Genomics Data 12 (2017) 109–110

Contents lists available at ScienceDirect

Genomics Data

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

Sequencing and de novo transcriptome assembly of the Chinese giant salamander (*Andrias davidianus*)

Yong Huang *, Xiao Chan Gao, Jian Li Xiong, Hong Tao Ren, Xi Hong Sun

College of Animal Science and Technology, Henan University of Science and Technology, Luoyang 471003, China

ARTICLE INFO

Article history:
Received 17 February 2017
Received in revised form 21 March 2017
Accepted 26 March 2017
Available online 27 March 2017

ABSTRACT

Next-generation technologies for determination of genomics and transcriptomics composition have a wide range of applications. *Andrias davidianus*, has become an endangered amphibian species of salamander endemic in China. However, there is a lack of the molecular information. In this study, we obtained the RNA-Seq data from a pool of *A. davidianus* tissue including spleen, liver, muscle, kidney, skin, testis, gut and heart using Illumina HiSeq 2500 platform. A total of 15,398,997,600 bp were obtained, corresponding to 102,659,984 raw reads. A total of 102,659,984 reads were filtered after removing low-quality reads and trimming the adapter sequences. The Trinity program was used to de novo assemble 132,912 unigenes with an average length of 690 bp and N50 of 1263 bp. Unigenes were annotated through number of databases. These transcriptomic data of *A. davidianus* should open the door to molecular evolution studies based on the entire transcriptome or targeted genes of interest to sequence. The raw data in this study can be available in NCBI SRA database with accession number of SRP099564.

Specifications	
Organism/cell line/tissue	*A. davidianus*/spleen, liver, muscle, kidney, skin, testis, gut and heart
Sex	Male
Sequencer or array type	Illumina HiSeq2500
Data format	Raw data: FASTQ files, analyzed data: txt files
Experimental factors	De novo transcriptome assembly of *A. davidianus*
Experimental features	Freshly and healthy collected spleen, liver, muscle, kidney, skin, testis, gut and heart were pooled for total RNA extraction, sequencing, de novo transcriptome assembly and annotation
Consent	Full consent
Sample source location	Luoyang, Henan, China

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/sra/SRR5260688

2. Introduction

The Chinese giant salamander (*A. davidianus*) belonging to one of the most primitive orders Caudata, family Cryptobranchidae, is the largest extant amphibian species in the world and found only in China [1,2]. It was called "babyfish" mainly lives in Chinese interior water area. Being crown as a living fossil from 350 million years ago, so it is considered to be a valuable model in edible value, medical value, view value, science value and so on [3]. The Chinese giant salamander is classified as critically endangered by the International Union for Conservation of Nature and Natural Resources (IUCN) Red list of Threatened Species, is a class II on the national list of protected animals in China. Nonetheless, only a limited number of transcriptomes from *A. davidianus* have been sequenced so far [4–8]. Meanwhile, the genetic information and gene sequences about the Chinese giant salamander in public databases are also scanty. Here we performed de novo transcriptome assembly for *A. davidianus* by next-generation sequencing. The obtained transcriptomic data will be useful for further studies of the genomes evolution of *A. davidianus* may have represented transition steps from aquatic to terrestrial life.

3. Experimental design, materials and methods

3.1. Animal materials

The healthy *A. davidianus* were obtained from a farm located in Luan chuan (Henan Province, China). Organ tissues including spleen, liver, muscle, kidney, skin, testis, gut and heart collected from three-year-old (3Y) male *A. davidianus* immediately after dissection, washed in sterile PBS. Animals tissues sample for RNA extraction were snap-

* Corresponding author.
E-mail address: huangyong1979111@163.com (Y. Huang).

http://dx.doi.org/10.1016/j.gdata.2017.03.011

frozen in liquid nitrogen and stored at −80 °C for further analysis. All experiments were performed in strict accordance with the Animal Protection Law of China, and were approved by the Review Committee for the Use of Animal Subjects of Henan University of Science and Technology.

3.2. Library construction and sequencing

The *A. davidianus* were pooled from 8 tissues (spleen, liver, muscle, kidney, skin, testis, gut and heart) and used for total RNAs library construction. RNA-Seq libraries were generated using the TruSeq RNA-Seq Sample Prep kit according to the manufacturer's protocol (Illumina Inc., San Diego, CA). Poly-A RNAs were isolated from total RNA and chemically fragmented. First and second strand cDNA synthesis were followed by end-repair and adenosines were added to the 3′ends. Adapters were ligated to the cDNA and 200 ± 25 bp fragments were gel purified and enriched by PCR. The libraries were quantified using Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA) and run on the Illumina HiSeq2500 platform (Illumina Inc.). Paired-read sequences, 125 nt in length, were collected.

3.3. De novo transcriptome assembly and functional annotation

A total of 15,398,997,600 bp were obtained from sequencing libraries. After cleaning and quality checks, the de novo assembly of all sequencing data using the Trinity program. It generated 158,103 all-transcripts with an average length of 810 bp and an N50 of 1659 bp; and 132,912 all unigenes were achieved with an average length of 690 bp and an N50 of 1263 bp (Table 1). Of these, 34,075 were >1 kb and 21,855 >1 kb, respectively. We annotated the all 71,443 unigenes to NR, NT, CDD, Swissprot, TrEMBL, Pfam, KEGG, KOG and GO databases. A microsatellite program MISA was also used to identify and localize microsatellite motifs.

Conflict of interest

The authors declare that they have no competing interests.

Table 1
Summary of *A. davidianus* assembly statistics.

Index	*A. davidianus* transcript	*A. davidianus* unigenes
All number	158,103	132,912
Length ≧ 500 bp	59,715	42,327
Length ≧ 1000 bp	34,075	21,855
N50 length	1659 bp	1263 bp
N90 length	285 bp	262 bp
Max length	16,067 bp	16,067 bp
Minor length	201 bp	201 bp
All length	128,175,999 bp	91,713,308 bp
Mean length	810 bp	690 bp

Acknowledgments

This work was funded by Natural Science Foundation of Henan Province of China (162300410069) and Educational Commission of Henan Province of China (16A240002).

References

[1] R.W. Murphy, J. Fu, D.E. Upton, T. de Lema, E.M. Zhao, Genetic variability among endangered Chinese giant salamanders, *Andrias davidianus*. Mol. Ecol. 9 (2000) 1539–1547.
[2] P. Zhang, Y.Q. Chen, Y.F. Liu, H. Zhou, L.H. Qu, The complete mitochondrial genome of the Chinese giant salamander, *Andrias davidianus* (Amphibia: Caudata). Gene 311 (2003) 93–98.
[3] K.Q. Gao, N.H. Shubin, Earliest known crown-group salamanders. Nature 422 (2003) 424–428.
[4] F. Li, L. Wang, Q. Lan, H. Yang, Y. Li, X. Liu, Z. Yang, RNA-Seq analysis and gene discovery of *Andrias davidianus* using Illumina short read sequencing. PLoS One 10 (2015), e0123730.
[5] Y. Fan, M.X. Chang, J. Ma, S.E. LaPatra, Y.W. Hu, L. Huang, P. Nie, L. Zeng, Transcriptomic analysis of the host response to an iridovirus infection in Chinese giant salamander, *Andrias davidianus*. Vet. Res. 46 (2015) 136.
[6] X. Jiang, Y. Wang, X. Zhang, Data set for transcriptome analysis of the Chinese giant salamander (*Andrias davidianus*). Data Brief 6 (2016) 12–14.
[7] Z. Qi, Q. Zhang, Z. Wang, T. Ma, J. Zhou, J.W. Holland, Q. Gao, Transcriptome analysis of the endangered Chinese giant salamander (*Andrias davidianus*): immune modulation in response to *Aeromonas hydrophila* infection. Vet. Immunol. Immunopathol. 169 (2016) 85–95.
[8] R. Che, Y. Sun, R. Wang, T. Xu, Transcriptomic analysis of endangered Chinese salamander: identification of immune, sex and reproduction-related genes and genetic markers. PLoS One 9 (2014), e87940.