

Data Management Plan

The research products generated in the proposed effort include biological samples, genetic and genomic data, and physiological rate measurements. Here I present a data management plan that is based upon the NSF Division of Ocean Sciences Sample and Data Policy (NSF 11-060). I will cooperate and consult with the Biological and Chemical Oceanography Data Management Office (BCO-DMO) as appropriate and adhere to the overall data sharing philosophy (section II).

In general, I will:

- a) provide both inventory metadata and primary data within specified time limits (section III, A), including any and all marine environmental and genomic data (as in section VI, D).
- b) address data sharing issues in annual and final reports (section IV).
- c) archive any and all biological samples collected during my expeditions in Dr. Virginia Edgcomb's laboratory at Woods Hole Oceanographic Institution and, after 2 years, make them freely accessible to other researchers upon request.

The specific products and data that will be collected in the proposed effort are listed in Table 1. They will be managed as follows:

1. 16S and 18S rRNA marker gene 454-pyrosequencing data will be stored indefinitely on Edgcomb laboratory computers and deposited into the Short Read Archive (NCBI) within 2 years of acquisition. Associated metadata for every sample will also be submitted to the BCO-DMO.
2. Longer marker gene sequences will be stored indefinitely on Edgcomb laboratory computers and deposited into GenBank (NCBI) within 2 years of acquisition. Associated metadata for every sample will also be submitted to the BCO-DMO.
3. Within two years of acquisition, assembled and annotated archaeal symbiont contig sequences will be made publicly available through the genome/metagenome repositories MG-RAST metagenomics analysis server, as well as the Joint Genome Institute's (Department of Energy) Integrated Microbial Genomes and Metagenomes (IMG/M) system. Metadata associated with the genomic data will be stored on MG-RAST, IMG/M and the BCO-DMO.
4. Assembled contigs that are identified as derived from the eukaryotic, protist host are ancillary to the proposed work, however will be made immediately publicly available as a GenBank (NCBI) Whole Genome Shotgun project for use by other researchers.
5. Methanogenesis rate data collected during laboratory experiments will be made available, after publication, to other researchers upon request.

Table 1: Description of the data that will be collected as part of the proposed research effort.

<i>Type of data</i>	<i>Brief description</i>
1. Symbiont and host population and diversity data	454-pyrosequencing sequences of 16S and 18S rRNA genes (archaeal symbiont and protist host, respectively) from natural populations of methanogen-protist symbioses.
2. Symbiont and host marker gene sequences	Longer sequences of taxonomic gene markers for individual symbiont-host pairs (16S and 18S rRNA genes, respectively).
3. Archaeal symbiont genome data	Genomic sequence data derived from the DNA from individual methanogen-protist symbioses.
4. Protist host genome data	Ancillary protistan genomic sequence data that will result from the sequencing of all of the DNA derived from an individual methanogen-protist symbiosis.
5. Methanogenesis rates	Laboratory measurements of methane production by natural populations of methanogen-protist symbioses.