



OPEN

## Intra-regional classification of Codonopsis Radix produced in Gansu province (China) by multi-elemental analysis and chemometric tools

Ruibin Bai<sup>1</sup>, Yanping Wang<sup>1</sup>, Jingmin Fan<sup>1</sup>, Jingjing Zhang<sup>1</sup>, Wen Li<sup>1</sup>, Yan Zhang<sup>2</sup> & Fangdi Hu<sup>1</sup>✉

Multi-elemental analysis is widely used to identify the geographical origins of plants. The purpose of this study was to explore the feasibility of combining chemometrics with multi-element analysis for classification of Codonopsis Radix from different producing regions of Gansu province (China). A total of 117 Codonopsis Radix samples from 7 counties of Gansu province were collected. Inductively coupled plasma mass spectrometry (ICP-MS) was used for the determination of 28 elements (<sup>39</sup>K, <sup>24</sup>Mg, <sup>44</sup>Ca, <sup>27</sup>Al, <sup>137</sup>Ba, <sup>57</sup>Fe, <sup>23</sup>Na, <sup>88</sup>Sr, <sup>55</sup>Mn, <sup>66</sup>Zn, <sup>65</sup>Cu, <sup>85</sup>Rb, <sup>61</sup>Ni, <sup>53</sup>Cr, <sup>51</sup>V, <sup>7</sup>Li, <sup>208</sup>Pb, <sup>59</sup>Co, <sup>75</sup>As, <sup>133</sup>Cs, <sup>71</sup>Ga, <sup>77</sup>Se, <sup>205</sup>Tl, <sup>114</sup>Cd, <sup>238</sup>U, <sup>107</sup>Ag, <sup>4</sup>Be and <sup>202</sup>Hg). Among macro elements, <sup>39</sup>K showed the highest level, whereas <sup>23</sup>Na was found to have the lowest content value. Micro elements showed the concentrations order of: <sup>88</sup>Sr > <sup>55</sup>Mn > <sup>66</sup>Zn > <sup>85</sup>Rb > <sup>65</sup>Cu. Among trace elements, <sup>53</sup>Cr and <sup>61</sup>Ni showed higher content and <sup>4</sup>Be was not detected in all samples. Intra-regions differentiation was performed by principal component analysis (PCA), cluster analysis (CA) and supervised learning algorithms such as linear discriminant analysis (LDA), k-nearest neighbors (k-NN), support vector machines (SVM), and random forests (RF). Among them, the RF model performed the best with an accuracy rate of 78.79%. Multi-elemental analysis combined with RF was a reliable method to identify the origins of Codonopsis Radix in Gansu province.

Codonopsis Radix is normally used as a replacement for expensive ginseng, which is one of the most common traditional Chinese medicines<sup>1</sup>. Codonopsis Radix has various medicinal properties, such as immunomodulatory<sup>2</sup>, anticancer<sup>3</sup>, gastrointestinal protection<sup>4</sup>, antidiabetic<sup>5</sup> and intestinal flora regulation activities<sup>6</sup>. Due to its good nutritional properties, Codonopsis Radix is also an important food material for tea, wine, soup, plaster, porridge, etc.<sup>7</sup> As a Chinese medicinal material with the same origin as medicine and food, Codonopsis Radix is absolutely popular in Asian countries such as China, Japan, South Korea, Singapore and et al<sup>8</sup>. China is a major producer and exports high quality Codonopsis Radix to other countries. In China, Codonopsis Radix has been cultivated in Gansu, Shanxi, Sichuan, Hubei, Yunnan and Chongqing provinces. Among them, Codonopsis Radix from Gansu production areas is of good quality and high price. Some counties in the southeast of Gansu Province are the most important production areas of Codonopsis Radix in China, and they account for more than 80% of the total output, such as Weiyuan, Lintao, Longxi, Zhang, Min, Tanchang and Wen<sup>9</sup>. The unique natural conditions of these places are the guarantee for the production of high quality Codonopsis Radix, such as altitude, soil, precipitation and sunshine<sup>10</sup>. The composition of inorganic elements in plants is influenced by genetic and environmental factors. Soil composition and climate variation can influence the content of elements in plants<sup>11</sup>. Therefore, it is necessary to understand the final expression levels of the elements in Codonopsis pilosula. However, the related researches are limited.

In recent years, consumers are paying more and more attention to the geographical traceability and authenticity of herbs, because these aspects have become indicators of their quality and safety<sup>12</sup>. Hence, it is of great

<sup>1</sup>School of Pharmacy @ the State Key Laboratory of Applied Organic Chemistry (SKLAOC), Lanzhou University, Lanzhou 730000, China. <sup>2</sup>National Engineering Research Center for Gelatin-Based Traditional Chinese Medicine, Dong-E-E-Jiao Co., Ltd., Liaocheng 252052, China. ✉email: hufd@lzu.edu.cn

Element	Lintao (n = 10)	Weiyuan (n = 31)	Longxi (n = 20)	Zhang (n = 12)	Min (n = 19)	Tanchang (n = 8)	Wen (n = 17)
<sup>39</sup> K	11,000 ± 1700	11,000 ± 1400	9200 ± 1300	8800 ± 1100	9600 ± 770	11,000 ± 2200	14,000 ± 2000
<sup>24</sup> Mg	1500 ± 130	1500 ± 200	1500 ± 220	1500 ± 220	1500 ± 180	1500 ± 87	1900 ± 190
<sup>44</sup> Ca	330 ± 59	370 ± 71	330 ± 70	290 ± 32	310 ± 34	340 ± 32	580 ± 120
<sup>27</sup> Al	290 ± 120	380 ± 280	250 ± 120	220 ± 28	350 ± 83	570 ± 170	780 ± 340
<sup>137</sup> Ba	250 ± 120	240 ± 130	140 ± 71	87 ± 49	110 ± 34	290 ± 180	370 ± 130
<sup>57</sup> Fe	180 ± 72	250 ± 180	160 ± 70	150 ± 17	220 ± 45	330 ± 95	430 ± 130
<sup>23</sup> Na	170 ± 210	190 ± 240	190 ± 140	180 ± 140	110 ± 62	140 ± 30	160 ± 77

**Table 1.** Macro elements content of *Codonopsis Radix* samples according to their geographical origin (Content are expressed as µg/g of dry matter basis).

importance to the economies of Gansu province to ensure the geographical origin of *Codonopsis Radix*. However, there was no report on the traceability of the origins of *Codonopsis Radix*.

Multi-elemental analysis is commonly used for origin traceability of plants, because inorganic elements are not metabolized during the metabolic process and closely related to the soil and climate environments of the planting areas. Inductively coupled plasma mass spectrometry (ICP-MS) is a stable and effective method for multi-element determination with higher sensitivities than common methods<sup>13–15</sup>. The elemental content combined with chemometrics has been used as a powerful tool for the differentiation of the geographical origins of plants<sup>16,17</sup>. Richter et al. used machine learning techniques and ICP-MS to predict the geographic origins of white asparagus samples from Poland, China, the Netherlands, Greece, Germany, Spain and Peru successfully<sup>18</sup>. Canizo et al. achieved the classification of grapes produced in different vineyards in Mendoza province (Argentina) based on multi-element analysis and chemometric techniques<sup>10</sup>. Therefore, it might be a robust strategy to distinguish the geographical origins of *Codonopsis Radix* in Gansu province.

This study aimed (1) to analyze the content of 28 elements (<sup>39</sup>K, <sup>24</sup>Mg, <sup>44</sup>Ca, <sup>27</sup>Al, <sup>137</sup>Ba, <sup>57</sup>Fe, <sup>23</sup>Na, <sup>88</sup>Sr, <sup>55</sup>Mn, <sup>66</sup>Zn, <sup>65</sup>Cu, <sup>85</sup>Rb, <sup>61</sup>Ni, <sup>53</sup>Cr, <sup>51</sup>V, <sup>7</sup>Li, <sup>208</sup>Pb, <sup>59</sup>Co, <sup>75</sup>As, <sup>133</sup>Cs, <sup>71</sup>Ga, <sup>77</sup>Se, <sup>205</sup>Tl, <sup>114</sup>Cd, <sup>238</sup>U, <sup>107</sup>Ag, <sup>4</sup>Be and <sup>202</sup>Hg) in 117 *Codonopsis Radix* samples from 7 counties of Gansu province by ICP-MS; (2) to assess the potential of multi-element analysis combined with chemometric tools to distinguish *Codonopsis Radix* samples based on their geographic origin.

## Results and discussion

**Validation of analytical methods.** The validation of the ICP-MS method was evaluated by linearity, sensitivity, accuracy and precision (see Supplementary Table S1 online). The observed R<sup>2</sup> values ranging from 0.9931 to 1.0000 and the experimental F value greater than the tabulated critical F conclude that the linearity of the analytical curves was good. The sensitivity is determined by detection (LOD) and quantification (LOQ) limits, which are according to the latest recommendations of IUPAC<sup>19</sup>, taking α and β (indicating types I and II errors, respectively) as the default value of 0.05. Except <sup>4</sup>Be and <sup>107</sup>Ag, the LOQ values of the remaining elements in the *Codonopsis Radix* samples in this study were all lower than their natural levels. Since there are no certified samples available for *Codonopsis Radix*, the accuracy of the method was determined by analyzing *Codonopsis Radix* samples (n = 3) fortified at their native levels before microwave digestion. The average recovery rate of most chemical substances is within the acceptable range (85–115%), which indicates that the loss during the digestion process is not obvious or almost negligible. The relative standard deviation (RSD) values for the elements were found to be less than 10%. Therefore, results of quality parameters confirmed that the used methods meet the standards required for the application of analytical methods.

**Analysis of macro elements.** The elements analyzed in the present study were classified into macro (≥ 100 µg/g), micro (10 ~ 100 µg/g) and trace (< 10 µg/g) according to their content. The total content of 7 macro elements in 117 *Codonopsis Radix* samples from different counties was summarized in Table 1. The order of the average content of macro elements was as follows: <sup>39</sup>K > <sup>24</sup>Mg > <sup>27</sup>Al > <sup>44</sup>Ca > <sup>57</sup>Fe > <sup>137</sup>Ba > <sup>23</sup>Na, of similar trends seen in the studies of Bai et al.<sup>20</sup>. Among them, K, Ca, Na, Mg and Fe were essential for normal human metabolism. They were important in the osmotic pressure balance, enzymatic reaction and hematopoiesis<sup>21,22</sup>. The mean levels were 11,000 µg/g for <sup>39</sup>K, 1600 µg/g for <sup>24</sup>Mg, 370 µg/g for <sup>44</sup>Ca, 240 µg/g for <sup>57</sup>Fe, 170 µg/g for <sup>23</sup>Na. Thus, *Codonopsis Radix* could be used as a dietary supplement to provide the body with important mineral elements, especially K. Surprisingly, the content of K element in *Codonopsis Radix* is far higher than that in some Chinese herbal medicines such *Rhizoma Coptidis*<sup>23</sup> and *Ephedrae herba*<sup>16</sup>, and even higher than some fruits rich in K, such as figs<sup>24</sup>. Furthermore, compared with the report of Sun et al., it was found that the content of Mg and Fe in Chinese *Angelica* from Min county is much higher than that in *Codonopsis Radix*, and the content of Na is lower than that in *Codonopsis Radix*<sup>25</sup>. It showed that the absorption capacity of different medicinal materials to the metal elements in the same soil might be different. Al and Ba were not essential elements of the human body, and excessive intake was harmful to nervous and kidney system<sup>26</sup>. The mean content of <sup>27</sup>Al and <sup>137</sup>Ba was found to be 400 and 207 µg/g in all *Codonopsis Radix* samples, which was higher than those of some reported traditional Chinese medicines, such as *Atractylodes macrocephala Koidz*<sup>17</sup> and *Rhizoma Coptidis*<sup>23</sup>. In addition, the highest <sup>39</sup>K, <sup>44</sup>Ca, <sup>24</sup>Mg, <sup>57</sup>Fe, <sup>27</sup>Al and <sup>137</sup>Ba content was all found in the samples from Wen County, which may be related to the abundance of mineral elements in the local soil.

Element	Lintao (n=10)	Weiyuan (n=31)	Longxi (n=20)	Zhang (n=12)	Min (n=19)	Tanchang (n=8)	Wen (n=17)
<sup>88</sup> Sr	49±9	61±34	56±24	28±8	26±7	26±4	65±29
<sup>55</sup> Mn	23±4	27±5	21±4	17±3	25±3	28±2	36±9
<sup>66</sup> Zn	18±4	21±8	15±3	35±8	17±2	21±5	23±4
<sup>65</sup> Cu	5.6±0.7	5.3±0.8	5.6±1.5	4.0±1.3	5.4±0.9	5.0±0.5	6.5±1.2
<sup>85</sup> Rb	5.5±1.6	6.9±2.8	8.3±3.0	6.0±4.0	3.4±1.9	9.2±2.7	8.7±4.2

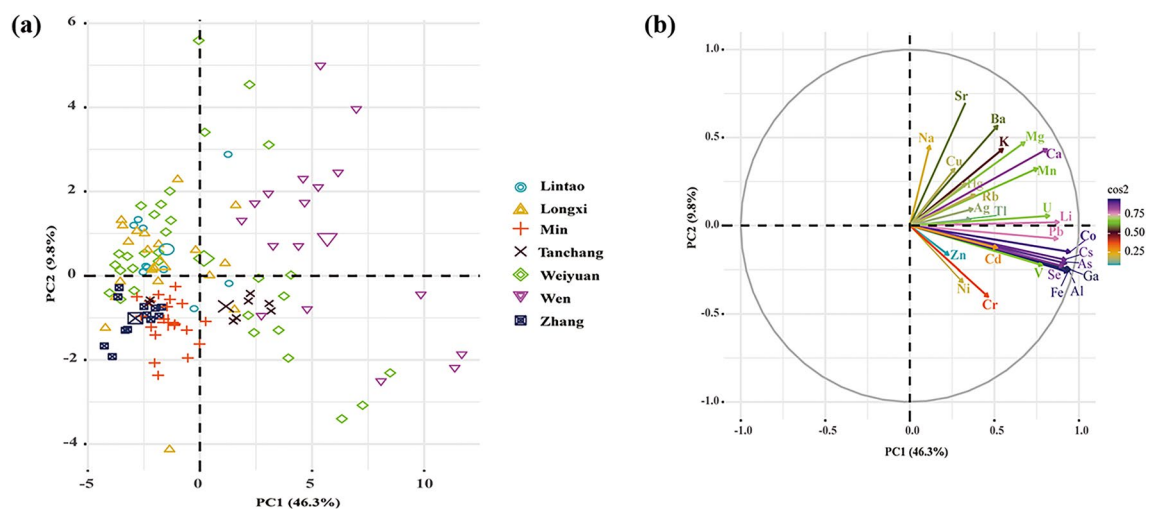
**Table 2.** Macro elements content of *Codonopsis Radix* samples according to their geographical origin (Content are expressed as µg/g of dry matter basis).

Element	Lintao (n=10)	Weiyuan (n=31)	Longxi (n=20)	Zhang (n=12)	Min (n=19)	Tanchang (n=8)	Wen (n=17)
<sup>61</sup> Ni	0.9±0.4	1.0±0.3	1.1±1.3	0.9±0.4	1.0±0.1	0.8±0.05	1.4±0.5
<sup>53</sup> Cr	0.8±0.3	1.1±1.1	1.3±1.4	0.5±0.4	1.7±0.6	1.0±0.3	1.3±0.6
<sup>51</sup> V	0.4±0.2	0.6±0.4	0.4±0.2	0.3±0.04	0.5±0.1	0.8±0.2	1.6±1
<sup>7</sup> Li	0.3±0.1	0.4±0.2	0.2±0.08	0.2±0.03	0.2±0.06	0.3±0.2	0.6±0.3
<sup>208</sup> Pb	0.2±0.1	0.3±0.1	0.3±0.08	0.2±0.03	0.3±0.05	0.3±0.07	0.4±0.08
<sup>59</sup> Co	0.1±0.03	0.2±0.07	0.1±0.03	0.1±0.02	0.1±0.01	0.2±0.03	0.2±0.05
<sup>75</sup> As	0.1±0.05	0.2±0.09	0.1±0.04	0.1±0.02	0.2±0.02	0.2±0.09	0.3±0.08
<sup>133</sup> Cs	0.06±0.02	0.08±0.05	0.06±0.02	0.06±0.02	0.07±0.02	0.1±0.03	0.2±0.06
<sup>71</sup> Ga	0.09±0.04	0.1±0.09	0.08±0.04	0.06±0.01	0.1±0.03	0.2±0.05	0.2±0.09
<sup>77</sup> Se	0.1±0.07	0.2±0.1	0.1±0.07	0.09±0.03	0.1±0.03	0.2±0.06	0.3±0.09
<sup>205</sup> Tl	0.04±0.02	0.05±0.02	0.05±0.02	0.04±0.03	0.02±0.01	0.06±0.02	0.05±0.03
<sup>114</sup> Cd	0.03±0.01	0.02±0.01	0.03±0.01	0.05±0.03	0.03±0.02	0.05±0.03	0.2±0.1
<sup>238</sup> U	0.02±0.01	0.02±0.01	0.01±0.01	0.01±0.003	0.01±0.002	0.02±0.003	0.03±0.01
<sup>107</sup> Ag	0.002±0.0004	0.002±0.0001	0.001±0.0001	0.002±0.0005	0.002±0.0004	0.001±0.0005	0.003±0.0004
<sup>202</sup> Hg	0.03±0.01	0.03±0.01	0.03±0.01	0.02±0.01	0.03±0.01	0.03±0.01	0.03±0.01

**Table 3.** Trace elements content of *Codonopsis Radix* samples according to their geographical origin (Content are expressed as µg/g of dry matter basis).

**Analysis of micro elements.** The content of 5 micro elements in 117 *Codonopsis Radix* samples from different counties was given in Table 2. The mean content of micro elements in the analyzed *Codonopsis Radix* samples showed the order: <sup>88</sup>Sr > <sup>55</sup>Mn > <sup>66</sup>Zn > <sup>85</sup>Rb > <sup>65</sup>Cu. Among them, Zn and Mn are essential elements of the human body. Zn was important in the growth and intellectual development of children<sup>27</sup>. Mn played a key role in the immune system and was also considered as potent antioxidants<sup>28</sup>. The mean content of <sup>55</sup>Mn and <sup>66</sup>Zn was 26 and 21 µg/g, respectively. The content of Mn element was lower than the mean content of Mn in 39 traditional Chinese medicines reported by Gyamfi, and the content of Zn is slightly higher<sup>29</sup>. The highest <sup>55</sup>Mn level was found in samples from Wen county and the highest <sup>66</sup>Zn level was found in Zhang county. In a certain concentration range, Cu is an essential element, which is related to the prosthetic groups of various enzymes and participates in key redox reactions. However, once Cu is excessive in the body, it will cause neurodegenerative diseases and impaired liver function<sup>30</sup>. The permissible limit set by the Green Trade Standard of Importing and Exporting Medicinal Plants and Preparations (WM-T2-2004) is 20 µg/g. The <sup>65</sup>Cu content in 117 *Codonopsis Radix* samples ranged from 2.2 to 9.6 µg/g, with an average content of 5.4 µg/g. The Cu content of all samples in this study was within the permissible limit and the highest content was found in samples from Wen county. Sr and Rb are not essential elements, the mean levels of <sup>88</sup>Sr and <sup>85</sup>Rb in the analyzed *Codonopsis Radix* samples were 48 and 6.8 µg/g, respectively. The highest <sup>88</sup>Sr level was found in samples from Wen county and the highest <sup>85</sup>Rb level was found in Tanchang county. Among them, the content of Sr was much higher than that of *Rhizoma Coptidis*<sup>23</sup> and Chinese *Angelica*<sup>25</sup> in China, and it was lower than that of some medicinal herbs in Turkey<sup>14</sup>.

**Analysis of trace elements.** The content of 15 trace elements in 117 *Codonopsis Radix* samples from different counties was given in Table 3. These include some essential trace elements such as Cr, Ni, Se, V and Co, which are essential nutrients that act as cofactors in metabolism and other biological processes. For example, chromium is important in the utilisation of glucose<sup>31</sup>. Selenium is an important component of the enzyme glutathione peroxidase<sup>32</sup>. Nickel is believed to act as a cofactor for iron absorption from the intestine during physiological processes<sup>33</sup>. Vanadium is an enzyme cofactor in hormone, glucose, lipid, bone and tooth metabolism<sup>34</sup>. Cobalt is an important component of vitamin B12<sup>35</sup>. The average content of <sup>53</sup>Cr, <sup>61</sup>Ni, <sup>51</sup>V, <sup>77</sup>Se and <sup>59</sup>Co was 1.2, 1.0, 0.7, 0.2 and 0.1 µg/g, respectively. The highest <sup>61</sup>Ni, <sup>51</sup>V, <sup>59</sup>Co and <sup>77</sup>Se content were all found in the samples from Wen County, and the highest <sup>53</sup>Cr content was found in Min county. The contents of non-toxic trace elements <sup>7</sup>Li, <sup>71</sup>Ga, <sup>133</sup>Cs and <sup>107</sup>Ag were also determined and their average contents were 0.3, 0.1, 0.09 and 0.002



**Figure 1.** Score (a) and loading (b) plots of the first principal component (PC1) versus the second principal component (PC2).

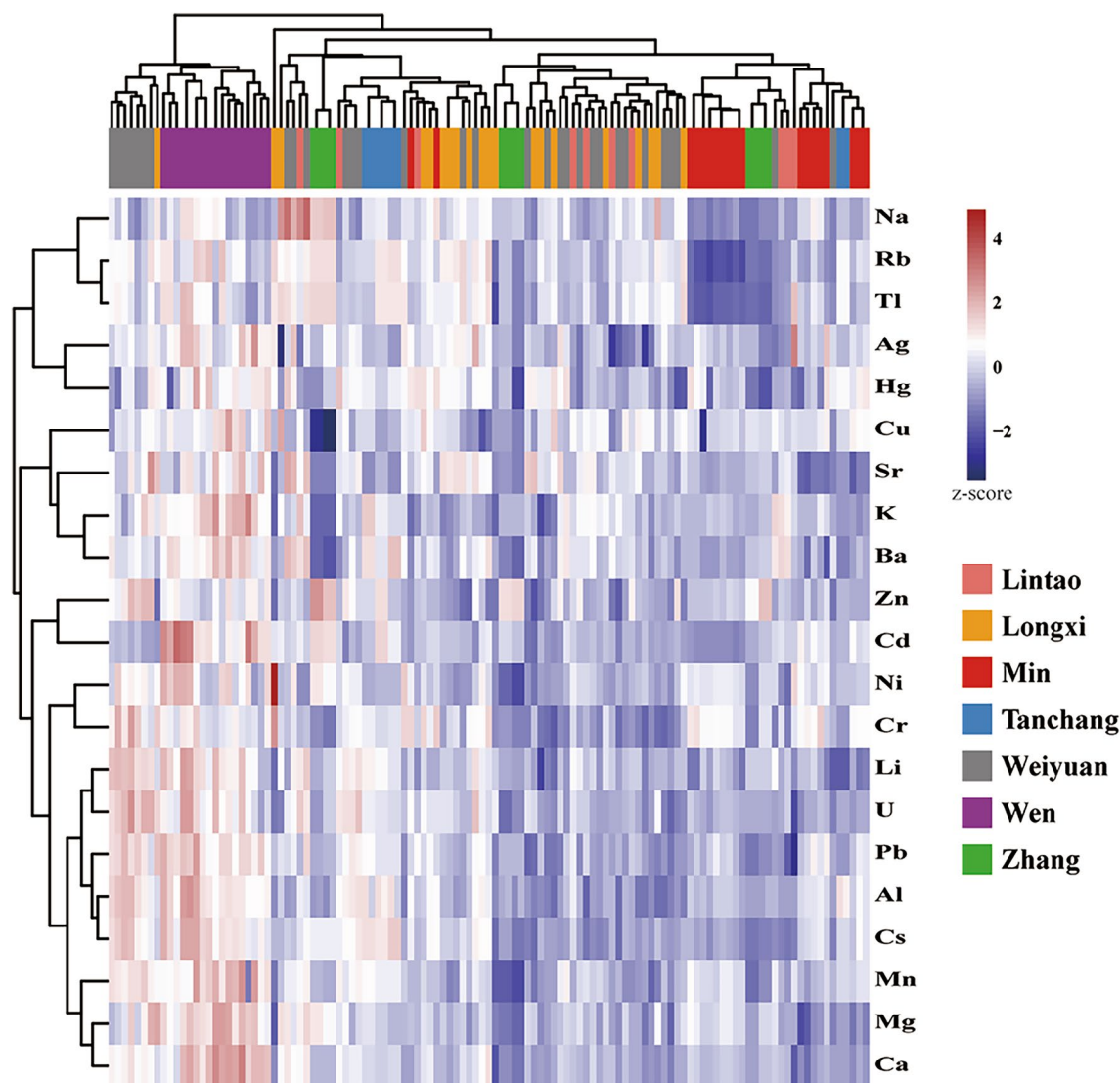
ug/g, respectively. The content of Ni, V, Co, Se Li, Cs and Ag was in good agreement with the values reported by Bai et al.<sup>20</sup>, and the content of Ga was lower.

Toxic trace elements, including As, Cd, Hg, Pb, U and Tl are recognized as toxic environmental contaminations and their contents needs to be carefully monitored, especially in herbs. The mean contents of toxic trace elements in all analyzed *Codonopsis Radix* samples were in the following order:  $^{208}\text{Pb} > ^{75}\text{As} > ^{114}\text{Cd} > ^{205}\text{Tl} > ^{202}\text{Hg} > ^{238}\text{U}$ , and the values were 0.3, 0.2, 0.04, 0.05, 0.03, and 0.02 ug/g, respectively. These toxic elements could cause irreversible damage to nervous, digestive and immune systems, lungs, kidneys, skin and eyes<sup>36</sup>. Based on the Green Trade Standard of Importing & Exporting Medicinal Plants & Preparations (WM-T2-2004), the permitted limit of Pb, As, Cd, and Hg established for TCM were 5.0, 2.0, 0.3 and 0.2  $\mu\text{g/g}$ , respectively. The highest levels of  $^{208}\text{Pb}$ ,  $^{75}\text{As}$ ,  $^{114}\text{Cd}$  and  $^{202}\text{Hg}$  in all *Codonopsis Radix* samples of the study are 0.6, 0.4, 0.4, 0.05  $\mu\text{g/g}$ . After comparison, it is found that 117 *Codonopsis Radix* samples had  $^{208}\text{Pb}$ ,  $^{75}\text{As}$  and  $^{202}\text{Hg}$  content below the permissible limit. However, the content of  $^{114}\text{Cd}$  in 4 samples from Wen county exceeded the permissible limits. It was found that the content of  $^{208}\text{Pb}$ ,  $^{75}\text{As}$ ,  $^{114}\text{Cd}$ , and  $^{202}\text{Hg}$  elements in the analyzed *Codonopsis Radix* from Gansu production areas was generally lower than the content of these heavy metal elements in other herbs, such as *Argy Wormwood Leaf*, *Morinda Root*, *Zedoray Rhizome*, *Indian Madder Root*<sup>37</sup>, which might be related to the soil, climate and less heavy metal pollution of planting areas. The content of Cr, As and Pb was lower than the values reported by Kong<sup>38</sup>.

**Analysis of variance.** A Kruskal–Wallis test after Bonferroni correction was performed to initially investigate the content of elements with significant differences in *Codonopsis Radix* samples from seven counties of Gansu province in China (indicated by  $p < 0.05$ ). The p-values showed that the content of 27 elements ( $^4\text{Be}$  was not detected) in *Codonopsis Radix* had significant differences. Therefore, all of the elements had a great influence on the provenance classification for one or more pairs of origins (see Supplementary Table S2 online). The box plots depicted the content profiles of the key elements with higher significance in *Codonopsis Radix* from different counties (see Supplementary Fig. S1 online). For  $^{55}\text{Mn}$ ,  $^{44}\text{Ca}$ ,  $^{24}\text{Mg}$ ,  $^{208}\text{Pb}$  and  $^{238}\text{U}$  elements, highest levels are observed in samples from Wen county. Levels of  $^{71}\text{Ga}$  for samples from Wen and Tanchang counties are significantly higher than other counties.  $^{66}\text{Zn}$  revealed higher levels for samples from Zhang county. Content of  $^{55}\text{Mn}$  and  $^{208}\text{Pb}$  were found to be very low in samples from Zhang county. The content of  $^{88}\text{Sr}$  was found to be low in samples from Min, Tanchang and Zhang counties.  $^{85}\text{Rb}$  revealed lower levels for samples from Min county.

**Principal component analysis (PCA).** In order to get an overview of the data set and visualize the differences between *Codonopsis Radix* samples from different counties, PCA was firstly performed with all 27 elements with significant differences ( $p < 0.05$ ). The total information content of a given number of principal components was represented by the cumulative percentage value (%) of the total variance. The first two principal components (PCs) represented 56.1% (PC1 represented for 46.3% and PC2 for 9.8%). Figure 1 exhibited PC1-PC2 score and loading plots. As shown in Fig. 1a, there was a large overlap between the scores corresponding to *Codonopsis Radix* samples from different origins. Nevertheless, *Codonopsis Radix* samples from the Wen county presented positive score on the considered PC1 and were distinguished with samples from the Zhang and Min counties with negative scores on PC1.

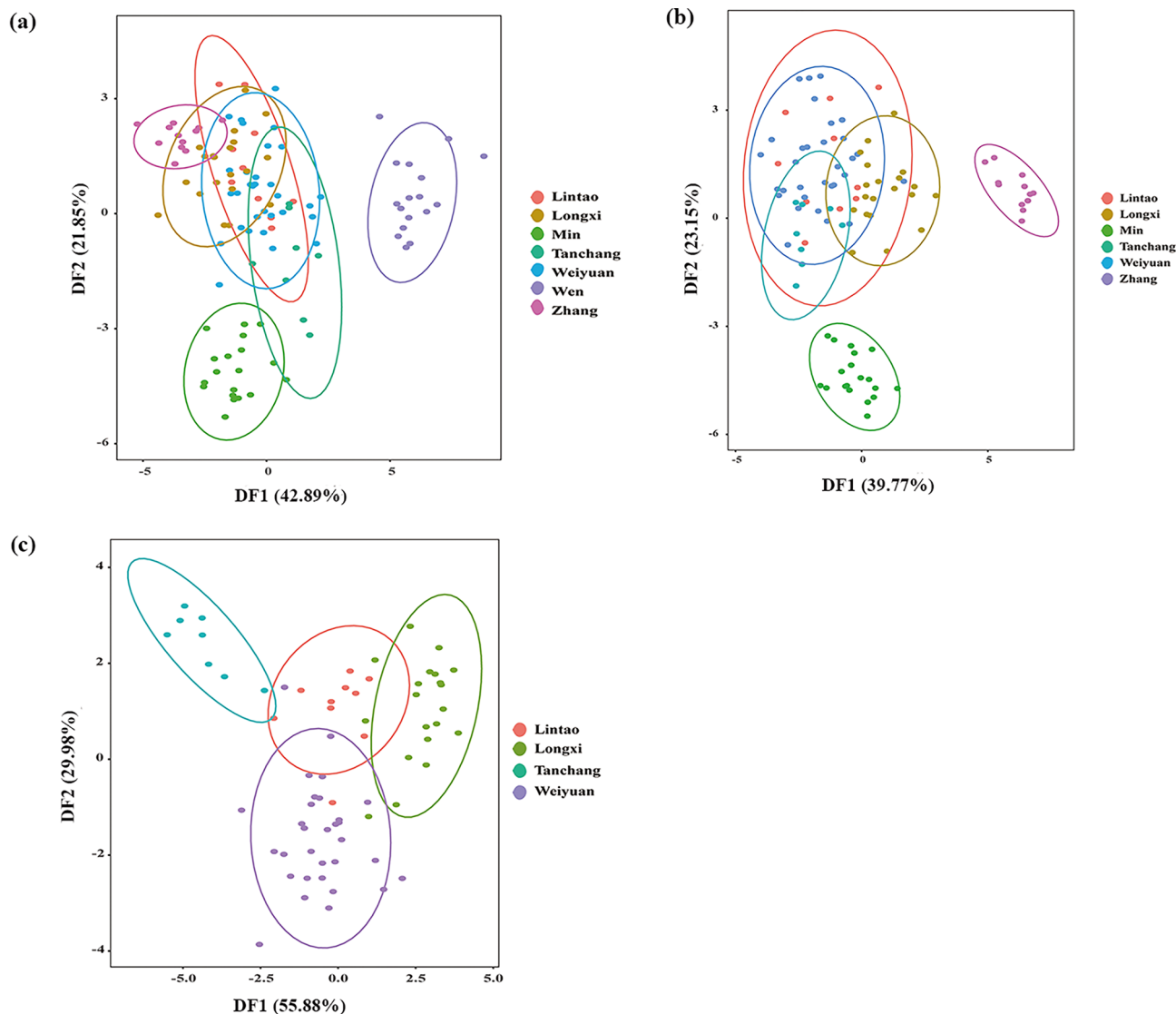
As shown in Fig. 1b, PC1 was strongly associated with the values of  $^{57}\text{Fe}$ ,  $^{27}\text{Al}$ ,  $^{71}\text{Ga}$ ,  $^{75}\text{As}$ ,  $^{133}\text{Cs}$ ,  $^{77}\text{Se}$  and  $^{59}\text{Co}$ . *Codonopsis Radix* samples with negative scores on PC1 indicated that lower content of these elements. On the other hand,  $^{88}\text{Sr}$ ,  $^{23}\text{Na}$ ,  $^{61}\text{Ni}$  and  $^{53}\text{Cr}$  were the main variables of PC2. Positive score corresponded to higher content of  $^{88}\text{Sr}$  and  $^{23}\text{Na}$  and negative scores indicate higher content of  $^{61}\text{Ni}$  and  $^{53}\text{Cr}$ .



**Figure 2.** Cluster heatmap showing the element concentrations in *Codonopsis Radix* samples. Color scale displays the range of concentrations from low (blue) to high (red).

**Cluster analysis.** Using Pearson correlation as a similarity measure, combined with hierarchical clustering to generate a heat map. Strong correlations were observed within the  $^{27}\text{Al}$ ,  $^{57}\text{Fe}$ ,  $^{51}\text{V}$ ,  $^{59}\text{Co}$ ,  $^{75}\text{As}$ ,  $^{71}\text{Ga}$  and  $^{77}\text{Se}$  elements, which indicated that the respective content profiles were quite similar (see Supplementary Fig. S2 online). In order to avoid generating excessive data, the heat map was recalculated containing only one of the highly related elements. As shown in Fig. 2, the horizontal tree diagram of *Codonopsis Radix* samples showed a completely separated cluster of samples from Wen county. Samples from other counties were not forming separated clusters. The indications of key elements related to provenance differentiation could be extracted from the content profiles. It displays not only high content of essential elements such as  $^{44}\text{Ca}$ ,  $^{24}\text{Mg}$ ,  $^{39}\text{K}$ , but also toxic elements such as  $^{208}\text{Pb}$  and  $^{114}\text{Cd}$  for the *Codonopsis Radix* samples from Wen county.

**Linear discriminant analysis (LDA).** In order to categorize the *Codonopsis Radix* samples in accordance with their geographical origins, a stepwise LDA was carried out. In the first step, the whole data set was explored to widely separate the *Codonopsis Radix* samples. As shown in Fig. 3a, the first two canonical discriminant functions (DFs) explained 64.74% of the variance. The plotted data showed that *Codonopsis Radix* samples from Wen county formed a distinct independent group. While samples from Lintao, Weiyuan, Longxi, Min, Zhang and Tanchang counties were clustered together with indistinct separation. In order to decide the accuracy of the samples in proximity, the elemental content of these counties was plotted separately. As shown in Fig. 3b, *Codonopsis Radix* samples from Min and Zhang counties were obviously separated. However, samples of Lintao, Weiyuan, Longxi and Tanchang clustered together. Undertaking this approach further, samples from Lintao, Weiyuan, Longxi and Tanchang counties grouped separately with unclear separation in Fig. 3c.



**Figure 3.** Scatter plot of the first two discriminant functions of linear discriminant analysis of *Codonopsis Radix* samples according to their geographical origin. (a) data from all regions, (b) data from Lintao, Weiyuan, Longxi, Min, Zhang and Tanchang, (c) data from Lintao, Weiyuan, Longxi and Tanchang.

**Statistical classification analysis.** So as to carry out a predictive classification analysis, data matrix was randomly divided into a training set (70% of the objects of the whole data matrix) and the test set (30%). Training set with known class membership was used to calculate the classifier. The test set contained objects not included in the training, and had known class membership to verify the model built.

In this work, four chemometric models, namely LDA, k-NN, SVM, and RF, were selected and tested to classify *Codonopsis Radix* samples according to their geographic origin. The LDA, k-NN, SVM and RF methods needed to optimize some parameters and build a model, and then evaluated it as a predictive tool. Trained each model by using k-fold cross-validation on the training set to build different classifiers. This process was repeated n times, so each subset must be tested at least once. In this work, the choices of number of neighbor k for k-NN; number of variables evaluated at each split (mtry) and number of trees (ntree) for RF; and penalty factor C and  $\epsilon$  of the  $\epsilon$ -insensitive loss function for SVM, were calculated by using ten-fold cross-validation technique repeated five times by which maximum accuracy was selected.

Once the best value for each model was selected, the sensitivity (samples that belonging to that category and correctly classified in that category), specificity (samples that do not belong to the modeled category and correctly classified as not belonging to the sample) and the mean accuracy rate were considered for evaluation of the classification achieved using chemometric methods. The results indicating the performance of the different classification methods were shown in Table 4.

According to the analysis of Table 4, it could be discovered that the four chemometric methods showed different degrees of success in the prediction of test samples. The sequence of successful recognition rate was as follows: RF > SVM > LDA > k-NN. RF showed the best performance in distinguishing the *Codonopsis Radix* samples based on their geographical origins, with total categorization accuracy of 78.79%. Great results were acquired

Groups	Number of samples		LDA		RF (ntree = 1000, mtry = 16) <sup>a</sup>		SVM (C = 10, $\epsilon = 0.01$ ) <sup>b</sup>		k-NN (k = 1) <sup>c</sup>	
	Train set	Test set	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
Lintao	6	3	0.00	96.67	0.00	96.67	0.00	100.00	0.00	90.00
Longxi	14	6	66.67	85.19	66.67	96.30	50.00	81.48	83.33	88.89
Min	14	5	100.00	96.43	100.00	100.00	100.00	100.00	100.00	92.86
Weiyuan	23	9	77.78	91.67	77.78	79.17	77.78	87.50	55.56	100.00
Wen	12	5	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Zhang	9	3	100.00	100.00	100.00	100.00	100.00	96.67	66.67	96.67
Tanchang	6	2	50.00	100.00	100.00	100.00	50.00	100.00	50.00	100.00
Mean accuracy (%)			75.76		78.79		75.76		72.73	

**Table 4.** Discrimination results obtained with the different chemometrics models. <sup>a</sup>ntree number of trees, mtry number of variables tried at each split. <sup>b</sup>C penalty factor,  $\epsilon$   $\epsilon$ -insensitive loss function. <sup>c</sup>k number of k neighbors.

for *Codonopsis Radix* samples from Wen, Min, Zhang and Tanchang counties with 100% accurate predictions. However, the *Codonopsis Radix* from Weiyuan, Lintao and Longxi origin could not be well predicted, which might be due to the close proximity of the three production areas and similar natural conditions and soil types.

In sum, the elemental analysis based on ICP-MS chosen in this study causes higher acquisition and maintenance costs compared to other elemental techniques. On the other hand, it can simultaneously determine multiple elements combined with low detection limits. In addition, the ICP-MS investigation did not reveal considerable disadvantages in terms of analysis time or cost compared to further establish identification methods. Therefore, our technology based on ICP-MS combined with chemometric analysis is a powerful tool for original traceability and identification of medicinal materials, which is consistent with results reported in many studies<sup>39–41</sup>. In the current study, differentiation through elemental composition was found reliable and satisfactory for *Codonopsis Radix* collected from 7 counties in Gansu province. Among all statistical tests, RF proved to be the most successful of origin differentiation of the analyzed samples. Our results might provide a new strategy for the origin traceability of Chinese herbal medicines. It should be emphasized that the 117 *Codonopsis Radix* samples this study was all harvested at their optimum harvesting period. Except that samples from Wen county had longer growth years, the samples from the other six producing areas had the same growth years. More samples need to be collected for further analysis to explore whether the growth period and harvest period will affect the content of elements in *Codonopsis*.

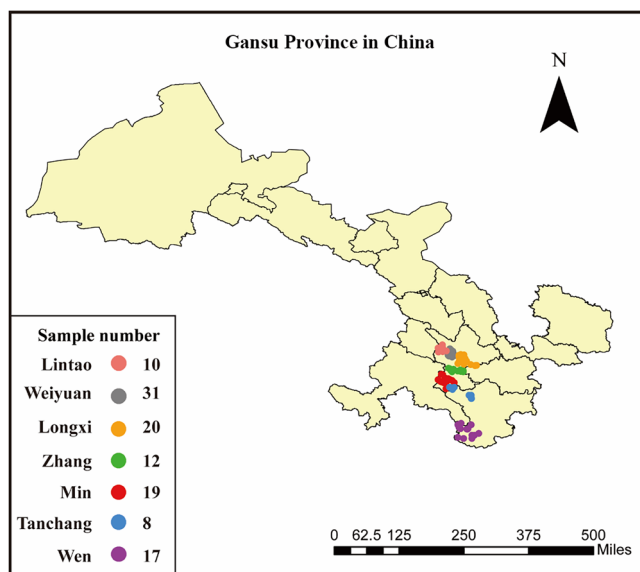
## Materials and methods

**Sample collection.** A total of 117 *Codonopsis Radix* samples were collected from seven counties of Gansu province (China): Lintao (35.39 N, 103.88E), Weiyuan (35.17 N, 104.19E), Longxi (34.98 N, 104.61E), Zhang (34.87 N, 104.48E), Min (34.41 N, 104.04E), Tanchang (34.06 N, 104.38E), Wen (32.95 N, 104.70E) during 2019 harvesting season (Fig. 4). The samples from Lintao, Weiyuan, Longxi, Zhang, Min and Tanchang were identified as *Codonopsis pilosula* (Franch.) Nannf. The samples from Wen were identified as *Codonopsis pilosula* Nannf. var. *modesta* (Nannf.) L. T. Shen. The *Codonopsis Radix* samples were all identified by Mr. Xicang Yang from the Department of Pharmacy, Affiliated Hospital of Gansu University of Chinese Medicine. All samples were subjected to the same process of rubbing, sweating and drying to obtain dried *Codonopsis Radix*. The experiments in this study were performed in accordance with relevant guidelines and regulations set by the Ministry of Agriculture of the People's Republic of China.

**Reagents.** Suprapure nitric acid (65%) from Merck (Darmstadt, Germany) was used to digest the samples. Ultrapure water (18.2M $\Omega$  cm resistivity at 25 °C) was obtained from a water purification system Milli Q (Millipore, Germany). Certified multi-element standard solutions (<sup>39</sup>K, <sup>24</sup>Mg, <sup>44</sup>Ca, <sup>27</sup>Al, <sup>137</sup>Ba, <sup>57</sup>Fe, <sup>23</sup>Na, <sup>88</sup>Sr, <sup>55</sup>Mn, <sup>66</sup>Zn, <sup>65</sup>Cu, <sup>85</sup>Rb, <sup>61</sup>Ni, <sup>53</sup>Cr, <sup>51</sup>V, <sup>7</sup>Li, <sup>208</sup>Pb, <sup>59</sup>Co, <sup>75</sup>As, <sup>133</sup>Cs, <sup>71</sup>Ga, <sup>77</sup>Se, <sup>205</sup>Tl, <sup>114</sup>Cd, <sup>238</sup>U, <sup>107</sup>Ag, <sup>4</sup>Be) was purchased from Inorganic Ventures (USA). Mercury(<sup>202</sup>Hg) mono-elemental standard solution was purchased from Institute of Metrology, China.

**Sample pretreatment and digestion.** The dried *Codonopsis Radix* (water content  $\leq 16.0\%$ ) was baked in an oven at 60 °C for 2 h, grounded using a pulverizer and stored in the plastic bags. About 300 mg of the samples were accurately weighed into a PTFE digestion vessel. 3 mL of concentrated HNO<sub>3</sub> and 1 mL of concentrated H<sub>2</sub>O were added to the vessel and waited for about 20 min before the vessel is closed. Digestion of *Codonopsis Radix* was performed using MARS microwave-assisted digestion system (CEM, United Kingdom). The digestion procedure was as follows: (1) 900 W at 110 °C for 10 min; (2) 1200 W at 200 °C for 16 min; (3) 1200 W at 240 °C for 25 min; (4) 0 W for 40 min to cool. After cooling, transferred the contents of the tubes to a volumetric flask and make up to 50 mL with double distilled water. Blank experiments ( $n = 3$ ) were performed in the same way.

**ICP-MS analysis.** ICP-MS analysis was carried out on an Agilent 7900 instrument (Agilent Technologies, Santa Clara, CA, USA). 28 elements (<sup>39</sup>K, <sup>24</sup>Mg, <sup>44</sup>Ca, <sup>27</sup>Al, <sup>137</sup>Ba, <sup>57</sup>Fe, <sup>23</sup>Na, <sup>88</sup>Sr, <sup>55</sup>Mn, <sup>66</sup>Zn, <sup>65</sup>Cu, <sup>85</sup>Rb, <sup>61</sup>Ni, <sup>53</sup>Cr, <sup>51</sup>V, <sup>7</sup>Li, <sup>208</sup>Pb, <sup>59</sup>Co, <sup>75</sup>As, <sup>133</sup>Cs, <sup>71</sup>Ga, <sup>77</sup>Se, <sup>205</sup>Tl, <sup>114</sup>Cd, <sup>238</sup>U, <sup>107</sup>Ag, <sup>202</sup>Hg and <sup>9</sup>Be) were determined. The



**Figure 4.** Geographical locations and number of *Codonopsis Radix* studied in this work (prepared by RB in ArcGIS Pro, <https://www.esri.com/zh-cn/arcgis/products/arcgis-pro/resources>).

frequency power was 1300 W. The collision, carrier and plasma gas flow rate flow rate were 5.00, 1.17 and 15.00 L/min, respectively. The spray chamber temperature was 2 °C. The collision cell was operated at a helium flow rate of 5.5 mL/min to separate out polyatomic interferences.  $^{209}\text{Bi}$ ,  $^{115}\text{In}$ ,  $^{103}\text{Rh}$ ,  $^{72}\text{Ge}$  and  $^{45}\text{Sc}$  were used as internal standards to correct for matrix effects and instrument drift.

**Statistical analysis.** First, the concentration of all elements was compared with the LOD and LOQ.  $^4\text{Be}$  was not detected in all samples. It was found that the  $^{107}\text{Ag}$  concentration contained a value lower than the LOD, which meant that there were no detectable  $^{107}\text{Ag}$  in some *Codonopsis Radix* samples. In order to avoid difficulties in applying the logarithmic function, the concentration of the  $^{107}\text{Ag}$  element located below the LOD was set to the LOD level instead of zero.

Three copies of all samples were averaged, and the logarithm of the average content of the elements to the base of 10 ( $\log_{10}$ ) was used for data analysis. The data matrix consisted of 27 columns (Be was not detected) and 117 rows for chemometrics analysis. The columns represented the content of elements and the rows corresponded each sample. Analysis of variance (ANOVA) was used to compare the differences of elemental content of the samples, with  $p < 0.05$  as the significance level. Principal component analysis (PCA) was carried out to reduce dimensionality and to visualize the dataset. Based on Pearson correlation, hierarchical cluster analysis (HCA) was executed to detect feature similarities in the dataset.

Four chemometric methods were performed to evaluated different models for classification of *Codonopsis Radix* grown in Gansu provinces according to their origins: linear discriminant analysis (LDA), k-nearest neighbor (k-NN), support vector machine (SVM), and random forest (RF). LDA is a classification method designed to maximize the ratio of between-class variance to the within-class variance for achieving maximum separability. The decision boundary created by LDA is called a discriminant function, and is a linear combination of the variables that can best distinguish categories<sup>42</sup>. k-NN is a classification technique that uses the Euclidean distance to calculate the k samples (neighbors) closest to the test sample to the test sample in the feature space, and then sets its category label to the most frequent category label that appears in the found neighbors<sup>43</sup>. SVM is a powerful method for building classifiers. It aims to create a decision boundary between two classes, so that labels can be predicted based on one or more feature vectors<sup>44</sup>. RF consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes the model's prediction<sup>45</sup>. All basic statistic and multivariate analysis were performed with R software version R 3.6.3.

## Conclusion

The elemental composition profile was assessed for the first time in detail for macro, micro and trace elements in the *Codonopsis Radix* samples collected from seven counties of Gansu province. Among the macro elements,  $^{39}\text{K}$  had the highest and  $^{23}\text{Na}$  the lowest content levels. Among the micro elements, the content order was found to be  $^{88}\text{Sr} > ^{55}\text{Mn} > ^{66}\text{Zn} > ^{85}\text{Rb} > ^{65}\text{Cu}$ . Among the trace elements,  $^{53}\text{Cr}$  and  $^{61}\text{Ni}$  showed higher content. The contents of trace toxic elements 208Pb, 75As and 202Hg were below the permissible limit in all samples. And, 114Cd was also below the permissible limit except the 4 samples from Wenxian counties. Statistical analyses of the data using RF successfully classified the *Codonopsis Radix* samples from Wenxian, Minxian, Zhangxian, Tanchang and Weiyuan (or Lintao or Longxi) production areas. The optimal parameters of this model are preferably  $\text{n tree} = 1000$  and  $\text{m try} = 16$ .



Received: 3 December 2021; Accepted: 4 May 2022

Published online: 20 May 2022

## References

- Jiang, Y. *et al.* Sesquiterpene glycosides from the roots of *Codonopsis pilosula*. *Acta Pharm. Sin. B.* **6**, 46–54 (2016).
- Bai, R. B. *et al.* Immune-enhancement effects of oligosaccharides from *Codonopsis pilosula* on cyclophosphamide induced immunosuppression in mice. *Food Funct.* **11**, 3306–3315 (2020).
- Bai, R. B. *et al.* Cytotoxicity of two water-soluble polysaccharides from *Codonopsis pilosula* Nannf. Var. *modesta* (Nannf.) L.T. Shen against human hepatocellular carcinoma HepG2 cells and its mechanism. *Int. J. Biol. Macromol.* **120**, 1544–1550 (2018).
- Li, J. *et al.* Structure features and anti-gastric ulcer effects of inulin-type fructan CP-A from the roots of *codonopsis pilosula* (Franch.) Nannf. *Molecules* **22**, 1–11 (2017).
- Liu, W., Lv, X., Huang, W., Yao, W. & Gao, X. Characterization and hypoglycemic effect of a neutral polysaccharide extracted from the residue of *Codonopsis pilosula*. *Carbohydr. Polym.* **197**, 215–226 (2018).
- Jing, Y. *et al.* Absorption of *Codonopsis pilosula* saponins by coexisting polysaccharides alleviates gut microbial dysbiosis with dextran sulfate sodium-induced colitis in model mice. *Biomed Res. Int.* 1781036, (2018).
- Zou, Y. F. *et al.* Prospects of *Codonopsis pilosula* polysaccharides: structural features and bioactivities diversity. *Trends Food Sci Technol.* **103**, 1–11 (2020).
- Gao, S. M. *et al.* Traditional uses, phytochemistry, pharmacology and toxicology of *codonopsis*: a review. *J. Ethnopharmacol.* **219**, 50–70 (2018).
- Wu, X. J., Zhang, X. B., Guo, L. P., Yu, Y. & Huang, L. Q. Study on distribution division of *Codonopsis Radix*. *J. Tradit. Chin. Med.* **42**, 4368–4372 (2017).
- Canizo, B. V., Escudero, L. B., Pérez, M. B., Pellerano, R. G. & Wuilloud, R. G. Intra-regional classification of grape seeds produced in Mendoza province (Argentina) by multi-elemental analysis and chemometrics tools. *Food Chem.* **242**, 272–278 (2018).
- Zhao, H. Origin traceability of peanut kernels based on multi-element fingerprinting combined with multivariate data analysis. *J. Sci. Food Agric.* **100**, 4040–4048 (2020).
- D'Archivio, A. A. *et al.* Geographical discrimination of red garlic (*Allium sativum* L.) produced in Italy by means of multivariate statistical analysis of ICP-OES data. *Food Chem.* **275**, 333–338 (2019).
- Llorent-Martínez, E. J., Ortega-Barrales, P., Fernández-De Córdova, M. L., Domínguez-Vidal, A. & Ruiz-Medina, A. Investigation by ICP-MS of trace element levels in vegetable edible oils produced in Spain. *Food Chem.* **127**, 1257–1262 (2011).
- Tokalioglu, Ş. Determination of trace elements in commonly consumed medicinal herbs by ICP-MS and multivariate analysis. *Food Chem.* **134**, 2504–2508 (2012).
- Khan, N. *et al.* Analysis of minor and trace elements in milk and yogurts by inductively coupled plasma-mass spectrometry (ICP-MS). *Food Chem.* **147**, 220–224 (2014).
- Ma, X. *et al.* Discrimination of three *Ephedra* species and their geographical origins based on multi-element fingerprinting by inductively coupled plasma mass spectrometry. *Sci. Rep.* **8**, 4–13 (2018).
- Hu, L., Chen, X., Yang, J. & Guo, L. Geographic authentication of the traditional Chinese medicine *Atractylodes macrocephala* Koidz. (Baizhu) using stable isotope and multielement analyses. *Rapid Commun. Mass Spectrom.* **33**, 1703–1710 (2019).
- Richter, B., Gurk, S., Wagner, D., Bockmayr, M. & Fischer, M. Food authentication: Multi-elemental analysis of white asparagus for provenance discrimination. *Food Chem.* **286**, 475–482 (2019).
- Olivieri, A. C. Analytical figures of merit: From univariate to multiway calibration. *Chem. Rev.* **114**, 5358–5378 (2014).
- Bai, J., Ji, L. & Wang, Y. Study on the content changes of 27 kinds of heavy metals and trace elements in *Glycyrrhizae radix et rhizoma*, *Astragali radix* and *Codonopsis radix*. *J. Food Saf.* **12**, 1448–1455 (2021).
- Stone, M. S., Martyn, L. & Weaver, C. M. Potassium intake, bioavailability, hypertension, and glucose control. *Nutrients* **8**, 1–13 (2016).
- Li, K. *et al.* The good, the bad, and the ugly of calcium supplementation: A review of calcium intake on human health. *Clin. Interv. Aging.* **13**, 2443–2452 (2018).
- Du, H. *et al.* Combination of stable isotopes and multi-elements analysis with chemometric for determining the geographical origins of *Rhizoma Coptidis*. *Microchem. J.* **152**, 104427 (2020).
- Lo Turco, V., Potorti, A. G., Tropea, A., Dugo, G. & Di Bella, G. Element analysis of dried figs (*Ficus carica* L.) from the Mediterranean areas. *J. Food Compos. Anal.* **90**, 103503 (2020).
- Sun, L. *et al.* Geographical origin differentiation of Chinese *Angelica* by specific metal element fingerprinting and risk assessment. *Environ. Sci. Pollut. Res.* **27**, 45018–45030 (2020).
- Kohzadi, S., Shahmoradi, B., Ghaderi, E., Loqmani, H. & Maleki, A. Concentration, source, and potential human health risk of heavy metals in the commonly consumed medicinal plants. *Biol. Trace Elem. Res.* **187**, 41–50 (2019).
- Chasapis, C. T., Spiliopoulou, C. A., Loutsidou, A. C. & Stefanidou, M. E. Zinc and human health: An update. *Arch. Toxicol.* **86**, 521–534 (2012).
- Lucchini, R. G., Aschner, M., Landrigan, P. J. & Cranmer, J. M. Neurotoxicity of manganese: Indications for future research and public health intervention from the Manganese 2016 conference. *Neurotoxicology* **64**, 1–4 (2018).
- Gyamfi, E. T. Assessment of essential and non-essential elements in selected traditional medicines from India, Ghana and China. *Environ. Sci. Pollut. Res.* **28**, 1812–1822 (2021).
- Rajeswari, S. & Swaminathan, S. Role of copper in health and diseases. *Int. J. Curr. Sci.* **10**, 94–107 (2014).
- Cefalu, W. T. & Hu, F. B. Role of chromium in human health and in diabetes. *Diabetes Care* **27**, 2741–2751 (2004).
- Rayman, M. P. The importance of selenium to human health. *Lancet* **356**, 233–241 (2000).
- Srivastava, S., Pant, M. & Agarwal, N. A review on role of nickel in the biological system. *Int. J. Curr. Microbiol. App. Sci.* **5**, 719–727 (2016).
- Rehder, D. Structure and function of vanadium compounds in living organisms. *Biometals* **5**, 3–12 (1992).
- Yamada, K. Cobalt: its role in health and disease. *Met. Ions Life Sci.* **13**, 295–320 (2013).
- Luo, L. *et al.* Heavy metal contaminations in herbal medicines: determination, comprehensive risk assessments, and solutions. *Front. Pharmacol.* **11**, 1–14 (2021).
- Zuo, T. T. *et al.* Innovative health risk assessment of heavy metals in Chinese herbal medicines based on extensive data. *Pharmacol. Res.* **159**, 104987 (2020).
- Kong, D. *et al.* Establishment of health risk assessment model for assessing medicinal and edible plants contaminated by heavy metals take *Astragali Radix*, *Codonopsis Radix* and *Laminariae Thallus* as examples. *J. Tradit. Chin. Med.* **44**, 5042–5050 (2019).
- Hidalgo, M. J., Fechner, D. C., Marchevsky, E. J. & Pellerano, R. G. Determining the geographical origin of *Sechium edule* fruits by multielement analysis and advanced chemometric techniques. *Food Chem.* **210**, 228–234 (2016).
- Zhang, J. *et al.* Geographical origin discrimination of pepper (*Capsicum annum* L.) based on multi-elemental concentrations combined with chemometrics. *Food Sci. Biotechnol.* **28**, 1627–1635 (2019).

41. Kang, X. *et al.* Elemental analysis of sea cucumber from five major production sites in China: a chemometric approach. *Food Control* **94**, 361–367 (2018).
42. Maione, C. & Barbosa, R. M. Recent applications of multivariate data analysis methods in the authentication of rice and the most analyzed parameters: a review. *Crit. Rev. Food Sci. Nutr.* **59**, 1868–1879 (2018).
43. Zhang, Z. Introduction to machine learning: k-nearest neighbors. *Ann. Transl. Med.* **4**, 218 (2016).
44. Manitoba, C. & Genetics, M. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics Proteomics*. **15**, 41–51 (2018).
45. Schonlau, M. & Zou, R. Y. The random forest algorithm for statistical learning. *Stata J.* **20**, 3–29 (2020).

## Acknowledgements

This work was supported by the National Key Research and Development Program of China (2018YFC1706300, 2018YFC17063003), the Major Projects in Gansu Province (21ZD4FA013), the People's Livelihood Special Project from Technical Innovation Guide Plan of Gansu Province (20CX4FK014), the Talent Innovation and Entrepreneurship Project of Lanzhou (2017-RC115, 2020-RC-41).

## Author contributions

All authors contributed to the study conception and design. R.B.: Conceptualization; Data curation; Methodology; Writing - original draft; Writing - review and editing. Y.W.: Investigation; Formal analysis; Writing - original draft. J.F.: Investigation; Resources. Y.Z.: Supervision; Resources. J.Z.: Methodology; Software. W.L.: Resources. F.H.: Funding acquisition; Project administration; Resources; Formal analysis; Conceptualization; Writing - review and editing. All authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-12556-z>.

**Correspondence** and requests for materials should be addressed to F.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022