Characterization, quantitation and similarity evaluation of *Codonopsis lanceolata* from different regions in China by HPLC-Q-TQF-MS and chemometrics

Yuanyuan Xia, Fulei Liu, Feng Feng, Wenyuan Liu

**A R T I C L E   I N F O**

**Keywords:**
- *Codonopsis lanceolata*
- Characterization
- Quality evaluation
- Food analysis
- Food composition
- Similarity analysis
- Quadrupole time-of-flight mass spectrometry
- Hierarchical clustering analysis
- Principal component analysis

**A B S T R A C T**

*Codonopsis lanceolata*, a valuable medicinal plant, has been cultivated in various areas, but lacks effective comparison and evaluation methods. An approach was proposed on qualitative and quantitative analysis of cultured *C. lanceolata* by high-performance liquid chromatography coupled with tandem mass spectrometry in this work. Samples from different regions were compared by chemometric analysis. As a result, twenty-two peaks, including 7 triterpenoid saponins, 4 phenylpropanoids, 2 polyacetylenes, 3 amino acids, 2 nucleosides, 1 organic acid and three other kinds of compounds, were assigned based either on reference standards or literature. Among them, 9 compounds were reported for the first time in *C. lanceolata*. In addition, 5 constituents were quantified simultaneously in 18 samples of *C. lanceolata*. Amino acids, phenylpropanoids and polyacetylenes could be undoubtedly determined, to comprehensively evaluate the quality of *C. lanceolata*. The chemometric methods, including similarity analysis, hierarchical clustering analysis and principal component analysis, were introduced to classify the samples from different regions of China. This work was expected to obtain a comprehensive understanding of characterization, quantitation and similarity evaluation of cultured *C. lanceolata* from different regions in China, which could provide a basis for the further study of this herb.

1. Introduction

*Codonopsis lanceolata* (Sieb et Zucc) Bentham et Hooker (Campanulaceae) is a valuable medicinal plant and has been widely used in the fields of food and traditional Chinese medicine (TCM). Dried root of *C. lanceolata* (also called deodeok) is becoming popular as a special herb in some Asian countries and has been used to treat bronchitis, rheumatism and inflammation in China (Nanjing University of Chinese Medicine, 2006). A supplement containing its root extract has been reported to alleviate partial androgen deficiency of the aging male (PADAM)-like symptoms (Ushijima et al., 2007). In Korea, *C. lanceolata* powder has been added to traditional *kochujang* to improve its quality (Hong et al., 2010; Sung et al., 2011). Nowadays, due to lack of wild resources, *C. lanceolata* has been cultivated artificially in various areas in China.

Certain phytochemical and pharmacological studies have been published for *C. lanceolata*. To be specific, modern pharmacological research indicates that it has various pharmacological activities, including anti-tumor (Li et al., 2015), antiobesity (Lee et al., 2014), antioxidant, antimicrobial and antitumorigenic activities (He et al., 2010), cognitive-enhancing (Weon et al., 2014a) and neuroprotective effects (Weon et al., 2014b). Previous phytochemical investigation has reported the presence of benzofuranylpropanoids (lanceolunes A–C) (Hu et al., 2012), tangshenosides (Ren et al., 2013) and triterpene glycosides (Ushijima et al., 2008) in root of *C. lanceolata*.

Additionally, some researches about extraction optimization, qualitative or quantitative analyses has been reported for the study of *C. lanceolata*. For instance, the extraction methods for total saponins (Zhao et al., 2012), total phenols, flavonoids and phenolic acids (He et al., 2011) were investigated respectively. Ichikawa et al. (2008, 2009), established an HPLC–MS method for qualitative and quantitative analyses of triterpenoid saponins in *C. lanceolata*. However, most of the

---

* Abbreviations: HCA, hierarchical clustering analysis; HPLC-DAD-Q-TOF-MS, high-performance liquid chromatography-diode array detector-quadrupole-time of flight-mass spectrometry; PCA, principal component analysis; SA, similarity analysis; TCMs, traditional Chinese medicines
* Corresponding author.
* E-mail address: liuwenyuan8506@163.com (W. Liu).

http://dx.doi.org/10.1016/j.jfca.2017.05.009
Received 6 January 2017; Received in revised form 10 May 2017; Accepted 25 May 2017
Available online 26 May 2017
0889-1575/ © 2017 Elsevier Inc. All rights reserved.
studies were merely restricted to the investigation of partial components rather than overall constituents. In addition, literature of systematic determination and origin discrimination of *C. lanceolata* root is still limited.

TCMs consist of a variety of compounds, which form the basis of their pharmacological or biological activities. The chemical compositions of TCMs are dependent on geographical origin, as well as different cultivation, harvest and storage conditions (Drašar and Moravcová, 2004). Furthermore, chemical constituents and herbal authenticity of TCMs are critical to pharmacological research and clinical application. Counterfeit or adulterated herbs in the commercial market are producing an unpredicted but considerable risk to public safety. Therefore, it is necessary to investigate quantitative and chemical fingerprinting analysis on cultured *C. lanceolata* from various origins in China.

Analysis of merely several bioactive constituents or markers of TCMs is insufficient. Fingerprint analysis is a comprehensive and effective technique to evaluate the authenticity and quality of TCMs by the State Food and Drug Administration of China and the World Health Organization (Gong et al., 2003). Chemometrics is a typical statistical tool, which could overview complex data obtained from TCMs chromatographic profiles. Similarity analysis (SA), hierarchical clustering analysis (HCA) and principal component analysis (PCA) are widely used methods to monitor groups of herbal samples (Yudthavorasit et al., 2014).

In this paper, characterization, quantitation and similarity evaluation of 18 batches of cultured *C. lanceolata* from different regions in China were investigated by high-performance liquid chromatography coupled with quadrupole time-of-flight mass spectrometry (HPLC-Q-TOF-MS) and chemometrics. Firstly, an HPLC-Q-TOF-MS method was developed for quantitative and qualitative analysis of *C. lanceolata*. Then chemometrics including SA, HCA and PCA were utilized for chemometric fingerprint and classification of samples from different regions of China. This study provided comprehensive understanding of chemical profile, quantitative analysis and similarity evaluation of cultured *C. lanceolata* from different regions, which could provide a basis for the quality evaluation of this herb.

2. Experimental

2.1. Materials and reagents

HPLC-grade acetonitrile and methanol were purchased from Tedia (Fairfield, OH). Other reagents were of analytical grade and provided from Nanjing Chemical Reagent Corp. (Nanjing, China). High purity water for HPLC was purified using a Milli Q-Plus system (Millipore, Bedford, MA).

Standards of arginine (1), guanosine (2), adenosine (3), phenylalanine (4), tryptophan (5), chlorogenic acid (6) syringin (7) and lobe-tyolin (14) were purchased from Shanghai Yuanye Bio-Technology Co. (Shanghai, China). The purities of these standards were all over 98%, with adenosine being over 99%.

Eighteen batches of dried roots of *C. lanceolata* (S1-S18) from different geographical origins in China are summarized in Table 1. All the samples were authenticated in our laboratory by Professor Feng Feng. The voucher specimens were deposited at the Herbarium of Medicinal Plants of China Pharmaceutical University.

2.2. Preparation of standard solutions and sample solutions

Each accurately weighed reference compound was dissolved and diluted with 70% methanol (v/v) to obtain a series of stock standard solutions. A mixed stock standard solution containing 5 standards was prepared by adding and diluting each stock standard solution with 70% methanol (v/v) to a series of working standard solutions.

One gram of ground powder of *C. lanceolate*, which had been passed through a 40 mesh screen, was accurately weighed and extracted ultrasonically with 10 mL 70% (v/v) methanol at room temperature for 60 min. The extracted solution was then centrifuged at 16,000 rpm for 5 min and the supernatant solution was filtered through a 0.45-μm membrane before HPLC analysis.

2.3. Chromatographic conditions

HPLC analysis was acquired on a Shimadzu series 2010 HPLC instrument (Shimadzu Corp., Kyoto, Japan) equipped with a quaternary pump, an online vacuum degasser, a column compartment, an autosampler and a UV detector. All separation was conducted on a Hedera ODS-2 C18 column (250 mm × 4.6 mm, 5 μm) from Jiangsu Hanbon Sci. & Tech. Corp. (Huaian, China) in gradient elution mode with a mobile phase consisted of 0.1 mL L⁻¹ formic acid in acetonitrile (A) and 0.1 mL L⁻¹ formic acid in water (B). The flow rate was maintained at 0.8 mL min⁻¹. The gradient elution program for *C. lanceolata* was carried out as follows: 0–20 min, isocratic 7% A; 20–25 min, 7%–10% A; 25–27 min, 10–12% A; 27–45 min, 12–16% A; 45–80 min, 16–37% A; 80–85 min, 37–42% A. The column temperature was maintained at 30 °C, the injection volume was 20 μL and the UV detector was set at 218 nm. The data were processed with Shimadzu 2.0 ChemStation software (Shimadzu Corp., Kyoto, Japan).

2.4. HPLC–Q–TOF–MS conditions

HPLC-Q-TOF-MS analysis for quantitative analysis was performed on an Agilent 6520 Q-TOF mass spectrometer equipped with a diode array detector (DAD) and electrospray interface (ESI) (Agilent Technologies, Santa Clara, CA). The MS system was operated both in positive and negative ionization modes with the mass spectra scan range set at m/z 50–1700. The typical ionization source conditions were as follows: nebulizer pressure 40 psi, drying gas temperature 325 °C, drying gas flow rate 8.0 L min⁻¹, sheath gas temperature 400 °C, sheath gas flow rate 10 L min⁻¹, fragmentor voltage 100 V, capillary voltage 4000 V (+ ESI)/3500 V (− ESI), skimmer voltage 65 V and collision energy 25–60 eV. Data acquisition and analysis were processed by Agilent Mass Hunter Workstation Data Acquisition Software Version B.04.00 (Agilent Technologies, Santa Clara, CA).

2.5. Data acquisition and processing

The software “Similarity Evaluation System for Chromatographic Fingerprint of TCM” was published by Chinese Pharmacopoeia Committee (Version 2004A) (Beijing, China) and applied in the SA of chromatographic patterns. A simulative mean chromatogram as a representative standard fingerprint chromatogram was generated (R, Fig. 1). Ten peaks (4, 5, 6, unidentified 1, unidentified 2, 10, 12, 14, 16 + 17, 20 + 21) existing in all 18 samples of *C. lanceolata* were selected and designated as “characteristic common peaks” based on their identification or tentative identification and larger peak area (Fig. 1). Hence, these 10 peaks were selected as the variables and 18 samples of *C. lanceolata* (S1–18) from different geographical origins were analyzed by HCA and PCA using SPSS software (SPSS 22; SPSS, Chicago, IL). Quantitation of 5 constituents was performed in triplicate and the results were expressed as means. Significant differences of means at p < 0.05 were determined by analysis of variance (ANOVA) using SPSS.

3. Results and conclusion

3.1. Optimization of sample preparation and chromatographic conditions

To achieve satisfactory extraction efficiency of *C. lanceolata*, variables involved in the procedure were as follows: extraction methods (ultrasonic and refluxing), extracting solvent (40%, 70% and 100% methanol, v/v), extraction time (30, 60 and 90 min) and ratio of drug to...
solvent (10, 15 and 20 mL g\(^{-1}\)) were optimized. The optimized conditions for extraction were to extract 1.0 g powder with 10 mL of 70% (v/v) methanol in an ultrasonic device for 60 min.

To achieve better chromatographic behavior of C. lanceolata, various HPLC parameters were also investigated. Mobile phase (methanol-water, acetonitrile-water, acetonitrile-formic acid and acetonitrile (with formic acid)-water (with formic acid)), detection wavelength (207, 218, 254, 280 and 320 nm), column temperature (25 °C, 30 °C and 35 °C) and flow rate (0.7, 0.8 and 0.9 mL min\(^{-1}\)) were optimized. The detection wavelength of 218 nm, column temperature of 30 °C and flow rate of 0.8 mL min\(^{-1}\) provided stronger UV absorption, a sufficiently large number of detectable peaks and better peak resolution on the HPLC.

Table 1

Samples of Codonopsis lanceolata: geographical origin, acquisition time, content and similarity.

<table>
<thead>
<tr>
<th>sample no.</th>
<th>origin</th>
<th>acquisition time</th>
<th>content (mg g(^{-1}))</th>
<th>similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>phenylalanine</td>
<td>tryptophan</td>
</tr>
<tr>
<td>S1</td>
<td>Shandong</td>
<td>October 2015</td>
<td>0.490</td>
<td>0.474</td>
</tr>
<tr>
<td>S2</td>
<td>Anhui</td>
<td>July 2015</td>
<td>0.190</td>
<td>0.437</td>
</tr>
<tr>
<td>S3</td>
<td>Anhui</td>
<td>October 2015</td>
<td>0.092</td>
<td>0.203</td>
</tr>
<tr>
<td>S4</td>
<td>Shandong</td>
<td>October 2015</td>
<td>0.116</td>
<td>0.318</td>
</tr>
<tr>
<td>S5</td>
<td>Guangxi</td>
<td>October 2015</td>
<td>0.030</td>
<td>0.380</td>
</tr>
<tr>
<td>S6</td>
<td>Guangxi</td>
<td>October 2015</td>
<td>0.127</td>
<td>0.462</td>
</tr>
<tr>
<td>S7</td>
<td>Hunan</td>
<td>October 2015</td>
<td>0.247</td>
<td>0.325</td>
</tr>
<tr>
<td>S8</td>
<td>Hunan</td>
<td>October 2015</td>
<td>0.193</td>
<td>0.460</td>
</tr>
<tr>
<td>S9</td>
<td>Hunan</td>
<td>August 2015</td>
<td>0.128</td>
<td>0.100</td>
</tr>
<tr>
<td>S10</td>
<td>Jiangxi</td>
<td>August 2015</td>
<td>0.204</td>
<td>0.381</td>
</tr>
<tr>
<td>S11</td>
<td>Jiangxi</td>
<td>October 2015</td>
<td>0.140</td>
<td>0.398</td>
</tr>
<tr>
<td>S12</td>
<td>Yunnan</td>
<td>October 2015</td>
<td>0.252</td>
<td>0.456</td>
</tr>
<tr>
<td>S13</td>
<td>Zhejiang</td>
<td>October 2015</td>
<td>0.160</td>
<td>0.254</td>
</tr>
<tr>
<td>S14</td>
<td>Zhejiang</td>
<td>October 2015</td>
<td>0.168</td>
<td>0.223</td>
</tr>
<tr>
<td>S15</td>
<td>Zhejiang</td>
<td>October 2015</td>
<td>0.150</td>
<td>0.265</td>
</tr>
<tr>
<td>S16</td>
<td>Northeast</td>
<td>October 2015</td>
<td>0.264</td>
<td>0.387</td>
</tr>
<tr>
<td>S17</td>
<td>Guangdong</td>
<td>October 2015</td>
<td>0.315</td>
<td>0.120</td>
</tr>
<tr>
<td>S18</td>
<td>Shandong</td>
<td>September 2014</td>
<td>0.067</td>
<td>0.007</td>
</tr>
</tbody>
</table>

Contents reported were the means of three replications. Data have been analyzed by ANOVA, instead of reporting the standard error.

For each constituent, significant difference of mean content within the same column was less than 0.05 (p < 0.05).

Fig. 1. The chromatographic fingerprints for the 18 batches of Codonopsis lanceolata. Reference (R) was produced by the median of 18 chromatograms.
For mobile phase, elution ability of acetonitrile is superior to methanol, and formic acid can get better baseline resolution and peak shape. Hence, formic acid was added into acetonitrile and water to obtain a stable baseline because of the low wavelength. As a result, the best chromatographic behavior was achieved as follows: a Hedera ODS-2 C18 column (250 mm × 4.6 mm, 5 \(\mu\)m), 0.1 mL L\(^{-1}\) formic acid in acetonitrile (A) and 0.1 mL L\(^{-1}\) formic acid in water (B) as mobile phase, wavelength of 218 nm, column temperature at 30 °C and flow rate of 0.8 mL min\(^{-1}\).

3.2. Identification of chemical constituents of Codonopsis lanceolata

Chemical constituents of C. lanceolata were identified by the developed HPLC-DAD-Q-TOF method both in positive and negative ion modes. A combination ion mode was adopted and representative total ion chromatograms (TICs) with numbered peaks are illustrated in Fig. 2. A total of 22 compounds, including 7 triterpenoid saponins, 4 phenylpropanoids, 2 polyacetylenes, 3 amino acids, 2 nucleosides, 1 organic acid and three other kinds of compounds, were identified or tentatively identified from C. lanceolata (Table 2). Their chemical structures are shown in the Supporting information. Among them, 9 compounds were reported for the first time in C. lanceolata. Besides, 8 compounds (1, 2, 3, 4, 5, 6, 7 and 14) were unambiguously identified by comparison with their reference standards on retention time and MS/MS data, while others 14 compounds were tentatively inferred based on their fragmentation pathways and previous reports.

3.2.1. Identification of triterpenoid saponins

Triterpenoid saponins are the major group of bioactive components from C. lanceolata. The molecular weight of the compounds was determined by the predominant ion [M−H]\(^{-}\) or [M + Na]\(^{+}\) and/or [M + NH\(_4\)]\(^{+}\) in full-scan mass spectra. Further structural information was obtained by referring to previous reports on fragmentation behaviors of triterpenoid saponins.

Peaks 16 and 19 both exhibited the deprotonated molecule [M−H]\(^{-}\) at m/z 1351 and sodiated molecule [M + Na]\(^{+}\) at m/z 1375, being tentatively assigned as lancemaside B and lancemaside E, respectively (Ichikawa et al., 2008). In the positive ion mode of peak 16, the m/z 705 ion [M + H − C\(_{36}\)H\(_{56}\)O\(_{10}\)]\(^{+}\) was produced by loss of 648 Da, corresponding to a glucuronic acid unit and an aglycone unit. Peaks 16 and 19 both yielded a product ion at m/z 411 [M + H − C\(_{47}\)H\(_{74}\)O\(_{19}\)]\(^{+}\), which was produced by loss of 942 Da, corresponding to a glucuronic acid, an aglycone, a hexose and an arabinose from the [M+H]\(^{+}\) ion. For peaks 16 and 19, the mass differences between 411 [M + H − C\(_{47}\)H\(_{74}\)O\(_{19}\)]\(^{+}\) ion and their product ions 279 [M + H − C\(_{30}\)H\(_{56}\)O\(_{19}\)−C\(_{6}\)H\(_{10}\)O\(_{4}\)]\(^{+}\) and 265 [M + H − C\(_{47}\)H\(_{74}\)O\(_{19}\)−C\(_{6}\)H\(_{10}\)O\(_{4}\)]\(^{+}\) were 132 Da and 146 Da, corresponding to loss of a xylose unit and a rhamnose unit, respectively. The MS information of other triterpenoid saponins, compounds 17–18, 20–22, are recorded in Table 2.

As described above, the molecular weights of triterpenoid saponins...
<table>
<thead>
<tr>
<th>NO.</th>
<th>( t_0 ) (min)</th>
<th>formula</th>
<th>([M + H]^+) ((m/z))</th>
<th>([M - H]^+) ((m/z))</th>
<th>Fragment ions in positive (+) ion mode</th>
<th>Fragment ions in negative (-) ion mode</th>
<th>Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>3.09</td>
<td>C(_4)H(_4)N(_2)O(_2)</td>
<td>175.1147</td>
<td>173.1045</td>
<td>349.2315 ([2M + H]^+), 158.0914 ([M + H - \text{NH}_3]^+)</td>
<td></td>
<td>arginine</td>
</tr>
<tr>
<td>2a</td>
<td>5.57</td>
<td>C(_6)H(_8)N(_2)O(_2)</td>
<td>284.1020</td>
<td>282.0862</td>
<td>567.1919 ([2M + H]^+), 306.0816 ([M + Na]^+)</td>
<td></td>
<td>guanosine</td>
</tr>
<tr>
<td>3a</td>
<td>6.33</td>
<td>C(_8)H(_8)N(_2)O(_4)</td>
<td>268.0983</td>
<td>266.0901</td>
<td>152.0565 ([M + H - \text{rbose}]^+), 135.0300 ([M + \text{H} - \text{rbose} - \text{NH}_3]^+)</td>
<td></td>
<td>adenosine</td>
</tr>
<tr>
<td>4a</td>
<td>8.55</td>
<td>C(_7)H(_7)NO(_2)</td>
<td>166.0822</td>
<td>164.0720</td>
<td>209.0868 ([M + Na]^+), 136.0586 ([M + H - \text{rbose} - \text{NH}_3]^+)</td>
<td></td>
<td>phenylalanine</td>
</tr>
<tr>
<td>5a</td>
<td>16.40</td>
<td>C(<em>{11})H(</em>{12})O(_2)</td>
<td>205.0919</td>
<td>203.0821</td>
<td>188.0712 ([M + H - \text{NH}_3]^+), 170.0596 ([M + \text{H} - \text{NH}_3 - \text{H}_2\text{O}]^+)</td>
<td>130.0910 ([M + H - \text{HCOOH}]^+), 159.0932 ([M - \text{H} - \text{CO}_2]^+)</td>
<td>hypoxanthine</td>
</tr>
<tr>
<td>6a</td>
<td>35.1032</td>
<td>C(_6)H(_8)O(_2)</td>
<td>353.1087</td>
<td>353.0897</td>
<td>395.1327 ([M + Na]^+), 364.1124 ([M + \text{Na} + \text{CH}_3\text{OH}]^+)</td>
<td></td>
<td>chlorogenic acid</td>
</tr>
<tr>
<td>7a</td>
<td>33.48</td>
<td>C(_7)H(_8)O(_2)</td>
<td>397.1673</td>
<td>396.1577</td>
<td>276.0721 ([M + Na]^+), 296.1254 ([M + Na]^+)</td>
<td></td>
<td>syringin</td>
</tr>
<tr>
<td>8</td>
<td>34.02</td>
<td>C(_9)H(_8)O(_2)</td>
<td>391.1763</td>
<td>390.1667</td>
<td>209.0819 ([M - \text{H}_2\text{O}]^+), 161.0201 ([M - \text{H}_2\text{O}]^+)</td>
<td></td>
<td>tanghensoside II</td>
</tr>
<tr>
<td>9</td>
<td>49.08</td>
<td>C(<em>{11})H(</em>{12})O(_3)</td>
<td>393.1792</td>
<td>392.1684</td>
<td>411.3559 ([M + Na]^+), 417.1741 ([M + Na]^+)</td>
<td></td>
<td>(E)-2-hexenyl-a-L-arabinopyranosyl</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>149.0456 ([C_6H_5O_4]^+)</td>
<td></td>
<td>(1→2)-β-D-glucopyranoside</td>
</tr>
<tr>
<td>10</td>
<td>50.62</td>
<td>C(<em>{13})H(</em>{14})O(_2)</td>
<td>677.2388</td>
<td>676.2282</td>
<td>701.2191 ([M + Na]^+)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>52.42</td>
<td>C(<em>{13})H(</em>{14})O(_2)</td>
<td>393.1807</td>
<td>392.1651</td>
<td>411.3559 ([M + Na]^+), 417.1741 ([M + Na]^+)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>60.74</td>
<td>C(<em>{13})H(</em>{14})O(_2)</td>
<td>1031.3761</td>
<td>1030.3663</td>
<td>547.2164 ([M + H - 3\text{C}_6\text{H}_12\text{O}_5]^+)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13a</td>
<td>61.41</td>
<td>C(<em>{13})H(</em>{14})O(_2)</td>
<td>581.2229</td>
<td>580.2123</td>
<td>569.2065 ([M + Na]^+), 576.1965 ([M + \text{Na} + \text{CH}_3\text{OH}]^+)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>67.45</td>
<td>C(<em>{13})H(</em>{12})O(_3)</td>
<td>395.1753</td>
<td>394.1657</td>
<td>815.3398 ([2M + Na]^+), 419.1676 ([M + Na]^+)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>70.64</td>
<td>C(<em>{13})H(</em>{12})O(_3)</td>
<td>373.1478</td>
<td>372.1372</td>
<td>329.1555 ([M + Na]^+), 328.1455 ([M + Na]^+)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>78.89</td>
<td>C(<em>{13})H(</em>{16})O(_3)</td>
<td>1351.6310</td>
<td>1350.6213</td>
<td>1375.6174 ([M + Na]^+), 705.3929 ([M + \text{H} - \text{C}_6\text{H}_12\text{O}_5]^+)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>78.90</td>
<td>C(<em>{13})H(</em>{16})O(_2)</td>
<td>1205.5772</td>
<td>1204.5675</td>
<td>141.1495 ([M + H - \text{C}_6\text{H}_12\text{O}_5]^+)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>79.36</td>
<td>C(<em>{13})H(</em>{16})O(_2)</td>
<td>1087.5456</td>
<td>1086.5359</td>
<td>414.2089 ([M + NH_4]^+), 257.1151 ([M + Na - \text{C}_6\text{H}_12\text{O}_5]^+)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>80.56</td>
<td>C(<em>{13})H(</em>{16})O(_2)</td>
<td>1351.6379</td>
<td>1350.6283</td>
<td>1229.5576 ([M + Na]^+), 411.1508 ([M + \text{H} - \text{C}_6\text{H}_12\text{O}_5]^+)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>82.05</td>
<td>C(<em>{13})H(</em>{16})O(_2)</td>
<td>1189.5842</td>
<td>1188.5745</td>
<td>129.0543 ([M + \text{H} - \text{C}_6\text{H}_12\text{O}_5 - \text{C}_6\text{H}_12\text{O}_5 - \text{H}_2\text{O}]^+)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued on next page)
could be determined by predominant ions \([\text{M} - \text{H}]^+\) and \([\text{M} + \text{Na}]^+\) . Also in the positive ion mode of all the triterpenoid saponins (compounds 16–22), fragment ions were always produced by loss of sugar moieties at the C-3 position, an aglycone with the sugar moiety (C-3) and oligosaccharide moiety at the C-28 position (Ichikawa et al., 2008).

### 3.2.2. Identification of Phenylpropanoids and Polycyctenes

Phenylpropanoids and polycyctenes are unique constituents originating from *C. lanceolata*. A total of four phenylpropanoids and two polycyctenes were identified in *C. lanceolata* based on reference standards and literature reports.

Phenylpropanoids are a group of phenolic compounds with additional derivationisation by a phenylethanol side-chain (Hanhineva et al., 2009). In negative ion mode, peak 7 and 8 both displayed formate adduct molecular ion \([\text{M} + \text{HCOO}]^-\) at \(m/z\) 417 and fragment ion \([\text{M} - \text{H} - \text{C}_6\text{H}_10\text{O}_4]^+\) at \(m/z\) 209 due to loss of a hexose residue. Meanwhile, in positive ion mode, peaks 7 and 8 both presented \([\text{M} + \text{Na}]^+\) at \(m/z\) 395. Product ions \([\text{M} + \text{Na} - \text{CH}_3\text{O}]^+\) at \(m/z\) 364 and \([\text{M} + \text{Na} - \text{C}_6\text{H}_10\text{O}_4]^+\) at \(m/z\) 233 were produced for peak 7 and 8, respectively. Peak 7 was precisely characterized as syringin by comparing with its reference standard and reported reference (Lee et al., 2005; Ren et al., 2013). Comparing with the literature (Zhu et al., 2001; Song et al., 2008), peak 8 was tentatively identified as tangshenoside II.

Peak 10 showed \([\text{M} + \text{Na}]^+\) ion at \(m/z\) 701 in positive ion mode. Product ion \([\text{M} + \text{Na} - \text{C}_6\text{H}_12\text{O}_3 - \text{C}_6\text{H}_10\text{O}_4]^+\) at \(m/z\) 347 was produced from \([\text{M} + \text{Na}]^+\) . Peak 10 also presented molecular ions at \(m/z\) 677 \([\text{M} - \text{H}]^-\) and \(m/z\) 1355 \([2\text{M} - \text{H}]^-\) in negative ion mode. Two fragment ions \([\text{C}_6\text{H}_12\text{O}_7]^+\) and \([\text{C}_6\text{H}_10\text{O}_4]^+\) were detected. Therefore, peak 10 was tentatively assigned as tangshenoside I based on MS/MS data from the literature (Jung et al., 2006; Liu et al., 2013), but still needs to be further confirmed by reference standard.

Compound 12 presented an \([\text{M} + \text{Na}]^+\) ion at \(m/z\) 1055, \([\text{M} - \text{H}]^-\) ion at \(m/z\) 1031 and an MS/MS fragment \([\text{M} + \text{H} - 3\text{C}_6\text{H}_10\text{O}_4]^+\) at \(m/z\) 547, being tentatively characterized as tangshenoside IV (Yuda et al., 1990; Ma et al., 2014).

Lobetylolinin (peak 13) and lobetylolin (peak 14), polycyctene compounds, are widely used as marker compounds to assess the quality of *C. lanceolata* (Tada et al., 1996). Peak 14 showed precursor ion \([\text{M} + \text{Na}]^+\) at \(m/z\) 419 with product ion \([\text{M} + \text{Na} - \text{C}_6\text{H}_10\text{O}_4]^+\) at \(m/z\) 257, \([\text{M} + \text{HCOO}]^-\) at \(m/z\) 441 with product ion \([\text{C}_6\text{H}_11\text{O}_3]^+\) at \(m/z\) 143. On retention time and mass spectrum of the reference standard, peak 14 was confirmed as lobetylolin. Peak 13 exhibited \([\text{M} + \text{Na}]^+\) ion at \(m/z\) 581 and \([\text{M} + \text{HCOO}]^-\) ion at \(m/z\) 603. The molecular weight of \([\text{M} + \text{Na} - \text{C}_6\text{H}_10\text{O}_4]^+\) ion at \(m/z\) 419 was equal to the quasi-molecular ion of lobetylolin. In addition, some fragments (e.g., \(m/z\) 215, 143 and 119) were the same as those of lobetylolin. Therefore, peak 13 was inferred as lobetylolin (Ma et al., 2014).

### 3.2.3. Identification of Amino Acids and Nucleosides

Amino acids are the typical nutrients in *C. lanceolata*. Their common fragments, generated by the losses of NH$_3$, H$_2$O, CO$_2$ and HCOOH because of the presence of amino and acid groups, have been extensively studied in positive ion mode (Qu et al., 2002). Peaks 1, 4 and 5 were confirmed as arginine, phenylalanine and tryptophan, respectively, by comparison with the peak retention times and MS spectra with those of standards (Liu et al., 2013). Peaks 2 and 3 were confirmed as guanosine and adenosine respectively by comparison with their authentic standards (Liu et al., 2013).

### 3.2.4. Identification of Organic Acid and Other Types of Compounds

Peak 6 was identified as chlorogenic acid by comparing with a reference standard. Peaks 9 and 11 both yielded a series of ions \([\text{M} + \text{Na}]^+\) at \(m/z\) 417, \([\text{M} + \text{HCOO}]^-\) at \(m/z\) 439 and \([\text{C}_6\text{H}_10\text{O}_4]^+\) at \(m/z\) 149. Based on the MS data in the literature (Yuda et al., 1990; Ma et al., 2014), they were inferred as \((E)-2$-$hexenyl$-a$-$arabinopyranosyl(1$\rightarrow$2)$-b$-$o$-$glucopyranoside and \((E)-2$-$hexenyl$-a$-$arabinopyranosyl(1$\rightarrow$6)$-b$-$o$-$glucopyranoside, respectively. But the elution order and their structures need to be further confirmed by more support information.
Table 3
Method validation for quantification of 5 constituents in *Codonopsis lanceolata*.

<table>
<thead>
<tr>
<th>Constituent</th>
<th>Calibration curves</th>
<th>LOD (μg·mL⁻¹)</th>
<th>LOQ (μg·mL⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phenylalanine</td>
<td>y = 21370x − 101</td>
<td>0.999</td>
<td>0.42</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>y = 246246x − 4150</td>
<td>1.000</td>
<td>1.26</td>
</tr>
<tr>
<td>Chlorogenic acid</td>
<td>y = 48973x + 322</td>
<td>0.998</td>
<td>0.32</td>
</tr>
<tr>
<td>Syringin</td>
<td>y = 123483x − 102</td>
<td>0.999</td>
<td>0.11</td>
</tr>
<tr>
<td>Lobetyolin</td>
<td>y = 23578x + 312</td>
<td>0.999</td>
<td>0.26</td>
</tr>
</tbody>
</table>

- Peak, peak area; x, concentration of each compound (μg/g).
- RSD values of intra- and inter-day precision, repeatability and stability were calculated based on analysis of one sample solution for six replicate determinations on the same day and on three consecutive days for intra-day and inter-day precision, respectively. Six independently prepared samples were assessed for the repeatability test. A single sample solution was analyzed at 0, 2, 4, 8, 12 and 24 h to assess the stability. The RSD values of intra- and inter-day precision, repeatability and stability of the 5 compounds were all lower than 2% (Table 3). The recovery was evaluated by assessing the accuracy of this developed method. The accurate amounts of 5 standard solutions were added into 0.5 g of *C. lanceolata* powder and then processed and analyzed by the developed method. An unspiked sample (0.5 g) was also prepared and analyzed simultaneously for comparison. Six replicates were prepared and the mean recoveries were calculated. As shown in Table 3, all the mean recoveries were between 96% and 103% and the RSD values were all less than 2%. The results indicated that the developed method was validated and applicable for *C. lanceolata* sample analysis. All tests described above were carried out on one sample from Shandong province (S1).

3.4. Quantitative application

The established analytical method was subsequently applied to analyze 18 samples of cultured *C. lanceolata*. Amounts of 5 compounds in different samples were determined and shown in Table 1, which varied significantly in different samples of *C. lanceolata*. The results showed that samples collected from different provinces and samples from the same province at different harvesting time had different contents of these 5 constituents, which would indicate various qualities and bioactivities of this herb. Polycyclotenes showed high contents in all samples. As shown in Table 1, lobetyolin (14) was the predominant polycyclotene. Amino acids (phenylalanine (4) and tryptophan (5)) were the second most abundant in *C. lanceolata*. Phenylpropanoids (syringin (7)) were also typical constituents in *C. lanceolata*. Tangshenoside I (10), another phenylpropanoid but without standard compound, had the highest HPLC and MS responses in all samples. As a result, amino acids, phenylpropanoids and polycyclotenes could be undoubtedly determined to comprehensively evaluate the quality of *C. lanceolata*.

3.5. Similarity analysis

The similarities of these 18 chromatograms compared with their mean chromatographic fingerprint were processed and written in
As revealed by the table, samples from Shandong province (S1, S4 and S18) had relatively low similarities (less than 0.8) and the others were over 0.8, implying that except for Shandong province, C. lanceolata samples from other origins were closely related, which indicated that quality consistency of C. lanceolata produced was different among different provinces.

### 3.6. Hierarchical clustering analysis

The aim of HCA is to calculate the degree of association among sample objects, which is expressed as distance. The smallest distance indicates the highest degree of association. Namely, those sample objects are considered to belong to the same group (Giacomino et al., 2011). The sample clusters were shown in a dendrogram (Fig. 3A) with a high-dimensional matrix (18 sample objects × 10 variables). As shown in Fig. 3A, the 18 samples could be divided into 2 clusters. Cluster I consisted of C. lanceolata from Shandong province (S1, S4 and S18), and the 15 samples from other origins were grouped into cluster II. In summary, C. lanceolata from Shandong province was relatively not close to other origins, which was in accordance with and used to support the result of HPLC fingerprint similarity.

### 3.7. Principal components analysis

PCA is one of the typical chemometric tools to monitor the outline of all data. The target of it is to reduce the dimensionality of complex data sets, namely allow projection of data from a higher to a lower dimensional space (defined by principal components, PCs) and then reconstruct them without any preliminary assumption about their distribution (Jolliffe, 2002). PCA was processed by using the same data matrix as HCA. Fig. 3B showed PCA score plots of C. lanceolata from various geographical origins. The first three PCs (PC1, PC2 and PC3) with 80.655% of cumulative variance were major contributors to discrimination. The 18 samples of C. lanceolata can be classified into two groups. Group I contained C. lanceolata from Shandong province (S1, S4 and S18), and group II was marked by the 15 samples from other origins. Above all, results of SA, HCA and PCA were consistent. The sample S18 in Cluster I was harvested in 2014 from Shandong province and samples S1 and S4 in Cluster I, with shorter distance than S18, were harvested in 2015 from Shandong province, while all the samples in cluster II were harvested in 2015 from the other eight provinces. These results showed that, harvesting time and geographical source are the major effects on sample classification.

### 4. Conclusion

In this study, an HPLC-DAD-Q-TOF-MS analytical method was successfully developed for the comprehensive qualification and quantification of cultured C. lanceolata, and the method feature was simple, specific and reliable. Twenty-two peaks were assigned, including 7 triterpenoid saponins, 4 phenylpropanoids, 2 polyacetylenes, 3 amino acids, 2 nucleosides, 1 organic acid and three other kinds of compounds. Among them, 9 compounds were identified for the first time. In addition, 5 constituents were quantified simultaneously in 18 samples of C. lanceolata. Furthermore, the chemical fingerprint and origin discrimination of 18 batches of cultured C. lanceolata from different areas in China were successfully analyzed by chemometric methods (SA, HCA and PCA). The results of SA, HCA and PCA were consistent, which could be validated against each other. The work has obtained a comprehensive understanding of chemical profile, quantitative analysis and similarity evaluation of cultured C. lanceolata from different regions. Amino acids, phenylpropanoids and polyacetylenes could be used to comprehensively evaluate the quality of C. lanceolata, which could provide a basis for the further study of this herb.

### Acknowledgement

This work was financially supported by Taian Science and Technology Program (No. 2015NS2110).

### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jfca.2017.05.009.
References


