Journal of Visualized Experiments

Gene Mining and Sequence Analysis of Purine Nucleosidase Based on RNA-Seq

--Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE61561R3
Full Title:	Gene Mining and Sequence Analysis of Purine Nucleosidase Based on RNA-Seq
Corresponding Author:	Shan Lin, Ph.D. Beijing University of Chinese Medicine Shenzhen, Guangdong CHINA
Corresponding Author's Institution:	Beijing University of Chinese Medicine
Corresponding Author E-Mail:	biotechlin@foxmail.com
Order of Authors:	Shan Lin, Ph.D.
	Fenfang Wu
	Shiping Hu
	Yun Ran
	Xiaoni Chen
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US$2,400)
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	Shenzhen, Guangdong, China
Please confirm that you have read and agree to the terms and conditions of the author license agreement that applies below:	I agree to the Author License Agreement
Please specify the section of the submitted manuscript.	Genetics
Please provide any comments to the journal here.	

TITLE:

Gene Mining and Sequence Analysis of Purine Nucleosidase Based on RNA-Seq

AUTHORS AND AFFILIATIONS:

Fenfang Wu[1], Shiping Hu[2], Yun Ran[2], Xiaoni Chen[1], Shan Lin[1]*

[1]Department of Central Laboratory, Shenzhen Hospital, Beijing University of Chinese Medicine, Shenzhen, Guangdong, China

[2]Department of Hepatopathy, Shenzhen Hospital, Beijing University of Chinese Medicine, Shenzhen, Guangdong, China

Email Addresses of Co-authors:

Fenfang Wu (wufenfang6768@gmail.com)

Shiping Hu (shipinghu18@qq.com)

Yun Ran (ranyun99@sina.com)

Xiaoni Chen (chenxiaoni902@163.com)

Corresponding Author:

Shan Lin (biotechlin@foxmail.com)

KEYWORDS:

gene mining, sequence analysis, RNA-Seq, purine nucleosidase, adenosine, biosynthesis

SUMMARY:

In this protocol, a method for gene mining and sequence analysis of purine nucleosidase (PN, EC:3.2.2.1) based on RNA-Seq was described. ProtProm analysis was applied to show the unique secondary and tertiary structures of PN. Furthermore, the *PN* gene was cloned from transcriptome to verify the reliability of RNA-Seq results.

ABSTRACT:

Caterpillar fungus (*Ophiocordyceps sinensis*) is one of the most valued fungal Traditional Chinese medicine (TCM), and it contains plenty of active ingredients such as adenosine. Adenosine is considered as a biologically effective ingredient that has a variety of anti-tumor and immunomodulatory activities. In order to further elucidate the mechanism of purine nucleosidase (PN) in adenosine biosynthesis, a gene encoding PN was successfully mined and further analyzed based on the RNA-Seq database of caterpillar fungus. The full-length cDNA of *PN* was 855 bp, which encoded 284 amino acids. BLAST analysis showed the highest homology of 85.06% with nucleoside hydrolase in NCBI. ProtProm analysis showed that the relative molecular weight was 30.69 kDa and the isoelectric point was 11.55. The secondary structure of PN was predicted by Predict Protein; the results showed that alpha helix structure accounted for 28.17%, strand structure accounted for 11.97%, and loop structure accounted for 59.86%. Moreover, *PN* gene was further cloned from transcriptome and detected by agarose gel electrophoresis for verification. This study provides more sufficient scientific basis and new ideas for the genetic regulation of adenosine biosynthesis in fungal

TCM.

INTRODUCTION:

Fungal Traditional Chinese medicine (TCM) has abundant species resources[1,2]. Caterpillar fungus (*Ophiocordyceps sinensis*) is a well-known fungal TCM and is regarded as a source of innovative drugs[3,4]. Caterpillar fungus is a worm and fungus combined mixture that is found on the Tibetan plateau in southwestern China, where *Hirsutella sinensis* is parasitic on the caterpillar body[5]. Currently, *H. sinensis* is reported as the only anamorph of caterpillar fungus according to molecular and morphological biology evidence[6,7], and it has less associated toxicity and similar clinical efficacy compared to wild caterpillar fungus[8]. It was revealed that *H. sinensis* possesses a variety of biologically effective ingredients, such as nucleosides, polysaccharides, and ergosterols, with extensive pharmacological effects such as repairing a liver injury[9–11]. Adenosine is a typical active ingredient isolated from caterpillar fungus, and it is a kind of purine alkaloid[12]. Adenosine has a variety of biological activities: anti-tumor, antibacterial, and immunomodulatory activities[13,14]. Unfortunately, the biosynthetic mechanism of adenosine as well as the key genes involved is still unclear[15,16].

Adenosine mainly shows its anti-tumor effect through immunosuppressive actions in the tumor microenvironment[17]. It was reported that adenosine showed immunosuppressive functions, which was critical to initiate tissue repair after injury and to protect tissues against excessive inflammation[18,19]. Moreover, it was demonstrated that adenosine-mediated repression of immunity could severely impair cancer immunosurveillance as well as promote tumor growth[20]. Thus, it is urgent to study the mechanism of adenosine biosynthesis for its wide application in anti-tumor.

It was reported that a complete view of expressed genes and their expression levels could be systematically conducted by next-generation sequencing of transcriptome[21]. Furthermore, transcriptome sequencing and analysis was applied to predict the genes involved in the biosynthetic pathway of the active ingredients, and further investigate the interaction of different biosynthetic pathways[22]. Purine nucleosidase (PN, EC 3.2.2.1) is a class of nucleosidase with substrate specificity for purine nucleosides, which can hydrolyze the glycoside bonds of purine nucleosides into sugars and bases[23]. It typically plays important roles in adenosine biosynthesis. It was reported that the biosynthetic pathway of adenosine in fungal TCM was predicted; qPCR and gene expression showed that the increased adenosine accumulation is a result of down-regulation of *PN* gene, indicating that the *PN* gene may play an important role in adenosine biosynthesis[15]. Therefore, the mechanism of *PN* in adenosine biosynthesis must be urgently clarified. However, the sequence information and protein structure of PN as well as other key genes involved in adenosine biosynthesis of fungal TCM have not been further studied.

In this study, a novel sequence of *PN* gene was mined from RNA-Seq data of caterpillar fungus and verified by gene cloning. Furthermore, the molecular characteristics and protein structure of PN were comprehensively analyzed, which could provide new directions and ideas for the gene regulation of adenosine biosynthesis.

PROTOCOL:

NOTE: A strain of anamorph of caterpillar fungus (*H. sinensis*) was deposited in our laboratory. *Escherichia coli* DH5 were preserved by Shenzhen Hospital, Beijing University of Chinese Medicine.

1. Preparing for RNA-Seq

1.1 Harvesting of mycelia

1.1.1 Prepare fermentation medium for fermentation of *H. sinensis*: powdered corn flour (1%), silkworm pupae (1.5%), yeast extract (0.5%), tryptone (1%), glucose (1.5%), bran (1.5%), dextrin (0.5%), KH_2PO_4 (0.02%) and $MgSO_4$ (0.01%).

1.1.2 Prepare inoculation by 10% fermentation medium for scale-up culture (add 10 mL medium per 100 mL medium). Conduct submerged fermentation by the condition of 16 °C on a rotary shaker at 150 rpm for 10 days.

1.1.3 Asexually reproduce and harvest mycelia of the anamorph of caterpillar fungus for 10 days. Centrifuge the fermented medium and discard the supernatant after centrifugation. Suspend the mycelia by adding 100 mL of ultrapure water for 3 times and remove the supernatant by centrifugation. Grind the cleaned mycelia into a powder using liquid nitrogen.

1.2 RNA-Seq

1.2.1 Extract total RNA of the anamorph of caterpillar fungus according to the manufacturer's protocols (**Table of Materials**) and further treat the sample with RNase-free DNase I (**Table of Materials**).

1.2.2 Isolate the mRNA from total RNA PolyATtract mRNA Isolation Systems, and isolate poly(A) mRNA using beads with oligo(dT) according to the manufacturer's protocols (**Table of Materials**).

1.2.3 Take the short fragments as templates to synthesize the first-strand cDNA by random hexamer-primers according to the manufacturer's protocols (**Table of Materials**). Perform the synthesis of second-strand cDNA according to the manufacturer's protocols.

1.2.4 Subsequently, generate the sequencing libraries using the Ultra RNA Library Prep Kit according to the manufacturer's protocols (**Table of Materials**).

1.2.5 Purify short fragments by PCR extraction kit according to the manufacturer's protocols (**Table of Materials**) and resolve it by EB buffer, respectively.

1.2.6 Connect the short fragments (threshold of 300 bp) with sequencing adapters according to the result of agarose gel electrophoresis.

1.2.7 Subsequently, conduct amplification with PCR using the templates selected from suitable fragments.

1.2.8 Sequence the library by Illumina HiSeq 4000 with paired-end sequencing according to the manufacturer's protocols. Filter dirty raw reads from the raw sequence data to obtain clean data. Adopt *denovo* assembly to get Unigenes with the least Ns that cannot be extended on either end.

1.2.9 Align Unigene sequences by blastx to protein databases such as nr, Swiss-Prot, KEGG, and COG (e-value < 0.00001). Retrieve proteins with the highest sequence similarity with the given Unigenes along with their protein functional annotations. Summarize the RNA-Seq results (**Table of Materials**).

NOTE: Commercial kits were used in the above steps, and all operations were done according to the manufacturer's protocol.

2. Gene mining of purine nucleosidase

2.1 Download the files of RNA-Seq results on the computer. Find the annotation result files of assembled Unigenes from the RNA-Seq results.

NOTE: Paired-end reads were used again for gap filling of scaffolds to obtain sequences with least Ns that cannot be extended on either end. Such sequences were defined as Unigenes. Unigene annotation provides information of expression and functional annotation of Unigene.

2.2 Open the **Annotation Files** path and enter **map 00230** in the search bar; then search purine metabolism (map00230) in the KEGG classification of annotation files.

2.3 Mark EC:3.2.2.1 (PN) red in the annotated map00230 and indicate that there were assembled Unigenes that have been annotated to PN.

NOTE: There were three Unigenes (Unigene10777, Unigene14697, and Unigene17827) annotated to PN after clicking on the EC number 3.2.2.1 and shown in the annotated map00230.

2.4 Click on **EC number 3.2.2.1** and show the annotated Unigenes information.

2.5 Open the LTFViewer software and import Unigene.fa file with a **Ctrl-O** shortcut and show the sequence information of assembled Unigenes.

2.6 Search for sequence information of Unigene10777, Unigene14697, and Unigene17827 with a **Ctrl-F** shortcut.

2.7 Download the sequence information of Unigene10777, Unigene14697, and Unigene17827 with Ctrl-C and Ctrl-V shortcuts.

2.8 Eliminate Unigene10777 and Unigene17827 with excessively short sequences of open reading frame (ORF).

NOTE: Basic sequence information was displayed in LTFViewer software.

2.9 Select Unigene14697 (size 1,705 bp, gap 0 0%) with suitable length of ORF for further study.

3. Bioinformatic analysis

3.1 Analyze the ORF of *PN* gene by ORFfinder (https://www.ncbi.nlm.nih.gov/orffinder/).

3.1.1. Paste the sequence into the box. Choose parameters as follows, minimal ORF length (nt): 75, genetic Code: 1. standard, ORF start codon to use: ATG only. Click on the **Submit** button to obtain the ORF information.

3.2 Use the ProtParam tool (http://us.expasy.org/tools/protparam.html) to calculate the theoretical molecular mass and isoelectric point.

3.2.1. Paste the amino acid sequence (in one-letter code) into the box and click on the **Compute Parameters** button to obtain the results.

3.3 Apply SignalP5.0 Server (http://www.cbs.dtu.dk/services/SignalP/) to predict the signal peptides.

3.3.1. Enter protein sequences in FASTA format. Choose parameters as follows, organism group: Eukarya, output format: Long output. Click on the **Submit** button to obtain the results.

3.4 Apply BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to analyze the homology of protein sequences.

3.4.1. Click on the **Protein Blast** button and enter the sequence into the box. Choose parameters as follows, database: Non-redundant protein sequences (nr), algorithm: blastp (protein-protein BLAST). Click on the **Blast** button to obtain the results.

3.5 Apply Clustal X program (http://www.clustal.org/) to align the acid sequences of PN

from different fungi.

3.5.1. Upload a file or paste the sequences into the box. Set the parameters as follows, output format: ClustalW with character counts. Click on the **Submit** button to obtain the results. Clustal X can only recognize files in the FASTA format, and the path of files can only include English names.

3.6 Use MEGA 4.0 (https://www.megasoftware.net/mega4/) to conduct the phylogenetic tree.

3.6.1. Open the software and click on the **File** button to upload the sequences. Select the data type as **Protein Sequences**, click on the **OK** button to proceed to the next step. Subsequently, click on the **Phylogeny** button and select **Bootstrap Test Phylogeny**, and then click on **Neighbor Joining Tree**. Select the default parameters and click on the **Compute** button to obtain the results.

3.7 Apply InterProScan (http://www.ebi.ac.uk/interpro/search/sequence/) to identify the catalytic domain of PN.

3.7.1. Enter the sequence into the box. Select the default parameters and click on the **Search** button to obtain the results.

3.8 Apply Predict Protein (http://www.predictprotein.org/) online to predict the protein secondary structure.

3.8.1. Enter an amino-acid sequence (one letter code) into the box, and then click on the **predictProtein** button to obtain the results.

3.9 Apply Online tools SWISS-MODEL (http://swissmodel.expasy.org/) to evaluate the three-dimensional structure of PN[24].

3.9.1. Click on the **Start Modeling** button and paste the target sequence into the box. Fill in the **Project Title** and **Email** information and click on the **Search for Templates** button to obtain the results.

4. Gene cloning and construction of recombinant plasmid

4.1 Design primers whose reverse primer contained a *Not*I site and the forward primer had an *Eco*RI site.

4.2 Show the forward primer as: AGA**GAATTC**ATGACCATGCCAGATTCT (5'–3'), and the reverse primer as: ATA**GCGGCCGC**CTAACGCGTGCCGTTAGA (5'–3') by Primer Express.

4.3 Prepare the primers as well as the cDNA of caterpillar fungus for cloning of *PN* gene.

Conduct PCR as follows: pre-denaturation at 95 °C for 5 min, denaturation at 94 °C for 45 s, renaturation at 55 °C for 60 s, extension at 72 °C for 90 s, repeat for 35 cycles, and extension at 72 °C for 10 min.

4.4 Obtain the PCR fragments and detect it by agarose gel electrophoresis for verification. Ligate the PCR fragments with pMD18-T. Conduct ligation system of PMD18-T as follows: 1 µL PMD18-T, 4 µL Solution1, and 5 µL Target gene. Set the conditions as follows: maintaining at 16 °C for 16 h, inactivation at 65 °C for 15 min.

4.5 Transfer the recombinant plasmids to the competent *E. coli* JM109 cells according to the operation manual[25].

4.6 Digest the recombinant pMD18-T/*PN* plasmids and vector ppic9K with *Eco*RI and *Not*I. Ligate the fragments after digestion by T4 DNA ligase.

4.7 Construct the recombinant plasmid ppic9K/*PN* for further heterologous expression.

REPRESENTATIVE RESULTS:

The ORF sequence of *PN* gene was 855 bp in length, which encoded 284 amino acids with a calculated molecular mass of 30.69 kDa and a predicted isoelectric point of 11.55, indicating that PN is an alkaline protein. Application of SignalP4.0 Server was conducted to identify signal peptide, and the results indicated that PN has no signal peptides. Moreover, the results of BLASTP search indicated that PN originated from caterpillar fungus shared the highest identity (85.06%, *E* value = 1e-88) with nucleoside hydrolase from *Purpureocillium lilacinum* (OAQ81830.1). Furthermore, the ClustalX program was applied to perform multiple sequence alignment of PN and the results were shown in **Figure 1**, which revealed that 11–166 amino acids were the conserved amino acid sequences of inosine/uridine hydrolase domain. Subsequently, the result of phylogenetic tree showed that PN from caterpillar fungus shared the closest phylogenetic relationship with other nucleoside hydrolase from entomogenous fungus such as *Purpureocillium lilacinum* (OAA82129.1, XP 018708456.1) based on the amino acid sequences similarity (**Figure 2**). Meanwhile, the analysis result of InterPro Scan revealed that PN had a catalytic domain of inosine/uridine-preferring nucleoside hydrolase (IPR023186).

Subsequently, PN protein secondary structure was predicted by Predict Protein, the results were shown in **Figure 3,** indicating that alpha helix structure accounted for 28.17%, strand structure accounted for 11.97%, and loop structure accounted for 59.86%. The tertiary structure of PN protein was constructed by Wiss-model simulation (**Figure 4**), and the results were similar to the ones predicted by Predict Protein. According to CDS online analysis software, PN belongs to nucleoside hydrolase family and catalyzes the hydrolysis of all of the commonly occurring purine and pyrimidine nucleosides into ribose and the associated base but has a preference for inosine and uridine as substrates.

The ORF of *PN* gene was amplified by PCR; the PCR products were detected by agarose gel

electrophoresis (**Figure 5**). The results indicated that PCR products with the correct sizes were successfully amplified.

FIGURE AND TABLE LEGENDS:

Figure 1. Multiple alignment of amino acid sequences for PN from fungal TCM and other nucleoside hydrolases. The sequences were those from *Trichoderma guizhouense* (OPB46800.1), *Purpureocillium lilacinum* (OAQ81830.1), and *Purpureocillium lilacinum* (XP_018180602.1).

Figure 2. Phylogenetic tree of PN showing the relationship with other species on amino acid sequences of nucleoside hydrolase. Phylogenetic tree was constructed with MEGA 4.0 with the method Neighbor-Joining. Test of inferred phylogeny was Bootstrap for 1,000 replications.

Figure 3. Prediction of secondary structure for PN. Blue stands for strand, and dark red stands for helix.

Figure 4. The tertiary structure of PN protein predicted by Wiss-model. The family type of PN belongs to inosine/uridine-preferring nucleoside hydrolase, which has a preference for inosine and uridine as substrates.

Figure 5. Agarose gel electrophoresis of *PN* gene cloned from the transcriptome of caterpillar fungus. Lane M: Trans2K Plus II DNA Marker; lane 1, PCR products of *PN* gene.

DISCUSSION:

Human health is facing a series of major medical problems such as tumor, cardiovascular, and cerebrovascular diseases[26,27]. TCM has been regarded as the source of research and development of innovative medicine, because of its rich species resources and diverse structure and functions of active ingredients[28,29]. Caterpillar fungus is a fungal parasite on the larvae of Lepidoptera, and it is an invigorant in Chinese tradition and considered as one of the best invigorants with Panax and Pilose antlers[30]. A variety of active ingredients such as adenosine, sterols, nucleosides, terpenes, and peptides can be extracted from TCM[29,31]. The active ingredients have a variety of physiological activities and structural types, and can be used as a source for the research and application of innovative drugs[32].

So far, there were many reports on the pharmacological effects of adenosine. However, the studies on the adenosine biosynthesis as well as the genes involved in were few[16,33]. Nevertheless, KEGG annotation of functional genes in *Cordyceps militaris* was carried out, and biosynthetic pathway of adenosine was speculated; it was found that *5'-nucleotidase* may be a key gene in adenosine biosynthesis[33]. Other studies speculated the biosynthetic pathway of adenosine; it was indicated that *adenosine kinase* and *5'-nucleotidase* genes were involved in the phosphorylation as well as dephosphorylation processes in metabolic pathway of adenosine[34,35]. In addition, the biosynthetic pathway of adenosine in fungal TCM

was predicted; *PN* gene was proved to play an important role in adenosine biosynthesis since down-regulation of *PN* gene was consistent with adenosine accumulation[15]. Unfortunately, the key genes involved in adenosine biosynthesis were lacking in-depth mining and analysis. Therefore, it is urgent to conduct the study of gene mining and sequence analysis of the key genes involved in adenosine biosynthesis.

Generally, the development of biotechnology requires more and more genetic resources[36]. Compared to traditional methods of gene mining, including microbial screening for obtaining genetic resources by molecular biological[37], metagenomic techniques for mining new genetic resources[38], and cloning of natural protein sequence after purification[39], the protocol of gene mining applied in this study is more efficient and accurate. Furthermore, the focus of this paper is on how to perform gene mining and sequence analysis of functional enzyme involved in biosynthesis of active ingredients based on RNA-Seq. This protocol could be very helpful to study the biosynthesis mechanism of other active ingredients of TCM. At the same time, other researchers could also refer to this protocol to mine functional proteins with research value and conduct in-depth research on them. However, this protocol also has some limitations. Firstly, gene mining relies on annotated RNA-Seq data, and RNA-Seq appears to be somewhat costly. Secondly, the results of sequence analysis based on bioinformatics analysis are predictive and need to be further verified by experiments.

In conclusion, the protocol of gene mining and sequence analysis provided an important theoretical basis to study the mechanism of adenosine biosynthesis, as well as the key role of PN in adenosine biosynthesis. Taken collectively, this study also would provide a more adequate scientific basis for gene regulation of adenosine biosynthesis and provide a new idea for promoting the modern industrial development of active ingredients in TCM.

DISCLOSURES:

The authors report no conflicts of interest.

ACKNOWLEDGMENTS:

This study was supported by National Natural Science Foundation of China (31871244, 81973733, 81803652), Natural Science Foundation of Guangdong Province (2019A1515011555, 2018A0303100007), Shenzhen Foundation of Health and Family Planning Commission (SZBC2018016), Special Fund for Economic and Technological Development of Longgang District of Shenzhen City (LGKCYLWS2020001571, LGKCYLWS2019000361).

REFERENCES:

1. Dong, C. J. The traditional Chinese medicine fungus Cordyceps and its biotechnological production. *Research Journal of Biotechnology*. **8**, 1–2 (2013).
2. Xia, E. H. et al. The caterpillar fungus, *Ophiocordyceps sinensis*, genome provides insights into highland adaptation of fungal pathogenicity. *Scientific Reports*. **7** (1), 1806 (2017).

3. Koganti, P. et al. *Cordyceps sinensis* increases hypoxia tolerance by inducing heme oxygenase-1 and metallothionein via Nrf2 activation in human lung epithelial cells. *BioMed Research International*. **2013**, 569206 (2013).
4. Shen, C. Y., Jiang, J. G., Li, Y., Wang, D. W., Wei, Z. Anti-ageing active ingredients from herbs and nutraceuticals used in traditional Chinese medicine: pharmacological mechanisms and implications for drug discovery. *British Journal of Pharmacology*. **174** (2017).
5. Jiang, Y., Yao, Y. J. Names related to *Cordyceps sinensis* anamorph. *Mycotaxon*. **84**, 245–254 (2002).
6. Chen, Y. Q., Wang, N., Qu, L. H., Li, T. H., Zhang, W. M. 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*. **29**, 597–607 (2001).
7. Liu, Z. Y. et al. Molecular evidence for the anamorph-teleomorph connection in *Cordyceps sinensis*. *Mycological Research*. **105**, 827–832 (2001).
8. Yu, S. J., Zhang, Y., Fan, M. Z. Analysis of volatile compounds of mycelia of *Hirsutella sinensis*, the anamorph of *Ophiocordyceps sinensis*. *Applied Mechanics and Materials*. **140**, 253–257 (2012).
9. Singh, M. et al. *Cordyceps sinensis* increases hypoxia tolerance by inducing heme oxygenase-1 and metallothionein via Nrf2 activation in human lung epithelial cells. *BioMed Research International*. **2013**, 569206 (2013).
10. Cha et al. Production of mycelia and exo-biopolymer from molasses by *Cordyceps sinensis* 16 in submerged culture. *Bioresource Technology*. **98**, 165–168 (2007).
11. Lin, S. et al. Enhancement of cordyceps polysaccharide production via biosynthetic pathway analysis in *Hirsutella sinensis*. *International Journal of Biological Macromolecules*. **92**, 872–880 (2016).
12. Xia, E. H. et al. The caterpillar fungus, *Ophiocordyceps sinensis*, genome provides insights into highland adaptation of fungal pathogenicity. *Scientific Reports*. **7**, 1806 (2017).
13. Cha, S. H. et al. Production of mycelia and exo-biopolymer from molasses by *Cordyceps sinensis* 16 in submerged culture. *Bioresource Technology*. **98**, 165–168 (2007).
14. Antonioli, L., Blandizzi, C., Pacher, P., Hasko, G. Immunity, inflammation and cancer: a leading role for adenosine. *Nature Reviews. Cancer*. **13**, 842–857 (2013).
15. Lin, S., Zou, Z., Zhou, C., Zhang, H., Cai, Z. Transcriptome analysis reveals the molecular mechanisms underlying adenosine biosynthesis in anamorph Strain of caterpillar fungus. *BioMed Research International*. **2019**, 1864168 (2019).
16. Lin, S. et al. Biosynthetic pathway analysis for improving the cordycepin and cordycepic aid production in *Hirsutella sinensis*. *Applied Biochemistry and Biotechnology*. **179**, 633–649 (2016).
17. Allard, B., Beavis, P. A., Darcy, P. K., Stagg, J. Immunosuppressive activities of adenosine in cancer. *Current Opinion in Pharmacology*. **29**, 7–16 (2016).
18. Fredholm, B. B. Adenosine, an endogenous distress signal, modulates tissue damage and repair. *Cell Death and Differentiation*. **14**, 1315–1323 (2007).

19. Ohta, A., Sitkovsky, M. Role of G-protein-coupled adenosine receptors in downregulation of inflammation and protection from tissue damage. *Nature*. **414**, 916–920 (2001).
20. Allard, B., Turcotte, M., Stagg, J. CD73-generated adenosine: orchestrating the tumor-stroma interplay to promote cancer growth. *Journal of Biomedicine & Biotechnology*. **2012**, 485156 (2012).
21. Zhang, Y., Wang, X., Nan, P., Li, J., Jin, L. De novo transcriptome sequencing of genome analysis provides insights into *Solidago canadensis* invasive capability via photosynthesis. *Journal of Plant Interactions*. **14**, 572–579 (2019).
22. Liu, Z. Q. et al. Transcriptome sequencing and analysis of the entomopathogenic fungus *Hirsutella sinensis* isolated from *Ophiocordyceps sinensis*. *BMC Genomics*. **16**, 106 (2015).
23. Ogawa et al. Purification, characterization, and gene cloning of purine nucleosidase from *Ochrobactrum anthropi*. *Applied and Environmental Microbiology*. **67**, 1783–1787 (2001).
24. Konstantin, A. et al. The swiss-model workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*. **16**, 195–201(2006).
25. Chung, C., Niemela, S. L., Miller, R. H. One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. *Proceedings of the National Academy of Sciences*. **86**, 2172–2175 (1989).
26. Lin, S. et al. Association between aldose reductase gene C(-106)T polymorphism and diabetic retinopathy: A systematic review and Meta-Analysis. *Ophthalmic Research*. **63,** 1–10 (2020).
27. Peng, Y. et al. Inguinal subcutaneous white adipose tissue (ISWAT) transplantation model of murine islets. *Journal of Visualized Experiments: JoVE*. **16,** 156 (2020).
28. Zhang, T. T., Jiang, J. G. Active ingredients of traditional Chinese medicine in the treatment of diabetes and diabetic complications. *Expert Opinion on Investigational Drugs*. **21**, 1625–1642 (2012).
29. Lin, S., Zhou, C., Zhang, H., Cai, Z. Expression, purification and characterization of 5'-nucleotidase from caterpillar fungus by efficient genome-mining. *Protein Expression and Purification*. **168**, 105566 (2020).
30. Kinjo, N., Mu, Z. Morphological and phylogenetic studies on *Cordyceps sinensis* distributed in southwestern China. *Mycoence*. **42**, 567–574 (2001).
31. Xiao, J. H., Ying, Q., Xiong, Q. Nucleosides, a valuable chemical marker for quality control in traditional Chinese medicine Cordyceps. *Recent Patents on Biotechnology*. **7**, 2 (2013).
32. Tu, P., Yong, J., Guo, X. Discovery, research and development for innovative drug of traditional Chinese medicine under new situations. *Zhongguo Zhong Yao Za Zhi*. **40**, 3423–3428 (2015).
33. Zheng, P. et al. Genome sequence of the insect pathogenic fungus *Cordyceps militaris*, a valued traditional Chinese medicine. *Genome Biology*. **12**, R116 (2011).
34. Covarrubias, R. et al. Role of the CD39/CD73 purinergic pathway in modulating arterial thrombosis in mice. *Arteriosclerosis, Thrombosis, and Vascular Biology*. **36**, 1809–1820 (2016).

35. Ogawa, Y., Murayama, N., Yanoshita, R. Molecular cloning and characterization of ecto-5'-nucleotidase from the venoms of *Gloydius blomhoffi*. *Toxicon*. **54**, 408–412 (2009).
36. Lin, S. et al. Mining and characterization of two novel chitinases from *Hirsutella sinensis* using an efficient transcriptome-mining approach. *Protein Expresion and Purification*. **133**, 81–89 (2017).
37. Ueda, M., Hirano, Y., Fukuhara, H. Gene cloning, expression, and X-ray crystallographic analysis of a β-mannanase from *Eisenia fetida*. *Enzyme and Microbial Technology*. **117**, 15–22 (2018).
38. Rebets, Y., Kormanec, J., Luzhetskyy, A., Bernaerts, K., Anné, J. Cloning and expression of metagenomic DNA in *Streptomyces lividans* and subsequent fermentation for optimized production. *Methods in Molecular Biology*. **1539**, 99 (2017).
39. Wang, S. S., Ning, Y. J., Wang, S. N., Zhang, J., Chen, Q. J. Purification, characterization, and cloning of an extracellular laccase with potent dye decolorizing ability from white rot fungus *Cerrena unicolor* GSM-01. *International Journal of Biological Macromolecules*. **95**, 920–927 (2016).

Figure 1.

Click here to access/download;Figure;Figure 1.tif

Figure 2.

Click here to access/download;Figure;Figure 2.tif

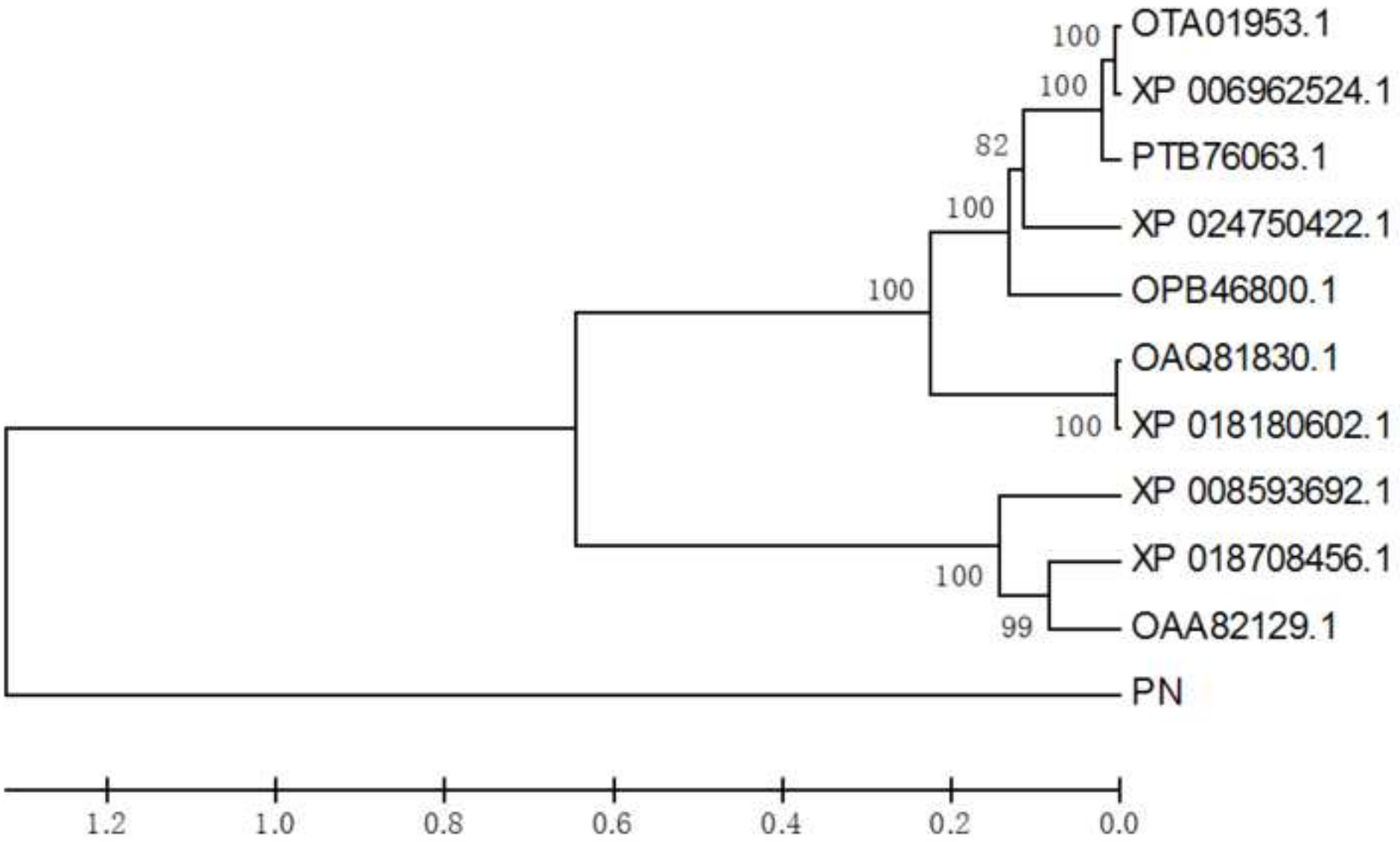


Figure 3.

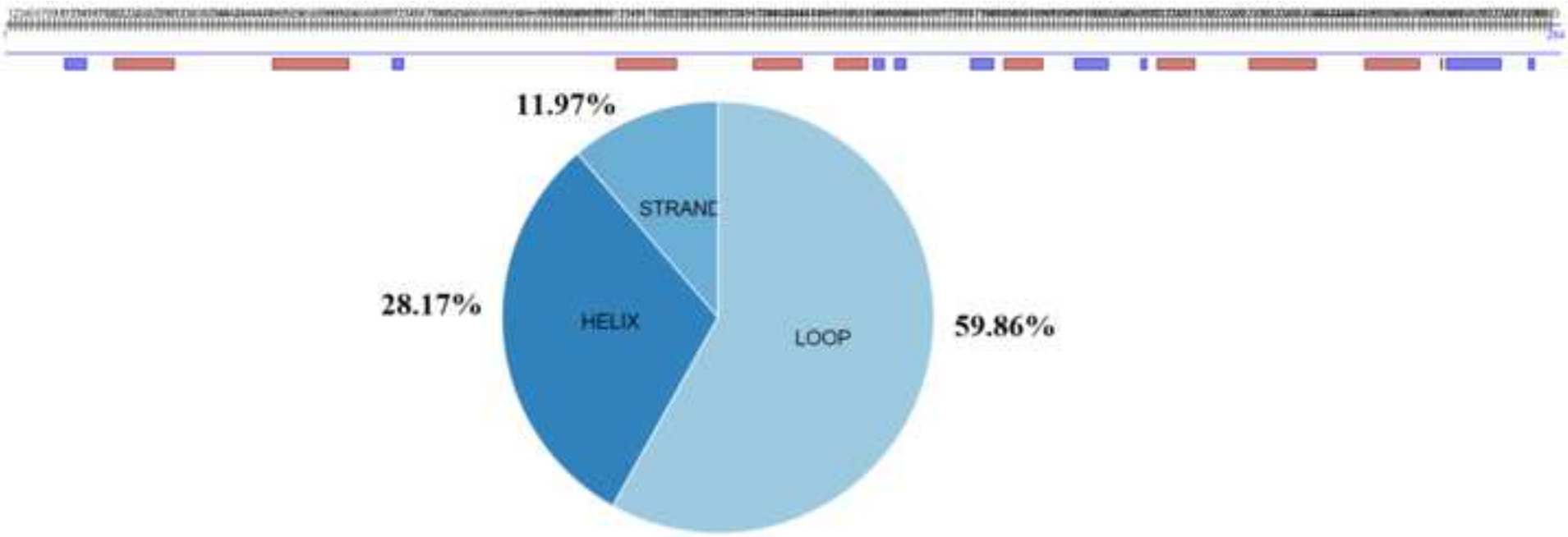


Figure 4.

Click here to access/download;Figure;Figure 4.tif

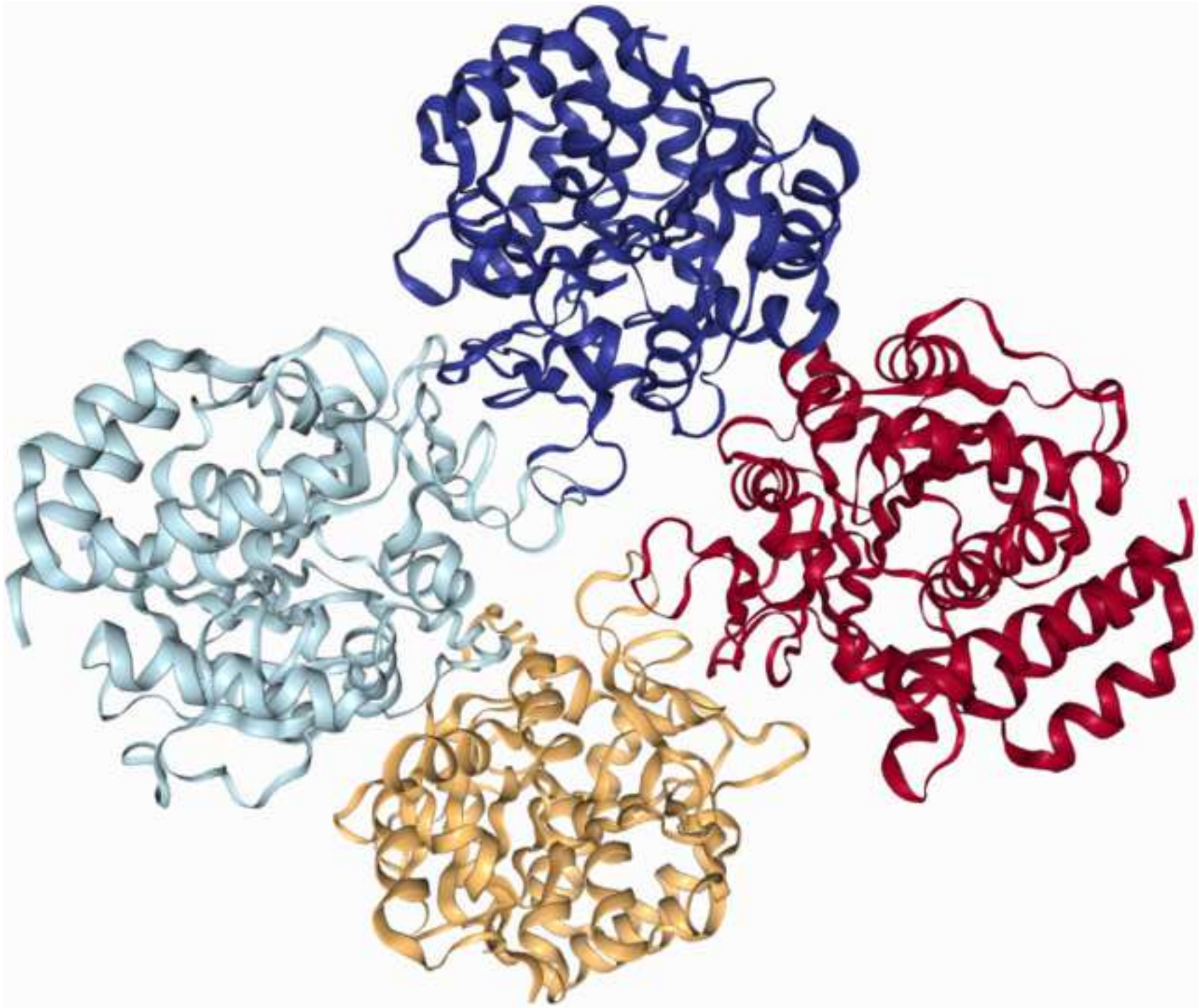
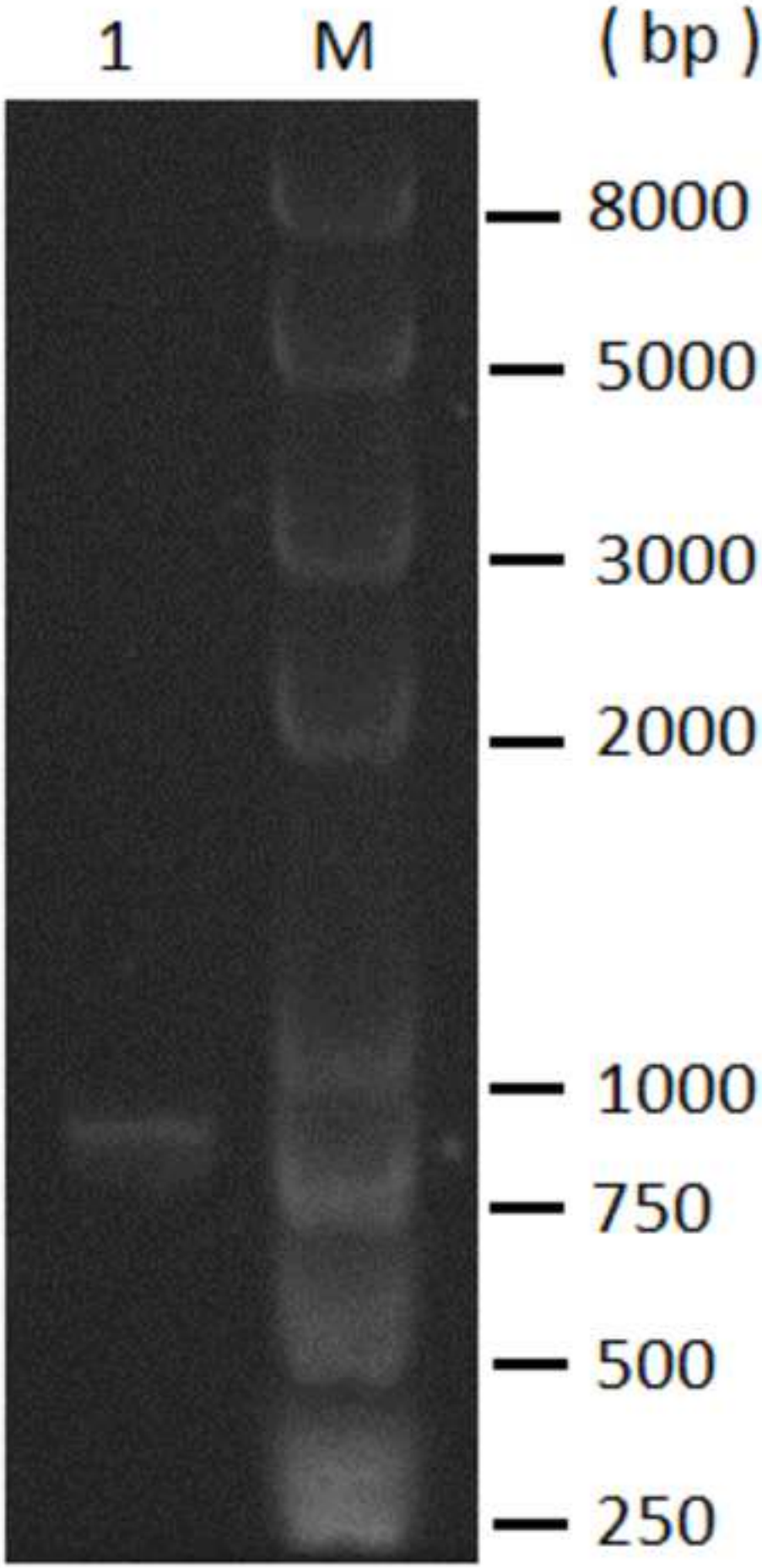


Figure 5.

Click here to access/download;Figure;Figure 5.tif



Name of Material/Equipment	Company	Catalog Number	Comments/Description
RNase-free DNase I	TaKaRa	2270B	
PolyATtract mRNA Isolation Systems	Promega	III	
Random hexamer-primers	Thermo Scientific	SO142	
NEBNext1 Ultra RNA Library Prep Kit	NEB	E7530S	
PCR extraction kit	QiaQuick		
Agarose	TransGen Biotech	GS201-01	
High-throughput sequencer	Illumina	HiSeq™ 4,000	
LTF Viewer	LTF	V5.2	
ORF program	NCBI		
ProtParam tool	SIB Swiss Institute of Bioinformatics		
SignalP Server	DTU Health Tech	5.0	
BLAST	NCBI		
Clustal X program	UCD Dublin Center for Evolutionary Medicine and Informatics		
MEGA		4.0	
InterProScan	European Molecular Biology Laboratory		
Predict Protein	Technical University of Munich		
WISS-MODEL	Swiss Institute of Bioinformatics		
Primer Express	Applied Biosystems	3.0	
EcoRI	NEB	R0101V	
NotI	NEB	ER0591	

pMD18-T Vector	TaKaRa	6011
agarose	Sigma-Aldrich	GS201-01
Trans2K® Plus II DNA Marker	Sigma-Aldrich	BM121-01
6×DNA Loading Buffer	Sigma-Aldrich	GH101-01
GelStain	Sigma-Aldrich	GS101-02
50 x TAE	Sigma-Aldrich	T1060
Gel imaginganalysis system	Syngene	G:BOX F3
E. coli JM109	Promega	
T4 DNA ligase	EarthOx	BE004A-02
pPIC9K	Genloci	GP0983

Journal of Visualized Experiments

No.: JoVE submission JoVE61561

Title: Gene Mining and Sequence Analysis of Purine Nucleosidase Based on RNA-Seq

Dear Chief Editor:

I am resubmitting here with the revised manuscript entitled “Gene Mining and Sequence Analysis of Purine Nucleosidase Based on RNA-Seq” along with the coauthors Fenfang Wu, Shiping Hu, Yun Ran, Xiaoni Chen and Shan Lin.

We thank you and reviewers for the critical feedback. We are delighted to be given the opportunity to revise our manuscript. The valuable comments helped us with the improvement of structure and focus of the manuscript. Please do forward our heartfelt thanks to these experts.

We carefully considered the comments offered by the reviewers, and the amendments have been combined into the revised manuscript. The modifications in the revised manuscript have been highlighted red. In addition, we have listed these modifications and changes below for your checking. We believe that the new revision reads more smoothly and the significance of our work can be conveyed more effectively.

We attach herein our responses to the reviewers’ comments/questions, and please find it as below.

Many thanks for considering our contribution. We hope this revised manuscript will meet the standard of your journal.

If you have any questions about this revision, please feel free to contact me, thanks.

Best regards.

Dr. Shan Lin
Department of Central Laboratory,

Shenzhen Hospital,
Beijing University of Chinese Medicine,
Shenzhen 518100, Guangdong, China,
biotechlin@foxmail.com,
Phone number: +86-755-89911830,
Fax number: +86-755-89911830.

PS1: List of changes in the revised manuscript

The changes which have been made in manuscript are listed as follows:

1. Page 1, line 28; page 2, line 47, "(*Ophiocordyceps sinensis*)" has been added behind Caterpillar fungus.
2. Page 1, line 32, "PN" has been modified to "purine nucleosidase (PN)".
3. Page 2, line 48-50, "Caterpillar fungus is a worm and fungus combined mixture, and found on the Tibetan plateau in the southwestern China, which *Hirsutella sinensis* is parasitic on the caterpillar body" has been added.
4. Page 3, line 107-108, "Prepare inoculation by 10% seed medium (add 10 ml seed medium into per 100 ml fermentation medium)" has been modified to "Prepare inoculation by 10% fermentation medium for scale-up culture (add 10 ml medium into per 100 ml medium)".
5. Page 3, line 127, "see **Table of Materials**" has been removed.
6. Page 4, line 142-145, "Sequence the library by Illumina HiSeq™ 4000 with paired-end sequencing. Filter dirty raw reads from the raw sequence data to obtain the clean data. Adopt *denovo* assembly to get Unigenes with least Ns and cannot be extended on either end." has been added.
7. Page 4, line 147-149, "Align Unigene sequences by blastx to protein databases like nr, Swiss-Prot, KEGG and COG (e-value<0.00001). Retrieve proteins with the highest sequence similarity with the given Unigenes along with their protein functional annotations" has been added.
8. Page 6, line 252, "WISS-MODEL" has been modified to "SWISS-MODEL".
9. Page 7, line 265, "by Primer Express" has been added.

10. Page 8, line 345-347, “Caterpillar fungus is a fungal parasite on the larvae of Lepidoptera, and it is an invigorant in Chinese tradition and honored as one of the three greatest invigorants together with Panax and Pilose antlers” has been added.
11. Page 10-12, references 5, 24, 30 have been added.

PS2: Responses to the editor and reviewers’ comments

Reviewer #1:

Accept

Response: Dear reviewer, thank you so much for your evaluations on our paper.

Reviewer #2:

Manuscript Summary:

Since the first review round of the paper on "Gene Mining and Sequence Analysis of Purine Nucleosidase Based on RNA-Seq" the manuscript still lacks a lot of data.

Response: Dear reviewer, thank you so much for your evaluations on our paper, the manuscript has been improved according to your valuable comments, and careful modifications have been made and combined to the revised manuscript.

Major Concerns:

Question 1: First of all, in the Abstract, the scientific name of the caterpillar fungus should be used. While the fungus is now named in the introduction I would recommend focusing more on the fungus and its traits than discussing TCM in detail. The same is true for the discussion e.g. in L366: why can this be just used for TCM and not for other applications? Please define PN not only in the summary but also in the Abstract and Introduction to increase readability.

Please reference all the databases used with the right literature reference. I did check for example for Swiss Model and the references are not cited in your manuscript.

Response: Thank you for your valuable comment. Follow your advice, the scientific name of the caterpillar fungus (*Ophiocordyceps sinensis*) has been used in the

Abstract. Furthermore, we have focused more on the fungus and its traits in Introduction and Discussion. Moreover, this not only can be used for TCM, but also for other applications. In addition, we have defined PN not only in the summary but also in the Abstract and Introduction to increase readability. Meanwhile, we have added the references for Swiss Model and checked all the databases used with the right literature reference.

Question 2: The figures 1 and 3 are still of very poor quality. The table of materials needs a lot more formatting to increase readability. Figure 1 Please calculate a Bootstrap consensus tree with the highest log likelihood and describe the precise method and values used. Figure 4: please include a In the protocols I find a lot of things confusing and not well documented. (see comments below). Especially the steps between sequencing and the final annotation are missing and are crucial steps which are also important to understand the final results. Crucial information such as Data cleaning are also missing.

Response: Thank you for your valuable suggestion. We have tried our best to increase the quality of Figure 1 and 3, and ensure they meet the requirements required for publication. Furthermore, we have used MEGA 4.0 to calculate the Bootstrap consensus tree of the sequences in Figure 1 (500 replicates; seed=64238) with the highest log likelihood. Moreover, we have refined the crucial steps in the Protocols, and complemented the crucial steps between sequencing and final annotation, as well as rewritten 1.2.8 and 1.2.9 section.

Minor Concerns:

Question 3: L100: please rewrite the medium section (1.1.1. and 1.1.2) and use one name for the medium or two if they are not the same. At the moment it doesn't make any sense if the seed medium and the fermentation medium are made from the same but then for a 10 % medium it is diluted with fermentation medium?

Response: Thank you for your valuable suggestion. We have rewritten the medium

section (1.1.1. and 1.1.2) and use one name for the medium. The growth state of mycelia is variable when the volume of medium increased. In addition, inoculation was prepared by 10% fermentation medium for scale-up culture.

Question 4: L109: Please explain how?

Response: Thank you for your valuable comment. After 10 days of asexual reproduction, the fermented medium was centrifuged, and the supernatant was discarded after centrifugation. Subsequently, the mycelia were suspended by adding 100 ml ultrapure water for 3 times, and supernatant was removed by centrifugation. Then, the cleaned mycelia were grinded into a powder using liquid nitrogen.

Question 5: L113: don't you mean pellet here? biomass? where do you add the water to?

Response: Thank you for your valuable comment. Biomass of mycilia was harvested by centrifuging the fermented medium. Subsequently, ultrapure water was added to resuspend the harvested mycelia for cleaning.

Question 6: L124: Either use no names of kits and mention all of them in the table of material or mention all in the text. In the text mention what kind of technique e.g. mRNA enrichment; add how much total rna; how do you control your successful RNA extraction, mRNA enrichment, library prep?

Response: Thank you for your valuable suggestion. We have mentioned all of them in the Table of Materials. In addition, beads with Oligo(dT) are used to isolate poly(A) mRNA after total RNA is collected from eukaryote. Fragmentation buffer is added for interrupting mRNA to short fragments. Taking these short fragments as templates, random hexamer-primer is used to synthesize the first-strand cDNA. The second-strand cDNA is synthesized using buffer, dNTPs, RNaseH and DNA polymerase I, respectively. These steps might be generally considered routine operations, and so they are not covered in great detail.

Question 7: L128: refer to your kit.

Response: Thank you for your valuable suggestion. Ultra™ RNA Library Prep Kit (NEB, United States) has been referred to generate sequencing libraries.

Question 8: L142: it should be added what kind of sequencing platform, paired end etc.

Response: Thank you for your valuable comment. The library was sequenced by Illumina HiSeq™ 4000 with paired-end sequencing in BGI. This part of the explanation has been combined to the manuscript.

Question 9: L150: Especially also for this step it is necessary to know which platform was used, which parameters. As you receive already annotated files- please refer to the software used.

Response: Thank you for your valuable suggestion. The BGI platform was used to conduct the sequencing and de novo assembly. Blastx alignment (evalue < 0.00001) between Unigenes and protein databases like nr, Swiss-Prot, KEGG and COG is performed, and the best aligning results are used to decide sequence direction of Unigenes. In addition, annotated files are in Excel format, and Microsoft Office software was referred.

Question 10: L245: Don't you mean Swiss model?

Response: Thank you for your valuable comment. We apologize for the spelling mistake. Indeed, we mean SWISS-MODEL.

Question 11: L257: with which software`?

Response: Thank you for your valuable comment. Primer Express was used to design primers, and it was mentioned in the Table of materials.

Question 12: L343: refer to report examples.

Response: Thank you for your valuable suggestion. References have been added

behind the description.

Reviewer #3:

After revision, the manuscript matches the standard for publishing in Jove.

Response: Dear reviewer, thank you so much for your evaluations on our paper.

Responses to the editor and reviewers' comments

Title: Gene Mining and Sequence Analysis of Purine Nucleosidase Based on RNA-Seq

Dear Chief Editor:
Thank you for your kind work on our paper, and thank the reviewers' comments. We have made the response to the editorial and reviewers' comments as follows.

Reviewer #1:

Accept

Response: Dear reviewer, thank you so much for your evaluations on our paper.

Reviewer #2:

Manuscript Summary:

Since the first review round of the paper on "Gene Mining and Sequence Analysis of Purine Nucleosidase Based on RNA-Seq" the manuscript still lacks a lot of data.

Response: Dear reviewer, thank you so much for your evaluations on our paper, the manuscript has been improved according to your valuable comments, and careful modifications have been made and combined to the revised manuscript.

Major Concerns:

Question 1: First of all, in the Abstract, the scientific name of the caterpillar fungus should be used. While the fungus is now named in the introduction I would recommend focusing more on the fungus and its traits than discussing TCM in detail. The same is true for the discussion e.g. in L366: why can this be just used for TCM and not for other applications? Please define PN not only in the summary but also in the Abstract and Introduction to increase readability.

Please reference all the databases used with the right literature reference. I did check for example for Swiss Model and the references are not cited in your manuscript.

Response: Thank you for your valuable comment. Follow your advice, the scientific

name of the caterpillar fungus (*Ophiocordyceps sinensis*) has been used in the Abstract. Furthermore, we have focused more on the fungus and its traits in Introduction and Discussion. Moreover, this not only can be used for TCM, but also for other applications. In addition, we have defined PN not only in the summary but also in the Abstract and Introduction to increase readability. Meanwhile, we have added the references for Swiss Model and checked all the databases used with the right literature reference.

Question 2: The figures 1 and 3 are still of very poor quality. The table of materials needs a lot more formatting to increase readability. Figure 1 Please calculate a Bootstrap consensus tree with the highest log likelihood and describe the precise method and values used. Figure 4: please include a In the protocols I find a lot of things confusing and not well documented. (see comments below). Especially the steps between sequencing and the final annotation are missing and are crucial steps which are also important to understand the final results. Crucial information such as Data cleaning are also missing.

Response: Thank you for your valuable suggestion. We have tried our best to increase the quality of Figure 1 and 3, and ensure they meet the requirements required for publication. Furthermore, we have used MEGA 4.0 to calculate the Bootstrap consensus tree of the sequences in Figure 1 (500 replicates; seed=64238) with the highest log likelihood. Moreover, we have refined the crucial steps in the Protocols, and complemented the crucial steps between sequencing and final annotation, as well as rewritten 1.2.8 and 1.2.9 section.

Minor Concerns:

Question 3: L100: please rewrite the medium section (1.1.1. and 1.1.2) and use one name for the medium or two if they are not the same. At the moment it doesn't make any sense if the seed medium and the fermentation medium are made from the same but then for a 10 % medium it is diluted with fermentation medium?

Response: Thank you for your valuable suggestion. We have rewritten the medium section (1.1.1. and 1.1.2) and use one name for the medium. The growth state of mycelia is variable when the volume of medium increased. In addition, inoculation was prepared by 10% fermentation medium for scale-up culture.

Question 4: L109: Please explain how?

Response: Thank you for your valuable comment. After 10 days of asexual reproduction, the fermented medium was centrifuged, and the supernatant was discarded after centrifugation. Subsequently, the mycelia were suspended by adding 100 ml ultrapure water for 3 times, and supernatant was removed by centrifugation. Then, the cleaned mycelia were grinded into a powder using liquid nitrogen.

Question 5: L113: don't you mean pellet here? biomass? where do you add the water to?

Response: Thank you for your valuable comment. Biomass of mycilia was harvested by centrifuging the fermented medium. Subsequently, ultrapure water was added to resuspend the harvested mycelia for cleaning.

Question 6: L124: Either use no names of kits and mention all of them in the table of material or mention all in the text. In the text mention what kind of technique e.g. mRNA enrichment; add how much total rna; how do you control your successful RNA extraction, mRNA enrichment, library prep?

Response: Thank you for your valuable suggestion. We have mentioned all of them in the Table of Materials. In addition, beads with Oligo(dT) are used to isolate poly(A) mRNA after total RNA is collected from eukaryote. Fragmentation buffer is added for interrupting mRNA to short fragments. Taking these short fragments as templates, random hexamer-primer is used to synthesize the first-strand cDNA. The second-strand cDNA is synthesized using buffer, dNTPs, RNaseH and DNA polymerase I, respectively. These steps might be generally considered routine operations, and so they are not covered in great detail.

Question 7: L128: refer to your kit.

Response: Thank you for your valuable suggestion. Ultra™ RNA Library Prep Kit (NEB, United States) has been referred to generate sequencing libraries.

Question 8: L142: it should be added what kind of sequencing platform, paired end etc.

Response: Thank you for your valuable comment. The library was sequenced by Illumina HiSeq™ 4000 with paired-end sequencing in BGI. This part of the explanation has been combined to the manuscript.

Question 9: L150: Especially also for this step it is necessary to know which platform was used, which parameters. As you receive already annotated files- please refer to the software used.

Response: Thank you for your valuable suggestion. The BGI platform was used to conduct the sequencing and de novo assembly. Blastx alignment (evalue < 0.00001) between Unigenes and protein databases like nr, Swiss-Prot, KEGG and COG is performed, and the best aligning results are used to decide sequence direction of Unigenes. In addition, annotated files are in Excel format, and Microsoft Office software was referred.

Question 10: L245: Don't you mean Swiss model?

Response: Thank you for your valuable comment. We apologize for the spelling mistake. Indeed, we mean SWISS-MODEL.

Question 11: L257: with which software`?

Response: Thank you for your valuable comment. Primer Express was used to design primers, and it was mentioned in the Table of materials.

Question 12: L343: refer to report examples.

Response: Thank you for your valuable suggestion. References have been added behind the description.

Reviewer #3:

After revision, the manuscript matches the standard for publishing in Jove.

Response: Dear reviewer, thank you so much for your evaluations on our paper.

Responses to the reviewers' comments

Title: Gene Mining and Sequence Analysis of Purine Nucleosidase Based on RNA-Seq

Reviewer #1:

Accept

Response: Dear reviewer, thank you so much for your evaluations on our paper.

Responses to the reviewers' comments

Title: Gene Mining and Sequence Analysis of Purine Nucleosidase Based on RNA-Seq

Reviewer #2:

Manuscript Summary:

Since the first review round of the paper on "Gene Mining and Sequence Analysis of Purine Nucleosidase Based on RNA-Seq" the manuscript still lacks a lot of data.

Response: Dear reviewer, thank you so much for your evaluations on our paper, the manuscript has been improved according to your valuable comments, and careful modifications have been made and combined to the revised manuscript.

Major Concerns:

Question 1: First of all, in the Abstract, the scientific name of the caterpillar fungus should be used. While the fungus is now named in the introduction I would recommend focusing more on the fungus and its traits than discussing TCM in detail. The same is true for the discussion e.g. in L366: why can this be just used for TCM and not for other applications? Please define PN not only in the summary but also in the Abstract and Introduction to increase readability.

Please reference all the databases used with the right literature reference. I did check for example for Swiss Model and the references are not cited in your manuscript.

Response: Thank you for your valuable comment. Follow your advice, the scientific name of the caterpillar fungus (*Ophiocordyceps sinensis*) has been used in the Abstract. Furthermore, we have focused more on the fungus and its traits in Introduction and Discussion. Moreover, this not only can be used for TCM, but also for other applications. In addition, we have defined PN not only in the summary but also in the Abstract and Introduction to increase readability. Meanwhile, we have added the references for Swiss Model and checked all the databases used with the right literature reference.

Question 2: The figures 1 and 3 are still of very poor quality. The table of materials needs a lot more formatting to increase readability. Figure 1 Please calculate a Bootstrap consensus tree with the highest log likelihood and describe the precise method and values used. Figure 4: please include a In the protocols I find a lot of things confusing and not well documented. (see comments below). Especially the steps between sequencing and the final annotation are missing and are crucial steps which are also important to understand the final results. Crucial information such as Data cleaning are also missing.

Response: Thank you for your valuable suggestion. We have tried our best to increase the quality of Figure 1 and 3, and ensure they meet the requirements required for publication. Furthermore, we have used MEGA 4.0 to calculate the Bootstrap consensus tree of the sequences in Figure 1 (500 replicates; seed=64238) with the highest log likelihood. Moreover, we have refined the crucial steps in the Protocols, and complemented the crucial steps between sequencing and final annotation, as well as rewritten 1.2.8 and 1.2.9 section.

Minor Concerns:

Question 3: L100: please rewrite the medium section (1.1.1. and 1.1.2) and use one name for the medium or two if they are not the same. At the moment it doesn't make any sense if the seed medium and the fermentation medium are made from the same but then for a 10 % medium it is diluted with fermentation medium?

Response: Thank you for your valuable suggestion. We have rewritten the medium section (1.1.1. and 1.1.2) and use one name for the medium. The growth state of mycelia is variable when the volume of medium increased. In addition, inoculation was prepared by 10% fermentation medium for scale-up culture.

Question 4: L109: Please explain how?

Response: Thank you for your valuable comment. After 10 days of asexual

reproduction, the fermented medium was centrifuged, and the supernatant was discarded after centrifugation. Subsequently, the mycelia were suspended by adding 100 ml ultrapure water for 3 times, and supernatant was removed by centrifugation. Then, the cleaned mycelia were grinded into a powder using liquid nitrogen.

Question 5: L113: don't you mean pellet here? biomass? where do you add the water to?

Response: Thank you for your valuable comment. Biomass of mycilia was harvested by centrifuging the fermented medium. Subsequently, ultrapure water was added to resuspend the harvested mycelia for cleaning.

Question 6: L124: Either use no names of kits and mention all of them in the table of material or mention all in the text. In the text mention what kind of technique e.g. mRNA enrichment; add how much total rna; how do you control your successful RNA extraction, mRNA enrichment, library prep?

Response: Thank you for your valuable suggestion. We have mentioned all of them in the Table of Materials. In addition, beads with Oligo(dT) are used to isolate poly(A) mRNA after total RNA is collected from eukaryote. Fragmentation buffer is added for interrupting mRNA to short fragments. Taking these short fragments as templates, random hexamer-primer is used to synthesize the first-strand cDNA. The second-strand cDNA is synthesized using buffer, dNTPs, RNaseH and DNA polymerase I, respectively. These steps might be generally considered routine operations, and so they are not covered in great detail.

Question 7: L128: refer to your kit.

Response: Thank you for your valuable suggestion. Ultra^{TM} RNA Library Prep Kit (NEB, United States) has been referred to generate sequencing libraries.

Question 8: L142: it should be added what kind of sequencing platform, paired end etc.

Response: Thank you for your valuable comment. The library was sequenced by Illumina HiSeq™ 4000 with paired-end sequencing in BGI. This part of the explanation has been combined to the manuscript.

Question 9: L150: Especially also for this step it is necessary to know which platform was used, which parameters. As you receive already annotated files- please refer to the software used.

Response: Thank you for your valuable suggestion. The BGI platform was used to conduct the sequencing and de novo assembly. Blastx alignment (evalue < 0.00001) between Unigenes and protein databases like nr, Swiss-Prot, KEGG and COG is performed, and the best aligning results are used to decide sequence direction of Unigenes. In addition, annotated files are in Excel format, and Microsoft Office software was referred.

Question 10: L245: Don't you mean Swiss model?

Response: Thank you for your valuable comment. We apologize for the spelling mistake. Indeed, we mean SWISS-MODEL.

Question 11: L257: with which software`?

Response: Thank you for your valuable comment. Primer Express was used to design primers, and it was mentioned in the Table of materials.

Question 12: L343: refer to report examples.

Response: Thank you for your valuable suggestion. References have been added behind the description.

Responses to the reviewers' comments

Title: Gene Mining and Sequence Analysis of Purine Nucleosidase Based on RNA-Seq

Reviewer #3:

After revision, the manuscript matches the standard for publishing in Jove.

Response: Dear reviewer, thank you so much for your evaluations on our paper.