Experimental Parasitology xxx (2017) 1–7

Contents lists available at ScienceDirect

Experimental Parasitology

journal homepage: www.elsevier.com/locate/yexpr

De novo assembly and functional annotations of the transcriptome of *Metorchis orientalis* (trematoda: Opisthorchiidae)

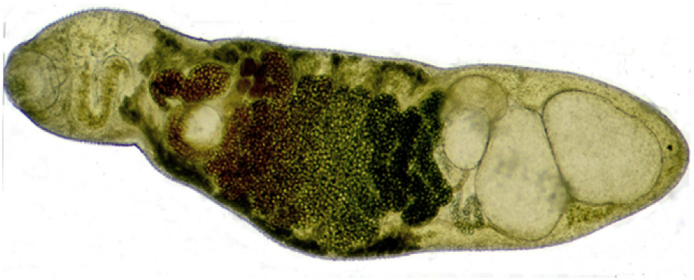
Jun-Feng Gao [a, b], Yuan Gao [a], Jian-Hua Qiu [a], Qiao-cheng Chang [a], Yan Zhang [a], Ming Fang [c, **], Chun-Ren Wang [a, c, *]

[a] College of Animal Science and Veterinary Medicine, Heilongjiang Bayi Agricultural University, Daqing, Heilongjiang Province 163319, PR China
[b] Department of Parasitology, Heilongjiang Institute of Veterinary Science, Qiqihar, Heilongjiang Province, PR China
[c] College of Life Science and Biotechnology, Heilongjiang Bayi Agricultural University, Daqing, Heilongjiang Province 163319, PR China

HIGHLIGHTS

- *De novo* sequencing, transcriptome assembly and functional annotations of the adult *M. orientalis*.
- The first report about high throughput transcriptome dataset for *M. orientalis*.
- A lot of putative proteins of *M. orientalis* are involved in pathways associated with the immune system or immune diseases.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 13 July 2017
Received in revised form
25 October 2017
Accepted 3 December 2017
Available online xxx

Keywords:
Metorchis orientalis
Transcriptome
De novo
Next-generation sequencing

ABSTRACT

Metorchis orientalis is a neglected zoonotic parasite, living in the gallbladder and bile duct of poultry and some mammals as well as humans. In spite of its economic and medical importance, the information known about the transcriptome and genome of *M. orientalis* is limited. In this study, we performed *de novo* sequencing, transcriptome assembly and functional annotations of the adult *M. orientalis*, obtained about 77.4 million high-quality clean reads, among which the length of the transcript contigs ranged from 100 to 11,249 nt with mean length of 373 nt and N50 length of 919 nt. We then assembled 31,943 unigenes, of which 20,009 (62.6%) were annotated by BLASTn and BLASTx searches against the available database. Among these unigenes, 19,795 (62.0%), 3407 (10.7%), 10,620 (33.2%) of them had significant similarity in the NR, NT and Swiss-Prot databases, respectively; 5744 (18.0%) and 4678 (14.6%) unigenes were assigned to GO and COG, respectively; and 9099 (28.5%) unigenes were identified and mapped onto 256 pathways in the KEGG Pathway database. Furthermore, we found that 98 (1.08%) unigenes were related to bile secretion and 5 (0.05%) to primary bile acid biosynthesis pathways category. The characterization of these transcriptomic data has implications for the better understanding of the biology of *M. orientalis*, and will facilitate the development of intervention agents for this and other pathogenic flukes of human and animal health significance.

Abbreviations: M. orientalis, *Metorchis orientalis*; *C. sinensis*, *Clonorchis sinensis*; *S. mansoni*, *Schistosoma mansoni*; *S. japonicum*, *Schistosoma japonicum*; *O felineus*, *Opisthorchis felineus*; *O. viverrini*, *Opisthorchis viverrini*; QC, quality control; NT, the non-redundant nucleotide database; NR, the non-redundant protein database; Swiss-Prot, Swiss protein database; COG, Cluster of Orthologous Groups; KEGG, Kyoto Encyclopedia of genes and genomes; GO, Gene Ontology; SSR, Simple sequence repeats; MISA, MIcrosatellite; BP, biological process; MF, molecular function; CC, cellular component.

* Corresponding author. College of Animal Science and Veterinary Medicine, Heilongjiang Bayi Agricultural University, Daqing, Heilongjiang Province 163319, PR China.
** Corresponding author.
E-mail addresses: fangming618@126.com (M. Fang), chunrenwang@sohu.com (C.-R. Wang).

https://doi.org/10.1016/j.exppara.2017.12.001
0014-4894/

1. Introduction

Fish-borne trematode infections affect the human health and ~750 million people are at risk of infection around Asia (Keiser and Utzinger, 2009). These flukes provoke remarkable morbidity and cause serious damage to aquaculture, which is a valuable source of food and employment in developing countries. *Metorchis orientalis* is a neglected zoonotic parasite, which has a complex life cycles, its first-intermediate was *Parafossarulus striatulus*, and its second-intermediate hosts were freshwater fishes, such as *Pseudorasbora parva*, *Abbottina rivularis*, *Cichlasoma managuense* and *Pseudogobio rivularis* (Zhan et al., 2017). *Metorchis orientalis* is mainly harbored in the gallbladder and bile duct of definitive host, includes birds (peafowl, red-crowned crane, kingfisher), poultry (chicken, duck, goose), mammal (cats, dogs, *Mus musculus*, cavies) including humans (Lin et al., 2001). It can cause gallbladder enlargement, gallbladder wall thickening and biliary tract congestion. Despite recently some reports about *M. orientalis* were published in Korea and China (Sohn, 2009; Qiu et al., 2017; Zhan et al., 2017), but it mainly focus on its biology and pathology, the genetic information about *M. orientalis* is still rare. Notably, the complete mitochondrial genome of *M. orientalis* has been sequenced, and its phylogenetic implications with closely related species has been clarified (Na et al., 2016), but it is still not sufficient to investigate the genetics, epidemiology and biology aspect of *M. orientalis*. Until now, no commercial vaccines are available, with only one reliable drug (praziquantel) being available to treat *M. orientalis* infections (http://www.waterpathogens.org/book/liver-flukes).

With the development of next-generation sequencing technologies and bioinformatic methods (Mangiola et al., 2013), researches are able to efficiently sequence the transcritomes of new organisms, which provides an effective tool for gene discovery, genome annotation, drug target identification and vaccine development (Nagaraj et al., 2007; Ranganathan et al., 2009). Although many trematode transcriptomes have been sequenced (Young et al., 2011; Jex et al., 2012; Wang et al., 2015; Huang et al., 2013; Pomaznoy et al., 2015; Liu et al., 2016a, 2016b), most trematode species have no available transcriptomic information, including *M. orientalis*; molecular studies on the metabolic regulation, signal transduction, and immune response of *M. orientalis* are still scarce.

In present study, we take advantage of next-generation sequencing and *de novo* short-read assembly technology to uncover a global view of the transcriptome of adult *M. orientalis*, in order to understand the biological functions of *M. orientalis*, which should facilitate the identification of intervention targets for *M. orientalis* and other medically and veterinary important trematodes.

2. Materials and methods

2.1. Ethics statement

Animals from which specimens were collected were handled in accordance with animal protection law of the People's Republic of China (a draft of an animal protection law in China released on September 18, 2009). This study was approved by the National Institute of Animal Health Animal Care and Use Committee at Heilongjiang Bayi Agricultural University (approval number 2016−015).

2.2. Parasite material

Adult *M. orientalis* flukes were collected by necropsy from the gallbladder and bile duct of naturally infected ducks in Daqing, Heilongjiang Province, China. Adults were washed in physiological saline for five times to avoid contamination from hosts, then the specimen were identified by species morphology according to existing keys and descriptions (Chen and Tang, 1981). Materials were immediately frozen in liquid nitrogen and stored at liquid nitrogen and −80 °C until use.

2.3. RNA isolation and Illumina sequencing

Total RNA was isolated from 30 adults of *M. orientalis* using Trizol reagent (Invitrogen, Life Technologies, Carlsbad, CA, USA) according to the manufacturer's protocol. Total RNA of independent adults of *M. orientalis* were stored at −80 °C until use. Oligo (dT) was used to isolate poly(A) mRNA from total RNA. Mixed with the fragmentation buffer, the mRNA was fragmented into short fragments. Then cDNA was synthesized using the mRNA fragments as templates. Short fragments were purified and resolved with EB buffer for end reparation and single nucleotide A (adenine) addition. After that, the short fragments were connected with adapters. Suitable fragments were selected for PCR amplification as templates. During the quality control (QC) steps, Agilent 2100 Bioanaylzer and ABI StepOnePlus Real-Time PCR System were used in quantification and qualification of the sample library. Illumina HiSeqTM 2000 was applied to sequencing at the BGI-Shenzhen, Shenzhen, China according to the manufacturer's instructions (Illumina, San Diego, CA, USA).

2.4. De novo assembly

We performed quality control on the raw reads. We removed adaptor sequences, highly redundant sequences, reads containing more than 10% 'N' residues (ambiguous bases in reads), and low quality reads containing more than 50% bases with Q-value $\leq$ 20. *De novo* assembly of clean reads was performed with Trinity (Grabherr et al., 2011). Briefly, Trinity first combines reads with a certain length of overlap to form longer fragments, called contigs; then maps the reads back to contigs. Trinity connects the contigs and gets sequences that cannot be extended on either end. The optional parameters to achieve this is "–seqType fq –min_contig_length 100 –min_glue 3 –group_pairs_distance 250 –path_reinforcement_distance 85 –min_kmer_cov 3". Such sequences are defined as unigenes. Unigenes from the assembly can be used for further processes of sequence splicing and redundancy removal with the sequence clustering software TGICL with parameters "-l 40 -c 10 -v 20" (Pertea et al., 2003).

2.5. Bioinformatics analysis

Unigene sequences were aligned with the NCBI non-redundant nucleotide (NT) database (Liu et al., 2015) by BLASTn (Zhang et al., 2017) with an e-value threshold of 0.00001; and aligned with the NCBI non-redundant protein (NR) database (Liu et al., 2015), Swiss protein (Swiss-Prot) database (Liu et al., 2015), Cluster of Orthologous Groups of proteins (COG) database (Liu et al., 2015) and Kyoto Encyclopedia of genes and genomes (KEGG) database (Liu et al., 2015) by BLASTx (Zhang et al., 2017) to assign the predicted function with an e-value threshold of 0.00001. ESTScan (Iseli et al., 1999) was used to predict protein coding sequences and their direction if unigenes were unaligned to any of the databases with default setting. Software Blast2GO (Götz et al., 2008) was employed to classify unigenes to Interpro and Gene Ontology (GO) terms including molecular function, biological processes, and cellular components (Conesa et al., 2005) for NR annotation. After predicting GO annotations for all unigenes with default setting. We

Please cite this article in press as: Gao, J.-F., et al., *De novo* assembly and functional annotations of the transcriptome of *Metorchis orientalis* (trematoda: Opisthorchiidae), Experimental Parasitology (2017), https://doi.org/10.1016/j.exppara.2017.12.001