

Interactive Annotation of Bacteriophage Genomes

Eric Rasche

Center for Phage Technology
Texas A&M University

2016-03-18



Who Am I / CPT

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

Solutions

Annotation

Demo

Q&A

- Systems Administrator, Programmer, Bioinformatician
- Worked at the CPT for ≈ 3 years
- Completely rewrote org's practices on genome annotation
- ≈ 160 tools/wrappers developed
 - 94 Phage specific
 - 28 for external services (NCBI, Apollo)
 - 26 Other (Small utilities, NGS, comparative genomics, scripts)
- Working on next generation of genome annotation infrastructure

Who Am I / Galaxy

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

Solutions

Annotation

Demo

Q&A

- Co-developer of Galaxy Interactive Environments
- Galaxy Committer
 - Involved in: cargo-port, starforge, planemo, pulsar, galaxy
- Galaxy Tool Developer
 - JBrowse in Galaxy
 - NCBI Entrez Suite
 - HMMER3 Suite
 - ART
 - Seqtk suite
 - progressiveMauve

JBrowse

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

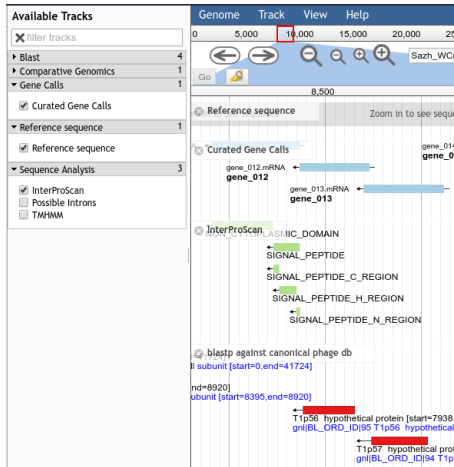
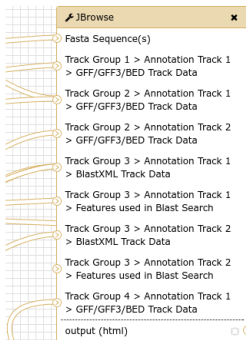
Problems

Solutions

Annotation

Demo

Q&A



Bacteriophages / Background

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

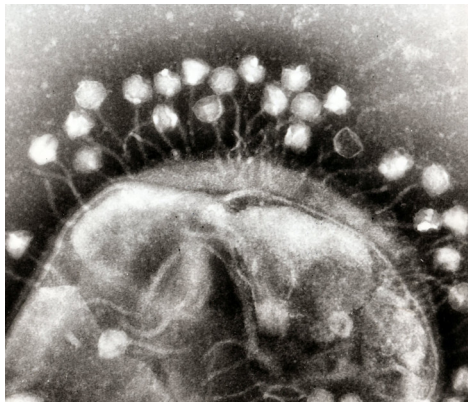
Solutions

Annotation

Demo

Q&A

- Viruses which attack bacteria
- Some are lytic, some are lysogenic



Caudovirales

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

Solutions

Annotation

Demo

Q&A



Myoviridae, Podoviridae, and Siphoviridae

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

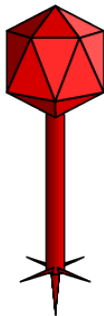
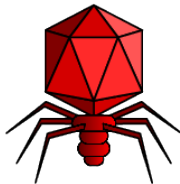
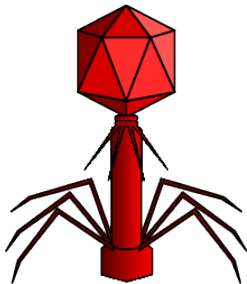
Problems

Solutions

Annotation

Demo

Q&A



Why they're great

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

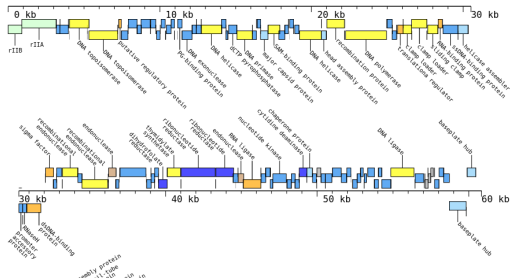
Solutions

Annotation

Demo

Q&A

- Very practical organism to study
- Grow quickly (200 virions per 20 minutes)
- Small, easy to annotate



Coming Wave of Phage Genomics

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

Solutions

Annotation

Demo

Q&A

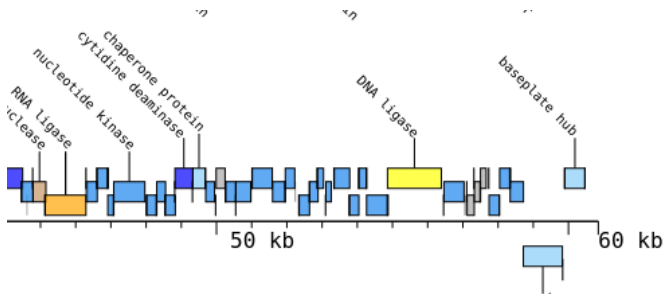
- Phages are incredibly diverse
- Relatively poor representation in databases
- Very practical organism
- Phage Therapy & end of antibiotic era

Phage Therapy

Phage Annotation

Eric Rasche

- Cannot contain generalized transducers
- Cannot be lysogenic
- Cannot contain toxin genes
- \implies need well annotated phages



Special Problems of Phage Genomics

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

Solutions

Annotation

Demo

Q&A

- High gene density, lots of overlap
- Small genomes, cost effective sequencing?



Special Problems of Phage Genomics

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

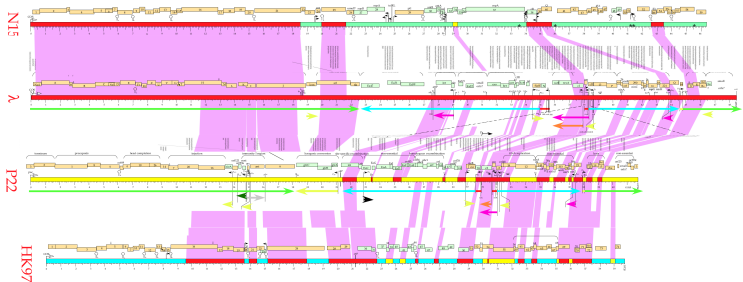
Solutions

Annotation

Demo

Q&A

- Mosaic structure issues
- High mutation rates, 10-100X bacteria
- High diversity + high recombination \Rightarrow nearly total lack of reference genomes



Special Problems of Phage Genomics

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

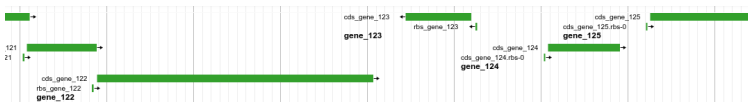
Solutions

Annotation

Demo

Q&A

- Modify RNAPol or encode own, identifying promoters impossible
- Morons inserted in transcripts with independent promoters/terminators



More Problems: Prophages

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

Solutions

Annotation

Demo

Q&A

- Vast majority of “phage” gene sequences in prophages
- ≈ 6 prophage elements per bacterial genome
- Skewed BlastP results
- Functionality of prophage genes unknown

More Problems: Historical

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

Solutions

Annotation

Demo

Q&A

- Inch wide, mile deep
 - Almost all actual experiments done on few paradigms
 - Everything else “known” via computation methods

More Problems: Historical

Phage Annotation

Eric Rasche

■ Within that inch...

- Each paradigm phage had own community
- Each community had own (conflicting) terminology



It wouldn't be a bioinformatics presentation without this graphic

Who

Phages

Who Cares

Problems

Solutions

Annotation

Demo

Q&A

More Problems: Competing Groups

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

Solutions

Annotation

Demo

Q&A

Term	Phage	Meaning
LTF	T4	Long Tail Fiber
	T5	L-shaped Tail Fiber
STF	T4	Short Tail Fiber
	T5	Side Tail Fiber

More Problems: Automation

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

Solutions

Annotation

Demo

Q&A

- Historical gene naming issues
- Renaming fiasco (lambda, T7), cross referencing papers is tough
- Mutation rate
- Little to no ontology based annotation

Redeeming Qualities for Annotation?

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

Solutions

Annotation

Demo

Q&A

- Every caudovirales has list of mandated genes (MCP, portal, scaffolding protein, holin, endolysin, spanin)
- Proteins usually clustered (to avoid segregation by high rec rate)
- Genomic position/neighbours are important and preserved
- High rec rate forces domains close to each other
- Mosaicism can help infer functionality of regions

Undergraduates!

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

Solutions

Annotation

Demo

Q&A

- We train them as phage annotators, they annotate a phage genome in a semester
- Take advantage of human reasoning until software develops further



Complete Genome Sequence of *Klebsiella pneumoniae* Carbapenemase-Producing *K. pneumoniae* Siphophage Sushi

Dat T. Nguyen, Lauren E. Lessor, Jesse L. Cahill, Eric S. Rasche, Gabriel F. Kutry Everett

Center for Phage Technology, Texas A&M University, College Station, Texas, USA

Klebsiella pneumoniae is a Gram-negative bacterium in the family *Enterobacteriaceae*. It is associated with numerous nosocomial infections, including respiratory and urinary tract infections in humans. The following reports the complete genome sequence of *K. pneumoniae* carbapenemase-producing *K. pneumoniae* T1-like siphophage Sushi and describes its major features.

Received 20 July 2015 Accepted 24 July 2015 Published 3 September 2015

Citation: Nguyen DT, Lessor LE, Cahill JL, Rasche ES, Kutry Everett GF. 2015. Complete genome sequence of *Klebsiella pneumoniae* carbapenemase-producing *K. pneumoniae* siphophage Sushi. *Genome Announcements* 3(3):e00994-15. doi:10.1128/genomea.00994-15.

Copyright © 2015 Nguyen et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported License](#).

Address correspondence to Gabriel F. Kutry Everett, cp1@tamu.edu.

Automation

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

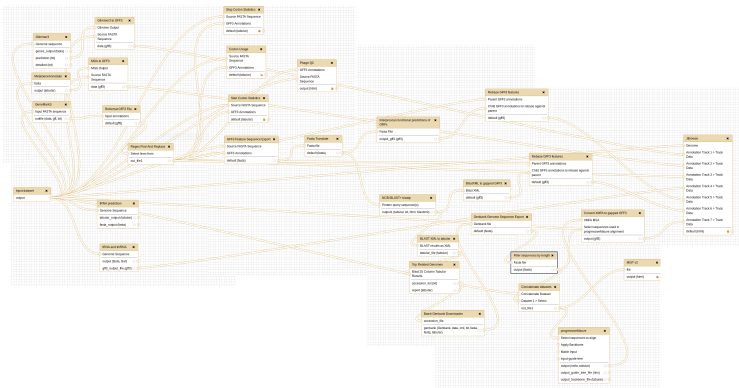
Solutions

Annotation

Demo

Q&A

- Great test bed for automation
- Software is *not* time/memory/CPU constrained
- GO can be incredibly useful



Apollo

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

Solutions

Annotation

Demo

Q&A

- Genome Annotation Client
- “Google Docs” of Genome annotation
- Only way forward for annotation efforts. Everyone should use it.

Apollo

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

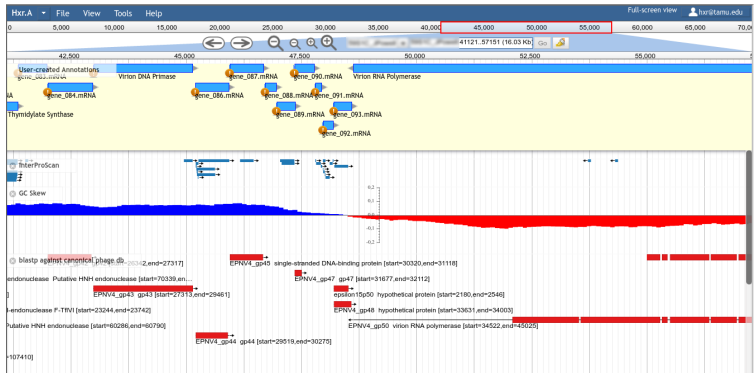
Problems

Solutions

Annotation

Demo

Q&A



Annotation in Practice

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

Solutions

Annotation

Demo

Q&A

Theorem

$$Galaxy + Apollo = <3 \quad (1)$$

Proof.

- 1 Have shaved 5 weeks off of course
- 2 Replaced with deeper dives into software
- 3 Added new activities



Annotation in the Future

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

Solutions

Annotation

Demo

Q&A

Theorem

$$Galaxy + Apollo = < 3 \cdot 10^{10^{10}} \quad (2)$$

via Galaxy Credentials for Remote Services.

Proof is left as an exercise for the reader.

Annotation in our Course

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

Solutions

Annotation

Demo

Q&A

Two main parts:

- Structural annotation
- Functional annotation

Data Movement

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

Solutions

Annotation

Demo

Q&A

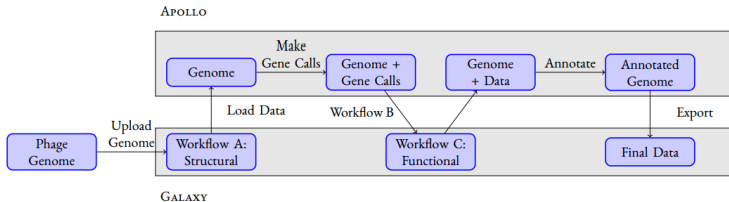


Figure: The data pathway through Apollo, Galaxy, and time

Demo

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

Solutions

Annotation

Demo

Q&A

■ Quick demo of Apollo's functionality

Q&A

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Problems

Solutions

Annotation

Demo

Q&A

- Thank you for your time!
- Questions?