Plant Molecular Biology
https://doi.org/10.1007/s11103-019-00944-5

Metabolome and transcriptome profiling reveals quality variation and underlying regulation of three ecotypes for *Cistanche deserticola*

Xiao Sun[1] · Lin Li[1] · Jin Pei[2] · Chang Liu[1] · Lin-Fang Huang[1,2]

Received: 14 July 2019 / Accepted: 4 December 2019

Abstract

Cistanche deserticola is a plant used both as food and medicine. We are interested in understanding how *C. deserticola* responds to environmental conditions. Samples were collected from three ecotypes grown in saline–alkali land, grassland and sandy land. Transcriptome and metabolome analysis were performed by using RNA-seq and LC–ESI–MS/MS. Among 578 metabolites identified, 218, 209 and 215 compounds were found differentially produced among the three ecotypes. Particularly, 2′-acetylacteoside, belonging to phenylethanoid glycosides (PhGs) is the most significantly differentially produced with a VIP > 0.5 and fold change > 2, representing a potential chemical marker to distinguish the three ecotypes. RNA-Seq analysis revealed 52,043 unigenes, and 947, 632 and 97 of them were found differentially expressed among the three ecotypes. Analysis of the correlation between the metabolome profiles and transcriptome profiles among three ecotypes identified that the 12 key genes related to PhGs biosynthesis were differentially expressed. Particularly, the expression of *PAL*, *ALDH* and *GOT* genes were significantly up-regulated in saline–alkali land compared to the other two. In summary, we found PhGs content was higher in saline–alkali land compared with other ecotypes. This is likely due to the up-regulation of the PhGs biosynthetic genes in response to the saline–alkali conditions.

Keywords *Cistanche deserticola* · Quality variation · Phenylethanoid glycosides (PhGs) · Phenylalanine ammonia-lyase (*PAL*) · Metabolomics · RNA-seq

Abbreviations

PhGs	Phenylethanoid glycosides
PAL, EC:4.3.1.24	Phenylalanine ammonia-lyase
C4H, EC:1.14.14.91	Cinnamic 4-hydroxylase
4CL, EC:6.2.1.12	4-Coumarate-CoA ligase
GOT/PAT, EC:2.6.1.1)	Aspartate aminotransferase
ALDH, EC:1.2.1.3	Aldehyde dehydrogenase
TyDC, EC: 4.1.1.25	Tyrosine decarboxylase
ADT2, EC:4.2.1.91	Arogenate dehydratase
AOC3, EC:1.4.3.21	Monoamine oxidase
DDC, EC:4.1.1.26	Aromatic-L-amino-acid decarboxylase
CuAO, EC:1.4.3.22	Copper amine oxidase
SBP, EC:1.11.1.7	Peroxidase
CYP98A3, EC:1.14.14.96	5-O-(4-coumaroyl)-D-quinate 3′-monooxygenase

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11103-019-00944-5) contains supplementary material, which is available to authorized users.

✉ Jin Pei
peixjin@163.com

✉ Chang Liu
cliu6688@yahoo.com

✉ Lin-Fang Huang
lfhuang@implad.ac.cn; 15801545922@139.com

Xiao Sun
934305103@qq.com

[1] Engineering Research Center of Tradition Chinese Medicine Resource, Ministry of Education, Key Research Laboratory of Traditional Chinese Medicine Resources Protection, Administration of Traditional Chinese Medicine, National Administration of Traditional Chinese Medicine, Institute of Medicinal Plant Development, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing 100193, China

[2] Chengdu University of Traditional Chinese Medicine, Chengdu 61137, Sichuan, China

Published online: 16 December 2019

Introduction

Cistanche deserticola is an edible and medicinal plant distributed primarily in North Africa, Arabic, and Asian countries. It is also eaten as food in Iran, India, Japan and Mongolia (Wang et al. 2017). The fleshy stem of *C. deserticola* has been commonly used as a tonic in China, Korea and Japan to improve memory, enhance sexual function, protect liver, tonify kidney, etc. (Gu et al. 2013; Li et al. 2016; Wang et al. 2012; Xu et al. 2017). Modern phytochemical studies on *C. deserticola* led to the identification and isolation of various components including phenylethanoid glycosides (PhGs), iridoids, lignans, and polysaccharides (Jiang and Tu 2009). Among them, the PhDs were found to be the active components (REF). The international market demand for the *C. deserticola* products has skyrocked in recent years, leading to the over-exploitation and near exhaustion of the *C. deserticola* natural resources (Zheng et al. 2014). Consequently, determine the suitable and optimal environment for *C. deserticola* growth and the understanding of the underlying regulation are in urgent need.

C. deserticola has an extensive range of adaptability and mainly grows in habitats with an altitude of 225–1150 m. It can grow in dry habitats such as sandy land and dry river ditch, but also under stress conditions such as saline–alkali land. *C. deserticola* mainly naturally distribute in three ecotypes of saline–alkali land, grassland and sandy land. On one hand, *C. deserticola* plays an important role in the improvement of ecological environment of the arid regions. On the other hand, *C. deserticola* adapted to is habitats and forms unique ecotypes that might present unique chemical profiles and gene expression profiles. This properties will in turn define various pharmacological activity of the ecotype, which are translated into the quality of the products (Huang and Juan 2012).

Previously, Fourier transform infrared spectroscopy (FTIR) combined with two-dimensional correlation spectroscopy (2D-IR) has been used to analyze and evaluate water-soluble polysaccharide, galactitol and phenylethanoid glycosides in the different parts of *C. deserticola* (Rong et al. 2009). However, the distribution of PhGs and other main active components in different ecotypes of *C. deserticola* has not been reported. *C. deserticola* in Xinjiang and Inner Mongolia has differentiated into two ecotypes by RAPD technique (Cui et al. 2004). It has been hypothesized that these ecotypes possess distinct characteristics, chemical compositions and pharmacological difference, however, detailed analysis has not been conducted. Our previous study shows that the different qualities of *C. deserticola* in different locations, but the mechanism is unclear (Zheng et al. 2014).

Molecular profiling approaches, such as metabolomics and transcriptomics have been used to explore the underlying molecular mechanisms responsible for the plant's differential metabolomes. The metabolite profiles reflect the biological changes caused by genetic variations and/or environmental disturbances (Liu et al. 2019). For non-model plants without reference gene sequences, transcriptome sequencing (RNA-seq) is an efficient approach based on next-generation high-throughput sequencing technology, providing broad coverage and unbiased expression of transcription abundance (Zhang et al. 2019). Therefore, it can be used to understand the relationship between gene function and metabolic phenotype (Abdelrahman et al. 2019; Chen et al. 2018). Relevant reports indicate that the Pearson correlation and spearman correlation coefficients can be used to analyze the correlation between transcriptome and metabolome (Cavill et al. 2016; Hao et al. 2019). Integrated metabolomics and transcriptome analysis are powerful methods for understanding the molecular mechanisms in various biological processes (Cho et al. 2016; Lou et al. 2014).

In this study, we carried out qualitative and quantitative analysis of chemical constituents such as PhGs from different ecotypes by metabolomics analysis. Comparative analyses then revealed the quality variation of *C. deserticola* from the three ecotypes. Simultaneously, we analyzed the genetic diversity and differential gene expression of *C. deserticola* from the three ecotypes. Particularly, we identified and analyzed the genes involved in the biosynthesis of PhGs. For the first time, we used transcriptome and metabolome correlative analysis to assess the quality differences of *C. deserticola* in different ecotypes. Our findings provide a new insight into the quality variation and underlying regulation of *C. deserticola* from three ecotypes. Results obtained from this study have laid the foundation for molecular breeding, optimized cultivation and the selection of most suitable sites for *C. deserticola*.

Materials and methods

Plant materials

The fresh succulent stems for *C. deserticola* in excavation stage were collected from three different ecotypes, Ebinur Lake of Xinjiang (A: saline–alkali land), Tula Village of Xinjiang (B: grassland) and Alxa Left Banner of Inner Mongolia (C: sandy land) in northwestern China (Fig. 1; Table 1). The voucher specimens were deposited in the herbarium of the Institute of Medicinal Plant Development at the Chinese Academy of Medical Sciences in Beijing, China. After cleaning, the succulent stem tissues were cut into small pieces and immediately frozen in liquid nitrogen, and then stored at − 80 °C until further processing.

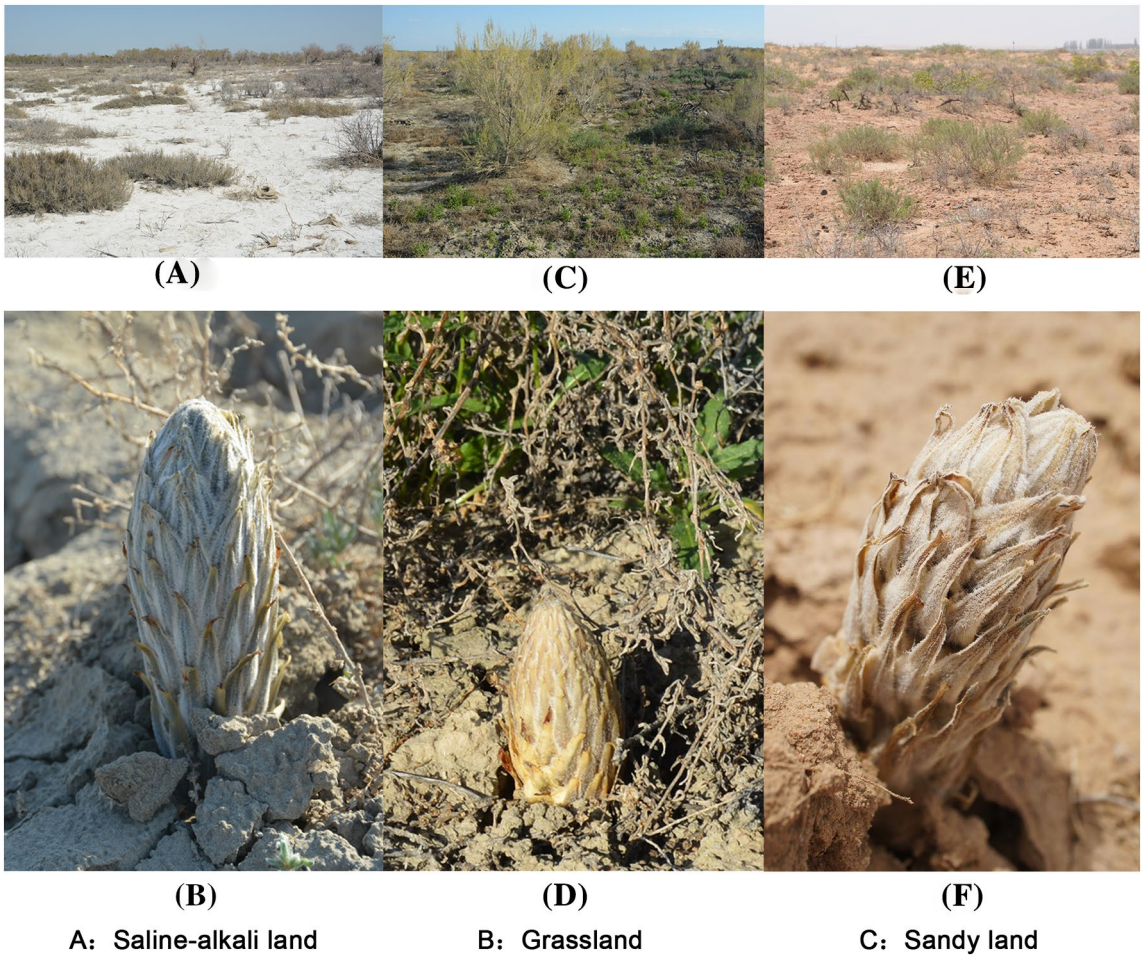

Fig. 1 Photograph shows the *Cistanche deserticola* in three ecotypes. **a**, **b** saline–alkali land; **c**, **d** grassland; **e**, **f** sandy land

Table 1 List of *Cistanche deserticola* samples used in this study

Sample number	Tissue	Sampling site	Ecotypes	Longitude	Latitude	Altitude (m)	Ecotype
HM1	Meat stem	Ebinur Lake, Xinjiang	Saline–alkali land	83.358675	44.881659	211	A
HM2	Meat stem	Ebinur Lake, Xinjiang	Saline–alkali land	83.15277	44.74575788	199	A
HM3	Meat stem	Ebinur Lake, Xinjiang	Saline–alkali land	83.356425	44.825635	215.43	A
HM4	Meat stem	Tula Village,Xinjiang	Grassland	85.540477	46.498027	824.76	B
HM5	Meat stem	Tula Village,Xinjiang	Grassland	85.548162	46.493541	797.3	B
HM6	Meat stem	Tula Village,Xinjiang	Grassland	85.556225	46.483256	767.32	B
HM26	Meat stem	Alxa Left Banner, Inner Mongolia	Sandy land	105.848988	38.834672	2221.87	C
HM27	Meat stem	Alxa Left Banner, Inner Mongolia	Sandy land	105.383916	38.828163	1316.97	C
HM28	Meat stem	Alxa Left Banner, Inner Mongolia	Sandy land	105.437577	38.725391	1307.6	C

Sample Preparation and Extraction

The freeze-dried stem was crushed using a mixer mill (MM 400, Retsch) with a zirconia bead for 1.5 min at 30 Hz. Stem (100 mg) powder was weighted and extracted overnight at 4 °C with 1.0 ml 70% aqueous methanol. Following centrifugation at 10,000×*g* for 10 min, the extracts were absorbed (CNWBOND Carbon-GCB SPE Cartridge, 250 mg, 3 ml; ANPEL, Shanghai, China, www.anpel.com.cn/cnw) and filtrated (SCAA-104, 0.22 μm pore size; ANPEL) before LC–MS analysis.

High-performance liquid chromatography (HPLC) conditions

The sample extracts were analyzed using an LC–ESI–MS/MS system (HPLC, Shim-pack UFLC SHIMADZU CBM30A system). The analytical conditions were as follows, HPLC column: Waters ACQUITY UPLC HSS T3 C18 (1.8 µm, 2.1 mm × 100 mm); Solvent: water (0.04% acetic acid): acetonitrile (0.04% acetic acid); Gradient program: 100:0 v/v at 0 min, 5:95 v/v at 11.0 min, 5:95 v/v at 12.0 min, 95:5 v/v at 12.1 min, 95:5 v/v at 15.0 min; flow rate: 0.40 ml/min; temperature: 40 °C; injection volume: 2 µl. The effluent was alternatively connected to an ESI-triple quadrupole-linear ion trap (Q TRAP)-MS. A quality-control sample was prepared by equal blending of all samples; during the assay, the quality control sample was run every 10 injections to monitor the stability of the analytical conditions (Gu et al. 2013; Li et al. 2016; Wang et al. 2012).

ESI-Q TRAP-MS/MS

LIT and triple quadrupole (QQQ) scans were acquired on a triple quadrupole-linear ion trap mass spectrometer (Q TRAP), API 6500 Q TRAP LC/MS/MS System, equipped with an ESI Turbo Ion-Spray interface, operating in a positive ion mode and controlled by Analyst 1.6 software (AB Sciex). The ESI source operation parameters were as follows: ion source, turbo spray; source temperature 500 °C; ion spray voltage (IS) 5500 V; ion source gas I (GSI), gas II (GSII), curtain gas (CUR) were set at 55, 60, and 25.0 psi, respectively; the collision gas (CAD) was high. Instrument tuning and mass calibration were performed with 10 and 100 µ mol/l polypropylene glycol solutions in QQQ and LIT modes, respectively. QQQ scans were acquired as MRM experiments with collision gas (nitrogen) set to 5 psi. DP and CE for individual MRM transitions was done with further DP and CE optimization. A specific set of MRM transitions were monitored for each period according to the metabolites eluted within this period.

Metabolite Identification and Quantification

Qualitative analysis of primary and secondary MS data was carried out by comparison of the precursor ions (Q1), product ions (Q3) values, the retention time (RT), and the fragmentation patterns with those obtained by injecting standards using the same conditions if the standards were available (Sigma-Aldrich, USA http://www.sigmaaldrich.com/united-states.html) or conducted using a self-compiled database MWDB (MetWare biological science and Technology Co., Ltd., Wuhan, China) and publicly available metabolite databases if the standards were unavailable. Repeated signals of K^+, Na^+, NH_4^+, and other large molecular weight substances were eliminated during identification. The quantitative analysis of metabolites was based on the MRM mode. The characteristic ions of each metabolite were screened through the QQQ mass spectrometer to obtain the signal strengths. Integration and correction of chromatographic peaks was performed using MultiQuant version 3.0.2 (AB SCIEX, Concord, Ontario, Canada). The corresponding relative metabolite contents were represented as chromatographic peak area integrals.

The intensities of mass peaks for each sample were sum-normalized and Pareto-scaled using the SIMCA-P software package. Principal component analysis (PCA) and orthogonal partial least-squares discriminant analysis (OPLS-DA) with data from 9 samples (three ecotypes × three biological replicates) were performed to observe differences in metabolic composition among the three ecotypes of *C. deserticola*. Metabolites with significant differences in content were set with thresholds of variable importance in projection (VIP) ≥ 0.5 and fold change ≥ 2 or ≤ 0.5.

RNA extraction, sequencing, and de novo assembly

Stems were sampled at three different ecotypes for transcriptome sequencing experiments, and three biological replicates were used. The RNA-Seq analysis was performed by the Allwegene Company (Beijing). RNA was extracted using TRIzol-A + reagent (TIANGEN BIOTECH, Beijing) and treated with RNase-free DNase I (TaKaRa). RNA quantity was measured using a Nanodrop Quibt 2.0 Flurometer (Life Technologies, CA, USA). RNA quality was evaluated with an Agilent Bioanalyzer Model 2100 (Agilent Technologies, Palo Alto, CA). Samples with an RNA Integrity Number (RIN) value greater than 7.5 were deemed acceptable according to the Illumina transcriptome sequencing protocol. All sequencing reads from the nine libraries have been submitted to the Sequence Read Archive (SRA) in GenBank (BioProject PRJNA528225).

Samples were cleaned using the QIAquick PCR kit, and were eluted by eluent buffer for end repairing and sequencing adapter joining. Then the cDNA fragments were separated by agarose gel electrophoresis and fragments of 100 ± 300 bp were enriched by PCR amplification to create cDNA libraries. The constructed cDNA libraries were then sequenced on the Illumina HiSeq 2500 platform. The raw reads in FASTQ format were first processed using in-house perl scripts. In this step, clean reads were obtained by removing reads containing adapter, ploy-N, and low quality reads from raw reads. At the same time, Q20 and GC content of the clean data were calculated. Assembly of the clean reads was performed using Trinity (Grabherr et al. 2011).

Validation of RNA-seq data by qRT-PCR

Total RNA was reverse transcribed to generate the first strand cDNA using a PrimeScript™ RT reagent Kit with gDNA Eraser (Takara, Kyoto, Japan, RR047B). The cDNA was diluted 1:4 in distilled water and used as template for qPCR. The *C. deserticola* actin-97 gene was used as an internal control for normalization. The qPCR reaction was performed on an ABI 7500 qPCR instrument. Real-time PCR reactions were performed using SYBR® Premix Ex Taq™ II (Tli RNase H Plus, Takara Bio, Kusatsu, Japan), ROX plus (RR82LR, TaKaRa Bio, Kusatsu, Japan). The $2^{-\Delta\Delta C_T}$ method (Livak and Schmittgen 2001) was used for relative expression quantitative analysis. The primer sequences used in the qPCR analysis are shown in Table S1.

Gene annotation, analysis of differential gene expression and SNP calling

All expressed genes were functionally annotated using seven databases, including Nr (NCBI non-redundant protein sequences); Nt (NCBI non-redundant nucleotide sequences); Pfam (protein family); KOG/COG (clusters of orthologous groups of proteins); Swiss-Prot (a manually annotated and reviewed protein sequence database); KO (KEGG ortholog database); GO (gene ontology). The sequence reads for each sample were remapped to the reference sequences using RSEM software (Bo and Dewey 2011). Gene expression levels were measured using the FPKM (fragments per kilobase of transcript per million fragments) (Cole et al. 2010) based on the number of uniquely mapped reads. For genes with more than one alternative transcript, the longest transcript was selected to calculate the FPKM. Differential expression analysis of the three groups was performed using the DESeq R package (1.10.1). DESeq provides statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. The resulting P values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with an adjusted P value < 0.05 found by DESeq were assigned as differentially expressed. The differences of gene expression in the succulence stem of *C. deserticola* in three ecotypes were screened by Audic (Audic and Claverie 1997) method. Picard—tools v1.41 and samtools v0.1.18 were used to sort, remove duplicated reads and merge the bam alignment results of each sample. GATK2 software was used to perform SNP calling. Raw vcf files were filtered with GATK standard filter method and other parameters (cluster Window Size: 10; MQ0 > = 4 and (MQ0/(1.0*DP)) > 0.1; QUAL < 10; QUAL < 30.0 or QD < 5.0 or H Run > 5), and only SNPs with distance > 5 were retained.

Correlation analysis between metabolome and transcriptome data

Pearson correlation coefficient was calculated for metabolome and transcriptome data integration. In this study, Log_2 conversion of data was performed uniformly before analysis. For the joint analysis between the metabolome and transcriptome, cor program from R was used in this experiment, and the screening criteria were Pearson correlation coefficient > 0.8. As shown in File S1, the correlation between phenylethanoid glycosides metabolites and their key genes. The relationships between metabolome and transcriptome data were visualized by using Cytoscape (The Cytoscape Consortium, San Diego, CA, USA, version 3.7.0).

PAL, ALDH and GOT enzymes activity measurements

The succulent stem of *C. deserticola* in three ecotypes were weighed by 0.3 g respectively, and immediately homogenized in 3 ml extract by ice bath. The homogenate was centrifuged at 10,000×*g* at 4 °C for 15 min. The supernatant obtained was used for enzyme determination. Enzyme activity was determined using the Solarbio kits. PAL catalyzed the decomposition of L-phenylalanine into trans-cinnamic acid and ammonia. The trans-cinnamic acid had the maximum absorption value at 290 nm. In the presence of coenzyme I, catalytic acetaldehyde and acetaldehyde dehydrogenase NAD + is converted into acetic acid and NADH, at 340 nm absorbance value will increase, determine the absorbance at 340 nm value change, can calculate the activity of ALDH. GOT catalyzed the transamination of alpha-ketoglutaric acid and aspartic acid to form glutamic acid and oxaloacetic acid, and oxaloacetic acid was further decarboxylated to form pyruvate. Pyruvate can react with 2,4-dinitrophenylhydrazine to produce 2,4-dinitrophenylhydrazone, which is brownish red in alkaline condition. By measuring the change of absorbance at 505 nm, the activity of GOT enzyme could be calculated.

Results

Comparison of metabolites produced by three ecotypes of *C. deserticola*

From the total ion chromatograms of UPLC–MS/MS, positive and negative ions were obtained (Fig. S1). The contents of 578 metabolites were determined, of which 221 were primary metabolites and 357 were secondary metabolites. The most abundant metabolites of among the three ecotypes are nucleotides and their derivatives, amino acids and derivatives, fatty acids, phenylpropanoid, phytohormone, organic acids and derivatives, Hydroxycinnamoyl derivatives and

flavonoids, etc. (File S2a; Table S2; File S2). The differences of the metabolites are mostly seen for lipids and amino acids and their derivatives accounting for a large proportion (Fig. S2b).

As the results of PCA analysis (Fig. 2a) showed that the biological replicates of three ecotypes clustered together in different areas, indicating that there were significant differences in the metabolism. The samples from the saline–alkali

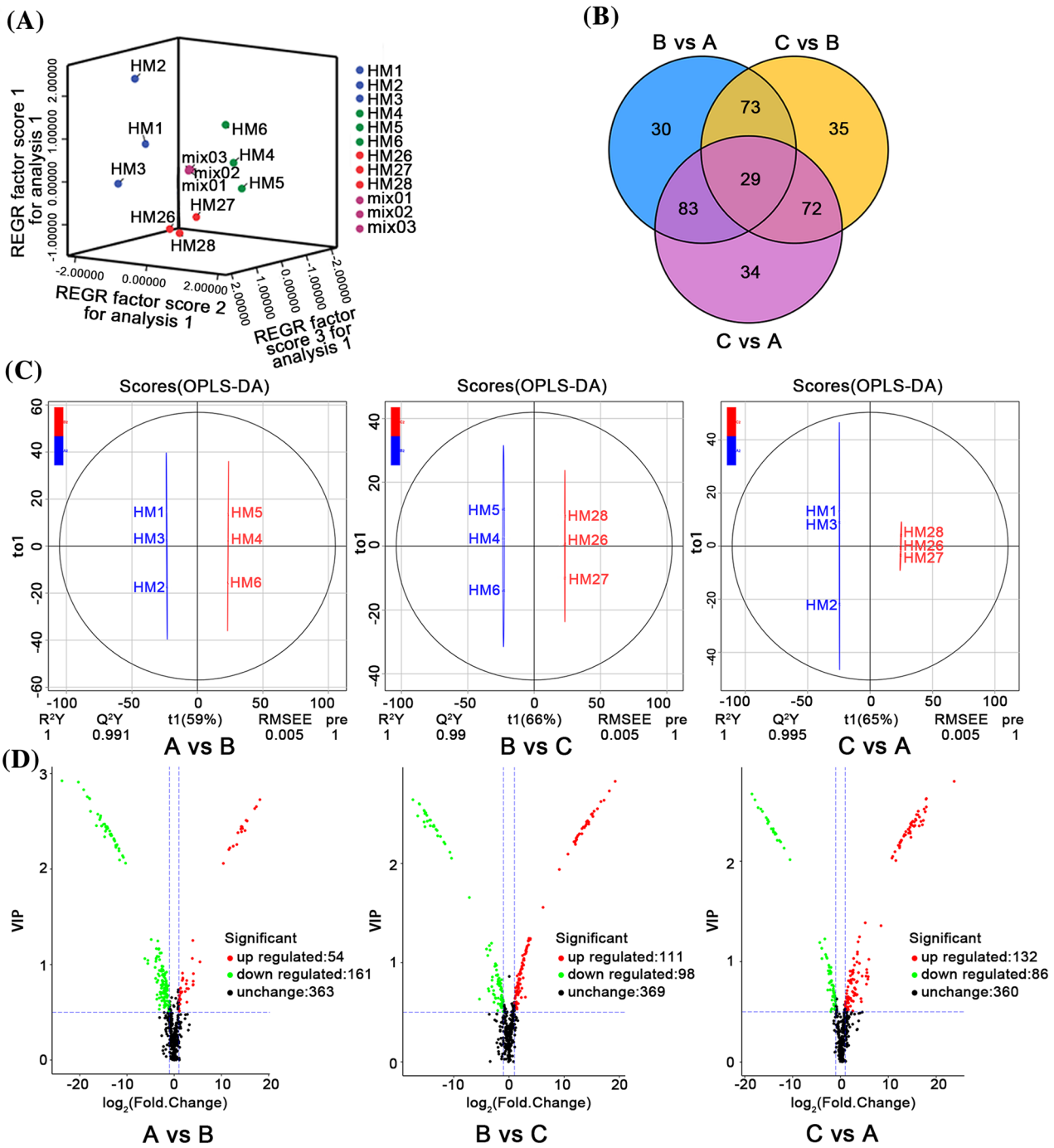

Fig. 2 Metabolite profiles across the three ecotypes of *Cistanche deserticola*. **a** PCA score map. **b** The similarities and differences of metabolite composition. **c** OPLS-DA score map. High predictability (Q2) of the OPLS-DA models was observed in the comparison between saline–alkali land versus grassland (Q2 = 0.991; **c**), as well as between grassland versus sandy land (Q2 = 0.99; **c**), and sandy land versus saline–alkali land (Q2 = 0.995; **c**). **d** Volcano map. HM1–3 are samples from saline–alkali land, HM4–6 are samples from grassland, HM7–9 are samples from sandy land, and mix01–04 are samples for quality control. A: saline–alkali land; B: grassland and C: sandy land

land mainly concentrated in the upper left part of the plot, suggesting that principle components 1 (PC1) and PC2 contributed to the separation of the samples. The samples from grassland mainly concentrated in the right half of the plot, indicating that PC2 had a great influence on them. The samples in the sandy land were far apart from the other two groups. Meanwhile, high predictability (Q^2) of the OPLS-DA models was observed in the comparison of the three ecotypes (Fig. 2c). Differential metabolites were screened based on fold-change ≥ 2 or ≤ 0.5 among the metabolites with a VIP value > 0.5 (File S3).

The screening results have been illustrated using Venn diagrams (Fig. 2b) and Volcano plots (Fig. 2d), there were 215 differentially produced metabolites (54 upregulated and 161 downregulated) between saline–alkali land and grassland. In contrast, 209 metabolites were found to have been differentially produced (111 upregulated and 98 downregulated) between grassland and sandy land respectively. There were 218 differential metabolites (132 upregulated and 86 downregulated) between sandy land and saline–alkali land (Fig. 2d). These differential metabolites in three ecotypes were mapped to the KEGG database (Fig. S3a). The differential metabolites were mainly enriched in "Metabolic pathways", "Biosynthesis of secondary metabolites", etc.

Differentially expressed genes among the three ecotypes of *C. deserticola*

By using RNA-seq technology, 7.54G, 7.50G and 7.52G of raw data were obtained from samples of the three ecotypes with the total numbers of clean reads are 24,557,662, 24,440,365 and 24,429,502 (Table S3). After assembly, 52,043 unigenes were obtained, 71.51% of them have lengths mainly distributed between 300 and 2000 bp, while 8.64% of them have lengths greater than 2000 bp (Fig. S4). By using BLASTx, these unigenes were compared to five public databases to determine their potential functions. A total of 26,775 (51.45%) unigenes had hits in at least one database, and 49% of unigenes had no hits in any of the database (Table S4). In particular, there were 23,977, 15,556, 7367, 13,333, and 14,292 having hits in the Nr, Nt, KEGG, GO and KOG databases respectively (Fig. S5). 4587 of them had hits in all five databases. The sequence derived from *Sesamum indicum* (43.65%) in the database had the most numbers of homologous sequence with the *C. deserticola*.

The results of transcriptome analysis showed that the gene expression levels were different for *C. deserticola* in three ecotypes. We found the DEGs were significantly enriched in the KEGG pathways that was related to metabolism. The results of expression profiles (Fig. 3a) showed that saline–alkali land (A) and grassland (B) were more similar, while sandy land (C) showed a slightly different expression profile with other ecotypes. It suggested that the *C. deserticola* in saline–alkali land and grassland ecotypes have more similarity in terms of the functions and expression of genes. DEGs results (Fig. 3b, c; File S4) revealed that the expression level of 435 and 512 DEGs were higher or lower in saline–alkali land than that in sandy land. The expression level of 242 and 390 DEGs were higher or lower in grassland ecotypes than that in sandy land, while the expression level of 43 and 54 DEGs were higher or lower in saline–alkali land than that in grassland ecotypes (Fig. 3c).

Regarding the KEGG classification (Fig. S3b), it indicated that the DEGs of *C. deserticola* between saline–alkali land and sandy land were enriched in the 67 metabolic pathways, while DEGs of *C. deserticola* between grassland and sandy land were enriched in the 52 metabolic pathways including "Phenylalanine metabolism", "Plant hormone signal transduction". DEGs of *C. deserticola* between saline–alkali land and grassland were enriched in the 8 metabolic pathways, "Spliceosome", etc. Compared with grassland ecotypes and sandy ecotypes, the most abundant were "Metabolic pathways" and "Biosynthesis of Secondary Metabolites", which were consistent with the enrichment results of differential metabolites (Table S2), illustrating that the synthesis of Secondary Metabolites of *C. deserticola* in three ecotypes were quite different. Metabolic pathway analysis was performed on the transcriptome data of the stem of *C. deserticola* and 295 KEGG pathways were compared (Fig. S6). There were 21 pathways involved in secondary metabolism, including phenylpropanoid biosynthesis, flavonoid biosynthesis, flavone and flavonol biosynthesis, etc. There were 138 Unigenes involved in biosynthesis of phenylpropanoids. About 25,288, 22,563 and 16,940 SNP loci were detected in *C. deserticola* in saline–alkali land, grassland and sandy land, respectively (Table S5). As shown in Fig. S7, the first 20 metabolic pathways. The SNP enrichment pathways in the three ecotypes of *C. deserticola* were basically the same.

qPCR Validation

To validate the key RNA-Seq results, we selected six DEGs (PhGs biosynthetic genes) (Fig. S8a–f) and analyzed their expression levels in three ecotypes using qPCR. The expression patterns of these genes were similar to the RNA-Seq results, with correlation coefficients (R^2) > 0.5429 (Fig. S8g). The results validated the relevance of the RNA-Seq data and qPCR showed good consistency for both up- and down-regulated gene expression.

Analysis of phenylethanoid glycosides (PhGs)

The result turned out that the *C. deserticola* in saline–alkali land had the best quality, as well as a consistency with the significantly up-regulated expression of key genes such as

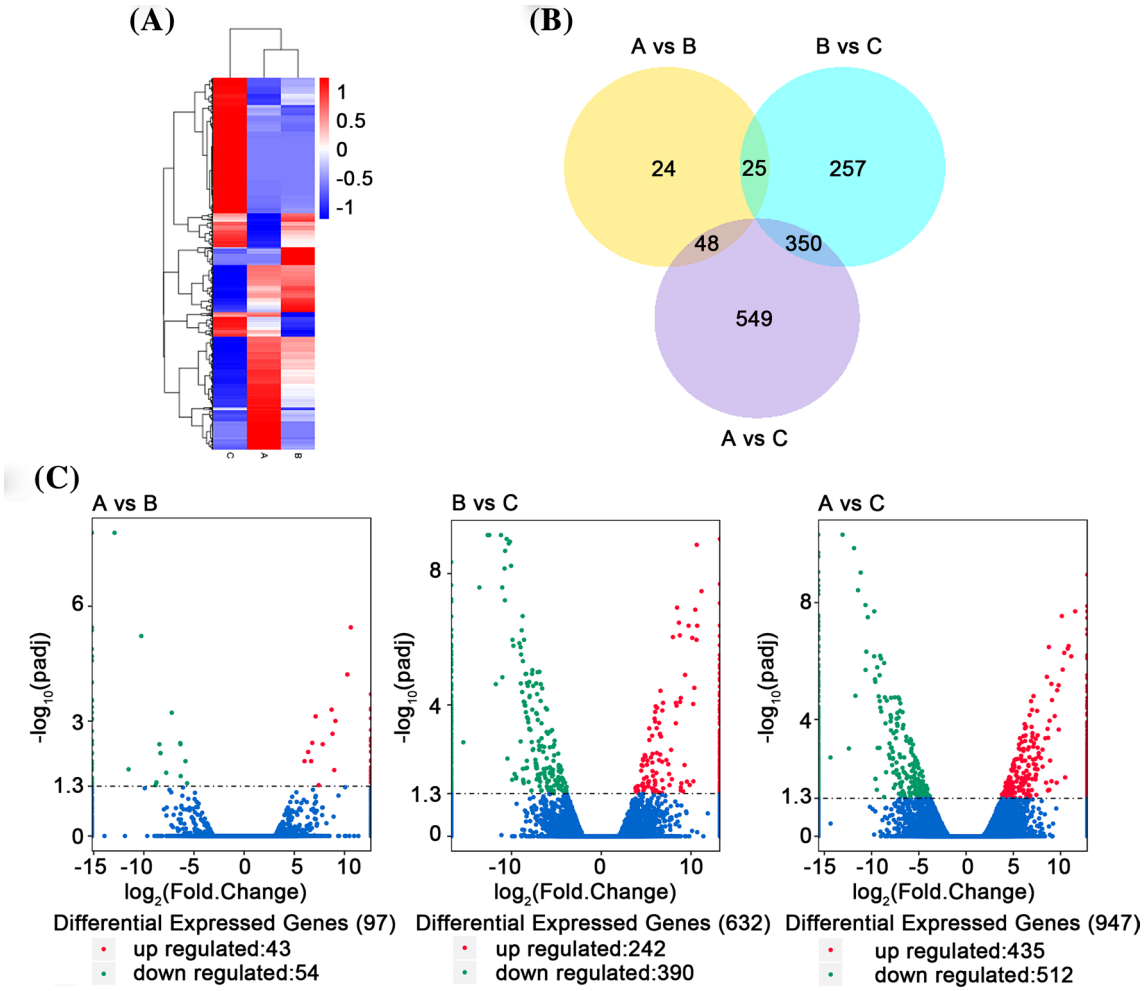

Fig. 3 Transcription profiles in the meat stems of *Cistanche deserticola* from three ecotypes. **a** Heatmap and hierarchial clustering showing differentially expressed genes (DEGs); **b** Venn diagram showing the distribution of DEGs from the three ecotypes; **c** volcano map showing the $\log_{10}$(P value) and $\log_2$(fold change) of DEGs. A: saline–alkali land; B: grassland and C: sandy land

PAL, GOT and *ALDH* genes related to PhGs biosynthesis in saline–alkali land. Meanwhile, twelve candidate genes related to PhGs were screened. A total of five substances most representative of PhGs in *C. deserticola* including echinacoside, cistanoside A, isoacteoside, acteoside and 2′-acetylacteoside were detected (Fig. 4; Table S6). These five PhGs in saline–alkali land, grassland and sandy land showed a decreasing trend. There were significant differences in the PhGs between saline–alkali land and sandy land. The contents of PhGs in grassland were closer to that of saline–alkali land. The contents of 2′-acetylacteoside, isoacteoside and acteoside in the saline–alkaline land were higher than the other two ecotypes. The above three metabolites were the lowest in the sandy land. In contrast, the content of cistanche A and echinacoside in sandy land *C. deserticola* was significantly higher than in the other two ecotypes. Among them, the above two metabolites in saline–alkali land were the lowest.

According to KEGG enrichment results based on differential metabolites, the contents of 2′-acetylacteoside, isoacteoside, acteoside and cistanoside A were significantly different between saline–alkali land and sandy land (Table S7). Sandy land compared with saline–alkali land, 17, 6 and 2 differential metabolites were enriched in Phenylpropanoid biosynthesis, Phenylalanine metabolism, Phenylalanine, tyrosine and tryptophan biosynthesis, respectively. Sandy land compared with grassland, 10, 7 and 4 different metabolites were enriched in the above three metabolic pathways, respectively. saline–alkali land compared with grassland, 5, 3 and 2 different metabolites were enriched in the above three metabolic pathways, respectively. This result was consistent with the difference in the PhGs contents from the three ecotypes (Fig. S3b).

Based on the annotation information of the transcriptome of *C. deserticola*, 12 PhGs candidate genes were screened, including Phenylalanine ammonia-lyase

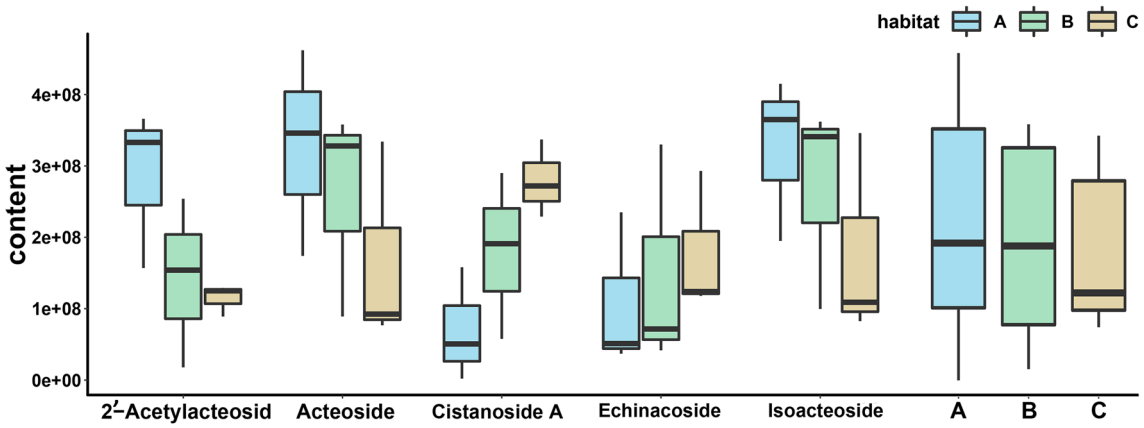

Fig. 4 Box plots of phenylethanoid glycosides (PhGs) contents in the stems of *Cistanche deserticola* from the three ecotypes. Maximum and minimum values of a metabolite among three biological replicates are represented at the upper and lower ends of the whisker, respectively, and their 75th and 25th percentiles are represented at upper and lower ends of the box, respectively. On the right is the box diagram of total PhGs of three ecotypes *Cistanche deserticola.* A: saline–alkali land; B: grassland and C: sandy land

(*PAL*, EC:4.3.1.24), cinnamic 4-hydroxylase (*C4H*, EC:1.14.14.91), 4-coumarate-CoA ligase (*4CL*, EC:6.2.1.12), Aspartate aminotransferase (*GOT*, EC:2.6.1.1), Aldehyde dehydrogenase (*ALDH*, EC:1.2.1.3), Tyrosine decarboxylase (*TyDC*, EC: 4.1.1.25), Arogenate dehydratase (*ADT2*, EC:4.2.1.91), Monoamine oxidase (*AOC3*, EC:1.4.3.21), Aromatic-L-amino-acid decarboxylase (*DDC*, EC:4.1.1.26), Copper amine oxidase (*CuAO*, EC:1.4.3.22), Peroxidase (*SBP*, EC:1.11.1.7) and 5-O-(4-coumaroyl)-D-quinate 3′-monooxygenase (*CYP98A3*, EC:1.14.14.96) (Table 2). According to the candidate genes expression profiles of PhGs from three ecotypes *C. deserticola,* the expression levels of most candidate PhGs gene were the highest in the saline–alkali land, except for the *ADT2, DDC, CuAO, SBP and CYP98A3* (Fig. 5; File S7). However, the expression of *ADT2*, *DDC*, *SBP* and *CYP98A3* were the highest in sandy land.

Eight *PAL* candidate genes were identified, of which six were greater than 10 (FPKM > 10). Through the KEGG search, c123670_g1 was enriched into the "phenylalanine metabolism" pathway, which was significantly up-regulated in saline–alkali land. There were 5 candidate genes encoding *GOT*. Four of them were highly expressed in saline–alkali land, followed by the other two ecotypes. c133601_g1 is annotated into the pathway of "biosynthesis of phenylalanine, tyrosine and tryptophan", which was up-regulated in saline–alkali land (Table 2). In addition, we had functional annotations on the genes of SNPs, there were two related pathways for the biosynthetic of PhGs, including phenylpropane biosynthetic pathway and the phenylalanine metabolic pathway (File S6).

Correlation analysis of PAL, ALDH and GOT enzyme activities with gene expression

The results (Fig. 6a) of PAL enzyme activity assay showed that the PAL enzyme activity trend in three ecotypes was grassland > sandy land > saline land and there were significant differences. The results were similar to those of GOT enzyme activity. ALDH enzyme activity was slightly different, the highest in saline–alkali land, followed by grassland, and the lowest in sandy land. However, the trends of enzyme activity were inconsistent with the trends of genes expression. Transcriptome results reflect the transcriptional level, while genes need to undergo complex processes such as translation and modification after transcriptional translation to proteins. Furthermore, enzyme activity is also related to the activation of active groups. Correlation analysis (Fig. 6b) combined the enzyme activity and FPKM value of key genes showed that c123670_g1 in saline–alkali land was negatively correlated with PAL enzyme activity. In ALDH gene, c130727_g2 had a significant negative correlation in sandy land, while c130703_g1 had a significant positive correlation in grassland. For GOT, there was a significant negative correlation in c126983_g1 and c123830_g1 in sandy land and c133601_g1 in grassland.

Correlation analysis between DEGs and metabolites reveals the differential regulatory network of PhGs biosynthesis in three ecotypes of *C. deserticola*

The result of correlative analysis indicated that up-regulated *PAL, GOT and ALDH* genes in saline–alkali land was the potential reasons for the best quality of *C. deserticola* in

Table 2 Candidate genes related to PhGs biosynthesis

	Unigene ID	A	B	C	E value	Identity (%)	Reference sequence ID
PAL	c115887_g2	340.47	288.81	219.16	0	99	GU434100.1
	c115887_g1	1123.95	788.42	754.44	2E−146	100	GU451308.1
	c123670_g1	173.34	52.02	4.94	0	81	HM062775.1
	c131638_g2	2804.72	1205.81	962.27	1E−73	86	DQ408636.1
	c131638_g3	2024.62	1099.36	808.91	6E−80	89	GU434100.1
	c88669_g1	744.90	602.45	621.06	2E−65	90	AF401636.1
	c94257_g1	18.08	5.51	8.86	2E−71	90	JQ277717.1
GOT	c123830_g1	73.28	45.68	46.33	0	88	XM_020695127.1
	c126983_g1	211.47	212.26	81.80	0	84	XM_011092164.2
	c132549_g1	359.66	187.55	142.03	0	89	XM_011091860.2
	c133601_g1	1.85	1.73	0.00	0	86	XR_002701846.1
	c134631_g1	167.14	68.40	48.45	0	86	XM_011092164.2
ALDH	c113557_g1	7.57	3.28	0.65	3.00E−172	86	GQ245974.1
	c117064_g1	16.14	0.00	0.00	0	85	GQ245973.1
	c120686_g2	83.03	66.20	65.31	0	87	AF162665.1
	c121665_g1	6.01	7.02	0.00	7.00E−134	82	AJ306960.1
	c125616_g1	501.39	434.98	117.27	0	86	KP968844.1
	c128761_g1	43.63	30.42	27.31	0	83	EF492045.1
	c129177_g2	2.55	0.08	1.11	6.00E−169	81	AF323586.1
	c130703_g1	273.79	116.59	75.96	5.00E−165	80	DQ286963.1
	c130727_g2	1.07	2.96	3.33	0.00E+00	79	KC109794.1
	c131433_g4	40.70	20.23	18.39	0.00E+00	81	AJ417789.1
	c133796_g4	13.87	6.91	8.14	0.00E+00	82	EF631997.1
	c133919_g1	13.56	15.24	8.93	0.00E+00	83	GQ245973.1
4CL	c96837_g1	112.70	71.21	81.01	2.00E−70	83	FJ230968.1
	c112223_g1	66.97	55.87	61.38	1.00E−68	85	XM_011101255.2
	c116219_g1	86.44	92.60	115.52	2.00E−66	81	KP070830.1
	c122834_g4	17.31	3.03	9.66	0.00E+00	81	KF220557.1
	c130244_g3	161.11	160.19	107.75	2.00E−60	84	KT963457.1
	c133994_g1	364.49	653.84	983.26	0	83	KT963456.1
	c135925_g1	81.64	20.09	14.97	0	83	AY587891.1
	c96837_g3	60.23	50.93	94.22	0	82	KF220556.1
	c134708_g2	86.26	59.34	48.64	0	82	KC576841.1
C4H	c130384_g1	159.71	24.87	12.22	3E−172	81	EU676019.1
	c140054_g1	1.83	0.00	0.00	6E−169	75	XM_020415803.1
TyDC	c111529_g1	12.69	8.01	3.53	0	88	XM_011081824.2
AOC3	c111256_g1	5081.51	3156.65	1811.31	0	83	XM_011082824.2
ADT2	c122819_g1	14.82	22.32	26.23	0	87	XM_011091062.2
	c127251_g3	3.24	3.28	4.23	0	81	XM_020696326.1
DDC	c133291_g1	908.99	1034.24	1760.16	0	84	KU640395.1
CuAO	c120660_g1	12.91	19.63	2.20	0	86	XM_011082825.2
SBP	c122586_g1	1.205	0.15	7.49	4.00E−10	0.9259	XM_006304388.2
CYP98A3	c129155_g1	2.42	0.49	12.10	0.00E+00	81.53	KP070826.1

Columns 3, 4 and 5 represent FPKM values of three ecotypes of *C. deserticola*. A: saline–alkali land; B: grassland and C: sandy land

saline–alkali land. 2′-acetylacteoside could be used as an indicator to judge three ecotypes of *C. deserticola*. In this study, we constructed a metabolic pathway map related to PhGs biosynthetic based on the KEGG database, and labeled differential metabolites and DEGs in the three ecotypes of *C. deserticola* in the pathway (Fig. 7). It mainly included three KEGG pathways: "Phenylpropanoid biosynthesis (Ko00940)", "Phenylalanine, tyrosine and

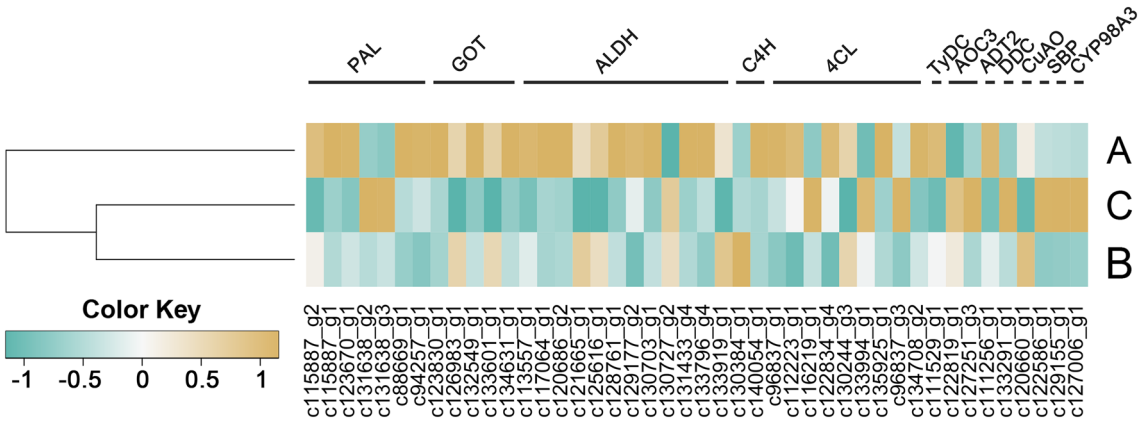

Fig. 5 Heatmap showing the expression profiles of genes related to PhGs across the three ecotypes. The FPKM values of key genes of the three ecotypes were standardized with Z score. The cluster heatmap showed that the gene expression patterns of grassland and sandy land ecotypes were more similar, and the expression amount of key genes was up-regulated in saline–alkali land ecotype. A: saline–alkali land; B: grassland and C: sandy land

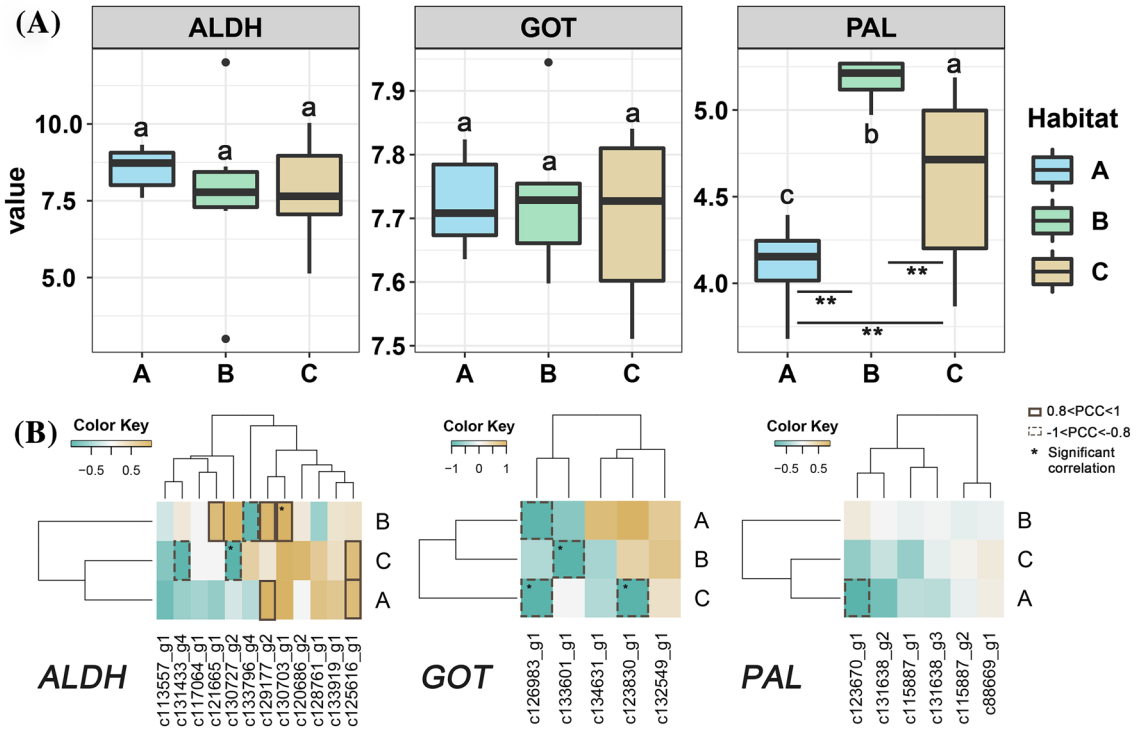

Fig. 6 Correlation analysis of PAL, ALDH and GOT enzyme activities and gene expression. **a** Box plots of PAL, ALDH and GOT enzyme activity. Maximum and minimum values of a metabolite among three biological replicates are represented at the upper and lower ends of the whisker, respectively, and their 75th and 25th percentiles are represented at upper and lower ends of the box, respectively. **Represent $P < 0.01$. **b** Hierarchical cluster heat map based on the results of correlation analysis between enzyme activities and gene expression. A: saline–alkali land C: sandy land

tryptophan biosynthesis (Ko00400)" and "tyrosine metabolism (Ko00350)". Figure 8 showed that the content of metabolites and DEGs in the biosynthetic of PhGs were specifically different depending on the *C. deserticola* (saline–alkali land, grassland and sandy land) in different ecotypes. The expression of *PAL* (EC:4.3.1.24; c123670_g1, c131638_g1) and *GOT* (EC:2.6.1.1; c133601_g1) in *C. deserticola* in saline–alkali land was significantly higher than that in sandy land. Grassland compared with sandy land, the *CYP98A3* (EC:1.14.14.96; c129115_g1) genes was down regulated.

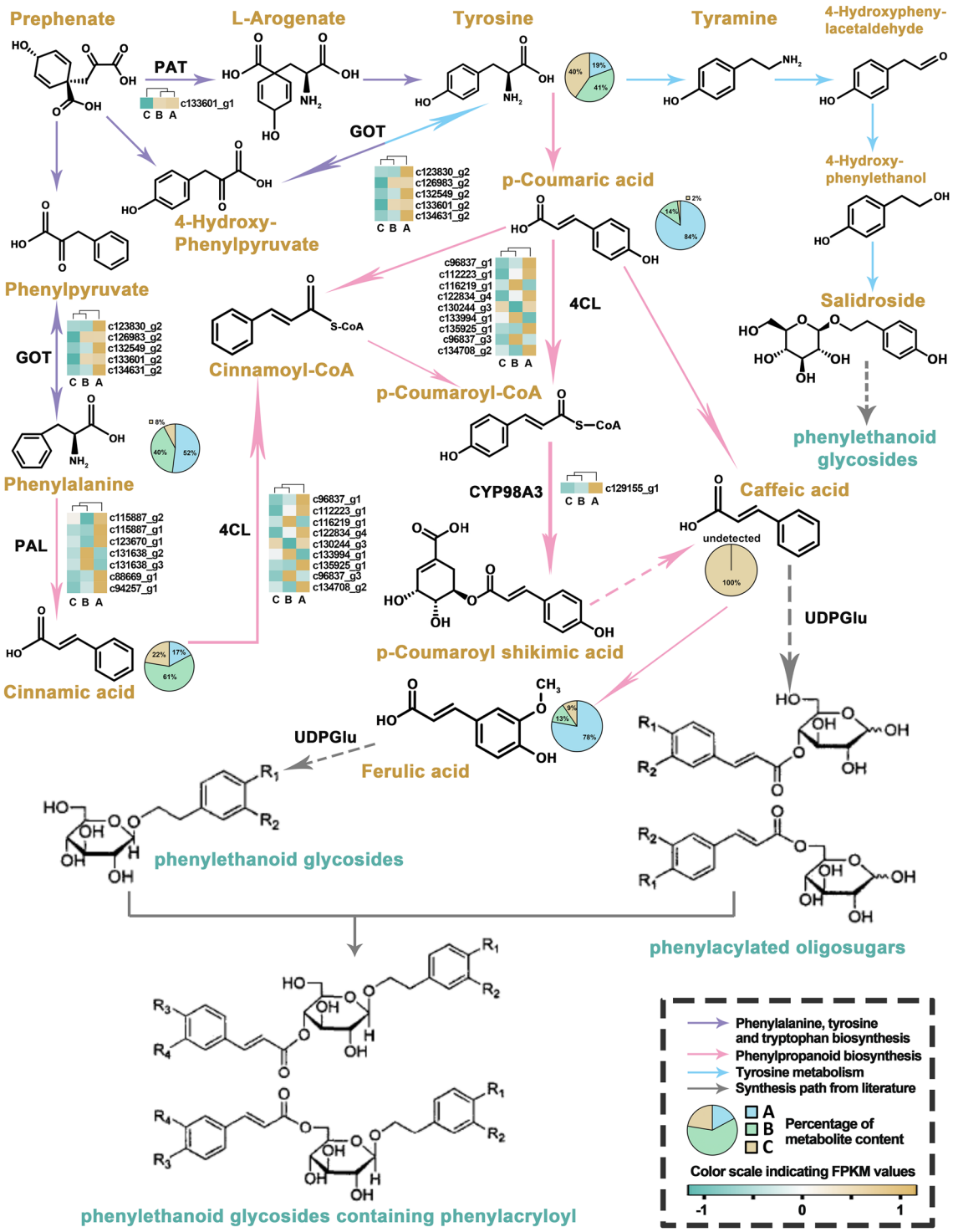

Fig. 7 Pathways and genes involved in the biosynthesis of PhGs in three ecotypes of *Cistanche deserticola*. The purple arrows represent the phenylalanine, tyrosine and tryptophan biosynthetic pathway. The pink arrow represents the phenylpropanoid biosynthetic pathway. The blue arrow represents tyrosine metabolic pathway. The gray arrow represents pathway from literature. Phenylalanine ammonia-lyase (PAL, EC: 4.3.1.24), 4-coumarate-CoA ligase (4CL, EC: 6.2.1.12), Aspartate aminotransferase (GOT/PAT, EC: 2.6.1.1), 5-O-(4-coumaroyl)-D-quinate 3′-monooxygenase (CYP98A3, EC: 1.14.14.96). A: saline–alkali land; B: grassland and C: sandy land

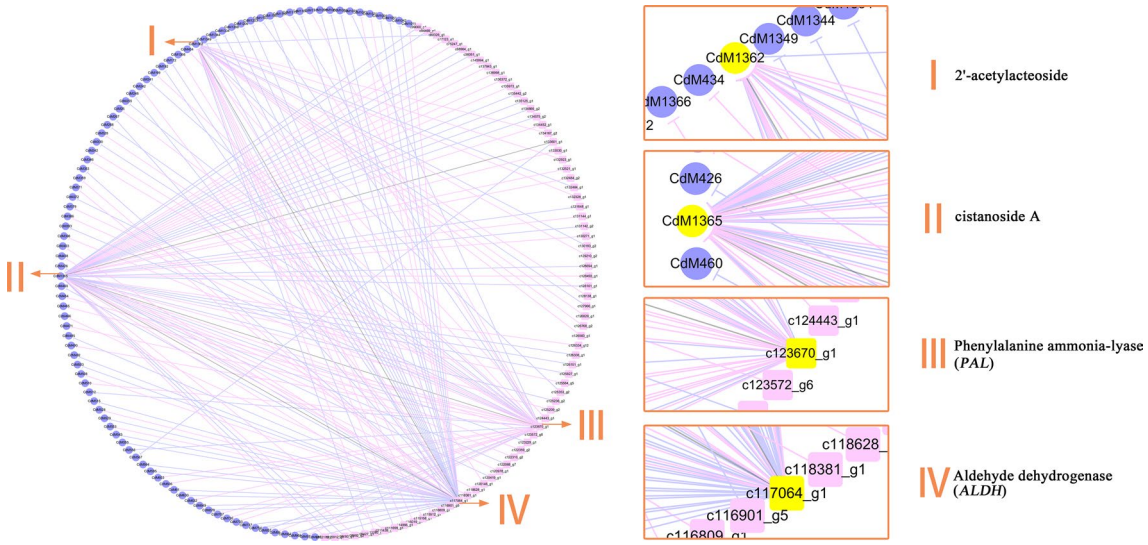

Fig. 8 Connection network of regulatory genes and metabolites between saline and sandy land. Cytoscape software (The Cytoscape Consortium, San Diego, CA, USA, version 3.7.0) was used to visualize the network among the metabolome and transcriptome data in saline–alkali land and sandy land. Solid line represent positive correlation, dotted line represent negative correlation. A: saline–alkali land C: sandy land

Compared saline–alkali land with sandy land, the contents of phenylalanine, p-coumaric acid and ferulic acid were up regulated. On the contrary, the contents of tyrosine and caffeic acid were down regulated.

Venn results (Fig. 2b) inferred the most significant difference between saline–alkali land and sandy land (218 differential metabolites and 947 DEGs). The correlation tests between quantitative changes of PhGs related metabolites and transcripts in the two different ecotypes were carried out. Altogether, 74 transcripts with strong correlation coefficient values (Pearson correlation coefficient >0.8) with 91 metabolites were identified (File S1). The networks showed that the 74 transcripts were grouped into two clusters (I-II), and the 91 metabolites were clustered into two groups (III-IV) (Fig. 7). *PAL* (c123670_g1, cluster III) had a strong positive correlation with the content of *N*-acetyl-5-hydroxytryptamine, apigenin and 4-hydroxy-3,5-diisopropylbenzaldehyde, while having a negative correlation with the content of 3-o-p-coumaroyl shikimic acid o-hexoside and syringic acid o-feruloyl-o-hexoside. *ALDH* (c117064_g1, cluster IV) had a strong positive correlation with the content of *N*-phenylacetylglycine, 3-(4-hydroxyphenyl) propionic acid, etc., while having a negative correlation with the content of cyanidin 3-o-rutinoside (keracyanin), MAG (18:4) isomer1, etc. Cistanoside A (CdM1362, cluster I) had a strong positive correlation with c120146_g1, c132923_g1, c124443_g1, etc. Cistanoside A had a strong negative correlation with c125353_g2, c116901_g5, etc. 2′-acetylacteoside (CdM1365, cluster II) had a strong positive correlation with c137843_g1, c136998_g1, etc. 2′-Acetylacteoside had a strong negative correlation with c122068_g7, c99060_g1, etc. It is worth mentioning that the results of correlation analysis between DEGs and metabolites in grassland and sandy land showed that 2′-acetylacteoside was negatively correlated with c111699_g1 gene and positively correlated with c126768_g2 and c115158_g1 gene.

Discussion

In this study, as an effort to elucidate the quality variation and their underlying regulation in three different ecotypes of *C. deserticola,* growing on saline–alkali land, grassland and sandy land, a correlation analysis of transcriptome and metabolome was performed.

Comparing the quality differences of *C. deserticola* in three ecotypes

To date, various PhGs have been isolated from *C. deserticola*, which are reported to have a strong antioxidative effect (Jiang and Tu 2009). The PhGs in *C. deserticola* mainly include echinacoside, cistanoside A, isoacteoside, acteoside and 2′-acetylacteoside (Yong and Peng-Fei 2009). As their names suggest, PhGs are characterized by a phenethyl alcohol (C_6–C_2) moiety attached to a β-glucopyranose/β-allopyranose via a glycosidic bond. The core structures are often abundantly decorated with substituents such as various saccharides (e.g. allose, arabinose, xylose, lyxose, apiose, glucose, and rhamnose) and aromatic acids (e.g., cinnamic

acid, caffeic acid, isoferulic acid, ferulic acid, and coumaric acid) through ester or glycosidic linkages, respectively (Fu et al. 2008).

In general, among three ecotypes, the contents of five PhGs in saline–alkaline land are the highest, followed by grassland, and the lowest content of *C. deserticola* is in sandy land. The contents of isoacteoside, acteoside and 2′-acetylacteoside are the highest in saline–alkali land, while the contents of echinacoside and cistanoside A are the highest in sandy land. These results are almost consistent with the expression heat map of candidate genes responsible for the biosynthesis of PhGs. On one hand, most of the candidate genes for biosynthesis of PhGs have the highest expression levels in saline–alkaline land, including *PAL, ALDH, GOT,* etc. The up-regulation of homologous genes encoding *PAL, ALDH* and *GOT* genes in saline–alkali land indicate that the accumulation of acteoside in *C. deserticola* might be regulated by the *PAL, ALDH* and GOT mediated pathways. The high expression levels of these genes may be the reason for the high content of acteoside in saline–alkali land of *C. deserticola.* On the other hand, some genes, such as *ADT2*, *DDC*, *CYP988A3* and *SBP*, are highly expressed in the sandy land. This observation is consistent with the high content of echinacoside and cistanoside A in *C. deserticola,* whether or not there are any mechanistic connections remains to be determined in future studies. The pathway annotation diagram shows that PAL and GOT genes are significantly up-regulated in saline–alkali land, which echoes that the content of PhGs in *C. deserticola* in saline–alkali land is significantly higher than that in grassland and sandy land.

In order to further study the underlying regulation of PhGs biosynthesis from different habitats, we selected four metabolic pathways for KEGG annotation, including phenylpropanoid biosynthesis, phenylalanine, tyrosine and tryptophan biosynthesis, tyrosine metabolism and phenylalanine metabolism. Phenylalanine and tyrosine are derived from chorismate, the final product of the shikimate pathway. Chorismate is converted by CM to prephenate, whose subsequent conversion to Phe and Tyr may occur via two alternative pathways (Maeda and Dudareva 2012; Rudolph et al. 2018).

Study on acteoside biosynthesis indicated that the caffeoyl moiety of acteoside is biosynthesized from phenylalanine via a cinnamate pathway, whereas the hydroxytyrosol moiety of acteoside is biosynthesized from tyrosine through dopamine (Ellis 1983). Dopamine is incorporated into acteoside through oxidation to the corresponding aldehyde, reduction to the alcohol, and then β-glycosylation (Saimaru and Orihara 2010). In addition, an intermediate pathway related to the synthesis of acteoside in *Rehmannia glutinosa*, and screened for Aspartate aminotransferase (*GOT*), Primary amine oxidase (*AOC3*), Coumaroylquinate (coumaroylshikimate) 3′-monooxygenase (*C3′H*), Dopa decarboxylase (*DDC*), Amine oxidase copper-containing (*CuAO*) and Aldehyde dehydrogenase (*ALDH*) and other key enzyme genes for the synthesis of acteoside (Zhou et al. 2016).

Explore the molecular mechanism leading to the best quality of saline–alkali land

In saline–alkali land, the expression of *PAL* gene of *C. deserticola* is significantly up-regulated (Figs. 5 and 7). *PAL* is the entrance enzyme in the phenylpropanoid biosynthetic pathway and an important regulation point between primary and secondary metabolism. Moreover, *PAL* is the key enzyme in the biosynthetic pathway of many medicinal compounds such as flavonoids, coumarins, and polyphenols (Junli et al. 2010). Phenylpropanoid compounds have important roles in several different pathways such as protection from abiotic stresses, signal transduction and plant defense against pathogens. Therefore, the evolutionary emergence of the phenylpropanoid pathway in plants is an important adaptation that enables plant defense against abiotic and biotic stresses (Kim and Hwang 2014). *PAL* is of great significance in plant cell differentiation and woody processes, resistance to pests and diseases, and plant stress resistance. Induction of adventitious root growth of *Morinda citrifolia* by medium salt intensity showed positive correlation between *PAL* activity and salt strength (Md et al. 2010). The selective expression of *PAL* gene is controlled by a complex regulatory network for the development and environmental control of phenylpropane biosynthesis (Liang et al. 1989). Salt stress treatment can induce the expression of *PAL* gene in *Salvia* and subsequently increase *PAL* activity and total phenolic accumulation in the early hours after stress treatment. Promoter analysis demonstrated that *PAL* includes *cis*-acting elements such as ABRE, ARE, HSE, TCA-elements, TC-rich repeat and GA, CGTCA, TGACG, WUN motif in the upstream region which respond to stress hormones and might show its vital role in biotic and abiotic stresses (Valifard et al. 2015). Using different concentrations of neutral salts (NaCl and Na_2SO_4) and alkaline salts ($NaHCO_3$, Na_2CO_3) to simulate the treatment of wheat with salt-alkali stress, the *PAL* activity of leaves increased significantly (Wei and Yu 2012). *PAL* enzyme activity is enhanced after salt treatment of medicinal orchid (Nag and Kumaria 2018).

The expression level of *ALDH* gene in *C. deserticola* is also significantly up-regulated in saline–alkali land. Aldehyde dehydrogenases (ALDH) are enzymes which oxidize a wide variety of aliphatic and aromatic aldehydes. Previous studies have showed that the gene expression of *ALDH* was increased under salt injury and water shortage (Chen and Murata 2002; Li et al. 2013). *ALDH* plays an important role in the salt and alkali resistance of grapes. The *ALDH* gene promoter region of the grape contains the stress response element TC-rich repeats, which may be related

to the stress response. Reactive oxygen species (ROS) in plants are by-products of photosynthesis and respiration. However, under high stress, drought, cold damage, heavy metals, mechanical damage, nutrient deficiencies, pathogens and other stress conditions, plants will produce excessive ROS, which will cause damage to plant cells (Bartels and Sunkar 2005). In the process of evolution, plants produce a variety of highly efficient ROS scavenging mechanisms, among which the enhancement of aldehyde dehydrogenase activity to resist ROS toxicity is one of them (Gomes et al. 2006). Glutamic–aspartic transaminase (GOT) catalyzes the reversible transfer of the amino group from L-aspartate to 2-oxoglutarate to form oxaloacetate and L-glutamate. In eukaryotes, there are two AAT isozymes: one is located in the mitochondrial matrix, the second is cytoplasmic. 4-coumarate–CoA ligase (4CL), a plant enzyme that catalyzes the formation of 4-coumarate-CoA from 4-coumarate and coenzyme A. The branchpoint reactions between general phenylpropanoid metabolism and pathways lead to various specific end products. Trans-cinnamate 4-monooxygenase (C4H) is a cytochrome P450 (heme-thiolate) protein found in plants. 5-O-(4-coumaroyl)-D-quinate 3′-monooxygenase (CYP98A3) is a cytochrome P450 (heme-thiolate) protein, found in plants. It also acts on trans-5-O-(4-coumaroyl) shikimate.

In summary, saline–alkali stress induces up-regulation of *PAL* and *ALDH* genes expression of *C. deserticola*, increases *PAL* and *ALDH* enzymes activity, and leads to enhance content of phenylpropane compounds, thus resisting external stress stimulation. *C. deserticola* from different habitats has formed a relatively stable ecological type under the special natural ecological environment. Different ecotypes show differences in gene expression, biochemical and other phenotypes, and such differences are the essence of quality variation. *C. deserticola* from different habitats have formed relatively stable ecological types under the special natural ecological environment. Different ecotypes show significant change in gene expression, biochemical and such differences are the essence of quality variation (Huang and Juan 2012).

Correlation network analysis mining important genes and chemical markers

Based on the results on the correlated compound production and gene expression, we identified 12 key genes related to PhGs biosynthesis, *PAL*, *ALDH* and *GOT* genes as responsible for the over-prodcution of PhDs. The expression profiles of these two genes are highly correlated with the abundance profiles of 38 and 53 metabolites, respectively. *PAL* and *ALDH* are important genes for plants to resist the external abiotic stress. Their expressions are affected by environmental factors. In particular, 2′-acetylacteoside is the most significant with a VIP > 0.5 and fold change > 2 (Table S7). Meanwhile, in the saline–alkali land vs sandy land and saline–alkali land vs grassland, 2′-acetylacteoside is differentially clustered into a clusters, suggesting that 2′-acetylacteoside is a major contributors and might be used as an indicator to distinguish saline–alkali land, grassland and sandy land of *C. deserticola*, consistent with the previous results (Huang and Juan 2012). As Ecologist Sharapov (Huang and Juan 2012) pointed out: "The evolution of the plant kingdom is in the direction of adapting to the lack of water, and the plant's drought and biochemistry have a certain reaction in the development of certain special substances that prevent plants from losing water." The ecological type of *C. deserticola* is closely related to the ecological environment driving its arid climate and geography. Unique interrelationship and genetic differentiation between the metabolites of *C. deserticola* is an ecological strategy and ecological evolution of *C. deserticola* in extreme arid environment.

Currently, studies have shown that *PAL* and *ALDH* genes are affected by environmental stress, but it is unknown whether *GOT* genes are affected by the environment stimulus. Therefore, we plan to focus on the response relationship between *GOT* gene and environment. At the same time, the mechanism of *PAL* and *ALDH* affected by environment in *C. deserticola* will be further explored to provide a better theoretical basis for the breeding of *C. deserticola* varieties and the preservation of excellent genes.

Conclusion

Through combined metablomic and transcriptomic analysis, we found that (1) the ecotype of *C. deserticola* from the saline–alkali land has the best quality demonstrated by its highest contents of bioactive compounds, PhGs; (2) 2′-acetylacteoside can be used as an marker to distinguish the three ecotypes of *C. deserticola*; (3) the high levels of PhGs might result from the up-regulated expression of *PAL*, *ALDH*, *GOT* and other key genes related to the biosynthesis of PhGs in response to the saline–alkali environmental factors.

Acknowledgements This work was supported by National Natural Science Foundation of China (No. 81473315), National Science & Technology Fundamental Resources Investigation Program of China (2018FY100701), the Open Research Fund of Chengdu University of Traditional Chinese Medicine Key Laboratory of Systematic Research of Distinctive Chinese Medicine Resources in Southwest China (003109034001) and CAMS Innovation Fund for Medical Sciences (CIFMS) No. 2016-I2M-3-015, which are gratefully acknowledged.

Author's contribution Xiao Sun: conceptualization, data curation, formal analysis, methodology, software, validation, visualization, writing—original draft, writing—review and editing; Lin Li: conceptualization, data curation, formal analysis, visualization, writing—original draft; Jin Pei: funding acquisition, resources, supervision; Chang Liu: software, supervision, writing—review and editing; Linfang Huang:

conceptualization, data curation, funding acquisition, project administration, writing—review and editing.

References

Abdelrahman M, Hirata S, Sawada Y, Hirai MY, Sato S, Hirakawa H, Mine Y, Tanaka K, Shigyo M (2019) Widely targeted metabolome and transcriptome landscapes of *Allium fistulosum–A. cepa* chromosome addition lines revealed a flavonoid hot spot on chromosome 5A. Sci Rep 9:3541

Audic S, Claverie JM (1997) The significance of digital gene expression profiles. Genome Res 7:986–995

Bartels D, Sunkar R (2005) Drought and salt tolerance in plants. Crit Rev Plant Sci 24:23–58

Bo L, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinform 12:323

Cavill R, Jennen D, Kleinjans J, Briede JJ (2016) Transcriptomic and metabolomic data integration. Brief Bioinform 17:891–901

Chen TH, Murata N (2002) Enhancement of tolerance of abiotic stress by metabolic engineering of betaines and other compatible solutes. Curr Opin Plant Biol 5:250–257

Chen Q, Li M, Wang C, Li Z, Xu J, Zheng Q, Liu P, Zhou H (2018) Combining targeted metabolites analysis and transcriptomics to reveal chemical composition difference and underlying transcriptional regulation in Maca (*Lepidium Meyenii* Walp.) ecotypes. Genes (Basel). https://doi.org/10.3390/genes9070335

Cho K, Cho KS, Sohn HB, Ha IJ, Hong SY, Lee H, Kim YM, Nam MH (2016) Network analysis of the metabolome and transcriptome reveals novel regulation of potato pigmentation. J Exp Bot 67:1519

Cole T, Williams BA, Geo P, Ali M, Gordon K, Baren MJ, Van Salzberg SL, Wold BJ, Lior P (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511–515

Cui GH, Min C, Huang LQ, Xiao SP, Da LI (2004) Study on genetic diversity of Herba Cistanches by RAPD. Zhongguo Zhong Yao Za Zhi 29:727–730

Ellis BE (1983) Production of hydroxyphenylethanol glycosides in suspension cultures of *Syringa vulgaris*. Phytochemistry 22:1941–1943

Fu G, Pang H, Wong YH (2008) Naturally occurring phenylethanoid glycosides: potential leads for new therapeutics. Curr Med Chem 15:2592–2613

Gomes APS, Baracat-Pereira MC, Andrade MO, Rodrigues SM, Fontes EPB, DaMatta FM (2006) Arabidopsis and tobacco plants ectopically expressing the soybean antiquitin-like ALDH7 gene display enhanced tolerance to drought, salinity, and oxidative stress. J Exp Bot 57:1909–1918

Grabherr MG, Haas BJ, Moran Y, Levin JZ, Thompson DA, Ido A, Xian A, Lin F, Raktima R, Qiandong Z (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644

Gu L, Xiong WT, Wang C, Sun HX, Li GF, Liu X (2013) Cistanche deserticola decoction alleviates the testicular toxicity induced by hydroxyurea in male mice. Asian J Androl 15:838–840

Hao R, Du X, Yang C, Deng Y, Zheng Z, Wang Q (2019) Integrated application of transcriptomics and metabolomics provides insights into unsynchronized growth in pearl oyster *Pinctada fucata* martensii. Sci Total Environ 666:46–56

Huang LF, Juan FU (2012) Academic study on ecological variation of traditional Chinese medicinal materials. Chin Tradit Herbal Drugs 43:1249–1258

Jiang Y, Tu PF (2009) Analysis of chemical constituents in Cistanche species. J Chromatogr A 1216:1970–1979

Junli H, Min G, Zhibing L, Baofang F, Kai S, Yan-Hong Z, Jing-Quan Y, Zhixiang C (2010) Functional analysis of the Arabidopsis PAL gene family in plant growth, development, and response to environmental stress. Plant Physiol 153:1526–1538

Kim DS, Hwang BK (2014) An important role of the pepper phenylalanine ammonia-lyase gene (PAL1) in salicylic acid-dependent signalling of the defence response to microbial pathogens. J Exp Bot 65:2295

Li X, Guo R, Li J, Singer SD, Zhang Y, Yin X, Zheng Y, Fan C, Wang X (2013) Genome-wide identification and analysis of the aldehyde dehydrogenase (ALDH) gene superfamily in apple (Malus × domestica Borkh). Plant Physiol Biochem 71:268–282

Li Z, Lin H, Gu L, Gao J, Tzeng CM (2016) Herba Cistanche (Rou Cong-Rong): one of the best pharmaceutical gifts of traditional Chinese medicine. Front Pharmacol 7:41

Liang XW, Dron M, Cramer CL, Dixon RA, Lamb CJ (1989) Differential regulation of phenylalanine ammonia-lyase genes during plant development and by environmental cues. J Biol Chem 264:14486–14492

Liu W, Song Q, Cao Y, Xie N, Li Z, Jiang Y, Zheng J, Tu P, Song Y, Li J (2019) From (1)H NMR-based non-targeted to LC–MS-based targeted metabolomics strategy for in-depth chemome comparisons among four Cistanche species. J Pharm Biomed Anal 162:16–27

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Delta Delta C(T)) method. Methods 25:402–408

Lou Q, Liu Y, Qi Y, Jiao S, Tian F, Jiang L, Wang Y (2014) Transcriptome sequencing and metabolite analysis reveals the role of delphinidin metabolism in flower colour in grape hyacinth. J Exp Bot 65:3157–3164

Maeda H, Dudareva N (2012) The shikimate pathway and aromatic amino acid biosynthesis in plants. Annu Rev Plant Biol 63:73–105

Md AB, Eun-Jung L, Kee-Yoeup P (2010) Medium salt strength induced changes in growth, physiology and secondary metabolite content in adventitious roots of *Morinda citrifolia*: the role of antioxidant enzymes and phenylalanine ammonia lyase. Plant Cell Rep 29:685–694

Nag S, Kumaria S (2018) In silico characterization and transcriptional modulation of phenylalanine ammonia lyase (PAL) by abiotic stresses in the medicinal orchid *Vanda coerulea* Griff. ex Lindl. Phytochemistry 156:176–183

Rong XU, Sun SQ, Liu YG, Chen J, Chen SL, Zhou F (2009) Analysis and evaluation of different radial part of *Cistanche deserticola* by Fourier transform infrared spectroscopy and two-dimensional infrared correlation spectroscopy. Chin J Anal Chem 37:221–226

Rudolph S, Riedel E, Henle T (2018) Studies on the interaction of the aromatic amino acids tryptophan, tyrosine and phenylalanine as well as tryptophan-containing dipeptides with cyclodextrins. Eur Food Res Technol 244:1–9

Saimaru H, Orihara Y (2010) Biosynthesis of acteoside in cultured cells of Olea europaea. J Nat Med 64:139–145

Valifard M, Mohsenzadeh S, Niazi A, Moghadam A (2015) Phenylalanine ammonia lyase isolation and functional analysis of phenylpropanoid pathway under salinity stress in Salvia species. Aust J Crop Sci 9:656–665

Wang T, Zhang X, Xie W (2012) *Cistanche deserticola* Y. C. Ma, "Desert Ginseng": a review. Am J Chin Med 40:1123–1141

Wang X, Wang J, Guan H, Xu R, Luo X, Su M, Chang X, Tan W, Chen J, Shi Y (2017) Comparison of the chemical profiles and antioxidant activities of different parts of cultivated *Cistanche deserticola* using ultra performance liquid chromatography-quadrupole time-of-flight mass spectrometry and a 1,1-diphenyl-2-picrylhydrazyl-based assay. Molecules 22:2011

Wei G, Yu L (2012) Effects of salinity–alkalinity stress on root activity and phenylalanine ammonia-lyase activity of wheat seedlings. Crops 1:31–34

Xu X, Zhang Z, Wang W, Yao H, Ma X (2017) Therapeutic effect of cistanoside A on bone metabolism of ovariectomized mice. Molecules. https://doi.org/10.3390/molecules22020197

Zhang Y, Li D, Zhou R, Wang X, Dossa K, Wang L, Zhang Y, Yu J, Gong H, Zhang X, You J (2019) Transcriptome and metabolome analyses of two contrasting sesame genotypes reveal the crucial biological pathways involved in rapid adaptive response to salt stress. BMC Plant Biol 19:66

Zheng S, Jiang X, Wu L, Wang Z, Huang L (2014) Chemical and genetic discrimination of Cistanches Herba based on UPLC-QTOF/MS and DNA Barcoding. PLoS ONE 9:e98061

Zhou Y, Wang X, Wang W, Duan H (2016) De novo transcriptome sequencing-based discovery and expression analyses of verbascoside biosynthesis-associated genes in *Rehmannia glutinosa* tuberous roots. Mol Breed 36:139