

Enhancing Student Engagement with Online Annotation of Bacteriophage Genomes

Eric Rasche¹, Jason Gill², Ryland Young^{1, 3}

1. Center for Phage Technology, Texas A&M University, College Station, United States
2. Department of Animal Science, Texas A&M University, College Station, United States
3. Department of Biochemistry and Biophysics, Texas A&M University, College Station, United States

BICH464: Phage Genomics for undergraduates

Semester long, 3 credit course taught for 12 years (pre-dates HHMI Phage Hunters). During the course, each student:

- Annotates a complete, novel phage genome
- Prepares and polishes a GenBank file for submission
- Writes a 500 word submission for ASM Genome Announcements

The students additionally isolate novel phages from the environment for future years.

The ultimate goal of the course is to expose students to real research, including the time demands, requirement for initiative, phage biology, and fundamental genomics required.

Why Undergraduate Phage Genomics?

- Phages are ideal—small size, high coding density
- Can be annotated within a single semester, by a single undergraduate
- Sequencing and assembly are no longer the throughput-limiting step.
- Not computationally intensive, analysis steps run in fraction of time of bacteria/eukaryotes

Previous Course Iterations

Students:

- Students hated the command line!
- Struggled to copy command line statements and failed to understand the relationship between their printed protocols and their screen.

Staff:

- Documenting parameters students used in software was impossible (student typos, closed-source websites).
- Dependence on mutable third party services was quite problematic for reproducibility.

These issues resulted in sub-optimal experience for everyone involved, and some steps were not strictly reproducible.

CPT Viral Annotation Innovations

Use Galaxy:

- Galaxy provides a single interface to all software, from BLAST to InterProScan to TMHMM.
- Galaxy is a robust system with extremely large community supporting development and features. Not just for eukaryotic genomics!
- *100% reproducible science*.
- Automatic tracking of user's files and tool run history.
- Perfect documentation of commands run and parameters used.

Workflows!

- Students run complete, automated workflows instead of running individual tools, or typing in complex command line statements.
- Everyone runs identical workflows, and admins can pre-configure steps, eliminating user error rate.
- More time for teaching and bioinformatics.

Apollo is the Future:

- Can adapt the widely used eukaryotic system for viral annotation needs.
- Students only see the most up-to-date copy, which we transparently and automatically back up.
- Students cannot lose files or progress. They cannot accidentally switch to an old version.
- Unified display of outputs from very different tools.

Current Status & Future Plans

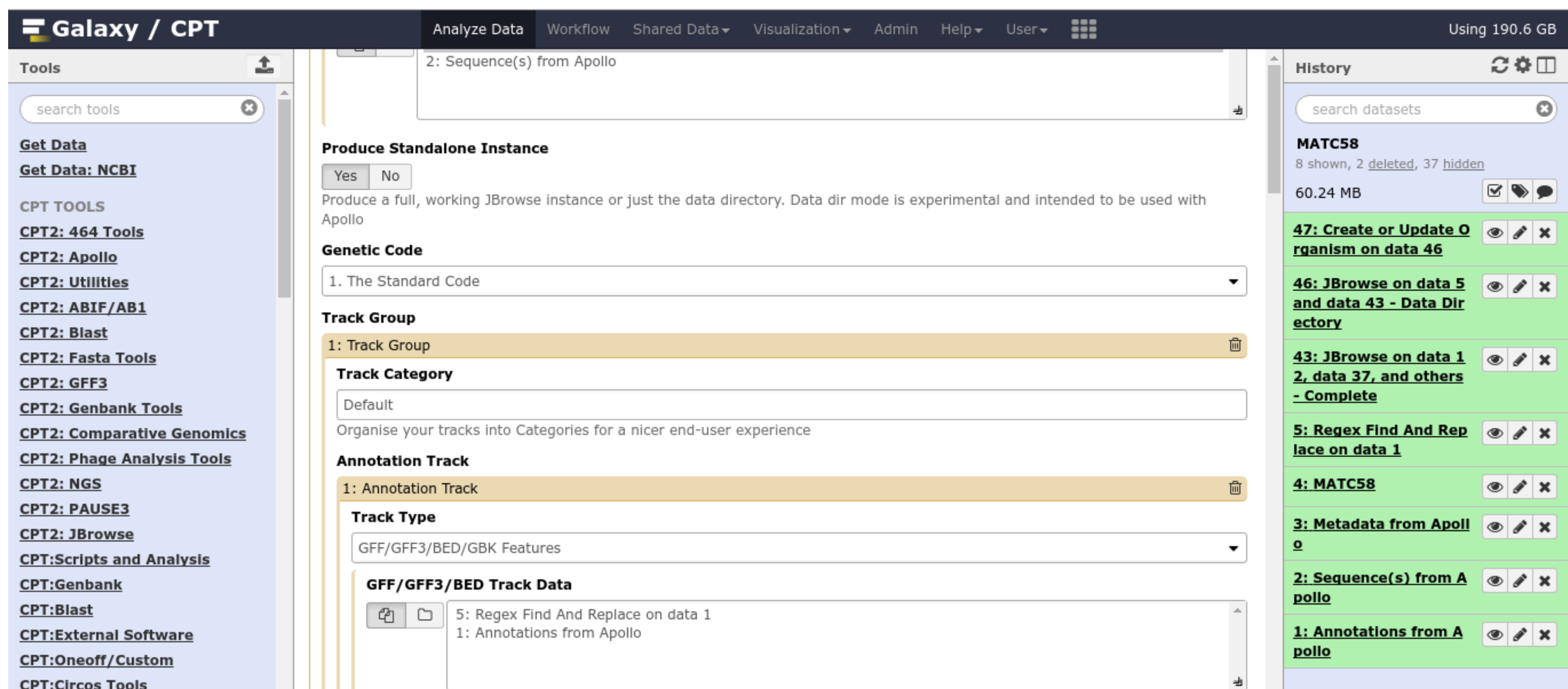
Try it Today! <https://cpt.tamu.edu/beta>

- Open Beta for Galaxy & Apollo annotation workflow.
- All work is open source, and community-focused.
- NSF funding for further development of community viral annotation (*de novo* and re-annotation) efforts.

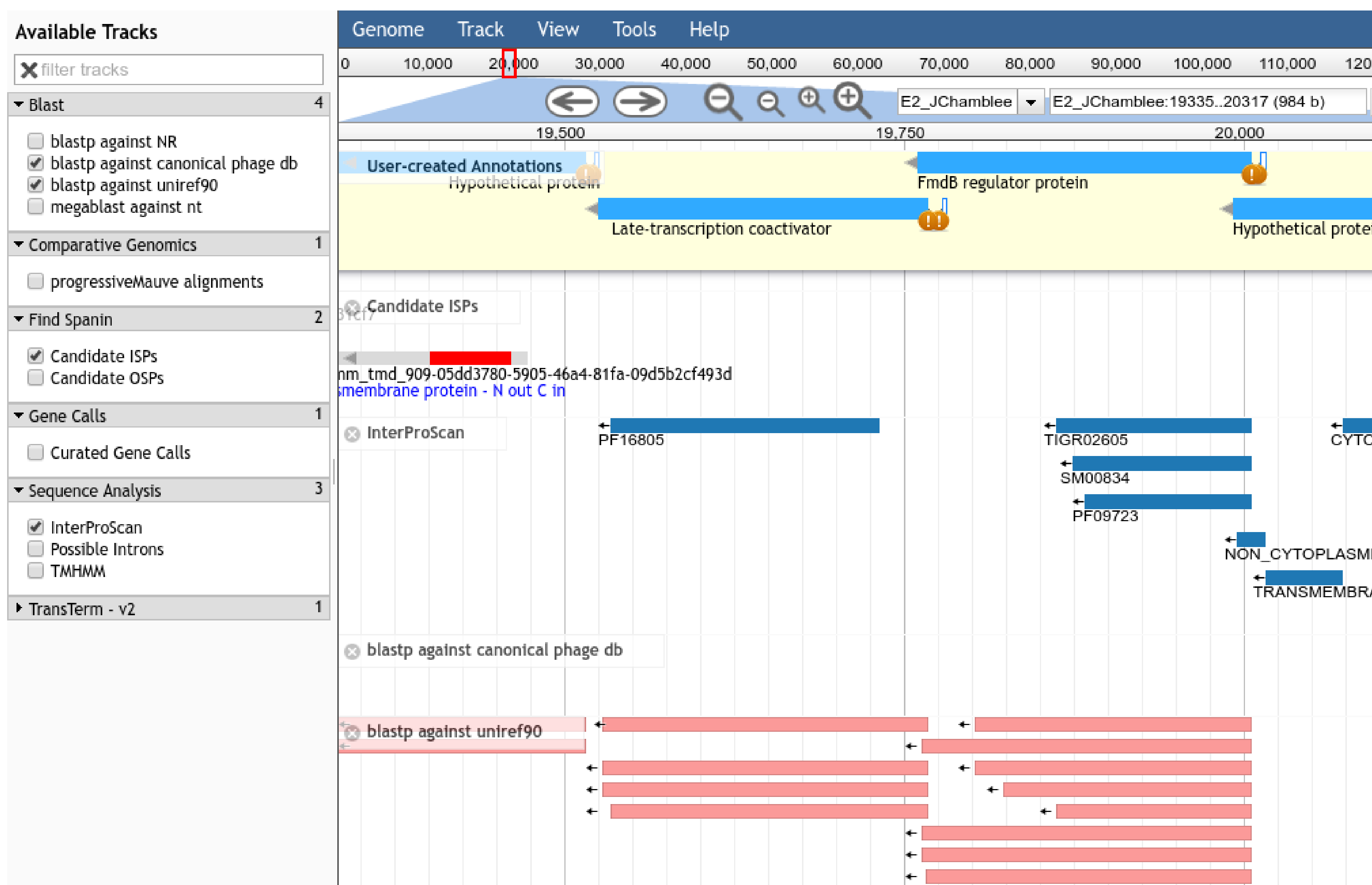
Future

- Easily clone our infrastructure, and deploy your own annotation community!

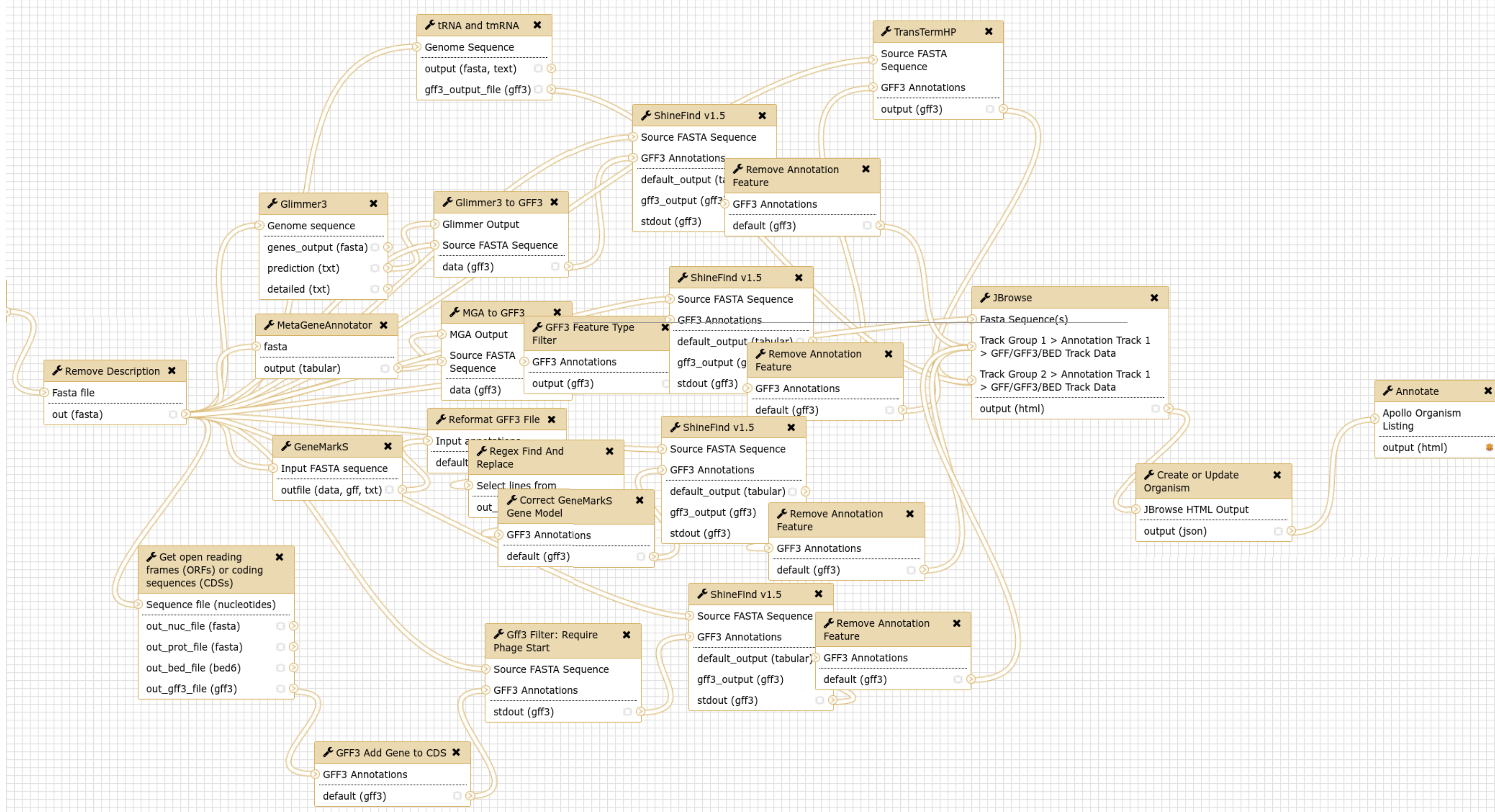
Galaxy & Apollo: Automation & Unified Results Display



In **Galaxy**, every tool has a standardized interface. Once students learn the interface, they can use any of the ≈ 600 tools available in the CPT's Galaxy, just by understanding the tool's use case. Tool runs are automatically recorded in their *history*, like an automated lab notebook.



Apollo, the online, real-time, collaborative genome annotation suite, has significantly increased student/staff productivity and decreased IT burden. Results from highly disparate HTML pages, text files, blast reports, etc. are visually unified, allowing researchers to focus on biology rather than on interpreting reports.



Galaxy Workflows enable researchers to develop common analysis units such as “Gene Calling”, and to adjust these workflows whenever new and improved methods become available. We recently added MetaGeneAnnotator to our gene callers. Students and researchers could then compare identically-styled GeneMarkS, Glimmer3, and MGA outputs in Apollo. They could bypass the opaqueness of large text reports, and cut right to the biology.