Antarctic invertebrate collection locations and NCBI SRA accessions from Halanych and Kocot (2014) Biol Bull (Antarctic Inverts project)

Website: https://www.bco-dmo.org/dataset/671638
Data Type: Cruise Results
Version: 1
Version Date: 2016-12-22

Project
» Genetic connectivity and biogeographic patterns of Antarctic benthic invertebrates (Antarctic Inverts)

Contributors	Affiliation	Role
Halanych, Kenneth M.	Auburn University	Principal Investigator
Mahon, Andrew	Central Michigan University	Co-Principal Investigator
Copley, Nancy	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract
This dataset was published as Table 1 from Halanych and Kocot (2014). It contains GenBank SRA accession links of Lophotrochozoan specimens collected globally.

Table of Contents

- Coverage
- Dataset Description
  - Methods & Sampling
  - Data Processing Description
- Data Files
- Related Publications
- Related Datasets
- Parameters
- Instruments
- Deployments
- Project Information
- Funding

Coverage

Spatial Extent: **N**:66.55 **E**:33.1 **S**:-65.65 **W**:-123.08

Methods & Sampling

From Halanych and Kocot (2014):

Taxon sampling
Transcriptomes were obtained for the 18 invertebrate species listed in Table 1. For most animals, adult body wall (epidermis, muscles, and underlying tissue) was used for transcriptomes. For entoprocts and small brachiopods (Novocrania anomala and Macandrevia cranium), whole-body tissue was employed; for larger brachiopods (Glottidia pyramidata, Hemithris psittacea, and Laqueus californicus), extractions focused on mantle and muscle; for the solenogaster (a.k.a neomenioid) aplacophoran mollusc Proneomenia custodiens, unhatched juveniles were used. Collection locality information is also given in Table 1. The three annelid transcriptomes (Magelona beckleyi, Paramphinome jeffreysii, and Phyllochaetopterus prolifica) have been previously published (Weigert et al., 2014) in a phylogenomics paper detailing basal annelid relations. The remainder were part of a phylogenomics project on lophotrochozoan relationships (Kocot, 2013).

Transcriptomic data collection

We follow the methods described in Weigert et al. (2014) and Kocot (2013), which are briefly summarized here. RNA was extracted with TRIzol (Invitrogen) and purified with the RNeasy kit (Qiagen; Valencia, CA) using oncolumn DNAse digestion. In the case of the small juveniles of Proneomenia custodiens, RNA was extracted with the RNeasy Micro kit (Qiagen; Valencia, CA). The SMART cDNA library construction kit (Clontech Laboratories, Mountain View, CA) was used to construct cDNA libraries. These libraries were sequenced by The Genomic Services Lab at the Hudson Alpha Institute in Huntsville, Alabama, on the Illumina (San Diego, CA) HiSeq 2000 platform with 2x100 paired-end sequencing. To reduce memory usage during transcriptome assembly, sequencing data were digitally normalized to a k-mer coverage depth of 30 using the normalize-by-median.py script (Brown et al., 2012). Data were then assembled using Trinity (Grabherr et al., 2011; 8 June 2012 version) with default settings.

Identification of genes
UniProt accession numbers for human toll-like receptor genes were obtained from InnateDB (Innate DB, 2014; Lynn et al., 2008; Breuer et al., 2013). Amino acid sequences for TLR1-TLR10 were obtained from then National Center for Biotechnology Information (NCBI; Table 2) and used for bait in a TBLASTN search (Altschul et al., 1990) against the assembled nucleotide transcriptomes. For these searches, an e-value of 10^-5 was employed and only top hits longer than 800 nucleotides (to filter out smaller gene fragments) were retained. Significant hits were translated to amino acid sequences with TransDecoder in the Trinity package (Grabherr et al., 2011), and redundant sequences were removed with cd-hit (Li and Godzik, 2006) using an identity (-c flag) of 1.0. To further confirm their identity, the translated hits were searched against the Swiss Prot database using BLASTP with an e-value cutoff of 10^-5. As TransDecoder can yield multiple translations per nucleotide sequence, all translations were subject to BLASTP, and only the best hits were retained. Sequences that returned the top hit with a significant e-value and matched to a TLR gene were retained and compared to the SMART database (SMART, 2014; Schultz et al., 1998; Letunic et al., 2012) using the "normal" search setting to identify protein domains.

Previous phylogenetic analyses (e.g., Zheng et al., 2005; Davidson et al., 2008) on TLR genes typically used only the TIR domain for analyses. These analyses are limited to about 140-180 amino acids in length. Here, we aligned recovered TLR contigs from lophotrochozoans and Priapulus with human and Drosophilia TLR genes using MAFFT ver. 7 (Katoh and Standley, 2013). Even after reducing the alignment to the conserved TIR domain, sequences were highly variable, resulting in a questionable alignment. Considering that the TIR domain is shared among several gene families (Aravind et al., 2001) and the variability of the alignment, we question whether a tree produced from these sequences accurately represents gene genealogy. This suspicion was confirmed by using the T-REX server (Boc et al., 2012) for a RAxML-VI-HPC (Stamatakis, 2006) analysis (parameters included PROTGAMMA, WAG, 100 bootstraps) that yielded a tree with very low nodal support value. Roughly 50% of the nodes reported a bootstrap value less than 40%, and exploring different parameters did little to improve results. We used a much broader taxon sampling than for previous TIR trees, which likely influenced variability of the alignment. Due to the questionable reliability of such a phylogenetic analysis, results are not included herein.

Data Processing Description

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 links to NCBI accession pages
- converted lat and lon to decimal degrees
- replaced commas with underscores or semi-colons
- reformatted to flat table by adding columns for taxon1 and taxon2

[table of contents | back to top]

Data Files

File	
Hal_Kocot_2014_T1.csv	(Comma Separated Values (.csv), 3.03 KB) MD5:0f3e76e477823aa0577f396750551605
Primary data file for dataset ID 671638	

[table of contents | back to top]

Related Publications

Halanych, K. M., & Kocot, K. M. (2014). Repurposed Transcriptomic Data Facilitate Discovery of Innate Immunity Toll-Like Receptor (TLR) Genes Across Lophotrochozoa. The Biological Bulletin, 227(2), 201–209. doi:10.1086/bblv227n2p201 https://doi.org/10.1086/BBLv227n2p201
Results

[table of contents | back to top]

Related Datasets

IsRelatedTo

Halanych, K. M., Mahon, A. (2016) **Best other leucine-rich repeat (LRR) genes from Lophotrochozoa BLASTP search, from Table 4, Halanych and Kocot (2014) Biol. Bull. (Antarctic Inverts project).** Biological and Chemical Oceanography Data Management Office (BCO-DMO). Version Date 2016-12-27 http://lod.bco-dmo.org/id/dataset/671685 [view at BCO-DMO]

Halanych, K. M., Mahon, A. (2016) **Best toll-like receptors (TLR) genes from Lophotrochozoa BLASTP search, from Table 3, Halanych and Kocot (2014) Biol. Bull. (Antarctic Inverts project).** Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2016-12-27 http://lod.bco-dmo.org/id/dataset/671662 [view at BCO-DMO]

[table of contents | back to top]

Parameters

Parameter	Description	Units
taxon1	taxonomic group	unitless
taxon2	more specific taxonomic group	unitless
species	taxonomic genus and species name	unitless
locality	location of specimen collection	unitless
latitude	latitude; north is positive	decimal degrees
longitude	longitude; east is positive	decimal degrees
NCBI_SRA_accession	SRA NCBI GenBank accession number	unitless
link_NCBI_SRA_accession	link to SRA NCBI GenBank accession	unitless

[table of contents | back to top]

Instruments

Dataset-specific Instrument Name	Illumina (San Diego, CA) HiSeq 2000 platform at The Genomic Services Lab at the Hudson Alpha Institute in Huntsville, Alabama,
Generic Instrument Name	Automated DNA Sequencer
Generic Instrument Description	A DNA sequencer is an instrument that determines the order of deoxynucleotides in deoxyribonucleic acid sequences.

Dataset-specific Instrument Name	
Generic Instrument Name	Thermal Cycler
Generic Instrument Description	A thermal cycler or "thermocycler" is a general term for a type of laboratory apparatus, commonly used for performing polymerase chain reaction (PCR), that is capable of repeatedly altering and maintaining specific temperatures for defined periods of time. 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. They can also be used to facilitate other temperature-sensitive reactions, including restriction enzyme digestion or rapid diagnostics. (adapted from http://serc.carleton.edu/microbelife/research_methods/genomics/pcr.html)

[table of contents | back to top]

Deployments

Halanych_lab_2011-16

Website	https://www.bco-dmo.org/deployment/671488
Platform	Auburn University lab
Start Date	2011-08-01
End Date	2016-07-31
Description	Invertebrate genomics

[table of contents | back to top]

Project Information

Genetic connectivity and biogeographic patterns of Antarctic benthic invertebrates (Antarctic Inverts)

Coverage: Antarctica

Extracted from the NSF award abstract:

The research will explore the genetics, diversity, and biogeography of Antarctic marine benthic invertebrates, seeking to overturn the widely accepted suggestion that benthic fauna do not constitute a large, panmictic

population. The investigators will sample adults and larvae from undersampled regions of West Antarctica that, combined with existing samples, will provide significant coverage of the western hemisphere of the Southern Ocean. The objectives are: 1) To assess the degree of genetic connectivity (or isolation) of benthic invertebrate species in the Western Antarctic using high-resolution genetic markers. 2) To begin exploring planktonic larvae spatial and bathymetric distributions for benthic shelf invertebrates in the Bellinghausen, Amundsen and Ross Seas. 3) To continue to develop a Marine Antarctic Genetic Inventory (MAGI) that relates larval and adult forms via DNA barcoding.

[table of contents | back to top]

Funding

Funding Source	Award
NSF Office of Polar Programs (formerly NSF PLR) (NSF OPP)	PLR-1043745
NSF Office of Polar Programs (formerly NSF PLR) (NSF OPP)	PLR-1043670

[table of contents | back to top]