Tehran University of Medical Sciences Publication
http://tums.ac.ir

Iran J Parasitol

Open access Journal at
http://ijpa.tums.ac.ir

Iranian Society of Parasitology
http://isp.tums.ac.ir

Original Article

Next-Generation Sequencing Reveals that Oxidative Phosphorylation Might Be a Key Pathway Differently Expressed in the Third and Fourth Stages Larvae of *Angiostrongylus cantonensis*

***Yue GUO [1,2], Hong Chang ZHOU [1,2], Ying DONG [1], Hai Yan DONG [1,2], Yun Liang YAO [1,2], Jing QIAN [1,2], Hui ZHANG [1], Xiao Yu LI [1], Zhong Shan ZHANG [2], Han Bing LIN [1], Tian ZHOU [1], Meng Jia ZHAO [1], Tang Qin JI [1], Run Ze WANG [1], Feng Ping ZHANG [1]**

1. *School of Medicine, Huzhou University, Huzhou Cent Hosp, 759 Er Huan Rd, Huzhou, Zhejiang, China*
2. *Key Laboratory of Vector Biology and Pathogen Control of Zhejiang Province, Huzhou University, Huzhou, Zhejiang, China*

Received 16 Apr 2020
Accepted 11 Jun 2020

Keywords:
A. cantonensis;
Transcriptional sequencing;
Third stage larvae;
Fourth stage larvae;
Next-generation sequencing

***Correspondence**
Email:
guoyue66@126.com

Abstract

Background: When *Angiostrongylus cantonensis* develops from the third and fourth stage, it needs to change its host from the middle host, snail to the final host, rat. However, the mechanism involved in this change remains unclear.

Methods: The transcriptome differences of the third and fourth stages of *A. cantonensis* were explored by next-generation Illumina Hiseq/Miseq sequencing in China, in 2018.

Results: Overall, 137 956 488 clean reads and 20 406 213 373 clean bases of the two stages larvae were produced. Based on the queries against the Gene Ontology (GO), NCBI non-redundant protein sequences (Nr), Swissprot, and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases, 14 204 differentially expressed genes (DEGs) were predicted. GO enrichment analysis revealed 5660 DEGs with the top s categories as followings: biological process (GO:0008150, related to 5345 DEGs), cellular component (GO:0005575, related to 5297 DEGs), molecular function (GO:0003674, related to 5290 DEGs). In KEGG enrichment analysis, 116 genes were related to oxidative phosphorylation and 49 genes involved in the glycolytic process.

Conclusion: Metabolism changes, especially oxidative phosphorylation and glycolysis, might play a key role in *A. cantonensis* infection of its final rat host. Many other pathways might also contribute to the transcriptome changes between these two life stages. Overall, additional studies are needed for further details.

Introduction

The worm *Angiostrongylus cantonensis* is an animal parasite first identified in the lung of *Rattus rattus* and *R. norvegicus* in China, in 1933(1). Human infection of *A. cantonensis* can cause angiostrongyliasis. Recent reports have indicated that this disease has been spreading to many other non-endemic countries and regions such as France, Germany, the Caribbean (including Jamaica), Brazil, Ecuador, and South Africa. Until 2018, at least 3,000 cases of angiostrongyliasis were reported in 31 countries (2, 3). Thus, *A. cantonensis* has become a new threat to health worldwide.

In nature, *A. cantonensis* usually employs some species of rats as the final host and some snails as the middle host. Its life cycle includes egg, five stages of larvae, and adult worms (3, 4). The five stages of larvae are the first, second, third, fourth, and fifth stage, respectively. Among them, the third stage (L3) lives in the lung sac of its intermediate snail host and fourth stage (L4) parasites reside in the brain of the final host; rat. From L3 to L4, *A. cantonensis* changes its host from snail to rat, adapting to a new living environment while facing new threats from the rat immune system. During this period of infection, the parasite transcriptome would be changed rapidly, but the details of these changes remain unclear.

In recent years, next-generation sequencing (NGS) has provided a popular and powerful tool to reveal the transcriptome of parasites (5, 6). For example, in nematode, NGS was used to diagnose infections and the control of parasitism. For some species of nematodes with agricultural, aquacultural, and medical importance, NGS provides an opportunity to further understand these pathogens with respect to taxonomy, epidemiology, and evolutionary history (7).

Here we employed Illumina Hiseq/Miseq sequencing to reveal the transcriptome differences between the third and fourth stages of *A. cantonensis*.

Materials and Methods

Animals

To complete the life cycle of *A. cantonensis* in the laboratory, the apple snail *Pomacea canaliculate* (*P. canaliculate*) was used as the middle host and Sprague-Dawley rats (SD rats) were used as the final host. Rats were housed in clean animal rooms and provided with food and water for 24 h.

This research was conducted in strict accordance with the Regulations for the Administration of Affairs Concerning Experimental Animals (#201800527-SGY03， as approved by the Animal Ethics Committee of Huzhou University).

Sample collection

The lung sac of *A. cantonensis* positive *P. canaliculate* was collected and torn into pieces, then L3 was collected manually using a microscope. In total, 700 L3 were used for transcriptional sequencing. Two *A. cantonensis* positive SD rats were anesthetized to collect the fourth stage larvae. Each rat provided 200 L4 for sequencing. All the live worms were washed thrice with phosphate-buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4), immersed in RNA hold reagent (Transgene Biotech, Beijing, China) and stored in liquid nitrogen.

RNA extraction, cDNA library preparation and RNA sequencing

RNA from all samples was extracted using Trizol reagent (Invitrogen) after being ground by a homogenizer. RNA was extracted according to the manufacturer's instructions. The concentration and OD values of extracted RNA were determined using Nanodrop 2000

(Thermo Scientific CA). RNA agarose gel electrophoresis was also performed to examine the integrity of total RNA, and Agilent 2100 was used for RIN.

At least 5 μg of total RNA, with concentration higher than 200 ng/μL and OD values of 1.8–2.2 were used for the following tests. mRNA was enriched using Oligo (dT) beads which can specifically bind to the poly A tail of mRNA. Enriched mRNA was then placed into fragmentation buffer to be randomly digested into 1.5–2 kb fragments. Then mRNA was reverse transcribed to cDNA, followed by double-stranded DNA synthesis. After the double-stranded DNA was blunt-ended using End Repair Mix, Illumina Hiseq/Miseq sequencing was performed.

Gene expression quantification

Each cycle of all sequencing reads underwent statistical analysis to obtain the base distribution and quality fluctuation, which can intuitively reflect the quality of sequencing data and library construction. The quality of raw data was tested by the Q (Quality) value (8, 9). Raw data was filtered to obtain clean data using SeqPrep (https://github.com/jstjohn/SeqPrep) and Sickle (https://github.com/najoshi/sickle). In this step, the adapter sequences, low quality reads, high N-rate sequences, and other data that might seriously affect the quality of subsequent assembly were removed (10).

De novo assembly and annotation of clean data

De novo assembly of all clean data was performed to generate two groups, contigs and singletons, using the software Trinity (http://trinityrnaseq.sourceforge.net/, version number: trinityrnaseq-r2013-02-25). After *de novo* assembly, *de novo* annotation was performed. Briefly, the ORF prediction process provided by Trinity software (Http://trinityrnaseq.sourceforge.net/analysis/extract_proteins_from_trinity_transcripts.html) was used to predict the genes of interest. Then Blast X (Version 2.2.25) was employed to obtain the annotation information of all assembled transcript sequences. The querying databases included NCBI non-redundant protein sequences (Nr), Swissprot, Gene Ontology (GO), and the Kyoto Encyclopedia of Genes and Genomes (KEGG). The E-value cutoff was set as 1.0×10^{-5}.

Differentially expressed gene (DEGs) screening

Differential expressed genes between L3 and L4 were obtained by edgeR (http://www.bioconductor.org/packages/2.12/bioc/html/edgeR.html), a software used for differential gene expression analysis. EdgeR uses the gene read count data provided by RSEM to perform differential expression calculations (11). The analysis method was based on a negative binomial distribution model, and the screening criteria for significant differentially expressed genes were set as: FDR < 0.05 & | $\log_2$FC (fold change) | ≥ 1.

GO and KEGG pathway enrichment analysis of DEGs

GO and KEGG pathway enrichment analysis was used to obtain the annotation information of DEGs. The two analyses was implemented using the GOseq R package (v1.10.0) based on a Wallenius non-central hyper-geometric distribution and the KEGG database (http://www.genome.jp/kegg/) (12).

Results

Raw and clean data of sequencing, unigenes, and transcripts

Overall, 13.09 Gb (3.47 Gb from L3, 4.94 Gb from L4, and 4.68 Gb from 3-L4) of clean data were used for *de novo* assembly. In total, 139 009 448 raw reads and 20 851 417 200 raw bases of the two stages of worms were obtained by Illumina sequencing. After being filtered by SeqPrep (https://github.com/jstjohn/SeqPrep) and

Sickle (https://github.com/najoshi/sickle), 137 956 488 clean reads and 20 406 213 373 clean bases were produced, as shown in Table 1. Using Trinity (http://trinityrnaseq.sourceforge.net/, version number: trinityrnaseq-r2013-02-25) analysis, 10135 unigenes and 10646 transcripts were obtained in L3, and 128668 unigenes and 193060 transcripts were obtained in L4.

Table 1: Total reads and bases obtained by Illumina sequencing (L3 represents the third stage larvae, L4 refers to the fourth stage larvae.)

Sample	*Raw Reads*	*Clean Reads*	*Raw Bases*	*Clean Bases*	*Q20 of clean bases*
L3	40656794	40284860	6098519100	5939940748	94.62
L4	58352654	57718612	8752898100	8514071361	94.62
2-1L4	40000000	39953016	6000000000	5952201264	98.15
Total	139009448	137956488	20851417200	20406213373	

Differentially expressed genes (DEGs)

After *de novo* assembly and annotation of clean data followed by gene expression estimation, DEGs between L3 and L4 stages were obtained using edgeR (http://www.bioconductor.org/packages/2.12/bioc/html/edgeR.html). In total, 14 024 differentially expressed genes between L3 and L4 were obtained. Gene expression of third and fourth stage *A. cantonensis* was showed in Fig. 1A.

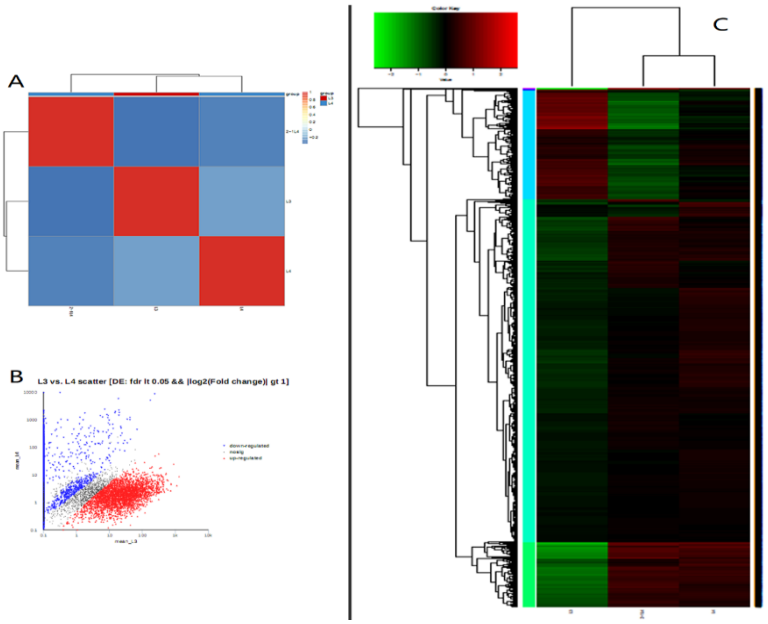

Fig. 1: A. Heatmap showing gene expression of third and fourth stage *A. cantonensis*. L3 refers to the third stage, L4 is the fourth stage (L4 has two biological replicates, 2-1L4 and L4). Gene expression levels were determined by RSEM.

B. Scatter plot of differentially expressed genes between L3 and L4. Red, blue, and black splashes represent downregulated, non-significant, and upregulated genes.

C. Heatmap of differential mRNA expression between samples. Each column and row represent a sample and a gene respectively. The color indicates the gene expression level in the sample. Red and green indicate high and low expression, respectively

DEG annotation by GO and KEGG

The GO assigned terms belonged to 3 categories: Biological Process, Cellular Component, Molecular Function. The top 5 upregulated categories obtained by GO annotation were cell (1393 genes related), cell part (1392 genes related), cellular process (1359 genes related), metabolic process (1259 genes related), and catalytic activity (1255 genes related), and the top 5 downregulated categories were cell (3718 genes concerning), cell part (3714 genes concerning), cellular process (3590 genes concerning), organelle (3395 genes concerning), and binding (3220 genes concerning). The top related items of Biological Process, Cellular Component, and Molecular Function are visualized and shown in Fig. 2A.

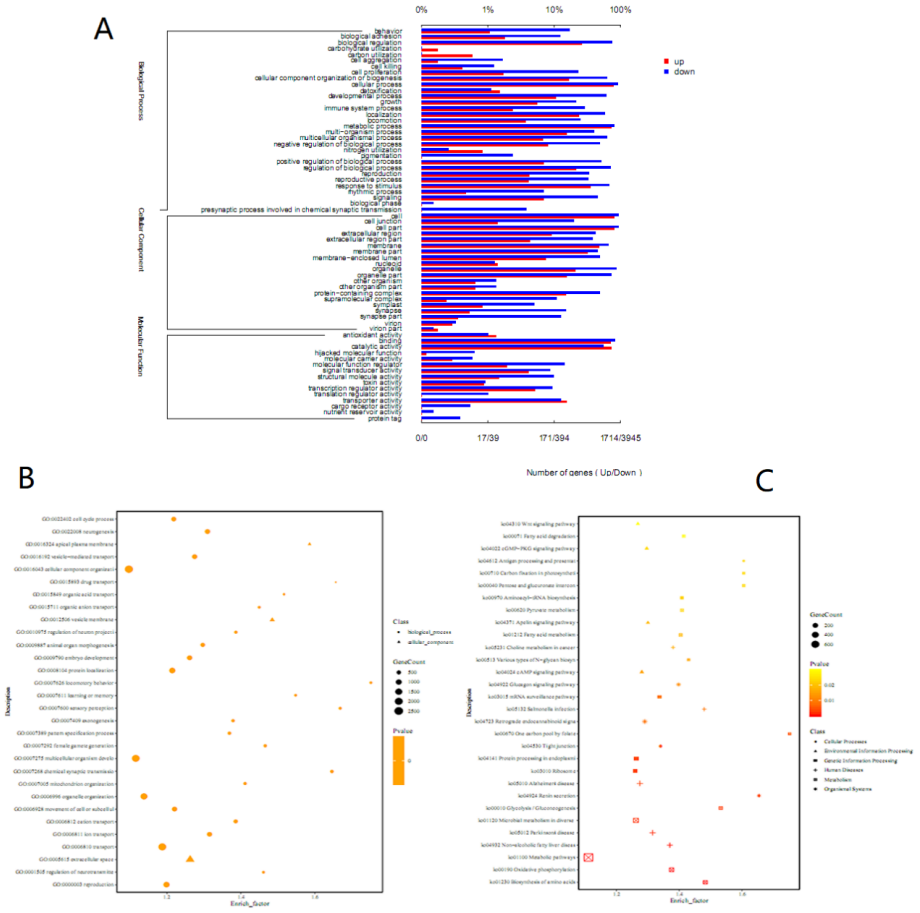

Fig. 2: A. GO annotation of differentially expressed genes. The lower abscissa indicates the number of genes annotated to a GO term, and the upper abscissa indicates the proportion of the number of genes among all the genes involved in all GO terms. Red and blue represent upregulated and downregulated genes, respectively.
B. Scatter plot of differentially expressed genes by GO enrichment analysis. Each icon represents one GO term. Abscissa indicates the enrichment rate. Size of the icons indicates the number of genes involved. The color indicates the significance of enrichment, P value.
C. Scatter plot of differentially expressed genes by KEGG enrichment analysis. Each icon represents one KEGG pathway. Abscissa indicates enrichment rate. Size of the icons represents the number of genes involved. The color indicates the significance of enrichment, *P* value

The top 5 pathways obtained from KEGG analysis were related to metabolic pathways (9587 pathways related, path:ko01100), Biosynthesis of secondary metabolites (4119 pathways related, path:ko01110), Biosynthesis of antibiotics (3173 pathways related, path:ko01130), Microbial metabolism in diverse environments (2781 pathways related, path:ko01120), and Carbon metabolism (2020 pathways related, path:ko01200).

GO and KEGG enrichment analysis of DEGs

GO enrichment analysis revealed 5660 DEGs and the top 5 categories were as follows: biological process (GO:0008150, related to 5345 DEGs), cellular component (GO:0005575, related to 5297 DEGs), molecular function (GO:0003674, related to 5290 DEGs), cell (GO:0005623, related to 5112 DEGs), and cell part (GO:0044464, related to 5107 DEGs). In total, 2,345 DEGs were functionally assigned to 364 pathways by KEGG enrichment pathway analysis. The top five pathways were as follows: Metabolic pathways (ko01100, concerning 780 DEGs), Biosynthesis of secondary metabolites (ko01110, concerning 297 DEGs)，Biosynthesis of antibiotics (ko01130, concerning 213 DEGs), Microbial metabolism in diverse environments (ko01120, concerning 171 DEGs), and Huntington's disease (ko05016, concerning 122 DEGs). GO and KEGG enrichment analysis results are exhibited as scatter figures in Fig. 2B and Fig. 2C.

Oxidative phosphorylation and glycolysis

In KEGG enrichment analysis, 116 genes were related to oxidative phosphorylation (ko00190, p value: 0.000116674455475099) suggesting that DEGs related to oxidative phosphorylation were statistically significant. GO enrichment analysis indicated oxidative phosphorylation as an important biological process among all DEGs. Nine members of oxidative phosphorylation related IDs were GO:0017077, GO:0090324, GO:2000276, GO:2000277, GO:0006119, GO:2000275, GO:0002082, and GO:1903862.

KEGG enrichment analysis exhibited 49 genes involved in the glycolytic process (ko00010), and the P value was 0.002338682, indicating that glycolysis was different in the two life stages with statistical significance (Fig. 3). GO enrichment analysis also exhibited 1 gene concerning glycolysis (GO:0093001), with statistical significance between two stages (p value 0.67715768). These 2 analysis results indicated that glycolysis was not among the top 5 or 10 pathways or classed in KEGG enrichment or GO enrichment analysis, but this process might play an important role in parasite metabolism.

Discussion

As a parasitic worm, *A. cantonensis* has a complex life history compared to other opportunistic worm parasites. Among all life stages of *A. cantonensis*, the third and fourth stages are typically special.

The L3 of *A. cantonensis* are about 0.45 by 0.03 mm in body size. The L4 of *A. cantonensis* has a length around 1.0 mm and 0.04 mm width (4). *A. cantonensis* is parasitic in the lung sac of the middle host snail at the third stage and the brain of SD rats at the fourth stage. Apparently, the body size and living environment of *A. cantonensis* change dramatically when it completes development in snail host and successfully infects the rat host. This shift in hosts brings new changes to *A. cantonensis*. The snail host is an invertebrate with a relatively low immune system, whereas the rat is an advanced mammal with a complex immune system.

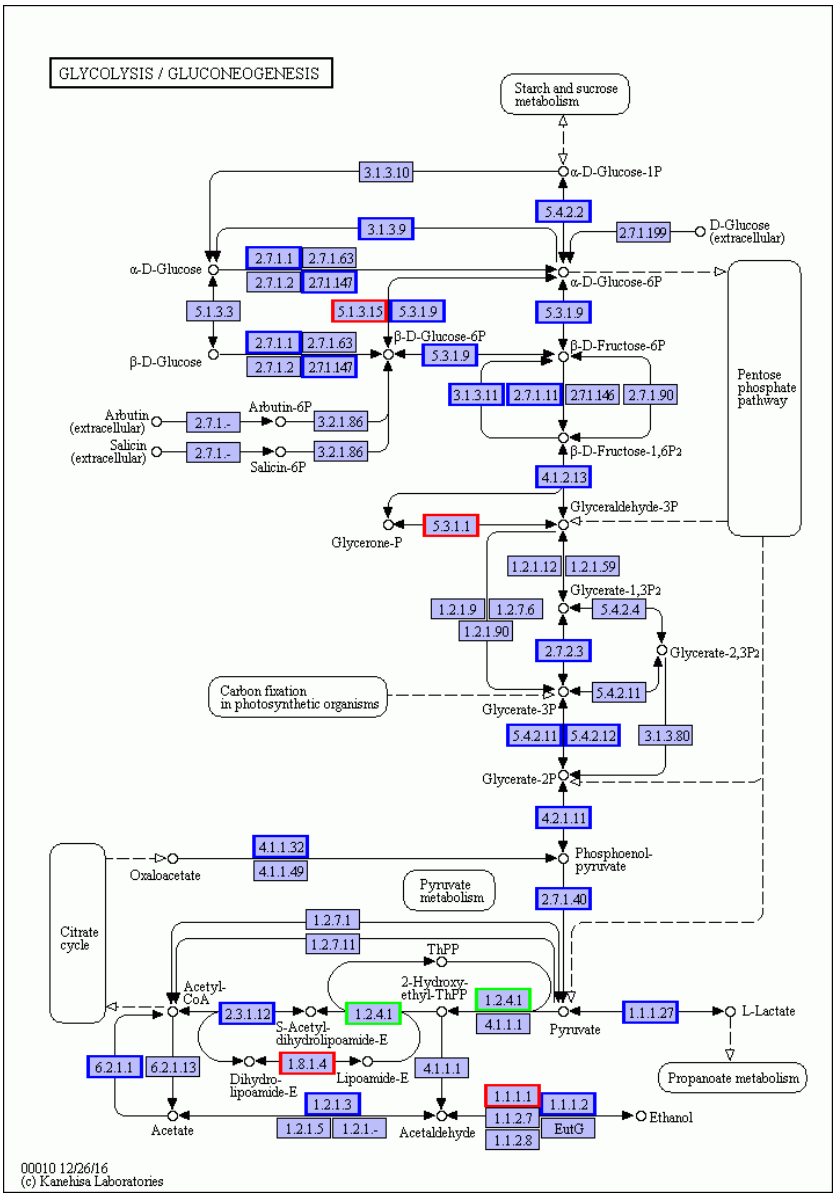

Fig. 3: Glycolysis KEGG annotation pathway of upregulated and downregulated genes. The blue and white background of the box indicates the genes detected and not-detected in this sequencing. Blue and red framed lines of boxes represent downregulated and upregulated genes

Those two hosts are also different in body temperature; the snail is a cold-blooded animal whereas the rat is a homeothermic animal. In addition, the blood circulation system, blood components, tissue components, and oxygen content of the two hosts differ from each other. Once successfully invaded into the rat body, *A. cantonensis* needs to face large changes in oxygen concentration, temperature, moisture, nutrition, and immune evasion.

Here, our sequencing results and both GO and KEGG enrichment analysis of DEGs suggested that metabolism was the most involved term from L3 to L4, especially oxidative phosphorylation and glycolysis. As mentioned in the results, eight GO terms and 1

KEGG enriched pathway were involved in oxidative phosphorylation.

Oxidative phosphorylation occurs in the mitochondrial inner membrane of eukaryotic cells.

Oxidation refers to the decomposition process of organic substances including sugars, lipids, amino acids, and so on. Phosphorylation refers to the role of ATP generation in biological oxidation. There are two types of phosphorylation, metabolite-linked phosphorylation and respiratory chain-linked phosphorylation. Oxidative phosphorylation involves different substrates and different enzymes in different oxygen concentration situations. Many reports indicate that oxidative phosphorylation could occur in its special way by utilizing different enzyme systems to face the hypoxic environment in some parasites including *Ascaris suum*, *Trichuris muris*, *Schimdtea mediterranea*, and *Fasciola hepatica* (13) (14). Other research studies have shown that cancer cells use different oxidative phosphorylation methods to satisfy their rapid growth and division (15-17). Our results also suggest that *A. cantonensis* might use a different pathway to produce energy for survival in the new host.

Glycolysis was also indicated by GO and KEGG enrichment analysis. Glycolysis is a pathway of oxidative phosphorylation under anaerobic or hypoxic conditions. Inside the body of organisms, there are three main pathways for the oxidative decomposition of sugars, anaerobic oxidation of sugars, aerobic oxidation of sugars, and the pentose phosphate pathway. Among them, the anaerobic oxidation of sugar is also called glycolysis. During glycolysis, glucose is degraded into lactic acid, and then a small amount of ATP under anaerobic or hypoxic conditions is produced. DEGs of glycolysis represented the survival strategies of third and fourth stage larvae were different. This might be caused by the different oxygen concentration between the snail lung sac and rat brain. After successful invasion into the rat host, *A. cantonensis* probably chooses a new metabolism strategy to survive under anaerobic conditions.

Here we used next-generation Illumina sequencing to analyze the transcriptome differences between L3 and L4 of *A. cantonensis*. Although, some biological differences were observed, the support was not solid, due to the lack of other strong evidence by qPCR (Real-time Quantitative PCR Detecting System), which will be performed in further studies. However, Illumina sequencing also provided some invalid information, possibly due to the following 2 reasons. First, although the genome information of *A. cantonensis* was reported, the assembly of this genome and the annotation information of this parasite are still incomplete. Henceforth, we used *de novo* assembly, which also provided information from other species of organisms (18-20). Second, Illumina sequencing outcomes would be affected by many other reasons and even small differences in samples could be enlarged and reflected in the sequencing results and analysis. However, this research provided some valuable information to distinguish the transcriptome differences between L3 and L4 of *A. cantonensis*.

Conclusion

Metabolism changes, especially oxidative phosphorylation and glycolysis, might play a key role in *A. cantonensis* infection of its final rat host. Many other pathways might also contribute to the transcriptome changes between these two life stages.

Acknowledgements

This research was supported by the Natural Science Funds of Zhejiang Province (LY15H190007& LQ15C040003), grants from the Key Laboratory of Vector Biology and Pathogen Control of Zhejiang Province, Huzhou University (Grant No. HUZUL201902), as well as research funds of Huzhou Universi-

ty (2019XJKJ41&2019XJKJ43).

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. HT C. A preliminary report on a survey of animal parasites of Canton, China. Lingnan Sci J. 1933;12:65-74.
2. Morassutti AL, Perelygin A, De Carvalho MO, et al. High throughput sequencing of the Angiostrongylus cantonensis genome: a parasite spreading worldwide. Parasitology. 2013;140(10):1304-1309.
3. Barratt J, Chan D, Sandaradura I, et al. *Angiostrongylus cantonensis:* a review of its distribution, molecular biology and clinical significance as a human pathogen. Parasitology. 2016;143(9):1087-1118.
4. Jarvi SI, Quarta S, Jacquier S, et al. High prevalence of *Angiostrongylus cantonensis* (rat lungworm) on Eastern Hawai island: A closer look at life cycle traits and patterns of infection in wild rats (*Rattus* spp.). PLoS One. 2017;12(12):e0189458.
5. Wain J, Mavrogiorgou E. Next-generation sequencing in clinical microbiology. Expert Rev Mol Diagn. 2013;13(3):225-7.
6. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010;11(1):31-46.
7. Jex AR, Littlewood DT, Gasser RB. Toward next-generation sequencing of mitochondrial genomes--focus on parasitic worms of animals and biotechnological implications. Biotechnol Adv. 2010;28(1):151-159.
8. Cock PJ, Fields CJ, Goto N, et al. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res. 2010;38(6):1767-1771.
9. Erlich Y, Mitra PP, McCombie WR, et al. Alta-Cyclic: a self-optimizing base caller for next-generation sequencing. Nat Methods. 2008;5(8):679-82.
10. Grabherr MG, Haas BJ, Yassour M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644-52.
11. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.
12. Kanehisa M, Araki M, Goto S, et al. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 2008;36(Database issue):D480-4.
13. Takamiya S, Hashimoto M, Mita T, et al. Bioinformatic identification of cytochrome b5 homologues from the parasitic nematode Ascaris suum and the free-living nematode *Caenorhabditis elegans* highlights the crucial role of A. suum adult-specific secretory cytochrome b_5 in parasitic adaptation. Parasitol Int. 2016;65(2):113-120.
14. Salinas G, Otero L, Martinez-Rosales C, et al. Complex I and II Subunit Gene Duplications Provide Increased Fitness to Worms. Front Genet. 2019;10:1043.
15. Solaini G, Sgarbi G, Baracca A. Oxidative phosphorylation in cancer cells. Biochim Biophys Acta. 2011;1807(6):534-542.
16. Weber GF. Metabolism in cancer metastasis. Int J Cancer. 2016;138:2061-2066.
17. Giampazolias E, Tait SW. Mitochondria and the hallmarks of cancer. FEBS J. 2016;283(5):803-814.
18. Liu J, Li G, Chang Z, et al. BinPacker: Packing-Based De Novo Transcriptome Assembly from RNA-seq Data. PLoS Comput Biol. 2016;12(2):e1004772.
19. Moreton J, Izquierdo A, Emes RD. Assembly, Assessment, and Availability of De novo Generated Eukaryotic Transcriptomes . Front Genet. 2016;6:361.
20. Fukushima A, Nakamura M, Suzuki H, et al. High-Throughput Sequencing and De Novo Assembly of Red and Green Forms of the Perilla frutescens var. crispa Transcriptome. PLoS One. 2015;10(6):e0129154.