Genomics Data 11 (2017) 92–94

Contents lists available at ScienceDirect

Genomics Data

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

Data in Brief

Sequencing and *de novo* transcriptome assembly of *Anthopleura dowii* Verrill (1869), from Mexico

Jorge-Tonatiuh Ayala-Sumuano [a], Alexei Licea-Navarro [b], Enrique Rudiño-Piñera [c], Estefanía Rodríguez [d], Claudia Rodríguez-Almazán [c,*]

[a] Instituto de Neurobilogía, Universidad Nacional Autónoma de México, Blvd. Juriquilla 30001, Juriquilla, Querétaro, México
[b] Departamento de Innovación Biomédica, Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE), Baja California, C.P. 22860, México
[c] Departamento de Medicina Molecular y Bioprocesos, Instituto de Biotecnología, Universidad Nacional Autónoma de México. Av. Universidad 2001, C.P. 62210, Cuernavaca, Morelos, México
[d] Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York NY 10024, USA

ARTICLE INFO

Article history:
Received 12 November 2016
Accepted 18 November 2016
Available online 5 December 2016

Keywords:
Anthopleura dowii
Sea anemone
Illumina
Transcriptome

ABSTRACT

Next-generation technologies for determination of genomics and transcriptomics composition have a wide range of applications. Moreover, the development of tools for big data set analysis has allowed the identification of molecules and networks involved in metabolism, evolution or behavior. By natural habitats aquatic organisms have implemented molecular strategies for survival, including the production and secretion of toxic compounds for their predators; therefore these organisms are possible sources of proteins or peptides with potential biotechnological application. In the last decade anthozoans, mainly octocorals but also sea anemones, have been proben to be a source of natural products. Members of the genus *Anthopleura* are one of the best known and most studied sea anemones because they are common constituents of rocky intertidal communities and show interesting ecological and biological phenomena (e.g. intraespecific competition, symbiosis, etc.); however, many aspects of these taxa remain in need to be analyzed. This work describes the transcriptome sequencing of *Anthopleura dowii* Verrill, 1869 (Cnidaria: Anthozoa: Actiniaria); this is the first report of this kind for these species. The data set used to construct the transcriptome has been deposited on NCBI's database. Illumina sequence reads are available under BioProject accession number PRJNA329297 and Sequence Read Archive under accession number SRP078992.

Specifications	
Organism/cell line/tissue	*Anthopleura dowii Verrill, 1869 (Cnidaria: Anthozoa: Hexacorallia: Actiniaria: Actiniidae)*
Sex	N/A
Sequencer or array type	Illumina Genome Analyzer IIx
Data format	Raw data
Experimental factors	Transcriptome profiling of sea anemone *Anthopleura dowii* Verrill, (1869)
Experimental features	Live specimens of the sea anemone *Anthopleura dowii Verrill*, (1869) was collected in intertidal zone in Ensenada, Baja California, México. Tentacles were dissected and macerated for RNA isolation and de novo transcriptome assembly.
Consent	N/A
Sample source location	Intertidal zone of Ensenada, Baja California, México

* Corresponding author.
E-mail address: claudiar@comunidad.unam.mx (C. Rodríguez-Almazán).

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/sra/SRP078992

2. Introduction

Sea anemones (Cnidaria: Anthozoa: Hexacorallia: Actiniaria) are generally opportunistic predators; they produce stinging capsules generally called nematocyts, which are differentially distributed across the different tissues of the polyp. Historically, tentacles have been the main source of tissue from which toxin peptides have been isolated because of easy access to them and the relatively high contraction of nematocysts in this tissue. Nematocysts have a variety of biologically active compounds such as proteins and peptides with activity of ion channels blockers and modulators, cytolysins, phospholipases and protease inhibitors, all these bioactive molecules of interest for biotechnological applications [1]. These compounds have been isolated mainly from tentacles and acrorhagi using diverse physical or chemical methods, and

http://dx.doi.org/10.1016/j.gdata.2016.11.022

isolated from extracts by fractions guided by bioassay, purify them and determine their chemical composition and action mechanism [2–4].

The use of deep-sequencing technologies as RNA-seq have allowed the identification and quantification of bioactive compounds of biotechnological importance [5]. Nonetheless, only a limited number of transcriptomes from sea anemones have been sequenced so far [6–11] Here we performed *de novo* transcriptome assembly for the sea anemone *Anthopleura dowii* Verrill (1869) (Actiniaria: Enthemonae: Actinioidea: Actiniidae) by next-generation sequencing.

3. Experimental design, materials and methods

3.1. Specimens

Two specimens of the actiniarian *A. dowii* Verril (1869) were collected in the intertidal zone of Ensenada, Baja California, Mexico. Live specimens were transported to the laboratory and tentacles were dissected for RNA isolation.

Specimens were identified using polyp anatomy and the distribution and size of cnidae in various regions of the polyps; in addition, we used two partial mitochondrial markers (partial 12S rDNA and16S rDNA) following molecular methodologies detailed in [12] for specimen identification. Voucher specimens fixed in ethanol have been deposited at the American Museum of Natural History (AMNH).

3.2. RNA extraction, RNA-seq and transcriptome assembly

RNA was isolated using the SV Total RNA Isolation System (Promega) following the protocol provided by the manufacturer. Briefly, 30 mg of tentacle tissue were manually macerated to homogeneity with a Kontes microtube pellet pestle rod (Daigger) in a 1.5 ml microcentrifuge tube with 175 μl of the provided RNA Lysis Buffer. After dilution with 350 μl of the RNA Dilution Buffer the sample was heated for 3 min at 70 °C. Cellular debris were then discarded by centrifugation. The cleared lysate was mixed with 95% ethanol and transferred to one of the spin baskets supplied with the kit. After washing with the RNA Wash Solution, the sample was treated with the provided DNAse for 15 min and then washed twice with the RNA Wash Solution. After centrifugation, total RNA was recovered in nuclease-free water. The RNA was quantitated with a Nanodrop 1000 (Thermo Scientific) and its integrity was confirmed using a 2100 Bioanalyzer (Agilent Technologies). RIN values (RNA integrity number) of 8.5 were obtained, indicating that the RNA had the needed quality to proceed to library construction and high-throughput sequencing.

A complementary DNA (cDNA) library was constructed from the obtained total RNA, using the Illumina TruSeq Stranded mRNA Sample Preparation Kit, following the protocol provided by the supplier. Automated DNA sequencing was performed at the Massive DNA Sequencing facility in the Institute of Biotechnology (Cuernavaca, Mexico) with a Genome Analyzer IIx (Illumina), using a 72 bp paired-end sequencing scheme over cDNA fragments ranging in size of 200–400 bp. Each library consisted of two fastq files (forward and reverse reads), from which the adaptors were clipped-off. The quality of cleaned raw reads was verified with the FastQC program (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/).

3.3. Transcriptome de novo assembly

A. dowii Verrill (1869) RNA-seq raw data was *de novo* assembled using Trinity [13], a program based in De Brujin graphs for the assembly of short reads. Trinity was executed using default parameters for the assembly of paired-end reads. For mapping and abundance estimation, reads were mapped with Bowtie [14], using the reconstructed transcriptome as a reference. Abundance of transcripts were estimated by RSEM [15], as described in the Trinity protocol [16]. Global GC content of the sequences was determined using Emboss GeeCee tool [17].

Table 1
Statistics of the *A. dowii* Verrill (1869) assembly.

No. of raw data bases	5,989,351,536
No. of read pairs	41,592,719
No. of assembled bases	67,045,873
No. of assembled genes	76,580
No. of assembled transcripts	94,775
Assembly GC percent	40.65
Contig N50	1179
Average contig	707.42

A total of 70,097,332 raw reads from Illumina technology were produced for *A. dowii* Verrill (1869) transcriptome sequencing. These reads were assembled by Trinity pipeline resulting in 72,684 contigs with a N50 = 1179 bp, average length of 707 bp. (Table 1). This work provides the first transcriptome assembly for the sea anemone *A. dowii* Verrill (1869). The information presented here may be useful to identify new molecules for biotechnology and pharmaceutical relevance.

3.4. Availability of supporting data

All data used to constuct the transcriptome have been deposited on NCBI's database under the BioProject accession number PRJNA329297. Raw Illumina sequence reads are available in the NCBI's Sequence Read Archive (SRA) under accession number SRR3932753.

Conflict of interest

The authors declare no conflicts of interest regarding the publication of this paper.

Acknowledgements

This work was supported by Consejo Nacional de Ciencia y Tecnología (CONACYT) [grant number 178151]. We thank to Samanta Jimenez (CICESE), Ernesto Ortiz (IBT-UNAM), Itzel Amaro Estrada (INIFAP) and Luciana Gusmão (AMNH) for technical assistance.

References

[1] B. Frazao, V. Vasconcelos, A. Antunes, Se anemone (Cnidaria, Anthozoa, Actinaria) toxins: an overview. Mar. Drug 10 (2012) 1812–1851.
[2] V.H. Junior, F. Zara, S. Marangoni, D.O. de Toyama, A.J. de Souza, S.C. de Oliveira, M.H. Toyama, Identification of two novel cytolysin from the hydrozoan *Olindias sambaquiensis* (Cnidaria). J. Venom Anim. Toxins. Incl. Trop. Dis. 20 (2014), http://dx.doi.org/10.1186/1678-9199-20-10.
[3] W.H. Liu, L. Wang, Y.L. Wang, L.S. Peng, W.Y. Wu, W.L. Peng, X.Y. Jiang, H.B. Tu, H.P. Chen, P. Ou-Yang, A.L. Xu, Cloning and characterization of a novel neurotoxin from the sea anemone *Anthopleura* sp. Toxicon 41 (2003) 793–801.
[4] T. Honma, S. Minagawa, H. Nagai, M. Ishida, Y. Nagashima, K. Shiomi, Novel peptide toxins from acrorhangi, agfressive organs of the sea anemone. Actinia equina 46 (2005) 768–774.
[5] P.F. Campos, D. Andrade-Silva, A. Zelanis, A.F. Paes Leme, M.M. Rocha, M.C. Menezes, S.M. Serrano, I.L. Junqueira-de-Azevedo, Trends in the evolution of snake toxins underscored by an integrative omics approach to profile the venom of the colubrid phalotris mertensi. Genome Biol. Evol. 8 (8) (2016) 2266–2287.
[6] R.R. Helm, S. Siebert, S. Tulin, J. Smith, C.W. Dunn, Characterization of differential transcript abundance through time during *Nematostella vectensis* development. BMC Genomics 14 (2013) 266, http://dx.doi.org/10.1186/1471-2164-14-266.
[7] S.A. Kitchen, C.M. Crowder, A.Z. Poole, V.M. Weis, E. Meyer, De novo assembly and characterization of four Anthozoan (phylum Cnidaria). Transcriptomes. G3 (Bethesda) 5 (2015) 2441–2452, http://dx.doi.org/10.1534/g3.115.020164.
[8] J. Macrander, M.R. Brugler, M. Daly, A RNA-seq approach to identify putative toxins from acrorhagi in aggressive and non-aggressive *Anthopleura elegantissima* polyps. BMC Genomics 16 (2015) 221, http://dx.doi.org/10.1186/s12864-015-1417-4.
[9] M. Laura, C. Borghi, V. Bobbio, A. Allavena, The effect on the transcriptome of *Anemone coronaria* following infection with rust (Tranzschelia discolor). PLoS One 10 (2015) e0118565, http://dx.doi.org/10.1371/journal.pone.0118565.
[10] J. Macrander, M. Broe, M. Daly, Tissue-specific venom composition and differential gene expression in sea anemones. Genome Biol. Evol. 8 (2016) 2358–2375, http://dx.doi.org/10.1093/gbe/evw155.
[11] D.J. Stefanik, T.J. Lubinski, B.R. Granger, A.L. Byrd, A.M. Reitzel, L. DeFilippo, A. Lorenc, J.R. Finnerty, Production of a reference transcriptome and transcriptomic database

(EdwardsiellaBase) for the lined sea anemone, *Edwardsiella lineata*, a parasitic cnidarian. BMC Genomics 15 (71) (2014)http://dx.doi.org/10.1186/1471-2164-15-71.
[12] D. Lauretta, V. Haüssermann, M.R. Brugler, E. Rodríguez, *Isoparactis fionae* sp. nov. (Cnidaria: Anthozoa: Actiniaria) from southern Patagonia, with a discussion of the family Isanthidae. Org. Divers. Evol. 14 (1) (2013) 31–42, http://dx.doi.org/10.1007/s13127–013–0149-z.
[13] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di 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, http://dx.doi.org/10.1038/nbt.1883.
[14] 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 (2009) R25, http://dx.doi.org/10.1186/gb-2009-10-3-r25.
[15] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. BMC Bioinf. 12 (2011) 323, http://dx.doi.org/10.1186/1471-2105-12-323.
[16] B.J. Haas, A. Papanicolaou, M. Yassour, M. Grabherr, P.D. Blood, J. Bowden, M.B. Couger, D. Eccles, B. Li, M. Lieber, M.D. Macmanes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C.N. Dewey, R. Henschel, R.D. Leduc, N. Friedman, A. Regev, De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. Nat. Protoc. 8 (2013) 1494–1512, http://dx.doi.org/10.1038/nprot.2013.084.
[17] G.C. Tool, http://www.bioinformatics.nl/cgi-bin/emboss/geecee2016 (accessed September 7, 2016).