Genomics Data 13 (2017) 18–20

Contents lists available at ScienceDirect

Genomics Data

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

Data in Brief

De novo transcriptome assembly of loggerhead sea turtle nesting of the Colombian Caribbean

Javier Hernández-Fernández[a,b,*], Andrés Pinzón[c], Leonardo Mariño-Ramírez[d]

[a] 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
[b] Facultad de Ciencias, Pontificia Universidad Javeriana, Bogotá, Colombia
[c] Bioinformatics and Systems Biology Group, National University of Colombia, Bogotá, Colombia
[d] NCBI, NLM, NIH Computational Biology Branch, Building 38A, Room 6S614M 8600 Rockville Pike, MSC 6075 Bethesda, MD 20894-6075, United States

ARTICLE INFO

Keywords:
Caretta caretta turtle
Trinity
RNAseq
Illumina
N50

ABSTRACT

Loggerhead sea turtle *Caretta caretta* is widely distributed in the oceans of tropical and subtropical latitude. This turtle is an endangered species due to anthropic and natural factors that have decreased their population levels. In this study, RNA sequencing and de-novo assembly of genes expressed in blood were performed. The raw FASTQ files have been deposited on NCBI's SRA database with accession number SRX2629512. A total of 5.4 Gb raw sequence data were obtained, corresponding to 48,257,019 raw reads. Trinity pipeline was used to perform a de-novo assembly, we were able to identify 64,930 transcripts for female loggerhead turtle transcriptome with an N50 of 1131 bp. The obtained transcriptome data will be useful for further studies of the physiology, biochemistry and evolution in this species.

Specifications	
Organism/cell line/tissue	*Caretta caretta*/blood tissue
Sex	Female
Sequencer or array type	Illumina Hiseq2000
Data format	Raw data FASTQ file
Experimental factors	De novo transcriptome assembly of *Caretta caretta*
Experimental features	A blood sample of a living specimen of the sea turtle *Caretta caretta* 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.
Consent	N/A
Sample source location	Islas del Rosario, Bolivar, Colombia

1. Direct link to deposited data

https://www.ncbi.nlm.nih.gov/sra/SRX2629512

2. Introduction

The loggerhead turtle *Caretta caretta* [19], is distributed around the oceans of the world in tropical and subtropical latitudes [1]. The main nesting beaches are found in the Florida Peninsula in North America [2], Brazil, Japan and Greece and on the Mediterranean Sea, Oman in the Arabian Sea, and on the Madagascar Island [3–5]. It is an important member of complex ecological marine and coastal systems [6]. It contributes to the resilience of marine environments maintaining the balance in ecosystems and food chains they occupy, through the control of mollusks, crustaceans and other invertebrate marine populations [7]. Its main contribution is to transfer large amounts of biomass to abyssal zones [8]. In the Colombian Caribbean, the presence of this species has been recognized, and nesting beaches in the departments of Magdalena, Guajira, Bolivar and San Andres have been identified [9]. *Caretta caretta* is protected by national laws and international agreements that hopefully will mitigate the decrease of nesting females caused by anthropic causes [10,11]. However, Colombia has the second highest level of *Caretta caretta* capturing per year in the world, with an approximate

* Corresponding author at: 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.
E-mail address: javier.hernandez@utadeo.edu.co (J. Hernández-Fernández).

http://dx.doi.org/10.1016/j.gdata.2017.06.005
Received 3 April 2017; Received in revised form 31 May 2017; Accepted 7 June 2017
Available online 08 June 2017

number/year of 600 turtles [12]. In 1986, the International Union for Conservation of Nature [IUCN] reported the *Caretta caretta* turtle in its red list and denoted as vulnerable; in 1996, it entered to the endangered (EN) category; currently it can be found in the category of vulnerable A2b (11). The population decline of the *Caretta caretta* turtle is attributed to factors such as contamination, modification of the habitat, introduction of predators, by-catch and excessive clandestine fishing [13]. In addition, pathologies such as pneumonia, hepatitis, meningitis, eye conditions and other conditions associated with exposure to hydrocarbons and organochlorine pesticides have been described, which cause the death of loggerhead turtles [7]. Other pathologies such as malformations, depigmentation and embryonic mortality have not yet been studied, but could be related to genetic factors [14].

The metabolic adaptations and physiological mechanisms underlying their ability to move long distances, prolonged anoxia times and proper water maintenance have been the aim of intense interest for many years [15]. Loggerhead turtles generally exhibit contrary behavior to divers who breathe air, they rest in the background, regularly for long periods during the demersal stages of their life cycle. They have high anaerobic capacity and tolerance to anoxia. Given that lungs collapse during deep dives, their storage capacity is expendable, species increase their blood volume and concentrations of respiratory pigments [15]. Additionally, the tolerance to anoxia is developed as they increase in size, which defines their behavior in the water column.

In this study, we performed de novo transcriptome assembly for a blood sample from *Caretta caretta* sea turtle from young female individual by next-generation sequencing. This transcriptome data will be useful for further studies of the physiology, biochemistry and evolution.

3. Experimental design, materials and methods

3.1. Animal materials

Blood tissue sample from a *Caretta caretta* individual was obtained from the CEINER Oceanarium in San Martin de Pajares Island, Cartagena. The blood was obtained from the dorsal cervical breasts in accordance with [20] 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 lab of the Universidad Jorge Tadeo Lozano, Bogota 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 *Caretta caretta* sea turtle was used for total RNA extraction using RNeasy Mini Kit (Quiagen, Hilden, Germany). For mRNA library preparation, we use a TruSeq RNA Library Prep Kit v2 according to manufacturer's instructions (Illumina, San Diego, U.S.A.). The poly-A containing mRNAs were isolated using poly-T oligo-attached magnetic beads. The first strand cDNA followed by a second strand cDNA was synthetized from purified mRNAs. End repair was performed followed by adenylation of 3 ends. Adapters were ligated and PCR was done to selectively enrich DNA fragments with adapters and to amplify the amount of DNA in the library, respectively. The quality control of generated libraries was done using the 2100 bioanalyzer (Agilent, Santa Clara, U.S.A.). RIN values (RNA integrity number) of 7.5 were obtained. The library was paired-end sequenced by Macrogen Co. (Seoul, South Korea) using Hiseq 2000 Platform. The quality of cleaned raw reads was verified with the fastQC program (http://www.bioinformatics.bbrc.ac.uk/projects/fastqc/).

3.2. De novo transcriptome assembly

The quality of sequencing reads was performed by means of FastQC [16]. Read trimming on quality (Q50) and sequencing adaptors removal was run with Trimmomatic [17]. We obtained a total of 5,4 Gb of raw data corresponding to 4,873,958,919 bp.

Table 1
Statistics of *Caretta caretta* transcriptome assembly.

No of raw data bases	4,873,958,919
No of reads of pairs	48,257,019
No of assembled transcripts	64,930
Assembly GC percent	45.8
Contig N50	1731
Contig minimum	165
Contig maximum	11,046
Average contig	731

De novo transcriptome assembly was performed using Trinity [18], this program, was executed using default parameters for the assembly of paired end reads. Mapping and abundance estimation was performed by means of Bowtie [21] using the constructed transcriptome as a reference. Transcriptome sequencing and read processing are summarized in Table 1. We obtained a total of 64,930 assembled transcripts with a N50 = 1731 bp, average length of 731 bp.

In conclusion, hereby we present the first sequencing effort and the novo assembly of the transcriptome of the loggerhead sea turtle *Caretta caretta*. This transcriptome data will be useful for further studies of the evolution, phylogenomics, physiology, biochemistry and gene regulation of anoxia tolerance.

Conflict of interest

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by the Office of Research, Creation and Extension of the Universidad Jorge Tadeo Lozano (693-12-15). We are grateful to CEINER Oceanarium on the island of St. Martin Pajares for their collaboration in obtaining and providing sample of *Caretta caretta* 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).

References

[1] P.C.H. Pritchard, J.A. Mortimer, Taxonomy external morphoology, and species identification, p-21-38, in: K.L. Eckert, A. Bjornald Abreu-Grobois, M. Donelly (Eds.), Research and Management Techniques for the Conservation of Sea Turtles, IUCN/SSC. Marine Turtle Specialist Group Public, Washington. D.C, 1999(No.4.).
[2] FWC, A Statistical Analysis of Trends in Florida's Loggerhead Nest Counts with Data Through 2012, Florida Fish & Wildlife Conservation Commission, 2015 (http://myfwc.com/research/wildlife/sea-turtles/nesting/loggerhead-trends [consulted, February 2017]).
[3] C.K. Dodd, Synopsis of the biological data on the loggerhead sea turtle *Caretta caretta* (Linnaeus 1758), Fish and Wildlife Service Biological Report, vol. 88(14), 1988, pp. 1–110.
[4] European Environment Agency, *Caretta caretta* - Dssessemts of conservation status at the European level, Habitat Directive Article 17 Reporting. Copenhagen: European Environment Agency, 2009, p. 2 (13-7-2009).
[5] D. Lancheros-Piliego, J. Hernández Fernández, AMDAR and PCR-extra-fast for molecular identification of the loggerhead sea turtle *Caretta caretta* (Testudines: Cheloniidae) using the mitochondrial gene cytochrome c oxidase I (COI), Univ. Sci. 18 (3) (2013) 321–330.
[6] 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, (4), (2000).
[7] A. Machado, J.A. Bermejo, Estado de conservación de la tortuga Boba (*Caretta caretta*) en las Islas Canarias. Plan de Seguimiento de la tortuga Boba en Canarias, Observatorio Ambiental Granadilla, Santa Cruz de Tenerife, 2012 (154 pp.).
[8] P.C.H. Pritchard, Estado global de las tortugas marinas. Un análisis. Margarita Island: Venezuela. Convención Interamericana para la Protección y Conservación de las Tortugas Marinas, Primera Conferencia de las Partes, (2004) (COP1CIT).
[9] C. Ceballos, Distribution of nesting beaches and sea turtle feeding areas and their threats in the Colombian Caribbean, Bol. Invest. Mar. Cost. 33 (2004) 79–99

(Spanish).
[10] 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, p. 48.
[11] UICN, IUCN Red List of Threatened Species. Versión 2016.4. www.iucnredlist.org, (2017) (Consulted: 02-II- 2017).
[12] F. Humber, B.J. Godley, A.C. Broderick, So excellent a fishe: a global overview of legal marine turtle fisheries, Divers. Distrib. 20 (2014) 579–590, http://dx.doi.org/10.1111/ddi.12183.
[13] J. Frazier, S. Salas, The status of marine turtles in the Egyptian Red Sea, Biol. Conserv. 30 (1) (1984) 41–67.
[14] J. Azanza, M. Ruisanchez, A. Ibarra, L.C. Ruiz-Urquiola, Ríos-Tamayo, Indicadores del éxito reproductivo de la tortuga verde (*Chelonia mydas*) en tres playas de la península de Guanahacabibes, Pinar del río, Cuba, Rev. Investig. Mar. 27 (1) (2006) 69–78 (69).
[15] A. Krivoruchko, K.B. Storey, Forever young: mechanisms of natural anoxia tolerance and potential links to longevity, Oxidative Med. Cell. Longev. 3 (3) (2010) 186–198.
[16] R. Hobza, A. Widmer, Efficient molecular sexing in dioecious *Silene latifolia* and *S. dioica* and paternity analysis in F1 hybrids, Mol. Ecol. Resour. 8 (2008) 1274–1276, http://dx.doi.org/10.1111/j.1755-0998.2008.02344.x.
[17] 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.
[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.
[19] C. Linnaeus, Systema naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis, reformata, Editio decima, Laurentii Salvii, Holmiae [Stockholm], 1758-1759.
[20] P.H. Dutton, Methods for collection and preservation of samples for sea turtle genetic studies, Proceedings of the international symposium on sea turtle conservation genetics, NOAA Tech. Memo, Miami, Florida, 1996, pp. 17–24 NMFS-SEFSC-396.
[21] B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome Biol. 10 (3) (2009) R25.