Collaborative research: Characterization of *Synechococcus*-cyanophage interactions across phylogenetic and temporal scales

Resource Sharing and Data Management Plan

All data collected during this project will be archived in accordance with NSF policy (see AAG Chapter VI.D.4) on the dissemination and sharing of research results.

<u>Data and resources to be generated by this project</u> The following data and resources will be generated by this study: *Synechococcus* isolates from Narragansett Bay (NB) and flow cytometric profiles of those isolates; cyanophage isolates from NB; ITS sequences and complete or nearly complete genome sequences of *Synechococcus*; DNA polymerase marker gene sequences and complete genomes of cyanophage isolates; *Synechococcus* and cyanophage amplicon sequence data for environmental samples taken in NB; *Synechococcus* and cyanophage abundances, total chlorophyll concentrations, temperature, salinity, and inorganic nutrient concentrations in NB taken during the proposed two year time series and temperature, salinity, and cyanophage titers from the previously collected 10 year time-series. A syllabus for the undergraduate genome project (Ahlgren) and virology course (Marston) course will be generated from the broader impacts activities. Details about the format, storage, and accessibility plans of these data and materials are provided below.

General policies for data sharing and redistribution of data
To facilitate collaboration between Ahlgren and Marston, data generated from this project will be made freely available between PIs and their students and postdoctoral fellows via Google Drive or a similar file-sharing platform. For researchers outside of the Ahlgren and Marston labs, we will follow a policy of free distribution of data to be provided upon request to the PIs Ahlgren or Marston. Policies regarding the public dissemination of sequencing and proteomics data are noted below, but these raw and processed data will be archived on public databases upon publication. Reuse of other processed data following publication will fall under the prevue of the policies of the publishing journals, but as much as possible we intend to provide such data to requesting parties to facilitate on-going research.

Strains
Frozen stocks of *Synechococcus* strains isolated from Narragansett Bay will be archived at -80°C (with 2% DMSO) and cyanophage will be stored at 4°C (known to be stable for at least 20 years). These strains will be made available upon request following publication of our work.

Flow cytometry data
Flow cytometry data for *Synechococcus* isolates will be stored as FCS 3.1 format accessible by other flow cytometry programs like FloJo.

Oceanographic measurements
Measurements of the physical, biological, and chemical properties of NB for the two year time-series will be deposited in Biological and Chemical Oceanography Data Management Office (BCO-DMO) within 2 years of collection following the NSF OCE's

guidelines.

DNA sequence data
DNA sequencing data generated (ITS and genome sequences of *Synechococcus*, cyanophage, and environmental amplicon sequences) will be stored as FASTQ format files and submitted to the Short Read Archive (SRA) at the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/sra) and released for public access upon publication of our research. Assembled and annotated *Synechococcus* genome sequences will be submitted to NCBI's Genome database and made available for public use following publication of these genomes.

Data storage
All data products will be backed up on several computer hard drives or portable hard drives, including those with Redundant Array of Inexpensive Disks (RAID), and if possible for smaller datasets on compact discs (CDs). For data archived on hard drives, data will be transferred onto new drives every ~ 4 yrs recognizing the typical lifetime of this type of media storage. Lab members will record details about the experiments they conduct in physical lab notebooks, which will be stored in the labs or the offices of Ahlgren and Marston.

Code
While this project will not generate any new significant bioinformatics or statistical tools, code used for comparative genomics of cyanophage and *Synechococcus* genomes and statistical analysis of data (namely analysis of infection networks, time-series and variance partitioning analysis of environmental amplicon sequencing) will be made available at Github and/or with journal publications as supplemental material.

Genome project course syllabus
The syllabus for Ahlgren's The Genome Project course and Marston's Virology course described in the Broader Impacts section will be made publically available on their respective university websites in PDF format.