Unveiling the authentic secrets of traditional herbal medicines: A comprehensive investigation into metagenomic, metabolomic, and quality assessment of two categories of Codonopsis Pilosula from Shanxi Province

Yuhui Wu
Shanxi University of Chinese Medicine
Xianglong Meng
Shanxi University of Chinese Medicine
Xiang Han
Shanxi University of Chinese Medicine
Zhulin Bu
Shanxi University of Chinese Medicine
Yuting Li
Shanxi University of Chinese Medicine
Haixian Zhan
Shanxi University of Chinese Medicine
Shuosheng Zhang
zhangshuosheng@aliyun.com

Shanxi University of Chinese Medicine

Research Article

Keywords: genuine traditional herbal medicines, Tai Codonopsis pilosula, Lu Codonopsis pilosula, secondary metabolites, soil microorganisms, soil metal elements

Posted Date: September 10th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4843710/v1

Additional Declarations: No competing interests reported.

Title Page

Original article

Unveiling the authentic secrets of traditional herbal medicines: A comprehensive investigation into metagenomic, metabolomic, and quality assessment of two categories of Codonopsis Pilosula from Shanxi Province

Yuhui Wu[a,b,c#], Xianglong Meng[a,b,c#], Xiang Han[a,b,c], Zhulin Bu[a,b,c] , Yuting Li[a,b,c], Haixian Zhan[a], Shuosheng Zhang[a,b,c,*]

[a]College of Chinese Materia Medica and Food Engineering, Shanxi University of Chinese Medicine, Jinzhong 030619, Shanxi, China

[b]Shanxi Key Laboratory of Traditional Herbal Medicines Processing, Jinzhong 030619, Shanxi, China

[c]Chinese herbal medicine processing inheritance and innovation key research laboratory

*Corresponding authors. Tel.: +86-13603585965 (S.S. Zhang).

E-mail addresses: 1141546657@qq.com (S.S. Zhang).

Unveiling the authentic secrets of traditional herbal medicines: A comprehensive investigation into the metagenomic, metabolomic, and quality assessment of two categories of *Codonopsis pilosula* from Shanxi Province, China

Abstract:

Background The main source of *Codonopsis pilosula* (CP) is the Campanulaceae family *Codonopsis pilosula* (Franch.) Nannf., which is one of the traditional herbal medicines in Shanxi Province, China. In Shanxi Province, CP can be divided into two major categories, Lu CP (Lu dangshen in Chinese, LCP) and Tai CP (Tai dangshen in Chinese, TCP), because of their geographical conditions and traits. Since LCP is cultivated on a large scale, it has emerged as one of the commonly prescribed medicines in TCM clinics. TCP predominantly exists as a wild species. The artificial cultivation technology for TCP has not yet been fully developed, and the availability of wild resources for TCP is limited. There have been few comparative studies on the authentic evaluation of the two types of CPs.

Results The findings revealed that the two types of CPs met the quality standards. With respect to the growing environment, TCP and LCP grow in the same longitudinal extent, with a maximum longitudinal span of only 53'. Both regions presented similar natural environments conducive to the growth of CP. Additionally, the lobetyolin content in TCP was greater than that in LCP. The soil microbial abundance in the TCP group was considerably significantly greater than that in the LCP group, and a new genus, Ascobolus, was found in TCP. Additionally, compared with TCP, LCP materials presented elevated levels of metals. These results not only indicate the main ecological and environmental parameters of Shanxi Province as a suitable production area for CP but also offer differential information for comparisons between the two types of CP specifications in Shanxi Province.

Conclusions TCP and LCP share similar geographical environments for growth.

However, TCP features a dense ring pattern beneath the stalk, a high content of lobetyolin, and a new flora, Ascobolus, in the root soil.

Keywords genuine traditional herbal medicines, Tai *Codonopsis pilosula*, Lu *Codonopsis pilosula*, secondary metabolites, soil microorganisms, soil metal elements

1. Introduction

Traditional herbal medicines (THMs) are derived from natural sources. Prior to the introduction of synthetic drugs in Western medicine, Chinese people frequently relied on natural remedies known as THMs to address illnesses, following traditional Chinese medical (TCM) theories and principles. The assessment of THM quality and clinical effectiveness is influenced by various factors, including the cultivation environment, different processing methods and techniques employed in distinct planting areas or regions. These factors are crucial in determining the extrinsic characteristics of THMs as well as the content of intrinsic bioactive compounds or secondary metabolites. The aforementioned factors can also be encapsulated by an authentic TCM concept known as *Daodi*.

Daodi medicines are synonymous with genuine THMs (GTHMs). *Daodi* medicines refer to those that have been selected through long-term application in TCM clinical practice. Compared with similar THMs grown elsewhere, it is cultivated in specific regions; has superior properties, quality, and efficacy; has consistent quality; and is more popular.

China possesses an extensive land area, and its distinctive natural and geographical conditions have fostered a wide range of GTHM resources. Northeast China GTHMs are referred to as *Guanyao*, such as Asarum (*Xixin* in Chinese), Antler (*Lurong* in Chinese), and *Schisandrae fructus* (*Wuweizi* in Chinese). Similarly, the Henan Province THMs, such as Yam (*Shanyao* in Chinese), Achyranthes bidentata (*Nixi* in Chinese), Rehmannia (*Dihuang* in Chinese), and Chrysanthemum (*Juhua* in Chinese), are known as *Huaiyao*. The THMs produced in Zhejiang Province and the coastal continental shelf, such as *Zhejiang fritillaria* (*Zhebeimu* in Chinese), Hangbai chrysanthemum (*Hang baiju* in Chinese), and Scrophularia (*Xuanshen* in Chinese),

are called *Zheyao*. Shanxi Province is situated in an area that lies between the North China Plain and the Inner Mongolia Highlands and is located across the Yellow River system. This advantageous geographical position provides perfect conditions for the growth of GTHMs, including *Astragalus membranaceus* (*Huangqi* in Chinese), *Codonopsis pilosula* (*Dangshen* in Chinses, CP), Forsythiae (*Lianqiao* in Chinese), Polygala (*Yuanzhi* in Chinese), *Radix bupleuri* (*Chaihu* in Chinese), *Scutellaria baicalensis* (*Huangqin* in Chinese), Jujube seed (Suanzaoren in Chinese), Kushen (*Kushen* in Chinese), Hawthorn (*Shanzha* in Chinese), and Peach kernel (*Taoren* in Chinese). CP can be divided into many categories, such as Fang CP (*Fang Dangshen* in Chinese), Wen Yuan CP (*Wenyuan Dangshen* in Chinese), and Tiao CP (*Tiao Dangshen* in Chinese). According to its origin in Shanxi Province, geographical conditions and traits can be divided into two major categories: Lu CP (*Lu dangshen* in Chinese, LCP) and Tai CP (*Tai dangshen* in Chinese, TCP). Both of these categories are derived from *Codonopsis pilosula* (Franch.) Nannf., which belongs to the Campanulaceae family. LCP is cultivated mostly in Changzhi city in the Taihang Mountains area, with the most famous LCP found in Pingshun County. There are two main basins inside the county: the Zhuozhang River and the Wei River. The presence of mountains, hills, and river valleys in the area provides suitable habitat for the growth of the LCP. TCP is produced mainly in Xinzhou in the Wutai Mountain area of Wutai County. The TCP growing area is characterized by a network of intersecting gullies, and the TCP that grows in high-altitude cold mountain locations above 1,500 m is of exceptional quality. Currently, there is a lack of research on the evaluation of real regions via modern analytical techniques and methods.

The large-scale cultivation of LCP has been extensively practised in Shanxi, while TCP predominantly exists as a wild species. The artificial cultivation technology for TCP has not yet been fully developed, and the availability of wild resources for TCP is limited. Admittedly, there have been few comparative studies on the authentic evaluation of the two types of CPs. The present study investigated the relationships of TCP and LCP from Shanxi Province with the environment, secondary metabolites, and THM authenticity via metagenomic, metabolomic, and

comprehensive quality assessment methods. These findings reveal the details of THMs and serve as valuable references for artificial cultivation, site selection, and processing in the Shanxi Province, which specializes in two types of CP.

2 Results

2.1 TCP and LCP geographical environment analysis

The TCP and LCP acquisition sites are specified in Section 2.1, and a comprehensive analysis of the growth environments of the two CP species in these regions was performed. The geographical environment map indicated that the transition from blue to orange depicted the increasing distributions of precipitation, temperature, and sunshine time. The comparison is as follows.

Wutai County is situated in the northeastern part of Shanxi Province between 112°57′41″ and 113°50′56″ east longitude and 38°28′ and 39°4′49″ north latitude. The precipitation in Wutai County shows a spatial distribution that is characterized by unevenness, with a pattern that extends from low in the south to high in the north. The maximum recorded precipitation is 69.5 mm. Precipitation is mostly concentrated in the Wutai Mountain region, particularly in the northern part of Wutai County, where the altitude is relatively high. The annual rainfall in this area can reach approximately 500 mm. The county experiences moderate temperatures during the summer and frigid temperatures during the winter. The primary factor contributing to the temperature variation in the county is the decrease in altitude. The temperature inversely correlates with altitude, exhibiting latitudinal and altitudinal distributions from high in the south to low in the north. The low-temperature area is mainly concentrated in the Wutai Mountain area, which is located at relatively high altitudes in the northern part of the county. The minimum temperature can decrease to as low as -3.14 ℃. The distribution of sunshine duration in Wutai County exhibited spatial disparities, varying from shorter in the south to longer in the north. The regions characterized by extended periods of sunlight were found primarily in the elevated terrain to the north of the county. The annual total duration of sunshine was 2400–2700 h. TCP growth may be facilitated by the natural conditions of Mount Wutai in Wutai County, such as its high altitude of 3061 m, which results in relatively low

temperatures and lengthy periods of sunshine (Fig. 1A–C).

Pingshun County is located in southeastern Shanxi Province and spans from 113°11′45″ to 113°44′04″ east longitude and from 35°56′37″ to 36°27′44″ north latitude. Pingshun County experiences uneven precipitation distributions, with the rainfall density increasing from southwest to northeast. The northeastern mountainous area receives the highest amount of precipitation, reaching 65.38 mm, while the annual precipitation in the county is 584 mm. The temperature in the county is mild in summer and cold in winter, and the temperature gradient increases from south to north. The region with the lowest temperature is located primarily in the southern mountainous area of the county, where the minimum temperature is 7.66 °C. The distribution of sunshine duration varies from east to west, with longer durations predominantly found in the eastern portion of the county. The maximum duration recorded was 6.22 hours, and the yearly average duration of sunshine was 2,517.8 h (Fig. 1D–F).

In conclusion, both Wutai and Pingshun Counties share the same longitude, with a maximum longitude difference of only 53’. TCP thrives in regions characterized by high elevation, low temperatures, approximately 500 mm of rainfall, and extended periods of sunlight. The natural environments of the two locations are similar and are conducive to CP cultivation. However, there is a considerable difference in height within the Wutai Mountain area of Wutai County.

2.2 TCP and LCP quality complies with regulations

Both CPs complied with the requirements stated in the Chinese Pharmacopoeia (2020 edition). An analysis of the properties revealed that the circular patterns beneath the root head of the LCP decreased in thickness as they moved downwards, but the circular patterns beneath the root head of the TCP were denser. Microscopic analysis revealed that the LCP powder was a light yellow colour with many inulins, many fragmented laticiferous vessel containing light yellow particles, and reticular ducts that were easily observable. The similarity of the TCPs suggests that both varieties conform to the specifications outlined in the "Chinese Pharmacopoeia" (2020 edition)

(Fig. 1G–H).

The lobetyolin content in TCP was determined to be 1.68%, whereas that in LCP was found to be 1.17%. These results indicate a higher concentration of codontoside in TCP than in LCP (Fig. 1I).

2.3 The secondary metabolites of LCP and TCP are similar

The total ion chromatograms (TICs) of the QC samples were compared via spectral overlap analysis. The results indicated that the response intensity and rete ntion time of each chromatographic peak were essentially the same. PCA was performed on each sample, revealing a clear separation between LCP and TCP.

The PCA model parameters of TCP compared to LCP, which were obtained after 7-fold cross-validation, yielded $R^2X=0.998$. This finding indicates that the model is reliable and that there are significant differences between th e groups (Fig. 2C).

Simultaneously, OPLS-DA was employed to further validate that the degree of dissociation between TCP and LCP was remarkable and that the experimental results were reliable. By comparing the secondary spectra of the database, the components with relatively high contents of TCP and LCP were identified. The peak labelling results for different species are as follows. The BPC diagram of TCP shows 17 chromatographic peaks (14 in the positive ion mo de and 3 in the negative ion mode), and the BPC diagram of LCP shows 15 chromatographic peaks (11 in the positive ion mo de and 4 in the negative ion mode) (Fig. 2A–B).

Through a comparative analysis of the differential substances found in TCP and LCP, using a screening condition of a response intensity greater than 3 times, one unknown component was found in wild TCP, and two unknown components were determined in LCP. Isovalerylcarnitine, 8-methoxyatractylenolide I, and atractylenolide I were identified as shared compounds f ound in both wild Tai *Codonopsis pilosula* and cultivated Tai *Codonopsis pilosula*. Bufotenine, 3-

192 indolyllactic acid, and sinapyl alcohols are the common compounds in wild varieties o
193 f TCP and LCP. Anileridine, 9-octadecenoic acid, 5,8,11-trihydroxy-, and hexyl 6-O-
194 pentopyranosylhexopyranoside were found in both cultivated TCP and LCP.
195 Lobetyolin and anileridine are common compounds found in all the
196 three varieties of *Codonopsis pilosula*. Both lobetyolin and hexyl 6-O-
197 pentopyranosylhexopyranoside
198 were detected in both positive and negative ion modes (Fig. 2E).
199
200 **2.4 The abundance of microbial flora in TCP nonrooted soil was high**
201 The locations for the collection of TCP and LCP soils are described in Section 2.1.
202 Nonroot soil was investigated through metagenomics. The results revealed that the
203 community in TCP (YF Group) presented more significant changes in diversity and
204 species composition than did that in LCP (PF Group) (Fig. 3A).
205 The analysis of alpha diversity in the soil samples from the two regions revealed
206 similar distribution patterns but varying richness levels. Compared with the LCP
207 samples, the TCP samples presented greater diversity (Fig. 3B).
208 Beta diversity analysis was employed to examine alterations in species
209 composition across different time and spatial scales. Principal component analysis
210 was conducted on bacteria, fungi, and archaea in each sample at the sublevel. The
211 PCoA map clearly shows distinct separation of the YF and PF groups in the bacterial
212 community, indicating significant variations in the soil microbial composition across
213 locations. In the PCoA diagram, there is a noticeable tendency for the YF and PF
214 groups to separate in terms of distribution. The proximity between each group is
215 minimal, and the disparity between the groups is both slight and noteworthy (Fig. 3C).
216 In this study, LEfSe (LDA> 3.5, P<0.05) difference analysis of the microbial
217 community was used to analyse the species with significant differences in the relative
218 abundance of nonroot soil microbial colonies in different regions. LEfSe revealed 101
219 significantly different populations, including 52 in the YF group and 22 in the PF
220 group. (Fig. 3D)
221 To investigate the effects of soil microbial diversity on CP growth in various

production regions, the high-throughput sequencing results of bacteria, fungi, and archaea were annotated to 2663, 201, and 171 genera, respectively. The top 10 genera with clear annotation information were defined as the dominant genera.

The abundances of the bacterial genera Branchiibius, Candidatus Rokubacteria, Chloroflexi, Gemmatimonadetes, and Verrucomicrobia in the YF group significantly increased, and the abundances of Betaproteobacteria, Sphingosinicella, and Verrucomicrobia in the PF group significantly increased. The YF group presented a considerable increase in the abundance of the fungal genera Rhizophagus, Aspergillus, and Fusarium. Similarly, the PF group presented significant increases in the abundances of Rhizophagus, Fusarium, and Diversispora. The PF group presented an increase in the abundance of Thermoplasmata and Nitrosopumilales, whereas the abundance of the remaining genera decreased (Fig. 3E).

Moreover, the functional metabolism of microorganisms in non-root system soil was analysed via eggNOG annotation. The eggNOG functional classification, which is based on microorganisms from various production areas, revealed that the most abundant categories were function unknown, energy production and conversion, amino acid transport and metabolism. This process is followed by signal transduction mechanisms, translation, ribosomal structure and biogenesis, cell wall/membrane/envelope biogenesis, and transcription. The least abundant categories were nuclear structure, RNA processing and modification.

When the PF group was compared with the YF group, there was a significant increase in the abundance of genes related to RNA processing and modification, replication, recombination and repair, translation, ribosomal structure and biogenesis, and energy production and conversion in the PF group. On the other hand, the YF group presented an increase in carbohydrate transport and metabolism, inorganic ion transport and metabolism, translation, ribosomal structure and biogenesis, and amino acid transport and metabolism (Fig. 3F).

Finally, the degree of species variation in nonroot system soils may differ across different regions. Among all the groups, the YF group presented the greatest number of different colonies, suggesting that the soil microbial diversity of TCP was both

abundant and distinctive. The expression of functional genes in soil bacteria from various places exhibited similar patterns, although the relative abundance of individual genes varied.

2.5 The abundance of microbial flora in TCP-rooted soil was high

The analysis of soil microorganisms in the non-root system soil revealed that the community of TCP (YG group) presented more pronounced changes in community diversity and species composition than did the communities of LCP (PG group).

Alpha diversity analysis indicated that the bacterial populations in the soil samples from the root system soils in the two regions presented similar distribution patterns but varied in terms of abundance. However, there were distinct variations in the distribution and abundance of bacterial populations in the soil across different regions. The diversity of the soil microflora in TCP (YG group) was greater than that in the LCP soil (PG group). Beta diversity analysis revealed that the principal component analysis of bacteria, fungi and archaea in each sample was performed at the genus level, indicating that the YG and PG groups were obviously separated on the PCoA diagram in terms of the community distributions of bacteria, fungi and archaea, indicating that there were significant differences in the structure of rhizosphere soil microbial colonies in different regions. LEfSe (LDA>3.5, P<0.05) was employed to analyse the species that exhibited significant differences in the relative abundance of root soil microbial colonies across several locations. LEfSe analysis revealed 171 distinct bacterial populations at the genus level, including 81 in the YG group and 37 in the PG group (Fig. 4A–C).

In the analysis of nonroot soil microorganisms described in Section 3.4, a new strain of Ascobolus was identified among the fungal species in the YG group (Fig. 4D).

The soil metabolism in the root system was analysed via eggNOG annotation. Compared with the YG group, the G group presented significant increases in energy production, inorganic ion transport, and nuclear structure (Fig. 4E).

Ultimately, the degree of variation in root soil composition among different

places varies. Among all the groups, the YG group presented the greatest variation in the number of colonies, suggesting that the soil microbial richness of TCP was high and possessed distinct characteristics. The expression of functional genes in soil bacteria from various places exhibited similar overall patterns, although the relative abundance of specific genes varied.

2.6 Ecological function analysis

The top 10 genes were identified as the dominant species by analysing the nitrogen cycling of non-root and root system soils at the genus level, for example. The abundances of different species increased or decreased to different degrees during nitrogen cycling. Nitrospira and Bradyrhizobium are likely the primary genes involved. All of them exhibited an increasing pattern during nitrogen cycling, as depicted below (Fig. 5A–B)

2.7 Horizontal abundance of TCP metallic elements

The findings of the examination of metallic elements in the soil are presented in Table 1. A comparison of the metallic element data of the two groups of non-root system soils revealed that the contents of Ti, Mg, Cr ($P < 0.01$), and Ni ($P < 0.001$) in the YF group were significantly greater than those in the PF group. The contents of Al ($P < 0.05$), Ca, Cu, Mn, V and Ba ($P < 0.001$) in the PF group significantly increased. The contents of the other elements in each group were identical. The content of metallic elements in Group F was the highest, and the types were the most abundant.

Analysis of the metallic element data from the two groups of root system soils clearly revealed that the YG group presented significantly higher concentrations of Ti, Zn, Ni, and Cr ($P < 0.001$) compared with the PG group. The PG group presented increased levels of K, Na, Mn, Mo ($P < 0.001$), Al ($P < 0.05$), Ca, and V ($P < 0.01$), whereas the concentrations of the remaining elements in the other groups were similar. The YG group presented the highest concentrations of metallic elements and the most diverse range of metallic elements.

312 **Table 1 Metallic element 1 ($\bar{x}\pm s$, n=6)**

Metallic elements (mg/g)	YF group	PF group	YG group	PG group
K	10.18±0.42**	13.6±0.31***	10.28±0.61	14.44±0.94***
Ti	1.78±0.01***	1.65±0.01***	1.82±0.05***	1.7±0.08***
Na	1.03±0.19***	1.85±0.02***	1.07±0.1*	2.06±0.16***
Al	27.55±0.67***	43.99±0.4*	27.66±0.64***	46.86±1.51*
Ca	1.97±0.2***	4.17±0.23***	2.08±0.2***	7.39±0.55**
Mg	8.83±0.4***	8.2±0.21***	8.87±0.66***	8.41±0.37***
Fe	20.67±0.6*	25.86±0.24***	20.96±0.91	26.77±1.41***
Cu	46.79±0.85***	65.75±0.68***	45.74±1.45***	71.44±1.8*
Zn	127.35±7.24***	128.32±3.19***	159.3±5.19	123.59±5.24
Mn	649.42±21.63***	725.15±4.35***	684.17±10.06***	734.78±15.86***
V	65.23±1.04***	88.86±0.26***	67.15±1.47***	93±1.63***
Co	12.66±0.2***	13.8±0.17***	12.55±0.63*	14.93±0.63***
Ni	63.67±1.08***	37.49±0.49***	65.45±1.43***	14.93±0.63***
Mo	0.94±0.04***	0.91±0.03***	0.91±0.03	1.17±0.09***
Ba	198.32±5.07***	306.25±0.74***	202.55±2.67***	324.93±11.9**
Sb	0.38±0.01	0.39±0.01	0.39±0.01	0.39±0.01
Cr	155.2±2.61	68.78±0.77	165.14±2.46***	72.49±1.82***
Se	0.51±0.05***	0.57±0.01***	0.43±0.05	0.5±0.04
Sr	60.27±1.72	93.13±0.53*	63.67±0.89***	79.34±2.21***

313 Note: Compared with the TF group, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$; Compared with the TG

group, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

3 Correlation analysis

Through the study of the correlation of soil bacteria and metabolic material, in the top 20 bacterial groups, Bradyrhizobiaceae, Bradyrhizobium, Hyphomicrobiaceae, Rhodoplanes, Gaiellaceae, Gaiella, Streptomycetaceae, Streptomyces, Phyllobacteriaceae, Mesorhizobium and dangshen Bufotenine, and hexyl 6-O-pentopyranosylhexopyranoside components were positively correlated and negatively correlated with 1-isopropylbenzimidazole-2-sulfonic acid, 8-benzylcanadine, and 1-(2-chlorobenzyl)pyrrole-2-carboxamide in CP. Sphingomonadaceae, Sphingosinicella, Solirubrobacteraceae, Solirubrobacter, Geodermatophilaceae, Blastococcus, Gemmatimonadaceae, Gemmatirosa, Sinobacteraceae, Steroidobacter, Sphingomonadaceae, Sphingomonas, Xanthobacteraceae, Pseudolabrys and 1-isopropylbenzimidazole-2-sulfonic acid, 8-benzylcanadine, and 1-(2-chlorobenzyl) pyrrole-2-carboxamide components were positively correlated; however, with the dangshen Bufotenine peak, hexyl 6-O-pentopyranosylhexopyranoside was negatively correlated. Through correlation analysis of metal elements and metabolic substances, the contents of the metal elements zinc, Mg, Mn, Ti, Mo, Co, Ni, and Cr in the soil and the hexyl dangshen composition 6-O- pentopyranosylhexopyranoside and Bufotenine were positively correlated and negatively correlated with the remaining dangshen composition. K, Al, Ca, Na, Cu, Ba, Cd, V, Sr and lobetyolin, alpha-D-xylopyranose, 8-benzylcanadine, 1-isopropylbenzimidazole-2-sulfonic acid, 7-methoxy-2-propylquinolin-4-ol, and 1-(2-chlorobenzyl) pyrrole-2-carboxamide were positively correlated and negatively correlated with the other components of Codonginseng (Fig. 6 A–C).

4 Discussion

CP, a medicinal food that is homologous to traditional Chinese medicine, has been widely utilized in clinical practice, health product development, and dietary recommendations since ancient times. The primary sources of CP include three

varieties: *Codonopsis pilosula* (Franch.) Nannf., *Codonopsis pilosula* Nannf. var. *modesta* (Nannf.) L. T. Shen, and *Codonopsis tangshen* Oliv. The quality of these herbs is closely linked to their origin and environmental conditions during growth. Comparisons of precipitation and temperature can help determine the quality of CP. From a soil microbial perspective, various components displayed comparable trends in functional gene expression among soil bacteria, with variations in the relative abundance of specific genes. Previous research has consistently demonstrated a greater prevalence of ring markings in TCPs than in LCPs, with empirical data corroborating this finding. Additionally, the examination of traits revealed a potential link between the horizontal grain pattern observed on the root head rings of TCP and their capacity to acclimate to low-temperature plateaus.

The present study also included an analysis of TCP and LCP cultivated in mountainous regions at a longitude of 113°. Microscopic identification revealed the presence of stone cells, cork cells, latex tubes, and inulin. Targeted metabolomic analysis identified common secondary metabolites such as 9-octadecenoic acid, 5,8,11-trihydroxy, and 6-O-hexyl pentopyranosylhexopyranoside for TCP and LCP. Furthermore, investigations of the soil microbial flora led to the discovery of a new fungal genus, Ascobolus, in rhizosphere soil associated with TCP.

In addition to natural environmental conditions, soil properties play a significant role in reflecting soil fertility and serve as the primary source of nutrients for plants, thereby influencing the quality of medicinal materials. Previous studies have shown that various physical and chemical indices of soil content considerably affect the metabolism and accumulation of the main active components of plants. Our research revealed positive correlations between the levels of specific bacterial species (Blastococcus, Gemmatirosa, Steroidobacter, and Lysobacter) and metal elements (Na, Ca, and Sr) with the level of lobetyolin in CP.

In contrast, Bradyrhizobium bacteria displayed adverse associations with the key efficacy component Lobetyolin, which was likely attributable to lower metabolic accumulation rates. Consequently, increasing the abundances of Blastococcus, Gemmatirosa, Steroidobacter, and Lysobacter bacterial species, in conjunction with

increasing Na, Ca, and Sr metal concentrations during CP cultivation, may increase both potency levels and overall quality. These research results provide a foundation for the systematic evaluation of CP production sites and cultivation techniques to meet high-quality standards.

This study examined the distinctions between TCP and LCP with respect to their shape characteristics, content, growth environment perspective, soil composition, and flora. The findings suggest that, with the exception of differences in acetylene CP content, the remaining variations are negligible. The TCP and LCP identified in the Shanxi Province area serve as large reservoirs of GTHMs, exhibiting attributes similar to those of soil flora and microorganisms. In particular, TCP presented a more concentrated ring of horizontal grains and increased density at higher elevations, as did the differences in soil genus levels and acetylene content in CP roots.

5 Conclusion

TCP and LCP share similar geographical environments for growth. However, TCP features a dense ring pattern beneath the stalk, a high content of Lobetyolin, and a new flora, Ascobolus, in the root soil.

6 Materials and methods

6.1 Materials

On the basis of the findings of the fourth Chinese materia medica resource survey, Wutai County in Shanxi Province was selected as the representative area for TCP sampling, and Pingshun in Changzhi city was designated for LCP collection. The specific sampling sites are detailed in Table 1.

The soil samples were collected in accordance with the Association Standard of the People's Republic of China-2022 technical specifications for soil sampling, preparation and long-term storage (T/GSSF004). With respect to the S-shaped route, 10 soil samples and root samples were obtained from several sites in the sampling region, all of which had the same position and maturity. The soil samples from the rhizosphere and root system were obtained from four different vertical directions, approximately 1 m from the root, and were then deposited in sterile centrifuge tubes.

The samples were promptly immersed in liquid nitrogen and preserved in dry ice, after which they were stored at -80 ℃.

The fresh CP samples from which the soil was collected were removed, and the various parts were then separated. These pieces were placed in sterile self-sealing bags, numbered, promptly immersed in liquid nitrogen, preserved on dry ice and finally stored -80 ℃. The collected samples were assigned numbers, and the precise information is displayed in Table 2.

The CP was authenticated by Professor Zhang Shuosheng from Shanxi University of Chinese Medicine as the dried root of *Codonopsis pilosula* (Franch.) Nannf. of the Campanulaceae family. The voucher samples were stored in the plant specimen room of the Shanxi College of Traditional Chinese Medicine (specimen number: SXTCM-Zhang-2023001–2023012).

Table 2 Numbering of the soil and root samples

Col. Time	Col. Loc.	Lat/Long	Group	No. of no-root sym. so. smp.	No. of root sym. so. smp.	No. of root smp.
2023.08.11	Heiyatang Vlg., Menjiaoshi Twp., Wutai Cty., Shanxi	113°43′47.73″ 38°48′17.40″	YF group	TCP - NR1	TCP - R1	TCP -1
				TCP - NR2	TCP - R2	TCP -2
				TCP - NR3	TCP - R3	TCP - 3
				TCP - NR4	TCP - R4	TCP - 4
	Wengzigou Vlg., Menjiaoshi Twp., Wutai Cty., Shanxi	113°58′18.55″ 38°70′63.89		TCP - NR5	TCP - R5	TCP - 5
				TCP - NR6	TCP - R6	TCP - 6
2023.08.12	Xincheng	113°29′11.1714″	PF	LCP -	LCP - R1	LCP -

Vlg., Pingshun city, Changzhi, Shanxi	36°00′50.0110″	group	NR1		1
			LCP - NR2	LCP - R2	LCP - 2
			LCP - NR3	LCP - R3	LCP - 3
Zhaodian Vlg., Pingshun Cty, Changzhi, Shanxi	113°44′12.43″ 36°14′508702°		LCP - NR4	LCP - R4	LCP - 4
			LCP - NR5	LCP - R5	LCP - 5
			LCP - NR6	LCP - R6	LCP - 6

6.2 Instruments and reagents

Instruments: The instruments used in this study were as follows: a Q Exactive HF-X mass spectrometer (Thermo Fisher Technology Co., Ltd., CA, USA); a Vanquish UHPLC (Thermo Scientific, Waltham, MA); a low-temperature high-speed Eppendorf 5430R centrifuge (Eppendorf Innovation Co., LTD., GER); a Mettler–Toledo one-thousandth LE204E balance (METTLER TOLEDO Co., USA); a NexlON 5000 ICP-MS (PerkinElmer Co., LTD, CA, USA); an Agilent 710 ICP-OE (Agilent Technologies, Palo Alto, Calif.); a Qubit 2.0 fluorometer (r Life Technologies, CA, USA); an Agilent 2100 bioanalyzer (Agilent Technologies, Palo Alto, Calif.); CFX96 Touch Real-Time PCR (Bio-Rad, Co., CA, USA); a KZ-III-F homogenizer (Servicebio Technology Co., Ltd., Wuhan, CN); an MX-S vortex mixer (Dalong Xingchuang Experimental Instrument Co., Ltd., Beijing, CN); an SB-5200 ultrasonic cleaner (Xinzhi Biotechnology Co., Ltd., Ningbo, CN); and a vacuum freeze dryer (Kangyibo Medical Devices Co., Ltd., Beijing, CN).

Chromatographic column: ACQUITY UPLC HSS T3 (2.1 mm×100 mm, 1.8 μm, Waters Co., Ltd., MA, USA).

Reagents: Ultra-pure water (MS grade, W6-4, Thermo Fisher Technology Co., LTD, CA, USA); acetonitrile (MS grade, A955-4, Thermo Fisher Technology Co.,

Ltd., CA, USA); methanol (MS grade, A456-4, Thermo Fisher Technology Co., LTD, CA, USA); formic acid (GR, F0654-25 ml, Tixi Ai Chemical Industry Development Co., Ltd., Shanghai); hydrogen peroxide (GR, 20240103, National Pharmaceutical Group Chemical Reagent Co., Ltd., Shanghai, CN); concentrated nitric acid (GR, 20231220, National Pharmaceutical Group Chemical Reagent Co., Ltd., Shanghai, CN); and hydrofluoric acid (GR, C12897954, McLean Biochemical Technology Stock Co., Ltd., Shanghai, CN).

Test kits: Magnetic bead-based soil and faeces genomic DNA Extraction Kit (TIANGEN Biochemical Technology Co., Ltd., CN); RNase A (10 mg/ml, Cat: RT405-02, TIANGEN Biochemical Technology Co., Ltd., CN); Qubit dsDNA Assay Kit (Q32854, Life Technologies, CA, USA); and NEBNext Ultra DNA Library Prep Kit for Illumina (NEB, Ipswich, MA, USA).

6.3 Software and database

Geographical Database: National Tibetan Plateau Data Centre (http://data.tpdc.ac.cn).

Geographic information system: ArcGIS software (version 10.8, http://www.esri.com/zh-cn/home).

Bioinformatics software: fastp (https://github.com/OpenGene/fastp), Prodigal (https://github.com/hyattpd/Prodigal/releases/tag/v2.6.3), Bowtie2 (Bowtie download|SourceForge.net), DIAMOND (version 0.9.110, http://www.crystalimpact.com), MEGAN (https://software-ab.informatik.uni-tuebingen.de), Gephi (https://gephi.org), BWA (http://maq.sourceforge.net), cBot cluster generation system (https://support.illumina.com.cn), CD-HIT (https://github.com/weizhongli/cdhit), and ARG-OAP (http://smile.hku.hk/ARGs/indexing).

Functional annotation database: Non-redundant protein database (NR database, version 2021.11, wget-c fto://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr.gz), Kyoto Encyclopedia of Genes and Genomes database (KEGG database, http://www.genome.jp/), evolutionary genealogy of genes: Non-supervised

Orthologous Groups database (eggNOG database, http://eggnog5.embl.de), and Carbohydrate-Active enZYmes database (CAZy database, http://www.cazy.org).

6.4 Natural environmental data processing methods

The natural environmental data for Wutai and Pingshun Counties in Shanxi Province, including precipitation, sunshine, and temperature data, were obtained from the National Tibetan Plateau Data Centre. The acquired data were subsequently analysed via ArcGIS software. The processing methodology employed was as follows. A total of ten nc-formatted datasets spanning from 2012 to 2022 were selected, with each dataset containing twelve months of data. The mean temperature data for these years were extracted and stored in .tif format. Additionally, the R language was utilized to retrieve sunshine duration data from the Global Surface Summary of the Day (GSOD) meteorological stations in China covering the period between 2012 and 2022. The IDW inverse weight matrix was utilized for data interpolation, with the inclusion of only 5 adjacent stations to mitigate excessive smoothing. A downscaling operation using the delta algorithm was applied to increase the precision postinterpolation. The final raster data had a spatial resolution of 0.1°*0.1°.

6.5 UPLC–MS/MS analysis[1]

CP root sample preparation: After the sample was pulverized in liquid nitrogen, 100 mg of the resulting powder was weighed and placed in a 1.5 mL centrifuge tube. Then, 1 mL of 70% methanol was added to the tube, and the mixture was vigorously mixed for 30 s by vortexing. The tube was subsequently subjected to sonication in a water bath for 90 min, followed by centrifugation for 10 min at a speed of 16,000 × *g* at a temperature of 4 ℃. The supernatant was transferred to a 96-well protein filter plate and filtered with nitrogen under positive pressure, and the filtrate was transferred to a 2-mL EOP tube and dried under vacuum. Next, 300 μL of 40% methanol solution was added to the lyophilized sample to dissolve it. The mixture was then vigorously agitated for 30 s by vortexing, followed by centrifugation for 10 min (16000 × *g*, 4 ℃), and the supernatant was collected.

Chromatographic conditions: The samples were separated via ultrahigh-

performance liquid chromatography (UHPLC) in combination with an ACQUITY UPLC HSS T3 column (2.1 mm × 100 mm, 1.8 μm) with the column temperature maintained at 35 °C and a flow rate of 0.3 mL/min. The elution gradient comprised mobile phase A, which was a 0.1% aqueous solution of formic acid, and mobile phase B, which was a 0.1% acetonitrile solution of formic acid. The process of gradient elution was executed in accordance with the information provided in Table 3.

MS conditions: The primary and secondary spectra of the samples were collected via a Q Exactive HFX mass spectrometer, which was connected to a UHPLC instrument and equipped with an electrospray ionization (ESI) source operating in both positive and negative ionization modes. The spray voltage was 3800 V (ESI+)/3500 V (ESI-), the sheath gas flow rate was 40 L/min, the ion transfer tube temperature was 320 ℃, and the atomization temperature was 350 ℃. The detection mode was full-MS/dd-MS2 mode, MS/MS spectra were obtained from the top 10 MS1 ions, and the collision energy was normalized by stepwise energy levels 20, 40, and 60. The mass–charge ratio of the first stage was scanned from 90 to 1300.

Table 3 UHPLC gradient elution

Time (min)	Mobile phase A (%)	Mobile phase B (%)
Initial	95	5
17.0	2	98
17.2	95	5
20	95	5

6.6 Properties and microscopic identification

The CP properties were determined in accordance with the criteria specified for the CP traits in the Chinese Pharmacopoeia (2020 edition). The microscopic identification was conducted following the guidelines outlined in the Microscopic Identification Method 2001 of the Chinese Pharmacopoeia (2020 edition).

6.7 HPLC method[2]

Chromatographic conditions: The analysis was performed using a Hypersil GOLD aQ column at a temperature of 30 ℃. The detection wavelength was set to 220 nm. The mobile phase consisted of acetonitrile (phase A) and an aqueous solution containing 0.3% phosphoric acid (phase B). Gradient elution, as shown in Table 1, was employed for separation. The flow rate was maintained at 1 mL/min, and the injection volume was set to 10 μL.The process of gradient elution was executed in accordance with the information provided in Table 4.

Table 4 HPLC gradient elution

Time/min	A/%	B/%
0	10	90
10	25	75
30	60	40
35	10	90
40	10	90

Three milligrams of the reference substance codontoside was precisely weighed and dissolved in 75% methanol. The solution was transferred to dry and clean volumetric flasks with a capacity of 10 mL, ensuring a constant volume. The mixture was shaken well to obtain the reference substance solution. Each reference solution was removed into the same volumetric flask and diluted with 75% methanol to prepare the reference mixture.

6.8 Metagenomic sequencing

DNA extraction and quality control: DNA was extracted from the soil samples via a magnetic soil and stool DNA kit (Cat: DP712, TIANGEN Biochemical Technology Co., Ltd., CN). The DNA concentration was determined via a Qubit dsDNA Assay Kit with a Qubit 2.0 fluorometer (Life Technologies, CA, USA). Only the DNA that successfully underwent the quality check was eligible for use in the construction of

the metagenomic libraries. The NEBNext Ultra DNA Library Prep Kit for Illumina (NEB) was used to create metagenomic libraries. The attribute sequence of each sample was then supplemented with an index. The index sample generated clusters on the cBot cluster generation system. Following the generation of clusters, library preparation was carried out via a NovaSeq 6000.

Bioinformatics analysis: Data generated from the Illumina platform were used for bioinformatics analysis. The software and parameters are as follows.

Sequencing results preprocessing: The default parameters of fastp[3] software were utilized to conduct preprocessing on the raw data acquired from the Illumina sequencing platform. The splicing software MEGAHIT[4] was used to assemble the clean data. A de Brujin graph was constructed according to the overlap relationship between k-mers, resulting in the generation of contigs. Contigs exceeding a length of 800 bp were selected for data statistics and further analysis.

Gene prediction and abundance analyses: Prodigal[5] was used to predict the ORF of the spliced contig sequence and then translate it into an amino acid sequence. The ORF prediction results were subsequently used to calculate the abundance information for genes in the corresponding samples. The abundance information, measured in transcripts per million (TPM), for each gene in each sample was computed via the following formula (where r represents the number of reads aligned to the gene and L represents the length of the gene)[6-10]:

$$G_k = \frac{r_k}{L_k} \cdot \frac{1}{\sum_{i=1}^{n} \frac{r_i}{L_i}}$$

.

Species annotation: DIAMOND[11] (V0.9.9.110) was used to align the sequences of bacteria, fungi, Archaea and viruses. The alignment results of each sequence were filtered to include only those with an eigenvalue ≤1e-5[12]. The species annotation information of the sequence was then determined via the LCA algorithm, which was applied to the taxonomic assignment of the MEGAN[13]. On the basis of

the abundance tables at each taxonomic level, Krona analysis[14] was used to determine the relative abundance profile. Gephi software was then employed to create a species association network map, with a threshold set at a species correlation coefficient >0.7 and $P<0.01$.

Functional annotation: DIAMOND (Version 0.9.9.110) was used to compare the nonredundant gene sets with the KEGG[15, 16], eggNOG[17] and CAZy[18] databases. For each sequence alignment result, the selected BestBlastHit results were utilized for further analysis[19-21]. From the results of the alignment, the relative abundance of different functional levels was determined. The number of annotated genes was counted, and the relative abundance profiles are displayed.

Statistical analysis: PCA[22] (Rade4 package, Version 2.15.3) and NMDS [23](Rvegan package, Version 2.15.3) dimension reduction analyses were applied to compare the species and functional compositions between different groups. ANOSIM and Adonis were employed to assess the dissimilarities between the groups. Then, the Wilcoxon (two-group)/Kruskal–Wallis (multigroup) rank sum test (or STAMP analysis)[24] and LEfSe analysis[25] were used to identify distinct species or functions at each level between the groups.

6.9 Metallic element detection[26, 27]

After 0.1 g of the sample was weighed precisely, it was placed into the digesting inner tank. Then, 5 mL of concentrated nitric acid was added, and the mixture was left undisturbed overnight. The next day, 2 mL of hydrogen peroxide and 1 mL of hydrofluoric acid were introduced into the inner tank, and a stainless-steel jacket was firmly sealed. The tank was placed in a constant-temperature drying oven for digestion at 150–170 °C for 4 h. After cooling, the stainless-steel jacket was carefully unscrewed, and the digesting inner tank was subsequently removed. The inner tank was heated on an electric hot plate at 160 °C for 30 min, 1% nitric acid was added to 25 mL, and the content of each element was measured by the instrument. The results were calculated via the following formula:

$$\text{Content of metallic element (mg/kg)} = C*V/m*D,$$

where C is the concentration of elements in the solution, mg/L; V is the extraction

volume, mL; D is the dilution factor; and m is the sample mass, g.

Abbreviation

Abbreviation	Full name
nc format	Network Common Data Format
IDW	Inverse Distance Weighted
GSOD	Global Surface Summary of the Day
MS	Mass Spectroscopy
EIS	Electrochemical impedance spectroscopy
HPLC	High Performance Liquid Chromatography
DNA	Deoxyribonucleic acid
ORF	open reading frame
LCA	Lowest Common Ancestor
KEGG	Kyoto Encyclopedia of Genes and Genomes
NMDS	Nonmetric multidimensional scaling
BPC	Base Peak Chromatogram
PCoA	Principal co-ordinates analysis
LEfSe	Linear discriminant analysis effect size
EggNOG	Evolutionary gene genealogy Nonsupervised Orthologous Groups
UPLC-MS/MS	Ultraperformance liquid chromatography tandem mass Spectrometry
PCA	Principal component analysis
VIP	Variable importance in projection

Supplementary Information

All date generated and analysed during this study are included in this pulished article and its supplementary information files.

Acknowledgements

Thanks to the village committee of Heiyatang Vlg., Menjiaoshi Twp., Wutai Cty., Wengzigou Vlg., Menjiaoshi Twp., Wutai Cty., Xincheng Vlg., Pingshun city and Zhaodian Vlg., Pingshun Cty for the help in collecting soil and *Codonopsis pilosula* samples needed in this experiment, and thanks for the permission for these samples collection.

The authors also thank the National Sect. Plateau Data Center (http://data.tpdc.ac.cn), and the process (https://www.processon.com) provides part of the graphics data.

Authors' contributions

Y.W. conceived and designed the study; Y.W., X.H. and L.B helped performed the experiments; Y.W. analyse the data； Y.W. and T.L. drew a picture of the article； Y.W. wrote the manuscript; X.M. and S.Z. modified the manuscript. All authors have read and agreed to the published version of the manuscript.

Declarations

Ethics approval and consent to participate

All local, national or international guidelines and legislation were adhered to in the production of this study. Permission to collect plant specimens has been obtained from Changzhi City and Xinzhou City of Shanxi Province for this study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[a]College of Chinese Materia Medica and Food Engineering, Shanxi University of Chinese Medicine, Jinzhong 030619, Shanxi, China

[b]Shanxi Key Laboratory of Traditional Herbal Medicines Processing, Jinzhong 030619, Shanxi, China

[c]Chinese herbal medicine processing inheritance and innovation key research laboratory

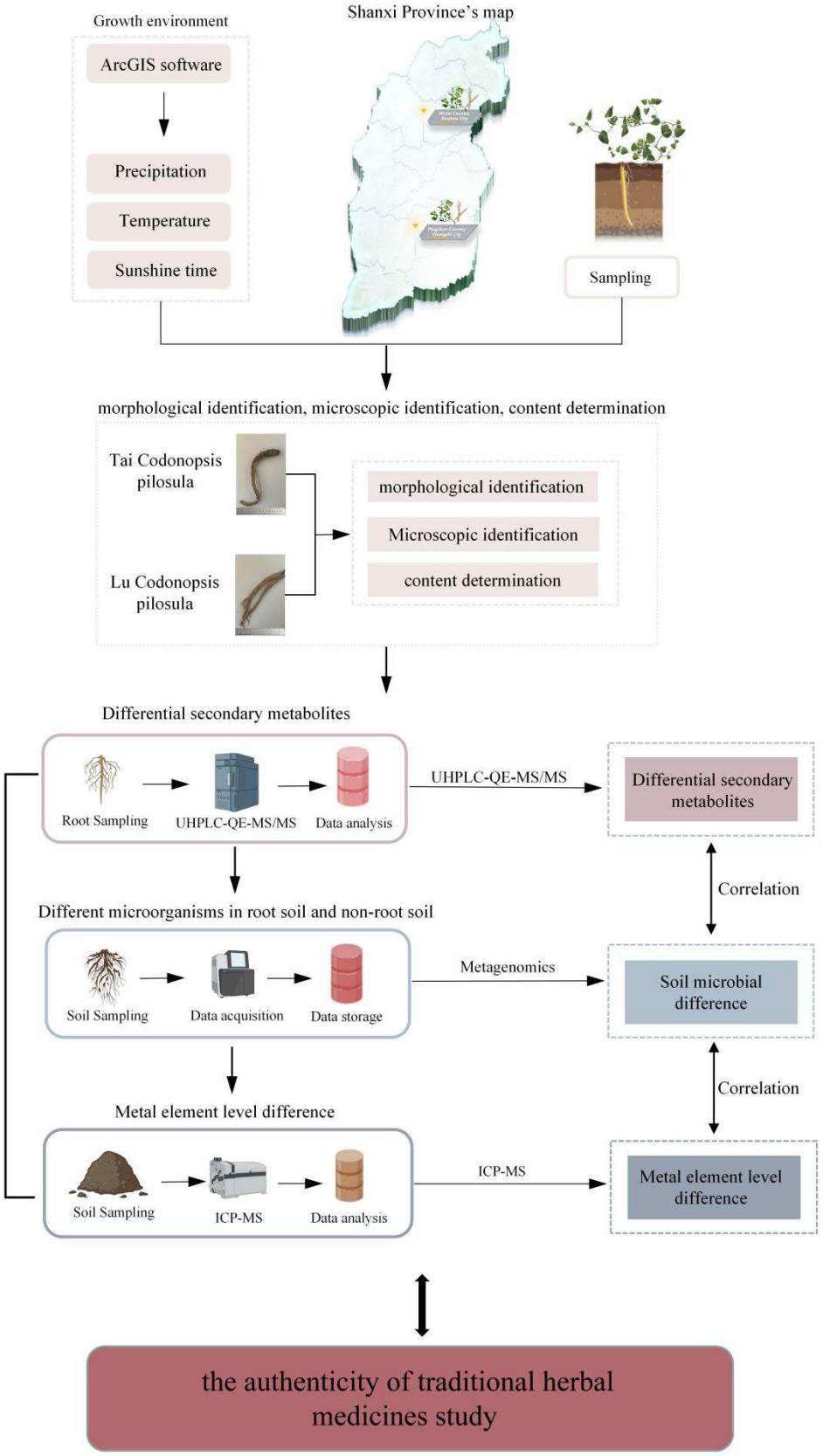
Funding

This research was funded by the characterize and correlate the changes in the properties of medicinal materials during the manufacturing process of "fire/temperature" in Shanxi Province, China (CZ2023041_019). Additionally, this research aimed to demonstrate the research and application of these materials, establish a TCM processing technology foundation, conduct a survey of TCM resources (CZ2023016), establish the Shanxi Key Laboratory of Chinese Medicinal Materials Processing (202104010910029), develop the Shanxi Modern Agricultural Industry Traditional Chinese Medicine Technology System (CZ2023018), collaborate with Jinyao Comprehensive Development and Utilization Collaborative Innovation Center (CZ2023056_003)(XN2024040_006), support pharmaceutical processing discipline construction at Shanxi University of Chinese Medicine (2023XKJS-22) through their innovation team (2022TD1014) , and the Shanxi Key Laboratory of Traditional Chinese Medicine Processing (Shanxi Medicine Comprehensive Development and Utilization Collaborative Center) (XN2024040_006), and Construction of TCM processing discipline (XN2024093_024) with the Chinese Medicine - Key Research Laboratory (CZ2024010_002)

Reference

1. Kim HW, Wang M, Leber CA, Nothias LF, Reher R, Kang KB, van der Hooft JJJ, Dorrestein PC, Gerwick WH, Cottrell GW: NPClassifier: A Deep Neural Network-Based Structural Classification Tool for Natural Products. J Nat Prod 2021, 84(11):2795-2807.
2. Zhang Hui, Shi Shao-ming, WANG Zhi-Qi, Yang Lu-Jie, Zhang Xin-Xin: Comparison of the content of active components of two Codonopsis Codonopsis under soil culture and hydroculture.Journal of Northwest Normal University (Natural Science Edition)2024, 60(02):86-92.
3. Chen S, Zhou Y, Chen Y, Gu J: fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018, 34(17):i884-i890.
4. Li D, Liu CM, Luo R, Sadakane K, Lam TW: MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 2015, 31(10):1674-1676.
5. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ: Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010, 11:119.
6. Karlsson FH, Fåk F, Nookaew I, Tremaroli V, Fagerberg B, Petranovic D, Bäckhed F, Nielsen J: Symptomatic atherosclerosis is associated with an altered gut metagenome. Nat Commun 2012, 3:1245.
7. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T et al: A human gut microbial gene catalogue established by metagenomic sequencing. Nature 2010, 464(7285):59-65.
8. Villar E, Farrant GK, Follows M, Garczarek L, Speich S, Audic S, Bittner L, Blanke B, Brum JR, Brunet C et al: Ocean plankton. Environmental characteristics of Agulhas rings affect interocean plankton transport. Science 2015, 348(6237):1261447.
9. Cotillard A, Kennedy SP, Kong LC, Prifti E, Pons N, Le Chatelier E, Almeida M, Quinquis B, Levenez F, Galleron N et al: Dietary intervention impact on gut microbial gene richness. Nature 2013, 500(7464):585-588.
10. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M, Arumugam M, Batto JM, Kennedy S et al: Richness of human gut microbiome correlates with metabolic markers. Nature 2013, 500(7464):541-546.
11. Buchfink B, Xie C, Huson DH: Fast and sensitive protein alignment using DIAMOND. Nat Methods 2015, 12(1):59-60.
12. Oh J, Byrd AL, Deming C, Conlan S, Kong HH, Segre JA: Biogeography and individuality shape function in the human skin metagenome. Nature 2014, 514(7520):59-64.
13. Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC: Integrative analysis of environmental sequences using MEGAN4. Genome Res 2011, 21(9):1552-1560.
14. Ondov BD, Bergman NH, Phillippy AM: Interactive metagenomic visualization in a Web browser. BMC Bioinformatics 2011, 12:385.
15. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M: From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res 2006, 34(Database issue):D354-357.

16. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M: Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res 2014, 42(Database issue):D199-205.
17. Powell S, Forslund K, Szklarczyk D, Trachana K, Roth A, Huerta-Cepas J, Gabaldón T, Rattei T, Creevey C, Kuhn M et al: eggNOG v4.0: nested orthology inference across 3686 organisms. Nucleic Acids Res 2014, 42(Database issue):D231-239.
18. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B: The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res 2009, 37(Database issue):D233-238.
19. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, Arumugam M, Kultima JR, Prifti E, Nielsen T et al: An integrated catalog of reference genes in the human gut microbiome. Nat Biotechnol 2014, 32(8):834-841.
20. Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, Zhang D, Xia H, Xu X, Jie Z et al: Gut microbiome development along the colorectal adenoma-carcinoma sequence. Nat Commun 2015, 6:6528.
21. Bäckhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, Li Y, Xia Y, Xie H, Zhong H et al: Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. Cell Host Microbe 2015, 17(6):852.
22. Avershina E, Frisli T, Rudi K: De novo semi-alignment of 16S rRNA gene sequences for deep phylogenetic characterization of next generation sequencing data. Microbes Environ 2013, 28(2):211-216.
23. Noval Rivas M, Burton OT, Wise P, Zhang YQ, Hobson SA, Garcia Lloret M, Chehoud C, Kuczynski J, DeSantis T, Warrington J et al: A microbiota signature associated with experimental food allergy promotes allergic sensitization and anaphylaxis. J Allergy Clin Immunol 2013, 131(1):201-212.
24. Parks DH, Tyson GW, Hugenholtz P, Beiko RG: STAMP: statistical analysis of taxonomic and functional profiles. Bioinformatics 2014, 30(21):3123-3124.
25. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C: Metagenomic biomarker discovery and explanation. Genome Biol 2011, 12(6):R60.
26. Wang B H, Ma Z H, Fu W L: Determination of heavy metals in soil by atomic absorption spectrometry in sealed high-pressure digestion tank. In: International Symposium on Quality and Safety Management, Testing and Traceability of Agricultural Products: 2008; Beijing, China. 5.
27. Lu Rukun: Methods for agricultural chemical analysis of soil. In.: Beijing: China Agricultural Science and Technology Press; 2000.


Growth environment
ArcGIS software
Precipitation
Temperature
Sunshine time
Shanxi Province's map
Sampling
morphological identification, microscopic identification, content determination
Tai Codonopsis pilosula
Lu Codonopsis pilosula
morphological identification
Microscopic identification
content determination
Differential secondary metabolites
Root Sampling
UHPLC-QE-MS/MS
Data analysis
UHPLC-QE-MS/MS
Differential secondary metabolites
Correlation
Different microorganisms in root soil and non-root soil
Soil Sampling
Data acquisition
Data storage
Metagenomics
Soil microbial difference
Correlation
Metal element level difference
Soil Sampling
ICP-MS
Data analysis
ICP-MS
Metal element level difference
the authenticity of traditional herbal medicines study

738

739

740

Figure legends

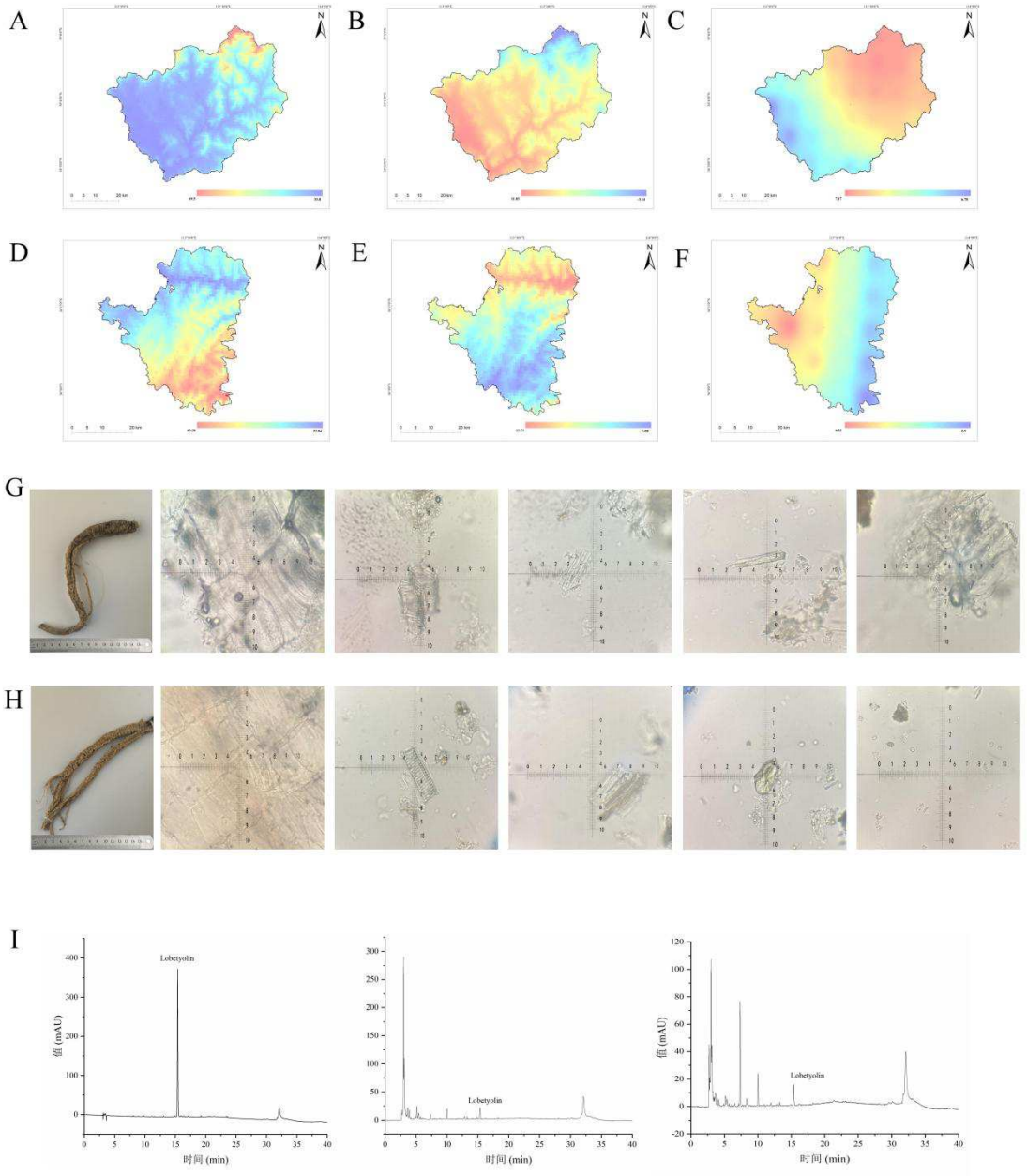


Fig. 1 Precipitation distribution map of Wutai County, Shanxi Province **(A)**; temperature distribution map of Wutai County, Shanxi Province **(B)**; sunshine distribution map of Wutai County, Shanxi Province **(C)**; precipitation distribution map of Pingshun County, Shanxi Province **(D)**; temperature distribution map of Pingshun County, Shanxi Province **(E)**; sunshine duration distribution map of Pingshun County, Shanxi Province **(F)**; characteristics and microscopic identification map of wild TCPs and characteristics **(G)**; microscopic identification map of LCPs **(H)**; Lobetyolin content map of control substance, TCP and LCP **(I)**

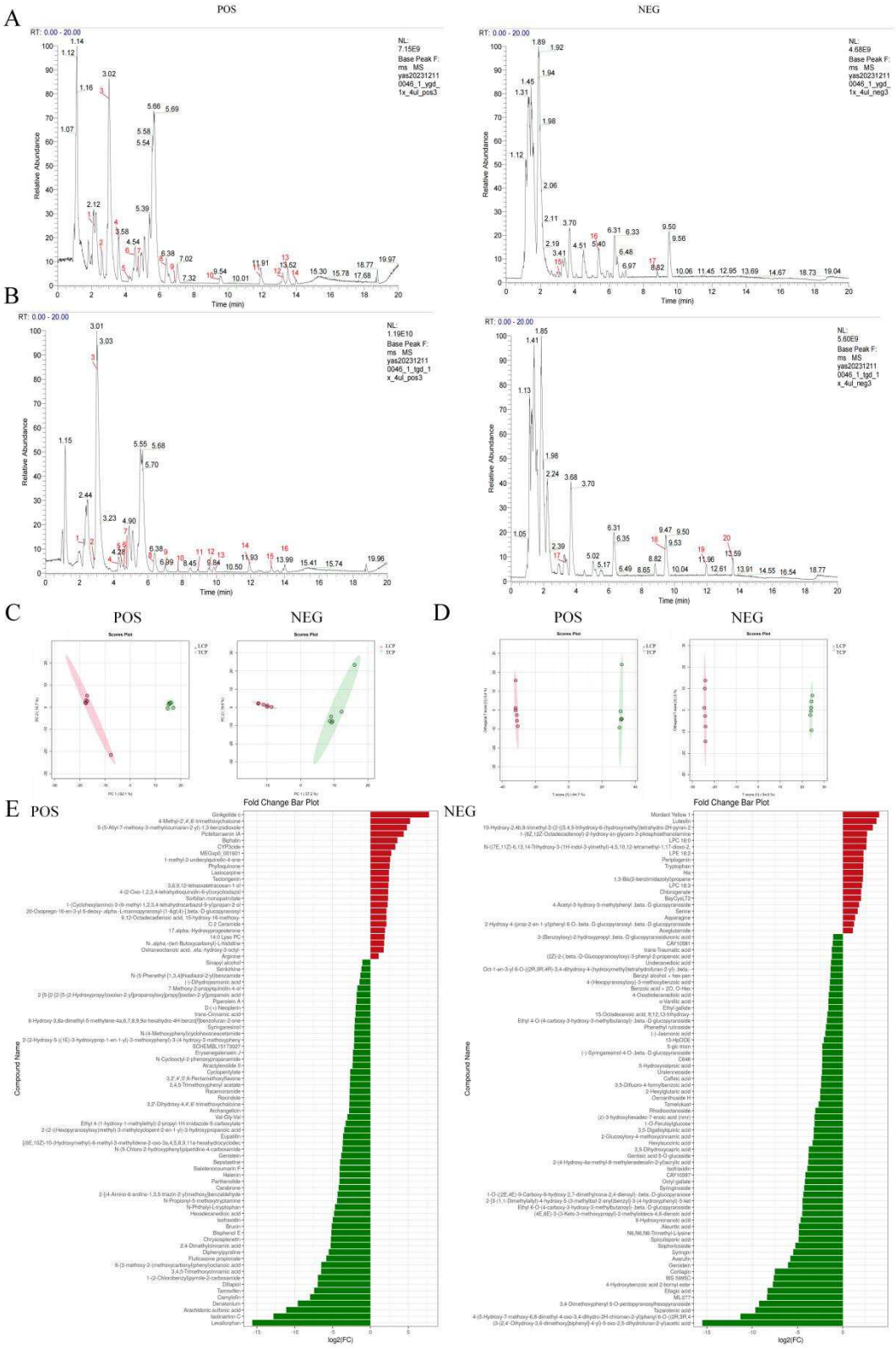


Fig. 2 TCP active ingredient BPC standard peak figure **(A)**; LCP effective component BPC standard peak figure **(B)**; TCP and LCP positive ion mode PCA scoring figure, PCA score plots of TCP and LCP in the negative ion mode of the implant platform**(C)**; TCP and LCP positive ion mode OPLS-DA scoring figure, OPLS-DA score plots of TCP and LCP in the negative ion mode

of the implant platform **(D)**; Differentially abundant metabolites in the positive ion mode of TCP and LCP; F: TCP and LCP anion pattern difference metabolites **(E)**

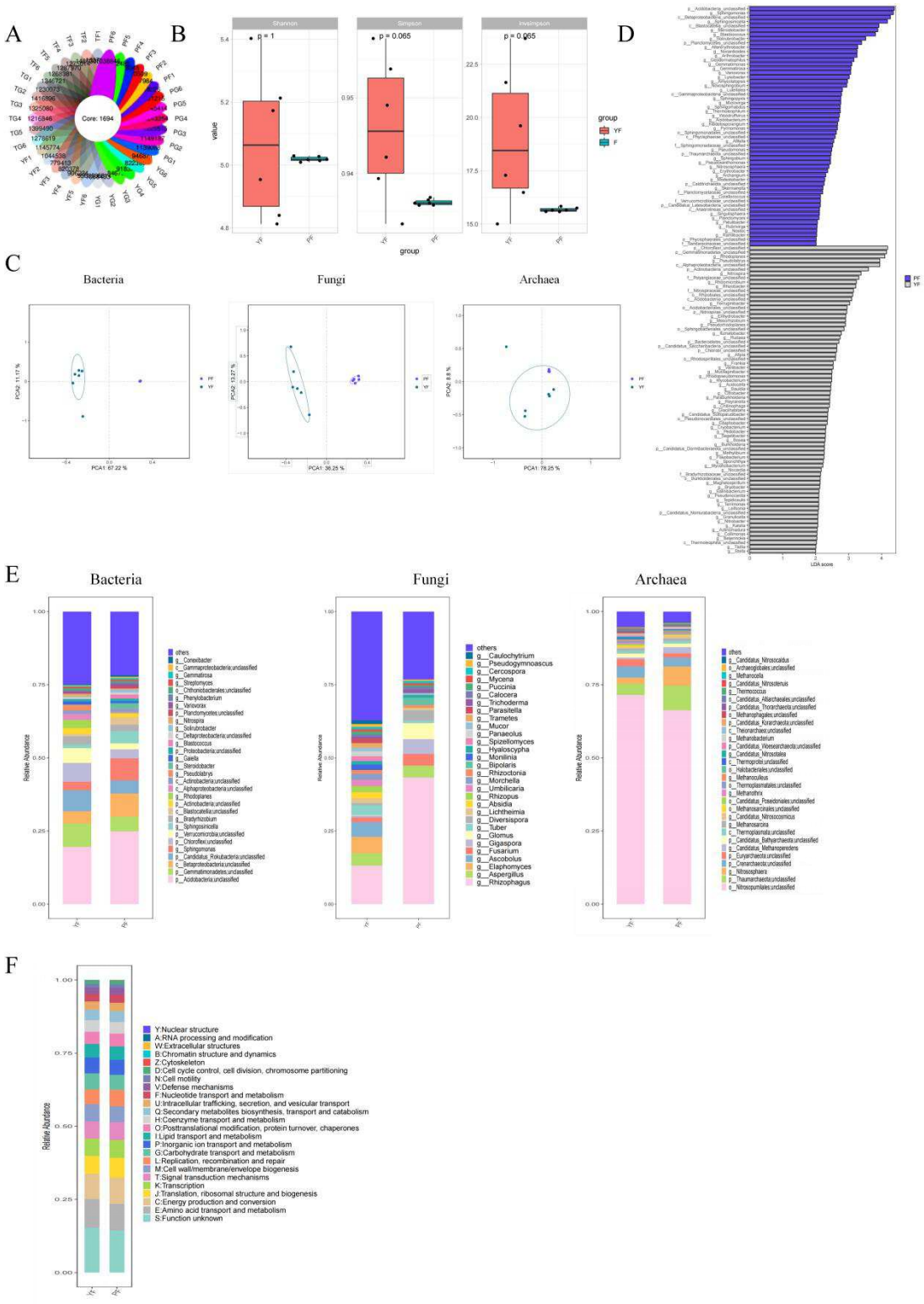
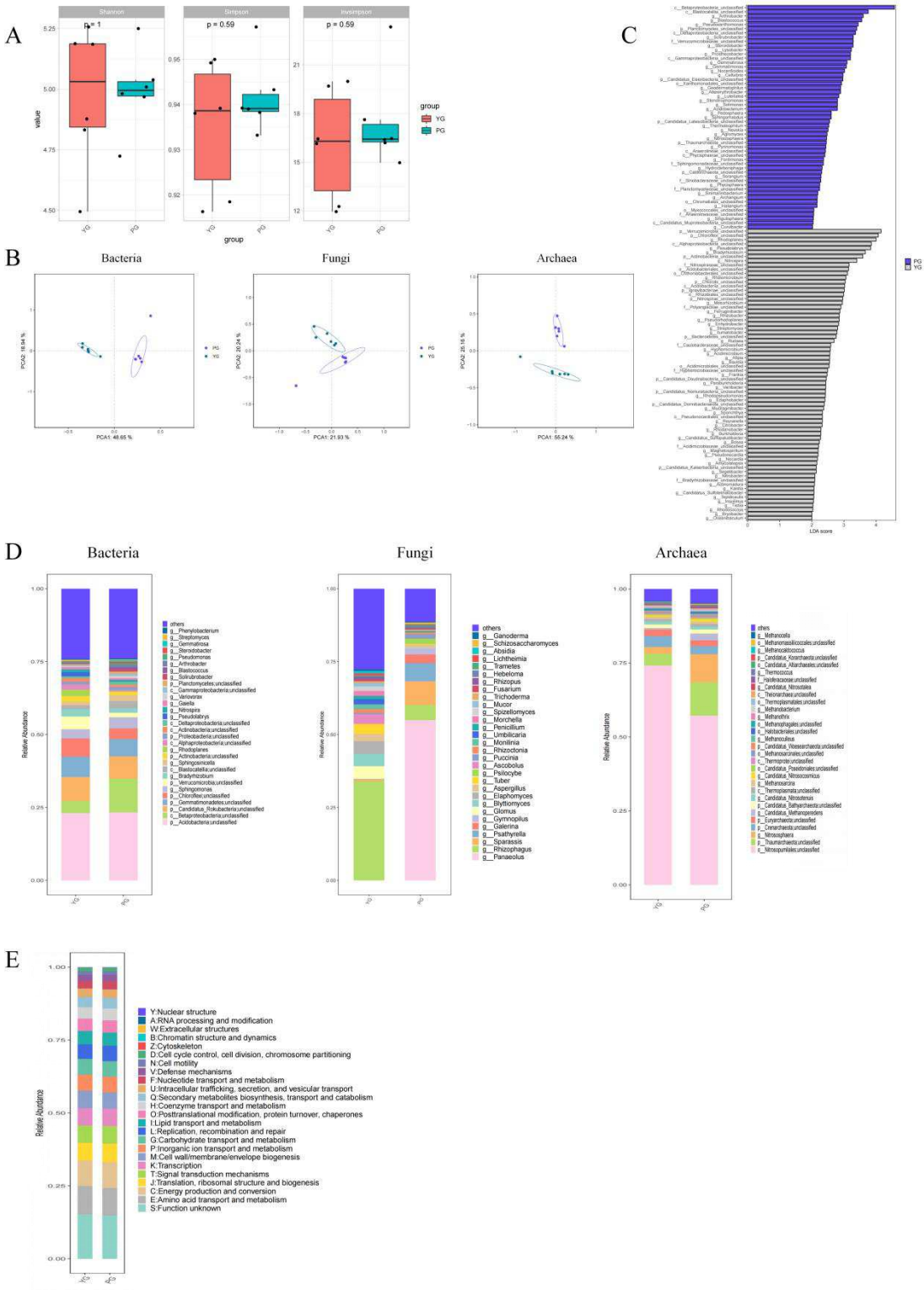


Fig. 3 Soil abundance Wayne figure **(A)**; Nonroot soil bacteria, archaea, and fungi PCA scoring figure **(B)**; α diversity analysis of nonroot soil **(C)**; LEfSe difference analysis of nonroot soil

760 microorganisms **(D)**; Difference map of the microbial abundances of bacteria, archaea and fungi in

761 nonroot soil **(E)**; EggNOG annotation diagram **(F)**



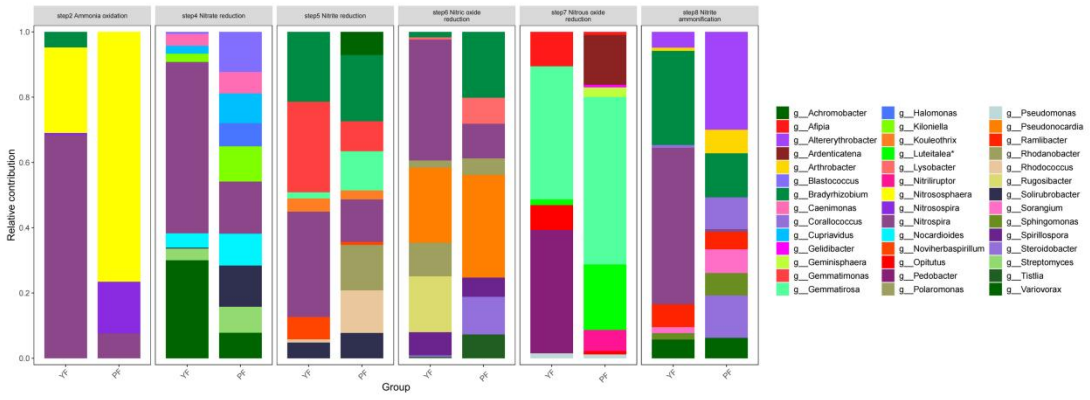
762

763 **Fig. 4** PCA score plot of bacteria, archaea and fungi in root soil **(A)**; Root soil alpha diversity

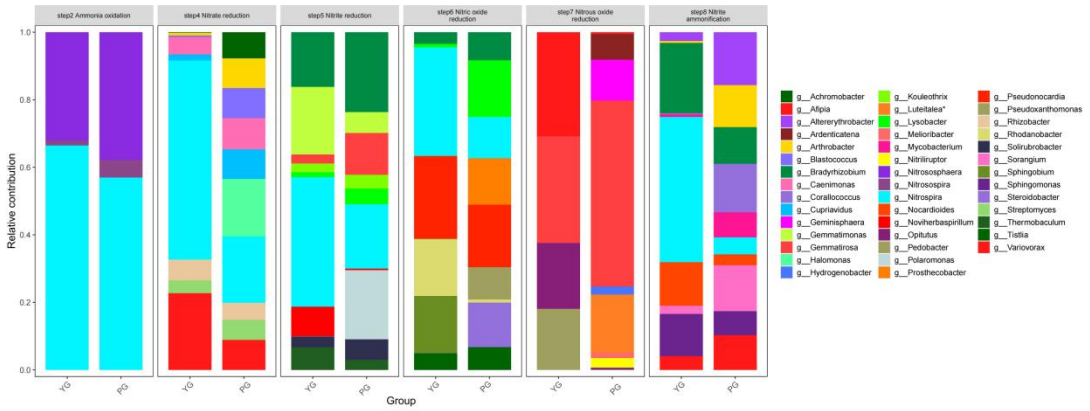
764 analysis diagram **(B)**; Root soil microbial LEfSe difference analysis diagram **(C)**; Difference map

765 of the microbial abundances of bacteria, archaea and fungi in nonroot soil **(D)**; EggNOG
766 annotation diagram **(E)**

A



B



767

768 **Fig. 5** Functional map of metabolic genes in nonroot soil and nitrogen cycling ecosystems **(A)**;
769 Functional map of soil-nitrogen cycling ecosystem metabolic genes in roots
770 **(B)**

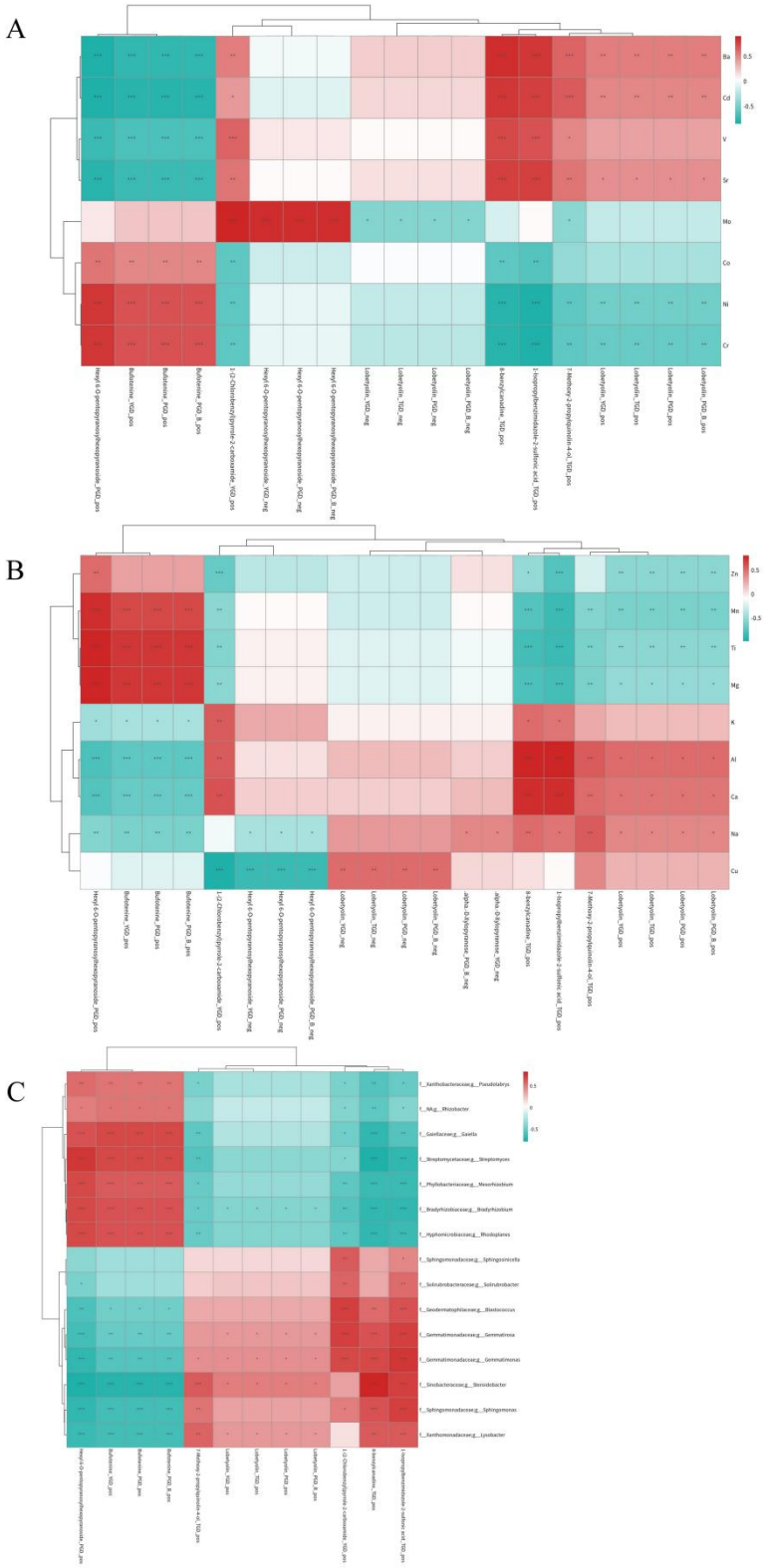


Fig. 6 Correlation analysis of soil microorganisms and pharmacodynamic components in CPs **(A)**; correlation analysis diagram of soil metal elements and pharmacodynamic components in CPs **(B, C)**

Supplementary Files

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

- 2.3SupportingmaterialofthesecondarymetabolitesofTCPandLCPinNEGmode.xlsx
- 2.3SupportingmaterialofthesecondarymetabolitesofTCPandLCPinPOSmode.xlsx
- 2.7SupportingmaterialofthemetalelementofTCPandLCP.xlsx