



ELSEVIER

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 imbricata*, nesting of the Colombian Caribbean



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.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Specifications Table

Subject area	Genetics and Transcriptomics
More specific subject area	Transcriptomics of sea turtles <i>Eretmochelys imbricata</i>

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

2352-3409/© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Type of data	Raw reads of DNA sequences
How data was acquired	A blood sample of a living specimen of the sea turtle <i>Eretmochelys imbricata</i> 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 hawksbill 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. imbricata* sea turtle published
- The obtained transcriptome data will be useful for further studies of the physiology, biochemistry evolution and others of *E. imbricata* 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 transcripts	47,586
Assembly GC percent	46,2
Contig N50	1100
Contig minimum	147
Contig maximum	8,666
Averag contig	724

2. Experimental design, materials and methods

2.1. Animal materials

A blood sample from a *Eretmochelys imbricata* individual was obtained from the CEINER Oceanarium in San Martín de Pajares Island, Cartagena. The blood was obtained from the dorsal cervical breasts in accordance with Dutton [14] methodology. The sample was placed in sterilized tubes with Tris-EDTA buffer 0.1 M (GreinerBio-one®, Kremsmünster, Austria) solution and was transported at 4 °C to the Molecular Biology laboratory of the Universidad Jorge Tadeo Lozano, Bogotá campus. The sample was collected following the ethical standards established by the legislation and the study obtained permission from the Ministry of the Environment for the development of the Biodiversity research (No 24 of June 22, 2012) and the Genetic Resources Access contract (No 64 of April 2013).

A blood sample of a *Eretmochelys imbricata* sea turtle was used for total RNA extraction using the RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. RNA integrity was confirmed using a 2100 Bioanalyzer (Agilent Technologies). For mRNA library preparation, we used a TruSeq RNA Library Prep Kit v2 according to manufacturer's instructions (Illumina, San Diego, U.S.A.). The quality control of generated libraries was done using the 2100 bio-analyzer (Agilent, Santa Clara, U.S.A.). The library was paired-end sequenced by Macrogen Co. (Seoul, South Korea) using the HiSeq. 2000 Platform. The quality of cleaned raw reads was verified with the FastQC program (<http://www.bioinformatics.bbrc.ac.uk/projects/fastqc/>). FastQC delivered quality metrics that were used to identify if the data required initial pre-processing before the transcriptome assembly.

2.2. De novo transcriptome assembly

The quality of sequencing reads was performed by means of FastQC [2]. Read trimming on quality (Q50) and sequencing adaptors removal was run with Trimmomatic [16,17].

De novo transcriptome assembly was performed using Trinity [18], using default parameters for the assembly of paired end reads. Mapping and abundance estimation was performed by means of Bowtie [19] using the constructed transcriptome as a reference.

Acknowledgments

We are grateful to CEINER Oceanarium on the island of St. Martín Pajares for their collaboration in obtaining and providing sample of *Eretmochelys imbricata* for the development of this study. Samples were obtained under a research permit that was granted by the Ministry of Environment and Territorial Development (#24 of June 22, 2012) and Contract for Access to Genetic Resources (#64 of April 23, 2013).

Funding sources

This work was supported by the Office of Research, Creation and Extension of the Universidad Jorge Tadeo Lozano.

Transparency document. Supporting information

Transparency data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2017.10.015>.

References

- [1] Linnaeus, *Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis, reformata*, Editio decima, Laurentii Salvii, Holmiae [Stockholm], 1766, pp. 1758–1759.
- [2] R. Márquez, M. del Carmen-Farías, *Las Tortugas Marinas y Nuestro Tiempo*, Fondo de Cultura Económica, Carretera Picacho-Ajusco 227; 14200 México, D.F. ISBN 968-16-4436-0 Impreso en México, 2000.
- [3] A.L. Bass, Genetic analysis to elucidate the natural history and behavior of Hawksbill Turtles (*Eretmochelys imbricata*) in the wider Caribbean: a review and re-analysis, *Chelonian Conserv. Biol.* 3 (2) (1999) 195–199.
- [4] A.B. Meylan, M. Donnelly, Status justification for listing the hawksbill turtle (*Eretmochelys imbricata*) as critically endangered on the 1996 IUCN Red List of Threatened Animals, *Chelonian Conserv. Biol.* 3 (2) (1999) 200–224.
- [5] H. Zolgharnein, M.A. Salari-Aliabadi, A.M. Foroughmand, S. Roshani, Genetic population structure of Hawksbill turtle (*Eretmochelys imbricata*) using microsatellite analysis, *Iran. J. Biotechnol.* 9 (1) (2011) 56–62.
- [6] M.A. Marcovaldi, G.G. Lopez, L.S. Soares, A.J.B. Santos, C. Bellini, P.C.R. Barata, Fifteen years of hawksbill sea turtle (*Eretmochelys imbricata*) nesting in Northern Brazil, *Chelonian Conserv. Biol.* 6 (2) (2007) 223–228.
- [7] N. Trujillo-Arias, D.F. Amorocho, D. López-Álvarez, L.M. Mejía-Ladino, Relaciones filogeográficas de algunas colonias de alimentación y anidación de la tortuga carey (*Eretmochelys imbricata*) en el Pacífico y Caribe Colombianos, *Bol. Invest. Mar. Cost.* 43 (2014) 159–182.
- [8] IUCN, IUCN Red List of Threatened Species. Versión 2016.4. (www.iucnredlist.org). (Consulted: 02-II- 2017).
- [9] CITES. Available: (https://cites.org/eng/dec/valid16/16_127.php). Consulted: 1 April 2017.
- [10] J.V. Rueda, Programa Nacional Para La Conservación de Las Tortugas Marinas y Continentales de Colombia (Propuesta para concertar Plan Estratégico Para la Recuperación de Las Especies Colombianas de Fauna Silvestre Amenazadas y en vías de Extinción), Ministerio del Medio Ambiente-Corpoguajira. Dirección General de Ecosistemas, Bogotá, Colombia, 2001.
- [11] SWOT. The state of the world's turtles. The world's most (and least) threatened sea turtles. SWOT Report VII, vol. 7, SeaturtleStatus.org, Arlington, 2012, 48.
- [12] K.L. Eckert, K.A. Bjorndal, F.A. Abreu-Grobois, M. Donnelly, Técnicas de Investigación y Manejo para la Conservación de las Tortugas Marinas. Grupo especialista en Tortugas Marinas. Unión Internacional para la Conservación de la Naturaleza y Comisión de Supervivencia de Especies, Publicación, 2000.
- [13] L. Barreto, Diagnóstico del estado actual de las tortugas en el Pacífico colombiano. Informe de País. Fundación Conservación Ambiente Colombia. 2011. Available (<http://www.natura.org.co/estacion-septiembre/ver-categoria.html>). Consulted: 10 June 2015.
- [14] P. Dutton Methods for collection and preservation of sample for sea turtle genetic studies, p. 17–24. En: B. W. Bowen y W. N. Witzell (Editores), Proceedings of the International Symposium on Sea Turtles Conservation Genetics, NOAA Technical memorandum. 1996 NMFS-SEFSC-396.173.
- [15] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics* 30 (2014) 2114–2120. <http://dx.doi.org/10.1093/bioinformatics/btu170>.
- [16] J. Hernández-Fernández, A. Pinzón, L. Mariño-Ramírez, De novo transcriptome assembly of loggerhead sea turtle nesting of the Colombian Caribbean, *Genom. Data* 13 (2017) 18–20. <http://dx.doi.org/10.1016/j.gdata.2017.06.005>.
- [17] D. Chacón, La tortuga carey del Caribe: Introducción a su biología y estado de conservación. Available at [www.hawksbillwwf.org]. WWF-Programa Regional Para América Latina y El Caribe, San José, Costa Rica, 3, 8, 2004, Consulted: July 20, 2013.
- [18] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, et al., Fulllength transcriptome assembly from RNA-Seq data without a reference genome, *Nat. Biotechnol.* 29 (2011) 644–652. <http://dx.doi.org/10.1038/nbt.1883>.