

# Operational taxonomic unit (OTU) table for 18S rRNA gene tag sequences from DNA and RNA from samples collected in coastal California in 2013 and 2014

**Website:** <https://www.bco-dmo.org/dataset/748064>

**Data Type:** Other Field Results

**Version:** 1

**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

This dataset is a raw output operational taxonomic unit (OTU) table generated by processing and clustering raw 18S rRNA gene tag sequences from extracted DNA and RNA. Columns represent samples, including month sampled, material (either extracted RNA or DNA), and depth (in meters); thus values in each column represent the number of sequences in that sample that belong to a given OTU (OTUs by row). Each row represents a single OTU. The last column lists the taxonomic identifier assigned to each OTU. The raw sequence data can be found in the NCBI SRA database under accession number SRP070577 with the associated BioProject PRJNA311248.

## Table of Contents

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

## Coverage

**Spatial Extent:** N:33.7125 E:-118.259167 S:33.452833 W:-118.475167

**Temporal Extent:** 2013-04-24 - 2014-01-15

## Dataset Description

This dataset is a raw output operational taxonomic unit (OTU) table generated by processing and clustering raw 18S rRNA gene tag sequences from extracted DNA and RNA. Columns represent samples, including month sampled, material (either extracted RNA or DNA), and depth (in meters); thus values in each column represent the number of sequences in that sample that belong to a given OTU (OTUs by row). Each row represents a single OTU. The last column lists the taxonomic identifier assigned to each OTU. The raw sequence data can be found in the NCBI SRA database under accession number SRP070577 with the associated BioProject PRJNA311248. Metadata for these sequences can be found in the dataset: "18S rRNA gene tag sequences from DNA and RNA": <https://www.bco-dmo.org/dataset/745527>

## Methods & Sampling

These data were published in Hu et al., 2016.

This dataset is a raw output operational taxonomic unit (OTU) table generated by processing and clustering raw 18S rRNA gene tag sequences from DNA and RNA. The numbers in each column represent the number of sequences from that sample belonging to a given OTU (row), with the last column listing the taxonomic ID assigned to each OTU. The raw sequence data can be found in the NCBI SRA database under accession number [SRP070577](#) with the associated BioProject [PRJNA311248](#). Metadata for these sequences can be found in the dataset:

"18S rRNA gene tag sequences from DNA and RNA": <https://www.bco-dmo.org/dataset/745527>

Nucleotide bases with a Q score lower than 20 for the last 30 bp of each sequence were trimmed. Paired-end sequences were merged using FLASH (Magoc and Salzberg 2011) with a minimum of 10 bp and maximum of 150 bp overlap between each sequence pair. Sequences shorter than 350 bp, longer than 460 bp, or which had an average quality score lower than 25 were discarded using QIIME v1.8 (Caporaso *et al.* 2010). Chimeric sequences were identified and removed, by either *de novo* or reference-based chimera checking (identify chimeric seqs.py in QIIME, intersection method).

The code release v2 associated with this version of the dataset can be downloaded as a .zip file from the Supplemental Documents section of this page. Future code updates will be accessible from the GitHub repository [https://github.com/shu251/V4\\_tagsequencing\\_18Sdiversity\\_q1](https://github.com/shu251/V4_tagsequencing_18Sdiversity_q1).

## Data Processing Description

BCO-DMO Data Manager Processing Notes:

- \* data extracted from xlsx sheet to csv
- \* 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.

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

## Data Files

File
<b>otu_table.csv</b> (Comma Separated Values (.csv), 3.08 MB) MD5:52da5f36a7ab848ec625ca7efd75d764 Primary data file for dataset ID 748064

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

## Parameters

Parameter	Description	Units
OTU_ID	Taxonomic designations called Operational Taxonomic Units	unitless
April_150m_DNA	DNA sequences from April at 150m depth at SPOT	unitless
July_DCM_DNA	DNA sequences from July at the DCM at SPOT	unitless

April_890m_DNA	DNA sequences from April at 890m depth at SPOT	unitless
Oct_DCM_DNA	DNA sequences from Oct at the DCM at SPOT	unitless
Oct_150m_DNA	DNA sequences from Oct at 150m depth at SPOT	unitless
July_150m_DNA	DNA sequences from July at 150m depth at SPOT	unitless
April_150m_cDNA	RNA(cDNA) sequences from April at 150m depth at SPOT	unitless
July_150m_cDNA	RNA(cDNA) sequences from July at 150m depth at SPOT	unitless
Jan_DCM_DNA	DNA sequences from Jan at the DCM at SPOT	unitless
Jan_150m_DNA	DNA sequences from Jan at 150m depth at SPOT	unitless
April_CAT_DNA	DNA sequences from April at the surface at Catalina Island	unitless
April_5m_DNA	DNA sequences from April at 5m depth at SPOT	unitless
July_POLA_DNA	DNA sequences from July at the surface at the Port of Los Angeles	unitless
July_CAT_DNA	DNA sequences from July at the surface at Catalina Island	unitless
July_5m_DNA	DNA sequences from July at 5m depth at SPOT	unitless
April_5m_cDNA	RNA(cDNA) sequences from April at 5m depth at SPOT	unitless
July_CAT_cDNA	RNA(cDNA) sequences from July at the surface at Catalina Island	unitless
July_5m_cDNA	RNA(cDNA) sequences from July at 5m depth at SPOT	unitless
Oct_CAT_DNA	DNA sequences from Oct at the surface at Catalina Island	unitless
April_CAT_cDNA	RNA(cDNA) sequences from April at the surface at Catalina Island	unitless
Oct_POLA_DNA	DNA sequences from Oct at the surface at the Port of Los Angeles	unitless

Oct_5m_DNA	DNA sequences from Oct at 5m depth at SPOT	unitless
Oct_CAT_cDNA	RNA(cDNA) sequences from Oct at the surface at Catalina Island	unitless
Oct_5m_cDNA	RNA(cDNA) sequences from Oct at 5m depth at SPOT	unitless
Jan_CAT_DNA	DNA sequences from Jan at the surface at Catalina Island	unitless
Jan_5m_DNA	DNA sequences from Jan at 5m depth at SPOT	unitless
Jan_POLA_DNA	DNA sequences from Jan at the surface at the Port of Los Angeles	unitless
Jan_5m_cDNA	RNA(cDNA) sequences from Jan at 5m depth at SPOT	unitless
April_DCM_DNA	DNA sequences from April at the DCM at SPOT	unitless
Jan_CAT_cDNA	RNA(cDNA) sequences from Jan at the surface at Catalina Island	unitless
July_DCM_cDNA	RNA(cDNA) sequences from July at the DCM at SPOT	unitless
April_DCM_cDNA	RNA(cDNA) sequences from April at the DCM at SPOT	unitless
Oct_150m_cDNA	RNA(cDNA) sequences from Oct at 150m depth at SPOT	unitless
Jan_DCM_cDNA	RNA(cDNA) sequences from Jan at the DCM at SPOT	unitless
Jan_150m_cDNA	RNA(cDNA) sequences from Jan at 150m depth at SPOT	unitless
April_POLA_cDNA	RNA(cDNA) sequences from April at the surface at the Port of Los Angeles	unitless
July_POLA_cDNA	RNA(cDNA) sequences from July at the surface at the Port of Los Angeles	unitless
April_POLA_DNA	DNA sequences from April at the surface at the Port of Los Angeles	unitless
Oct_POLA_cDNA	RNA(cDNA) sequences from Oct at the surface at the Port of Los Angeles	unitless
Jan_POLA_cDNA	RNA(cDNA) sequences from Jan at the surface at the Port of Los Angeles	unitless

July_890m_DNA	DNA sequences from July at 890m depth at SPOT	unitless
July_890m_cDNA	RNA(cDNA) sequences from July at 890m depth at SPOT	unitless
Oct_DCM_cDNA	RNA(cDNA) sequences from Oct at the DCM at SPOT	unitless
Jan_890m_DNA	DNA sequences from Jan at 890m depth at SPOT	unitless
Jan_890m_cDNA	RNA(cDNA) sequences from Jan at 890m depth at SPOT	unitless
Oct_890m_DNA	DNA sequences from Oct at 890m depth at SPOT	unitless
Oct_890m_cDNA	RNA(cDNA) sequences from Oct at 890m depth at SPOT	unitless
April_890m_cDNA	RNA(cDNA) sequences from April at 890m depth at SPOT	unitless
taxonomy	Full taxonomic description from SILVA v111 database	unitless

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

## Instruments

<b>Dataset-specific Instrument Name</b>	Illumina MiSeq
<b>Generic Instrument Name</b>	Automated DNA Sequencer
<b>Generic Instrument Description</b>	A DNA sequencer is an instrument that determines the order of deoxynucleotides in deoxyribonucleic acid sequences.

<b>Dataset-specific Instrument Name</b>	
<b>Generic Instrument Name</b>	CTD Sea-Bird SBE 911plus
<b>Generic Instrument Description</b>	The Sea-Bird SBE 911 plus is a type of CTD instrument package for continuous measurement of conductivity, temperature and pressure. The SBE 911 plus includes the SBE 9plus Underwater Unit and the SBE 11plus Deck Unit (for real-time readout using conductive wire) for deployment from a vessel. The combination of the SBE 9 plus and SBE 11 plus is called a SBE 911 plus. The SBE 9 plus uses Sea-Bird's standard modular temperature and conductivity sensors (SBE 3 plus and SBE 4). The SBE 9 plus CTD can be configured with up to eight auxiliary sensors to measure other parameters including dissolved oxygen, pH, turbidity, fluorescence, light (PAR), light transmission, etc.). more information from Sea-Bird Electronics

<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](#) ]