

# Microbial eukaryotic focused metatranscriptome data from seawater collected in coastal California in May of 2015

**Website:** <https://www.bco-dmo.org/dataset/745518>

**Data Type:** Other Field Results

**Version:** 2

**Version Date:** 2018-10-15

## Project

» [Protistan, prokaryotic, and viral processes at the San Pedro Ocean Time-series](#) (SPOT)

Contributors	Affiliation	Role
<a href="#">Caron, David</a>	University of Southern California (USC)	Principal Investigator
<a href="#">Hu, Sarah K.</a>	University of Southern California (USC)	Co-Principal Investigator, Contact
<a href="#">York, Amber D.</a>	Woods Hole Oceanographic Institution (WHOI BCO-DMO)	BCO-DMO Data Manager

## Abstract

Seawater was collected via Niskin bottles mounted with a CTD from the San Pedro Ocean Time-series (SPOT) station off the coast of Southern California near the surface (5 m), 150 and 890 m, in late May 2015. Raw sequence data was generated as part of a metatranscriptome study targeting the protistan community. Raw sequences are available at the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database (SRA Study ID: SRP110974, BioProject: PRJNA391503). Sequences for BioProject PRJNA608423 will be available at NCBI on Jan 1st, 2021. These data were published in Hu et al. (2018).

## Table of Contents

- [Coverage](#)
- [Dataset Description](#)
  - [Methods & Sampling](#)
  - [Data Processing Description](#)
  - [BCO-DMO Processing Description](#)
- [Data Files](#)
- [Supplemental Files](#)
- [Related Publications](#)
- [Parameters](#)
- [Instruments](#)
- [Deployments](#)
- [Project Information](#)
- [Funding](#)

## Coverage

**Spatial Extent:** Lat:33.55 Lon:-118.4

**Temporal Extent:** 2015-05-20

## Dataset Description

Seawater was collected via Niskin bottles mounted with a CTD from the San Pedro Ocean Time-series (SPOT) station off the coast of Southern California near the surface (5 m), 150 and 890 m, in late May 2015. Raw sequence data was generated as part of a metatranscriptome study targeting the protistan community. Raw sequences are available at the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database (SRA Study ID: SRP110974, BioProject: PRJNA391503). These data were published in Hu et al. (2018).

## Methods & Sampling

Seawater was collected from the San Pedro Ocean Time-series (SPOT) station off the coast of Southern California near the surface (5 m), 150 and 890 m, in late May 2015. Briefly, seawater was pre-filtered (80 mm) into 20 L carboys to minimize the presence of multicellular eukaryotes. Replicate samples (ranging in volume from 1.5-3.5 L) from each depth were filtered onto sterile GF/F filters (nominal pore size 0.7 mm, Whatman, International Ltd. Florham Park, NJ). While we cannot avoid some impact that sample handling (i.e., bringing samples to the surface) may have had on our results, filters were immediately placed in 1.5 mL of lysis buffer and flash frozen in liquid nitrogen in < 40 min and away from light to minimize RNA degradation.

Total RNA was extracted from each filter using a DNA/RNA AllPrep kit (Qiagen, Valencia, CA, #80204) with an in-line genomic DNA removal step (RNase-free DNase reagents, Qiagen #79254) ([dx.doi.org/10.17504/protocols.io.hk3b4yn](https://doi.org/10.17504/protocols.io.hk3b4yn)). Extracted RNA was quality checked and low biomass samples were pooled. Six replicates were processed and sequenced from the surface, while pairs of filters were pooled for either 150 or 890 m, yielding 3 and 4 replicates respectively (Supporting Information Table S1). RNA concentrations were normalized before library preparation (Supporting Information). ERCC spike-in was added before sequence library preparation with Kapa's Stranded mRNA library preparation kit using poly-A tail selection beads to select for eukaryotic mRNA (Kapa Biosystems, Inc., Wilmington, MA, #KK8420).

Also see:

<https://www.protocols.io/view/sample-collection-from-the-field-for-downstream-mo-hisb4ee><https://www.protocols.io/view/rna-and-optional-dna-extraction-from-environmental-hk3b4yn>

The associated assembly files can be found at Zenodo (see Hu, S. K. (2017), DOI: 10.5281/zenodo.1202041). The assembly files were also published in the journal publication Hu, et al. (2018).

Related code can be found in the github repository [https://github.com/shu251/SPOT\\_metatranscriptome](https://github.com/shu251/SPOT_metatranscriptome). The version of the code used for these publications can be found in the Supplemental Files section of this page.

## Data Processing Description

Sequence adapters, low quality (phred score < 10, from 5' and 3' ends, and within a 25 bp sliding window) or short sequences (< 50 bps), and sequences containing more than 50 consecutive As or Ts were removed using Trimmomatic v. 0.32 (Bolger et al., 2014). All quality trimmed sequences were aligned to ERCC sequences using 'align\_and\_estimate\_abundance.pl' in the Trinity v. 2.1.1 (Grabherr et al., 2011) package. Reads that were aligned to ERCC sequences were removed using a custom PERL script (available: [https://github.com/shu251/SPOT\\_metatranscriptome](https://github.com/shu251/SPOT_metatranscriptome)).

## BCO-DMO Processing Description

[Changes discussed with and reviewed by the data submitter]

- \* added a conventional header with dataset name, PI name, version date
- \* modified parameter names to conform with BCO-DMO naming conventions
- \* blank values in this dataset are displayed as "nd" for "no data." nd is the default missing data identifier in the BCO-DMO system.
- \* removed columns "object\_status" and blank columns "filename3" and "filename4."
- \* Added column SRA\_ID\_link to the SRA run at NCBI
- \* removed column "assembly" which had values of "See related publication for access to assembly files" This information was included in the Methods & Sampling section of the methodology and further explained.
- \* For curatorial purposes BCO-DMO forked the github code repository [https://github.com/shu251/SPOT\\_metatranscriptome](https://github.com/shu251/SPOT_metatranscriptome) and created a github release (see [https://github.com/BCODMO/SPOT\\_metatranscriptome/releases/tag/bcodmo\\_v1](https://github.com/BCODMO/SPOT_metatranscriptome/releases/tag/bcodmo_v1)). The release .zip file was downloaded to BCO-DMO's servers and added to the dataset landing page as a supplemental file to satisfy NSF OCE sharing requirements.
- \* changes in version 2: data for bioproject PRJNA608423 added to the dataset.

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

---

## Data Files

File
<b>metaT.csv</b> (Comma Separated Values (.csv), 11.80 KB) MD5:cbec0cc726f0b64619a035aba3abdbfa
Primary data file for dataset ID 745518

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

---

## Supplemental Files

File
<b>SPOT metatranscriptome code from Github, release bco-dmo_v1</b> filename: SPOT_metatranscriptome-bcodmo_v1.zip (ZIP Archive (ZIP), 810.06 KB) MD5:369ccce50b2d833723c7f9ea7607e78d
Zip file containing required code for data compilation and analysis for a eukaryotic-focused metatranscriptome survey in the North Pacific. This is release "bcodmo_v1" <a href="https://github.com/BCODMO/SPOT_metatranscriptome/tree/bcodmo_v1">https://github.com/BCODMO/SPOT_metatranscriptome/tree/bcodmo_v1</a> .

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

---

## Related Publications

Hu, S. (2017). RNA (and optional DNA) extraction from environmental samples (filters) v2 (protocols.io.hk3b4yn). Protocols.io. doi:[10.17504/protocols.io.hk3b4yn](https://doi.org/10.17504/protocols.io.hk3b4yn)  
*Methods*

Hu, S. (2017). Sample collection from the field for downstream molecular analysis - microbial eukaryote-focused v1. Protocols.io. doi:[10.17504/protocols.io.hisb4ee](https://doi.org/10.17504/protocols.io.hisb4ee)  
*Methods*

Hu, S. (2018) Required code for data compilation and analysis for a eukaryotic-focused metatranscriptome survey in the North Pacific. SPOT\_metatranscriptome version bcodmo\_v1. GitHub. Date accessed 2018-10-15. [https://github.com/BCODMO/SPOT\\_metatranscriptome/tree/bcodmo\\_v1](https://github.com/BCODMO/SPOT_metatranscriptome/tree/bcodmo_v1)  
*Software*

Hu, S. K. (2017). Shifting Metabolic Priorities Among Key Protistan Taxa Within And Below The Euphotic Zone [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.1202041>  
*Software*

Hu, S. K., Liu, Z., Alexander, H., Campbell, V., Connell, P. E., Dyhrman, S. T., ... Caron, D. A. (2018). Shifting metabolic priorities among key protistan taxa within and below the euphotic zone. Environmental Microbiology. doi:[10.1111/1462-2920.14259](https://doi.org/10.1111/1462-2920.14259)  
*Results*

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

---

## Parameters

Parameter	Description	Units
SRA_run	SRA Run identifier at NCBI	unitless
SRA_run_link	URL for SRA Run Page at NCBI	unitless
SRA_study	SRA study identifier at NCBI	unitless
bioproject_accession	BioProject accession number at NCBI	unitless
biosample_accession	BioSample accession number at NCBI	unitless
library_ID	SRA title	unitless
title	Descriptive title of SRA accession	unitless
sample_name	Sample name	unitless
library_strategy	Library strategy ("AMPLICON")	unitless
library_source	Library source ("TRANSCRIPTOMIC" or "GENOMIC")	unitless
library_selection	Library selection ("PCR")	unitless
library_layout	Library layout ("paired")	unitless
platform	Sequencing platform ("Illumina")	unitless
instrument_model	Sequencing instrument model ("Illumina MiSeq")	unitless
design_description	Sequencing design description	unitless
filetype	Type of files	unitless
filename	Name of file 1 (see NCBI for access)	unitless
filename2	Name of file 2 (see NCBI for access)	unitless

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

## Instruments

<b>Dataset-specific Instrument Name</b>	HiSeq
<b>Generic Instrument Name</b>	Automated DNA Sequencer
<b>Dataset-specific Description</b>	HiSeq High Output 125 bp PE sequencing was performed at UPC Genome Core at University of Southern California, Los Angeles, CA (BioProject: PRJNA391503).
<b>Generic Instrument Description</b>	General term for a laboratory instrument used for deciphering the order of bases in a strand of DNA. Sanger sequencers detect fluorescence from different dyes that are used to identify the A, C, G, and T extension reactions. Contemporary or Pyrosequencer methods are based on detecting the activity of DNA polymerase (a DNA synthesizing enzyme) with another chemoluminescent enzyme. Essentially, the method allows sequencing of a single strand of DNA by synthesizing the complementary strand along it, one base pair at a time, and detecting which base was actually added at each step.

<b>Dataset-specific Instrument Name</b>	
<b>Generic Instrument Name</b>	Niskin bottle
<b>Generic Instrument Description</b>	A Niskin bottle (a next generation water sampler based on the Nansen bottle) is a cylindrical, non-metallic water collection device with stoppers at both ends. The bottles can be attached individually on a hydrowire or deployed in 12, 24, or 36 bottle Rosette systems mounted on a frame and combined with a CTD. Niskin bottles are used to collect discrete water samples for a range of measurements including pigments, nutrients, plankton, etc.

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

## Deployments

### SPOT\_Yellowfin\_Cruises

<b>Website</b>	<a href="https://www.bco-dmo.org/deployment/754348">https://www.bco-dmo.org/deployment/754348</a>
<b>Platform</b>	R/V Yellowfin
<b>Start Date</b>	2005-01-19
<b>End Date</b>	2018-07-18
<b>Description</b>	San Pedro Ocean Time Series (SPOT) station (33°33'N, 118°24'W) R/V Yellowfin, monthly SPOT cruises in the San Pedro Channel Deployment: SPOT Platform: RV Yellowfin Platform Type: vessel

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

## Project Information

### Protistan, prokaryotic, and viral processes at the San Pedro Ocean Time-series (SPOT)

**Coverage:** San Pedro Channel off the coast of Los Angeles

Planktonic marine microbial communities consist of a diverse collection of bacteria, archaea, viruses, protists (phytoplankton and protozoa) and small animals (metazoan). Collectively, these species are responsible for virtually all marine pelagic primary production where they form the basis of food webs and carry out a large fraction of respiratory processes. Microbial interactions include the traditional role of predation, but recent research recognizes the importance of parasitism, symbiosis and viral infection. Characterizing the response of pelagic microbial communities and processes to environmental influences is fundamental to understanding and modeling carbon flow and energy utilization in the ocean, but very few studies have attempted to study all of these assemblages in the same study. This project is comprised of long-term (monthly) and short-term (daily) sampling at the San Pedro Ocean Time-series (SPOT) site. Analysis of the resulting datasets investigates co-occurrence patterns of microbial taxa (e.g. protist-virus and protist-prokaryote interactions, both positive and negative) indicating which species consistently co-occur and potentially interact, followed by examination gene expression to help define the underlying mechanisms. This study augments 20 years of baseline studies of microbial abundance, diversity, rates at the site, and will enable detection of low-frequency changes in composition and potential ecological interactions among microbes, and their responses to changing environmental forcing factors. These responses have important consequences for higher trophic levels and ocean-atmosphere feedbacks. The broader impacts of this project include training graduate and undergraduate students, providing local high school student with summer lab experiences, and PI presentations at local K-12 schools, museums, aquaria and informal learning centers in the region. Additionally, the PIs advise at the local, county and state level regarding coastal marine water quality.

This research project is unique in that it is a holistic study (including all microbes from viruses to small metazoa) of microbial species diversity and ecological activities, carried out at the SPOT site off the coast of southern California. In studying all microbes simultaneously, this work aims to identify important ecological interactions among microbial species, and identify the basis(es) for those interactions. This research involves (1) extensive analyses of prokaryote (archaeal and bacterial) and eukaryote (protistan and micro-metazoan) diversity via the sequencing of marker genes, (2) studies of whole-community gene expression by eukaryotes and prokaryotes in order to identify key functional characteristics of microorganismal groups and the detection of active viral infections, and (3) metagenomic analysis of viruses and bacteria to aid interpretation of transcriptomic analyses using genome-encoded information. The project includes exploratory metatranscriptomic analysis of poorly-understood aphotic and hypoxic-zone protists, to examine their stratification, functions and hypothesized prokaryotic symbioses.

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

---

## Funding

Funding Source	Award
<a href="#">NSF Division of Ocean Sciences (NSF OCE)</a>	<a href="#">OCE-1737409</a>

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