# Phage Annotation Tools and Methods

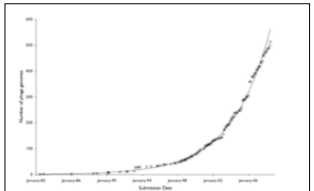
**Ramy K. Aziz[1*], Bhakti Dwivedi[2], Joe Anderson[3], Bonnie Hurwitz[4],Brad Hull[1], JP Massar[5], Mya Breitbart[2], Matthew Sullivan[4], Jeff Elhai[3], Robert A. Edwards[1*]**

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

PHANTOME

## MOTIVATION

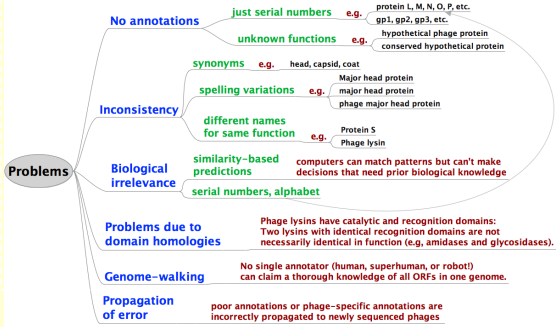
- Phages are the most ubiquitous and diverse biological entities on Earth. Understanding the function and evolution of their genomes is thus critical.
- The number of publicly available complete phage genome sequences is growing exponentially
- Nevertheless, the interpretation of these genomes suffers from the lack of a systematic, consistent, well-maintained system for phage genome annotation.

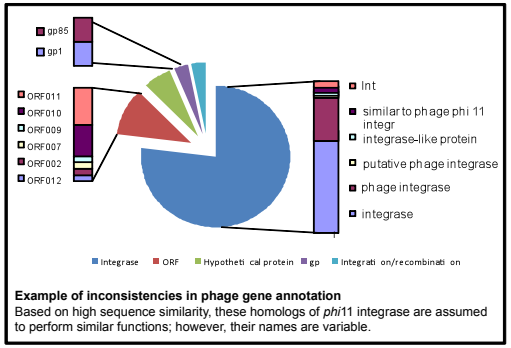


Growth in the number of phage genomes in the last several years.

- Additionally, a "biologist-friendly" resource for bioinformatics analysis of phage genomes is desperately needed.

## PROBLEMS WITH CURRENT PHAGE ANNOTATIONS



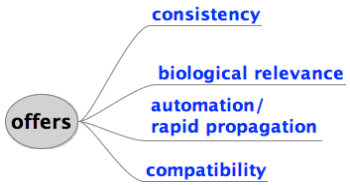

**Example of inconsistencies in phage gene annotation**
Based on high sequence similarity, these homologs of *phi*11 integrase are assumed to perform similar functions; however, their names are variable.

## PROJECT 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, that allow researchers to extract and manipulate information in creative ways
- To create a rapid automated phage annotation pipeline using subsystems technology (Phi-RAST)

## NEEDED ANNOTATION SYSTEM

- PhAnToMe (**http://www.phantome.org**) 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**




## SUBSYSTEMS-BASED ANNOTATION

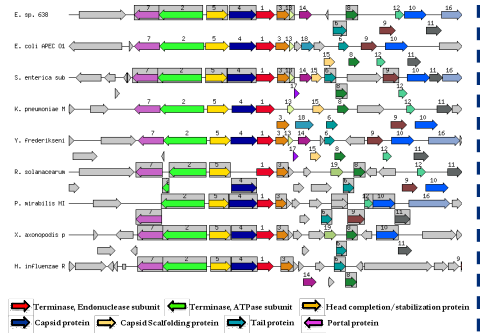
- A subsystem is simply a table of protein-encoding genes (Pegs) with related functional roles conserved across genomes.

| | Peg 1 | Peg 2 | … | Peg n |
|---|---|---|---|---|
| Genome a | | | | |
| Genome b | | | | |
| … | | | | |
| Genome z | | | | |

- Phage-specific subsystems are built for the universal phage biological functions (e.g., capsids, tails, integrases, lysins), but this set is being extended to include family-specific roles.
- **Current subsystems:**
- Universal functions:
  - structural modules (capsids, tails, tail fibers, etc.)
  - functional modules (replication, lysis, integration, etc.)
- Phage family-specific functions:
  - T4-like phage core proteins, T4-like cyanophage core proteins, T7-like phage core proteins, etc.

## EXAMPLE APPLICATION

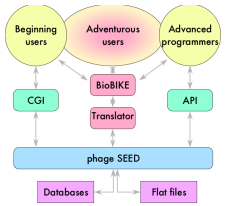
- Identifying conserved gene clusters of phage packaging genes across phage genomes.




**Phage Packaging Clusters.** Conserved clusters of terminase, portal, and scaffolding protein-encoding genes in phage genomes

## ACCESSING ANNOTATIONS & TOOLS

- PhAnToMe's phage annotations and analytical tools will be accessed via BioBIKE (Biological Integrated Knowledge/ Programming Environment, http://biobike.csbc.vcu.edu).




- ❑ Easy data access and display
- ❑ Build and run your own code
- ❑ Perform computational, bioinformatics analyses
- ❑ Annotate your favorite phage genes/genomes

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

## PHANTOME LABS

- PhAnToMe Labs (http://www.phantome.org/index.php/labs) is a collection of projects in development by the PhAnToMe team or outside programmers.
- The Labs currently include:
  - Phage Eco-locator (assessing phage distribution in metagenomes)
  - A repeat finder for identifying prophage attachment sites
  - A prophage-finding pipeline