Phage Annotation

Eric Rasche

VVIIO

Phages

Annotation

Solutions

Annotation

Demo

A&Q

Interactive Annotation of Bacteriophage Genomes

Eric Rasche

Center for Phage Technology
Texas A&M University

2016-03-18





Who Am I / CPT

Phage Annotation

Fric Rasche

Who

Phages

.

. .

Annotation

Solutions

Annotation

Demo

A&Q

- Systems Administrator, Programmer, Bioinformatician
- Worked at the CPT for \approx 3 years
- Completely rewrote org's practices on genome annotation
- $\blacksquare \approx 160 \text{ tools/wrappers developed}$
 - 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

_

Annotation

Caludiana

Annotation

D

O&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

Fric Rasche

Who

Phages

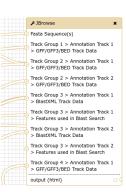
.

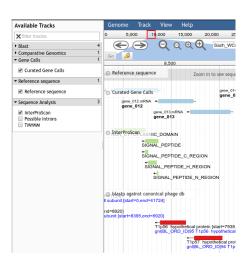
Annotation

- - - - - -

Annotatio

Demo





Bacteriophages / Background

Phage Annotation

Eric Rasche

\//ho

Phages

_

Annotation

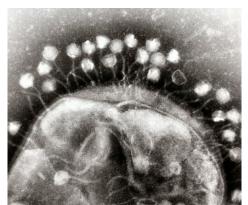
Annotation

Solutions

Annotation

Dama

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



Caudovirales

Phage Annotation

Eric Rasche

Who

Phages

Who Cares

Annotation

Caludiana

Jointions

Annotation

Demo



Myoviridae, Podoviridae, and Siphoviridae

Phage Annotation

Eric Rasche

VA /I. .

Phages

\//h= C==

Annotation

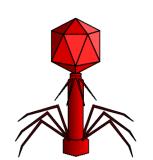
7 (11110) (4) (1)

Solution

Annotation

D

08/0







Why they're great

Phage Annotation

Eric Rasche

Who

Phages

Who Care

Annotation

7 (1111010101

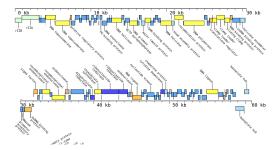
Solutions

Annotation

D

A&Q

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



Coming Wave of Phage Genomics

Phage Annotation

Fric Rasche

Who

Phages

Who Cares

Annotation

Caludiana

Annotation

_

O&A

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

Phage Therapy

Phage Annotation

Fric Rasche

Who

Who Cares

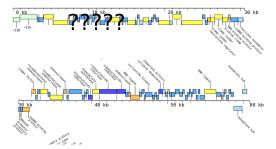
Annotation

Annotatio

Solutions

Annotation

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



Special Problems of Phage Genomics

Phage Annotation

Fric Rasche

Who

Phages

Annotation

Annotation

Solutions

Annotation

_

A&Q

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



Special Problems of Phage Genomics

Phage Annotation

Eric Rasche

Who

Phages

Who Co

Annotation

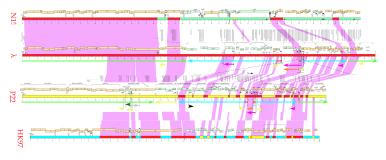
C = 1...+:====

- - - - - -

Annotation

Demo

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



Special Problems of Phage Genomics

Phage Annotation

Fric Rasche

Who

Phages

\//ba Ca

Annotation

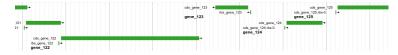
, uniocació

Solutions

Annotation

D

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



More Problems: Prophages

Phage Annotation

Fric Rasche

Who

Phages

Who Care

Annotation

Solutions

Annotation

D

O&A

- Vast majority of "phage" gene sequencse 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 Care

Annotation

C = 1...+:====

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

Fric Rasche

vvno

rnages

Annotation

Solutions

_

Q&A

- Within that inch...
 - Each paradigm phage had own community
 - Each community had own (conflicting) terminology

HOW STANDARDS PROLIFERATE: (SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC.) 500N: 14?! RIDICULOUS! WE NEED TO DEVELOP ONE UNIVERSAL STANDARD SITUATION: SITUATION: THAT COVERS EVERYONE'S THERE ARE THERE ARE USE CASES. 14 COMPETING 15 COMPETING STANDARDS. STANDARDS.

More Problems: Competing Groups

Phage Annotation

Eric Rasche

VVIIO

Phage:

Who Cares

Annotation

Joidtions

Annotation

Demo

A&Q

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

Fric Rasche

Who

Phages

Who Care

Annotation

.

Solutions

Annotation

Demo

- 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

Fric Rasche

Who

Pnages

Annotation

Annotation

Demo

- 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

Annotation

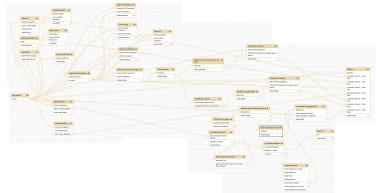
Annotation

Solutions

Annotation

Demo

- We train them as phage annotators, they annotate a phage genome in a semester
- Human brain power until Al/software develops further



Automation

Phage Annotation

Eric Rasche

Who

Pnages

Who Car

Annotation

Solutions

Annotation

D

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

Apollo

Phage Annotation

Eric Rasche

VVIIO

rnages

Annotation

. . . .

Annotation

_

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

Apollo

Phage Annotation

Eric Rasche

\//ho

Phages

.

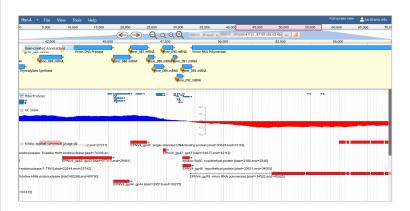
Annotation

Annotatio

Solution

Annotation

Demo



Annotation in Practice

Phage Annotation

Eric Rasche

Who

Phages

\//ba Caus

Annotation

7 (111101010101

Annotation

Demo

A&C

Theorem

$$Galaxy + Apollo = <3$$
 (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

Annotation

Alliotatio

Solutions

Annotation

Dem

A&C

Theorem

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

via Galaxy Credentials for Remote Services

Annotation in our Course

Phage Annotation

Eric Rasche

vvno

i ilages

Annotation

Annotation

.

Two main parts:

- Structural annotation
- Functional annotation

Data Movement

Phage Annotation

Eric Rasche

Who

Phages

....

Annotatio

Annotatio

Solutions

Annotation

Demo

A&C

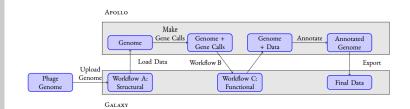


Figure: The data pathway through Apollo, Galaxy, and time

Demo

Phage Annotation

Eric Rasche

Who

Phages

Who Care

Annotation

7 (11110121101

Solutions

Annotation

Demo

A&Q

■ Quick demo of Apollo's functionality

Q&A

Phage Annotation

Eric Rasche

VVho

Phages

Who Care

Annotation

C = 1...t. = = =

- - - - - -

Annotation

Demo

- Thank you for your time!
- Questions?