Marine Genomics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Marine Genomics

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

De novo assembly and annotation of the *Acropora gemmifera* transcriptome

Matthew J. Oldach, Peter D. Vize*

Department of Biological Science, University of Calgary, 2500 University Drive NW, Calgary, Alberta T2N1N4, Canada

A B S T R A C T

Stony corals from the genus *Acropora* are widely distributed, important reef-builders and have become increasingly utilized for investigating links between genetics and spawning behaviour. We assembled and annotated a composite transcriptome from *Acropora gemmifera* using Illumina HiSeq2500 analysis of two libraries from different lunar and solar phases to identify genes that have potential functional roles in reproductive-related traits. A total of 31.6 million combined raw reads were assembled using Trinity and built into 104,000 contigs. Functional gene annotation was performed using dammit, Gene Ontology (GO), KOG (WebMGA) and KEGG pathway analyses (Kaas). This resource will be valuable for researchers studying gene expression patterns in coral reproductive cycles and evolution of the genus *Acropora*.

1. Introduction

Reef-building corals provide habitat for numerous species, support fisheries, promote tourism and protect infrastructure along coastlines. However, the future of coral reefs is precarious, having lost at least 50% coral cover over the past 40–50 years, a rate of change on coral reefs not seen for tens of million years (Hoegh-Guldberg, 2014). Understanding the coral reproductive process is therefore critical for their management and conservation. A library of gene transcripts from key stages of the lunar cycle preceding broadcast spawning can provide valuable information about which genes are important for processes that are turned on at a particular stage. To our knowledge no previous assemblies have undertaken this aim.

Broadcast spawning is the most prevalent mode of sexual reproduction in reef building corals and most species have a single annual gametogenesis cycle (Harrison and Wallace, 1990). Timing of reproduction is extremely accurate as most species only spawn on one or two nights per year (Babcock et al., 1986) and this is predictable to within a half-hour (Vize et al., 2005; Levitan et al., 2011). Despite the simple anatomy of corals –lack of eyes and central nervous system- their spawning behaviour is intimately linked to environmental light. While seasonal cycles of local weather and temperature patterns set the month of spawning (Penland et al., 2004; van Woesik, 2010; Keith et al., 2016), later stages of the cycle are controlled by light. The date of spawning is set by the lunar cycle (Harrison et al., 1984; Babcock et al., 1986; Oliver et al., 1988) and the hour at which they spawn is set by the daily solar cycle (Levitan et al., 2004; Brady et al., 2009).

There is a growing body of research on the molecular mechanisms underlying coral reproductive cycles, specifically in *Acropora* species as a model (Kaniewska et al., 2015; Brady et al., 2016; Oldach et al., 2017; Rosenberg et al., 2017). Numerous biological processes are regulated by entrained biological clocks (reviewed by Dunlap, 1999). Biological clock gene orthologs are conserved in cnidarians and recent evidence suggests that some of these show different cycles of transcription at different phases of the moon (Levy et al., 2007; Hoadley et al., 2011; Kaniewska et al., 2015; Brady et al., 2016; Oldach et al., 2017; Rosenberg et al., 2017). Corals from the genus *Acropora* (Scleractinia, Acroporidae) are widely distributed around the globe and extremely speciose (approximately 180 spp. according to Vernon, 2000). As a model, we used the species *Acropora gemmifera* (Brook, 1892), a common and rapidly growing species of stony coral found throughout the Pacific, Red Sea and the Gulf of Aden. This species is classified as least concern species on the IUCN Red List and Appendix II of CITES. We assembled and characterized a composite transcriptome for two different time periods of the lunar phase, in the month preceding spawning, to maximize the diversity and completeness of the *de novo* assembled transcripts. This composite transcriptome composed from key stages in the lunar cycle should provide valuable information about which genes are important for processes involved in coral broadcast spawning.

2. Data description

2.1. Study site and subjects (sample collection and maintenance)

An *A. gemmifera* colony (measuring 20–25 cm in diameter) was

* Corresponding author.
E-mail address: pvize@ucalgary.ca (P.D. Vize).

https://doi.org/10.1016/j.margen.2017.12.007
Received 24 November 2017; Received in revised form 20 December 2017; Accepted 21 December 2017

Table 1
MlxS descriptors of the study.

Item	Description
Submitted_to_insdc	Yes (SRA)
Investigation_type	Eukaryote
Project_name	*Acropora gemmifera* transcriptome
Geo_loc_name	Houbihu Harbor, Taiwan
Lat_lon	21.95032°N/120.754606°E
Depth	2 m
Alt_elev	0 m
Collection_date	2014-04-15
Collected_by	Peter Vize
Env_biome	Sea water (ENVO:00002149)
Env_feature	Bay (ENVO:00000032)
Env_material	Sea water (ENVO:00002149)
Env_package	Water
Temp	NA
Salinity	NA
Sequencing method	Illumina HiSeq
Assembly method	Trinity v2.4.0

collected from the southwest coast of Taiwan (Houbihu Harbor, 21.95032°N, 120.754606°E) April 18, 2014 under local permit and a CITES export permit (Permit Number: FTS503W0040611). The colony was housed in an indoor raceway tank and was exposed to natural outdoor light. Lunar irradiance, equivalent to that at 5 m depth, was simulated to mimic moonrise and moonset using a Mr. Aqua 15-Watt bulb (Lamp model number EA063) according to the US Navy Observatory website (http://aa.usno.navy.mil/data/) using the Complete Sun and Moon Data Available for One Day option.

2.2. RNA isolation, library preparation and Illumina sequencing

Samples were collected at two different lunar phases and time of day (midnight new moon and noon first quarter moon) from the same individual colony to maximize the diversity and completeness of clock regulated genes in the *de novo* assembled transcripts. MlsX data on collected material is available in Table 1. Total RNA was extracted using the TRIZOL RNA Isolation Method (Chomczynski and Mackey, 1995) following the manufacturer's instructions for proteoglycan and polysaccharide rich samples. Nucleic acid quality and quantity were assessed using both a NanoDrop 1000 spectrophotometer (Thermo Scientific) and RNA Tape station (Agilent) assay at the University of Calgary Core DNA Services facility. Samples were sent to a commercial service provider (BGI, Hong Kong, China) where library preparation was performed using Poly-A mRNA capture methods. Sequencing was performed on an Illumina HiSeq2500 platform with 250 bp paired-end reads. After sequencing, adapter sequences were removed and raw reads were filtered for contamination and low-quality reads (performed by BGI).

2.3. Transcriptome assembly

We obtained 15,769,950 and 15,857,844 raw paired end-reads for each library respectively, yielding a total of 31,627,794 reads. Transcriptome assembly and annotation was conducted using Amazon Web Services Elastic Cloud Compute service (AWS EC2). The *de novo* transcriptome assembler Trinity v2.4.0 (Grabherr et al., 2011) was used to assemble reads. Transcriptome assembly quality and completeness was assessed using Transrate v1.0.3 (Smith-Unna et al., 2016). Transrate gives a metric of *de novo* transcriptome assembly quality;

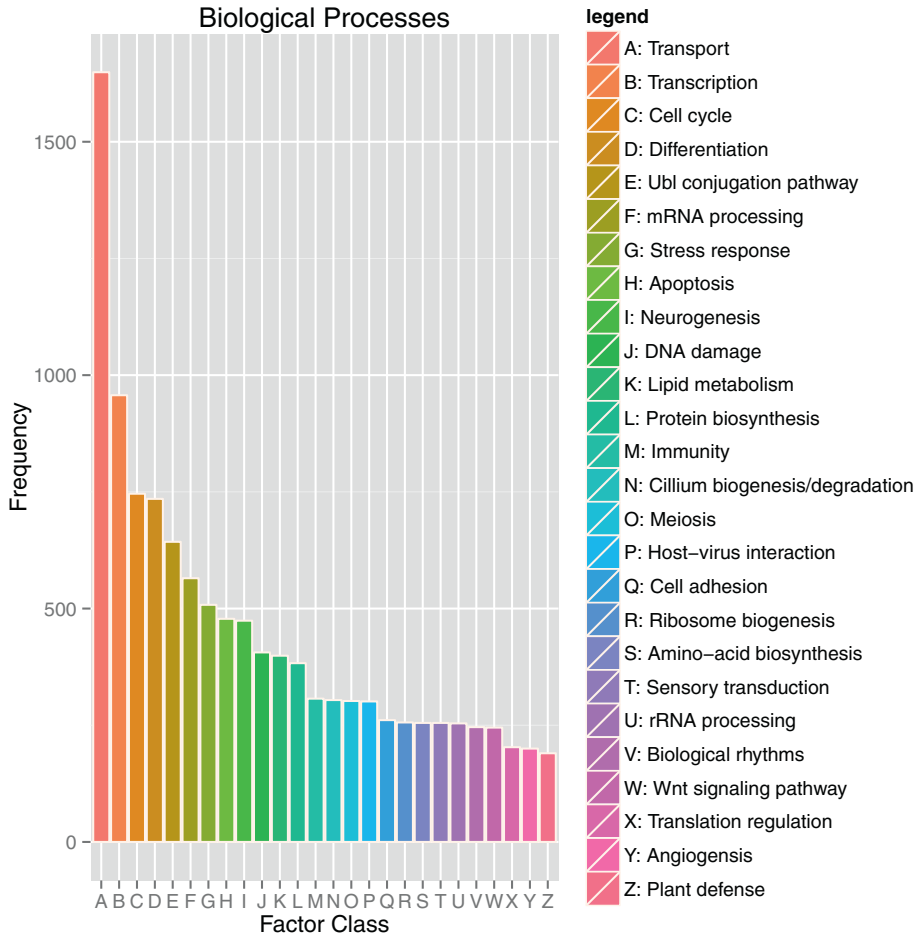

Fig. 1. GO biological processes represented in the *A.gemmifera* transcriptome.