

Article

# A Comparative Transcriptome Analysis Reveals Physiological Maturation Properties of Mycelia in *Pleurotus tuoliensis*

Fang Du <sup>1</sup>, Nu Er zi ya·Ya Li mai mai ti <sup>2</sup>, Qingxiu Hu <sup>1,\*</sup>, Yajie Zou <sup>1</sup>, Dou Ye <sup>1</sup>, and Haijun Zhang <sup>1</sup>

<sup>1</sup> Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China

<sup>2</sup> Institute of Plant Protection, Xinjiang Academy of Agricultural Sciences/Key Laboratory of Integrated Pest Management on Crop in Northwestern Oasis, Urumqi 830091, China

\* Correspondence: Huqingxiu@caas.cn

Received: 22 August 2019; Accepted: 10 September 2019; Published: 11 September 2019

**Abstract:** *Pleurotus tuoliensis* is a precious edible fungus with extremely high nutritive and medicinal value. The cultivation period of *P. tuoliensis* is longer than those of other *Pleurotus* species, which is mainly due to a longer mycelium physiological maturation period (30–60 days). Currently, the molecular processes underlying physiological maturation of the mycelium remain unclear. We performed a comparative transcriptomic analysis of immature and mature mycelia using RNA-seq. De novo transcriptome assembly resulted in identification of 17,030 unigenes. 451 differentially expressed genes—including those encoding nucleoside diphosphate kinase (NDPK), glycoside hydrolase family proteins, exopolysaccharidase, and versatile peroxidases—were identified. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses revealed that nucleotide synthesis and energy metabolism are highly active during the physiological maturation of mycelia, and genes related to these pathways were significantly upregulated in mature mycelia. NDPK is predicted to be essential for mycelia maturation. Our findings contribute to a comprehensive understanding of mycelia maturation in a commercially important fungal species. Future efforts will focus on the function of NDPK and the mechanism by which it regulates mycelia maturation.

**Keywords:** *Pleurotus tuoliensis*; mycelium physiological maturation; comparative transcriptomic analysis; nucleoside diphosphate kinase

## 1. Introduction

*Pleurotus tuoliensis*, also known as Bailinggu, is one of the most commercially important species of edible mushrooms [1]. *P. tuoliensis* was initially found in Xinjiang province [2] and is now widely commercially cultivated in China, Japan, and Korea, due to its desirable nutritional and medicinal value [3].

In recent years, *P. tuoliensis* yield decreased year by year. In 2017, only 70,000 tons were produced, which represents a 77% decrease from 2012 (<http://hz.cefa.org.cn/2019/03/18/10498.html>). This decrease may be due to this fungus' biological characteristics, such as long cultivation period, a high proportion of deformed fruiting bodies, poor uniformity of fruiting bodies, and strain degeneration [4]. Together, these factors present challenges for mushroom farmers and thus lead to a decline in cultivation enthusiasm. Physiological maturation of *P. tuoliensis* mycelia is essential for its fruiting. This maturation takes 30–60 days and is responsible for the above-mentioned long cultivation period. In this physiological maturation stage, mycelia accumulate and metabolize

nutrients [5]. Cold stimulation of physiologically mature mycelia can result in improved consistency in budding of primordia in culture bottles and ensures a high degree of fruiting body uniformity [6]. Our previous study demonstrated that N-carbamoyl-L-aspartate (CA-asp) can be used as an indicator of physiological maturation of mycelia. A mycelial CA-asp content of 0.9–1.2 g/L is thought to indicate physiological maturation and induces primordia formation and fruiting body growth [7]. The identification of such an indicator makes *P. tuoliensis* cultivation more tractable and aids efforts to improve *P. tuoliensis* quality and yield. However, the molecular mechanisms underlying physiological maturation of *P. tuoliensis* mycelia remain undefined.

RNA-seq is a rapid and cost-effective technology for discovery of genes expressed in a given growth stage, especially in non-model species for which reference genome sequences are not available [8]. Many transcriptome studies have been performed to explore growth and development of mushroom species, including *Pleurotus eryngii* [9], *Cordyceps militaris* [10], and *Ganoderma lucidum* [11]. This high-throughput deep sequencing approach has also been successfully applied to investigate the mechanisms underlying response to cold stimulation in *P. tuoliensis* mycelia [6] as well as the identification of genes associated with *P. tuoliensis* fruiting body formation [12].

To our knowledge, no transcriptome analyses have been performed to investigate the transcriptional changes that occur during physiological maturation of *P. tuoliensis* mycelia. To identify genes and pathways that are differentially expressed during physiological mycelium maturation and explore molecular mechanisms underlying this maturation, a comparative transcriptome study on immature and mature *P. tuoliensis* mycelia was performed in this study.

## 2. Materials and Methods

### 2.1. Collection of *Pleurotus Tuoliensis* Mycelia

*P. tuoliensis* (ACCC 50869) was obtained from the Agricultural Culture Collection of China. Mycelia were grown at 25 °C in plastic bottles containing 820 g of cultivation substrate (40.43% cotton seed hull, 21.56% corncob, 25.16% wheat bran, 9.88% corn flour, 0.99% calcium carbonate, and 1.98% lime, and 65% water content, pH 8.5–8.8) in the dark for 37 days, after which the substrate was fully covered by mycelia. We randomly selected three cultivation bottles and took the mycelia from cultivation substrate into three corresponding tubes and froze them at –80 °C until RNA extraction (control samples, A). The remaining plastic bottles were maintained at a temperature of 17 °C in the dark to induce physiological maturation of mycelia. After 20 days of induction, CA-asp content in mycelia were determined every two days. After CA-asp content in mycelia reached 0.9g/L, mycelia reached physiological maturation [7]. We then randomly selected three cultivation bottles and transferred the mycelia to three corresponding tubes. All mycelia samples were stored at –80 °C until RNA extraction (case samples, B). The remaining plastic bottles were transferred to a mushroom house for fruiting management.

### 2.2. Library Preparation and RNA-Seq

Mycelia total RNA was extracted using TRIzol reagent (Life Technologies, New York, NY, USA) according to the manufacturer's protocol. The concentration, quality, and integrity of the extracted RNA samples were determined by agarose gel electrophoresis and a NanoDrop spectrophotometer (Thermo Scientific, USA). Six cDNA libraries were constructed using the Illumina TruSeq Stranded mRNA LT Library Preparation Kit (Illumina, USA) and then paired-end sequenced on an Illumina HiSeq 2000 platform by Shanghai Personal Biotechnology Co. Ltd. (Shanghai, China).

### 2.3. De Novo Transcriptome Assembly and Annotation

Raw reads were assessed for quality using FastQC (<https://www.bibsonomy.org/bibtex/f230a919c34360709aa298734d63dca3>) and have been deposited in the National Center for Biotechnology Information (NCBI) database under the the BioProject accession PRJNA551095. Low-quality reads, duplicate sequences and adapter sequences were removed to obtain high-quality clean reads using cutadapt

(<https://cutadapt.readthedocs.io/en/stable/>). Clean paired-end reads were assembled into contigs based on the similarity between overlapping regions using Trinity [13]. The resulting contigs were clustered with TGICL-2.1 [14] with the parameters “-l40-v20”. The transcript isoforms with the longest sequence in each Trinity subcomponent were extracted as a representative sequence of the Trinity subcomponent, called Unigene. All unigenes were pooled to yield the final unigene library.

Based on sequence homology, all assembled unigenes were compared to the NCBI non-redundant protein (Nr) and Swiss-Prot databases using the BLASTX program ( $e$ -value  $\leq 1e-5$ ). The Blast2GO program [15] was used to obtain gene ontology (GO) annotations for the unigenes, and GO-term classification was conducted based on the Nr annotations. Pathway annotations were assigned using the BBH (bi-directional best hit) method of the Kyoto Encyclopedia of Genes and Genomes (KEGG) Automatic Annotation Server (KAAS) online tool [16]. Unigenes were also aligned to the Evolutionary Genealogy of Genes: Non-Supervised Orthologous Groups (eggNOG) database for further functional prediction and classification.

#### 2.4. Identification and Functional Analysis of Differentially Expressed Genes (DEGs)

To identify the differentially expressed genes, transcript abundance were first calculated by mapping clean reads to unigenes database using RSEM [17]. DEGs were then identified by R Bio-package DESeq2. [18]. The significance of gene expression differences was assessed using the  $p$ -value  $\leq 0.05$  and  $|\log_2 \text{ratio}| \geq 1$ . These differentially expressed data have been deposited in the Gene Expression Omnibus (GEO) database at the NCBI archives (<https://www.ncbi.nlm.nih.gov/geo/>) under accession GSE135839.

The online analysis tool, the Database for Annotation, Visualization and Integrated Discovery (DAVID; <https://david.ncifcrf.gov/>) were used to analysis DEGs for GO term and KEGG pathway enrichment.

#### 2.5. Reverse Transcription-Quantitative PCR (RT-qPCR) Analysis

We selected 12 DEGs identified by RNA-seq for validation by RT-qPCR. Total RNA from immature and mature mycelia was extracted using the same procedures as used for RNA-seq. RNA from each sample was reverse transcribed using TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix (TransGen Biotech, Beijing, China). Primers for the 12 DEGs were designed in Primer 6.0 software (Premier, Ottawa, Canada) and are shown in Table 1. Reaction mixtures (20  $\mu$ L) contained 10  $\mu$ L of 2 $\times$  SYBR Green Master Mix, 0.4  $\mu$ L of 10 nM each primer, 2  $\mu$ L of cDNA template, and 7.2  $\mu$ L of ddH<sub>2</sub>O. The qPCR reactions were cycled on an ABI Prism 7500 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA) with SYBR green fluorescent dye (TaKaRa, Dalian, China). Amplification conditions were: 95 °C for 3 min, followed by 40 cycles of 94 °C for 3 s, 60 °C for 32 s, and 72 °C for 30 s. Each reaction was performed in triplicate using the glyceraldehyde-phosphate dehydrogenase gene (*gapdh*) as an internal reference. Relative gene expression levels were calculated using the  $2^{-\Delta\Delta C_t}$  method [19].

**Table 1.** Primers for qRT-PCR of 12 DEGs

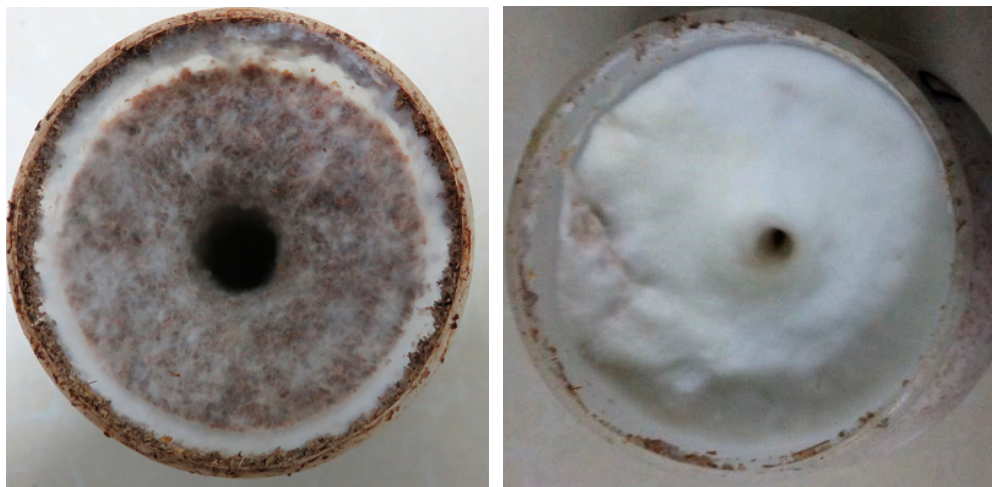
Gene ID	Annotation	Primer
DN2903_c0_g1	Exopolygalacturonase (EPS)	F:GTGAGTTGGCTTGCTTCT R:CGATTCCTATCTGTCTTGGT
DN3768_c0_g2	Pseudouridine synthase (PS)	F:TCGTCTACCTCCATATTGTC R:CTCTCAGCACCATACCTAG
DN4800_c0_g2	Nucleoside diphosphate kinase (NDPK)	F:GCTGTTGGACGCAATATC R:CACACTTCCTAAGACTTCAC
DN8870_c0_g2	Versatile peroxidase (VP)	F:GCAATGATGAAGACACAGAC R:GCAACTACGCAGCCTAAT
DN5369_c0_g1	Fruiting body protein SC3 (SC3)	F:CTTCATCGCCTATCTTCACT R:CGACCAATAACAGTAACG

DN5495_c0_g1	Cytochrome P450 monooxygenase (P450)	F: TCCTTCGTATTCGTCTTCAG R: CGTCGCTACCTTACATCAA
DN8865_c1_g1	Cell cycle checkpoint control protein RAD9A (RAD9A)	F: GCATCACCGTCCATCATT R: CTTCCGCAATCTCGTCTT
DN556_c0_g1	Carbonic anhydrase (CA)	F:CGTCTCATCTCTGTCACTAAT R: CCGAGGTGTTCAGGAATC
DN10455_c0_g1	$\alpha$ -galactosidase ( $\alpha$ -GAL)	F: CAGAATACGACCGCATGG R: AAGCACCACCAACCAATG
DN8848_c0_g1	NADPH dehydrogenase afvA (AFVA)	F: CCTTCCTATCGTCTTCTCTG R: CCACTTCAACTGCCTCTC
DN7771_c0_g2	Dual specificity protein phosphatase (DUSP)	F: AGGAACGACTGAACTGAG R: GGTATGGTGGAGGTGAAG
DN10260_c0_g1	ATP-dependent DNA helicase Q4 (RECQL4)	F: GTCTGTTGCTGCTTGTCT R: CTCTATCCATTGCTTGATGTC

### 3. Results

#### 3.1. Growth of *P. Tuoliensis* Mycelia

The cultivation substrate was fully covered by white fungal mycelia after incubation at 25 °C for 37 days in the dark. These cultivation bottles were then maintained at a temperature of 17 °C in the dark to induce mycelia maturation. The sawdust media surface was gradually covered with plump and elastic hyphae (Figure 1). At the 35th day, CA-asp content in mycelia was determined to be 0.982 g/L, which means the physiological maturation of mycelia. Primordia were initiated within 7 days under low temperature and light stimulation conditions and within 15 days, developed reproductive fruiting bodies that could be harvested. To investigate the molecular basis of physiological maturation of mycelia, we collected immature mycelia and mature mycelia for RNA-seq analysis.

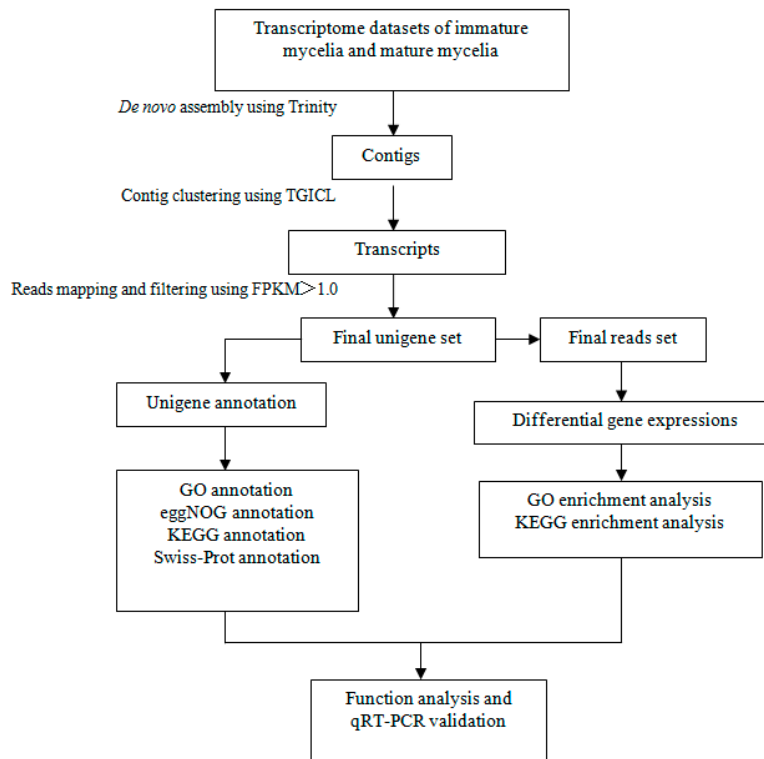


**Figure 1.** The morphological characteristics of immature mycelia (Left image) and mature mycelia (Right image).

#### 3.2. Sequencing and Gene Annotation

We performed paired-end sequencing of six cDNA libraries (three biological replicates each for immature and mature mycelia). We obtained an average of 45,688,081 raw reads per library. After removing adapter sequences and low-quality reads, clean reads represented approximately 99.53% of the raw reads. Clean reads were combined into a single dataset for de novo assembly. A total of 60,330 transcripts, including sequence isoforms, were obtained. Transcripts displayed a mean length

of 1897 bp. To reduce redundancy, only the longest sequence isoform within each subcomponent was selected. In total, we obtained 17,030 unigenes representing unique transcripts for further analysis (Figure 2).



**Figure 2.** Analytical framework of the *P. tuoliensis* mycelia transcriptome. Raw reads were filtered and then de novo assembled into contigs by Trinity. Unigenes were obtained by clustering the contigs with TGICL. FPKM values were calculated for each of the unigenes, and only unigenes with FPKM > 1.0 were used for further analyses and annotation against the public databases.

This unigene dataset displayed a total size of 21,120,061 bp, an average length of 1240 bp, and an N50 of 1891 and served as the mRNA transcriptome for mycelia (Table 2).

**Table 2.** Assembly and annotation statistics of *P. tuoliensis* mycelia transcriptome

<b>Samples</b>	<b>PE01</b>	<b>PE02</b>	<b>PE03</b>	<b>PE351</b>	<b>PE352</b>	<b>PE353</b>
Raw reads	41,410,374	46,175,662	43,409,086	45,505,956	48,825,340	48,802,066
Clean reads	41,222,978	45,964,590	43,208,980	45,306,508	48,616,554	48,565,980
Percentage of clean reads (%) <sup>a</sup>	99.54	99.54	99.53	99.56	99.57	99.51
<b>Assembled statistics</b>						
Number of assembled transcripts	60,330					
Mean length of assembled transcripts (bp)	1897					
Number of assembled unigene	17,030					
Total assembled unigene size (bp)	21,120,061					
Mean length of assembled unigene (bp)	1240					
N50 (bp) <sup>b</sup>	1891					
The number of N50 transcripts	3412					
<b>Annotation</b>						

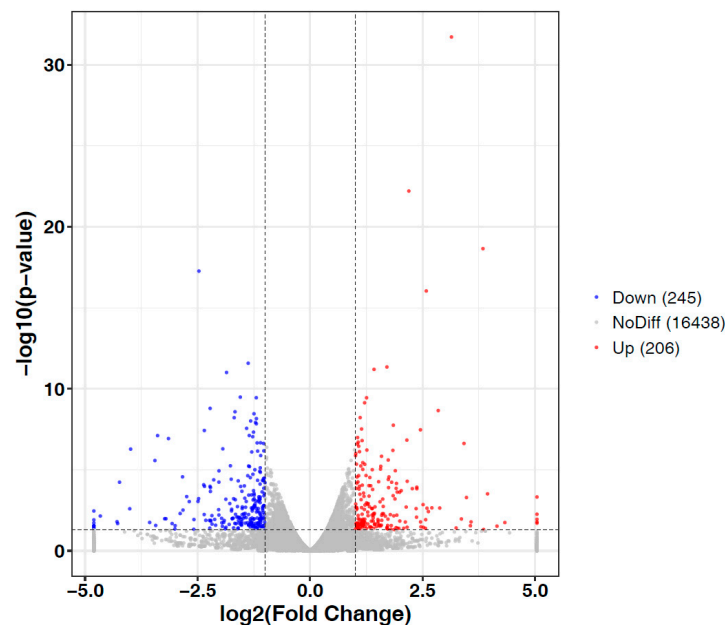
NR	13,611
GO	8804
KEGG	2942

<sup>a</sup> Percentage of the number of clean reads in the number of total raw reads. <sup>b</sup> All sequences are arranged from long to short sequence, and the sequence lengths are sequentially added in this order. When the added length reaches 50% of the total length of the sequence, the length of the last sequence is recorded as N50 (bp).

In total, 14,300 and 7115 unigenes were mapped using the eggNOG (Table S1) and Swiss-Prot (Table S2) databases, respectively. In addition, 13,611 (79.92%) unigenes matched known proteins in NR (NCBI non-redundant protein sequences) database (Table S3), and 95% of these were homologous with proteins from *P. ostreatus*, a species evolutionarily related to *P. tuoliensis* (Table S4). Additionally, 8,804 (51.70%) unigenes assigned to more than one GO term were classified into 67 functional groups from the Gene Ontology database (Table S5). Using KAAS (KEGG Automatic Annotation Server) with *P. ostreatus* as the reference, a total of 2,942 unigenes were mapped to 35 level-2 pathways (Table S6). Furthermore, a total of 2603 unigenes were annotated in all of these databases.

### 3.3. Identification and Functional Enrichment Analysis of Differentially Expressed Genes (DEGs)

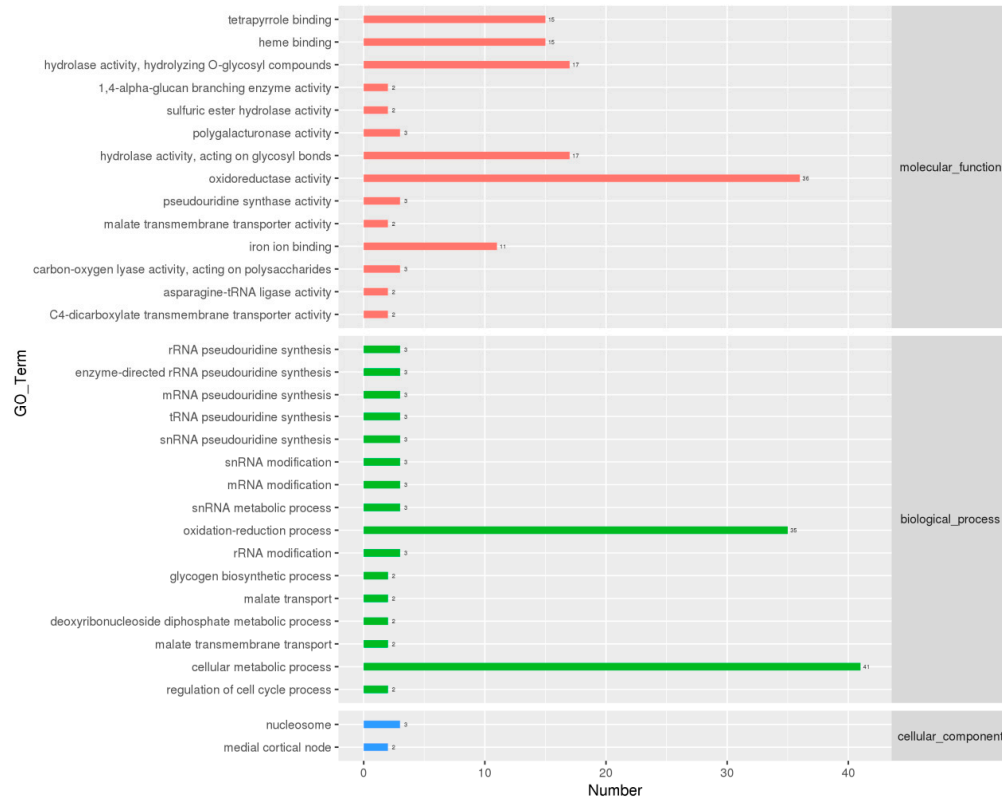
To identify significantly differentially expressed genes (DEGs), we compared the normalized read counts (FPKM values) of the immature and mature mycelia samples. This comparative analysis revealed 451 significantly DEGs ( $p$ -value  $\leq 0.05$ ,  $|\log_2 \text{ratio}| \geq 1$ ), which included 206 upregulated and 245 downregulated genes (Figure 3). The analyzed data of differential expression analysis for each unigene are shown in Table S7.



**Figure 3.** The volcano plots showing expression level of each unigene. Limits defined by  $p$ -value  $\leq 0.05$  and  $|\log_2 \text{ratio}| \geq 1$ . Red points represent upregulated genes; blue points represent downregulated genes; gray area represents insignificant gene.

We next performed GO analysis on the DEGs to gain insight into their biological functions. A total of 197 DEGs were enriched. Enrichment result of all DEGs is shown in Table S8. Figure 4 shows the top 32 enriched GO terms assigned to our DEGs. The cellular metabolic process (GO:0044237) annotation was enriched with most DEGs (False discovery rate (FDR)  $> 0.05$ ). Oxidation-reduction

process (GO:0055114) and oxidoreductase activity (GO:0016491) were also enriched (FDR > 0.05), indicating that oxidation-reduction processes may be highly active in mature mycelia. Other enriched GO terms were mainly involved in the synthesis and processing of genetic information, including rRNA modification (GO:0000154) (FDR > 0.05), mRNA pseudouridine synthesis (GO:1990481) (FDR < 0.05), snRNA metabolic process (GO:0016073) (FDR > 0.05), and deoxyribonucleoside diphosphate metabolic process (GO:0009186) (FDR > 0.05). Within the molecular function category, terms associated with carbohydrate metabolism, such as hydrolase activity acting on O-glycosyl compounds (GO:0004553) (FDR > 0.05) and hydrolase activity acting on glycosyl bonds (GO:0016798) (FDR > 0.05), were enriched. In general, GO terms associated with genetic information synthesis and processing were largely enriched.

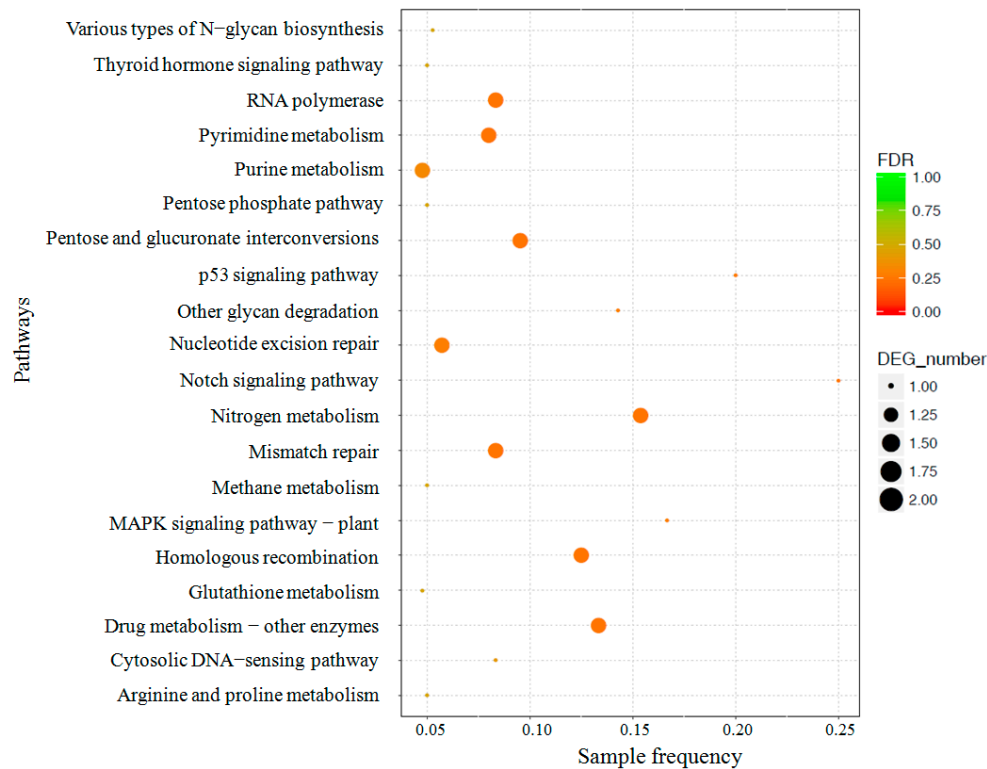


**Figure 4.** GO functional classification of differentially expressed genes. The green bars represent biological process; yellow bars represent cellular component; blue bars represent molecular function.

To better understand the functions and interactions of our DEGs, all DEGs underwent a pathway-based analysis by mapping against the KEGG database. Pathway enrichment analysis demonstrated that only 35 pathways were enriched, which included 45 DEGs (Table S9). We found that these pathways were mainly associated with nucleotide metabolism, DNA replication and repair, and energy metabolism, such as pyrimidine metabolism (ko00240), nucleotide excision repair (ko03420), pyruvate metabolism (ko00620), and the pentose phosphate pathway (ko00030). The results of our KEGG analysis corresponded with those obtained by our GO enrichment analysis. The 20 enriched pathways with FDR < 1 are shown in Figure 5.

In addition, the total annotation results of Unigenes and DEGs are listed in Table S10.



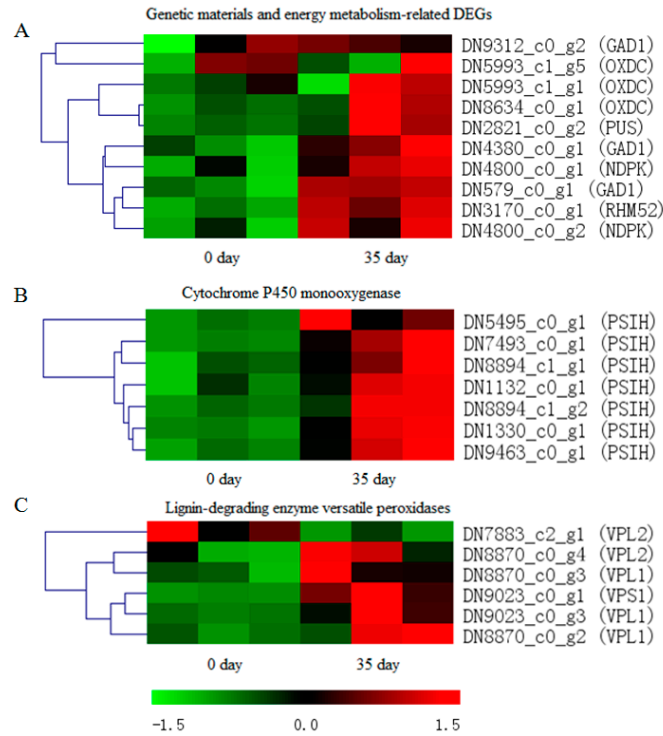


**Figure 5.** The top 20 enriched KEGG pathways of differentially expressed genes between immature mycelia and mature mycelia. ‘Sample frequency’ means the ratio of the number of differentially expressed genes in a pathway to the number of all annotated genes in this pathway in the figure legend. DEG number indicates the number of differential genes annotated in the pathway, indicated by a dot, the larger the dot, the more differential genes are enriched. The color gradient indicates the corrected  $p$ -value by Benjamini–Hochberg FDR method. Only pathways with FDR <1 were shown.

### 3.4. Identification of DEGs Related to Physiological Maturation of Mycelia

To identify genes related to physiological maturation of mycelia, we analyzed significantly upregulated DEGs in mature mycelia. Overall, these upregulated DEGs are mainly involved in the synthesis of genetic materials and energy metabolism, and include unigenes encoding glutamate decarboxylase Gad1 (DN579\_c0\_g1), oxalate decarboxylase OxdC (DN8634\_c0\_g1), pseudouridylate synthase Pus (DN2821\_c0\_g2), DNA repair and recombination protein RHM52 (DN3170\_c0\_g1), and nucleoside diphosphate kinase Ndpk (DN4800\_c0\_g2), demonstrating that this complex developmental program requires not only fast mycelial vegetative growth, but also more energy to accumulate nutrients and thus nurture developing fruit bodies during the reproductive stage. Versatile peroxidases (Vps) are key lignin-degrading enzymes, as they possess both manganese and lignin peroxidase activities. We observed upregulation of Vp-encoding genes (DN9023\_c0\_g1, DN9023\_c0\_g3, and DN8870\_c0\_g2) in mature mycelia, consistent with nutrient accumulation. Another important class of DEGs identified in our mature mycelia were those encoding post-modification enzymes, such as cytochrome P450 monooxygenase and methyltransferases, which are important for the synthesis of secondary metabolites. The expression of genes encoding cytochrome P450s monooxygenase Psih (DN1132\_c0\_g1, DN1330\_c0\_g1, DN5495\_c0\_g1, DN7493\_c0\_g1, DN8894\_c1\_g1, DN8894\_c1\_g2, and DN9463\_c0\_g1) and a methyltransferase (DN4644\_c0\_g2) were found to be significantly upregulated in mature mycelia. Finally, upregulated fruiting body protein SC3 (DN5369\_c0\_g1) was enriched in mature mycelia, probably indicating the initiation and formation of fruiting bodies (Figure 6).





**Figure 6.** Genes involved in the regulation of mycelia physiological maturation in *P. tuoliensis*. (A) DEGs associated with genetic material and energy metabolism. (B) DEGs associated with metabolic processes and stress responses. (C) Lignin-degrading enzymes versatile peroxidases. The right side of the heatmap indicates the gene ID in our transcriptomic data for *P. tuoliensis* and the homologous gene name. The gene expression values (FPKMs) were transformed to z-score values.

### 3.5. Distinct Distribution of Carbohydrate-Active Enzymes during Mycelial Maturation

In this study, we also searched our DEGs for families related to carbohydrate-active enzymes (CAZymes) using the Carbohydrate Active Enzymes (CAZy) database [20]. BLASTX indicated that a total of 21 DEGs from the mycelia transcriptome were CAZymes. These upregulated CAZymes could be classified into superfamilies, such as glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate-binding modules (CBMs), and carbohydrate esterase (CE) superfamily (Figure 7A), which likely indicates a functional change in plant cell wall polysaccharide degradation in cultivation substrate. The differences in expression of these CAZymes between immature and mature mycelia are summarized in Figure 7B. The genes encoding exopolysaccharidase (DN2903\_c0\_g1, GH28),  $\beta$ -glucosidase (DN3476\_c0\_g1, GH1), 1,4- $\alpha$ -glucan-branching enzyme (DN4658\_c0\_g1, GH13), glucoamylase (GH15), and mannan endo-1,4- $\beta$ -mannosidase (DN5333\_c0\_g4, GH5) were upregulated in mature mycelia, and  $\beta$ -mannosidase (GH2) and  $\alpha$ -galactosidase (GH27) were downregulated.

