

Animal Genomics with Galaxy: Analyze, Publish, and Visualize

Dan Blankenberg

<http://galaxyproject.org>

<http://usegalaxy.org>

<http://getgalaxy.org>

Overview

- What is Galaxy?
- Galaxy for Experimental Biologists
- Galaxy for Bioinformaticians

Galaxy, a web-based genome analysis platform

- An open-source **framework** for integrating various computational tools and databases into a cohesive workspace
- A web-based **service** we provide, integrating many popular tools and resources for comparative genomics
- A completely **self-contained application** for building your own **Galaxy** style sites

Overview

- What is Galaxy?
- Galaxy for Experimental Biologists
- Galaxy for Bioinformaticians

Galaxy: the one-stop shop for Genome Analysis

- Analyze
 - Retrieve data directly from popular data resources or upload your own
 - Interactively manipulate genomic data with a comprehensive and expanding “best-practices” toolset
- Publish and Share
 - Results and step-by-step analysis record (**Data Libraries** and **Histories**)
 - Customizable pipelines (**Workflows**)
 - Complete protocols (**Pages**)
- Visualize
 - Send data results to external Genome Browsers
 - Build reusable AJAX-based custom Genome Browsers (**Trackster**)

Example Demo

- A simple goal: identification of exons containing the largest number of single nucleotide polymorphisms (SNPs)
 - Retrieve genomic data from external data resource
 - Exon locations
 - SNP locations
 - Determine which SNPs overlap exons (**Join**)
 - Determine the number of SNPs for each exon (**Group**)
 - **Sort** by number of SNPs per exon
 - **Filter** to keep exons with highest SNP counts
 - View results in UCSC Genome Browser
 - Create a reusable analysis **Workflow**
 - Share analysis **Results, History** and **Workflow**

Overview

- What is Galaxy?
- Galaxy for Experimental Biologists
- Galaxy for Bioinformaticians

Galaxy: the instant web-based tool and data resource integration platform

- Open Source downloadable package that can be deployed in individual labs
- Zero Configuration, but highly configurable
- Painlessly
 - Run your own private Galaxy Server
 - Add new Tools
 - Integrate new Data Sources
- Secure your private instance for using highly sensitive data

The Problem

- You have written a Perl script to analyze genomic data and you want to share it with command-line averse colleagues

The Galaxy Solution

- Solution: Integrate the script as a new Tool into your own Galaxy server
- Demo How-to:
 - Obtain and install Galaxy source code (GetGalaxy.org)
 - Write an XML file describing the inputs and outputs and how to execute the script
 - Instruct Galaxy to load the tool

Using Galaxy

- Project homepage: GalaxyProject.org
- Use public Galaxy server: UseGalaxy.org
- Download Galaxy source: GetGalaxy.org
- Screencasts: GalaxyCast.org
- Public Mailing Lists
 - galaxy-bugs@bx.psu.edu
 - galaxy-user@bx.psu.edu
 - galaxy-dev@bx.psu.edu

Acknowledgments

- All Members of the Galaxy Team (see them at GalaxyProject.org)
- Thousands of our users
- GMOD Team
- UCSC Genome Informatics Team
- BioMart Team
- FlyMine/InterMine Teams
- Funding sources
 - NSF-ABI
 - NIH-NHGRI
 - Beckman Foundation
 - Huck Institutes at Penn State
 - Pennsylvania Department of Public Health
 - Emory University

The First Galaxy Developer Conference

A community gathering for people interested in tool integration, deployment, development, and extension of the Galaxy framework to meet their data analysis goals

May 15 - 17, 2010

Cold Spring Harbor Laboratory
<http://galaxyproject.org>