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

Short Communication

Transcriptome Profiling and Bioinformatic Analysis of the Fourth-Stage Larvae of *Angiostrongylus cantonensis*

***Yue Guo [1,2], Hui Zhang [1], Haiyan Dong [1,2], Ming Ke Lu [3], Xiaoyu Li [1]**

1. *School of Medicine, Huzhou University, Huzhou, Zhejiang, China*
2. *Key Laboratory of Vector Biology and Pathogen Control of Zhejiang Province, Huzhou University, Huzhou, Zhejiang, China*
3. *School of Life Sciences, Xiamen University, Fujian, China*

Received 17 Jul 2022
Accepted 25 Sep 2022

Keywords:
Angiostrongylus cantonensis;
Fourth-stage larvae;
Transcriptome;
Next-generation sequencing

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

Abstract

Background: To explore the transcriptome profiling of the fourth-stage larvae of *Angiostrongylus cantonensis*.

Methods: Two groups of fourth-stage larvae were collected to extract total RNA in Zhejiang, China 2020. Then, mRNA was separated and reverse transcribed into cDNA. Next-generation sequencing was used to explore the transcriptome information. Finally, to obtain the biological annotation information, the transcriptome information was run against the related databases, including Nr, GO, COG, KOG and ORF.

Results: Overall, 128667 unigenes and 193059 transcripts were obtained. The Nr annotations of unigenes and transcripts showed that *A. cantonensis* was the 5th and 4th most related species, respectively. Meanwhile, the annotation of unigenes and transcripts by querying GO, COG, KOG and ORF showed that L4 was extremely active in gene expression, concerning signal transduction, transcription, posttranslational modification, metabolism, etc.

Conclusion: The fourth-stage larvae of *A. cantonensis* have their own profiling in the transcriptome, which is related to signal transduction, transcription, posttranslational modification, metabolism, etc.

Introduction

The nematode lungworm of rats *Angiostrongylus cantonensis* is the causative agent of angiostrongyliasis (1), which is a world distributed infection. Although the naturally suitable final host of *A. cantonensis* is rats (2), human infection occurred in recent years.

A. cantonensis has a complex life cycle composed of 5 stages of larvae and an adult stage. The 3rd larvae (L3) stage is infectious to the final host. After invading the body of the final host, the larvae break thought the brain blood barrier (BBB) to develop into the 4th stage larvae (L4). In the end, the larvae leave and settle in the lung to mature into the adult stage. As for multicellular parasites, worms are scarcely seen living in the brain of the final mammalian host. In order to adapt to the special living environment, L4 *A. cantonensis* expresses different transcriptional profile (2).

In recent years, next-generation sequencing (NGS) provided an opportunity to explore the transcriptome profiling. In order to unlock the transcriptional profile of L4 *A. cantonensis*, this work used Illumina hiseq TM 2000 to perform genome sequencing. Then, transcriptional annotation was performed on the results using the databases of Nr, GO, COG, KOG and ORF.

Methods and Materials

L3 *A. cantonensis* was used to infect the SD rats in Zhejiang, China 2020 (#20201011201-SGY03, approved by the Animal Ethics Committee of Huzhou University).

Anesthetized infected rats were humanely euthanized to harvest the L4 *A. cantonensis*. Three biological repetitions were included in this study. Each rat represented 1 group of *A. cantonensis* with 100 worms.

Total RNA extraction

RNA was extracted according to general protocol of RNA extraction by Trizol. L3 groups were washed 3 times by PBS and transferred into 1.5 ml Trizol to ground on ice. Extracted RNA was re-dissolved by 20-40 µl of DEPC-treated-water and then stored in -80 ℃ freezer. Meanwhile, the quality of total RNA was tested. The purity and concentration of RNA were tested using a Nana-drop, the integrity of RNA was detected through gel electrophoresis, and the RIN value was obtained by Agilent 2100. The sample concentration was ≥ 200 ng/µL, while the OD value was in the range of 1.8~2.2. At least 5 µg of RNA was needed.

Sequencing

We separated mRNA from total RNA by a magnetic bead with Oligo (dT). Long mRNA was divided into 200 bp fragments by a fragmentation buffer. Then, cDNA was synthesized using fragmented mRNA as a template. Finally, dsDNA was synthesized, and all samples were sequenced using Illumina Hiseq/Miseq.

Annotation of sequencing results and expression calculation

The original read sequence produced by Illumina Hiseq/Miseq was filtered to generate raw data and clean data. Then, the Trinity method was used to assemble data into transcriptome information, including the unigenes and transcripts; both were annotated using Gene Ontology (GO) as well as the NR and COG/ KOG databases.

Results

Reading assembling

Information of unigenes and transcripts are showed in Table 1. The length of transcripts was mainly distributed in 201-600 bp, with 101204 transcripts (52.42%) in the range of 201-400 bp and 24010 transcripts (12.44%) in the range of 401-600 bp. On the other hand,

the length of assembled unigenes was mainly in the range of 1-600 bp, with 89545 unigenes (69.59%) in the range of 1-400 bp and 17256 unigenes (13.41%) in the range of 401-600 bp.

Table 1: Unigenes and transcripts resulting from the assembly

Type	*Unigene*	*Transcripts*
Total sequence num	128667	193059
Total sequence base	69644863	181206178
Percent GC	44.57	43.55
Largest	26376	26376
Smallest	201	182
Average	541.28	938.61
N50	694	2133
N90	241	301

Annotations of unigenes and transcripts

To annotate the assembled unigenes and transcripts, the following databases were quarried: Nr, Swissprot, String, COG, KOG, NOG and GO databases.

Nr annotations of unigenes and transcripts

Regarding the unigenes on the species level, the top 5 most related species were: *Ancylostoma ceylanicum* with 5056 annotated unigenes, *Dictyocaulus viviparus* with 1921 annotated unigenes, *Haemonchus contortus* with 1876 annotated unigenes, *Diploscapter pachys* with 1684 annotated unigenes and *A. cantonensis* with 1554 annotated unigenes, as shown in Fig. 1A. For transcripts, the top 5 most related species were: *Ancylostoma ceylanicum* (14357 annotated transcripts), *D. viviparus* (8649), *H. contortus* (7415), *A. cantonensis* (2598) and *Necator americanus* (2528), as shown in Fig. 1C.

GO annotations of unigenes and transcripts

Unigene annotation by GO level 2 resulted in the following top 5 most relevant terms. Cellular components (CC) terms: GO:0005623 concerning 12933 unigenes, GO:0044464 linked with 12926 unigenes. Biological process (BP) terms: GO:0009987 annotated with 12452 unigenes, GO:0008152 concerning 11086 unigenes. Molecular function (MF) terms: GO:0005488 related to 11234 unigenes.

Regarding the transcripts, the following 5 GO terms were the most highly annotated. Cellular components (CC) terms: GO:0005623 annotated with 28076 transcripts, GO:0044464 related with 28060 transcripts. Biological process (BP) terms: GO:0009987 linked with 27068 transcripts, GO:0008152 concerning 23963 transcripts. Molecular function (MF) terms: GO:0005488 annotated with 24231 transcripts. The results are shown in Fig. 1B,D.

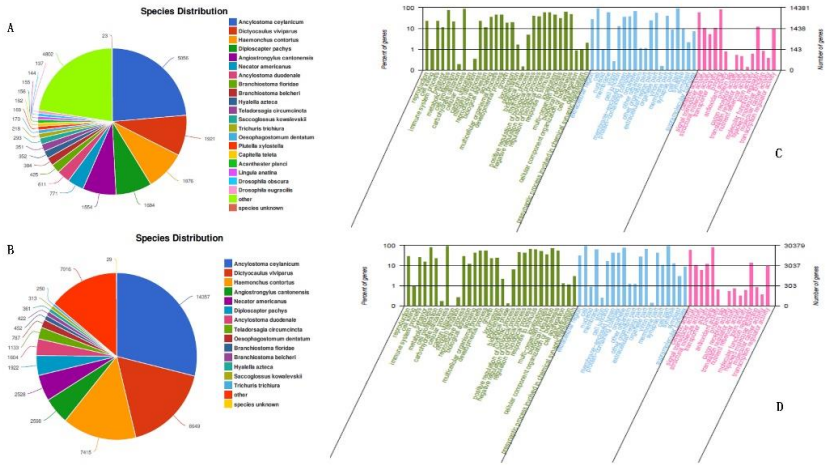

Fig. 1: Annotations of unigenes and transcripts provided by Nr and GO. **A.** Unigene annotation by the Nr database on the species level; **B.** Unigene annotation by the GO database; **C.** Transcript annotation by the Nr database on the species level; **D.** Transcript annotation by the GO database

COG annotation of unigenes and transcripts

The functional annotation of unigenes by the COG database exhibited the following top 5 items: 1469 unigenes involved in [R] General function prediction only of POORLY CHARACTERIZED, 1291 unigenes concerning the [T] Signal transduction mechanisms of CELLULAR PROCESSES AND SIGNALING, 883 unigenes related to [L] Replication, recombination and repair of INFORMATION STORAGE AND PROCESSING, 875 unigenes linked with [E] Amino acid transport and metabolism of METABOLISM, 793 unigenes related with [C] Energy production and conversion of METABOLISM.

By querying the COG database, the following functional annotations of transcripts were showed as the top 5 most related ones: 3185 transcripts with a function of [R] General function prediction only of POORLY CHARACTERIZED type, 2593 transcripts related to the function of [T] Signal transduction mechanisms of CELLULAR PROCESSES AND SIGNALING, 1742 transcripts with the function of [L] Replication, recombination and repair of INFORMATION STORAGE AND PROCESSING, 1519 transcripts involved in [E] Amino acid transport and metabolism of METABOLISM, 1429 transcripts related to [O] Posttranslational modification, protein turnover, chaperones of CELLULAR PROCESSES AND SIGNALING. The results are shown in Fig. 2B,D.

KOG annotation of unigenes and transcripts

Querying the KOG database resulted in the following: [T] Signal transduction mechanisms of CELLULAR PROCESSES AND SIGNALING, [R] General function prediction only of POORLY CHARACTERIZED, [K] Transcription of INFORMATION STORAGE AND PROCESSING.

As for the transcripts, the KOG database querying results were as following: [T] Signal transduction mechanisms of CELLULAR PROCESSES AND SIGNALING, [R] General function prediction only of POORLY CHARACTERIZED, [K] Transcription of INFORMATION STORAGE AND PROCESSING, [O] Posttranslational modification, protein turnover etc. The results are shown in Fig. 2A,C.

The top 10 highly annotated unigenes and transcripts provided by ORF (open reading frame) annotation are listed in Tables 2 and 3.

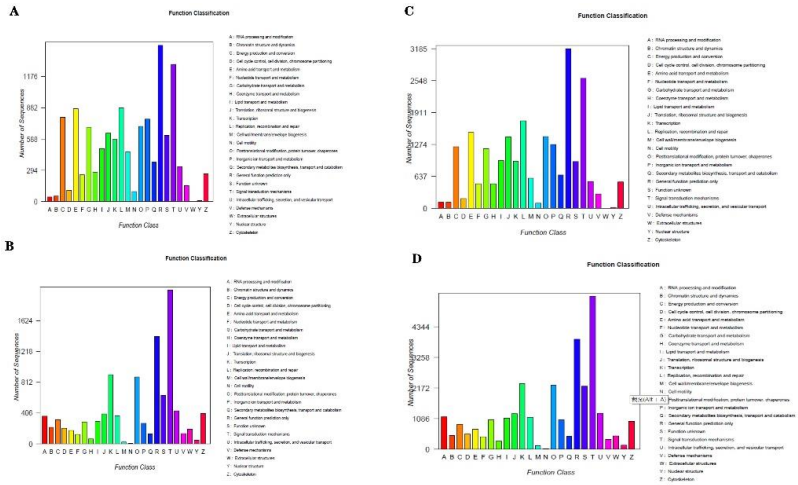

Fig. 2: Annotations of unigenes and transcripts provided by COG and KOG. **A.** Unigene annotation by COG; **B.** Unigene annotation by KOG; **C.** Transcript annotation by COG; **D.** Transcript annotation by KOG

Table 2: The top 10 most-related unigene provided by ORF annotation

Seq_id	*Protein_id*	*Pfam_id*	*Domain*	*DomainDescription*	*ProteinStart*	*ProteinEnd*	*PfamStart*
TRINITY_DN703_c0_g1	TRINITY_DN703_c0_g1_i3|m.1524	PF13634.3	Nucleoporin_FG	Nucleoporin FG repeat region	22	133	3
TRINITY_DN86386_c0_g1	TRINITY_DN86386_c0_g1_i2|m.1138	PF08238.9	Sel1	Sel1 repeat	280	291	27
TRINITY_DN1209_c0_g3	TRINITY_DN1209_c0_g3_i8|m.7588	PF03160.11	Calx-beta	Calx-beta domain	105	121	72
TRINITY_DN47674_c0_g1	TRINITY_DN47674_c0_g1_i1|m.1512	PF01442.15	Apolipoprotein	Apolipoprotein A1/A4/E domain	5	18	168
TRINITY_DN123814_c0_g1	TRINITY_DN123814_c0_g1_i1|m.41	PF01442.15	Apolipoprotein	Apolipoprotein A1/A4/E domain	53	61	38
TRINITY_DN9327_c0_g2	TRINITY_DN9327_c0_g2_i1|m.2176	PF01581.13	FARP	FMRFamide related peptide family	70	74	7
TRINITY_DN9327_c0_g2	TRINITY_DN9327_c0_g2_i1|m.2176	PF01581.13	FARP	FMRFamide related peptide family	87	91	7
TRINITY_DN9327_c0_g2	TRINITY_DN9327_c0_g2_i1|m.2176	PF01581.13	FARP	FMRFamide related peptide family	135	139	7
TRINITY_DN9327_c0_g2	TRINITY_DN9327_c0_g2_i1|m.2176	PF01581.13	FARP	FMRFamide related peptide family	102	106	7
TRINITY_DN9327_c0_g2	TRINITY_DN9327_c0_g2_i1|m.2176	PF01581.13	FARP	FMRFamide related peptide family	172	176	7

Table 3: The top 10 most-related transcript provided by ORF annotation

Seq_id	*Protein_id*	*Pfam_id*	*Domain*	*DomainDescription*	*ProteinStart*	*ProteinEnd*	*Pfam Start*	*Pfam End*
TRINITY_DN703_c0_g1_i3	TRINITY_DN703_c0_g1_i3|m.1524	PF13634.3	Nucleoporin_FG	Nucleoporin FG repeat region	22	133	3	91
TRINITY_DN703_c0_g1_i11	TRINITY_DN703_c0_g1_i11|m.1539	PF13634.3	Nucleoporin_FG	Nucleoporin FG repeat region	215	326	3	91
TRINITY_DN703_c0_g1_i7	TRINITY_DN703_c0_g1_i7|m.1519	PF13634.3	Nucleoporin_FG	Nucleoporin FG repeat region	198	286	8	91
TRINITY_DN1209_c0_g3_i9	TRINITY_DN1209_c0_g3_i9|m.7591	PF03160.11	Calx-beta	Calx-beta domain	105	121	72	88
TRINITY_DN1209_c0_g3_i6	TRINITY_DN1209_c0_g3_i6|m.7593	PF03160.11	Calx-beta	Calx-beta domain	105	121	72	88
TRINITY_DN703_c0_g1_i13	TRINITY_DN703_c0_g1_i13|m.1535	PF13634.3	Nucleoporin_FG	Nucleoporin FG repeat region	203	286	13	91
TRINITY_DN1209_c0_g3_i13	TRINITY_DN1209_c0_g3_i13|m.7604	PF03160.11	Calx-beta	Calx-beta domain	105	121	72	88
TRINITY_DN1209_c0_g3_i11	TRINITY_DN1209_c0_g3_i11|m.7585	PF03160.11	Calx-beta	Calx-beta domain	105	121	72	88
TRINITY_DN1209_c0_g3_i14	TRINITY_DN1209_c0_g3_i14|m.7587	PF03160.11	Calx-beta	Calx-beta domain	105	121	72	88
TRINITY_DN86386_c0_g1_i2	TRINITY_DN86386_c0_g1_i2|m.1138	PF08238.9	Sel1	Sel1 repeat	280	291	27	38

Discussion

In recent years, next-generation sequencing has become a powerful tool to analyze the transcriptome characteristics of worms (3). In this study, we sequenced the transcriptome of fourth-stage (L4) *A. cantonensis* and compared the obtained unigenes and transcripts with related databases.

As for L4 *A. cantonensis*, the Nr annotation of unigenes and that of transcripts were nearly similar. *A. cantonensis* was the 4th top ranked species in transcripts and the 5th top ranked species in unigenes according to Nr annotation. The top 3 most related species were *A. ceylanicum*, *D. viviparus* and *H. contortus* in both unigene and transcript groups. This might be explained by 2 reasons. Firstly, the genome and transcriptome annotation of *A. cantonensis* was insufficient, and more information are required. Secondly, in comparison, *A. ceylanicum*, *D. viviparus* and *H. contortus* were sufficiently annotated. Thirdly, *A. ceylanicum*, *D. viviparus* and *H. contortus* were all parasitic worms of mammals (4-7), which implied that the 3 worms might have similar transcriptome information to that of *A. cantonensis*.

Although GO annotation exhibited different numbers of concerning unigenes and transcripts, the general tendency of GO items was similar among unigenes and transcripts. It is notable that GO:0008152 might play a significant biological function, because this item was closely related to transcripts, pertained to the metabolic process.

COG annotations of unigenes and transcripts were identical; this was also the case of KOG annotations. COG annotation indicated

that signal transduction mechanisms, replication, recombination and repair, amino acid transport and metabolism as well as energy production and conversion were highly related to unigenes and transcripts. KOG annotation showed that signal transduction mechanisms, transcription, posttranslational modification, protein turnover, and chaperones were mostly related.

The characteristics of the COG and KOG results might be explained by the fact that L4 was extremely active in the genetic transcription to enlarge its body size and combat with the immune system of the final host. As for L3 *A. cantonensis*, its body size was in the range of 0.45 by 0.03 mm (8). Meanwhile, the size for L4 was increased to 1.0 mm by 0.04 mm. In addition, L4 developed in the brain of the final host, so it also needed to avoid the immunity attack.

The top 10 highly annotated unigenes and transcripts provided by ORF annotation also showed that nucleoporin-related protein had a high level, which might also be caused by active metabolism and signal transduction of L4 *A. cantonensis*.

Conclusion

Although the complete genome sequencing of *A. cantonensis* was previously performed, the annotation of genome and transcriptome still needs further investigation. To the best of our knowledge, this work is the first to analyze the transcriptional profile of L4 *A. cantonensis*, exhibiting its gene expression characteristics.

Acknowledgements

This research was supported by the Key Laboratory of Vector Biology and Pathogen Control of Zhejiang Province, Huzhou University (Grant No. HUZUL202101) and Huzhou Natural Science Foundation (Grant No. 2019YZ06).

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. Pien FD, Pien BC. *Angiostrongylus cantonensis* eosinophilic meningitis. Int J Infect Dis, 1999;3(3):161-3.
2. Mackerras MJ, Sandars DF. The life history of the rat lung-worm, *Angiostrongylus cantonensis* (Chen) (Nematoda: Metastrongylidae). Aust J Zool, 1955;3(1): 1-21.
3. Guo Y, Chang Zhou H, Dong Y, et al. Next-Generation Sequencing Reveals That Oxidative Phosphorylation Might Be a Key Pathway Differently Expressed in the Third and Fourth Stages Larvae of *Angiostrongylus cantonensis*. Iran J Parasitol, 2020; 15(4): 587-595.
4. Guo Y, Yan Dong H, Chang Zhou H, et al. Mechanism of the Passage of *Angiostrongylus cantonensis* across the Final Host Blood-Brain Barrier Using the Next-Generation Sequencing. Iran J Parasitol, 2021; 16(3): 454-463.
5. Yoshida Y, Kondo K, Kurimoto H, et al. Comparative studies on *Ancylostoma braziliense* and *Ancylostoma ceylanicum*. III. Life history in the definitive host. J Parasitol, 1974; 60(4):636-41.
6. Traub RJ. *Ancylostoma ceylanicum*, a re-emerging but neglected parasitic zoonosis. Int J Parasitol, 2013; 43(12-13):1009-15.
7. Porter D, Cauthen G. Experiments on the life history of the cattle lungworm, *Dictyocaulus viviparus*. Am J Vet Res, 1942; 3(9): 395-400.
8. Ransom BH. The life history of the twisted wireworm (*Haemonchus contortus*) of sheep and other ruminants :(Preliminary report). 1906; US Government Printing Office.