Total spectral count of proteins from R/V Kilo Moana cruise KM1128 for the METZYME expedition in the tropical North Pacific in 2011.

Website: https://www.bco-dmo.org/dataset/730722
Data Type: Cruise Results
Version: 5
Version Date: 2022-06-06

Project
» Connecting Trace Elements and Metalloenzymes Across Marine Biogeochemical Gradients (GPc03) (MetZyme)

Program
» U.S. GEOTRACES (U.S. GEOTRACES)

Contributors	Affiliation	Role
Saito, Mak A.	Woods Hole Oceanographic Institution (WHOI)	Principal Investigator
Ake, Hannah	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager
York, Amber D.	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Abstract
Global metaproteomic dataset for KM1128 the METZYME expedition. The global proteome dataset presented here included 16951 protein identifications and 31994 unique peptide identifications compiled from 37 samples and 74 raw files (CID and HCD modes for each file) and ### total spectra.

Table of Contents

- Coverage
- Dataset Description
  - Methods & Sampling
  - Data Processing Description
  - BCO-DMO Curation Notes
- Related Publications
- Related Datasets
- Parameters
- Instruments
- Deployments
- Project Information
- Program Information
- Funding

Coverage

Location: Tropical North Pacific
Spatial Extent: **N**:20.6372 **E**:175.9503 **S**:-15.7878 **W**:153.2745
Temporal Extent: 2011-10-01 - 2011-10-25

Dataset Description

These data are part of the Ocean Protein Portal "MetZyme 0.2" dataset version 5 (https://proteinportal.whoi.edu/; Saito et al., 2020).

Methods & Sampling

Ocean metaproteomes were sampled and extracted as described in Saito et al., 2014. The protocols are reproduced here:

Samples were collected during the KM1128 METZYME research expedition (Metals and Enzymes in the Pacific) on the R/V Kilo Moana October 1-25, 2011 from Oahu, Hawaii to Apia, Samoa, with Carl Lamborg and Mak Saito as Chief Scientists. Microbial biomass for protein analyses was collected on vertical profiles using in situ high volume particle filtration pumps with a focus on the North Pacific and Equatorial regions. Protein samples were collected by a suite of 4 L/min and 8 L/min McLane Pumps (WTS-LV; McLane Research Laboratories Inc., Falmouth MA, USA) outfitted with custom Mini-MULVS multiple filter head systems. The 0.2-3.0 micron size fraction was collected on 142 mm filters (Supor, Pall Corp.) for analyses used in this study. The volume of water was pumped until a minimal flow rate was achieved or the allotted cast time period expired, typically ~300-500 L. Filters were sectioned immediately after pump retrieval, and protein samples (¼ filter) were stored in RNAlater reagent (Ambion, Life Technologies), which has been shown to be an effective preservative for cyanobacterial biomass (Saito et al., 2011), frozen at -80C, transported back to the laboratory on dry ice, and stored at -80C until analysis.

Total microbial protein (0.2-3.0um fraction) was extracted using detergent-based methods, described below. Total protein showed enhanced concentrations in the photic zone, particularly in the Equatorial and South Pacific portions of the transect (Fig. S1F). For protein extraction, samples were thawed and the filter and RNAlater preservative (Ambion Life Technologies) were separated. The removed preservative was spin-concentrated by a 5K MWCO membrane (Sartorius Stedim Biotech 6 mL, 5 K MWCO Vivaspin units; Goettingen, Germany), and rinsed with 0.1M Tris buffer to recover, desalt, and concentrate any suspended material. The sample filter was unfolded and placed in a larger tube to which 1% SDS extraction buffer (1% SDS, 0.1M Tris/HCL pH 7.5, 10mM EDTA) and the rinsed/desalted RNAlater fraction was added back. Each sample was incubated at room temperature for 15 minutes, heated at 95C for 10 minutes, and shaken at room temperature for 350 rpm for 1 h. The protein extract was decanted and placed in a new tube and centrifuged for 30 min x 3220 g at room temperature. The supernatants were removed and filtered through a 5 um low protein binding syringe filter (Fisher Scientific), and the filter rinsed with extraction buffer. The extracts were concentrated by 5 kD membrane centrifugation to a small volume, washed with extraction buffer, and concentrated again. Each sample was precipitated with cold 50% methanol (MeOH) 50% acetone 0.5 mM HCL for 3 days at –20C, centrifuged at 14100 x g (14500 rpm) for 30 m at 4C, decanted and dried by vacuum concentration (Thermo Savant Speedvac) for 10 min or until dry. Pellets were resuspended in 1% SDS extraction buffer and left at room temperature (RT) for 1 h to redissolve. Total protein was quantified (Bio-Rad DC protein assay, Hercules, CA) with BSA as a standard.

Extracted proteins were purified from SDS detergent, reduced, alkylated, and trypsin digested, while embedded within a polyacrylamide tube gel, using a modified protocol from a previously published method (38). The tube gel approach allowed all proteins including membrane proteins to be solubilized by detergent and purified while immobilized in the gel matrix. A gel premix was made by combining 1 M Tris HCL (pH 7.5) and 40% Bis-acrylimide L 29:1 (Acros Organics) at a ratio of 1:3. The premix (103 ul) was combined with an extracted protein sample (usually 25 ug-200 ug), TE, 7 ul 1% APS and 3 ul of TEMED (Acros Organics) to a final volume of 200 ul. After 1 h of polymerization at RT, 200 ul of gel fix solution (50% ETOH, 10% acetic acid in LC/MS grade water) was added to the top of the gel and incubated at RT for 20 minutes. Liquid was then removed and the tube gel was transferred into a new 1.5 mL microtube.

Data Processing Description

BCO-DMO Curation Notes

Data Version 5 (2022-06-06) replaces data version 4.
* This version updates the to match the data ingested into the Ocean Protein Portal as version 5 at OPP.
* Both the OPP and the BCO-DMO datasets page imported data from source file "METZYME_02_proteins_for_OPP.csv" file date 2022-06-06. These data will appear as file "730722_v5_metzyme_0_pt_2_proteins.csv" on the BCO-DMO dataset page.
* This version of the BCO-DMO dataset did not join with the related station information dataset

(https://www.bco-dmo.org/dataset/730891 version 3) as was done with past versions since the collection information (lat,lon, date,time (local), cruise) was included in source file METZYME_02_peptides_for_OPP.csv.
* Additional datetime column added in ISO Format (utc time zone), coverted from local HST time provided.

Data Version 4 (2021-03-30) replaces data version 2 (2018-05-17). There was never a published version 3 of this dataset.

- Data version 4 uses updated station information (dataset 730891 version 3) joined with the submitted file Metzyme_Proteins_V1_targetdecoyPSMvalidator_030918.xlsx. Also several transformation and find and replace operations performed as described below.
- Sheet "Final" imported into the BCO-DMO data system from file
Metzyme_Proteins_V1_targetdecoyPSMvalidator_030918.xlsx
#- N/A interpreted as no data value
- skipped row that started "END"
- Data unpivoted to transform data in the header into depth and station columns and put all spectral counts in one "spectral_count" column.
- Columns renamed to fit the OPP template (see https://github.com/oceanproteinportal/data-file-templates).
- Added minimum_filter_size_microns column with value '0.2'
- Added maximum_filter_size_microns with 'value 3.0'
- Added station information from dataset 730891 version 3 using a join operation station and depth as keys. - Added dates times (local and UTC), and lat, lon.
- removed prefixes and suffixes (e.g. \|\|_GAP_) and changed delimiter within a value to ; from || in kegg_id, kegg_descriptoin, pframs_id, and pframs_name, uniprot_id, enzyme_comm_id.
stripped whitespace around internal cell delimiter ;
- file saved as csv format.

BCO-DMO Processing Notes Data version 2 (2018-05-17):

- Added column KO_link for the protein links at NCBI
- Replaced blank cells with nd
- Replaced #N/A with NA
- Replaced commas in cells with ;
- Removed first column which contained the row number
- Added cruise column for relevant deployment information

[table of contents | back to top]

Related Publications

Saito, M. A., Bertrand, E. M., Duffy, M. E., Gaylord, D. A., Held, N. A., Hervey, W. J., Hettich, R. L., Jagtap, P. D., Janech, M. G., Kinkade, D. B., Leary, D. H., McIlvin, M. R., Moore, E. K., Morris, R. M., Neely, B. A., Nunn, B. L., Saunders, J. K., Shepherd, A. I., Symmonds, N. I., & Walsh, D. A. (2019). Progress and Challenges in Ocean Metaproteomics and Proposed Best Practices for Data Sharing. Journal of Proteome Research, 18(4), 1461–1476. https://doi.org/10.1021/acs.jproteome.8b00761
Methods

Saito, M. A., Bulygin, V. V., Moran, D. M., Taylor, C., & Scholin, C. (2011). Examination of Microbial Proteome Preservation Techniques Applicable to Autonomous Environmental Sample Collection. Frontiers in Microbiology, 2. doi:10.3389/fmicb.2011.00215
Related Research

Saito, M. A., McIlvin, M. R., Moran, D. M., Goepfert, T. J., DiTullio, G. R., Post, A. F., & Lamborg, C. H. (2014). Multiple nutrient stresses at intersecting Pacific Ocean biomes detected by protein biomarkers. Science, 345(6201), 1173–1177. https://doi.org/10.1126/science.1256450
Methods

Saito, M. A., Saunders, J. K., Chagnon, M., Gaylord, D. A., Shepherd, A., Held, N. A., Dupont, C., Symmonds, N., York, A., Charron, M., & Kinkade, D. B. (2020). Development of an Ocean Protein Portal for Interactive Discovery and Education. Journal of Proteome Research, 20(1), 326–336. https://doi.org/10.1021/acs.jproteome.0c00382
Related Research

[table of contents | back to top]

Related Datasets

IsRelatedTo

Saito, M. (2019) **Station locations for the global metaproteomic dataset from R/V Kilo Moana cruise KM1128, the METZYME expedition in the tropical North Pacific in 2011.** Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 3) Version Date 2019-04-12 doi:10.1575/1912/bco-dmo.730891.3 [view at BCO-DMO]
Relationship Description: Station locations and sampling information for these data.

Saito, M. A. (2017) **FASTA file of protein identifications from R/V Kilo Moana KM1128 in the North Pacific during 2011 (MetZyme project).** Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2017-07-14 http://lod.bco-dmo.org/id/dataset/708816 [view at BCO-DMO]
Relationship Description: This dataset is part of the same "MetZyme 0.2" dataset at the Ocean Protein Portal (https://proteinportal.whoi.edu/).

Saito, M. A. (2022) **Peptides and their spectral counts from KM1128 the METZYME expedition on R/V Kilo Moana in the tropical North Pacific in 2011.** Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 5) Version Date 2022-06-03 http://lod.bco-dmo.org/id/dataset/731314 [view at BCO-DMO]
Relationship Description: The peptides related to these proteins. This dataset is part of the same "MetZyme 0.2" dataset at the Ocean Protein Portal (https://proteinportal.whoi.edu/).

IsPartOf

Protein Portal (2024). Ocean Protein Portal. Available from https://proteinportal.whoi.edu/

[table of contents | back to top]

Parameters

Parameter	Description	Units
sample_id	Sample identifier	unitless
MS_MS_sample_name	Unique name for the mass spec sample and run	unitless
station_id	Station number	unitless
depth_m	Depth of sampling event	meters
latitude_dd	Latitude	decimal degrees
longitude_dd	Longitude	decimal degrees
date_y_m_d	Date (local) of sampling at station and depth (HST, UTC-10)	unitless
time_h_m_s	Time (local) of sampling at station and depth (HST, UTC-10)	unitless

ISO_DateTime_UTC	Sampling datetime with timezone (UTC) in ISO 8601 format	unitless
minimum_filter_size_microns	Minimum filter size	microns
maximum_filter_size_microns	Maximum filter size	microns
cruise_id	Cruise deployment number	unitless
protein_id	The specific name of the full protein length sequence assembled in the metagenome that was used for peptide identification. An identifier that uniquely identifies this protein within this dataset and the FASTA file (see Related Datasets).	unitless
spectral_count_sum	Spectral count of proteins at each station and depth	count
molecular_weight_kDa	Molecular weight	kilodalton (kDa)
protein_name	Descriptive name of the function of the protein	unitless
ncbi_id	NCBI Taxonomy organism identifier (for ncbi_name)	unitless
ncbi_name	NCBI Taxonomy name (corresponding to ncbi_id)	unitless
kegg_id	The Kegg Orthology Entry identifier for the best Kegg match.	unitless
kegg_description	Description of the function of the specific KEGG protein group	unitless
kegg_pathway	Decription of the cellular pathway that the KEGG protein is a part of.	unitless
pfams_id	Protein family (Pfam) ID number	unitless
pfams_name	Protein family (Pfam) description	unitless
uniprot_id	Uniprot database ID number	unitless
enzyme_comm_id	Enzyme Commission ID number	unitless

[table of contents | back to top]

Instruments

Dataset-specific Instrument Name	5 kD membrane centrifuge
Generic Instrument Name	Centrifuge
Dataset-specific Description	Used to concentrate the extracts
Generic Instrument Description	A machine with a rapidly rotating container that applies centrifugal force to its contents, typically to separate fluids of different densities (e.g., cream from milk) or liquids from solids.

Dataset-specific Instrument Name	McLane Pumps (WTS-LV; McLane Research Laboratories Inc., Falmouth MA, USA)
Generic Instrument Name	McLane Pump
Dataset-specific Description	Used to collect microbial biomass for protein analyses
Generic Instrument Description	McLane pumps sample large volumes of seawater at depth. They are attached to a wire and lowered to different depths in the ocean. As the water is pumped through the filter, particles suspended in the ocean are collected on the filters. The pumps are then retrieved and the contents of the filters are analyzed in a lab.

[table of contents | back to top]

Deployments

KM1128

Website	https://www.bco-dmo.org/deployment/59053
Platform	R/V Kilo Moana
Start Date	2011-10-01
End Date	2011-10-25
Description	This is a MetZyme project cruise. The original cruise data are available from the NSF R2R data catalog.

[table of contents | back to top]

Project Information

Connecting Trace Elements and Metalloenzymes Across Marine Biogeochemical Gradients (GPc03) (MetZyme)

Coverage: Tropical North Pacific along 150 degrees West from 18 degrees North to the equator

MetZyme project researchers will determine the role of enzymatic activity in the cycling of trace metals. Specifically the research will address the following questions: (1) degradation of sinking particulate organic material in the Tropical North Pacific can be influenced by the ability of microbes to synthesize zinc proteases, which in turn is controlled by the abundance or availability of zinc, and (2) methylation of mercury is controlled, in part, by the activity of cobalt-containing enzymes, and therefore the supply of labile cobalt to the corrinoid-containing enzymes or co-factors responsible for methylation. To attain their goal, they will collect dissolved and particulate samples for trace metals and metalloenzymes from three stations along a biogeochemical gradient in the Tropical North Pacific (along 150 degrees West from 18 degrees North to the equator). Sinking particles from metal clean sediment traps will also be obtained. The samples will also be used to carry out shipboard incubation experiments using amendments of metals, metal-chelators, B12, and proteases to examine the sensitivity and metal limitation of heterotrophic, enzymatic degradation of organic matter within the oceanic "Twilight Zone" (100-500 m). This study will result in a novel metaproteomic/metalloenzyme datasets that should provide insights into the biogeochemical cycling of metals, as well as co-limitation of primary productivity and controls on the export of carbon from the photic zone. In addition to the final data being contributed to BCO-DMO, an online metaproteomic data server will be created so the community has access to the raw data files generated by this research.

[table of contents | back to top]

Program Information

U.S. GEOTRACES (U.S. GEOTRACES)

Website: http://www.geotraces.org/

Coverage: Global

GEOTRACES is a SCOR sponsored program; and funding for program infrastructure development is provided by the U.S. National Science Foundation.

GEOTRACES gained momentum following a special symposium, S02: Biogeochemical cycling of trace elements and isotopes in the ocean and applications to constrain contemporary marine processes (GEOSECS II), at a 2003 Goldschmidt meeting convened in Japan. The GEOSECS II acronym referred to the Geochemical Ocean Section Studies To determine full water column distributions of selected trace elements and isotopes, including their concentration, chemical speciation, and physical form, along a sufficient number of sections in each ocean basin to establish the principal relationships between these distributions and with more traditional hydrographic parameters;

* To evaluate the sources, sinks, and internal cycling of these species and thereby characterize more completely the physical, chemical and biological processes regulating their distributions, and the sensitivity of these processes to global change; and

* To understand the processes that control the concentrations of geochemical species used for proxies of the past environment, both in the water column and in the substrates that reflect the water column.

GEOTRACES will be global in scope, consisting of ocean sections complemented by regional process studies. Sections and process studies will combine fieldwork, laboratory experiments and modelling. Beyond realizing the scientific objectives identified above, a natural outcome of this work will be to build a community of marine scientists who understand the processes regulating trace element cycles sufficiently well to exploit this knowledge reliably in future interdisciplinary studies.

Expand "Projects" below for information about and data resulting from individual US GEOTRACES research projects.

[table of contents | back to top]

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-1260233
Gordon and Betty Moore Foundation: Marine Microbiology Initiative (MMI)	GBMF3782

[table of contents | back to top]