Data Policy Compliance

This project will comply with the NSF OCE Data and Sample Policy

Description of Data Types

(1) Observational. Water column physico-chemical data: Time of collection, date, location, depth, temperature, salinity, dissolved oxygen, nitrate, nitrite, ammonia, total carbon, total organic carbon, dissolved organic and inorganic carbon, total nitrogen, silicate, and particulate organic carbon. To maximize access to these data, data will be stored in BCO-DMO by the end of this project and in PANGAEA (http://www.pangaea.de/about/). PANGAEA is a member of the ICSU World Data System, created by the International Council for Science, is jointly hosted by the Alfred Wegener Institute and the University of Bremen, and is established as a repository for the long-term availability and accessibility of archived data and metadata in secure and machine readable formats.

(2) Experimental.

   a. Fluorescence in situ hybridization (FISH) counts for particular Group II Syndiniales clades: these data will include counts for all probe targets for all water samples collected. In addition, we will count total prokaryotic and eukaryotic cells on DAPI stained filters for all samples, numbers of infected hosts, and free dinospores. Data will be deposited as for (1).
   b. Sequence data: MiSeq PE 300bp iTAG data will be generated for the eukaryotic domain of life by start of Yr 2 from all samples in our collection. These data will be deposited to GenBank and made publicly available by the end of the project period. Links to data will be provided through BCO-DMO. Draft genome data for single cells (assemblies of Illumina NextSeq data) for selected cells will be generated by end of Yr. 2 and deposited to GenBank and Integrated Microbial Genomes & Microbiome Samples Expert Review (IMG/MER). These partial genomes of parasites will also be submitted to EuPathDB, a bioinformatics resource center for accessing genomic-scale datasets associated with pathogenic eukaryotic microorganisms. Links to all data and associated water column metadata will be provided through BCO-DMO.
   c. Image data: Microscopy images (light and fluorescence) of host cells exhibiting infections by Group II Syndiniales will be posted to BCO-DMO and selected images used in publications.
   d. Experimental protocols: DNA extraction protocols and protocols for FISH hybridization with all probes used in the study will be made publicly available through protocols.io by the end of the project period.
   e. Educational outreach materials: Protocols for the laboratory exercise on FISH hybridization for high school students will be posted on Amazon Inspire, the new education resource for teachers dedicated to the free search, discovery and sharing of digital education resources (http://www.amazoninspire.com).
   f. All data will be shared through peer reviewed publications

(3) Model code and model output:

   a. The parasite-host model will be shared as open source, allowing the scientific community to continue to test and refine its numerical experiments and adapt it to other systems. Code will be shared through GitHub, a free web-based repository for open source projects. There will be no embargoes, restrictions, or privacy concerns associated with the data, provided users indicate a citation for the product. Model output will be stored in a local THREDDS server. Documentation will be provided which will allow users to identify relevant variables and explain the data in full.

b. The model will also be submitted to protocols.io by the end of the project period. Links to this location will be provided through BCO-DMO. Modeling approaches and results will also be reported in a publication

Data and Metadata Formats and Standards

(1) Observational data will be submitted as spreadsheet files (.xlsx).
(2) Genetic data formatting will follow the standards of NCBI, IMG, and MG-RAST and links to these data will be provided through BCO-DMO.
(3) Cell counts will be submitted as spreadsheet files, and image data will be submitted as JPEG files or format demanded by repositories.
(4) Model formulae will be sent to BCO-DMO as MS-Word documents, along with links to a GitHub site for the model code and a THREDDS site for the model output. Model output will be stored in ASCII format for readability over different platforms. Binary data sets also will be served, with full public access and instructions. Model-generated figures will be stored in pdf and eps/jpg formats respectively for the same purpose.

Data Storage and Access During the Project

All data will be shared electronically among all collaborators by posting as data are generated to a dedicated project page on Basecamp (https://basecamp.com) or Google Drive.
All experimental data generated at WHOI will be stored on our laboratory computers and associated Network Accessed Storage devices that are backed up daily by WHOI's automated backup service.

Mechanisms and Policies for Access, Sharing, Re-Use, and Re-Distribution

Our project data will be publicly available by the end of the project period through NCBI GenBank/ Gene Expression Omnibus (GEO) database, BCO-DMO, MG-RAST, and IMG. Our data will not involve security concerns, nor does access need to restricted in any way.

Data Sharing via BCO-DMO and Archiving

BCO-DMO staff will work with us to manage our project data, and data generated during the proposed research project will be contributed to the BCO-DMO system. Links to data repositories for our genomic and image data will be provided through BCO-DMO (see above). After data contributed to BCO-DMO are online and fully documented, BCO-DMO ensures that the data are archived properly at the appropriate National Data Center (e.g. NODC) for long-term archive preservation.

Roles and Responsibilities

Lead PI Gast will ensure compliance with this Data Management Plan. Edgcomb will be responsible for deposition of all iTAG data and water sample physico-chemical, modeling, and image data generated at WHOI. Co-PI Pachiadaki will be responsible for deposition of all genomic data. Gast will work with Follows and the student on the deposition of model code and output.