Data in Brief 15 (2017) 573–576

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

Data of first *de-novo* transcriptome assembly of a non-model species, hawksbill sea turtle, *Eretmochelys imbricate*, nesting of the Colombian Caribean

Javier Hernández-Fernández

Genetics, Molecular Biology and Bioinformatics Lab, Department of Natural and Environmental Sciences, Universidad Jorge Tadeo Lozano, Cra. 4 N° 22-61 modulo 7 Piso 6, Bogotá D.C., Colombia

ARTICLE INFO

Article history:
Received 4 August 2017
Received in revised form
26 September 2017
Accepted 4 October 2017
Available online 11 October 2017

Keywords:
Hawksbill turtle
Trinity
RNAseq
illumina
N50

ABSTRACT

The hawksbill sea turtle, *Eretmochelys imbricata,* is an endangered species of the Caribbean Colombian coast due to anthropic and natural factors that have decreased their population levels. Little is known about the genes that are involved in their immune system, sex determination, aging and others important functions. The data generated represents RNA sequencing and the first de-novo assembly of transcripts expressed in the blood of the hawksbill sea turtle. The raw FASTQ files were deposited in the NCBI SRA database with accession number SRX2653641. A total of 5.7 Gb raw sequence data were obtained, corresponding to 47,555,108 raw reads. Trinity was used to perform a first de-novo assembly, and we were able to identify 47,586 transcripts of the female hawksbill turtle transcriptome with an N50 of 1100 bp. The obtained transcriptome data will be useful for further studies of the physiology, biochemistry and evolution in this species.

Specifications Table

Subject area	Genetics and Transcriptomics
More specific subject area	Transcriptomics of sea turtles *Eretmochelys imbricata*

http://dx.doi.org/10.1016/j.dib.2017.10.015

Type of data	Raw reads of DNA sequences
How data was acquired	A blood sample of a living specimen of the sea turtle *Eretmochelys imbricata* was collected for total RNA isolations. Prepared a paired-end library, sequenced by the Hiseq. 2000 system. The obtained data was subjected to de novo assembly.
Data format	Raw data FASTQ file
Experimental factors	The hawksbill turtle is in captivity in swimming pools with sea water. The concentration of oxygen is 5 mg/L, temperature of 30 °C and salinity of 30.5
Experimental features	The de novo assembling of the transcriptome and the functional identification of the genes expressed by hawsbill turtle was performed.
Data source location	Islas del Rosario, Bolívar, Colombia 10°17’67.95” N – 75°77’16.27” W
Data accessibility	The raw FASTQ files were deposited in the NCBI SRA database with accession number SRX2653641 (https://www.ncbi.nlm.nih.gov/sra/SRX2653641/)

Value of the data

- This is the first de novo transcriptome of *E. imbricate* sea turtle published
- The obtained transcriptome data will be useful for further studies of the physiology, biochemistry evolution and others of *E. imbricate* sea turtle.
- It is possible to know and analyze the metabolic pathways in which the genes identified are involved.

1. Data

The hawksbill turtle, *Eretmochelys imbricata* [1], is a non-model species that is found throughout the tropics in the central Atlantic and Indo-Pacific regions [2–5]. Marcovaldi et al. [6] reported that this species nests from the state of Florida (USA) in the wider Caribbean, to the south coast of Espirito Santo in Brazil. Trujillo-Arias et al. [7] located foraging areas for the hawksbill turtle in Colombia on the Islas del Rosario and in the National Natural Park of Cabo de la Vela.

This turtle is listed by the Union for Conservation of Nature as critically endangered A2bd [8] and in Appendix I of CITES [9]. The main causes of the population decline include meat and shells marketing, egg consumption and oil production [10,11], bycatch in industrial and artisanal fisheries and habitat loss [12,13]. The demand for shells continues today on the black market [3] with the dramatically decreasing their population [4].

The data of this article are represented by the raw FASTQ files deposited in the NCBI SRA database with accession number SRX2653641 (https://www.ncbi.nlm.nih.gov/sra/SRX2653641/).

The transcriptome sequencing and read processing are summarized in Table 1. We obtained a total of 47,586 assembled transcripts with a N50 = 1100 bp, average length of 724 bp.

Table 1
Statistics of *Eretmochelys imbricata* transcriptome assembly.

No of raw data bases	4,803,065,908
No of reads of pairs	47,555,108
No of assembled transcrips	47,586
Assembly GC percent	46,2
Contig N50	1100
Contig minimum	147
Contig maximum	8,666
Averag contig	724