

DATA MANAGEMENT PLAN

1. Types of data and samples

With this project we will generate laboratory-derived data on the biological and physiological characteristics of one to three related mixotrophic flagellates and the replication characteristics of a giant virus that infects them. We will also be generating field-based analyses of flagellate and virus abundance, and community 16S rRNA gene sequence data of prokaryotes ingested by the flagellate. In particular, we anticipate the following general types of data: (1) Electron micrographs of flagellate ingestion process and of the viral replication cycle (2) Data on growth/replication characteristics of the host and virus (3) Model code and output (4) DNA sequences (18S rRNA sequences of the host, 16S rRNA tag sequencing data of cells ingested by the flagellate) (5) abundance of virus and host in the field by qPCR. For the field sampling component, we will rely on the Hawaii Ocean Time-series shipboard measurement program for ancillary environmental data, and these data are publicly available via the HOT program data server (<http://hahana.soest.hawaii.edu/hot/hot-dogs/interface.html>).

Data from field and laboratory analyses will be generated according to established analytical protocols, with methodologies for each procedure included with the accompanying metadata. Quantitative PCR (qPCR) and quantitative reverse transcriptase (qRT) PCR will be utilized to assess variability in viruses (both DNA and RNA) abundances from discrete water column samples. 18S rRNA gene sequences will be generated from laboratory isolates of protists. Samples and subsamples will be physically archived in appropriate locations (*e.g.*, freezers, refrigerators) until analysis. Types of samples will include: plankton concentrates (on filters) for subsequent nucleic acid extractions, nucleic acid extracts, cryopreserved plankton cells (for subsequent cultivation efforts), and isolated planktonic cells in appropriate seawater-based media.

All code used in model analyses will be stored as annotated, fully reproducible R scripts. These scripts will be included as supplemental data in the resulting publications, and will also be available by personally contacting the investigators.

2. Data and Metadata Standards

Data quality will be assured through analysis of replicate samples, and proper accounting of standards, and controls. Data will be archived in multiple locations, including hard copies, laboratory computers, local servers and cloud-based servers. Our project will be registered with the Biological and Chemical Oceanography Data Management Office (BCO-DMO) established through the Ocean Carbon and Biogeochemistry program at the Woods Hole Oceanographic Institution. We will follow the appropriate metadata standards outlined in the document “Data Management and Guidelines Manual” from the BCO-DMO project office. We will be generating a large amount of experimental and sequence data in this project that is not adequately served though BCO-DMO. DNA sequence data will be deposited into the GenBank public archive in accordance with the NSF Sample and Data Policy. Information on how to access the data (database location and accession numbers) will be posted with other

project data in BCO-DMO. Both metadata and primary data will be submitted to the appropriate public data repositories (BCO-DMO, GenBank). In addition, to increase accessibility to project data and the dissemination of our research findings, we have budgeted funds for publications deriving from this project to appear as open-access articles in peer-reviewed journals. Ph.D. or M.S. theses associated with the project will also be made available electronically.

3. Data Sharing, Reuse and Redistribution Policies

We will share and archive data collected as part of this research project in compliance with the Division of Ocean Science Data and Sample Policy. Biological data will also be maintained on computers in the PIs lab with backups on a portable external hard drive and on a server maintained by the Center for Microbial Oceanography-Research and Education. Data will be made publicly available via GenBank and BCO-DMO as soon as possible after collection, but within 1 year of collection, and all project data will be made publicly available within six months of the project end date. The original data collector/creator/principal investigator does not retain the right to use the data before opening it up to wider use.

4. Policies and provisions for re-use, re-distribution

All data from this project are considered within the public domain and the datasets deriving from the project will not be copyrighted. Hence, we do not anticipate intellectual property issues associated with the acquisition of the data. The data acquired and preserved in the context of this proposal will be further governed by the University of Hawaii's policies pertaining to intellectual property, record retention, and data management.

5. Plans for archiving and preservation of access

Data submitted to BCO-DMO are maintained in perpetuity by archiving at the National Oceanographic Data Center (NODC). We will expand our culture collection of phytoplankton along with confirmed and putative viruses that infect a significant number of these protistan hosts. For those virus-host systems where the virus is successfully purified and characterized, the hosts and viruses will be submitted as soon as practical (no later than the date on which their descriptions are published) to the National Center for Marine Algae and Microbiota (NCMA) or American Type Culture Collection (ATCC). Pure cultures not accepted for archiving will be maintained at UH for as long as feasible but for no less than three years following description/publication.