Marine Genomics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Marine Genomics

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

Method paper

De novo transcriptome assembly of the coral *Agaricia lamarcki* (Lamarck's sheet coral) from mesophotic depth in southwest Puerto Rico

Alex J. Veglia[a], Nicholas M. Hammerman[a,b], Ramón E. Rivera-Vicéns[a,c], Nikolaos V. Schizas[a,*]

[a] University of Puerto Rico at Mayagüez, Department of Marine Sciences, PO Box 9000, Mayagüez, Puerto Rico 00681, USA
[b] School of Biological Sciences, University of Queensland, Gehrmann Laboratories, Level 8, Research Road, St. Lucia, QLD 4072, Australia
[c] Department of Earth and Environmental Sciences, Paleontology and Geobiology, Ludwig-Maximilians-Universität München, Munich, Germany

ARTICLE INFO

Keywords:
Gene ontology
NCGAS
Scleractinian
MCEs
Caribbean
RNA-seq

ABSTRACT

The plating coral, *Agaricia lamarcki* is a widely distributed species inhabiting reefs across the Caribbean basin and Florida. This species is of interest since it is considered a depth-generalist, found from 10 to 70 m. Given the scope of contemporary studies on this coral's population dynamics and physiology, as well as, the potential of mesophotic reefs to be refuge habitats for deteriorated shallow water reefs, we present the first *de novo* transcriptome assembly of an important mesophotic coral. Using next-generation paired-end sequencing (Illumina Hiseq4000; 2 × 150 bp), we obtained a total of 82,506,058 raw reads. The novel transcriptome assembly strategy included the recently developed National Center for Genome Analysis Support *de novo* transcriptome assembly pipeline. Assembly produced a total of 101,322 biologically true, non-redundant transcripts with an average contig length of 959 and N50 of 1830. EvidentialGene and TransDecoder were used to identify open reading frames (ORFs) with homology insight provided by the UniProtKb and PFAM databases. ORF prediction resulted in 38,517 putative ORFs of which 12,107 ORFs were annotated as genes dealing with molecular function, 1266 with biological processes and 416 with cellular components.

1. Introduction

The family Agariciidae in the Caribbean is comprised of two genera and eight taxonomically accepted species with a wide morphological range, including encrusting sheets, thick vertical projecting leaves, flat plates, and rounded surfaces (WoRMS, 2018). *Agaricia* spp. occupy many different ecological niches on Caribbean reefs, inhabiting fore reef, slope and deep channel habitats from shallow to mesophotic depths. Within many western Atlantic nations, they are the primary scleractinians encountered at deep mesophotic reefs (Bongaerts et al., 2013, 2015; Appeldoorn et al., 2016, 2019; Hoeksema et al., 2017). *Agaricia* spp. also facilitates many species interactions, notably those between other corals, associated micro-invertebrates (Veglia et al., 2018), sponges (Garcia-Hernandez et al., 2016) and small fish (Jackson and Buss, 1975; Aerts, 1998).

Agaricia lamarcki (Milne Edwards & Haime, 1851) is of special interest since it is a gonochoric brooder and considered a depth-generalist (*e.g.* occurring from 10 to 75 m) with a wide geographic distribution throughout the Caribbean basin, Gulf of Mexico, Florida and Bahamas (Aronson et al., 2008). A wide depth distribution is possible due to both morphological and physiological adaptations such as depth-differentiated clades of *Symbiodinium* C and D (Bongaerts et al., 2013; Lucas et al., 2016), and variable plate thickness. Shallow populations of *Agaricia lamarcki* have experienced drastic declines in the past few decades leading to its listing as vulnerable on the IUCN red list of threatened species (Aronson et al., 2008). Concordantly, transplantation work found that relocation depth had the strongest effect on coral fragment mortality, and that shallow *A. lamarcki* fragments were far more likely to experience bleaching than fragments at mesophotic depths (Laverick and Rogers, 2018). Within southwest Puerto Rico relatively healthy populations persist, perhaps due to genetic panmixia found between mesophotic and adjacent shallow colonies, plausibly forming one continuous interbreeding population (Hammerman et al., 2018).

Coral genomes and transcriptomes produced over the past decade (Shinzato et al., 2014; Ortiz-González et al., 2017; Mansour et al., 2016; Kenkel and Bay, 2017) have been crucial for providing insight into coral evolution and stress-response physiology (Kenkel et al., 2013; Bhattacharya et al., 2016; Lin et al., 2015). Here we present the first high quality reference transcriptome of the scleractinian *Agaricia lamarcki* collected from mesophotic depth in southwestern Puerto Rico. *Agaricia lamarcki* is an emerging model species in coral genomics and

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

https://doi.org/10.1016/j.margen.2018.08.003
Received 4 May 2018; Received in revised form 6 August 2018; Accepted 21 August 2018

ecology and there is significant interest in how this organism sustains healthy populations on mesophotic reefs. The sequencing and characterization of the *A. lamarcki* transcriptome facilitates future conservational genomic studies promoting long-term survival of the species. There are currently no published coral transcriptomes associated with mesophotic habitats and thus the *A. lamarcki* transcriptome reported here represents a unique resource for inferring coral gene functioning at these depths. Furthermore, we present a novel coral transcriptome assembly strategy that performs biological contamination removal prior to the assembly process that incorporates multiple *de novo* assembly programs. By presenting this different methodology we hope to ignite a discussion for coral transcriptome assembly method standardization to facilitate more accurate and appropriate cross-species comparisons in the future. We also report novel coral-microbe associations through taxonomic analysis of reads discarded as contamination to provide an example of significant biological insight that can be inferred from next generation sequencing contamination.

2. Methods and analysis

2.1. Sampling collection

A colony of *Agaricia lamarcki* was sampled at 45 m depth on the shelf edge of Guánica, Puerto Rico (Table 1, Supplementary Fig. 1). The sample was collected with a research permit from the Department of Natural and Environmental Resources of Puerto Rico (O-*VS*-PVS15-MA-00016-26,092,014). Open circuit SCUBA diving was used to collect the *A. lamarcki* sample with a hammer and chisel. The coral fragment was immediately brought to the lab rinsed and placed in a dry zip-lock bag in a −80 °C freezer until molecular processing.

2.2. Molecular techniques and sequencing

Coral tissue near the edge of the colony was removed from the $CaCO_3$ skeleton with a sterile razor blade and triturated with a pestle and total RNA was extracted using the TRIZOL RNA Isolation method (Chomczynski and Mackey, 1995). Concentrations of extracted RNA were verified and quantified using the Nanodrop2000 (Thermo Fisher Scientific). Samples were sent in dry ice to the Duke Center for Genomic and Computational Biology at Duke University for further quality and quantity assessment (2100 Agilent Bioanalyzer) and library preparation using the Kapa stranded mRNA seq kit (Cat # KK8421). Total RNA sequencing with poly-A tail selection was then performed on an Illumina Hiseq4000 with 150 bp paired-end reads.

Table 1
MixS data description.

Item	Definition
General feature of classification	
Classification	Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Cnidaria; Anthozoa; Hexacorallia; Scleractinia; Agariciidae; *Agaricia*; *Agaricia lamarcki*
Investigation type	Eukaryote transcriptome
Project name	*Agaricia lamarcki* transcriptome
Environment	
Geographic location	Caribbean Sea: Puerto Rico, Guánica: Falling Rock Reef
Latitude, longitude	Falling Rock Reef: 17° 53.922′N, 66° 56.622′W
Collection date	2/18/2017
Environment properties	Shelf edge: mesophotic coral ecosystem
Depth	45 m
Collector	Nicholas M. Hammerman
Sequencing	
Sequencing method	Illumina Hiseq4000; Paired-end (2 × 150)
Assembly	
Method	*De novo* assembly
Program	Trinity v2.6.6, SOAPdenovo-Trans, Velvet v1.2.10, Oases v0.2.09, Trans-ABySS, EvidentialGene v2013.07.27
Finishing strategy	High quality transcriptome assembly
Accessibility	
DDBJ/ENA/GenBank	GGLC00000000

2.3. Transcriptome assembly

To ensure the highest quality reference transcriptome possible, we followed several assembly strategies differing in contamination removal procedure and assembly of the RNA reads. The different methods are further detailed in Table 2. Here, we present the methodology that produced the highest quality assembly based off general transcriptome quality assessment metrics (*e.g.* N50, L50, BUSCO, *etc.*; Honaas et al., 2016).

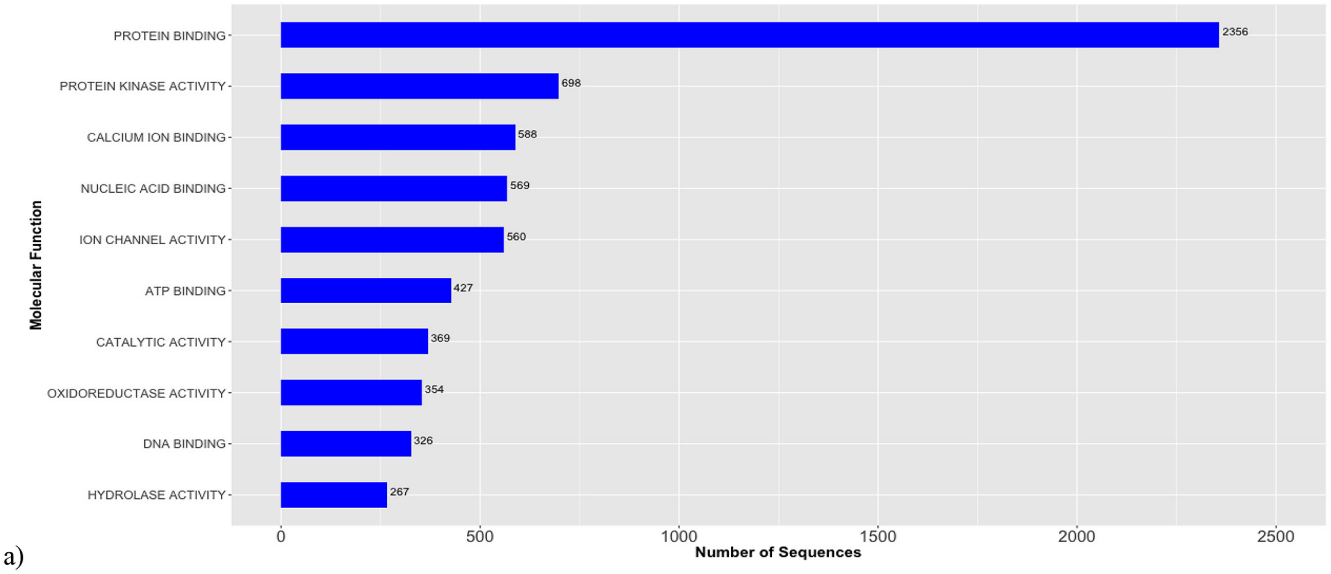
Sequencing produced 82.5 million raw paired-end reads, and initial library quality checks were done using FastQC v0.11.5(Andrews, 2010). Adapter sequences were removed from libraries by performing sequential filtering with increasing stringency using the bbduk.sh script found within the Bestus Bioinformaticus (BB) Tool Kit (http://sourceforge.net/projects/bbmap). A final check and removal step with AdaptorRemoval v2 (Schubert et al., 2016) was done prior to further filtering. Quality filtering was then performed using bbduk.sh removing all reads with an average per base Phred score below 10. Before assembling the reads, read libraries were scanned for biological contamination using NCBI BLAST v2.7.1 (Altschil et al., 1990) and two contamination databases: i. Custom *Symbiodinium* genome database [Shoguchi et al. (2013; *S. minutum*); Lin et al. (2015; *S. kawagutii*); Aranda et al. (2016; *S. microadriaticum*); unpublished data from the ReFuGe 2020 Consortium, C.X. Chan, P. Lundgren, C.R. Voolstra (clades C1and C15)] and a ii. Marine bacteria database [MarDB V2 (https://mmp.sfb.uit.no/databases/)]. To improve the accuracy and computational efficiency of the *de novo* assembly (Martin and Wang, 2011), any read validated as biological contamination was removed prior to assembly and stored for downstream taxonomic analysis. Processed read libraries were then quality checked again with FastQC v0.11.5 before proceeding to the *de novo* assembly.

The remaining high quality reads were assembled following the National Center for Genome Analysis Support (NCGAS) *de novo* transcriptome assembly pipeline (github.com/NCGAS/de-novo-transcriptome-assembly-pipeline). First, multiple *de novo* assemblies with varying kmer values using different assembly software [Trinity v2.6.6 (Grabherr et al., 2011), SOAPdenovo-Trans v1.03 (Xie et al., 2014), Velvet v1.2.10 (Zerbino and Birney, 2008), Oases v0.2.09 (Schulz et al., 2012), Trans-ABySS v2.0.1 (Robertson et al., 2010)] were generated and combined. The program EvidentialGene v2013.07.27 (Gilbert, 2012) was then used to generate a clean consensus *Agaricia lamarcki* transcriptome from the aforementioned set of *de novo* assemblies. EvidentialGene is a genome informatics project for high quality and accurate gene sets for animals and plants. Using a combination of standard gene prediction and contemporary genome assembly methods, EvidentialGene accurately predicts biologically real and unique transcripts and significantly reduces the presence of false transcripts within the final assembly of isogroups. False transcripts can be a problem in single-assembler assemblies, especially when prior genomic information is absent. Once a high quality consensus transcriptome assembly was acquired, basic statistical analyses with QUAST v4.6.3 (Mikheenko et al., 2016) showed an average contig length of 959, N50 of 1830, and a total of 101,322 biologically true, non-redundant transcripts (Table 2).

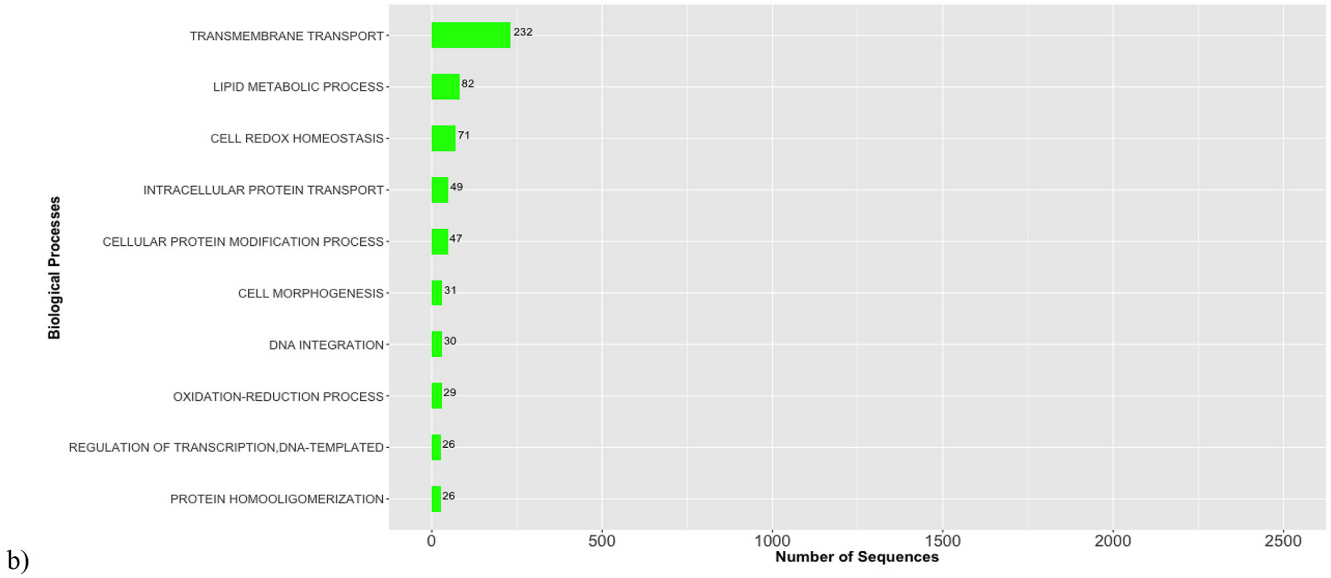
The Transcriptome Shotgun Assembly check steps for submission were used to confirm that the *A. lamarcki* assembly was contamination-free prior to transcriptome completeness analysis and gene ontology calling. BUSCO v3.0.2 (Simão et al., 2015; Waterhouse et al., 2017) with the BUSCO metazoans single-copy orthologs reference database

Table 2
Transcriptome assembly comparison table including *Agaricia lamarcki* assemblies and the three coral transcriptomes produced by Kenkel and Bay (2017): i. *Galaxea ateata;* ii. *Galaxea archelia*; iii. *Goniopora columna*. All *Agaricia lamarcki* assemblies were produced in this study with assembly a being the one described in the body of the manuscript. BUSCO results are shown for each assembly with complete single-copy (S) and duplicated (D) BUSCO percentages for *A. lamarcki* assemblies.

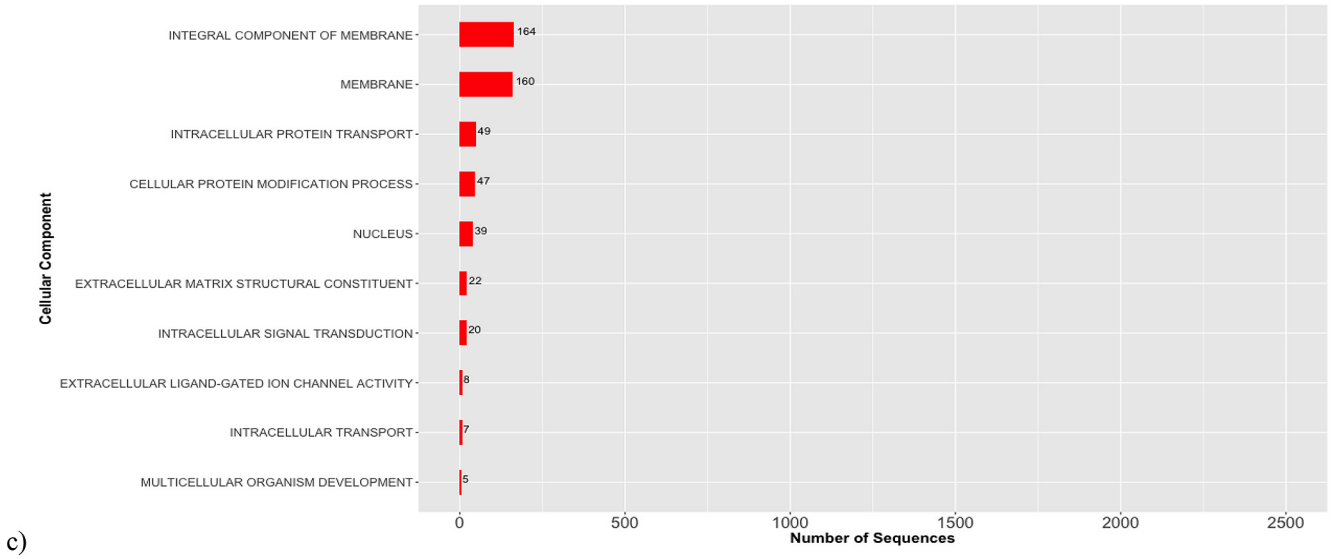
	Agaricia lamarcki (assembly a)	*Agaricia lamarcki* (assembly b)	*Agaricia lamarcki* (assembly c)	*Agaricia lamarcki* (assembly d)	i. *Galaxea atreata*; ii. *Galaxea archelia*; iii. *Goniopora columna*
Extraction method	TRIZOL	TRIZOL	TRIZOL	TRIZOL	Aurum Total RNA mini kit
Illumina sequencer	Hiseq4000	Hiseq4000	Hiseq4000	Hiseq4000	Hiseq3000/4000
Assembly program	NCGAS pipeline	Trinity v2.6.6	NCGAS pipeline	Trinity v2.6.6	Trinity v.2.0.6
Contamination filtering	Pre-assembly	Pre-assembly	Post-assembly	Post-assembly	Post-assembly
BLASTx Filtering (marine bacteria)	MarDB V2	MarDB V2	MarDB V2	MarDB V2	N/A
BLASTx Filtering (*Symbiodinium*)	*S. minutum*, *S. kawagutii*, *S. microadriaticum*, and unpublished genomic data (clades C1and C15)	*S. minutum*, *S. kawagutii*, *S. microadriaticum*, and unpublished genomic data (clades C1and C15)	*S. minutum*, *S. kawagutii*, *S. microadriaticum*, and unpublished genomic data (clades C1and C15)	*S. minutum*, *S. kawagutii*, *S. microadriaticum*, and unpublished genomic data (clades C1and C15)	*S. kawagutii*
N raw reads ($\times 10^6$)	82.5	82.5	82.5	82.5	i. 92.8; ii. 96; iii. 102.8
N quality filtered: PE, SE ($x10^6$)	56.9, 0	56.9, 0	82.3, 0	82.3, 0	i. 35, 5.8; ii. 33.3, 6.0; iii. 26.9, 4.7
N contigs/transcripts (before BLASTx filtering)	N/A	N/A	114,181	395,600	i. 173,883; ii. 164,996; iii.185,625
N contigs/transcripts (after BLASTx filtering)	101,322 (contigs > 200 bp)	339,208 (contig > 200)	58,762 (contig > 200 bp)	264,079 (contig > 200 bp)	i. 65,460; ii. 67,127; iii. 72,405;
N isogroups/genes (% annotated)	38,517	N/A	N/A	N/A	i. 29,145; ii. 26,693; iii. 37,894
Mean contig length (bp)	959	661	Pre-BLASTX filtering: 1007; post-BLASTx filtering: 834	Pre-BLASTX filtering: 708; post-BLASTx filtering: 583	i. 1754; ii. 1894; iii. 1492
N50 (bp)	1830	1691	Pre-BLASTX filtering: 1917; post-BLASTx filtering: 2099	Pre-BLASTX filtering: 1770; post-BLASTx filtering: 1714	i. 2300; ii. 2480; iii. 1984;
GC (%)	48.62	45.1	Pre-BLASTX filtering: 49.13; post-BLASTx filtering: 41.95	Pre-BLASTX filtering: 46.45; post-BLASTx filtering: 40.55	i. 42.3%; ii. 42.1%; iii. 42.2%
BUSCOs searched	978	978	978	978	978
Complete BUSCOs	899 [S:75.2%; D:16.8%]	916 [S:47.8%, D:45.9%]	Pre-BLASTX filtering: 901 [S:73.1%; D:19.0%]; post-BLASTx filtering: 834 [S:69.5%; D:15.7%]	Pre-BLASTX filtering: 919 [S:46.4%; D:47.5%]; post-BLASTx filtering: 875 [S:43.6%; D:45.9%]	i. 880; ii. 899; iii. 881
Fragmented BUSCOs	43	55	Pre-BLASTX filtering: 41; post-BLASTx filtering: 74	Pre-BLASTX filtering: 52; post-BLASTx filtering: 90	i. 36; ii. 30; iii. 31
Missing BUSCOs	36	7	Pre-BLASTX filtering: 36; post-BLASTx filtering: 70	Pre-BLASTX filtering: 9; post-BLASTx filtering: 13	i. 62; ii. 49; iii. 66



a)



b)



c)

Fig. 1. Top 10 gene ontology (GO) assignments associated with: a. molecular function. b. Biological processes. c. Cellular component.

was used to assess completeness. Out of a total of 978 BUSCO groups searched, the *A. lamarcki* transcriptome contained 899 (92%) and 43 (4.4%) complete and fragmented BUSCOs, respectively (Table 2).

2.4. Functional annotation

Open reading frame (ORF) identification and extraction was performed using both EvidentialGene and TransDecoder v5.2.0 (http://transdecoder.github.io) with homology insight provided by UniProtKb/Swiss-Prot and the PFAM database (2018-04; Finn et al., 2016). ORF prediction analysis identified a total of 38,517 putative ORFs, in which 16,283 were complete ORFs while 22,243 were incomplete with the 5′, 3′, or internal segment present. Following the Trinotate pipeline (http://trinotate.github.io), the SwissProt and PFAM databases were utilized to determine protein function and gene ontology of identified ORFs. Homologies were obtained with BLAST v2.7.1 and protein domains were identified with the program HMMER v3.1 (Finn et al., 2011). The Signalp Server v4.1 (Petersen et al., 2011) was then used to predict signal peptides followed by transmembrane region prediction using the TMHMM Server V2.0 (Krogh et al., 2001) with a rRNA transcript identification step using RNAmmer v1.2 (Lagesen et al., 2007). Of the 38,517 ORFs identified, 12,107 of them were determined to be associated with molecular processes, 1266 of them were associated with biological processes, and 416 of them were associated with cellular components (Fig. 1).

2.5. Taxonomic analysis of contamination

Taxonomic information of reads flagged and removed as bacterial contamination from read libraries was retrieved from the marine bacteria database (MarDB V2). Of the 711,081 reads flagged as bacterial contamination, 26% (190,116) matched a plasmid scaffold of *Candidatus Entotheonella* sp. TSY1, a known endosymbiont for the marine sponge *Theonella swinhoei* (yellow chemotype; Wilson et al., 2014). Roughly 5% of contaminant reads matched an alpha-proteobacteria *Tistrella mobilis* strain, which is an aerobe isolated from Indian Ocean seawater (tinyurl.com/TmobilisMCCC). Another 5% of contaminant reads mapped to a heavy-metal arsenic tolerant strain of *Thalassospira xiamenensis* that is associated with deep-sea sediments in the Indian Ocean (Li et al., 2017). The taxonomic analysis of contamination in this study has resulted in previously unrecorded bacterial associations with a coral and elucidates the potential of using biological sequence contamination to obtain interesting insights into coral holobionts.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.margen.2018.08.003.

Acknowledgments

Funding for RNA-seq was provided through Seed Money of Puerto Rico Sea Grant awarded to L Rivera Garcia. We thank J Garcia-Hernandez, P Sanchez for assistance with diving. Computational resources were provided by Puerto Rico Sea Grant Project # R-101-1-14 awarded to NVS and Monica Alfaro. The authors thank the National Center for Genome Analysis Support (NCGAS) at Indiana University for access to their Carbonate Cluster used for the de novo transcriptome assembly. All opinions, findings, conclusions or recommendations detailed in this article are those of the authors and do not reflect the views of NCGAS or Indiana University.

Data accessibility

Raw reads can be accessed through the NBCI Sequence Read Archive (SRP133494). This Transcriptome Shotgun Assembly project has been deposited at DDBJ/ENA/GenBank under the accession GGLC00000000. The transcriptome version described in this paper is the first version, GGLC03000000. Both data sources are linked to the NCBI BioSample and BioProject numbers SAMN08606487 and PRJNA435937, respectively.

References

Aerts, L.A.M., 1998. Sponge/coral interactions in Caribbean reefs: analysis of overgrowth patterns in relation to species identity and cover. Mar. Ecol. 175, 241–249.
Altschil, S.F., Gish, W., Miller, W., Meyers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
Andrews, S., 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online at. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
Appeldoorn, R., Ballantine, D., Bejarano, I., Carlo, M., Nemeth, M., Otero, E., Pagan, F., Ruiz, H., Schizas, N., Sherman, C., Weil, E., 2016. Mesophotic coral ecosystems under anthropogenic stress: a case study at Ponce, Puerto Rico. Coral Reefs 35, 63–75.
Appeldoorn, R., Alfaro, M., Ballantine, D.L., Bejarano, I., Ruiz, H.J., Schizas, N.V., Schmidt, W.E., Sherman, C.E., Weil, E., 2019. Puerto Rico. In: Loya, Y., Puglise, K.A., Bridge, T.C.L. (Eds.), Mesophotic Coral Ecosystems. Springer, New York.
Aranda, M., Li, Y., Liew, Y.J., Baumgarten, S., Simakov, O., Wilson, M.C., Piel, J., Ashoor, H., Bougouffa, S., Bajic, V.B., Ryu, T., Ravasi, T., Bayer, T., Micklem, G., Kim, H., Bhak, J., Lajeunesse, T.C., Voolstra, C.R., 2016. Genomes of coral dinoflagellate symbionts highlight evolutionary adaptions conducive to a symbiotic lifestyle. Sci. Rep. 6, 39734.
Aronson, R., Bruckner, A., Moore, J., Precht, B., Weil, E., 2008. Agaricia lamarcki. The IUCN Red List of Threatened Species 2008. (e.T132970A3515504).
Bhattacharya, D., Agrawai, S., Aranda, M., Baumgarten, S., Belcaid, M., Drake, J.L., Erwin, D., Foret, S., Gates, R.D., Gruber, D.F., Kamel, B., Lesser, M.P., Levy, O., Liew, Y.J., MacManes, M., Mass, T., Medina, M., Mehr, S., Meyer, E., Price, D.C., Putnam, H.M., Qiu, H., Shinzato, C., Shogucki, E., Stokes, A.J., Tambutte, S., Tchernov, D., Voolstra, C.R., Wagner, N., Walker, C.W., Weber, A.P.M., Weis, V., Zelzion, E., Zoccola, D., Falkowski, P.G., 2016. Comparative genomics explains the evolutionary success of reef-forming corals. eLife 5, e13288.
Bongaerts, P., Frade, P.R., Ogier, J.J., Hay, K.B., van Bleijswijk, J., Englebert, N., Vermeij, M.J., Bak, R.P., Visser, P.M., Hoegh-Guldberg, O., 2013. Sharing the slope: depth partitioning of agariciid corals and associated *Symbiodinium* across shallow and mesophotic habitats (2-60 m) on a Caribbean reef. BMC Evol. Biol. 13, 205.
Bongaerts, P., Frade, P.R., Hay, K.B., Englebert, N., Latijnhouwers, R.W., Bak, R.P.M., Vermij, M.J.A., Hoegh-Guldberg, O., 2015. Deep down on a Caribbean reef: lower mesophotic depths harbor a specialized coral-endosymbiont community. Sci. Rep. 5, 7652.
Chomczynski, C., Mackey, K., 1995. Modification of the TRI reagent procedure for isolation of RNA from polysaccharide- and proteoglycan-rich sources. BioTechniques 6, 942–945.
Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39, W29–W37 Web Server Issue.
Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., Bateman, A., 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44, D279–D285 Database Issue.
Garcia-Hernandez, J.E., van Moorsel, G.W.N.M., Hoeksema, B.W., 2016. Lettuce corals overgrowing tube sponges at St. Eustatius, Dutch Caribbean. Mar. Biodiverse. https://doi.org/10.1007/s12526-016-0467-4.
Gilbert, D., 2012. Perfect Arthropod Genes Constructed With Gigabases of RNA6th Annual Arthropod Genomics Symposium. Kansas State U. https://doi.org/10.7490/f1000research.1112595.1.
Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Fd, Palma, Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference. Nat. Biotechnol. 29, 644–652.
Hammerman, N.M., Rivera-Vicens, R.E., Galaska, M.P., Weil, E., Appeldoorn, R.S., Alfaro, M., Schizas, N.V., 2018. Population connectivity of the plating coral *Agaricia lamarcki* from southwest Puerto Rico. Coral Reefs 37, 183–191.
Hoeksema, B.W., Bongaerts, P., Baldwin, C.C., 2017. High coral cover at lower mesophotic depths: a dense *Agaricia* community at the leeward side of Curaçao, Dutch Caribbean. Mar. Biodivers. 47, 67–70.
Honaas, L.A., Wafula, E.K., Wickett, N.J., Der, J.P., Zhang, Y., et al., 2016. Selecting superior de novo transcriptome assemblies: lessons learned by leveraging the best plant genome. PLoS One 11 (1), e0146062. https://doi.org/10.1371/journal.pone.0146062.
Jackson, J.B.C., Buss, L., 1975. Allelopathy and spatial competition among coral reef invertebrates. PNAS 72, 5160–5163.
Kenkel, C.D., Bay, L.K., 2017. Novel transcriptome resources for three scleractinian coral species from the Indo-Pacific. Gigascience 6, 1–4.
Kenkel, C.D., Meyer, E., Matz, M.V., 2013. Gene expression under chronic heat stress in populations of the mustard hill coral (*Porites astreoides*) from different thermal environments. Mol. Ecol. 22, 4322–4334.
Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305 (3), 567–580.
Lagesen, K., Hallin, P.F., Rodland, E., Staerfeldt, H.H., Rognes, T., Ussery, D.W., 2007. RNAammer: consistent annotation of rRNA genes in genomic sequences. Nucleic Acids Res. 35, 3100–3108.
Laverick, J.H., Rogers, A.D., 2018. Experimental evidence for reduced mortality of

Agaricia lamarcki on a mesophotic reef. Mar. Environ. Res. 134, 37–43.
Li, M., Yang, S., Lai, Q., Shao, Z., 2017. Draft genome sequence of *Thalassospira xiamenensis* strain MCCC 1A03042. Genome Announc. 5 (e01702-16).
Lin, S., Cheng, S., Song, B., Zhong, X., Lin, X., Li, W., Li, L., Zhang, Y., Zhang, H., et al., 2015. The *Symbiodinium kawagutii* genome illuminates dinoflagellate gene expression and coral symbiosis. Science 350, 691–694.
Lucas, M.Q., Stat, M., Smith, M.C., Weil, E., Schizas, N.V., 2015. *Symbiodinium* (internal transcribed spacer 2) diversity in the coral host *Agaricia lamarcki* (Cnidaria: Scleractinia) between shallow and mesophotic reefs in the Northern Caribbean (20–70 m). Mar. Ecol. 37 (5), 1079–1087.
Mansour, T.A., Rosenthal, J.J.C., Brown, C.T., Roberson, L.M., 2016. Transcriptome of the Caribbean stony coral *Porites astreoides* from three developmental stages. Gigascience 5, 33.
Martin, J.A., Wang, Z., 2011. Next-generation transcriptome assembly. Nat. Rev. Genet. 12 (10), 671.
Mikheenko, A., Valin, G., Prjibelski, A., Saveliev, V., Gurevich, A., 2016. Icarus: visualizer for de novo assembly evaluation. Bioinformatics 32 (21), 3321–3323. https://doi.org/10.1093/bioinformatics/btw379.
Ortiz-González, I.C., Rivera-Vicéns, R.E., Schizas, N.V., 2017. *De novo* transcriptome assembly of the hydrocoral *Millepora alcicornis* (branching fire coral) from the Caribbean. Mar. Genomics 32, 27–30.
Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786.
Robertson, G., Schein, J., Chiu, R., et al., 2010. De novo assembly and analysis of RNA-seq data. Nat. Methods 7, 909–912.
Schubert, M., Lindgreen, S., Orlando, L., 2016. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes 9, 88.
Schulz, M.H., Zerbino, D.R., Vingron, M., Birney, E., 2012. Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics. https://doi.org/10.1093/bioinformatics/bts094.
Shinzato, C., Inoue, M., Kusakabe, M., 2014. A snapshot of a coral "Holobiont": a transcriptome assembly of the Scleractinian coral, *Porites*, captures a wide variety of genes from both the host and symbiotic zooxanthellae. PLoS One 9, e85182.
Shoguchi, E., Shinzato, C., Kawashima, T., Gyoja, F., Mungpakdee, S., Koyanagi, R., Takeuchi, T., Hisata, K., Tanaka, M., Fujiwara, M., Hamada, M., Seidi, A., Fujie, M., Usami, T., Goto, H., Yamasaki, S., Arakaki, N., Suzuki, Y., Sugano, S., Toyoda, A., Kuroki, Y., Fujiyama, A., Medina, M., Coffroth, M.A., Bhattacharya, D., Satoh, N., 2013. Draft assembly of the Symbiodinium minutum nuclear genome reveals dinoflagellate gene structure. Curr. Biol. 23, 1399–1408.
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. Bioinformatics. https://doi.org/10.1093/bioinformatics/btv351.
Veglia, A.J., Hammerman, N.H., Rivera Rosaly, C.R., Lucas, M.Q., Galindo Estronza, A., Corgosinho, P.H., Schizas, N.V., 2018. Characterizing population structure of coral-associated fauna from mesophotic and shallow habitats in the Caribbean. J. Mar. Biol. Assoc. UK 1–11. https://doi.org/10.1017/S0025315418000413.
Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., Zdobnov, E.M., 2017. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol. Biol. Evol. https://doi.org/10.1093/molbev/msx319.
Wilson, M.C., Mori, T., Rückert, C., Uria, A.R., Helf, M.J., Takada, K., Gernert, C., Steffens, U.A.E., Heycke, N., Schmitt, S., Rinke, C., Helfrich, E.J.N., Brachmann, A.O., Gurgui, C., Wakimoto, T., Kracht, M., Crüsemann, M., Hentschel, U., Abe, I., Matsunaga, S., Kalinowski, J., Takeyama, H., Piel, J., 2014. An environmental bacterial taxon with a large and distinct metabolic repertoire. Nature 506 (7486), 58.
WoRMS Editorial Board, 2018. World Register of Marine Species. Available from. http://www.marinespecies.org, Accessed date: 12 April 2018.
Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., Huang, W., He, G., Gu, S., Li, S., Zhou, X., Lam, T., Li, Y., Xu, X., Wong, G.K., Wang, J., 2014. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics 30 (12), 1660–1666.
Zerbino, D.R., Birney, E., 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18 (5), 821–829. https://doi.org/10.1101/gr.074492.107.