Using genome browsers constructed by G-OnRamp to provide students with a Course-based Undergraduate Research Experience in genome annotation

Wilson Leung[1], Luke Sargent[2], Yating Liu[1], Nathan T. Mortimer[3], David Lopatto[4], Jeremy Goecks[2], Sarah C. R. Elgin[1]

[1]Washington University in St. Louis, MO; [2]Oregon Health & Science University, OR; [3]Illinois State University, IL; [4]Grinnell College, IA

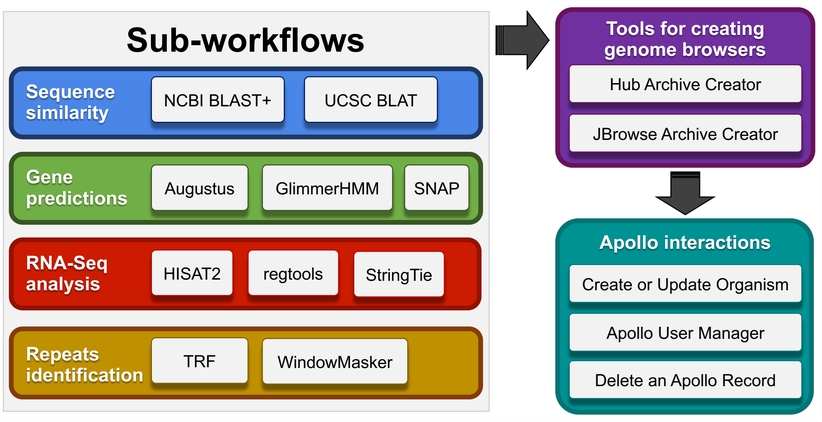
G-OnRamp (http://g-onramp.org) provides an easy-to-use web platform for educators to create genome browsers to engage undergraduate students in research projects, both collaborative annotation of eukaryotic genes/genomes and "big data" biomedical analyses

Abstract

Course-based Undergraduate Research Experiences (CUREs) based on genome annotation are beneficial to researchers, educators, and students alike. They provide researchers with high quality gene models and provide educators with an effective way to teach students about eukaryotic genes/genomes. Genome browsers provide visualizations that facilitate the synthesis of multiple types of experimental and computational evidence for constructing gene models. To reduce the technical expertise required to construct genome browsers, the Genomics Education Partnership (GEP) and the Galaxy Project (https://galaxyproject.org) have developed G-OnRamp (http://g-onramp.org), a web-based platform for constructing UCSC Assembly Hubs and JBrowse genome browsers with evidence tracks for sequence alignments, gene predictions, RNA-Seq data, and repeats identification. G-OnRamp also provides tools to create and manage Apollo instances for collaborative genome annotations. G-OnRamp has been used to create genome browsers for >20 species (http://g-onramp.org/genome-browsers), including those for a CURE that examined lipid synthesis pathway genes in four parasitoid wasp species. This CURE engaged more than 200 students from 15 diverse institutions. Results from an anonymous survey of G-OnRamp users showed that most respondents find G-OnRamp useful in their research and their teaching; some plan to use it to develop new CUREs. Version 1.1 of G-OnRamp added the capability to incorporate extrinsic evidence into the Augustus gene predictions, and improved compatibility with new versions of Apollo, JBrowse, and Galaxy. G-OnRamp can be deployed locally via a virtual appliance or on the Cloud (Amazon EC2) via CloudLaunch (http://g-onramp.org/deployments). Faculty interested in developing a CURE using G-OnRamp can contact us at http://gep.wustl.edu/contact_us.

G-OnRamp has a modular and flexible architecture

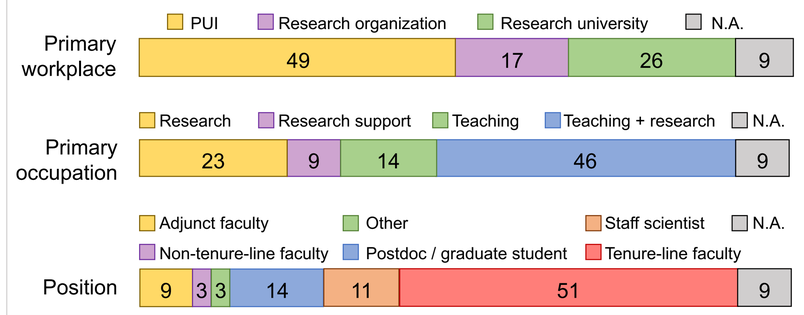
- Add tools and workflows to Galaxy for creating genome browsers
- Analyze genome assemblies using four sub-workflows
- Provide tools for managing and interacting with Apollo
- Use the **Workflow Canvas** to to add tools and customize workflows



G-OnRamp training workshops

- 6 workshops from 2016-2018
- 65 participants from 40+ institutions
- Produced genome browsers for 18 eukaryotic genomes

Demographics of G-OnRamp Workshop Participants (%)

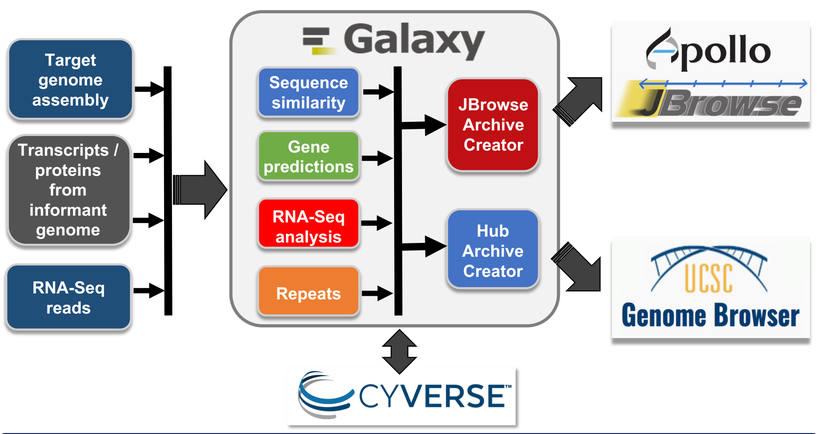


Genome browsers: http://g-onramp.org/genome-browsers

Training materials: http://g-onramp.org/training

G-OnRamp: create genome browsers for eukaryotic genomes

- Create **UCSC Assembly Hubs** and **JBrowse** genome browsers for eukaryotic genomes
- Create **Apollo** instances for real-time collaborative genome annotation in research and education settings
- Upload Assembly Hubs to **CyVerse** for long-term storage



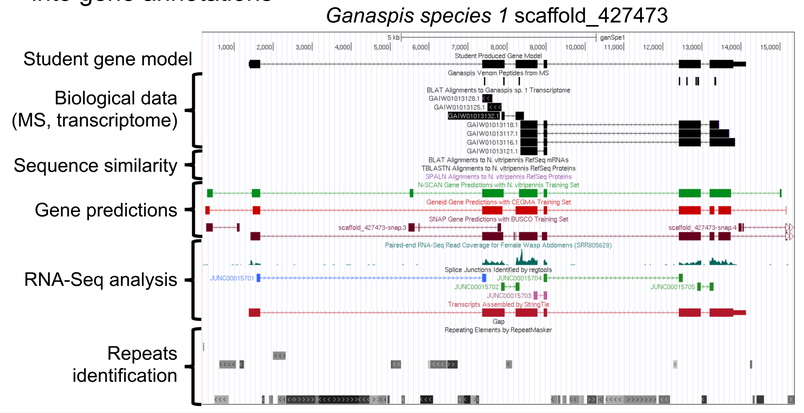
Session on 1/12 @ 8:30am
(Terrace Room - Handlery Hotel)

Comparative gene annotations of four parasitoid wasp species

Research goal: understand how venom proteins from parasitoid wasps manipulate the signal transduction pathways and second messenger system of their hosts

- Engaged more than 200 students from 15 diverse institutions
  - 7 Primarily Undergraduate Institutions; 4 Minority-Serving Institutions
- Incorporate RNA-Seq and protein mass spectrometry (MS) data into gene annotations

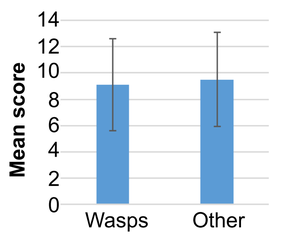
Ganaspis species 1 scaffold_427473



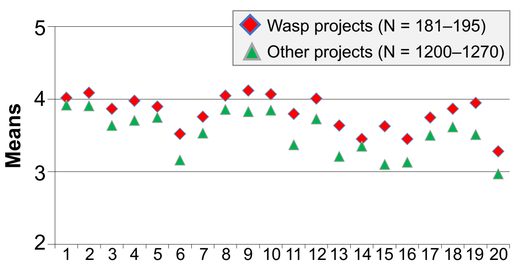
Contact: Nathan T. Mortimer (ntmorti@ilstu.edu)

Students who participated in the wasp project show similar gains compared to other GEP students

Mean post-course test scores



Responses to SURE survey questions



1. Understanding the research process
2. Knowledge construction
3. Readiness for research
4. Tolerance for obstacles
5. Skill interpreting results
6. Clarifying career choices
7. Integrating theory/practice
8. Tackling real problems
9. Assertions need evidence
10. Ability to analyze data
11. Reading / understanding Primary science literature
12. Understanding science
13. Ethical conduct
14. Lab techniques
15. Skill – oral presentation
16. Skill in scientific writing
17. Understanding how scientists think
18. Independence
19. Learning community
20. Teaching potential

GEP + Galaxy = G-OnRamp

The Genomics Education Partnership (http://gep.wustl.edu)

- Nationwide collaboration of **100+ institutions**
- Engages **>1300 students** annually in bioinformatics and genomics
- Integrates active learning into the curriculum through Course-based Undergraduate Research Experiences (**CUREs**)

Poster PE0138

Galaxy PROJECT

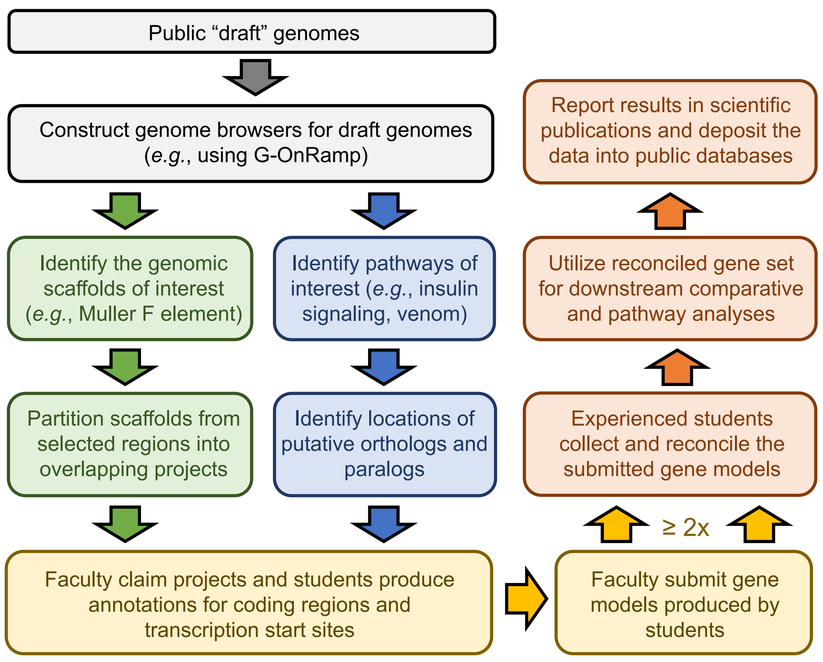
https://galaxyproject.org
open, web-based platform for bioinformatics analyses

- **Accessible:** does not require programming experience
- **Reproducible:** repeat analyses that contain multiple steps
- **Transparent:** share and publish workflows and results

Posters PE0141 and PE0142
Galaxy Session on 1/14 @ 4:00pm (California)

Workflows for creating student annotation projects

- Partition genomic regions or genes of interest into student projects
- Each project done by at least two students working independently, and then reconciled by experienced students for **quality control**



Deployment options

http://g-onramp.org/deployments

G-OnRamp virtual appliance

- Use for local testing and training

G-OnRamp on the Cloud

- Use for production analysis of whole genome assemblies
- Run G-OnRamp on Amazon Web Services via CloudLaunch (https://launch.usegalaxy.org)

G-OnRamp publications

- Liu Y *et al.* Bioinformatics. 2019 Nov 1;35(21):4422-4423
- Sargent L *et al.* bioRxiv 781658; doi: 10.1101/781658

Acknowledgements

- G-OnRamp was supported by NIH grant 1R25 GM119157 awarded to SCRE. The work on parasitoid wasps is supported by NIH grants 1R35 GM133760 and 1R03 AG063314 awarded to NTM.
- GEP is supported by NSF grant #1915544, NIH grant #GM130517, and hosted by The University of Alabama and WUSTL. Galaxy is supported by NIH grant HG006620-04 and OHSU.