

NCBI accession metadata 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/745527>

Data Type: Other Field Results

Version: 1

Version Date: 2018-09-04

Project

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

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

Abstract

Raw DNA and RNA V4 tag sequences include spatially and temporally distinct samples from coastal California. Samples were collected in Niskin bottles with a CTD rosette at the San Pedro Ocean Time-series (SPOT) between April of 2013 and January of 2014. This dataset contains sequence data accession numbers and metadata for the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database (SRA Study ID: SRP070577, BioProject: PRJNA311248).

Table of Contents

- [Coverage](#)
- [Dataset Description](#)
 - [Methods & Sampling](#)
 - [Data Processing Description](#)
- [Data Files](#)
- [Supplemental Files](#)
- [Related Publications](#)
- [Related Datasets](#)
- [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

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

Sequence data can be found in the NCBI SRA database under accession number [SRP070577](#) with the associated BioProject [PRJNA311248](#).

Methods & Sampling

Samples were collected seasonally at the San Pedro Ocean Time-series (SPOT) station at four depths (surface,

subsurface chlorophyll maximum, 150 and 890 m). The SPOT station was sampled from 5 m, the subsurface chlorophyll maximum (SCM), 150 and 890 m using 10 L Niskin bottles mounted on a CTD rosette, during regularly scheduled cruises (<https://dornsife.usc.edu/spot/>).

Seawater from all samples was sequentially pre-filtered through 200 µm and 80 µm Nitex mesh to reduce abundances of multicellular eukaryotes (metazoa). Near-surface and SCM seawater (2 L) and 150 and 890 m seawater (4 L) was filtered onto GF/F filters (nominal pore size 0.7 µm; Whatman, International Ltd, Florham Park, NJ, USA) and immediately flash frozen in liquid nitrogen for later DNA and RNA extraction.

Total DNA and RNA were extracted simultaneously from each sample using the All Prep DNA/RNA Mini kit (Qiagen, Valencia, CA, USA, #80204). Genomic DNA was removed during the RNA extraction with RNase-Free DNase reagents (Qiagen, #79254). Total extracted RNA was checked for residual genomic DNA by performing a polymerase chain reaction (PCR) using DNA specific primers to ensure that no amplified products appeared when run on an agarose gel. RNA was reverse transcribed into cDNA using iScript Reverse Transcription Supermix with random hexamers (Bio-Rad Laboratories, Hercules, CA, USA, #170-8840).

The resulting cDNA and DNA from each sample were PCR amplified using V4 forward (5' - CCAGCA[GC]C[CT]GCGGTA ATTCC-3') and reverse (5' -ACTTTCGTTCTTGAT[CT][AG]A-3') primers (Stoeck *et al.* 2010). Duplicate PCR reactions were performed in 50 µL volumes of: 1X Phusion High-Fidelity DNA buffer, 1 unit of Phusion DNA polymerase (New England Biolabs, Ipswich, MA, USA, #M0530S), 200 µM of dNTPs, 0.5 µM of each V4 forward and reverse primer, 3% DMSO, 50 mM of MgCl and 5 ng of either DNA or cDNA template per reaction. The PCR thermal cycler program consisted of a 98°C denaturation step for 30 s, followed by 10 cycles of 10 s at 98°C, 30 s at 53°C and 30 s at 72°C, and then 15 cycles of 10 s at 98°C, 30 s at 48°C and 30 s at 72°C, and a final elongation step at 72°C for 10 min, as described in Rodríguez-Martínez *et al.* (2012). PCR products were purified (Qiagen, #28104) and duplicate samples were pooled. The ~400 bp cDNA and DNA PCR products were quality checked on an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA).

Sampling Locations:

SPOT (33° 33' N, 118° 24' W) - surface, DCM, 150 m, and 890 m

Port of LA (33° 42.75' N, 118° 15.55' W) - surface

Catalina (33° 27.17' N, 118° 28.51' W)-surface

For protocols 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>

<https://www.protocols.io/view/18s-v4-tag-sequencing-pcr-amplification-and-librar-hdmb246>

Data Processing Description

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.

BCO-DMO Data Manager Processing Notes:

- * File "SRA_metadata_finalSHU.xlsx" imported into the BCO-DMO data system.
- * Added SRA_run_ids from file "Listof_SRA_IDs.xlsx."
- * 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.
- * Split lat_lon column into "lat" and "lon" in decimal degrees.
- * Added active html links to dataset for SRA Run IDs
- * For curatorial purposes and to satisfy the OCE Sample and Data Policy requirements, the contributor's github was forked to BCO-DMO's github, a release was made, and the .zip file containing that release

downloaded to BCO-DMO's data servers.

- * BCO-DMO/V4_tagsequencing_18Sdiversity_q1 forked from shu251/V4_tagsequencing_18Sdiversity_q1
- * Release v2 created: https://github.com/BCODMO/V4_tagsequencing_18Sdiversity_q1/tree/v2

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

Data Files

File
SRA_metadata.csv (Comma Separated Values (.csv), 27.82 KB) MD5:53720e5a5d4feefc745d738acdc9e5af Primary data file for dataset ID 745527

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

Supplemental Files

File
18S rRNA gene tag-sequencing code from GitHub, release v2 filename: V4_tagsequencing_18Sdiversity_q1-v2.zip (ZIP Archive (ZIP), 813.68 KB) MD5:984c586d20c83b159971cc26067b678c Zip file containing release https://github.com/BCODMO/V4_tagsequencing_18Sdiversity_q1/tree/v2

[[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) 18S rRNA gene tag-sequencing code - V4 hypervariable region. version v2. GitHub. Date accessed 2018-09-16. https://github.com/BCODMO/V4_tagsequencing_18Sdiversity_q1/tree/v2
Software

Hu, S. K., Campbell, V., Connell, P., Gellene, A. G., Liu, Z., Terrado, R., & Caron, D. A. (2016). Protistan diversity and activity inferred from RNA and DNA at a coastal ocean site in the eastern North Pacific. FEMS Microbiology Ecology, fiw050. doi:[10.1093/femsec/fiw050](https://doi.org/10.1093/femsec/fiw050)
Results

Mesrop, L., & Hu, S. (2017). 18S V4 tag sequencing PCR amplification and library prep (Illumina) v1. Protocols.io. doi:[10.17504/protocols.io.hdmb246](https://doi.org/10.17504/protocols.io.hdmb246)
Methods

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

Related Datasets

References

University of Southern California (2016). marine metagenome, DNA and RNA based tag sequences raw sequence reads. 2016/02. In: BioProject [Internet]. Bethesda, MD: National Library of Medicine (US), National Center for Biotechnology Information; 2011-. Available from:

Parameters

Parameter	Description	Units
bioproject_accession	NCBI BioProject identifier	unitless
sample_name	Sample name	unitless
SRA_run_ID	SRA Run identifier	unitless
SRA_run_link	URL for SRA Run Page at NCBI	unitless
library_ID	SRA title	unitless
SRA_study_ID	SRA study identifier	unitless
SRA_title	Descriptive title of SRA accession	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 file 1	unitless
filename	Name of file 1	unitless
filetype2	Type of file 2	unitless
filename2	Name of file 2	unitless
Depth	Nominal depth ("Surface" or number of meters) of sample collection	various
lat	Latitude of sample collection	decimal degrees
lon	Longitude of sample collection	decimal degrees

Instruments

Dataset-specific Instrument Name	Illumina MiSeq
Generic Instrument Name	Automated DNA Sequencer
Generic Instrument Description	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.

Dataset-specific Instrument Name	
Generic Instrument Name	CTD Sea-Bird SBE 911plus
Generic Instrument Description	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

Dataset-specific Instrument Name	
Generic Instrument Name	Niskin bottle
Generic Instrument Description	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

Website	https://www.bco-dmo.org/deployment/754348
Platform	R/V Yellowfin
Start Date	2005-01-19
End Date	2018-07-18
Description	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 metazoan) 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
NSF Division of Ocean Sciences (NSF OCE)	OCE-1737409

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