Dataset: Antarctic salp genome and RNAseq transcriptome from ARSV Laurence M. Gould, Umitaka-Maru, R/V Polarstern LMG1110, UM-08-09, ANT-XXVII-2 in the Southern Ocean from 2009-2011 (Salp_Antarctic project)

Project(s): Population ecology of Salpa thompsoni based on molecular indicators (Salp_Antarctic)

Abstract: A preliminary genome sequence and complete reference transcriptome have been assembled for the Southern Ocean salp, Salpa thompsoni (Urochordata, Thaliacea). The reference transcriptome contains 216,931 sequences; 41,210 (18%) were associated with predicted, hypothetical, or known proteins; 13,058 (6%) were mapped and annotated. Whole-transcriptome (RNA-seq) analysis of 39 samples collected during austral spring and summer 2011 in the WAP, and in summer 2009 in the Indian Sector revealed clustering of samples by regions, seasons, and areas (Bray-Curtis similarity). Spring versus summer samples showed significant differential expression of 77 genes associated with environmental stress response and 51 genes associated with sexual reproduction (paired t-tests, $p<0.05$). Gene Ontology (GO) term enrichment analysis identified 41 GO terms responsible for spring versus summer differences, including 156 genes associated with translation (i.e., protein synthesis). The genome sequence of 318,767,936 bp covers >50% of the estimated 602 MB (±173 MB) genome size for S. thompsoni, with >50% (16,823) of sequences showing significant homology to known proteins and ~38% (12,151) of the total protein predictions associated with Gene Ontology functional information. A total of 109,958 SNP variants and 9,782 indel predictions were generated, serving as a resource for future phylogenomic and population genomic studies. Salpa thompsoni exhibits rapid rates of evolution (>1.5 times that observed for vertebrates) typical of other urochordates examined. An initial survey of small RNAs revealed the presence of known, conserved miRNAs, as well as novel miRNA genes; unique piRNAs; and mature miRNA signatures for varying developmental stages. For a complete list of measurements, refer to the supplemental document 'Field_names.pdf', and a full dataset description is included in the supplemental file 'Dataset_description.pdf'. The most current version of this dataset is available at: http://www.bco-dmo.org/dataset/675040

Description: Antarctic salp genome and RNAseq transcriptome

This dataset reports Salpa thompsoni specimens used for genomics/transcriptomics with their GenBank accession links.

Related Dataset: Salp sample log

Acquisition Description: 1) Cruise R/V LM GOULD (LMG1110): Samples collected using Multiple Opening/Closing Net and Environmental Sensing System (MOCNESS) with a

mouth opening of 1-m2 and nine 335µm mesh nets; upper 200 m were sampled with a 2.3 m2 Isaacs-Kidd Midwater Trawl (IKMT) with a 505 µm mesh net. Western Antarctic Peninsula region, Southern Ocean (November-December 2011)

2) Cruise R/V Polarstern (PS-ANT-XVII/2): Samples collected using a Rectangular Midwater Trawl (RMT 1+8) from the upper 200 m. Western Antarctic Peninsula region, Southern Ocean (January 2011)

3) Cruise R/V Umitaka Maru (UM-08-09): Samples collected using a RMT 1+8 from 2,000 m to surface. Indian Sector, Southern Ocean (January 2009)

Processing Description: Molecular genetic (genomic and transcriptomic) data resulting from this project have been submitted to appropriate sections of the NCBI GenBank database, as follows:

The Salpa thompsoni Whole Genome Shotgun project (see Jue et al., 2016) has been deposited at DDBJ/ENA/GenBank Genome section under the Accession No. MKHR00000000. The version described in this paper is version MKHR01000000. The genomic data include the final assembly and annotations, as well as all of the raw data files (genomic and small RNAs). Accession Nos. are assigned to the genomic data submission, as follows: SUBID, SUB1479239; BioProject, PRJNA318929; BioSample, SAMN04870323. The Salp Genome Bioproject information and data are available at https://www.ncbi.nlm.nih.gov/genome/?term=MKHR00000000.

The Salpa thompsoni Transcriptome Shotgun Assembly project (see Batta-Lona et al., 2016) has been deposited at DDBJ/ENA/GenBank under the Accession No. GFCC00000000. The version described in this paper is version GFCC01000000. The transcriptomic data includes raw sequencing data and alignment information. The Bioproject Accession No. is PRJNA279245, which includes RNA-seq data from 48 field-collected specimens, which are assigned Biosample Accession Nos. SAMN05604989 - SAMN05605036. The Salp Transcriptome Bioproject information and data are available at https://www.ncbi.nlm.nih.gov/nuccore/GFCC00000000.

References cited:

Batta-Lona, P.G., A. Maas, R. O'Neill, P.H. Wiebe, and A. Bucklin (2016) Transcriptomic profiles of spring and summer populations of the Southern Ocean salp, Salpa thompsoni, in the Western Antarctic Peninsula region. Polar Biol. doi:10.1007/s00300-016-2051-6

Jue, N., Batta-Lona, P.G., S. Trusiak, C. Obergfell, A. Bucklin, M.J. O'Neill, and

R.J. O'Neill (2016) Rapid evolutionary rates and unique genomic signatures discovered in the first reference genome for the Southern Ocean salp, Salpa thompsoni (Urochordata, Thaliacea). Genome Biol. Evol. 8: 3171-3186 doi:10.1093/gbe/evw215

BCO-DMO Processing notes:
- added conventional header with dataset name, PI name, version date
- modified parameter names to conform with BCO-DMO naming conventions
- added cruise deployment identifiers
- removed SRA accessions from data display

Project Information

Population ecology of Salpa thompsoni based on molecular indicators

The Antarctic salp, Salpa thompsoni, is an increasingly important player in the vulnerable Antarctic Peninsula pelagic ecosystem. Observations of high abundance of Salpa thompsoni during the summer in the Southern Ocean suggest that this species is capable of rapid somatic and population growth, and frequently forms dense blooms under favorable environmental conditions. The proposed research will examine genome-wide patterns of gene expression, target gene expression levels, and patterns of population genetic diversity and structure of the target salp species. Our preliminary results and data analysis have provided a promising basis for transcriptomic studies of S. thompsoni in the Southern Ocean. The proposed next steps in our genomic/transcriptomic analysis of Salpa thompsoni are: 1) completion of a reference transcriptome as a basis for genome-wide analysis of gene expression; 2) whole transcriptome shotgun sequencing (RNA-Seq) analysis to characterize gene expression in relation to individual characteristics and environmental conditions; 3) quantitative real-time PCR (qRT-PCR) characterization and validation of gene expression for 10-20 top differentially-expressed genes; and 4) detection of strand-specific allelic variation at SNP (Single Nucleotide Polymorphic) sites to analyze clonal diversity and population genetic diversity and structure. We hypothesize that: 1) deep analysis of the Salpa thompsoni transcriptome will reveal significant associations among selected set of differentially-expressed genes and critical life history stages and events (e.g., ontogenetic maturation, sexual reproduction, senescence) of the salp; and 2) the species will show variable levels of clonal diversity and significant genetic differentiation among salp populations in different regions of the Southern Ocean. Samples will be obtained from research cruises during 2011-2013 in diverse regions of the Southern Ocean; dedicated sample and data collection will be carried out during a cruise of the R/V LM GOULD (LMG11-10) to the Western Antarctic Peninsula region in November, 2011. The significance of this effort lies in new understanding of the molecular processes underlying the complex life history and population dynamics of S. thompsoni in relation to the Antarctic pelagic ecosystem and extreme and variable environmental conditions of the Southern Ocean. Most of the data

from this project are available from the Marine Geoscience Data System (MGDS), part of IEDA and is available at http://www.marine-geo.org/tools/search/Files.php?data_set_uid=18148.

Deployment Information

Deployment description for ARSV Laurence M. Gould LMG1110

UNOLS STRS record: http://strs.unols.org/Public/diu_cruise_view.aspx?cruise_id=127242 The primary science objectives of the cruise are to examine genome-wide patterns of gene expression, target gene expression levels, and patterns of population genetic diversity and structure of the Antarctic salp, Salpa thompsoni in relation to biological and physical environmental parameters in the Western Antarctic Peninsula region. High-frequency acoustics data will be used to provide information about the distribution of salps, krill, and other zooplankton. Sampling from shelf and oceanic waters between 0 and 2,000 meters will take place at selected stations using a 1-meter^2 MOCNESS to characterize the planktonic assemblage, and a Reeve net to collect live material for molecular and biochemical analysis. Environmental parameters to be measured include standard hydrographic variables (temperature, salinity, and depth), as well as fluorescence and turbidity. Water samples will be collected using a CTD rosette to determine chlorophyll concentration. An additional science objective is to develop a method of using acoustics to assess the abundance and distribution of salps in the Southern Ocean. Cruise Data Report

---

Deployment description for Umitaka-Maru UM-08-09

A cruise of the Japanese Research Vessel Umitaka-Maru in the Indian sector of the Southern Ocean, from Cape Town, South Africa to Fremantle, Australia.

Instrument Information

Instrument	Multiple Opening/Closing Net and Environmental Sensing System (MOCNESS) with a mouth opening of 1-m2 and nine 335µm mesh nets
Description	*local description not specified*
Generic Instrument	MOCNESS1

Name	
Generic Instrument Description	The Multiple Opening/Closing Net and Environmental Sensing System or MOCNESS is a family of net systems based on the Tucker Trawl principle. The MOCNESS-1 carries nine 1-m2 nets usually of 335 micrometer mesh and is intended for use with the macrozooplankton. All nets are black to reduce contrast with the background. A motor/toggle release assembly is mounted on the top portion of the frame and stainless steel cables with swaged fittings are used to attach the net bar to the toggle release. A stepping motor in a pressure compensated case filled with oil turns the escapement crankshaft of the toggle release which sequentially releases the nets to an open then closed position on command from the surface. -- from the MOCNESS Operations Manual (1999 + 2003).

Instrument	Rectangular Midwater Trawl (RMT 1+8)
Description	*local description not specified*
Generic Instrument Name	Midwater Trawl
Generic Instrument Description	A mid-water or pelagic trawl is a net towed at a chosen depth in the water column to catch schooling fish such as herring and mackerel. Midwater trawl nets have very large front openings to herd schooling fish toward the back end where they become trapped in the narrow "broiler". The sides of the deployed net are spread horizontally with two large metal foils, called "doors," positioned in front of the net. As the trawler moves forward, the doors, and therefore the net, are forced outward, keeping the net open. This instrument designation is used when specific make and model are not known.

Instrument	Multiple sequencing platforms: Ion Torrent Proton (Life Technologies, Grand Island, NY), a 454 FLX WGS (Roche Applied Science, Branford, CT), and an Illumina HiSeq 200 (Illumina, San Diego, CA)
Description	For genome sequencing
Generic Instrument Name	Automated DNA Sequencer
Generic	General term for a laboratory instrument used for deciphering the order of

Instrument Description	bases in a strand of DNA. Sanger sequencers detect fluorescence from different dyes that are used to identify the A, C, G, and T extension reactions. Contemporary or Pyrosequencer methods are based on detecting the activity of DNA polymerase (a DNA synthesizing enzyme) with another chemoluminescent enzyme. Essentially, the method allows sequencing of a single strand of DNA by synthesizing the complementary strand along it, one base pair at a time, and detecting which base was actually added at each step.

Instrument	2.3 m2 Isaacs-Kidd Midwater Trawl (IKMT) with a 505 µm mesh net
Description	*local description not specified*
Generic Instrument Name	Isaacs-Kidd Midwater Trawl
Generic Instrument Description	A trawl with a pentagonal mouth opening and a dihedral depressor vane as part of the mouth opening. IKMTs come in various dimensions (refer to individual dataset documentation). The original IKMTs were 10 foot (304 cm) and 15 foot (457 cm) at the mouth. The 10 foot IKMT net was 31 feet (9.45 m) in length (Wiebe and Benfield 2003).

Instrument	
Description	A PCR product ~500 bp was generated, followed by development of internal primers to produce a smaller PCR product amenable to qPCR.
Generic Instrument Name	PCR Thermal Cycler
Generic Instrument Description	General term for a laboratory apparatus commonly used for performing polymerase chain reaction (PCR). The device has a thermal block with holes where tubes with the PCR reaction mixtures can be inserted. The cycler then raises and lowers the temperature of the block in discrete, pre-programmed steps. (adapted from http://serc.carleton.edu/microbelife/research_methods/genomics/pcr.html)

Instrument	Agilent 2100 Bioanalyzer
Description	For RNA quality control, and to assess the size distribution of library fragments.

Generic Instrument Name	Bioanalyzer
Generic Instrument Description	A Bioanalyzer is a laboratory instrument that provides the sizing and quantification of DNA, RNA, and proteins. One example is the Agilent Bioanalyzer 2100.