

Data Management Plan

Summary

The analysis of phage-bacteria interaction networks (PBINs) and their effects on ocean ecosystems will involve: (i) development of theoretical models; (ii) implementation of theoretical models as software; (iii) analysis of PBIN data and other ecological data. All software and data will be released for community usage in at least two of three cyber “locations”:

- 1) Supplementary File on journal homepages accompanying publication.
- 2) Weitz group home page or on a dedicated website maintained by the Weitz group.
- 3) Third party resource pages specific to the data or software tool.

I address the management of each kind of “data” below. Please note that release of all aggregated data and software tools will be completely free and without restrictions on usage. To facilitate this goal, all software tools will be released according to the GNU Public License v3 or the Creative Commons 1.0 license. Both licenses facilitate unrestricted use and modification of materials.

Theoretical Models

Theoretical models will be published with complete information necessary to recapitulate the theory. Theoretical models are collections of mathematical analyses and arguments that are best described in a standard publication. We follow this practice, usually by including detailed theoretical models in the Supplementary Information of publications. Links to all publications are available on the PI’s group website: <http://ecothery.biology.gatech.edu>. However, most models are implemented as software tools for end-users, which we discuss below.

Software

We will release all software in this project as open-source software, freely available for download, and with no restriction on use or modification. Software will be released as supplementary files in journal publications and on the Weitz group homepage here:

<http://ecothery.biology.gatech.edu/downloads/>

In addition, we will distribute software on community sites to maximize their use and accessibility. We anticipate the release of the following software tools and dissemination methods.

1. *Estimating network properties of PBINs.* We will develop BiWeb-ml, a Matlab-based software tool for the analysis of bipartite networks, such as PBINs. We will distribute BiWeb-ml using GitHub. The code development for the Python component of the project is led by Dr. Timothy Poisot, U of Quebec.

<https://github.com/tpoisot/BiWeb>

In addition, we will distribute Matlab software on the “User Community” File Exchange:

http://www.mathworks.com/matlabcentral/fileexchange/?s_cid=global_nav

2. *Multi-scale models of phage-bacteria ecological dynamics.* We have developed a preliminary version of software to model the coevolutionary dynamics of phages and bacteria. This software is part of a forthcoming publication in *Evolution on CRISPR/Cas induced coevolution*. The entire Matlab code has been released for public use:

<http://dryad2.lib.ncsu.edu/handle/10255/dryad.37461>

Future development of this software as part of Aims 2 and 3 will be released on Dryad and/or GitHub. Both provide environments for release and maintenance of biological models.

Phage-Bacteria Interaction Network Data

We will analyze PBIN data from the Sullivan and Wilhelm groups, in addition to analyze previously collected data. In all cases when we digitize PBIN data, we will release it as supplementary material with the original publication and release it on our website. We did this with our first publication (Flores et al., PNAS 2011), in which all 38 data sets were released as a single Excel file:

<http://ecothery.biology.gatech.edu/downloads/host-phage-infection-network-datasets>

Metadata will be released using Excel or a well-defined database format.