

HOW TO TRAIN YOUR STUDENT TO DO BIOINFORMATICS IN THE GALAXY ECOSYSTEM USING PHAGE

Jolene Ramsey^{1,2}, Curtis Ross^{1,2}, Anthony Criscione^{1,2}, Mei Liu^{1,2}, Ry Young^{1,2}, Jason J. Gill^{1,3}

¹Center for Phage Technology, Texas A&M University, College Station, Texas, USA

²Department of Biochemistry & Biophysics, Texas A&M University, College Station, Texas, USA

³Department of Animal Science, Texas A&M University, College Station, Texas, USA



CENTER FOR PHAGE TECHNOLOGY

BICH464: BACTERIOPHAGE GENOMICS

A full semester upper-level phage research course, suitable for in-person or online synchronous or asynchronous delivery

Traditional lectures
Front-loaded phage biology content

Wet bench lab
Phage isolation training

Computer lab
Scaffold hands-on bioinformatic exercises in web browser

★ 3 student products ★

- 1- Genbank file of their annotated phage genome
- 2- Manuscript to submit to peer-reviewed journal
- 3- Novel phage they isolated from environment

SELF-PACED TUTORIALS, VIDEOS, & FAQ

<https://cpt.tamu.edu/training-material/>

Phage Annotation Pipeline in CPT Galaxy

Additional Analyses

Center for Phage Technology at Texas A&M University

YouTube

ABSTRACT

In the modern genomic era, students without computer programming skills can perform advanced computational analyses to answer biological questions. Bacteriophage, the viruses of bacteria, represent an affordable, historically powerful, and currently relevant model for training in computational research. Unlike the genomes of most cellular life forms, phage genomes are usually a single contiguous molecule <200,000 bases in length. Such a small genome size permits speedy analyses and enforcement of strict high-quality standards for research products in the genomics education setting. At the Center for Phage Technology (CPT), we use two open source, web-based platforms: Galaxy, for reproducible computational analyses, and Apollo, a collaborative genome annotation editor, to facilitate annotation of phage genomes by undergraduate students (<https://cpt.tamu.edu/galaxy-pub/>). The CPT has trained a steady stream of undergraduate and graduate students, as well as scientists, both informally and through formal university course offerings to use semi-automated workflows in Galaxy and Apollo for collaborative annotation of genomes, including feature calling, contextualized functional prediction, and comparative genomics. A series of tutorials for self-paced training in the Galaxy Training Network format accompanies our platform at <https://cpt.tamu.edu/training-material/>.

During computer lab time in our semester-long upper level undergraduate course called Bacteriophage Genomics, students are provided with assembled, complete DNA contigs. In the user-friendly Galaxy web browser front-end for powerful command-line bioinformatics tools, students collaboratively annotate in the Apollo genome browser interface. Their data and activities are recorded in the platform history which facilitates reproducibility, reanalysis, instructor assessment, and constructing the methods section for a manuscript. Being a legitimate guided research experience, Bacteriophage Genomics demands creativity, persistence in the face of uncertainty, and application of disciplinary knowledge and skills as they interpret the bioinformatic results for their project. Ultimately, the course culminates in two student-owned products: the Genbank file of an annotated phage genome for deposition into the NCBI public database and a manuscript describing that genome ready for submission to a peer-reviewed journal. The resulting data and student publications from this training effort are continuously collated on our BioProject page (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA222858>). We note that, unsurprisingly, the fully online nature of the system lent itself to a seamless transition to virtual instructional delivery during the COVID-19 pandemic. In the long-term, our trained student researchers often go on to participate in additional laboratory research or apply to graduate or medical school, while others seek employment in the biotechnology industry.

START HERE

STUDENTS & SCIENTISTS

+ >Phage genome
GGTGTGCGCCCACTTCGTGCACTC
CGAACCCTCAGGATCTACCTT...

CPT PHAGE GALAXY

<https://cpt.tamu.edu/galaxy-pub/>

Input & output data in an ordered, annotatable history

Internal & external bioinformatic tool selection

Web-based portal for command line tools with point-and-click parameter setting that chains into automatic & reproducible workflows

Queued job, Running job, Complete jobs, Inline preview

Status feedback

Upload, retrieve via API, or download to save data
Run tools directly from web browser
Perform new analyses iteratively
Rerun jobs with same or tweaked settings
Easily convert file formats
Send data to and from Apollo for annotation

CPT Apollo Collaborative genome annotation editor

Contextualized functional prediction from evidence

Dynamic evidence tracks from Galaxy tools

Analyze contextualized data dynamically
Make informed annotation decisions
Real-time online collaboration in browser

Data exchange

STRUCTURAL WORKFLOW

calling gene features, terminators

23 steps
~15 min to complete

Visual output

Structural Order of Operations:

- tRNAs
- Protein-coding genes
- MetaGeneAnnotator/Glimmer > Sixpack
- Check starts, trade-offs for maximizing coverage, orientation/morons
- Terminators

Inspect sequence

FUNCTIONAL WORKFLOW

Predicting protein function by sequence properties & similarity to known proteins

57 steps, completion time genome size-dependent

Functional Order of Operations:

- BLASTp/BLASTn
- Canonical phages > SwissProt > NR
- InterProScan (conserved domains), transmembrane domains
- Tricky features: Introns, Inteins, Frameshifts, Lysis genes

EXIT

PUBLIC DATABASES

Deposit curated data

Retrieve public data for re-analysis

RE-ENTER

FIND MORE

Center for Phage Technology (CPT)
www.cpt.tamu.edu
@TAMU_CPT
Center for Phage Technology at Texas A&M University

CPT Galaxy & Apollo instances **apollo**
www.cpt.tamu.edu/galaxy-pub/

Ramsey & Rasche *et al.* (2020) PLoS Comput Biol
16(11):e1008214. doi: [10.1371/journal.pcbi.1008214](https://doi.org/10.1371/journal.pcbi.1008214)

CPT Training Materials
www.cpt.tamu.edu/training-material/

CPT BioProject data & publications list
www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA222858

FUNDING

NSF Awards EF-0949351 & DBI-1565146 to R.Y. and J.J.G.

CENTER FOR PHAGE TECHNOLOGY TEXAS A&M UNIVERSITY

TEXAS A&M AGRILIFE RESEARCH

ACKNOWLEDGEMENTS

CPT lab members Apollo team Galaxy community <https://training.galaxyproject.org/>

COMPARATIVE WORKFLOW

Phage relationships @ nucleotide & amino acid levels

11 steps
<20 min to complete

Nucleotide Similarity Table

Genome Dot Plot

Protein Similarity Table

STANDALONE TOOLS

Draw genome maps

Predict termini