

Research Article

Transcriptome Analysis Reveals the Molecular Mechanisms Underlying Adenosine Biosynthesis in Anamorph Strain of Caterpillar Fungus

Shan Lin ^{1,2,3}, Zhicheng Zou,² Cuibing Zhou,² Hancheng Zhang,² and Zhiming Cai²

¹Department of Central Laboratory, Shenzhen Hospital, Beijing University of Chinese Medicine, Shenzhen University School of Medicine, Shenzhen 518100, Guangdong, China

²Institute of Translational Medicine, Shenzhen Second People's Hospital, First Affiliated Hospital of Shenzhen University, Shenzhen 518035, China

³Zhongshan School of Medicine, Shenzhen Second People's Hospital, Guangzhou 510080, China

Correspondence should be addressed to Shan Lin; biotechlin@foxmail.com

Received 25 January 2019; Accepted 28 July 2019; Published 11 December 2019

Academic Editor: Young-Mi Lee

Copyright © 2019 Shan Lin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Caterpillar fungus is a well-known fungal Chinese medicine. To reveal molecular changes during early and late stages of adenosine biosynthesis, transcriptome analysis was performed with the anamorph strain of caterpillar fungus. A total of 2,764 differentially expressed genes (DEGs) were identified ($p \leq 0.05$, $|\log_2 \text{Ratio}| \geq 1$), of which 1,737 were up-regulated and 1,027 were down-regulated. Gene expression profiling on 4–10 d revealed a distinct shift in expression of the purine metabolism pathway. Differential expression of 17 selected DEGs which involved in purine metabolism (map00230) were validated by qPCR, and the expression trends were consistent with the RNA-Seq results. Subsequently, the predicted adenosine biosynthesis pathway combined with qPCR and gene expression data of RNA-Seq indicated that the increased adenosine accumulation is a result of down-regulation of *ndk*, *ADK*, and *APRT* genes combined with up-regulation of *AK* gene. This study will be valuable for understanding the molecular mechanisms of the adenosine biosynthesis in caterpillar fungus.

1. Introduction

The Chinese caterpillar fungus, *Ophiocordyceps sinensis* (renamed from *Cordyceps sinensis*) [1], is one of the most outstandingly valued traditional Chinese medicinal fungi [2], and generally found on the Tibetan Plateau with high altitude ranges from 3,600 to 5,400 m [3]. Previous studies have revealed that caterpillar fungus and its anamorph possess a variety of biologically effective ingredients, such as purines, adenosine, polysaccharides, cordycepic acid, ergosterol, with extensive pharmacological effects [4, 5]. Adenosine is considered as an ancient extracellular signalling molecule, which could regulate almost all aspects of tissue function, and many previous studies have reported that adenosine plays a positive role of immunity, inflammation, and cancer [6, 7]. Due to the unique medicinal value, the natural sources of shortage and increasing demand, mycelia fermentation of *Cordyceps* fungal species has become a feasible and sustainable mean for producing the medicinal fungus and its bioactive

compounds [8, 9]. On the basis of morphological and molecular biology evidences, *Hirsutella sinensis* is currently considered as the only correct anamorph of caterpillar fungus [10, 11]. It was reported that mycelia of *H. sinensis* have similar clinical efficacy and less associated toxicity compared with wild caterpillar fungus, and they are expected to be substitutes of caterpillar fungus [12]. Therefore, with the increasing interests of caterpillar fungus both on mycology and pharmacology, it is becoming more and more urgent to investigate the hereditary information or functional genes of caterpillar fungus by omics sequencing technology.

RNA-Seq is considered as one of the most frequently used methods for transcriptome analysis and gene expression profiling [13], it has many advantages compared with other gene expression profiling technologies, such as allowing a comprehensive assay that does not require probes for targets to be specified in advance [14]. In recent years, the high-throughput RNA-Seq technique has emerged as a useful tool for transcriptome analysis and exploring unknown genes [15].

Gene expression profiles during secondary metabolism are complex. RNA-Seq has been applied to investigate active ingredient accumulation of several traditional Chinese medicine, such as *Ganoderma lucidum* [16], *Paris polyphylla* [17], and *Panax japonicas* [18]. The use of the RNA-Seq technique identified many DEGs that were associated with secondary metabolism. These studies have provided extensive insights into the understanding of significant genetic differences in secondary metabolism. However, the basic molecular mechanism of the active ingredient accumulation in mycelia of fungal medicines, particularly in the different stages of secondary metabolite accumulation, requires further study [19].

Previous studies have been investigated to identify DEGs [20–24]. High-throughput cDNA synthesis and sequencing of poly(A)-enriched RNA were applied to replace microarrays as a quantitative platform for measuring gene expression, and full length cDNA sequencing to 2-channel gene expression microarrays in the context of measuring differential gene expression, as well as sequencing data to a highly replicated microarray experiment profiling two divergent strains of *Saccharomyces cerevisiae* were compared [20]. The hippocampal expression profiles of wild-type mice and mice transgenic for delta-C-doublecortin-like kinase were compared with Solexa/Illumina deep sequencing technology and five different microarray platforms, approximately 2.4 million sequence tags per sample were obtained, and the changes in expression observed by deep sequencing were larger than observed by microarrays or quantitative PCR [21]. In addition, Robinson et al. [22] developed a bioconductor package (edgeR) for examining differential expression of replicated count data, and an over-dispersed Poisson model was conducted to account for both biological and technical variability. These previous studies indicated that it is viable to identify DEGs by different methods with different efficiencies, and developing an efficient method seems meaningful for identifying DEGs.

In the present study, to better understand the molecular factors and their regulatory genes involved in accumulation of active ingredients, the mRNA expression profiles of mycelia between early and late stages of secondary metabolite were compared. We gain insights into the purine accumulation mechanism of caterpillar fungus, particularly the expression of genes in adenosine biosynthesis, as cells transition from early into late stages of secondary metabolite. Physiological observations such as growth and adenosine biosynthesis were linked to transcriptional data obtained first by transcriptomic sequencing, followed by quantitative real-time PCR (qPCR). In addition, GO enrichment as well as KEGG pathway analyses showed that these DEGs were involved in cellular metabolic process, catalytic activity, and biosynthesis of secondary metabolites. These results provide novel insight into understanding the molecular mechanisms of adenosine accumulation and aid in understanding its biosynthesis pathways, and developing future studies on the metabolic regulation of caterpillar fungus.

2. Materials and Methods

2.1. Strains and Materials. A strain of anamorph of caterpillar fungus was isolated and deposited in our laboratory. Submerged

fermentation was performed at 16°C on a rotary shaker at 150 rpm, and mycelia were asexual reproduced and harvested for 10 days. The medium was consisted of tryptone (1%), powdered corn flour (1%), silkworm pupae (1.5%), glucose (1.5%), bran (1.5%), dextrin (0.5%), yeast extract (0.5%), KH_2PO_4 (0.02%), and MgSO_4 (0.01%). The early stage samples of secondary metabolite accumulation were collected for 4-day fermentation, and the late stage samples of secondary metabolite accumulation were collected for 10-day fermentation.

2.2. Determination of Mycelia Biomass and Purine Contents. Caterpillar fungus mycelia were cultured at 16°C and collected after cultivation in a shake flask. After washing three times with ultrapure water, the mycelia were dried at 60°C to a constant weight. Assay of purine contents were carried out by HPLC according to a reported procedure with some modifications [25], the column temperature was maintained at 35°C. The standards or samples were separated using a gradient mobile phase consisting of methyl alcohol (A) and ultrapure water (B). The gradient condition is: 0–3 min, 15% A; 3.0–3.5 min, 15–25% A; 3.5–8.55 min, 24% A; 8.5–9.0 min, 24–35% A; 9.0–15.0 min, 35% A; 15.0–16.0 min, 35–85% A; 16.0–22.0 min, 85% A; 22.0–22.5 min, 85–15% A, and 22.5–27.5 min, 15% A. The column was cleaned by 100% methyl alcohol for every 10 runs. The flow rate was set at 1.0 mL/min. The peaks were detected at 260 nm and identified by comparing the retention times with the standard. Standard curves were prepared and the linear regression equation was obtained. The percentage purine extraction yield (mg/g) was calculated as the purine content of extraction divided by dried sample weight.

2.3. RNA Isolation, Library Construction, and Sequencing. Total RNA of the anamorph of caterpillar fungus was extracted using TRIZOL and treated with RNase-free DNase I (TaKaRa) according to the manufacturer's protocols (Invitrogen, CA, USA). The mRNA was isolated from total RNA using Promega PolyATtract mRNA Isolation Systems; beads with oligo(dT) were used to isolate poly(A) mRNA. Subsequently, random hexamer-primers were applied to synthesize the first-strand cDNA taking these short fragments as templates, and the second-strand cDNA was synthesized. Sequencing libraries were generated by NEBNext1 Ultra RNA Library Prep Kit for Illumina (NEB, MA, USA). Short fragments were purified with QiaQuick PCR extraction kit and resolved with EB buffer. And then the short fragments were connected with sequencing adapters with respect to the result of agarose gel electrophoresis, and suitable fragments were selected as templates for amplification with PCR. Finally, the library was sequenced using Illumina HiSeq™ 4,000 (Illumina, CA, USA).

2.4. Analysis of RNA-Seq Data. The following criteria were applied to remove the sequences: more than 10% unknown nucleotides (N) reads and adapter, and low quality sequence (more than 30% of <Q20 bases). Subsequently, all the clean reads were mapped to the genome by HISAT software. Transcripts assembly was developed by Trinity (<https://github.com/trinityrnaseq/trinityrnaseq/wiki>) [26]. Then, the assembly results were optimally filtered by TransRate (

hibberdlab.com/transrate/) [27] and re-evaluated by BUSCO (Benchmarking Universal Single-Copy Orthologs, <http://busco.ezlab.org>) [28].

The gene expression level was calculated by the normalized number of fragments per kb per million reads (FPKM) method [29]. The DEGs between the early and late stages were identified by RSEM (RNA-Seq by Expectation-Maximization) software by the following filter criteria: p -value ≤ 0.05 as well as absolute value of \log_2 (FPKM_early/FPKM_late) ≥ 1 [30].

2.5. Gene Ontology and KEGG Pathway Enrichment Analysis. DEGs were annotated by GO database (<http://www.geneontology.org/>) using hypergeometric test to examine the biological functions and pathways of these genes. GO functional enrichment analysis provides GO terms which are significantly enriched in DEGs compared to the genome background, showing which DEGs are connected to the wanted biological functions. The analysis firstly maps all DEGs to GO terms in the database (<http://www.geneontology.org/>), calculating gene numbers for every term, then using ultra-geometric test to find significantly enriched GO terms in DEGs comparing to the genome background [31]. The calculating formula is shown as follows:

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{N-M}{n-i} \binom{M}{i}}{\binom{N}{n}}, \quad (1)$$

where N is the number of all genes with GO annotation, n is the number of DEGs in N , M is the number of all genes that are annotated to the certain GO terms, m is the number of DEGs in M . The calculated p -value went through Bonferroni Correction, taking corrected- p -value ≤ 0.05 as a threshold. GO terms fulfilling this condition were defined as significantly enriched GO terms in DEGs.

KEGG is the major public pathway-related database of biological systems that integrates genomic, chemical, and systemic functional information [32]. KEGG pathway analyses were performed by the KEGG database (<http://www.genome.jp/kegg/>), the calculating formula is shown as formula (1). Those with a p value < 0.05 were considered the significant pathways.

2.6. qPCR for Verifying DEGs. Verification of RNA-Seq data was performed by qPCR, and $2^{-\Delta\Delta Ct}$ method was conducted to calculate the relative expression levels by comparing the cycle thresholds (CTs) of the target genes with that of the 18S rRNA gene. 17 candidate genes which involved in purine metabolism were selected and validated by qPCR. Differences in relative transcript expression levels were compared at $p < 0.05$ level among different secondary metabolite accumulation periods (early-VS-late) using the Student's T -test. Primer pairs of the candidate genes were designed by Primer Express tool (Applied Biosystems, Foster City, USA), and 18S rRNA gene was selected as the internal control (Supplementary Table S1).

qPCR mixture (10 μ L) was prepared and consisted of 1 μ L of cDNA from early and late stage samples, respectively, 5 μ L of SYBR Green PCR Master Mix (2 \times) (Promega, Wisconsin, USA), and 0.5 μ L (100 μ mol/L) of each forward and reverse

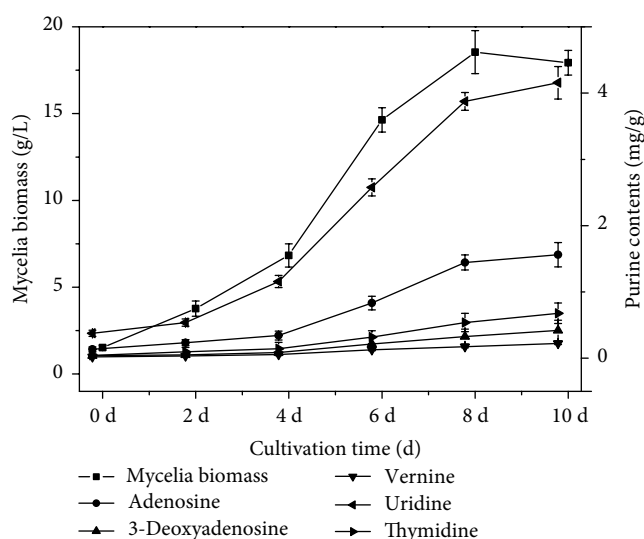


FIGURE 1: Time course of purine accumulation. Mycelia biomass and purine contents both sharply increased from the early stage (4 d) of secondary metabolite accumulation to and late stage (10 d).

primer. qPCR analyses were performed three times with independent RNA samples according to the temperature-time profile as follows: denaturation of 95°C for 2 min, 40 cycles of 95°C for 15 sec, 60°C for 1 min.

2.7. Statistical Analysis. All experiments in this study were performed in triplicate if not specifically noted. The experimental data were analyzed by the statistical software SPSS (version 9.0, IBM, Chicago). Student's T -test and the analysis of variance (ANOVA) test were performed ($p < 0.05$).

3. Results and Discussion

3.1. Growth Characteristics and Purine Accumulation. Caterpillar fungus grows slowly by artificial culture under suitable conditions, and people attempt to cultivate this fungus for producing its fruiting bodies have frequently failed [33, 34]. To meet the requirement of market, submerged cultivation of caterpillar fungus mycelia provided an environmental-friendly way to resolve this demand [35]. The volatile compound profiles from caterpillar fungus mycelia by submerged cultivation were extracted, and many kinds of active ingredients in the mycelia were observed more abundant than that those in wild caterpillar fungus, indicating submerged cultivation of caterpillar fungus has the trend of gradually replacing the position of caterpillar fungus in market [12].

In this study, *H. sinensis* was subjected to growth under optimal culture conditions, and the identified optimal culture conditions were adopted to perform the dynamic profiles of cell growth and purine production. As shown in Figure 1, the mycelia biomass slightly increased until 4 d with 6.83 g/L, and significantly increased to 18.54 g/L at 8 d, then maintained a relatively stable level until 10 d. Furthermore, adenosine and 3-deoxyadenosine production slightly increased until 4 d with 0.342 mg/g and 0.086 mg/g, respectively, and significantly increased to 1.562 mg/g and 0.419 mg/g at 10 d, respectively. Moreover,

TABLE 1: Summary of RNA-Seq data statistics and annotation information.

Samples	Early_1	Early_2	Early_3	Late_1	Late_2	Late_3
Raw reads number	54,028,688	50,541,530	51,518,828	50,358,610	49,729,670	51,270,900
Raw bases	8,158,331,888	7,631,771,030	7,779,343,028	7,604,150,110	7,509,180,170	7,741,905,900
Clean reads number	53,390,400	48,871,682	50,822,684	49,823,872	49,082,804	50,458,096
Clean bases	7,955,067,632	7,261,806,751	7,576,879,005	7,446,100,035	7,324,579,677	7,530,275,238
Clean rate (%)	98.82	96.70	98.65	98.94	98.70	98.41
Q20 (%)	98.72	98.69	98.65	98.72	98.67	98.63
Q30 (%)	95.92	95.88	95.75	95.9	95.8	95.68
GC content (%)	59.75	59.97	60.04	59.79	59.93	59.91
Filtered clean reads	26,695,200	24,435,841	25,411,342	24,911,936	24,541,402	25,229,048
Mapped reads	23,511,904	21,470,372	22,389,649	21,661,602	21,354,130	21,818,903
Mapped ratio	88.08%	87.86%	88.11%	86.95%	87.01%	86.48%

uridine, vernine, and thymidine production also slightly increased until 4 d and then sharply increased until 10 d. The dynamic profiles of mycelia biomass and purine contents were similar, which all have the same tendency of significant increase after 4 d cultivation (Figure 1). Therefore, the dynamic profiles of cell growth and purine production support our sampling time for early stage of secondary metabolite accumulation at 4 d fermentation, as well as late stage at 10 d fermentation.

A previous study set the sampling time of mycelia at 3 d for growth period of secondary metabolite accumulation and 9 d for stable period, which was earlier than that in our study [31]. However, the present results showed that secondary metabolite increased dramatically from 4 d and maintained a steady level at the end of fermentation until 10 d. Furthermore, a fungal strain UM01 isolated from natural *C. sinensis* was sampled after inoculating for 5 days, while it lacked the investigation on time course of mycelia biomass or production [36]. Therefore, it is not difficult to perceive that 4 d and 10 d were better sampling time points for early and late stages, respectively.

3.2. Summary of RNA-Seq Data for Transcriptome Analysis. Secondary metabolite accumulation is crucial for active ingredients and pesticide effects of traditional Chinese medicine. It is well-known that RNA-Seq technique is a powerful approach for transcriptome analysis and exploring unknown genes [37]. Currently, the RNA-Seq technique has been performed in various fungus medicines, including caterpillar fungus, *G. lucidum* [16] and *Cordyceps militaris* [38]. The overriding aim of these studies is to elucidate transcriptome profile changes caused by metabolite accumulation, through comparing results from early and late stages of secondary metabolite accumulation.

In this study, six cDNA libraries from two groups (three from early stage, and three from late stage) were constructed and sequenced. The major characteristics of the sequencing and annotation data are described in Table 1. Subsequently, more than 48 million clean reads for six libraries were obtained after low quality and adaptor sequences were filtered out. Among these clean reads, more than 98.63% and 95.75% had quality scores at the ratio of Q20 and Q30 level, respectively. Moreover, there were 86.48–88.11% of the clean reads mapped onto the reference genome. As shown in Table 2, a total of

TABLE 2: The evaluation of assembly result.

Source	Original assembly	Optimized assembly
Total transcripts num	68,661	52,923
Total unigenes num	42,289	39,336
Total sequence base	125,951,552	67,697,788
Largest	20,659	20,280
Smallest	201	201
Average length	1,834.40	1,279.18
N50	5,014	3,524
E90N50	3,490	3,914
GC percent	58.44	57.56
Mean mapped reads	1,914.28125609	2,699.98525834
TransRate score	0.1929	0.40042
BUSCO score	92.7% (15.5%)	93.4% (15.5%)

68,661 original assembly transcripts and 52,923 optimized assembly transcripts were obtained from the six libraries, and the average transcript length was approximately 1,834.40 for original assembly and 1,279.18 for optimized assembly, respectively. Furthermore, the TransRate score for original assembly and optimized assembly was 0.1929 and 0.40042, respectively. The BUSCO score for original assembly and optimized assembly was 92.7% (15.5%) and 93.4% (15.5%), respectively. Moreover, all transcripts obtained by this RNA-Seq were compared with six databases (NR, swiss-prot, Pfam, COG, GO, and KEGG databases) for functional annotation, and the Venn diagram of functional annotation of transcripts is shown (Supplementary Figure S1). The annotation results showed that number of common comments to the six database was 3,744, and number of unique comments to NR, swiss-prot, Pfam, COG, GO, and KEGG were 3,099, 599, 386, 10, 0, and 137, respectively.

Several studies have reported the transcriptome analysis of entomogenous fungi, and conducted the functional annotation of transcripts. Characterization of the *O. sinensis* transcriptome among three stages of the life cycle was investigated, and a total of 14,922 unigenes were identified and categorized under three gene ontology categories, which were obviously less than those in our study [36]. A previous study detected and analyzed the DEGs of *H. sinensis* growing during

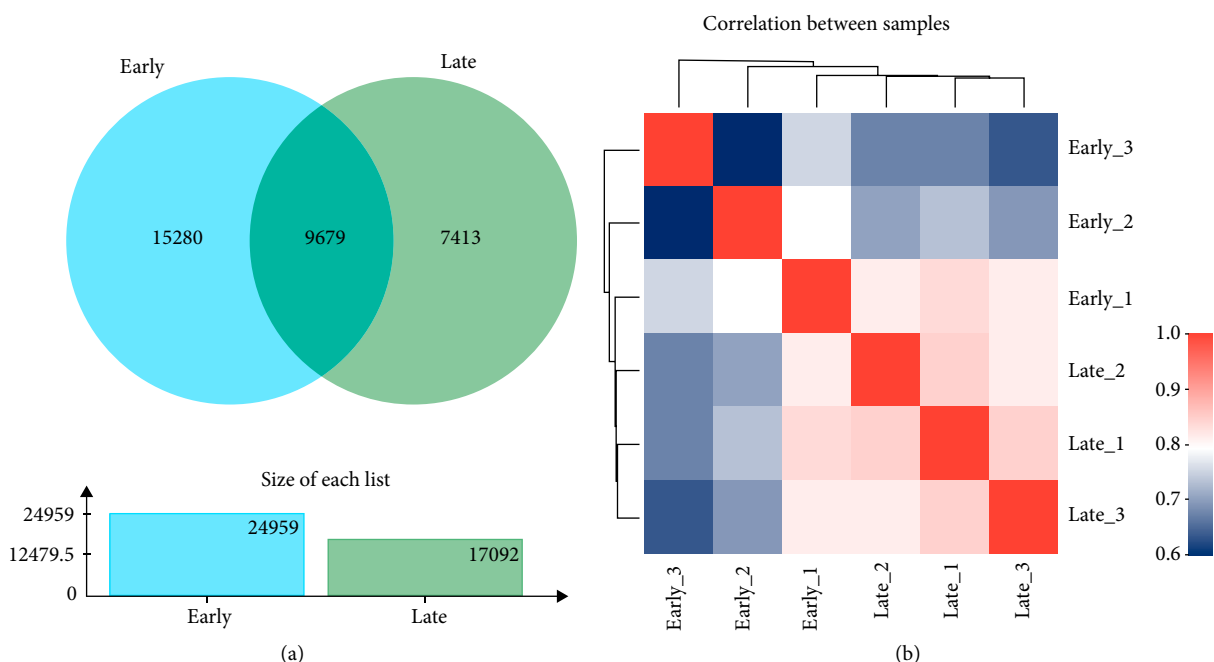


FIGURE 2: Expression analysis between early and late samples based on the expression matrix. (a) Venn analysis between early and late samples, the intersecting areas of the circles represent the number of unigene/transcript common to each group, inter-sample Venn analysis can obtain co-expression and specific expression genes/transcripts between samples and groups. (b) Inter-sample correlation analysis, different colors represent the size of the correlation coefficient between samples. Correlation analysis helps to understand the correlation between samples, especially among biological duplicates.

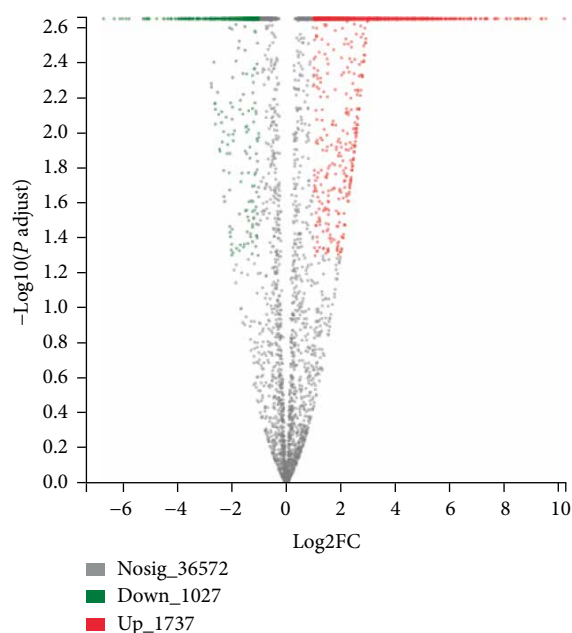


FIGURE 3: Volcano plot of differentially expressed genes. The abscissa is the multiple change value of gene expression difference between two samples. The ordinate is the statistical test value of the variation of gene amount. The red dots indicate significantly upregulated unigene/transcript, the green dots indicate significantly downregulated unigene/transcript, the black dots indicate no significantly expressed unigene/transcript.

different days, while each of the growth and development stages did not be compared [31]. In another study, the transcriptome of the medicinal *O. sinensis* fruiting body was

analyzed and a total of 34,289 unigenes were obtained, but the unigenes involved in growth and development stages were not analyzed [39]. In this study, the DEGs from early and late stages were identified and analyzed, which would be useful for the further study of secondary metabolite accumulation.

3.3. Expression Analysis and Identification of DEGs. Because of the importance of RNA-Seq, a lot of methods have been conducted to analyze RNA-Seq data for identification of DEGs in recent years, including edgeR [22], bay_Seq [40], DE_Seq [41], and NBP_Seq [42]. The majority of these methods are based on Poisson or negative binomial distributions when they are dealing with RNA-Seq count data [43]. However, FPKM is the most frequently used measure of mRNA abundance based on RNA-Seq data [44], it is calculated from the number of fragments mapped to a particular gene region with a feature length, which is the number of nucleotides in a capable region of a gene [24].

In this study, a FPKM method for identification of DEGs with RNA-Seq data was developed, and the FPKM of each unigene/transcript for early and late samples was calculated, as well as the value of $\log_2(\text{late_FPKM}/\text{early_FPKM})$, p -value of statistical test and FDR were calculated, if the value of $\log_2(\text{late_FPKM}/\text{early_FPKM}) > 0$, then this gene is up-regulated at late stage, or else it is down. Overall distribution of unigene/transcript expression in each sample is shown (Supplementary Figure S2). The box plot of expression distribution presented that each box graph corresponds to five expression statistics (maximum, upper quartile, median, lower quartile, and minimum). Meanwhile, the violin plot of expression distribution

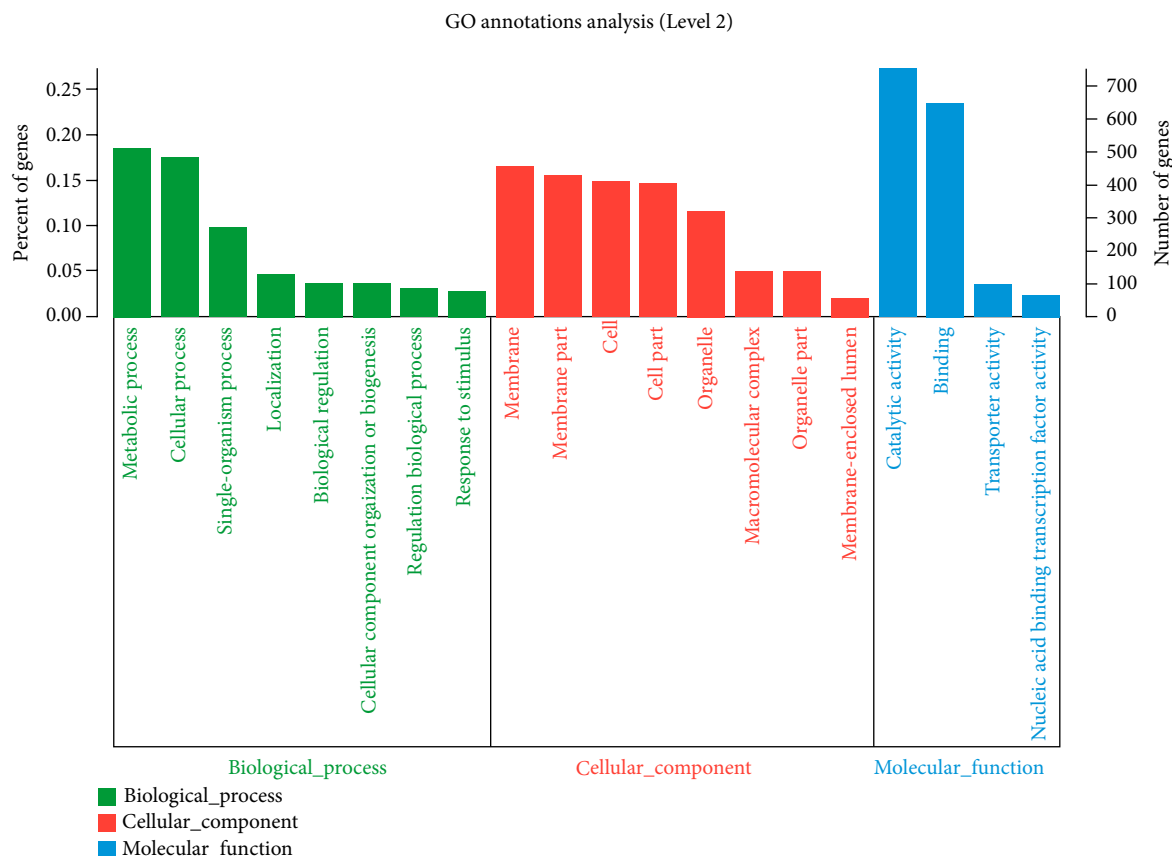


FIGURE 4: GO classification statistics of DEGs. The abscissa in the figure represents the secondary classification of GO. The vertical axis represents the percentage of the total unigene/transcript contained in this secondary classification.

presented that enlarged portion of the image represents the region with the highest concentration of unigene/transcript expression in the sample.

Based on the expression matrix, Venn and correlation between samples were analyzed. As shown in Figure 2(a), there were 9,679 common elements between early and late stages, indicating co-expression and specific expression genes/transcripts between samples could be obtained by inter-sample Venn analysis shown. As shown in Figure 2(b), correlation analysis helps to understand the correlation between samples, especially among biological duplicates. Among the six cDNA libraries, a total of 39,336 genes were detected and the FPKM method was utilized to evaluate the gene expression level (Figure 3). In order to analyze the transcriptome differences between early and late stages of secondary metabolite accumulation, the late stage was compared to the early stage. A total of 2,764 significant DEGs were identified, of which 1,737 genes were up-regulated and 1,027 genes were down-regulated (p value ≤ 0.05 and $|\log_2 FC| \geq 1$).

3.4. Gene Ontology Functional Annotation and Enrichment of DEGs. GO is an international standardized gene functional classification system which offers a dynamic-updated controlled vocabulary and a strictly defined concept to comprehensively describe properties of genes and their products in any organism [30]. In order to offer a dynamic-

updated controlled vocabulary and a strictly defined concept to comprehensively describe properties of DEGs and their products in any organism, GO functional classification annotation, which includes three ontologies (molecular function, cellular component and biological process) were conducted [45].

As shown in Figure 4, in the GO category of biological process, DEGs were involved in metabolic process (513 DEGs), cellular process (485 DEGs), single-organism process (268 DEGs), localization (127 DEGs), biological regulation (99 DEGs), cellular component organization or biogenesis (98 DEGs), regulation of biological process (84 DEGs), and response to stimulus (75 DEGs). Among the DEGs related to the biological process, the most significant term was metabolic process, indicating that metabolic process was extremely active during secondary metabolite accumulation. In the GO category of molecular functions, DEGs were involved in membrane (452 DEGs), membrane part (430 DEGs), cell (408 DEGs), cell part (404 DEGs), organelle (317 DEGs), macromolecular complex (135 DEGs), organelle part (135 DEGs), and membrane-enclosed lumen (55 DEGs). It was indicated that the enriched terms were potentially associated with the secondary metabolite accumulation, and the most significant term located in the membrane played the most important role during secondary metabolite accumulation. In the GO category of molecular functions, DEGs were involved

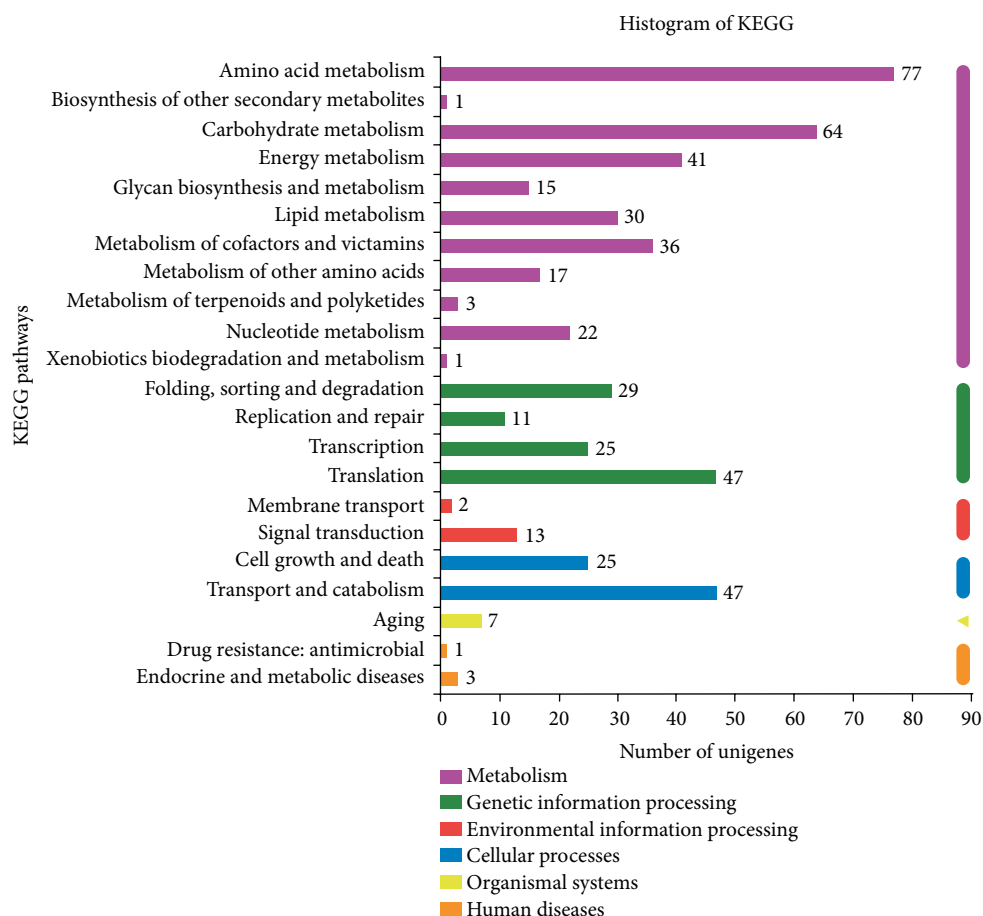


FIGURE 5: Histogram of KEGG pathway classification statistics. The ordinate is the name of KEGG metabolic pathway, the abscissa is the number of transcript/unigene annotated into the pathway.

in catalytic activity (754 DEGs), binding (646 DEGs), transporter activity (96 DEGs), and nucleic acid binding transcription factor activity (63 DEGs). The major molecular function category was catalytic activity, indicating that a large number of enzymes were involved in the synthesis of secondary metabolites.

Furthermore, GO functional enrichment of genes/transcripts in gene concentration was analyzed by Fisher's exact test. When the adjusted p value (p adjust) was <0.05 , this GO function was considered to be significantly enriched. As shown in Supplementary Figure S3, DNA integration, peptide biosynthetic process, cellular protein metabolic process, and biosynthetic process were significantly enriched. Among them, DNA integration had the maximal rich factor, indicating that DNA integration played an important role in cell growth and secondary metabolite accumulation.

3.5. KEGG Pathway Annotation and Enrichment of DEGs. Different genes usually cooperate with each other to exercise their biological functions, and pathway-based analysis helps to further understand genes biological functions [46]. KEGG is the major public pathway-related database. Pathway enrichment analysis could significantly enrich metabolic pathways or signal transduction pathways in DEGs compared with the

whole genome background [43]. A KEGG pathway analysis was developed to identify the pathways of the DEGs involved in secondary metabolite accumulation. Among the 2,764 late-VS-early DEGs, there were 636 DEGs with pathway annotation, and the results of main pathway enrichment analysis of late-VS-early DEGs are shown in Figure 5. Totally, 636 DEGs were mapped to 104 KEGG pathways, and 22 pathways were significantly enriched ($p \leq 0.05$). In the significant pathways, several main pathways were represented, including amino acid metabolism, carbohydrate metabolism, energy metabolism, metabolism of cofactors and vitamins, translation, transport, and catabolism.

KEGG pathway enrichment analysis was carried out on genes/transcripts in gene concentration by Fisher's exact test. When the adjusted p value (p adjust) was <0.05 , this KEGG function was considered to be significantly enriched. As shown in Supplementary Figure S4, cell cycle-yeast, phenylalanine metabolism, tyrosine metabolism, glycerophospholipid metabolism, pantothenate and CoA biosynthesis, and various types of N -glycan biosynthesis were significantly enriched. Among them, cell cycle-yeast had the minimum p value, indicating that cell cycle-yeast was most significantly enriched and played the most important role in cell growth and secondary metabolite accumulation.

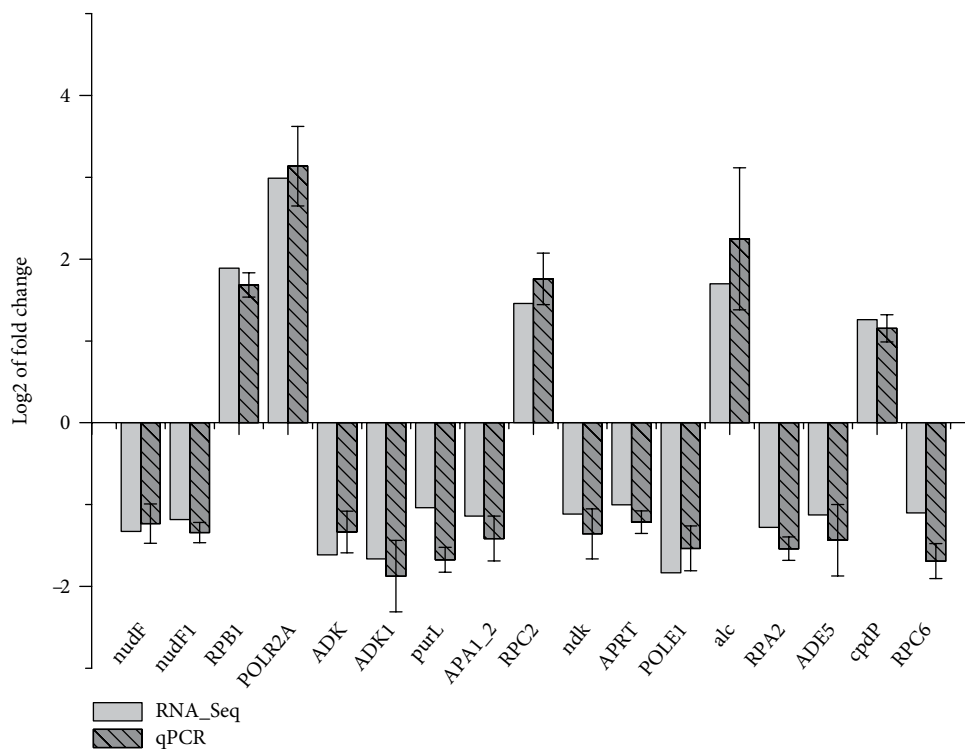


FIGURE 6: Validation of DEGs which involved in purine metabolism by qPCR. Light gray columns represent the expression level of the DEGs obtained by RNA-Seq, and the dark gray columns represent the qPCR results.

3.6. Verification of DEGs Involved in Purine Metabolism. Differential expression analysis was frequently conducted to screen DEGs, and qPCR was commonly applied in relative expression levels analysis for decades to verify DEGs [47–49]. Gene annotation and differential expression analysis by qPCR identified 464 transcripts that may be involved in catabolism and metabolism of phytohormone, and relative expression levels analysis showed that eleven phytohormone-related genes have different expression patterns in the seed stratification process of *P. polyphylla* [47]. Two proteases which are known to be directly involved in the process of pathogenesis in entomopathogenic fungi *Beauveria bassiana* were identified through a comparative analysis of gene expression patterns and verified them by qPCR [48]. The differential expression of thirteen PHB accumulation related genes was investigated by qPCR, indicating thirteen most up-regulated genes played important roles in PHB metabolism in *Acidiphilium cryptum* [49]. These previous studies suggested that qPCR is a reliable way to verify the DEGs involved in metabolic pathway.

In this study, verification of the selected DEGs was conducted by qPCR, and the results are shown in Figure 6. The 17 candidate genes which involved in purine metabolism (map00230), including ADP-ribose pyrophosphatase (*nudF*, *nudF1*), RNA polymerase (*RPB1*, *POLR2A*), adenosine kinase (*ADK*, *ADK1*), phosphoribosylformylglycinamide synthase (*purL*), ATP adenylyltransferase (*APA1_2*), RNA polymerase III subunit *RPC2* (*RPC2*), nucleoside-diphosphate kinase (*ndk*), adenine phosphoribosyltransferase (*APRT*), DNA polymerase epsilon subunit 1 (*POLE*), allantoinase (*alc*), RNA polymerase I subunit *RPA2* (*RPA2*), phosphoribosylamine-glycine ligase

(*ADE5*), 3',5'-cyclic-nucleotide phosphodiesterase (*cpdP*) and RNA polymerase III subunit *RPC6* (*RPC6*), were validated by qPCR. Among them, there were 5 up-regulated DEGs, including *RPB1* (3.70-fold), *POLR2A* (7.94-fold), *RPC2* (2.75-fold), *alc* (3.25-fold) and *cpdP* (2.39-fold), and 12 down-regulated DEGs, including *nudF* (0.40-fold), *nudF1* (0.44-fold), *ADK* (0.33-fold), *ADK1* (0.32-fold), *purL* (0.49-fold), *APA1_2* (0.45-fold), *ndk* (0.46-fold), *APRT* (0.50-fold), *POLE1* (0.28-fold), *RPA2* (0.41-fold), *ADE5* (0.46-fold), and *RPC6* (0.47-fold). Although the fold change varied between the two methods, the trends in the expression of the 17 genes were consistent with the RNA-Seq results, suggesting that the RNA-Seq results were reliable.

Indeed, there were few studies that focused on the purine metabolism in the anamorph of caterpillar fungus. The purine metabolic pathway in *C. militaris* was constructed based on the KEGG annotations, and the genes putatively involved in purine metabolism were obtained, while the verification of the putative genes was not conducted [50]. In addition, purine biosynthesis pathway of *H. sinensis* was predicted, which starts from adenosine and ends with urate after 7 steps of catalysis, but the pathway was not systematic and the DEGs involved in purine biosynthesis were not analyzed [31]. In the present study, purine metabolism was more systematically analyzed, and the DEGs involved in purine metabolism were investigated and validated, which could provide useful information for further metabolic regulation.

3.7. Construction of Adenosine Metabolic Pathway. Furthermore, based on KEGG purine metabolism (map00230), the predicted adenosine metabolic pathway and gene expression profiles in

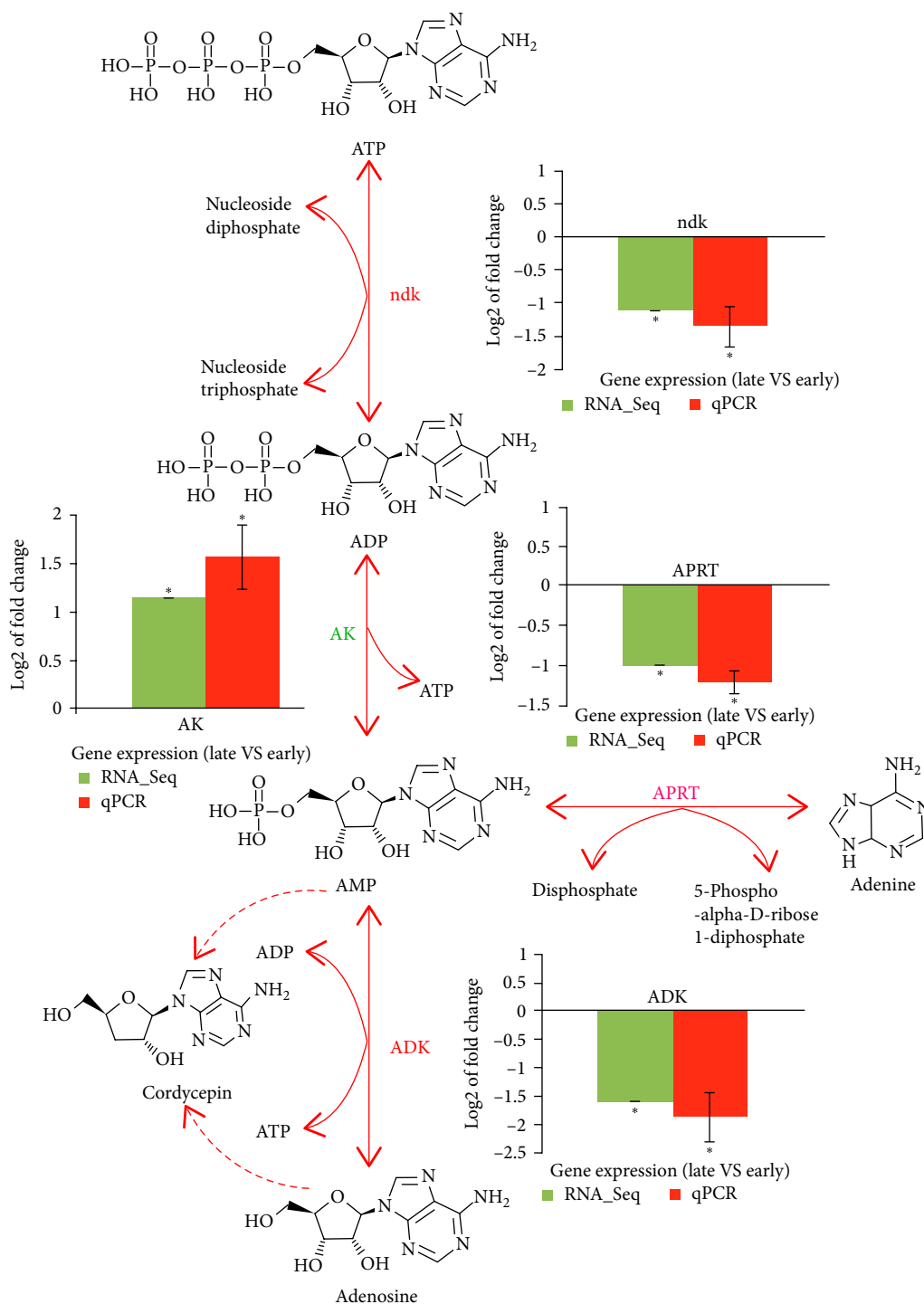


FIGURE 7: The predicted adenosine biosynthesis pathway and gene expression profiles in caterpillar fungus. Graphs show RNA-Seq and qPCR expression analysis of DEGs at late stage compared with early stage of secondary metabolites biosynthesis. Genes in red font were down-regulated and in green font were up-regulated, with asterisks (*) indicating significant differences ($p < 0.05$).

caterpillar fungus were conducted according to the results of annotation and expression analysis. RNA-Seq and qPCR expression analysis of DEGs at late stage compared with early stage of secondary metabolites biosynthesis were carried out. In adenosine biosynthesis pathway (Figure 7), ATP is converted to ADP by *ndk*, and ADP is converted to AMP by *AK*, AMP as an important intermediate in purine metabolism

pathway could be converted to adenosine, adenine, and cordycepin, respectively. Among these key genes involved in this pathway, *ndk*, *ADK* and, *APRT* genes were significantly down-regulated, while *AK* gene was significantly up-regulated, which indicated that *ndk*, *ADK*, and *APRT* genes play roles of retro-regulation, while *AK* gene plays the role of positive regulation in the process of adenosine accumulation.

Previously, Zheng et al. constructed metabolic pathways of purine and adenosine based on the KEGG annotation of *C. militaris* to model the cordycepin biosynthesis, and found that adenosine kinase was an important enzyme involved in cordycepin biosynthesis [50]. In addition, a putative biosynthetic pathway was proposed for adenosine in *O. sinensis*, indicating adenosine kinase, adenylate kinase, and 5'-nucleotidase participated in phosphorylation and dephosphorylation in adenosine metabolic pathway [39]. The above previous studies supported the result that *AK* gene plays the role of positive regulation for adenosine biosynthesis, while our results elucidated more detail for the role of other enzymes in the process of adenosine biosynthesis.

4. Conclusion

This study provides comprehensive transcriptome data on the early and late stages of secondary metabolite accumulation through RNA-Seq technology. Transcriptome analysis indicated that 2,764 DEGs participated in secondary metabolic synthesis, and 17 of them enhanced purine biosynthesis in caterpillar fungus. The predicted adenosine biosynthesis pathway combined with real-time PCR and RNA-Seq gene expression data indicated that the increased adenosine accumulation is a result of down-regulated of *ndk*, *ADK*, and *APRT* genes combined with up-regulated *AK* gene. This study provides useful information for understanding the molecular mechanisms of accumulation of adenosine, and guiding the metabolic regulation of caterpillar fungus.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by National Natural Science Foundation of China (81803652), Natural Science Foundation of Guangdong Province (2018A0303100007), Foundation of Health and Family Planning Commission (SZBC2018016), China Postdoctoral Science Foundation (2018M633261).

Supplementary Materials

Supplementary 1. Table S1: The primers used for real-time PCR. Supplementary 2. Figure S1: Venn diagram of functional annotation of transcripts. Supplementary 3. Figure S2: Overall distribution of unigene/transcript expression in each sample. Supplementary 4. Figure S3: Bubble diagram of GO enrichment analysis. Supplementary 5. Figure S4: Bubble diagram of KEGG enrichment analysis. (*Supplementary Materials*)

References

- [1] Y. Zhang, E. Li, C. Wang, Y. Li, and X. Liu, "Ophiocordyceps sinensis, the flagship fungus of China: terminology, life strategy and ecology," *Mycology*, vol. 3, no. 1, pp. 2–10, 2012.
- [2] E. H. Xia, D. R. Yang, J. J. Jiang et al., "The caterpillar fungus, *Ophiocordyceps sinensis*, genome provides insights into highland adaptation of fungal pathogenicity," *Scientific Reports*, vol. 7, no. 1, p. 1806, 2017.
- [3] C.-H. Dong and Y.-J. Yao, "Nutritional requirements of mycelial growth of *Cordyceps sinensis* in submerged culture," *Journal of Applied Microbiology*, vol. 99, no. 3, pp. 483–492, 2005.
- [4] M. Singh, R. Tulsawani, P. Koganti, A. Chauhan, M. Manickam, and K. Misra, "Cordyceps sinensis increases hypoxia tolerance by inducing heme oxygenase-1 and metallothionein via Nrf2 activation in human lung epithelial cells," *BioMed Research International*, vol. 2013, pp. 1–13, 2013.
- [5] S. Cha, J. Lim, C. Yoon, J. Koh, and H. Chang, S. Kim, "Production of mycelia and exo-biopolymer from molasses by *Cordyceps sinensis* 16 in submerged culture," *Bioresource Technology*, vol. 98, no. 1, pp. 165–168, 2007.
- [6] L. Antonioli, C. Blandizzi, P. Pacher, and G. Haskó, "Immunity, inflammation and cancer: a leading role for adenosine," *Nature Reviews Cancer*, vol. 13, no. 12, pp. 842–857, 2013.
- [7] D. Vijayan, A. Young, M. W. L. Teng, and M. J. Smyth, "Targeting immunosuppressive adenosine in cancer," *Nature Reviews Cancer*, vol. 17, no. 12, pp. 709–724, 2017.
- [8] Y. Dong, T. Jing, Q. Meng et al., "Studies on the antidiabetic activities of *Cordyceps militaris* extract in diet-streptozotocin-induced diabetic Sprague-Dawley rats," *BioMed Research International*, vol. 2014, pp. 1–11, 2014.
- [9] X. Zhao, X. Yu, G. Zhang et al., "Aqueous extracts of *Cordyceps kyushuensis* kob induce apoptosis to exert anticancer activity," *BioMed Research International*, vol. 2018, pp. 1–8, 2018.
- [10] Y.-Q. Chen, N. Wang, L.-H. Qu, T.-H. Li, and W.-M. Zhang, "Determination of the anamorph of *Cordyceps sinensis* inferred from the analysis of the ribosomal DNA internal transcribed spacers and 5.8S rDNA," *Biochemical Systematics and Ecology*, vol. 29, no. 6, pp. 597–607, 2001.
- [11] Z.-Y. Liu, Y.-J. Yao, Z. Q. Liang, A.-Y. Liu, D. N. Pegler, and M. W. Chase, "Molecular evidence for the anamorph-teleomorph connection in *Cordyceps sinensis*," *Mycological Research*, vol. 105, no. 7, pp. 827–832, 2001.
- [12] S. J. Yu, Y. Zhang, and M. Z. Fan, "Analysis of volatile compounds of mycelia of *Hirsutella sinensis*, the anamorph of *Ophiocordyceps sinensis*," *Applied Mechanics and Materials*, vol. 140, pp. 253–257, 2012.
- [13] Y. Guo, S. Zhao, Q. Sheng et al., "RNAseq by total RNA library identifies additional RNAs compared to Poly(A) RNA library," *BioMed Research International*, vol. 2015, no. 6, pp. 1–9, 2015.
- [14] R. Gupta, I. Dewan, R. Bharti, and A. Bhattacharya, "Differential expression analysis for RNA-Seq data," *ISRN Bioinformatics*, vol. 2012, Article ID 817508, pp. 1–8, 2012.
- [15] Y. Jin, Y. Fang, M. Huang et al., "Combination of RNA sequencing and metabolite data to elucidate improved toxic compound tolerance and butanol fermentation of *Clostridium acetobutylicum* from wheat straw hydrolysate by supplying sodium sulfide," *Bioresource Technology*, vol. 198, pp. 77–86, 2015.
- [16] J. Shao, H. Chen, D. Yang et al., "Genome-wide identification and characterization of natural antisense transcripts by strand-

- specific RNA sequencing in *Ganoderma lucidum*,” *Scientific Reports*, vol. 7, no. 1, 2017.
- [17] L.-Z. Ling, S.-D. Zhang, F. Zhao et al., “Transcriptome-wide identification and prediction of miRNAs and their targets in *Paris polyphylla* var. *yunnanensis* by high-throughput sequencing analysis,” *International Journal of Molecular Sciences*, vol. 18, no. 1, p. 219, 2017.
- [18] A. Rai, M. Yamazaki, H. Takahashi et al., “RNA-seq transcriptome analysis of *Panax japonicus*, and its comparison with other panax species to identify potential genes involved in the saponins biosynthesis,” *Frontiers in Plant Science*, vol. 7, p. e0144, 2016.
- [19] S. Lin, Z.-Q. Liu, P. J. Baker et al., “Enhancement of cordyceps polysaccharide production via biosynthetic pathway analysis in *Hirsutella sinensis*,” *International Journal of Biological Macromolecules*, vol. 92, pp. 872–880, 2016.
- [20] S. Bloom, Z. Khan, L. Kruglyak, M. Singh, and A. A. Caudy, “Measuring differential gene expression by short read sequencing: quantitative comparison to 2-channel gene expression microarrays,” *BMC Genomics*, vol. 10, no. 1, p. 221, 2009.
- [21] P. A. C. T. Hoen, Y. Ariyurek, H. H. Thygesen et al., “Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms,” *Nucleic Acids Research*, vol. 36, no. 21, pp. e141–e141, 2008.
- [22] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, “edgeR: a bioconductor package for differential expression analysis of digital gene expression data,” *Bioinformatics*, vol. 26, no. 1, pp. 139–140, 2010.
- [23] S. Audic and J.-M. Claverie, “The significance of digital gene expression profiles,” *Genome Research*, vol. 7, no. 10, pp. 986–995, 1997.
- [24] A. Mortazavi, B. A. Williams, K. Mccue, L. Schaeffer, and B. Wold, “Mapping and quantifying mammalian transcriptomes by RNA-Seq,” *Nature Methods*, vol. 5, no. 7, pp. 621–628, 2008.
- [25] L. Deng, T.-Y. Zhou, L. Pi et al., “Optimization of microwave-assisted extraction of cordycepic acid and cordycepin from cultured *Cordyceps militaris* by response surface methodology,” *Asian Journal of Chemistry*, vol. 25, no. 14, pp. 8065–8071, 2013.
- [26] M. G. Grabherr, B. J. Haas, M. Yassour et al., “Full-length transcriptome assembly from RNA-Seq data without a reference genome,” *Nature Biotechnology*, vol. 29, no. 7, pp. 644–652, 2011.
- [27] R. Smith-Unna, C. Bournsnel, R. Patro, J. M. Hibberd, and S. Kelly, “TransRate: reference-free quality assessment of de novo transcriptome assemblies,” *Genome Research*, vol. 26, no. 8, pp. 1134–1144, 2016.
- [28] F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov, “BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs,” *Bioinformatics*, vol. 31, no. 19, pp. 3210–3212, 2015.
- [29] C. Trapnell, B. A. Williams, G. Pertea et al., “Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms,” *Nature Biotechnology*, vol. 28, no. 5, pp. 511–515, 2010.
- [30] C. Trapnell, A. Roberts, L. Goff et al., “Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks,” *Nature Protocols*, vol. 7, no. 3, pp. 562–578, 2012.
- [31] Z.-Q. Liu, S. Lin, P. Baker et al., “Transcriptome sequencing and analysis of the entomopathogenic fungus *Hirsutella sinensis* isolated from *Ophiocordyceps sinensis*,” *BMC Genomics*, vol. 16, no. 1, p. 106, 2015.
- [32] M. Kanehisa, M. Araki, S. Goto et al., “KEGG for linking genomes to life and the environment,” *Nucleic Acids Research*, vol. 36, no. 1, pp. D480–D484, 2008.
- [33] J. C. Holliday and M. P. Cleaver, “Medicinal value of the caterpillar fungi species of the genus *Cordyceps* (Fr.) link (ascomycetes). A review,” *International Journal of Medicinal Mushrooms*, vol. 10, no. 3, pp. 219–234, 2008.
- [34] R. Stone, “Improbable partners aim to bring biotechnology to a himalayan kingdom,” *Science*, vol. 327, no. 5968, pp. 940–941, 2010.
- [35] S. Lin, Z.-Q. Liu, Y.-P. Xue et al., “Biosynthetic pathway analysis for improving the cordycepin and cordycepic acid production in *Hirsutella sinensis*,” *Applied Biochemistry and Biotechnology*, vol. 179, no. 4, pp. 633–649, 2016.
- [36] X. Zhong, L. Gu, H. Wang et al., “Profile of *Ophiocordyceps sinensis* transcriptome and differentially expressed genes in three different mycelia, sclerotium and fruiting body developmental stages,” *Fungal Biology*, vol. 122, no. 10, pp. 943–951, 2018.
- [37] J. Cheng, K. Li, Y. Zhu, W. Yang, J. Zhou, and K. Cen, “Transcriptome sequencing and metabolic pathways of astaxanthin accumulated in *Haematococcus pluvialis* mutant under 15% CO₂,” *Bioresource Technology*, vol. 228, pp. 99–105, 2017.
- [38] Y. Wang, Z. Wang, C. Liu, S. Wang, and B. Huang, “Genome-wide analysis of DNA methylation in the sexual stage of the insect pathogenic fungus *Cordyceps militaris*,” *Fungal Biology*, vol. 119, no. 12, pp. 1246–1254, 2015.
- [39] L. Xiang, Y. Li, Y. Zhu et al., “Transcriptome analysis of the *Ophiocordyceps sinensis* fruiting body reveals putative genes involved in fruiting body development and cordycepin biosynthesis,” *Genomics*, vol. 103, no. 1, pp. 154–159, 2014.
- [40] T. J. Hardcastle and K. A. Kelly, “baySeq: empirical Bayesian methods for identifying differential expression in sequence count data,” *BMC Bioinformatics*, vol. 11, no. 1, 2010.
- [41] S. Anders and W. Huber, “Differential expression analysis for sequence count data,” *Genome Biology*, vol. 11, no. 10, p. R106, 2010.
- [42] P. Glaus, A. Honkela, and M. Rattray, “Identifying differentially expressed transcripts from RNA-seq data with biological variation,” *Bioinformatics*, vol. 28, no. 13, pp. 1721–1728, 2012.
- [43] D. K. Y. Lim, H. Schuhmann, S. R. Thomas-Hall et al., “RNA-Seq and metabolic flux analysis of *Tetraselmis* sp. M8 during nitrogen starvation reveals a two-stage lipid accumulation mechanism,” *Bioresource Technology*, vol. 244, pp. 1281–1293, 2017.
- [44] B. J. Haas, A. Papanicolaou, M. Yassour et al., “De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis,” *Nature Protocols*, vol. 8, no. 8, pp. 1494–1512, 2013.
- [45] B. Wang, J. Zhu, E. Pierson, D. Ramazzotti, and S. Batzoglou, “Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning,” *Nature Methods*, vol. 14, no. 4, pp. 414–416, 2017.
- [46] J. Liang, X. Wang, J. Shang et al., “Draft genome sequence of the mulberry tree *Morus notabilis*,” *Nature Communications*, vol. 4, no. 9, 2013.
- [47] J. Qi, N. Zheng, B. Zhang et al., “Mining genes involved in the stratification of *Paris polyphylla* seeds using high-throughput embryo transcriptome sequencing,” *BMC Genomics*, vol. 14, no. 1, p. 358, 2013.

- [48] A. A. K. Pathan, K. Uma Devi, H. Vogel, and A. Reineke, "Analysis of differential gene expression in the generalist entomopathogenic fungus *Beauveria bassiana* (Bals.) Vuillemin grown on different insect cuticular extracts and synthetic medium through cDNA-AFLPs," *Fungal Genetics and Biology*, vol. 44, no. 12, pp. 1231–1241, 2007.
- [49] A.-L. Xu, "Real-time PCR analysis of metabolic pathway of PHB in *Acidiphilium cryptum* DX1-1," *Journal of Microbiology and Biotechnology*, vol. 20, no. 1, pp. 71–77, 2010.
- [50] P. Zheng, Y. Xia, G. Xiao et al., "Genome sequence of the insect pathogenic fungus *Cordyceps militaris*, a valued traditional Chinese medicine," *Genome Biology*, vol. 12, no. 11, p. R116, 2011.