Integrated Metabolomic and Transcriptomic Analysis of triterpenoid Accumulation in the Roots of Codonopsis pilosula. var. modesta (Nannf.) L.T.Shen at Different Altitudes

Zi-xia Wang  
Lanzhou University

Peng-peng Li  
Lanzhou University

Yan-jun Jia  
Lanzhou University

Long-xia Wen  
Lanzhou University

Zhuo-shi Tang  
Lanzhou University

Yan-ping Wang  
Lanzhou University

Fang Cui  
Lanzhou University

Ke Sun  
Lanzhou Foci Pharmaceutical Co., Ltd. Research Institute

Fang-Di Hu  
Email: hufd@lzu.edu.cn  
Lanzhou University

Research Article

Keywords: triterpenoid, metabolome, transcriptome, biosynthesis, altitude, Codonopsis pilosula. var. modesta (Nannf.) L.T.Shen

Posted Date: September 7th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3219638/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Abstract

Background: *Codonopsis Radix* is a benefit Traditional Chinese Medicine and triterpenoid are the major bioactive constituents of *Codonopsis* Radix. *Codonopsis pilosula* var. *modesta* (Nannf.) L.T. Shen (CPM) is a precious variety and the legal source of *Codonopsis Radix*, and it roots are known as Wen Danshen, which is distribute in high mountains area and the altitudes are ranging from 1300 to 4300 m. Environment plays an important role in the synthesis and metabolism of active ingredients in medicinal plants, but there is no report elaborate the effect of altitude on terpenoid metabolites accumulation in CPM. In this study, we integrated metabolomic and transcriptomic to explain the effects of altitude on terpenoid biosynthetic pathways and secondary metabolite accumulation in CPM, fresh root samples from CPM grown at low altitude (1480 m) and high altitude (2300m) at the same harvest stage were selected for analysis.

Results: Untargeted metabolic results were shown that there were significant differences in the total secondary metabolites between high altitude and low altitude CPM group. Based on our laboratory previously established UPLC-Q-TOF-MS method, 10 triterpenoids in the above two altitude CPM fresh roots were quantitatively analyzed. According to their chemical structure and isomerism, they can be divided into 6 categories including Gansuidine-type tetracyclic triterpenes (codopitirol A), Cycloaneurane tetracyclic triterpenes (24-methylenecycloartanol), Xylorane-type pentacyclic triterpenes (kakoonol, friedel-1-en-3-one, friedelin), Dandelion pentacyclic triterpenoids (codopimodol A, taraxerol), Oleanocarpine pentacyclic triterpenoids (bryonolol, glut 5-en-3β-ol), Ursulane-type pentacyclic triterpenes (α-amyrin), our result showed that there are significant differences in the content and the types of terpenoids between different altitude CPM and the content were higher in the high altitude samples of CPM. The results of transcriptome study showed that CPM could significantly up-regulate the gene expression levels of seven key enzymes in the biosynthetic pathway of triterpenoid precursor substances. These enzymes include isoprenyl diphosphate isomerase (IDI), geraniol pyrophosphate synthase (GPPS), farnesyl pyrophosphate synthase (FPPS), farnesyl pyroacyltransferase (SS), squalene synthase (SE), beta-amyrin synthase (β-AS) and lupiol synthase (LS). The quantitative polymerase chain reaction was used to further verify the accuracy of the transcriptome data.

Conclusions: This study shows that there were significant differences between high altitude and low altitude groups of CPM, which was manifested in the following aspects including overall secondary metabolites, content and types of triterpenoids, and gene expression levels. The CPM at high altitude were more likely to accumulate triterpenes than those at low altitude, which was related to the up-regulation of the gene expression levels of seven key enzymes in the triterpenoid precursor biosynthetic pathway, thereby increasing the accumulation of triterpenoids. These results expand our understanding of how altitude affects plant metabolite biosynthesis.

1. Introduction
Codonopsis Radix as a benefit Traditional Chinese Medicine (TCM) has the functions of invigorating spleen and lung, nourishing blood and boosting fluid. Since the records of materia medica in the Qing Dynasty, it has been used in medicine and edible for nearly 300 years. Codonopsis pilosula var. modesta (Nannf.) L.T. Shen (CPM) is a precious variety of Codonopsis pilosula and the legal source of Codonopsis Radix in the Pharmacopoeia of the People's Republic of China (R. Bai et al., 2022), and it roots are known as Wen Dangshen. In addition, modern studies have confirmed the prominent pharmacodynamic and nutritional value of CPM in Codonopsis Radix from the aspects of pharmacodynamic material basis and efficacy. Terpenoids are a class of natural hydrocarbons which structural unit is isoprene, and they are also the largest secondary metabolites in plants. The effective constituents of terpenoids in CPM mainly included triterpenoids such as taraxerol and friedelin (Wakana, Kawahara, & Goda, 2011). Studies have shown that triterpenoids in CPM have significant anti-inflammatory activities (Fu et al., 2016).

Biosynthesis of terpenoids in plants includes two pathways: One is the mevalonic acid pathway (MVA) located in the cytoplasm, the other is the methylerythritol phosphate pathway (MEP) located in plastids, both of which formed isopenteny diphosphate (IPP) and its isomer dimethyl allyl diphosphate (DMAPP) followed by further synthesis of terpenoids (Bergman, Davis, & Phillips, 2019). Firstly, IPP combined with DMAPP generate geranyl diphosphate (GPP), a precursor of monopterpenoid. Then GPP combines with another one molecule IPP to form farnesyl diphosphate (FPP), which is a precursor of triterpenes (Yi et al., 2021; K. Zhang, Wang, Gao, & Ma, 2022).

Geranyl diphosphate synthase (GPPS) and farnesyl diphosphate synthase (FPPS) are important synthases for the formation of sesquiterpene and triterpene precursor. Biosynthesis induced by jasmonic acid (JA) and methyl dihydrojasmonate (MDJ) has been studied. In Panax ginseng C. A. Mey (J. Li et al., 2016), Panax notoginseng F.H. Chen (J. Li et al., 2017) found that the up-regulated expression of GPPS and FPPS genes was positively correlated with the accumulation of terpenoids. Meanwhile, the method of gene cloning has been found in Artemisia annua L. (Mercke, Crock, Croteau, & Brodelius, 1999) and Poria cocos (Wang, Li, & Liu, 2014). The expression abundance of FPPS genes was positively correlated with the contents of triterpenoids. Through gene cloning, it has been confirmed that IDI plays a very important regulatory role in the biosynthesis pathway of triterpenoids in Ganoderma Lucidum Karst (Wu et al., 2013). Cloning the cDNA of isopentenyl pyrophosphate isomerase from Artemisia annua (Aa) and overexpressing it in Artemisia annua can lead to metabolic changes in terpene biosynthesis, and significantly increase the contents of arteannuin B and artemisinin (Ma et al., 2017). AATC and mevalonate kinase (MVK) are key enzymes in the biosynthetic pathway of terpenoids MVA. AATC has been shown to mediate the conversion of acetyl mevalonate CoA to acetyl CoA in Ginkgo biloba. MVK mediates the conversion of mevalerate to mevalerate phosphate, confirming that AATC and MVK are functional genes in the MVA biosynthesis pathway (Chen et al., 2017).

The chemical component in medicinal plants are the material basis for the effect of TCM, most of them are plant secondary metabolites, which are greatly affected by environmental factors (Y. Li, Kong, Fu, Sussman, & Wu, 2020). The accumulation of the chemical component and gene expression in plant are all have spatiotemporal specific property, the differential analysis of gene expression levels in different
growth environments is an effective means to screen candidate genes involved in the synthesis of active ingredients in medicinal plant (Dong et al., 2022; Xia et al., 2022), At the meanwhile, environmental factors are closely related to the growth and quality of medicinal plants (Y. G. Chen et al., 2021). CPM is cultivated in high mountains with altitudes ranging from 1300 to 4300 m in Wen Count, which is the main producing area of CPM located in the southernmost area of China’s Gansu Province (Yuting, Juan, & Wenlong, 2021). The vertically structure of geography forms different biological communities and various climate characteristics (Y. Bai, Tian, Fang, Chen, & Liu, 2020; H. Zhang et al., 2023). The environment is closely related to the growth of medicinal plants and the quality of medicinal materials, and altitude is the most comprehensive and decisive among all the environmental factors affecting the distribution of plants (Earnshaw et al., 1987; Ye, Zhu, Gao, Jiang, & Wu, 2020). Considering the high quality of Wen Dangshen and the specificity of planting distribution, it is meaningful to revealed the effects of environment on biosynthesis and accumulation of sesquiterpenoids and triterpenoids in CPM, while there is still no article reported it.

In this study, the liquid chromatography-mass spectrometry (LC-MS) technology was used for non-targeted metabolomics to study the accumulation patterns of chemical components in the high and low altitude populations, and to observe the effects of different altitudes on the secondary metabolites in the CPM. And the accumulation of the 10 triterpenoids in CPM at high and low altitude were studied by UPLC-Q-TOF-MS/MS technique which was established previously in our laboratory (W. Li et al., 2023a). The key enzyme genes expression profiles involved in triterpenoids biosynthesis pathway in CPM were obtained through high-throughput RNA-sequencing (RNA-seq) data analysis, the quantitative polymerase chain reaction was used to further verify the accuracy of the transcriptome data. The aim of this study is to systematically study the influence of altitude on the biosynthesis pathway of triterpenoids in CPM.

2. Materials and methods

2.1 Statement

The experimental materials were collected in China Wen County of Gansu Province, we declare that there is no endangered or protected species involved in the experimental collection site and surrounding areas. This study was conducted at the State Key Laboratory of Functional Organic Molecular Chemistry, School of Pharmacy in Lanzhou University, and Long medicine Collaborative Innovation Center of Lanzhou University.

2.2 Plant materials

In October 2021, the Low altitude and High altitude samples were selected from Huangjia Village, Qiaotou Town (E104°47′58″, N33°5′38″, 1480m), and Songping Village, Zhongzhai Town (E104°28′19″, N33°17′23″, 2300m), both in China Wen County of Gansu Province. The root samples of 2-year old CPM in the harvest period were selected, and all samples followed three biological replicates, which were located in low altitude samples (L1-L3) and high altitude samples (H1-H3). All samples were certified by Professor
All tissue samples were stored at −80° C immediately after collection until RNA extraction.

### 2.3 Analysis of sesquiterpenoids and triterpenoids

Based on our laboratory previously established UPLC-Q-TOF-MS method (W. Li et al., 2023), 10 triterpenoids in the above two altitude CPM fresh roots were quantitatively analyzed, three independent analyses were performed in biological replicates. The 10 triterpenoids, including tirucallane (codopitirol A), cycloartane (24-methylene cycloartanol), friedelane (kokoanol, friedel-1-en-3-one, friedelin), taraxastane pentacyclic triterpenoids (codopimodol A, taraxerol), oleanane (bryonolol, glut-5-en-3β-ol), ursane (α-amyrin) were performed quantitative analysis. The analytical methods are briefly described as follows: First, fresh root samples of CPM were taken, and each sample was ground to a fine powder under liquid nitrogen, 2 g of powder was precisely weighed, and extracted with ethyl acetate for 6 h in a Soxhleter extractor. Then, the resulting residue was concentrated under reduced pressure and dissolved in 1 ml of methanol containing 1% chloroform. The solution was centrifuged at 12 000 rpm for 10 min and the supernatant was retained. The supernatant was filtered through a 0.22-µm membrane filter before injection into UHPLC, and all solutions were stored at 4 °C until further use.

The UHPLC-Q-TOF-MS system consisted of a UHPLC system (Agilent, Santa Clara, CA, USA) and a Q-TOF-MS(Agilent) equipped with an ESI source (Agilent) and APCI. An ACQUITY UPLC HSS T3 C18 (100 mm×2.1 mm, 1.8 µm) column was used. The mobile phase consisted of 0.1% formic acid (A) and 0.1% formic acid in methanol. Gradient elution: 0–3 min, 30%-80% B; 3–10 min, 80%-92% B; 10–20 min, 92% B; 20–21 min, 92%-95% B; 21–30 min, 95% B; 30–31 min, 95%-100% B; 31–32 min, 100% B. The flow rate was 0.4 mL/min. Column temperature was 40°C and injection volume was 2 µL. Mass spectrometry conditions was atmospheric pressure chemical ionization source, positive ion scanning mode, scanning range m/z 100–1000, drying temperature 350°C, evaporation temperature 450°C, drying gas flow rate 5 L/min; The atomizer pressure was 60 psi. The capillary voltage was 3500 V and Corona + was 4 µA. And Agilent Mass Hunter software was used to process the collected mass spectrum data, which could display the collected high-quality mass spectrum information data and export the original data. Origin software was used for mapping and analysis.

### 2.4 Transcriptome sequencing and analysis

#### 2.4.1 Library Construction and Sequencing (mRNA-Seq)

Total RNA was obtained from each sample using the RNAprep Pure Polysaccharide Polyphenol Plant Total RNA Isolation Kit (TIANGEN, Beijing, China) according to the manufacturer’s instructions. The quality and quantity of extracted RNA were assessed using Agilent Bioanalyzer 2000 (Agilent Technologies, Palo Alto, USA). Sequencing was performed on the Illumina Nova Seq 6000 sequencing platform and 150 bp paired-end reads were generated to construct RNA-Seq sequence libraries (Novogene, Beijing, China). The datasets generated during the current study are available in the SRA repository, the BioProject ID is PRJNA979796.
2.4.2 Identification and analysis of differentially expressed genes

The transcriptome sequences obtained after Trinity splicing (Sewe, Silva, Sicat, Seal, & Visendi, 2022). After filtered the low-quality data were removed, and then the quantity was normalized using the DESeq2 package in R(3.6.2). The RSEM (v1.2.15) software was used to calculate the gene expression levels of samples at different altitudes by FPKM values (B. Li & Dewey, 2011). The differentially expressed genes (DEGs) were screened. Use the q value to adjust the P value (q< 0.005 & | log2 (a fold change) | > 1) threshold is set to significantly differentially expressed. The expression levels of genes were visualized using the heat map package in R software (3.6.2). GOseq(1.10.0) and KOBAS(v2.0.12) software were used to perform GO function (Young, Wakefield, Smyth, & Oshlack, 2010) and KEGG pathway enrichment analysis (Bu et al., 2021) of DEGs, respectively.

2.5 Analysis of untargeted metabolomics

2.5.1 Extraction of metabolites

The fresh root samples of CPM ground in 100 mg liquid nitrogen was placed in an EP tube, and 500µL of 80% methanol aqueous solution was added. After vortexing, the cells were placed in an ice bath for 5min and centrifuged at 15000 g and 4°C for 20 min. A certain amount of supernatant was diluted to 53% methanol with mass spectrometric water. After centrifugation at 15000g for 20 min at 4°C, the supernatant was collected and samples were injected for LC-MS analysis(W. Li et al., 2023). Equal volumes of samples were mixed from each experimental sample as QC samples. The chromatographic column was Hypesil Gold column(C18). The mobile phase in the positive ion mode is composed of 0.1% formic acid (A) and methanol (B). The mobile phase in the negative ion mode consisted of 5 mM ammonium acetate (A) and methanol (B). The column temperature was 40°C. The flow rate was 0.2 mL/min, the injection volume was 2 µL, and the gradient elution was performed. The mass spectrum conditions were ESI source, positive and negative ion scanning modes, and the scanning range was selected from m/z 100 to 1500. ESI source Settings are as follows: Spray Voltage :3.5kV; Sheath gas flow rate :35psi; Aux Gas flow rate :10L/min; Capillary Temp :320°C; S-lens RF level: 60; Aux gas heater temp:350°C; MS/MS secondary scans were data-dependent scans.

The data were imported into CD 3.1 database search software for processing, and the retention time, mass-to-charge ratio and other parameters of each metabolite were simply screened. Then the retention time deviation was set to 0.2 min and the mass deviation was set to 5 ppm to align the peak of different samples, so that the identification was more accurate. Subsequently, the mass deviation of 5 ppm, signal intensity deviation of 30%, signal-to-noise ratio of 3, minimum signal intensity, plus ion and other information were set for peak extraction, while the peak area was quantified, and then the target ion was integrated. Then the molecular formula was predicted by molecular ion peaks and fragment ions, and compared with mzCloud(https://www.mzcloud.org/), mzVault and Masslist databases. The background ions were removed with blank samples, and the original quantitative results were standardized. Finally,
the identification and relative quantification results of metabolites were obtained. The data processing was based on Linux operating system (CentOS version 6.6) and R and Python software.

2.5.2 Functional annotation of metabolites

Using KEGG database (https://www.genome.jp/kegg/pathway.html), HMDB (https://hmdb.ca/metabolites) database and LIPIDMaps database (http://www.lipidmaps.org/) to identify the metabolites. The metabolomics data processing software metaX(Wen, Mei, Zeng, & Liu, 2017) was used to transform the data and perform principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) to obtain the VIP value of each metabolite. In the univariate analysis, the statistical significance (P value) of each metabolite between the two groups was calculated based on t test, and the fold change (FC value) of metabolites between the two groups was calculated. The default criterion for differential metabolites screening was VIP > 1, P-value < 0.05 and FC ≥ 2 or FC ≤ 0.5. KEGG database was used to study the function of metabolites and metabolic pathways when x/n > y/n, this metabolic pathway was considered to be enriched. A metabolic pathway was considered significantly enriched when its p-value was less than 0.05.

2.6 Integrative analysis of metabolomic and transcriptomic data

The correlation between the expression of differential genes and differential metabolites was analyzed by correlation heat map analysis and correlation network map analysis. KEGG enrichment analysis, iPath pathway visualization analysis and KEGG pathview visualization analysis were used to analyze the pathway enrichment of differential genes and differential metabolites. Based on the Pearson statistical method, the correlation coefficient $r^2$ and P value of differential genes and differential metabolites were calculated. The metabolites of Top5 and the genes of Top10 were selected, and the metabolism-transcriptional correlation network was plotted with the help of the mixOmics package of R language. KEGG Pathways were used as entries to obtain pathways enriched in metabolism and transcription. The ggplot2 package of R language was used to draw the metabolis-transcriptional KEGG enrichment bubble map for the pathways enriched by metabolism and transcription. The pathway entries of co-enrichment results of metabolomics and transcriptomics were selected, and the patiview enrichment map of metabolism-transcriptional KEGG was drawn for the pathways co-enriched by metabolism and transcription with the help of the pathview package of R language.

2.7 Real-Time Quantitative PCR Validation of gene expression

Real-time quantitative PCR (qPCR) was used to verify the differentially expressed genes in roots samples of CPM at high and low altitudes, and the results were compared with RNA-seq data. Complementary DNA (cDNA) was extracted from total RNA extracted with the Fast Quant RT SuperMix Kit (item number DP424) produced by Beijing Tiangen Biotechnology Co., LTD. Revert Aid First produced by Thermo Scientific Biotechnology, USA The results were obtained by reverse transcription with Strand cDNA
Synthesis Kit (K1622). The Ct values of each template were detected by Step One Plus Real-Time fluorescence quantitative PCR instrument, and calculated by the $2^{-\Delta\Delta Ct}$ method. The Ct value of the target gene was subtracted from the Ct value of the internal reference gene to obtain $\Delta Ct$. $\Delta\Delta Ct$ was obtained by subtracting the mean $\Delta Ct$ value of the control group from the $\Delta Ct$ value of the experimental group. $2^{-\Delta\Delta Ct}$ was used to calculate the relative expression changes of the target genes in the control and experimental groups. We selected GAPDH as an internal control, GenBank accession number CB970967.

3 Results

3.1 Untargeted metabolomic analysis

A total of 845 metabolites were identified in the samples, 303 of which were identified in the negative ion model, and 545 in the positive ion model. All metabolites showed differential accumulation in high and low altitude samples (Fig. 1A). The results of principal component analysis showed that PC1 and PC2 explained 27.47% and 22.76% of the variance, respectively. The samples at high and low altitudes had a certain separation trend, but did not completely distinguish (Fig. 1B). Through PLS-DA analysis, the effective separation of high and low altitude Costanopsis pilocarpa was achieved according to the metabolites (Fig. 1C), and the R2 value of cross test results was 1.00 (Fig. 1D), which further confirmed the accuracy of our analysis results.

A total of 404 compounds were annotated into 6 major categories and 32 sub-categories of metabolic pathways, of which 12 metabolites were annotated into the "metabolism of terpenoids and polyketones" pathway (Table S1). A total of 5 triterpenoids were obtained (Table S2), including dammarane tetracyclic triterpenoids (arenol B), cycloartenane tetracyclic triterpenoids (propanaxediol, quercitrinic acid), oleanane pentacyclic triterpenoids (oleanolic acid) and ursolic acid pentacyclic triterpenoids (ursolic acid). In negative ion mode, functional annotation through KEGG database revealed that metabolites were enriched in the "terpenoid and sterol biosynthesis" pathway (Fig. 1E), and metabolites were also enriched in the "sesquiterpene and triterpenoid biosynthesis" pathway in positive ion mode (Fig. 1F). These results suggested that altitude had an important effect on the synthesis and accumulation of secondary metabolites, especially terpenoids.

3.2 Differences of triterpenoids in high and low altitude

The results of the determination of 10 triterpenoids in CPM at high and low altitudes (Table S3) show that they can be divided into 6 groups according to their chemical structure and isomerism, they can be divided into 6 categories: Gansuidine-type tetracyclic triterpenes (codopitirol A), Cycloaneurane tetracyclic triterpenes (24-methylenecycloartanol), Xylorane-type pentacyclic triterpenes (kokoonol, friedel-1-en-3-one, friedelin), Dandelion pentacyclic triterpenoids (codopimodol A, taraxerol), Oleanocarpine pentacyclic triterpenoids (bryonolol, glut 5-en-3β-ol), Ursulane-type pentacyclic triterpenes (α-amyrin) (Fig. 2A and Fig. 2C). Further analysis showed that the contents of sesquiterpenes and triterpenes in high-altitude samples were significantly higher than those in low-altitude samples, and the contents of Xylorane-type
pentacyclic triterpenes, Dandelion pentacyclic triterpenoids, Oleanocarpine pentacyclic triterpenoids, and Ursulane-type pentacyclic triterpenes in high-altitude samples were significantly higher than those in low-altitude samples ($P < 0.05$).

Principal component analysis of high altitude (H1-H3) and low altitude (L1-L3) samples was performed with the contents of sesquiterpenoids and triterpenoids as variables. PC1 and PC2 explained 84.96% of the variation (Fig. 2B). Principal component analysis showed that there were significant differences in sesquiterpenoids and triterpenoids in high altitude and low altitude CPM, which could be separated by PC1. These results suggested that altitude had an important effect on the biosynthesis of triterpenoids in CPM, and the content of triterpenoids in CPM increased significantly with the altitude.

### 3.3 Transcriptome analysis of C. pilosula var. modesta in high and low altitude

Using RNA-seq technology, we performed transcriptome analysis of CPM samples from high and low altitude populations to screen potential genes affecting the composition of sesquiterpenoids and triterpenoids. After raw data filtering, sequencing error rate checking, and GC content distribution checking, clean reads for subsequent analysis were obtained (Table S4). About 75.81 to 76.77% of the clean reads from each library were mapped to the reference sequence. A total of 58,620 unigenes were generated by this sequencing. The average length of these unigenes was 1,203bp. Among them, 68.5% (40,160) and 38.9% (22,777) unigenes were longer than 500bp and 1000bp, respectively (Table S5). The intraclass correlation ranged from 0.826 to 0.916 (Fig. 3). The expression level of each gene was analyzed by calculating the FPKM value. The gene expression levels of high-altitude samples were significantly higher than those of low-altitude samples (Fig. 3). A total of 17,351 DEGs were identified in the samples of CPM at high and low altitudes (Fig. 3), among which 11,993 DEGs were up-regulated and 5,358 DEGs were down-regulated in the samples of CPM at high altitude compared with those at low altitude. The heatmap cluster analysis of DEGs could significantly distinguish the high altitude and low altitude samples (Fig. 3). The expression patterns of the six differentially expressed genes were confirmed by qPCR to verify the accuracy of RNA-seq data. The results show that the expression patterns of the qRT-PCR genes are highly consistent with the RNA-seq data (Figure S2), which means that our RNA-seq data are reproducible and reliable.

### 3.4 Expression analysis of genes involved in triterpenoid biosynthesis

Based on KEGG enrichment analysis, the terpenoids biosynthesis of CPM mainly includes two biological pathways: “terpenoid skeleton biosynthesis” (map00900) and “sesquiterpene and triterpene biosynthesis” (map00909). The differentially expressed genes (DEGs) is vital to the biosynthesis of triterpenoids. With further screening of FPKM value >0.3, there are 93 Unigenes encode 35 key enzymes (Table 1). 25
Unigenes of these are deeply identified as DEGs. These genes were considered as possible genes involved in the synthesis of triterpenoids in CPM (Table S6).
## Table 1
Number of unigenes encoding key enzymes involved in triterpenes biosynthesis in *C. pilosula* var. *modesta* (FPKM > 0.3)

<table>
<thead>
<tr>
<th>Enzyme name</th>
<th>EC number</th>
<th>Unigene number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR)</td>
<td>1.1.1.267</td>
<td>1</td>
</tr>
<tr>
<td>hydroxymethylglutaryl-CoA reductase (HMGR)</td>
<td>1.1.1.34</td>
<td>3</td>
</tr>
<tr>
<td>NAD+-dependent farnesol dehydrogenase</td>
<td>1.1.1.354</td>
<td>2</td>
</tr>
<tr>
<td>squalene monooxygenase (SE)</td>
<td>1.14.14.17</td>
<td>4</td>
</tr>
<tr>
<td>(E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (ispG)</td>
<td>1.17.7.1, 1.17.7.3</td>
<td>2</td>
</tr>
<tr>
<td>4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase (ispH)</td>
<td>1.17.7.4</td>
<td>2</td>
</tr>
<tr>
<td>geranylgeranyl diphosphate/geranylgeranyl-bacteriochlorophyllide a reductase</td>
<td>1.3.1.83,1.3.1.111</td>
<td>3</td>
</tr>
<tr>
<td>prenylcysteine oxidase / farnesylcysteine lyase (PCYOX)</td>
<td>1.8.3.5, 1.8.3.6</td>
<td>2</td>
</tr>
<tr>
<td>protein-S-isoprenylcysteine O-methyltransferase (CMT)</td>
<td>2.1.1.100</td>
<td>3</td>
</tr>
<tr>
<td>1-deoxy-D-xylulose-5-phosphate synthase (DXS)</td>
<td>2.2.1.7</td>
<td>5</td>
</tr>
<tr>
<td>acetyl-CoA C-acetyltransferase (AATC)</td>
<td>2.3.1.9</td>
<td>4</td>
</tr>
<tr>
<td>hydroxymethylglutaryl-CoA synthase (HMGS)</td>
<td>2.3.3.10</td>
<td>2</td>
</tr>
<tr>
<td>geranyl diphosphate synthase (GPS)</td>
<td>2.5.1.1</td>
<td>2</td>
</tr>
<tr>
<td>farnesyl diphosphate synthase (FPPS)</td>
<td>2.5.1.1,2.5.1.10</td>
<td>2</td>
</tr>
<tr>
<td>geranylgeranyl diphosphate synthase, type II (GGPS)</td>
<td>2.5.1.1,2.5.1.10,2.5.1.29</td>
<td>7</td>
</tr>
<tr>
<td>farnesyl-diphosphate farnesyltransferase (SS)</td>
<td>2.5.1.21</td>
<td>3</td>
</tr>
<tr>
<td>protein farnesyltransferase subunit beta (FNTB)</td>
<td>2.5.1.58</td>
<td>3</td>
</tr>
<tr>
<td>hexaprenyl-diphosphate synthase</td>
<td>2.5.1.82, 2.5.1.83</td>
<td>1</td>
</tr>
<tr>
<td>protein farnesyltransferase/geranylgeranyltransferase type-1 subunit alpha</td>
<td>2.5.1.58,2.5.1.59</td>
<td>1</td>
</tr>
<tr>
<td>all-trans-nonaprenyl-diphosphate synthase</td>
<td>2.5.1.84,2.5.1.85</td>
<td>2</td>
</tr>
<tr>
<td>ditrans,polycis-polyprenyl diphosphate synthase (DHDSS)</td>
<td>2.5.1.87</td>
<td>8</td>
</tr>
<tr>
<td>4-diphosphocytidyl-2-C-methyl-D-erythritol kinase</td>
<td>2.7.1.148</td>
<td>1</td>
</tr>
<tr>
<td>farnesol kinase</td>
<td>2.7.1.216</td>
<td>1</td>
</tr>
<tr>
<td>mevalonate kinase (MVK)</td>
<td>2.7.1.36</td>
<td>2</td>
</tr>
<tr>
<td>phosphomevalonate kinase (PMVK)</td>
<td>2.7.4.2</td>
<td>1</td>
</tr>
<tr>
<td>Enzyme name</td>
<td>EC number</td>
<td>Unigene number</td>
</tr>
<tr>
<td>----------------------------------------------------</td>
<td>-------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>2-C-methyl-D-erythritol 4-phosphate cytidyltransferase(ispE)</td>
<td>2.7.7.60</td>
<td>1</td>
</tr>
<tr>
<td>prenylcysteine alpha-carboxyl methylesterase(PCME)</td>
<td>3.1.1.</td>
<td>1</td>
</tr>
<tr>
<td>prenyl protein peptidase</td>
<td>3.4.22.</td>
<td>2</td>
</tr>
<tr>
<td>STE24 endopeptidase(STE24)</td>
<td>3.4.24.84</td>
<td>5</td>
</tr>
<tr>
<td>diphosphomevalonate decarboxylase(MVD)</td>
<td>4.1.1.33</td>
<td>3</td>
</tr>
<tr>
<td>(−)-germacrene D synthase</td>
<td>4.2.3.22,4.2.3.75</td>
<td>3</td>
</tr>
<tr>
<td>isoprene synthase</td>
<td>4.2.3.27</td>
<td>1</td>
</tr>
<tr>
<td>isopentenyl-diphosphate Delta-isomerase(IDI)</td>
<td>5.3.3.2</td>
<td>3</td>
</tr>
<tr>
<td>beta-amyrin synthase</td>
<td>5.4.99.39</td>
<td>5</td>
</tr>
<tr>
<td>lupeol synthase (LS)</td>
<td>5.4.99.41</td>
<td>2</td>
</tr>
</tbody>
</table>

Based on the biosynthetic pathways of triterpenoids in CPM at different altitudes and the gene expression levels of key enzymes, it was found that the biological pathways for the synthesis of precursors of triterpenoids include both MVA and MEP pathways (Table S6). Through the expression levels of key enzyme genes in CPM at high and low altitudes, it was found that CPM samples at different altitudes could be distinguished according to the expression levels of key enzyme genes in the MVA and MEP pathways, but the effects of altitude on the expression levels of key enzyme genes were inconsistent, and the effects of altitude on the MVA and MEP biological pathways could not be determined.

Both MVA and MEP biosynthetic pathways are generate IPP and DMAPP, both of which are active and can further bind to larger molecules. First, IPP and DMAPP can bind to generate GPP, which can then binds to another IPP to form FPP, which is the precursor of triterpenes. IDI, GGPS and FPPS are the key enzymes for the synthesis of triterpenoid precursor FPP(Yi et al., 2021; Zhang, Wang, Gao, & Ma, 2022). By comparing the gene expression levels in CPM at high and low altitudes, differentially expressed genes (DEGs) were found which was associated with IDI, GGPS and FPPS enzymes, and these DEGs could be used to distinguish CPM samples at different altitudes (Fig. 4). At the same time, altitude could significantly increase the expression levels of these genes which was associated with IDI, GGPS and FPPS.

Beta-amyrin synthase (β-AS) is an important modification enzyme for the formation of oleanolic acid from 2,3-oxidized squalene(Zhao et al., 2015). β-AS also a precursor for the synthesis of various triterpenoids, and it plays a pivotal role in the synthesis of triterpenoids(Confalonieri et al., 2009). Three β-AS coding genes were DEGs between the high and low altitude populations, and all of them were highly expressed in the high altitude CPM. These results suggested that CPM at high altitude increased the
accumulation of 10 triterpenoids by up-regulating the gene expression levels of seven key enzymes in the biosynthetic pathway of triterpenoids precursors, which was consistent with the results of triterpenoids content measurement in the populations at high and low altitudes.

3.5 Association analysis of transcriptomics and metabolomics

The association analysis of transcriptome and metabolome information based on signal pathway enrichment is an notably method to understand metabolites the pathways and key enzymes related to their synthesis(Cavill, Jennen, Kleinjans, & Briedé, 2016). In this study, all the obtained DEGs and DAMs were mapped to the KEGG pathway database to obtain their common pathway information, and the main biochemical pathways and signal transduction pathways in which the DEGs and DAMs were shared (results are shown in Table S7).

Transcriptomics and metabolomics mapped together to 65 pathways in the KEGG database, and a total of 8 genes and 1 metabolite were co-enriched in the "sesquiterpene and triterpene biosynthesis pathway" (results are shown in Table S7). Beta-Caryophyllene is a differential metabolite of high altitude and low altitude samples enriched in the "sesquiterpene and triterpene biosynthesis pathway", a total of 500 differentially expressed genes were associated with it, among them 390 were significantly correlated with it ($P<0.05$). The visual analysis results of pathway enrichment in KEGG database showed (Figure S4) that farnesyl pyrophosphate farnesyl transferase was up-regulated with altitude in CPM, and was the key catalytic enzyme in the conversion of triterpenes precursors in CPM. At the same time, beta-amyrin (EC number 5.4.99.39) was up-regulated with altitude, in triterpenoid biosynthesis, beta-amyrin can catalyze polycyclizations, resulting in the formation of new carbon-carbon bonds and chiral centers, which in turn form various types of triterpenoids(Hoshino, 2017). These results suggest that altitude can increase the accumulation of triterpenoids in CPM by up-regulating the gene expression levels of key enzymes in the triterpenoid biosynthetic pathway.

4. Discussion

4.1 Effects of altitude on sesquiterpene and triterpene accumulation

A total of 54 differential metabolites were obtained from the CPM between the high altitude and low altitude, among which 31 DAMs were significantly up-regulated in the CPM at high altitude ($P<0.05$), and 23 DAMs were significantly up-regulated in the CPM at low altitude ($P<0.05$). Further analysis showed that the content of 10 triterpenes in 6 categories was significantly higher in high-altitude samples. The results show that altitude has a significant impact on the accumulation of triterpenes, and can promote the accumulation of triterpenes which is consistent with the accumulation of carotenoids in Tibetan peach fruit in high altitude areas(Zheng et al., 2023).
Triterpenoids such as taraxerol, friedelin and 24-methylenecycloartanol were found to be anti-inflammatory (Ji, Chen, & Wang, 2016), anti-diabetic (Sunil et al., 2021), hypolipidemic (Jiao, Zhang, Lou, Wu, & Zhang, 2007), anti-tumor (Hong et al., 2016), and neuroprotective (Berté et al., 2018) showed good activity. In this study, the contents of dandelion taraxerol and friedelin in high altitude samples were 2.97 and 1.41 times of those in low altitude samples. These results indicated that CPM showed higher sesquiterpenoids and triterpenoids accumulation at high altitude, so it was speculated that CPM growing at low altitude had more active anti-inflammatory, anti-diabetes, hypolipidemia, anti-tumor and neuroprotective effects than C. pilosula var. modesta growing at low altitude.

4.2 Effect of altitude on sesquiterpene and triterpene biosynthesis

Environmental factors such as temperature, atmospheric pressure, ultraviolet radiation, humidity, wind speed and geology in high altitude area are significantly different from those in low altitude area, which results in severe environmental conditions such as high intensity ultraviolet radiation, low oxygen and low temperature in high altitude area and low incidence of pathogens (Arias, Berli, Fontana, Bottini, & Piccoli, 2022). The harsh natural environment at high altitude created favorable conditions for the accumulation of a large number of secondary metabolites in medicinal plants (Burlando, Comara, & Boggia, 2022; Du, Lin, Yu, Zhu, & Li, 2021; X. Liu et al., 2022; Zhu et al., 2022). The accumulation of such compounds is closely related to abiotic stresses such as extreme temperature, ultraviolet radiation and drought, and is crucial for species to regulate growth and development, protect plants from pests and diseases, and adapt to extreme environments (Mechri, Tekaya, Attia, Hammami, & Chehab, 2020; Tholl, 2015; Xu et al., 2019). β-caryophyllene plays an important role in plant resistance to biotic/abiotic stresses (Bidabadi, VanderWeide, & Sabbatini, 2020; Vitiello et al., 2020).

Extreme ecosystems may be the source of important new microorganisms that can adapt to the environment and regulate the accumulation of secondary metabolites in plants. It has been studied that cold-loving and cold-tolerant microorganisms that can promote plant growth have been isolated from plants growing in cold environments (Tapia-Vázquez et al., 2020). The "microbial metabolism in different environments" pathway is the pathway obtained by enrichment of 184 upregulated DEGs in KEGG database (Figure S2). The upregulation of DEGs involved in microbial metabolism in different environments indicated that C. pilosula var. modesta at high altitude may have undergone changes in the expression of specific genes related to microbial metabolism in different environments, and these gene expression changes may be caused by harsh environmental factors in high altitude areas such as alpine and cold.

Oxidative stress reaction may be induced by high intensity ultraviolet radiation in natural environment. A total of 5,128 upregulated DEGs were annotated in the GO database, and 645 upregulated DEGs in high-altitude C. pilosula var. modesta samples were annotated in the function of oxidoreductase activity in the GO database (Figure S2). Studies have shown that stress-induced ROS increased the synthesis and metabolism of terpenoids, which are ginseng with Panax notosides (Liao et al., 2018). The results of this
study suggest that *C. pilosula.* var. *modesta* at high altitude may increase the accumulation of sesquiterpenoids and triterpenoids through ROS pathway induced in plants due to high intensity UV radiation under harsh environment.

Kinase plays an important role in the regulation of plant resistance to various stresses and can increase plant tolerance to abiotic stresses(Y. Liu et al., 2023). Kinases play an important role in regulating the biosynthesis of terpenoids(Abbas et al., 2021), while in this study, a total of 344 upregulated DEGs were annotated to the "kinase activity" function in the GO database.In conclusion, the results of this study suggest that high altitude *C. pilosula.* var. *modesta* may promote the tolerance of *C. pilosula.* var. *modesta* to the harsh environment at high altitude by up-regulating the expression of genes related to kinase activity, and further promote the synthesis of sesquiterpene and triterpene compounds.

5. Conclusion

This study shows that there were significant differences between high altitude and low altitude groups of CPM fresh roots, which was manifested in the following aspects including overall secondary metabolites, contents and types of terpenoids, and gene expression levels. The CPM at high altitude were more likely to accumulate triterpenes than those at low altitude. This was associated with increased accumulation of triterpenoids by up-regulating the gene expression levels of seven key enzymes in the triterpenoid precursor biosynthetic pathway. This possible occurs due to the responses of CPM to harsh environmental factors at high altitude. The mechanisms underlying the differences in the accumulation of sesquiterpenes and triterpenes at different altitudes may be related to microbial metabolism in different environments, oxidative stress response induced by high intensity ultraviolet radiation, and up-regulation of kinase activity under abiotic stress. Meanwhile, future studies should focus on exploring the functions of candidate DEGs to accurately identify the major genes affecting the biosynthesis and accumulation of sesquiterpenoids and triterpenoids at high altitude. These results expand our understanding of how altitude affects plant metabolite biosynthesis, and the functions of genes involved in triterpene biosynthesis and their possible pathways need to be further investigated.

Declarations

Ethics approval and consent to participate

We confirm that all methods in this study were performed in accordance with the relevant guidelines and regulations on plants (either cultivated or wild), and we obtained specific permission to collect *Codonopsis Radix*.

Consent for publication

Not applicable

Availability of data and materials
The datasets generated and/or analysed during the current study are available in the SRA repository, the BioProject ID is PRJNA979796. [http://www.ncbi.nlm.nih.gov/bioproject/979796].

**Competing interests**

The authors declare that they have no competing interests.

**Funding**

This work was supported by the National Key Research and Development Program of China (2018YFC1706300, 2018YFC1706303), the National Natural Science Foundation of China (81972018), the Key Research and Development Project of Gansu Province (21ZD4FA013), the People's Livelihood Special Project from Technical Innovation Guide Plan of Gansu Province (20CX4FK014), the Key Talent Project of Organization Department of Gansu Provincial Party Committee ([2022] Gan-71), the Central Government Guiding Local Science and Technology Development Fund Project (22ZY1QA012), the Talent Innovation and Entrepreneurship Project of Lanzhou (2017-RC-115, 2020-RC-41).

**Authors' contributions**

**Zi-xia Wang**: Writing-original draft, Methodology; Editing, Data Curation & Formal analysis. **Peng-peng Li**: Investigation, Writing-Review & editing. **Yan-jun Jia**: Resources, Methodology & Formal analysis. **Long-xia Wen**: Data Curation, Formal analysis & Validation. **Zhuo-shi Tang**: Investigation, Formal analysis. **Yan-ping Wang**: Methodology, Editing. **Fang Cui**: Editing, Data Curation & Formal analysis. **Ke Sun**: Writing-Review, Editing. **Fangdi Hu**: Conceptualization, Validation; Supervision & Project administration.

**Acknowledgements**

Not applicable

**References**


doi:10.3109/08923973.2015.1122617  

doi:10.1002/ptr.2223  


doi:10.1002/bab.1477  

doi:10.1002/jssc.202200835  


Non-targeted metabolomics analysis of *Codonopsis pilosula* var. *modesta* (Nannf.) L.T.Shen in in high and low altitude. (A) Heatmap of metabolomics in high and low altitude *Codonopsis pilosula* var. *modesta* (Nannf.) (B) Principal component analysis score plot of C. pilosula var. modesta using metabolomics in high and low altitude. (C) Partial Least Squares Discriminant Analysis and (D) cross-validation of C. pilosula var. modesta using metabolomics in high and low altitude. Metabolites enriched in KEGG pathways under negative ion mode (E) and positive ion mode (F).
Figure 2

Different concentrations of triterpenoids in high and low altitude *Codonopsis pilosula*. var. *modesta*(Nannf.) L.T.Shen. (A) Heatmap of triterpenoids level in high and low altitude *Codonopsis pilosula*. var. *modesta*(Nannf.) L.T.Shen.(L1~L3 represent low altitude *C. pilosula*. var. *modesta*, H1~H3 represent high altitude *C. pilosula*. var. *modesta*). Darker red represents higher content of compounds, and darker blue represents lower content of compounds. Each row represents an individual compound, and each column represents an individual sample. Dendrograms indicate the correlation between groups of terpenes. (B) Principal component analysis score plot of *C. pilosula*. var. *modesta* using triterpenoids composition in high and low altitude. (C) Bar graph showing the concentration of triterpenoids compounds of each chemotype in high and low altitude *C. pilosula*. var. *modesta*. 
Figure 3

Transcriptome analysis of *Codonopsis pilosula. var. modesta*(Nannf.) L.T.Shen in high and low altitude. (A) Pearson correlation analysis of gene expression level. (B) FPKM distribution box plot of *C. pilosula. var. modesta* in high and low altitude. (C) Venn diagram representation of the number of differentially expressed genes (DEGs) identified by transcriptome analysis. (D) Heatmap of DEGs. The horizontal coordinate is the sample and the clustering result of the sample, and the vertical coordinate is the clustering result of differential genes and genes. The color represents the expression level of genes in the sample. The redder the color, the higher the expression level, and the greener the expression level, the lower the expression level.
Figure 4

Schematic of triterpenoids in high and low altitude *Codonopsis pilosula* var. *modesta*(Nannf.) L.T.Shen. Heatmap showing the normalized expression (FPKM) of genes involved in the terpenoid backbone biosynthesis pathway. The data were normalized by rows using the function “scale”. The colors are installed in deep red and deep blue that depict the relative expression of genes from high to low.
Supplementary Files

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

- 20230904Supplementarymaterials.docx
- GraphicalAbstract.png