Enhancing Student Engagement with Online Annotation of Bacteriophage Genomes

CPT — Center for PHAGE TECHNOLOGY

Eric Rasche[1], Jason Gill[2], 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. Biochemistry and Biophysics, Texas A&M, College Station, United States

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.
- Great test bed for teaching bioinformatics. Many computations are extremely fast (seconds to hours), which allows for more interactive exploration.

Previous Course Iterations

Students

- **Students hated the command line!**
- They also struggled to correctly copy command line statements and failed to understand the relationship between their printed protocols and what they were doing on the screen.
- Keeping track of a huge number of files from various tool outputs was tough for untrained undergraduates.

Staff

- Could not keep tabs on student progress throughout the semester.
- Documenting the exact command line and parameters with which students ran software was impossible (typos, websites not under our control).
- Strong dependence on mutable third party services was concerning for reproducibility.

These issues resulted in *Completely Unreproducible Annotations* and a sub-optimal experience for everyone involved.

CPT Innovations

Galaxy

- Allowed the switch from a variety of third party websites to a local Galaxy deployment.
- Galaxy provides a single interface to all software, from BLAST to InterProScan to TMHMM.
- Automatic tracking of user's files and tool run history.
- Automatic, standardized file organisation across all students.
- Perfect documentation of precise commands run.
- We could non-intrusively observe student annotation progress.

Workflows

- Students run workflows instead of tools one-by-one.
- Everyone runs identical workflows, and admins can pre-configure steps, reducing user error rate.

Apollo

- 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.
- Staff could non-intrusively observe student annotation progress.
- Unified display of outputs of very different tools.

Galaxy & Apollo: Automation & Unified Results Display
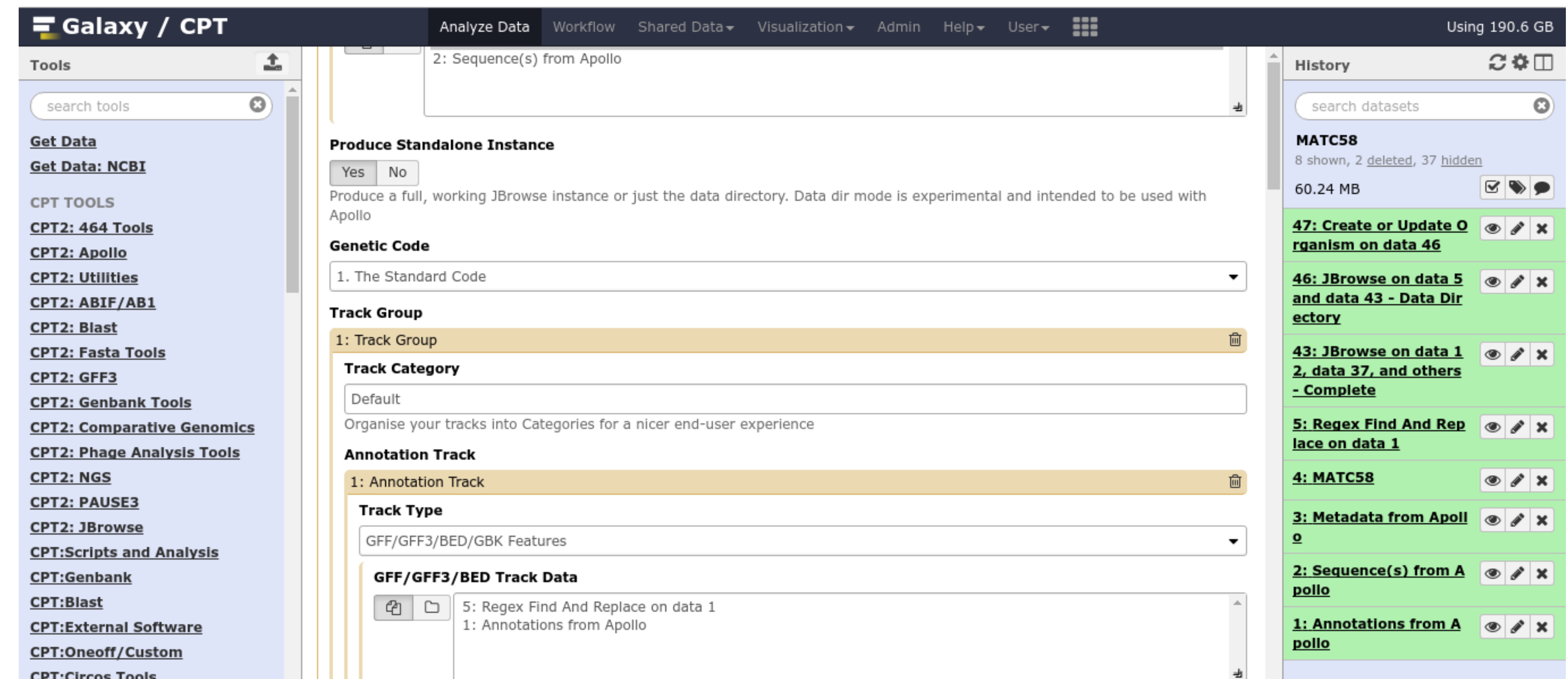
In **Galaxy**, every tool has a standardized interface. Once students learn the interface, they can use any of the ≈600 tools available in the CPT's Galaxy, just by understanding the tool's use case.
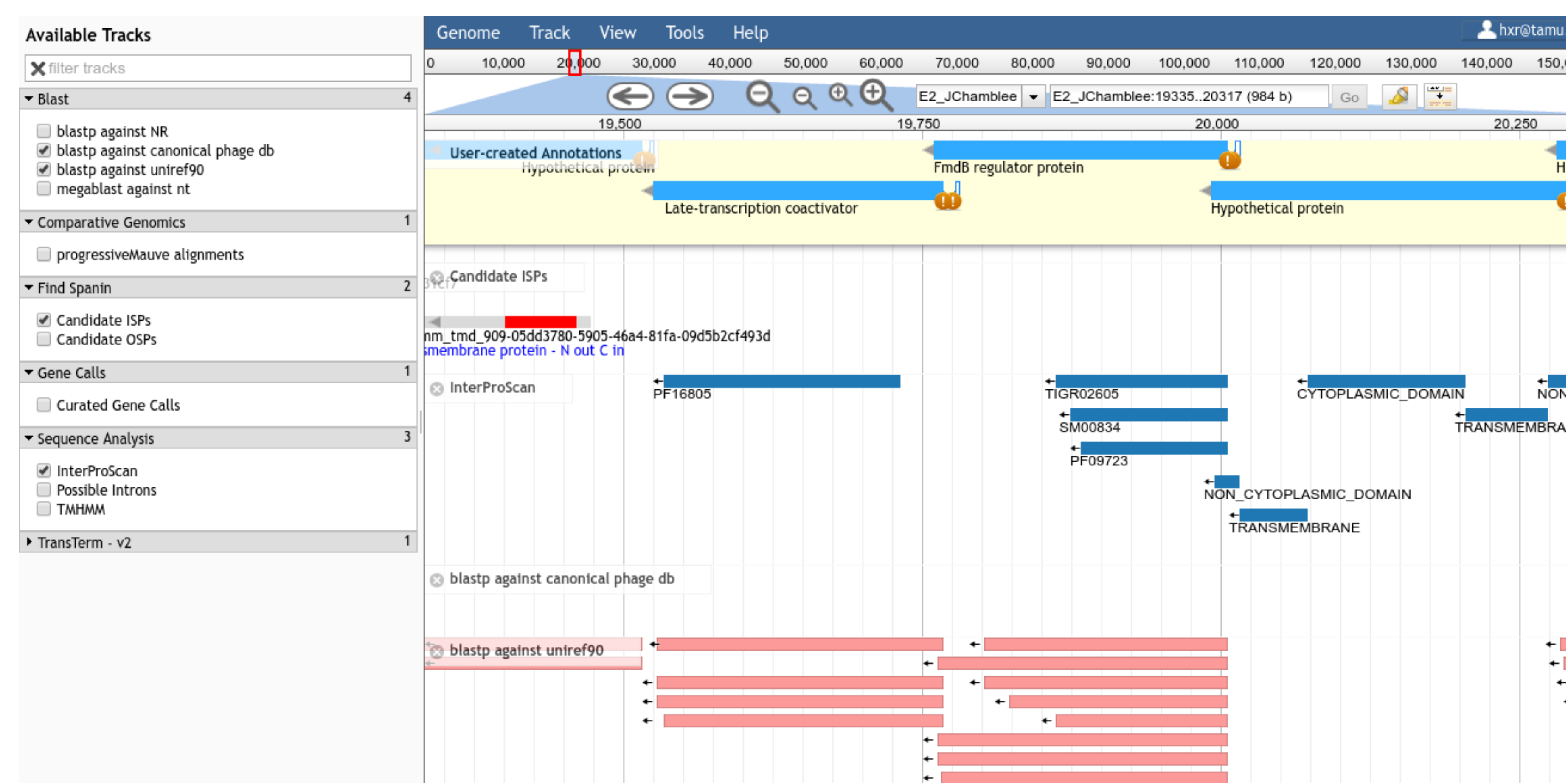
Apollo, the online collaborative genome annotation suite, has significantly increased student/staff collaboration and productivity and decreased IT burden. Also visible: display of heterogeneous data within the Apollo Genome Browser. Results from highly disparate HTML pages, text files, etc. are visually unified, allowing researchers to focus on biology.
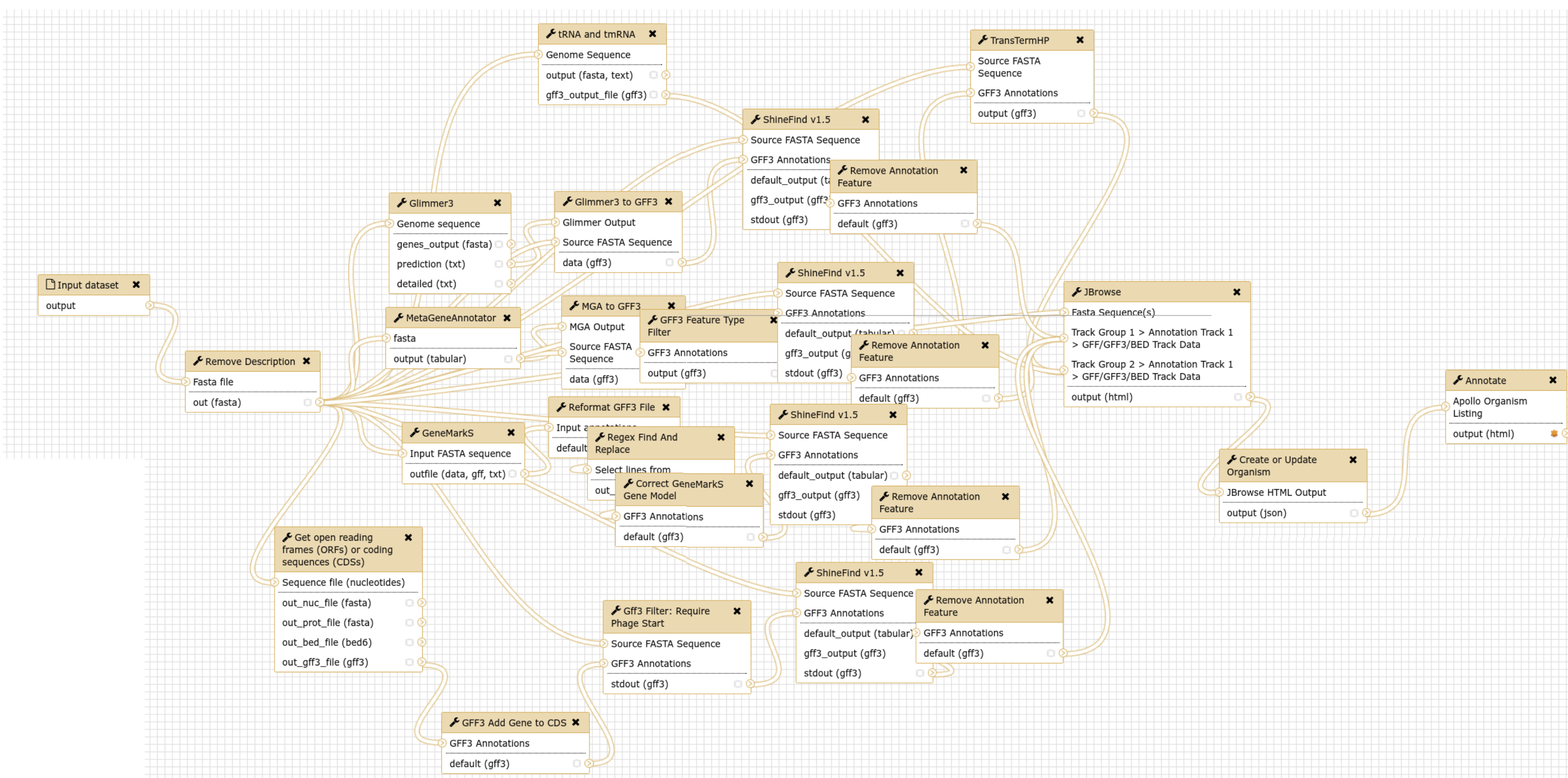
Galaxy Workflows enable researchers to develop common analysis units such as "Gene Calling", and to adjust these workflows whenever new and improved methods become available. We recently added MetaGeneAnnotator to our gene callers. Students and researchers could then compare identically-styled GeneMarkS, Glimmer3, and MGA outputs in Apollo, bypassing the opaqueness of large text files and cutting right to the biology.