, T. W., Yang, H. J., Kim, Y. W., Ma, J. Y., and Park, K. I. (2019). Ethanol extract of Chrysanthemum zawadskii Herbich induces autophagy and apoptosis in mouse colon cancer cells through the regulation of reactive oxygen species. *BMC Complement. Altern. Med.* 19:274. doi: 10.1186/s12906-019-2688-0

Kim, D. Y., Won, K. J., Hwang, D. I., Kim, N. Y., Kim, B., and Lee, H. M. (2022). 1-Iodohexadecane alleviates 2,4-dinitrochlorobenzene-induced atopic dermatitis in mice: Possible involvements of the skin barrier and mast cell SNARE proteins. *Molecules* 27:1560. doi: 10.3390/molecules27051560

Kishimoto, S., and Ohmiya, A. (2006). Regulation of carotenoid biosynthesis in petals and leaves of chrysanthemum (*Chrysanthemum morifolium*). *Physiologia Plant.* 128, 436–447. doi: 10.1111/j.1399-3054.2006.00761.x

Kurimoto, S. I., Fujita, H., Kawaguchi, S., Sasaki, Y. F., Nakamura, T., and Kubota, T. (2021). Kamiohnoyneosides A and B, two new polyacetylene glycosides from flowers of edible Chrysanthemum "Kamiohno". *J. Nat. Med.* 75, 167–172. doi: 10.1007/s11418-020-01443-4

Lai, J. P., Lim, Y. H., Su, J., Shen, H. M., and Ong, C. N. (2007). Identification and characterization of major flavonoids and caffeoylquinic acids in three Compositae plants by LC/DAD-APCI/MS. *J. Chromatography B.* 848, 215–225. doi: 10.1016/j.jchromb.2006.10.028

Lee, H., Kim, Y. I., Nirmala, F. S., Jeong, H. Y., Seo, H. D., Ha, T. Y., et al. (2021). Chrysanthemum zawadskil Herbich attenuates dexamethasone-induced muscle atrophy through the regulation of proteostasis and mitochondrial function. *Biomed. Pharmacother.* 136:111226. doi: 10.1016/j.biopha.2021.111226

Lee, M., and Shim, S. Y. (2020). Inhibitory effects of eriodictyol-7-O-β-d-glucuronide and 5,7-dihydroxy-4-chromene isolated from Chrysanthemum zawadskii var. latilobum in FcεRI-mediated human basophilic KU812F cell activation. *Molecules* 25, 994. doi: 10.3390/molecules25040994

Li, H. (1993). *Chinese Chrysanthemum*. Nanjing: Jiangsu Scientific and Technological Publishing Co.

Li, T. (2014). Construction of CDDP fingerprints and genetic diversity analysis of Chrysanthemum. MSc thesis. [Taian(Shandong)]: Shandong Agricultural University, 1–64.

Li, P., Huang, Z., She, Y., Qin, S., Gao, W., Cao, Y., et al. (2019). An assessment of the interaction for three *Chrysanthemum indicum* flavonoids and α-amylase by surface plasmon resonance. *Food Sci. Nutri.* 8, 620–628. doi: 10.1002/fsn3.1349

Li, Y., Liu, X. J., Su, S. L., Yan, H., Guo, S., Qian, D. W., et al. (2022). Evaluation of anti-inflammatory and antioxidant effects of Chrysanthemum stem and leaf extract on zebrafish inflammatory bowel disease model. *Molecules* 27, 2114. doi: 10.3390/molecules27072114

Li, B., Ma, C., Zhao, X., Hu, Z., Du, T., Xu, X., et al. (2018). YaTCM: Yet another traditional Chinese medicine database for drug discovery. *Comput. Struct. Biotechnol. J.* 16, 600–610. doi: 10.1016/j.csbj.2018.11.002

Li, J., Wan, Q., Abbott, R. J., and Rao, G. Y. (2013). Geographical distribution of cytotypes in the *Chrysanthemum indicum* complex as evidenced by ploidy level and genome-size variation. *J. Syst. Evol.* 51, 196–204. doi: 10.1111/j.1759-6831.2012.00241.x

Li, J., Wan, Q., Guo, Y. P., Abbott, R. J., and Rao, G. Y. (2014). Should I stay or should I go: biogeographic and evolutionary history of a polyploid complex (*Chrysanthemum indicum* complex) in response to Pleistocene climate change in China. *New Phytol.* 201, 1031–1044. doi: 10.1111/nph.12585

Li, Y., Wang, J., Zhong, S., Li, J., and Du, W. (2020). Scutellarein inhibits the development of colon cancer via CDC4-mediated RAGE ubiquitination. *Int. J. Mol. Med.* 45, 1059–1072. doi: 10.3892/ijmm.2020.4496

Li, D. L., Zhu, H. W., Ren, Q. J., and Xu, Z. L. (2010). Comparative analysis on the configuration of vegetative organs of medicinal Chrysanthemum from different original locations and species. *Zhong Yao Cai* 33, 1845–1849. doi: 10.13863/j.issn1001-4454.2010.12.005

Liu, R. (2010). A study on the relationship between some wild species of Chrysanthemum and cultivated chrysanthemum. MSc thesis. [Baoding(Hebei)]: Hebei Agricultural University, 1–58.

Liu, Q. J. (2013). Accumulation and antioxidant activity of phenolic substances in Opisthopappus taihangensis and related *Chrysanthemum indicum*. PhD dissertation. [Zhengzhou(Henan)]: Henan Agricultural University, 1–92.

Liu, Y. (2020). Comparative study on agronomic characters, chemical components and pharmacological effects of different chrysanthemum germplasm resources. MSc thesis. [Wuhan(Hubei)]: Hubei University of Chinese Medicine, 1–124.

Liu, H., Chen, X., Chen, H., Lu, J., Chen, D., Luo, C., et al. (2021a). Transcriptome and metabolome analyses of the flowers and leaves of Chrysanthemum dichrum. *Front. Genet.* 12:716163. doi: 10.3389/fgene.2021.716163

Liu, Y., Dai, M., Bao, W. Z., Huang, B. S., Guo, L. P., and Liu, D. H. (2021c). Characteristics of mineral elements in chrysanthemums from different origins in Macheng and their correlation with soil nutrients and effective components. *China J. Chin. Mate. Med.* 46, 281–289. doi: 10.19540/j.cnki.cjcmm.20200717.101

Liu, Y., Gong, W. L., Bao, W. Z., Guo, L. P., Xu, Y., Liu, Y. M., et al. (2019). Establishment of *Chrysanthemum morifolium* HPLC fingerprint of Hubei and comparison of active components among varieties. *China J. Chin. Mate. Med.* 44, 3711–3717. doi: 10.19540/j.cnki.cjcmm.20190701.106

Liu, C. X., Guo, D. A., and Liu, L. (2018). Quality transitivity and traceability system of herbal medicine products based on quality markers. *Phytomedicine* 44, 247–257. doi: 10.1016/j.phymed.2018.03.006

Liu, H. B., Ma, P., Xu, L. J., and Xiao, P. G. (2021b). Informatics and big data: A new stage of pharmacophylogeny in medicinal plants. *Mod. Chin. Med.* 23, 1506–1511. doi: 10.13313/j.issn.1673-4890.20210901002

Liu, P. L., Wan, Q., Guo, Y. P., Yang, J., and Rao, G. Y. (2012). Phylogeny of the genus Chrysanthemum L.: evidence from single-copy nuclear gene and chloroplast DNA sequences. *PLoS One* 7:e48970. doi: 10.1371/journal.pone.0048970

Lu, W. X., Hu, X. Y., Wang, Z. Z., and Rao, G. Y. (2022a). Hyb-Seq provides new insights into the phylogeny and evolution of the Chrysanthemum zawadskii species complex in China. *Cladistics*. doi: 10.1111/cla.12514

Lu, Y. F., Li, D. X., Zhang, R., Zhao, L. L., Qiu, Z., Du, Y., et al. (2022b). Chemical antioxidant quality markers of *Chrysanthemum morifolium* using a spectrum-effect approach. *Front. Pharmacol.* 13:809482. doi: 10.3389/fphar.2022.809482

Luo, C., Chen, D. L., Cheng, X., Zhao, H. E., and Huang, C. L. (2017). Genome size estimations in Chrysanthemum and correlations with molecular phylogenies. *Genet. Resour. Crop. Evol.* 64, 1451–1463. doi: 10.1007/s10722-016-0448-2

Luo, X. Y., Wang, C., Dai, S. L., Li, B. Q., Liu, Q. Q., Zhu, J., et al. (2013). Genetic diversity of large-flowered Chrysanthemum cultivars revealed by ISSR analysis. *Sci. Agri. Sin.* 46, 2394–2402.

Lybrand, D. B., Xu, H., Last, R. L., and Pichersky, E. (2020). How plants synthesize pyrethrins: Safe and biodegradable insecticides. *Trends Plant Sci.* 25, 1240–1251. doi: 10.1016/j.tplants.2020.06.012

Lyu, L., Qin, M. J., He, D. X., and Gu, Y. H. (2008). Analyses of ISSR molecular marker and genetic relationship of different provenances of *Dendranthema morifolium*, D. indicum and D. nankingese. *J. Plant Resour. Environ.* 17, 7–12.

Ma, Y. P., Chen, M. M., Wei, J. X., Zhao, L., Liu, P. L., Dai, S. L., et al. (2016). Origin of Chrysanthemum cultivars—Evidence from nuclear low-copy LFY gene sequences. *Biochem. Syst. Ecol.* 65, 129–136. doi: 10.1016/j.bse.2016.02.010

Mandel, J. R., Dikow, R. B., Siniscalchi, C. M., Thapa, R., Watson, L. E., and Funk, V. A. (2019). A fully resolved backbone phylogeny reveals numerous dispersals and explosive diversifications throughout the history of Asteraceae. *Proc. Natl. Acad. Sci. U. S. A.* 116, 14083–14088. doi: 10.1073/pnas.1903871116

Mao, C. Y. (2020). Study on quality evaluation standard of the flower of *Chrysanthemum morifolium* Ramat based on the correlation of ingredients and efficacy. [PhD dissertation. [Beijing]: China Academy of Chinese Medical Sciences, 1–246.

Masuda, Y., Nakano, M., and Kusaba, M. (2022). The complete sequence of the chloroplast genome of Chrysanthemum rupestre, a diploid disciform capitula species of Chrysanthemum. *Mitochondrial DNA B Resour.* 7, 603–605. doi: 10.1080/23802359.2022.2057252

Matsuda, K. (2012). Pyrethrin biosynthesis and its regulation in Chrysanthemum cinerariaefolium. *Top. Curr. Chem.* 314, 73–81. doi: 10.1007/128_2011_271

Mekapogu, M., Vasamsetti, B. M. K., Kwon, O. K., Ahn, M. S., Lim, S. H., and Jung, J. A. (2020). Anthocyanins in floral colors: Biosynthesis and regulation in Chrysanthemum flowers. *Int. J. Mol. Sci.* 21, 6537. doi: 10.3390/ijms21186537

Mérot, C. (2020). Making the most of population genomic data to understand the importance of chromosomal inversions for adaptation and speciation. *Mol. Ecol.* 29, 2513–2516. doi: 10.1111/mec.15500

Nakano, M., Hirakawa, H., Fukai, E., Toyoda, A., Kajitani, R., Minakuchi, Y., et al. (2021). A chromosome-level genome sequence of Chrysanthemum seticuspe, a model species for hexaploid cultivated chrysanthemum. *Commun. Biol.* 4, 1167. doi: 10.1038/s42003-021-02704-y

Nie, J., Xiao, L., Zheng, L., Du, Z., Liu, D., Zhou, J., et al. (2019). An integration of UPLC-DAD/ESI-Q-TOF MS, GC-MS, and PCA analysis for quality evaluation and identification of cultivars of Chrysanthemi Flos (Juhua). *Phytomedicine* 59:152803. doi: 10.1016/j.phymed.2018.12.026

Nishina, A., Sato, D., Yamamoto, J., Kobayashi-Hattori, K., Hirai, Y., and Kimura, H. (2019). Antidiabetic-like effects of naringenin-7-O-glucoside from edible Chrysanthemum 'Kotobuki' and naringenin by activation of the PI3K/Akt pathway and PPARγ. *Chem. Biodivers.* 16:e1800434. doi: 10.1002/cbdv.201800434

Oberprieler, C., Himmelreich, S., and Vogt, R. (2007). A new subtribal classification of the tribe Anthemideae (Compositae). *Willdenowia* 37, 89–114. doi: 10.3372/wi.37.37104

Park, C. H., Chae, S. C., Park, S. Y., Kim, J. K., Kim, Y. J., Chung, S. O., et al. (2015). Anthocyanin and carotenoid contents in different cultivars of Chrysanthemum (*Dendranthema grandiflorum* Ramat.) flower. *Molecules* 20, 11090–11102. doi: 10.3390/molecules200611090

Park, J., Choi, S., Park, S., Yoon, J., Park, A. Y., and Choe, S. (2019). DNA-free genome editing via ribonucleoprotein (RNP) delivery of CRISPR/Cas in lettuce. *Methods Mol. Biol.* 1917, 337–354. doi: 10.1007/978-1-4939-8991-1_25

Peng, A., Lin, L., Zhao, M., and Sun, B. (2019). Classification of edible chrysanthemums based on phenolic profiles and mechanisms underlying the protective effects of characteristic phenolics on oxidatively damaged erythrocyte. *Food Res. Int.* 123, 64–74. doi: 10.1016/j.foodres.2019.04.046

Peng, Y. R., Shi, L., and Luo, Y. H. (2006). Protective effect of the total flavones from Chrysanthemum on isoprenaline-induced myocardial ischemiain rats. *Lishizhen Med. Materia Med. Res.* 2006, 1131–1132.

Poe, S., and Swofford, D. L. (1999). Taxon sampling revisited. *Nature* 398, 299–300. doi: 10.1038/18592

Rajic, A., Akihisa, T., Ukiya, M., Yasukawa, K., Sandeman, R. M., Chandler, D. S., et al. (2001). Inhibition of trypsin and chymotrypsin by anti-inflammatory triterpenoids from Compositae flowers. *Planta Med* 67, 599–604. doi: 10.1055/s-2001-17350

Ryu, J., Nam, B., Kim, B. R., Kim, S. H., Jo, Y. D., Ahn, J. W., et al. (2019). Comparative analysis of phytochemical composition of gamma-irradiated mutant cultivars of *Chrysanthemum morifolium*. *Molecules* 24:3003. doi: 10.3390/molecules24163003

Sawada, Y., Sato, M., Okamoto, M., Masuda, J., Yamaki, S., Tamari, M., et al. (2019). Metabolome-based discrimination of chrysanthemum cultivars for the efficient generation of flower color variations in mutation breeding. *Metabolomics* 15, 118. doi: 10.1007/s11306-019-1573-7

Shao, B. J. (2018). Preliminary study on germplasm resources innovation of forage species within Chrysanthemum in broad sense. MSc thesis. [Beijing]: Beijing Forestry University, 1–114.

Shao, Q. S., Guo, Q. S., Li, Y. C., and Mao, P. F. (2011). Quantitative analysis of morphological variation of medicinal Chrysanthemum germplasm resources. *China J. Chin. Mate. Med.* 36, 1261–1265.

Shao, Y. H., Sun, Y. D., Li, D., and Chen, Y. P. (2020). *Chrysanthemum indicum* L.: A comprehensive review of its botany, phytochemistry and pharmacology. *Am. J. Chin. Med.* 48, 871–897. doi: 10.1142/S0192415X20500421

Shen, C. Z., Zhang, C. J., Chen, J., and Guo, Y. P. (2021). Clarifying recent adaptive diversification of the Chrysanthemum-group on the basis of an updated multilocus phylogeny of subtribe Artemisiinae (Asteraceae: Anthemideae). *Front. Plant Sci.* 12:648026. doi: 10.3389/fpls.2021.648026

Song, C., Liu, Y. F., Song, A., Dong, G., Zhao, H. B., Sun, W., et al. (2018). The Chrysanthemum nankingense genome provides insights into the evolution and diversification of Chrysanthemum flowers and medicinal traits. *Mol. Plant* 11, 1482–1491. doi: 10.1016/j.molp.2018.10.003

Song, Y. J., Xu, L. J., Miao, J. H., and Xiao, P. G. (2020). Research progress in *Chrysanthemum indicum*. *Modern Chin. Med.* 22, 1751–1756. doi: 10.13313/j.issn.1673-4890.20190722006

Sun, C. Q., Chen, F. D., Fang, W. M., Liu, Z. L., and Teng, N. J. (2010a). Advances in research on distant hybridization of Chrysanthemum. *Sci. Agri. Sin.* 43, 2508–2517. doi: 10.3864/j.issn.0578-1752.2010.12.015

Sun, Q. L., Hua, S., Ye, J. H., Zheng, X. Q., and Liang, Y. R. (2010b). Flavonoids and volatiles in *Chrysanthemum morifolium* Ramat flower from Tongxiang County in China. *Afr. J. Biotechnol.* 9, 3817–3821.

Sun, J., Wang, Z., Chen, L., and Sun, G. (2021). Hypolipidemic effects and preliminary mechanism of Chrysanthemum flavonoids, its main components luteolin and luteoloside in hyperlipidemia rats. *Antioxidants.* 10:1309. doi: 10.3390/antiox10081309

Suo, F. M., Chen, S. L., Yu, H., Xie, C. X., and Sun, C. Z. (2011). Study on the origin suitability of four famous Chrysanthemums in China. *Moderni. Trad. Chin. Med. Mater. Med.-World Sci. Technol.* 13, 332–339.

Tang, F. P. (2009). Study on distant hybridization between Chrysanthemum and four related genera. PhD dissertation. [Nanjing(Jiangsu)]: Nanjing Agricultural University, 1–121.

Tao, J. H. (2017). Study on the effect mechanism of polysaccharides from non-medicinal parts of chrysanthemum on inflammatory bowel disease. PhD dissertation. [Nanjing(Jiangsu)]: Nanjing University of Chinese Medicine, 1–174.

Tian, Z., Jia, H., Jin, Y., Wang, M., Kou, J., Wang, C., et al. (2019). Chrysanthemum extract attenuates hepatotoxicity via inhibiting oxidative stress in vivo and in vitro. *Food Nutr. Res.* 63:1667. doi: 10.29219/fnr.v63.1667

Tian, D., Yang, Y., Yu, M., Han, Z. Z., Wei, M., Zhang, H. W., et al. (2020). Anti-inflammatory chemical constituents of Flos Chrysanthemi Indici determined by UPLC-MS/MS integrated with network pharmacology. *Food Funct.* 11, 6340–6351. doi: 10.1039/D0FO01000F

Tyagi, S., Jung, J. A., Kim, J. S., and Won, S. Y. (2020). A comparative analysis of the complete chloroplast genomes of three Chrysanthemum boreale strains. *PeerJ.* 8:e9448. doi: 10.7717/peerj.9448

van Lieshout, N., van Kaauwen, M., Kodde, L., Arens, P., Smulders, M. J. M., Visser, R. G. F., et al. (2022). De novo whole-genome assembly of Chrysanthemum makinoi, a key wild chrysanthemum. *G3.* 12, jkab358. doi: 10.1093/g3journal/jkab358

Wang, L. Z. (2020). Preliminary identification of natural hybrids between *Chrysanthemum indicum* and Chrysanthemum vestitum. MSc thesis. [Taian(Shandong)]: Shandong Agricultural University, 1–56.

Wang, T. Y., Guo, Q. S., and Wang, T. (2012). Karyotype analysis of 21 medicinal Chrysanthemum cultivation types. *J. Nanjing Agri. Univ.* 35, 13–18.

Wang, Y. J., Guo, Q. S., Yang, X. W., Xu, W. B., and Tao, H. Y. (2008b). Characterization of chemical components of essential oil from flowers of *Chrysanthemum morifolium* produced in Anhui province. *China J. Chin. Materia Med.* 33, 2207–2211.

Wang, L., Hao, D. C., Fan, S. S., Xie, H. T., Bao, X. L., Jia, Z. J., et al. (2022). N2O emission and nitrification/denitrification bacterial communities in upland black soil under combined effects of early and immediate moisture. *Agriculture* 12:330. doi: 10.3390/agriculture12030330

Wang, J., Liang, Q., Zhao, Q., Tang, Q., Ahmed, A. F., Zhang, Y., et al. (2021b). The effect of microbial composition and proteomic on improvement of functional constipation by *Chrysanthemum morifolium* polysaccharide. *Food Chem. Toxicol.* 153:112305. doi: 10.1016/j.fct.2021.112305

Wang, D. Q., Liu, S. J., and Liang, Y. M. (1999). A study on producing areas of Chinese flos Dendranthematis. *China J. Chin. Mate. Medi.* 24:573.

Wang, D. Q., Liu, S. J., and Liang, Y. M. (2001). Research on medicinal taxa of Chinese chrysanthemum. *J. Anhui College TCM.* 20, 45–48.

Wang, Y. J., Su, J., Yu, J. J., Yan, M. Q., Shi, M. L., Huang, Q. D., et al. (2021c). Buddleoside-rich *Chrysanthemum indicum* L. extract has a beneficial effect on metabolic hypertensive rats by inhibiting the enteric-origin LPS/TLR4 pathway. *Fronti. Pharmacol.* 12:755140. doi: 10.3389/fphar.2021.755140

Wang, Y. S., Wang, M. X., Han, S., Zhang, J., Shen, X., Zhou, J., et al. (2016). Differences in amino acid contents in fresh flowers and manufactured goods among four Chrysanthemum cultivars in Tongxiang City. *Zhejiang Province. J. Anhui Agri. Univ.* 43, 1024–1028. doi: 10.13610/j.cnki.1672-352x.20161205.011

Wang, Y. J., Yang, X. W., and Guo, Q. S. (2008a). Studies on chemical constituents in Huangjuhua (flowers of *Chrysanthemum morifolium*). *China J. Chin. Materia Med.* 33, 526–530.

Wang, Z., Yuan, Y., Hong, B., Zhao, X., and Gu, Z. (2021a). Characteristic volatile fingerprints of four Chrysanthemum teas determined by HS-GC-IMS. *Molecules* 26:7113. doi: 10.3390/molecules26237113

Watson, L. E., Bates, P. L., Evans, T. M., Unwin, M. M., and Estes, J. R. (2002). Molecular phylogeny of Subtribe Artemisiinae (Asteraceae), including Artemisia and its allied and segregate genera. *BMC Evol. Biol.* 2:17. doi: 10.1186/1471-2148-2-17

Wei, M., Zhang, Y. J., Wang, T., Guo, Q. S., Zou, Q. J., Chen, F. R., et al. (2021). Correlations between content of linarin in Chrysanthemum indicum and climatic factors in habitats. *China J. Chin. Materia Med.* 46, 2167–2172. doi: 10.19540/j.cnki.cjcmm.20210320.104

Won, S. Y., Kwon, S. J., Lee, T. H., Jung, J. A., Kim, J. S., Kang, S. H., et al. (2017). Comparative transcriptome analysis reveals whole-genome duplications and gene selection patterns in cultivated and wild Chrysanthemum species. *Plant Mol. Biol.* 95, 451–461. doi: 10.1007/s11103-017-0663-z

Wu, G. S. (2007). Study on the genetic relationship between some Chrysanthemum and Ajania plants. MSc thesis. [Nanjing (Jiangsu)]: Nanjing Agricultural University, 1–101.

Wu, X. B. (2014). Preliminary study on distant hybridization of Chrysanthemum sensu lato (VII). MSc thesis. [Beijing]: Beijing Forestry University, 1–62.

Wu, X., Sun, Y., Shen, X., and Wang, Z. (2015). Study on combined effects of chemical components for different flowers blossoming degree of yellow medicinal *Chrysanthemum morifolium* from Zhejiang. *China J. Chin. Mate. Med.* 40, 3174–3178.

Xiao, Y. (2019). Research on nutritional and functional properties of Chrysanthesum mongolicum tea and its application in whey tea beverage. MSc thesis. [Hohhot (Inner Mongolia)]: Inner Mongolia Agricultural University, 1–62.

Xie, F. (2016). Preliminary study on distant hybridization of Chrysanthemum (IX): Germplasm utilization of Hippolytia and other genera. MSc thesis. [Beijing]: Beijing Forestry University, 1–82.

Xie, Y., Qu, J., Wang, Q., Wang, Y., Yoshikawa, M., and Yuan, D. (2012). Comparative evaluation of cultivars of *Chrysanthemum morifolium* flowers by HPLC-DAD-ESI/MS analysis and antiallergic assay. *J. Agri. Food Chem.* 60, 12574–12583. doi: 10.1021/jf304080v

Xiong, Y. X. (2014). Research on the germplasm resources of Fubaiju and the selection of excellent lines. MSc thesis. [Wuhan(Hubei)]: Hubei University of Chinese Medicine, 1–125.

Xu, W. B., Guo, Q. S., Li, Y. N., and Wang, T. (2005). Comparative study on internal quality of various *Chrysanthemum morifolium*. *China J. Chin. Mate. Med.* 30, 1645–1648.

Xu, W. B., Guo, Q. S., and Wang, C. L. (2006). RAPD analysis of genetic diversity of medicinal *Chrysanthemum morifolium*. *China J. Chin. Mate. Medi.* 2006, 18–21.

Xu, M., Jiang, Y., Chen, S., Chen, F., and Chen, F. (2021). Herbivory-induced emission of volatile terpenes in *Chrysanthemum morifolium* functions as an indirect defense against *Spodoptera litura* larvae by attracting natural enemies. *J. Agri. Food Chem.* 69, 9743–9753. doi: 10.1021/acs.jafc.1c02637

Xue, H., Jiang, Y., Zhao, H., Köllner, T. G., Chen, S., Chen, F. D., et al. (2019). Characterization of composition and antifungal properties of leaf secondary metabolites from thirteen cultivars of *Chrysanthemum morifolium* Ramat. *Molecules* 24:4202. doi: 10.3390/molecules24234202

Xue, G. M., Li, X. Q., Chen, C., Chen, K., Wang, X. B., Gu, Y. C., et al. (2018). Highly oxidized guaianolide sesquiterpenoids with potential anti-inflammatory activity from *Chrysanthemum indicum*. *J. Nat. Prod.* 81, 378–386. doi: 10.1021/acs.jnatprod.7b00867

Xue, L. M., Qin, X. M., and Guo, J. G. (2007). Comparative study on main components of *Chrysanthemum morifolium* introduced in Ruicheng in Shanxi Province. *Chin. Trad. Herb. Drug.* 2007, 751–754.

Xue, G. M., Xue, J. F., Zhao, C. G., Zhao, Z. Z., Zhi, Y. L., Du, K., et al. (2021). 1,10-seco guaianolide-type sesquiterpenoids from *Chrysanthemum indicum*. *J. Asian Nat. Prod. Res.* 23, 877–883. doi: 10.1080/10286020.2020.1787388

Yan, H. Y., Liu, Y., Xu, Y., Fang, Y., Guo, L. P., and Liu, D. H. (2021). Analysis and evaluation of mineral elements of *Chrysanthemum morifolium* for medicinal and tea

use of different germplasm resources. *China J. Chin. Mate. Med.* 46, 272–280. doi: 10.19540/j.cnki.cjcmm.20201023.101

Yang, L., Aobulikasimu, N., Cheng, P., Wang, J. H., and Li, H. (2017a). Analysis of floral volatile components and antioxidant activity of different varieties of *Chrysanthemum morifolium*. *Molecules* 22:1790. doi: 10.3390/molecules22101790

Yang, C. F., Dong, C. M., Xing, B., Xia, W., and Tian, Y. J. (2018a). Study on quality of *Chrysanthemum morifolium* from different habitats. *Mod. Chin. Med.* 20, 716–720. doi: 10.13313/j.issn.1673-4890.20170519008

Yang, P. F., Feng, Z. M., Yang, Y. N., Jiang, J. S., and Zhang, P. C. (2017b). Neuroprotective caffeoylquinic acid derivatives from the flowers of *Chrysanthemum morifolium*. *J. Nat. Prod.* 80, 1028–1033. doi: 10.1021/acs.jnatprod.6b01026

Yang, W., Glover, B. J., Rao, G. Y., and Yang, J. (2006). Molecular evidence for multiple polyploidization and lineage recombination in the *Chrysanthemum indicum* polyploid complex (Asteraceae). *New Phytol.* 171, 875–886. doi: 10.1111/j.1469-8137.2006.01779.x

Yang, X. W., Han, M. H., Tao, H. Y., Wang, Z. A., Yang, Z., and Xiao, S. Y. (2007). GC-MS analysis of essential oil from anthodiums of *Chrysanthemum morifolium* processed by microwave-airflow and steam calefaction. *China J. Chin. Mate. Medi.* 32, 227–230. doi: 10.1021/acs.jnatprod.6b01026

Yang, Y. J., Jiang, Z. F., Guo, J., Yang, X., Xu, N., Chen, Z., et al. (2018b). Transcriptomic analyses of *Chrysanthemum morifolium* Ramat under UV-B radiation treatment reveal variations in the metabolisms associated with bioactive components. *Indus. Crop Prod.* 124, 475–486. doi: 10.1016/j.indcrop.2018.08.011

Yang, P. F., Yang, Y. N., Feng, Z. M., Jiang, J. S., and Zhang, P. C. (2019). Six new compounds from the flowers of Chrysanthemum morifolium and their biological activities. *Bioorg. Chem.* 82, 139–144. doi: 10.1016/j.bioorg.2018.10.007

Zhang, R. D. (2008). The Chrysanthemum culture research in ancient China. PhD dissertation. [Nanjing(Jiangsu)]: Nanjing Normal University, 16–25.

Zhang, L. J., and Dai, S. L. (2009). Research advance on germplasm resources of Chrysanthemum morifolium. *Chin. Bull. Bot.* 44, 526–535. doi: 10.3969/j.issn.1674-3466.2009.05.002

Zhang, E. X., Fang, L., Zhang, J., Yu, L. J., and Xiao, X. (2000). The antioxidant activity research of Chrysanthemum extract. *Food Sci.* 21, 6–9.

Zhang, N., He, Z., He, S., and Jing, P. (2019). Insights into the importance of dietary chrysanthemum flower (*Chrysanthemum morifolium* cv. Hangju)-wolfberry (*Lycium barbarum* fruit) combination in antioxidant and anti-inflammatory properties. *Food Res. Int.* 116, 810–818. doi: 10.1016/j.foodres.2018.09.015

Zhang, W., Jiang, Y., Chen, S., Chen, F. D., and Chen, F. (2021b). Concentration-dependent emission of floral scent terpenoids from diverse cultivars of Chrysanthemum morifolium and their wild relatives. *Plant Sci.* 309:110959. doi: 10.1016/j.plantsci.2021.110959

Zhang, K., Jiang, Y., Zhao, H., Köllner, T. G., Chen, S., Chen, F. D., et al. (2020). Diverse terpenoids and their associated antifungal properties from roots of different cultivars of *Chrysanthemum morifolium* Ramat. *Molecules* 25:2083. doi: 10.3390/molecules25092083

Zhang, W., Xu, H., Duan, X., Hu, J., Li, J., Zhao, L., et al. (2021a). Characterizing the leaf transcriptome of Chrysanthemum rhombifolium (Ling et C. Shih), a drought resistant, endemic plant from China. *Front. Genet.* 12:625985. doi: 10.3389/fgene.2021.625985

Zhang, Z., Zhang, Y., Wang, L., Cui, T., Wang, Y., Chen, J., et al. (2022). On-line screening of natural antioxidants and the antioxidant activity prediction for the extracts from flowers of *Chrysanthemum morifolium* ramat. *J. Ethnopharmacol.* 294:115336. doi: 10.1016/j.jep.2022.115336

Zhao, H. B. (2007). Phylogeny of tribe Anthemideae (Asteraceae) from East Asia and intergeneric cross between Dendranthema×grandiflorum (Ramat.) Kitam. and Ajania pcifica (Nakai) K. Bremer & Humphries. PhD dissertation. [Nanjing (Jiangsu)]: Nanjing Agricultural University, 1–181.

Zhao, S. H. (2015). Analysis of nutrient composition of six kinds of chrysanthemum. MSc thesis. [Kaifeng(Henan)]: Henan University, 1–81

Zhao, H. B., Chen, F. D., Chen, S. M., Wu, G. S., and Guo, W. M. (2010). Molecular phylogeny of Chrysanthemum, Ajania and its allies (Anthemideae, Asteraceae) as inferred from nuclear ribosomal ITS and chloroplast trnL-F IGS sequences. *Plant Syst. Evol.* 284, 153–169. doi: 10.1007/s00606-009-0242-0

Zhao, H. B., Chen, F. D., Guo, W. M., Tang, F. P., and Fang, W. M. (2008). Preliminary study on hybrid compatibility between Chrysanthemum and some genera of Anthemideae. *J. Nanjing Agri. Univ.* 2008, 139–143.

Zheng, C. P. (2015). Isolation, purification, structure identification and bioactivity of chrysanthemum polysaccharide. MSc thesis. [Nanchang(Jiangxi)]: Nanchang University, 1–107

Zhou, J. (2009). Studies on the problem of origin of Chinese garden chrysanthemum. PhD dissertation. [Beijing]: Beijing Forestry University, 1–148.

Zhou, L., Liu, Q., Hong, G., Song, F., Zhao, J., Yuan, J., et al. (2019a). Cumambrin A prevents OVX-induced osteoporosis via the inhibition of osteoclastogenesis, bone resorption, and RANKL signaling pathways. *FASEB J.* 33, 6726–6735. doi: 10.1096/fj.201800883RRR

Zhou, H. P., Ren, M. X., Guan, J. Q., Liu, Y. L., Xiong, Y. X., Zhong, Q. F., et al. (2019b). Research progress on chemical constituents and pharmacological effects of Chrysanthemum morifolium and predictive analysis on quality markers. *Chin. Trad. Herb. Drug.* 50, 4785–4795.

Zhou, Z., Xian, J., Wei, W., Xu, C., Yang, J., Zhan, R., et al. (2021). Volatile metabolic profiling and functional characterization of four terpene synthases reveal terpenoid diversity in different tissues of *Chrysanthemum indicum* L. *Phytochemistry* 185:112687. doi: 10.1016/j.phytochem.2021.112687

Zhuo, F. F., Zhang, C., Zhang, H., Xia, Y., Xue, G. M., Yang, L., et al. (2019). Chrysanthemulide A induces apoptosis through DR5 upregulation via JNK-mediated autophagosome accumulation in human osteosarcoma cells. *J. Cellu. Physiol.* 234, 13191–13208. doi: 10.1002/jcp.27991

Zou, Q., Guo, Q., Wang, T., Chen, J., Yang, F., and Yang, C. (2022). Comparison of metabolome characteristics and screening of chemical markers in *Chrysanthemum indicum* from different habitats. *Physiol. Mol. Biol. Plants* 28, 65–76. doi: 10.1007/s12298-022-01137-z

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 17 August 2022
DOI 10.3389/fpls.2022.951824

OPEN ACCESS

EDITED BY
Wansheng Chen,
Second Military Medical University, China

REVIEWED BY
Ruibing Chen,
Naval Medical University, China
Yun-peng Du,
Beijing Academy of Agricultural and Forestry Sciences, China

*CORRESPONDENCE
Chunnian He
cnhe@implad.ac.cn

SPECIALTY SECTION
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

RECEIVED 24 May 2022
ACCEPTED 18 July 2022
PUBLISHED 17 August 2022

CITATION
Shen J, Li P, Wang Y, Yang K, Li Y, Yao H, Wang Q, Xiao P and He C (2022) Pharmacophylogenetic study of *Scutellaria baicalensis* and its substitute medicinal species based on the chloroplast genomics, metabolomics, and active ingredient. *Front. Plant Sci.* 13:951824. doi: 10.3389/fpls.2022.951824

Pharmacophylogenetic study of *Scutellaria baicalensis* and its substitute medicinal species based on the chloroplast genomics, metabolomics, and active ingredient

Jie Shen[1,2], Pei Li[1], Yue Wang[1], Kailing Yang[1], Yue Li[1], Hui Yao[1], Qiang Wang[3], Peigen Xiao[1] and Chunnian He[1]*

[1]Key Laboratory of Bioactive Substances and Resources Utilization of Chinese Herbal Medicine, Ministry of Education, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China, [2]School of Medical Laboratory, Weifang Medical University, Weifang, China, [3]State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, China

The genetic relationships among the species in *Scutellaria* genus remain unclear because of the variation in the number of species and complex trait. The usage of *S. baicalensis* and its four substitute medicinal species (*S. amoena*, *S. hypericifolia*, *S. likiangensis*, and *S. viscidula*) in traditional medicines make their specialized metabolism important in China, but interspecific genetic and chemical differences have rarely been reported for these species. In this study, the chloroplast genomes of four substitute species for *S. baicalensis* were assembled, and comparative and phylogenetic analyses were performed with these species and other *Scutellaria* relatives. In addition, metabolomics analyses were performed and the contents of the main active compounds were determined to reveal the interspecific chemical diversity of *S. baicalensis* and its four substitute species. The full lengths of their chloroplast genomes ranged from 151,574 to 151,816 bp with an average GC content of 38.34%, and a total of 113 genes were annotated. In the chloroplast genomes of *S. baicalensis* and its four substitutes, one hypervariable region (*petA-psbL*) is proposed as a potential DNA barcode. Phylogenetic analysis showed that the subdivision of the genus *Scutellaria* should be reconsidered. The metabolomics and content determination analyses showed that the four species exhibit a metabolism similar to that of *S. baicalensis* in different parts. Except for the roots of *S. likiangensis*, all parts of the substitute species showed high contents of baicalin. Genetic and chemical analyses of four substitute

medicinal species for *S. baicalensis* were performed here for the first time, and their pharmacophylogenetic relationships were further explored, providing a scientific basis for the subsequent development of the medicinal value and resource utilization of *Scutellaria*.

KEYWORDS

***Scutellaria baicalensis*, pharmacophylogeny, substitute medicinal species, chloroplast genomics, metabolomics**

Introduction

Scutellaria is a significant genus in the family Lamiaceae, which includes approximately 360–469 accepted species worldwide (Ranjbar and Mahmoudi, 2017; Safikhani et al., 2018; Zhao et al., 2020; Shen et al., 2021). According to records, 72 species in this genus have been used with a long history as traditional herbal medicines to treat various diseases, such as cancer and inflammatory, hepatic, gastric, neurological, and cardiovascular diseases (Marsh et al., 2014; Grzegorczyk-Karolak et al., 2016; Wang et al., 2018, 2020). *Scutellaria* species are rich in flavonoids and terpenoids, and the main compounds baicalein, baicalin, and wogonin have been evaluated in clinical trials (Shen et al., 2021).

In China, *S. baicalensis* is the most widely studied species in *Scutellaria*. Scutellariae Radix, consisting of the roots of *S. baicalensis,* has been used as a traditional Chinese medicine for thousands of years and is officially recorded in successive editions of the *Chinese Pharmacopoeia* (Wang et al., 2018). According to research, the roots of some closely related species, such as *S. amoena, S. hypericifolia, S. likiangensis,* and *S. viscidula*, are similar in shape to the roots of *S. baicalensis*, and they have abundant natural medicinal resources and have been widely used as substitutes for Scutellariae Radix in China (Shen et al., 2021). These alternatives have been used as herbal folk medicines for a long time and are listed in some local flora and local standards for traditional Chinese medicine. In the current *Sichuan Standard for Traditional Chinese Medicine* (2010), Scutellariae Amoenae Radix (Chuan Huang Qin) originating from the roots of *S. amoena*, *S. hypericifolia*, and *S. tenax* var. *patentipilosa* is used to prevent miscarriage and bleeding. In addition, in the current standard *Quality standard of Chinese and ethnic medicinal materials in Guizhou Province* (2003), the root of *S. amoena* is listed as Scutellariae Amoenae Radix (Xi'nan Huangqin) and described as useful for treating fever, cough, hemoptoe, jaundice, diarrhea, and carbuncles and for preventing miscarriage. Furthermore, the root of *S. viscidula* was listed in the *Drug Standard of Jilin Province* (1977) as Radix Scutellariae Viscidula and described as useful for treating cough, jaundice, diarrhea, and swelling and pain of the eye and for preventing miscarriage. Despite their important medicinal value, the genetic and chemical diversity of these substitutes for *S. baicalensis* remains unclear.

According to the Flora of China (Wu and Li, 1977), S. baicalensis and its substitute species (S. amoena, S. hypericifolia, S. likiangensis, and S. viscidula) all belong to the Scutellaria subgen. Scutellaria sect. Stachymacris subsect. Angustifoliae clade, which has high similarity in terms of morphology and root shape. To date, only one study has used the plastid molecular marker trnL-F to discriminate between S. baicalensis and its adulterants (*S. amoena* and *S. viscidula*) (Wang et al., 2012). In recent years, chloroplast (cp) genomes have played an essential role in identifying closely related plant species and performing phylogenetic analysis since cp DNA is more discriminating and versatile than nuclear DNA sequences (Cui et al., 2019a; Zhao et al., 2020). The cp is an essential photosynthetic organelle in plants (Daniell et al., 2016). The cp is a uniparentally inherited plastid that contains circular double-stranded DNA (Zhao et al., 2020). A typical angiosperm cp genome has a four-part structure that includes two identical regions in opposite orientations called inverted repeats (IRs), separated by large single-copy (LSC) and small single-copy (SSC) regions (Ruhlman and Jansen, 2014). *Scutellaria* is a very isolated genus with unsatisfactory traditional divisions, and complete cp genomes have been published for only 13 species in the genus (Jiang et al., 2017; Lee and Kim, 2019; Zhao et al., 2020). In addition, little is known regarding the plastome structural variation among S. baicalensis and its substitutes. Thus, sequencing the cp genomes of these four Scutellariae Radix substitute species is beneficial not only for accurately identifying closely related species but also for greatly contributing to medicinal cp genetic development in *Scutellaria*. In addition, until now, among the substitute species for S. baicalensis, only one complete cp genome (S. amoena) of the subsect. Angustifoliae clade has been published. The availability of complete cp genome sequences of S. hypericifolia, S. likiangensis, and S. viscidula will help reveal the evolutionary pattern of the complete cp genome in the subsect. Angustifoliae clade.

Plant secondary metabolites usually provide the ultimate features for distinguishing among similar species (Nguyen et al., 2016). The phytochemical diversity of plant metabolites could

provide a basis for our understanding of the evolution and classification of different plant species (Wink, 2015). Currently, metabolomics is considered the most common method for profiling and comparing phytochemical compositions among species (Kushalappa and Gunnaiah, 2013). To date, a large number of chemical components have been identified from the genus *Scutellaria*, and flavonoids and diterpenes have been confirmed to be the two main groups of active constituents in this genus. Based on the pharmacophylogenetic theory, plants with similar therapeutic effects are phylogenetically related and often similar in chemical composition. However, due to the lack of adequate chemical composition and content determination studies, additional evidence is needed to confirm the relationship between chemical diversity and phylogeny in *Scutellaria* (Shen et al., 2021). Previous studies have indicated that the major flavonoids baicalin, baicalein, wogonoside, and wogonin are 4′-deoxyflavones unique to the roots of *S. baicalensis* with a wide range of pharmacological activities, and high concentrations of these compounds have also been detected in many *Scutellaria* plants. However, only two substitutes (*S. amoena* and *S. viscidula*) for *S. baicalensis* have been reported to contain compounds similar to those in *S. baicalensis*, such as baicalein, baicalin, and wogonin (Wang et al., 2012; Ling et al., 2016). Moreover, the chemical profiles of these four substitutes (*S. amoena*, *S. hypericifolia*, *S. likiangensis*, and *S. viscidula*) for *S. baicalensis* remain to be completely explored. Currently, metabolomics is considered the most common method for profiling and comparing phytochemical compositions among species (Kushalappa and Gunnaiah, 2013). Due to the significant differences in chemical composition between the aerial parts and roots of *S. baicalensis* previously reported (Shen et al., 2019), in this study of the four species mentioned above, we divided *S. baicalensis* and its four substitute species into aerial parts and roots for separate metabolomics determination.

In this study, we first sequenced the complete cp genomes of *S. amoena*, *S. hypericifolia*, *S. likiangensis,* and *S. viscidula* by using Illumina sequencing technology. Comparative analyses were conducted to explore the different characteristics of the cp genomes of these four substitutes for *S. baicalensis*. In addition, a phylogenetic tree was constructed to reveal the phylogenetic positions of these medicinal plants in *Scutellaria*. Furthermore, the chemical profiles of metabolites in different plant parts were determined by ultra-performance liquid chromatography/quadrupole time-of-flight mass spectrometry (UPLC-QTOF-MS), and chemometric analyses (PCA, PLS-DA, and HCA) were conducted. In addition, high-performance liquid chromatography (HPLC) analysis was used to determine the 15 main flavonoids in different parts of *S. baicalensis* and its four substitute species. Analyses of phylogenetic relationships and phytochemical diversity among the four substitute species for *S. baicalensis*, in this study, can expand our understanding of the genetics and chemistry of the genus *Scutellaria*.

Materials and methods

Plant materials

Fresh whole plants of *S. amoena, S. baicalensis, S. hypericifolia, S. likiangensis,* and *S. viscidula* were collected from the Gucheng District (Lijiang City, Yunnan Province), Baihua Mountain (Beijing), Luoshage (Dali Bai Autonomous Prefecture, Yunnan Province), Wenhai Road (Yulong Naxi Autonomous County, Lijiang City, Yunnan Province), and Wulan Halaga (West Ujimqin, Inner Mongolia Autonomous Region) in China, respectively. Three samples were collected at each site in August 2019 and were divided into two parts: aerial parts and roots. The samples were authenticated by Professor Qiang Wang (State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences), and voucher specimens (HQS-01, *S. amoena*; HQS-02, *S. baicalensis*; HQS-03, *S. hypericifolia*; HQS-04, *S. likiangensis*; and HQS-05, *S. viscidula*) were deposited in the pharmacophylogeny research center (Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China).

DNA extraction and sequencing

Fresh leaves were collected and dried using silica gel for DNA extraction. Genomic DNA was extracted using the Plant Genomic DNA Kit (Huayueyang, Beijing, China) according to the instructions. Then, 1% (w/v) agarose gel electrophoresis was used to test DNA integrity, and concentrations were determined using a NanoPhotometer® spectrophotometer (IMPLEN, Munchen, Germany) and a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, United States). Finally, high-quality DNA was used for library construction and sequencing on any Illumina platform, and 150-bp paired-end reads were generated.

Chloroplast genome assembly, annotation, and structural analysis

K-values of 21, 55, 85, and 115 were obtained using GetOrganelle software (Jin et al., 2018) to obtain the desired results, and the entire genome of *S. baicalensis* Georgi (NC_027262.1) from the National Center for Biotechnology Information (NCBI) was used as the reference for cp genome assembly.

Chloroplast annotation of the assembled sequences was performed using GeSeq[1] (Michael et al., 2017). These sequences

1 https://chlorobox.mpimp-golm.mpg.de/geseq.html

were compared with the known database, and sequences of the protein-coding gene BLAT and the rRNA/tRNA gene BLAT were identified based on similarity values exceeding 85%. In addition to the default known database in GeSeq, the cp genome annotation process also included the addition of the entire genome of *S. baicalensis* from the NCBI database. Based on the GeSeq annotation results, the software used to create a circular physical map of the cp genome was OGDRAW[2] (Stephan et al., 2019).

The GC content of the cp genome was analyzed using MEGA-X (Kumar et al., 2018). The MISA Perl script[3] (Beier et al., 2017) was used to identify simple sequence repeats (SSRs). The definition of microsatellites (unit size/minimum repeats) was (1/10) (2/5) (3/4) (4/3) (5/3) (6/3), and the maximum sequence length between two SSRs was set at 100 (Zhao et al., 2020). REPuter (Kurtz et al., 2001) was used to identify the sizes and locations of repeat sequences, including forwarding, palindromic, reverse, and complement repeats, in the cp genomes. The minimal size for all repeat types was 30 bp, with a Hamming distance of 3 (Zhao et al., 2020). The spliced cp genome sequence file was uploaded to the Tandem Repeats Finder v4.09 website[4] to identify the tandem repeats in cp DNA with default parameters (Benson, 1999).

Interspecific comparison and phylogenetic analysis

Thirteen reported complete cp genomes of *Scutellaria* were downloaded from the NCBI database: *S. baicalensis* (MF521633), *S. calcarata* (MN128385), *S. indica* var. *coccinea* (MN047312), *S. insignis* (KT750009), *S. lateriflora* (KY085900), *S. mollifolia* (MN128384), *S. orthocalyx* (MN128383), *S. quadrilobulata* (MN128381), *S. kingiana* (MN128389), *S. altaica* (MN128387), *S. przewalskii* (MN128382), *S. scordifolia* (MT712016), and *S. tsinyunensis* (MT544405). The mVISTA web interface[5] was used to align and compare the five cp genome [*S. amoena, S. baicalensis* (MF521633), *S. hypericifolia, S. likiangensis, S. viscidula*] sequences (Mayor et al., 2000). Shuffle-LAGAN was selected as the alignment program for detecting the rearrangements and inversions in sequences. Multiple sequence alignment of cp genomes was conducted using Condon Code Aligner V.9.0.1 and MAFFT v.7.471[6] (Frazer et al., 2004). The divergence among different cp genomes and the identification of mutational hotspots were performed by quantifying nucleotide variability in DnaSP v6.12.03 (Rozas et al., 2017). The window length was set to 600 bp, with a 200-bp step size. IRscope[7] was used to visualize the gene differences at the boundaries of the junction sites of the 14 cp genomes (Amiryousefi et al., 2018). In this study, in addition to *S. baicalensis* and its four substitute species (*S. amoena*, *S. hypericifolia, S. likiangensis, and S. viscidula*), an additional 12 cp genomes were downloaded from NCBI to construct a cp phylogenetic tree. Maximum likelihood (ML) phylogenetic inference was performed using IQ-TREE software (Nguyen et al., 2015) with 1000 bootstrap replicates based on the TVM+F+R6 nucleotide substitution model to assess branch support. MrModeltest 2.3 (Nylander, 2004) was applied to calculate the best-fitting model based on Akaike's information criterion. Bayesian inference (BI) phylogenetic trees were analyzed in MrBayes version v3.2.7 (Ronquist and Huelsenbeck, 2003) based on the complete cp genome tree.

Utra-performance liquid chromatography/quadrupole time-of-flight mass spectrometry analysis of specialized metabolites

Dividing each whole plant into the aerial part and root part, all dried samples of the two parts were crushed and screened through a 40-mesh sieve. Samples (25 mg) were macerated with MeOH (25 ml) and placed in an ultrasonic water bath for 30 min. After cooling to room temperature and complementing the reduced volume, each extract was filtered through a 0.25-μm membrane filter, and 1 μL extracts were injected for each analysis. Three biological replicates of the samples were completed using different individuals of the same species.

The UPLC-QTOF-MS analysis was performed on a Waters Acquity Ultra High-Performance LC system (Waters Co., Milford, MA, United States) coupled with a Waters Xevo G2-XS TOF mass spectrometer (Waters Co., Milford, MA, United States) equipped with an electrospray ionization interface. Chromatographic separations were conducted on a Waters Acquity BEH C18 column (2.1 mm × 100 mm, 1.7 μm). The mobile phase was comprised of acidified water with 0.1% formic acid (A) and MeOH (B). The column temperature was maintained at 35°C. The elution program was as follows: 0–3 min: 10–35% B; 3–9 min: 35–70% B; 9–12 min: 70–100% B; and 12–14 min: 100% B. The flow rate was 0.3 mL/min, and the photodiode array detector scanning range was from 200 to 400 nm. The negative mode was used for ionizing the metabolites, and the mass spectra were scanned from 1,000 to 1,200 m/z. The parameters were set as follows: the cone voltage and capillary voltage were set at 40 V and 2.4 kV, respectively. The source temperature and the desolvation temperature were

2 http://ogdraw.mpimp-golm.mpg.de/
3 http://pgrc.ipk-gatersleben.de/misa
4 http://tandem.bu.edu/trf/trf.html
5 http://genome.lbl.gov/vista/mvista/submit.shtml
6 https://mafft.cbrc.jp/alignment/server/index.html
7 https://irscope.shinyapps.io/irapp/

100°C and 350°C, respectively. The desolvation gas and cone gas rates were 900 and 50 L/h, respectively. Leucine-enkephalin was used as the lock mass in the analyses (negative ion mode: $[M-H]^- = 554.2615$).

High-performance liquid chromatography analysis of chemical components in the five *Scutellaria* species

The aerial parts and root samples were all processed according to a previously reported method (Shen et al., 2019). These samples were also analyzed following our previously established protocol (Shen et al., 2019). Three biological replicates of the samples were completed using different individuals of the same species.

Data processing and analysis

Progenesis QI 2.3 software (Waters, Milford, MA, United States) was used to process the centroid MS^E raw data from UPLC-QTOF-MS, including import data, alignment review, experimental design setup, peak selection, deconvolution, and compound identification. The MetaboAnalyst 5.0[8] webserver was used for multivariate statistical analyses, including principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) in the Pareto scaling mode, hierarchical clustering analysis (HCA) with Euclidean distances, and heatmap analysis.

Results

Assembly of chloroplast genomes and comparative analysis of the four substitute medicinal species for *Scutellaria baicalensis*

Using Illumina HiSeq/MiSeq sequencing platforms, we obtained 3137–3953 M (Illumina raw reads) for four substitute medicinal species of *S. baicalensis* (*S. amoena*, *S. hypericifolia*, *S. likiangensis*, and *S. viscidula*), and 3127–3838 M reads were finally assembled to generate complete cp genomes. The complete cp genomes of the four species were 151,574–151,816 bp in total size (**Supplementary Table S1**). The lengths were similar in *S. amoena* and *S. likiangensis*. *S. viscidula* was the longest, and *S. hypericifolia* was the shortest. The cp genomes all exhibited the typical quadripartite structure with a large single-copy region (LSC) (83,810–83,995 bp), a small single-copy region (SSC) (17,321–17,326 bp), and a pair of inverted repeats (IRs) (25,253–25,265 bp) (**Supplementary Table S1** and **Figure 1**). The sizes of the SSC and IR regions were identical among *S. amoena*, *S. hypericifolia*, and *S. likiangensis* but differed in *S. viscidula*. The average GC content of the genomes was 38.34%, which is similar to values in other *Scutellaria* species (Lukas and Novak, 2013; Zhao et al., 2020). The GC content was the highest (43.61–43.64%) in the IR regions, the lowest (32.68–32.72%) in the SSC region, and was in the range of 36.34–36.36% in the LSC region (**Supplementary Table S1**).

In four *Scutellaria* species, a total of 113 genes were identified, including 80 protein-coding genes, 29 tRNA genes, and 4 rRNA genes. Of these, 18 genes were duplicated in the IR regions, including seven protein-coding genes (*rps12, rps7, ndhB, ycf15, ycf2, rpl23,* and *rpl2*), seven tRNAs (*trnN-GUU, trnR-ACG, trnA-UGC, trnI-GAU, trnV-GAC, trnL-CAA,* and *trnI-CAU*), and four rRNAs (*rrn5, rrn4.5, rrn23,* and *rrn16*) (**Supplementary Table S2**). There were five SSR types (mono-, di-, tri-, tetra-, and pentanucleotides) in these four *Scutellaria* species (**Supplementary Figure S1**) and 40, 43, 40, and 43 repeat sequences in the *S. amoena*, *S. hypericifolia, S. likiangensis,* and *S. viscidula* cp genomes, respectively. The number of mononucleotide repeats was the highest, accounting for 58.1% of the total number of SSRs, and the number of pentanucleotide repeats was the lowest, accounting for 3.73%. On average, there were approximately 22 tandem repeats, 13 forward repeats, and 17 palindrome repeats. The highest total number of repetitive sequences was found in *S. viscidula,* with 61 duplicate sequences, and the lowest total was found in *S. hypericifolia,* with 50 repetitive sequences (**Supplementary Figure S1**).

Interspecific comparisons of sequence identity among the cp genomes of *S. baicalensis* and its four substitute medicinal species were conducted with mVISTA, and the annotated *S. baicalensis* sequence was used as a reference. The mVISTA results indicated that the cp genomes of the species of *Scutellaria* all showed a high degree of conservation (**Supplementary Figure S2**). The variation was more pronounced in the single copy region than in the IR region, and the protein-coding region was more conserved in terms of sequence than the non-coding region. In these five cp genomes of *Scutellaria* species, the non-coding region was more variable than the coding region, and the LSC and SSC regions were more variable than the IR region. Based on the mVISTA results, 6 highly variable regions were found in four species of *Scutellaria*: *trnK* (*UUU*)-*rps16, rps16-trnQ* (*UUG*), *atpH-atpI*, *petN-psbM, trnT* (*GGU*)-*psbD,* and *petA-psbJ* (**Supplementary Figure S2**). The average nucleotide variability (PI) of the cp genomes of the four *Scutellaria* species was 0.00136 (**Supplementary Figure S3**), and mutation hotspots with high PI values (>0.01) were distributed in the LSC region with a maximum PI value of 0.0155 (*petA-psbL*). The mutational hotspots within these *Scutellaria* species were commonly located

8 https://www.metaboanalyst.ca/

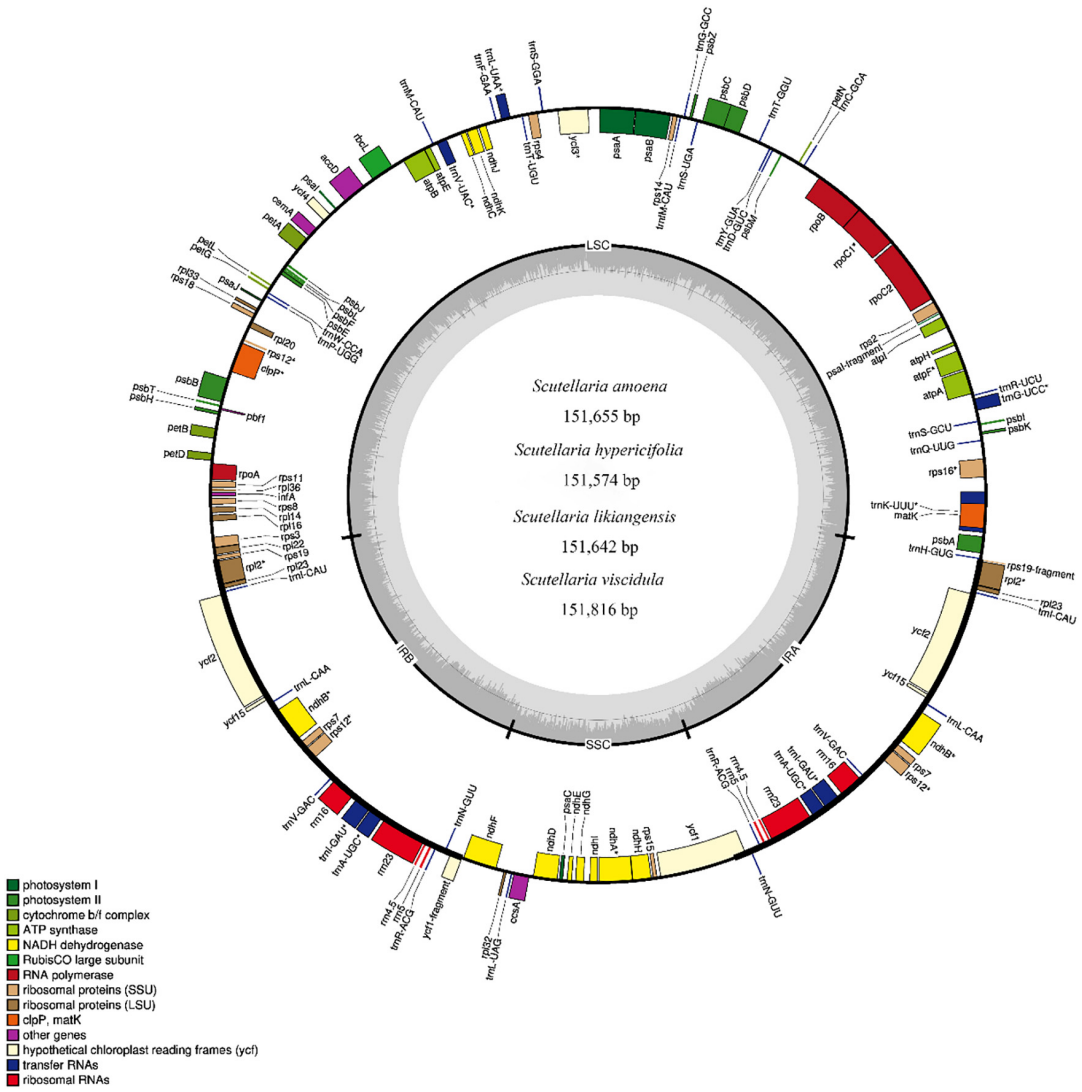


FIGURE 1
Gene map of the complete cp genomes of *S. amoena*, *S. hypericifolia*, *S. likiangensis*, and *S. viscidula*. Only one map is shown here due to the negligible differences among the four species. Genes inside the circle are transcribed clockwise, while genes outside the circle are transcribed counterclockwise. The different colors indicate different functional groups of genes. Dark gray areas inside the circle indicate the GC content, while light gray areas indicate the AT content of the genome.

in the LSC and SSC regions, as well as the IR region with PI values lower than 0.002. The development and utilization of these highly variable sites are beneficial to phylogenetic research on the subspecies and related genera of *Scutellaria* (**Supplementary Figure S3**).

The expansion and contraction of IR regions in the cp genomes of these five *Scutellaria* species and 12 other species of *Scutellaria* were compared (**Supplementary Figure S4**). In all species, the IRa/LSC junction was located within the *rps19* gene. The *rps19* gene had 41–46 bp projections into the IRa region, resulting in the presence of a portion of the *rps19* gene (*rps19* fragment) in the IRb region. The *ycf1* gene straddles the IRb/SSC boundary, but only a small fraction of the gene (771–783 bp) is located in the IRb region, and the pseudogene *ycf1* was detected in the IRa region. The length of *trnH* from the IRa/LSC border was 0 bp.

Phylogenetic analysis based on chloroplast genomes

In this study, based on complete cp genome sequences, 17 species (including the four *Scutellaria* cp genomes newly sequenced in this study) from *Scutellaria* were used for the construction of an ML tree and BI trees (with *Pogostemon*

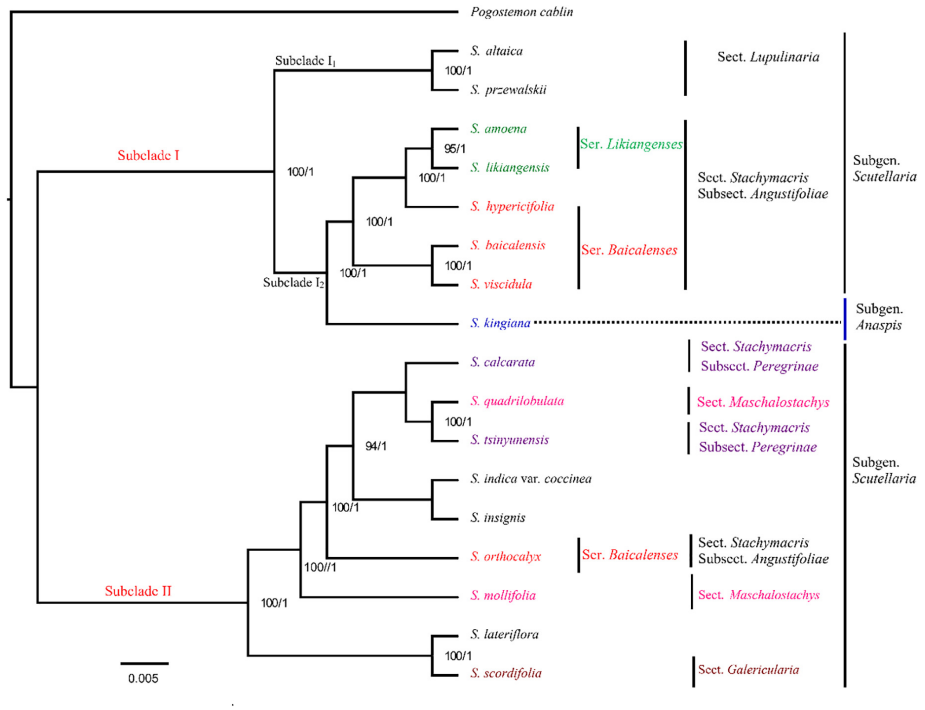


FIGURE 2
Phylogenetic relationships of the 17 species by maximum likelihood (ML) and Bayesian inference (BI) analyses (*Pogostemon cablin* as an outgroup) based on alignments of complete chloroplast genomes. The numbers above nodes are supported values with ML bootstrap values on the left and Bayesian posterior probability (PP) values on the right.

cablin as an outgroup, **Figure 2** and **Supplementary Figure S5**). All species of *Scutellaria* were recovered in two subclades (ML/BS 100, BI/PP 1) (**Figure 2**). Subclade I (100, 1) comprised eight species. Subclade I is divided into Subclade I_1 and Subclade I_2, with 100% support, and these clustered species all have well-developed roots in their morphology. Subclade I_1 contains two species, *S. altaica* and *S. przewalskii*, both of which are perennial subshrubs with developed woody rhizomes and belong to Subg. Scutellaria Sect. Lupulinaria in the *Flora of China* (Wu and Li, 1977). Subclade I_2 consists of six species *S. amoena, S. likiangensis, S. hypericifolia, S. baicalensis, S. viscidula,* and *S. kingian*a. In the maximum likelihood tree constructed based on the cp genome, *S. kingiana* was found to have high support (100, 1) for clustering these five species into one, indicating a close affinity. Moreover, *S. kingiana* is a dwarf perennial herb with creeping woody developed rhizomes and belongs to Subgen. Anaspis in the *Flora of China* (Wu and Li, 1977). The species of *S. amoena*, *S. baicalensis*, *S. hypericifolia*, *S. likiangensis,* and *S. viscidula* are clustered together and are all perennial herbs with developed fleshy rhizomes and belong to *Scutellaria* subgen. *Scutellaria* sect. *Stachymacris* subsect. *Angustifoliae* clade in the *flora of China* (Wu and Li, 1977; Lou and Qin, 1995; He et al., 2012). In addition, *S. viscidula* and *S. baicalensis* clustered in one clade with high support (BS value 100), which indicated that they are more closely related. *Scutellaria amoena, S. likiangensis,* and *S. hypericifolia* were clustered in one clade with high support (100,1). Moreover, as substitutes for *S. baicalensis*, the roots of *S. amoena*, *S. likiangensis*, *S. hypericifolia*, and *S. viscidula* are yellow and well developed (Lou and Qin, 1995; Li et al., 2003; He et al., 2012). Subclade II (100, 1) consists of nine species: *S. calcarata, S. lateriflora, S. indica* var. *coccinea, S. insignis, S. mollifolia, S. orthocalyx, S. quadrilobulata, S. tsinyunensis,* and *S. scordifolia* (**Figure 2**). In Subclade II, all cp sequences from the same clade had high support, and as in Subclade I, the positions of these nine species in phylogenetic trees are also different from the traditional Chinese plant classification system in *Flora of China*. In addition, according to the morphological description of the *Flora of China* (Wu and Li, 1977), most of them are perennial herbs and have no well-developed roots (**Figure 2**). Therefore, not only the taxonomic positions of the subgenera but also the taxonomic positions of sections and series need further confirmation in the genus *Scutellaria*.

Specialized metabolites in the *Scutellaria* species

The UPLC-QTOF-MS analyses of the aerial parts and roots of *S. baicalensis* and its four substitute medicinal species

(*S. amoena*, *S. hypericifolia*, *S. likiangensis*, and *S. viscidula*) were performed in negative ion mode using electrospray ionization (ESI^-). Progenesis QI is the most commonly used and reliable software for the online identification of metabolomics data (Gu et al., 2019; Wang et al., 2019). In this study, the chemical composition of these five species was tentatively identified by an online search of an in-house library of the genus *Scutellaria* and a Progenesis MetaScope internal repository through Progenesis QI software and by comparison with literature and standards, as shown in **Table 1**. Progenesis QI was used with m/z and typical fragment ions through comparison with candidate compounds from various databases. For example, in the online identification by Progenesis QI, baicalin was identified based on molecular weight, adduct ion (m/z 467.0583, M+Na-2H), and typical fragment ion (m/z 269.0431, $C_{15}H_9O_5^-$). The compound apigenin was identified by two adduct ions, M-H ($C_{15}H_{10}O_5^-$, m/z 269.0699) and 2M-H ($C_{30}H_{18}O_{10}^-$, m/z 539.0989), and the fragment ion was identified as $C_7H_3O_4^-$ (m/z 151.0039). We further demonstrated the reliability of Progenesis QI identification with reference substances. In the negative ion mode, most of the compounds exhibited deprotonation. The most common compounds were -H_2O and -CO_2. Compounds containing carboxylic acid groups commonly lost CO_2 by negative ionization. Flavonoid glycosides were mainly identified by the loss of character classes of molecules, including the loss of CO and the loss of molecules, by cleavage through the RDA reaction. The loss of sugars such as rhamnose, arabinose, and glucuronide mainly occurred in flavonoid glycosidic compounds. However, similar to other plant species, the species of *Scutellaria* also have specific metabolites, which may contain many isomeric compounds. Given their similar or identical molecular weights, mass spectrometry fragmentation behavior, and minor differences in retention times, it remains challenging to resolve these structural isomers using only mass spectrometry-based metabolomics approaches. Therefore, when identifying compounds by searching for isomers online with Progenesis QI, this study identified compounds based on their polarity, ion fragmentation patterns, and UV absorption wavelengths and referred to literature reports for further confirmation. For example, compounds with retention times of 8.52 min, M-H^- ($C_{18}H_{15}O_7^-$) ions, and 343.0809 m/z were tentatively identified as nevadensin (Liang et al., 2018). The major fragment ions of nevadensin were 241.0496 (M-2CH_2O-CH_2-CO), 313.0352 (M-CH_3), and 328.0572 (M-H-CH_3). All compounds were identified by the same method. Finally, a total of 75 compounds were putatively identified in negative ion mode, 16 of which were identified by Progenesis QI in agreement with the results of the standards comparison. These compounds included 1 aromatic aldehyde, 2 phenylethanol glycosides, 4 pyrone glycosides, 9 diterpenoids, 55 flavonoids, and 4 other compounds.

Principal component analysis was first applied to analyze the chemical diversity of different parts among the tested *Scutellaria* species (**Figure 3A**). In this model, differences in the chemical composition of the five *Scutellaria* species were observed. The three substitute species other than *S. likiangensis* (*S. amoena, S. hypericifolia*, and *S. viscidula*) were similar in distance from *S. baicalensis* for both the aerial parts and roots, but *S. likiangensis* was more distant. In addition, the distance between *S. hypericifolia* and *S. amoena* was shorter for aerial parts.

Furthermore, partial least squares discriminant analysis (PLS-DA) was used to better describe the different chemical metabolic profiles of *S. baicalensis* and its four substitute species in different parts (**Figure 3**). The variable importance in the project (VIP) value reflects the influence of metabolites on classification. Therefore, metabolites with VIP values ranking in the top 10 were selected as potential markers to distinguish species (**Figures 3B,C**). In the aerial parts, the PLS-DA loading plot revealed that the separation of the species was mainly caused by pyrone glycoside, flavonoids, and diterpenes. Moreover, the compound scutevulin-2′-*O*-β-D-glucuronopyranoside (**40**) showed the highest relative content in *S. amoena*, the compounds apigenin-6-C-glucoside-8-C-arabinoside (**15**) and isoscutellarein-8-glucuronide (**47**) were more abundant in *S. baicalensis*, and the compounds luteolin-4′-*O*-glucoside (**42**) and chrysin-7-*O*-D-glucuronopyranoside (**50**) displayed the highest content in *S. hypericifolia*. The compounds isocarthamidin-7-*O*-D-glucuronide (**17**) and 5,7,2′-trihydroxy-6-methoxyflavanone-7-*O*-β-D-glucuronide (**35**) were more abundant in *S. likiangensis*, and the compounds scusalvioside A (**1**), scutebarbatine H (**52**) and unknown (**81**) demonstrated relatively high contents in *S. viscidula*. In the PLS-DA plot for the roots, *S. amoena* showed relatively high concentrations of baicalein (**63**), unknown (**20**), chrysin 6-C-alpha-L-arabinopyranoside-8-C-glucoside (**27**), 5,7,2′-trihydroxy-6′-methoxyflavone (**67**), and unknown (**36**); *S. hypericifolia* showed the highest concentration of luteolin-4′-*O*-glucoside (**42**); *S. likiangensis* showed the highest concentration of chrysin-8-C-D-glucopyranoside (**30**); and *S. viscidula* displayed the highest concentrations of isocarthamidin-7-*O*-D-glucuronide (**17**) and unknown (**28, 54**). The unknown metabolites **20**, **28, 36, 54,** and **81** were not annotated because their m/z and possible molecular formulas did not match any of the previously reported compounds.

Quantitative determination of the 15 assessed chemical components in the *Scutellaria* species

The HPLC analytical method described in our previous reports (Shen et al., 2019) was used to simultaneously quantify 15 compounds in every sample. The analytical results are shown in **Figure 4** and **Supplementary Table S3**. In the aerial parts

TABLE 1 Compounds putatively identified from the *Scutellaria* species.

No.	t_R (Min)-m/z (n)	Compound type	Identification	Adduct ion	Molecular formula	UV λ max (nm)	δ (ppm)	$[M-H]^-$ measured	$[M-H]^-$ predicted	Key MS^E fragment ions (Da)
1	2.16_303.0720m/z	Pyrone glycoside	Scusalvioside A	M-H	$C_{12}H_{16}O_9$	290	−0.375	303.0720	303.0716	174.9530
2	2.52_462.1711n	Pyrone glycoside	2-(3,4-Dihydroxyphenyl) ethyl 3-*O*-(6-Deoxy-α-L-mannopyranosyl)-β-D-glucopyranosi	M-H, M+Cl	$C_{20}H_{30}O_{12}$	289	−5.7623	461.1638	461.1659	341.0840, 447.1124
3	2.66_465.1020m/z	Pyrone glycoside	Scusalvioside B	M-H	$C_{21}H_{22}O_{12}$	289	−3.9333	465.102	465.1033	359.0962, 403.1203
4	2.76_300.1187n	Phenylethanoid glycosides	Salidroside	M-H, M+FA-H	$C_{14}H_{20}O_7$	279	−7.2744	345.117	345.1186	299.1122, 174.9557
5	3.05_464.0940n	Flavonol	5,7,2′6′-Tetrahydroxyflavonol-2′-*O*-D-glucopyranside	M-H, 2M-H	$C_{21}H_{20}O_{12}$	278	−3.1591	463.0867	463.0877	287.0537
6	3.25_325.0911m/z	Others	(*E*)-3-[2-[(2*S*,3*R*,4*S*,5*S*,6*R*)-3,4,5-trihydroxy-6-(hydroxymethyl) oxan-2-yl] oxyphenyl] prop-2-enoic acid	M-H	$C_{15}H_{18}O_8$	272	−5.3478	325.0911	325.0923	163.0389, 145.0300, 119.0491
7	3.29_465.1021m/z	Flavononol	Amoenin D	M-H	$C_{21}H_{22}O_{12}$	289	−3.7262	465.1021	465.1033	347.0784, 285.0387
8	3.39_303.0501m/z	Flavonoid	(2*R*,3*R*)-3,5,7,2′6′-pentahydroxyflavanone	M-H	$C_{15}H_{12}O_7$	288	−3.0489	303.0501	303.0505	285.0378, 259.0575, 77.0169
9	3.54_449.1081m/z	Flavonoid	2,2′-Dihydroxy-3,4,5,6′-tetramethoxy-4′5′-Methylenedioxychalcone	M-H	$C_{21}H_{22}O_{11}$	281	−1.9534	449.1081	449.1084	287.0534
10	3.61_121.0285m/z	Aromatic aldehydes	2-Hydroxybenzaldehyde*	M-H	$C_7H_6O_2$	284	−8.0915	121.0285	121.0290	122.0368
11	3.65_476.1875n	Phenylethanoid glycosides	Darendoside B	M-H, M+Cl, M+K-2H, M+FA-H, 2M-H	$C_{21}H_{32}O_{12}$	–	−3.9779	475.1802	475.1816	387.1263, 341.1218, 235.9249
12	3.88_480.0892n	Pyrone glycoside	Gossypetin-8-C-glucoside	M-H, 2M-H	$C_{21}H_{20}O_{13}$	290, 347	−2.5323	479.0819	479.0819	303.0514, 167.0327
13	3.92_593.1501m/z	Flavonoid	Apigenin-6,8-digalactoside	M-H	$C_{27}H_{30}O_{15}$	322	−1.8072	593.1501	593.1507	473.1048, 353.0662
14	4.11_449.1075m/z	Flavonoid	(*cis*)-5,7,2′-Trihydroxyflavanonol-3-*O*-β-D-glucopyranoside	M-H	$C_{21}H_{22}O_{11}$	282	−3.179	449.1075	449.1084	287.0552
15	4.31_563.1398m/z	Flavonoid	Apigenin-6-C-glucoside 8-C-arabinoside	M-H	$C_{26}H_{28}O_{14}$	271, 333	−1.4863	563.1398	563.1401	473.1092, 353.0658
16	4.50_449.1078m/z	Flavonoid	Flavanomarein	M-H	$C_{21}H_{22}O_{11}$	282, 333	−2.5914	449.1078	449.1084	287.0558, 269.0471
17	4.50_463.0880m/z	Flavonoid	Isocarthamidin-7-*O*-D-glucuronide*	M-H	$C_{21}H_{20}O_{12}$	246, 283	−0.4871	463.088	463.0877	285.0387
18	4.80_450.1150n	Dihydroflavones	Isookanin-7-*O*-glucoside	M-H2O-H, M+Na-2H	$C_{21}H_{22}O_{11}$	289, 334	−2.6806	431.0972	431.0978	311.0541, 174.9557
19	4.84_464.0939n	Dihydroflavones	Carthamidin-7-*O*-D-glucuronide*	M+Na-2H, 2M-H, 2M+Hac-H, M-H	$C_{21}H_{20}O_{12}$	246, 283, 347	−3.4338	463.0884	463.0877	287.0531

(Continued)

10.3389/fpls.2022.951824

TABLE 1 Continued

No.	t_R (Min)-m/z (n)	Compound type	Identification	Adduct ion	Molecular formula	UV λ max (nm)	δ (ppm)	[M-H]$^-$ measured	[M-H]$^-$ predicted	Key MS^E fragment ions (Da)
20	4.87_244.0658n		Unknown							
21	5.09_432.1076n	Flavonol	Kaempferol-7-*O*-rhamnoside	M-H2O-H, M-H	$C_{21}H_{20}O_{10}$	282, 333	4.4983	431.0973	431.0978	285.0401
22	5.09_462.0788n	Flavonoid	Scutellarin*	M-H, 2M-H	$C_{21}H_{18}O_{12}$	280, 331	−2.1404	461.0716	461.072	283.0239
23	5.13_285.0393m/z	Flavone	5,7,2′,5′-Tetrahydroxyflavone	M-H	$C_{15}H_{10}O_6$	273	−4.0318	285.0393	285.0399	271.0351, 243.0292, 227.0345
24	5.26_581.1913m/z	Flavononol	Amoenin A	M-H	$C_{27}H_{34}O_{14}$	287	6.4136	581.1913	581.1871	551.1733, 433.1083, 285.0387, 174.9557
25	5.33_447.0965m/z	Flavonoid	Luteolin-7-*O*-D-glucopyranside*	M-H	$C_{21}H_{20}O_{11}$	282, 347	7.2842	447.0965	447.0928	271.0615
26	5.33_478.1097n	Flavonoid	5-(5,7-Dihydroxy-3-methoxy-4-oxo-4H-chromen-2-yl)-2-hydroxyphenyl β-D-glucopyranoside	M-H, M+Na-2H	$C_{22}H_{22}O_{12}$	285, 334	−2.9882	477.1024	477.1033	301.0693, 271.0615, 174.9557
27	5.36_547.1452m/z	Flavonoid	Chrysin 6-C-*alpha*-L-arabinopyranoside-8-C-glucoside	M-H	$C_{26}H_{30}O_{14}$	272, 319	−0.8791	547.1452	547.1452	427.1052, 367.0847, 337.0724
28	5.43_310.1041n		Unknown							
29	5.49_638.2191n	Flavonoid	Leucosceptoside A	M-H, M+K-2H	$C_{30}H_{38}O_{15}$	275, 325	−3.0709	637.2118	637.2133	561.1579, 437.0855
30	5.55_415.1022m/z	Flavonoid	Chrysin-8-C-D-glucopyranside	M-H	$C_{21}H_{20}O_9$	279, 329	−3.0133	415.1022	415.1029	295.0581, 267.0623
31	5.59_419.1329m/z	Others	2,3,8,9,10-Pentamethoxy-6a,11a-Dihydro-6H-[1] benzofuro[3,2-c]chromene	M+FA-H	$C_{20}H_{22}O_7$	302	−5.0173	419.1329	419.1342	285.0395
32	5.62_462.0786n	Flavonoid	Kaempferol-3-glucuronide	M-H, M+Na-2H	$C_{21}H_{18}O_{12}$	288, 322	−2.6553	461.0713	461.072	285.0387
33	5.65_432.1050n	Flavonoid	Apigenin-7-*O*-glucoside*	M-H, 2M-H	$C_{21}H_{20}O_{10}$	280, 335	−1.4329	5.6545	431.0978	367.0850, 309.0404
34	5.69_288.0617n	Flavonoid	(2*S*) 5,7,2′6′-Tetrahydroxyflavanone	M-H, M+Na-2H, 2M-H, 2M+Hac-H	$C_{15}H_{12}O_6$	288	−5.7107	287.0545	287.0556	125.0236
35	5.69_478.1095n	Flavonoid	5,7,2′-Trihydroxy-6-methoxyflavanone-7-*O*-β-D-glucuronoide	M-H, M+Na-2H, 2M-H	$C_{22}H_{22}O_{12}$	285, 347	−3.3199	477.1023	477.1033	301.0706
36	5.69_547.1441m/		Unknown							
37	6.11_446.0844n	Flavonoid	Norwogonin-7-*O*-D-glucuronopyranoside	M-H, M+Na-2H	$C_{21}H_{18}O_{11}$	288	−1.2162	445.0771	445.0771	291.0272
38	6.15_522.1367n	Flavonoid	(2*S*)5,2-Dihydroxy-7,8,6′-trimethoxyflavanone-2-*O*-D-glucuronide	M-H, M+Na-2H	$C_{24}H_{26}O_{13}$	279, 333	−1.2292	543.0775	543.1115	413.0862, 287.0537
39	6.25_302.0775n	Others	Hematoxylin	M-H, 2M-H	$C_{16}H_{14}O_6$	293	−5.2188	301.0701	301.0712	299.0553, 205.0492, 119.0491
40	6.25_475.0875m/z	Flavonoid	Scutevulin 2′-*O*-β-D-glucuronopyranoside	M-H	$C_{22}H_{20}O_{12}$	272, 324	−1.3927	475.0875	475.0877	321.0374, 299.0553

(Continued)

10.3389/fpls.2022.951824

TABLE 1 Continued

No.	t_R (Min)-m/z (n)	Compound type	Identification	Adduct ion	Molecular formula	UV λ max (nm)	δ (ppm)	$[M-H]^-$ measured	$[M-H]^-$ predicted	Key MS^E fragment ions (Da)
41	6.39_652.2351n	Others	[5-Hydroxy-6-[2-(4-hydroxy-3-methoxyphenyl) ethoxy]-2-(hydroxymethyl)-4-(3,4,5-trihydroxy-6-methyloxan-2-yl) oxyoxan-3-yl] (E)-3-(4-hydroxy-3-methoxyphenyl) prop-2-enoate	M-H, M+Cl, M+FA-H, M-H2O-H	$C_{31}H_{40}O_{15}$	288	−2.5089	651.2278	651.2289	475.1784, 175.0397
42	6.42_448.0991n	Flavonoid	Luteolin-4′-*O*-glucoside	M-H, M+Na-2H, M+K-2H, 2M+Hac-H, 2M-H	$C_{21}H_{20}O_{11}$	282	−3.177	447.0919	447.0928	271.0615
43	6.54_467.0583m/z	Flavonoid	Baicalin*	M+Na-2H	$C_{21}H_{18}O_{11}$	277, 314	−2.7721	467.0583	467.059	239.0354, 269.0431
44	6.64_286.0465n	Flavonoid	Scutellarein*	M-H, M+K-2H, 3M-H	$C_{15}H_{10}O_6$	281, 322	−4.4541	285.0399	285.0399	117.0346
45	6.68_415.1024m/z	Flavonoid	Puerarin	M-H	$C_{21}H_{20}O_9$	275, 347	−2.4492	415.1024	415.1029	325.0722, 295.0615, 267.0652
46	6.71_433.0766m/z	Flavonoid	Guajavarin	M-H	$C_{20}H_{18}O_{11}$	273, 332	−2.3514	433.0766	433.0771	301.0353
47	6.78_462.0795n	Flavonoid	Isoscutellarein 8-glucuronide*	M-H, M+Na-2H, 2M-H	$C_{21}H_{18}O_{12}$	288, 333	−0.8026	461.0721	461.072	285.0387, 174.9557
48	7.00_359.0763m/z	Flavonoid	5,6,2′-Trihydroxy-7,8,6′-trimethoxyflavone	M-H	$C_{18}H_{16}O_8$	279	−2.6723	359.0763	359.0767	299.0551
49	7.10_452.2213n	Diterpenes	Scutebata J	M-H, M+Cl, M+FA-H	$C_{27}H_{32}O_6$	–	3.227	497.2271	497.2175	371.1123
50	7.24_430.0906n	Flavonoid	Chrysin-7-*O*-D-glucuronopyranoside*	2M-H, 2M+FA-H, M-H, M+FA-H	$C_{21}H_{18}O_{10}$	271, 309	1.4184	429.0829	429.0822	253.0500, 174.9557
51	7.37_857.1339m/z	Flavonoid	Luteolin*	3M-H	$C_{15}H_{10}O_6$	256, 348	−2.0835	857.1339	857.1354	285.0406
52	7.46_467.2127m/z	Diterpenes	Scutebarbatine H	M-H	$C_{27}H_{32}O_7$	–	11.0066	467.2127	467.2070	–
53	7.46_475.0900m/z	Flavonoid	Hispidulin-7-*O*-D-glucuronide	M-H	$C_{22}H_{20}O_{12}$	273, 313	3.8674	475.09	475.0877	342.0745, 314.0643, 283.0606
54	7.56_230.0500n		Unknown							
55	7.56_283.0597m/z	Flavonoid	Acacetin	M-H	$C_{16}H_{12}O_5$	–	−5.1247	283.0597	283.0607	268.0370, 163.0025
56	7.56_503.2461m/z	Diterpenes	Scutolide C	M-H2O-H	$C_{31}H_{38}O_7$	–	4.1236	503.246	503.2434	462.2048, 225.0558
57	7.60_459.0952m/z	Flavonoid	Wogonoside*	M-H	$C_{22}H_{20}O_{11}$	276	−1.3095	459.0952	459.0928	283.0617
58	7.83_467.0581m/z	Flavonoid	Apigenin-7-*O*-glucuronide	M+Na-2H	$C_{21}H_{18}O_{11}$	267, 315	−3.257	467.0581	467.059	445.0802, 313.0707
59	8.12_270.0519n	Flavonoid	Apigenin*	M-H, 2M-H	$C_{15}H_{10}O_5$	267, 339	−3.5697	269.0445	269.045	227.0343, 143.0473, 117.0339,
60	8.19_300.0620n	Flavonoid	5,7,2′-Trihydroxy-6′-methoxyflavone	M-H, M+K-2H	$C_{16}H_{12}O_6$	281	−4.7514	299.0547	299.0556	269.0447, 151.0036, 143.0473, 117.0339

(Continued)

10.3389/fpls.2022.951824

TABLE 1 Continued

No.	t_R (Min)-m/z (n)	Compound type	Identification	Adduct ion	Molecular formula	UV λ max (nm)	δ (ppm)	$[M-H]^-$ measured	$[M-H]^-$ predicted	Key MS^E fragment ions (Da)
61	8.26_270.0888n	Flavonoid	Alpinetin*	M-H, 2M-H	$C_{16}H_{14}O_4$	286	−1.4766		269.0814	227.0776, 184.0518
62	8.29_493.2254m/z	Diterpenes	Scutolide E	M-H	$C_{29}H_{34}O_7$	–	−4.5466	493.2254	493.2227	447.2200
63	8.52_270.0528n	Flavonoid	Baicalein*	M-H, M+Na-2H, M+FA-H, 2M-H, 2M+Hac-H, M-H2O-H	$C_{15}H_{10}O_5$	275, 324	0.0951	269.0455	269.045	239.0337, 195.0444
64	8.52_285.0388m/z	Flavonoid	3,6,3′,4′-Tetrahydroxyflavone	M-H	$C_{15}H_{10}O_6$	274, 322	−5.833	285.0388	285.0399	239.0342, 211.0396, 195.0444
65	8.52_343.0809m/z	Flavonoid	Nevadensin	M-H	$C_{18}H_{16}O_7$	272, 343	−4.1316	343.0809	343.0818	313.0352, 269.0456, 232.0572, 141.0496
66	8.55_299.0545m/z	Flavonoid	Scutellarein 4′-methyl ether	M-H	$C_{16}H_{12}O_6$	280	−5.3997	299.0545	299.0556	269.0450
67	8.94_299.0550m/z	Flavonoid	5,7,2′-Trihydroxy-6′-methoxyflavone	M-H	$C_{16}H_{12}O_6$	270, 335	−3.8571	299.055	299.0556	284.032
68	8.98_374.1003n	Flavonoid	Skullcapflavone II	M-H, M+Cl, M+K-2H, 2M-H	$C_{19}H_{18}O_8$	281	0.3282	373.093	373.0924	373.093, 358.0685, 328.0220
69	9.31_255.0657m/z	Flavonoid	Pinocembrin*	M-H	$C_{15}H_{12}O_4$	289, 348	−2.2057	255.0657	255.0658	169.0648
70	9.38_283.0618m/z	Flavonoid	Wogonin*	M-H	$C_{16}H_{12}O_5$	276	−1.3095	283.0617	283.0618	163.0036
71	9.71_254.0586n	Flavonoid	Chrysin*	M-H, M+K-2H, M+Cl, 2M-H, 3M-H	$C_{15}H_{10}O_4$	268, 313	2.5939	254.0586	254.0579	209.0603, 143.0502, 107.0122
72	9.77_283.0604m/z	Flavonoid	Oroxylin A*	M-H	$C_{16}H_{12}O_5$	272, 320	−2.6747	283.0604	283.066	239.0354, 165.9893, 109.9997
73	10.06_538.0890n	Flavonoid	8,8′-Bibaiaclein	M-H, M+K-2H	$C_{30}H_{18}O_{10}$	277	−1.9444	537.0817	537.0822	391.0448, 373.0354
74	10.17_581.2945m/z	Diterpenes	Scutefolide F	M+FA-H	$C_{29}H_{44}O_9$	–	−4.2526	581.2945	581.2962	535.2905
75	10.23_343.0813m/z	Flavonoid	5,2′-Dihydroxy-6,7,8-trimethoxyflavone	M-H	$C_{18}H_{16}O_7$	–	−2.8737	343.0813	343.0818	328.0576
76	10.46_459.2012m/z	Diterpenes	Scutegalerin E	M+Na-2H	$C_{23}H_{34}O_8$	–	2.7113	459.2012	459.1995	365.1947
77	10.76_406.3089n	Diterpenes	5-[1,2,4a-Trimethyl-5-(3-methylbutanoyloxymethyl)-2,3,4,7,8,8a-hexahydronaphthalen-1-yl]-3-methylpentanoic acid	M+FA-H, 2M+FA-H, M+Cl	$C_{25}H_{42}O_4$	–	1.4408	451.3071	451.306	405.1945, 117.055
78	10.80_405.1919m/z	Diterpenes	Scutebarbolide C	M-H2O-H	$C_{22}H_{32}O_8$	–	−0.0388	405.1919	405.1913	311.1657, 301.1785
79	11.19_329.1746m/z	Diterpenes	Barbatin C	M-H2O-H	$C_{20}H_{28}O_5$	–	−3.4979	329.1746	329.1753	329.1744, 285.1492, 139.1102
80	11.55_492.1656n	Diterpenes	Scutellarioside I	M-H, M+Na-2H, M-H2O-H	$C_{24}H_{28}O_{11}$	–	4.8679	491.1583	491.1554	513.1390, 491.1558, 445.1524
81	12.99_476.3382n		Unknown							

*Identifications were confirmed by comparing t_R and MS spectra to standard compounds.

10.3389/fpls.2022.951824

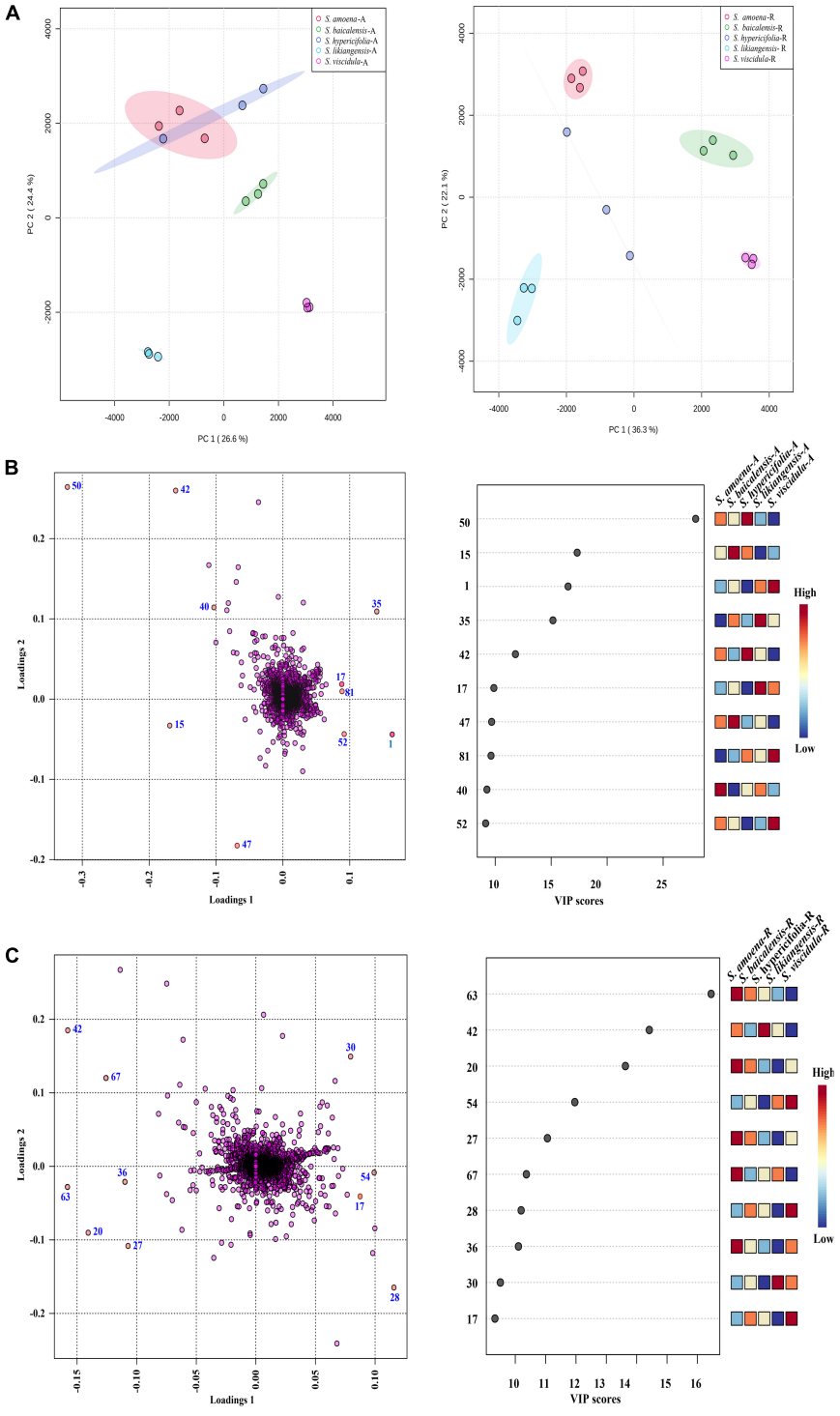


FIGURE 3
The multivariate analysis of the metabolite data was derived from five *Scutellaria* species. **(A)** PCA score map of the different parts of *Scutellaria* species (-A: aerial parts, -R: roots). **(B)** The loading plot and variable importance in the projection of the PLS-DA model for the aerial parts of *Scutellaria* species. **(C)** The loading plot and variable importance in the projection of the PLS-DA model for the roots of *Scutellaria* species. (The colored boxes on the right indicate the relative concentrations of the corresponding metabolite in each group under study; the number of the metabolites corresponds to **Table 1**).

of five *Scutellaria* species, the contents of isocarthamidin-7-*O*-D-glucuronide (**17**), carthamidin-7-*O*-D-glucuronide (**19**), scutellarin (**22**), apigenin-7-*O*-β-D-glucopyranoside (**33**), isoscutellarein-8-glucuronide (**47**), and apigenin (**59**) were higher than those in roots. Of these, compounds **17, 19, and 22** were higher in the aboveground parts of the five species of *Scutellaria*. In addition, the contents of some compounds were consistent with the relative abundances of compounds identified in the PLS-DA pattern. In the aerial parts, the concentration of isocarthamidin-7-*O*-D-glucuronide (**17**) was significantly lower in *S. hypericifolia* (53.6195 ± 18.9497 mg/g) and that of isoscutellarein-8-glucuronide (**47**) was significantly higher in *S. baicalensis* (5.1336 ± 1.6451 mg/g). In addition, carthamidin-7-*O*-D-glucuronide (**19,** 120.5610 ± 5.0985 mg/g) and scutellarin (**22,** 26.5083 ± 3.4493 mg/g) displayed the highest contents in *S. likiangensis*. Apigenin-7-*O*-β-D-glucopyranoside (**33**) showed the lowest content (1.6602 ± 0.2706 mg/g) in *S. viscidula*. Apigenin (**59**) displayed a relatively high content in *S. hypericifolia* (0.0802 ± 0.0175 mg/g).

Nine compounds, baicalin (**43**), chrysin-7-*O*-D-glucuronopyranoside (**50**), luteolin (**51**), wogonoside (**57**), alpinetin (**61**), baicalein (**63**), wogonin (**70**), chrysin (**71**), and oroxylin A (**72**), were more abundant in the roots of these five *Scutellaria* species. In addition, the contents of chrysin-7-*O*-D-glucuronopyranoside (**50**), chrysin (**71**), and oroxylin A (**72**) in *S. likiangensis* were higher at 69.7029 ± 19.0448, 1.6112 ± 0.6629, and 2.7468 ± 0.6434 mg/g, respectively. Wogonoside (**57**), alpinetin (**61**), baicalein (**63**), and wogonin (**70**) were more abundant in *S. baicalensis,* with contents of 31.6445 ± 0.9172, 1.8570 ± 0.1488, 14.6276 ± 0.9966, and 0.0077 ± 0.0021 mg/g, respectively. The compounds baicalin (**43**) and luteolin (**51**) were more abundant in *S. amoena*. Furthermore, the dried root of *S. baicalensis* (Scutellariae Radix) is a typical Chinese materia medica. According to the Chinese pharmacopoeia (2020 edition), the content of baicalin in Scutellariae Radix should not be less than 9.0%. In this study, in contrast to the content in *S. likiangensis* (50.1253 ± 14.5487 mg/g), the contents of baicalin (**43**) were higher than 9.0% in the roots of *S. amoena* (148.8557 ± 47.1488 mg/g), *S. baicalensis* (130.6066 ± 9.3707 mg/g), *S. hypericifolia* (96.1738 ± 28.3923 mg/g), and *S. viscidula* (93.5047 ± 8.3265 mg/g) (**Supplementary Table S3**).

Discussion

In this study, the complete cp genomes of four species of *Scutellaria* (*S. amoena*, *S. hypericifolia*, *S. likiangensis*, *S. viscidula*) were assembled for the first time, and comparative and phylogenetic analyses with other related *Scutellaria* species were also performed (Lee and Kim, 2019; Liu et al., 2020; Zhao et al., 2020). The four *Scutellaria* cp genomes have high similarities in genome structure and size, gene number, and the distribution of repeat sequences. Furthermore, one high-frequency sequence variation intergenic region (*petA-psbJ*) was found among the above four *Scutellaria* species and *S. baicalensis*. The fragment *petA-psbJ* has been used in previous phylogenetic studies (Latvis et al., 2017) but not in Lamiaceae. *petA-psbJ* may represent a potential barcode specific to these *Scutellaria* species that are currently available. In addition, the phenomena of expansion and contraction of the IR region of the among the species of *Scutellaria* have been observed. Changes in the boundaries of LSCs and IRs are relatively common in plants (Li et al., 2018; Yan et al., 2019; Zhao et al., 2020). Similar to the findings of Zhao et al. (2020), the changes in the LSC region were larger, which indicates that divergence in LSC length leads to variation in the size of the cp genomes based on IR contraction or expansion. During the evolution of cp genomes, differences in the four IR boundaries among species have been frequently observed, leading to further variation in cp genome size. Thus, although IR regions are the most conserved regions in the cp genome sequence, they have been used to explain differences in size among cp genomes due to contraction and expansion at their boundaries (Liu et al., 2019; Xue et al., 2019). In addition, the location of the IR boundary can be used to study phylogeny; in closely related species, small expansions or contractions tend to have similar endpoints (Cui et al., 2019b).

Morphological polymorphisms make the classification of *Scutellaria* extremely difficult, and it is also challenging to define species, groups, and phylogenetic relationships within the genus. *Scutellaria* is a very isolated genus with unsatisfactory traditional divisions. Currently, the commonly used classification method in China is based on the *Flora of China* (1977) (Wu and Li, 1977). In this study, phylogenetic analyses showed that the cp-based phylogenetic tree and traditional plant classification of *Scutellaria* in China were incongruent. According to the *Flora of China* (1977), the species of *S. hypericifolia*, *S. baicalensis*, and *S. viscidula* are grouped, while in the cp-based phylogenetic tree, *S. hypericifolia* is clustered with *S. likiangensis* and *S. viscidula*. In addition, it was observed that *S. baicalensis* and its substitute species may be a special group in the genus *Scutellaria*.

Phylogenetically closely related species often share similar chemical characteristics and even similar clinical efficacy, which is the core idea of pharmacophylogeny (Gong et al., 2020; Hao and Xiao, 2020). The 17 species of genus *Scutellaria* involved in the cp genome phylogenetic tree are classified into two main branches, where the presence or absence of a well-developed root is a significant morphological difference between Subclade I and Subclade II. Subclade I can be divided into two clades, Subclade I_1 and Subclade I_2. Subclade I_1 contains the species *S. altaica* and *S. przewalskii*, both of which are perennial subshrubs with developed woody rhizomes and no phytochemical studies. In Subclade I_2, *S. kingiana* is a dwarf perennial herb and has not been chemically studied. The species of *S. amoena*, *S. baicalensis*, *S. hypericifolia*, *S. likiangensis,* and

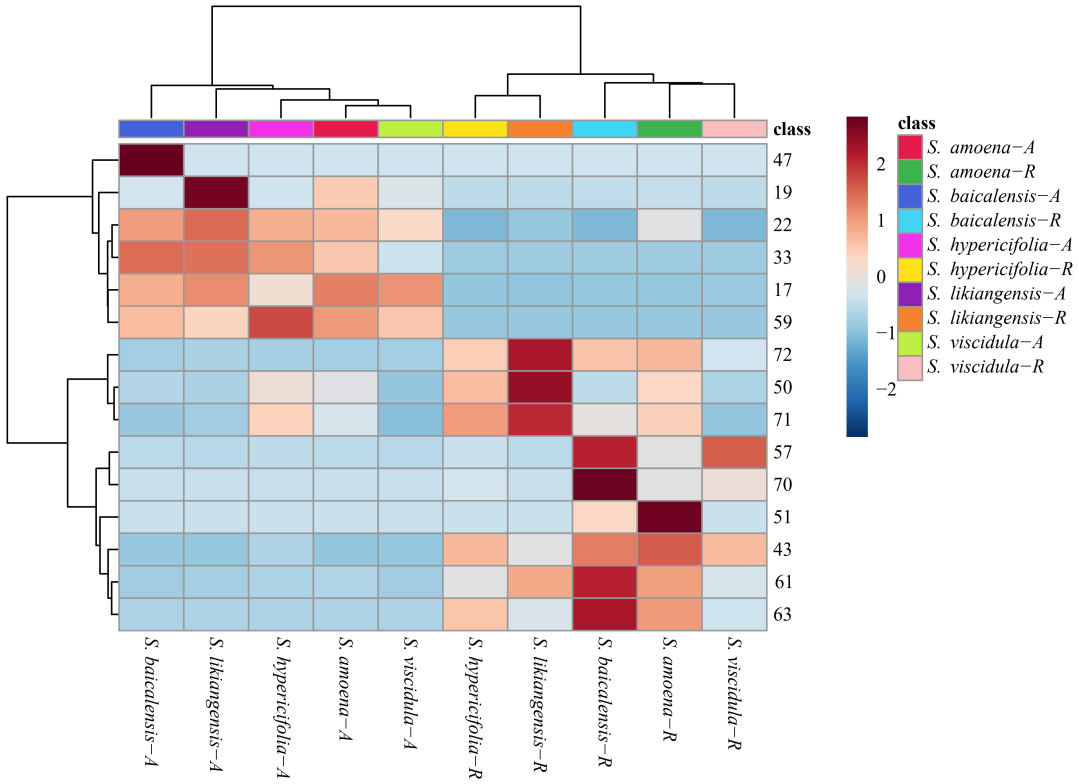


FIGURE 4
Clustering heatmap of 15 compounds in different parts of five *Scutellaria* species (the number of analytes corresponds to **Table 1**; -A: aerial parts, -R: roots).

S. viscidula are clustered together and are all erect or ascending erect perennial herbs with fleshy rhizomes. As in previous reports of *S. baicalensis*, there are significant differences in the chemical composition between their aerial parts and roots (Shen et al., 2019), while both are reported to have various benefits for human health, such as anti-aging, anti-inflammation, and antitumor activities, which are mainly attributed to flavonoids (Gao et al., 2016; Zhang et al., 2017; Shen et al., 2019, 2020; Song et al., 2020). In this study, we further explored the chemical characteristics of different parts (aerial parts and roots) of these species. The four substitute medicinal species and *S. baicalensis* share similarity in chemical composition both in aerial parts and roots. It is reported that the 15 accessed chemical components were the active or main compounds in the species of *S. baicalensis*. Among these compounds, **19, 22, 33, 43, 47, 50, 51, 57,** and **59** were all detected in the aerial parts of the four substitute medicinal species for *S. baicalensis*. The compounds isocarthamidin-7-*O*-D-glucuronide (**17**), scutellarin (**22**), and apigenin (**59**) showed higher contents and various activities (Li et al., 2017; Wang and Ma, 2018; Oaarowski et al., 2019) in the aerial parts. Moreover, the content of scutellarin (**22**, 26.5083 ± 3.4493) was highest in the aerial parts of *S. likiangensis*, which could be a good choice as a natural source of scutellarin for making cerebrovascular and cardiovascular drugs (Breviscapin) (Wen et al., 2021). In particular, it has been reported that the pharmacological activities of Scutellariae Radix are closely related to flavones lacking 4'-OH groups on the B-rings (Zhao et al., 2018; Song et al., 2020; Sun et al., 2020), such as baicalin (**43**), wogonoside (**57**), baicalein (**63**), wogonin (**70**), chrysin (**71**), and oroxylin A (**72**). The roots of the four substitute species all contain these special compounds. The compounds in the roots of *S. amoena* were similar to those in the roots of *S. baicalensis*. In addition, according to the standard of the Chinese pharmacopoeia (baicalin > 9%), *S. likiangensis* could not be used as a substitute species for *S. baicalensis*. Subclade II consists of nine species: *S. calcarata, S. lateriflora, S. indica* var. *coccinea, S. insignis, S. mollifolia, S. orthocalyx, S. quadrilobulata, S. tsinyunensis,* and *S. scordifolia*, and all of them have short roots and are small plants. Among them, only *S. lateriflora* and *S. scordifolia* have had limited chemical studies. *S. lateriflora* contains both flavonoids and diterpenoids, and only flavonoids have been reported for *S. scordifolia*. Currently, flavonoids and diterpenes are thought to be the two main active constituent groups in the genus *Scutellaria*. The relationship of patterns of major chemical diversity to the known phylogeny of these *Scutellaria* species are still unclear due to the lack

of adequate studies on phylogeny, chemical composition and constituent contents. In further studies, additional evidence is needed to confirm the relationship between the chemical diversity and the known phylogeny in *Scutellaria*.

The species *S. amoena*, *S. hypericifolia*, *S. likiangensis,* and *S. viscidula* have been traditionally used for a long time as substitutes for *S. baicalensis* in their places of origin. In recent years, this substitution has gradually decreased due to the expansion of areas of artificial cultivation of *S. baicalensis* and declines in the availability of wild resources. However, as these substitutes for *S. baicalensis* have been included in the local standards, long-term use habits are not applied differently. Considering the differences in geographical distribution, growth environment, and species genetics, our attention to the possible nuances in efficacy is still needed. Although the present study revealed that their primary chemical composition is similar, there are differences in the types of compounds and relative contents of the main active ingredients. Therefore, the similarities and differences in their pharmacological activities still need to be further investigated to ensure the validity and stability of the quality of the medicinal herbs.

Conclusion

In this study, the complete cp genomes of four species of *Scutellaria* (*S. amoena, S. hypericifolia, S. likiangensis,* and *S. viscidula*) were sequenced by high-throughput sequencing. In addition, the complete cp genomes and chemical analysis of *S. hypericifolia, S. likiangensis,* and *S. viscidula* are reported for the first time. Through comparison, we found that the cp genomes of these four species have similar structural characteristics and the typical four-part structure of other land plants. The nucleotide variation analysis showed that the mutational hot spot of *petA-psbL* could be used as an effective molecular marker for distinguishing *S. baicalensis* from its substitutes (*S. amoena* and *S. hypericifolia, S. likiangensis,* and *S. viscidula*). The phylogenetic analysis showed that the traditional morphological classification method in China is different from the cp-based phylogenetic tree, which suggests that the subdivision of the genus *Scutellaria* should be reconsidered. Additionally, the chemical distributions of four *Scutellaria* species determined using UPLC-QTOF-MS and HPLC analyses revealed that each *Scutellaria* plant exhibits a chemical metabolism similar to that of *S. baicalensis*. Further investigations are needed to reveal the links between the genetic and chemical relationships of *Scutellaria* species; however, information on these four complete cp genomes of *Scutellaria* will provide a valuable medicinal genetic resource for *Scutellaria* species authentication and phylogenetic analysis, and this comparative study will be a basis for further studies on the genus *Scutellaria.*

Data availability statement

The data presented in this study are deposited in GenBank, accession numbers MW286269–MW286272.

Author contributions

JS wrote the manuscript. CH and PX systemically revised the manuscript for important content. JS, YW, and PL helped to complete the data analysis. JS, YL, and KY completed the figures and tables. QW helped to collect and identify the samples. HY and QW reviewed and edited the original manuscript. CH proposed the conception. JS designed the structure of the manuscript. All authors read and approved the final manuscript.

Funding

This research was funded by the Innovation Team and Talents Cultivation Program of the National Administration of Traditional Chinese Medicine (Grant No: ZYYCXTD-D-202005) and the CAMS Innovation Fund for Medical Sciences (CIFMS) ID: 2021-I2M-1–071. The Major Scientific and Technological Special Project for The Drug Innovation Major project (No. 2018ZX09711001-008).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary Material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.951824/full#supplementary-material

References

Amiryousefi, A., Hyvonen, J., and Poczai, P. (2018). IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34, 3030–3031. doi: 10.1093/bioinformatics/bty220

Beier, S., Thiel, T., Munch, T., Scholz, U., and Mascher, M. (2017). MISA-web: A web server for microsatellite prediction. *Bioinformatics* 33, 2583–2585. doi: 10.1093/bioinformatics/btx198

Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* 27, 573–580. doi: 10.1093/nar/27.2.573

Cui, Y., Chen, X., Nie, L., Sun, W., Hu, H., Lin, Y., et al. (2019a). Comparison and phylogenetic analysis of chloroplast genomes of three medicinal and edible *Amomum* species. *Int. J. Mol. Sci.* 20:4040. doi: 10.3390/ijms20164040

Cui, Y., Nie, L., Sun, W., Xu, Z., Wang, Y., Yu, J., et al. (2019b). Comparative and phylogenetic analyses of ginger (*Zingiber officinale*) in the Family Zingiberaceae based on the complete chloroplast genome. *Plants* 8:283. doi: 10.3390/plants8080283

Daniell, H., Lin, C. S., Yu, M., and Chang, W. J. (2016). Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17:134. doi: 10.1186/s13059-016-1004-2

Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., and Dubchak, I. (2004). VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* 32, W273–W279.

Gao, L., Duan, D. D., Zhang, J. Q., Zhou, Y. Z., Qin, X. M., and Du, G. H. (2016). A bioinformatic approach for the discovery of antiaging effects of baicalein from *Scutellaria baicalensis* Georgi. *Rejuvenation Res.* 19, 414–422. doi: 10.1089/rej.2015.1760

Gong, X., Yang, M., He, C. N., Bi, Y. Q., Zhang, C. H., Li, M. H., et al. (2020). Plant pharmacophylogeny: Review and future directions. *Chin. J. Integr. Med.* 28, 567–574. doi: 10.1007/s11655-020-3270-9

Grzegorczyk-Karolak, I., Golab, K., Gburek, J., Wysokinska, H., and Matkowski, A. (2016). Inhibition of advanced glycation end-product formation and antioxidant activity by extracts and polyphenols from *Scutellaria alpina* L. and *S. altissima* L. *Molecules* 21, 739–839. doi: 10.3390/molecules21060739

Gu, R., Rybalov, L., Negrin, A., Morcol, T., Long, W., Myers, A. K., et al. (2019). Metabolic profiling of different parts of *Acer truncatum* from the Mongolian Plateau using UPLC-QTOF-MS with comparative bioactivity assays. *J Agric. Food Chem.* 67, 1585–1597. doi: 10.1021/acs.jafc.8b04035

Hao, D. C., and Xiao, P. G. (2020). Pharmaceutical resource discovery from traditional medicinal plants: Pharmacophylogeny and pharmacophylogenomics. *Chin. Herb. Med.* 12, 104–117.

He, C. N., Peng, Y., Xiao, W., and Xiao, P. G. (2012). The ethnopharmacological investigation of Chinese *Scutellaria* plants. *Mod. Chin. Med.* 14, 16–20.

Jiang, D., Zhao, Z., Zhang, T., Zhong, W., Liu, C., Yuan, Q., et al. (2017). The chloroplast genome sequence of *Scutellaria baicalensis* provides insight into intraspecific and interspecific chloroplast genome diversity in *Scutellaria*. *Genes* 8:277. doi: 10.3390/genes8090227

Jin, J., Yu, W., Yang, J., Song, Y., dePamphilis, C. W., Yi, T., et al. (2018). GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genome. *bioRxiv* [Preprint]. doi: 10.1101/256479

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001). REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 29, 4633–4642. doi: 10.1093/nar/29.22.4633

Kushalappa, A. C., and Gunnaiah, R. (2013). Metabolo-proteomics to discover plant biotic stress resistance genes. *Trends Plant Sci.* 18, 522–531. doi: 10.1016/j.tplants.2013.05.002

Latvis, M., Mortimer, S. M. E., Morales-Briones, D. F., Torpey, S., Uribe-Convers, S., Jacobs, S. J., et al. (2017). Primers for *Castilleja* and their utility across Orobanchaceae: I. Chloroplast primers. *Appl. Plant Sci.* 5:1700038. doi: 10.3732/apps.1700020

Lee, Y., and Kim, S. (2019). The complete chloroplast genome of *Scutellaria indica* var. *coccinea* (Lamiaceae), an endemic taxon in Korea. *Mitochondrial DNA B Resour.* 4, 2539–2540. doi: 10.1080/23802359.2019.1640649

Li, K., Fan, H., Yin, P., Yang, L., Xue, Q., Li, X., et al. (2017). Structure-activity relationship of eight high content flavonoids analyzed with a preliminary assign-score method and their contribution to antioxidant ability of flavonoids-rich extract from *Scutellaria baicalensis* shoots. *Arab. J. Chem.* 11, 159–170.

Li, X., Huang, L. Q., Shao, A. J., Lin, S. F., and Zheng, X. (2003). A survey of studies on germplasm resources of *Scutellaria baicalensis* Georgi. *World Sci. Technol.* 5, 54–58.

Li, X., Li, Y., Zang, M., Li, M., and Fang, Y. (2018). Complete chloroplast genome sequence and phylogenetic analysis of *Quercus acutissima*. *Int. J. Mol. Sci.* 19:2443. doi: 10.3390/ijms19082443

Liang, C., Zhang, X., Diao, X., Liao, M., Sun, Y., and Zhang, L. (2018). Metabolism profiling of nevadensin in vitro and in vivo by UHPLC-Q-TOF-MS/MS. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 1084, 69–79. doi: 10.1016/j.jchromb.2018.03.032

Ling, G., Chengkang, L., Fanli, Y., Chongyang, D., and Chengke, B. (2016). Similarity and diversity evaluation of bioactive ingredients in *S. baicalensis* and *S. viscidula* by HPLC. *Northwest Pharm.* 31, 115–118.

Liu, X., Chang, E.-M., Liu, J.-F., Huang, Y.-N., Wang, Y., and Yao, N. (2019). Complete chloroplast genome sequence and phylogenetic analysis of *Quercus bawanglingensis* Huang, Li et Xing, a vulnerable oak tree in China. *Forests* 10:587.

Liu, X., Zuo, Y., Lin, L., Li, W., Wang, Q., and Deng, H. (2020). The complete chloroplast genome of *Scutellaria tsinyunensis* (Lamiaceae), an endemic species from China. *Mitochondrial DNA B. Resour.* 5, 2568–2570. doi: 10.1080/23802359.2020.1781562

Lou, Z., and Qin, B. (1995). *Species systematization and quality evaluation of commonly used Chinese traditional drugs*, Vol. 2. Beijing: Peking Union Medical College and Beijing Medical University Press.

Lukas, B., and Novak, J. (2013). The complete chloroplast genome of *Origanum vulgare* L. (Lamiaceae). *Gene* 528, 163–169. doi: 10.1016/j.gene.2013.07.026

Marsh, Z., Yang, T., Nopo-Olazabal, L., Wu, S., Ingle, T., Joshee, N., et al. (2014). Effect of light, methyl jasmonate and cyclodextrin on production of phenolic compounds in hairy root cultures of *Scutellaria lateriflora*. *Phytochemistry* 107, 50–60. doi: 10.1016/j.phytochem.2014.08.020

Mayor, C., Brudno, M., Schwartz, J. R., Poliakov, A., Rubin, E. M., Frazer, K. A., et al. (2000). VISTA : Visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* 16, 1046–1047. doi: 10.1093/bioinformatics/16.11.1046

Michael, T., Pascal, L., Tommaso, P., Ulbricht-Jones, E. S., Axel, F., Ralph, B., et al. (2017). GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45, W1–W11. doi: 10.1093/nar/gkx391

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 Analyt. Technol. Biomed. Life Sci.* 1012–1013, 61–68. doi: 10.1016/j.jchromb.2016.01.002

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300

Nylander, J. A. A. (2004). *MrModeltest v.2. Program distributed by the author*. Uppsala: Uppsala University.

Oaarowski, M., Kujawski, R., Mikoajczak, P. A., Wielgus, K., Klejewski, A., Wolski, H., et al. (2019). In vitro and in vivo activities of flavonoids – apigenin, baicalin, chrysin, scutellarin – in regulation of hypertension – a review for their possible effects in pregnancy-induced hypertension. *Herb. Polonica* 65, 55–70.

Ranjbar, M., and Mahmoudi, C. (2017). A taxonomic revision of *Scutellaria* sect. *Lupulinaria* subsect. *Lupulinaria* (Lamiaceae) in Iran. *Feddes Repertorium* 128, 63–101.

Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574.

Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248

Ruhlman, T. A., and Jansen, R. K. (2014). The plastid genomes of flowering plants. *Methods Mol. Biol.* 1132, 3–38.

Safikhani, K., Jamzad, Z., and Saeidi, H. (2018). Phylogenetic relationships in Iranian *Scutellaria* (Lamiaceae) based on nuclear ribosomal ITS and chloroplast trnL-F DNA data. *Plant Syst. Evol.* 304, 1077–1089.

Shen, J., Li, P., He, C.-N., Liu, H.-T., Liu, Y.-Z., Sun, X.-B., et al. (2019). Simultaneous determination of 15 flavonoids from different parts of *Scutellaria baicalensis* and its chemometrics analysis. *Chin. Herb. Med.* 11, 20–27.

Shen, J., Li, P., Liu, S., Liu, Q., Li, Y., Sun, Y., et al. (2021). Traditional uses, ten-years research progress on phytochemistry and pharmacology, and clinical studies

of the genus *Scutellaria*. *J. Ethnopharmacol.* 265:113198. doi: 10.1016/j.jep.2020.113198

Shen, J., Li, P., Liu, S., Liu, Q., Li, Y., Zhang, Z., et al. (2020). The chemopreventive effects of Huangqin-tea against AOM-induced preneoplastic colonic aberrant crypt foci in rats and omics analysis. *Food Funct.* 11, 9634–9650.

Song, J. W., Long, J. Y., Xie, L., Zhang, L. L., Xie, Q. X., Chen, H. J., et al. (2020). Applications, phytochemistry, pharmacological effects, pharmacokinetics, toxicity of *Scutellaria baicalensis* Georgi. and its probably potential therapeutic effects on COVID-19: A review. *Chin. Med.* 15:102. doi: 10.1186/s13020-020-00384-0

Stephan, G., Pascal, L., and Ralph, B. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47, W59–W64. doi: 10.1093/nar/gkz238

Sun, C., Zhang, M., Dong, H., Liu, W., Guo, L., and Wang, X. (2020). A spatially-resolved approach to visualize the distribution and biosynthesis of flavones in *Scutellaria baicalensis* Georgi. *J. Pharm. Biomed. Anal.* 179:113014. doi: 10.1016/j.jpba.2019.113014

Wang, L., Chen, W., Li, M., Zhang, F., Chen, K., and Chen, W. (2020). A review of the ethnopharmacology, phytochemistry, pharmacology, and quality control of *Scutellaria barbata* D. Don. *J. Ethnopharmacol.* 254:112260. doi: 10.1016/j.jep.2019.112260

Wang, L., and Ma, Q. (2018). Clinical benefits and pharmacology of scutellarin: A comprehensive review. *Pharmacol. Ther.* 190, 105–127. doi: 10.1016/j.pharmthera.2018.05.006

Wang, X. G. H., M, R., and Liu, T. (2012). Analysis of trnL-F sequences of the medicinal plants *Scutellaria baicalensis* and its adulterants. *Chin. Agric. Sci. Bull.* 28, 173–177.

Wang, Z., Zhu, C., Liu, S., He, C., Chen, F., and Xiao, P. (2019). Comprehensive metabolic profile analysis of the root bark of different species of tree peonies (*Paeonia* Sect. Moutan). *Phytochemistry* 163, 118–125. doi: 10.1016/j.phytochem.2019.04.005

Wang, Z. L., Wang, S., Kuang, Y., Hu, Z. M., Qiao, X., and Ye, M. (2018). A comprehensive review on phytochemistry, pharmacology, and flavonoid biosynthesis of *Scutellaria baicalensis*. *Pharm. Biol.* 56, 465–484. doi: 10.1080/13880209.2018.1492620

Wen, L., He, T., Yu, A., Sun, S., Li, X., Wei, J., et al. (2021). Breviscapine: A review on its phytochemistry, pharmacokinetics and therapeutic effects. *Am. J. Chin. Med.* 49, 1369–1397. doi: 10.1142/S0192415X21500646

Wink, M. (2015). "Introduction: Biochemistry, physiology and ecological functions of secondary metabolites," in *Biochemistry of plant secondary metabolism*, Vol. 40, ed. M. Wink (Chichester: Wiley-Bleckwell). doi: 10.1080/10256016.2015.1014355

Wu, C. Y., and Li, X. W. (1977). "*Scutellaria* L," in *Flora reipublicae popularis sinicae*, Vol. 65, eds C. Y. Wu and X. W. Li (Beijing: Science Press), 124–248.

Xue, S., Shi, T., Luo, W., Ni, X., Iqbal, S., Ni, Z., et al. (2019). Comparative analysis of the complete chloroplast genome among *Prunus mume*, *P. armeniaca*, and *P. salicina*. *Hortic. Res.* 6:89.

Yan, M., Zhao, X., Zhou, J., Huo, Y., Ding, Y., and Yuan, Z. (2019). The complete chloroplast genomes of *Punica granatum* and a comparison with other species in Lythraceae. *Int. J Mol. Sci.* 20:2886. doi: 10.3390/ijms20122886

Zhang, F., Zhang, B., Shen, R., Xu, X., Guo, L., Wang, Y., et al. (2017). The *Scutellaria baicalensis* stem-leaf total flavonoid regulates the balance of Th17/Treg in EAE rats. *Int. J. Clin. Exp. Med.* 10, 2408–2418.

Zhao, F., Li, B., Drew, B. T., Chen, Y. P., Wang, Q., Yu, W. B., et al. (2020). Leveraging plastomes for comparative analysis and phylogenomic inference within Scutellarioideae (Lamiaceae). *PLoS One* 15:e0232602. doi: 10.1371/journal.pone.0232602

Zhao, Q., Cui, M. Y., Levsh, O., Yang, D., Liu, J., Li, J., et al. (2018). Two CYP82D enzymes function as flavone hydroxylases in the biosynthesis of root-specific 4'-deoxyflavones in *Scutellaria baicalensis*. *Mol. Plant* 11, 135–148. doi: 10.1016/j.molp.2017.08.009

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 18 August 2022
DOI 10.3389/fpls.2022.949743

OPEN ACCESS

EDITED BY
Chunnian He,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

REVIEWED BY
Dahui Liu,
Hubei University of Chinese Medicine, China
Li Xiang,
Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, China

*CORRESPONDENCE
Xianzhang Huang
nylgxyhxz@126.com
Zhongyi Zhang
zyzhang@fafu.edu.cn
Minhui Li
prof_liminhui@yeah.net

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

RECEIVED 21 May 2022
ACCEPTED 26 July 2022
PUBLISHED 18 August 2022

CITATION
Cui Z, Li S, Chang J, Zang E, Liu Q, Zhou B, Li C, Li M, Huang X, Zhang Z and Li M (2022) The pharmacophylogenetic relationships of two edible medicinal plants in the genus *Artemisia*. *Front. Plant Sci.* 13:949743. doi: 10.3389/fpls.2022.949743

The pharmacophylogenetic relationships of two edible medicinal plants in the genus *Artemisia*

Zhanhu Cui[1,2†], Siqi Li[3†], Jiayin Chang[3†], Erhuan Zang[3], Qian Liu[3], Baochang Zhou[4], Chao Li[5], Mengzhi Li[5], Xianzhang Huang[5*], Zhongyi Zhang[1*] and Minhui Li[3,4,6*]

[1]College of Agriculture, Fujian Agriculture and Forestry University, Fuzhou, China, [2]The First People's Hospital of Nanyang Affiliated to Henan University, Nanyang, China, [3]Inner Mongolia Key Laboratory of Characteristic Geoherbs Resources Protection and Utilization, Baotou Medical College, Baotou, China, [4]College of Traditional Chinese Medicine, Inner Mongolia Medical University, Hohhot, China, [5]Nanyang Institute of Technology, Nanyang, China, [6]Inner Mongolia Hospital of Traditional Chinese Medicine, Hohhot, China

Artemisia argyi and *Artemisia indica* are edible medicinal plants belonging to the genus *Artemisia* in the Asteraceae. There are many similarities in their morphology, traditional curative effect, and modern pharmacological treatment. In this study, we built distribution maps of *A. argyi* and *A. indica* in China and a phylogenetic tree of common medicinal plants in Asteraceae. Then, we verified the chemical composition changes of *A. argyi* and *A. indica* via their metabolome. Traditional efficacy and modern pharmacological action were verified by network pharmacology and *in vitro* using RAW264.7 cells. The results showed that *A. argyi* and *A. indica* are widely distributed in China, and they shared pharmaphylogeny, which provides theoretical support for the mixed use of *A. argyi* and *A. indica* in most regions of China. Furthermore, there were both similarities and differences in volatile oil and flavonoid composition between *A. argyi* and *A. indica*. The network pharmacology results showed that *A. argyi* and *A. indica* had 23 common active compounds and that both had pharmacological effects on chronic gastritis (CG). Molecular docking analyses showed that quercetin, luteolin, and kaempferol have strong binding affinities with the target proteins JUN, TP53, AKT1, MAPK3, TNF, MAPK, and IL6. The cell experiment results further demonstrated that *A. argyi* and *A. indica* treat CG *via* the NOD-like receptor pathway. Based on the theory of pharmaphylogeny, this study explored the pharmaphylogeny between *A. argyi* and *A. indica* from various perspectives to provide a basis for the substitution of *A. argyi* and *A. indica*.

KEYWORDS

***Artemisia argyi, Artemisia indica*, pharmaphylogeny, metabolomics, network pharmacology, inflammation**

Introduction

The genus *Artemisia* belongs to the family Asteraceae, which includes more than 500 species worldwide (Abad et al., 2012), among which 186 species and 44 varieties are distributed all over China.[1] It was recorded that there are more than 20 species of *Artemisia* that have been used as medicinal plants (Editorial Board of China Bencao, 1999). *Artemisia argyi* (*A. argyi*) and *Artemisia indica* (*A. indica*) are two edible and medicinal plants (Yang et al., 2020). On the one hand, *A. argyi* and *A. indica*, as widely eaten traditional foods, and are used to make Chinese dishes such as qingtuan and dumplings. Therefore, *A. argyi* and *A. indica* are liked and accepted by most people. On the other hand, as a medicinal plant, according to the Chinese Pharmacopeia, *A. argyi* has the effects of warming the meridian, stopping bleeding, dispersing cold, and relieving pain (Chinese Pharmacopoeia Commission, 2020). It is recorded that *A. argyi* has the effect of warming the stomach (Editorial Board of China Bencao, 1999). As a medicinal plant, *A. indica* is often used instead of *A. argyi*, mixed with it, or as part of moxibustion therapy (Huang and Liu, 1999; Qin et al., 2012). People in some areas eat *A. argyi* or *A. indica* leaves to relieve abdominal pain caused by chronic gastritis (CG) (Li et al., 2018). Essential oils, flavonoids, and terpenoids have been found in *A. argyi* (Zhang et al., 2017; Chen et al., 2021), and the chemical components in *A. indica* are similar to those in *A. argyi* (Wei et al., 2009; Qin et al., 2012). In terms of pharmacology, these compounds have anti-tumor, anti-oxidation, anti-inflammatory and other effects (Chen et al., 2021). Flavonoids are the main components of *A. argyi*, and they are also the important material basis for *A. argyi* to exert pharmacological effects. The content of flavonoids is high and easy to detect. It is often used as an index component in the content determination and fingerprint study of *A. argyi* (Xue et al., 2019; Lan et al., 2020). However, there are few studies on flavonoids in *A. indica*. Therefore, the metabolite composition and the mechanism underlying the similar pharmacological effects of the two plants are worthy of further exploration. According to the theory of pharmaphylogeny, there is a certain correlation between medicinal plant affinities, chemical composition, and curative effect (pharmacological action and traditional curative effect) (Zhou et al., 2022). Therefore, due to the pharmaphylogeny between *A. argyi* and *A. indica*, we guess that their pharmacological action, chemical composition, flavor, and meridian tropism are similar.

In view of the wide potential of *A. argyi* and *A. indica* in the food and medical industries, in this study, we investigated the pharmaphylogeny between *A. argyi* and *A. indica* from multiple perspectives using metabonomics, network pharmacology, and *in vitro* experiments. We aimed to compare the geographic distribution of the two *Artemisia* species, and the flavonoid and volatile metabolites in the plants at three growth stages. Further *in vitro* cell experiments were used to reveal their pharmacological properties. The pharmacophylogenetic study of the two *Artemisia* species was preliminarily explored by the above methods, which provided. This preliminary study provides a scientific basis for the substitution of *A. argyi* and *A. indica*.

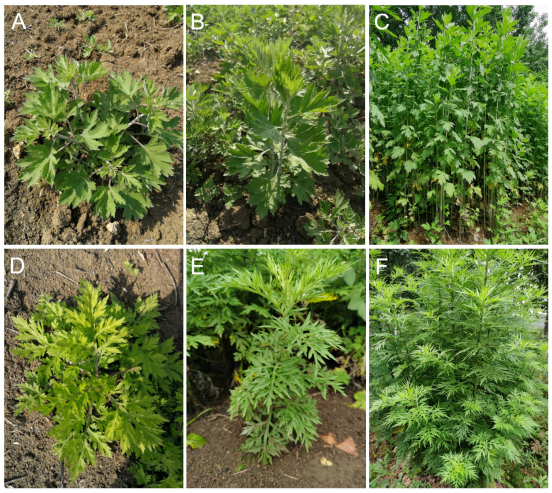


FIGURE 1
Morphological characteristics of *A. argyi* and *A. indica*. *A. argyi* **(A–C)** and *A. indica* **(D–F)** at three different growth stages (April, May, and June).

Materials and methods

Investigation and collection

A. argyi and *A. indica* were collected during the growing season of 2020 from an experimental field in Nanyang County (33°03′6.56″N, 112°49′36.91″E), Henan Province, China. *A. argyi* and *A. indica* leaves were obtained at three different growth stages: the early stage (April 6, 2020), middle stage (May 16, 2020), and later stage (June 26, 2020) and were labeled ARAE, ARAM, and ARAL for *A. argyi* and ARIE, ARIM, and ARIL for *A. indica* for each growth stage, respectively. The morphological characteristics of *A. argyi* and *A. indica* at the three growth stages are shown in **Figure 1**.

The distribution data of *A. argyi* and *A. indica* in China were obtained from an herbarium (Chinese Virtual Herbarium),[2] and the latitude and longitude were corrected in Google Maps and saved in a.csv format. Additionally, 19 environmental factors and 1 altitude factor, nationwide, were converted into.asc format

1 http://www.iplant.cn/frps

2 https://www.cvh.ac.cn/

in ArcGIS. Furthermore, 20 environmental variables, 195 points of *A. argyi,* and 133 points of *A. indica* were collected and imported into MaxEnt software for modeling analysis. The Jackknife method was used to evaluate the contribution of variables in the MaxEnt model, and the crossvalidate method was used for 10 repeated runs. The model output mode was "Logisitc," so that the model prediction results were between 0 and 1, and the interpretation of the results was closer to the distribution probability of *A. argyi* and *A. indica*. The other parameters were default, and the output of the model was an ASCII file. The model output result file was imported into ArcGIS software and converted into a layer file in tif format. The layers were reclassified by spatial analysis tools into four levels: "unsuitable area," "low suitable area," "medium suitable area," and "high suitable area," which reflects the regional situation of *A. argyi* and *A. indica* suitable for growth in China.

Phylogenetic tree construction

Internal transcribed spacer (ITS) gene sequences of *Artemisia* were collected for phylogenetic analysis. The ITS serves as a research region in phylogeny, biogeography, and population genetics (Liu and Christer, 2017; Ren et al., 2017). *Chrysanthemum*, *Atractylodes,* and *Cirsium* in Asteraceae were selected as outgroups as they are relatively common genera in Asteraceae and are commonly used as medicinal plants. The sequence of *Artemisia* and other Asteraceae were downloaded from GenBank for analysis. Mega11.0 was used to align the sequences, the genetic distance was calculated using the Kimura 2-parameter model, and the neighbor-joining (NJ) method was used to build a phylogenetic tree. Bootstrapping was performed with 1,000 iterations on all NJ trees to calculate the support for each branch.

Flavonoid metabolite analysis

Freeze-dried leaves were crushed using a mixer mill for 1.5 min at 30 Hz. Then, 100 mg powder was extracted overnight at 4°C using 1.0 mL 70% aqueous methanol. The extracts were filtered and analyzed using an ultra-performance liquid chromatography-electrospray ionization tandem mass spectrometry (UPLC-ESI-MS/MS) system (UPLC SHIMADZU Nexera X2; MS, Applied Biosystems 4500 Q TRAP). The analytical conditions were as follows: UPLC column, Agilent SB-C18 (1.8 μm, 2.1 mm × 100 mm); solvent system, water (0.1% formic acid), acetonitrile (0.1% formic acid); gradient program, 95:5 v/v at 0 min, 5:95 v/v at 9.0 min, 5:95 v/v at 10.0 min, 95:5 v/v at 10.1 min, 95:5 v/v at 14 min; flow velocity, 0.35 mL/min; column oven, 40°C; injection volume, 4 μL. The effluent was connected to an ESI-triple quadrupole-linear ion trap (QTRAP)-MS. The ESI source operation parameters were as follows: ion source, turbo spray; temperature, 550°C; ion spray voltage, 5,500–4,500 V; ion source gas I, gas II, and curtain gas were set at 50, 60 and 25.0 psi, respectively. Instrument tuning and mass calibration were performed with 10 and 100 μmol/L polypropylene glycol solutions in the triple quadrupole (QQQ) and linear ion trap (LIT) modes, respectively. The declustering potential and collision energy for individual multiple reaction monitoring (MRM) transitions were determined. A specific set of MRM transitions was monitored for each period according to the metabolites eluted during this period. Metabolite analyses were performed according to the reported method (Wang et al., 2018). Based on the self-built database MWDB (Metware Biotechnology Co., Ltd., Wuhan, China), and a public database of metabolite information, MS was used for qualitative and quantitative analysis of the samples' metabolites.

Volatile metabolite analysis

One gram of the leaves' powder was placed into a 20 mL head-space vial. The vials were sealed using crimp-top caps with TFE-silicone headspace septa (Agilent). Each vial was placed at 60°C for 10 min, and then a 65 μm divinylbenzene/carboxen/polydimethylsiloxane fiber (Supelco, Bellefonte, PA, United States) was exposed to the headspace of the sample for 20 min at 60°C. The volatile compound identification and quantification were carried out using an Agilent Model 7890 BGC and a 7000D mass spectrometer (Agilent), equipped with a 30 m × 0.25 mm × 1.0 μm DB-5 ms capillary column. Helium was used as the carrier gas at a flow rate of 1.0 ml/min. The injector temperature and detector temperature were 250°C and 280°C, respectively. The oven temperature was programmed at 40°C for 5 min, increased by 6°C/min to 280°C, and held for 5 min. The ionization voltage was 70 eV. Mass spectra were scanned from 30 to 350 amu. The compounds were identified by comparing the mass spectra with the MWGC database (Metware Biotechnology Co., Ltd., Wuhan, China) and a linear retention index.

Network pharmacology and molecular docking analysis

A number of common potential active compounds from *A. argyi* and *A. indica* were screened according to the selected criteria of bioavailability (OB) ≥ 30% and oral drug-like (DL) ≥ 0.18. The PubChem database[3] was used to determine the compound names and molecular structures. The target genes of these compounds were predicted by the TCMSP database and Universal Protein database (UniProt).[4] We used "CG"

3 https://pubchem.ncbi.nlm.nih.gov/

4 https://www.uniprot.org/

as a keyword in the Online Mendelian Inheritance in Man database (OMIM),[5] GeneCards database (GeneCards),[6] and gene-disease associations database (DisGeNET)[7] to search for CG-related targets and convert the target protein name to a gene name using the UniProt database. Subsequently, after removing repetitive targets, we acquired all CG targets. The target genes of *A. argyi* and *A. indica* against CG were obtained by the intersection of these target genes and the CG target gene using the online Wayne analysis tool. The candidate targets of *A. argyi* and *A. indica* used in CG therapy were predicted. The STRING database[8] was used to construct a protein–protein interaction (PPI) regulation network. The minimum required interaction score was set to 0.9, and a two-plant-active ingredient compound-gene target-CG network was constructed using the Cytoscape 3.6.9 software by integrating the target genes for CG, the active ingredients in *A. argyi* and *A. indica*, and their corresponding targets. To determine functional term enrichment, the intersected target genes from *A. argyi* and *A. indica,* and CG were uploaded to the database for annotation, visualization, and integrated discovery (DAVID 6.8).[9] Gene Ontology (GO) enrichment analysis ($p < 0.05$) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis ($p < 0.05$) were performed using the bioinformatics platform.[10] Based on the above compound and target information, the protein structure of the corresponding target was obtained from the PDB database.[11] The AutoDock Vina software[12] was used for molecular docking analysis.

Determination of active compound content

A. argyi and *A. indica* at three different growth stages were crushed with a high-throughput tissue grinder (Scientz-48 L). Then, 100 mg powder was extracted overnight at 4°C using 4.0 mL 70% aqueous methanol, and then subjected to ultrasonic extraction for 20 min to obtain the sample solution. Then, 1 mg of quercetin, luteolin, and kaempferol was added to a 10 mL volumetric flask, with 70% methanol constant volume to the scale line, and the mixed control solution (reference solution) was prepared. The reference and sample solutions were filtered through a 0.45 μm filter and analyzed by high-performance liquid chromatography (HPLC) (Thermo Scientific Dionex UltiMate 3000). Referring to the method of others (Zheng et al., 2012), the mobile phase components (methanol: water; 60: 40) were filtered through a 0.2 μm filter before use and were pumped from the solvent reservoir at a flow rate of 1 ml/min. The column temperature was 35°C, and the measured wavelength was 360 nm. HPLC chromatograms of the control and sample solutions are shown in **Supplementary Figure 1**.

TABLE 1 Oligonucleotide primers used in RT-PCR experiments.

Gene	Forward	Reverse
NLRP3	CTGGACCACCCCCTGCTGAGA	GGAAGAAGCCCTTGCACCCCTCA
IL-1β	CCAGGATGAGGACATGAGCA	CGGAGCCTGTAGTGCAGTTG
β-actin	CATGTACGTTGCTATCCAGGC	CTCCTTAATGTCACGCACGAT

Cell culture and cytotoxicity analysis

RAW264.7 cells were obtained from the cell bank of the BeNa Culture Collection (Beijing, China) and cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum, penicillin, and streptomycin (100 mg/mL). The cells were cultured at 37°C in an atmosphere of 5% CO_2 for all experiments. Cytotoxicity was assessed using the MTT assay. RAW264.7 cells were seeded at a density of 1×10^4 cells/well in 96-well plates (Corning Inc., Corning, NY, United States). After overnight incubation, the cells were treated with various concentrations (50, 100, 200, and 400 μg/mL) of *A. argyi* and *A. indica* for 24 h. After that, 10 μL MTT solution (0.5 mg/mL) was added to each well, and samples were incubated for 4 h. The resulting colored formazan crystals were dissolved in 150 μL dimethyl sulfoxide (DMSO) by horizontal shaking. The absorption values were measured at 570 nm with a Multimode Microplate Reader (Enspire, PerkinElmer, MA, United States). Cell viability was determined relative to the untreated cells in the control group.

Assay to measure cell viability

Cell viability was assessed by the MTT assay. Briefly, RAW264.7 cells were seeded at a density of 1×10^4 cells/well in 96-well plates. After overnight incubation, cells were treated with various concentrations (25, 50, and 100 μg/mL) of *A. argyi* and *A. indica* for 12 h, and then LPS (1 μg/mL) was added or not for the next 24 h. After that, 10 μL MTT solution (0.5 mg/mL) was added to each well, and the cells were incubated for 4 h. The resulting colored formazan crystals were dissolved in 150 μL DMSO by horizontal shaking.

5 https://www.omim.org/
6 https://www.genecards.org/
7 http://www.disgenet.org/
8 https://string-db.org/
9 https://david.ncifcrf.gov/
10 http://www.bioinformatics.com.cn/
11 http://www.rcsb.org/
12 http://vina.scripps.edu/

The absorption values were measured at 490 nm with the Multimode Microplate Reader. Cell viability was determined relative to the untreated cells in the control group, and then cell survival was calculated.

Nitric oxide determination in the supernatant of RAW264.7 cells

RAW264.7 cells were seeded at a density of 5×10^5 cells/well in 6-well plates (Corning Inc., Corning, NY, United States). After overnight incubation, cells were treated with various concentrations (25, 50, and 100 μg/mL) of *A. argyi* and *A. indica* for 12 h, and then LPS (1 μg/mL) was added or not for the next 24 h. After 24 h, the cell supernatant was absorbed for NO determination. The level of NO was determined by assaying the concentration of nitrite in the cell supernatant through the Griess reaction using the NO assay kit (Nanjing Jiancheng Institute of Biological Engineering), according to the manufacturer's instructions.

Real-time quantitative polymerase chain reaction

Total RNA was extracted from cultured RAW264.7 cells using the TRIzol reagent (Solarbio, China) and reverse-transcribed using the PrimeScript™ RT reagent kit with the gDNA Eraser (Perfect Real Time) (TaKaRa, Japan), according to the manufacturer's instructions. Aliquots of the obtained cDNA samples were then amplified using PCR with the following schedule: 40 cycles at 95°C for 30 s, primer annealing at 95°C for 5 s, and extension at 60°C for 34 s. The specifically designed primers are listed in **Table 1**. All primers were tested, the fluorescence signals were recorded, and the relative values were compared to those of the control group.

Statistical analysis

Cluster analysis and principal component analysis (PCA) were carried out using R[13] in accordance with previously described methods (Wang et al., 2016). K-means analysis and heatmap analysis based on hierarchical clustering were performed in R. The normalization method used for the K-means analysis was the unit variance scaling method, which involves standardizing the data according to the mean and standard deviation of the original data. Sub-classes were classified according to the trends in the relative content of metabolites in different samples. Metabolites with a fold change ≥ 2 or ≤ 0.5, and VIP ≥ 1 were defined as differentially accumulated metabolites between the compared samples. All experiments were repeated 3 times. The data was analyzed using GraphPad Prism 9 statistical software. The experimental results are expressed as means ± standard deviation (*SD*), and the comparison between groups was performed using the *t*-test or one-way ANOVA of variance. $P < 0.05$ was considered statistically significant.

13 http://www.r-project.org/

Results

Pharmaphylogeny between *Artemisia argyi and Artemisia indica*

A. argyi and *A. indica* are mixed for use in some areas, but the specific locations have not been recorded. Therefore, we visualized the distribution of *A. argyi* and *A. indica* in China and found these species to be distributed in almost all provinces and cities as shown in **Figure 2**. *A. argyi* is mainly distributed in the eastern coastal areas, and *A. indica* is mainly distributed in the southeastern coastal areas. In addition, we divided four types of regions according to the distribution of the two plants in each region. Plants are widely distributed in the "high suitability area" and less distributed in the "low suitability area." The widespread distribution of these two plants in China explains the mixed use of *A. argyi* and *A. indica* in many areas.

As shown in **Figure 3**, the results of the phylogenetic analysis showed that four strains, *A. argyi*, *A. indica*, *A. lavandulifolia*, and *A. mongolica*, separately formed a small branch. Thus, these species are closely phylogenetically related and can be clearly distinguished.

Flavonoid metabolite analysis

A typical total ion content (TIC) diagram of a quality control (QC) sample is shown in **Supplementary Figure 2**. The TIC diagram shows a continuous description of the total intensity of all ions in the mass spectrum. The detection diagram of metabolites in the MRM mode is shown in **Supplementary Figure 3**, which shows the ion current diagram of various substances. **Supplementary Figure 4** shows the high degree of overlap of TIC plots between multiple QC samples. The consistency of retention time (RT) and peak intensity of the two QC samples showed good signal stability when detecting the same sample at different times. High correlation coefficients (r = 0.989–0.998) were obtained between the three biological replicates of each sample (**Supplementary Figure 5**), indicating good sample uniformity. These results suggest that the data obtained in this study have good reproducibility and reliability.

The leaf samples of *A. argyi* and *A. indica* obtained at different growth stages were analyzed using LC-MS. A total of

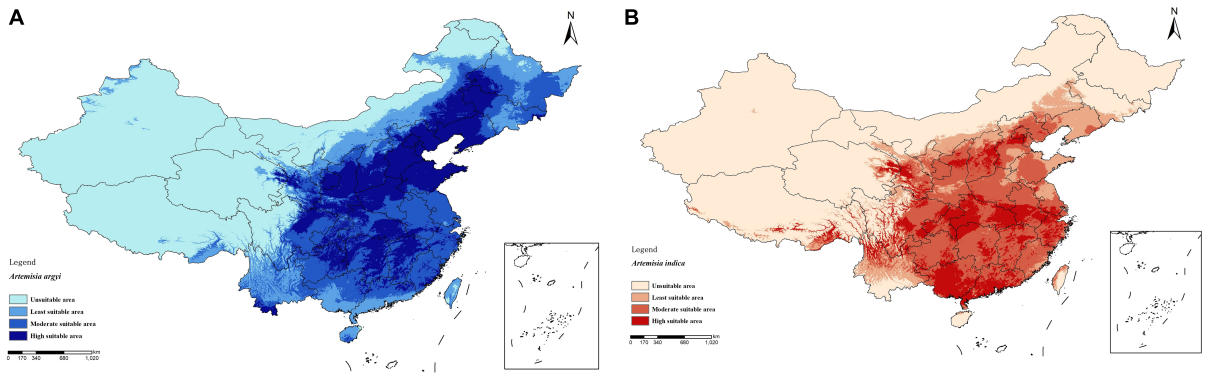


FIGURE 2
Distribution of *A. argyi* and *A. indica*. Darker color means more suitable for growth. Distribution of *A. argyi* **(A)** and *A. indica* **(B)** in China. *A. argyi* is mainly distributed in the eastern coastal areas, and *A. indica* is mainly distributed in the southeastern coastal areas.

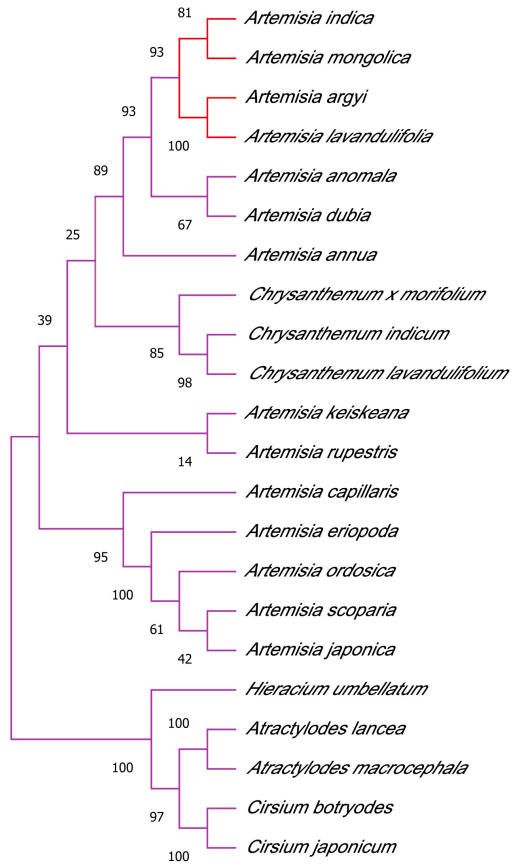


FIGURE 3
Phylogenetic tree of Asteraceae species using the neighbor-joining method. *A. argyi*, *A. indica*, *A. lavandulifolia*, and *A. mongolica* are closely related.

311 flavonoids were identified, including 160 flavones (51.45%), 61 flavonols (19.61%), 21 dihydroflavones (6.75%), 17 flavonoid carbonosides (5.47%), 16 isoflavones (5.14%), 10 tannins (3.22%), eight flavanols (2.57%), eight dihydroflavonols (2.57%), seven chalcones (2.25%), two sinensetins (0.64%), and one dihydroisoflavone (0.32%). Information about the metabolite sequence number, molecular weight, formula, ionization model, metabolite name, classification, CAS number, level A or B identification confidence, CPD ID, and KEGG map is listed in Supplementary Table 1. A total of 236 compounds were identified in *A. argyi*, 216 in the early stage, 227 in the middle stage, and 230 in the later stage. A total of 214 compounds were identified in *A. indica*, 183 in the early stage, 189 in the middle stage, and 191 in the late stage.

We used PCA to cluster all the flavonoid metabolites identified in the different samples. This comparison showed that there was a difference between the six groups, and all samples were in the 95% confidence interval (Hotelling's T-squared ellipse). In the PCA score chart, the metabolic phenotypes of *A. argyi* and *A. indica* were separated, and *A. argyi* and *A. indica* could also be distinguished in each growth stage (Figure 4A). In addition, the metabolic profile differences of *A. argyi* and *A. indica* during the different growth stages were further estimated using hierarchical cluster analysis (HCA). The metabolic profile differences during the different growth stages were clearly divided into six groups on the heat map; the results showed that there were some differences in metabolite content between the two cultivars in different growth stages (Figure 4B).

The difference in flavonoid metabolites between *A. argyi* and *A. indica* were analyzed using a LC-MS/MS detection platform and multivariate analysis. K-means analysis divided the 299 differential flavonoid metabolites into 12 clusters (Figure 5). Among them, 94 metabolites were significantly increased in ARAL, and mainly included butin, acacetin, quercetin, pectolinarigenin, and kumatakenin. A total of 30 metabolites were significantly accumulated in ARAM. Fifty-four metabolites were significantly increased in ARIM, primarily isorhamnetin, luteolin-7-O-glucuronide, and scutellarin (Supplementary Table 2). During the same period, the metabolites of the two

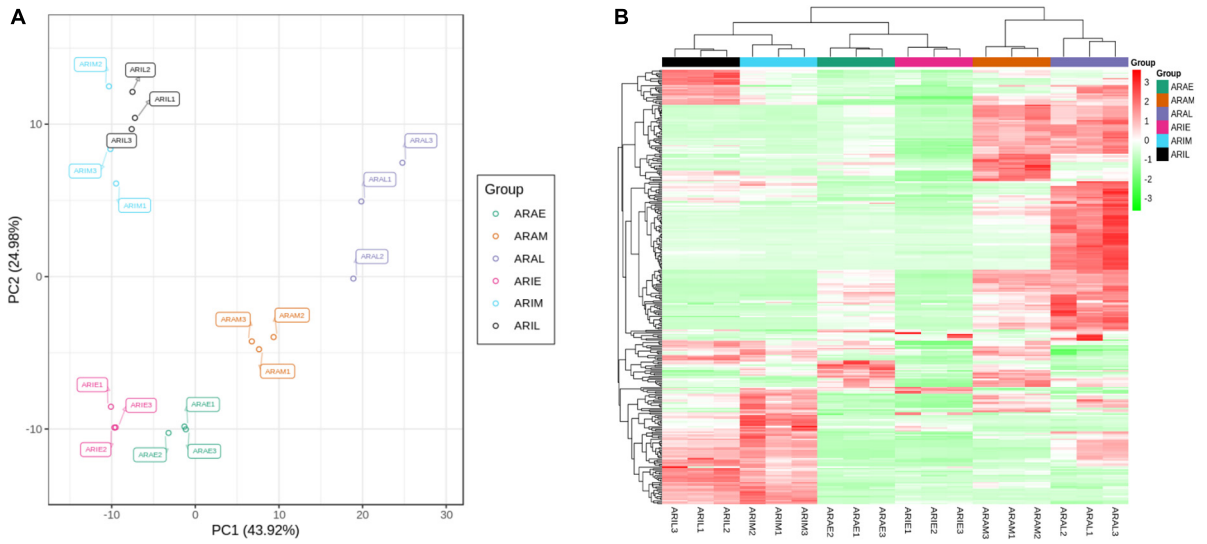


FIGURE 4
Differential flavonoid metabolites analysis. **(A)** PCA of flavonoid metabolites of *A. argyi* and *A. indica*. Each point in the figure represents a sample, and samples of the same group are represented by the same color. **(B)** Clustering heat map of flavonoid metabolites of *A. argyi* and *A. indica*. The X-axis represents the name of the sample, and the Y-axis represents the metabolite information. Different colors are the values obtained after the standardization of the relative content, in which red represents the high content and green represents the low content.

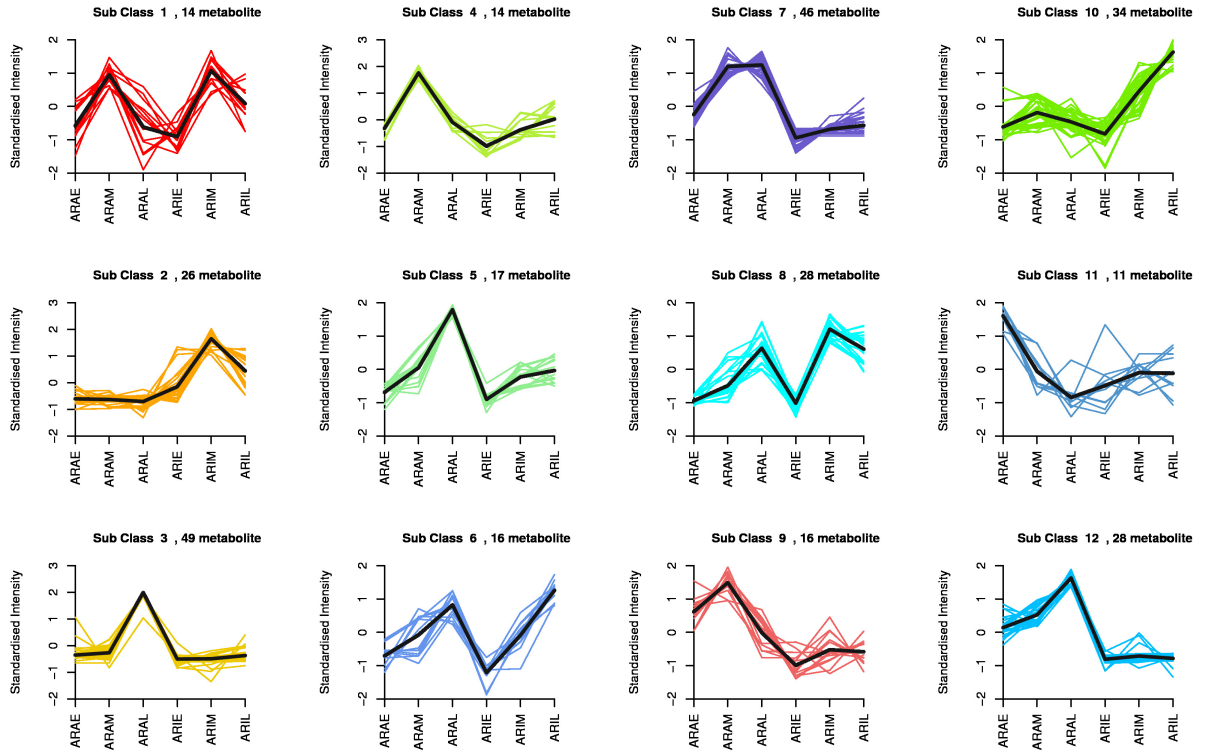


FIGURE 5
K-means analysis of differential flavonoid metabolites of *A. argyi* and *A. indica*. The X-axis represents the name of the sample, the Y-axis represents the standardized relative content of metabolites, the subclass represents the number of metabolites with the same trend.

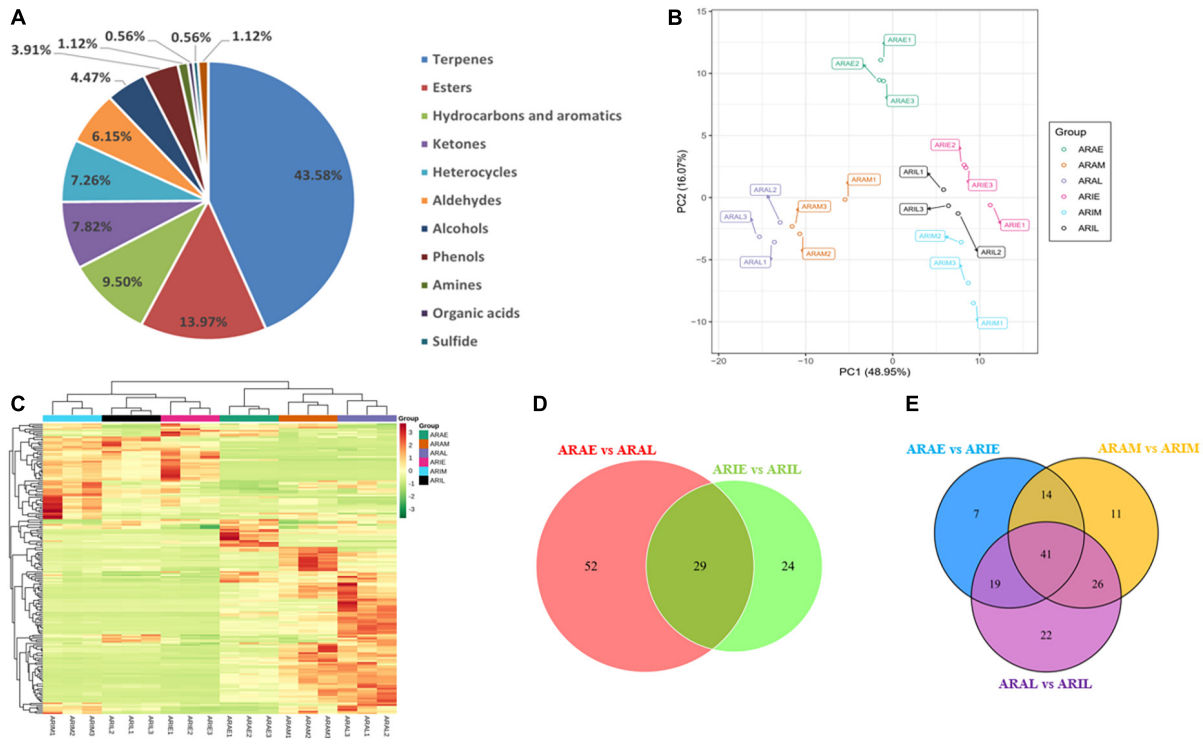


FIGURE 6
Differential volatile metabolites analysis. **(A)** Volatile compounds identified by GC-MS. **(B)** PCA of volatile metabolites. **(C)** Hierarchical cluster analysis of volatile metabolites. **(D,E)** Venn diagrams of *A. argyi* and *A. indica* in different growth stages.

varieties showed different accumulations. There were 185 different metabolites in ARAE vs. ARIE (30 upregulated and 155 downregulated), 181 between ARAM and ARIM (59 upregulated and 122 downregulated), and 189 between ARAL and ARIL (55 upregulated, 134 downregulated). The trends of these different metabolites are clearly displayed in the volcano plots (Supplementary Figures 6A–C). In the same variety, there were 173 significantly different metabolites between ARIE vs. ARIL (173 upregulated, 8 downregulated, Supplementary Figure 6D), and 182 between ARAE vs. ARAL (156 upregulated, 26 downregulated, Supplementary Figure 6E). Most of the metabolites were upregulated.

Volatile metabolite analysis

To investigate the dynamic changes in the metabolic profiles of the volatile substances, a metabolomics method was used to analyze the samples of *A. argyi* and *A. indica* at different growth stages. The mass spectrum results of QC samples were overlapped and analyzed, as shown in Supplementary Figure 7. The results showed that the TIC curve of the QC samples has a high overlap rate, which indicates that the test results are in good agreement with the experimental results.

A total of 179 volatile metabolites in the two varieties were identified using GC-MS/MS, including 44% terpenoids, 14.4% esters, 9% hydrocarbons and aromatics, 8% ketones, 7% heterocycles, 6.4% aldehydes, 4% alcohols, 4% phenols, 1% amines, 0.6% organic acids, 0.6% sulfides, and 1% other metabolites (Supplementary Tables 3, 4 and Figure 6A). In total, 174 compounds were obtained in *A. argyi*, including 166 in ARAE, 156 in ARAM, and 158 in ARAL; 165 compounds were identified in *A. indica*, including 144 in ARIE, 149 in ARIM, and 147 in ARIL. A total of 117 compounds were common between *A. argyi* and *A. indica*. PCA models based on 18 test samples showed differences in each group. The PCA results of the six sample groups showed significant metabolic differences among the different varieties in the different periods (Figure 6B). The accumulation pattern of metabolites in *A. argyi* and *A. indica* is directly displayed in the hierarchical cluster analysis of the heat map, which shows that the different varieties have different metabolite accumulations in different growth stages and that they may have different metabolic patterns (Figure 6C).

Based on the GC-MS/MS detection platform and multivariate analysis, the differences in volatile metabolites in *A. argyi* and *A. indica* were analyzed. A Venn diagram was used to describe the accumulation of different metabolites at different growth stages (Figures 6D,E). In the comparison between ARAE and ARAL, there were 81 differential metabolites,

TABLE 2 Identification of common potential active compounds from *A. argyi* and *A. indica*.

ID	MOL_ID	Compounds	Class	OB (%)	DL
A1	MOL006436	Okanin	Chalcones	98.81	0.20
A2	MOL005190	Eriodictyol	Dihydroflavones	71.79	0.24
A3	MOL002975	Butin	Dihydroflavones	69.94	0.21
A4	MOL002844	Pinocembrin	Dihydroflavones	64.72	0.18
A5	MOL004328	Naringenin	Dihydroflavones	59.29	0.21
A6	MOL004093	Azaleatin	Flavonols	54.28	0.30
A7	MOL004112	Patuletin	Flavonols	53.11	0.34
A8	MOL000239	Kumatakenin	Flavonols	50.83	0.29
A9	MOL000354	Isorhamnetin	Flavonols	49.60	0.31
A10	MOL005229	Artemetin	Flavones	49.55	0.48
A11	MOL000098	Quercetin	Flavonols	46.43	0.28
A12	MOL000422	Kaempferol	Flavonols	41.88	0.24
A13	MOL005842	Pectolinarigenin	Flavones	41.17	0.30
A14	MOL002776	Baicalin	Flavones	40.12	0.75
A15	MOL011604	Syringetin	Flavones	36.82	0.37
A16	MOL000006	Luteolin	Flavones	36.16	0.25
A17	MOL003044	Chrysoeriol	Flavones	35.85	0.27
A18	MOL001689	Acacetin	Flavones	34.97	0.24
A19	MOL004083	Tamarixetin	Flavones	32.86	0.31
A20	MOL002322	Isovitexin	Flavones	31.29	0.72
A21	MOL002881	Diosmetin	Flavones	31.14	0.27
A22	MOL001735	Hispidulin	Flavones	30.97	0.27
A23	MOL007274	Cirsimaritin	Flavones	30.35	0.30

of which 57 were upregulated and 24 were downregulated. In the comparison between ARIE and ARIL, 53 differential metabolites were identified, of which 51 were upregulated and two were downregulated. Comparing the data of ARAE and ARIE, we identified 81 differential metabolites, of which 16 were upregulated and 65 downregulated. In the comparison between ARAM and ARIM, 92 differential metabolites were found, of which 35 were upregulated and 57 downregulated, whereas in the comparison between ARAL and ARIL, 108 differential metabolites were identified, of which 33 were upregulated and 75 downregulated. These trends are depicted in the volcano plot (**Supplementary Figure 8**).

Network pharmacology analysis of *Artemisia argyi and Artemisia indica* in the treatment of chronic gastritis

Based on the above flavonoid metabolite information, the TCMSP database, and OB $\geq$ 30% and DL $\geq$ 0.18 as the screening conditions, a total of 23 active compounds from *A. argyi* and *A. indica* were identified (**Table 2**). These active compounds were obtained from the TCMSP database to identify the corresponding target proteins, which were converted into standard gene names using the UniProt database. After deduplication, 202 potential targets of the above compounds were identified.

By integrating the CG targets collected in the OMIM database, GeneCards database, and DiGSeE database, and eliminating duplicate targets, a total of 1,267 target proteins related to CG were obtained (**Supplementary Figure 9A**). After matching the targets of the common active compounds with the targets of CG, the intersection was determined, leading to 100 common target genes being identified, as shown in **Figure 7A**.

We used Cytoscape 3.6.9 to construct the "active components-target-disease" network. The compounds and overlapped targets between *A. argyi* and *A. indica* and CG were imported into the system, and the "*A. argyi* and *A. indica* component-target-CG" network was constructed by connecting to the predicted targets. The core nodes were screened according to network topology characteristics, such as the node degree value. The results showed that quercetin, luteolin, and kaempferol play a critical role in the entire network, and that they may be significant active compounds for the treatment of CG (**Figure 7B**).

Based on the overlapping targets between *A. argyi* and *A. indica* and CG, a PPI network was established. There were 92 nodes and 421 edges in the network diagram, and the average node degree was 9.15 (**Supplementary Figure 9B**). Proteins are represented as nodes with different colors, and the lines connecting the nodes represent the functional association between proteins. The line thickness represents the confidence in the reported association. The number of connections of a node represents the degree of the node. The higher the degree, the greater the position of the gene in the regulatory network, indicating that it is a central gene. According to the topological network analysis results, the top nine core genes with a degree value greater than 20 are shown in **Figure 7C**. The results showed that JUN, TP53, AKT1, RELA, MAPK3, TNF, MAPK1, IL6, and MAPK14 were the most connected key targets with other genes.

GO and KEGG enrichment analyses were used to analyze the anti-CG function of *A. argyi* and *A. indica*. The target genes were mostly enriched in negative regulation of apoptotic process (GO:0043066), response to drug (GO:0042493), and extrinsic apoptotic signaling pathway in absence of ligands (GO:0097192) in BP enrichment analysis, extracellular space (GO:0005615), cytosol (GO:0005829), and caveola (GO:0005901) in CC analysis, and enzyme binding (GO:0019899), identical protein binding (GO:0042802), and protein binding (GO:0005515) in MF analysis (**Figure 8A**). The most important pathways mediating the effects of *A. argyi* and *A. indica* on CG treatment were the NOD-like receptor, TNF, hypoxia-inducible factor (HIF)-1, T cell receptor, Toll-like receptor, and PI3K-AKT signaling pathways (**Figure 8B**).

We selected the top five targets (JUN, TP53, AKT1, RELA, and MAPK3) in the PPI network and three key compounds

10.3389/fpls.2022.949743

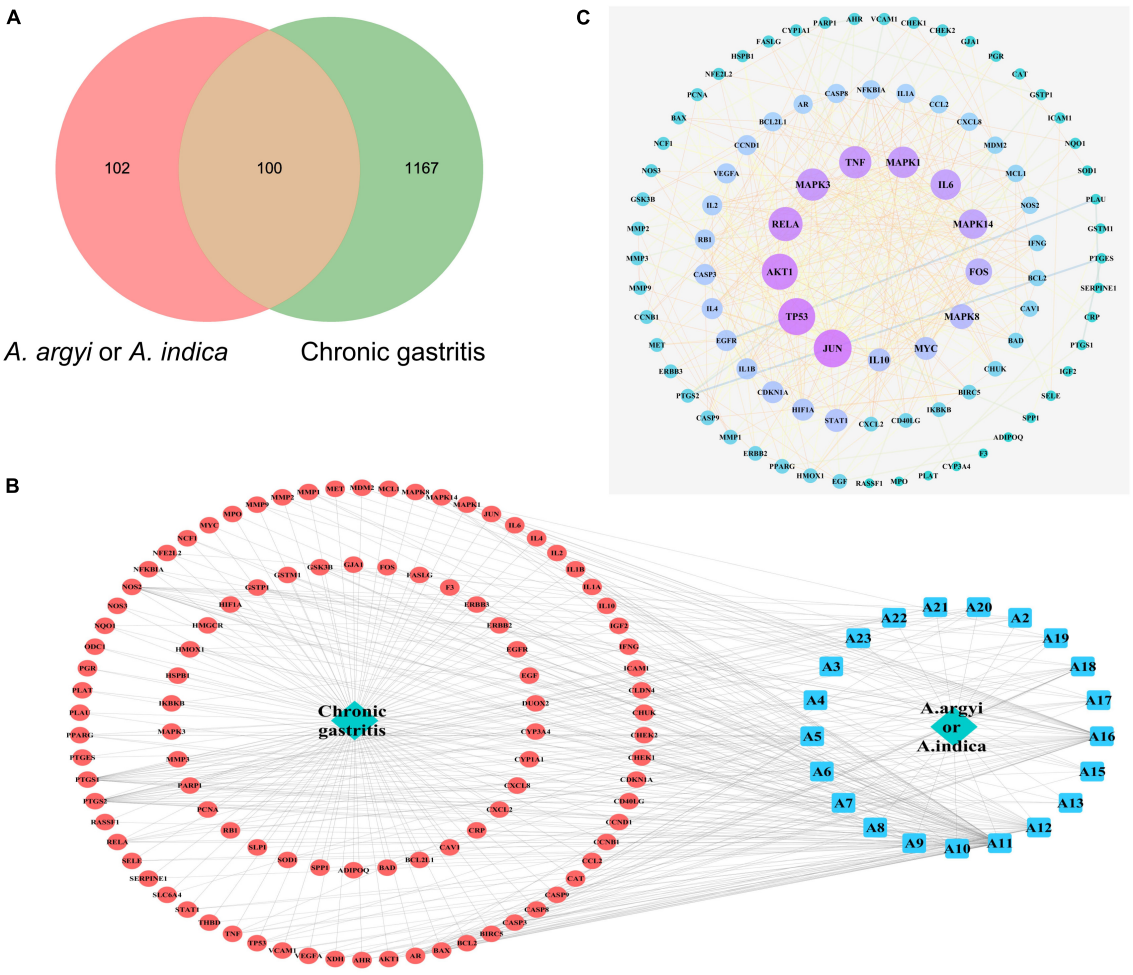


FIGURE 7
Diagram of interactions between drug targets and disease targets. **(A)** Venn diagram of the intersection of target genes of *A. argyi* and *A. indica* and chronic gastritis (CG)-related genes. **(B)** Topological network of CG target genes of *A. argyi* and *A. indica*. **(C)** Network map of *A. argyi* and *A. indica*-compound-drug action target-disease. The larger node represents the higher degree.

(quercetin, luteolin, and kaempferol) for the molecular docking analysis. Figure 9A shows a heatmap depicting the docking results, and the docking visualization results are shown in Figure 9B. The results showed that the binding energy of each compound to protein was < -5 kcal/mol, indicating that each compound could bind well to the protein. The binding energies of quercetin, luteolin, and kaempferol to JUN were -8.9, -8.6, and -8.8 kcal/mol, respectively.

Content analysis of active compound

The mixed reference solution was injected in the following sequence: 0.1, 0.2, 0.5, 1, 2, and 5 μL. The sample quantity (μg) was taken as the abscissa and the peak area as the ordinate, and the standard curve was drawn. The linear equation is: y = 66.921x -0.2352, R^2 = 0.9995. The whole grass powder of *A. argyi* and *A. indica* was carefully weighed and numbered as: ARAE 100.4 mg, ARAM 100.4 mg, ARAL 100.3 mg, ARIE 100.2 mg, ARIM 100.0 mg, and ARIL 100.2 mg. There was a good linear relationship between sample quantity and peak area in the range of 0.04-0.2 mg., and the chromatographic peaks of these two compounds were not detected. The content of luteolin in the samples is shown in Figure 10. In different growth periods, the content of luteolin in *A. argyi* was the highest in April and that of *A. indica* was highest in June.

Cytotoxicity analysis and protective effects of the extracts on RAW264.7 cells

First, we determined the effects of different doses of *A. argyi* and *A. indica* on the viability of RAW264.7 cells using the

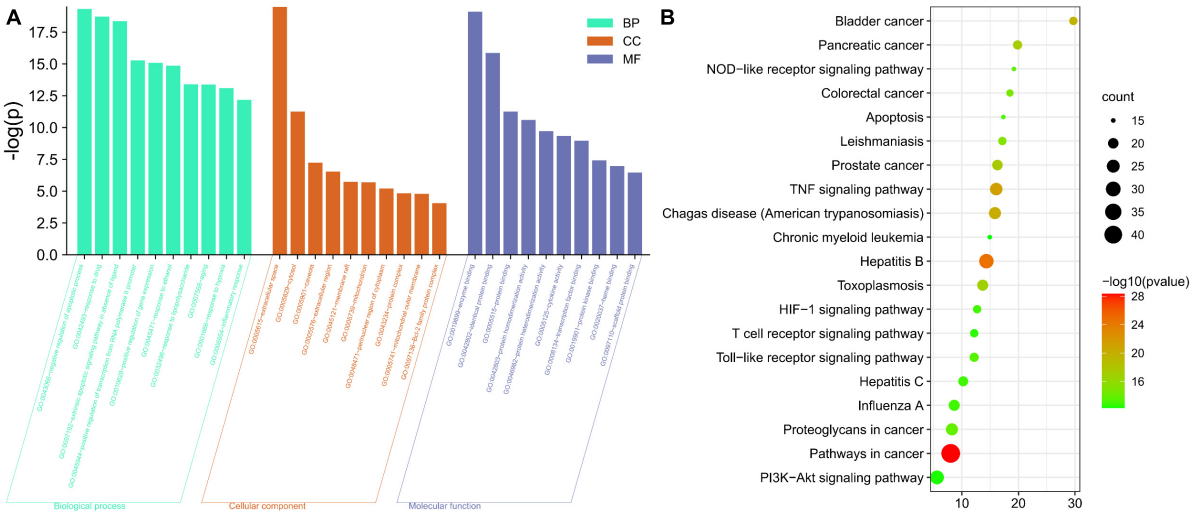


FIGURE 8
Diagram of enrichment analysis results. **(A)** Gene Ontology enrichment analysis for key targets (top 10 were listed). **(B)** KEGG pathway enrichment analysis of key targets (top 20 were listed); the abscissa label represents fold enrichment of pathways.

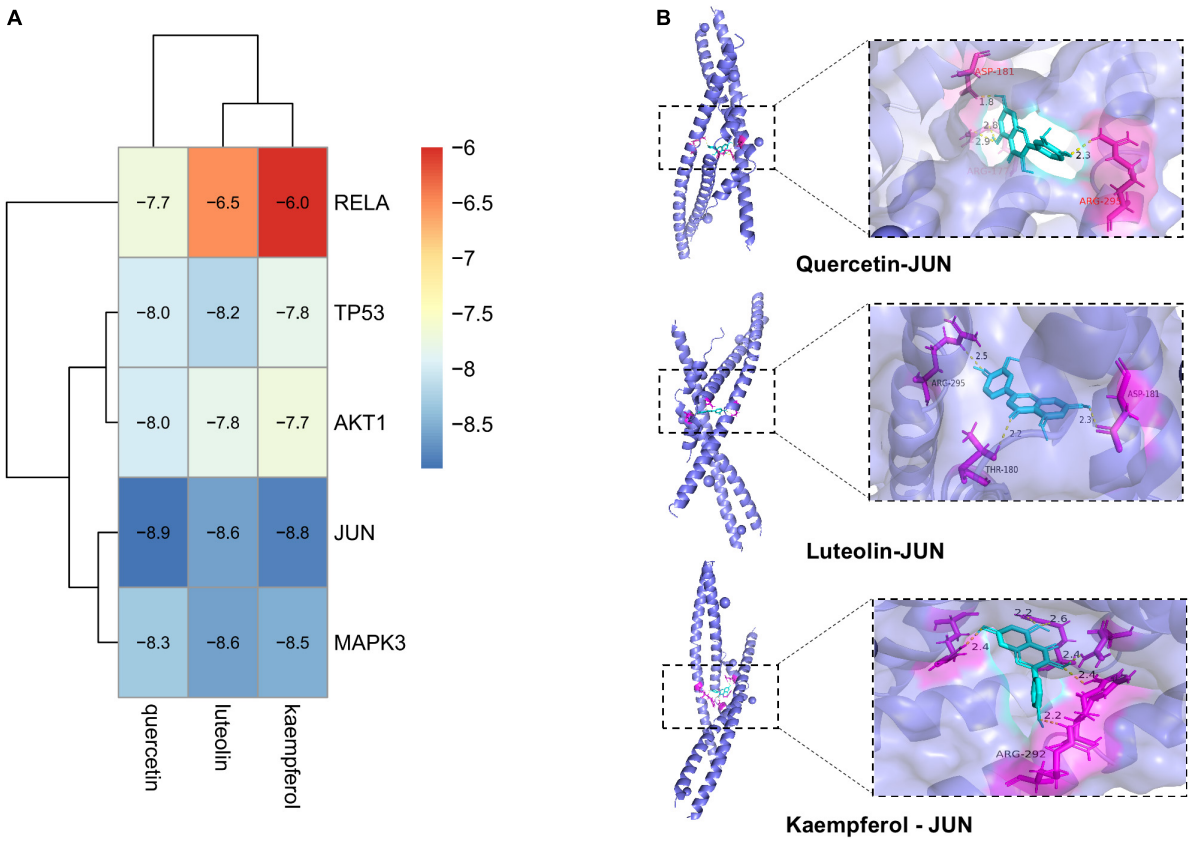


FIGURE 9
Diagram of molecular docking results. **(A)** Heat map of binding energy between the compounds of *A. argyi* and *A. indica* and core target molecule docking. The smaller the value the stronger the combination. **(B)** Docking model diagram of the compound and key target molecule.

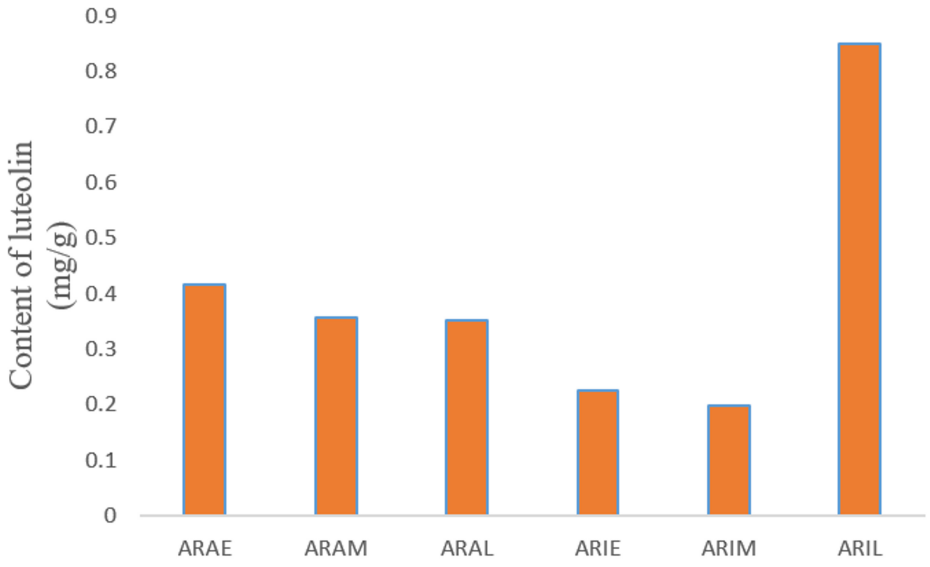


FIGURE 10
Diagram of luteolin content. X-axis is the sample name, and Y-axis is the content of luteolin in the sample powder.

MTT assay to determine the effect of the drug on cytotoxicity. Figure 11A shows that *A. argyi* and *A. indica* had no cytotoxicity when the dosage was $\leq$ 100 μg/mL in all three growth stages stage. Therefore, three concentrations were selected (25, 50, and 100 μg/mL) for subsequent experiments.

As shown in Figure 11B, compared with the model group, ARAE, ARAM, and ARAL had significantly improved cell survival rate and ARAL had the best effect. Interestingly, *A. indica* was consistent with *A. argyi,* as *A. indica* harvested in June (ARIL) significantly improved cell survival compared to that of the model group.

Artemisia argyi and *Artemisia indica* alleviated LPS-induced RAW264.7 cell inflammation via the NOD-like receptor Signaling pathway

Based on the cell viability results, we used ARAL and ARIL to validate the network pharmacology results. The KEGG results revealed that the inflammatory pathways that may be involved in the treatment of CG using *A. argyi* and *A. indica* include the NOD-like receptor signaling pathway, TNF signaling pathway, Toll-like receptor signaling pathway, and the PI3K-Akt signal pathway. Since the NOD-like receptor signaling pathway is closely related to CG and plays a key anti-inflammatory role, two key factors in this pathway, NLRP3 and IL-1β, were selected to verify the anti-inflammatory mechanism of *A. argyi* and *A. indica*.

First, the cell supernatant was absorbed for NO determination (Figure 12); after adding LPS, the NO content was significantly increased, and by giving positive medicine (indomethacin) and different concentrations of alcohol extract of *A. argyi* and *A. indica*, the NO content decreased significantly in a dose-dependent manner (i.e., at 100 μg/ml dosage of *A. argyi* and *A. indica*, the NO content was the lowest). Based on the results of the NO kit, we then analyzed the mRNA expression of two inflammatory factors, NLRP3 and IL-1β, in RAW264.7 cells. As shown in Figure 13, compared with that in the control group, the mRNA expression of NLRP3 and IL-1β increased in the model group, and showed a significant downward trend after the administration of three concentrations of *A. argyi* extract, but it did not show a dose-dependent relationship. In addition, the results shown in Figure 13 suggest that the result of *A. indica* was basically consistent with that of *A. argyi*; compared with that in the control group, the mRNA in the model group increased and tended to decrease after *A. argyi* administration.

Discussion

Artemisia is a large genus of family Asteraceae, and *A. argyi* and *A. indica* are two edible and medicinal plants with similar morphology, traditional therapeutic effects, physicochemical properties, and flowering period (see text footnote 1). In addition, *A. argyi* and *A. indica* are distributed in almost all areas of China, mostly in coastal areas, and studies have reported mixed use in some areas (Song et al., 1994; Huang and Liu, 1999; Wu et al., 2008; Wei et al., 2009; Liang et al., 2010; Qin et al., 2012). We conducted a systematic analysis of the ITS sequences of Asteraceae, and according to the results of

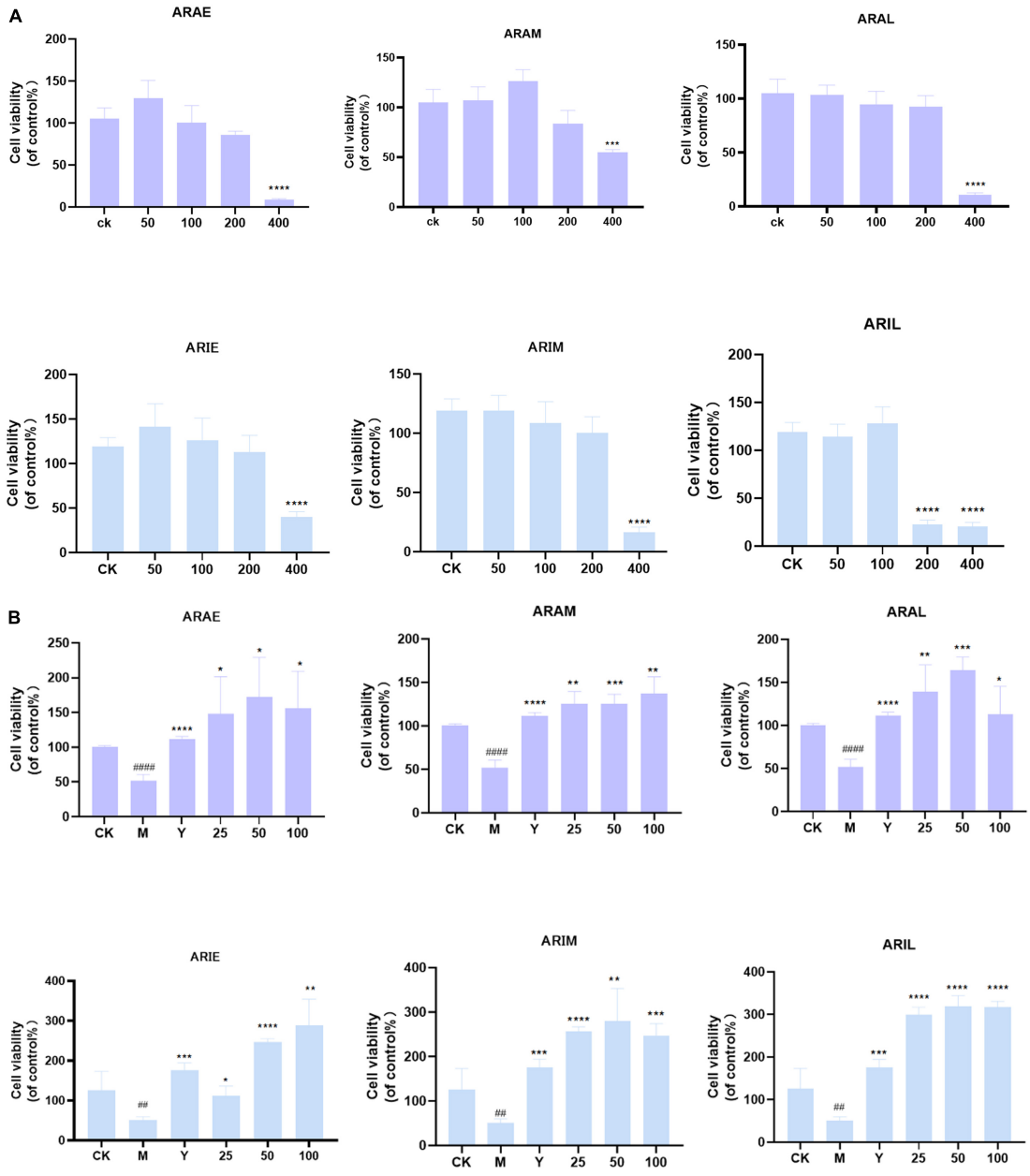


FIGURE 11
Effect of toxicity **(A)** and activity **(B)** of *A. argyi* and *A. indica* on survival rate of RAW264.7 macrophages. *A. argyi* and *A. indica* between 0 and 100 μg/ml had no toxic effect on RAW264.7 cells. The activity of ARAL was higher than that of ARAE and ARAM, and that of ARIL was higher than that of ARIE and ARIM in RAW2647 macrophages (means ± SD, $n = 3$). $^{****}P < 0.0001$, $^{***}P < 0.001$, $^{**}P < 0.01$, $^{*}P < 0.1$ compared with control group. $^{\#\#\#\#}P < 0.0001$, $^{\#}P < 0.1$ compared with that of the control group.

the phylogenetic tree, *A. argyi*, *A. indica*, *A. lavandulifolia*, and *A. mongolica* are the most closely related species. A new study showed that *A. argyi* and *A. indica* are most closely related based on phylogenetic relationships constructed from chloroplast genome data of 19 species in the family Asteraceae (Lan et al., 2022). The four species belong to Sect. *Artemisia* (see text footnote 1). *A. argyi* and *A. indica* are similar in flavor, meridian tropism, and function. In pharmacophylogeny, closely related species are similar not only in morphology but also in biochemical characteristics, so their chemical compositions, including secondary plant metabolites, are often similar (Li, 2008). It has been reported that flavonoids and volatile oils are common in *Artemisia*. Metabolomics was used to analyze the changes in the contents of total flavonoids and essential oils of *A. argyi* and *A. indica*. There were 299 and 179 total flavonoids and essential oils, respectively, among which 117

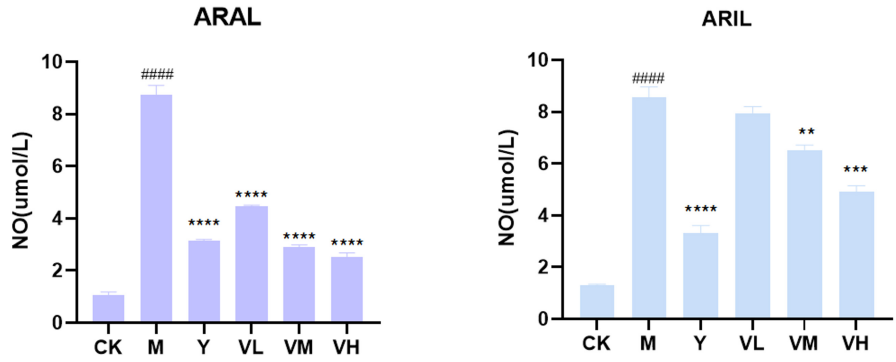


FIGURE 12
Effect of *A. argyi* and *A. indica* on NO production in RAW264.7 macrophages. The content of LPS group increased and decreased after administration (means ± *SD*, *n* = 3). ****$P < 0.0001$, ***$P < 0.001$, **$P < 0.01$ compared with that of the LPS group; ####$P < 0.0001$ compared with that of the control group.

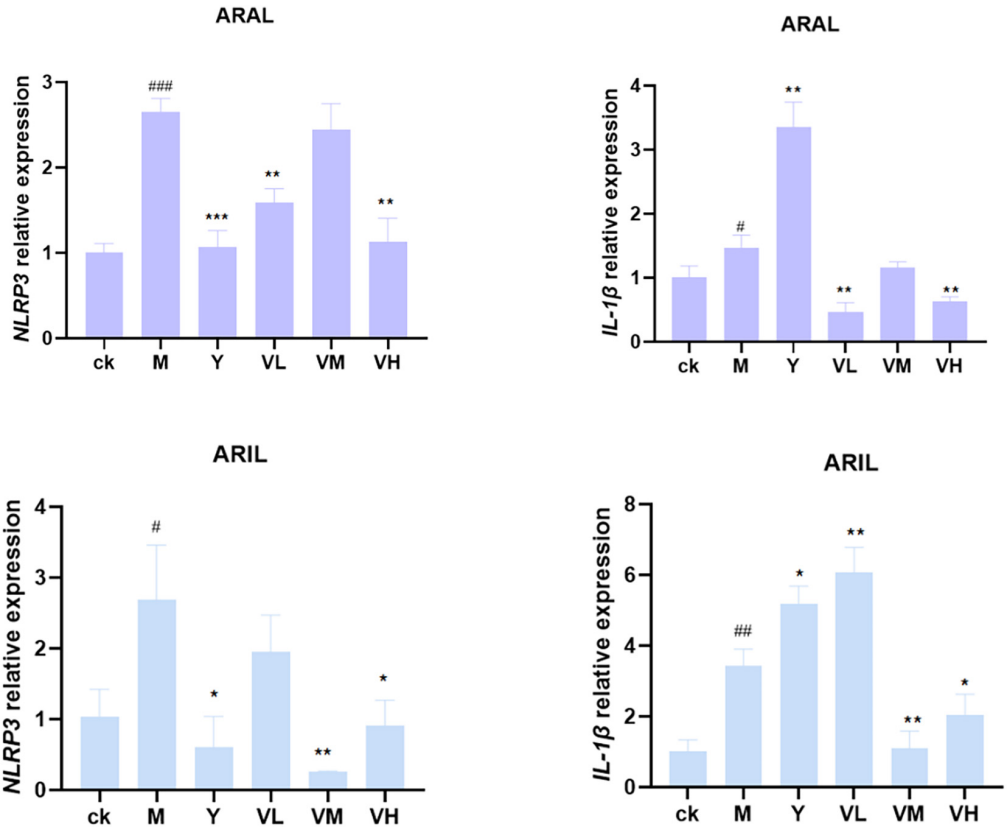


FIGURE 13
Relative expression of NLRP3 and IL-1β mRNA in RAW264.7 cells. After administration, NLRP3 and IL-1 β mRNA expression decreased (mean ± *SD*, *n* = 3). ***$P < 0.001$, **$P < 0.01$, *$P < 0.1$ compared with that of the LPS group; ###$P < 0.001$, ##$P < 0.01$, #$P < 0.1$ compared with that of the control group.

essential oils were the same between *A. argyi* and *A. indica*. The total flavonoids varied in different growth stages, but there were more than 180 flavonoids similar between the two species. In addition, the content of volatile oils and flavonoids, such as eupatilin, prunetin, and artemitin, will increase with the growth of plants.

The network pharmacology results showed that both *A. argyi* and *A. indica* can treat CG. Three flavonoids in

A. argyi and *A. indicia,* quercetin, luteolin, and kaempferol, play critical roles in the entire network, and they may be significant active compounds for the treatment of CG. Quercetin and kaempferol have previously been shown to be active ingredients in the treatment of gastric ulcers (Wang et al., 2019). Studies have shown that quercetin has good anti-inflammatory and antibacterial effects (Yao et al., 2016). It can effectively inhibit *Helicobacter pylori*, which is closely related to CG treatment. Overman et al. (2011) found that quercetin in macrophages and adipocytes can decrease the expression levels of the inflammatory genes *TNF-α, IL-6, IL-1*β, and *COX-2*, suppressing the activation of nuclear factor (NF-κB) and c-Jun N-terminal kinase (JNK). Luteolin, a flavonoid commonly found in medicinal plants, also has anti-inflammatory properties (Skiba et al., 2016). Wang et al. investigated the expression of inflammatory cytokines in macrophages. Luteolin can inhibit the expression of iNOS, IL-1β, IL-6, TNF-a, and CD86 (Wang et al., 2017). Kaempferol can inhibit the translocation of CagA and VacA proteins *in vitro* and downregulate pro-inflammatory cytokines (Miguel, 2009). Tang et al. found that kaempferol exerts an inhibitory effect on the activation of NF-κB and Akt in LPS plus ATP-stimulated cardiac fibroblasts and thus decreases the release of TNF-α, IL-1β, IL-6, and IL-18 (Tang et al., 2015). According to our content analysis, luteolin content was not the highest in ARAL and ARIL, which indicated that the anti-CG effect of *A. argyi* and *A. indica* is the result of multiple pathways, multiple targets, and synergistic effects of multiple compounds. The anti-CG action and mechanism of luteolin need further experimental verification.

In addition, *A. argyi* and *A. indica* and Dragon Boat Festival seem to be inseparable. In some places, people harvest *A. argyi* and *A. indica* on the Day of Dragon Boat Festival (Xue et al., 2019). Cell experiments have shown that *A. argyi* harvested in June had the strongest therapeutic effect on CG. Combined with the metabolomics and content determination results, we speculated that this might be the result of a combination of components. According to the similarities and differences between *A. argyi* and *A. indica*, we believe that these species can be used together in food. However, if used as medicine, we must pay close attention to the appropriate dosage and administration when mixing these species.

Conclusion

To the best of our knowledge, this study was the first to explore the pharmaphylogeny of *A. argyi* and *A. indica* from multiple perspectives. The correlation between geographical distribution, traditional therapeutic effect, pharmacological action, and chemical composition of *A. argyi* and *A. indica* were analyzed. Our results clarified the reasons for the mixed use of *A. argyi* and *A. indica* in folk remedies and showed that the treatment of CG of *A. argyi* was better than that of *A. indica*, which may be the result of the accumulation of various compounds. The results provided a basis for the substitution or mixed use of *A. argyi* and *A. indica*.

Data availability statement

The data presented in this study are deposited in the figshare repository, doi: 10.6084/m9.figshare.19915000.

Author contributions

ML, ZZ, and XH conceived and managed the project and reviewed the manuscript. ZC, SL, and JC conducted the experiments and wrote the manuscript. EZ, QL, BZ, CL, and ML interpreted the data. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the National Key Research and Development Program of China (grant no. 2017YFC1700704), the National Natural Science Foundation of China (grant no. 81803661), the Science and Technology Open Cooperation Project of Henan Province of China (grant no. 172106000053), and the China Agriculture Research System of MOF and MARA (grant no. CARS-21).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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.2022.949743/full#supplementary-material

References

Abad, M. J., Bedoya, L. M., Apaza, L., and Bermejo, P. (2012). The *Artemisia* L. Genus: A review of bioactive essential oils. *Molecules* 17, 2542–2566. doi: 10.3390/molecules17032542

Chen, C. J., Luo, D. D., Miao, Y. H., Guo, L. P., and Liu, D. H. (2021). Analysis and evaluation on leaf quality of different Artemisia argyi germplasm resources. *Chin. J. Exp. Tradit. Med. Form.* 27, 129–136. doi: 10.13422/j.cnki.syfjx.20210511

Chinese Pharmacopoeia Commission (2020). *Pharmacopoeia Of The People's Republic Of China*. Beijing: Chinese Medical Science and Technology Press.

Editorial Board of China Bencao (1999). *China Bencao*. Shanghai: Shanghai Science and Technology Press.

Huang, H. B., and Liu, X. C. (1999). Pharmacognostical identification of *Artemisia argyi* and *Artemisia indica*. *J. Chin. Med. Mater.* 22, 283–287. doi: 10.13863/j.issn1001-4454.1999.06.006

Lan, X. Y., Zhang, Y., Zhu, L. B., Liu, D. H., Huang, X. Z., Zhou, L., et al. (2020). Research progress on chemical constituents from Artemisiae Argyi Folium and their pharmacological activities and quality control. *China J. Chin. Mater. Med.* 45, 4017–4030. doi: 10.19540/j.cnki.cjcmm.20200714.201

Lan, Z. H., Tian, X. F., Shi, Y. H., Gao, R. R., Yin, Q. G., Xiang, L., et al. (2022). Chloroplast genome structure characteristics and phylogenetic analysis of *Artemisia indica*. *China J. Chin. Mater. Med* doi: 10.19540/j.cnki.cjcmm.20220713.101[Epub ahead of print].

Li, M. H. (2008). *A Pharmacophylogenetic Study of Salvia l .(lamiaceae) in China*. Beijing: Peking Union Medical College.

Li, S., Zhou, S. B., Yang, W., and Meng, D. L. (2018). Gastro-protective effect of edible plant *Artemisia argyi* in ethanol-induced rats via normalizing inflammatory responses and oxidative stress. *J. Ethnopharmacol.* 214, 207–217. doi: 10.1016/j.jep.2017.12.023

Liang, Z. L. L. K. S., Zhu, Y. L., Fang, F., and Ye, Y. Q. (2010). Study on germination characteristics of Artemisia indica Willd. seeds. *J. Anhui Agric. Sci.* 38, 13034–13036. doi: 10.13989/j.cnki.0517-6611.2010.24.087

Liu, Y. K., and Christer, E. (2017). New specific primers for amplification of the Internal Transcribed Spacer region in Clitellata (*Annelida*). *Ecol. Evol.* 7, 10421–10439. doi: 10.1002/ece3.3212

Miguel, L. L. (2009). Distribution and biological activities of the flavonoid luteolin. *Mini Rev. Med. Chem.* 9, 31–59. doi: 10.2174/138955709787001712

Overman, A., Chuang, C. C., and Mclntosh, M. (2011). Quercetin attenuates inflammation in human macrophages and adipocytes exposed to macrophageconditioned media. *Int. J. Obes.* 35, 1165–1172. doi: 10.1038/ijo.2010.272

Qin, W. H., Huang, K. N., and Huang, H. X. (2012). Affect of different processing methods on flavonoids content and analgesic effect in *Artemisiae indica* from Guangxi. *Chin. J. Exp. Tradit. Med. Form.* 18, 51–53. doi: 10.13422/j.cnki.syfjx.2012.12.025

Ren, Y. Y., Jiang, N. P., Liu, R. Y., Song, L. K., Tan, R., and Gu, J. (2017). ITS2 sequence analysis and identification of medicinal *Artemisia* plants. *China J. Chin. Mater. Med.* 42, 1395–1400. doi: 10.19540/j.cnki.cjcmm.20170222.016

Skiba, M. A., Szendzielorz, K., Mazur, B., and Krol, W. (2016). The inhibitory effect of flavonoids on interleukin-8 release by human gastric adenocarcinoma (AGS) cells infected with cag PAI (+) *Helicobacter* pylori. *Cent. Eur. J. Immunol.* 41, 229–235. doi: 10.5114/ceji.2016.63119

Song, P. S., Zhang, B. C., Wei, Y. L., and Liu, H. F. (1994). Investigation and resource distribution of original plants of Gansu Artemisia argyi medicinal materials. *J. Chin. Med. Mater* 17, 15–16. doi: 10.13863/j.issn1001-4454.1994.09.006

Tang, X. L., Liu, J. X., Dong, W., Li, P., Li, L., Hou, J. C., et al. (2015). Protective effect of kaempferol on LPS plus ATP-induced inflammatory response in cardiac fibroblasts. *Inflammation* 38, 94–101. doi: 10.1007/s10753-014-0011-2

Wang, A. M., Li, R. S., Ren, L., Gao, X. L., Zhang, Y. G., Ma, Z. M., et al. (2018). A comparative metabolomics study of flavonoids in sweet potato with different flesh colors (Ipomoea batatas (L.) Lam). *Food Chem.* 260, 124–134. doi: 10.1016/j.foodchem.2018.03.125

Wang, S. C., Tu, H., Wan, J., Chen, W., Liu, X. Q., 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, S., Cao, M., Xu, S., Zhang, J., Wang, Z., and Mao, X. (2017). Effect of luteolin on inflammatory responses in RAW264.7 macrophages activated with LPS and IFN-γ. *J. Funct. Foods* 32, 123–130. doi: 10.1016/j.jff.2017.02.018

Wang, Y., Sun, Y. W., Wang, Y. M., Ju, Y., and Meng, D. L. (2019). Virtual screening of active compounds from *Artemisia argyi* and potential targets against gastric ulcer based on Network pharmacology. *Bioorg. Chem.* 88:102924. doi: 10.1016/j.bioorg.2019.102924

Wei, Z. Y., Wu, H. E., and Liang, H. Y. (2009). Analysis of chemical constituents of volatile oil from *Artemisia indica* of Guangxi by GC–MS. *Chin. J. Ethnomedicine Ethnopharmacy* 18, 27–29.

Wu, H. E., Li, Y. H., Wei, Z. Y., and Liang, H. Y. (2008). A comparative study on the chemical constituents of volatile oil from *Artemisia indica*, *Artemisia feddei* and *Artemisia argyi* in Guangxi province. *China Med. Herald* 5, 23–26.

Xue, Z. Q., Guo, L. X., Guo, M., Yang, G. Y., Zhang, D., Guo, L., et al. (2019). Study on difference of chemical constituents of Qiai in different harvest periods. *China J. Chin. Mater. Med.* 44, 5433–5440. doi: 10.19540/j.cnki.cjcmm.20190830.202

Yang, M. T., Kuo, T. F., Chung, K. F., Liang, Y. C., Yang, C. W., Lin, C. Y., et al. (2020). Authentication, phytochemical characterization and anti-bacterial activity of two *Artemisia* species. *Food Chem.* 333:127458. doi: 10.1016/j.foodchem.2020.127458

Yao, L., Yao, J. Y., Han, C. Y., Yang, J. X., Chaudhry, M. T., Wang, S. N., et al. (2016). Quercetin. Inflammation and Immunity. *Nutrients* 8:167. doi: 10.3390/nu8030167

Zhang, Y., Kang, L. P., Guo, L. P., and Zhou, A. K. (2017). Herbalogical study of Mugwort leaves and research advances in its application. *Shanghai J. Acupunct. Moxibustion* 36, 245–255. doi: 10.13460/j.issn.1005-0957.2017.03.0245

Zheng, G. H., Xu, Y., Geng, C., and Jiang, Y. Z. (2012). Quantification of effective ingredients in the *Artemisia sacrorum* Ledeb by reverse phase-high performance liquid chromatography. *J. Med. Sci. Yanbian Univ.* 35, 113–116. doi: 10.16068/j.1000-1824.2012.02.003

Zhou, B. C., Liu, Y. B. G. Y., Bi, Y. Q., Xiao, H., Li, M. H. (2022). Mongolian medicine "Bashaga" variety systematization based on pharmacognosy and pharmacophylogeny. *Mod. Chin. Med.* [Epub ahead of print].

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 20 October 2022
DOI 10.3389/fpls.2022.958460

OPEN ACCESS

EDITED BY
Chunnian He,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

REVIEWED BY
Mario Juan Simirgiotis,
Austral University of Chile, Chile
Kunming Qin,
Jiangsu Ocean Universiity, China

*CORRESPONDENCE
Xiaoqin Wang
nywangxiaoqin@163.com

SPECIALTY SECTION
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

RECEIVED 31 May 2022
ACCEPTED 11 August 2022
PUBLISHED 20 October 2022

CITATION
Jia X, Liu Y, Wang SW, Ma J, Yu J, Yue X, Zhang Y and Wang XQ (2022) Screening of metabolic markers present in *Oxytropis* by UHPLC-Q-TOF/MS and preliminary pharmacophylogenetic investigation. *Front. Plant Sci.* 13:958460. doi: 10.3389/fpls.2022.958460

Screening of metabolic markers present in *Oxytropis* by UHPLC-Q-TOF/MS and preliminary pharmacophylogenetic investigation

Xin Jia, Yang Liu, Suwei Wang, Jiannan Ma, Juan Yu, Xin Yue, Ying Zhang and Xiaoqin Wang*

College of Pharmacy, Inner Mongolia Medical University, Hohhot, China

Plants belonging to the *Oxytropis* genus, family Leguminosae, are found throughout the world, with about 80 species mainly distributed in northwest and northeast China. The plants have medicinal properties and many plants have been used as folk medicine for the treatment of colds, inflammation of carbuncle swelling, pain, and different types of bleeding. In recent years, due to the reduced availability of wild resources and increased clinical demand, additional *Oxytropis* species have been used in Mongolian medicine. This study explored the medicinal potential of four *Oxytropis* species, investigating their phylogeny, chemical components, and pharmacological activities. *Oxytropis myriophylla* (Pall) DC., *Oxytropis hirta* Bunge, and *Oxytropis bicolor* Bge. were found to be closely related at the taxonomic level. While previous investigations on the bioactive constituents of *Oxytropis* have been limited and have concentrated largely on flavonoids and saponins, the present study established a novel UHPLC-Q-TOF/MS based on metabolite profiling to comprehensively analyze the chemical composition of the four *Oxytropis* species and to identify marker compounds. A total of 75 compounds were identified from the four species, with 23 identified as characteristic marker components. Twenty-six marker compounds were identified in *O. myriophylla* from different geographical regions. Analysis of pharmacological activity showed that extracts of *O. myriophylla* and *O. hirta* had stronger anti-inflammatory activity than the extracts from the other species. The relationships between the chemical components, traditional curative uses, and pharmacological activities were analyzed to provide a preliminary documentation of the pharmacophylogenetic characteristics of the *Oxytropis* family as a whole. Several marker compounds, including licoricesaponin G2, licoricesaponin J2, and glycyrrhizic acid found in *O. hirta* were found to have effective anti-inflammatory activity, consistent with the traditional application of reducing swelling and healing wounds. This preliminary investigation

into the pharmacophylogeny of the genus *Oxytropis* will contribute to the conservation and exploitation of the medicinal resources of this genus.

KEYWORDS

chemotaxonomy, ethnopharmacology, metabolomics, pharmacophylogenetics, *Oxytropis*

Introduction

There are ∼350 *Oxytropis* species (family Leguminosae) found throughout the world, of which about 80 species are mainly distributed in northwest and northeast China (Elisens and Denford, 1982; Sun and Xu, 1992). Many of these plants have been used as folk medicine for the treatment of colds, inflammation of carbuncle swelling, pain, and different types of bleeding (Batsuren et al., 1992). In recent years, due to the dwindling availability of wild resources and increasing clinical demand, additional *Oxitropis* species are being used in Mongolian medicine. Two species, in particular, *Oxytropis bicolor* Bunge and *Oxytropis racemosa* Turcz., are abundant and are often used as substitutes in clinical applications. All four *Oxytropis* species are widely distributed in Inner Mongolia, especially *O. myriophylla*, which is the most commonly used in medicinal applications and is prevalent in the eastern, central, and western regions of Inner Mongolia. Table 1 shows details of the ethnopharmacology of the four *Oxytropis* species, with information on geographical distribution, synonyms (common names), medicinal parts, traditional uses, and pharmacological activities, drawn from both herbal books and published literature.

Despite their extensive usage, there has been extremely limited research on these *Oxytropis* species, especially in terms of their chemical composition. To date, the major bioactive constituents described in *Oxytropis* have been limited to flavonoids and saponins, with minimal investigation into other components. *O. myriophylla* (DY) and *O. hirta* (YM) are the most used *Oxytropis* varieties in Mongolian medicine. The two species have similar morphological characteristics but no specific chemical markers distinguishing them have been reported. The major bioactive components in the two species, based on HPLC-UV and HPLC-ELSD analysis were found to be saponins, flavonoids, and alkaloids (Okawa et al., 2002). However, a literature review (Baimukhambetov, 1973) observed that only a few flavonoids have been found in YM while several studies found that *Oxytropis racemosa* Turcz. (SZ) contained mostly flavonoids with few alkaloids (Song et al., 2010, 2013). In contrast, the major chemical components in *Oxytropis bicolor* Bge. (ES) were triterpenoids (Sun et al., 1991), with little evidence of flavonoids and alkaloids. This lack of information has significantly limited the clinical application of these medicinal plants.

Metabolomics is a powerful method that can detect global metabolite variations and discover specific markers in plant species. Metabolomic investigations are based on liquid chromatography-mass spectrometry (LC-MS), and in recent years, the use of ultra-high performance liquid chromatography-quadrupole time-of-flight mass spectrometry UHPLC-Q-TOF-MS has proved to be extremely useful for the rapid identification of metabolites in herbs due to its unsurpassed sensitivity and high resolution (Zhang et al., 2014; Yu et al., 2017; Zhao et al., 2018; Cui et al., 2020; Liu et al., 2020a). The use of UHPLC-Q-TOF-MS in metabolomics and in combination with different chemometric statistical tools is a versatile technique that can be effectively utilized to discover quality markers in the authentication of diverse herbal medicines (Masson et al., 2014; Wu et al., 2018; Liu et al., 2020b; Pan et al., 2020).

In this study, UHPLC-Q-TOF-MS was used as a rapid and accurate analytical method for the detection and characterization of the chemical constituents of *Oxytropis*. Principal component analysis (PCA) and orthogonal partial least squares discriminant analysis (OPLS-DA) were used to distinguish between the four *Oxytropis* species as well as between *O. myriophylla* obtained from different geographical regions based on specific markers. This method can accurately identify the different metabolites of the different plants, allowing a comprehensive analysis of their chemical components as well as the identification of specific marker compounds for distinguishing between the four species. The anti-inflammatory activity of the four *Oxytropis* species was analyzed *in vitro*. In combination with a review of the literature, this study investigated the chemical composition, pharmacological activities, and preliminary pharmacophylogeny of the species, proposing relationships between ethnopharmacology, pharmacology, and bioactive components. This study can provide a theoretical basis for a better understanding of *Oxytropis* and its utilization.

Materials and methods

Materials and reagents

The whole grass of *Oxytropis myriophylla* (Pall) DC. (DY), *Oxytropis hirta* Bunge. (YM), *Oxytropis racemosa* Turcz. (SZ),

TABLE 1 The distribution, traditional curative effect, and pharmacological activities of four *Oxytropis*.

Plant	Synonyms/Common names	Distribution	Medicinal part	Traditional curative effect	Pharmacological activities	Ref.
O. bicolor	"Rentoucao", "Diding", "Maozhuahua"	Inner Mongolia, Hebei, Shanxi, Shaanxi, Ningxia, Gansu, Qinghai, Henan	Seed	Detoxification and analgesia	Antibacterial	(Chen et al., 2005)
O. hirta	"Oxytropishirta Bunge"	Heilongjiang, Jilin, Liaoning, Inner Mongolia, Hebei, Shanxi, Shandong, Shaanxi, Gansu, Henan, Hubei	Whole grass	Kill "sticky", clear away heat, dry "xieriwusu", callus, regenerate muscle, lock pulse, stop bleeding, reduce swelling and defecate	Antibacterial	(Ye et al., 2022)
O. myriophylla	"*Oxytropis myriophylloides* Hurus."	Inner Mongolia, Heilongjiang, Jilin, Liaoning, Hebei, Shaanxi, Shanxi, Ningxia	Whole grass	Kill "sticky", clear away heat, dry "xieriwusu", heal wounds, regenerate muscles, stop bleeding and reduce swelling	Anti inflammatory and analgesic, antibacterial, antioxidant	(She et al., 2010; Saiyin, 2014; Meng et al., 2016)
O. racemosa	"Paopaocao", "Maozhuazhua" "Yazuidou", "OxytropismandshuricaBge."	Hebei, Shanxi, Inner Mongolia, Shaanxi, Gansu	Whole grass	Promoting digestion	Improve digestive function	(Zhang et al., 2019)

and *Oxytropis bicolor* Bge. (ES) were collected by Yang Liu from the Saihanwula Nature Reserve of Chifeng city in China (44°20′N, 118°30′E, elevation 1,440 m) on 12 August 2019. A further sample of *O.myriophylla* (DYWC) was collected in Wuchuan city, Inner Mongolia, China (41°16′N, 110°08′E, elevation 1,640 m) on 18 August 2019. Each *Oxytropis* contained six samples and were identified by associate professor Bi Qu of the college of pharmacy in Inner Mongolia medical university. The preserved leaf specimens are stored in the medicinal herbarium of the college of pharmacy at Inner Mongolia Medical University.

Reagents for metabolomic analysis included formic acid and methanol (LC-MS grade), purchased from Concord Technology (Tianjin, China) and acetonitrile from Sigma-Aldrich (St. Louis, MO, USA). Ultrapure water was used in all experiments.

For analyzing anti-inflammatory activity *in vitro*, RPMI 1,640 medium and Penicillin-Streptomycin (10,000 U/mL) were purchased from Gibco (USA), fetal bovine serum was purchased from ExCell Bio (Australia), NO kit was purchased from Nanjing Jiancheng (Nanjing, China), Mouse TNF-α ELISA KIT and Mouse IL-6 ELISA KIT purchased from Solarbio (Beijing, China). Mouse RAW 264.7 macrophages were obtained from the cell bank of the Chinese Academy of Sciences (Shanghai, China).

Metabolomic analysis

Dried whole grass was ground to a fine powder and passed through a No. 100 mesh sieve. The powder sample (0.10 g) was dissolved in 5 mL of a 1:1 (v/v) methanol: water solution and soaked at room temperature for 12 h. The mixture was then ultrasonicated for 45 min and centrifuged (13,000 rpm, 10 min, 4°C) to obtain the supernatant.

The chromatographic separation was performed on an ExionLC system (AB Sciex, Foster City, CA, USA). A Waters Acquity BEH C_{18} column (2.1 × 100 mm, 1.7 μm) was used at a temperature of 35°C. The mobile phase consisted of 0.1% formic acid (A) and acetonitrile (B). The gradient conditions were: 0–2 min, 15 → 25% B; 2–6 min, 25 → 40% B; 6–9 min, 40 → 70% B; 9–11 min, 70 → 95% B; 11–13 min, 95% B; maintained at 15% B for an additional 10 min for re-equilibration.

For the high-resolution detection, a 5,600 Q-TOF mass spectrometer (AB Sciex) equipped with an electrospray ionization source (Turbo Ionspray) was used. MS detection was implemented in both the negative and positive ion modes. The analytical conditions used were as follows: gas1 and gas2, 55 psi; curtain gas, 35 psi. heat block temperature, 550°C; ion spray voltage, −4.5 kV in negative ion mode and 5.5 kV in positive ion

mode, respectively; declustering potential, 50 V; collision energy, 40 V. QC samples were used to assess the system reproducibility and stability of the acquisition method by pooling small aliquots of each sample.

Anti-inflammatory activity *in vitro*

For the extraction of total extracts, 10 g of dried medicinal powder from each of the four species was added to 150 mL of 70% ethanol solution. The solutions were extracted three times using heating and reflux (2 h per extraction). The extracts were then combined and concentrated under reduced pressure The total extracts of DY, YM, SZ, and ES were obtained as 1.7723, 1.7963, 1.4352, and 2.7994 g, and the extraction rates were 17.723, 17.963, 14.352, and 27.994%, respectively.

For cell culture and passaging, frozen RAW 264.7 cells were removed from liquid nitrogen and quickly thawed in a 37°C water bath. The supernatant was then removed by centrifugation, 1 mL of RPMI 1,640 was added to the complete medium containing 10% fetal bovine serum to resuspend the cells, after which the cell suspension was transferred to a Petri dish to be cultured in a incubator at 37°C, 5% CO_2. The cell status was observed the next day. When the cells grew to 80–90% of the culture dish, they could be passaged. From the beginning of recovery, it is recorded as the first generation. Generally, cells from three to eight generations are selected for experiments. When the cell density was moderate, they were removed from the medium and fresh medium was added after centrifugation to make a cell suspension. The cells were then transferred to a Petri dish in a certain proportion to continue culturing in the incubator at 37°C, 5% CO_2 for later use.

For MTT assays, cells in the logarithmic growth phase were selected, and the cell concentration was adjusted to 1×10^4 cells/mL, then inoculated into a 96-well plate, 100 μL/well, and cultured in a 37°C, 5% CO_2 incubator for 24 h. Different concentrations of the total extracts in the medium were added to cells in the experimental group, while equal volumes of the medium were added to cells in the control group. The cells were incubated with the extracts for 24 h. About 10 μL of MTT solution at 5 mg/mL was added per well and incubated in the dark for 4 h. The supernatant was then removed and 100 μL/well dimethyl sulfoxide (DMSO) was added and mixed evenly. After the purple crystal (formazan) was completely dissolved, the optical density (OD) of each well was measured using a microplate reader at the wavelength of 570 nm. The experiment was repeated three times to calculate the cell survival rate of each group.

$$\text{Cell Viability} = \left[\left(\text{As} - \text{Ab}\right) / \left(\text{Ac} - \text{Ab}\right)\right] \times 100 \quad (1)$$

As: OD_{570} of experimental wells (medium containing cells, substance to be tested).
Ac: OD_{570} of control wells (medium containing cells, no substance to be tested).
Ab: OD_{570} of blank wells (medium without cells and substances to be tested).

For the establishment of the inflammatory cell model, RAW 264.7 cells in the logarithmic growth phase were selected, and the cell concentration was adjusted to 5×10^5 cells/mL, and then inoculated in a 24-well plate, 500 μL/well, and cultured in a 37°C, 5% CO_2 incubator for 24 h. The experimental group was incubated with different concentrations of lipopolysaccharide (LPS) (0.01, 0.1, 1, and 10 μg/mL), while the control group received only a culture medium. The NO concentrations were measured at 12, 24, and 48 h using a kit, according to the instructions. The experiment was repeated three times.

For the effect of LPS-induced RAW 264.7 cells on NO production and inflammatory factors, RAW 264.7 cells in the logarithmic growth phase were selected, and the cell concentration was adjusted to 5×10^5 cells/mL, and then inoculated in a 24-well plate, 500 μL/well, and cultured in a 37°C, 5% CO_2 incubator for 24 h. The cells were pretreated with different concentrations of total extracts for 1 h and then LPS was added for a total of 24 h. The positive drug group was added with LPS and indomethacin (INM, 100 μM). The model control group received only LPS, the blank control group received only the medium, and the NO concentrations were determined according to the instructions of the NO, TNF-α, and IL-6 kit. The experiment was repeated three times.

Data processing and statistical analysis

The original obtained map was converted by XCMS. This included baseline filtering, peak identification, peak alignment, and other steps, and provided the data matrix, including the mass charge ratio (m/z), retention time (Rt), peak area (intensity), and other details. All data were normalized by the total peak area, and the generated Excel table was used for subsequent metabolome analysis. To reduce the signal interference caused by accidental error, the variables with RSD $\geq$ 40% in the quality control (QC) assessments were eliminated in Excel. The excel file was imported into SIMCA 14.1 (Umetrics, Umeå, Sweden) software for multivariate analysis, mainly PCA and OPLS-DA. The variable importance in projection (VIP) and S-plot were used to identify different components. The data of *in-vitro* anti-inflammatory activity experiments were statistically analyzed and plotted using GraphPad Prism 9.0 software. T-test was used for comparison between groups, and $P < 0.05$ was considered statistically different, $P < 0.01$ was

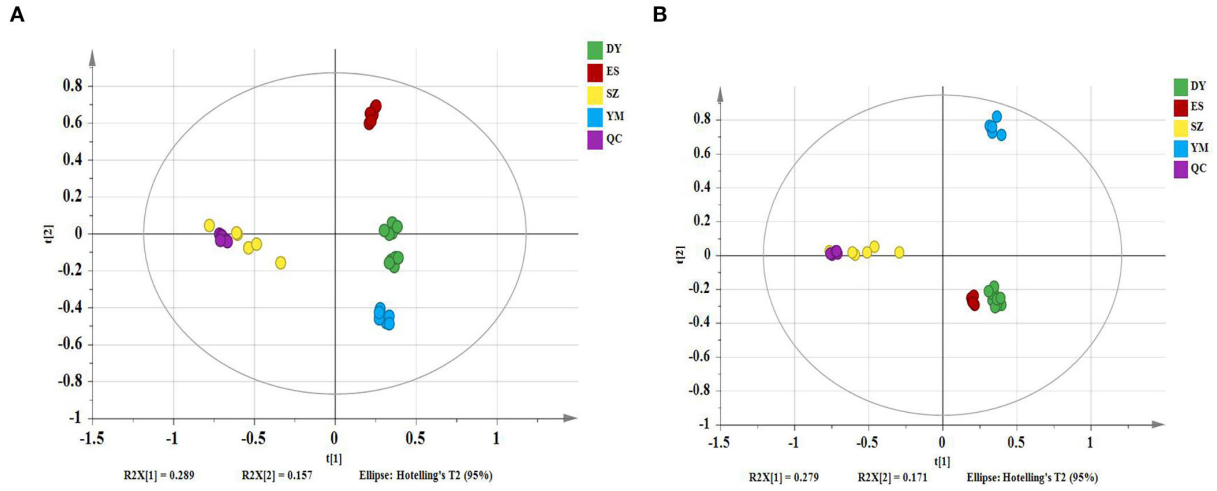


FIGURE 1
The principal component analysis (PCA) scatter plot of different *Oxytropis* species [**(A)** positive mode; **(B)** negative mode].

considered statistically significantly different, and $P < 0.001$ was considered highly statistically different.

Results

Assignment of metabolic markers present in *Oxytropis* by UPLC-Q-TOF-MS

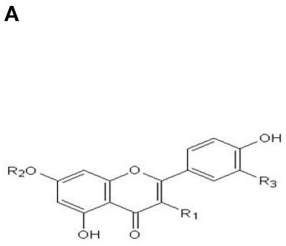
PCA analysis of four *Oxytropis*

The PCA models containing all samples were first established to assess the distribution of the QC samples and the distances between the four species. The model showed 28.9% of variations in X (R2X[1] = 0.289), 15.7% of variations in response X (class) (R2X[2] = 0.157), and 60.3% of variations in response Y (Q2Y = 0.603) in the positive mode (Figure 1A). The model described 27.9% of the variations in X (R2X[1] = 0.279), 17.1% of the variations in response X (class)(R2X[2] = 0.171), and 62.2% of the variations in response Y (Q2Y = 0.622) in the negative mode (Figure 1B). It can be seen from the figures that under the positive and negative ion modes, the QC samples are closely gathered together, confirming both the stability and repeatability of the experiment and that the data were stable, reliable, and effective. In addition, it was clear that the different *Oxytropis* samples could be distinguished, indicating significant differences in the chemical components of the different *Oxytropis* varieties.

Global identification of chemical components in four *Oxytropis*

The four different varieties of *Oxytropis* were globally characterized by UHPLC-QTOF-MS. As the differences between the four species could be directly observed from base peak chromatography (BPC), Supplementary Figure S1 shows the positive and negative ion modes in the BPC diagram of the *Oxytropis* samples. Additionally, using the information from MS/MS chromatogram together with information retrieved from the literature (Sun and Chen, 1997; Song et al., 2010, 2013; Li et al., 2012; Masafumi et al., 2022) and public databases, a total of 75 compounds were identified (Supplementary Table S1). Compounds in each *Oxytropis* species were classified by their retention times, fragment ions, accurate molecular mass, and credible molecular formulas and chemical names. Compounds belonging to various classes were identified in the four species, including 42 flavonoid glycosides, 14 saponins, 5 alkaloids, 5 amino acids, 3 organic acids, and 6 other compounds. Figure 2 shows the typical chemical structures of the flavonoid glycosides and saponins identified in *Oxytropis*. The flavonoid glycosides and saponins formed the two largest categories among the 75 identified compounds.

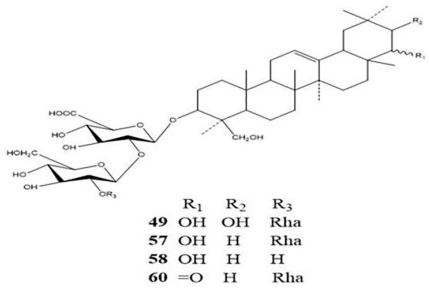
Oxytropis myriophylla (DY) is the most commonly used species in Mongolian medicine, and it contained compounds present in all four *Oxytropis* varieties**.** In all, 43 compounds were detected in DY, including 21 flavonoid glycosides, 6 saponins, 3 alkaloids, and 5 amino acids, among other compounds (Supplementary Table S1). The majority of these compounds have been reported to have significant biological activities. For instance, the flavonoid glycoside kaempferol 3-caffeylrobinobioside-7-rhamnoside (Compound 23) was only detected in DY. Figure 3A shows the MS/MS mass spectrum of compound 23. The precursor ion $[M-H]^-$ was at *m/z* 901.2424 found at 3.258 min. The compound formula was predicted to be $C_{42}H_{46}O_{22}$. The different daughter ions at *m/z*



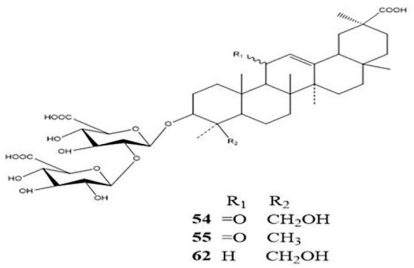
A

No.	R_1	R_2	R_3
6	*O*-glu-(1→6)-glu	rha	OH
7	*O*-[glu-(1→2)gal]-(1→6)-rha	rha	H
9	*O*-glu-(1→6)-glu	rha	H
11	*O*-glu-(1→6)-rha	rha	H
12	*O*-6-*trans*-caf-glu-(1→2)-glu	rha	OH
16	*O*-2-*trans*-caf-glu-(1→3)-glu/*O*-4-*trans*-caf-glu-(1→3)-glu	rha	H
18	*O*-6-*trans*-caf- glu-(1→2)-rha	rha	OH
19	*O*-6-acetyl-gal-(1→2)-rha	rha	H
20	*O*-glu-(1→6)-rha-(1→2)-api/xyl	H	OCH_3
22	*O*-6-*trans*-cin-glu-(1→2)-rha	glu	H
23	*O*-4-*trans*-caf-glu-(1→6)-rha	rha	H
32	*O*-glu-(1→6)-xyl	H	OCH_3
33	*O*-glu	H	OH
34	*O*-gluA	H	H
35	6-malonyl- glu-(1→2)-rha	H	H
38	*O*-glu	H	OCH_3
41	*O*-glu-(1→2)-glu	CH_3/H	OH/OCH_3
42	*O*-6-malonyl- glu	CH_3/H	OH/OCH_3
47	*O*-glu	CH_3	OH
50	H	glu	OCH_3
52	H	6-malonyl-glu	OCH_3
70	*O*-glu-(1→6)-rha	H	H
72	*O*-rha-(1→4)acetyl-glu	H	H
73	*O*-gal-2-acetyl-ara	H	OCH3

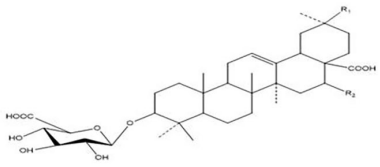
B



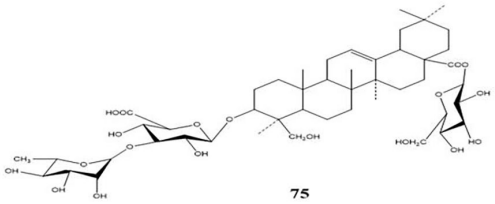
	R_1	R_2	R_3
49	OH	OH	Rha
57	OH	H	Rha
58	OH	H	H
60	=O	H	Rha



	R_1	R_2
54	=O	CH_2OH
55	=O	CH_3
62	H	CH_2OH



	R_1	R_2
66	$COOCH_3$	H
67	CH_3	OH



75

FIGURE 2
Typical chemical structures of flavonoid glycosides and saponins identified from *Oxytropis* [**(A)** flavonoids; **(B)** saponins].

755.1830, 609.1457, and 284.0312 were $[M–H–Rha]^-$, $[M–H–Rha–Rha],^-$, and $[M–H–Rha–Rha–Glc–Caffeyl]^-$, respectively.

Oxytropis hirta (YM) is also a commonly used *Oxytropis* species in Mongolian medicine, and to date, few compounds have been isolated from YM. In this study, ∼30 different compounds were detected in YM. These included 3 alkaloids and 8 saponins as well as other compounds including flavonoid glycosides. Three saponins, namely, licoricesaponin G2, licoricesaponin J2, and glycyrrhizic acid were only found in YM (Supplementary Table S1). Recent studies have reported that licorice, which is a triterpene saponin, shows significant anti-inflammatory, anticancer, and hepatoprotective activities (Tang et al., 2015; Zheng et al., 2015; Yang et al., 2017).

In *Oxytropis racemosa* (SZ), we identified 10 saponins besides flavonoid glycosides and alkaloids that have been previously reported. For instance, triterpenoid-xyl-rha-gluA (Compound 51) was only detected in SZ. The general fragmentation pathways are illustrated in Figure 3B. The precursor ion $[M + H]^+$ *m/z* 943.4882 was observed at 6.831 min, while the other fragments were seen at *m/z* 811.4482, 635.4145, 489.3559, 471.3471, and 453.3365. Interestingly, the presence of *m/z* 811.4482 fragments indicated the loss of xylose from the precursor ion, MS/MS fragments *m/z* 635.4145 suggested the loss of glucuronic acid from the product ion at *m/z* 811.4482, and *m/z* 489.3559 indicated a loss of rhamnose monohydrate from the product ion at *m/z* 635.4145, whereas *m/z* 453.3365 indicated loss of H_2O from the product ion at *m/z* 471.3471.

Thirty compounds were detected in *Oxytropis bicolor* (ES), principally 9 flavonoid glycosides, 8 saponins, 2 alkaloids, and other compounds. Of these, 6 flavonoid glycosides were detected only in ES (Supplementary Table S1). These results may provide

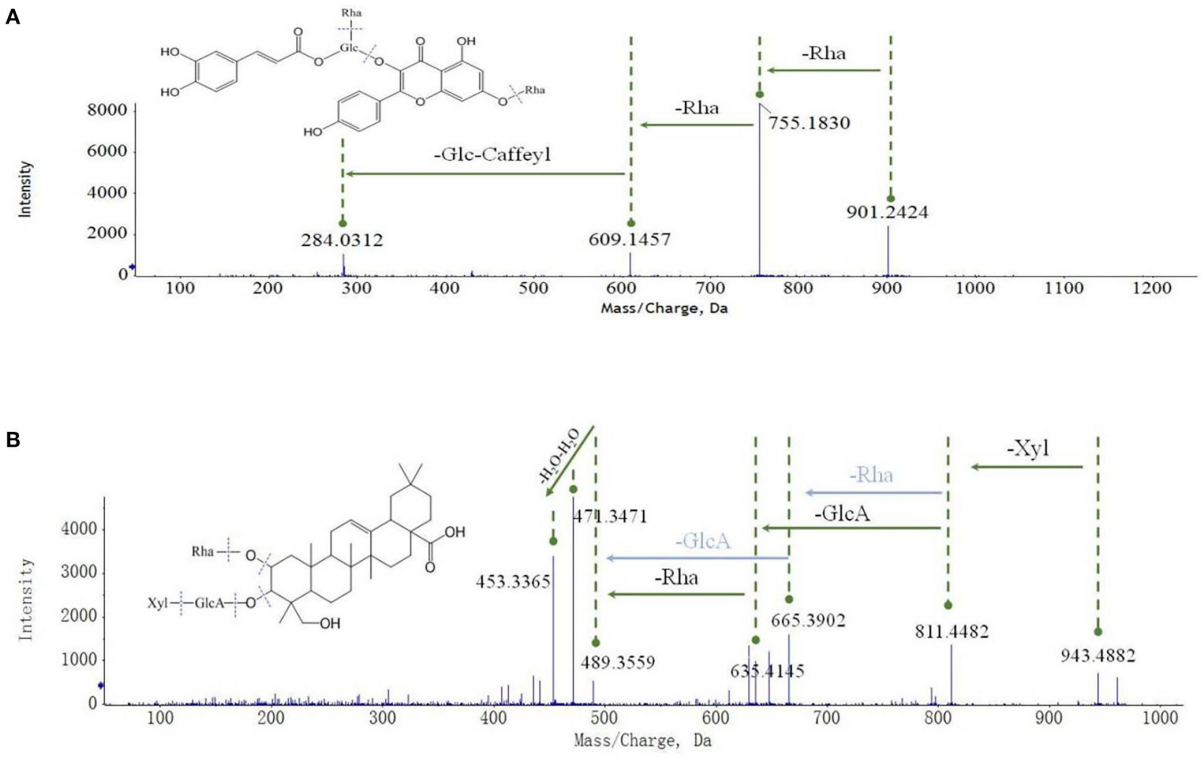


FIGURE 3
Mass spectrum cleavage law of flavonoids and saponins [**(A)** flavonoids; **(B)** saponins].

a meaningful basis to develop and promote the pharmacological use of ES in the future.

Visualization of the distribution of marker compounds by heatmap

A heatmap visualization was generated to provide detailed information on the distribution of the marker compounds present in *Oxytropis*. As shown in Figure 4, the four *Oxytropis* species were clearly classified with the "unsupervised" clustering analysis. The samples were further classified automatically. The results indicated that the identified 75 components served as representatives in classifying *Oxytropis*. In the heatmap, each cell represents an individual compound, with darker colors indicating higher contents. Each sample shows a dark color block, indicating the specific compounds present in the different *Oxytropis* species. For example, quercetin3-*O*-gentiobioside-7-*O*-rha, sarmenoside III, and kaempferol 3-caffeylrobinobioside-7-rhamnoside were highly enriched in DY, while licoricesaponin G2, glycyrrhizic acid, and licoricesaponin J2 were significantly enriched in YM (Figure 4). DY and ES contained the highest proportions of the 75 identified compounds and the compound proportion was higher in DY than in DYWC. According to the distribution of each component in the heatmap, 23 potential marker compounds (Table 2) were identified to establish a Support Vector Machine (SVM) model for classification and prediction. In the model, black is the classification result (training set) of modeling, and red is the prediction result (test set) after modeling. The classification accuracy is 100%. The two results overlap, indicating that the model has good prediction ability and can be used for the identification of different basic sources of *Oxytropis*. The results are shown in Supplementary Figure S2.

As reported earlier, flavonoid glycosides and saponins were found to be the main bioactive compounds in *Oxytropis*. Moreover, in our study, the relative concentrations of some flavonoid glycoside and saponin compounds were different in DY plants from different geographical origins. Therefore, further studies are required to understand the basis of the difference.

Comparison of marker compounds in *O. myriophylla* obtained from different geographical regions

To identify potential marker compounds and distinguish between *O. myriophylla* plants from different geographical regions, a supervised OPLS-DA was conducted. The grouping variable area was adopted with pair-wise analysis which can

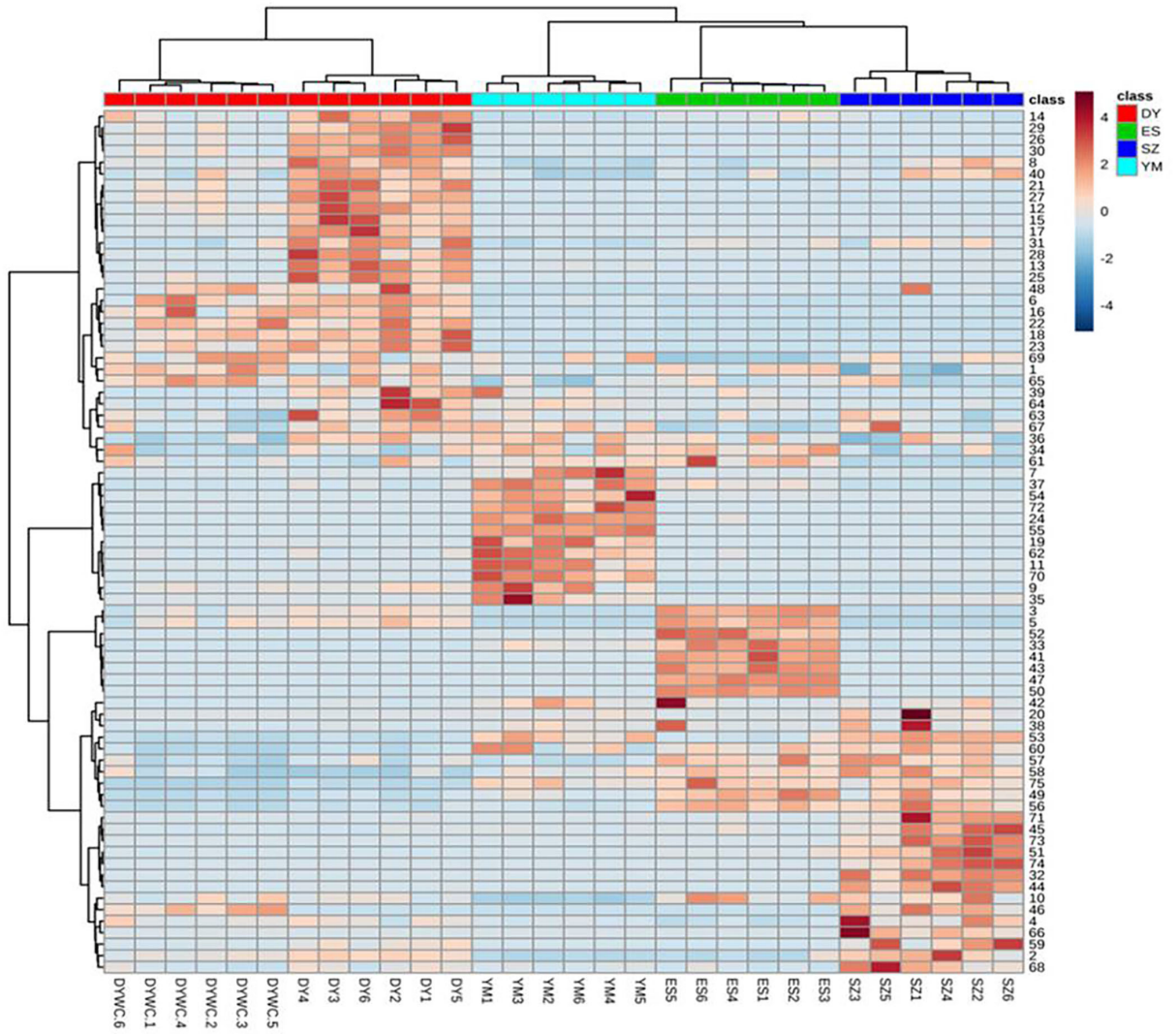


FIGURE 4
The heat map of identified differential metabolites of *Oxytropis*. (Darker brown indicates higher contents. The figures on the right correspond to compound number order in Supplementary Table S1).

make the differences clearer and more straightforward. OPLS-DA has been regularly used to screen different metabolites. Figure 5 shows the OPLS-DA of DY and DYWC in both the positive and negative modes, indicating that the two groups could be clearly distinguished.

Variables with VIPs > 1 in the model were considered to contribute significantly to differences between the plants. Partial correlation coefficients (Pcorrs) were also used to identify the most influential and important variables. While the Pcorr value itself does not indicate any uniform screening principles, higher Pcorr values indicate greater contributions to the model. In this experiment, Pcorr > 0.5 was set as the screening criterion. As seen in Figure 6, the S-plot can also visually indicate the exact contribution of each variable. The points further from the origin of the e-coordinates in the "S" segment clearly showed that the compounds from different geographical origins were significantly different (Li et al., 2018).

After screening variables using VIP and Pcorr, the Mann-Whitney *U* test was used to calculate the *p*-values. A total of 26 potential marker compounds were identified in the two DY plant samples from different geographical regions using the *S* scatter plot. The results are shown in Supplementary Table S2. Interestingly, while no major differences in the types of compounds identified were observed between DY and DYWC, significant differences in their concentrations were found with higher concentrations apparent in DY plants from Saihanwula. For example, the concentrations of several prominent amino acids such as DL-phenylalanine

10.3389/fpls.2022.958460

TABLE 2 Components for SVM establishment.

No.	Identification	Formula	Class	Source
6	Quercetin 3-*O*-gentiobioside-7-*O*-rha	$C_{33}H_{40}O_{21}$	Flavonoid glycoside	DY
16	Kaempferol 3-(2”-caffeyllaminaribioside)-7-rhamnoside/kaempferol 3-(4”-caffeyllaminaribioside)-7-rhamnoside	$C_{42}H_{46}O_{23}$	Flavonoid glycoside	DY
18	Sarmenoside III	$C_{42}H_{46}O_{23}$	Flavonoid glycoside	DY
22	3-(*β*-D-Glucopyranosyloxy)-4',5-dihydroxy-7-[2-*O*-[6-*O*-(3-methoxy-4-hydroxy-trans-cinnamoyl)-*β*-D-glu]-*α*-L-rhamnopyranosyloxy]flavone	$C_{43}H_{48}O_{23}$	Flavonoid glycoside	DY
23	Kaempferol 3-caffeylrobinobioside-7-rhamnoside	$C_{42}H_{46}O_{22}$	Flavonoid glycoside	DY
41	Rhamnetin 3-sophoroside/isorhamnetin 3-laminaribioside/rhamnetin 3-laminaribioside	$C_{28}H_{32}O_{17}$	Flavonoid glycoside	ES
43	Chrysoeriol-7-*O*-(2”-*O*-mannopyranosyl)allopyranoside/complanatuside	$C_{28}H_{32}O_{16}$	Flavonoid glycoside	ES
45	3-[6-*O*-[4-*O*-(4-Oxo-4-hydroxybutyryl)-6-deoxy-*α*-L-mannopyranosyl]-*β*-D-glu]-3',4',5,7-tetrahydroxyflavone	$C_{31}H_{34}O_{19}$	Flavonoid glycoside	ES
47	Rhamnetin 3-*O*-*β*-glucopyranoside	$C_{22}H_{22}O_{12}$	Flavonoid glycoside	ES
50	Thermopsoside/kaempferide 7-glucoside	$C_{22}H_{22}O_{11}$	Flavonoid glycoside	ES
52	Luteolin 3'-methyl ether 7-malonylglucoside/quercetin 3-(3”,6”-diacetylgalactoside)	$C_{25}H_{24}O_{14}$	Flavonoid glycoside	ES
32	Isorhamnetin 3-*O*-[*β*-D-xylopyranosyl-(1 → 6)-*β*-D-glucopyranoside]	$C_{27}H_{30}O_{16}$	Flavonoid glycoside	SZ
44	[(2R,3R)-2-(3,4-dihydroxyphenyl)-5,7-dihydroxy-3,4-dihydro-2H-chromen-3-yl]oxymethyl hydrogen carbonate	$C_{17}H_{16}O_9$	Flavonoid glycoside	SZ
51	Triterpenoid-xyl-rha-gluA	$C_{47}H_{74}O_{19}$	Saponin	SZ
71	Herbacetin 7-(6”-quinoylglucoside)	$C_{28}H_{30}O_{17}$	Flavonoid glycoside	SZ
73	Isorhamnetin 3-(2”'-acetyl-α-arabinopyranosyl)-(1 → 6)-galactoside/tricetin 4'-methyl ether 7-apiosyl-(1 → 2)-(6”-acetylglucoside)	$C_{29}H_{32}O_{17}$	Flavonoid glycoside	SZ
74	Unknown	$C_{31}H_{34}O_{18}$	Flavonoid glycoside	SZ
19	Kaempferol 3-(2”-rhamnosyl-6”-acetylgalactoside) 7-rha	$C_{35}H_{42}O_{20}$	Flavonoid glycoside	YM
24	Unknown	$C_{15}H_{19}NO_8$	Glucoside	YM
54	Licoricesaponin G2	$C_{42}H_{62}O_{17}$	Saponin	YM
55	Glycyrrhizic acid	$C_{42}H_{62}O_{16}$	Saponin	YM
62	Licoricesaponin J2	$C_{42}H_{64}O_{16}$	Saponin	YM
72	3-[4-*O*-(6-*O*-acetyl-*β*-D-glu)-α-L-rhamnopyranosyloxy]-4',5,7-trihydroxyflavone/kaempferol 3-*O*-(6”-*O*-acetyl)glucoside-7-*O*-rhamnoside/multiflorin A	$C_{29}H_{32}O_{16}$	Flavonoid glycoside	YM

and proline were almost 0.7-fold higher in DY compared with DYWC. Proline is widely considered to play a key role in maintaining osmotic balance and thus has an important osmoprotective function in plant defense mechanisms. As an osmotic agent, the different concentrations of this amino acid in plants from different regions suggest a reflection of

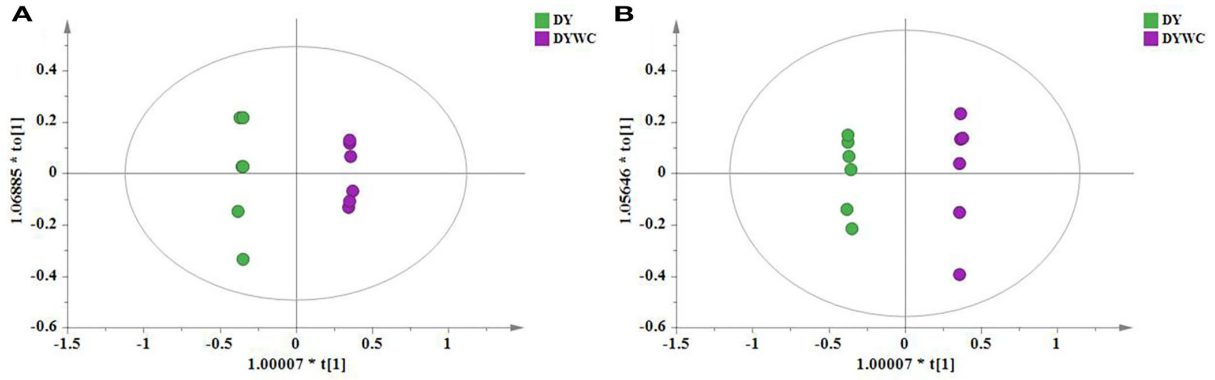


FIGURE 5
OPLS-DA of DY and DYWC [**(A)** positive mode; **(B)** negative mode].

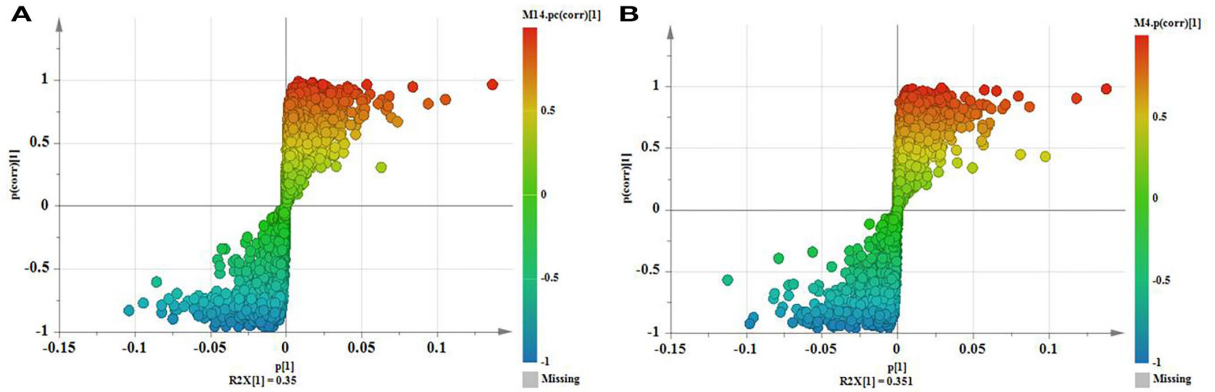


FIGURE 6
S-Plot of DY and DYWC [**(A)** positive mode; **(B)** negative mode].

different environmental conditions between the areas (Szabados and Savoure, 2010; Sharma et al., 2011). Aromatic amino acids, such as phenylalanine, serve as precursors for the biosynthesis of polyphenols and flavonoids. This can possibly explain the reason for the higher concentrations of most components seen in DY plants from Saihanwula compared with Wuchuan. For instance, 13 flavonoid glycoside compounds were found to be present in higher concentrations in DY from Saihanwula with kaempferol 3-caffeylrobinobioside-7-rhamnoside and additional unknown compounds serving as representatives. Additionally, few saponins, alkaloids, and other compounds were found. Similarly, the SVM model was established for the classification and prediction of DY plants from the two different geographical regions. The results showed a 100% classification accuracy, and the predicted results can overlap with the actual results, thereby indicating that the model has good predictive power (Supplementary Figure S3). This model can potentially be used to distinguish between DY plants from different geographical regions.

Anti-inflammatory activity of *Oxytropis* extracts

Effects of extracts on cell proliferation

All four *Oxytropis* extracts were non-cytotoxic to RAW 264.7 cells at mass concentrations up to 200 μg/mL, and there was no significant difference compared with the control group. Therefore, the extract concentrations of 12.5, 50, and 200 μg/mL (low-, medium-, and high-dose groups) were selected for subsequent experiments. The MTT experimental results of the four *Oxytropis* extracts are shown in Figure 7.

Establishment of LPS-induced RAW 264.7 macrophage inflammation model in mouse

LPS concentrations of 0.01, 0.1, 1, and 10 μg/mL were used. The NO production in RAW 264.7 cells at 12, 24, and 48 h is shown in Figure 8. When the LPS concentration was 1 μg/mL for 24 h, NO production by the cells was highest, showing

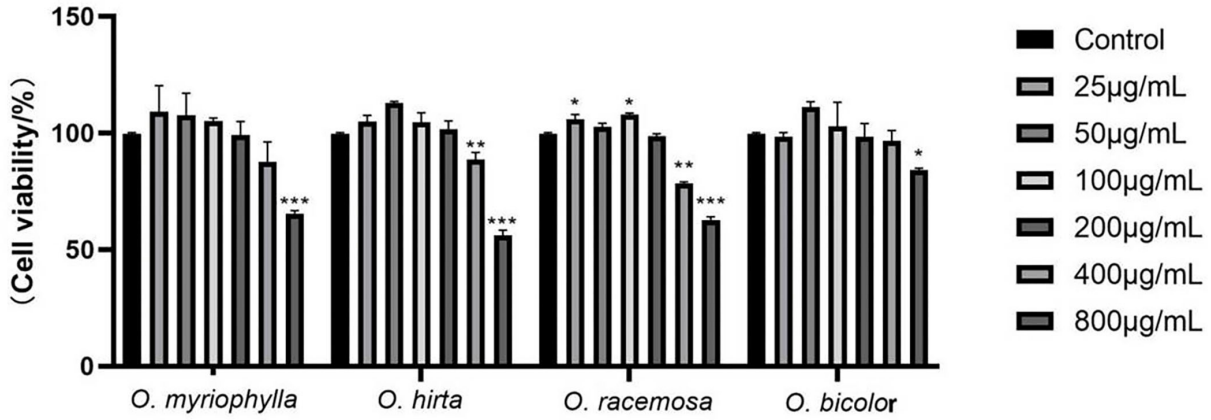


FIGURE 7
Effects of the total extracts of the four *Oxytropis* on the proliferation of mouse RAW264.7 macrophages (Compared with the control group, $^{*}P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$).

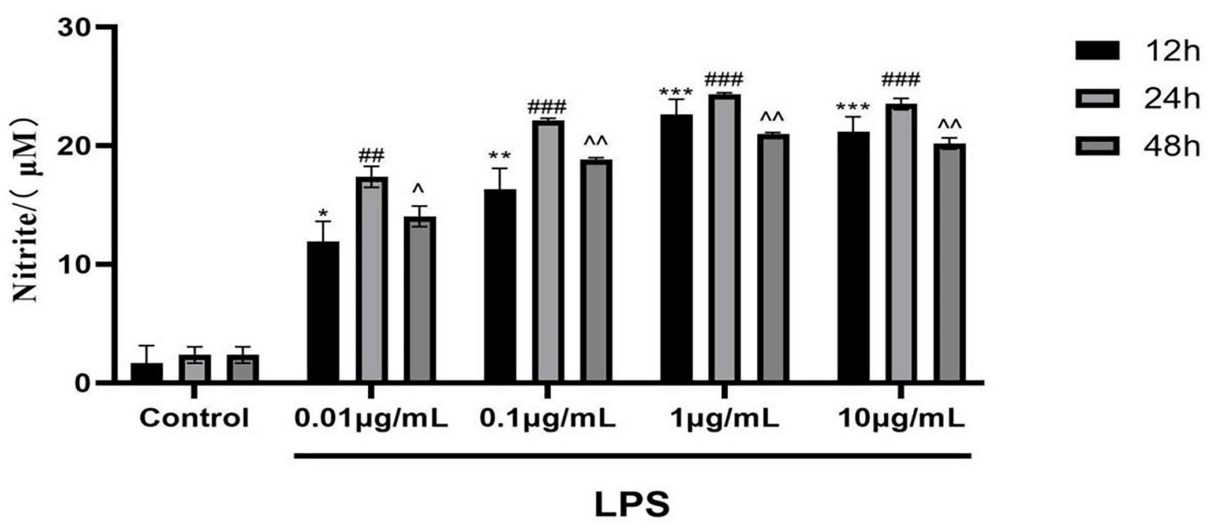


FIGURE 8
Effects of different concentrations of LPS for different times on NO production in mouse RAW 264.7 macrophages (Compared with the control group, at 12, 24 and 48 h respectively, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$; $^{\#\#}P < 0.01$, $^{\#\#\#}P < 0.001$; $^{\wedge}P < 0.05$, $^{\wedge\wedge}P < 0.01$).

a highly significant difference in comparison with the blank control group ($P < 0.001$). These results indicate the successful establishment of the LPS-induced RAW 264.7 macrophage inflammation model. Therefore, follow-up experiments were carried out with an LPS concentration of 1 μg/mL and an incubation time of 24 h as the optimal modeling conditions.

Effects of LPS-induced NO production in mouse RAW 264.7 macrophages

The effects of the four total extracts on LPS-induced NO production in RAW 264.7 cells are shown in Figure 9. The NO levels in the cultured supernatants of each group were significantly higher than those in the blank control group, indicating the successful establishment of the RAW 264.7 macrophage inflammation model. At low, medium, and high doses, the different total extracts could significantly reduce NO levels in the cultured supernatants ($P < 0.001$, $P < 0.01$, or $P < 0.05$), and had different anti-inflammatory effects. Specifically, *O. myriophylla* and *O. hirta* reduced NO production to a greater extent compared to the LPS group, exhibited stronger anti-inflammatory activity, and the effects of *O. myriophylla* and *O. hirta* on inflammatory factor TNF-α and IL-6 expression were examined (Figure 10). The inhibition rates of NO secretion at low, medium, and high doses are shown in Table 3.

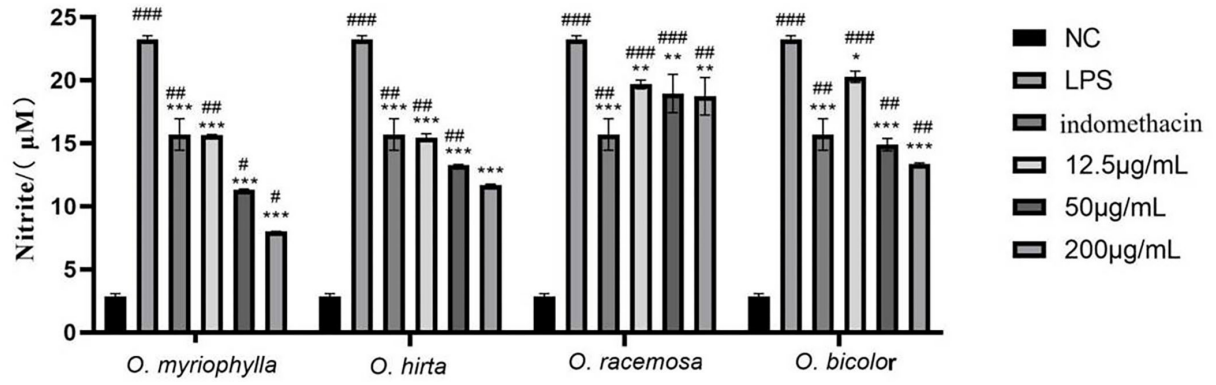


FIGURE 9
Effects of the total extracts of the four *Oxytropis* on NO production in LPS-induced mouse RAW 264.7 macrophages (NC, blank control group; LPS, 1 μg/mL LPS group; indomethacin: 100 μM indomethacin + 1 μg/mL LPS group. Compared with the NC group, $^{\#}P < 0.05$,$^{\#\#}P < 0.01$,$^{\#\#\#}P < 0.001$;compared with the model group (LPS group),$^{*}P < 0.05$, $^{**}P < 0.01$,$^{***}P < 0.001$).

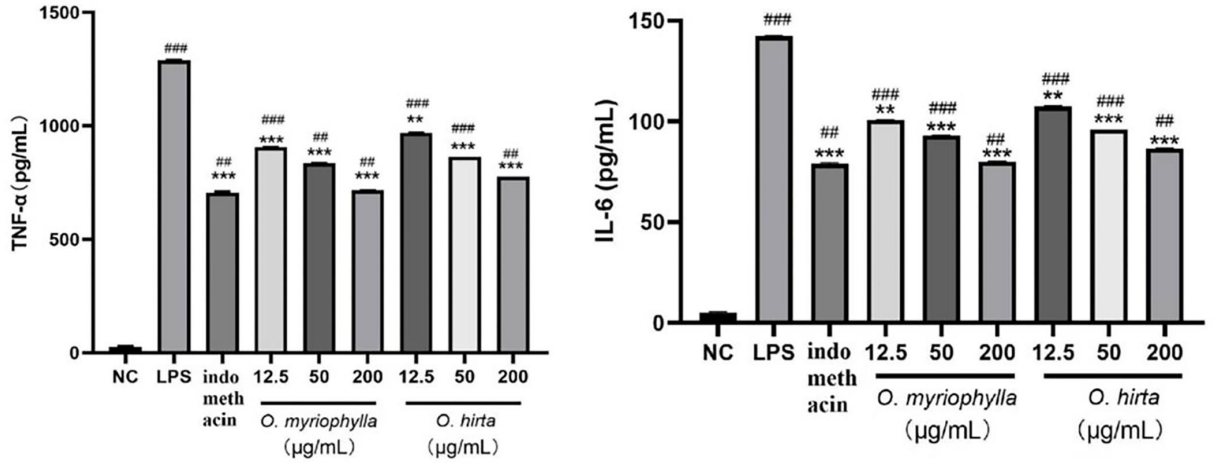


FIGURE 10
Effects of *O. myriophylla* and *O. hirta* on LPS-induced TNF-α and IL-6 expression in mouse RAW 264.7 macrophages (NC, blank control group; LPS, 1 μg/mL LPS group; indomethacin: 100 μM indomethacin + 1 μg/mL LPS group. Compared with the NC group, $^{\#\#}P < 0.01$, $^{\#\#\#}P < 0.001$; compared with the model group (LPS group), $^{**}P < 0.01$, $^{***}P < 0.001$).

Discussion

Characteristic marker components of *Oxytropis*

At present, with the shortage of wild resources, more *Oxytropis* species are currently being used in clinical applications. The analysis of overall chemical composition is thus of great significance to compare the differences between *Oxytropis* varieties. However, to date, only one or a few indices have been used to assess the compositions of the plant extracts, either qualitatively or quantitatively, which cannot fully reflect the quality. Our preliminary findings on the chemical constituents of *Oxytropis* using traditional chemical methods only identified a few compounds. Further studies were thus needed. In this study, we conducted a metabolomics investigation using UHPLC–Q-TOF–MS to characterize the compositions of four different *Oxytropis* species and assessing the differences between *O. myriophylla* from two different geographical regions. Metabolomics is a highly sensitive high-throughput technique that is widely used in medical research (Yao et al., 2019; Xia et al., 2021). However, there have been no reports on *Oxytrop*is quality control and evaluation with metabolomics. In the present study, we used positive and negative ion analysis to study the overall chemical composition of the *Oxytropis* species. This, combined with multivariate statistical analysis, was able to identify the different chemical components of the plant species.

TABLE 3 NO inhibition rate of total extracts from *Oxytropis*.

Total extracts	NO inhibition rate (%)		
	Low dose	Medium dose	High dose
O. myriophylla	33.51	51.30	65.54
O. hirta	32.74	42.92	49.64
O. bicolor	12.79	35.90	42.51
O. racemosa	15.36	18.45	19.43

The metabolites of medicinal plants have different types and structures. They often vary significantly according to time and space (Anne et al., 2020) and are influenced by different environmental conditions. The Saihanwula Nature Reserve of Chifeng city in China and Wuchuan city in Inner Mongolia of China are the main areas where *O. myriophylla* grows. We compared the chemical compositions and analyzed the relative contents of differential metabolites from the two different geographical origins with multivariate statistical analysis. Although no significant differences in the types of metabolites were found between DY and DYWC, there were significant differences in their concentrations. Most of the components showed higher concentrations in DY from the Saihanwula Nature Reserve of Chifeng city. The metabolic pathways and bioactive compounds present in the plants can be significantly affected by different environmental conditions (Ballhorn et al., 2011). For instance, different environmental conditions, such as variations in light, temperature, pH, and soil conditions, can all lead to substantial alterations in the metabolic pathways and the subsequent accumulation of secondary metabolites (Ramakrishna and Ravishankar, 2014; Jia et al., 2015). Saihanwula Nature Reserve and Wuchuan city are geographically distant. The significant differences in metabolite concentrations between DY plants from Saihanwula and Wuchuan may thus be attributed to the different environmental conditions in these two regions. Furthermore, these variations may also lead to marked differences in both the quality and pharmacological activities of the plant extracts.

Preliminary pharmacophylogenetic investigation in *Oxytropis*

Oxytropis is usually divided into six subgenera (Figure 11). They are Subgen. *Ptiloxytropis* Bunge, Subgen. *Triticaria* Vass, Subgen. *Physoxytropis* Bunge, Subgen. *Tragacanth oxytropis* Vass., Subgen. *Oxytropis*, and Subgen. *Orobia* (Bunge) C. W. Chang. The four *Oxytropis* species investigated in this study belong to the Subgen. *Orobia* (Bunge) C. W. Chang, which contains 13 Sect. *Oxytropis myriophylla*, *Oxytropis hirta,* and *Oxytropis bicolor* belong to Sect. *Baicalia* Stell. ex Bunge, while *Oxytropis racemosa* belongs to Sect. *Gobicola* Bunge. From the perspective of plant taxonomy, *O. myriophylla*, *O. hirta,* and *O. bicolor* are more closely related.

Flavones are the major components of this genus. From the isolated flavonoids in these four *Oxytropis* mainly exist in the form of flavonoid glycosides. The aglycons mainly include quercetin, kaempferol, and isorhamnetin, most of which are flavones. The sugar substituents are mainly glucose and rhamnose, in addition to galactose, xylose, arabinose, and mannose. The results are in agreement with the findings of Li et al. (2012) and Song et al. (2013). Although the structure of flavonoid glycosides is relatively simple in the four *Oxytropis*, there is one distinguishing feature; in all oxyglycosides, the glycosidic bonds were mostly in positions three and seven.

Besides flavonoids, triterpenoids exist in the form of triterpenoid saponins in the four *Oxytropis*, most of which are oleanane type with 3-*O* substitution, and the sugar substituents are glucose and rhamnose. Alkaloids are mainly indole alkaloids in the four *Oxytropis*.

Investigating and summarizing the ethnomedical applications, pharmacological activities, and chemical components of the four *Oxytropis* species showed that there was a good correlation between the pharmacological activities and the ethnomedical applications of the four plant species, with many pharmacological activities having a corresponding material basis (Figure 12). At the same time, it also suggested to us that the differences in efficacies between the *Oxytropis* species may be related to their specific components. Recent research has confirmed the pharmacological efficacies of both *O. myriophylla* and *O. hirta*. For instance, *O. myriophylla* has anti-inflammatory, analgesic, antibacterial, and antioxidant effects while *O. hirta* has only antibacterial and anti-inflammatory effects. In addition, the pharmacological experiments of this study also preliminarily confirmed that the total extracts of both plants had good anti-inflammatory effects. This is closely related to the flavonoids and saponins they contain. Glycyrrhizic acid, licorice saponins J2, and licoricesaponin G2 are marker compounds specific to *O. hirta* which have been demonstrated to have anti-inflammatory effects. The effect of *O. racemosa* on digestion and strengthening the spleen has also demonstrated a relationship with pharmacological research on improving digestive function, but its material basis is not clear. In addition, the antibacterial, anti-inflammatory, antioxidant, and anti-viral activities of *O. bicolor* are mostly related to flavonoids. According to the records, *O. myriophylla* and *O. hirta* have similar traditional applications, specifically, killing "viscosity", clearing heat, drying "xieriwusu", healing, muscle regeneration, and hemostasis. *O. racemosa* is used for digestion and spleen strengthening, and *O. bicolor* for detoxification and analgesia. From the perspective of pharmacological activity and traditional applications, *O. myriophylla* and *O. hirta* can be used instead. To fully tap and make use of the medicinal value of this genus, further investigation on the chemical composition and activities

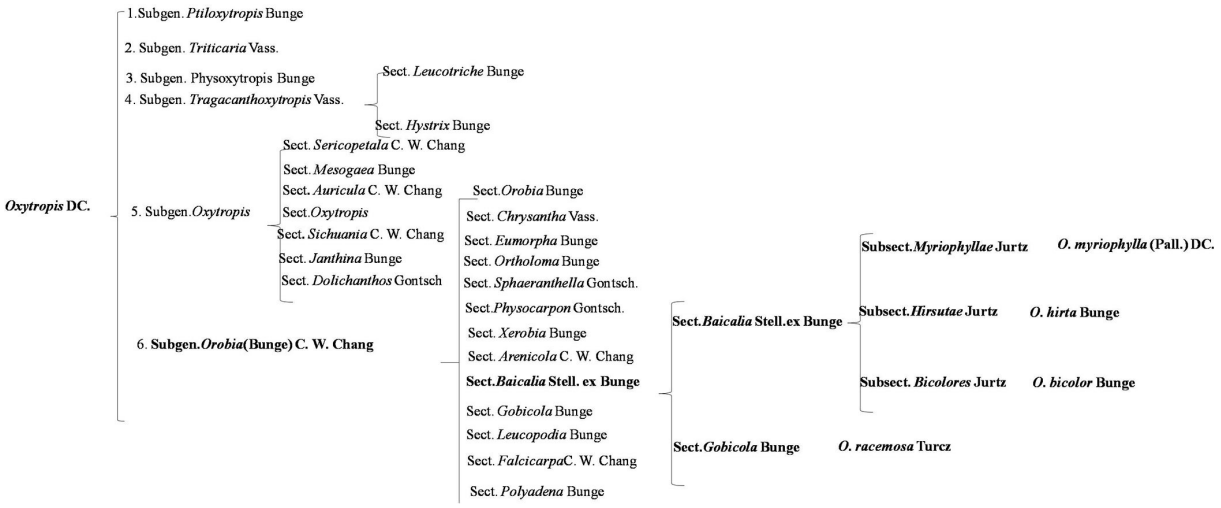


FIGURE 11
Plant chemotaxonomy of the four *Oxytropis*.

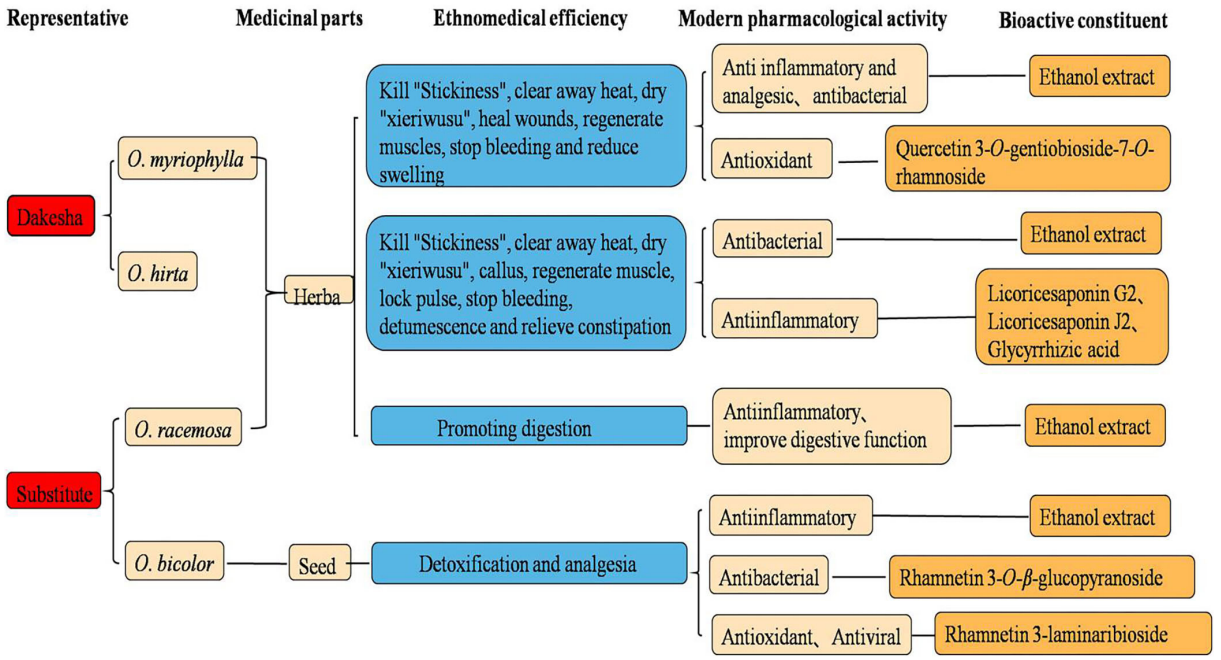


FIGURE 12
Proposed relationships between ethnopharmacology, pharmacology, and bioactive constituents from *Oxytropis*.

of the components is required along with an understanding of its material basis.

Conclusion

Oxytropis species are rich in resources and are widely distributed throughout the world. Although plants belonging to this genus have significant medicinal value, there has been, to date, little in-depth research on them, specifically, in the areas of classification, chemical composition, and pharmacological activities. The present study undertook to investigate its medicinal potential, summarizing the taxonomy of the genus based on relevant research and the literature. From the perspective of plant taxonomy, *O. myriophylla*, *O. hirta*, and *O. bicolor* are more closely related. UPLC-Q-TOF/MS combined

with PCA and OPLS-DA multivariate statistical analysis was used to characterize the overall chemical compositions of four *Oxytropis* species as well as of plants from two different geographical regions. Twenty-three differential metabolites were identified, with specific metabolic markers found in the four species. An analysis of pharmacological activities showed that, compared with two other *Oxytropis* species, extracts of *O. Myriophylla* and *O. hirta* had stronger anti-inflammatory activity. The study analyzed the relationships between the specific chemical components, traditional applications, and pharmacological activities to present a preliminary study of the pharmacophylogenetics of the *Oxytropis* genus as a whole. The study showed that *O. myriophylla* and *O. hirta* are more closely related to each other than to *O. bicolor* and *O. racemose,* suggesting that it may not be effective to use the latter two species as substitutes in Mongolian medicine. This preliminary investigation of the pharmacophylogeny of the genus *Oxytropis* will contribute to the better exploitation of the medicinal potential of this genus.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

XQW and XJ designed the experiment. XJ, SW, JNM, JY, XY, and YZ all performed the experiment work and prepared figures and tables. XJ and YL edited the final version of the manuscript. All the authors have studied and approved the final manuscript.

Funding

This work was supported by grants from the National Natural Science Foundation of China (Grant Nos. 81760686), Natural Science Foundation of Inner Mongolia (Grant Nos. 2018BS08011), and Team Project Foundation of Inner Mongolia Medical University (QNLC-2020048).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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.2022.958460/full#supplementary-material

References

Anne, R., Lee, B. K., Rui, M. C., Portela, R. J., and Stosch, M. (2020). Analysis of transformed upstream bioprocess data provides insights into biological system variation. *Plant Biotechnol. J.* 15, 2000113. doi: 10.1002/biot.202000113

Baimukhambetov, M. A. (1973). Flavonoids of the genus *Oxytropis*. *Chem. Nat. Compd.* 9, 564396. doi: 10.1007/BF00564396

Ballhorn, D. J., Stefanie, K., Manfred, J., Imke, S., Martin, H., and Adrian, D. H. (2011). Genetic and environmental interactions determine plant defences against herbivores. *J. Ecol.* 99, 313–326. doi: 10.1111/j.1365-2745.2010.01747.x

Batsuren, D., Tsetsegmaa, S., Batbayar, N., Dungerdorzh, D., Akhmedzhanova, V. I., Mil'grom, Y. M., et al. (1992). Alkaloids of *Oxytropis*. I. *Chem. Nat. Compd.* 28, 340–344. doi: 10.1007/BF00630255

Chen, X. W., Wang, M. L., and Wu, W. J. (2005). Preliminary study on bacteriostatic effect of Oxytropis bicolor in Ningxia. *Ningxia Agric. For. Sci. Technol.* 8, 15–16. doi: 10.3969/j.issn.1002-204X.2005.06.008

Cui, Y., Yang, H. H., Jing, J. X., Liu, T., Wang, R. J., Di, F. Y., et al. (2020). Rapid characterization of chemical constituents of Gansuibanxia decoction by UHPLC-FT-ICR-MS analysis. *J. Pharm. Biomed. Anal.* 179, 113029. doi: 10.1016/j.jpba.2019.113029

Elisens, W. J., and Denford, K. E. (1982). Flavonoids studies in four species of the *Oxytropis campestis* (Fabaceae-Galegeae). *Can. J. Bot.* 60, 1431. doi: 10.1139/b82-183

Jia, X., Sun, C. S., Li, G. Y., Li, G. B., and Chen, G. L. (2015). Effects of progressive drought stress on the physiology, antioxidative enzymes and secondary metabolites of Radix Astragali. *Acta Physiol. Plant.* 37, 4. doi: 10.1007/s11738-015-2015-4

Li, C. R., Li, M. N., Yang, H., Li, P., and Gao, W. (2018). Rapid characterization of chemical markers for discrimination of Moutan Cortex and its processed products by direct injection-based mass spectrometry profiling and metabolomic method. *Phytomedicine* 45, 76–83. doi: 10.1016/j.phymed.2018.04.003

Li, M. X., Lan, Z. H., Wei, L. L., Zhang, W. J., Ru, X., Zhang, R. X., et al. (2012). Phytochemical and biological studies of plants from the genus *Oxytropis*. *Rec. Nat. Prod.* 6, 1–20. doi: 10.1002/chin.201021261

Liu, F. J., Jiang, Y., Li, P., Liu, Y. D., Yao, Z. P., Xin, G. Z., et al. (2020a). Untargeted metabolomics coupled with chemometric analysis reveals species-specific steroidal alkaloids for the authentication of medicinal Fritillariae Bulbus and relevant products. *J. Chromatogr. A.* 1612, 460630. doi: 10.1016/j.chroma.2019.460630

Liu, W. X., Cheng, X. L., Guo, X. H., Hu, X. R., Wei, F., and Ma, S. C. (2020b). Identification of Calculus Bovis and its mixed varieties by ultra-high-performance liquid chromatography/quadrupole time-of-flight mass spectrometry (UHPLC-Q/TOF-MS) combined with the principal component analysis (PCA) method. *J. Pharm. Biomed. Anal.* 179, 112979. doi: 10.1016/j.jpba.2019.112979

Masafumi, O., Rie, Y., Hurlad, D., Ryota, T., Toshihiro, N., Junei, K., et al. (2022). Five triterpene glycosides from *O. myriophylla*. *Chem. Pharm. Bull.* 50, 1097–1099. doi: 10.1248/cpb.50.1097

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

Meng, G. X., Ao, Wuliji., Huo, W. X., Wang, X. L., Wang, H. (2016). Study on screening for antibacterial active site of *Oxytropis myriophylla* in vitro. *Chin. J. Inf. Tradit. Chin. Med.* 23, 51-54. doi: 10.3969/j.issn.1005-5304.2016.12.013

Okawa, M., Yamaguchi, R., Delger, H., Tsuchihashi, R., Nohara, T., Kinjo, J., et al. (2002). Five triterpene glycosides from *O. myriophylla*. *ChemInform*. 34, 3148. doi: 10.1002/chin.200303148

Pan, H. Q., Yao, C. L., Yao, S., Yang, W. Z., Wu, W. Y., and Guo, D. (2020). A metabolomics strategy for authentication of plant medicines with multiple botanical origins, a case study of Uncariae Rammulus Cum Uncis. *J. Sep. Sci*. 43, 1043–1050. doi: 10.1002/jssc.201901064

Ramakrishna, A., and Ravishankar, G. A. (2014). Influence of abiotic stress signals on secondary metabolites in plants. *Plant Signal Behav*. 6, 1720–1731. doi: 10.4161/psb.6.11.17613

Saiyin B. (2014). Research progress of Oxytropis myriophylla. *Chin. J. Ethn. Med.* 10, 56-58. doi: 10.16041/j.cnki.cn15-1175.2014.10.034

Sharma, S., Villamor, J. G., and Verslues, P. E. (2011). Essential role of tissue-specific proline synthesis and catabolism in growth and redox balance at low water potential. *Plant Physiol.* 157, 292–304. doi: 10.1104/pp.111.183210

She, G. M., Sun, F. F., Lv, H. N., and Liu, B. (2010). Radical scavenging activity of Oxtropis myriophylla. *Chin. J. Exp. Tradit. Med. Formulae.* 16, 91–94. doi: 10.13422/j.cnki.syfjx.2010.18.010

Song, S., Zheng, X. P., Liu, W. D., Du, R. F., Bi, L. F., and Zhang, P. C. (2010). 3-hydroxymethylglutaryl flavonol glycosides from a Mongolian and Tibetan medicine, *Oxytropis racemosa*. *Chem. Pharm. Bull.* 58, 1587–1590. doi: 10.1248/cpb.58.1587

Song, S., Zheng, X. P., Liu, W. D., Du, R. F., Feng, Z. M., Zhang, P. C., et al. (2013). Rapid identification of unstable acyl glucoside flavonoids of *Oxytropis* racemosa Turcz by high-performance liquid chromatography-diode array detection-electrospray ionisation/multi-stage mass spectrometry. *Phytochem. Anal.* 24, 155–161. doi: 10.1002/pca.2395

Sun, J. Z., and Xu, J. L. (1992). Classification and distribution of the genus *Oxytropis* DC. from Gansu. *J. Gansu Sci.* 3, 34–40. doi: 10.16468/j.cnki.issn1004-0366.1992.03.006

Sun, R. Q., and Chen, J. C. (1997). Saponins from *Oxytropis Bzcolor*. *Phytochemistry* 44, 505–507. doi: 10.1016/S0031-9422(96)00550-X

Sun, R. Q., Jia, Z. J., and Liang, C. D. (1991). Three saponins from *Oxytropis* species. *Phytochemistry* 30, 2707–2709. doi: 10.1016/0031-9422(91)85128-M

Szabados, L., and Savoure, A. (2010). Proline: a multifunctional amino acid. *Trends Plant Sci.* 15, 89–97. doi: 10.1016/j.tplants.2009.11.009

Tang, Z. H., Li, T., Tong, Y. G., Chen, X. J., Chen, X. P., Wang, Y. T., et al. (2015). A systematic review of the anticancer properties of compounds isolated from Licorice (Gancao). *Planta Med.* 81, 1670–1687. doi: 10.1055/s-0035-1558227

Wu, J., Yang, W. Z., Pan, H. Q., Yao, S., Wu, W. Y., and Guo, D. (2018). Geographic impact evaluation of the quality of Alismatis Rhizoma by untargeted metabolomics and quantitative assay. *J. Sep. Sci.* 41, 839–846. doi: 10.1002/jssc.201700902

Xia, Z. D., Liu, X., Tong, L. G., Wang, H., Feng, M. L., Xi, X. H., et al. (2021). Comparison of chemical constituents of Bupleurum marginatum var.stenophyllum and Bupleurum chinense DC. using UHPLC–Q-TOF–MS based on a metabonomics approach. *Biomed. Chromatogr.* 35, 5133. doi: 10.1002/bmc.5133

Yang, R., Yuan, B. C., Ma, Y. S., Zhou, S., and Liu, Y. (2017). The anti-inflammatory activity of licorice, a widely used Chinese herb. *Pharm. Biol.* 55, 5–18. doi: 10.1080/13880209.2016.1225775

Yao, H., Gao, T. H., Li, J., Chen, Z. J., Wang, L. J., Zhang, J. M., et al. (2019). Metabonomics study on the effect of Siwu decoction for blood deficiency syndrome in rats using UPLC–Q/TOF–MS analysis. *Biomed. Chromatogr.* 33, e4617. doi: 10.1002/bmc.4617

Ye, S.W., Hu, J. J., Hu, Z. L., Zhao, L., Hao, N., and Tian, X. R. (2022). Study on the antimicrobial constituents from the roots of *Oxytropis hirta* Bunge. *Chin. J. Pestic. Sci.* 24, 289-297. doi: 10.16801/j.issn.1008-7303.2021.0158

Yu, X. A., Ge, A. H., Zhang, L., Li, J., An, M., Cao, J., et al. (2017). Influence of different processing times on the quality of Polygoni Multiflora Radix by metabolomics based on ultra high performance liquid chromatography with quadrupole time-of-flight mass spectrometry. *J. Sep. Sci.* 40, 1928–1941. doi: 10.1002/jssc.201600913

Zhang, A. H., Zou, D., Yan, G. L., Tan, Y. L., Sun, H., and Wang, X. J. (2014). Identification and characterization of the chemical constituents of Simiao Wan by ultra high performance liquid chromatography with mass spectrometry coupled to an automated multiple data processing method. *J. Sep. Sci.* 37, 1742–1747. doi: 10.1002/jssc.201400105

Zhang, G., Zhou, L., Zhang, S.P., Guo, Y. Z., Gao, D., Shi, F. Y., et al. (2019). Chemical composition analysis and nutritional value evaluation of Oxytropis racemosa. *Prog. Vet. Med.* 40, 62–68. doi: 10.16437/j.cnki.1007-5038.2019.09.012

Zhao, G., Liu, Y. F., Gou, Y., Zhou, Q. M., He, C. J., Guo, L., et al. (2018). Metabolomics study of cultivated bulbus fritillariae cirrhosae at different growth stages using UHPLC-QTOF-MS coupled with multivariate data analysis. *Phytochem. Anal.* 29, 290–299. doi: 10.1002/pca.2742

Zheng, Y. F., Wei, J. H., Fang, S. Q., Tang, Y. P., Cheng, H. B., Wang, T. L., et al. (2015). Hepatoprotective triterpene saponins from the roots of Glycyrrhiza inflata. *Molecules* 20, 6273–6283. doi: 10.3390/molecules20046273

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 28 November 2022
DOI 10.3389/fpls.2022.970535

OPEN ACCESS

EDITED BY
Da-Cheng Hao,
Dalian Jiaotong University, China

REVIEWED BY
Yan Lin,
Guizhou Medical University, China
Min Xu,
Kunming University of Science and Technology, China

*CORRESPONDENCE
Haitao Liu
htliu0718@126.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Plant Metabolism and Chemodiversity, a section of the journal Frontiers in Plant Science

RECEIVED 16 June 2022
ACCEPTED 31 August 2022
PUBLISHED 28 November 2022

CITATION
Liu J, Mu X, Liang J, Zhang J, Qiang T, Li H, Li B, Liu H and Zhang B (2022) Metabolic profiling on the analysis of different parts of *Schisandra chinensis* based on UPLC-QTOF-MS with comparative bioactivity assays. *Front. Plant Sci.* 13:970535. doi: 10.3389/fpls.2022.970535

Metabolic profiling on the analysis of different parts of *Schisandra chinensis* based on UPLC-QTOF-MS with comparative bioactivity assays

Jiushi Liu[1†], Xinlu Mu[1†], Jinmei Liang[2], Jianuo Zhang[1], Tingyan Qiang[1], Hongbo Li[3], Bin Li[1], Haitao Liu[1*] and Bengang Zhang[1]

[1]Key Laboratory of Bioactive Substances and Resources Utilization of Chinese Herbal Medicine, Ministry of Education, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [2]Department of Pharmacy, Medical Guarantee Center Pla General Hospital, Beijing, China, [3]College of Horticulture, Shenyang Agricultural University, Shenyang, China

The *Schisandra chinensis* is an important edible plant, and previous phytochemical research focused on the *S. chinensis* fruit (SF) due to its long history as traditional Chinese medicine. *Schisandra chinensis* fruit was used as an astringent tonic to astringe the lungs and the kidneys, replenish energy, promote the production of body fluids, tonify the kidney, and induce sedation. The components of *S. chinensis*, such as its stems (SS), leaves (SL), and roots (SR), have drawn little attention regarding their metabolites and bioactivities. In this study, a strategy of combining a chemical database with the Progenesis QI informatics platform was applied to characterize the metabolites. A total of 332 compounds were tentatively identified, including lignans, triterpenoids, flavonoids, tannins, and other compound classes. Heatmap and principal component analysis (PCA) showed remarkable differences in different parts of the plants. By multiple orthogonal partial least-squares discriminant analyses (OPLS-DA), 76 compounds were identified as potential marker compounds that differentiate these different plant parts. Based on the variable influence on the projection score from OPLS-DA, the active substances including gomisin D, schisandrol B, schisantherin C, kadsuranin, and kadlongilactone F supported the fact that the biological activity of the roots was higher than that of the fruit. These substances can be used as marker compounds in the plant roots, which likely contribute to their antioxidant and anti-inflammatory activities. The plant roots could be a new medicinal source that exhibits better activity than that of traditional medicinal parts, which makes them worth exploring.

KEYWORDS

***Schisandra chinensis*, different parts, metabolomics, bioactivity assays, chemometric**

Introduction

Schisandra chinensis is an important plant resource that is distributed in North China, D.P.R. Korea, R.O. Korea, Japan, and most Eastern parts of Russia; in addition, the plant has a very large market and offers significant economic and medicinal values (Panossian and Wikman, 2008; Ye et al., 2019). *Schisandra chinensis* is a special cash crop in rural areas of Northeast China, and the annual demand for *S. chinensis* as a traditional tonic herb and food is more than 30,000 tons. A key method for farmers to increase income and economic development was established by planting *S. chinensis* crops, and the planting area has expanded yearly. It is necessary to strengthen the research and utilization of *S. chinensis*, promote its industrial transformation, and upgrade the intensive method of processing. The *S. chinensis* fruit (SF), which is called Wuweizi in China, has been used for thousands of years in traditional Chinese medicine as a superior drug and has been included in the list of available healthy foods by the Ministry of Health of China since 2002 (http://www.nhc.gov.cn/sps/s3593/200810/bc239ea3d226449b86379f645dfd881d.shtml) (Medica Editorial Board of National Institute of Chinese Medicine, 1999). *Schisandra chinensis* fruit was used as an astringent tonic to astringe the lungs and the kidneys, replenish energy, promote the production of body fluids, tonify the kidney, and induce sedation (Chinese Pharmacopoeia Commission, 2020). Our ethnobotanical survey found that, aside from fruits, other parts of *S. chinensis* have been used in foods and herbal drugs for a long time, which contain nutrients and biologically active phytochemicals. Due to the lemon flavor that is usually absent in the most commonly used seasonings, local individuals in North China often dry *S. chinensis* stems (SS) and use them as a seasoning for stewed meat and as a substitute of pepper. The use of leaves and roots as traditional herbal tea is well known, and they have been used to delay the senescence process in Chinese folk medicine (Chen et al., 2011).

Gomisin A, schisandrol B, schisandrin, and schisandrin C are natural dibenzocyclooctadiene lignans that are isolated from SF and are considered to be the main active compounds responsible for the bioactivity of SF (Szopa et al., 2017). The compounds exhibit antihepatotoxic, antioxidant, antitumor activity, and anti-human immunodeficiency virus (HIV) activities as well as have effects on physical performance and the central nervous system (Xu et al., 2015; Chen et al., 2019; Yan et al., 2021). More recently, studies reported the therapeutic effects of *S. chinensis* on alleviating cough, liver injury, kidney injury, lung injury, platelet aggregation hepatitis, and cardiovascular disease (Gui et al., 2020; Xu et al., 2020; Lin et al., 2021). Gomisin M_1 was found to modulate miRNA biogenesis to inhibit the proliferation, migration, and invasion of hepatocellular carcinoma (HCC) cells (Xu et al., 2021). Schisandrol B may play a role in the treatment of carbon tetrachloride (CCl_4)-induced liver injury by downregulating the expression of iNOS and COX-2 and regulating the expression of the NF-κB and IL-17 signaling pathways to inhibit the expression of proinflammatory factors. Schisandrin B alleviated CCl_4-induced liver inflammation and fibrosis by inhibiting macrophage polarization by targeting peroxisome proliferator-activated receptor gamma (PPARγ) (Chen et al., 2021). However, other parts of *S. chinensis,* such as its stems (SS), roots (SR), and leaves (SL), have been scarcely reported on their phytochemistry and bioactivity (Liu et al., 2020). Indeed, different parts of *S. chinensis* have different medicinal values; however, research on the chemical composition and biological activity of *S. chinensis* is limited, thus further development and utilization of this species are still challenging.

Because of the large number, structural diversity, and content range (10% to sub-ppm level) of chemical compounds in plants, performing secondary metabolomics analysis has been a great challenge (Hur et al., 2013; Lu et al., 2016). The development of liquid chromatography-tandem mass spectrometry (LC-MS) provides strong support for the metabolic profiling analysis of plant extracts (Chaleckis et al., 2019). Plant metabolomics was widely used in the analysis of metabolites from different geographical sources, different growth periods, and different plant parts (Dai et al., 2015; Han et al., 2015; Jandric et al., 2017). Until now, the use of ultra-performance liquid chromatography coupled with quadrupole/time-of-flight MS (UPLC-Q-TOF-MS/MS) to analyze *S. chinensis* samples has been used for only 28 marker compounds of SF extracts. Moreover, no report is available to compare the differential metabolites in the four parts of *S. chinensis*.

In the present study, the local database of metabolites from the genus *Schisandra* was established by using Progenesis SDF Studio. The metabolites of SL, SF, SS, and SR were compared, and the marker compounds were identified by UPLC-QTOF-MS coupled with chemometric analysis. Previous biological activity studies mainly focused on SF, and the results showed that SF exhibited antioxidant and anti-inflammatory activities in the cell line RAW 264.7 (Hu et al., 2012; Wang et al., 2018). The antioxidative and antiphlogistic are the basis for its liver protection activity and anti-neurodegenerative diseases (Guo et al., 2008). However, there have been no reports of antioxidant and anti-inflammatory activities with other parts of *S. chinensis*. Therefore, through screening the antioxidant and anti-inflammatory activities of different parts of *S. chinensis in vitro*, we may find alternative sources of the antioxidants and anti-inflammatory compounds; thus, we can study the correlation between chemical composition and biological activities in the plant parts of *S. chinensis*.

Materials and methods

Materials

The four different parts, SF, SR, SS, and SL, were collected from plants in Shenyang, Liaoning province, China, in September 2020. The certificate specimen was identified by Professor Bengang Zhang from the Institute of Medicinal Plant Development (IMPLAD) and deposited at the Medical Plant Resource Center in the IMPLAD.

Chemicals and reagents

Reference substances were used to compare the MS data, and retention time (RT) of the identified compounds consist of schisandrol A, pregomisin, schisantherin B, schisantherin A, schisantherin D, schisandrin B, benzoylgomisin O, interiotherin A, angeloylgomisin O, gomisin D, angeloylisogomisin O, schisanhenol, schisandrol B, gomisin J, gomisin O, gomisin G, gomisin K_2, gomisin K_3, schisandrin C, schisandrin A, and gomisin N. All of the above compounds were isolated from *S. chinensis* and stored in the laboratory. For UPLC-MS analysis, LC/MS-grade acetonitrile, methanol, and formic acid were purchased from Thermo Fisher (USA), and water was purchased from Guangzhou Watsons Food & Beverage Co., Ltd. (GuangZhou, China). The dimethyl sulfoxide (DMSO) and 2,2′-diphenyl-1-picrylhydrazyl (DPPH) were purchased from Solarbio, Beijing, China. Then, 96-well plates were purchased from Promega Corporation (Madison, WI, USA). The analytical grade of methanol was procured from Beijing Chemical Works Beijing, China). Pure water (18.2 MΩ) was obtained from a Milli-Q System (Millipore, Billerica, MA, USA). Dulbecco's modified Eagle medium (DMEM), fetal bovine serum (FBS), and Cell Counting Kit-8 (CCK-8) were purchased from Gibco Life Tech (Waltham, MA, USA) and Beyotime Bio-Technology Co., Ltd. (Shanghai, China). Lipopolysaccharide (LPS) was purchased from Sigma Aldrich Corporation (St. Louis, USA). Nitric Oxide Assay Kit was obtained from the Beyotime Institute of Biotechnology (Shanghai, China).

Sample preparation

The fruit, root, stem, and leaf samples were air dried, powdered, and sifted through a 50-mesh sieve. The extraction method in Chinese Pharmacopoeia 2020 was applied: each powdered sample (0.25 g) was added into a 20 ml volumetric flask, methanol was poured, sonication treatment was carried out for 30 min (250 W, 20 KHz), and the sample solution was filtered through a 0.22 μm membrane filter prior to injection into the UPLC-QTOF-MS system. Each part has three samples, and every sample was repeated two times. The stability and repeatability of the methodology, which employed gradient elution, were determined by the repeat analysis of a pooled quality control (QC) sample which was mixed with the 24 samples of *S. chinensis*. Finally, the filtrate was diluted 10-fold with methanol for the UPLC-QTOF-MS analysis.

UPLC-QTOF/MS conditions

Ultra-performance liquid chromatography was performed on a Waters Acquity UPLC system (Waters, Milford, MA, USA), equipped with a diode array detection (DAD) system, which was recorded in the range of 200–400 nm. A Waters ACQUITY BEH C18 Column (2.1 × 100 mm, 1.7 μm, Waters, Milford, MA, USA) was applied for UPLC separation. Phase A was water and phase B was acetonitrile. The UPLC gradient elution was applied as follows: 49–49% B (0–2 min), 49–52% B (2–4 min), 52–60% B (4–6 min), 60–62% B (6–9 min), 62–62% B (9–10 min), 62–66% B (10–13 min), 66–70% B (13–15 min), and 70–95% B (15–17 min). The flow rate was 0.3 mL/min, and the injected volume of the sample solution was 1 μl. The column and autosampler were maintained at 30 and 20°C, respectively.

A Waters Xevo G2 QTOF (Waters, Milford, MA, USA) was used for MS detection. The instrument equipped with an electrospray ionization (ESI) source controlled by MassLynx 4.1 software. The conditions of the MS detector were as follows: capillary of 3 kV, sampling cone of 30 V, source temperature of 100°C, desolvation temperature of 300°C, cone gas flow of 50 L/h, desolvation gas flow of 600 L/h, and collision energy of 6 kV. Then, nitrogen was used as a nebulizer and auxiliary gas. MS^E data were obtained in centroid mode with a mass range of 100–1,200 Da, and the scanning time is 1 s. The instrument was performed in both low-energy (function 1) and high-energy (function 2) scan functions, and the collision energy was 6 and 20–40 V, respectively. The lock mass compound was 200 pg/μL of leucine enkephalin (*m/z* 556.2771 in positive mode).

Main lignans content detection

We detected the following 9 common lignans in *S. chinensis* and investigated their distribution in different parts: schisandrol A (1), gomisin D (2), gomisin J (3), schisandrol B (4), angeloylgomisin H (5), gomisin G (6), schisantherin A (7), schisandrin B (8), and schisandrin C (9). The extract method and methodology for content detection were determined according to Mu et al. (2022). A sample (1 g) was mixed with 30 mL of methanol in a capped conical flask and was extracted three times by using the ultrasonic method

(60 kHZ, 250 W), for 20 min each time. The weight loss in the ultrasonic procedure was compensated and filtered through a 0.22-μm membrane filter prior to the UPLC system. The chromatographic method is the same as the above "2.4 UPLC-QTOF/MS conditions."

Construction of the *Schisandra* compound database

A local database of the genus *Schisandra* was created using Progenesis SDF Studio. The compounds from *Schisandra* species were sorted and summarized by searching SciFinder, Pubmed, CNKI, and Google Scholar. The CAS number was obtained by searching SciFinder for its name or structure (found in the original article) and verifying the single structure file (·mol) of the selected compound and transfer. All ·mol files were consolidated into an internal database by Progenesis SDF Studio. The established database was used to identify the compounds.

Data processing and analysis

The data were profiled using Progenesis QI 2.3 software, which was obtained from positive ion modes of different parts (Waters, USA). The operation steps by Progenesis QI are as follows: data import, peak alignment, experimental grouping, peak extraction, normalization, deconvolution, identification, and statistical analysis. The 30 runs imported were calibrated according to automatically selected QC samples. Peak extraction setting parameters were as follows: limits automatic, sensitivity default set to 3, chromatographic minimum peak width, and 1.0–17.0 min of retention time (RT). Structures of the chemical constituents were identified by the *Schisandra* compound database, which was constructed by a ·mol file in Progenesis SDF Studio. In addition, an online compound library (Nature Chemical Biology) was used to identify more chemical compounds. The adduct ion in Progression QI 2.3 was setted as follows: $[M+H]^+$, $[M+Na]^+$, $[M+K]^+$, $[M+NH_4]^+$, $[2M+H]^+$, and $[2M+Na]^+$. The data were imported in a .raw format, and chromatographic peak alignment was used to set RT limits, including ignoring ions before 1 min and after 17 min; then, the automatic processing was completed, the experimental design was setup, the compound identification was carried out, and compound statistics were applied. SIMCA 14.1 was employed to process exported data, using principal component analysis (PCA) and orthogonal partial least-squares discriminant analysis (OPLS-DA), combined with S-plot to analyze potential marker compounds. In S-plots, the significant differential retention time-exact mass (RT-EM) pairs were chosen and marked on the Progenesis QI for compound identification. The RT-EM pairs were finally determined by variance analysis, with a p-value of ≤ 0.05 and a max fold change of ≥ 2.

Antioxidant activities

To evaluate the DPPH radical scavenging activity, the Solarbio Kit was used. Following the method (Usman et al., 2022), 10.0000 mg from different parts were accurately weighed, 10 ml of extract solution was added and extracted in a 40°C water bath for 30 min, and then, the gradient was diluted to 1.000, 0.5000, 0.2500, 0.1250, 0.0625, and 0.0313 mg/ml. After centrifugation at room temperature of 10,000 rpm for 10 min, the supernatant was put on ice for testing. The following reagents were added into a 1.5 ml EP tube: blank tube (25 μl extract + 975 μl working solution), determination tube (25 μl supernatant + 975 μl working solution), and control tube (25 μl supernatant + 975 μl absolute ethanol); DPPH free radical scavenging rate is defined as D% = {[A blank − (A determination − A control)]/A blank} * 100%.

Anti-inflammatory activities

Cell culture and the determination of cell viability was the first step to assay the anti-inflammatory activities of different parts. RAW 264.7 macrophages were purchased from the cell bank of the Type Culture Preservation Committee of the Chinese Academy of Science (Shanghai, China) and were selected to verify cytotoxicity. Then, to measure the content of Nitric oxide (NO), the NO standard curve was drawn and determined by the Griess kit. The specific methods are shown in Supplementary material 1.

Finally, the appropriate dose and concentration were selected according to cell viability to determine NO. To establish the inflammatory cell model, a total of 200 μL of RWA264.7 cells (2 × 106 cells/mL) were added into 96-well plates and cultured at 37°C in 5% CO_2 for 24 h. Next, 100 μL of different concentrations extracts (25, 50, 100, and 200 μg/mL) from SR, SS, SL, and SF were added to the culture plate and incubated for 1 h. Afterward, 1 μg/mL of LPS was added to the culture plate and incubated for 24 h, and the supernatant was obtained. Finally, 50 μL of supernatant, 50 μL of Griess Reagent I, and 50 μL of Griess Reagent II were added to the culture plate, and the absorbance at 540 nm was determined by a microplate reader. The NO concentration of each group was calculated according to the obtained standard curve.

Results and discussion

The database of the genus *Schisandra*

A total of 237 compounds, including lignans, triterpenoids, flavonoids, tannins, and other compound classes, were searched from the phytochemistry and pharmacology literature of the genus *S. chinensis* (last updated December 2021). The database has been the most complete MS database for the genus *Schisandra* thus far. The *Schisandra* compound database included an additional 308 compounds from *Schisandra sphenanthera* and other species, in which many studies on phytochemistry as traditional Materia medica were performed (Szopa et al., 2019; Yang and Yuan, 2021; Yang et al., 2021). Thus, the 545 compounds were confirmed, and the ·mol files were stored.

The *Schisandra* compound database included 545 compounds from *Schisandra*, and each compound contained the following information: scientific name, CAS number, neutral mass, formula, and structure. The *Schisandra* compound database is conducive to the metabolic identification of *S. chinensis* and promotes the characterization of compounds from other *Schisandra* species. A new research model can be used as a reference to establish personalized databases for other species, and the size of these chemical databases may vary greatly because of the number of species.

Metabolite characterization and compound identification

Based on the local database, which was premade, and the online compound library, 332 compounds were detected by combining the Progenesis MetScope in Progenesis QI, including lignans, flavonoids, triterpenoids, and alkaloids. A total list of detected compounds, which consists of the formula and adducts, is available in the Supplementary material (Table S1). Lignans are the most important compounds in the genus *Schisandra*, especially dibenzocyclooctene lignans, such as schisantherrin A, schizandrin, deoxyschizandrin, γ-schizandrin, and gomisin J (Szopa et al., 2017). Most dibenzocyclooctene lignans show a variety of biological activities, including anti-inflammatory, antioxidant, antihepatotoxic, and antitumor activities. Several other compounds, including triterpenoids, flavonoids, tannins, and precursor organic acids, were also identified.

A total of 30 UPLC-MS samples were imported to Progenesis QI for data analysis. The ion at 416.1839 *m/z* (RT 3.51 min) was identified as a reliable example. A strong peak of $[M+Na]^+$ (*m/z* 559.1979) was observed during the full scan MS at 6.56 min, which is characterized by a loss of neutral molecules, such as C_6H_5COOH (−122 Da), OCH_2 (−30 Da), and H_2O (−18 Da), and the presence of some fragment ions, including $[M+H\text{-}H_2O]^+$, $[M+H\text{-}C_6H_5COOH]^+$, and $[M+H\text{-}C_6H_5COOH\text{-}CH_2O]^+$. Finally, the compound was identified as schisantherin A, and the fragmentation is shown in Figure 1C. These fragments corresponded with inferred Progenesis QI and the standard ion fragment (Figure 1).

Chemical comparison and biomarker probe between different parts

To study the metabolic diversity of different parts of *S. chinensis*, the extracts of fruits, leaves, stems, and roots were comparatively studied by UPLC-QTOF-MS. The basic peak ion (BPI) chromatogram of fruits, leaves, stems, and roots revealed significant differences in their overall composition (Figure 2). To fully understand the chemical differences between samples, all compounds were used for chemometrics without any filtration.

To understand the distribution characteristics of all compounds in different sections, heatmapping, the most common visualization method is widely used in metabonomics because of its vivid information expression (Trygg et al., 2007). It is worth noting that all peaks detected by QI in the heatmap have remarkable differences in different parts, and few metabolites had a higher relative content in all parts of the same plant. In addition, the clustering results showed that there was significant diversity among groups but almost no differences within groups (Figure 3A). The peaks varied greatly among different parts; for example, the content of deoxyschizandrin was in the order of SF > SR > SS > SL and of γ-schizandrin was SR > SF > SS > SL.

The PCA is a multivariate statistical analysis method that uses the linear transformation of multiple variables to select fewer significant variables. It is widely used in metabonomics to find the molecules that cause the differences between samples or groups and to further analyze the metabolic pathways, biomarkers, and biological significance (Lever et al., 2017). The QC samples were gathered by a PCA score plot (Figure 3B). There were significant discrepancies between the different parts, and except for within the SL group, the degree of polymerization was good among the groups, which was consistent with the heatmap results. The PCA score plot showed that the distance between roots and leaves was the farthest, which indicated that the divergence between them was the most obvious, which was consistent with the result of the BPI chromatogram. Obvious differences were observed in the metabolic spectrum of each part of *S. chinensis* by the PCA. The selected markers showed different variable trends between the different parts of *S. chinensis* (Figure 3C). Leaves, roots, stems, and fruits have different chemical constituents. For instance, the contents of gomisin D (530.2153n, RT 3.12 min), schisandrol B (416.1839n, RT 3.51 min), and schisantherin C (515.2279n, RT 7.66 min)

A

	M/Z	Adducts/fragments	Formula	Mass error(ppm)
MS	554.1893	M+NH_4	$C_{30}H_{36}O_9N$	0.56
	559.1944	M+Na	$C_{30}H_{32}O_9Na$	0.98
MS^E	415.1747	M+H−C_6H_5COOH	$C_{23}H_{27}O_7$	-0.96
	385.1640	M+H−C_6H_5COOH-CH_2O	$C_{22}H_{25}O_6$	-2.86
	371.1482	M+H−C_6H_5COOH-CH_2O-CH_2	$C_{21}H_{23}O_6$	-1.89

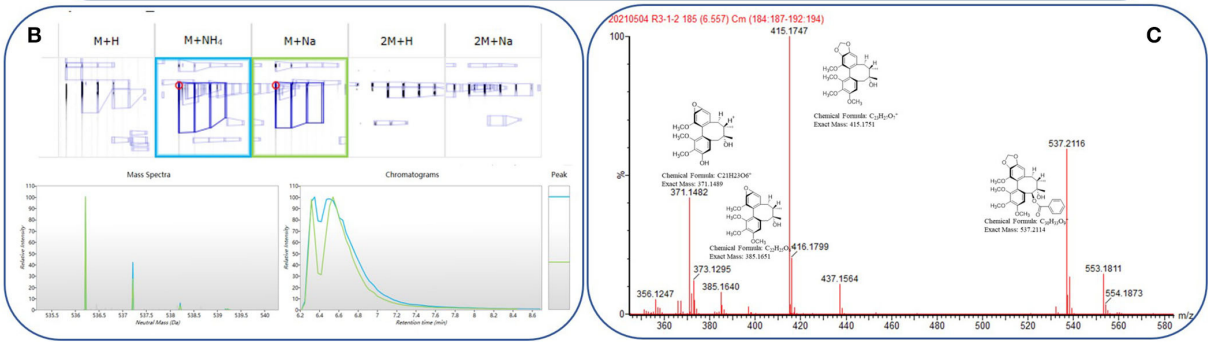


FIGURE 1
Progenesis QI showing the identification of schisantherin A from the filtered metabolic features. **(A)** Low-energy major precursor exact mass schisantherin A adduct ions and corresponding high-energy fragment ions. **(B)** Common adduct ions for schisantherin A. **(C)** Structure and exact mass number of fragment ions.

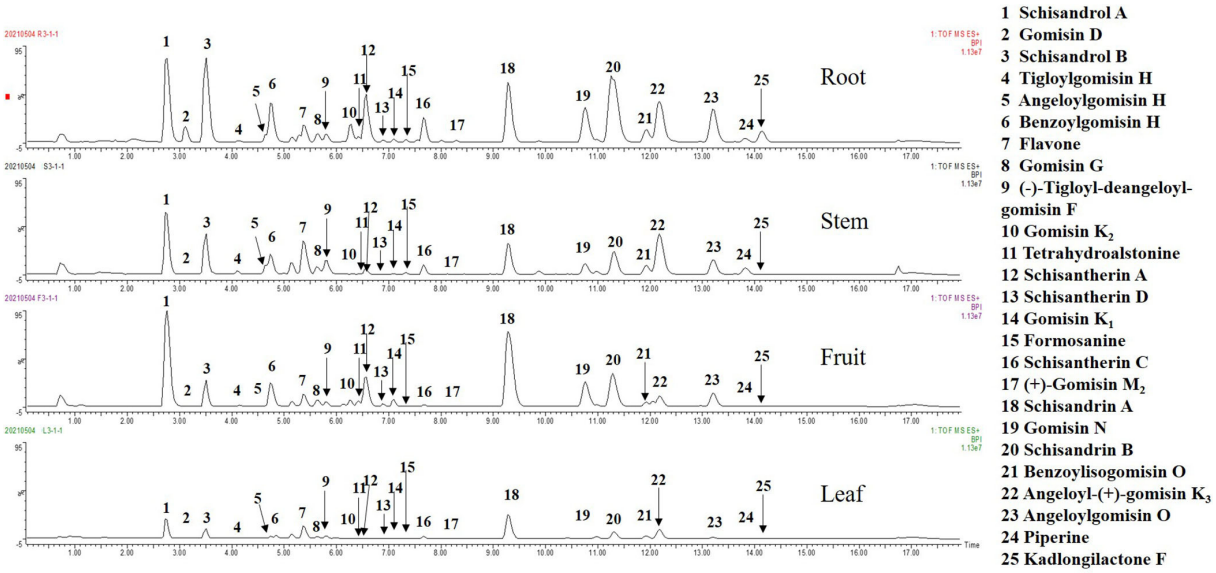


FIGURE 2
The basic peak ion (BPI) chromatograms in positive ion mode.

were higher in roots, while the contents were lower in leaves and stems. The difference in the metabolite spectrum may also help to explain the different efficacies of different parts of *S. chinensis*.

Orthogonal partial least-squares discriminant analysis was used to perform further discriminant analysis of metabolomics data in the two groups (Brereton, 2009). Four groups were compared by OPLS-DA with an S-plot to find marker compounds representing the difference between groups. These four groups were compared by OPLS-DA with an S-plot to further identify marker compounds causing the difference

10.3389/fpls.2022.970535

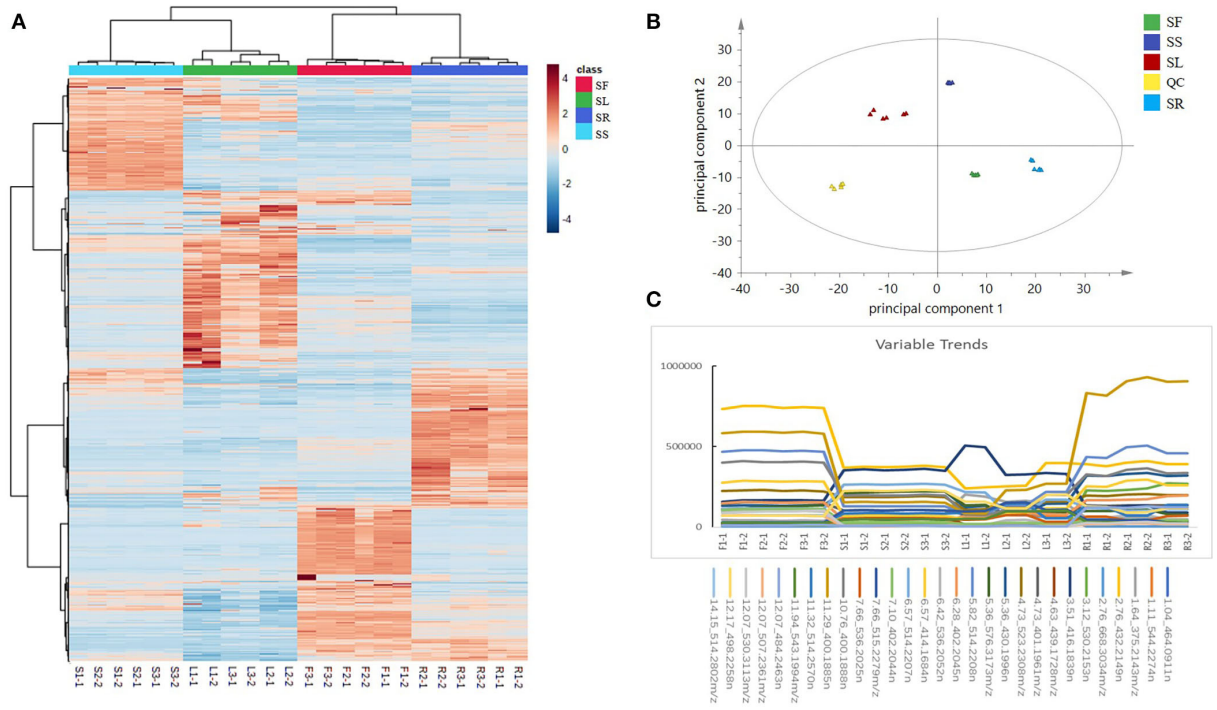


FIGURE 3
(A) A heatmap based on the relative abundance of 3,302 identified compounds in different parts of *Schisandra chinensis*. **(B)** A PCA score plot of five parts of *S. chinensis*: stem (SS), leaf (SL), fruit (SF), root (SR), and quality control (QC). **(C)** Examples of variation trends of different compounds in different parts of *S. chinensis*.

TABLE 1 Statistical parameter values of different orthogonal partial least-squares discriminant analysis (OPLS-DA) models based on ultra-performance liquid chromatography coupled with quadrupole/time-of-flight mass spectrometry (UPLC-Q-TOF-MS/MS) data in positive mode and the number of markers selected from S-plot.

	OPLS-DA				S-plot	
Model classes	**Scaling**	**Components**	**R2 (cum[a])[b] (%)**	**Q2 (cum)[c] (%)**	**Markers in −1**	**Markers in 1**
F vs. S	Pareto	2	0.991	1.000	9	8
F vs. R	Pareto	2	0.908	0.990	14	6
F vs. L	Pareto	2	0.929	0.994	14	5
S vs. R	Pareto	2	0.954	0.994	5	8
S vs. L	Pareto	2	0.992	0.988	8	8
R vs. L	Pareto	2	0.905	0.995	4	10

[a]cum: cumulative.
[b]R2 (cum): the variation displayed by all components in the model.
[c]Q2 (cum): the accuracy of the predicted class membership by the model.

among groups. The comparison of parameters and models is shown in Table 1. Each model was reliable according to high R2 and Q2 values. The S-plot has a higher research value and shows the observation variables clearly on the two-dimensional plane. It was conducive to screening the correlation between chemical composition and the model type. The variable correlation and contributions were shown separately on the X- and Y-axes. The variable results displayed distinct differences among the groups and were found at the top right (1) and the bottom left (−1), and the ions with no significant difference were in the middle of the S-plot. Six model classes were applied to screen the biomarkers (Figure 4). The potential biomarkers were screened

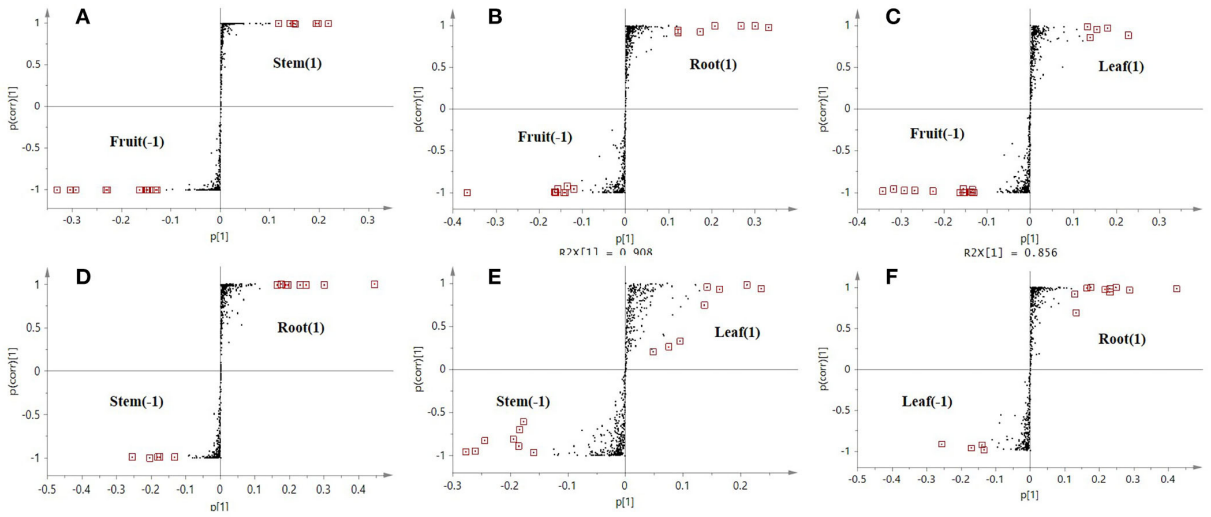


FIGURE 4
S-plot score from the six different comparison models along with selected candidate marker compounds from each model **(A–F)**.

by the ions at both ends of the S-plot (the red box marked in Figure 4).

The results showed that 193 candidate marker compounds were screened and marked. These marker compounds were introduced into Progenesis QI for further identification. Of 193 compounds, 76 marker compounds were inferred by filtering defining parameters with an ANOVA p-value of $\leq$0.05 and a max fold change of $\geq$2 and retrieving from multiple databases (Table 2). Furthermore, 54 of 76 compounds were reported from other *Schisandra* species, and 9 compounds were identified for the first time from *Schisandra*. An example in Figure 5 shows that the marker compounds were identified. The S-plot score of the comparison between SS and SL is shown in Figure 5B. The screened marker ions of the top S-plot are enlarged in Figure 5B1, and the screened ions of the bottom S-plot are displayed in Figure 5B2. The trends of the content variable are shown in Figure 5C. The averages of the selected markers showed the differences between SS and SL.

Evaluation of antioxidant activities

2,2′-Diphenyl-1-picrylhydrazyl (DPPH) free radicals are stable nitrogen center free radicals, and their ethanolic solution is purple and has the largest absorption peak at 517 nm. With the addition of antioxidants, DPPH captures an electron and pairs with free electrons, and the purple fades and becomes colorless. Based on this principle, the antioxidant ability can be detected and is widely used in the research of antioxidant foods, health products, and drugs. Figure 6A shows the free radical scavenging activity of the SR, SS, SL, and SF from *S. chinensis* at different concentrations. 2,2′-Diphenyl-1-picrylhydrazyl radicals exhibited a dose-dependent effect; at concentrations above 0.5 mg/mL, the antioxidant activity showed no distinction. However, at low concentrations (0.031 mg/mL), the antioxidant activity of roots and stems was much higher than that of leaves and fruits. The IC_{50} (half maximal inhibitory concentration) is used as an indicator of antioxidant activities. Figure 6B shows that SR had the strongest antioxidant capacity (IC_{50} 0.064 $\pm$ 0.005 mg/mL), followed by SS (IC_{50} 0.069 $\pm$ 0.006 mg/mL), SL (IC_{50} 0.112 $\pm$ 0.007 mg/mL), and SF (IC_{50} 0.140 $\pm$ 0.013 mg/mL). One-way ANOVA revealed significant differences between the four groups ($p < 0.01$). Overall, all the extracts from different parts of 95% ethanol exhibited DPPH scavenging activity, with the root activity being the highest (Figure 6).

Evaluation of anti-inflammatory activity on RWA264.7 cells

The anti-inflammatory properties of the alcohol extracts from the selected medicinal herbs were evaluated based on their ability to inhibit the production of NO in LPS-activated mouse macrophages. The effects of the different parts on cell viability were evaluated by the CCK-8 assay.

The results of the cytotoxicity test showed that, below 200 μg/mL, the SR, SS, SL, and SF of *S. chinensis* were not cytotoxic; therefore, 200, 100, 50, and 25 μg/mL were selected to verify the anti-inflammatory activity. The NO content results

TABLE 2 A total of 76 metabolites identified in positive ion mode.

No.	RT-EM	Tentative identification	Adducts	Formula	Mass error (ppm)	UPLC-QTOF-MS			
						F	L	S	R
1	1.04_371.1627*m/z*	(+)-Chelidonine	M+NH_4	$C_{20}H_{19}NO_5$	5.10	√	√	√	√
2	1.04_448.1017n	Plantaginin	M+Na, 2M+Na, M+H	$C_{21}H_{20}O_{11}$	1.04	√	√	√	√
3	1.04_464.0911n	Quercetin-3-O-glucoside	M+Na, 2M+Na, M+H	$C_{21}H_{20}O_{12}$	−6.42		√		
4	1.04_579.1526*m/z*	Procyanidin B_2	M+H	$C_{30}H_{26}O_{12}$	5.03	√	√	√	√
5	1.11_544.2274n	Schindilactone C	M+Na, 2M+Na, M+NH_4	$C_{29}H_{36}O_{10}$	−6.25		√		
6	1.60_819.4292*m/z*	Mitragynine	2M+Na	$C_{23}H_{30}N_2O_4$	−1.43	√	√	√	√
7	1.64_375.2143*m/z*	Schineolignins B	M+H	$C_{22}H_{30}O_5$	−6.22		√		
8	1.70_328.1397*m/z*	N-Fructosyl phenylalanine	M+H	$C_{15}H_{21}NO_7$	1.79	√	√	√	√
9	1.70_364.1524n	Coralyne	M+H, 2M+H	$C_{22}H_{22}NO_4$	−6.92	√	√	√	√
10	1.84_657.2447*m/z*	Piperlongumine	2M+Na	$C_{17}H_{19}NO_5$	4.48	√	√	√	√
11	2.43_733.3994*m/z*	Hirsuteine	2M+H	$C_{22}H_{26}N_2O_3$	4.65	√	√	√	√
12	2.76_284.0696n	Wogonin	2M+H, 2M+Na, M+H, M+Na	$C_{16}H_{12}O_5$	3.99	√	√	√	√
13	2.76_432.2149n	Schisandrol A*	M+H, M+NH_4, 2M+Na	$C_{24}H_{32}O_7$	0.14			√	√
14	2.76_668.3034*m/z*	Aconitine	M+Na	$C_{34}H_{47}NO_{11}$	−1.10	√			
15	3.12_530.2153n	Gomisin D*	M+Na, 2M+Na, 2M+H, M+NH_4, M+H	$C_{28}H_{34}O_{10}$	0.12	√	√	√	√
16	3.22_497.2856*m/z*	Atractylenolide III	2M+H	$C_{15}H_{20}O_3$	−6.39	√	√	√	√
17	3.29_389.1980*m/z*	Gomisin J*	M+H	$C_{22}H_{28}O_6$	−5.97	√	√	√	√
18	3.29_580.1782n	Naringin	M+NH_4, M+Na	$C_{27}H_{32}O_{14}$	−1.66	√	√	√	√
19	3.51_254.0595n	Daidzein	M+H, 2M+H, M+NH_4	$C_{15}H_{10}O_4$	6.36	√	√	√	√
20	3.51_416.1839n	Schisandrol B*	M+Na, 2M+Na, M+NH_4	$C_{23}H_{28}O_7$	0.98			√	√
21	4.14_523.2269*m/z*	Tigloylgomisin H	M+Na	$C_{28}H_{36}O_8$	−6.67	√	√	√	√
22	4.31_390.2037n	Pregomisin*	M+H, M+Na	$C_{22}H_{30}O_6$	−1.44	√	√	√	√
23	4.63_240.0872*m/z*	Isofraxidin	M+NH_4	$C_{11}H_{10}O_5$	2.69	√	√	√	√
24	4.63_357.1353*m/z*	3-Methoxycinnamic acid	2M+H	$C_{10}H_{10}O_3$	5.79	√	√	√	√
25	4.63_439.1728*m/z*	5,6,7,8-Tetrahydro-1,2,3,13-tetramethoxy-6,7-Dimethylbenzo[3,4]cycloocta[1,2-f][1,3]benzodioxol-5-ol	M+Na	$C_{23}H_{28}O_7$	0.23			√	√
26	4.73_401.1961*m/z*	Kadsuranin	M+H	$C_{23}H_{28}O_6$	0.54	√			√
27	4.73_523.2308*m/z*	Angeloylgomisin H*	M+Na	$C_{28}H_{36}O_8$	1.08	√		√	√
28	4.94_522.2256n	Benzoylgomisin H	M+Na, 2M+Na, M+NH_4	$C_{30}H_{34}O_8$	0.48	√	√	√	√
29	5.36_206.0599n	Scoparone	M+Na, 2M+Na	$C_{11}H_{10}O_4$	5.78	√	√	√	√
30	5.36_223.0743*m/z*	Flavone	M+H	$C_{15}H_{10}O_2$	−4.86	√	√	√	√
31	5.36_225.0905*m/z*	Flavanone	M+H	$C_{15}H_{10}O_2$	−2.25	√	√	√	√
32	5.36_430.1996n	Schisanchinins C	M+H, M+Na, 2M+H, 2M+Na	$C_{24}H_{30}O_7$	0.96		√	√	
33	5.36_576.3173*m/z*	Propindilactone J	M+NH_4	$C_{31}H_{42}O_9$	1.09			√	
34	5.40_727.3490*m/z*	Gelsevirine	2M+Na	$C_{21}H_{24}N_2O_3$	3.43	√	√	√	√
35	5.62_536.2053n	Gomisin G*	M+Na, M+NH_4	$C_{30}H_{32}O_9$	1.33	√	√	√	√
36	5.82_514.2208n	(-)-Tigloyl-deangeloyl-gomisin F	M+Na, 2M+Na, M+NH_4	$C_{28}H_{34}O_9$	1.03			√	

(Continued)

TABLE 2 (Continued)

No.	RT-EM	Tentative identification	Adducts	Formula	Mass error (ppm)	UPLC-QTOF-MS			
						F	L	S	R
37	6.28_402.2045n	Gomisin K_2^*	M+H, M+NH_4, M+Na, 2M+Na	$C_{23}H_{30}O_6$	0.66	√			√
38	6.38_993.3792*m/z*	Tetrahydroalstonine	2M+H	$C_{26}H_{28}N_2O_8$	2.78	√	√	√	√
39	6.56_536.2052n	Schisantherin A*	M+Na, 2M+Na, M+NH_4	$C_{30}H_{32}O_9$	0.98	√			√
40	6.56_654.2913*m/z*	Mesaconitine	M+Na	$C_{33}H_{45}NO_{11}$	4.44	√	√	√	√
41	6.57_271.0599*m/z*	Baicalein	M+H	$C_{15}H_{10}O_5$	−0.75	√	√	√	√
42	6.57_414.1684n	Kadsulignan L	M+H, M+Na, 2M+H	$C_{23}H_{26}O_7$	1.21	√			√
43	6.57_514.2207n	Tigloylgomisin P	M+Na, 2M+Na, M+NH_4	$C_{28}H_{34}O_9$	0.83	√			√
44	6.57_616.3108*m/z*	Hypaconitine	M+H	$C_{33}H_{45}NO_{10}$	−1.28	√	√	√	√
45	6.71_543.1633*m/z*	Schisantherin D*	M+Na	$C_{29}H_{28}O_9$	1.38	√	√	√	√
46	6.71_559.1378*m/z*	Formononetin	2M+Na	$C_{16}H_{12}O_4$	2.66	√	√	√	√
47	6.78_364.1502*m/z*	Isocorydine	M+Na	$C_{20}H_{23}NO_4$	−5.06	√	√	√	√
48	7.10_402.2044n	Gomisin K_1	M+H, M+Na, M+NH_4, 2M+Na	$C_{23}H_{30}O_6$	0.23	√			
49	7.20_425.1457*m/z*	Ginkgolide B	M+H	$C_{20}H_{24}O_{10}$	3.52	√	√	√	√
50	7.34_759.3292*m/z*	Formosanine	2M+Na	$C_{21}H_{24}N_2O_4$	−5.76	√	√	√	√
51	7.56_359.1485*m/z*	Matairesinol	M+H	$C_{20}H_{22}O_6$	−1.06	√	√	√	√
52	7.66_515.2279*m/z*	Schisantherin C	M+H	$C_{28}H_{34}O_9$	0.83		√		√
53	7.66_536.2025n	Gomisin C	M+H, M+Na	$C_{30}H_{32}O_9$	−4.01		√		
54	7.99_386.1728n	(+)-Gomisin M_1	M+H, M+Na, 2M+Na	$C_{22}H_{26}O_6$	−2.10	√	√	√	√
55	7.99_653.4038*m/z*	Ajmaline	2M+H	$C_{20}H_{26}N_2O_2$	−3.57	√	√	√	√
56	8.29_386.1721n	(+)-Gomisin M_2	M+H, M+Na, 2M+Na	$C_{22}H_{26}O_6$	−2.10	√	√	√	√
57	9.06_833.4501*m/z*	Schisandrin A*	2M+H	$C_{24}H_{32}O_6$	3.65	√	√	√	√
58	10.76_284.0696n	Acacetin	M+H, 2M+Na	$C_{16}H_{12}O_5$	3.85	√	√	√	√
59	10.76_290.0927n	Biochanin A	M+H, 2M+Na	$C_{16}H_{12}O_5$	3.85	√	√	√	√
60	10.76_400.1888n	Gomisin N*	M+H, M+Na, 2M+Na	$C_{23}H_{28}O_6$	0.52	√			√
61	11.29_228.0795n	3,4,5-Trihydroxystilbene	M+H, M+Na, 2M+H	$C_{14}H_{12}O_3$	3.69	√	√	√	√
62	11.29_400.1885n	Schisandrin B*	M+H, M+Na, M+NH_4, 2M+Na	$C_{23}H_{28}O_6$	0.87	√			√
63	11.32_514.2570n	Kadsufolin A	M+Na, 2M+Na, M+NH_4	$C_{29}H_{38}O_8$	0.71		√	√	
64	11.94_543.1994*m/z*	Benzoylisogomisin O	M+Na	$C_{30}H_{32}O_8$	0.72		√		
65	12.07_484.2463n	Angeloyl-(+)-gomisin K_3^*	M+H, M+NH_4, 2M+Na	$C_{28}H_{36}O_7$	0.28	√			
66	12.07_507.2361*m/z*	Xuetongdilactone E	M+Na	$C_{28}H_{36}O_7$	1.66	√			
67	12.07_530.3113*m/z*	Schinchinenins E	M+NH_4	$C_{30}H_{40}O_7$	0.17	√			
68	12.17_498.2258n	Angeloylisogomisin O*	M+Na, 2M+Na	$C_{28}H_{34}O_8$	0.82		√	√	
69	12.60_329.1769*m/z*	Anwuligan	M+H	$C_{20}H_{24}O4$	6.71	√	√	√	√
70	12.60_372.1572n	Arctigenin	M+H, M+Na	$C_{21}H_{24}O_6$	−0.19	√	√	√	√
71	12.60_384.1557n	Schisandrin C*	M+H, M+Na, 2M+Na	$C_{22}H_{24}O_6$	−4.19	√	√	√	√
72	12.96_520.2100n	Benzoylgomisin O*	M+Na, 2M+Na, M+NH_4	$C_{30}H_{32}O_8$	0.59	√	√	√	√
73	13.06_505.1838*m/z*	Interiotherin A*	M+H	$C_{29}H_{28}O_8$	−3.67	√	√	√	√
74	13.19_498.2257n	Angeloylgomisin O*	M+Na, 2M+Na, M+NH_4	$C_{28}H_{34}O_8$	0.70	√	√	√	√
75	14.15_514.2802*m/z*	Kadlongilactone F	M+NH_4	$C_{29}H_{36}O_7$	0.63				√
76	14.15_653.3302*m/z*	Gelsenicine	2M+H	$C_{19}H_{22}N_2O_3$	−4.87	√	√	√	√

*Indicates that it is identified by the standard.

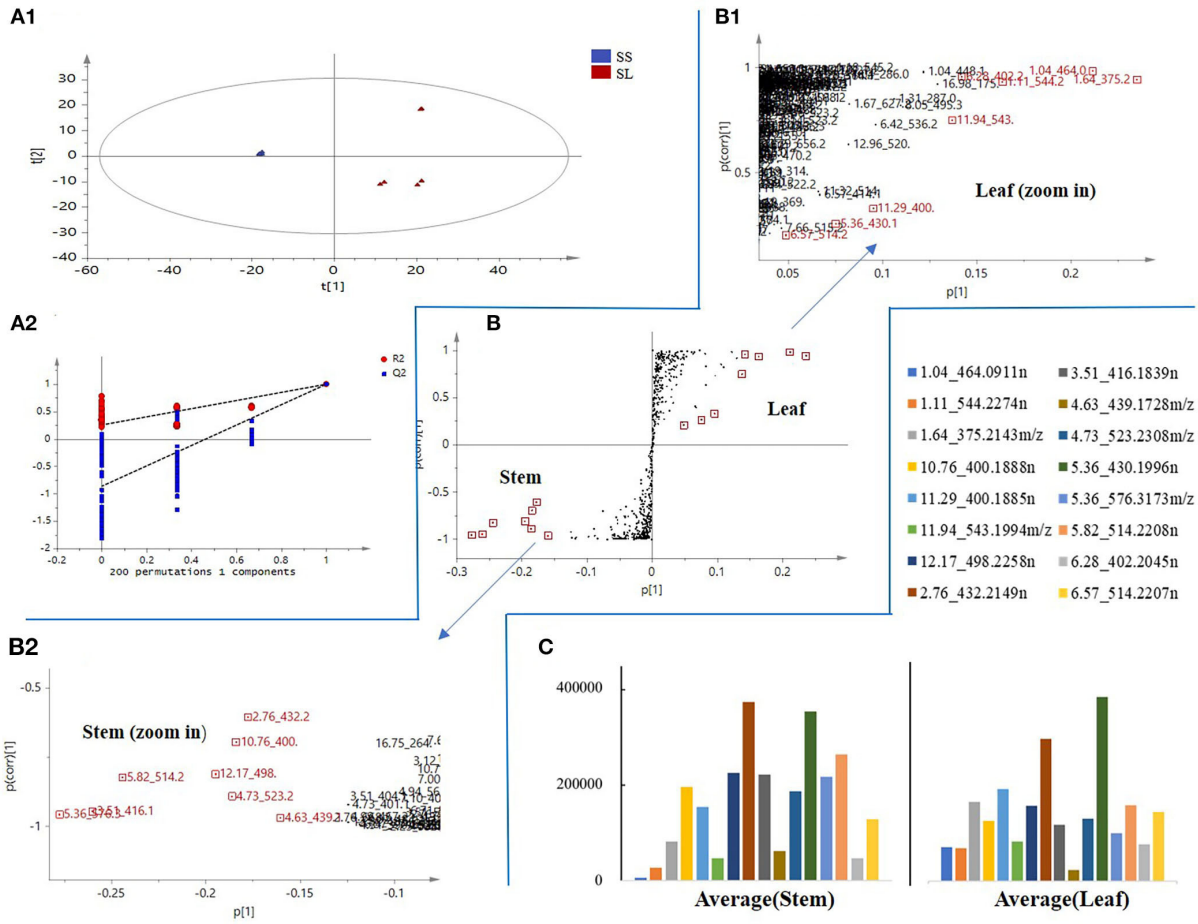


FIGURE 5
An example of marker compound selection in *S. chinensis*. **(A)** orthogonal partial least-squares discriminant analysis (OPLS-DA) score plot and the reliability of the OPLS-DA model. **(B)** S-plot score of SS and SL. **(C)** The trend of the relative contents of different compounds in SS and SL.

showed that there was a very significant difference between the model and the blank (##, $P < 0.01$), suggesting that LPS (1 μg/mL) could induce an increase in NO produced by RAW264.7 cells, and 100 and 200 μg/mL of alcohol extracts of roots, stems, leaves, and fruits could significantly reduce the expression of NO (**, $P < 0.01$). Below 100 μg/mL, the effects of alcohol extracts from four different parts were SR > SS > SF > SL (Figure 6).

Potential biomarkers from the root extracts

Recently, research on the antioxidant and anti-inflammatory activities of *S. chinensis* has mainly focused on SF extract. In our study, other parts of *S. chinensis* also showed significant antioxidant and anti-inflammatory activities. In addition, SR extracts exhibited higher antioxidant and anti-inflammatory activities than SF extracts. The chemical compositions of SR and SF were significantly different according to the PCA score plot (Figure 2). The potential biomarkers were identified by OPLS-DA together with an S-plot based on the difference shown in the PCA score plot (Figure 7).

The OPLS-DA statistical results showed the difference between SR and SF. The statistical model verified that the fit was good, with a cumulative R2 Y value of 0.99 and a Q2 value of 0.99 (Figure 7A). Twelve potential biomarkers were selected through the S-plot analysis of the SR and SF (Figure 7B). From the VIP plot analysis, the 12 selected potential biomarkers in SR had a higher VIP value (VIP > 3), indicating that these marker compounds were significantly different between SR and SF. In addition, the mean values of variables in each group showed that the selected compounds in SR and SF had different levels (Figure 7C). As shown in Figures 7B–D and Table 2, the top five potential biomarkers between SR and SF were 530.2153n (RT 3.12 min), 416.1839n (RT 3.51 min), 515.2279n (RT 7.66 min), 400.1885n (RT 11.29 min), and 514.2802 *m/z* (RT 14.15 min). The related compounds may

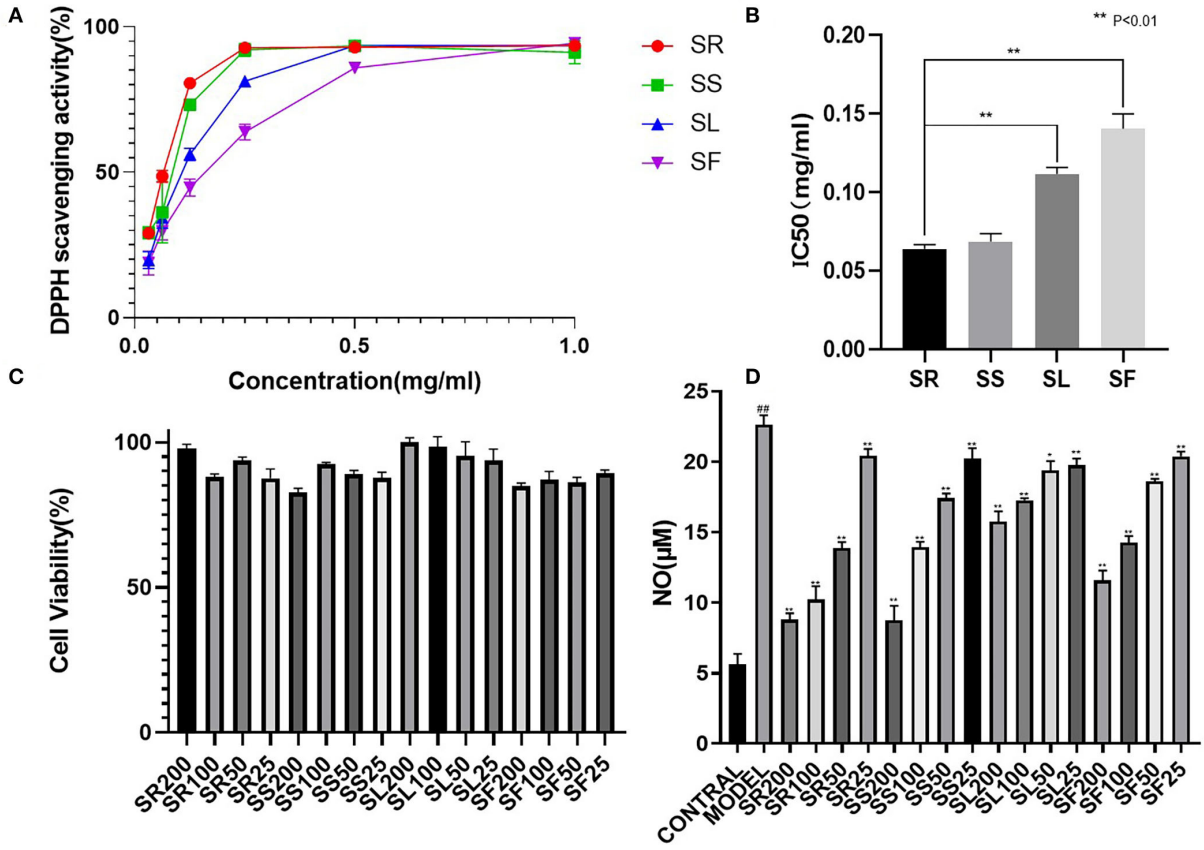


FIGURE 6
Comparative 2,2′-diphenyl-1-picrylhydrazyl (DPPH) scavenging capacity and anti-inflammatory activity of *S. chinensis*. **(A)** DPPH scavenging activity of different plant parts. **(B)** The half maximal inhibitory concentration (IC_{50}) values of different plant parts. **(C)** Cell viability of different plant parts. **(D)** Anti-inflammatory activity of different plant parts. The results were expressed as the mean ± standard deviation, and triplicate experiments were performed for each sample.

play a key role in the higher biological activity of SR relative to SF.

The ion at 416.1839 *m/z* (RT 3.51 min) was identified as schisandrol B. The compound identification was based on the major matched fragment ions, including *m/z* 439.1736 $[M+Na]^+$; in the process of substituent cracking with schisandrol B, H_2O molecules were lost, and the fragment ions of $[M+H-H_2O]^+$ (*m/z* 399.1820) were found. Fragment ions including $[M+H-H_2O-OCH_3]^+$ (*m/z* 368.1626) and $[M+H_2O-C_4H_6]^+$ (*m/z* 345.13) were found *via* further skeleton cracking reported in *S. chinensis* and were confirmed by reference standards (Figure 7D). The ion 530.2153n (RT 3.12 min) was identified as gomisin D, and the formula was $C_{24}H_{32}O_6$. In the positive ion mode, fragment ions at *m/z* 531 $[M+H]^+$, *m/z* 553 $[M+Na]^+$, *m/z* 569 $[M+K]^+$, *m/z* 485 $[M+H-CO_2]^+$, *m/z* 285 $[M+H-CO_2-C_4H_7COOH-C_4H_8]^+$, and *m/z* 401 $[M+H-C_4H_7COOH-CH_2O]^+$ were obtained. The main fragment ions produced were at *m/z* 485.2156 due to a loss of CO_2 (44 Da), at *m/z* 401.1588 due to a further loss of C_4H_7COOH (100 Da) and CH_2O (30 Da), and at *m/z* 285.0888 due to a loss of C_4H_7COOH (100 Da) and C_4H_8 (56 Da). The parent ions and the main fragmentation were consistent with the literature MS data for gomisin D (Figure 7E).

We detected the contents of 9 common lignans including schisandrol A (1), gomisin D (2), gomisin J (3), schisandrol B (4), angeloylgomisin H (5), gomisin G (6), schisantherin A (7), schisandrin B (8), and schisandrin C (8) in different parts of *S. chinensis*. The results indicated that the content of nine lignans varied greatly from different parts, and the total content of nine lignans was the highest in roots and the lowest in leaves. The content of schisandrol A and gomisin J in SF was higher than that in other parts. The content of gomisin D, schisandrol B, and schisantherin C in SR was significantly higher than that in other parts (Table 3).

Gomisin D is a lignan found in SFs and has been demonstrated to exert an antidiabetic effect and inhibit

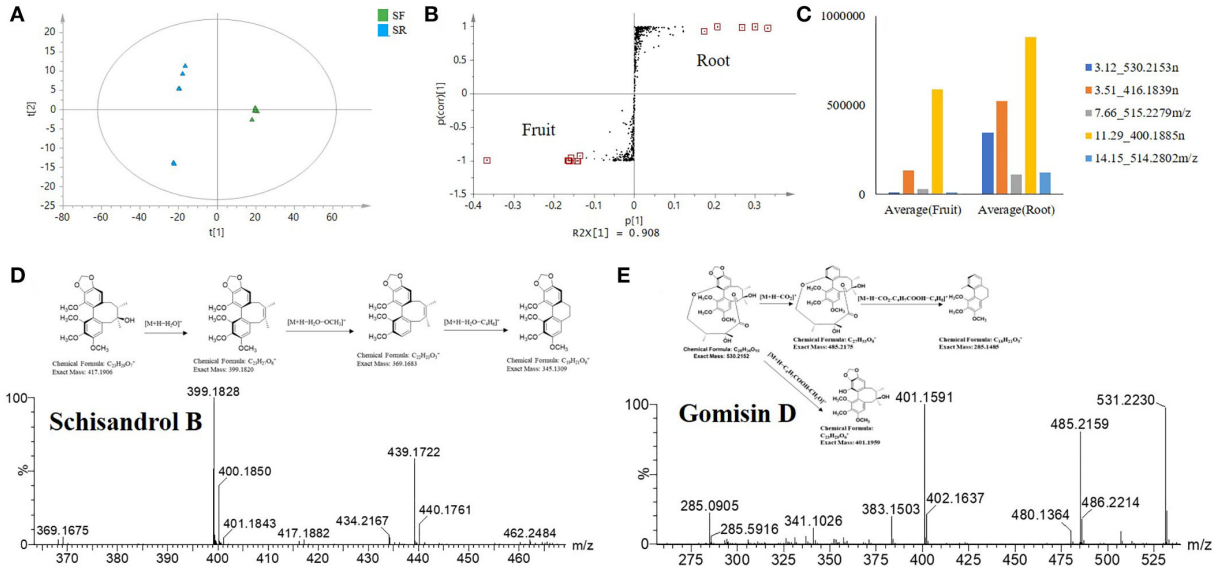


FIGURE 7
Selected *S. chinensis* potential biomarkers from the comparative analysis of SR and SF based on the S-plot. **(A)** OPLS-DA. **(B)** S-plot showing the selected markers. **(C)** Variable averages by a group of selected potential marker compounds. **(D)** Cracking process of schisandrol B. **(E)** Cracking process of gomisin D.

UDP-glucuronosyltransferase activity (Zhang et al., 2010; Song et al., 2015). In addition, gomisin D can scavenge ABTS (+) radicals and treat Alzheimer's disease (Mocan et al., 2016). Recently, gomisin D was used as a quality marker of Shengmai San and Shenqi Jiangtang Granule (Zhang et al., 2018). Schisandrol B is a lignan that can be isolated from dried SFs and exhibits hepatoprotective, cardioprotective, renoprotective, and memory-enhancing properties (Kim et al., 2008; Panossian and Wikman, 2008; Jiang et al., 2015, 2016; Zeng et al., 2017). There is also prior evidence that schisandrol B can partially suppress or prevent vascular fibrotic disorders by disrupting TGFβ1-assisted NF-κB signaling (Chun et al., 2018). In mice, schisandrol B can also drive benign liver enlargement, which is consistent with enhanced hepatocyte energy metabolism and energy utilization (Zhao et al., 2020). Kadsuranin, which is an anti-inflammatory marker that was identified, is also dibenzocyclooctene lignan derivative, and further experimental study may need to be performed to obtain the structure-activity relationship. Other researchers, who studied the antioxidant and anti-inflammatory activities of SF extracts, found that galloylated compounds, mainly dibenzocyclooctene lignan, are important bioactive constituents. However, gomisin D can be used as an important chemical marker of SR, which has significantly different contents in SF and SR. Gomisin D, schisandrol B, schisantherin C, kadsuranin, and kadlongilactone F could be considered chemical markers in the roots, which could support the root extracts has higher antioxidant and anti-inflammatory activities, and provide a new application direction for studying the parts with the antioxidant and anti-inflammatory activities from *S. chinensis*.

Conclusion

A local compounds database from the genus *Schisandra* was established by Progenesis SDF Studio. The metabolite characterization in the root, stem, leaf and fruit of *Schisandra chinensis* was performed by a UPLC-QTOF-MS method using chemometrics tools to identify biomarkers of different parts of the plant. Through the screening of the antioxidant and anti-inflammatory activities of different parts of *S. chinensis in vitro*, we found alternative sources of antioxidants and anti-inflammatory compounds to study the correlation between chemical composition and biological activities of the plant parts of *S. chinensis*. Gomisin D, schisandrol B, schisantherin C, kadsuranin, and kadlongilactone F, as biomarkers from roots could support the material basis for the higher antioxidant and anti-inflammatory activities that are found in root extracts compared to fruits and provide a new application direction for studying the parts that exhibit the antioxidant and anti-inflammatory activities from *S. chinensis*. The plant roots could be a new medicinal source that exhibits better activity than that of traditional medicinal parts, which makes them worth exploring. In this study, the metabolic profiling of different parts of *S. chinensis* was characterized by using the theory of pharmacophylogeny, to expand the search for

TABLE 3 Contents of nine lignans in different parts of *Schisandra chinensis* ($n = 3$, mg/g).

		1	2	3	4	5	6	7	8	9	Total
S1	Fruit (SF)	12.79	1.08	2.13	5.31	2.40	0.57	0.54	7.22	1.35	33.39
	Stems (SS)	6.44	1.89	0.04	8.92	1.98	0.86	0.12	1.91	0.83	22.99
	Leaves (SL)	4.19	1.49	0.10	4.89	1.15	0.00	0.09	1.06	0.09	13.06
	Roots (SR)	6.79	24.13	0.02	12.97	2.09	0.67	0.49	10.34	0.96	58.46
S2	Fruit (SF)	11.09	1.12	1.72	4.80	2.01	0.58	0.36	8.87	2.27	32.82
	Stems (SS)	5.55	1.90	0.01	7.98	1.66	0.80	0.08	2.33	1.33	21.64
	Leaves (SL)	3.64	1.53	0.05	4.66	1.00	0.03	0.06	1.22	0.14	12.35
	Roots (SR)	5.59	23.27	0.00	11.29	1.72	0.64	0.31	12.20	1.46	56.48
S3	Fruit (SF)	12.72	0.78	1.27	3.29	1.90	0.33	0.22	6.34	1.69	28.54
	Stems (SS)	6.26	1.34	0.00	5.41	1.53	0.54	0.05	1.64	1.01	17.77
	Leaves (SL)	4.25	1.08	0.04	2.06	1.01	0.02	0.06	2.43	0.59	11.54
	Roots (SR)	6.74	17.16	0.00	8.10	1.68	0.43	0.22	9.71	1.00	45.05
S4	Fruit (SF)	13.81	1.13	1.59	3.80	2.38	0.52	0.43	7.24	1.37	32.27
	Stems (SS)	6.97	1.88	0.01	6.41	2.01	0.80	0.09	1.89	0.88	20.95
	Leaves (SL)	4.76	1.58	0.03	2.48	1.36	0.02	0.13	2.83	0.52	13.72
	Roots (SR)	7.64	25.36	0.00	9.65	2.19	0.65	0.44	11.47	0.89	58.29
S5	Fruit (SF)	12.35	0.91	1.61	2.98	1.90	0.60	0.32	5.51	1.05	27.23
	Stems (SS)	6.31	1.54	0.01	5.03	1.60	0.97	0.07	1.47	0.69	17.68
	Leaves (SL)	6.57	0.59	0.02	3.18	1.31	0.50	0.12	2.51	0.53	15.34
	Roots (SR)	6.48	22.74	0.00	7.00	1.66	0.42	0.29	8.38	0.62	47.58
S6	Fruit (SF)	10.08	1.08	1.05	2.67	1.67	0.51	0.35	4.67	0.85	22.93
	Stems (SS)	5.09	1.83	0.01	4.40	1.41	0.87	0.08	1.22	0.52	15.44
	Leaves (SL)	5.44	0.72	0.02	2.90	1.21	0.52	0.13	2.16	0.46	13.56
	Roots (SR)	5.35	27.56	0.00	6.36	1.50	0.32	0.32	7.32	0.54	49.25

alternative drug sources. This study explored the chemical components and biological activities of the nonmedicinal parts of *S. chinensis*, which can be used to maximize the comprehensive utilization benefits of *S. chinensis* resources. Our next step is to isolate the labeled compounds of *S. chinensis* and further confirm the biological activity of the compound.

Credit authority statement

The work described in the manuscript has not been submitted elsewhere for publication as a whole. The relevant contents and data of the paper met the requirements of integrity. All authors have contributed and read the manuscript and agreed to its submission.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author/s.

Author contributions

JiuL was responsible for data analysis and writing of the manuscript. XM, JinL, JZ, and TQ did experimental work. BL performed data analysis work. BZ and HoL collected and identified samples. HaL conceived the design of the study. All authors contributed to the article and approved the submitted version.

Funding

The authors gratefully acknowledge the financial support by the National Natural Science Foundation of China (82204576), CAMS Innovation Fund for Medical Sciences (2021-I2M-1-031), and the Ability Establishment of Sustainable Use for Valuable Chinese Medicine Resources (2060302-2002-05).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships

that could be construed as a potential conflict of interest.

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.2022.970535/full#supplementary-material

References

Brereton, R. G. (2009). *Chemometrics for Pattern Recognition*. Hoboken, NJ: John Wiley and Sons.

Chaleckis, R., Meister, I., Zhang, P., and Wheelock, C. E. (2019). Challenges, progress and promises of metabolite annotation for LC-MS-based metabolomics. *Curr. Opin. Biotechnol.* 55, 44–50. doi: 10.1016/j.copbio.2018.07.010

Chen, L. J., Hu, P. P., and Xue, W. L. (2011). Research on tea beverages form fresh twigs and leaves of *Schisandra Chinense*. *J. Anhui Agri. Sci.* 39, 2647–2678. doi: 10.13989/j.cnki.0517-6611.2011.05.004

Chen, Q., Bao, L., Lv, L., Xie, F., Zhou, X., Zhang, H., et al. (2021). Schisandrin B regulates macrophage polarization and alleviates liver fibrosis via activation of PPARγ. *Ann. Transl. Med.* 9, 1500. doi: 10.21037/atm-21-4602

Chen, X., Tang, R., Liu, T., Dai, W., Liu, Q., Gong, G., et al. (2019). Physicochemical properties, antioxidant activity and immunological effects *in vitro* of polysaccharides from *Schisandra sphenanthera* and *Schisandra chinensis*. *Int. J. Biol. Macromol.* 131, 744–751. doi: 10.1016/j.ijbiomac.2019.03.129

Chinese Pharmacopoeia Commission (2020). *The Pharmacopoeia of the People's Republic of China, 2020 Edition Part I*. Beijing: China Medical Science Press.

Chun, J. N., Park, S., Lee, S., Kim, J. K., Park, E. J., Kang, M. J., et al. (2018). Schisandrol B and schisandrin B inhibit TGFβ1-mediated NF-κB activation via a Smad-independent mechanism. *Oncotarget* 9, 3121–3130. doi: 10.18632/oncotarget.23213

Dai, W. D., Qi, D. D., Yang, T., Lv, H. P., Guo, L., Zhang, Y., et al. (2015). Nontargeted analysis using ultraperformance liquid chromatography-quadrupole time-of-flight mass spectrometry uncovers the effects of harvest season on the metabolites and taste quality of tea (*Camellia sinensis* L.). *J. Agric. Food Chem.* 63, 9869–9878. doi: 10.1021/acs.jafc.5b03967

Gui, Y., Yang, Y., Xu, D., Tao, S., and Li, J. (2020). Schisantherin A attenuates sepsis-induced acute kidney injury by suppressing inflammation via regulating the NRF2 pathway. *Life Sci.* 258, 118161. doi: 10.1016/j.lfs.2020.118161

Guo, L. Y., Hung, T. M., Bae, K. H., Shin, E. M., Zhou, H. Y., Hong, Y. N., et al. (2008). Anti-inflammatory effects of schisandrin isolated from the fruit of *Schisandra chinensis* Baill. *Eur. J. Pharmacol.* 591, 293–299. doi: 10.1016/j.ejphar.2008.06.074

Han, J. S., Lee, S., Kim, H. Y., and Lee, C. H. (2015). MS-based metabolite profiling of aboveground and root components of *Zingiber mioga* and *officinale*. *Molecules* 20, 16170–16185. doi: 10.3390/molecules200916170

Hu, D., Cao, Y., He, R., Han, N., Liu, Z., Miao, L., et al. (2012). Schizandrin, an antioxidant lignan from *Schisandra chinensis*, ameliorates Aβ1-42-induced memory impairment in mice. *Oxid. Med. Cell. Longev.* 2012, 721721. doi: 10.1155/2012/721721

Hur, M., Campbell, A. A., Almeida-de-Macedo, M., Li, L., Ransom, N., Jose, A., et al. (2013). A global approach to analysis and interpretation of metabolic data for plant natural product discovery. *Nat. Prod. Rep.* 30, 565–583. doi: 10.1039/c3np20111b

Jandric, Z., Frew, R. D., Fernandez-Cedi, L. N., and Cannavan, A. (2017). An investigative study on discrimination of honey of various floral and geographical origins using UPLC-QTOF-MS and multivariate data analysis. *Food Control* 72, 189–197. doi: 10.1016/j.foodcont.2015.10.010

Jiang, Y. M., Fan, X. M., Wang, Y., Chen, P., Zeng, H., Tan, H., et al. (2015). Schisandrol B protects against acetaminophen-induced hepatotoxicity by inhibition of CYP-mediated bioactivation and regulation of liver regeneration. *Toxicol. Sci.* 143, 107–115. doi: 10.1093/toxsci/kfu216

Jiang, Y. M., Wang, Y., Tan, H. S., Yu, T., Fan, X. M., Chen, P., et al. (2016). Schisandrol B protects against acetaminophen-induced acute hepatotoxicity in mice via activation of the NRF2/ARE signaling pathway. *Acta Pharmacol. Sin.* 37, 382–389. doi: 10.1038/aps.2015.120

Kim, S. H., Kim, Y. S., Kang, S. S., Bae, K., Hung, T. M., and Lee, S. M. (2008). Anti-apoptotic and hepatoprotective effects of gomisin A on fulminant hepatic failure induced by D-galactosamine and lipopolysaccharide in mice. *J. Pharmacol. Sci.* 106, 225–233. doi: 10.1254/jphs.fp0071738

Lever, J., Krzywinski, M., and Altman, N. (2017). Points of significance: principal component analysis. *Nat. Methods* 14, 641–643. doi: 10.1038/nmeth.4346

Lin, C. C., Xu, Z. Y., Wang, B. H., Zhuang, W. Y., Sun, J. H., Li, H., et al. (2021). Relaxation effect of *Schisandra chinensis* lignans on the isolated tracheal smooth muscle in rats and its mechanism. *J. Med. Food* 24, 825–832. doi: 10.1089/jmf.2021.K.0037

Liu, G. Z., Liu, Y., Sun, Y. P., Li, X. M., Xu, Z. P., Jiang, P., et al. (2020). Lignans and terpenoids from the leaves of *Schisandra chinensis*. *Chem. Biodivers.* 17, e2000035. doi: 10.1002/cbdv.202000035

Lu, B. Y., Li, M. Q., and Yin, R. (2016). Phytochemical content, health benefits, and toxicology of common edible flowers: a review (2000–2015). *Crit. Rev. Food Sci. Nutr.* 56, 130–148. doi: 10.1080/10408398.2015.1078276

Medica Editorial Board of National Institute of Chinese Medicine (1999). *Chinese Materia Medica*. Shanghai: Shanghai Scientific and Technical Publishers.

Mocan, A., Schafberg, M., Cri?an, G., and Rohn, S. (2016). Determination of lignans and phenolic components of *Schisandra chinensis* (Turcz.) Baill. using HPLC-ESI-ToF-MS and HPLC-online TEAC: contribution of individual components to overall antioxidant activity and comparison with traditional antioxidant assays. *J. Funct. Foods* 24, 579–594. doi: 10.1016/j.jff.2016.05.007

Mu, X. L., Liu, J. S., Li, B., Wei, X. P., Qi,. Y. D., Zhang, B. G., et al. (2022). A comparative study on chemical characteristics, antioxidant and hepatoprotective activity from different parts of *Schisandrae chinensis* Fructus. *J. Food Process. Pres.* 2022, e16990. doi: 10.1111/jfpp.16990

Panossian, A., and Wikman, G. (2008). Pharmacology of *Schisandra chinensis* Bail: an overview of Russian research and uses in medicine. *J. Ethnopharmacol.* 118, 183–212. doi: 10.1016/j.jep.2008.04.020

Song, J. H., Cui, L., An, L. B., Li, W. T., Fang, Z. Z., Zhang, Y. Y., et al. (2015). Inhibition of UDP-glucuronosyltransferases (UGTS) activity by constituents of *Schisandra chinensis*. *Phytother. Res.* 29, 1658–1664. doi: 10.1002/ptr.5395

Szopa, A., Barnaś, M., and Ekiert, H. (2019). Phytochemical studies and biological activity of three Chinese Schisandra species (*Schisandra sphenanthera, Schisandra henryi* and *Schisandra rubriflora*): current findings and future applications. *Phytochem. Rev.* 18, 109–128. doi: 10.1007/s11101-018-9582-0

Szopa, A., Ekiert, R., and Ekiert, H. (2017). Current knowledge of *Schisandra chinensis* (Turcz.) Baill. (Chinese magnolia vine) as a medicinal plant species: a review on the bioactive components, pharmacological properties, analytical and biotechnological studies. *Phytochem. Rev.* 16, 195–218. doi: 10.1007/s11101-016-9470-4

Trygg, J., Holmes, E., and Lundstedt, T. (2007). Chemometrics in metabonomics. *J. Proteome Res.* 6, 469–479. doi: 10.1021/pr060594q

Usman, M., Bokhari, S. A. M., Fatima, B., Rashid, B., Nadeem, F., and Sarwar, M. B. (2022). Drought stress mitigating morphological, physiological, biochemical, and molecular responses of guava (*Psidium guajava* L.) cultivars. *Front. Plant Sci.* 13:878616. doi: 10.3389/fpls.2022.878616

Wang, X., Yu, J., Li, W., Wang, C., Li, H., Ju, W., et al. (2018). Characteristics and antioxidant activity of lignans in *Schisandra chinensis* and *Schisandra sphenanthera* from different locations. *Chem. Biodivers.* 15, e1800030. doi: 10.1002/cbdv.201800030

Xu, G. Y., Lv, X., Feng, Y. B., Li, H., Chen, C., Lin, H., et al. (2021). Study on the effect of active components of *Schisandra chinensis* on liver injury and its mechanisms in mice based on network pharmacology. *Eur. J. Pharmacol.* 910, 174442. doi: 10.1016/j.ejphar.2021.174442

Xu, J. B., Gao, G. C., Yuan, M. J., Huang, X., Zhou, H. Y., Zhang, Y., et al. (2020). Lignans from *Schisandra chinensis* ameliorate alcohol and CCl_4-induced long-term liver injury and reduce hepatocellular degeneration via blocking ETBR. *J. Ethnopharmacol.* 258, 112813. doi: 10.1016/j.jep.2020.112813

Xu, L., Grandi, N., Del, V. C., Mandas, D., Corona, A., Piano, D., et al. (2015). From the traditional Chinese medicine plant *Schisandra chinensis* new scaffolds effective on HIV-1 reverse transcriptase resistant to non-nucleoside inhibitors. *J. Microbiol.* 53, 288–293. doi: 10.1007/s12275-015-4652-0

Yan, T., Wang, N., Liu, B., Wu, B., Xiao, F., He, B., et al. (2021). *Schisandra chinensis* ameliorates depressive-like behaviors by regulating microbiota-gut-brain axis via its anti-inflammation activity. *Phytother. Res.* 35, 289–296. doi: 10.1002/ptr.6799

Yang, K., Qiu, J.ing., Huang, Z. C., Yu, Z. W., Wang, W. J., Hu, H. L., et al. (2021). A comprehensive review of ethnopharmacology, phytochemistry, pharmacology, and pharmacokinetics of *Schisandra chinensis* (Turcz.) Baill. and *Schisandra sphenanthera* Rehd. et Wils. *J. Ethnopharmacol.* 284, 114759. doi: 10.1016/j.jep.2021.114759

Yang, S., and Yuan, C. H. (2021). *Schisandra chinensis*: a comprehensive review on its phytochemicals and biological activities. *Arab. J. Chem.* 14, 103310. doi: 10.1016/j.arabjc.2021.103310

Ye, J. W., Zhang, Z. K., Wang, H. F., Bao, L., and Ge, J. P. (2019). Phylogeography of *Schisandra chinensis* (Magnoliaceae) reveal multiple refugia with ample gene flow in Northeast China. *Front. Plant Sci.* 10:199. doi: 10.3389/fpls.2019.00199

Zeng, H., Jiang, Y. M., Chen, P., Li, D. S., Li, D. S., Liu, D. S., et al. (2017). Schisandrol B protects against cholestatic liver injury through pregnane X receptors. *Brit. J. Pharmacol.* 174, 672–688. doi: 10.1111/bph.13729

Zhang, H., Zhang, X. J., Jiang, H. J., Xu, C., Tong, S. Q., and Yan, J. Z. (2018). Screening and identification of alpha-glucosidase inhibitors from shenqi jiangtang granule by ultrafiltration liquid chromatography and mass spectrometry. *J. Sep. Sci.* 41, 797–805. doi: 10.1002/jssc.201700835

Zhang, J., Shi, L. L., and Zheng, Y. N. (2010). Dibenzocyclooctadiene lignans from fructus *Schisandrae chinensis* improve glucose uptake *in vitro*. *Nat. Prod. Commun.* 5, 231–234. doi: 10.1177/1934578X1000500212

Zhao, Y. Y., Yao, X. P., Jiao, T. Y., Tian, J. N., Zhou, Y. Y., Gao, Y., et al. (2020). Lipidomics analysis on schisandrol B-induced liver enlargement in mice. *Acta Pharm. Sin.* 12, 922–929. doi: 10.16438/j.0513-4870.2019-1020

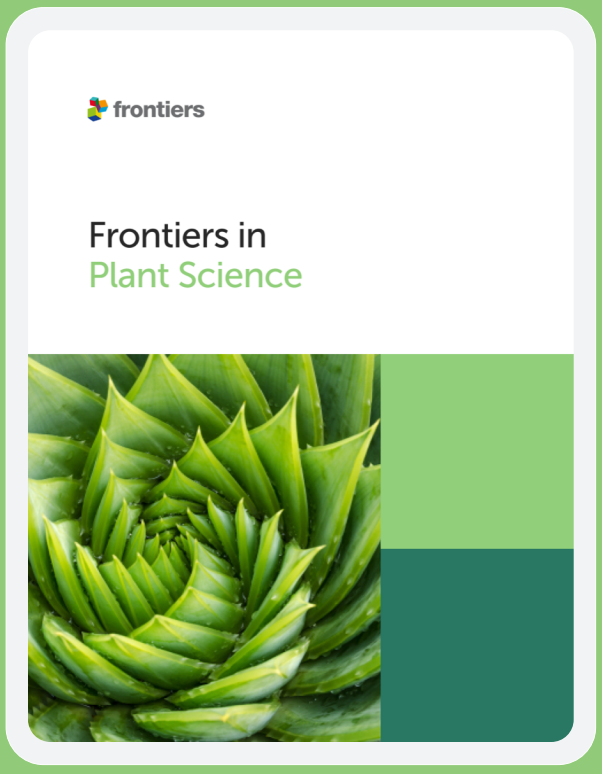
Frontiers in Plant Science

Cultivates the science of plant biology and its applications

The most cited plant science journal, which advances our understanding of plant biology for sustainable food security, functional ecosystems and human health.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact