

Total spectral count of proteins from R/V Falkor cruise 160115 for the ProteOMZ expedition in the Central Pacific in 2016

Website: <https://www.bco-dmo.org/dataset/737620>

Data Type: Cruise Results

Version: 3

Version Date: 2022-06-06

Project

» [The ProteOMZ Expedition: Investigating Life Without Oxygen in the Pacific Ocean](#) (ProteOMZ (Proteomics in an Oxygen Minimum Zone))

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

Relative protein abundance data of the upper 1000m water column from the ProteOMZ R/V Falkor expedition. There are 56,577 protein identifications over 102 samples, for ~6 million data points.

Table of Contents

- [Coverage](#)
 - [Dataset Description](#)
 - [Data Processing Description](#)
 - [BCO-DMO Processing Description](#)
 - [Data Files](#)
 - [Related Datasets](#)
 - [Parameters](#)
 - [Instruments](#)
 - [Deployments](#)
 - [Project Information](#)
 - [Funding](#)
-

Coverage

Location: Central Pacific

Spatial Extent: N:17.4465 E:-139.1089 S:-0.4708 W:-157.3022

Temporal Extent: 2016-01-19 - 2016-01-28

Dataset Description

These data are part of the Ocean Protein Portal "ProteOMZ" dataset (<https://proteinportal.whoi.edu/>; Saito et al., 2019).

Data Processing Description

The raw mass spectra files were searched against SEQUEST within Proteome Discoverer v2.2 software. Processed files were then loaded into Proteome Software and protein and peptide reports as well as and fasta files were exported. The files were modified slightly to map to the Protein Portal data model for submission to BCO-DMO.

BCO-DMO Processing Description

Preprocessing:

- Date, time, filter min, filter max, lat, lon, and cruise columns were added based on information from the Falkor 160115 Event log and CTD log.
- Column names reformatted to comply with BCO-DMO standards.

Dataset version 3 revision (2019-02-24): replaces the earlier data version from date (2018-05-25)

- * Values "#N/A" changed to "Unknown" which has a different meaning than blank values. Unknown = "Protein functional and taxonomic annotations are marked as "Unknown" for protein sequences which did not have any significant hits to known reference sequences or motifs in the metagenome annotation database. "
- * Event log and McLane pump log were updated to fix lat/lon, date/time issues. Since these were sources of information in this dataset, this dataset is also being updated. See respective "processing notes" sections for these two logs for detailed information about changes to those data sources. event log: <https://www.bco-dmo.org/dataset/708384> pump log: <https://www.bco-dmo.org/dataset/708495>
- * ISO_DateTime_UTC timestamp added from Date and time columns in the McLane log dataset. Date and time columns were HST time zone so 10 hours were added to make the time in UTC.
- * columns for max and min pump filter size added (min 0.2 max 3.0).
- * some station and target depths for station and target depth combinations did not exist in the Mclane pump log so the missing values were added to the mclane pump log. The following columns in this dataset come from the pump log: cruise, cast, date, time, lat, lon, depth (ISO_DateTime_UTC is derived from local date and time). Where there were two matched casts in the pump log for a station and target depth, the first cast for the station was used. The only differences between the two possible casts were the date/time columns.

Dataset Version 3: This file revision (2022-06-06) replaces a previous revision of dataset version 1 from 2019-02-24. This is still dataset version 3 because previous revisions of version 3 were not made public.

* Data from source file "ProteOMZ_proteins_for_OPP.csv" was imported into the BCO-DMO data system for this dataset. This file "ProteOMZ_proteins_for_OPP.csv" is from Ocean Protein Portal "ProteOMZ" dataset v3 (file version 2022-06-06)

** In the BCO-DMO data system missing data identifiers are displayed according to the format of data you access. For example, in csv files it will be blank (null) values. In Matlab .mat files it will be NaN values. When viewing data online at BCO-DMO, the missing value will be shown as blank (null) values.

* Column names adjusted to conform to BCO-DMO naming conventions designed to support broad re-use by a variety of research tools and scripting languages. [Only numbers, letters, and underscores. Can not start with a number] e.g. date_y-m-d changed to date_ymd

* ISO DateTime with timezone (UTC) column added in ISO 8601 format from local date and times in HST.

* ".0" removed from ncbi_id values to correspond to the integer identifier at NCBI.

* Data table attached to dataset as Data File:"737620_v3_proteomz-proteins.csv"

[[table of contents](#) | [back to top](#)]

Data Files

File
737620_v3_proteomz-proteins.csv (Comma Separated Values (.csv), 261.46 MB) MD5:f9e2a796920a457f8e89752bd74012cc
Primary data file for dataset ID 737620, version 3

[[table of contents](#) | [back to top](#)]

Related Datasets

IsSupplementedBy

Saito, M. A. (2024) **Total spectral counts of peptides from the R/V Falkor cruise 160115 in the Central Pacific for the ProteOMZ expedition in 2016.** Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2022-06-03 doi:10.26008/1912/bco-dmo.737596.1 [[view at BCO-DMO](#)]

Relationship Description: The peptides related to these proteins

IsRelatedTo

Saito, M. A. (2024) **FASTA file of identified protein sequences from the R/V Falkor cruise 160115 for the ProteOMZ expedition in the Central Pacific during 2016.** Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 3) Version Date 2019-02-15 doi:10.26008/1912/bco-dmo.737611.3 [[view at BCO-DMO](#)]

Relationship Description: These datasets are part of the Ocean Protein Portal "ProteOMZ" dataset (<https://proteinportal.whoi.edu/>; Saito et al., 2019).

Saito, M. A., Saunders, J. (2022) **Relative protein abundance from scaled and corrected exclusive peptide spectral counts from the ProteOMZ R/V Falkor expedition cruise FK160115 in the Pelagic central Pacific Ocean in 2016.** Biological and Chemical Oceanography Data Management Office (BCO-DMO). (Version 1) Version Date 2022-01-13 doi:10.26008/1912/bco-dmo.868030.1 [[view at BCO-DMO](#)]

Relationship Description: These datasets are part of the Ocean Protein Portal "ProteOMZ" dataset (<https://proteinportal.whoi.edu/>; Saito et al., 2019).

[[table of contents](#) | [back to top](#)]

Parameters

Parameter	Description	Units
sample_id	Unique sample name for the specific filter collected (station/depth/version if applicable)	unitless
MS_MS_sample_name	Unique name for the mass spec sample and run	unitless
station_id	The identifier for the station	unitless
depth_m	Cast depth where sample was taken	meters
latitude_dd	Latitude of station	decimal degrees
longitude_dd	Longitude of station	decimal degrees
date_ymd	Date of sampling (local time zone HST)	unitless
time_hms	Time of sampling (local time zone HST)	unitless
minimum_filter_size_microns	Minimum size of the collection filter	microns
maximum_filter_size_microns	Maximum size of the collection filter	microns

cruise_id	The unique cruise identifier	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 each protein	count
molecular_weight_kDa	Molecular weight of the full length protein sequences	kDa
protein_name	Descriptive name of the function of the protein	unitless
ncbi_id	NCBI Taxonomy organism identifier (for ncbi_name)	taxon
ncbi_name	NCBI Taxonomy name (corresponding to ncbi_id)	verbatimIdentification
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	Description 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
ISO_DateTime_UTC	Datetime with timezone (UTC) of sampling in ISO 8601 format	unitless

[[table of contents](#) | [back to top](#)]

Instruments

Dataset-specific Instrument Name	Alpkem Autosampler
Generic Instrument Name	Alpkem RFA300
Dataset-specific Description	Used in nutrient analysis
Generic Instrument Description	A rapid flow analyser (RFA) that may be used to measure nutrient concentrations in seawater. It is an air-segmented, continuous flow instrument comprising a sampler, a peristaltic pump which simultaneously pumps samples, reagents and air bubbles through the system, analytical cartridge, heating bath, colorimeter, data station, and printer. The RFA-300 was a precursor to the smaller Alpkem RFA/2 (also RFA II or RFA-2).

Dataset-specific Instrument Name	SeaBird SBE19 CTD
Generic Instrument Name	CTD Sea-Bird
Dataset-specific Description	Used for water sampling
Generic Instrument Description	A Conductivity, Temperature, Depth (CTD) sensor package from SeaBird Electronics. This instrument designation is used when specific make and model are not known or when a more specific term is not available in the BCO-DMO vocabulary. Refer to the dataset-specific metadata for more information about the specific CTD used. More information from: http://www.seabird.com/

Dataset-specific Instrument Name	Technicon AutoAnalyzer II
Generic Instrument Name	Technicon AutoAnalyzer II
Dataset-specific Description	Used to measure phosphate and ammonium
Generic Instrument Description	A rapid flow analyzer that may be used to measure nutrient concentrations in seawater. It is a continuous segmented flow instrument consisting of a sampler, peristaltic pump, analytical cartridge, heating bath, and colorimeter. See more information about this instrument from the manufacturer.

Dataset-specific Instrument Name	Trace Metal Rosette
Generic Instrument Name	Trace Metal Bottle
Dataset-specific Description	Used for nutrient sampling
Generic Instrument Description	Trace metal (TM) clean rosette bottle used for collecting trace metal clean seawater samples.

Deployments

FK160115

Website	https://www.bco-dmo.org/deployment/708387
Platform	R/V Falkor
Report	https://service.rvdata.us/data/cruise/FK160115/doc/FK160115_OfficialCruiseReport_Saito_v3.pdf
Start Date	2016-01-16
End Date	2016-02-11
Description	Project: Using Proteomics to Understand Oxygen Minimum Zones (ProteOMZ) More information is available from the ship operator at https://schmidtocean.org/cruise/investigating-life-without-oxygen-in-the... Additional cruise information is available from the Rolling Deck to Repository (R2R): https://www.rvdata.us/search/cruise/FK160115

[[table of contents](#) | [back to top](#)]

Project Information

The ProteOMZ Expedition: Investigating Life Without Oxygen in the Pacific Ocean (ProteOMZ (Proteomics in an Oxygen Minimum Zone))

Website: <https://schmidtocean.org/cruise/investigating-life-without-oxygen-in-the-tropical-pacific/#team>

Coverage: Central Pacific Ocean (Hawaii to Tahiti)

From Schmidt Ocean Institute's ProteOMZ Project page:

Rising temperatures, ocean acidification, and overfishing have now gained widespread notoriety as human-caused phenomena that are changing our seas. In recent years, scientists have increasingly recognized that there is yet another ingredient in that deleterious mix: a process called deoxygenation that results in less oxygen available in our seas.

Large-scale ocean circulation naturally results in low-oxygen areas of the ocean called oxygen deficient zones (ODZs). The cycling of carbon and nutrients – the foundation of marine life, called biogeochemistry – is fundamentally different in ODZs than in oxygen-rich areas. Because researchers think deoxygenation will greatly expand the total area of ODZs over the next 100 years, studying how these areas function now is important in predicting and understanding the oceans of the future. This first expedition of 2016 led by Dr. Mak Saito from the Woods Hole Oceanographic Institution (WHOI) along with scientists from University of Maryland Center for Environmental Science, University of California Santa Cruz, and University of Washington aimed to do just that, investigate ODZs.

During the 28 day voyage named “ProteOMZ,” researchers aboard R/V *Falkor* traveled from Honolulu, Hawaii to Tahiti to describe the biogeochemical processes that occur within this particular swath of the ocean’s ODZs. By doing so, they contributed to our greater understanding of ODZs, gathered a database of baseline measurements to which future measurements can be compared, and established a new methodology that could be used in future research on these expanding ODZs.

[[table of contents](#) | [back to top](#)]

Funding

Funding Source	Award
Gordon and Betty Moore Foundation: Marine Microbiology Initiative (MMI)	GBMF3782
Schmidt Ocean Institute (SOI)	R/V Falkor 160115 SOI ProteOMZ Expedition

[[table of contents](#) | [back to top](#)]