Shotgun Metaproteomics of Bering Strait surface water and Chukchi Sea bottom water.

Website: https://www.bco-dmo.org/dataset/719067
Data Type: experimental
Version:
Version Date: 2017-11-13

Project
» Collaborative Research: Linking geochemistry and proteomics to reveal the impact of bacteria on protein cycling in the ocean (Bacterial Recyclers)

Contributors	Affiliation	Role
Noble, William S.	University of Washington (UW)	Principal Investigator
Nunn, Brook L.	University of Washington (UW)	Principal Investigator
Switzer, Megan	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

Table of Contents

- Dataset Description
  - Methods & Sampling
  - Data Processing Description
- Data Files
- Parameters
- Instruments
- Project Information
- Funding

Dataset Description

Location: Water samples were collected in August of 2013 from the Bering Strait (BSt) chlorophyll maximum layer (7 m depth, 65°43.44″ N, 168°57.42″ W) and from the more northern Chukchi Sea (CS) bottom waters (55.5 m depth, 72°47.624″ N, 16°53.89″ W) using a 24-bottle CTD (conductivity, temperature, and depth) rosette (10 L General Oceanics Niskin X). The measurement of integrated water column chlorophyll was 226.88 mg/m2 at station BSt and 2.64 mg/m2 at station CS.

Water was collected on ship, filtered, and bacterial fractions were lysed, digested and analyzed using proteomic mass spectrometry.

Cruise = BEST Cruise 2013

Data are available for download at the EBI PRIDE Archive. Project number = PXD006472.

Homepage: http://www.ebi.ac.uk/pride/archive
Project URL: http://www.ebi.ac.uk/pride/archive/projects/PXD006472
Data URL: http://www.ebi.ac.uk/pride/archive/projects/PXD006472/files

Data are published in May, D.H., Timmins-Schiffman, E., Mikan, M.P., Harvey, H.R., Borenstein, E., Nunn, B.L., Noble, W.S. (2016) An alignment-free "metapeptide" strategy for metaproteomic characterization of microbiome samples using shotgun metagenomic sequencing. Journal of Proteome Research 15, 2697-2705. DOI: 10.1021/acs.jproteome.6b00239

Methods & Sampling

Water samples were collected in August of 2013 from the Bering Strait (BSt) chlorophyll maximum layer (7 m depth, 65°43.44″ N, 168°57.42″ W) and from the more northern Chukchi Sea (CS) bottom waters (55.5 m

depth, 72°47.624″ N, 16°53.89″ W) using a 24-bottle CTD (conductivity, temperature, and depth) rosette (10 L General Oceanics Niskin X). The measurement of integrated water column chlorophyll was 226.88 mg/m2 at station BSt and 2.64 mg/m2 at station CS. As our previous work has shown, to examine bacterial contributions, it is essential to remove the very high background contribution from algal inhabitants.(23) Also, oceanic marine bacteria are typically smaller than bacteria in gut biomes or freshwater systems, with the majority passing a 1.0 μm filter.(24, 25) Accordingly, a 15 L water sample was prefiltered through two high-volume cartridges (10 μm and then 1 μm) to remove larger eukaryotes, and the filtrate comprising the bacterial microbiome was then collected on a glass fiber filter (GF/F) with nominal pore size of 0.7 μm. Filters were flash frozen and stored at −80 °C until extraction. Filters were sliced, and

GF/F filters with the bacterial fraction were placed in 1.5 mL tubes with 100 μL of 0.5 mm glass beads, 100 μL of 6 M urea, and 500 μL of nanopure water. Filters were shaken on a bead beater for 1 min and then placed in ice for 5 min. This process was repeated 10 times to ensure cell lysis and filter breakup. A needle was then heated by flame and used to create a <0.5 mm hole at the bottom of the 1.5 mL sample tube. The sample tubes were then placed atop an open 1.5 mL tube and centrifuged (3000g, 10 min). This process was completed to isolate protein lysate from extracted particles and glass beads. Protein concentrations were determined using BCA colormetric assay; 100 μg of total protein was used for digestion. Each 100 μg protein sample received 300 ng of purified human ApoA1 to monitor protein digesion. Samples were reduced, alkylated, enzymatically digested with trypsin, and desalted. Prior to MS injections, 50 fmol of the Pierce Peptide Retention Time Standard (ThermoFisher Scientific) was added to each autosample vial at 50 fmol per 2 μg of total protein.

Peptides were separated using an inline NanoAquity HPLC with a 4 cm precolumn (5 μm; 200A; Magic C18) and 30 cm Reprosil-Pur Basic 3 μm C18 analytical column (Dr. Maisch GmbH, Germany). Peptides were eluted using a 2–30% ACN, 0.1% formic acid nonlinear gradient in 120 min at 300 nL/min. LC-MS/MS was performed with a Q-Exactive-HF (ThermoScientific) on technical triplicates for each sample. The instrument was operated in Top 20 data-dependent acquisition mode, collecting data on 400–1600 m/z range with a 5 s dynamic exclusion.

Data Processing Description

All computation was performed on a Univa Grid Engine cluster with 1.90 GHz AMD Opteron processors. The MOCAT pipeline was used to assemble a metagenome and predict genes as follows. Trimmed and filtered reads from both BSt and CS samples were aligned to the human hg19 reference using SOAPaligner v2.21, and aligned reads were removed. The remaining reads were assembled into contigs and scaftigs with SOAPdenovo v1.06. The assembly was revised, correcting for indels and chimeric regions, with SOAPdenovo v1.06 and BWA v0.7.5a-r16. Genes were predicted using Prodigal v2.60. We used three well-established gene fragment prediction tools to predict gene fragments directly from shotgun metagenomic sequencing reads from each sample: MetaGeneAnnotator (in multiple species mode), FragGeneScan version 1.2.0 (illumina_10 model parameters), and Orphelia (with Net300 prediction model). Separate metapeptide databases were constructed from the BSt and CS sequencing runs, from either predicted gene fragments or raw read sequences. When starting from raw read sequences, each read was translated in all six reading frames, and reading frames containing a stop codon were discarded. The results described in section 3 were obtained by starting with predicted gene fragments from MetaGeneAnnotator. Whether starting from gene fragments or raw read sequences, amino acid sequences from each nucleotide sequence were trimmed to the first and last tryptic cleavage site (or discarded if fewer than two sites), and the remaining ends were discarded. This was done in order to remove partial tryptic peptide sequences that are unlikely to be detected by LC-MS/MS of a trypsinized metaproteome. The resulting candidate sequences were discarded if they were less than 10 amino acids long, if they contained no tryptic peptides with seven or more amino acids, or if the minimum Phred quality score over the length of the sequence was less than 30. Finally, metapeptide candidates meeting all the above criteria were discarded if they were represented by fewer than two reads. A FASTA database was constructed from the remaining metapeptides. For purposes of comparison, we also made use of a metagenome-derived database of translated genes from the metagenome described above and the NCBI nonredundant database of protein sequences from large environmental sequencing projects (‘env_nr’, downloaded from ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/env_nr.gz on December 1, 2015). All database searches were performed using Comet version 2015.01 rev. 2, using a concatenated decoy database in which peptide sequences were reversed but C-terminal amino acids were left in place. Search parameters included a static modification for cysteine carbamidomethylation (57.021464) and a variable modification for methionine oxidation (15.9949). Enzyme specificity was trypsin, with one missed cleavage allowed. Parent ion mass tolerance was set to 10 ppm around five isotopic peaks, and fragment ion binning was 0.02, with offset 0.0.

Peptide-spectrum matches (PSMs) from all technical replicates were combined into a single data set. As described previously, after each unique peptide was associated with its top-scoring spectrum, irrespective of charge state, we used the widely used target–decoy search strategy of estimating the false discovery rate (FDR) associated with a given set of accepted peptides. In this context, the FDR is defined as the proportion of the accepted peptides that are not responsible for generating observed spectra. We then empirically examined the trade-off between FDR and the number of accepted peptides, since in practice the mass spectrometrist is typically interested in accepting as many peptides as possible while maintaining an acceptable FDR. Note that this trade-off is similar to the distinction between precision (1 – FDR) and recall or sensitivity. Results of searches of individual samples against multiple databases were integrated as follows. PSMs from searches against all databases were combined into a single tab-delimited file of features for input to Percolator. For each database, a new binary feature was added to the combined feature file indicating whether the PSM was derived from a search against that database. Percolator was then used to analyze the combined set, thereby computing a discriminant score for each PSM. For each scan with multiple PSMs (from multiple databases), all but the highest-scoring PSM were removed. Peptide-level FDR was then calculated as described above.

May, D.H., Timmins-Schiffman, E., Mikan, M.P., Harvey, H.R., Borenstein, E., Nunn, B.L., Noble, W.S. (2016) An alignment-free "metapeptide" strategy for metaproteomic characterization of microbiome samples using shotgun metagenomic sequencing. Journal of Proteome Research 15, 2697-2705.

[table of contents | back to top]

Data Files

File
Metaproteomics_PRIDE_2192_May.csv (Comma Separated Values (.csv), 210 bytes) MD5:5952603845f6489204ebe7e3da80de2e Primary data file for dataset ID 719067

[table of contents | back to top]

Parameters

Parameter	Description	Units
Repository	Name of database where data are currently served	unitless
Project	Unique project identifier for the database where data are currently served	unitless
Experiment_ID	unknown	unitless
Project_URL	Link to project page where data are currently served.	unitless

[table of contents | back to top]

Instruments

Dataset-specific Instrument Name	10L General Oceanics Niskin X
Generic Instrument Name	CTD - profiler
Generic Instrument Description	The Conductivity, Temperature, Depth (CTD) unit is an integrated instrument package designed to measure the conductivity, temperature, and pressure (depth) of the water column. The instrument is lowered via cable through the water column. It permits scientists to observe the physical properties in real-time via a conducting cable, which is typically connected to a CTD to a deck unit and computer on a ship. The CTD is often configured with additional optional sensors including fluorometers, transmissometers and/or radiometers. It is often combined with a Rosette of water sampling bottles (e.g. Niskin, GO-FLO) for collecting discrete water samples during the cast. This term applies to profiling CTDs. For fixed CTDs, see https://www.bco-dmo.org/instrument/869934.

Dataset-specific Instrument Name	Thermo Scientific Q-Exactive-HF
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]

Project Information

Collaborative Research: Linking geochemistry and proteomics to reveal the impact of bacteria on protein cycling in the ocean (Bacterial Recyclers)

Text from NSF award abstract:

Although proteins represent the primary source of new organic nitrogen in the ocean, the identification of individual proteins and mechanisms modulating their preservation has faced analytical and computational challenges in deciphering the vast suite of possible sequences and degradation by-products. Recent efforts to link geochemical cycling, biomedical proteomics and bioinformatics has demonstrated that only a small subset of the suite of proteins produced by marine diatoms appear to survive the degradation process, and those that do are largely protected by physical and enthalpic barriers to microbial attack. Although these discoveries help to explain the survival of individual proteins, they also generate multiple questions regarding bacteria as the dominant recyclers of organic nitrogen and carbon and needs for specific approaches to characterize modified protein products. Bacteria dominate the water column and sedimentary systems in both numbers and diversity, yet their relative contribution to the preserved proteomic pool appears low.

In this project, researchers at Old Dominion Universityand the University of Washington will join forces to decipher the bacterial role in protein recycling and their potential contribution. By integrating high mass accuracy tandem mass spectrometry-based proteomics with stable isotope-based geochemical analysis, they hope to identify those bacterial proteins initially synthesized during organic matter recycling. Three research objectives drive this investigation: (1) to determine the potential contribution of bacteria proteins to marine organic matter; (2) to identify those protein(s) synthesized by heterotrophic marine bacteria during initial stages of organic matter degradation; (3) to determine if glycan (carbohydrate) modifications represent an important component of preserved, yet unidentified, peptides seen in our analysis of oceanic particles and sediments.

Broader Impacts: This project will provide multiple opportunities for interdisciplinary student training in marine chemistry and proteomics as well as address the goal of disseminating results and tools to a broad audience. In the more traditional role, this project will expand the career for a female principal investigator in marine proteomics, support both graduate and undergraduate students at ODU which include opportunities for minority enrichment and provide training for a postdoctoral fellow at UW. On the broader level, the ODU PI participates in high school outreach programs for high achieving students in the local school which provides for summer internships and enrichment programs.

Relevant Links:

Old Dominion University: Marine Organic Geochemistry and Ecology Laboratory (MOGEL) Lab Website
Bering Sea Ecosystem Study: Data Archive
Environmental Proteomics: Bacteria Recyclers in the Ocean
Environmental Proteomics: Proteomics of Colwellia psychretheca at subzero temperatures

[table of contents | back to top]

Funding

Funding Source	Award
NSF Division of Ocean Sciences (NSF OCE)	OCE-1233589

[table of contents | back to top]