Marine Genomics xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Marine Genomics

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

Short communication

A transcriptome resource for the Antarctic pteropod *Limacina helicina antarctica*

Kevin M. Johnson *, Gretchen E. Hofmann

Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, Santa Barbara, CA 93106-9620, USA

ARTICLE INFO

Article history:
Received 20 January 2016
Received in revised form 6 April 2016
Accepted 6 April 2016
Available online xxxx

Keywords:
Southern Ocean
RNA sequencing
De novo assembly
Ocean acidification
Ocean warming

ABSTRACT

The pteropod *Limacina helicina antarctica* is a dominant member of the zooplankton assemblage in the Antarctic marine ecosystem, and is part of a relatively simple food web in nearshore marine Antarctic waters. As a shelled pteropod, *Limacina* has been suggested as a candidate sentinel organism for the impacts of ocean acidification, due to the potential for shell dissolution in undersaturated waters. In this study, our goal was to develop a transcriptomic resource for *Limacina* that would support mechanistic studies to explore the physiological response of *Limacina* to abiotic stressors such as ocean acidification and ocean warming. To this end, RNA sequencing libraries were prepared from *Limacina* that had been exposed to a range of pH levels and an elevated temperature to maximize the diversity of expressed genes. RNA sequencing (RNA-seq) was conducted on an Illumina NextSeq500 which produced 339,000,000 150 bp paired-end reads. The de novo transcriptome was produced using Trinity and annotation of the assembled transcriptome resulted in the identification of 81,229 transcripts in 137 KEGG pathways. This RNA-seq effort resulted in a transcriptome for the Antarctic pteropod, *Limacina helicina antarctica*, that is a major resource for an international marine science research community studying these pelagic molluscs in a global change context.

1. Introduction

The shell-forming (thecosome) pteropods have been identified as an ecologically important epipelagic zooplankton species and are predicted to be vulnerable to future ocean acidification (Bednaršek et al., 2014a, 2012a; Busch et al., 2014; Comeau et al., 2009). Ocean acidification, the dissolution of CO_2 into surface oceans that is accompanied by a decline in pH, drives changes in carbonate chemistry that alters the saturation state of seawater with respect to forming calcium carbonate minerals such as aragonite and calcite. Both of these mineral forms are common constituents of the shells and skeletons of marine invertebrates. Ocean acidification may present a physiological challenge to biogenic calcification, and many species, such as the shell-forming pteropods, are proposed to be negatively impacted by this shift in ocean chemistry (Comeau et al., 2009; Kroeker et al., 2010; Maas et al., 2015; Byrne, 2011; Lischka et al., 2011; Fabry et al., 2008). Recently, pteropods in general have been proposed as an indicator species for the progression and impacts of ocean acidification in the world oceans; however, there are possible difficulties with attributing visible shell damage with true physiological stress (Peck et al., 2015; Bednaršek et al., 2016; Peck et al., 2016). Since pteropods are ecologically significant in many regions, more information regarding the physiological response to future ocean conditions will benefit an international research community seeking to use *Limacina* as an indicator species in both polar and temperate marine sciences (Bednaršek et al., 2012a, 2012b, 2012c; Comeau et al., 2009; Fabry et al., 2009).

Given this research focus, tools to assess the physiological tolerances and population differences in pteropods are needed. However, genomic resources for pteropods are scarce and additional transcriptomic data would support the study of the Antarctic form and further, would facilitate the comparison of *Limacina* populations and subspecies collected in other oceanic regions (Maas et al., 2015; Koh et al., 2015). Here, using high-throughput strand-specific RNA sequencing (RNA-seq) and the de novo assembler, Trinity, we present the first de novo transcriptome of the Antarctic pteropod *Limacina helicina antarctica*. The overall goals for this sequencing project were: (1) to sequence the transcriptome of the Antarctic pteropod *L. helicina antarctica*, and (2) to develop a resource to support the exploration of gene expression changes in response to anthropogenic environmental stressors such as ocean acidification and ocean warming.

In order to capture a transcriptome that reflects different physiological conditions, the *Limacina* transcriptome was prepared using RNA sourced from animals from three different treatment groups: (1) wild-caught individuals exposed only to ambient environmental conditions (field acclimatized), (2) animals that were briefly heat stressed in the laboratory, and (3) animals that were exposed to variable levels of

* Corresponding author.
E-mail addresses: kevin.johnson@lifesci.ucsb.edu (K.M. Johnson), gretchen.hofmann@lifesci.ucsb.edu (G.E. Hofmann).

http://dx.doi.org/10.1016/j.margen.2016.04.002
1874-7787/

pCO_2 that mimic both current and future pH ranges for McMurdo Sound (Kapsenberg et al., 2015).

2. Data description

2.1. Methods

For this study, *L. helicina antarctica* were collected via 50 m vertical plankton tows at a near-shore site (near Cape Royds, McMurdo Sound, Antarctica) on November 18th, 2014 (77.5649°S, 166.18712°E) using a small collapsible net with 100 μm mesh. To capture the field-acclimatized transcriptome profile, 10 individuals were preserved in RNAlater® immediately after collection in the field. Additional *Limacina* collected from the same tows were transported back to the laboratory facilities at McMurdo Station. All individuals were subsequently placed into a 3 L flow-through holding tank in seawater at −0.7 °C, and monitored for 48 h to allow for clearing of the gut prior to use in exposure trials. Pteropods were collected and transported under permits and permissions associated with the U.S. Antarctic Program.

Following the holding period, actively swimming individuals were selected, and either heat stressed to induce genes associated with the defensome, or were exposed to elevated pCO_2. For the heat stress treatment, animals contained in 20 ml scintillation vials were placed into an +8 °C water bath for 5 h after which they were quickly de-shelled in a temperature controlled room (−1 °C), and flash frozen in liquid nitrogen in groups of 10 per cryovial. The three-pCO_2 treatments were chosen to represent 4 unique scenarios that ranged from present-day pH conditions to pH conditions that mimic future, projected acidification (Supplementary file 1). Following 24 h of exposure to each pCO_2 treatment, *Limacina* individuals were preserved in the same manner as the heat stressed individuals. All individuals were subsequently stored at −80 °C in cryovials prior to RNA extraction.

RNA was extracted from the 10 pooled pteropods from each of the 5 groups (Field-caught, +8 °C heat stress, low pCO_2, mid pCO_2 and high pCO_2) using 500 μl of Trizol® reagent according to the manufacturer's instructions (Invitrogen). Five separate libraries were generated from high-quality RNA using the New England BioLabs Ultra-Directional RNA Library Prep Kit for Illumina. Concentrations of each final library was quantified using a Qubit® 3.0 fluorometer (Life Technologies) and sequenced on an Illumina NextSeq500 sequencer using the mid-output mode with 300 cycles and 150 base-pair (bp) paired-end reads. Outputs for each lane were treated independently throughout the assembly pipeline.

The sequence reads for *L. helicina antarctica* were trimmed using Trimmomatic (version 0.33) with a 4-base sliding window to remove base pairs with quality scores below 30, and to remove reads that were shorter than 36 bases in length (Bolger et al., 2014). Following quality trimming, sequence quality was checked using FastQC (version 0.11.3). The de novo assembly was created using the de novo assembler Trinity (version 2.0.6) (Haas et al., 2013). Transcriptome annotation was completed using Blast + software (version 2.2.31) to execute a BLAST-X search against the NCBI non-redundant protein database (E-value ≤ 1e−5). InterPROscan and Gene Ontology terms were mapped to these annotations using the Blast2GO program.

2.2. Results and discussion

Transcriptome sequencing of the 5 libraries yielded 339,300,000 150-bp paired-end reads, with an average of 18.8 (±4.9) million paired-end reads for each library (Table 1). Prior to de novo assembly, reads were trimmed to omit transcripts with low quality Phred scores (Phred < 20) (Bolger et al., 2014). In order to assemble a de novo transcriptome with no reference genome, we pooled all reads (approx. 189 million) before running the de novo assembler, Trinity (Haas et al., 2013). The transcriptome generated by Trinity was comprised of 402,273 transcripts with an N50 of 500 bp and GC percent of 36.69% (Table 1). Transcripts ranged from 200 bp to 7941 bp with 32,015 transcripts exceeding 1000 bp in length.

Table 1
Transcriptome assembly statistics.

Statistic	Total
Total number of reads	339300000 Mb
Reads post-trimming	189628950 Mb
Total number of contigs	402,273
Contig N50	500 bp
Average contig length	471 bp
Median contig length	321 bp
Max contig length	7941 bp
Min contig length	200 bp
Percent GC	36.69%
Percent annotated contigs	20.2%

A total of 81,229 of the 402,273 transcripts generated by Trinity (20.2%) were annotated using a combination of BLAST-X, InterProScan and Blast2GO (Supplementary file 2). A BLAST-X homology search against the NCBI (Genbank) non-redundant protein database resulted in annotation of 21,276 transcripts. Next, gene ontologies (GO) were assigned using Blast2GO, and a total of 32,813 transcripts were successfully mapped to existing gene categories. In addition, incorporating GO-slim terms from Blast2GO added GO terms to an additional 27,146 transcripts. Finally, InterProScan was run through Blast2GO, and this analysis identified 87,001 conserved protein domains that were merged with the mapped gene ontologies. Comparison of top level gene ontology terms identified 102,323 transcripts in the top 10 gene ontology terms for "Molecular Function", 50,663 transcripts identified in the top 10 gene ontology terms for "Biological Process", and 56,842 transcripts identified in the top 10 gene ontology terms for "Cellular Component".

A taxonomic analysis of the annotated sequence completed using Megan5 showed the dominant source (62%) of annotation matches were from the sequenced genome of another gastropod mollusk, the California sea hare, *Aplysia californica* (Huson et al., 2011) (Supplementary Fig. 3). Analysis with Megan5 also identified 0.3% of the annotated transcripts with at least one BLAST-X hit that could be considered contaminants, and removed from the assembly; the approximate diversity for these transcripts was bacterial (0.14%), algal (0.09%) and fungal (0.07%).

Table 2
Top 10 KEGG pathways in the *Limacina helicina antarctica* transcriptome.

KEGG pathway	Pathway ID	Transcripts in pathway	Number of enzymes
Biosynthesis of antibiotics	01130	966	135
Purine metabolism	00230	3166	54
Glycine, serine and threonine metabolism	00260	155	35
Pyrimidine metabolism	00240	393	34
Amino sugar and nucleotide sugar metabolism	00520	176	30
Cysteine and methionine metabolism	00270	179	28
Alanine, aspartate and glutamate metabolism	00250	201	27
Valine, leucine and isoleucine degradation	00280	188	23
Arginine and proline metabolism	00330	144	23
Glycolysis/gluconeogenesis	00010	315	22
Pyruvate metabolism	00620	239	21
Aminoacyl-tRNA biosynthesis	00970	202	21
Starch and sucrose metabolism	00500	133	21
Phosphatidylinositol signaling system	04070	105	21
Inositol phosphate metabolism	00562	95	21
Carbon fixation pathways in prokaryotes	00720	200	20
Fructose and mannose metabolism	00051	147	20
Glycerophospholipid metabolism	00564	105	20

Please cite this article as: Johnson, K.M., Hofmann, G.E., A transcriptome resource for the Antarctic pteropod *Limacina helicina antarctica*, Mar. Genomics (2016), http://dx.doi.org/10.1016/j.margen.2016.04.002