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

Kevin M. Johnson[§] and Gretchen E. Hofmann

Kevin M. Johnson (corresponding author)

Department of Ecology, Evolution and Marine Biology, University of California, Santa

Barbara, Santa Barbara CA 93106-9620 USA

kevin.johnson@lifesci.ucsb.edu

Gretchen E. Hofmann

Department of Ecology, Evolution and Marine Biology, University of California, Santa

Barbara, Santa Barbara CA 93106-9620 USA

gretchen.hofmann@lifesci.ucsb.edu

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 stress 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 million 150 base pair, 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.

Keywords: Southern Ocean, RNA-sequencing, de novo assembly, Ocean acidification, Ocean warming

1. Introduction

The shell-forming (thecosome) pteropods have been identified as an ecologically important epipelagic zooplankton and are predicted to be vulnerable to future ocean acidification [1-4]. Ocean acidification, the dissolution of CO₂ 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 [4-9]. Recently, pteropods in general have been proposed as an indicator species for the progression and impacts of ocean acidification in the world's oceans; however, there are possible difficulties with attributing visible shell damage with true physiological stress [10-12]. 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 [2, 4, 13-15].

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 [6, 16]. 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 *Limacina 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 pCO₂ that mimic both current and future pH ranges for McMurdo Sound [17].

2. Data Description

2.1 Methods

For this study, *Limacina 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 100µm mesh using field methods as described in [18]. 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 hours 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 pCO₂. For the heat stress treatment, animals contained in 20ml scintillation vials were placed into an +8°C water bath for 5 hours 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-pCO₂ 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 hours of exposure to each pCO₂ 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 pCO₂, mid pCO₂ and high pCO₂) using 500 ul 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 *Limacina 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 [19]. 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) [20]. 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 $\leq 1e-5$). 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 million 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) [19]. 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 [21]. 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 200bp to 7,941 bp with 32,015 transcripts exceeding 1000 bp in length.

A total of 81,229 of the 402,273 transcripts generated by Trinity (approx. 20%) 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[22] showed the dominant source (62%) of annotation matches were from the sequenced genome of another gastropod mollusc, the California sea hare, *Aplysia californica*[22] (Supplementary figure 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%).

Overall, the KEGG analysis identified 16,664 transcripts, which represented 1,320 enzymes in 137 unique pathways. This analysis provided 66 pathways with at least 10 associated enzymes. The 18 most represented KEGG pathways, with between 20 - 135 associated enzymes, included enzymes involved in gluconeogenesis, amino acid metabolism, nitrogenous base metabolism, and antibiotic biosynthesis pathways (Table 2). The comprehensive list of KEGG pathways identified and associated transcripts is available in supplementary file 4.

A central goal of the project was to develop a resource for the marine molecular

ecology research community by identifying genes and pathways that would be of use in studying the response of *Limacina* to environmental stress related to anthropogenic change. To develop this resource, we hand sorted these gene families into 21 groups that represent processes that could be affected, or groups of genes, that might be activated by stressors such as ocean acidification or ocean warming (Table 3). Briefly, these groups contain 476 transcripts greater than 1,000bp that are involved in calcification and calcium transport, metabolism, cellular integrity, and genes in the defenseome. Notably, these large transcripts provide an excellent resource for studying the interaction of multiple stressors in an ocean acidification context. A comprehensive list of these genes with transcript identifiers and enzyme codes is available in supplementary file 5. Additional environmental and sequence information is provided in supplementary file 6 following the MIxS guidelines.

2.3 Conclusion

The transcriptome presented here provides a key resource for future investigations into the physiological and genomic responses of the Antarctic pteropods *Limacina helicina antarctica* to future ocean change. This resource will ultimately support investigations into the physiological plasticity of contemporary populations, and how and whether pteropods can adjust to ocean acidification and ocean warming. Finally, this resource will be useful for populations of *Limacina* from other oceans, including the Arctic [4] and temperate regions such as the northeastern Pacific where pteropods are routinely found in episodically upwelled water [23].

2.4 Data Deposition

178 The *Limacina helicina antarctica* transcriptome can be found in the SRA database under
179 BioProject PRJNA295792, with accession number SRR2463656. The annotated
180 transcriptome has been deposited to the DDBJ/EMBL/GenBank transcriptome shotgun
181 assembly (TSA) database under the accession GDRM000000000. The version described in
182 this paper is the first version, GDRM01000000.

183 **Acknowledgements**

184 The authors would like to thank members of the U.S. Antarctic Program and Lockheed's
185 Antarctic Support Corporation (ASC) for support at McMurdo Station, Antarctica during the
186 2014-2015 Summer Field Season. In addition, we thank Professor Mary Sewell (University
187 of Auckland) for the loan of the collapsible plankton net. This research was supported by a
188 grant from the U.S. National Science Foundation (NSF) through the Division of Polar
189 Programs (PLR-1246202 to GEH). During the course of this project, KMJ was supported by a
190 NSF Graduate Research Fellowship. We acknowledge the use of the Biological
191 Nanostructures Laboratory within the California NanoSystems Institute, supported by UC
192 Santa Barbara, and the University of California, Office of the President. Lastly, we
193 acknowledge support from the Center for Scientific Computing at the CNSI and MRL: an
194 NSF MRSEC (DMR-1121053) and NSF CNS-0960316.

References

1. Bednaršek N, Tarling GA, Bakker DCE, Fielding S, Feely RA: **Dissolution dominating calcification process in polar pteropods close to the point of aragonite undersaturation.** *PloS one* 2014, **9**(10).
2. Bednaršek N, Tarling GA, Bakker DCE, Fielding S, Jones EM, Venables HJ, Ward P, Kuzirian A, Leze B, Feely RA *et al*: **Extensive dissolution of live pteropods in the Southern Ocean.** *Nat Geosci* 2012, **5**(12):881-885.
3. Busch DS, Maher M, Thibodeau P, McElhany P: **Shell condition and survival of Puget Sound pteropods are impaired by ocean acidification conditions.** *PloS one* 2014, **9**(8).
4. Comeau S, Gorsky G, Jeffree R, Gattuso J-P: **Impact of ocean acidification on a key Arctic pelagic mollusc (*Limacina helicina*).** *Biogeosciences* 2009, **6**(9):1877-1882.
5. Kroeker KJ, Kordas RL, Crim RN, Singh GG: **Meta-analysis reveals negative yet variable effects of ocean acidification on marine organisms.** *Ecology Letters* 2010, **13**(11):1419-1434.
6. Maas AE, Lawson GL, Tarrant AM: **Transcriptome-wide analysis of the response of the thecosome pteropod *Clio pyramidata* to short-term CO₂ exposure.** *Comparative biochemistry and physiology Part D, Genomics & proteomics* 2015, **16**:1-9.
7. Byrne M: **Impact of ocean warming and ocean acidification on marine invertebrate life history stages: vulnerabilities and potential for persistence in a changing ocean.** *Oceanography and Marine Biology: an Annual Review* 2011, **49**:1-42.

- 218 8. Lischka S, Büdenbender J, Boxhammer T, Leibniz-Riebesel U: **Impact of ocean**
219 **acidification and elevated temperatures on early juveniles of the polar shelled**
220 **pteropod *Limacina helicina*: mortality, shell degradation, and shell growth.**
221 *Biogeosciences* 2011, **8**(4):919-932.
- 222 9. Fabry VJ, Seibel BA, Feely RA, Orr JC: **Impacts of ocean acidification on marine fauna**
223 **and ecosystem processes.** *Ices Journal of Marine Science* 2008, **65**(3):414-432.
- 224 10. Peck VL, Tarling GA, Manno C, Harper E, Tynan E: **Outer organic layer and internal**
225 **repair mechanism protects pteropod *Limacina helicina* from ocean**
226 **acidification.** *Deep Sea Research II* 2015.
- 227 11. Bednaršek N, Johnson J, Feely RA: **Vulnerability of pteropod (*Limacina helicina*) to**
228 **ocean acidification: shell dissolution occurs despite an intact organic layer.**
229 *Deep Sea Research Part II: Topical Studies in Oceanography* 2016.
- 230 12. Peck VL, Tarling GA, Manno C, Harper EM: **Response to comment Vulnerability of**
231 **pteropod (*Limacina helicina*) to ocean acidification: Shell dissolution occurs**
232 **despite an intact organic layer By Bednarsek et al.** *Deep Sea Research Part II:*
233 *Topical Studies in Oceanography* 2016.
- 234 13. Fabry VJ, McClintock JB, Mathis JT, Grebmeier JM: **Ocean acidification at high**
235 **latitudes: The bellweather.** *Oceanography* 2009, **22**(4):160-171.
- 236 14. Bednaršek N, Tarling G, Fielding S, Bakker D: **Population dynamics and**
237 **biogeochemical significance of *Limacina helicina antarctica* in the Scotia Sea**
238 **(Southern Ocean).** *Deep Sea Research Part II: Topical Studies in Oceanography* 2012,
239 **59:105-116.**

15. Bednaršek N, Tarling GA, Bakker DC, Fielding S, Cohen A, Kuzirian A, McCorkle D, Lézé B, Montagna R: **Description and quantification of pteropod shell dissolution: a sensitive bioindicator of ocean acidification.** *Global Change Biology* 2012, **18**(7):2378-2388.
16. Koh HY, Lee JH, Han SJ, Park H, Shin SC, Lee SG: **A transcriptomic analysis of the response of the arctic pteropod *Limacina helicina* to carbon dioxide-driven seawater acidification** *Polar Biology* 2015(1738).
17. Kapsenberg L, Kelley AL, Shaw EC, Martz TR, Hofmann GE: **Near-shore Antarctic pH variability has implications for the design of ocean acidification experiments.** *Sci Rep* 2015, **5**:9638.
18. Johnson KM, Hofmann GE, Hoshijima U, Nguyen ATN: **Studies of *Limacina helicina antarctica* in McMurdo Sound Antarctica: Distributions, carbonate chemistry and experimental biology in an OA context.** *Polar Biology* In prep.
19. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data.** *Bioinformatics (Oxford, England)* 2014, **30**(15):2114-2120.
20. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M *et al*: **De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis.** *Nature Protocols* 2013, **8**:1494-1512.
21. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M *et al*: **De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis.** *Nat Protoc* 2013, **8**(8):1494-1512.

22. Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC: **Integrative analysis of environmental sequences using MEGAN4**. *Genome Res* 2011, **21**(9):1552-1560.
23. Bednaršek N, Feely RA, Reum JCP, Peterson B, Menkel J, Alin SR, Hales B: ***Limacina helicina* shell dissolution as an indicator of declining habitat suitability owing to ocean acidification in the California Current Ecosystem**. *Proceedings of the Royal Society of London B: Biological Sciences* 2014, **281**(1785).

286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308

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 %

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

Table 3. Genes of Interest

Acid-base regulation	<p>Sodium-potassium pump</p> <p>V-type proton ATPase</p> <p>Calcium ATPase</p> <p>ABC transport proteins</p> <p>Sodium-bicarbonate transport proteins</p> <p>Carbonic anhydrase</p>
Metabolism	<p>NADH dehydrogenase</p> <p>Cyclic adenosine monophosphate</p> <p>Sodium-glucose transport proteins</p>
Oxidative stress	<p>Catalase</p> <p>Superoxide dismutase</p> <p>Aldehyde dehydrogenase</p> <p>Cytochrome C oxidase</p> <p>Monoxygenase</p> <p>Peroxidase</p> <p>Peroxidoxin</p> <p>Thioredoxin</p>
Cellular components	<p>Calmodulin</p> <p>Molecular Chaperones</p> <p>Heat shock proteins</p> <p>Ubiquitin</p>