

PHANTOME: PHage ANnotation TOols and Methods

Rob Edwards, Mya Breitbart, Jeff Elhai, and Matt Sullivan

Intellectual Merit

Phages, viruses that infect bacteria, are the most abundant biological entities on the planet. Phages have been central to many molecular biology tools and discoveries, and serve important ecological functions, including structuring microbial communities, driving evolution through horizontal gene transfer, and playing major roles in biogeochemical cycling. Despite the indisputable importance of phages in the biosphere, the annotations of their genomes have suffered from a lack of comparative tools, which impedes the ability of researchers to interpret sequences and obtain scientifically meaningful answers. There is no concerted effort to annotate phage genomes by any of the major microbial genome annotation centers. Having a consistent and accurate phage gene nomenclature is critical to phage research, and knowing which genes are viral versus microbial will help all researchers struggling to understand microbial genomes and metagenomes.

A platform for phage genome annotations (PHANTOME: PHage ANnotation TOols and Methods) will be developed based on the successful SEED database, extending the SEED to handle the nuances of both phages and prophages, to establish a consistent nomenclature for phage genes, and to develop a new tool for the identification of prophages. This new resource will be used to provide high quality annotations to over 1,000 existing phage and prophage genomes and dozens of existing phage metagenomes. However, the great majority of biological researchers have little experience in computer programming. To place the tools and data available within the resource in their hands and to provide them with the ability to combine them in creative ways, PHANTOME will be accessed using BioBIKE (Biological Integrated Knowledge/programming Environment), which makes use of familiar graphical conventions to facilitate problem solving. The workflows devised through the project will be made available to users through this interface, and likewise, workflows that users devise may be readily packaged and made available to others in the research community. A novel interactive help facility will be developed to aid new users find examples, tools, and procedures they need.

PHANTOME will enable researchers who come to the resource to answer specific questions to add their insights to the annotation as thoughts occur to them. At the same time, those who wish to bring their own sequences will be able to analyze and annotate them with the advantage of comparative analysis provided by the resource. The tools and high quality annotation developed in this project will serve as a solid basis for future efforts to comprehend phages and microbial genomes.

Broader Impacts

The project has the potential to educate and transform at many levels. Those who study phages will be encouraged to use bioinformatics in more creative ways through meeting workshops that introduce PHANTOME. A post-doc engaged by the project will travel to labs in the field, thereby gaining a panoramic view of phage biology that few attain. Graduate students in collaboration with participating high school teachers will lead biology modules that present molecular biology to high school students through the lens of discovery, using PHANTOME as the primary tool. Online discussion forums to meet users' needs will be facilitated by a staff of undergraduates who have learned bioinformatics in classes that use PHANTOME to enable students to relate book knowledge to productive investigation as soon as possible. In this way, biology students will absorb the connection between computation and biology within the context of actual research. Finally, all will gain by a greater understanding of phages, Earth's most populous biological entity, and a framework in which to place future understanding.