



Contents lists available at ScienceDirect

## Chinese Herbal Medicines

journal homepage: [www.elsevier.com/locate/chmed](http://www.elsevier.com/locate/chmed)

## Original Article

## A strategy for rapid discovery of traceable chemical markers in herbal products using MZmine 2 data processing toolbox: A case of Jing Liqueur

Jing Zhou<sup>a</sup>, Feng-jie Liu<sup>a</sup>, Xin-xin Li<sup>a</sup>, Ping Li<sup>a</sup>, Hua Yang<sup>a</sup>, Yuan-cai Liu<sup>b,c,\*</sup>, Yan-he Chen<sup>b,c</sup>, Chao-dan Wei<sup>b,c</sup>, Hui-jun Li<sup>a,\*</sup><sup>a</sup>State Key Laboratory of Natural Medicines, China Pharmaceutical University, Nanjing 210009, China<sup>b</sup>Hubei Provincial Key Laboratory for Quality and Safety of Traditional Chinese Medicine Health Food, Jing Brand Co., Ltd., Daye 435100, China<sup>c</sup>Jing Brand Research Institute, Jing Brand Co., Ltd., Daye 435100, China

## ARTICLE INFO

## Article history:

Received 25 September 2020

Revised 8 December 2020

Accepted 22 March 2021

Available online 28 May 2021

## Keywords:

GC–MS

herbal products

HPLC–QTOF–MS

Jing Liqueur

MZmine 2

traceable markers

## ABSTRACT

**Objective:** The quality evaluation of herbal products remains a big challenge. Traceable markers are the core concept of the authentication of herbal products. However, the discovery of traceable markers is labor-intensive and time-consuming. The aim of this study is to develop a convenient approach to rapidly screen the traceable markers for herbal product authentication.

**Methods:** Commercial Jing Liqueur and its 22 species of herbal ingredients were analyzed using HPLC–QTOF–MS and GC–MS to characterize nonvolatile and volatile chemicals. The acquired data were imported into MZmine 2 software for mass detection, chromatogram building, deconvolution and alignment. The aligned data were exported into a csv file and then traceable markers were selected using the built-in filter function in Excel. Finally, the traceable markers were identified by searching against online databases or publications, some of which were confirmed by reference standards.

**Results:** A total of 288 chemical features transferred from herbal materials to Jing Liqueur product were rapidly screened out. Among them, 52 markers detected by HPLC–QTOF–MS were annotated, while nine volatile markers detected by GC–MS were annotated. Moreover, 30 of these markers were confirmed by comparing with reference standards. A chemical fingerprint consisting of traceable markers was finally generated to ensure the authentication and quality consistency of Jing Liqueur.

**Conclusion:** A strategy for rapid discovery of traceable markers in herbal products using MZmine 2 software was developed.

© 2021 Tianjin Press of Chinese Herbal Medicines. Published by ELSEVIER B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Herbal products have been widely used as both therapeutic medicines and functional foods in healthcare worldwide (Aparicio-Soto, Sánchez-Hidalgo, Rosillo, Castejón, & Alarcón-de-la-Lastra, 2016; Guo, Bian, Qiu, Wang, & Wang, 2017). As the application of herbal products is increasingly integrated into the modern health care system, their quality has been one of the most important concerns in this field (Cañigueral et al., 2008; Liang, Xie, & Chan, 2004). However, herbal products, commonly generated from multiple raw materials, have extremely complex chemical compositions and undergo sophistic manufacturing processes, which makes quality evaluation a challenging task.

Chemical markers are often the bioactive basis for both raw materials and their related products, become a core concept of her-

bal product authentication (Fan, Cheng, Ye, Lin, & Qian, 2006; Liu, Guo, & Liu, 2018; Wang et al., 2017; Zhao et al., 2019). Traceable chemical markers are compounds that exist in any form of an herbal medicine such as the raw materials, pieces, extracts and finished products. In general, traceable chemical markers are selected by analyzing the variation of targeted chemicals throughout the whole production process, which is labor-intensive and time-consuming (Liu et al., 2020; Yu, Wen, & Su, 2013; Zhao et al., 2019). Herein, we proposed that potential traceable markers could be rapidly discovered by comparing the chemical profiles in the raw and finished forms of herbal products.

Liquid/gas chromatography coupled with mass spectrometry (LC/GC–MS) has become the most powerful tool to profile the chemical components in herbal products because of its high selectivity, sensitivity and throughput, as well as the ability to generate specific information including molecular mass and structural characteristics (Gomathi, Kalaiselvi, Ravikumar, Devaki, & Uma, 2015; Lai, Cheng, & Tsai, 2015; Müller & Bracher, 2015; Weitzel,

\* Corresponding authors.

E-mail addresses: [lyc@jingpai.com](mailto:lyc@jingpai.com) (Y.-c. Liu), [cpuli@126.com](mailto:cpuli@126.com) (H.-j. Li).

2011). MZmine 2 is an open-source toolbox for the processing and visualization of mass spectrometry based on molecular profile data, which can automatically process a large number of mass spectra in numerous ways, including mass detection, chromatogram building, deconvolution, alignment and compound identification (Pluskal, Castillo, Villar-Briones, & Orešič, 2010). As mentioned above, traceable chemical markers exist in both raw materials and finished products. That is, after feature detection and alignment in MZmine 2, the potential traceable markers can be highlighted by comparing the feature distribution among the raw materials and their related products. Although there were numerous analogous free software suites like XCMS (Tautenhahn, Patti, Rinehart, & Siuzdak, 2012), OpenMS (Röst et al., 2016) or SMART (Liang et al., 2016), MZmine 2 is one of the most popular and user-friendly packages for its straightforward operations for nonexpert users and a flexible and modular platform for MS data processing (Olivon, Grelier, Roussi, Litaudon, & Touboul, 2017).

Jing Liqueur, composed of 22 species of herbal materials, is a very popular healthcare liqueur with anti-fatigue, anti-inflammation and immunity-enhancing activities (Feng, Shan, Lu, Liu, & He, 2013; Shan et al., 2018). The trace levels, large number, and diverse structures and properties of the chemical components in Jing Liqueur increase the difficulty of its authentication. In this study, LC-MS and GC-MS were complementarily applied to characterize nonvolatile and volatile compounds in both the herbal materials and finished product of Jing Liqueur. The MZmine 2 tool was utilized to detect the mass features in all the samples and those features were aligned according to the mass-to-charge ratio ( $m/z$ ) and retention time (rt). The chemical transitivity from raw materials to products was visualized to select potential traceable markers, which were then identified by comparing with reference standards or by searching against the available MS/MS databases and publications (Fig. 1). Consequently, a total of 288 features were transferred into Jing Liqueur from the 22 raw materials and 61 were annotated. In addition, considering the practicability of authentication, a chemical fingerprint consisting of traceable markers of Jing Liqueur was constructed.

## 2. Materials and methods

### 2.1. Regents and materials

Jing Liqueur samples (No. P1707122/05) were provided by Jing Brand Co., Ltd. (Hubei, China). The 22 species of herbal materials (Table 1) were collected from different areas of China and authen-

ticated by Prof. Hui-jun Li according to Chinese Pharmacopoeia (2015). Voucher specimens were deposited in the State Key Laboratory of Natural Medicines, China Pharmaceutical University, Nanjing, China.

Reference standards of betaine (4), 5-hydroxymethylfurfural (9), 3'-hydroxy puerarin, 3,4-dihydroxybenzaldehyde, puerarin (17), magnoflorine, 3'-methoxypuerarin (19), calycosin-7-glucoside,  $\beta$ -ecdysone, lithospermic acid, salvianolic acid A (33), isoflavones aglycone, sagittatoside B (39), 2''-O-rhamnosyl icarisperidol, epimedine C, icarisperidol I, baohuoside II, and diosgenin (52) (purity  $\geq 98\%$ ) were purchased from Shanghai Yuanye Bio-Technology Co., Ltd (Shanghai, China). The other reference standards were purchased from Shanghai Haoyuan Bio-Technology Co., Ltd (Shanghai, China). HPLC-grade acetonitrile (ACN) and methanol (MeOH) were supplied by Merck (Darmstadt, Germany). Formic acid of HPLC grade was purchased from ROE Co., Ltd. (Newark, USA). Ultra-pure water was purified by a Millipore water purification system (Millipore, Milford, MA, USA).

### 2.2. Sample preparation

Four different standard solutions, comprising a total of 52 reference standards at concentration of 20  $\mu\text{g/mL}$  in MeOH, were prepared. The details are summarized in Table 2.

For HPLC-QTOF-MS analysis, the samples of Jing Liqueur and herbal materials were directly analyzed after centrifugation at 13 000 rpm/min for 10 min. For GC-MS analysis, it was necessary to remove all water in the sample solutions prior to injection. In detail, 4 g of  $\text{Na}_2\text{SO}_4$  was added into 6 mL of the herbal material extracts or Jing Liqueur solution. After blending, the mixtures were settled at room temperature for 10 h and then centrifuged at 13 000 rpm/min for 10 min. The supernatant was collected for GC-MS analysis.

### 2.3. Chromatography and mass spectrometry conditions

#### 2.3.1. HPLC-QTOF-MS method

Chromatographic separation was performed on an Agilent 1290 LC system (Agilent Technologies, USA). The samples were separated on an Agilent Zorbax Extend- $\text{C}_{18}$  (octadecyl chemically bonded phase silica gel) analytical column (4.6 mm  $\times$  250 mm, 5  $\mu\text{m}$ ) at 25  $^\circ\text{C}$ . The mobile phase was composed of solvent A (water containing 0.1% formic acid) and solvent B (ACN containing 0.1% formic acid). The elution gradient was as follows: 0–2.5 min, 10% B; 2.5–20 min, 10%–20% B; 20–35 min, 20%–35% B; 35–40 min,

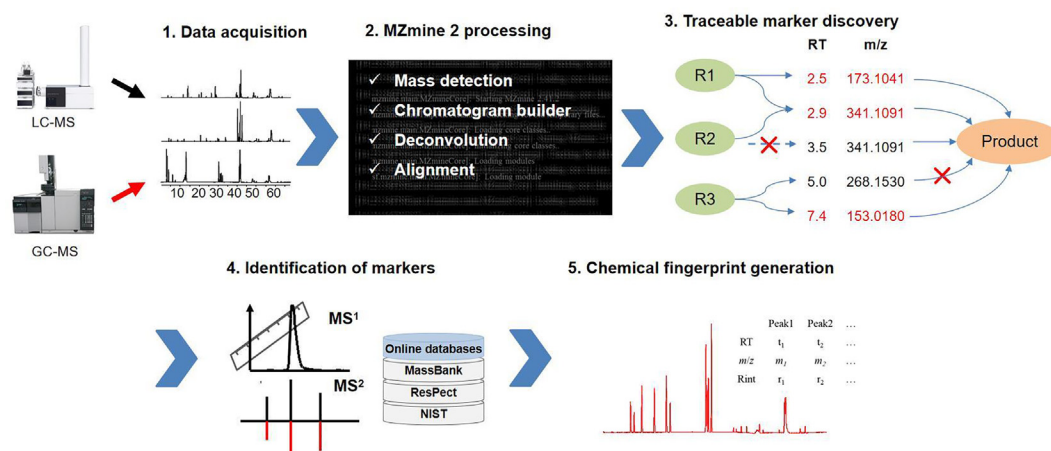


Fig. 1. Workflow of transitive chemical discovery.

**Table 1**  
Twenty-two species of raw materials involved in Jing Liqueur.

No.	Raw materials	No.	Raw materials
1	<i>Puerariae Lobatae Radix</i> (PLR)	12	<i>Epimedii Folium</i> (EF)
2	<i>Cuscutae Semen</i> (CS)	13	<i>Dioscoreae Rhizoma</i> (DR)
3	<i>Angelicae Sinensis Radix</i> (ASR)	14	<i>Curculiginis Rhizoma</i> (CRh)
4	<i>Caryophylli Flos</i> (CF)	15	<i>Codonopsis Radix</i> (CR)
5	<i>Cistanches Herba</i> (CH)	16	<i>Amomi Fructus</i> (AF)
6	<i>Salviae Miltiorrhizae Radix et Rhizoma</i> (SMRR)	17	<i>Astragali Radix</i> (AR)
7	<i>Achyranthis Bidentatae Radix</i> (ABR)	18	<i>Lycii Fructus</i> (LF)
8	<i>Alpiniae Oxyphyllae Fructus</i> (AOF)	19	<i>Allii Tuberosi Semen</i> (ATS)
9	<i>Morindae Officinalis Radix</i> (MOR)	20	<i>Euryales Semen</i> (ES)
10	<i>Imperatae Rhizoma</i> (IR)	21	<i>Rehmanniae Radix Praeparata</i> (RRP)
11	<i>Rosae Laevigatae Fructus</i> (RLF)	22	<i>Cinnamomi Cortex</i> (CC)

**Table 2**  
Preparation of reference solutions.

Reference solutions	Compositions	Concentration ( $\mu\text{g}\cdot\text{mL}^{-1}$ )
1	3'-hydroxy puerarin, chlorogenic acid, puerarin, daidzin, calycosin-7-glucoside, $\beta$ -ecdysone, isoflavones aglycone, epimedin B, epimedin C, icariin	20
2	betaine, adenosine, 5-hydroxymethylfurfural, magnoflorine, 3,2'-dihydroxyflavone, Ononin, formononetin, coumarin, epimedin A1, epimedin A, icariside I, salidroside, diosgenin sucrose, geniposide, geniposidic acid, 3,4-dihydroxybenzaldehyde, 3'-methoxypuerarin, purpureaside C, rutin, quercitrin, azelaic acid, rosmarinic acid, sagittoside A, 2''-O-rhamnosyl icariside II, astragaloside A, astragaloside II, ginsenoside Ro	20
3	nystose, citric acid, protocatechuic acid, 4-dicaffeoylquinic acid, echinacoside, acteoside, isoacteoside, lithospermic acid, salvianolic acid A, sagittoside B, astragaloside III, baohuoside II, chikusetsu saponin IVa	20

35%–60% B; 40–45 min, 60%–90% B; 45–50 min, 90%–95% B; 50–60 min, 95% B. The injection volume was set at 2  $\mu\text{L}$ , and the flow rate was set at 0.8 mL/min.

The above HPLC system was coupled to an Agilent 6545 QTOF-MS spectrometer (Agilent Technologies, USA) equipped with an electrospray ionization (ESI) source. The capillary voltage was set at 4000 V. The fragmentor voltage was set at 135 V. The flow rate of drying-gas ( $\text{N}_2$ ) was 10 L/min with a temperature of 350  $^\circ\text{C}$ , and the nebulizer pressure was 35 psi. The sheath gas was 11.0 L/min with a temperature of 350  $^\circ\text{C}$ , the cone voltage was 65 V and OCT 1 RF Vpp was set at 750 V. Mass spectra were recorded over a mass-to-charge ratio ( $m/z$ ) range of 50–1700 using both positive and negative ion detection modes. Multiple collision energies at 15, 25, 35, and 50 V were set to acquire MS/MS data.

### 2.3.2. GC–MS method

GC analyses were performed using an Agilent 7890B GC (Palo Alto, CA, USA) equipped with an Agilent-5977A MSD (Agilent Technologies, USA) and an Agilent 19091S-433UI HP-5MS (5% phenylmethylpolysiloxan) Ultra Inert (30 m  $\times$  250  $\mu\text{m}$ , 0.25  $\mu\text{m}$ ) column. Helium (purity  $\geq$  99.999%) was used as the carrier gas at a flow rate of 1 mL/min. A 1  $\mu\text{L}$  aliquot of sample was injected in split mode at a split ratio of 10:1 and at an injection temperature of 250  $^\circ\text{C}$ . The oven temperature was programmed to increase from an initial temperature of 40  $^\circ\text{C}$  (held for 5 min) to 250  $^\circ\text{C}$  at 8  $^\circ\text{C}/\text{min}$ . All samples were analyzed with electron ionization (70 eV) in full scan mode.

### 2.4. Data processing in MZmine 2 software

Raw data files were converted into the open format mzML using MSConvert software, which was a part of the cross-platform ProteoWizard program (<http://proteowizard.sourceforge.net/>). An optimized MZmine 2.41.2 workflow was developed for feature list generation.

For HPLC-QTOF-MS data, centroid mass detection was performed with a noise level of 0 and then the ADAP chromatogram builder was implemented including a minimum group size of 5, a group intensity threshold of  $10^3$ , and a  $m/z$  tolerance set to 0.01 Da or a relative tolerance of  $2 \times 10^{-5}$ . Then, the Wavelets (ADAP) algorithm was applied for chromatogram deconvolution including a signal-to-noise ratio of 10, a minimum feature height of  $5 \times 10^3$ , a coefficient threshold of 110, a peak duration range of 0–10 min and an rt wavelet range of 0–0.1 min. The alignment was performed using RANSCAN algorithm with an  $m/z$  tolerance of 0.01 Da or a relative tolerance of  $2 \times 10^{-5}$ , an rt tolerance of 0.5 min and an rt tolerance after correction of 0.1 min. RANSCAN iterations were estimated automatically and the threshold value was set at 4 s. In addition, isotopes were dismissed using an isotopic peak grouper with an  $m/z$  tolerance of 2, an rt tolerance of 0.1 and a maximum charge of 2; A duplicate peak filter was performed with an  $m/z$  tolerance of 0.01 and an rt tolerance of 0.1 min. The processed feature list was exported as a csv file containing rt,  $m/z$  and peak area data of different samples.

For GC–MS data, the workflow included the centroid mass detection algorithm, with a noise level of 0. ADAP chromatogram builder was performed with a minimum group size of 5, a group intensity threshold of  $10^3$ , and an  $m/z$  tolerance of 0.1. The next step was chromatogram deconvolution by applying the Wavelets (ADAP) algorithm, including a signal-to-noise ratio of 10, a minimum feature height of  $10^3$ , and a coefficient of 50. Finally, the detected features among different samples were aligned using the hierarchical aligner algorithm with an  $m/z$  tolerance of 0.1 and an rt tolerance of 0.02. The aligned rt,  $m/z$ , and peak area data were exported as a csv file.

## 3. Results

### 3.1. Analysis of nonvolatile chemicals based on HPLC-QTOF-MS

Due to the occurrence of complicated compounds in the 22 species of raw materials, HPLC-QTOF-MS analysis was performed in both positive and negative modes (Fig. 2 and Fig. S1). To reduce the false positive results from high-intensity noise signals, features with peak areas over 5000 were defined as real features in the samples. In positive scan mode, 3150 features were detected, only 120 features were discovered in Jing Liqueur. That is, over 90% of the chemical components in raw herbal materials were either lost or diluted to levels below the limits of detection after the complex manufacturing process. Similarly, 1816 features were observed in negative scan mode, but only 116 were observed in Jing Liqueur.

### 3.2. Analysis of volatile chemicals based on GC–MS

Because herbal materials with pungent taste, such as CRh, CF, CC, AF, ASR and AOF are rich in volatile components, GC–MS analysis was performed to trace volatile chemicals from the six herbal materials in the final Jing Liqueur (Figs. 3 and S2). As a result, a total of 1542 features were detected: 179 in CRh, 263 in CF, 201 in CC, 319 in AF, 325 in ASR, 246 in AOF and 481 in Jing Liqueur product.

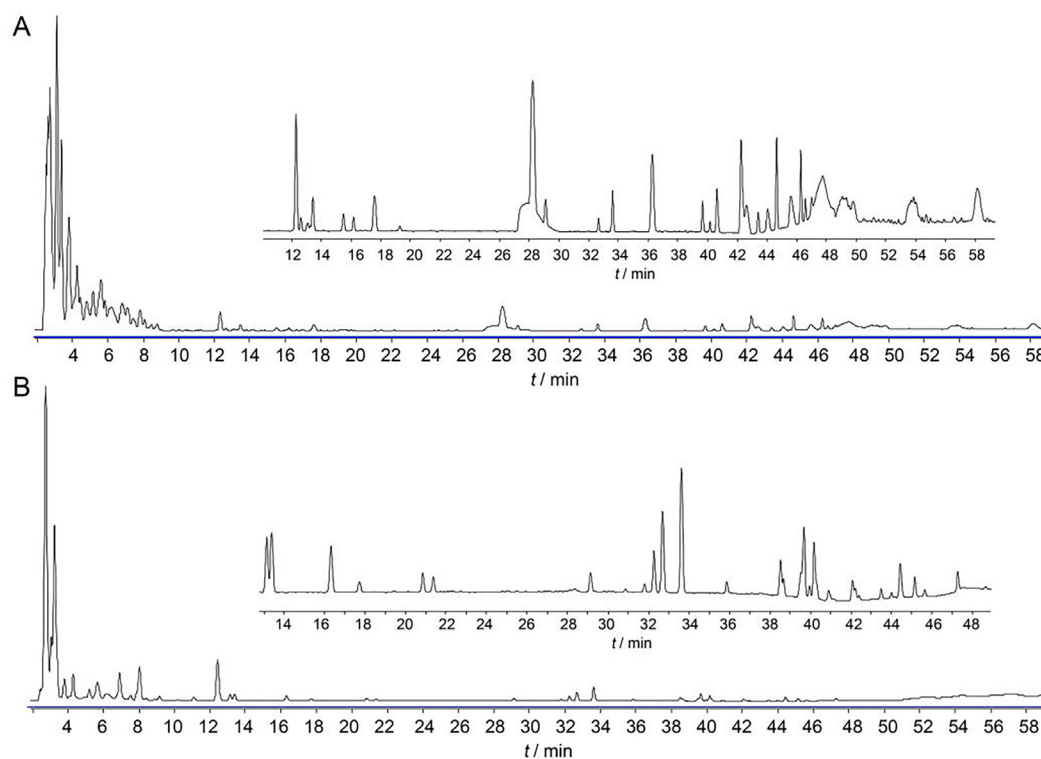


Fig. 2. BPCs of Jing Liqueur analyzed by HPLC-QTOF-MS in positive scan mode (A) and negative scan mode (B).

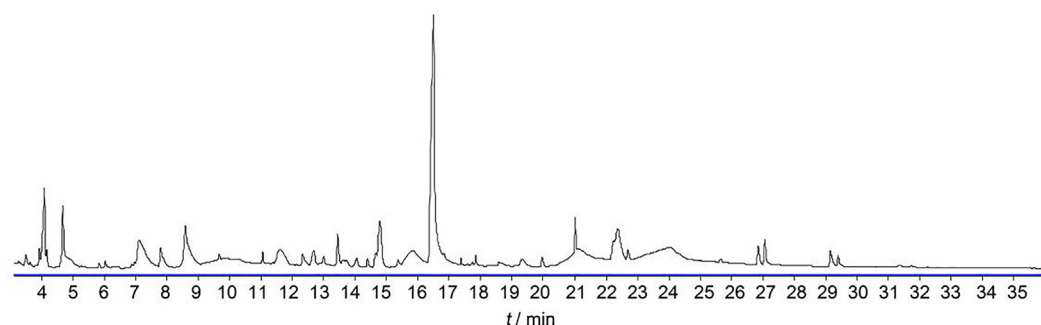


Fig. 3. BPCs of Jing Liqueur analyzed by GC-MS.

### 3.3. Discovery and identification of traceable markers in Jing Liqueur

The detected features were aligned using MZmine 2 software and a csv file containing retention time,  $m/z$  and peak area data of the different samples was generated. Traceable markers should exist both in some raw materials and in the Jing Liqueur product, which was used as a limitation to rapidly screen markers based on the filter function in Excel 2016. As a consequence, 205 traceable chemicals were discovered by HPLC-QTOF-MS analysis: 96 were detected in positive mode and 109 in negative mode; 83 volatile traceable chemicals were discovered by GC-MS analysis.

The 205 nonvolatile features detected by HPLC-QTOF-MS were identified by searching against online databases, including MassBank (Horai et al., 2010), ChemSpider (Pence & Williams, 2010) and ResPect (Sawada et al., 2012) with a relative  $m/z$  tolerance of  $10^{-5}$  and similarity score of 0.7. Due to the limited natural product data in online databases, a custom database consisting of the 50 reference standards was constructed, including their  $rt$ ,  $m/z$  and characteristic product ion data. The structures of other compounds that have been reported in the 22 species of herbal materials were

also involved in the database. Then, 21 features detected in positive mode and 39 in negative mode were annotated. Since there were 8 overlapping features, 52 unique features were finally annotated, including 18 flavonoids, 7 organic acids, 3 triterpenoid saponins, 3 steroids, 2 phenylethanoid glycosides, 2 phenolic acids, 2 alkaloids, 2 glycosides, 2 amino acids, and 11 other types of compounds (Table 3).

Flavonoids are a key group of compounds identified in Jing Liqueur. Of the 18 annotated flavonoids, 11 of them were present as flavonol glycosides and the others were present as isoflavone glycosides. Reference standards of flavonoids are easily obtained; Thus, 13 of the annotated flavonoids were identified by comparing with reference standards and the remaining were tentatively identified based on available databases or publications. For example, feature **17** (417.1164,  $[M + H]^+$ ,  $C_{21}H_{20}O_9$ ), a representative isoflavone glycoside, was identified as puerarin by its reference standard, and its typical fragmentation patterns were shown in Fig. 4A. Two diagnostic product ions at  $m/z$  297.0786 and 267.0678 were observed, and deduced to be generated by cleavage at the glucopyranosyl moiety. Feature **18** (Fig. 4B) with a precursor

**Table 3**  
Identification of nonvolatile chemical markers.

No.	RT/min	Formula	[M+X] <sup>+</sup> /[M-X] <sup>-</sup>	Experimental m/z	Theoretical m/z	Relative error / (×10 <sup>-6</sup> )	Product ions	Identification
1	2.537	C <sub>6</sub> H <sub>14</sub> N <sub>4</sub> O <sub>2</sub>	[M-H] <sup>-</sup>	173.1041	173.1044	1.72	131.0820	Arginine
2	2.602	C <sub>24</sub> H <sub>42</sub> O <sub>21</sub>	[M-H] <sup>-</sup>	665.2150	665.2146	-0.63	485.1512, 383.1195, 179.0561, 89.0242	Nystose #
3	2.668	C <sub>12</sub> H <sub>22</sub> O <sub>11</sub>	[M-H] <sup>-</sup>	341.1092	341.1091	-0.77	113.0237, 89.0237, 71.0133, 59.0134	Morindin
4	2.799	C <sub>5</sub> H <sub>11</sub> NO <sub>2</sub>	[M+H] <sup>+</sup>	118.0857	118.0863	4.74	72.0815, 59.0738, 58.0663	Betaine #
5	2.875	C <sub>10</sub> H <sub>13</sub> N <sub>5</sub> O <sub>4</sub>	[M+H] <sup>+</sup>	268.1036	268.104	1.61	136.0617	Adenosine #
6	3.027	C <sub>11</sub> H <sub>17</sub> NO <sub>8</sub>	[M-H] <sup>-</sup>	290.0881	290.0881	0.14	128.0344	N-(1-deoxy-D-fructos-1-yl) pyroglutamic acid
7	3.113	C <sub>6</sub> H <sub>8</sub> O <sub>7</sub>	[M-H] <sup>-</sup>	191.0192	191.0197	2.74	111.0096, 87.0081, 85.0287	Citric acid #
8	3.877	C <sub>4</sub> H <sub>6</sub> O <sub>4</sub>	[M-H] <sup>-</sup>	117.0184	117.0193	7.90	73.0288, 59.0120	Succinic acid
9	4.312	C <sub>6</sub> H <sub>6</sub> O <sub>3</sub>	[M+H] <sup>+</sup>	127.0385	127.0390	3.73	109.0280, 81.0338, 53.0396	5-hydroxymethylfurfural #
10	4.499	C <sub>16</sub> H <sub>22</sub> O <sub>10</sub>	[M-H] <sup>-</sup>	373.1138	373.1140	0.59	149.0595, 123.0444, 89.0235	Geniposidic acid #
11	5.021	C <sub>14</sub> H <sub>21</sub> NO <sub>4</sub>	[M+H] <sup>+</sup>	268.1530	268.1543	5.00	88.0758	Codonopsine
12	7.384	C <sub>7</sub> H <sub>6</sub> O <sub>4</sub>	[M-H] <sup>-</sup>	153.0183	153.0193	6.70	109.0294	Protocatechuic acid #
13	10.060	C <sub>16</sub> H <sub>18</sub> O <sub>9</sub>	[M+H] <sup>+</sup>	355.1014	355.1024	2.71	163.0364, 145.0272, 135.0432, 117.0336, 89.0358	Chlorogenic acid #
			[M-H] <sup>-</sup>	353.0872	353.0878	1.71	191.0553, 85.0294	
14	10.117	C <sub>9</sub> H <sub>6</sub> O <sub>3</sub>	[M+H] <sup>+</sup>	163.0373	163.0390	10.31	117.0345, 89.0391, 69.0331	Tribenzaldehyde
15	11.006	C <sub>7</sub> H <sub>6</sub> O <sub>3</sub>	[M-H] <sup>-</sup>	137.0236	137.0244	5.92	108.0198	Protocatechuic aldehyde #
16	11.156	C <sub>17</sub> H <sub>20</sub> O <sub>9</sub>	[M-H] <sup>-</sup>	367.1024	367.1035	2.87	193.0500, 134.0365	3-Feroyl-quinic acid
17	12.280	C <sub>21</sub> H <sub>20</sub> O <sub>9</sub>	[M+H] <sup>+</sup>	417.1163	417.1180	4.11	297.0744, 267.0636, 239.0694	Puerarin #
			[M-H] <sup>-</sup>	415.1032	415.1035	0.61	267.0654	
18	13.171	C <sub>26</sub> H <sub>28</sub> O <sub>13</sub>	[M+H] <sup>+</sup>	549.1584	549.1603	3.41	417.1172, 297.0747	Puerarin-7-O-xyloside
			[M-H] <sup>-</sup>	547.1455	547.1457	0.39	295.0605, 267.0655	
19	13.384	C <sub>22</sub> H <sub>22</sub> O <sub>10</sub>	[M-H] <sup>-</sup>	445.1133	445.1140	1.61	282.0526	3'-Methoxypuerarin #
20	14.332	C <sub>16</sub> H <sub>18</sub> O <sub>8</sub>	[M-H] <sup>-</sup>	337.0923	337.0929	1.75	191.0553, 93.0341	trans-5-p-Coumaroylquinic acid/4-O-p-Coumaroylquinic acid
21	16.463	C <sub>35</sub> H <sub>46</sub> O <sub>20</sub>	[M-H] <sup>-</sup>	785.2511	785.2510	-0.17	623.2178, 161.0239	Purpureaside C#
22	16.475	C <sub>7</sub> H <sub>6</sub> O <sub>2</sub>	[M-H] <sup>-</sup>	121.0287	121.0295	6.58	92.0273	Benzoic acid
23	19.796	C <sub>18</sub> H <sub>34</sub> O <sub>11</sub>	[M-H] <sup>-</sup>	425.2050	425.2028	-5.08	263.1492, 161.0441, 101.0239, 71.0141	Sophorosin hexanol Glycoside
24	20.665	C <sub>27</sub> H <sub>30</sub> O <sub>16</sub>	[M-H] <sup>-</sup>	609.1478	609.1461	-2.77	300.0249	Rutin #
25	20.971	C <sub>27</sub> H <sub>44</sub> O <sub>7</sub>	[M+H] <sup>+</sup>	481.3138	481.3160	4.54	445.2922, 371.2185, 165.1264	β-Ecdysone #
			[M+HCOOH-H] <sup>-</sup>	525.3062	525.3069	1.47	479.2998, 319.1917, 159.1014	
26	22.004	C <sub>27</sub> H <sub>44</sub> O <sub>7</sub>	[M+H] <sup>+</sup>	481.3138	481.3160	4.54	445.2922, 371.2185, 165.1264	25R-Inokosterone
27	22.605	C <sub>29</sub> H <sub>36</sub> O <sub>15</sub>	[M-H] <sup>-</sup>	623.1966	623.1981	2.47	461.1694, 161.0235	Acteoside #
28	24.053	C <sub>27</sub> H <sub>29</sub> O <sub>13</sub>	[M-H] <sup>-</sup>	561.1612	561.1614	0.29	309.0755	Formononetin-8-c-aposoy (1, 6)-O-glycoside
29	24.516	C <sub>29</sub> H <sub>36</sub> O <sub>16</sub>	[M-H] <sup>-</sup>	623.1978	623.1981	0.55	461.1675, 161.0254	Isoacteoside #
30	27.277	C <sub>23</sub> H <sub>22</sub> O <sub>10</sub>	[M+H] <sup>+</sup>	459.1265	459.1286	4.53	255.0632	Acetylaidzin
31	27.778	C <sub>9</sub> H <sub>16</sub> O <sub>4</sub>	[M-H] <sup>-</sup>	187.0968	187.0976	4.16	125.0972	Azelaic acid #
32	28.564	C <sub>18</sub> H <sub>16</sub> O <sub>8</sub>	[M-H] <sup>-</sup>	359.0756	359.0772	4.56	161.0245	Rosmarinic acid #
33	29.130	C <sub>26</sub> H <sub>22</sub> O <sub>10</sub>	[M-H] <sup>-</sup>	493.1129	493.1140	2.27	295.0628, 185.0234, 109.0289	Salvianolic acid A#
34	29.262	C <sub>24</sub> H <sub>26</sub> O <sub>10</sub>	[M+H] <sup>+</sup>	475.1589	475.1599	2.05	475.1756, 107.0514	Puerarinoid D
35	30.380	C <sub>16</sub> H <sub>12</sub> O <sub>4</sub>	[M+H] <sup>+</sup>	269.0805	269.0808	1.25	253.0487, 226.0616, 197.0591, 118.0399	Formononetin #
36	31.440	C <sub>39</sub> H <sub>50</sub> O <sub>20</sub>	[M+H] <sup>+</sup>	839.2930	839.2968	4.56	531.1808, 369.1310, 313.0640, 85.0276	Epimedin A1#
37	32.132	C <sub>33</sub> H <sub>40</sub> O <sub>15</sub>	[M-H] <sup>-</sup>	675.2293	675.2294	0.21	366.1103, 351.0852, 323.0872	Sagittatoside A #
38	32.225	C <sub>15</sub> H <sub>10</sub> O <sub>4</sub>	[M+H] <sup>+</sup>	255.0641	255.0652	4.27	152.0608, 91.0532	Daidzein #
			[M-H] <sup>-</sup>	253.0493	253.0506	5.24	223.0390, 195.0341, 132.0218, 91.0179	
39	32.454	C <sub>32</sub> H <sub>38</sub> O <sub>14</sub>	[M-H] <sup>-</sup>	645.2178	645.2189	1.67	366.1099, 351.0863, 323.0904	Sagittatoside B #
40	32.566	C <sub>38</sub> H <sub>48</sub> O <sub>19</sub>	[M+H] <sup>+</sup>	809.2834	809.2863	3.53	677.2334, 531.1820, 369.1248	Epimedin B #
			[M+HCOOH-H] <sup>-</sup>	853.2745	853.2772	3.32	645.2209	
41	32.855	C <sub>33</sub> H <sub>40</sub> O <sub>14</sub>	[M-H] <sup>-</sup>	659.2339	659.2345	0.95	366.1098, 351.0864, 323.0885	2'-O-Rhamnosyl icariin II#
42	32.969	C <sub>39</sub> H <sub>50</sub> O <sub>19</sub>	[M+H] <sup>+</sup>	823.2992	823.3019	3.29	677.2395, 531.1808, 369.1315	Epimedin C #
			[M+HCOOH-H] <sup>-</sup>	867.2909	867.2928	2.35	659.2337, 367.1164	
43	33.744	C <sub>33</sub> H <sub>40</sub> O <sub>15</sub>	[M+H] <sup>+</sup>	677.2418	677.2440	3.25	531.1831, 369.1308	Icariin #
			[M+HCOOH-H] <sup>-</sup>	721.2325	721.2349	3.58	513.1756, 367.1182	
44	34.942	C <sub>39</sub> H <sub>48</sub> O <sub>19</sub>	[M+H] <sup>+</sup>	821.2838	821.2863	2.99	532.1850, 369.1328, 313.0671, 211.0577, 145.0472, 129.0536, 99.0450, 85.0276	Epimedin A1 derivative
			[M+HCOOH-H] <sup>-</sup>	865.2771	865.2772	0.10	751.9058, 659.2384, 513.1701, 367.1136	
45	36.439	C <sub>11</sub> H <sub>12</sub> O <sub>4</sub>	[M-H] <sup>-</sup>	207.0654	207.0663	4.24	133.0288	Ethyl 2,4-dihydroxycinnamate
46	38.771	C <sub>53</sub> H <sub>82</sub> O <sub>25</sub>	[M-H] <sup>-</sup>	1117.5060	1117.5072	1.11	1117.5088, 997.5003, 955.4873, 793.4506	Achyranthosides D
47	39.578	C <sub>18</sub> H <sub>34</sub> O <sub>5</sub>	[M-H] <sup>-</sup>	329.2324	329.2333	2.87	229.1415, 211.1339, 171.1012	9, 12, 13-Trihydroxy-10-octadecenoic acid

Table 3 (continued)

No.	RT/min	Formula	$[M+X]^+/[M-X]^-$	Experimental $m/z$	Theoretical $m/z$	Relative error / ( $\times 10^{-6}$ )	Product ions	Identification
48	39.811	$C_{37}H_{46}O_{19}$	$[M-H]^-$	793.4361	—	—	793.4337, 631.3732	Epimedeside E
49	42.175	$C_{42}H_{66}O_{14}$	$[M-H]^-$	793.4361	793.4380	2.37	631.3886, 569.3873, 113.0246, 75.0289	Chikusetsusaponin IVA#
50	42.228	$C_{27}H_{30}O_{10}$	$[M+H]^+$	515.1893	515.1912	3.64	369.1317, 313.0692	Icarisid II
51	42.396	$C_{48}H_{76}O_{19}$	$[M-H]^-$	955.4527	955.4544	1.80	835.4529, 793.4492, 569.3802	Ginsenoside Ro
52	44.707	$C_{27}H_{42}O_3$	$[M+H]^+$	415.2101	—	—	119.0851	Diosgenin #

#Represents that this compound was confirmed by comparing with reference standard.

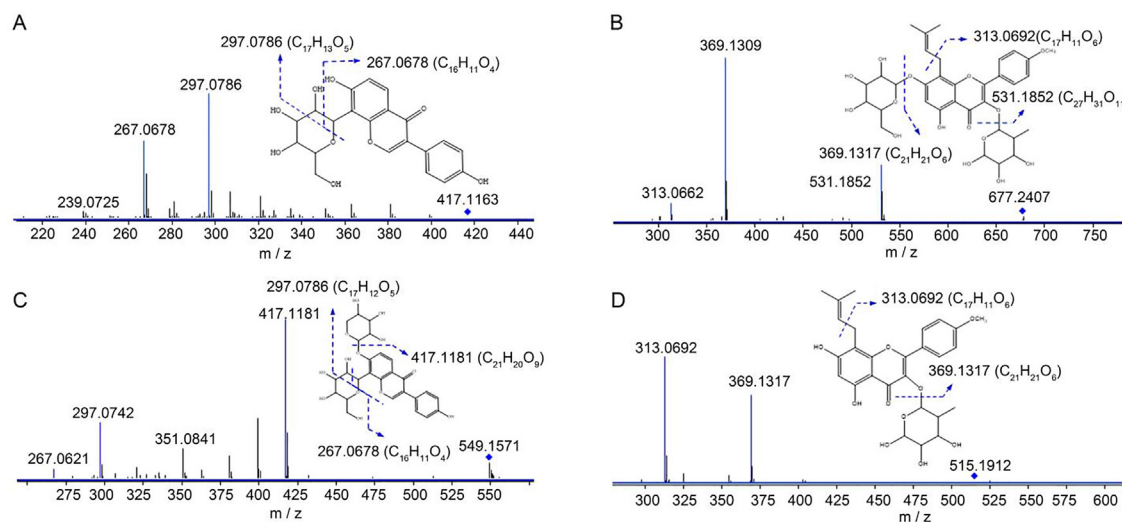


Fig. 4. Characteristic fragmentation patterns of typical flavonoids.

ion at  $m/z$  549.1584 ( $[M+H]^+$ ,  $C_{26}H_{28}O_{13}$ ) also generated product ions at  $m/z$  297.0743 and 267.0657, which indicated its structural similarity with feature 17.

The product ion at  $m/z$  417.1146 ( $[M+H]^+$  -  $C_5H_8O_4$ ) demonstrated that feature 18 was a conjugated O-xyloside. Combined with the results obtained by database searching, feature 18 was finally annotated as puerarin-7-O-xyloside.

Feature 43 (677.2418,  $[M+H]^+$ ,  $C_{33}H_{41}O_{15}$ ) was a representative flavonol glycoside and was identified as icariin by the corresponding reference standard. As shown in Fig. 4C, the characteristic product ions of  $m/z$  531.1852 ( $[M+H]^+$  -  $C_6H_{10}O_4$ ), 369.1309 ( $[M+H]^+$  -  $C_6H_{10}O_4$  -  $C_6H_{10}O_5$ ) and 313.0662 ( $[M+H]^+$  -  $C_6H_{10}O_4$  -  $C_6H_{10}O_5$  -  $C_4H_{10}$ ) were generated by the neutral loss of glycosides or the cleavage of the side chain of isoamylene. Feature 50 produced very similar product ions and its characteristic fragmentation patterns are shown in Fig. 4D. The product ions at  $m/z$  369.1317 and 313.0692 demonstrated the similar chemical scaffold of feature 50 to that of feature 43. The mass difference of 162 ( $C_6H_{10}O_5$ ) indicated that feature 43 could be derived from feature 50 by the loss of glucose. Thus, feature 43 was finally annotated as icariside II, which has been reported to be an important component in *Epimedium Folium* (EF) (Choi et al., 2008).

In the same way, the other nonvolatile components were annotated. Their characteristic product ions are summarized in Table 2.

Among the 83 unique volatile features in Jing Liqueur detected by GC-MS, 23 could be detected in CR, 24 detected in CF, 27 detected in CC, 27 detected in AF, 22 detected in AOF and 57 detected in ASR. The compounds were identified by comparing their mass spectra with those in the instrument's National Institute

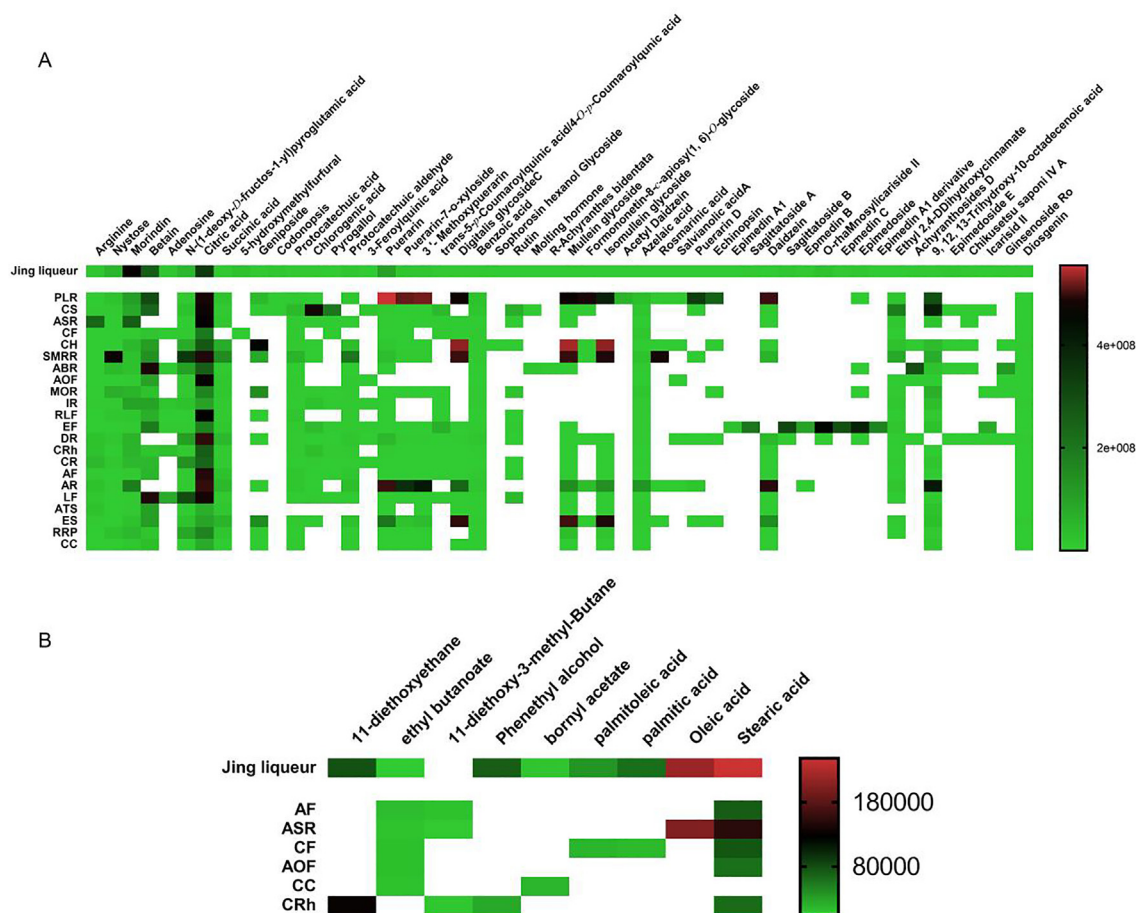
of Standards and Technology (NIST) library (Stein, 1995). A score value over 700 was considered necessary for a good match (Stein, 2011). To avoid false positive markers resulting from automatic data processing, the selected traceable features were confirmed by manual extraction of the raw data in MassHunter software. And then eight annotated volatile compounds were defined as traceable markers. The detailed information is shown in Table 4.

### 3.4. Generation of chemical marker fingerprint

As shown in Fig. 5, among the 61 annotated markers, some were exclusively present in certain herbal species, and the others coexisted in all 22 species of raw materials. The chemical profile of Jing Liqueur was integrated by those of all 22 herbal species. The specificity of chemical markers was evaluated by their distribution ratio among the 22 raw herbal species. Theoretically, specific markers of an herbal material contributed the most to the relative intensity of the MS profile of Jing Liqueur. By limiting the distribution ratio to 50%, 29 chemical markers derived from all the herbal species except *Allii Tuberosi Semen* (ATS) were selected. With regard to ATS, the contribution intensity of the 52 features was evaluated by the ratio of the intensity of ATS features to the total intensity of features of the 22 herbal species. Feature 1 contributed the most to the corresponding component in Jing Liqueur, which was added to the marker group. The chemical profiles of the 30 markers were extracted from the raw LC-MS data of Jing Liqueur to generate the chemical marker fingerprint, which could be utilized to guide its authentication (Fig. 6).

**Table 4**  
Identification of volatile chemical markers by searching against NIST database.

No	RT/min	Molecular weight	Formula	Match score	Identification
1	3.906	118	C <sub>6</sub> H <sub>14</sub> O <sub>2</sub>	887	1,1-Diethoxyethane
2	5.942	116	C <sub>6</sub> H <sub>12</sub> O <sub>2</sub>	772	Ethyl butanoate
3	10.400	160	C <sub>9</sub> H <sub>20</sub> O <sub>2</sub>	721	1,1-Diethoxy-3-methyl-Butane
4	15.355	174	C <sub>8</sub> H <sub>14</sub> O <sub>4</sub>	692	Phenethyl alcohol
5	17.370	196	C <sub>12</sub> H <sub>20</sub> O <sub>2</sub>	755	Bornyl acetate
6	26.857	254	C <sub>16</sub> H <sub>30</sub> O <sub>2</sub>	899	Palmitoleic acid
7	27.080	256	C <sub>16</sub> H <sub>32</sub> O <sub>2</sub>	864	Palmitic acid
8	29.163	282	C <sub>18</sub> H <sub>34</sub> O <sub>2</sub>	903	Oleic acid
9	29.409	284	C <sub>18</sub> H <sub>36</sub> O <sub>2</sub>	845	Stearic acid



**Fig. 5.** Visualization of traceable markers detected by HPLC-QTOF-MS (A) and GC-MS (B) from raw materials to Jing Liqueur.

### 4. Discussion

The quality evaluation of herbal products has troubled scientists and researchers. Quality markers with transitivity and traceability, proposed by Liu et al. (2018), have been comprehensively explored to control the quality of herbal products. However, the discovery of quality markers was very tedious, which included the characterization of chemical ingredients, biosynthetic pathway analysis, bioactivity evaluation and conformation of quality markers (Li et al., 2019). The transitivity and traceability of chemical markers are the core concepts of herbal product authentication. Based on this view, we developed a novel strategy to rapidly discover traceable markers of herbal products.

MZmine 2 software is usually applied for automatic mass detection, chromatogram building and deconvolution before batch identification of complex mixtures (Korf, Jeck, Schmid, Helmer, & Hayen, 2019; Pluskal et al., 2010). The alignment function in

MZmine 2 software was designed to align the features among different sample runs. Traceable markers exist both in certain materials and in the final product. Therefore, by aligning the chemical features among herbal materials and the product using MZmine 2, traceable chemical markers were rapidly discovered.

In this work, MZmine 2 was first used to discover traceable markers of herbal products. This approach is very convenient. Once the raw data of the materials and products were imported, mass detection, chromatogram building, deconvolution and alignment could be completed in a few minutes. The aligned features were exported into a csv file. The traceable features could be rapidly screened using the filter function in Excel. In addition, the online and custom database search functions built in MZmine 2 software lead to an expedient identify cation of chemical markers.

There are still some points that need to be improved. First, the identification of traceable markers is hampered by the limitation of reference standards and available databases, which is a common

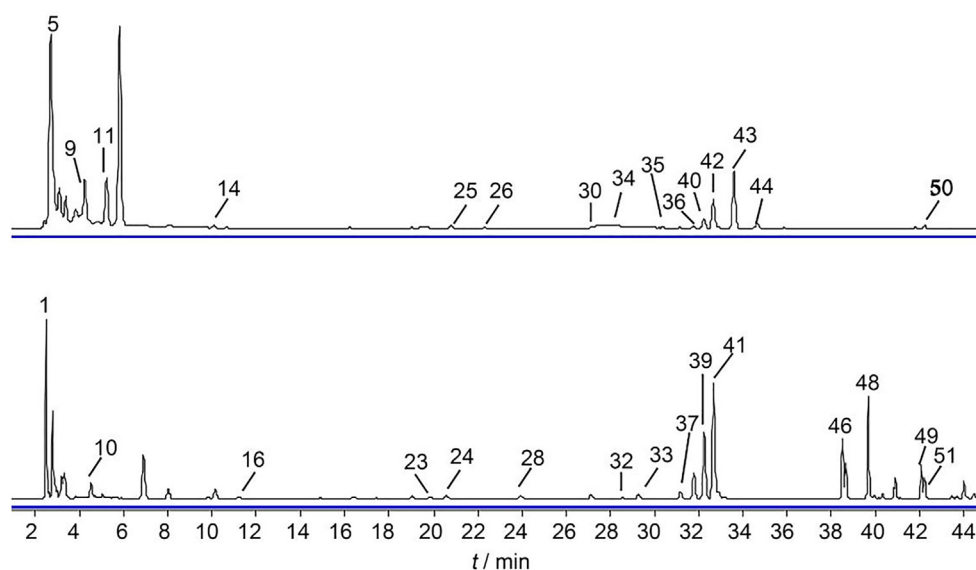


Fig. 6. Chemical fingerprints for authentication of Jing Liqueur.

issue in this field. With the increasing knowledge of natural products and the expanding databases, this issue will be resolved. Second, the chemical marker fingerprint of Jing Liqueur must be validated using additional commercial samples. Once the product fingerprint is confirmed, it can be embedded into data analysis software. Coupled with the fingerprint similarity evaluation system, real-time authentication of Jing Liqueur can be realized.

## 5. Conclusion

In this work, a novel strategy for the rapid discovery of traceable markers of herbal products was developed. Jing Liqueur, produced from 22 species of herbal materials, was taken as a case study. A total of 61 traceable markers were rapidly discovered and annotated. The chemical fingerprint of traceable markers was also generated, and could be used to authenticate Jing Liqueur products.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.chmed.2021.05.004>.

## References

- Aparicio-Soto, M., Sánchez-Hidalgo, M., Rosillo, M. Á., Castejón, M. L., & Alarcón-de-la-Lastra, C. (2016). Extra virgin olive oil: A key functional food for prevention of immune-inflammatory diseases. *Food and Function*, 7(11), 4492–4505.
- Cañigual, S., Tschopp, R., Ambrosetti, L., Vignutelli, A., Scaglione, F., & Petrini, O. (2008). The development of herbal medicinal products: Quality, safety, and efficacy as key factors. *Pharmaceutical Medicine*, 22(2), 107–118.
- Choi, H. J., Eun, J. S., Kim, D. K., Li, R. H., Shin, T. Y., Park, H., et al. (2008). Icariside II from *Epimedium koreanum* inhibits hypoxia-inducible factor-1 $\alpha$  in human osteosarcoma cells. *European Journal of Pharmacology*, 579(1–3), 58–65.
- Fan, X. H., Cheng, Y. Y., Ye, Z. L., Lin, R. C., & Qian, Z. Z. (2006). Multiple chromatographic fingerprinting and its application to the quality control of herbal medicines. *Analytica Chimica Acta*, 555(2), 217–224.
- Feng, S., Shan, Y., Lu, S., Liu, Y., & He, G. (2013). The anti-inflammatory effect of moderate drinking. *Liquor-Making Science and Technology*, 229, 121–124.

- Gomathi, D., Kalaiselvi, M., Ravikumar, G., Devaki, K., & Uma, C. (2015). GC-MS analysis of bioactive compounds from the whole plant ethanolic extract of *Evolvulus alsinoides* (L.) L. *Journal of Food Science and Technology*, 52(2), 1212–1217.
- Guo, B. J., Bian, Z. X., Qiu, H. C., Wang, Y. T., & Wang, Y. (2017). Biological and clinical implications of herbal medicine and natural products for the treatment of inflammatory bowel disease. *Annals of the New York Academy of Sciences*, 1401(1), 37–48.
- Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., et al. (2010). MassBank: A public repository for sharing mass spectral data for life sciences. *Journal of Mass Spectrometry*, 45(7), 703–714.
- Korf, A., Jeck, V., Schmid, R., Helmer, P. O., & Hayen, H. (2019). Lipid species annotation at double bond position level with custom databases by extension of the MZmine 2 open-source software package. *Analytical Chemistry*, 91(8), 5098–5105.
- Lai, K. M., Cheng, Y. Y., & Tsai, T. H. (2015). Integrated LC-MS/MS analytical systems and physical inspection for the analysis of a botanical herbal preparation. *Molecules*, 20(6), 10641–10656.
- Li, Y., Xie, Y., He, Y., Hou, W., Liao, M., & Liu, C. (2019). Quality markers of traditional Chinese medicine: Concept, progress, and perspective. *Engineering*, 5(5), 888–894.
- Liang, Y. J., Lin, Y. T., Chen, C. W., Lin, C. W., Chao, K. M., Pan, W. H., et al. (2016). SMART: Statistical metabolomics analysis-an R tool. *Analytical Chemistry*, 88(12), 6334–6341.
- Liang, Y., Xie, P., & Chan, K. (2004). Quality control of herbal medicines. *Journal of Chromatography B*, 812(1–2), 53–70.
- Liu, C., Guo, D. A., & Liu, L. (2018). Quality transitivity and traceability system of herbal medicine products based on quality markers. *Phytomedicine*, 44, 247–257.
- Liu, F. J., Jiang, Y., Li, P., Liu, Y. D., Yao, Z. P., Xin, G. Z., et al. (2020). Untargeted metabolomics coupled with chemometric analysis reveals species-specific steroidal alkaloids for the authentication of medicinal *Fritillariae Bulbus* and relevant products. *Journal of Chromatography A*, 1612, 460630. <https://doi.org/10.1016/j.chroma.2019.460630>.
- Müller, C., & Bracher, F. (2015). Determination by GC-IT/MS of phytosterols in herbal medicinal products for the treatment of lower urinary tract symptoms and food products marketed in Europe. *Planta Medica*, 81(07), 613–620.
- Olivon, F., Grellier, G., Roussi, F., Litaudon, M., & Touboul, D. (2017). MZmine 2 data-preprocessing to enhance molecular networking reliability. *Analytical Chemistry*, 89(18), 7836–7840.
- Pence, H. E., & Williams, A. (2010). Chemspider: An online chemical information resource. *Journal of Chemical Education*, 87(11), 1123–1124.
- Pluskal, T., Castillo, S., Villar-Briones, A., & Orešič, M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*, 11, 395.
- Röst, H. L., Sachsenberg, T., Aiche, S., Bielow, C., Weisser, H., Aicheler, F., et al. (2016). OpenMS: A flexible open-source software platform for mass spectrometry data analysis. *Nature Methods*, 13(9), 741–748.
- Sawada, Y., Nakabayashi, R., Yamada, Y., Suzuki, M., Sato, M., Sakata, A., et al. (2012). RIKEN tandem mass spectral database (ReSpect) for phytochemicals: A plant-specific MS/MS-based data resource and database. *Phytochemistry*, 82, 38–45.
- Shan, Y., Zhou, H., Chen, M., Chen, K., Liu, Y., & Wang, L. (2018). Study on anti-fatigue, regulating immunity and enhancing sexual function of Chinese Jing Liqueur. *Chinese Traditional Patent Medicine*, 40, 1600–1603.



- Stein, S. E. (1995). Chemical substructure identification by mass spectral library searching. *Journal of the American Society for Mass Spectrometry*, 6(8), 644–655.
- Stein, S. E. (2011). *National Institute and Standards and Technology (NIST) Mass Spectral Search Program, Version (2.0g)*. <https://chemdata.nist.gov/mass-spc/ms-search/>.
- Tautenhahn, R., Patti, G. J., Rinehart, D., & Siuzdak, G. (2012). XCMS online: A web-based platform to process untargeted metabolomic data. *Analytical Chemistry*, 84(11), 5035–5039.
- Wang, F., Wang, B. o., Wang, L., Xiong, Z. Y., Gao, W., Li, P., et al. (2017). Discovery of discriminatory quality control markers for Chinese herbal medicines and related processed products by combination of chromatographic analysis and chemometrics methods: *Radix Scutellariae* as a case study. *Journal of Pharmaceutical and Biomedical Analysis*, 138, 70–79.
- Weitzel, K. M. (2011). Bond-dissociation energies of cations—pushing the limits to quantum state resolution. *Mass Spectrometry Review*, 30(2), 221–235.
- Yu, H. L., Wen, C. J., & Su, H. Q. (2013). Establishment of traceable management information platform for quality safety of *Panax Ginseng*. *Chinese Traditional and Herbal Drugs*, 44(24), 3566–3574.
- Zhao, C., Liu, H., Miao, P., Wang, H., Yu, H., Wang, C., et al. (2019). A strategy for selecting “Q-Markers” of Chinese medical preparation via components transfer process analysis with application to the quality control of Shengmai Injection. *Molecules*, 24(9), 1811. <https://doi.org/10.3390/molecules24091811>.