Evidence for Regulation of Cordycepin Biosynthesis by Transcription Factors Krüppel-Like Factor 4 and Retinoid X Receptor Alpha in Caterpillar Medicinal Mushroom *Cordyceps militaris* (Ascomycetes)

by Zhang, H., Deng, L., Luo, S., Liu, L., Yang, G., Zhang, Y., Gao, B., Yang, D., Wang, X., Li, S., Li, X., Jiang, Y., Lao, W. and Vriesekoop, F.

Zhang, H., Deng, L., Luo, S., Liu, L., Yang, G., Zhang, Y., Gao, B., Yang, D., Wang, X., Li, S. and Li, X., Jiang, Y., Lao, W. and Vriesekoop, F. (2024) 'Evidence for Regulation of Cordycepin Biosynthesis by Transcription Factors Kräppel-Like Factor 4 and Retinoid X Receptor Alpha in Caterpillar Medicinal Mushroom Cordyceps militaris (Ascomycetes)', *International Journal of Medicinal Mushrooms*, *26* (10), pp. 19-40. Available at: https://doi.org/10.1615/IntJMedMushrooms.2024054952

Evidence for Regulation of Cordycepin Biosynthesis by Transcription Factors Krüppel-like factor 4 and Retinoid X receptor alpha in *Cordyceps militaris*

Hucheng Zhang[1], Lina Deng[2], Shuai Luo[1], Linying Liu[1], Guowei Yang[1], Yuning Zhang[1], Bo Gao[1], Dongqing Yang[1], Xiaojie Wang[1], Shuangshi Li[1], Xingjuan Li[1], Yaguang Jiang[1], Wenyan Lao[3]* & Frank Vriesekoop[4]*

[1] Bioengineering College, Beijing Polytechnic, Yizhuang Economic and Technological Development Zone, Daxing District, 100176 Beijing, People's Republic of China;
[2] Beijing Health Vocational College, Dadou Rd. 12, Fangshan District, 102402 Beijing, China;
[3]School of Biochemical Engineering of Beijing Union University, Beijing 100023, China
[4]Department of Food Science, Harper Adams University, United Kingdom.

*Corresponding author: Wenyan Lao, School of Biochemical Engineering of Beijing Union University, Beijing 100023, China, Address: Beijing Union University, No. 18, District 3, Fatouxili, Chaoyang District, Beijing, China. E-mail: laowenyan@buu.edu.cn, Fax: 86-010-52072174; Dr Frank Vriesekoop, Department of Food Science, Harper Adams University, Newport, TF10 8NB, United Kingdom; Tel: +44-19528115241, Fax: +44-1952814783, FVriesekoop@harper-adams.ac.uk

Other authors' information:
Hucheng Zhang, zhanghchbj@163.com
Shuai Luo, E-mail: luoshuai1985@sina.com
Linying Liu, E-mail: llyspo@163.com
Guowei Yang, E-mail: yguow@163.com
Yuning Zhang, E-mail: 14730201@qq.com
Bo Gao, E-mail: gb34929786@126.com
Dongqing Yang, E-mail: 13401100132@126.com
Xiaojie Wang, E-mail: 100169@bpi.edu.cn
Shuangshi Li, E-mail: lishuangshi.1980@163.com
Xingjuan Li, E-mail: lixingjuan@bpi.edu.cn
Yaguang Jiang, E-mail: jyg928@163.com

SHORT TITLE: Transcription Factors Klf4 and Rxra Involved in Cordycepin Biosynthesis

ABBREVIATIONS: **AMP**, Adenosine monophosphate; **ATP**, Adenosine Triphosphate; **cAMP**, Cyclic adenosine monophosphate; **CCM**, Cultured *Cordyceps militaris*; ***Cns***, key gene in cordycepin biosynthesis or transformation; **COG**, Clusters of orthologous groups of proteins; **DEG**, differentially expressed genes; **eggNOG**, evolutionary genealogy of genes: non-

supervised orthologous groups; **FDR**, false discovery rate; **FPKM**, Fragments Per Kilobase of exon model per Million fragments mapped; **GO**, gene ontology; **HPLC**, High Performance Liquid Chromatography; **KEGG**, Kyoto Encyclopedia of Genes and Genomes; **KOG**, Clusters of orthologous groups for eukaryotic complete genomes; **NR**, National center for biotechnology information reference sequence; **PDA**, Potato Dextrose Agar; **Pfam**, the protein family database, a large collection of protein families, each represented by multiple sequence alignments and hidden Markov models (HMMs); **qRT-PCR**, Quantitative real-time polymerase chain reaction; **RNA**, Ribonucleic Acid; **RSEM**, RNA-Seq by Expectation Maximization; **STEM**, short time-series expression miner; **Swiss-prot**, Protein sequence database; **TFs**, transcription factors.

ABSTRACT:

Cordyceps militaris, Chinese traditional medicinal fungus, has many bioactive properties. Cordycepin (3'-deoxyadenosine) is a major bioactive component of *C. militaris*. Various methods can significantly elevate cordycepin production, which suggests a diverse set of metabolic regulatory mechanisms. Thus, we aimed to identify transcription factors that regulate cordycepin biosynthesis pathways. Transcriptome analysis of wild type *C. militaris*, *C. militaris* GYS60, a cordycepin high-producing strain, and *C. militaris* GYS80, a low-producing strain, were used to measure expression and function of genes related to cordycepin biosynthesis. The transcriptome expression data were confirmed by qRT-PCR. We identified 155 relevant transcription factors in 19 families that included Fork head/winged helix factors, other C_4 zinc finger-type factors, C_2H_2 zinc finger factors, tryptophan cluster factors, nuclear receptors with C_4 zinc fingers, homeodomain factors, and Rel homology region factors. Energy generation and amino acid conversion pathways were activated in GYS60 so that abundance of cordycepin precursors was increased. Genes and transcription factors for rate-limiting enzymes in these pathways were identified. Overexpression of two key transcription factors, Krüppel-like factor 4 (Klf4) and Retinoid X receptor alpha (Rxra), promoted high cordycepin production in GYS60. In GYS60, Klf4 and Rxra were responsible for upregulation of genes in cordycepin biosynthesis, namely an oxidoreductase, 3',5'-cyclic AMP phosphodiesterase, a transferase, and adenylate cyclase. Upregulation of these genes increased 3'-AMP content, thereby elevating cordycepin synthesis.

Key words: cordycepin; transcription factor; *Cordyceps militaris*; metabolic network; biosynthesis

I. INTRODUCTION

For centuries, *Cordyceps militaris*, an entomopathogenic fungus, has been used as food and nutrition tonics worldwide, especially in eastern Asia.[1] *C. militaris* has many bioactive compounds, such as cordycepin,[2] adenosine, ergosterol, ergothioneine,[3] and cordyceps polysaccharides.[4] Thus, *C. militaris* has been extensively exploited in food, pharmaceutical, and cosmetics industries.[5-7] Among the biologically active ingredients, cordycepin has attracted the most attention for its multiple of therapeutic effects,[8] such as inhibition of cancer cell proliferation and induction of cell apoptosis,[9-12] and for its antibacterial,[13] antifungal,[14] antivirus,[15] antioxidant,[16] and immune system regulation properties.[17]

Cordycepin, 3'-deoxyadenosine, is the first adenosine analog isolated from *C. militaris*.[2] Various methods have been investigated to increase production of this medicinal agent. These methods include chemical synthesis,[18] varying carbon and nitrogen sources,[19, 20] enhancing production of putative precursors,[21] growth in the presence of mineral salts,[22] liquid fermentation,[23,24] extraction from fruiting bodies,[25] and genetics.[26] Yields have improved from 380 mg/L to 6,840 mg/L. We reported using a multifunctional plasma mutagenesis system to create a high cordycepin-producing strain, GYS60, whose cordycepin yield was 7,883 mg/L; by the same method, we created a low cordycepin-producing strain GYS80.[27] However, we did not establish detailed genetic mechanisms for the high- and low-producing phenotypes of these strains.

By focusing transcriptome and proteome analyses on the purine metabolism pathway, investigators have deduced the cordycepin biosynthetic pathway in *C. militaris*.[28-32] Cordycepin is synthesized from adenosine by a three-gene cluster.[33] *Cns*3 (CCM_04438) is the

first gene, a nucleotide kinase that contains two domains, an N-terminal (9-101 aa) nucleoside/nucleotide kinase and a C-terminal (681-851 aa) HisG family ATP phosphoribosyltransferase. *Cns*2 (CCM_04437) is the second gene, a phosphohydrolase member of the putative metal-dependent HDc family of phosphohydrolases. *Cns*1 (CCM_04436), the third gene, encodes an oxidoreductase of the GFO/IDH/MOCA family of oxidoreductases or the MviM family of dehydrogenases. The *Cns* gene cluster revealed that cordycepin metabolism overlaps with parts of the purine synthetic pathway.[33] Adenosine is phosphorylated to 3′-AMP by nucleotide kinase (Cns3), then dephosphorylated to 2′-C-3′-dA by phosphohydrolase (Cns2), and cordycepin is finally produced by oxidoreductase (Cns1). The cordycepin is transported out of the cell by Cns4, a member of the putative pleiotropic drug resistance family of ATP-binding cassette (ABC) transporters.

Two key genes, *Cns*1 and *Cns*2, were cloned in *Saccharomyces cerevisiae* to produce 137.27 mg/L cordycepin,[34] a yield that was lower than the 420.50 mg/L naturally produced by *C. militaris*.[20] Thus, the *S. cerevisiae* study indicated that there are other complementary pathways for cordycepin synthesis in *C. militaris*. However, to the best of our knowledge, research was still lacking about other cordycepin synthesis pathways and associated transcription factors.

Krüppel-like factor 4 (Klf4) and retinoid X receptor alpha (Rxrα) are two key transcription factors that are expressed and activated when cells incur mild DNA damage.[35] In *C. militaris*, Klf4 (CCM_08167) is annotated as high-affinity glucose transporter of the C_2H_2 zinc finger transcription factor family. Rxrα (CCM_07505) is annotated as DNA-binding protein eta2 in the family of nuclear receptors with C4 zinc fingers.

To better understand the molecular mechanism of cordycepin production, we performed transcriptome analysis of strains GYS60, GYS80, and wild type *C. militaris*. In GYS60, we found that Klf4 and Rxra were overexpressed, and we verified that they were responsible for upregulation of an oxidoreductase (CCM_07507), 3',5'-cyclic AMP phosphodiesterase (CCM_02777), a transferase (CCM_04722), and adenylate cyclase (CCM_06928) involved in cordycepin production.

II. MATERIALS AND METHODS

A. Chemicals and kits

Authentic cordycepin was purchased from Sigma-Aldrich (USA). Chromatographically pure methanol was from Merck (Darmstadt, Germany). Methanol and acetic acid HPLC grade were purchased from Tianjin Shield Company (Tianjin, China). All other chemicals (analytical grade) were purchased from Beijing Chemical Reagents Company (Beijing, China). RNA Nano 6000 kits were obtained from Agilent, and mRNA extraction kit DP411 and SYBR Green Super Real PreMixPlus kit and DP204-03 PCR Purification kit were from Tiangen Biochemical Technology Co., Ltd (Beijing, China). VAHTS Universal V6 RNA-seq Library Prep Kit NR604-02 for Illumina was from Nanjing Vazyme Biotech Co., Ltd, China. Other reagents were purchased from Sinopharm Chemical Reagent Co., Ltd, China.

B. Culture and fermentation of *C. militaris*

Wild-type *C. militaris* (WT) was isolated from fruiting bodies obtained from a supermarket in Beijing, China. High cordycepin-producing strain *C. militaris* GYS60 and low cordycepin-producing strain *C. militaris* GYS80 were obtained by a multifunctional plasma

mutagenesis procedure as we described.[27] All *C. militaris* strains were incubated on potato dextrose agar medium at 25 ◦C for 7 to 14 days until the mycelium almost filled the Petri dish. Potato dextrose liquid medium (modified as 1000 mL of 20.0% potato extract liquid + 2.0% peptone + 0.3% yeast extract powder + 4.0% glucose + 0.2% $K_2HPO_4•3H_2O$+ 0.02% $MgSO_4•7H_2O$) was used as fermentation medium. Six 1-cm diameter mycelium plugs were transferred into 250 mL flasks containing 70 mL liquid medium and shaken at 150 rpm for 8 days to expand the culture. Fermentation was performed in three 1-liter flasks. Thirty milliliter WT, GYS60, and GYS80 mycelium pellets were mixed with 270 ml of liquid medium in each flask to build the working samples. The flasks were shaken at 150 rpm for 30 days. From the first day, 2.0 mL fermentation liquid was taken every two days to measure the dry mycelium and for HPLC analysis of cordycepin in the fermentation medium and mycelium. A part of the mycelium pellets from the three samples was collected, frozen, and dried for transcriptome sequencing.

C. HPLC analyses of cordycepin

High performance liquid chromatography detection, standard curve preparation, and content calculation methods were performed as described.[23] The mycelia pellets from different sampling times were collected, weighed, lyophilized, and ground to powder. For the HPLC assays, the mobile phase was water (A) and methanol (B), and gradient elution was used. The elution procedure was 0-5 min, 5% B; 5-10 min, 5-15% B; 10-20 min, 15% B; 20-25 min, 15-100% B; 25-30 min, 100% B. The flow rate was 1.0 mL/min, and the sample volume was 10 μL. The detection wavelength was 260 nm. The cordycepin yield was calculated using the detected peak area according to the standard curve. The cordycepin

concentrations in mycelia and fermentation broth were calculated by normalizing to equal biomass by volume.

D. RNA extraction

After 15 days of culture, mycelia were collected at the late logarithmic growth stage and immediately frozen with liquid nitrogen. Total RNA of the mycelia was isolated using the RNA extraction kit DP411 according to the manufacturer's protocol (Tiangen, China). The RNA samples were assessed for quantity and quality with a NanoDrop ND-2000 spectrophotometer (Thermo Fisher Scientific, USA) and 1.0% ethidium bromide-stained agarose gels, respectively . RNA integrity was confirmed with an Agilent 2100 Bioanalyzer (Agilent, USA). Three biological replicates were obtained for each of the three strains.

E. cDNA synthesis, library preparation, and sequencing

After treating with RNase-free DNase I for 30 min at 37 °C to remove residual DNA, total mRNA was enriched with magnetic beads containing oligo(dT) followed by adding fragmentation buffer to randomly fragment mRNA. The first-strand cDNA synthesis was performed with 2.0 μg mRNA using the PrimeScriptTM First-Strand cDNA Synthesis Kit (Vazyme, China). Subsequently, double-stranded cDNA was synthesized using the SuperScript Double-Stranded cDNA Synthesis kit (Vazyme, China) and purified with the Pure-Link™ PCR purification kit (Invitrogen, USA).

According to the VAHTS Universal V6 RNA-seq Library Prep Kit protocol (Illumina, USA), 500 ng purified double stranded cDNA was repaired, followed by adding an A-tail and connecting to a sequencing adapter. cDNA of approximately 250-300 bp was screened using AMPure XP beads for PCR amplification, and the PCR product was purified using AMPure

XP beads to obtain the library. The DNA molecules in the library were amplified in 17 cycles (10 s at 98 °C, 30 s at 60 °C, and 30s at 72 °C) using the VAHTS RNA Adapters Set 1 - Set 2 (Illumina, USA). After amplification, the DNA library was analyzed by Qubit 2.0 for concentration and Agilent 2100 for insert size. Sequencing was performed by NovaSeq 6000 (Illumina, USA).

F. Data alignment and screening of differentially expressed genes

The raw data obtained by Illumina NovaSeq 6000 sequencing were used for quality analysis and filtration to obtain high-quality data (clean data), stored in FASTQ format. Trinity software was used to splice and merge the valid data. After the genes were reconstructed, DIAMOND software was used to compare the reconstructed transcripts with COG, GO, KEGG, KOG, Pfam, Swiss-Prot and NR databases. The similarity between three samples was greater than 30% and expectation value (E value) was less than 10^{-5}. Transcriptome information was analyzed with BMKCloud (www.biocloud.net) to predict the amino acid sequences of new genes. HMMER software was used to annotate new sequences according to the Pfam database. For gene expression analysis, clean reads were mapped to the complete reference to estimate gene and isoform expression levels from RNA-Seq data, to obtain the gene expression level for each sample. The gene expression level was determined by calculating the number of fragments in each sample normalized to FPKM (fragments per kilobase of exon model per million mapped reads). The level of transcriptome gene differential expression was expressed by $\log_2$ [fold change].

G. Enrichment analysis of differentially expressed genes

We obtained information about the functions of differentially expressed genes (DEGs)

using edgeR criteria according to the RSEM results. The detection of DEGs between three treatments (GYS60 *vs*. WT, GYS80 *vs*. WT, and GYS60 *vs*. GYS80) was based on the noisy distribution method (NOIseq). The corrected P > 0.8 and absolute value of $\log_2$ [fold change] >2.0 were considered as DEGs. Detection of DEGs between every single sample (GYS60 *vs*. WT, GYS80 *vs*. WT, and GYS60 *vs*. GYS80) was based on the Poisson distribution method (PoissonDis.) and the false discovery rate < 0.001; the absolute value of $\log_2$ [fold change] >2 was considered as DEGs. To further analyze information about the functions of DEGs, Gene Ontology (GO) functional annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed.

H. Trend analysis

To identify expression patterns of DEGs, trend analysis was used to cluster genes with similar expression patterns.[36] Differentially expressed genes were clustered by the short time-series expression miner (STEM) method (https://cloud.oebiotech.cn/task/detail/stem/), and the statistically significant (P≤0.05) profile boxes were colored.

I. Transcription factor analysis

To obtain information on the regulation of DEGs by transcription factors, the predicted gene sequences were aligned to the corresponding transcription factor database by using BMKCloud (www.biocloud.net).[37] The transcription factors were assigned to known factor families.

J. Quantitative real-time PCR

The *C. militaris* strains were cultured as described to the 15th day of the late logarithmic phase, then centrifuged at 10,000 x g at 4 ℃ for 5 min. Total mRNA was extracted from the

mycelia according to the instructions of the mRNA extraction kit (DP411). The concentration and purity of the mRNA were measured with the NanoDrop spectrophotometer and Fluorometer Qubit 4.0. The qRT-PCR reactions were performed on the CFX96 Real-Time PCR Detection System (Bio-Rad, CA, USA), according to the reaction parameters provided. Each reaction was performed in triplicate with a reaction mixture of 20 μL, containing 100 ng of total mRNA, 160 nmol/L primers, 10 μL 2× SYBR Green Super Real PreMixPlus (Tiangen, Beijing, China), and 0.6 μL iScript Reverse Transcriptase. The cycling conditions were 95°C for 4 min followed by 45 cycles of 95°C for 10 s, 57°C for 15 s and 72°C for 20 s. The cycle was first detected as statistically significant above background at a significant increase in fluorescence. Threshold cycle (C_t) was calculated automatically using the Bio-Rad iCycler iQ5 software V2.0. Negative controls without template and reverse transcriptase were included on each plate to ensure the absence of random and genomic DNA contamination. The standard curve method, which was performed by ten-fold dilution series, was used to determine the efficiency of every PCR reaction. Real-time PCR Miner identified and fitted the exponential phase of the curve using a four parameter logistic model, which fitted the raw fluorescence data as a function of PCR cycles, and a three-parameter simple exponent model, which used an interactive nonlinear regression algorithm. Candidate gene regression values were obtained using P-values from the regression data and an efficiency value was calculated from a weighted average. All primers were designed at https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome (see Table 1). The 18S rDNA was used as an internal control. The relative gene expression level was calculated by the $2^{-\triangle\triangle CT}$ method using the formula $\triangle\triangle Ct = (Ct_{target} - Ct_{rDNA})_{\text{interested sample}} - (Ct_{target} - Ct_{rDNA})_{\text{calibrated sample}}$. We selected the

expression of each gene on the first day of inoculation as the calibrated sample, and the 15th day as the interest sample.

III. RESULTS

A. Growth characteristics and cordycepin production of *C. militaris* strains

Strain GYS60 appeared as white fluff with an abundance of filamentous mycelia surrounding the colony on potato dextrose agar; strain GYS80 had more wrinkles than the wild type, which we reported earlier.[27] After 30 days of liquid culturing of WT, GYS60, and GYS80, the color of residual medium and mycelia did not show any significant change. This observation indicated the absence of large-scale cell autolysis and that the amount of carbon and nitrogen in the medium was enough for the long period of mycelial growth. The *C. militaris* strains showed similar growth characteristics and maximum specific growth rate (μ_{max}) during liquid fermentation (Table 2). The maximum biomass was reached at day 15 and remained unchanged thereafter. The cordycepin yield measured by HPLC of dry mycelia and extracellular liquid was calculated from the standard curve formula y=62.047x+6.5101 (R^2=0.9998). The three strains began to produce cordycepin on the 10th day; the maximum value was reached on the 15th day and remained stable until decreasing on the 19th day.

B. Characteristics of mRNA libraries

We constructed and sequenced nine mRNA libraries from the three *C. militaris* strains. After removing ambiguous "N" nucleotides (with a ratio of "N">10%) and low-quality sequences (quality score Q≤10 accounting for more than 50% of the entire read), we obtained approximately 68.03 Gb clean data. There were 27.19, 23.51, and 25.36 Mb average clean reads per sample in the WT group, GYS60 group, and GYS80 group, respectively (Table 3).

More than 97% of the clean bases had quality scores of Q20, and more than 94% of the clean bases had quality scores of Q30. In addition, a total average ratio of 94.36% reads was mapped successfully, of which an average of 93.87% was obtained for uniquely mapped reads. These values indicated high accuracy and high assembly integrity of the transcriptome data (Table 3) and that the data met the requirements for subsequent analysis.

C. Gene assembly and annotation

The clean reads of each sample were compared with the *C. militaris* reference genome (www.ncbi.nlm.nih.gov/home/?Term=cordyceps+militaris). We assembled 11,211 genes and then annotated genes based on the COG, GO, KEGG, KOG, Pfam, Swiss-Prot, eggNOG and NR databases (Table 4). The largest amount of annotation information was obtained from the NR database, 9,948 genes accounting for 88.75% of all unigenes. The annotation information from the KEGG database was in the middle, 5,871, accounting for 52.38%. The annotation information from the COG and KOG databases was relatively small, accounting for 31.90% and 38.70%, respectively. *C. militaris* accounted for more than 99.99% of the annotated sequence, which showed that the sequencing data were highly reliable.

We compared the entire gene set with the *C. militaris* standard CM01 in KOG, eggNOG, NR, Pfam, and Swiss-protein databases. We obtained 4,338, 8,684, 9,948, 7,244, and 5,113 genes, respectively (Fig. 1A). In addition, 0, 0, 1,142, 1, and 0 genes were annotated only using the KOG, eggNOG, NR, Pfam and Swiss-protein databases, respectively (Fig. 1A). We divided 4,338 genes into 26 KOG classification categories for homologous gene product classification (Fig. B). Among the functional classification categories, “general function prediction only” (14.67%) was the largest group, followed by “posttranslational modification,

protein turnover, chaperones" (9.12%) and "signal transduction mechanisms" (7.21%), which indicated that some modified proteins and signal transmission may be important in cordycepin biosynthesis.

D. Genes differentially expressed between WT, GYS60, and GYS80

After we merged the novel coding transcripts with the reference genome to obtain a complete reference, clean reads were mapped to this reference to calculate the gene expression level by FPKM values. On the basis of the mapping results, we analyzed global transcriptional changes from normalizing the differentially expressed gene (DEGs) data to identify DEGs (Fig. 2A). We calculated the read distributions (Fig. 2B), hierarchical clustering between nine samples (Fig. 2C), and the Pearson correlation coefficient between expression values across samples (Fig. 2D). Most of the squared Pearson correlation coefficients (R^2) were greater than 0.7, which indicated credible differential expression.

Fig. 2E shows a volcano plot of the global fold changes between the upregulated and downregulated genes. Compared with WT, we found 145 genes upregulated and 967 genes downregulated in GYS60 and 470 genes upregulated and 594 genes downregulated in GYS80. Compared with GYS80, GYS60 had 306 upregulated genes and 207 downregulated genes. As shown in the volcano plot, most of the DEGs represented a slight fold change, although some DEGs displayed large differences. These results revealed that some yet unknown genes regulated the cordycepin metabolic network. Additionally, Fig. 2F shows a heatmap of the hierarchical clustering of the DEGs. The expression profiles between the replicates were similar, which indicated that the transcriptome data were reliable.

E. GO analysis of differentially expressed genes

316 To identify the functions of genes, we used the GO database for enrichment analysis.
317 There were 7,050 differentially expressed genes significantly enriched in the GO entries (Fig.
318 3). The genes were classified into three major functional categories “cellular component”,
319 “molecular function”, and “biological process” and 49 subcategories. The biological process
320 contained 995 genes in 16 subcategories. The cell component category contained 14
321 subcategories with 2,010 genes. The molecular function category contained 4,045 genes in 19
322 subcategories. When we compared GYS60 with WT, genes related to detoxification,
323 antioxidant activity, and molecular transducer activity were significantly upregulated in
324 GYS60, whereas genes related to growth, supramolecular complex, structural molecular
325 activity, transcription factor activity, and electron carriers were significantly downregulated
326 in GYS60 (Fig. 3A). When we compared GYS80 with WT, genes related to molecular
327 transducer activity were significantly upregulated, and genes related to structural molecular
328 activity were significantly downregulated in GYS80 (Fig. 3B).
329 In GYS60 compared with GYS80, the rhythmic process, biological adhesion,
330 development process, detoxification reaction, and antioxidant activity related genes of
331 GYS60 were significantly upregulated, whereas genes related to supramolecular complexes
332 were significantly downregulated (Fig. 3C). These results suggested that detoxification and
333 antioxidant activities were important in the cordycepin biosynthesis network.
334 **F. KEGG pathway analysis of DEGs**
335 To understand the interactions between genes, we mapped all the differentially expressed
336 genes to the KEGG database to categorize their functions. We found 51 pathways differentially
337 enriched in the GYS60 compared with WT comparison. Of the top 20 pathways (Fig. 4A),

"Tryptophan metabolism" (8, 15.38%) contained the most genes and showed the most differential enrichment, followed by "Arginine and proline metabolism" (5, 9.62%). The "Biosynthesis of various secondary metabolites" pathway showed the greatest enrichment degree, and was upregulated significantly, followed by "Citrate cycle".

In GYS80, compared with WT (Fig. 4B), "Cyanoamino acid metabolism" (13, 4.17%) showed the most differential enrichment, and the "Arginine and proline metabolism" pathway was downregulated significantly. "Tryptophan metabolism", "Biosynthesis of various secondary metabolites" and "Citrate cycle" pathways were not in the top 20 of KEGG enriched pathways. In the GYS60 *vs*. GYS80 (Fig. 4C), "Tryptophan metabolism" (13, 8.90%) and "Fatty acid degradation" (10, 6.85%) showed the most differential enrichment, and "indole alkaloid biosynthesis" showed the greatest enrichment degree in GYS60.

The P values of 12 pathways in level 1 "Metabolism" were less than 0.05 (Table 5). In GYS60, compared with WT, the pathways "Tryptophan metabolism", "Purine metabolism" and "Arginine and proline metabolism" showed an even greater significant enrichment with $P < 0.01$.

G. Trend analysis of differentially expressed genes

We subjected the DEG expression patterns to trend analyses. We found 11,211 genes clustered into sixteen profiles, and profiles 2, 3, 0, 13, 7, 15, 8, and 12 showed significant enrichment ($P \leq 0.05$; Fig. 5). The expression of 85, 107 and 64 genes displayed an increasing trend after WT in profile 13, 15, and 12, respectively. The expression of 204, 124, and 105 genes were downregulated after WT in profile 2, 3, and 0, respectively. In profile 7, the expression of 143 genes showed consistently high expression from WT to GYS60, but a subsequent decrease

from GYS60 to GYS80. In profile 8, the expression of 93 genes showed consistently high expression from WT to GYS60, and an increase from GYS60 to GYS80.

H. Transcription factors and target genes in production of cordycepin

To further analyze the molecular mechanism of cordycepin overproduction in GYS60, we classified the total correspondence between transcription factors and target genes into nineteen transcription factor families (Fig. 6A). The transcription factors belonging to the basic helix-loop-helix factor family contained the most genes at 14,771 (22.54%), followed by the homeodomain factor family (13,633 genes, 20.80%), the C_2H_2 zinc finger factor family (6,096 genes, 9.30%), and the high-mobility group domain factor family (5,181 genes, 7.91%). The two most significantly up-regulated transcription factors were retinoid X receptor alpha (Rxra), belonging to nuclear receptors with C_4 zinc fingers, and Krüppel-like factor 4 (Klf4), belonging to the C_2H_2 zinc finger factor family (Table 6). The genes CCM_02777, CCM_03411, CCM_02755, CCM_04722, and CCM_07507 were significantly upregulated by Klf4 and Rxra, and the gene CCM_06928 was upregulated by Myb (Fig. 6B). The gene CCM_07507, annotated as oxidoreductase, was significantly upregulated by Klf4 and Rxra in GYS60 (Table 7, Fig. 6B). This highly upregulated transcription of Klf4 and Rxra likely led to enhanced ATP generation. Because adenosine-3',5'-cyclic monophosphate (3′,5′-cAMP) can be obtained by the over-dephosphorylation and cyclization of ATP, CCM_06928, an adenylate cyclase, directly increased the production of 3′,5′-cAMP, which is a plausible precursor of cordycepin (Fig. 6C). Adenine and AMP were also plausible cordycepin precursors.[38]

Five genes annotated as adenylate kinase (classified as EC 2.7.4.14, EC 2.7.4.3) did not show any significant difference among the three strains (Table 7). Another enzyme verified [38]

382 to be involved in cordycepin synthesis was 3', 5'-cyclic-AMP phosphodiesterase (EC 3.1.4.17),
383 annotated as gene CCM_02777, which transforms 3', 5'-cyclic-AMP into 3'-AMP. The
384 transcriptional level of CCM_02777 was significantly upregulated by the two transcription
385 factors Klf4 and Rxra in GYS60 (Table 7, Fig. 6C), but was unchanged in GYS80 (Table 7,
386 Fig. 6C).

387 Another pathway for the production of ATP in WT is by catalysis of several enzymes from
388 5'-phosphoadenylylsulfate. The transcriptional level of DEG CCM_06643, annotated as sulfate
389 adenylyltransferase (EC 2.7.7.4) that converts 5'-phosphoadenylylsulfate into adenylysulfate,
390 was significantly upregulated by Myb and Gata1 in GYS60 and unchanged in GYS80 (Table
391 7, Fig. 6C). The downstream gene CCM_02755, ATP adenylyltransferase (classified as EC
392 2.7.7.53), for P1, P4-diadenosine-5'-tetraphosphate (AppppA) formation from adenylysulfate
393 also showed significant upregulation by Klf4 and Rxra in GYS60 (Table 7, Fig. 6C) and
394 showed downregulation in GYS80 (Table 7). The gene CCM_03411, annotated as adenosine
395 kinase (EC 3.6.1.41) involved in conversion of AppppA into ATP, was significantly upregulated
396 by Klf4 and Rxra in GYS60 and unchanged in GYS80 (Table 7, Fig. 6B). Finally, the pathway
397 was connected to the cordycepin biosynthesis cluster *Cns*1 and *Cns*2. However, the
398 transcriptional levels of CCM_04436 (*Cns*1), annotated as oxidoreductase/dehydrogenase, and
399 CCM_04437 (*Cns*2), annotated as metal-dependent phosphohydrolase, were not upregulated
400 in GYS60, whereas the two genes were significantly upregulated in WT (Table 7). CCM_04438
401 (*Cns*3), annotated as nucleotide kinase, and CCM_04439 (*Cns*4), annotated as ATPase/ABC
402 transporter, were not upregulated in any of the three strains.

403 A phosphoesterase (EC 3.1.4.16) can also produce 3′-AMP from 2′,3′-cAMP, which was

generated during mRNA degradation by newGene_1809 and newGene_930 (Table 7). Adenine phosphoribosyltransferase (CCM_00088) and nucleoside triphosphate pyrophosphatase (CCM_03944) in the purine biosynthesis pathway were not upregulated (Table 7), in agreement with a report by Zheng et al.[28]

I. Expression of four key cordycepin synthesis genes

After transcriptome analysis of WT, GYS60 and GYS80, we also measured the expression levels of four key genes, *Cns*1, *Cns*2, *Cns*3 and *Cns*4, in the main pathway of cordycepin biosynthesis in the three strains. The four key genes encode oxidoreductase (*Cns*1), phosphorylase (*Cns*2), phosphotransferase (*Cns*3), and channel protein (*Cns*4).[33] In WT, the expression levels of the two most critical genes *Cns*1 and *Cns*2 were significantly higher than their expression in GYS60, whereas the expression levels of *Cns*3 and *Cns*4 were low (Fig. 7). However, compared with WT, the expression of these four key genes was significantly downregulated in GYS80 (Fig. 7). Compared with GYS80, the expression of these genes was significantly upregulated in GYS60 (Fig. 7). These results further indicated that cordycepin was not synthesized in GYS60 by the main synthetic pathway.

J. qRT-PCR validation of differentially expressed genes

We confirmed the transcriptome data by using qRT-PCR to measure expression of eight key genes, CCM_08167, CCM_07505, CCM_04436 (*Cns*1), CCM_04437 (*Cns*2), CCM_04438 (*Cns*3), CCM_04439 (*Cns*4), CCM_06928, CCM_04722, CCM_02777 and CCM_07507. Although the fold changes of these genes measured by qRT-PCR were slightly different from the transcriptome data, the two sets of measurements shared a similar tendency (Fig. 8). In GYS60, the expression of CCM_07507, CCM_02777, CCM_04722, and

CCM_06928 was the highest. In WT, CCM_04437 (*Cns*2) and CCM_04436 (*Cns*1) had the highest expression. In GYS60, the expression of CCM_04437 (*Cns*2) and CCM_04436 (*Cns*1) was not upregulated, whereas expression of CCM_04722, CCM_04439 (*Cns*4), and CCM_04438 (*Cns*3) was slightly higher than in WT.

These results indicated that in GYS60 the expression levels of genes CCM_07507, CCM_02777, CCM_06928 were strongly positively correlated with the biosynthesis of cordycepin, a finding that can explain why cordycepin production is significantly increased in GYS60. The expression patterns of CCM_04436 (*Cns*1), CCM_04437 (*Cns*2), CCM_04438 (*Cns*3) , and CCM_04439 (*Cns*4) in GYS60, related to the known main biosynthetic pathway of cordycepin, were different from expression in WT.

Expression of transcription factors genes CCM_07505 (Rxra) and CCM_08167 (Klf4) was distinctly different in GYS60 versus *C. militaris* (Fig. 8). CCM_07505, belonging to nuclear receptors with C_4 zinc fingers, participates in the mitogen-activated protein kinase signaling pathway (ko04011). CCM_08167, belonging to the C_2H_2 zinc finger factor family, is involved in mRNA surveillance (ko03015), RNA processing, and RNA modification. The qRT-PCR data for these transcription factors and other genes were consistent with the RNA-seq results.

IV. DISCUSSION

Cordycepin production can be increased by optimizing culture conditions.[20, 25, 41-44] Zhang et al. created a high cordycepin-producing strain GYS60, and a low cordycepin-producing strain GYS80.[27] The purpose of this study was to use transcriptome sequencing to identify the genes involved in the metabolic pathway of cordycepin synthesis and the likely means by

448 which the pathway is overexpressed in GYS60 and low-expressed in GYS80.

449 The transcription data were annotated from the reference database of *C. militaris* and
450 showed that more than 94% of the clean bases had quality scores of Q30 in the mycelial stage
451 of *C. militaris* (Table 3). To confirm the data, we calculated the read distributions, the Pearson
452 correlation coefficient between expression values across samples, and hierarchical clustering
453 between nine samples. These data showed the samples that shared the same culture condition,
454 which could be considered as reliable biological replicates. The DEG analysis results revealed
455 that only a few genes were regulated in GYS60, which indicated a possibility to narrow the
456 DEG targets to determine which genes were involved in the synthesis and metabolism of
457 cordycepin (Fig. 3 and 4).

458 We classified the DEGs with respect to the GO and KEGG databases. GO classification
459 indicated that the metabolic process might be improved to produce more metabolites and
460 consequently activate membrane transport to avoid the accumulation of high cordycepin
461 concentration in GYS60. The purine metabolism pathway of *C. militaris* may possess key
462 genes involved in cordycepin biosynthesis. Compared with WT, 10 genes were significantly
463 upregulated in GYS60, among which CCM_07507, CCM_06928, CCM_02777, and
464 CCM_04722 were strongly positively correlated with the biosynthesis of cordycepin, a finding
465 that can explain the significantly increased cordycepin production in GYS60 (Table 7, Fig. 8).
466 RT-PCR analysis confirmed that CCM_07507 had the highest expression level, followed by
467 CCM_06928, and CCM_02777 (Fig. 8). CCM_07507 is annotated as an oxidoreductase in the
468 GO database and shown as a gene related to energy production and transformation in the COG
469 database. CCM_06928 is annotated as adenylate cyclase in GO and KEGG and participates in

cAMP biosynthesis, which showed that ATP is transformed into 3', 5'-cyclic AMP in the purine metabolism pathway. CCM_02777 is annotated as a 3', 5'-cAMP phosphodiesterase in the GO and KEGG databases. CCM_04722 is annotated as a transferase in the GO database, whereas in the COG database it is shown as encoding a protein related to the synthesis, catalysis, and transport of secondary metabolites. KEGG classification showed that most DEGs were relevant to the primary pathway of metabolism, secondary metabolites, and citrate cycle. These results indicated that the mutagenesis system used to create the GYS60 strains may have affected amino acid and energy metabolism. Significant changes in KEGG pathways of tryptophan metabolism, purine metabolism, and arginine and proline metabolism were confirmed, as reported by Chen et al.[31]

Some transcription factors are identified as being involved in cordycepin production. CCM_05346, CCM_02568, and CCM_01481, belonging to Zn_2Cys_6-type transcription factor class, are verified to promote cordycepin production.[31,45] In this study, the expression of most transcription factor genes remained largely unchanged in the three strains, except two zinc fingers transcription factors. Compared with WT and GYS80, CCM_07505 (Rxra), belonging to the nuclear receptors with C_4 zinc fingers TF family, was highly regulated in GYS60 (Table 6, Fig. 8). Compared with WT, only CCM_08167 (Klf4), belonging to C_2H_2 zinc finger factors transcription factor family, was highly regulated in GYS60 (Table 6, Fig. 8). CCM_07505 participates in the mitogen-activated protein kinase signaling pathway (ko04011). CCM_08167 is involved in mRNA surveillance (ko03015), RNA processing, and RNA modification. The genes CCM_02777, CCM_03411, CCM_02755, CCM_04722, and CCM_07507 were significantly upregulated by Klf4 and Rxra, whereas CCM_06928 was upregulated by Myb

(Fig. 6B). The gene CCM_07507, annotated as oxidoreductase, was significantly upregulated by Klf4 and Rxra in GYS60 (Table 7, Fig. 6B). This highly upregulated transcriptional level likely led to enhanced ATP generation. Because adenosine-3',5'-cyclic monophosphate (3′,5′-cAMP) can be obtained by the over-dephosphorylation and cyclization of ATP, CCM_06928, an adenylate cyclase, directly increased the production of 3′,5′-cAMP, which is a plausible precursor of cordycepin (Fig. 6C). The expression of Klf4 and Rxra may be related to cellular DNA damage stress response induced by multifunctional plasma mutagenesis system in our previous study.[27] However, further study is needed to confirm that biosynthesis of cordycepin is regulated by Klf4 and Rxra.

We found two new genes related to RNA processing and splicing, newGene_1809 and newGene_930. NewGene_1809 is annotated in the GO database and participates in the processing and splicing of rRNA and tRNA (GO: 0006364, GO: 0008033); newGene_930 is annotated as RNA cleavage in both GO and KEGG databases (GO: 000375, K13124). In GYS60 compared with WT, the differential expression of the newGene_1809 and newGene_930 was significantly upregulated, $\log_2$(FC) reaching 2.64240 and 1.54194, respectively. When we compared GYS80 with WT, the $\log_2$(FC) of newGene_1809 and newGene_930 were only 0.30719 and 0.37827, respectively. In GYS60, upregulation of newGene_1809 and newGene_930 may lead to an increase in cordycepin production, which is a complementary pathway for cordycepin biosynthesis.[33, 39] In many organisms, RNA is degraded to 2',3'-cAMP, which is then converted by 2',3'-cyclonucleoside-2'-phospholipase (CCM_07683) into 3'-AMP.[39,46] In WT, cordycepin is produced by phosphohydrolase (CCM_04437, *Cns*2) and oxidoreductase (CCM_04436, *Cns*1). But compared with WT, the

514 differential expression level of CCM_07683 in GYS60 was only -0.35295. Therefore, in
515 GYS60, RNA degradation generates 2', 3'-cAMP, which is transformed into cordycepin by this
516 supplementary pathway. In GYS60, the expression levels of CCM_04439 (*Cns*4), CCM_04438
517 (*Cns*3), and CCM_04436 (*Cns*1), components of the main known biosynthetic pathway of
518 cordycepin, were not upregulated (Fig. 7, Table 7). Therefore, other genes are involved in the
519 biosynthesis of cordycepin in *C. militaris*.

520 **V. CONCLUSIONS**

521 We analyzed the expression of genes involved in cordycepin synthesis in *C. militaris* WT,
522 GYS60 and GYS80. Up-regulating genes in the bio-pathways of energy generation and purine
523 conversion was the major reason for the overproduction of cordycepin. Two zinc finger
524 transcription factors, Rxra and Klf4, were verified to increase the expression of four coding
525 genes in GYS60. In GYS60, the synthesis rate of ATP is accelerated by oxidoreductase (gene
526 CCM_07507), leading to more ATP production, which leads to more ATP being converted by
527 adenylate cyclase (CCM_06928) into 3', 5'-cAMP, then catalyzed by 3', 5'-cAMP
528 phosphodiesterase (CCM_02777) to produce 3'-AMP. Finally, 3'-AMP is catalyzed by
529 phosphohydrolase (CCM_04437, *Cns*2) and oxidoreductase (CCM_04436, *Cns*1) to produce
530 cordycepin. The cordycepin is transferred out of the cell by two transferases (CCM_04439 and
531 CCM_04722). Our findings reveal the flexibility of the cordycepin network and provide a
532 reference for genomic editing to improve cordycepin production as well as other compounds
533 of medicinal value in *C. militaris*.

534 **DECLARATION OF COMPETING INTEREST**

535 The authors declare that they have no known competing financial interests or personal

relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY STATEMENT

Data generated or analyzed during this study are provided in full within the published article.

DECLARATION OF GENERATIVE AI IN SCIENTIFIC WRITING

During the preparation of this report, the authors did not use any AI or AI-assisted technologies in the writing process.

DECLARATION OF ORIGINALITY

This article has not been published elsewhere and it has not been simultaneously submitted for publication elsewhere.

ACKNOWLEDGMENTS

The authors thank AiMi Academic Services (www.aimieditor.com) for English language editing and review services. The authors also sincerely thank the Reviewers and Editors for their helpful comments and suggestions.

AUTHOR CONTRIBUTION STATEMENT

Hucheng Zhang: Conceptualization, Supervision, Writing-Original draft, Formal Analysis. Lina Deng: Writing- Reviewing and Editing. Shuai Luo: Validation, Formal Analysis. Linying Liu: Software, Resources, Data curation. Guowei Yang: Data curation. Yuning Zhang and Yaguang Jiang: Visualization, Investigation. Bo Gao and Xiaojie Wang: Methodology. Xingjuan Li and Dongqing Yang: Software. Wenyan Lao and Frank Vriesekoop: Conceptualization, Project administration, Supervision.

FUNDING STATEMENT

This work was supported by Beijing Polytechnic project “Study on the immunomodulatory effect and mechanism of cordycepin on immunodeficient mice” [grant number 2024X020-KXZ] for Dr. Hucheng Zhang and R&D Program of Beijing Municipal Education Commission [grant number KM202210858001] for Dr. Linying Liu.

REFERENCES

1. Zhu JS, Halpern GM, Jones K. The scientific rediscovery of an ancient Chinese herbal medicine: *Cordyceps sinensis* Part I. J Altern Complem Med. 1998;4(3):289-303.
2. Cunningham KG, Manson W, Spring FS, Hutchinson SA. Cordycepin, a metabolic product from cultures of *Cordyceps militaris* (Linn.) link. Part I. isolation and characterization. J Chem Soc. 1951;2:2299-300.
3. Cohen N, Cohen J, Asatiani MD, Varshney VK, Yu HT, Yang YC, Li YH, Mau JL, Wasser SP. Chemical composition and nutritional and medicinal value of fruit bodies and submerged cultured mycelia of culinary-medicinal higher Basidiomycetes mushrooms. Int J Med Mushrooms. 2014;16(3):273-91.
4. Zhang GY, Yin QS, Han T, Zhao YX, Su JJ, Li MZ, Ling JY. Purification and antioxidant effect of novel fungal polysaccharides from the stroma of *Cordyceps kyushuensis*. Ind Crops Prod. 2015;69:485-91.
5. Paterson RRM. Cordyceps-A traditional Chinese medicine and another fungal therapeutic biofactory? Phytochemistry. 2008;69(7):1469-95.
6. Zhou X, Gong Z, Su Y, Lin J, Tang K. *Cordyceps* fungi: natural products, pharmacological functions and developmental products. J Pharm Pharmacol. 2009;61(3):279-91.
7. Zhang HC, Yang GW, Yang J, Fan HT, Luo S, Liu LY. Preparation of cordycepin nanoemulsion and its repair mechanism on the photoaged skin. China Surfactant Deterg Cosmet. 2022;52:1072-80. (in Chinese)
8. Tuli HS, Sandhu SS, Sharma AK. Pharmacological and therapeutic potential of *Cordyceps* with special reference to cordycepin. 3 Biotech. 2014;4(1):1-12.

9. Cui ZY, Park SJ, Jo E, Hwang IH, Lee KB, Kim SW, Kim DJ, Joo JC, Hong SH, Lee MG, Jang IS. Cordycepin induces apoptosis of human ovarian cancer cells by inhibiting CCL5-mediated Akt/NF-κB signaling pathway. Cell Death Discov. 2018;4:62.

10. Xu JC, Zhou XP, Wang XA, Xu MD, Chen T, Chen TY, Zhou PH, Zhang YQ.Cordycepin induces apoptosis and G2/M phase arrest through the ERK pathways in esophageal cancer cells. J Cancer. 2019;10(11):2415-2424.

11. Khan MA, Tania M. Cordycepin in anticancer research: Molecular mechanism of therapeutic effects. Curr Med Chem. 2020;27(6):983-996.

12. Zheng Q, Sun J, Li W, Li S, Zhang K. Cordycepin induces apoptosis in human tongue cancer cells in vitro and has antitumor effects in vivo. Arch Oral Biol. 2020;118:104846.

13. Jiang Q, Lou Z, Wang H, Chen C. Antimicrobial effect and proposed action mechanism of cordycepin against *Escherichia coli* and *Bacillus subtilis*. J Microbiol. 2019;57(4):288-297.

14. Sugar AM, McCaffrey RP. Antifungal activity of 3'-deoxyadenosine (cordycepin). Antimicrob Agents Chemother. 1998;42(6):1424-1427.

15. Verma AK. Cordycepin: a bioactive metabolite of *Cordyceps militaris* and polyadenylation inhibitor with therapeutic potential against COVID-19. J Biomol Struct Dyn. 2022;40(8):3745-3752.

16. Olatunji OJ, Feng Y, Olatunji OO, Tang J, Ouyang Z, Su Z. Cordycepin protects PC12 cells against 6-hydroxydopamine induced neurotoxicity via its antioxidant properties. Biomed Pharmacother. 2016;81:7-14.

17. Wang X, Xi D, Mo J, Wang K, Luo Y, Xia E, Huang R, Luo S, Wei J, Ren Z, Pang H, Yang

607 R. Cordycepin exhibits a suppressive effect on T cells through inhibiting TCR signaling
608 cascade in CFA-induced inflammation mice model. Immunopharmacol Immunotoxicol.
609 2020;42(2):119-127.
610 18. Huang S, Liu H, Sun YH, Chen J, Li XF, Xu JF, Hu YW, Li YQ, Deng ZW. Zhong SA 2018.
611 An effective and convenient synthesis of cordycepin from adenosine. Chem Pap. 2018;72:
612 149-160.
613 19. Mao XB, Eksriwong T, Chauvatcharin S, Zhong JJ. Optimization of carbon source and
614 carbon/nitrogen ratio for cordycepin production by submerged cultivation of medicinal
615 mushroom *Cordyceps militaris*. Process Biochem. 2005;40:1667-1672.
616 20. Mao XB, Zhong JJ. Significant effect of NH_4^+ on cordycepin production by submerged
617 cultivation of medicinal mushroom *Cordyceps militaris*. Enzyme Microb Technol.
618 2006;38: 343-350.
619 21. Gu YX, Wang ZS, Li SX, Yuan QS. Effect of multiple factors on accumulation of
620 nucleosides and bases in *Cordyceps militaris*. Food Chem. 2007;102(4): 1304-1309.
621 22. Fan D, Wang W, Zhong JJ. Enhancement of cordycepin production in submerged cultures
622 of Cordyceps militaris by addition of ferrous sulfate. Biochem. Eng J. 2012;60: 30-35.
623 23. Lee SK, Lee JH, Kim HR, Chun Y, Lee JH, Yoo HY, Park C, Kim SW. Improved
624 cordycepin production by *Cordyceps militaris* KYL05 using casein hydrolysate in
625 submerged conditions. Biomolecules. 2019;9(9):461.
626 24. Kang C, Wen TC, Kang JC, Meng ZB, Li GR, Hyde KD. Optimization of large-scale culture
627 conditions for the production of cordycepin with *Cordyceps militaris* by liquid static
628 culture. Sci World J. 2014;2014:510627.

25. Wen TC, Kang C, Meng ZB, Qi YB, Kang JC. Enhanced production of cordycepin by solid state fermentation of *Cordyceps militaris* using additives. Chiang Mai J Sci. 2016;43:972-984.

26. Das SK, Masuda M, Hatashita M, Sakurai A, Sakakibara M. Optimization of culture medium for cordycepin production using *Cordyceps militaris* mutant obtained by ion beam irradiation. Process Biochemistry. 2010;45(1):129-132.

27. Zhang H, Deng L, Zhang Z, Guan Y, Li B, Yang J, Fan H, Yang G, Chen X, Zhang J, Xin X, Vriesekoop F. Enhanced cordycepin production in caterpillar medicinal mushroom, *Cordyceps militaris* (ascomycetes), mutated by a multifunctional plasma mutagenesis system. Int J Med Mushrooms. 2020;22(12):1147-1159.

28. Zheng P, Xia Y, Xiao G, Xiong C, Hu X, Zhang S, Zheng H, Huang Y, Zhou Y, Wang S, Zhao GP, Liu X, St Leger RJ, Wang C. Genome sequence of the insect pathogenic fungus *Cordyceps militaris*, a valued traditional Chinese medicine. Genome Biol. 2011;12(11):R116.

29. Yin Y, Yu G, Chen Y, Jiang S, Wang M, Jin Y, Lan X, Liang Y, Sun H. Genome-wide transcriptome and proteome analysis on different developmental stages of *Cordyceps militaris*. PLoS One. 2012;7(12):e51853.

30. Raethong N, Laoteng K, Vongsangnak W. Uncovering global metabolic response to cordycepin production in *Cordyceps militaris* through transcriptome and genome-scale network-driven analysis. Sci Rep. 2018; 8(1):9250.

31. Chen BX, Wei T, Xue LN, Zheng QW, Ye ZW, Zou Y, Yang Y, Yun F, Guo LQ, Lin JF. Transcriptome analysis reveals the flexibility of cordycepin network in *Cordyceps*

militaris activated by L-alanine addition. Front. Microbiol. 2020;11: 577. doi: 10.3389/fmicb.2020.00577.

32. Suparmin A, Kato T, Dohra H, Park EY. Insight into cordycepin biosynthesis of *Cordyceps militaris*: Comparison between a liquid surface culture and a submerged culture through transcriptomic analysis. Plos One. 2017;12(11):e0187052.

33. Xia Y, Luo F, Shang Y, Chen P, Lu Y, Wang C. Fungal cordycepin biosynthesis is coupled with the production of the safeguard molecule pentostatin. Cell Chem Biol. 2017; 24(12):1479-1489.

34. Huo C, Li H, Li Q, Wang J, Li C, Wang L. Construction and optimization of cordycepin-producing *Saccharomyces cerevisiae*. Chin J Biotechnol. 2021;37:3334-4337. (in Chinese)

35. Beers J, Linask KL, Chen JA, Siniscalchi LI, Lin YS, Zheng W, Rao M, Chen GK. A cost-effective and efficient reprogramming platform for large-scale production of integration-free human induced pluripotent stem cells in chemically defined culture. Sci Rep. 2015;5:11319. DOi: 10.1038/srep11319

36. Kumar L, E Futschik M. Mfuzz: a software package for soft clustering of microarray data. Bioinformation. 2007;2(1):5-7.

37. Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, Zhang P, Banf M, Dai X, Martin GB, Giovannoni JJ, Zhao PX, Rhee SY, Fei Z. iTAK: A program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. Mol Plant. 2016;9(12):1667-1670.

38. Kuo HC, Huang IC, Chen TY. *Cordyceps* s.l. (Ascomycetes) species used as medicinal mushrooms are closely related with higher ability to produce cordycepin. Int J Med

673 Mushrooms. 2015;17(11):1077-1085.
674 39. Wongsa B, Raethong N, Chumnanpuen P, Wong-Ekkabut J, Laoteng K, Vongsangnak W.
675 Alternative metabolic routes in channeling xylose to cordycepin production of *Cordyceps*
676 *militaris* identified by comparative transcriptome analysis. Genomics. 2020;112(1):629-
677 636.
678 40. Liu T, Liu Z, Yao X, Huang Y, Qu Q, Shi X, Zhang H, Shi X.Identification of cordycepin
679 biosynthesis-related genes through de novo transcriptome assembly and analysis in
680 *Cordyceps cicadae*. R Soc Open Sci. 2018;5(12):181247.
681 41. Dong JZ, Liu MR, Lei C, Zheng XJ, Wang Y. Effects of selenium and light wavelengths on
682 liquid culture of *Cordyceps militaris* link. Appl Biochem Biotechnol. 2012;166(8):2030-
683 2036.
684 42. Sari N, Suparmin A, Kato T, Park EY. Improved cordycepin production in a liquid surface
685 culture of *Cordyceps militaris* isolated from wild strain. Biotechnol Bioproc Eng.
686 2016;21:595-600.
687 43. Wu FC, Chen YL, Chang SM, Shih IL. Cultivation of medicinal caterpillar fungus,
688 *Cordyceps militaris* (Ascomycetes), and production of cordycepin using the spent medium
689 from levan fermentation. Int J Med Mushrooms. 2013;15(4):393-405.
690 44. Kunhorm P, Chueaphromsri P, Chaicharoenaudomrung N, Noisa P. Enhancement of
691 cordycepin production from *Cordyceps militaris* culture by epigenetic modification.
692 Biotechnol Lett. 2022;44(4):581-593.

693 45.Wang F, Liu Q,Zhang JJ , Liu KB, Li K, Liu GJ, Dong CH. Comparative transcriptome
694 analysis between a spontaneous albino mutant and its sibling strain of *Cordyceps militaris* in
695 response to light stress. Front. Microbiol. 2018;9: 1237. doi: 10.3389/fmicb.2018.01237
696 46.. Young M, Chin J. Dinuclear Copper (I1) complex that hydrolyzes RNA. J Am Chem Soc.
697 1995;117(42):10577-10578.
698

699 **TABLE 1.** Primers for qRT-PCR

Name	Forward primers (5'→3')	Reverse primers (5'→3')
18S rDNA	AGGTTTCGGGAATGTGGCTC	AGAACATCAGGATCGGTCGG
CCM_07507 (oxidoreductase)	GAAGTCACACGATGGGAGCA	AGGCACGGAGGTAGTAGAGG
CCM_02777 (3',5'-cyclic-AMP phosphodiesterase)	GCACAAACCTGTGCCGTTTA	GTGACGTTGGATTCTTGCGG
CCM_04722 (transferase)	GATGCTTCGATGTTGGCGAC	AGCATGCAACGAACAACACC
CCM_06928 (adenylate cyclase)	GCGGTACCTCAACCTATCCG	CGCTTTGATCGGGAATGCTG
CCM_04436 (oxidoreductase/dehydrogenase)	GCTCACGAGTTCGTCACTCA	TCGAAGCTCGTGCTCATTGT
CCM_04437 (metal-dependent phosphohydrolase)	CATTTCCGAGTGCGACAACG	CACCTTGTCCTCGACAGAGC
CCM_04438 (nucleotide kinase)	CGAGACGGTCTACAAGGAGC	GTTGTAGACGTGGTTCGGGT
CCM_04439 (ABC transporter)	AACTATCCAAACTCGCCGCA	GCCACGGTGCTTTGCTAAAA
CCM_07505 (DNA-binding protein eta2)	ACTCAGAGGCAACTTCCGTG	TTGCTCTCGAAACTCGGCTT
CCM_08167 (MYB DNA-binding domain)	ATGCCCGAGTGCAACTTCTT	AGAAGCCGGCGAGGTAAAAG

700

701

TABLE 2. Growth characteristics of *C. militaris* strains (n=3，$\bar{x}\pm SD$)

Growth characteristics	*C. militaris*	GYS60	GYS80
Maximum specific growth rate (μ_{max})	0.200 ± 0.010	0.210 ± 0.020	0.190 ± 0.030
Biomass productivity (g/L/day)	3.420 ± 0.030	3.510 ± 0.050	3.340 ± 0.040
Mycelium total cordycepin titer (mg/g)	18.902 ± 0.315	41.245 ± 0.827	6.376 ± 0.142
Extracellular total cordycepin titer (g/L)	0.381 ± 0.010	7.883 ± 0.108	0.154 ± 0.016
Extracellular cordycepin productivity (g/L/day)	0.025 ± 0.001	0.526 ± 0.003	0.011 ± 0.001

The growth state of three *C. militaris* strains was approximately the same, and the hyphae were collected at the end of the log phase (15th day) for transcriptome analysis. From the first day, 2 mL of fermentation liquid was taken every 2 days to measure the hyphal dry weight and analyze the cordycepin by HPLC until the 30th day of culture.

709 **TABLE 3.** Overview of transcriptome sequencing and de novo assemble

Samples	Clean reads (Mb)	Clean base (Gb)	GC content (%)	Q20 (%)	Q30 (%)	Total Mapping Ratio (%)	Uniquely Mapping Ratio (%)
C.militaris-1	29.04	8.66	57.77	97.87	94.84	95.86	95.50
C.militaris-2	28.80	8.57	57.29	97.54	94.12	91.77	90.97
C.militaris-3	23.76	7.08	57.43	97.68	94.41	91.25	90.00
GYS60-1	23.99	7.15	57.36	97.65	94.30	93.34	92.98
GYS60-2	26.54	7.92	57.71	97.81	94.69	94.01	93.46
GYS60-3	20.00	5.98	57.46	97.91	94.87	95.90	95.67
GYS80-1	27.80	8.30	57.42	97.89	94.73	95.97	95.69
GYS80-2	26.12	7.80	57.80	97.88	94.73	95.87	95.57
GYS80-3	22.17	6.59	57.52	97.78	94.57	95.28	95.01
Mean	25.35	7.56	57.53	97.78	94.58	94.36	93.87

710

711

712 **TABLE 4.** Number of all annotated genes

Databases	COG	GO	KEGG	KOG	Pfam	Swiss-Prot	eggNOG	NR	All
Genes number*	3 576	7 050	5 871	4 338	7 244	5 113	8 684	9 948	9 950
Percent (%)	31.9	62.9	52.38	38.7	64.64	45.62	77.47	88.75	88.77
New gene number	47	161	93	64	124	73	149	299	302
Total gene number	11210	11208	11204	11209	11215	11212	11205	11215	11204

713 Note: COG: Clusters of orthologous groups of proteins; GO: gene ontology; KEGG: Kyoto
714 encyclopedia of genes and genome; KOG: Clusters of orthologous groups for eukaryotic
715 complete genomes; Pfam: the protein family database, a large collection of protein families,
716 each represented by multiple sequence alignments and hidden Markov models (HMMs);
717 Swiss-prot: Protein sequence database; eggNOG: evolutionary genealogy of genes: non-
718 supervised orthologous groups; NR: National center for biotechnology information reference
719 sequence. * means number of annotated genes with length between 300 and 1,000 bp.

720

721 **TABLE 5.** The smallest *P*-value ($P < 0.01$) pathway in KEGG in GYS60 *vs*. *C.militaris*.

Groups	Pathway ID	Pathways	DEGs	All genes	*P* value	Up or down
GYS60	ko00380	Tryptophan metabolism	8	101	0.000219	up
vs.	ko00230	Purine metabolism	7	83	0.000435	up
C.militaris	ko00330	Arginine and proline metabolism	5	78	0.009108	up
	ko00460	Cyanoamino acid metabolism	13	33	7.49E-06	up&down
	ko00052	Galactose metabolism	11	32	0.000165	down
GYS80	ko00040	Pentose and glucuronate interconversions	11	39	0.001127	up&down
vs.	ko00620	Pyruvate metabolism	16	70	0.001171	up&down
C.militaris	ko00010	Glycolysis/Gluconeogenesis	16	72	0.001612	down
	ko00561	Glycerolipid metabolism	13	53	0.001703	up&down
	ko00910	Nitrogen metabolism	7	21	0.003254	up&down
GYS60	ko00071	Fatty acid degradation	10	64	0.000663	up&down
vs.	ko00380	Tryptophan metabolism	13	101	0.000733	up&down
GYS80	ko00360	Phenylalanine metabolism	8	56	0.004122	up

722 DEGs: the number of genes with significant differences in a specific KEGG. All genes: the

723 number of genes in a specific KEGG

TABLE 6. Transcription factors (TF) of DEGs

TF name	TF genes	TF family	Function	*P* value	Log2(fold change)		
					GYS60 *vs.* *C. militaris*	GYS80 *vs.* *C. militaris*	GYS60 *vs.* GYS80
Rxra	gene7485 (CCM_07505)	Nuclear receptors with C4 zinc fingers	DNA-binding protein eta2, putative [*C. militaris* CM01]	0	2.80528↑	0.53786	1.43002↑
Klf4	gene8146 (CCM_08167)	C2H2 zinc finger factors	high-affinity glucose transporter, putative [*C. militaris* CM01]	0	1.85277↑	1.40517↑	0.02360
Klf4	gene3199 (CCM_03209)	C2H2 zinc finger factors	GATA transcription factor (Ams2), putative [*C. militaris* CM01]	0	1.26476↑	1.02621↑	0.52238
Gata1	gene2715 (CCM_02724)	Other C4 zinc finger-type factors	C6 and C2H2 transcription factor, putative [*C. militaris* CM01]	0	1.20148↑	1.38959↑	-0.01041
Myb	gene3249 (CCM_03259)	Tryptophan cluster factors	Basic helix-loop-helix dimerization region bHLH [*C. militaris* CM01]	0	0.38913	0.14671	1.30092 ↑
NR4A2	gene8516 (CCM_08539)	Nuclear receptors with C4 zinc fingers	Rhodanese-like protein [*C. militaris* CM01]	0	0.38583	1.25289↑	-0.82609
En1	gene1326 (CCM_01331)	Homeo domain factors	CP2 transcription factor, putative [*C. militaris* CM01]	0	0.22359	2.54221↑	-2.18188↓
FOXO3	gene5619 (CCM_05636)	Fork head / winged helix factors	C6 zinc finger domain protein [*C. militaris* CM01]	0	0.60971	-1.37073↓	1.83983↑
Gata1	gene9501 (CCM_09529)	Other C4 zinc finger-type factors	GATA transcription factor (Ams2), putative [*C. militaris* CM01]	0	0.05418	-1.05327↓	0.94037
Gata1	gene1103 (CCM_01106)	Other C4 zinc finger-type factors	C2H2 transcription factor (Egr2), putative [*C. militaris* CM01]	0	-0.03915	1.73459↑	-1.50263↓
Sox17	gene9538 (CCM_09566)	High-mobility group (HMG) domain factors	CP2 transcription factor, putative [*C. militaris* CM01]	0	-0.25248	-1.03203↓	0.74372
NFKB1	gene8881 (CCM_08906)	Rel homology region (RHR) factors	CCCH finger DNA binding protein, putative [*C. militaris* CM01]	0	-0.48412	-1.10727↓	0.52573
Gata1	gene4615 (CCM_04628)	Other C4 zinc finger-type factors	zinc finger protein 58 [*C. militaris* CM01]	0	-1.06162↓	0.79859	-1.88823↓
Gata1	gene1840 (CCM_01847)	Other C4 zinc finger-type factors	pH-response transcription factor pacC/RIM101 [*C. militaris* CM01]	0	-1.53152↓	-1.82507↓	0.03867

C. militaris CM01 is the standard *Cordyceps militaris* strain described in NCBI. *C. militaris* means wild type; GYS60 is cordycepin high-producing strain, and GYS80 is a low-producing strain. The selected data were taken from GYS60 with upregulation of $\log_2$ (fold change) ≥1.0 or downregulation of $\log_2$(fold change) ≤-1, and false discovery rate ≤0.05 when compared with *C. militaris* or GYS80. ↑ means upregulation and ↓ means downregulation. *P* value <0.01 is highly significant.

TABLE 7. Genes involved in putative cordycepin metabolic network

Gene ID	Enzymes (GO or KEGG annotation)	EC number	Log2(fold change)		
			GYS60 *vs.* *C. militaris*	GYS80 *vs.* *C. militaris*	GYS60 *vs.* GYS80
newGene_1809	RNA processing [39] (GO:0008033, GO:0006364)		2.64240↑	0.30719	1.85078↑
CCM_07507	Oxidoreductase (this study) (GO:0016491, K09242)		2.36455↑	0.52483↓	1.78870↑
CCM_06643	Sulfate adenylyltransferase (this study) (K13811)	EC:2.7.7.4	2.22528↑	1.37120	-0.22220↓
CCM_02777	3',5'-cyclic-AMP phosphodiesterase (this study) (GO:0004115, K01120)	EC:3.1.4.17	2.01875↑	0.28670	-0.51224
CCM_04722	Transferase (this study) (GO:0016740, K15394)		1.97760↑	0.60133	1.78197↑
CCM_06928	Adenylate cyclase (this study) (GO:0004016, K01768)	EC:4.6.1.1	1.90324↑	0.57041	-1.21869↓
CCM_02755	ATP adenylyltransferase (this study) (GO:0019915, K00988)	EC:2.7.7.53	1.84915↑	0.62800	-1.00530↓
CCM_03411	Bis(5'-nucleosyl)-tetraphosphatase (this study) (GO:0016787, K01525)	EC:3.6.1.41	1.72021↑	0.74780	-0.82103
newGene_930	RNA splicing (GO:0000375, K13124)		1.54194↑	0.37827	1.20816
CCM_02791	Steryl ester hydrolase (this study) (GO:1990748, K14674)	EC:3.1.1.13	1.34472↑	0.92269	-0.58570
CCM_07972	5'-nucleotidase (GO:0016787, K03787)	EC:3.1.3.5	0.92601↑	-0.52084↓	1.48938↓
CCM_00088	Adenine phosphoribosyltransferase [33] (GO:0003999, K00759)	EC:2.4.2.7	0.208830	0.51802	-0.32829↓
CCM_04439	ATPase/ABC transporter (Cns4) [38] (GO:0042626, K08712)		0.21510	-0.32659↓	0.50765
CCM_04438	Nucleotide kinase (Cns3) [33]		0.13103	-1.12442↓	1.23076↑
CCM_04437	Metal-dependent phosphohydrolase (Cns2) [33]	EC:6.3.2.6	-0.21138	-3.00718↓	2.72442↑
CCM_04436	Oxidoreductase/dehydrogenase(Cns1) [33] (GO:0016491)		-0.95862	-2.76622↓	1.83948↑
CCM_01353	5'-deoxynucleotidase [29] (GO:0002953, K13800)	EC:2.7.4.14	-0.27640	-1.21290↓	0.88545
CCM_02335	Cytidylate kinase [29] (GO:0004127, K13800)	EC:2.7.4.3	0.07140	0.36521	-0.23848
CCM_03940	Adenylate kinase [29] (GO:0004017, K18532)	EC:2.7.4.3	-0.58453	0.10143	-0.72196
CCM_04983	Adenylate kinase [29] (GO:0004017, K00939)	EC:2.7.4.3	0.11502	-0.17331	0.33991
CCM_07479	Adenylate kinase [29] (GO:0004017, K00939)	EC:2.7.4.3	0.34128	0.11587	0.20869
CCM_07683	2',3'-cyclic-nucleotide-2'-phosphodiesterase [39] (GO:1990838, K23093)	EC:3.1.4.16	-0.35295	-0.33384	-0.04010
CCM_03944	Nucleoside triphosphate pyrophosphatase [33] (GO:0035529, K01519)	EC: 3.6.1.-	-0.42974	-0.66420	0.20920
CCM_00622	5′-nucleotidase [40] (GO:0016787, K01081)	EC:3.1.3.5	-0.84821	-2.62540↓	1.58210↑

C. militaris is wild type *C. militaris*. GYS60 is cordycepin high-producing strain; GYS80 is a low-producing strain. The selected data were taken from GYS60 with upregulation of $\log_2$ (fold change)≥1.0 or downregulation of $\log_2$(fold change)≤-1 and false discovery rate ≤0.05 when compared with *C. militaris* or GYS80. ↑ means up-regulation and ↓ means down-regulation.

FIGURE LEGENDS

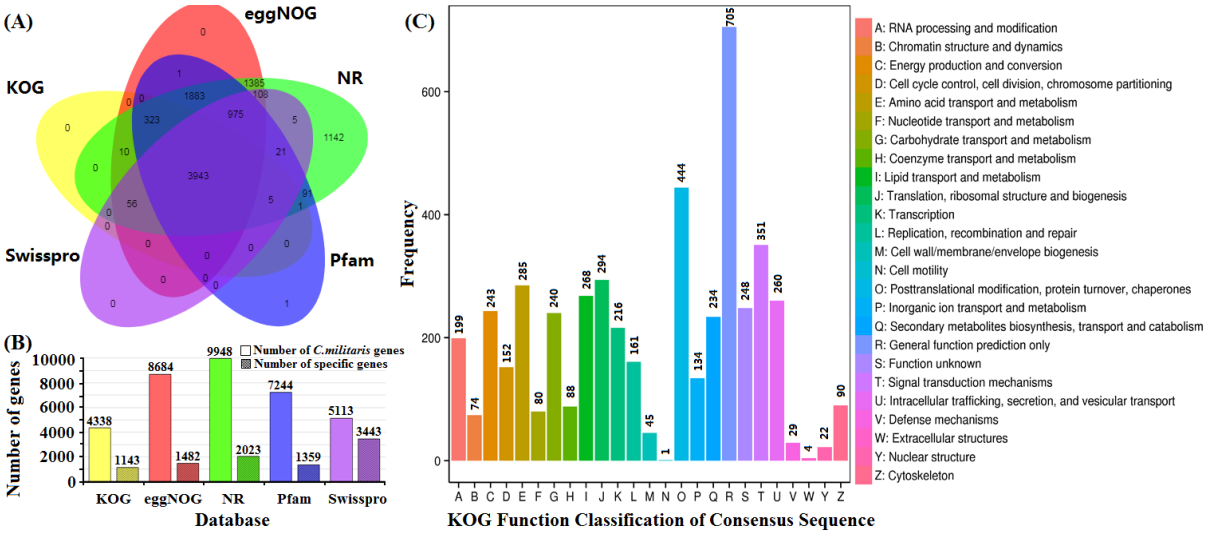

FIG. 1. Identification of all annotated genes. (A) Venn diagram showing the common and unique genes by five methods, including the KOG, eggNOG, NR, Pfam and Swiss-protein databases. (B) Number of *C.militaris* genes in each database. (C) Distribution of all genes based on KOG functional classification.

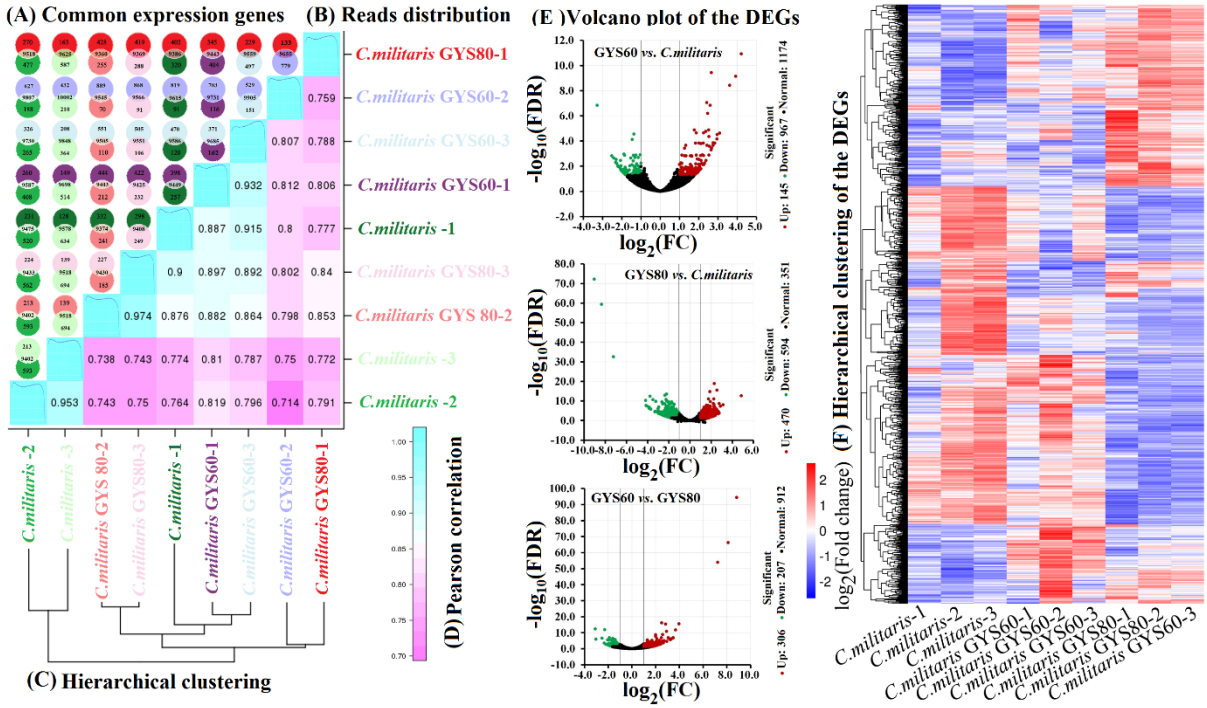

FIG. 2. Gene expression analysis. (A) Venn diagram of expressed genes in common between samples (above, the same coloring indicates the sample identity). (B) Rough graph of reads

distribution. (C) Hierarchical clustering of samples (below, a close distance indicates a similar expression profile of samples). (D) Heatmap of the Pearson correlations between samples. The x- and right y-axis represent each sample. The coloring indicates the Pearson correlation, high: cyan, low: pinkish-purple. (E) Volcano plot of the DEGs (GYS60 *vs*. *C.militaris*, GYS80 *vs*. *C.militaris*, GYS60 *vs*. GYS80). The x-axis represents the transformed $\log_2$foldchange. The y-axis represents the transformed -$\log_{10}$significance. Coloring indicates the fold change: upregulated DEGs, red; downregulated DEGs, green; nonDEGs, black. (F) Heatmap of the hierarchical clustering of the DEGs. The x-axis represents each compared sample; the y-axis represents the DEGs and gene clustering tree. In the corresponding sample, the coloring indicates the fold change: high, red; low, navy blue.

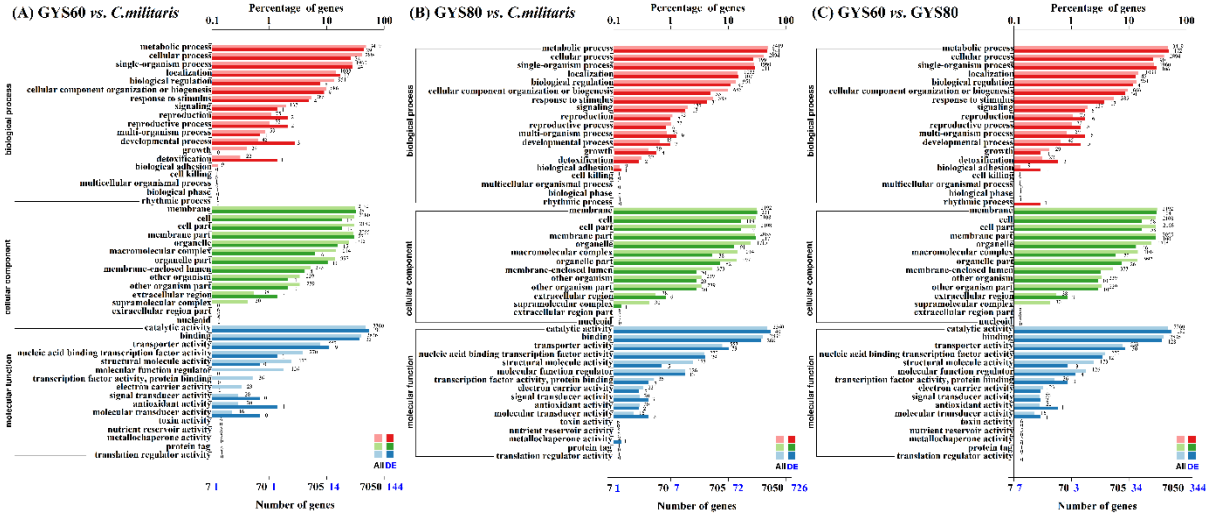

FIG. 3. Gene Ontology (GO) analysis of transcriptomes of *C. militaris* strains. (A) GYS60 *vs. C. militaris*. (B) GYS80 *vs. C. militaris*. (C) GYS60 *vs.* GYS80. The above and low x axis represent the percent and number of genes, respectively; the y axis represents the GO categories.

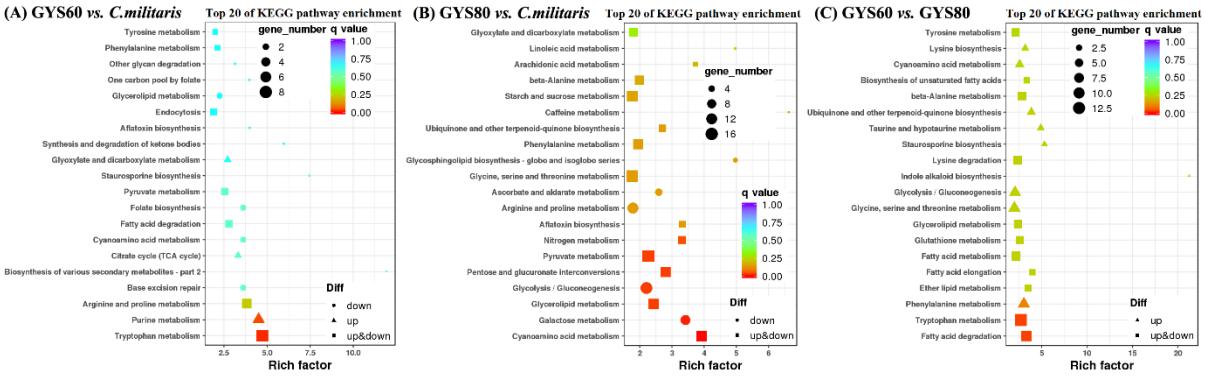

FIG. 4. Statistics of KEGG pathway enrichment. (A) GYS60 *vs*. *C.militaris*, (B) GYS80 *vs*. *C.militaris*, (C) GYS60 *vs*. GYS80. The x-axis represents the enrichment factor of DEGs, and the y-axis represents the KEGG pathway name. The enrichment factor is the ratio of the number of differential genes in a metabolic pathway to the number of all genes annotated for the pathway. The greater the value, the greater the enrichment degree. The degree of red indicates that the smaller the *q* value, the more significant the enrichment. The size of the dot indicates the number of differential genes in the pathway; a large dot indicates more genes.

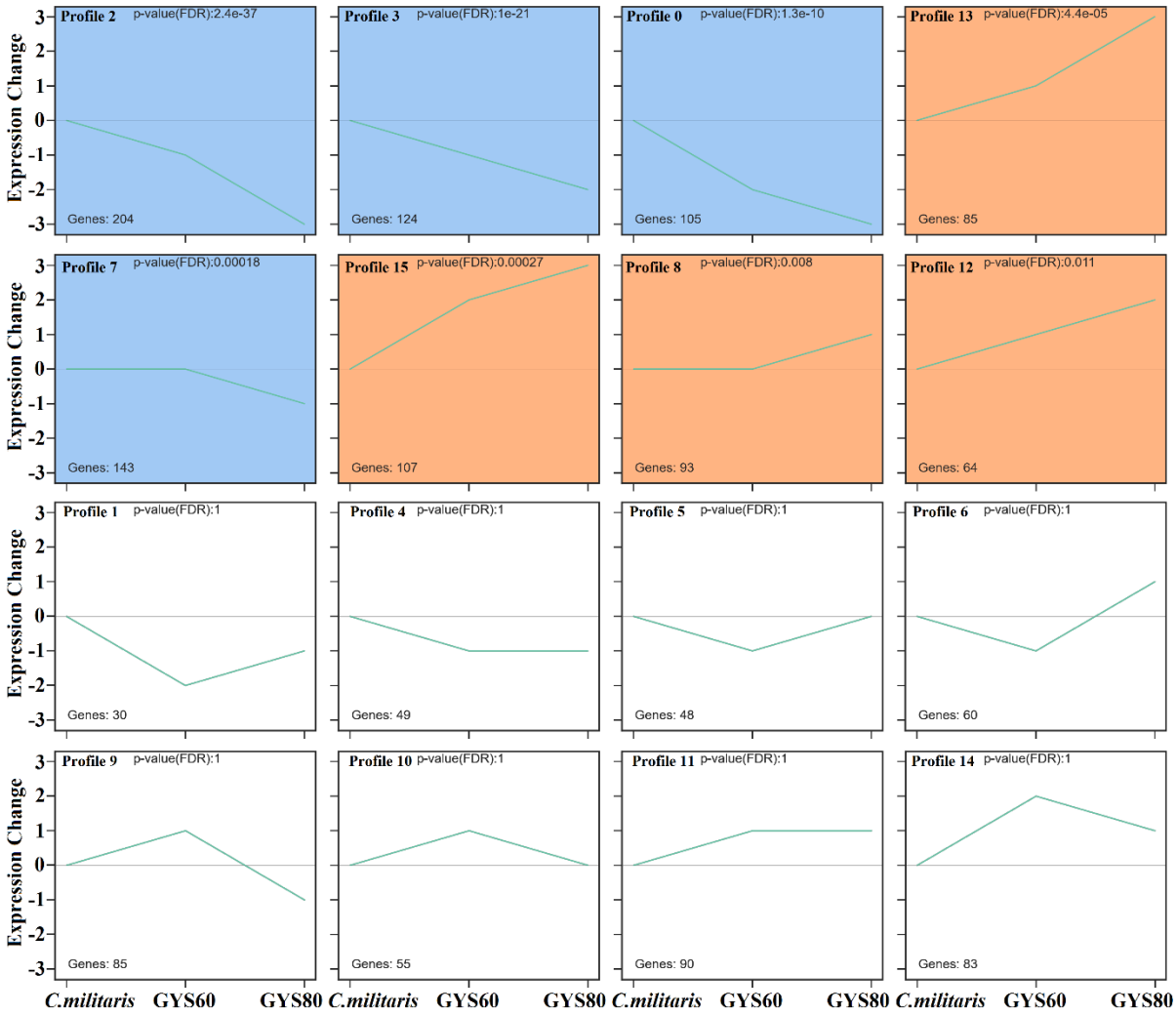

FIG. 5. STEM identification of the expression profiles of the number of DEGs in *C.militaris*, GYS60 and GYS80. Colored blocks indicate significant enrichment trends ($p \leq 0.05$), and different colors indicate different expression trends. The top-left number is the trend profile ID. The bottom-left number is the number of genes.

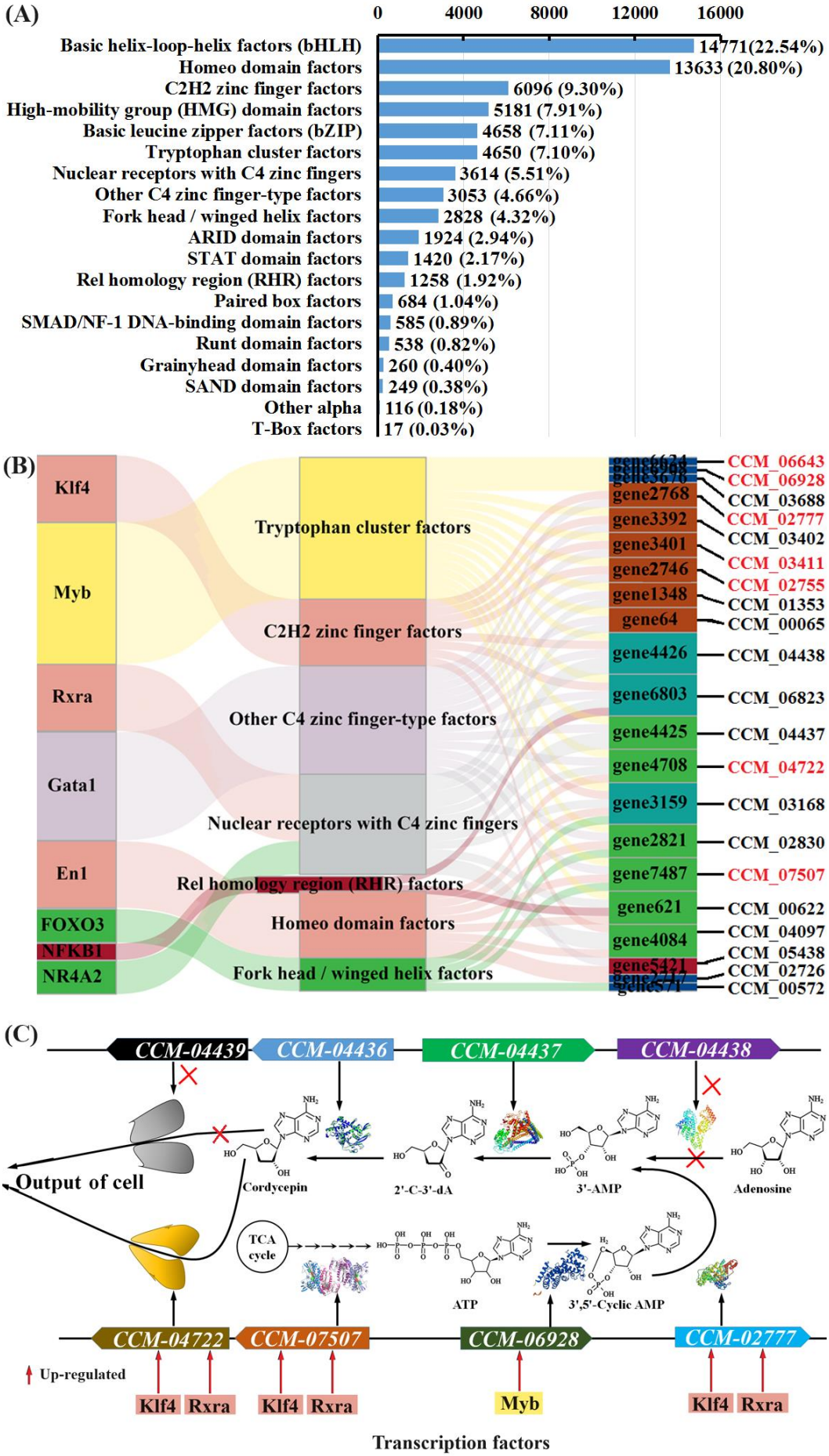
(A)
0
4000
8000
12000
16000
Basic helix-loop-helix factors (bHLH) 14771(22.54%)
Homeo domain factors 13633 (20.80%)
C2H2 zinc finger factors 6096 (9.30%)
High-mobility group (HMG) domain factors 5181 (7.91%)
Basic leucine zipper factors (bZIP) 4658 (7.11%)
Tryptophan cluster factors 4650 (7.10%)
Nuclear receptors with C4 zinc fingers 3614 (5.51%)
Other C4 zinc finger-type factors 3053 (4.66%)
Fork head / winged helix factors 2828 (4.32%)
ARID domain factors 1924 (2.94%)
STAT domain factors 1420 (2.17%)
Rel homology region (RHR) factors 1258 (1.92%)
Paired box factors 684 (1.04%)
SMAD/NF-1 DNA-binding domain factors 585 (0.89%)
Runt domain factors 538 (0.82%)
Grainyhead domain factors 260 (0.40%)
SAND domain factors 249 (0.38%)
Other alpha 116 (0.18%)
T-Box factors 17 (0.03%)
(B)
Klf4
Myb
Rxra
Gata1
En1
FOXO3
NFKB1
NR4A2
Tryptophan cluster factors
C2H2 zinc finger factors
Other C4 zinc finger-type factors
Nuclear receptors with C4 zinc fingers
Rel homology region (RHR) factors
Homeo domain factors
Fork head / winged helix factors
gene2768
gene3392
gene3401
gene2746
gene1348
gene64
gene4426
gene6803
gene4425
gene4708
gene3159
gene2821
gene7487
gene621
gene4084
gene5421
CCM_06643
CCM_06928
CCM_03688
CCM_02777
CCM_03402
CCM_03411
CCM_02755
CCM_01353
CCM_00065
CCM_04438
CCM_06823
CCM_04437
CCM_04722
CCM_03168
CCM_02830
CCM_07507
CCM_00622
CCM_04097
CCM_05438
CCM_02726
CCM_00572
(C)
CCM-04439
CCM-04436
CCM-04437
CCM-04438
Output of cell
Cordycepin
2'-C-3'-dA
3'-AMP
Adenosine
TCA cycle
ATP
3',5'-Cyclic AMP
CCM-04722
CCM-07507
CCM-06928
CCM-02777
Up-regulated
Klf4
Rxra
Klf4
Rxra
Myb
Klf4
Rxra
Transcription factors

FIG. 6. Transcription factors families and transcription factors related with cordycepin biosynthesis in *C.militaris*. (A) Types, quantities, and proportions of transcription factors in *C.militaris*. The x-axis represents the number of transcription factors, and the y-axis represents the transcription factor families. (B) The sankey diagram for the network of key transcription factors (left rectangle), transcription factor families (middle) and target genes (right). Each right rectangle represents a target gene, and the connection degree of each gene is visualized based on the size of the rectangle. The genes in red are the most significantly upregulated by transcription factors Klf4 and Rxra. (C) Cordycepin biosynthesis regulated by transcription factors Klf4 and Rxra in *Cordyceps militaris.* Red arrows means that the genes was up-regulated by transcription factors Klf4 and Rxra or Myb.

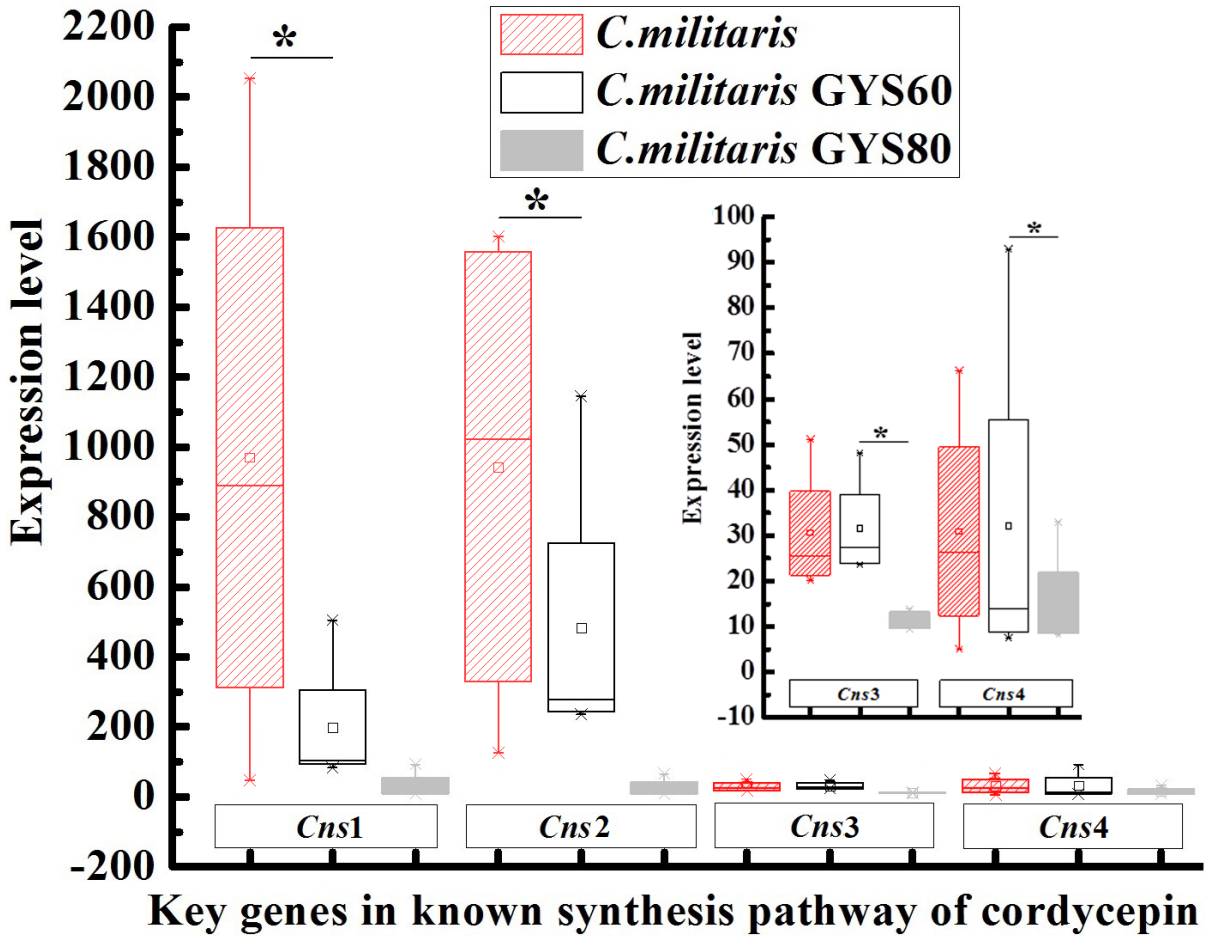

FIG. 7. Transcriptome analysis of four key synthetic genes of cordycepin. Asterisk (*)

indicates significance at $P<0.01$ compared with control *C. militaris*.

Note: *C. militaris* indicates wild type; GYS60 is high-cordycepin-producing; GYS80 is low-cordycepin-producing; *Cns1* encodes oxidoreductase/dehydrogenase; *Cns*2 encodes phosphohydrolases; *Cns*3 encodes a C-terminal phosphoribosyltransferase; *Cns*4 encodes ATP-binding cassette transporters.

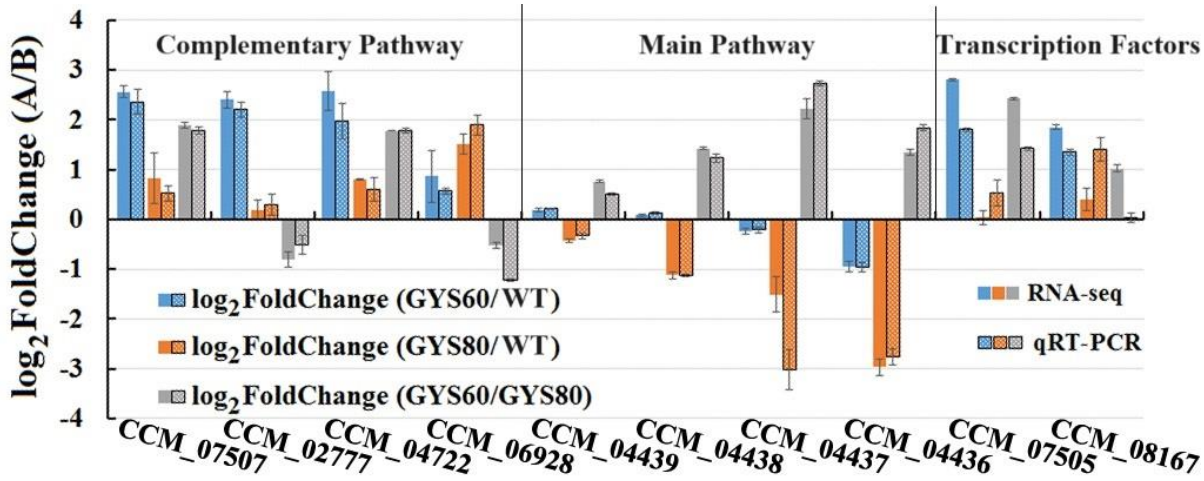

FIG. 8. Quantitative real-time PCR validation of DEGs. qRT-PCR was performed in biological triplicate. The smooth bars (RNA-Seq) represent the log_2fold change of transcriptome data; the rough bars (qRT-PCR) represent the log_2 regulation of qRT-PCR data. The blue, orange, and grey bars represent the log_2fold change (GYS60/*C.militaris*), the log_2fold change (GYS80/*C.militaris*), and the log_2fold change (GYS60/GYS80), respectively.