DATA MANAGEMENT PLAN

The main objectives of this proposal are to characterize the transcriptional response to infection of both virus and host (the diatoms *Pseudo-nitzschia* and *Thalassiosira pseudonana*) and to characterize the metavirome (with focus on ssDNA and RNA viruses) of diatom dominated waters in Pacific Northwest. Additional viruses and strains of *Pseudo-nitzschia* will also be isolated and characterized. Data management issues fall into three areas.

1) **Biological material:** Novel clonal diatom isolates will be deposited in the Provasoli-Guillard National Center for Culture of Marine Phytoplankton (https://ccmp.bigelow.org/). At present there is not an obvious analogous repository for marine viruses, although the Roscoff Culture Collection of marine phytoplankton now does include some viruses (http://www.sb-roscoff.fr/Phyto/RCC/index.php). If a suitable public collection cannot be arranged (preferred), the PI will make viral material available to other investigators on request.

2) **Sequence data:** Assembled and annotated viral genomes will be deposited in NCBI's GenBank database where they will be publicly available. Data from the RNA-seq experiments (raw data directly from the Illumina) will be deposited in the Sequence Read Archive (formerly the short read archive) at NCBI. For the metaviromic datasets, a BioProject will be created in GenBank with both the raw sequence reads deposited in the Sequence Read Archive and assembled and annotated viral metagenomic contigs and scaffolds submitted as a Whole Genome Sequencing project. In addition, sequences and associated environmental data will also be submitted to CAMERA (Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis).

3) **Environmental data:** The major field dataset for this proposal comes from the "GeoMICS" cruise along line P in May 2012 (Armbrust, Moffett, Cutter, Ingalls, Twining, Morris PIs). All investigators participating in the cruise have agreed to data sharing practices and timelines for data release. My lab has contributed total viral abundance measurements along the transect to this dataset. Sequences generated as part of the original GeoMICS effort (metagenomics of < 2 um fraction, metatranscriptomics for both < 2 um and >2 um size fractions and 16S and 18S rRNA libraries) will be submitted to NCBI by the primary investigators. Currently core oceanographic data is being prepared for deposition with the Biological and Chemical Oceanography Data Management Office (BCO-DMO) by the primary investigators. We will add a dataset to this project at the BCO-DMO with the viral metagenome methods and links to locations where the sequence data deposited. We also expect to collect samples for additional diatom culture and virus isolations on cruises of opportunity within Puget Sound. In these cases physical (CTD), chemical (nutrients) and biological (*Pseudo-nitzschia* and viral counts, domoic acid etc..) data will be deposited with the BCO-DMO.