

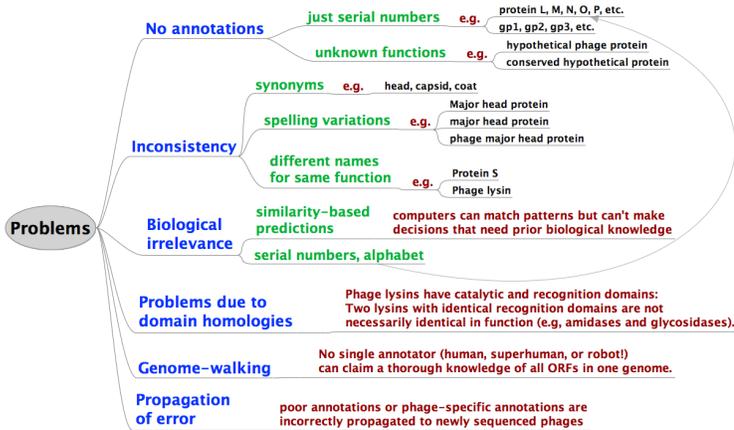
Ramy K. Aziz^a, Brad Hull^a, Bhakti Dwivedi^b, Joe Anderson^c, Bonnie Hurwitz^d, JP Massar^e, Matthew Sullivan^d, Jeff Elhai^c, Mya Breitbart^b, Ross Overbeek^f, Robert A. Edwards^a

^aSan Diego State University, San Diego, CA; ^bUniversity of South Florida, Saint Petersburg, FL; ^cVirginia Commonwealth University, Richmond, VA; ^dUniversity of Arizona, Tucson, AZ; ^eBerkeley, California; ^fThe Fellowship for Interpretation of Genomes, Burr Ridge, IL.

MOTIVATION

- Phages are the most ubiquitous and diverse biological entities on Earth. Understanding the function and evolution of their genomes is critical.
- The number of publicly available complete phage genome sequences is growing exponentially; nevertheless, the interpretation of these genomes suffers from the lack of systematic, consistent, well-maintained systems for phage genome annotation.
- Additionally, a "biologist-friendly" resource for bioinformatics analysis of phage genomes is desperately needed.

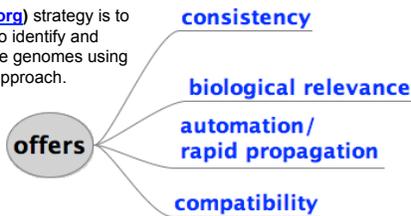
PROBLEMS WITH CURRENT PHAGE ANNOTATIONS



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.

➤ 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.

Example subsystems:

- Universal functions:
 - structural modules (capsids, tails, tail fibers, etc.)
 - functional modules (replication, lysis, integration, etc.)

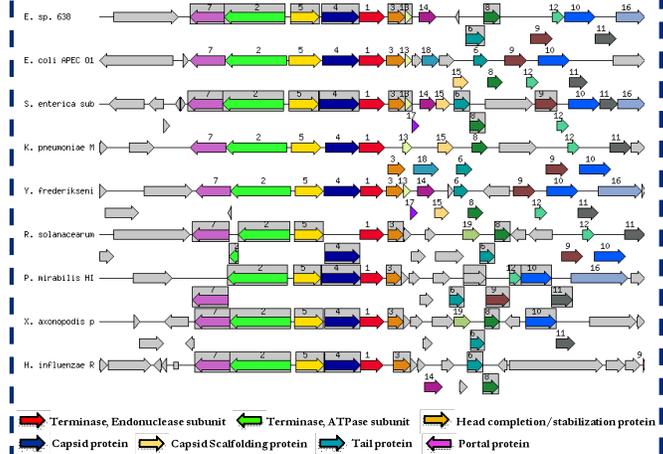
- Phage family-specific functions:
 - T4-like phage core proteins, T4-like cyanophage core proteins, T7-like phage core proteins, etc.

Genome	Lys	Hol	Rz/Rz1
Bacteriophage Lambda	+	+	+
Bacteriophage 933W	+	+	+
...			
<i>L. Lactis</i> bacteriophage ul36	+	+	-

Example Subsystem: Phage lysis modules
Lys = Phage lysin; Hol = Phage holin; Rz/Rz1 = Phage spanins

ENVIRONMENT: PHAGE SEED

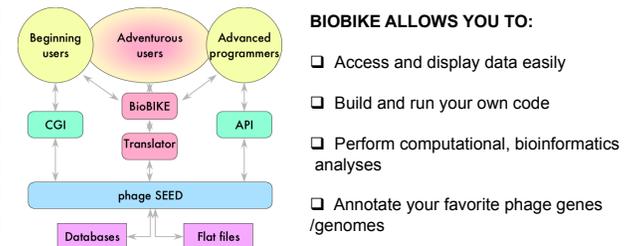
➤ The SEED database allows the analysis and visualization of multiple genomes, and allows easy-identification of functionally-coupled genes within chromosomal clusters.



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

PROGRAMMING INTERFACE: BioBIKE

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



➤ 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.

ANNOTATION PIPELINE: PHAGE RAST

➤ Phage RAST (PhiRAST, <http://www.phantome.org/phiRAST>) is an automatic annotation pipeline for Rapid Annotation using Subsystems Technology optimized to phage genomes.

➤ PhiRAST takes as an input a FASTA-formatted sequence file or a GBK-formatted pre-annotated file.

➤ The annotation pipeline includes optional gene calling and RNA finding steps, followed by gene annotation, subsystems assignment, and preliminary reconstruction of structural and functional modules in uploaded phages.