

BIOS-SCOPE RENEWABLE REAGENTS AND DATA SHARING PLAN:**What data will be generated by your research and where will it be deposited?**

Data Products: Throughout the duration of the project samples will be collected from BATS and BIOS-SCOPE process cruises, ship based experiments and laboratory based experiments. The core data produced in this project will be collected in conjunction with vertical hydrographic profiles and collaborative experiments. These data types will include hydrographic profiles (i.e. temperature, salinity, fluorescence); biogeochemical profiles (i.e. dissolved nutrients, bulk DOM, POM, dissolved oxygen etc.); biological profiles (i.e. microbial biomass and rate, zooplankton biomass etc.); nucleic acid sequences describing microbial, protist and zooplankton communities, metabolic potential and gene expression; and DOM chemical composition (i.e. FT-ICR Mass spectrometry of DOM compounds as well as neutral and amino sugar concentrations). Metadata from field sampling and experimental designs will be crucial to interpreting the data, and our management strategy ensures consistent formatting and archiving of linked data products.

Hydrographic and Biogeochemical data: All field and experimental biogeochemical data will be collected and analyzed according to oceanographic best practices (http://www.us-ocb.org/publications/TS_Workshop_report_FINAL.pdf). Metadata will include variable names, derived units, experimental set-ups, analytical methods, descriptions of synthesis and calibration procedures where appropriate, data location, season information, and quality control information. Data will be submitted to the Biological and Chemical Oceanography Data Management Office (BCO-DMO; <http://www.bco-dmo.org/>) for archiving and public dissemination. PIs of BIOS-SCOPE are familiar with BCO-DMO and understand the required metadata standards as well as the optimum formats for data submission. Data sets and associated metadata will be made available in Microsoft™ Excel™-compatible spreadsheets. In addition all meta oceanographic data from seawater samples collected during hydrographic casts concurrent to sampling will be supplied in association with the BATS data (<http://bats.bios.edu>) and will be submitted to BCO-DMO. Furthermore, physical and hydrographic data from BATS are deposited at OceanSITES: <http://www.whoi.edu/virtual/oceansites/> and are available to the public.

Experimental data: Data such as microbial enumeration, zooplankton counts, biological rate measurements, DOM concentration changes and chemical composition, nutrients, microbial growth efficiency and community structure generated from the various BIOS-SCOPE experiments (i.e. Microbial remineralization experiments, DOM bioavailability experiments, zooplankton/microbial experiments and zooplankton metabolism experiments) will also be stored in the BCO-DMO project. Metadata for the experiments will similarly be stored in BCO-DMO and a DOI created for the final dataset.

Genomic and transcriptomic sequence data: All sequence products will be permanently raw-archived with appropriate metadata in the NCBI Sequence Read Archive (SRA). These products will include genomic, metagenomic, 16S and 18S multiplexed phylogenetic and transcriptomic data. All sequence data will be collected and curated in accordance with the Genomic Standards Consortium's (<http://gensc.org>) Minimum Information about a (Meta)Genome Sequence (MIGS/MIMS/MIENS) standards. These outline a standardized format for the minimum information required to accurately describe genomic and metagenomic data, including metadata, with the goal of facilitating inter-study comparisons and transparency (Field et al. 2008). There is a new cyberinfrastructure project being developed for microbial data called iMicrobe. The BIOS-SCOPE PIs will work with the SCOPE PI and will plan to use iMicrobe data repository and pipelines as they are developed. iMicrobe and BCO-DMO have begun to make reciprocal links between the two sites in order to facilitate discovery and interoperability of "omic" and biogeochemical data streams. The project code capturing this BIOS-SCOPE project will be included in associated publications and the link will be posted to BCO-DMO.

Mass Spectrometry data from DOM characterization: Three types of mass spectral data will be generated. First, chromatographic mass spectral (LC/MS) data collected on the FT-ICR MS which includes the ion chromatograms and the resulting lists of retention times and high resolution mass:charge values for individual compounds. These data are processed through R-based scripts and exported to Matlab for further analysis. Second, fragmentation spectra (MS/MS) generated for selected ions in these analyses will

be stored on the RAID-V server at Woods Hole Oceanographic Institution (WHOI). Retention times, mass:charge values, MS/MS spectra and putative identifications for each mass spectral feature will be stored in an in-house database at WHOI (Longnecker et al., 2015). A third data type is LC/MS data collected on the triple-quadrupole (TSQ) mass spectrometer. These data include extracted ion chromatograms for selected precursor-fragment pairs as a function of retention time along the column. The primary data are processed by MAVEN (Melamud et al., 2010) and the quantity of each compound is exported to Matlab. All data will be stored in formats (e.g., spreadsheets, PDF files) that are readily accessible by common software. All data will be available from the Kujawinski group upon request. Kujawinski's group has established a relationship with the MetaboLights repository (<http://www.ebi.ac.uk/metabolights/>), which is sponsored by EMBL-EBI in the United Kingdom. Although they do not specialize in the storage of untargeted metabolomics data, they have welcomed past Kujawinski datasets. Datasets are given an accession number, which can be included in publications.

What is your plan for sharing the data? Data submitted to BCO-DMO will be available to all BIOS-SCOPE investigators throughout the duration of the project through a password protected portal. The data generated from the BIOS-SCOPE project will be made publicly available immediately upon publication in a peer-reviewed journal or within 2 years of the completion of the project. Data will be released to relevant publically available data repositories as described above (i.e. BCO-DMO, NCBI, iMicrobe, MetaboLights). Unpublished data of other types will be released at the discretion of the P.I.s, with the intent of maximizing the dissemination of data via peer-reviewed formats.

What renewable reagents will be generated by your research? Live bacterial cell cultures, fixed zooplankton specimens and nucleic acid samples will be archived throughout and after the closeout of the award. Storage and reposition will be as follows:

Cell cultures isolated during the project will be stored in a liquid nitrogen Dewar as part of the OSU High Throughput Culture Collection and will be designated with a unique identifier. Isolates will be maintained in two different Dewars for redundancy.

Microbial DNA- The remaining DNA extracts from the microbial time series and experimental work, after BIOSCOPE needs, will be archived at UCSB in Liquid Nitrogen or in Ultra low freezers.

Zooplankton Specimens and DNA- Any remaining zooplankton specimens and extracted DNA, after BIOSCOPE PI needs, will be archived at BIOS in alcohol Jars and in Ultra low freezers, respectively.

What is your plan for sharing the renewable reagents? Live bacterial isolates will be maintained throughout and after the award closes. Live isolates will be made available through the OSU Office for Commercialization and Corporate Development upon completion of a Uniform Biological Materials Transfer Agreement as detailed on the Giovannoni Laboratory webpage (<http://giovannonilab.science.oregonstate.edu/htcl-contributions-science-and-industry>). The cells are usually provided at no cost (beyond shipping expenses) to the requesting institution. The Giovannoni Laboratory fully supports and is committed to the timely distribution of scientific materials and data for the advancement of science

Nucleic acid samples from the Carlson laboratory and the Mass / Blanco-Bercial will be maintained throughout and after the award closes. Specimens and DNA samples will be made available upon request to academic research groups upon completion of a Uniform Biological Materials Transfer Agreement.

Citations

Field, D., Garrity, G., Gray, T., Morrison, N., Selengut, J., Sterk, P., ... & Methe, B. (2008). The minimum information about a genome sequence (MIGS) specification. *Nature biotechnology*, 26(5), 541-547.

Longnecker, K., Futrelle, J., Coburn, E., Kido Soule, M.C., and Kujawinski, E.B. (2015). Environmental metabolomics: databases and tools for data analysis. *Marine Chemistry* on-line: June 19.

Melamud, E., Vastag, L., and Rabinowitz, J.D. (2010). Metabolomic analysis and visualization engine for LC-MS data. *Analytical Chemistry* 82, 9818-9826.