Results of cluster analysis and gene ontology: carbonate Organic Matrix (COM) proteins from coral, mollusk, and sea urchin; analyzed in the Falkowski lab at Rutgers from 2010-2014 (CROA project)

Website: https://www.bco-dmo.org/dataset/536485
Data Type: Other Field Results
Version: 1
Version Date: 2014-10-29

Project
» The Molecular Basis of Ocean Acidification Effects on Calcification in Zooxanthellate Corals (CROA)

Program
» Science, Engineering and Education for Sustainability NSF-Wide Investment (SEES): Ocean Acidification (formerly CRI-OA) (SEES-OA)

Contributors	Affiliation	Role
Falkowski, Paul G.	Rutgers University (Rutgers IMCS)	Principal Investigator, Contact
Rosenthal, Yair	Rutgers University	Co-Principal Investigator
Schofield, Oscar M.E.	Rutgers University	Co-Principal Investigator
Sherrell, Robert M.	Rutgers University	Co-Principal Investigator
Rauch, Shannon	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract
This dataset includes results of cluster analysis and gene ontology: carbonate Organic Matrix (COM) proteins from coral, mollusk, and sea urchin; analyzed in the Falkowski lab at Rutgers from 2010-2014.

Table of Contents

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

Dataset Description

Cluster analysis and gene ontology to compare ~1500 proteins, from over 100 studies, extracted from calcium carbonates in stony corals, bivalve and gastropod mollusks, and adult and larval sea urchins. This dataset includes information presented in Supplemental Table S2 from Drake et al. 2014. Refer to Drake et al. (2014) for more information on methodology and results.

Methods & Sampling

Methodology described in Drake et al. 2014:
Sequences from over 100 biomineral proteome studies were grouped by hierarchical clustering using the CD-HIT suite web server (Li and Godzik, 2006; Huang et al., 2010; http://weizhong-lab.ucsd.edu/cd-hit/) and assigned gene ontology (GO) terms using Blast2Go software (Conesa et al., 2005). Although 1531 proteins reduced to 1051 clusters at 30% similarity, only 64 clusters showed sequence similarity across phyla. Studies published from the 1990s through June 2013, using N-terminal and mass spectrometry COM sequencing, RT-PCR, or GO and KEGG annotation of genomic and transcriptomic data sets are included. Mass spectrometry sequences were excluded if the experimental data were compared with gene models from a different species.

This dataset includes information presented in Supplemental Table S2 from Drake et al. 2014:
Proteins from coral, mollusk, and sea urchin direct COM sequencing, RT-PCR, or GO and KEGG annotation. 1076 proteins, including redundancy when noted by multiple sources, reduced to 1031 non-redundant sequences.

Data Processing Description

BCO-DMO Processing Notes:
- Replaced blanks with 'nd' to indicate 'no data'.
- Replaced alpha symbols with the word "alpha" and beta symbols with "beta".
- Created code list ("SI_Table_2_sequence_codes.csv") for long sequences contained in the 'detection_type' column.
- Created reference list ("SI_Table_2_references.csv") for full references of names in the 'reference' column.

[table of contents | back to top]

Data Files

File
SI_Table2.csv (Comma Separated Values (.csv), 131.92 KB) MD5:eef20978ad572b1e7ea5e6751c9a006d Primary data file for dataset ID 536485

[table of contents | back to top]

Supplemental Files

File	
SI_Table_2_references.csv	(Comma Separated Values (.csv), 17.34 KB) MD5:bc997b3b68e1f4363701c99256fca7c9
Complete reference information for "Carbonate Organic Matrix (COM) proteins from coral, mollusk, and sea urchin" dataset (dataset 536485; PI: Falkowski). The "abbrev" column contains the abbreviated reference as cited in dataset 536485. The "full_reference" column contains the complete citation for that paper.	
SI_Table_2_sequence_codes.csv	(Comma Separated Values (.csv), 1.68 KB) MD5:b3968320952b0f70c9409d1b32496fa9
Full sequences for those abbreviated in the "detection_type" column of "Carbonate Organic Matrix (COM) proteins from coral, mollusk, and sea urchin" dataset (dataset 536485; PI: Falkowski). The "code" column contains the sequence code as listed in dataset 536485. The "sequence" column contains the full sequence.	

[table of contents | back to top]

Related Publications

Drake, J. L., Mass, T., & Falkowski, P. G. (2014). The evolution and future of carbonate precipitation in marine invertebrates: Witnessing extinction or documenting resilience in the Anthropocene. Elementa: Science of the Anthropocene, 2. https://doi.org/10.12952/journal.elementa.000026
https://doi.org/10.12952/journal.elementa.000026#sthash.xkJXJlEd.dpuf
Results

[table of contents | back to top]

Parameters

Parameter	Description	Units
protein_name	Name of the protein.	text

cluster_num	Cluster numbers were assigned to each batch of proteins that clustered together with 30% similarity or higher based on a clustering method using the CD-HIT suite web server. This was done progressively from 100% sequence similarity down to 30% with the cluster numbers carrying through. However, if a protein showed 30% similarity to another batch of proteins, then, for the final cut-off of 30%, the proteins from the previously independent batch were grouped in with the new batch. For instance, Cluster005744 has a cluster_num of 12. That means that it has 30% sequence similarity with proteins from proteins 5955, 00152 1(a), and 19987. In a previous clustering (with a higher similarity cutoff), it also showed sequence similarity with proteins Cluster020987, Cluster020227, and Cluster015890 (so they were previously in cluster #12). However, now proteins Cluster020987, Cluster020227, and Cluster015890 show at least 30% sequence similarity with one of the Cluster000783, Cluster050768, Cluster050003, or lotgi1|230171 proteins (that were previously assigned to cluster #0). So now since at least 1 protein in cluster #12 is similar to cluster #0, all #12 proteins were assigned to cluster #0.	unitless
group_num	Group number.	unitless
genus	Genus name.	text
species	Species name.	text
taxon	Taxonomic grouping (coral, echinoderm, or mollusk).	text
reference	Reference paper. Refer to Supplemental Files for full citations.	text
ID1	The accession number, author's assigned protein number, or other note about a designator for the individual protein.	text
detection_type	Method used to sequence the protein.The strings of letters are the individual peptide sequences detected by LC-MS/MS.	text
type	Indicates if the protein can be assigned the name of a type of protein based on BLAST similarity. Also contains notes, such as if the protein was assigned its own accession number but was the same protein as found in another row.	text

[table of contents | back to top]

Instruments

Dataset-specific Instrument Name	Mass Spec
Generic Instrument Name	Mass Spectrometer
Generic Instrument Description	General term for instruments used to measure the mass-to-charge ratio of ions; generally used to find the composition of a sample by generating a mass spectrum representing the masses of sample components.

[table of contents | back to top]

Deployments

lab_Falkowski

Website	https://www.bco-dmo.org/deployment/537725
Platform	Rutgers_New_Brunswick
Start Date	2010-09-01
End Date	2014-08-01
Description	Laboratory-based research for the project "The Molecular Basis of Ocean Acidification Effects on Calcification in Zooxanthellate Corals" were conducted at Dr. Falkowski's lab at the Rutgers New Brunswick campus: 71 Dudley Road New Brunswick, NJ 08901

[table of contents | back to top]

Project Information

The Molecular Basis of Ocean Acidification Effects on Calcification in Zooxanthellate Corals (CROA)

Website: http://corals.marine.rutgers.edu/

Coverage: laboratory-based in New Brunswick, NJ

From the NSF award abstract:

Ocean acidification (the decrease in seawater pH) is driven by the increase in atmospheric CO2. This is expected to have a dramatic effect on organisms that precipitate calcium carbonate. Coral reefs are formed and maintained by calcifying organisms, particularly reef-building corals. Current predictions are that coral species will be negatively impacted; however the limited number of available measurements exhibit significant variability for reasons that are not understood. This is critically important as coral reef ecosystems hold significant cultural and economic values both nationally and internationally. This program is therefore focused on the molecular basis for calcification in corals in order to understand how corals will respond to ocean acidification in the next century. Rutgers University has a state-of-art coral culture facility that will be used to simulate future ocean conditions. The work will utilize a unique set of coral tissue cultures that will allow scientists to assess the cellular biology that underlies the responses of corals to ocean acidification. The laboratory measurements will also determine how geochemical signatures of corals are affected by varying environmental conditions. These results are important because coral geochemical signatures are used to understand how corals have responded to changes in the ocean pH in the historical past. The project will be conducted by a research team at Rutgers, in collaboration with scientists in Taiwan and Israel.

[table of contents | back to top]

Program Information

Science, Engineering and Education for Sustainability NSF-Wide Investment (SEES): Ocean Acidification (formerly CRI-OA) (SEES-OA)

Website: https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503477

Coverage: global

NSF Climate Research Investment (CRI) activities that were initiated in 2010 are now included under Science, Engineering and Education for Sustainability NSF-Wide Investment (SEES). SEES is a portfolio of activities that highlights NSF's unique role in helping society address the challenge(s) of achieving sustainability. Detailed information about the SEES program is available from NSF (https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504707).

In recognition of the need for basic research concerning the nature, extent and impact of ocean acidification on oceanic environments in the past, present and future, the goal of the SEES: OA program is to understand (a) the chemistry and physical chemistry of ocean acidification; (b) how ocean acidification interacts with processes at the organismal level; and (c) how the earth system history informs our understanding of the effects of ocean acidification on the present day and future ocean.

Solicitations issued under this program:
NSF 10-530, FY 2010-FY2011
NSF 12-500, FY 2012
NSF 12-600, FY 2013
NSF 13-586, FY 2014
NSF 13-586 was the final solicitation that will be released for this program.

PI Meetings:
1st U.S. Ocean Acidification PI Meeting(March 22-24, 2011, Woods Hole, MA)
2nd U.S. Ocean Acidification PI Meeting(Sept. 18-20, 2013, Washington, DC)
3rd U.S. Ocean Acidification PI Meeting (June 9-11, 2015, Woods Hole, MA – Tentative)

NSF media releases for the Ocean Acidification Program:

Press Release 10-186 NSF Awards Grants to Study Effects of Ocean Acidification

Discovery Blue Mussels "Hang On" Along Rocky Shores: For How Long?

Discovery nsf.gov - National Science Foundation (NSF) Discoveries - Trouble in Paradise: Ocean Acidification This Way Comes - US National Science Foundation (NSF)

Press Release 12-179 nsf.gov - National Science Foundation (NSF) News - Ocean Acidification: Finding New Answers Through National Science Foundation Research Grants - US National Science Foundation (NSF)

Press Release 13-102 World Oceans Month Brings Mixed News for Oysters

Press Release 13-108 nsf.gov - National Science Foundation (NSF) News - Natural Underwater Springs Show How Coral Reefs Respond to Ocean Acidification - US National Science Foundation (NSF)

Press Release 13-148 Ocean acidification: Making new discoveries through National Science Foundation research grants

Press Release 13-148 - Video nsf.gov - News - Video - NSF Ocean Sciences Division Director David Conover answers questions about ocean acidification. - US National Science Foundation (NSF)

Press Release 14-010 nsf.gov - National Science Foundation (NSF) News - Palau's coral reefs surprisingly resistant to ocean acidification - US National Science Foundation (NSF)

Press Release 14-116 nsf.gov - National Science Foundation (NSF) News - Ocean Acidification: NSF awards $11.4 million in new grants to study effects on marine ecosystems - US National Science Foundation (NSF)

[table of contents | back to top]

Funding

Funding Source	Award
NSF Emerging Frontiers Division (NSF EF)	EF-1041143

[table of contents | back to top]