

# Phage Annotation Tools and Methods



Ramy K. Aziz<sup>1\*</sup>, Bhakti Dwivedi<sup>2\*</sup>, Joe Anderson<sup>3</sup>, Bonnie Hurwitz<sup>4</sup>, JP Massar<sup>5</sup>,  
Mya Breitbart<sup>2</sup>, Matthew Sullivan<sup>4</sup>, Jeff Elhai<sup>3</sup>, Robert A. Edwards<sup>1\*</sup>

<sup>1</sup>San Diego State University, San Diego, CA; <sup>2</sup>University of South Florida, Saint Petersburg, FL; <sup>3</sup>Virginia Commonwealth University, Richmond, VA; <sup>4</sup>University of Arizona, Tucson, AZ; <sup>5</sup>Berkeley, California. \*Presenting authors. For correspondence, email: ramy.aziz@salmonella.org, redwards@sciences.sdsu.edu

# PHANTOME

## SUMMARY

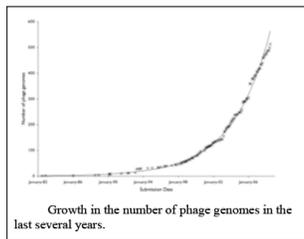
- Phages are the most abundant biological entities on Earth, yet their genomes are the most poorly annotated.
- Having a consistent and accurate phage gene nomenclature and knowing which genes are viral versus microbial will help researchers struggling to understand microbial genomes and metagenomes.
- PhAnToMe (Phage Annotation Tools and Methods, <http://www.phantome.org>) is a new platform for phage genome annotations and comparative phage genomics.
- PhAnToMe relies on the SEED database to handle the nuances of phages and prophages, develop a controlled vocabulary and a consistent nomenclature for phage genes, and create a new tool for the identification of prophages. It will provide high-quality annotations to over 1,000 existing phage and prophage genomes and dozens of existing phage metagenomes.
- Because most biologists have little experience in computer programming, PhAnToMe implements BioBIKE (Biological Integrated Knowledge/programming Environment), which makes use of familiar graphical conventions to facilitate problem solving.
- Researchers may analyze the currently available genomes, or bring their own sequences 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 phage and microbial genomes.

## INTEGRATED WEB SERVERS

SEED: <http://seed.sdsu.edu>  
RAST: <http://rast.nmpdr.org>  
BioBIKE: <http://biobike.csb.cvu.edu>

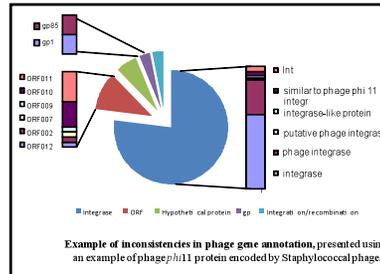
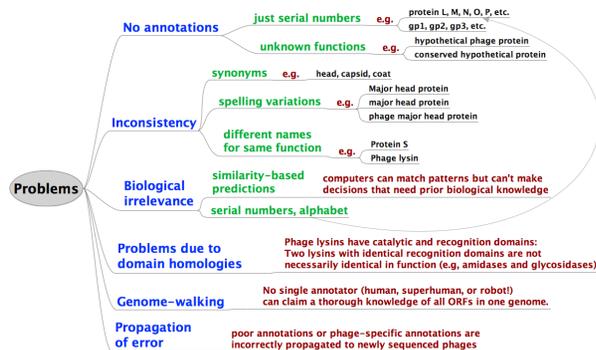
## MOTIVATION

- Phages are the most ubiquitous and diverse biological entities on Earth. Understanding the function and evolution of their genomes is thus critical.
- The number of publicly available complete phage genome sequences is growing exponentially



- Nevertheless, the interpretation of these genomes suffers from the lack of a systematic, consistent, well-maintained system for phage genome annotation.
- Additionally, a "biologist-friendly" resource for bioinformatics analysis of phage genomes is desperately needed.

## PROBLEMS WITH CURRENT PHAGE ANNOTATIONS



## AIMS

- To annotate phage genomes using SEED phage-specific subsystems for accurate, consistent, high-quality annotations
- To provide resources and tools based on these subsystems to allow researchers to extract and manipulate information in creative ways

## NEEDED ANNOTATION SYSTEM

- PhAnToMe strategy is to invest time and use human expertise to identify and correctly annotate phage and prophage genomes using the SEED phage-specific subsystem approach.

## SEED subsystem approach offers

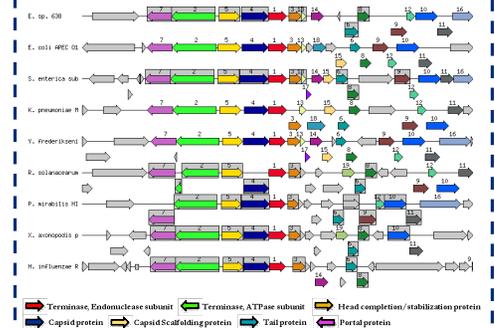
- consistency
- biological relevance
- automation/rapid propagation
- compatibility

## SUBSYSTEM-BASED ANNOTATION

- A subsystem is a simple table of protein-encoding genes (Pegs) with known, related functional roles across genomes.

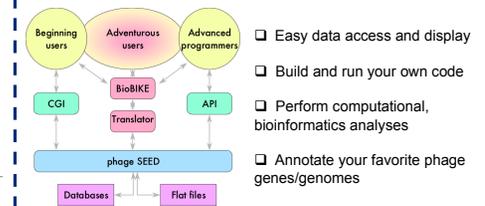
	Peg 1	Peg 2	...	Peg n
Genome a				
Genome b				
...				
Genome z				

- Phage-specific subsystems will be built on the core phage biological functions (e.g., capsids, tails, integrases, lysins), but this set will be extended to include other functional roles.
- Application:** Identifying conserved gene clusters of phage packaging genes across phage genomes.



## ACCESSING ANNOTATIONS & TOOLS

- PhAnToMe's phage annotations and analytical tools will be accessed via BioBIKE (Biological Integrated Knowledge/Programming Environment).



- BioBIKE functionality will allow users to extract and manipulate information in their own creative ways to answer specific questions and enable users to apply their expertise and insight to the annotations