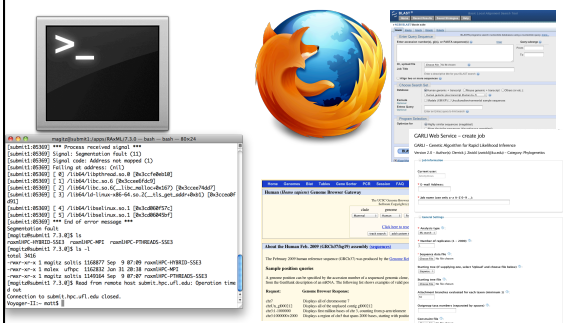
Introduction to Galaxy

Matt Gitzendanner magitz@ufl.edu

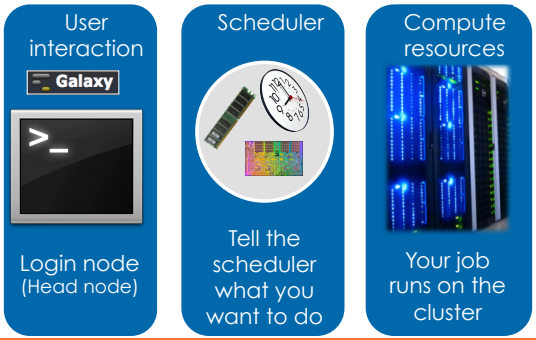
Oleksandr Moskalenko om@hpc.ufl.edu

Today's research computing

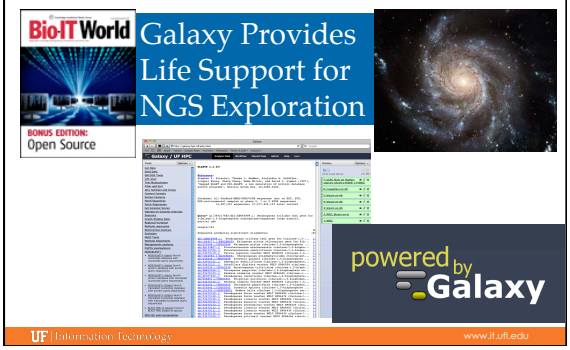
Approaches



Cluster basics



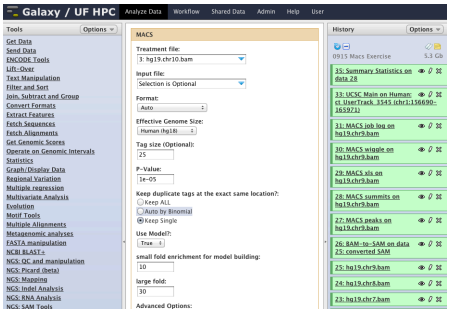
What is Galaxy?



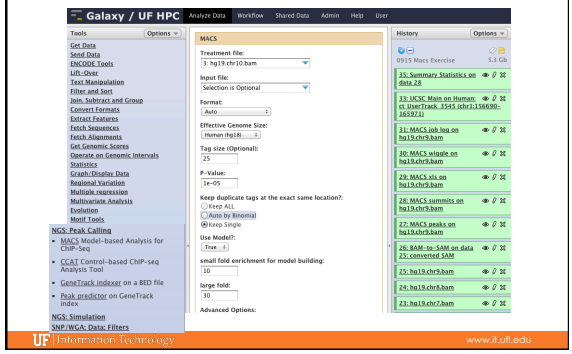
Galaxy: Data intensive biology *for everyone*

- Accessible, reproducible, transparent computational biology
- galaxy.hpc.ufl.edu
  - Local instance of Galaxy
    - Faster access to storage, easier upload
    - Local compute resources
    - Local control

Galaxy Analysis Workspace



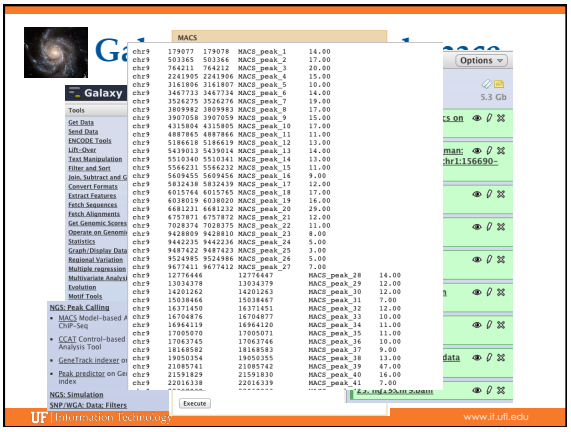
Galaxy Analysis Workspace



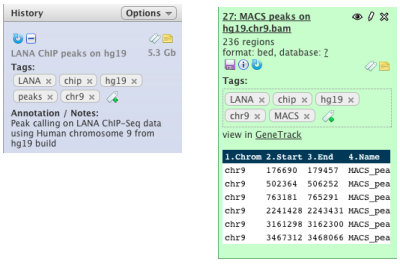
Galaxy Analysis Workspace

Galaxy Analysis Workspace

Galaxy Analysis Workspace



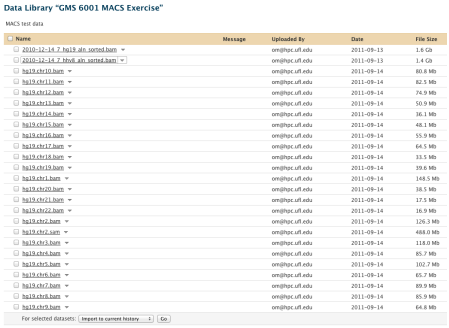
Metadata



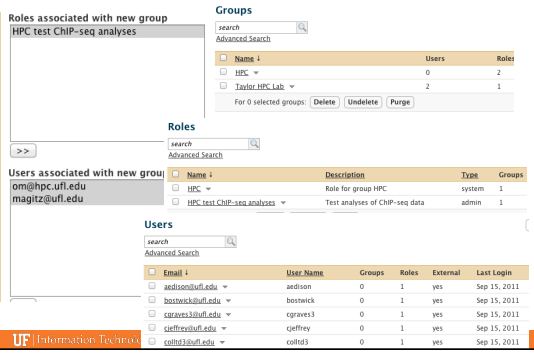
Getting Data into Galaxy

- Upload a file from your computer
  - Direct upload (<2GB)
  - For large files: scp or copy files to HPC
    - Load from within Galaxy
    - http://wiki.hpc.ufl.edu/index.php/Galaxy_Data_Import
- External data
  - UCSC table browser
  - Biomart
  - interMine / modMine
  - EuPathDB
  - EncodeDB
  - EpiGRAPH
  - FlyMine
  - GrameneMart…

Data libraries



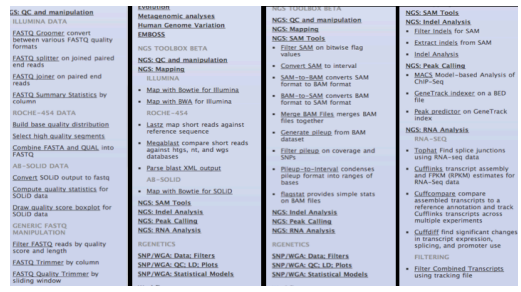
Data Access Control



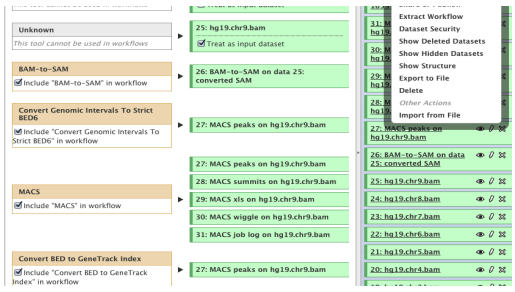
Galaxy Tool Suites

- Text Manipulation
- Format Converters
- Filtering and Sorting
- Join, Subtract, Group
- Sequence Tools
- Multi-species Alignment Tools
- Genomic Interval Operation
- Summary Statistics, graphing
- Regional Variation
- EMBOSS
- Evolution
- RNA-Seq
- ChIP-Seq
- GATK
- Phylogenetics

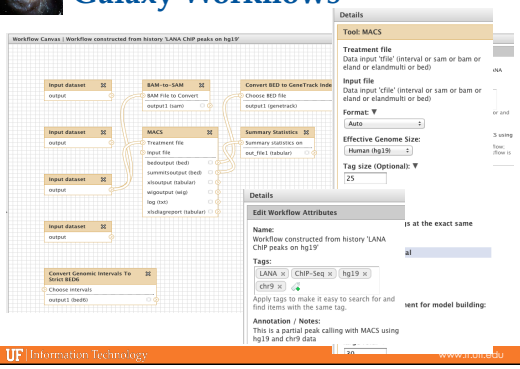
A galaxy of tools



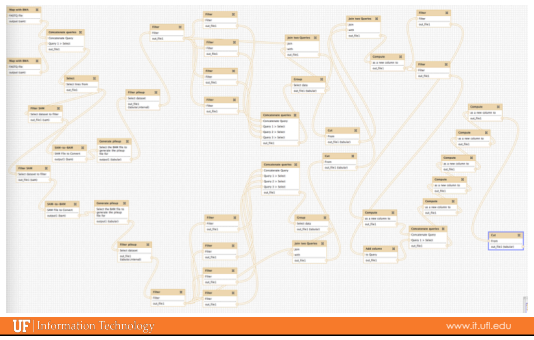
Galaxy Workflows



Galaxy Workflows



Galaxy Workflows



Sharing and publishing

Share or Publish History 'LANA ChIP peaks on hg19'

Making History Accessible via Link and Publishing It

This history is currently restricted so that only you and the users listed below can access it. You can:

Make History Accessible via Link

Generates a web link that you can share with other people so that they can view and import the history.

Make History Accessible and Publish

Makes the history accessible via link (see above) and publishes the history to Galaxy's Published Histories section, where it is publicly listed and searchable.

Sharing History with Specific Users

You have not shared this history with any users.

Share with a user

Back to Histories List

Sharing and publishing

Share or Publish History 'LANA ChIP peaks on hg19'

Making History Accessible via Link and Publishing It

This history is currently **accessible via link and published.**

Anyone can view and import this history by visiting the following URL:

http://galaxy.hpc.ufl.edu/u/moskalenko/h/lana-chip-peaks-on-hg19

This history is publicly listed and searchable in Galaxy's Published Histories section.

You can:

Unpublish History

Removes this history from Galaxy's Published Histories section so that it is not publicly listed or searchable.

Disable Access to History via Link and Unpublish

Disables this history's link so that it is not accessible and removes history from Galaxy's Published Histories section so that it is not publicly listed or searchable.

Sharing History with Specific Users

The following users will see this history in their history list and will be able to view, import, and run it.

Email

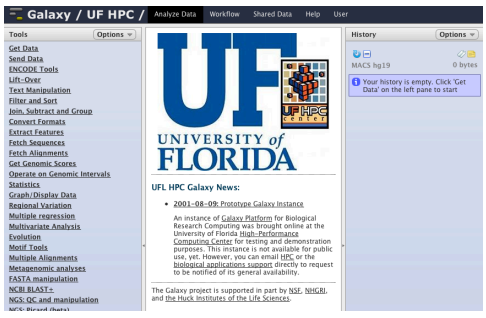
magitz@ufl.edu

Share with another user

Summary

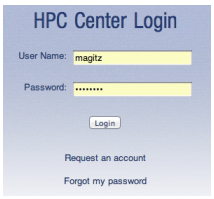
- Analyze data without the CLI
- Visualize the results
- Publish histories, workflows, and annotated pages
- Add new tools, get support @ HPC
- Focus on your science, not minutiae
- **UF Galaxy** – coming to a browser near you!

Demo



Galaxy demo

http://galaxy.hpc.ufl.edu



UF Research Computing

- Help and Support
  - Help Request Tickets
    - https://support.hpc.ufl.edu
    - For any kind of question or help requests
    - Searchable database of solutions
  - We are here to help!
    - support@hpc.ufl.edu

Training Schedule

- ✓ Aug 28: Intro to UFHPC, getting started
- ✓ Sept 10: Modules, RHEL6 Transition, User Q&A
- ✓ Sept 17: The Linux/Unix Shell - An Introduction
- ✓ Sept 24: Running Jobs, Submission Scripts, Modules
- ✓ Oct 1: Galaxy Overview, The Basics
- Oct 8: NGS Data Techniques: General Methods and Tools
- Oct 15: NGS Data Techniques: Reference Based Mapping and de Novo Assembly
- Oct 22: Phylogenetic Analyses
- ◆ Oct 29: Research Computing Day: Moving Big Data
- Nov 5: Multiprocessing at the HPC Center
- Nov 12: Using Git and CMake to Organize and Drive Data Analysis Pipelines
- Nov 19: Introduction to GPU Nodes
- Nov 29: NGS Data Techniques: RNA-Seq
- Dec 3: NGS Data Techniques: Alternative Splicing Analysis

UF Research Computing

- Help and Support (Continued)
  - http://wiki.hpc.ufl.edu
    - Documents on hardware and software resources
    - Various user guides
    - Many sample submission scripts
  - http://hpc.ufl.edu/support
    - Frequently Asked Questions
    - Account set up and maintenance