Nutritional, functional components, and geographical origin authentication of Forsythia suspensa from different geographical origins based on UPLC-MS/MS combined with chemometrics

Tao Zheng
Shaanxi University of Technology
Nian Liu
Shaanxi University of Technology
Daojun Li
Shaanxi University of Technology
Qi Tang
Shaanxi University of Technology
Guodong Wang
Shaanxi Normal University
Zhubing Hu
Henan University
Haitao Zeng
zenghaitao@snut.edu.cn

Shaanxi University of Technology
Tiantian Chen
Shaanxi University of Technology

Research Article

Keywords: Qingqiao (Forsythia suspensa), Geographical origins, UPLC-MS/MS, Nutritional ingredients, Functional components, Authentication and traceability

Posted Date: February 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8569369/v1

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Plant Biology on July 7th, 2026. See the published version at https://doi.org/10.1186/s12870-026-09384-y.

Abstract

Background Qingqiao (*Forsythia suspensa*) is a valuable traditional medicinal herb with considerable industrial potential owing to its rich profile of nutritional and functional components. However, systematic characterization of its metabolite composition across different growing regions remains limited, constraining effective quality control and origin-based standardization for industrial applications.

Results In this study, a comprehensive UPLC-MS/MS-based metabolomics approach combined with chemometric analysis was employed to compare Qingqiao samples from 3 distinct geographical origins: the Loess Plateau (LP), the Guanzhong Plain (GZP), and the Qinling Mountains (QM). 3,052 metabolites were identified, comprising 1,114 primary and 1,938 secondary metabolites. Distinct metabolic profiles were observed among the 3 regions: LP samples were abundant in organic acids, alkaloids, lignans, coumarins, flavonoids, quinones, amino acids, and derivatives; QM samples exhibited higher levels of lipids, steroids, terpenoids, and tannins; while GZP samples were enriched in phenolic acids, nucleotides, and their derivatives. Combined content-function analysis indicated that GZP1, GZP2, GZP3, QM3, QM4, LP1, LP2, LP5, and LP6 had strong potential to develop functional products. Comparative analysis revealed 557, 667, and 359 differentially accumulated metabolites between GZP-vs-LP, GZP-vs-QM, and QM-vs-LP, respectively, which were significantly enriched in 8 key metabolic pathways. Furthermore, 14 differential metabolites were identified as origin-specific biomarkers, including two rare compounds (xylosyl phellodendroside and 2-glucosyl-glucosyloxy-2-phenylacetic acid amide) which emerged as novel region-specific markers for Qingqiao. Validation using random forest and OPLS-DA demonstrated exceptional discriminative accuracy (93.33%), confirming the robustness of these markers.

Conclusions This study elucidated the region-specific variations in nutritional and functional components of Qingqiao, and established a robust, metabolomics-driven framework for authenticating its geographical origin, offering critical insights for quality control, functional application, and industrial standardization of Qingqiao.

Introduction

Qingqiao (*Forsythia suspensa*) is a common traditional Chinese medicinal herb plant with various effects and utilizations, and the natural distribution of wild Qingqiao is widely distributed in Shanxi, Hebei, Henan, Shaanxi provinces and other places [1]. Modern pharmacological research has demonstrated multiple pharmacological properties [2], including antibacterial and antiviral, hepatoprotective, antipyretic analgesia, antiemetic, and detoxification activities, highlighting its significant potential for health product development [3]. Additionally, *F. suspensa* is available in forms such as tea, beverages, mixed drinks, and dietary supplements [4]. During the COVID-19 pandemic, Qingqiao gained widespread international recognition as an oral prophylactic agent against viral infections [5].

Currently, commercial Qingqiao resources primarily originate from major production areas in Shanxi, Henan, and Shaanxi provinces [6]. However, different growing environments significantly impacted the quality of plants (vegetables or medicinal herbs) [7]. Specifically, varying environmental conditions in the

Loess Plateau (LP), Guanzhong Plain (GZP), and Qinling Mountains (QM), including temperature, light intensity, precipitation, and soil characteristics, might lead to differences in the nutritional quality of Qingqiao. Previous researches had demonstrated that differential growing environments induced significant variations in both nutritional and functional components profiles within the same plant species [8, 9]. Nevertheless, existing studies on a limited number of known active compounds (e.g., forsythin and forsythiaside A) have primarily centered on narrow compound categories or specific regions, leaving a gap in the comprehensive assessment of its nutritional and functional components across broad geographical scales [10]. This limitation hinders in-depth development and utilization of Qingqiao resources [11]. With advancing research on the pharmacological effects of Qingqiao, the comprehensive and rapid identification of nutritional, functional components and geographical origin authentication of Qingqiao in different geographical origins has emerged as a pressing issue requiring resolution.

Recently, analytical methods for investigating metabolites in medicinal and food homologous plants and exploring their nutritional and functional components have been described, including UPLC-MS/MS, GC-MS, and NIR combined with quantitative analysis [12, 13]. However, the remarkable diversity of plant metabolites presents a significant challenge for achieving both comprehensive analysis and discriminating geographical origin simultaneously [14]. Widely-targeted metabolomics represents an advanced analytical strategy that integrates the comprehensive coverage of untargeted approaches with the precision of targeted methods through UPLC-MS/MS platforms, which has demonstrated significant utility in pharmacological and food science research, such as in *Astragali radix* [15], *Pueraria lobata* [16], daylily [17], *Sesamum indicum* [18], and other species. Therefore, it has become feasible to establish a robust analytical framework that enables systematic characterization of phytochemical variations and precise geographical origin traceability in Qingqiao through the integration of UPLC-MS/MS with chemometrics.

This study utilized widely targeted metabolomics to systematically profile both primary and secondary metabolites in Qingqiao sourced from 3 different geographical locations: the LP, GZP, and QM. The analytical data were subsequently integrated with chemometrics approaches to discern geographical discrimination patterns and identify significantly differential metabolites among samples from different origins. Our findings establish a chemical basis for quality assessment of Qingqiao from major production areas, while providing critical insights for quality control standardization and geographical origin tracing through compositional analysis. This research advanced methodological frameworks for phytochemical evaluation and offered scientific support for the sustainable utilization of this medicinal plant resource.

Materials and methods

Qingqiao samples

The Qingqiao sampling was strategically conducted in the primary distribution regions of wild *F. suspensa*, encompassing Henan, Shaanxi, and Shanxi provinces. 20 Qingqiao samples from the LP (n = 11), GZP (n = 5), and QM (n = 4), were collected by our team from August 10 to August 15, 2024, as detailed in **Table S1**. To ensure statistical robustness, three replicate samples were collected from each sampling site, yielding a total of 60 Qingqiao specimens. All collected samples were taxonomically authenticated as *F. suspensa* by Professor Wu Zhenhai, a renowned botanist from Northwest A&F University. Following collection, the samples were processed through room temperature drying until they reached constant weight for subsequent analysis.

Sample preparation and extraction

The Qingqiao samples were subjected to vacuum freeze-drying using a Scientz-100F lyophilizer. The lyophilized material was pulverized in a Retsch MM 400 grinder (30 Hz for 1.5 min) to obtain homogeneous powder. For extraction, 50 mg of the powdered samples was combined with 1200 μL of pre-chilled (-20°C) 70% methanol aqueous solution containing internal standards. The mixture underwent intermittent vortex extraction with 6 times of 30 s agitation followed by 5-minute intervals. Following phase separation by centrifugation at 12,000 rpm for 3 min, the supernatant was carefully collected and filtered through a 0.22 μm microporous membrane (SCAA-104, ANPEL, Shanghai, China). The resulting filtrate was immediately transferred to HPLC vials and maintained at 4°C until UPLC-MS/MS analysis[19].

UPLC-MS/MS analysis

UPLC Conditions

An UPLC-ESI-MS/MS system (UPLC, ExionLCTM AD) and Tandem mass spectrometry system were chosen as UPLC -MS/MS platform for extensive targeted metabolomics analysis of Qingqiao samples extracts. The UPLC conditions were as follows: column, Agilent SB-C18 (1.8 um, 2.1 mm*100 mm); (2) solvent A, pure water with 0.1% formic acid; solvent B, acetonitrile with 0.1%formic acid; (3) elution gradient, starting conditions of 95% A, 5% B; within 9 min, a linear gradient to 5% A, 95% B, and 5% A, 95% B for 1 min; 95% A, 5.0% B within 1.1 min and kept for 2.9 min; (4) flow velocity, 0.35 mL per minute; (5) column oven, 40℃; injection volume, 2 μL. The effluent was alternatively connected to an ESI-triple quadrupole-linear ion trap (OTRAP)-MMS.

ESI-O TRAP-MS/MS

The ESI source was configured with optimized parameters described previously [20]: ionization temperature at 500°C; spray voltage in dual polarity mode (+ 5500 V/-4500 V); gas parameters including nebulizing GSI (ion source gas I, 50 psi), auxiliary gas GSII (60 psi), and CUR (curtain gas, 25 psi); collision- activated dissociation operated in high-energy mode. The experiment utilized the multiple reaction monitoring (MRM) scanning mode on a triple quadrupole mass spectrometer, with nitrogen as

the collision gas (moderate flow rate intensity). To address the characteristics of metabolites across different elution periods, specific MRM channels were assigned to each detection window, and the declustering voltage and collision energy parameters for each ion pair were determined through system optimization.

Data processing

Raw data acquired in both positive and negative ionization modes underwent total ion current normalization to account for inter-sample intensity variation prior to chemometric analysis. Following normalization, the refined datasets were systematically processed using Analyst 1.6.3 software (AB Sciex) for advanced spectral deconvolution and baseline correction. Subsequent chromatographic analysis was performed with MultiaQuant™ 3.0.1 (Waters Corporation) for automated peak detection and integration through adaptive threshold algorithms and peak area normalization against internal standards. The final quantified results, expressed as normalized peak areas (arbitrary units), were proportional to the relative abundance of detected analytes.

Chemometric analysis

The loading plot, score plot, and variable importance in projection (VIP) values of each variable in the OPLS-DA model were calculated to represent their contribution to classification. The metabolites with VIP ≥ 1, Fold Change ≥ 2 or ≤ 0.5, and *p*-value < 0.05 were recognized as differential compounds. The metabolites were further tested at the univariate level using the Duncan's test to measure the significance of each metabolite, and the *p*-value of less than 0.05 was considered significant. The relationships between the main compounds in Qingqiao samples and metabolite profiles were characterized by Mantel tests. Random forest method and external validation tests were applied for validation and optimization of the origin-specific biomarkers screened. Pie chart plot, heatmap plot, volcano plots, and histogram were completed through Origin Pro 2025b software.

Results

Metabolic profiling identification of Qingqiao

UPLC-MS/MS was employed to investigate the metabolites in Qingqiao forexploring nutritional ingredients and functional components to trace the geographical origins. The typical base peak chromatograms of the six QC samples in both ESI- (**Fig. S1A**) and ESI + modes (**Fig. S1B**) were well overlapped, indicating excellent repeatability and reliability of the data. A comprehensive analysis indicated that 3052 distinct compounds were identified in LP, GZP, and QM, which were grouped into 13 known classes (**Table S2**). The metabolite profiling revealed distinct compositional features in Qingqiao, with primary metabolites comprising 283 lipids, 282 amino acids and derivatives, 89 organic acids, and 77 nucleotides and derivatives, potentially contributing to its nutritional profile (Fig. 1A). Secondary metabolites (1,938 compounds, 63.49%) dominated the metabolic repertoire, indicating that the

distribution and content of secondary metabolites directly affected the medicinal and functional value of Qingqiao. These bioactive components were systematically categorized into 8 major classes: terpenoids (574 compounds, 29.62%), phenolic acids (431 compounds, 22.24%), flavonoids (383 compounds, 19.76%), alkaloids (257 compounds, 13.26%), lignans and coumarins (229 compounds, 11.82%), quinones (38 compounds, 1.96%), tannins (16 compounds, 0.83%), and 63.49% (10 compounds, 0.52%).

Distinct metabolic profiles were observed among the GZP, QM, and LP samples, highlighting region-specific phytochemical patterns (Fig. 1B). LP samples were characterized by markedly elevated levels of organic acids, alkaloids, lignans, coumarins, flavonoids, quinones, as well as amino acids and their derivatives. These metabolites are known to confer various pharmacological properties, including antibacterial, antiviral, antipyretic, anticancer, hepatoprotective, antioxidant, and anti-aging activities. In contrast, QM samples accumulated higher quantities of lipids, steroids, terpenoids, and tannins, which are associated with broad-spectrum antimicrobial effects. Meanwhile, GZP samples were particularly enriched in phenolic acids, nucleotides, and their derivatives compounds recognized for significant anti-inflammatory, antibacterial, antioxidant, and hepatoprotective function.

Additionally, the chord diagram illustrated the correlations between Qingqiao samples from different regions and their metabolite profiles (Fig. 1C). The results demonstrated that primary metabolites, such as lipids, amino acids, and their derivatives, were closely associated with samples from GZP, LP, and QM. Notably, secondary metabolites exhibited even stronger correlations with these samples. As one of the most widely used traditional Chinese herbal supplements, the nutritional and functional quality of Qingqiao is largely determined by the composition and abundance of these primary and secondary metabolites.

These findings underscored the substantial regional disparities in Qingqiao quality, providing a metabolic basis for the geographical discrimination and efficacy-oriented evaluation of this important TCM material.

Identification of terpenoids

The terpenoid composition of Qingqiao samples was primarily characterized by sesquiterpenes, diterpenes, and triterpenoids. The major identified constituents included 2-hydroxyoleanolic acid, 2-hydroxyursolic acid, 2,3-dihydroxy-12-ursen-28-oic acid, jujubogenin, betulinic acid, ursolic acid, oleanolic acid, and camaldulenic acid (**Fig. S2A, Table S3**). These findings were consistent with those reported in previous studies [2].

Comparative analysis revealed distinct accumulation patterns among different samples. 2-hydroxyoleanolic acid showed preferential accumulation in LP samples, with particularly high levels observed in LP1, LP4, and LP6. Similarly, several other compounds including 8-epiloganic acid, loganic acid, 11β-hydroxy-8,15-isopimaradiene-3-one, d-linalool 3-(6"-malonylglucoside), and 16-hydroxyferruginol demonstrated elevated expression levels in LP samples, especially in LP1, LP2, LP3,

and LP4. In contrast, QM samples exhibited higher concentrations of 2-hydroxyursolic acid, 2,3-dihydroxy-12-ursen-28-oic acid, and jujubogenin, particularly in QM2, QM3, and QM4. Furthermore, QM samples showed significant enrichment of betulinic acid, madasiatic acid, 3-epiursolic acid, ursolic acid, isomangiferolic acid, and camaldulenic acid, with QM3 samples containing the highest overall content of these compounds. Cannabifolin C, cordianol B, leoheteronin A, and leoheteronin B were evidently enriched in GZP4 samples (**Fig. S2B**). These findings demonstrate distinct terpenoid accumulation patterns among LP, GZP and QM samples, suggesting potential differences in their biosynthetic regulation or environmental influences.

Identification of phenolic acids

Phenolic acids are the iconic and pharmacologically active components of Qingqiao, with high content and significant biological activity. The major compounds, presented at higher concentrations, included calceolarioside B, desrhamnosylacteoside, calceolarioside A, tetrahydrorengyoxide-glucose-sinapoyl, rengyoside A, 3-hydroxybenzoic acid, forsythiaside A, magnoloside A, and forsythoside I (**Table S4**).

Phenylethanoid glycosides represent the characteristic and principal bioactive constituents, notable for their abundance and pharmacological significance. The major phenolic acids identified include rengyoside A, forsythiaside A, forsythoside I, forsythenside B, forsythoside H, forsythiaside C, rengyol-glucose-sinapoyl, and forsythenside A. Among these, forsythiaside A and rengyoside A were present in higher concentrations, with forsythiaside A being the most abundant component in Qingqiao and serving as the index compound for quality control as stipulated by the Chinese Pharmacopoeia [21]. The highest content of forsythiaside A was detected in GZP samples, followed by QM and LP samples. Notably, samples QM4, GZP1, GZP2, and GZP3 exhibited the highest concentrations of forsythiaside A.

Specifically, 3-(3,4-dihydroxybenzyl)-4-(3,4-dimethoxybenzyl) dihydrofuran-2(3H)-one, coumalic acid, and 3-hydroxybenzoic acid were detected at higher concentrations in LP samples. In contrast, calceolarioside B, forsythiaside B, forsythiaside C, and magnoloside D were more abundant in QM samples. Forsythiaside A, forsythoside E, forsythoside F, forsythiaside B, forsythiaside J, forsythoside H, and calceolarioside A were predominantly identified in GZP samples (**Fig. S2C**), with particularly high levels observed in GZP1, GZP2, and GZP3 (**Fig. S2D**). Among the 20 samples analyzed, forsythiaside C and isoforsythoside A were most concentrated in LP5, while 3-hydroxybenzoic acid was elevated in LP1 and LP2. Rengyoside A showed higher abundance in LP5, LP6, LP7, LP8, and LP9 samples. Additionally, cornoside, calceolarioside B, and calceolarioside A were notably elevated in the GZP1 sample, and forsythiaside A was also prominently detected in GZP samples (**Fig. S2D**). Among 20 samples, forsythiaside C and isoforsythoside A were more concentrated in LP5 samples, 3-hydroxybenzoic acid was higher in LP1 and LP2 samples, rengyoside A was higher in LP5, LP6, LP7, LP8, and LP9 samples. Additionally, cornoside, calceolarioside B, and calceolarioside A were notably higher in the GZP1 sample, while forsythiaside A, magnoloside A, and forsythoside I were enriched in QM4, GZP1, and GZP2 samples (**Fig. S2D**).

Identification of flavonoids

Flavonoids, significant products of plant secondary metabolism, were predominantly found in Qingqiao samples and primarily consist of flavonols and flavanones (**Table S5**).

The flavonoids with higher concentrations comprised kushenol S, quercetin glycoside derivatives, rutin, isohyperoside, cynaroside, and hyperin. As illustrated in **Fig. S2E**, the flavonoids identified were predominantly accumulated in LP samples, with the exception of kushenol S, nepetin-7-glucoside, and isorhamnetin 3,4'-diglucoside. Notably, phellodensin F, quercetin-3-O-robinobioside, rutin, kaempferol-4'-O-glucoside, luteolin-7-O-glucoside, and hyperin were highly expressed. Among the LP samples, LP5, LP6, LP8, LP9, and LP10 exhibited significant flavonoid accumulation (**Fig. S2F**). Conversely, kushenol S, isoorientin-7-O-glucoside, and nepetin-7-glucoside were more abundant in QM and GZP samples. Specifically, QM4 and GZP1 exhibited higher levels of kushenol S, while QM3 and GZP2 had elevated concentrations of nepetin-7-glucoside.

Identification of lignans and coumarins

In our study, we identified several key lignans and coumarins as characteristic components in Qingqiao, including pinoresinol, epipinoresinol, matairesinol, sanshodiol, phillygenin, O-Feruloyl 4-hydroxycoumarin, forsythialan A, forsythialan B, arctigenin, and matairesinoside (**Table S6**). Notably, pinoresinol, epipinoresinol, matairesinol, sanshodiol, phillygenin, forsythialan A, and arctigenin were predominantly accumulated in LP samples, with particularly high concentrations observed in LP1 and LP2 samples (**Fig. S2G**). In contrast, machilin A and licarin A were significantly enriched in LP10 samples (**Fig. S2H**). Furthermore, 1,2-Dihydro-7-hydroxy-(4-hydroxy-3-methoxyphenyl)-3-(hydroxymethyl)-6-methoxy-(1S,2R)-2-naphthalenecarboxaldehyde and vitexdoin were found in higher quantities in QM samples, especially in QM1 and QM4. Additionally, 5-[(2R,3S,4S,5S)-5-(2H-1,3-benzodioxol-5-yl)-3,4-dimethyloxolan-2-yl]-2-methoxyphenol, isoshonanin, mahaleboside, and skimmin (7-Hydroxycoumarin-7-O-glucoside) were markedly enriched in GZP samples, particularly in GZP2, GZP3, and GZP4.

Identification of lipids

The lipid components were categorized into 6 distinct groups, comprising 134 free fatty acids (FFAs), 47 lysophosphatidylethanolamines (LPEs), 41 lysophosphatidylcholines (LPCs), 40 glycerol esters (GLs), 1 phosphatidylcholine (PCs), and 20 sphingolipids (**Table S7**). The predominant lipids in Qingqiao samples were lysoPC 18:1, cis-8,11,14,17-eicosatetraenoic acid, 9-thiastearic acid, lysoPC 14:0, α-linolenic acid, and lysoPC 16:1 (**Fig. S2I**).

In LP samples, the most abundant compounds were cis-8,11,14,17-eicosatetraenoic acid, lysoPC 16:1, 5S-HpEPE, 14,15-dehydrocrepenynic acid, and tianshic acid. Conversely, GZP samples exhibited higher concentrations of lysoPE 16:0, lysoPE 16:0 (2n isomer), and lysoPC 18:0 (**Fig. S2I**). The remaining compounds were predominantly accumulated in QM samples. Notably, QM3 and QM4 samples contained a diverse array of significant lipids, including octadeca-9,12,15-trienoic acid, gorlic acid, lysoPC 18:0, 2-aminooctadecane-1,5,17-triol, octadec-6-enoic acid, 11-octadecanoic acid, and beta-hydroxypalmitic acid (**Fig. S2J**).

Summary and Quality Evaluation

In conclusion, GZP samples, particularly GZP1, GZP2, and GZP3, demonstrated the highest accumulation of pharmacologically relevant phenolic acids, most notably forsythiaside A, which serves as the official quality marker. These samples also exhibited significant enrichment in specific terpenoids, flavonoids, lignans, coumarins, and lipids, indicative of superior medicinal quality. QM samples, especially QM3 and QM4, contained the highest overall levels of terpenoids, including betulinic acid, ursolic acid, and camaldulenic acid, along with a diverse lipid profile and select flavonoids, suggesting strong biosynthetic activity and high material quality. LP samples showed elevated accumulation of certain terpenoids, flavonoids, lignans, and coumarins, with LP1, LP2, LP5, and LP6 being notable across multiple metabolite categories. Nevertheless, LP samples generally contained lower levels of key phenolic acids compared to those from GZP and QM.

Consequently, the lines of GZP1, GZP2, GZP3, QM3, QM4, LP1, LP2, LP5, and LP6 showed strong potential for development into various functional products.

Comparison of Qingqiao from different geographical origins based on multivariate chemometrics

The assessment of overall differences and within-group variations among Qingqiao from different producing areas required advanced chemometric interpretation of metabolomics data.

As evidenced by HCA analysis with a Euclidean distance threshold of 0.35, the 20 Qingqiao samples formed three distinct dendrogram clusters corresponding to their geographical origins (LP, GZP, and QM), demonstrating remarkable inter-regional metabolic divergence on the primary and secondary metabolite contents (Fig. 2A). This phytochemical stratification was reinforced by PCA, revealing 69.01% cumulative explained variance (PC1 = 57.84%, PC2 = 11.17%) in the 2D-PCA plot, where tight intra-group clustering and clear inter-group separation were observed (Fig. 2B). The concordance between unsupervised HCA and PCA methodologies substantiated the hypothesis that geographical origin and cultivar characteristics synergistically govern metabolite biosynthesis patterns. In OPLS-DA model, with R^2X = 0.387 and *P*-value < 0.01, samples from LP, GZP, and QM were distinctly grouped, consistent with the HCA and PCA findings (Fig. 2C). The OPLS-DA results further emphasized that the origin of the samples significantly influenced the metabolite composition and content of Qingqiao. Permutation tests (200 permutations) yielded Q^2 = 0.996 and R^2Y = 0.987, confirming that the model was not overfitted and demonstrating its strong predictive and fitting capabilities. This robustly validated the model's stability and reliability, making it highly suitable for distinguishing Qingqiao samples from different origins (Fig. 2D). These results indicated that the Qingqiao samples could be distinguished based on the untargeted metabolomic fingerprints data.

Screening for differential metabolites in Qingqiao from different geographical origins

To comprehensively investigate the compound alterations in Qingqiao samples induced by different geographical origins, an OPLS-DA model was established through pairwise comparisons among 3 geographical sources: GZP-vs-LP, GZP-vs-QM, and QM-vs-LP. The VIP value served as the key indicator for screening differential metabolites, with higher VIP values indicating a greater influence of chemical markers on group differences. Differential metabolites of Qingqiao samples from different geographical sources were identified using the criteria VIP > 1 and fold changes (FC ≥ 2 or ≤ 0.5).

A total of 1136 differential metabolites were identified across the 3 geographical origins. The volcano plot of these differential metabolites was presented in Fig. 3A. Specifically, 611 significant differential metabolites were identified between GZP and LP, with 248 up-regulated and 363 down-regulated (**Table S8**). Between GZP and QM, 573 differential metabolites were significant, comprising 225 up-regulated and 348 down-regulated (**Table S9**). For QM and LP, 327 metabolites were significantly up-regulated, and 287 were down-regulated (**Table S10**).

Differential primary metabolites among GZP-vs-LP, GZP-vs-QM, and QM-vs-LP

Comprehensive metabolomic profiling identified significant intergroup disparities in primary metabolites, with lipids and amino acids and derivatives alterations constituting the principal differentiating factors (**Fig. S3A**).

A comparative analysis of amino acid and derivatives revealed significant differential abundances across 3 experimental groups. Specifically, the GZP-vs-LP comparison exhibited 50 differentially abundant lipids (20 upregulated and 30 downregulated), followed by GZP-vs-QM with 37 differentials (19 upregulated, 18 downregulated), and QM-vs-LP showing 35 significant changes (10 upregulated, 25 downregulated). This tiered pattern of metabolic variation demonstrated group-specific regulatory patterns in amino acid metabolism, with the GZP-vs-LP pair displaying the most pronounced differential expression profile. The Gly-Phe-His, 4,5-Dehydro-L-leucine, cycloleucine, Lys-Lys-Val, pyroglutamic acid, and L-Homoglutamic acid exhibited pronounced upregulation in GZP samples compared to LP, and Lys-Lys-Val, L-Homoglutamic acid, His-Lys-Ser, and Gly-Phe-His were notable upregulated in GZP than that in QM samples. LP samples showed higher abundance of Gly-Ala-Asn, L-Asparagine, L-Tyrosine, Phe-arg, Ser-Ser-Glu, Met-Pro-Gln, and His-Ala-Met than that in GZP samples, and exhibited enhanced levels of aromatic amino acids and their derivatives, including DL-tryptophan and its L-isomer, L-tyrosine, 3-hydroxy-L-phenylalanine, and cyclo-(Gly-Phe) compared to QM samples. QM samples exhibited significantly higher levels of Gly-Ala-Asn, L-Asparagine, Met-Pro-Gln, Phe-Arg, and Ser-Ser-Glu compared to GZP, and revealed higher content of arginine methyl ester, pyroglutamic acid, N-monomethyl-L-arginine, 4,5-Dehydro-L-leucine, and cycloleucine were compared to LP.

Comparative analysis revealed that distinct differential expression patterns of lipids across the 3 groups (GZP-vs-LP, GZP-vs-QM, and QM-vs-LP) were GLs, LPCs, LPEs, and FFAs. The GZP-vs-LP comparison showed 46 significantly regulated lipids, with 11 upregulated and 35 downregulated. 3 lipids, 13((2R,3S)-3-Pentyloxiranyl)-8Z-tridecenoic acid (9.76-fold), 10-Hydroxydecanoic acid glucoside (3.20-

fold), and 1-O-sulfo 9,10-dihydroxyoctadecanoate (2.18-fold), demonstrated marked upregulation in GZP. Conversely, 6 lipid species, 2-hydroxyhexadecanoic acid, cis-8,11,14,17-Eicosatetraenoic acid, 14,15-dehydrocrepenynic acid, LysoPC 14:0, LysoPC 20:3, and 2-hydroxypentadecanoic acid, showed significant downregulation in GZP. More pronounced regulation was observed in the GZP-vs-QM comparison, exhibiting 101 differentially expressed lipids, including 5 upregulated and 96 downregulated. Particularly noteworthy were the marked increases in LysoPE 20:2 (2.90-fold), LysoPE 20:2(2n isomer) (3.00-fold), and 10-Hydroxydecanoic acid glucoside (2.04-fold) in the GZP compared to QM. Conversely, QM exhibited significant accumulation of several lipid species including DGMG 18:3, DGMG 18:3(2n isomer), 3-Hydroxy-palmitic acid methyl ester, 2-hydroxypentadecanoic acid, 2-Hydroxyhexadecanoic acid, LysoPC 20:3, and LysoPC 19:3. Conversely, the QM-vs-LP comparison demonstrated a contrasting pattern with 95 regulated lipids, comprising 83 upregulated and 12 downregulated. The samples in LP exhibited significant elevations in specific lysophospholipid species, particularly LysoPC 12:0 (2.36-fold) and LysoPC 14:0 (2.54-fold), along with notable accumulation of the oxidized fatty acid derivative 9-oxooctadeca-10,12-dienoic acid (3.71-fold). In contrast, the QM cohort demonstrated characteristic upregulation of structural lipid isomers, including LysoPE 18:3(2n isomer) (4.78-fold) and LysoPE 18:1(2n isomer) (2.61-fold). This group also showed elevated levels of several distinctive lipid components, including digalactosylmonoacylglycerol derivatives (DGMG 18:3 and its 2n isomer), LysoPC 17:1, along with modified fatty acids Beta-hydroxypalmitic acid and 2-hydroxyhexadecanoic acid.

These findings revealed group-specific lipids modulation patterns, particularly noting the predominant downregulation trend in GZP relative to both LP and QM groups. The differential expression patterns of GLs, LPCs and modified FFAs acid derivatives might indicate group-specific membrane remodeling activities, while the observed peptide variations suggested potential differences in protein turnover rates or signaling pathway activation.

Differential secondary metabolites among GZP-vs-LP, GZP-vs-QM, and QM-vs-LP

Significant differences in secondary metabolite profiles were identified across the 3 pairwise comparisons (GZP-vs-LP, GZP-vs-QM, and QM-vs-LP), encompassing 5 principal phytochemical classes: terpenoids, flavonoids, phenolic acids, alkaloids, with lignans and coumarins.

The terpenoid composition of Qingqiao samples exhibited significant variations across different geographical origins, with triterpenes, ditepenoids, monoterpenoids, and sesquiterpenoids constituting the predominant chemical classes (**Fig. S3B**). Detailed comparisons revealed distinct metabolic patterns among the analyzed groups. In the comparison of GZP-vs-LP, GZP samples demonstrated substantial upregulation of specific terpenoid derivatives, including 2 monoterpenoids (D-linalool 3-glucoside, mudanpioside A), 3 triterpenes (rosa aglycone glucoside, oleanderic acid, 2α,3α-Dihydroxyolean-12-en-28-oic acid), 2 sesquiterpenoids (staphylionoside A, blumenol B malonyl glucoside). Conversely, LP samples showed preferential accumulation of oxidized diterpenoids, particularly 16-hydroxyferruginol

and 11β-hydroxy-8,15-isopimaradiene-3-one, along with increased levels of steviol derivatives (steviol, isosteviol) and kolavenic acid. In the comparison of GZP-vs-QM, the triterpenoid fraction in GZP samples was characterized by enhanced production of morolic acid (2.12-fold), mudanpioside A (2.47-fold) and corosolic acid methyl ester (2.06-fold). Blumenol glycosylation products, including blumenol B and C malonyl glucosides, showed 6.96- and 7.21-fold increases respectively. QM specimens conversely exhibited elevated diterpenoid metabolism, with 11β-hydroxy-8,15-isopimaradiene-3-one, isopimaradieneone and 16-hydroxyferruginol showing marked increases; and exhibited elevated triterpenes levels, with displaying significant increases in betulonic acid, rosa aglycone glucoside, and hyptadienic acid. In the comparison of QM-vs-LP, QM samples exhibited marked upregulation of 6 triterpenes (2,19-dihydroxy-3-oxours-12-en-28-oic acid, 3-O-feruloyleuscaphic acid, norarjunolic acid, betulonic acid, 2α,3α-dihydroxyolean-12-en-28-oic acid, rosa aglycone glucoside) and 3 ditepenoids (isopimaric acid, triptonoterpene, isopimaradieneone). Conversely, LP samples demonstrated significant accumulation of 3 ditepenoids (2-oxokovalenic acid, andrographolide, jolkinolide D) and 3 monoterpenoids (oleoside 11-methyl ester, β-ionone, and geniposidic acid).

Phenolic acids, serving as the principal bioactive markers in Qingqiao, exhibited substantial quantitative and qualitative variations across different geographical origins (**Fig. S3B**). In the comparative analysis between GZP and LP specimens, GZP demonstrated selective enrichment of multiple phenolic derivatives including decaffeoylverbascoside, dracunculifoside B, 3-[4-(β-D-glucopyranosyloxy)phenyl]propionic acid, dihydrocoumaroyl hexoside, verbascoside, and the structurally complex 2-(3,4-dihydroxyphenyl)-ethyl-O-β-D-xylopranosyl-(1→6)-2-O-trans-caffeoyl-β-D-glucopyranoside-5'-caffeoyl-O-rha. Notably, forsythosides (Forsythoside E, F, and B) were particularly abundant in GZP samples. LP samples conversely exhibited preferential accumulation of simpler phenolic forms, dominated by 3-hydroxybenzoic acid, (4E,6E)-1,7-bis(4-hydroxy-3-methoxyphenyl)-hepta-4,6-dien-3-one, and 3-O-feruloylquinic acid-O-glucoside. When comparing GZP with QM, GZP specimens showed enhanced biosynthesis of chlorogenic acid isomers (including crypto-, neo-, and standard forms) along with various acylated glucose derivatives. Particularly distinctive were elevated levels of 2-β-D-glucopyranosyloxy-4-methoxy-cis-cinnamic acid and multiple O-acylated glucoses (1-O-caffeoyl, 6-O-p-coumaroyl, and 1-O-feruloyl forms). Conversely, reductions were observed in simple phenolic acids (3-hydroxybenzoic acid) and complex conjugates such as forsythenside A, rosmarinic acid-3'-O-glucoside, cannabidivarin, and various tartaric/caffeoyl derivatives. In the QM-vs-LP comparison, QM samples displayed preferential accumulation of sophisticated phenolic esters, including 2 sinapoylated rengyol derivatives (rengyoxide-glucose-sinapoyl and rengyol-glucose-sinapoyl), dracunculifoside B, forsythoside B, 3-O-feruloylquinic acid-O-glucoside, and the complex 2-(3,4-dihydroxyphenyl)-ethyl-O-β-D-xylopranosyl-(1→6)-2-O-trans-caffeoyl-β-D-glucopyranoside-5'-caffeoyl-O-rha. LP samples displayed dominance in chlorogenic acid isomers (neochlorogenic and cryptochlorogenic acids) along with 3-O-p-coumaroylquinic acid and feruloylmalic acid.

As prominent secondary metabolites in plants, the differemtial flavonoids in Qingqiao predominantly existed as flavonols and flavones (Fig. 3A). Our comparative metabolomic analysis revealed distinct accumulation patterns among different geographical origins. Within GZP-vs-LP comprasion, the

flavonoid profiles demonstrated distinct patterns: isoetin-7-O-glucoside-2'-O-arabinoside, 2'3'4'-trimethylcalythropsin 2-O-glucoside, delphinidin-3-O-rutinoside-7-O-glucoside, icariin, isorhamnetin-3-O-glucuronide-7-O-rhamnoside, and tamarixin were preferentially accumulated in GZP samples; LP specimens conversely enriched pinobanksin, naringenin, 5-hydroxy-7,8,3',4'-tetramethoxyflavone glucoside, isorhamnetin feruloyl glucoside, and centaurein. Within GZP-vs-QM comprasion, Notable upregulated compounds in GZP samples included 2',3',4'-Trimethylcalythropsin 2-O-glucoside, petunidin-3-O-(6"-O-p-coumaroyl)glucoside, tamarixetin-3-O-glucoside-7-O-rhamnoside, 3'-methoxyquercetin-3-O-L-rhamnosyl(1→2)-glucopyranoside, de-lphinidin-3-O-rutinoside-7-O-gluc–oside, and 3,3',4',5',7-pentahydroxyflavone. Conversely, GZP samples displayed substantial downregulation of methoxylated derivatives including 5-hydroxy-7,8,2',3'-tetramethoxyflavone 5-glucoside and 5-hydroxy-7,8,3',4'-tetramethoxyflavone glucoside, as well as complex conjugates like isorhamnetin feruloyl glucoside and centaurein. Within QM-vs-LP comprasion, the flavonoid spectrum showed QM-specific upregulation of 6 methoxylated and acylated derivatives: 5-hydroxy-7,8,3',4'-tetramethoxyflavone glucoside, dihydroxy-dimethoxyfla-vone-7-O-glucoside, 5,2'-dihydroxy-7,8-dimethoxyflavone glucoside, its aglycone form, centaurein, and isorhamnetin feruloyl glucoside. LP samples preferentially accumulated kaempferol derivatives (3-O-robinobioside and 4'-O-glucoside), luteolin glycosides (7-O-glucoside and 7-O-rutinoside), along with 2',4,4',6'-tetrahydroxychalcone and two flavanones (3,5,7-trihydroxyflavanone and 5,7,4'-trihydroxyflavanone).

The lignans were the main medicinal and functional ingredients in Qingqiao, with excellent antioxidant, hepatoprotective and hypolipidemic activities. Comparative analysis between revealed distinct phytochemical profiles in lignan and coumarin derivatives (Fig. 3B). Comparative phytochemical profiling revealed distinct lignan and coumarin derivative accumulation patterns among 3 comparisons. In the GZP-vs-LP comparison, GZP specimens displayed preferential accumulation of esculin (6,7-dihydroxycoumarin-6-O-glucoside), syringaresinol-4-O-β-D-glucopyranoside, and tortoside A. Conversely, LP samples exhibited significantly elevated concentrations of balanophonin, phillygenin, 5-methoxydehydrodiconiferyl alcohol 4-O-β-D-glucopyranoside, forsythialan B, arctigenin, matairesinoside, and chestnutlignansoide. When comparing GZP with QM, the GZP lignan fraction demonstrated substantially higher levels of sesamin, bursehernin, and isolariciresinol. Concurrently, GZP coumarin derivatives exhibited significant increases in isoscopolin, esculin (6,7-dihydroxycoumarin-6-O-glucoside), 7-hydroxy-8-methoxychromanon, and 6-hydroxymellein. Notable reductions were observed in icariside E5, matairesinoside, chestnutlignansoide, isoimperatorin, scopolin, and forsythialan B concentrations. The QM-vs-LP comparison revealed QM specimens with elevated levels of matairesinoside, N-sinapoylhydroxycoumarin, ostruthin, and chestnutlignanoside, suggesting a unique lignan biosynthetic profile. LP samples conversely demonstrated preferential accumulation of syringyl-type lignan derivatives, particularly dehydrodiconiferyl alcohol-γ'-O-glucoside and saikolignanoside D, along with symplocosin, (7'S,8R,8'R)-3,5'-dimethoxy-3',4,9'-trihydroxy-7',9-epoxy-8,8'-lignan, demethyl-anhydrosecoisolariciresinol, and the coumarin derivative isoscopolin.

Alkaloids are an extremely important class of functional components in Qingqiao, demonstrating significant physiological activities, among which alkaloids, plumeranes, and pyridine alkaloids exhibited

marked variations across different production regions (Fig. 3C). The GZP-vs-LP comparative analysis showed pronounced metabolic differentiation, with GZP specimens exhibiting preferential accumulation of specialized nitrogen-containing metabolites. These included o-aminobenzyl-β-D-glucopyranoside, pyridine-4-formyl-O-β-D-glucopyranoside, N-glucosylpico-

-linate, and heliocurassavicine derivatives. Conversely, LP samples demonstrated alkaloid dominance, particularly anatabine, stachydrine, and indolic acids (indole-5-carboxylic acid and indole-3-carboxylic acid). The GZP-vs-QM analysis identified unique metabolic patterns, and the GZP samples significant enrichment of o-aminobenzyl-β-D-glucopyranoside, indole-3-carboxaldehyde, and betanin. QM specimens presented a distinct chemotype characterized by elevated levels of anatabine, acetylcholine, betaine), and stachydrine, along with indole-5/3-carboxylic acids. The GZP-vs-LP analysis revealed that the LP samples uniquely accumulated 5 indole derivatives: pyran[3,4-b]indole-2-ketone, 3-indoleacrylic acid, naphthisoxazol A, along with indole-5/3-carboxylic acids. The QM samples displayed distinct accumulation of 3 (3S,3'S)-N,N'-((1R,2R)-5-((3,4,5-trihydroxy-6-methyltetrahydro-2H-pyran-2-yl)oxy)cyclohex-3-ene-1,2-diyl)bis

(2-methyl-2,3,4,9-tetrahydro-1H-pyrido[3,4-b]indole-3-carboxamide), 10-hydroxymethyllycaconi-

-tine, acetylcholine, and stachydrine.

Furthermore, we performed correlation analysis on all 7 major categories of metabolites identified in Qingqiao (Fig. 3D). The results revealed a strong correlation between flavonoids and lignans and coumarins, along with a significant positive correlation between alkaloids and amino acids and derivatives ($P < 0.05$). These observations could be explained by the biosynthetic pathways in plants, where lignans and coumarins compounds are synthesized downstream of flavonoids, while alkaloids are derived from amino acids and derivatives precursors. These findings offered novel insights into the co-regulatory mechanisms governing the coordinated biosynthesis and accumulation of flavonoids and lignans and coumarins, as well as amino acids and derivatives and alkaloids, across diverse geographical origins.

KEGG annotation and enrichment analysis

Differential metabolites were identified and enriched from three comparison groups, followed by systematic classification into specific KEGG pathways (Fig. 4). In the GZP-vs-LP, the KEGG bubble plot revealed that the biosynthesis of caffeic acid derivatives, biosynthesis of caffeoylquinic acid derivatives, biosynthesis of coumarins III, biosynthesis of quercetin aglycones I/II, and flavonoid biosynthesis were significantly enriched ($P < 0.05$) (Fig. 4A). For the comparison between GZP and QM, nicotinate and nicotinamide metabolism, biosynthesis of phenylacetic acid derivatives, diterpenoid biosynthesis, and glycerolipid metabolism were significantly enriched ($P < 0.05$) (Fig. 4B). When comparing QM and LP, the biosynthesis of caffeoylquinic acid derivatives, biosynthesis of coumarins III, flavonoid biosynthesis, diterpenoid biosynthesis, and glycerolipid metabolism pathways exhibited significant enrichment ($P <$

0.05) (Fig. 4C). These findings indicated that the distinct metabolite profiles observed in GZP, QM, and LP might be correlated with their different geographical origins.

Key significantly differential metabolites

A Venn diagram was applied to pinpoint unique metabolites in Qingqiao samples from different origins, revealing 38 overlapping differential metabolites as potential origin-specific biomarkers (Fig. 4D and **Table S11**). These included 9 flavonoids, 8 lipids, 5 phenolic acids, 4 alkaloids, 3 lignans and coumarins, 3 terpenoids, 2 organic acids, and one each of quinones, nucleotides and derivatives, amino acids and derivatives, and others. The relationship and relative content of these 38 metabolites were visualized in a heat map (Fig. 5A), which demonstrated that Group I metabolites were significantly higher in GZP samples, Group II in QM samples, and Group III in LP samples. This suggested that environmental factors such as temperature and precipitation may play a crucial role in the variation of Qingqiao metabolites (Fig. 5B).

Based on the content, fold change and *p*-value, 14 compounds were identified as candidate differential metabolites from the 38 common metabolites detected in GZP-vs-LP, GZP-vs-QM, and QM-vs-LP comparisons (**Table S12**). These metabolites were considered characteristic markers for distinguishing Qingqiao samples of different origins. To further investigate the variation patterns of these 14 candidate differential metabolites, a cluster heat map was constructed using their peak areas as variables. The heat map analysis revealed distinct metabolite profiles among the samples, with similar abundance patterns observed within samples of the same origin. Notably, significant differences were evident among LP, QM, and GZP samples, with QM samples showing markedly higher levels of most metabolites compared to LP and GZP samples (Fig. 6B). The 14 candidate differential metabolites were categorized into 3 distinct groups based on their accumulation patterns. The heat map analysis further demonstrated that Qingqiao samples were clearly separated into 3 major clusters corresponding to their origins (LP, QM, GZP), consistent with the HCA and PCA results (Fig. 4A and Fig. 4B). Group I metabolites, including 2-Glucosyl-glucosyloxy-2-phenylacetic acid amide and 4-(2-hydroxyvinyl) benzene-1,2-diol Glucoside, exhibited higher accumulation in GZP samples. Group II, comprising rosa aglycone glucoside, 2-(3,4-dihydroxyphenyl)-ethyl-O-p-D-xy-lopransy-(1-6-2-O-tras-caffeoyl-BD-glucopyranoside)-5'caffe-

-oyl-O-rha, matairesinoside, centaurein, 5-hydroxy-7,8,3',4'-tetramethoxyflavone glucoside, 2-hydroxypentadecanoic acid, 3-O-Feruloylquinic acid-O-glucoside, chestnutlignansoide, 2-Hydroxyhexadecanoic acid, and isorhamnetin feruloyl glucoside, exhibited predominant accumulation in QM samples. Group III metabolites, p-Coumaroylmalic acid and xylosyl phellodendroside, were predominantly expressed in LP samples. These findings indicated that the differential accumulation of specific metabolites played a crucial role in cultivar differentiation and could serve as reliable chemical markers for origin identification of Qingqiao samples.

Histogram analysis effectively visualized the content variations of 14 differential metabolites among Qingqiao samples from different geographical origins (Fig. 6), corroborating the patterns observed in the

heatmap analysis (Fig. 5C). Quantitative assessments revealed statistically significant differences in the concentrations of specific metabolites across sampling regions. Notably, samples from QM contained significantly higher levels of 2-hydroxyhexadecanoic acid, 5-hydroxy-7,8,3',4'-tetramethoxyflavone glucoside, and 2-hydroxypentadecanoic acid compared to those from LP and GZP ($p < 0.05$). In contrast, GZP samples exhibited markedly elevated concentrations of 2-glucosyl-glucosyloxy-2-phenylacetic acid amide and 4-(2-hydroxyvinyl)benzene-1,2-diol glucoside relative to both LP and QM samples. Furthermore, LP samples were characterized by significantly greater accumulation of p-coumaroylmalic acid and xylosyl phellodendroside compared to those from QM and GZP samples.

Matairesinoside and an isolariciresinol derivative (Chestnutlignanoside), both belonging to the lignan class, represented the characteristic and major bioactive constituents in Qingqiao. These compounds exhibited notable antiviral, antibacterial, and anti-inflammatory activities. Key phenolic acid components such as 3-O-feruloylquinic acid-O-glucoside and p-coumaroylmalic acid were also present, and their concentrations served as critical indicators for assessing the antioxidant capacity of Qingqiao. Rosa aglycone glucoside, a triterpenoid saponin, demonstrated diverse pharmacological effects, including antioxidant, anti-ischemic, and anti-hypoxic injury activities. Flavonol compounds, namely isorhamnetin feruloyl glucoside, xylosyl phellodendroside, and centaurein, contributed to a broad spectrum of biological functions such as antioxidant, anti-inflammatory, and anti-tumor actions. Notably, xylosyl phellodendroside was a structurally unique flavonoid glycoside, while 2-glucosyl-glucosyloxy-2-phenylacetic acid amide represented a distinct amide glycoside. The presence and abundance of these rare compounds could serve as origin-specific biomarkers, facilitating both authentication of Qingqiao medicinal materials and traceability of their geographical origins, thereby providing a scientific basis for quality evaluation across different producing regions.

To evaluate the identification and prediction ability of 14 origin-specific biomarkers, OPLS-DA analysis was performed on all Qingqiao samples from 3 geographical sources based on the 14 identified differential metabolites, and then a new OPLS-DA models were constructed. The OPLS-DA score plot was described in **Fig. S4A**. Chemometric analysis revealed significant differences among Qingqiao samples from 3 different regions, allowing their classification into 3 distinct categories. The new OPLS-DA model exhibited superior fitting ability and comparable predictive ability relative to the original model, as evidenced by the evaluation parameters $R^2X = 0.788$, $R^2Y = 0.977$, and $Q^2 = 0.966$. To ascertain whether the new OPLS-DA model was over-fitted, 200 permutation tests were performed. The results showed that the R^2 and Q^2 values of all 200 permutations on the left side were lower than the original points on the right side, strongly indicating the new model's effectiveness (**Fig. S4B**). Furthermore, among the 14 origin-specific biomarkers screened, all except 4-(2-hydroxyvinyl) benzene-1,2-diol Glucoside had VIP values greater than 1 (**Fig. S4C**).

To further validate the practicality of the newly established OPLS-DA model, comprehensive external validation tests were conducted. Specifically, two-thirds of the Qingqiao samples (n = 40) from 3 distinct regions were designated as unknown samples for blind testing. These samples were subsequently processed as non-classified cases within the newly developed OPLS-DA model to predict their potential

geographical origins. The external validation results demonstrated exceptional model performance, with all blind testing samples being accurately classified, achieving an overall recognition rate of 93.33%. This rigorous external validation process substantiated the model's robust discriminative capability. Consequently, the 14 origin-specific biomarkers identified through this study exhibited superior accuracy and specificity in characterizing Qingqiao samples, confirming their reliability as distinctive indicators for geographical origin determination.

ROC curves of 14 key differential metabolites

The ROC curve has been widely utilized as a robust analytical tool for assessing the clinical performance of diagnostic and prognostic models. In recent years, the application of ROC analysis has demonstrated significant utility in plant science research for identifying reliable biomarkers and effectively discriminating between different sample groups. **Fig. S5** displayed the area under the curve (AUC) estimated for 14 key differential metabolites. The AUC represented the probability that the key differential metabolites could accurately distinguish between two randomly selected samples. The AUC for all combined differential metabolites was 1, and the AUCs for the 14 individual key differential metabolites were all close to 1. Therefore, these 14 key differential metabolites were potential biomarkers that could be used to differentiate Qingqiao samples from the QM, GZP, and LP regions.

Discussion

Metabolites identified in Qingqiao

The quality of Chinese medicinal materials is intrinsically linked to variations in nutritional ingredients and functional components [22]. In this study, UPLC-MS/MS analysis was conducted on Qingqiao samples collected from 3 distinct geographical origins (LP, QM, and GZP) to determine the nutritional and functional components. Chemical analysis of the Qingqiao samples revealed 3,052 distinct compounds, including 1,114 primary metabolites and 1,938 secondary metabolites. Primary metabolites are particularly significant as they represent key determinants of plant nutritional quality [23].

282 amino acids were identified in Qingqiao, including those essential for human health that were reportedly synthesized exclusively by plants [24]. The sample also contained considerable quantities of non-essential amino acids, such as L-alloisoleucine, L-aspartic acid, L-glutamic acid, L-serine, and L-tyrosine, along with dipeptides and tripeptides. As fundamental structural units, amino acids played a critical role in metabolic pathways that generate diverse secondary metabolites, including phenolic compounds, flavonoids, and alkaloids. They also mediated key interactions between environmental factors and human health through direct and indirect mechanisms [25]. Lipids, another major class of primary metabolites, performed various physiological functions such as membrane formation, energy storage, and synthesis of signaling molecules [26, 27]. However, lipid composition in Qingqiao has been poorly characterized. This study identified 283 lipids, thereby addressing a significant knowledge gap. Notably, lipids were essential precursors for terpenoid biosynthesis. The predominant lipids in Qingqiao were free fatty acids, lysophosphatidylethanolamines (LPEs), lysophosphatidylcholines (LPCs), and

glycolipids (GLs), which contributed to its edible properties and nutritional value. These lipids have been associated with cardioprotective effects and inhibition of cardiovascular diseases, potentially compensating for deficiencies in other lipid classes [28]. Terpenoids were the most abundant class of compounds in Qingqiao [29]. In this study, 574 terpenoids were detected, comprising 172 triterpenoids, 132 diterpenes, 132 monoterpenes, and 110 sesquiterpenes. The major terpenoid constituents identified were 2-hydroxyoleanolic acid, 2-hydroxyursolic acid, 2,3-dihydroxy-12-ursen-28-oic acid, jujubogenin, betulinic acid, ursolic acid, oleanolic acid, and camaldulenic acid. Compounds such as oleanolic acid and ursolic acid were known for their antimicrobial activity and other health-promoting properties [30, 31]. Additionally, 12 triterpenoid saponins were prominent bioactive compounds in Qingqiao, demonstrating significant cardioprotective, antitumor, and hepatoprotective effects [32]. 431 phenolic acids were identified in Qingqiao, primarily derivatives with benzoic acid, phenylacetic acid, or cinnamic acid as core structures, mainly containing components such as caffeic acid, ferulic acid, syringic acid, gallic acid, and chlorogenic acid. The major phenylethanoid glycosides identified comprised rengyoside A, forsythiaside A, forsythoside I, forsythenside B, forsythoside H, forsythiaside C, rengyol-glucose-sinapoyl, and forsythenside A. These compounds demonstrated biological activities extending beyond antioxidant effects [33]. Notably, forsythiaside A and rengyoside A were present at comparatively higher concentrations and were found to modulate multiple signaling pathways, thereby exerting anti-inflammatory effects, and these properties suggested their potential as complementary anti-inflammatory agents. 383 flavonoids were identified in Qingqiao, including 141 flavonol derivatives (e.g., kaempferol and quercetin derivatives), 104 flavones (e.g., luteolin glycosides and nepetarioside), and 44 dihydroflavones (e.g., kushenol S and naringenin). Flavonol derivatives, along with flavones and dihydroflavonols, possessed diverse biological activities that were beneficial to human health [34]. Additionally, 257 alkaloids, 229 lignans and coumarins, as well as 38 quinones and 16 tannins 16, were identified in Qingqiao. Notably, despite distinct visual and morphological differences among Qingqiao samples from 3 geographical regions, their metabolic profiles were remarkably consistent.

These findings significantly advanced the current understanding of Qingqiao's metabolite composition and offered valuable insights for identifying bioactive compounds with potential nutritional and medicinal applications.

Metabolite differences between different geographic origins

Each region produced Qingqiao with distinct metabolic strengths: LP samples exceled in amino acids, lignans and coumarins, and alkaloids; GZP samples werer rich in phenolic acids; and QM samples stood out for its high lipid and terpenoid content. These region-specific phytochemical profiles reflected environmental adaptations,and defined the unique functional and nutritional qualities of Qingqiao from different origins.

Based on the HCA and PCA of metabolite contents in Qingqiao sanples from 3 different geographical origins, the metabolite contents were highly correlated within the same region, while significant

differences were observed among GZP, LP, and QM, indicating that metabolite accumulation was influenced by environmental factors. This phenomenon has been reported in daylily, *Zanthoxylum bungeanum* [35], *Astragali radix* [15], and jujube [14]. To quantitatively assess environmental impacts, we conducted Mantel tests comparing metabolite distance matrices with matrices of temperature, precipitation, and light differences between sampling sites (**Fig. S6**). The significant correlations observed for phenolic acids (AMAT: r = 0.311, p = 0.016; AAS: r = 0.405, p = 0.001), lignans and coumarins (AMIT: r = 0.285, p = 0.024), and lipids (AMT: r = 0.256, p = 0.012; MTWQ: r = 0.337, p = 0.007) supported the hypothesis that climatic factors contributed to the observed geographic variation in metabolites.

1,136 differential metabolites were identified among GZP, LP, and QM, with the most significant differences observed in amino acids and their derivatives, lipids, flavonoids, phenolic acids, lignans, alkaloids, and terpenoids, which were enriched in biosynthesis of caffeic acid derivatives, biosynthesis of caffeoylquinic acid derivatives, biosynthesis of coumarins III, biosynthesis of quercetin aglycones I/II, and flavonoid biosynthesis, nicotinate and nicotinamide metabolism, biosynthesis of phenylacetic acid derivatives, diterpenoid biosynthesis, and glycerolipid metabolism. The differentially expressed KEGG pathways revealed intricate metabolic networks in which phenylacetic acid derivatives serve as key precursors for caffeic acid derivative biosynthesis. These compounds subsequently underwent esterification to form caffeoylquinic acids or are channeled into coumarin III biosynthesis through hydroxylation and cyclization reactions. Notably, flavonoids, which shared early biosynthetic precursors (particularly p-coumaroyl-CoA) with caffeic acid derivatives through the phenylpropanoid pathway, demonstrate significant metabolic cross-talk between these branches. Meanwhile, diterpenoid biosynthesis, originating from either the methylerythritol phosphate (MEP) pathways, operated independently but might indirectly modulate phenylpropanoid metabolic flux through shared carbon allocation mechanisms. Additionally, glycerolipid metabolism, while primarily dedicated to membrane biosynthesis, intersects with phenylpropanoid metabolism by supplying acyl-CoA esters that could serve as modifying groups for flavonoid or coumarin structures.

The differences in metabolites between GZP and LP revealed significant variations in 50 amino acids and their derivatives. Some amino acids related to food flavor, such as L-asparagine and glycine-alanine-asparagine, were particularly abundant in LP [36]. The results indicated that while the lipid profiles of Qingqiao from different geographical origins were similar, their content varied significantly. The differential lipids identified in GZP, LP, and QM were predominantly GLs, LPCs, and free FFAs. Notably, FAAs within these lipids might function as signaling molecules that differentially regulated multiple cellular processes and physiological functions depending on their specific carbon chain lengths. Furthermore, LPCs emerged as critical signaling mediators demonstrating versatile biological roles, particularly in the modulation of cellular inflammatory responses through distinct molecular mechanisms [28]. Metabolic classification of the differential metabolites revealed that QM exhibited substantially elevated lipids content compared to GZP and LP, highlighting its superior nutritional profile as a functional food. Comparative metabolic pathway analysis of GZP vs QM and QM vs LP groups identified significant enrichment in the glycerolipid metabolism pathway ($P < 0.05$), suggesting this pathway might play a crucial role in mediating the observed lipid composition differences.

Flavonoids protect plants from environmental stress and provide significant health benefits [37]. LP exhibited high abundances of flavones, flavonols, and dihydroflavones. Pathway enrichment analysis confirmed that flavonoid biosynthesis was significantly altered ($P < 0.05$) in both GZP-vs-LP and QM-vs-LP comparisons, highlighting notable regional disparities in flavonoid accumulation. Additionally, the composition and content of phenolic acids varied greatly across different geographical regions. Temperature and other environmental factors could influence the phenolic acid content in plants [38]. Most phenolic acids were upregulated in GZP relative to LP and QM. These compounds, prevalent in fruits and vegetables, were known to mitigate oxidative stress-related diseases including coronary heart disease, stroke, and cancer [39], positioning GZP samples as a promising functional ingredient rich in phenolic acids [40].

In pairwise comparisons, QM samples displayed significantly higher terpenoid content than GZP and LP, with pronounced enrichment in diterpenoid biosynthesis. Importantly, triterpenoid saponins and triterpenes, key bioactive constituents of Qingqiao, were substantially upregulated in QM. Terpenoids are involved in plant defense mechanisms and possess therapeutic potential against cancer and neurological disorders [41]. Lignans and coumarins were the main active components of Qingqiao [42], and displayed significantly higher relative contents in LP compared to GZP and QM in the two-group comparison. Particularly, isofraxidin-7-O-glucoside, 8-methoxycoumarin, and 7-hydroxycoumarin-O-rhamnoside, as important active constituents of Qingqiao, were significantly upregulated in LP. Metabolic pathway analysis of the differential metabolites between GZP-vs-LP and QM-vs-LP revealed that they were also enriched in the “biosynthesis of coumarins III” ($P < 0.05$), indicating significant differences in coumarin content between LP and GZP/QM. In plants, tryptophan serves as a precursor in metabolic pathways that induce the production of plumeranes, such as betanin, indole-3-carboxylic acid, and indole-5-carboxylic acid. These indole alkaloids are crucial for human health, as they exhibit diverse medicinal properties. For example, acetylcholine is used to treat Alzheimer's disease, glaucoma (as a miotic agent), and myasthenia gravis, while betaine functions as a dietary supplement to support cardiovascular health and serves as an adjunct therapy for liver diseases. Compared to QM and GZP, the elevated levels of these bioactive alkaloids further endorse LP’s suitability as a high-quality functional material.

Based on content levels, fold change, and *p*-value significance, 14 metabolites were identified as origin-specific biomarkers capable of distinguishing Qingqiao samples from different geographical origins. An OPLS-DA model constructed using these 14 biomarkers exhibited strong predictive accuracy and specificity. Notably, xylosyl phellodendroside, a structurally distinctive flavonoid glycoside, and 2-glucosyl-glucosyloxy-2-phenylacetic acid amide, a characteristic amide glycoside, were recognized as definitive biomarkers for Qingqiao. These compounds enabled reliable authentication of the medicinal material and facilitated precise tracing of its geographical origin. The AUC values from random forest analysis and external validation further confirmed the suitability of the newly established OPLS-DA model for differentiating Qingqiao originating from LP, QM, and GZP.

This study established a comprehensive framework for the evaluation and screening of high-quality germplasm resources, geographical origin traceability, and differential component utilization in Qingqiao. The findings provided valuable insights for the conservation and sustainable utilization of this medicinal plant species. Future research directions should focus on expanding the collection scope and quantity of Qingqiao germplasms across its natural distribution range. From the selected elite plant strains, the volatile oil content and key functional components were quantitatively determined, enabling an in-depth investigation into the biosynthesis of bioactive compounds and the molecular mechanisms underlying their potential medicinal properties.

Conclusion

In our study, the nutritional and functional components, and geographical origin traceability of Qingqiao from 3 geographical origins: GZP, QM, and LP were explored through UPLC-MS/MS combined with chemometrics. Our high-resolution metabolomic analysis resolved 5 principal phytochemical classes: phenolic acids, flavonoids, lignans and coumarins, terpenoids, and lipids. Chemometric evaluation revealed marked phytochemical heterogeneity among specimens from LPs, QMs, and GZPs ecoregions. Multivariate analysis through heat mapping and compositional histograms demonstrated that spatial variations in flavonoids, lipids, terpenoids, phenolic acids, lignans and coumarins content served as critical determinants for geographical authentication. Specific lines, GZP1, GZP2, GZP3, QM3, QM4, LP1, LP2, LP5, and LP6, could be extended to develop functional products.14 structurally diverse differential metabolites were identified as origin-specific biomarkers for origin discrimination. Robust OPLS-DA and ROC curves models were constructed using these markers achieved 93.33% accuracy in external validation for geographical discrimination. These findings established a foundation for origin traceability of Qingqiao specimens and provided valuable insights for quality control standardization and germplasm optimization. The identified origin-specific biomarkers present practical utility for pharmaceutical quality assessment and strategic breeding programs targeting specific bioactive components.

Declarations

Authors contributions

Tao Zeng: Investigation, Experimental design, Writing-Review & Editing. **Nian Liu**: Investigation, Resources. **Daojun Li**: Data analysis, Validation. **Qi Tang**: Investigation, Visualization. **Guodong Wang:** Methodology, Supervision. **Zhubing Hu**: Software, Methodology. **Haitao Zeng:** Conceptualization, Investigation, Funding acquisition, Project administration. **Tiantian Chen:** Data curation, Picture editing, Funding acquisition.

Fundings

This investigation received research support from 3 distinguished funding programs: (1) Shaanxi Provincial Science and Technology Program Projects (No: 2025NC-YBXM-006, 2025NC-YBXM-061), (2) Talent Initiation Project of Shaanxi University of Technology (No: SLGRCQD-013, SLGRCQD-054); (3) The

“State Key Laboratory (Cultivation) of Qinba Mountains Biological Resources and Ecological Environment, ‘co-construction of city and school’ scientific research project industrialization project” (Project No. SXC-2309).

Data Availability Statements

All the metabolite data of Qingqiao demonstrated in our study were availably obtained from the corresponding author on reasonable request.

Ethics approval and consent to participate

This research on wild *Forsythia suspensa* fruits (Qingqiao) was approved by the local government. The collecting of these materials was allowed by the Convention on the Trade in Endangered Species of Wild Fauna and Flora and Regulations of the protection of wild plants in Shaanxi (Huanglong, Yichuan, Luonan, Danfeng, Tongguan, Weibin), Shanxi (Yongji, Jixian, Anze), and Henan (Lushi) provinces. All the methods were carried out in compliance with local and national regulations. The wild Qingqiao in the study were collected from natural populations, which did not affect the survival or reproduction of the species. All Qingqiao samples were authenticated by Professor Wu Zhenhai from Northwest A & F University, and the voucher specimens of all samples, with voucher ID number form SLG01 to SLG20, were stored in School of Biological Science and

Engineering, Shaanxi University of Technology, Hanzhong, Shaanxi Province, China.

Consent for publication

Not applicable.

Declaration of competing interest

All authors declared that they had no known competitive financial interests or personal relationships that could have appeared to influence the study reported in this paper.

References

1. Song Y, Du XR, Li AX, Fan AM, He LJ, Sun Z, Niu YB, Qiao YG. Assembly and analysis of the complete mitochondrial genome of *Forsythia suspensa* (Thunb.) Vahl. Bmc Genomics. 2023; 24(1):708.
2. Wang ZY, Xia Q, Liu X, Liu WX, Huang WZ, Mei X ,Luo J, Shon MX, Lin RC, Zou DX, Ma ZQ. Phytochemistry, pharmacology, quality control and future research of *Forsythia suspensa* (Thunb.) Vahl: A review. J. Ethnopharmacol. 2018; 210: 318-339.
3. Guo H, Liu AH, Ye M, Yang M, Guo DA. Characterization of phenolic compounds in the fruits of *Forsythia suspensa* by high-performance liquid chromatography coupled with electrospray ionization tandem mass spectrometry. Rapid Commun. Mass Sp. 2007; 21(5): 715-729.

4. Wang XM, Jia QW, Yao XY, Yang LQ, Pei K, Guo LL, Guo Y, Yang YK, Qin N. Analysis of *Forsythia suspensa* fruit and leaf extracts using UHPLC-Q-Exactive-Orbitrap/MS: In vivo antioxidant activity on D-galactose-induced aging mice. Food Chem. 2025; 462:141002.
5. Wu, XQ, Zhang WN, Hao MZ, Liu XP, Xiao J, Wang TF, Dong YZ, Zhao J. How Chinese herbal medicine prevents epidemics: from ancient pestilences to COVID-19 pandemic. Am. J. Chinese Med. 2021;49(05):1017-44.
6. Wang E, Lu ZR, Rohani ER, Ou JM, Tong XH, Han RC. Current and future distribution of *Forsythia suspensa* in China under climate change adopting the MaxEnt model. Front. Plant Sci. 2024;15:1394799.
7. Gao S, Liu XN, Liu Y, Cao BL, Chen ZJ, Xu K. The spectral irradiance, growth, photosynthetic characteristics, antioxidant system, and nutritional status of green onion (*Allium fistulosum* L.) grown under different photo-selective nets. Front. Plant Sci. 2021;12:650471.
8. Li Y, Shi LC, Pei NC, Cushman SA, Si YT. Transcriptomic responses to drought stress among natural populations provide insights into local adaptation of weeping forsythia. BMC Plant Biol. 2021; 21(1):273.
9. Mykhailenko O, Gudžinskas Z, Kovalyov V, Desenko V, Ivanauskas L, Bezruk I, Georgiyants V. Effect of ecological factors on the accumulation of phenolic compounds in *Iris* species from Latvia, Lithuania and Ukraine. Phytochem. Analysis. 2020; 31(5):545-63.
10. Zhou M, Huo J, Wang C, Wang W. UPLC/Q-TOF MS screening and identification of antibacterial compounds in *Forsythia suspensa* (Thunb.) Vahl leaves. Front. Pharmacol. 2022;12:704260.
11. Pang X, Zhang J, Yu Y, Sun G. Quality grade evaluation and the related research of *Forsythia suspensa* from different places on the market. New J. Chem. 2021; 45(37):17428-17437.
12. Wang H, Song W, Tao W, Zhang J, Zhang X, Zhao J, Yong J, Gao X, Guo L. Identification wild and cultivated licorice by multidimensional analysis. Food Chem. 2021; 339: 128111.
13. Wang HZ, Zhao MY, Wu ZZ, Qin NN, Fu YX, Guo S. Nutrient composition and functional constituents of daylily from different producing areas based on widely targeted metabolomics. Food Chem. X 2024; 21:101239.
14. Shi Q, Han G, Liu Y, Jiang J, Jia Y, Li X. Nutrient composition and quality traits of dried jujube fruits in seven producing areas based on metabolomics analysis. Food Chem. 2022; 385:132627.
15. Han X, Yin M, Fang Q, Tan X, Sun H, Cheng ME, Peng H, Huang L. Nutritional ingredients and functional components of cultivated and wild-simulated *Astragali radix* using widely targeted metabolomics. LWT-Food Sci. Technol. 2023;185:115186.
16. Shang X, Huang D, Wang Y, Xiao L, Ming R, Zeng W, Cao S, Lu L, Wu Z, Yan H. Identification of nutritional ingredients and medicinal components of *Pueraria lobata* and its varieties using UPLC-MS/MS-based metabolomics. Molecules. 2021; 26(21): 6587.
17. Wang H, Zhao M, Wu ZZ, Qin N, Fu Y, Guo S. Nutrient composition and functional constituents of daylily from different producing areas based on widely targeted metabolomics. Food Chem. X. 2024; 21:101239.

18. Dossou SS, Xu F, You J, Zhou R, Li D, Wang L. Widely targeted metabolome profiling of different colored sesame (*Sesamum indicum* L.) seeds provides new insight into their antioxidant activities. Food Res. Int. 2022;151:110850.
19. Huang H, Wei Y, Huang S, Lu S, Su H, Ma L, Huang W. Integrated metabolomic and transcriptomic analyses provide insights into regulation mechanisms during bulbous stem development in the Chinese medicinal herb plant, Stephania kwangsiensis. BMC Plant Biol. 2024; 24(1):276.
20. Zheng T, Yang J, Chen Q, Huang X, Xue Y, Tang Q, Wang G, Li Y, Hu Z, Zeng HT. Analysis of lipidomics profile of *Brassica napus* hybrid 'Fangyou 777'and its parents during ripening stages based on UPLC-MS/MS. BMC Plant Biol. 2025; 25(1):197.
21. Weng D, Tian R, Jin H, Zha S, Zhao Q, Zhao B. Study on extraction, purification, and biological activities of forsythiaside A from *Forsythia suspensa* (Thunb.) Vahl leaf. Biomass Conver. Bior. 2024;14(20):25015-31.
22. Chen Z, Liu L, Gao C, Chen W, Vong CT, Yao P, Yang Y, Li X, Tang X, Wang S, Wang Y. *Astragali Radix* (Huangqi): A promising edible immunomodulatory herbal medicine. J. Ethnopharmacol. 2020; 258:112895.
23. Adetunji C, Palai S, Ekwuabu C, Egbuna C, Mtewa A. General principle of primary and secondary plant metabolites: Biogenesis, metabolism, and extraction (Chapter 1). In Preparation of phytopharmaceuticals for the management of disorders 2020; (pp. 3-23).
24. Hu X, Guo F. Amino acid sensing in metabolic homeostasis and health. Endocr. Rev. 2021; 42(1):56-76.
25. Day L, Cakebread JA, Loveday SM. Food proteins from animals and plants: Differences in the nutritional and functional properties. Trends Food Sci. Tech. 2022;119:428-442.
26. Huang C, Li Y, Wang K, Xi J, Xu Y, Si X, Pei D, Lyu S, Xia G, Wang J, Li P. Analysis of lipidomics profile of *Carya cathayensis* nuts and lipid dynamic changes during embryonic development. Food Chem. 2022; 370:130975.
27. Zheng T, Tian M, Deng Z, Tang Q, Hu Z, Wang G, Zeng H. UPLC-MS/MS reveals the differences in lipids composition of *Camellia oleifera* from northern margin distribution area. Food Chem. X. 2024; 23:101629.
28. Zhang D, Li X, Duan X, Sun H, Cao Y. Lipidomics reveals the changes in lipid profile of flaxseed oil affected by roasting. Food Chem. 2021;364:130431.
29. Wang X, Wan X, Zhu T, Cui L, Xiao Q. Metabolomics combined with network pharmacology reveals the anti-hepatoma effects of terpenoids from *Polygonatum kingianum* var. grandifolium and *Polygonatum sibiricum* Redouté as well as differences in their terpenoid metabolites. BMC Plant Biol. 2025; 25(1):769.
30. Nguyen HN, Ullevig SL, Short JD, Wang L, Ahn YJ, Asmis R. Ursolic acid and related analogues: triterpenoids with broad health benefits. Antioxidants. 2021;10(8):1161.
31. Castellano JM, Ramos-Romero S, Perona JS. Oleanolic acid: extraction, characterization and biological activity. Nutrients. 2022;14(3):623.

32. Su HF, Shaker S, Kuang Y, Zhang M, Ye M, Qiao X. Phytochemistry and cardiovascular protective effects of Huang-Qi (*Astragali Radix*). Med. Res. Rev. 2021; 41(4):1999-2038.
33. Li L, Zhao Y, Zhang S, Zhang X, Duan G. Comparison of *Forsythia suspensa* leaf fermented tea in the four different regions of Shanxi Province: quality, functional ingredients, and bioactivity. Front. Sustain. Food S. 2024; 8:1482782.
34. Barreca D, Trombetta D, Smeriglio A, Mandalari G, Romeo O, Felice MR, Gattuso G, Nabavi SM. Food flavonols: Nutraceuticals with complex health benefits and functionalities. Trends Food Sci. Tech. 2021;117:194-204.
35. Zheng T, Zeng HT, Sun BY, Liu SM. Multi-environment evaluations across ecological regions reveal climate and soil effects on amides contents in Chinese prickly ash peels (*Zanthoxylum bungeanum* Maxim.). BMC Plant biol. 2023; 23(1):313.
36. Cai WQ, Jiang PF, Liu Y, Miao XQ, Liu AD. Distinct changes of taste quality and metabolite profile in different tomato varieties revealed by LC-MS metabolomics. Food Chem. 2024; 442:138456.
37. Oteiza PI, Fraga CG, Galleano MJ. Linking biomarkers of oxidative stress and disease with flavonoid consumption: from experimental models to humans. Redox Biol. 2021; 42: 101914.
38. Zhan X, Chen Z, Chen R, Shen C. Environmental and genetic factors involved in plant protection-associated secondary metabolite biosynthesis pathways. Front. Plant Sci. 2022;13:877304.
39. Hou J, Liang L, Su M, Yang T, Mao X, Wang Y. Variations in phenolic acids and antioxidant activity of navel orange at different growth stages. Food Chem. 2021;360:129980.
40. Khalid W, Ali A, Arshad MS, Afzal F, Akram R, Siddeeg A, Kousar S, Rahim MA, Aziz A, Maqbool Z, Saeed A. Nutrients and bioactive compounds of *Sorghum bicolor* L. used to prepare functional foods: a review on the efficacy against different chronic disorders. Int. J. Food Prop. 2022 Dec 31;25(1):1045-1062.
41. Luo Q, Tian Z, Zheng T, Xu S, Ma Y, Zou S, Zuo Z. Terpenoid composition and antioxidant activity of extracts from four chemotypes of *Cinnamomum camphora* and their main antioxidant agents. Biofuel. Bioprod. Bior. 2022;16(2):510-522.
42. Liu Y, Yang L, Wang J, Chen D. New lignans and phenylethanoid with antioxidant activity from aerial parts of *Forsythia suspensa* (Thunb.) Vahl. Nat. Prod. Res. 2023; 37(5):725-733.

Figures

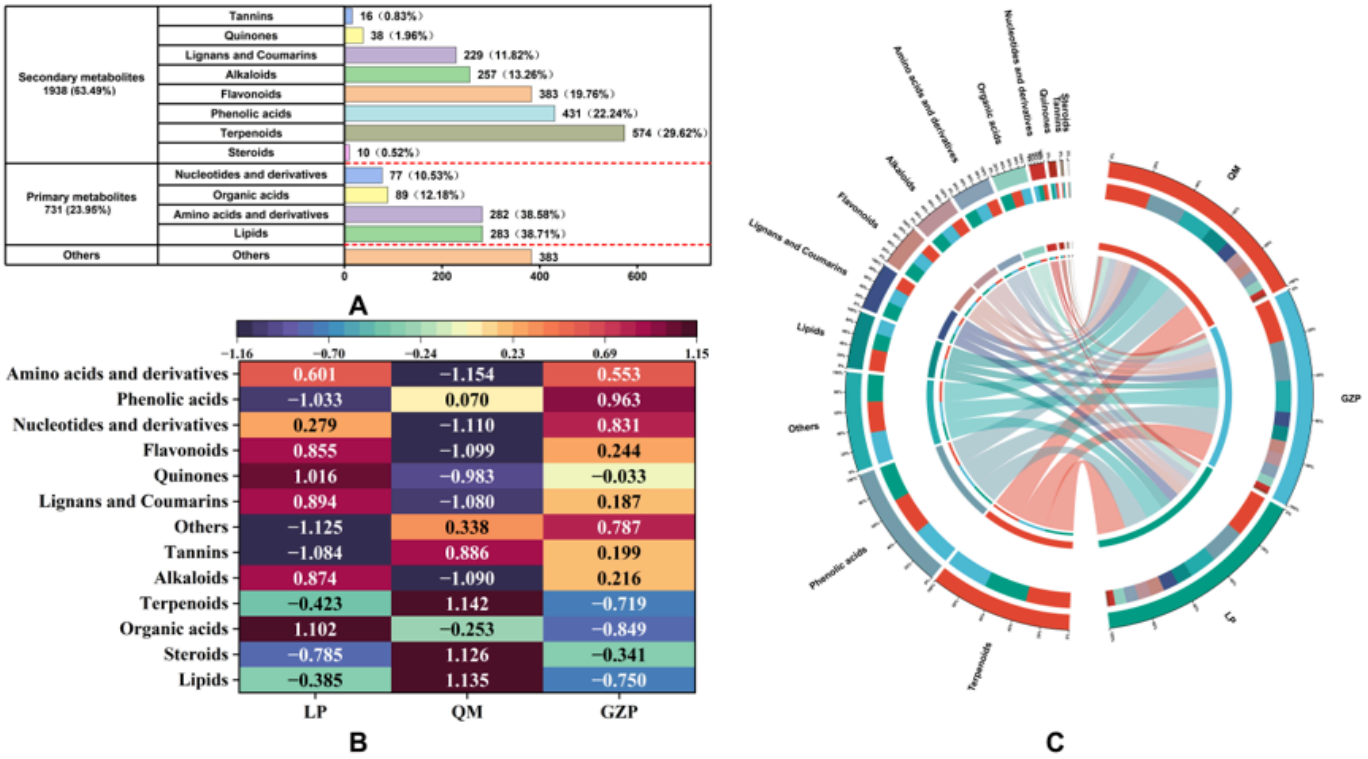

Figure 1

A: Grouped bar the 13 categories of 3052 metabolites. **B:** Chord diagram of GZP, LP, QM and metabolites. **C**: The content of 13 categories of Qingqiaofrom LP, QM, and GZP. The color of the heat map represented the abundance of metabolites, with red indicating a higher content and blue indicating a lower content.

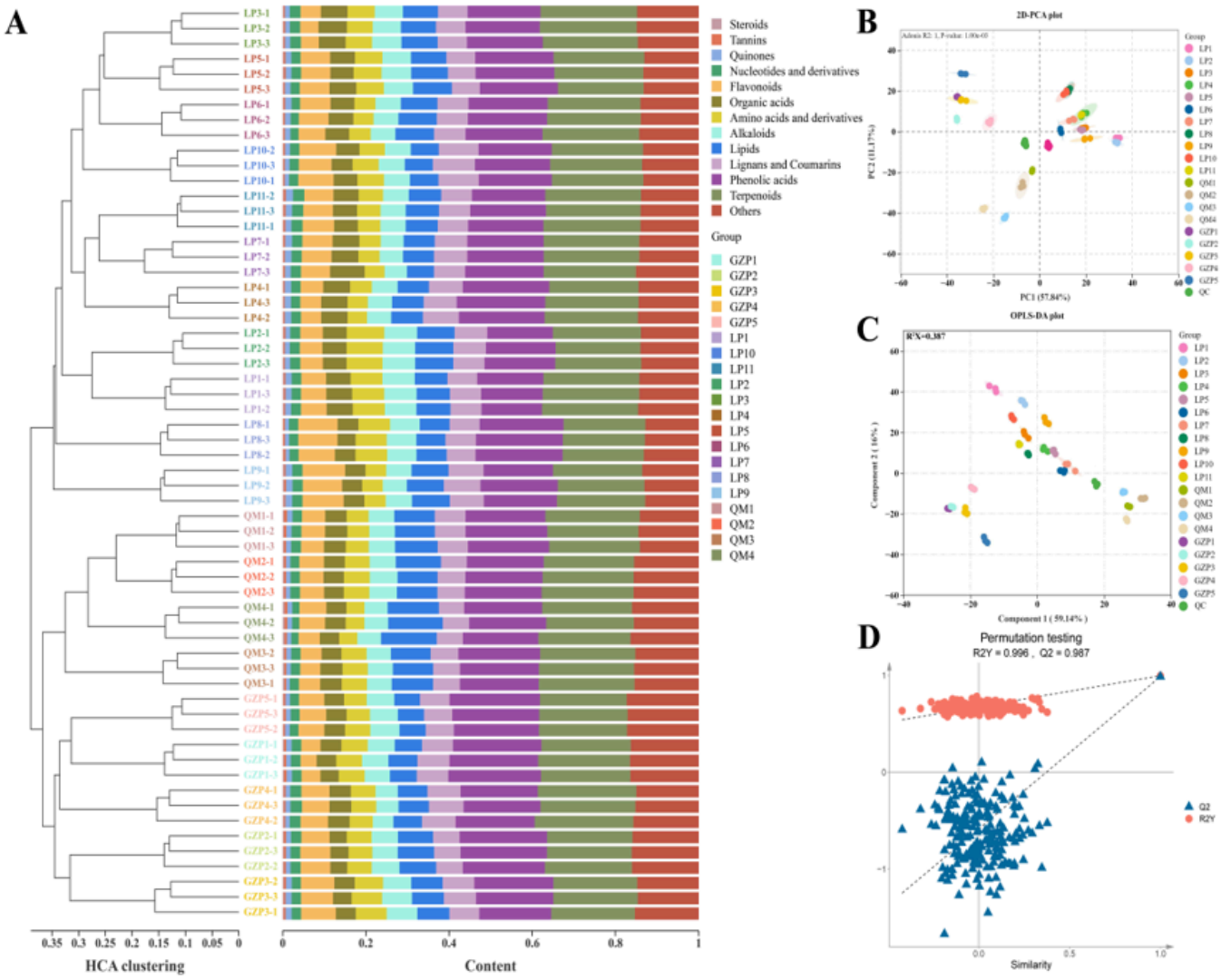

Figure 2

Chemometric analysis plots of the metabolomic data of 20 samples from different origins.

A: HCA plots; **B**: 2D-PCA plot; **C**: OPLS-DA analysis plot; **D**: Permutation testing plot.

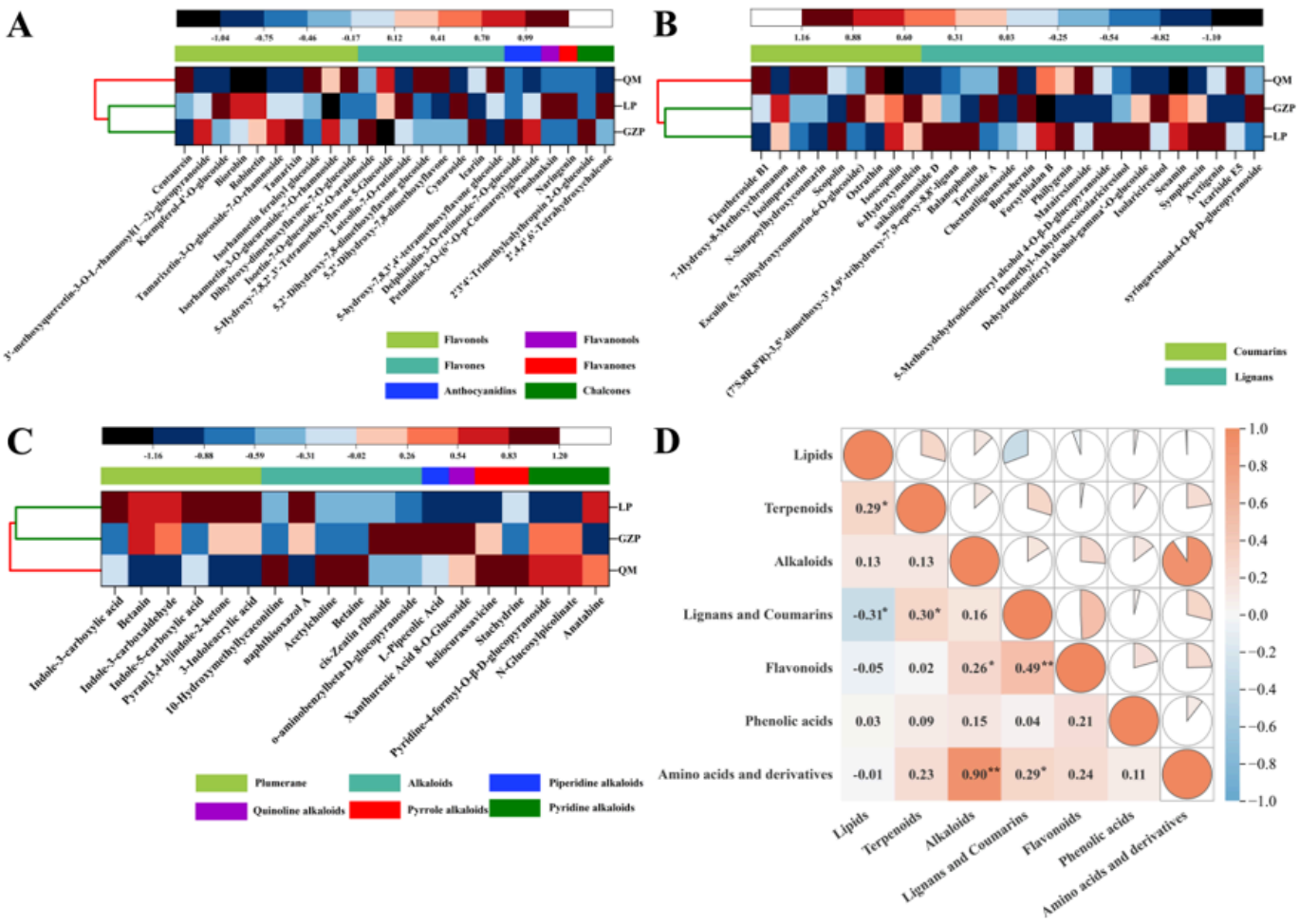

Figure 3

Heat map of the differential metabolites with higher content among GZP, LP, and QM. **A**: flavonoids. **B**: Lignans and coumarins. **C**: Alkaloids. **D**: Correlation analysis of lipids, amino acids and derivatives, terpenoids, flavonoids, phenolic acids, alkaloids, with lignans and coumarins, * indicated the significant difference at 0.05 level, ** indicated the significant difference at 0.01 level.

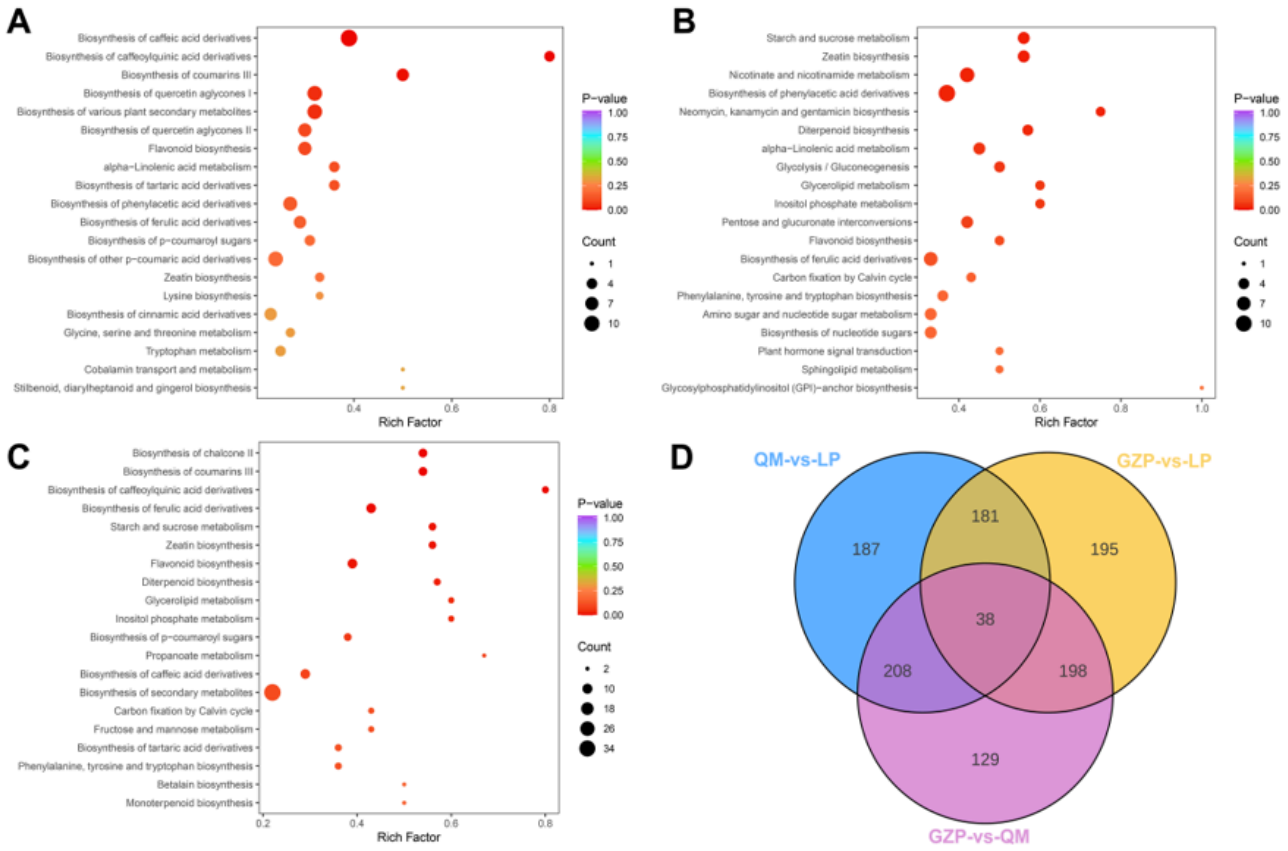

Figure 4

KEGG pathway and Venn diagram analysis of differential metabolites across 3 comparison groups.

A-C: KEGG annotation and enrichment results of differentially accumulated metabolites for the following pairwise comparisons: GZP-vs-LP, GZP-vs-QM, and QM-vs-LP. **D**: Venn diagram illustrating the overlapping metabolites shared among the 3 comparison groups.

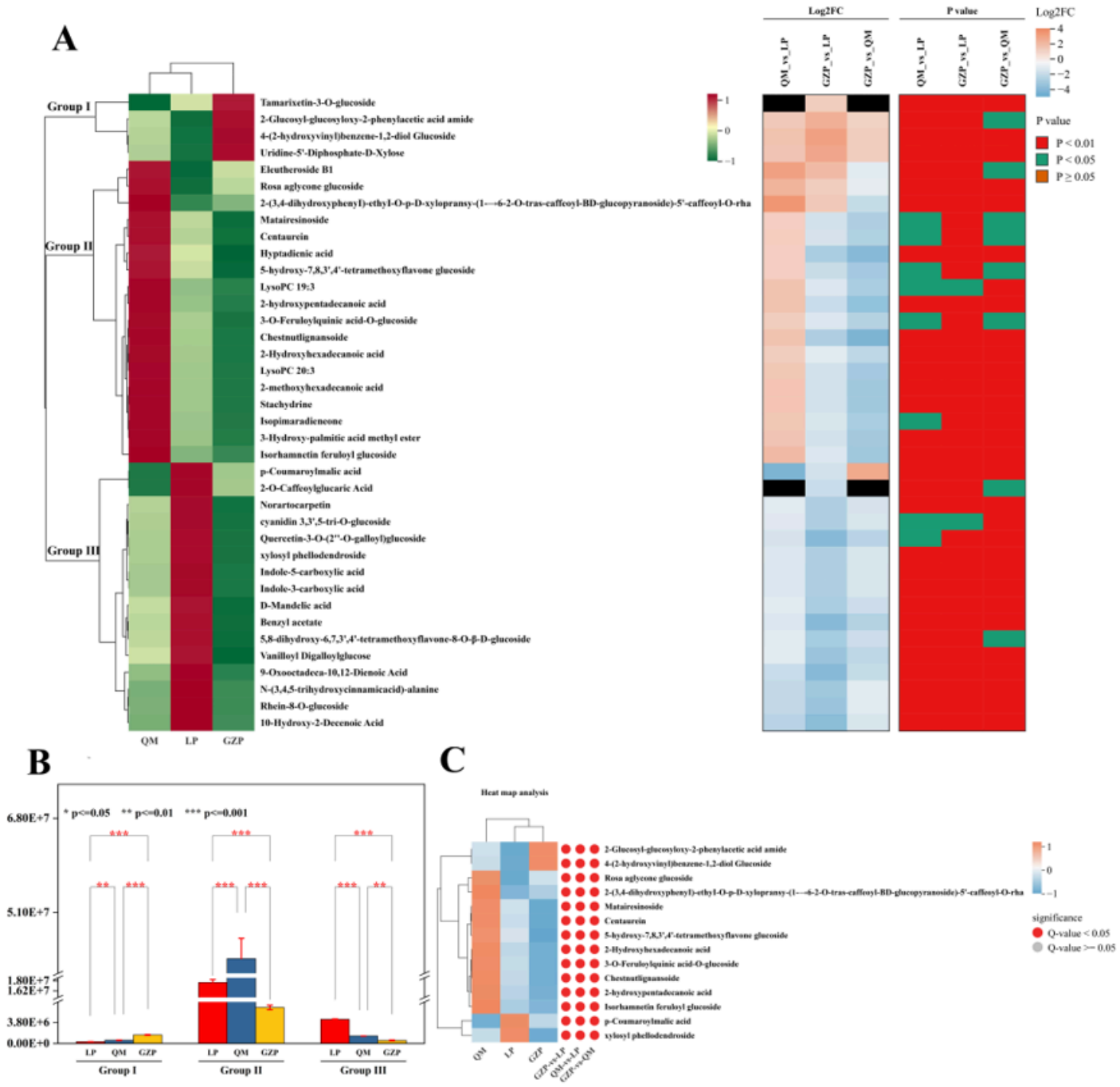

Figure 5

A: Complex heatmap of the 38 overlapping differential metabolites. The color of the heat map represented the abundance of metabolites, with red indicating a higher content and green indicating a lower content. **B**: Paired comparison plot of the compounds in Group I, Group II, and Group III. **C**: Heat map analysis from 14 candidate differential metabolites of the LP, QM, and GZP samples.

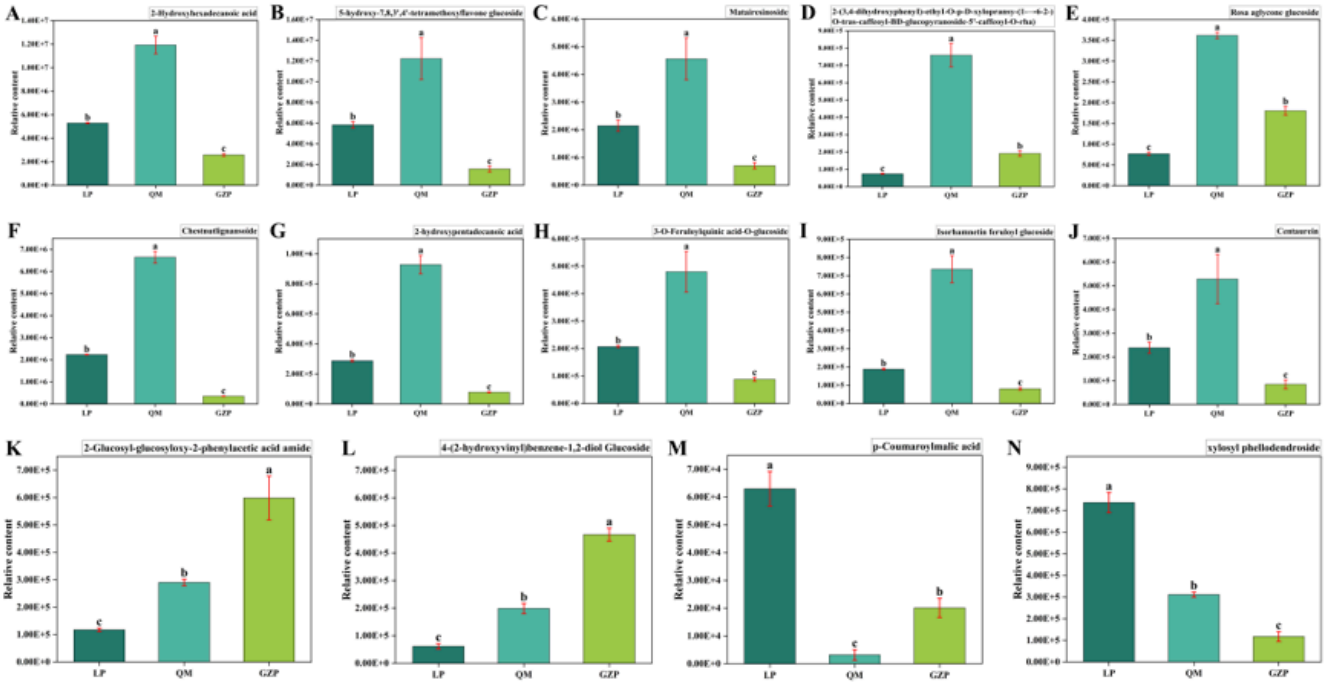

Figure 6

Content variation of those 14 key differential metabolites of Qingqiaosamples from LP, QM, and GZP.

Supplementary Files

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

- Supplementarymaterial.zip