Genomics Data 13 (2017) 44–45

Contents lists available at ScienceDirect

Genomics Data

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

Data in Brief

De novo transcriptome assembly for the lung of the ornamented pygmy frog (*Microhyla fissipes*)

Lusha Liu[a], Shouhong Wang[a,b], Lanying Zhao[c], Jianping Jiang[a,*]

[a] Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu 610041, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China
[c] Sichuan University, Chengdu 610041, China

ARTICLE INFO

Keywords:
Lung
Transcriptome
Microhyla fissipes
Unigenes

ABSTRACT

Microhyla fissipes, belonging to Neobatrachia, is a new model organism to study developmental biology, adaptive mechanisms from aquatic to terrestrial life, environmental toxicology, and human disease. *M. fissipes* use of lungs soon after hatching makes it extremely valuable for the study of lung function and development mechanisms. However, our knowledge of genes and pathways associated with lung development in *M. fissipes* is very limited. In this study, we conducted *de novo* transcriptome assembly for the lung of *M. fissipes* using the Illumina HiSeq4000 platform. We obtained approximately 9.0 GB clean data from the lung of the stage 28 tadpole with lung inflation (NCBI accession numbers: SRP107055). *De novo* transcriptome assembly identified 209,358 transcripts and 93,813 unigenes. In addition, BLASTX against NR, NT, KO, SwissProt, PFAM, GO and KOG databases were used to annotate all the 93,813 unigenes. This study provides the transcriptome and functional annotation of genes in *M. fissipes* lung development, which will be useful for comparative transcriptome analyses and promote research into mechanism of lung development in anuran.

Specifications	
Organism/cell line/tissue	*Microhyla fissipes*/lung
Sex	N.A.
Sequencer or array type	HiSeq4000
Data format	Raw and processed
Experimental factors	Transcriptome profiling of lung
Experimental features	Lung of the S28 tadpole was harvested for total RNA extraction. Prepared cDNA libraries were paired-end sequenced by HiSeq4000 system. The obtained data was subjected for *de novo* transcriptome assembly using Trinity, and coding regions were predicted by BLAST. We performed BLASTX against the NR, NT, KO, SwissProt, PFAM, GO and KOG database using BLAST with an e-value cut off of 1^{e-5} to annotate identified proteins.
Consent	N/A
Sample source location	Chengdu, Sichuan, China (30.5825° N, 103.8438° E)

1. Direct link to deposited data

https://www.ncbi.nlm.nih.gov/sra/?term=SRP107055.

2. Introduction

Microhyla fissipes is a typical tailless anuran from the family Microhylidae suborder Neobatrachia. Several characteristics including fast development, strong survivability, biphasic life cycle, small body size, diploid, and transparent tadpoles suggested that *M. fissipes* would be a good organism to study developmental biology, adaptive mechanisms from aquatic to terrestrial life, environmental toxicology, and human disease [1]. Amphibians use many different organs for gas exchange throughout their lives including lungs, internal and external gills, the internal skin of the mouth and pharynx, and the external skin [2]. Frogs delay lung inflation (meaning the initial act of inflating their lung rudiments with air) to different stages after hatching [3]. As a model metamorphosed from aquatic to terrestrial life, *M. fissipes* starts using lungs soon after hatching at stage 28 (S28). *M. fissipes* is therefore a good model to study the function and development mechanism of

* Corresponding author.
E-mail address: jiangjp@cib.ac.cn (J. Jiang).

http://dx.doi.org/10.1016/j.gdata.2017.07.002
Received 30 May 2017; Received in revised form 26 June 2017; Accepted 4 July 2017
Available online 04 July 2017

Table 1
Summary of *Microhyla fissipes* assembly statistics.

Index	Transcripts	Unigenes
All number	209,358	93,813
Length (500–1000 bp)	33,434	31,265
Length (≥ 1000 bp)	22,897	22,868
N50 length	663	1063
N90 length	239	387
Max length	7268	7268
Minor length	201	201
All length	107,212,015	75,405,112
Mean length	512	804

lung in anuran, and the evolution of lung function in the vertebrate. In this study, we performed *de novo* transcriptome assembly for the lung of *M. fissipes* by RNA-Seq.

3. Experimental design, materials, and methods

3.1. Sample collection

Mature female and male *M. fissipes* were collected from Shuangliu, Chengdu, China (30.5825° N, 103.8438° E) in June 2016. After acclimatized in our laboratory for 1 week, the male and female were injected LHRHa with 0.3 μg/g body weight resolving dosage. Tadpoles were subjected to a 12:12 h light:dark cycle at 25 ± 0.6 °C. Tadpole developmental stage was recorded using the *M. fissipes* developmental staging table. The lung tissue was dissected from one prometamorphosis (S28) tadpole, immediately frozen in liquid nitrogen and stored at − 80 °C for further analysis. All experiments were performed according to the Guideline for the Care and Use of Laboratory Animals in China, and proved by the Experimental Animal Use Ethics Committee of the Chengdu Institute of Biology.

3.2. RNA isolation, library preparation, and sequencing

Total RNAs were extracted by using the TRIZOL Kit (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions and DNAase I was used to remove DNA. Integrity and purity were checked with Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA) and Nanodrop (Thermo Fisher Scientific Inc., USA) with RIN number > 8.0. cDNA libraries were generated using the TruSeq RNA Sample Preparation Kit (Illumina Inc., San Diego, CA) following the standard protocols. In brief, the polyA mRNAs were isolated using polyT oligo-attached magnetic beads 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 PCR was conducted to selectively enrich DNA fragments with adapters. After quality control, the library was sequenced on the Illumina HiSeq4000 platform (Illumina Inc. San Diego, CA) with 150 nt paired-end reads in length.

3.3. De novo transcriptome assembly, identification protein coding regions, and annotation

A total of 62,036,628 raw paired-end reads were generated. After removal of the reads with adapters, low quality reads, reads with unknown nucleotides, 59,987,242 clean reads (96.7%) were obtained. Cleaned reads were *de novo* assembled into transcripts by Trinity software [4]. 209,358 transcripts were generated ranging from 201 to 7268 bp with an average length of 512 bp and an N50 of 663 bp. These represented 93,813 unique unigenes with an average length of 804 bp and an N50 of 1063 bp (Table 1). The 93,813 unigenes were annotated to NR, NT, KO, SwissProt, PFAM, GO and KOG databases with 33.8%, 18.6%, 17.9%, 29.2%, 27.8%, 28.8% and 12.1% annotation rate, respectively (Supplement Table). Furthermore, 6.6% of unigenes were annotated in all databases, while 42.2% were annotated in at least one database. In summary, the transcriptome sequences of the lung of *M. fissipes* provide the annotation genes and pathways associated with lung development and facilitate a better understanding of the molecular mechanisms of lung development. This study will lay a solid foundation for functional genomics studies on *M. fissipes* in the future.

Conflict of interest

The authors declare that they have no competing interests.

Acknowledgment

This study was supported by Important Research Project of Chinese Academy of Sciences (KJZG-EW-L13) and 2015 Western Light Talent Culture Project of the Chinese Academy of Sciences (Y6C3021). Animal procedures were approved by the Animal Care and Use Committee of Chengdu Institute of Biology.

Appendix A. Functional annotation of unigenes

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2017.07.002.

References

[1] L.S. Liu, L.Y. Zhao, S.H. Wang, J.P. Jiang, Research proceedings on amphibian model organisms, Zool. Res. 37 (2016) 237–245.
[2] W.E. Duellman, L. Trueb, Biology of Amphibians, McGraw-Hill Book Company, New York, NY, 1986.
[3] S.R. Christopher, J. Brandon, Plasticity of lung development in the amphibian, *Xenopus laevis*, Biol. Open. 2 (2012) 1324–1335.
[4] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q.D. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F.D. Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome, Nat. Biotechnol. 29 (2011) 644–652.