Marine Genomics xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Marine Genomics

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

De novo transcriptome assembly of the hydrocoral *Millepora alcicornis* (branching fire coral) from the Caribbean

Ingrid C. Ortiz-González, Ramón E. Rivera-Vicéns, Nikolaos V. Schizas *

University of Puerto Rico at Mayagüez, Department of Marine Sciences, P.O. Box 9000, Mayagüez, 00681, PR, USA

ARTICLE INFO

Article history:
Received 20 October 2016
Received in revised form 28 November 2016
Accepted 29 November 2016
Available online xxxx

Keywords:
Transcriptome
De novo assembly
Puerto Rico
Coral
Hydrozoan
Hydrocoral

ABSTRACT

The hydrocoral *Millepora* is found in shallow tropical/subtropical regions around the globe and is considered an important reef-building organism. *Millepora alcicornis* is the most common species in the Atlantic Ocean, and can be found from 0.5 to 50 m deep. It is distributed from the tropical/subtropical eastern western Atlantic Ocean, including Bermuda, Brazil, and on the east in Tenerife of the Canary Islands, the Cape Verde Archipelago and Ascension Island. No genomic information is available for this ecologically important group. Here, we report *de novo* transcriptome assembly of *M. alcicornis* sampled from Puerto Rico, Caribbean. We used paired-end sequencing (Illumina HiSeq4000, 2 × 150 bp) and obtained 76,518,693 reads. Transcriptome assembly was performed using Trinity, producing a total of 479,982 transcripts with an average size of 553 bp and a N50 of 749. Data was normalized using RSEM and filtered by a TPM of 3. Open reading frames (ORFs) from the filtered transcripts were obtained by TransDecoder using the hydrozoan *Hydra vulgaris* protein sequences as reference, generating 16,024 putative ORFs. Blast searches showed that 25.8% (4137) of the ORFs matched *H. vulgaris* and 24.6% matched other anthozoan cnidarians (*i.e. Nematostella vectensis* = 1621, *Exaiptasia pallida* = 1280, *Acropora digitifera* = 1050). Gene ontology generated by Blast2GO resulted in a total of 7220 ORFs associated with molecular function, 4917 with biological processes, and 2989 with cellular components. A general overview will be presented of the first assembled *M. alcicornis* transcriptome with emphasis on shared genes among Hydrozoans and Scyphozoans.

1. Introduction

Hydrocorals of the genus *Millepora* are a relatively rich-species genus with 18 species distributed around the globe. They inhabit reefs at depths between 0.5 m down to 50 m (Lewis, 2006) and usually grow upright in finger-like branches, leaf-like, or blade-like patterns. *Millepora* spp. grows also over other substrata such as rocks, piers, mangrove roots, living corals, gorgonians and sponges (Banaszak et al., 2003; Lewis, 2006; Wahle, 1980). Milleporids provide habitat for many species such as small fish, ophiuroids, barnacles, micro-crustaceans among others. In the Caribbean, *milleporids* are important reef builders and are commonly found in shallow and turbulent waters forming dense reef rims contributing to the stabilization and complexity of the carbonate structure (Edmunds, 1999; Lewis, 2006).

Millepora alcicornis (Linneaus 1758) is the most widely dispersed fire coral in the Atlantic in comparison with the other seven species in the same ocean. It has an extensive distribution from the tropical/subtropical western Atlantic Ocean, including Brazil, to the west coast of Africa in Cape Verde Archipelago, and Ascension Island (Amaral et al., 2008; Clemente et al., 2011; de Weerdt and Glynn, 1991; Hoeksema et al., 2014; Morri et al., 2000). More recently, *M. alcicornis* was reported from the southeastern coast of Tenerife but this record has been deemed as a recent invasion to the Canary Islands (Clemente et al., 2011; López et al., 2015). Dispersion during the medusoid stage may not be as effective due to the short pre-competency period time of the hydromedusae in the water column (Edmunds, 1999; Lewis, 2006). However, asexual reproduction through fragmentation of this branching hydrocoral is substantial during disturbances (Edmunds, 1999; Lewis, 2006) and may contribute more to the dispersal. In addition, the success of *M. alcicornis* to settle to distant locations can be attributed to transportation through ballast waters of large vessels and fouling of hulls (Clemente et al., 2011; López et al., 2015).

Despite the prevalence of *Millepora* in shallow water reefs, there is uncertainty on the number of true species in the genus. All the previous studies aiming to identify *Millepora* species are mainly based on morphological characters such as number, diameter and distance of gastropores and dactylopores, the presence or absence of the ampullae, and the shape, texture and character of the corallum (Boshma, 1948; de Weerdt and Glynn, 1991). This method has not produced consensus results to distinguish species because of the great morphological variability within and between species (Amaral et al., 2002; Lewis, 2006). In fact, the most recently described species from Brazil, *M. laboreli* (Amaral et al., 2008), was based exclusively on morphological

* Corresponding author.
E-mail address: nschizas@gmail.com (N.V. Schizas).

http://dx.doi.org/10.1016/j.margen.2016.11.005
1874-7787/

characters. Only recently, sequencing and molecular markers have been used to separate species of *Millepora*. Tepper et al. (2012) used the nuclear ribosomal region of 18S, ITS-1, 5.8S, ITS-2 and 28S to differentiate three groups of morphotypes: 1) *M. complanata*, 2) *M. alcicornis*, and 3) a set of intermediate morphotypes between these two species. The three morphotypes were grouped into two clades, but no differences were found among them using the nuclear ribosomal region. In a subsequent study by Ruiz-Ramos et al. (2014), they used the mitochondrial gene COI to differentiate the four recognized Caribbean species: *M. complanata*, *M. alcicornis*, *M. squarrosa* and *M. striata*. They proposed a "species complex" composed of *M. complanata*, *M. alcicornis* and *M. striata*, and identifying only *M. squarrosa* as a distinct species. It is clear that the existing morphological and molecular data has not reached consensus and the taxonomy of *Millepora* needs new methodologies and data to decipher the species differentiation problem. Here, we present the *M. alcicornis* transcriptome as a potential contribution to the resolution of the *Millepora* species problem. Differences in novel or low/high expressed genes could be used to further study speciation in *Millepora*.

2. Methods and analysis

2.1. Sampling collection and maintenance

Four colonies of *M. alcicornis* were collected from different reefs of the southwest coast of Puerto Rico (Table 1, Supplementary Fig. 1). All corals were obtained under a research permit from the Department of Natural and Environmental Resources of Puerto Rico (O-VS-PVS15-MA-00016-26092014) and sampling was conducted under their outlined auspices and regulations. Using a combination of SCUBA diving and snorkeling, coral fragments were collected from colonies between 0 and 3 m depth. After the four colonies were collected, we placed them in a closed aquarium system for a period of 8 weeks to limit gene expression driven by ambient variability. The aquarium was maintained with cycles of 12 h of broad light spectrum followed by 6 h of LED moonlight and 6 h of no light. Conditions in the aquarium were 27 °C, salinity of 34 ppt, pH of 8.2 and nitrate of 0 mg/L. Replacement of water lost by evaporation was performed when needed.

Table 1
MixS information of *Millepora alcicornis*.

Item	Definition
General feature of classification	
Classification	Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Cnidaria; Hydrozoa; Hydroidolina; Anthoathecata; Capitata; Milleporidae; *Millepora alcicornis*
Investigation type	Eukaryote transcriptome
Project name	*Millepora alcicornis* transcriptome
Environment	
Geographic location	Caribbean Sea: Puerto Rico, Guánica: Coral Reef Puerto Rico, Guánica: La Jungla Reef Puerto Rico, Lajas: Enrique Reef Puerto Rico, Cabo Rojo: Playita Azul Reef
Latitude, longitude	Coral reef: 17° 56.280′N, 66° 53.406′W La Jungla Reef: 17° 56.650′N, 66° 58.033′W Enrique Reef: 17° 57.288′N, 67° 3.155′W Playita Azul Reef: 18° 8.072′N, 67° 11.245′W
Collection date	2015-10
Environment properties	Shore reef
Depth	<3 m
Collector	Ingrid C. Ortiz González and Orlando Espinosa Ortiz
Sequencing	
Sequencing method	Illumina HiSeq4000; Paired-end (2 × 150)
Estimated size	500 Mbp
Assembly	
Method	*De novo* assembly
Program	Trinity 2.1.1
Finishing strategy	High quality transcriptome assembly
Data accessibility	
DDBJ/EMBL/GenBank	GFAS01000000

2.2. Molecular techniques and sequencing

After the 8 week period in the aquarium, *M. alcicornis* colonies were stored for one hour at −80 °C before RNA extractions. Total RNA was extracted using the TRIZOL RNA Isolation Method (Chomczynski and Mackey, 1995) and concentrations were measured using a NanoDrop2000. A total of four samples of *M. alcicornis* were sent to the Genomics Sequencing and Analysis Facilities (GSAF) in the University of Texas at Austin where sample quality assessment and library preparation was performed using RNA low cost high throughput and Poly-A mRNA capture methods. The quality and quantity of RNA before the library preparation step were checked with the 2100 Bioanalyzer Instrument (Agilent Inc.). Sequencing was performed on an Illumina HiSeq4000 platform with 150 bp paired-end reads.

2.3. Transcriptome assembly

We obtained 76.5 million paired-reads of *M. alcicornis*. FastQC v0.11.5 (Andrews, 2010) was used for quality assessment of reads, adapters and barcode contamination. Adapters were removed from sequences using Cutadapt 1.9.1 (Martin, 2011) followed by another run of FastQC. All reads were also trimmed to remove the last 10 bp of each read using FastX-Toolkit 0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/index.html) generating high quality reads. Assembly was performed using Trinity (2.1.1) (Grabherr et al., 2011) with a minimum contig length of 200 (k = 200). Statistical analysis of the assembled transcripts resulted in an average contig length of 553, N50 of 749, with a total of 479,982 sequences (Supplementary Table 1). Reads from the assembled transcriptome were checked for possible contaminant and vector sequences using NCBI Transcriptome Shotgun Assembly (TSA) check steps for submission of transcriptomes, before proceeding to the next analysis. Reads marked as contaminants or vectors were removed from the assembly before filtering steps (from 479,982 to 476,422 sequences). RSEM 1.2.31 (Li and Dewey, 2011) was used to normalize the transcripts and filter the data by a value of 3 TPM (Transcripts per Million) or higher. From a total of 476,422 sequences, only 38,731 had a value of TPM of 3 or more. The TPM method was selected instead of RPKM and FPKM because the mean expressed transcripts may vary between samples as explained by Li et al. (2010). BUSCO v1.22 (Simão et al., 2015) was also utilized to assess the transcriptome completeness using the BUSCO metazoans single-copy orthologs database. From a total of 843 sequences from the metazoan database, the *M. alcicornis* transcriptome had 414 and 29 complete and fragmented sequences, respectively (Supplementary Table 1).

2.4. Functional annotation

Open reading frames (ORFs) from the filtered transcripts were obtained using TransDecoder 3.0.0 (http://transdecoder.github.io) using the *Hydra vulgaris* protein sequences (UniProt) as reference and a minimum ORF length of 375 bp. A total of 16,024 putative ORFs were generated in which 7551 were complete ORFs and 8473 were incomplete ORFs with the 5′ or 3′ end only (minimum ORF length = 375 bp). All the ORFs were used to generate blast searches, annotation and gene ontology assignments with the program Blast2GO 4.0 (Conesa et al., 2005). Blast searches showed that 25.8% (4137) of the ORFs matched *H. vulgaris* (Fig. 1) and 24.6% matched other anthozoan cnidarians (*i.e.* the sea anemones *Nematostella vectensis* = 1621 and *Exaiptasia pallida* = 1280, and the scleractinian coral *Acropora digitifera* = 1050). These results are as expected because *M. alcicornis* is hydrocoral from the Class Hydrozoa, and not a coral of the Class Anthozoa.

Please cite this article as: Ortiz-González, I.C., et al., *De novo* transcriptome assembly of the hydrocoral *Millepora alcicornis* (branching fire coral) from the Caribbean, Mar. Genomics (2016), http://dx.doi.org/10.1016/j.margen.2016.11.005

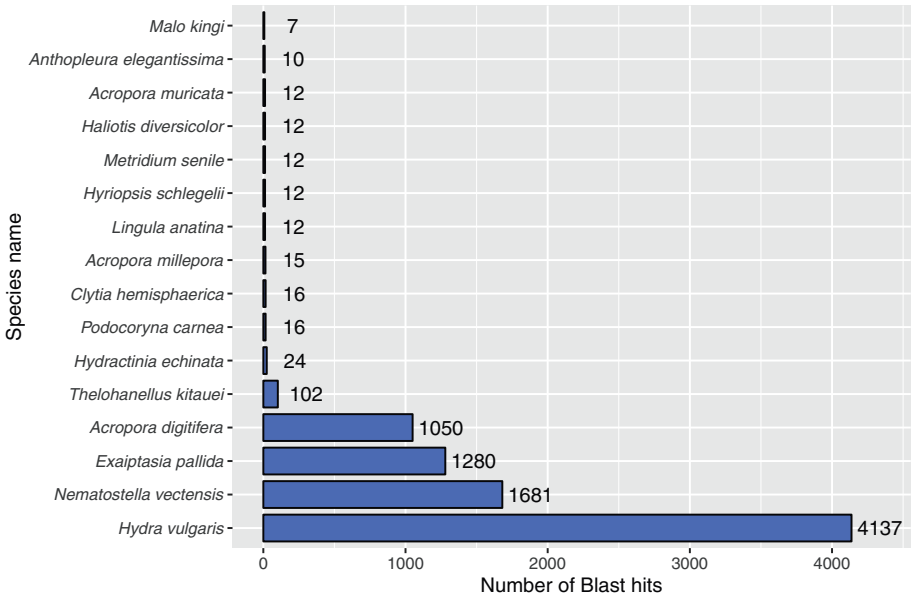

Fig. 1. Blast hits results against the NCBI database.

In terms of gene ontology, a total 15,126 (94.39%) ORFs were successfully assigned an ontology classification. The majority of the ORFs were assigned to Molecular Function with 7220, followed by Biological Processes and Cellular Component with 4917 and 2989, respectively (Fig. 2).

2.5. Expressed genes analysis and comparisons

In *de novo* assembled transcriptome of *M. alcicornis*, we found sequences of the *Millepora* nematocyst protein (MCTx-1), originally described by Iguchi et al. (2007). In addition, we found many proteins sequences that have been proposed to be part of the gastropore, dactylopores and developing medusa in other hydrozoans and scyphozoans transcriptome studies (Supplementary Table 2). Sanders et al. (2014) described proteins found in the dactylozooid, gastrozooid and gonozooid polyps of the hydrozoan *Hydractinia symbiolongicarpus*. Similarly, the *M. alcicornis* transcriptome yielded protein sequences of *Cnox-2*, found in the gastrozooid and dactylozooid polyps of *H. symbiolongicarpus*, and *pmp1*, found in the mouth of the polyp (also found in the manubrium of the medusa) of *Podocoryna carnea* (Supplementary Table 2). Many toxin sequences were also found in *M. alcicornis* transcriptome, but with no similarity to the toxin sequences described in the gastrozooid and dactylozooid polyps of *H. symbiolongicarpus* (Sanders et al., 2014), indicating the existence of a variety of toxins in hydrozoans. Other sequences related with the dactylozooid polyp of *H. symbiolongicarpus* found in the *M. alcicornis* transcriptome were: *cerebrus*, Wnt, myosin heavy chain. Some of the expressed genes in the gonozooid polyps of *H. symbiolongicarpus* were also found in our study. These sequences are the *hedgehog* and *capicua* proteins, related with the development of oocytes and the gastrodermis of the male gonopores, respectively. The BMP protein sequence, associated with the embryonic development of metazoans, was also expressed in the *M. alcicornis* transcriptome.

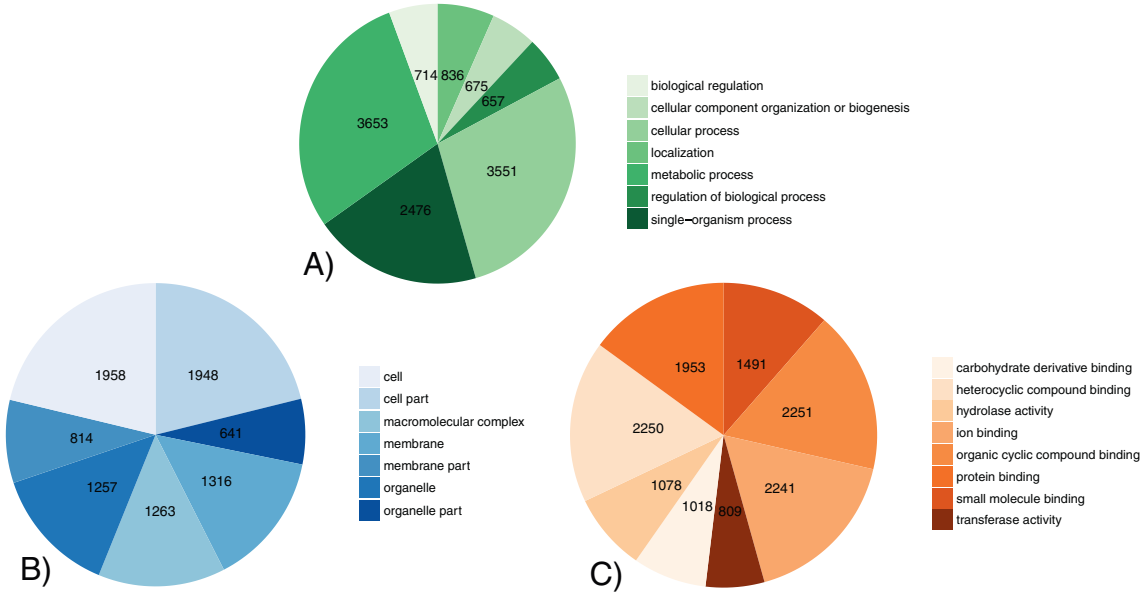

Fig. 2. Gene Ontology graphs. A) Molecular function (level3). B) Cellular component (level2). C) Biological processes (level2).

Please cite this article as: Ortiz-González, I.C., et al., *De novo* transcriptome assembly of the hydrocoral *Millepora alcicornis* (branching fire coral) from the Caribbean, Mar. Genomics (2016), http://dx.doi.org/10.1016/j.margen.2016.11.005

Millepora alcicornis has a brief medusoid stage before it settles to its sessile, benthic phase. We found several expressed genes (Wnt, vWFA, hemicentin, BHTM (betaine-homocysteine methyltransferase) and myosin heavy and light chains) that have been associated with the locomotive functions of the medusa stage in the common scyphozoan *Aurelia aurita* (Brekhman et al., 2015). The transcriptome of *M. alcicornis* is based on an adult colony (polypoid stage of the life cycle), however, it is possible that microscopic (~0.5 mm) *Millepora* medusae were present on the tissue we used for RNA extraction. Interestingly, the hydrozoan *H. symbiolongicarpus* exhibits no medusoid stage, and no sequences related with that developmental stage were found (Sanders et al., 2014). Our study presents the first hydrocoral transcriptome, which could serve as a potential tool to address taxonomic uncertainties in this group and as a base to measure responses of *M. alcicornis* to the rapidly changing environment of coral reefs of the Caribbean. The expression of proteins of different polyps of hydrozoans in our study demonstrates that the sequencing effort was effective and can be replicated for further comparison studies in these organisms.

3. Data accessibility

This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GFAS00000000. The version described in this paper is the first version, GFAS01000000.

Acknowledgments

We thank Orlando Espinosa for field assistance and the aquarium preparation. This project was supported by Arts and Sciences Faculty SEED money and Sea Grant Puerto Rico SEED money for ICOG. We also thank Dr. Richard Appeldoorn and the Caribbean Coral Reef Institute (CCRI), who partially supported the sequencing of the transcriptome. The sequence analysis was run in a workstation computer purchased through Sea Grant award R-101-1-14 to NVS.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.margen.2016.11.005.

References

Amaral, F.D., Broadhurst, M.K., Cairns, S.D., Schlenz, E., 2002. Skeletal morphometry of *Millepora* occurring in Brazil, including a previously undescribed species. Proc. Biol. Soc. Wash. 115, 681–695.

Amaral, F.D., Steiner, A.Q., Broadhurst, M.K., Cairns, S.D., 2008. An overview of the shallow-water calcified hydroids from Brazil (hydrozoa: Cnidaria), including a description of a new species. Zootaxa 68, 56–68.

Andrews, S., 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Banaszak, A.T., Ayala-Schiaffino, B.N., Rodríguez-Román, A., Enríquez, S., Iglesias-Prieto, R., 2003. Response of *Millepora alcicornis* (Milleporina: Milleporidae) to two bleaching events at Puerto Morelos reef, Mexican Caribbean. Rev. Biol. Trop. 51 (4), 57–66.

Boshma, H., 1948. The species problem in *Millepora*. Zool. Verh 1, 1–115.

Brekhman, V., Malik, A., Haas, B., Sher, N., Lotan, T., 2015. Transcriptome profiling of the dynamic life cycle of the scyphozoan jellyfish *Aurelia aurita*. BMC Genet. 16, 84.

Chomczynski, P., Mackey, K., 1995. TRIzol RNA isolation protocol. BioTechniques 19 (6), 942–945.

Clemente, S., Rodríguez, A., Brito, A., Ramos, A., Monterroso, Ó., Hernández, J.C., 2011. On the occurrence of the hydrocoral *Millepora* (hydrozoa: Milleporidae) in the subtropical eastern Atlantic (Canary Islands): is the colonization related to climatic events? Coral Reefs 30 (1), 237–240.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21 (18), 3674–3676.

de Weerdt, A., Glynn, P., 1991. A new and presumably now extinct species of *Millepora* (hydrozoa) in the eastern Pacific. Zool. Meded. 65 (15), 267–276.

Edmunds, P., 1999. The role of colony morphology and substratum inclination in the success of *Millepora alcicornis* on shallow coral reefs. Coral Reefs 18 (2), 133–140.

Grabherr, M.G., Hass, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amir, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, K., Regev, A., 2011. Trinity: reconstruction a full-length transcriptome without a genome from RNA-Seq data. Nat. Biotechnol. 29 (7), 644–652.

Hoeksema, B.W., Nunes, F.L.D., Lindner, A., De Souza, J.N., 2014. *Millepora alcicornis* (hydrozoa: Capitata) at Ascension Island: confirmed identity based on morphological and molecular analyses. J. Mar. Biol. Assoc. UK http://dx.doi.org/10.1017/S0025315414001283.

Iguchi, A., Iwanaga, S., Nagai, H., 2007. Isolation of a novel protein toxin from fire coral. Biochem. Biophys. Res. Commun. 365 (1), 107–112.

Lewis, J.B., 2006. Biology and ecology of the hydrocoral *Millepora* on coral reefs. Ad. Mar. Biol. 50 (05), 1–55.

Li, B., Dewey, C., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12 (1), 323.

Li, B., Ruotti, V., Stewart, R.M., Thomson, J.A., Dewey, C.N., 2010. RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics 26 (4), 493–500.

López, C., Clemente, S., Almeida, C., Brito, A., Hernández, M., 2015. A genetic approach to the origin of *Millepora* sp. in the eastern Atlantic. Coral Reefs 34 (2), 631–638.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. j. 17 (1), 10.

Morri, C., Bianchi, C.N., Cattaneo-Vietti, R., Sartoni, G., 2000. Shallow Epibenthic Communities of Ilha Do Sal (Cape Verde Archipelago, Eastern Atlantic). III Simpósio Fauna e Flora Das Ilhas Atlânticas, Supplement. pp. 157–165.

Ruiz-Ramos, D.V., Weil, E., Schizas, N.V., 2014. Morphological and genetic evaluation of the hydrocoral *Millepora* species complex in the Caribbean. Zool. Stud. 53, 4.

Sanders, S.M., Shcheglovitova, M., Cartwright, P., 2014. Differential gene expression between functionally specialized polyps of the colonial hydrozoan *Hydractinia symbiolongicarpus* (phylum Cnidaria). BMC Genomics 15 (1), 1.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinform. btv 351.

Tepper, C., Squiers, L., Hay, C., Gorbach, D., Friend, D., Black, B., Greenstein, B., Strychar, K., 2012. Cryptic species: a mismatch between genetics and morphology in *Millepora*. Mark. Sci. 2 (5), 57–65.

Wahle, C.M., 1980. Detection, pursuit, and overgrowth of tropical gorgonians by milleporid hydrocorals: Perseus and Medusa revisited. Science 209, 689–691.

Please cite this article as: Ortiz-González, I.C., et al., *De novo* transcriptome assembly of the hydrocoral *Millepora alcicornis* (branching fire coral) from the Caribbean, Mar. Genomics (2016), http://dx.doi.org/10.1016/j.margen.2016.11.005