Galaxy Overview
The Basics

Matt Gitzendanner
magitz@ufl.edu

6/11/14

UF | Research Computing
Information Technology
Home of High-Performance Computing and HiPerGator

UF Research Computing

UF | Research Computing
Information Technology
Home of High-Performance Computing and HiPerGator

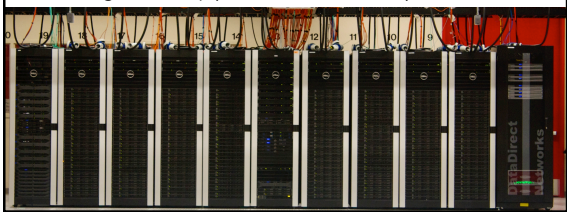
- Mission
  - Improve opportunities for research and scholarship
  - Improve competitiveness in securing external funding
  - Provide high-performance computing resources **and support** to UF researchers

UF RISING to National Preeminence

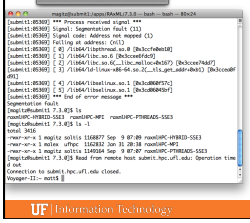
HiPerGator

The University of Florida Supercomputer for Research

- 16,384 cores–*total of about 21,000 cores today*
- Infiniband interconnect
- >3PB fast, high-availability, storage
- **GPGPUs**
- Large memory (**512GB to 1TB of RAM**) nodes



Approaches

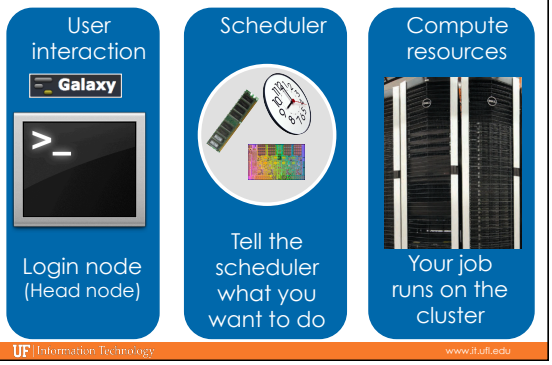


UNIVERSITY OF FLORIDA | High-Performance Computing

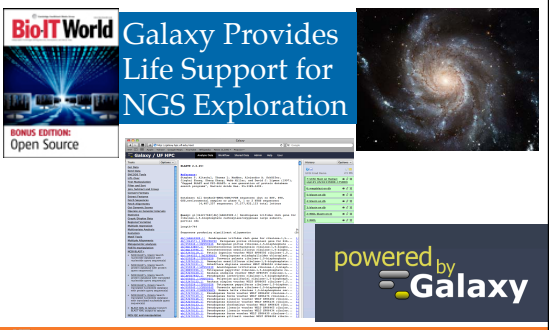
HiPerGator

The University of Florida Supercomputer for Research

Cluster basics



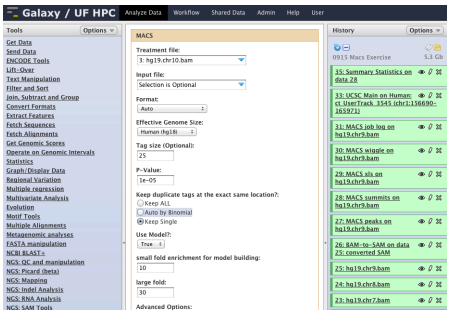
What is Galaxy?



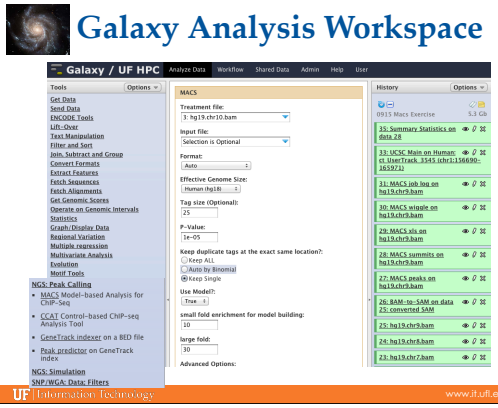
Galaxy: Data intensive biology *for everyone*

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

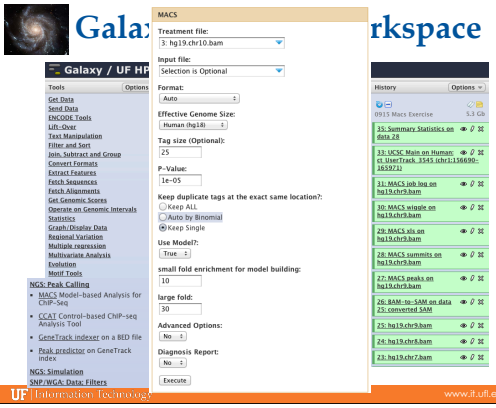
Galaxy Analysis Workspace

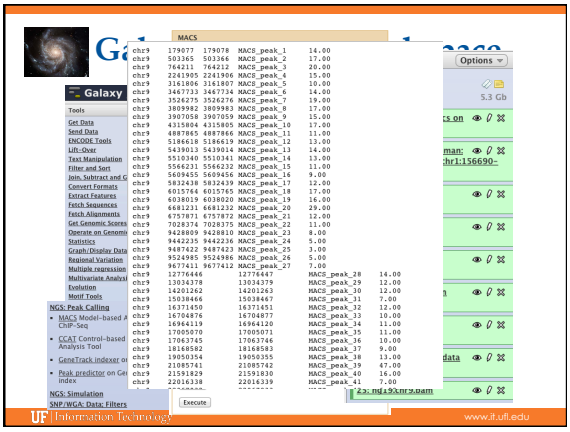


Galaxy Analysis Workspace

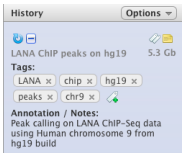


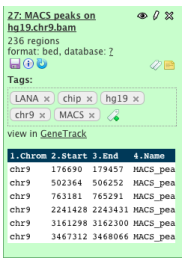
Gala rkspace





Metadata





UF Information Technology www.it.ufl.edu

Getting Data into Galaxy

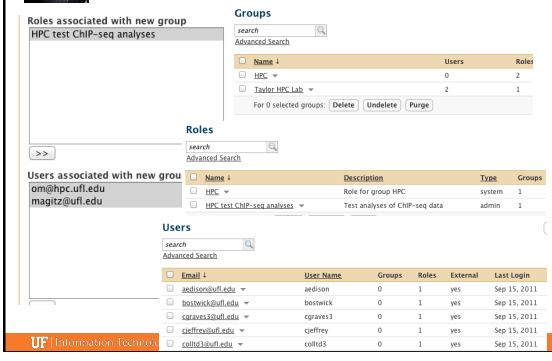
- Upload a file from your computer
  - Direct upload (<2GB)
  - For large files: scp or copy files to HPC
    - Load from within Galaxy
    - http://wiki.hpc.ufl.edu/index.php/Galaxy_Data_Import
- External data
  - UCSC table browser
  - Biomart
  - interMine / modMine
  - EuPathDB
  - EncodeDB
  - EpiGRAPH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

UF Information Technology www.it.ufl.edu

Data libraries

UF Information Technology www.it.ufl.edu

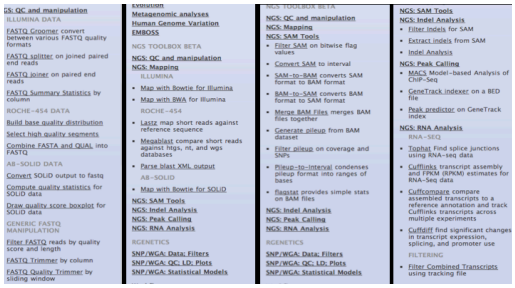
Data Access Control



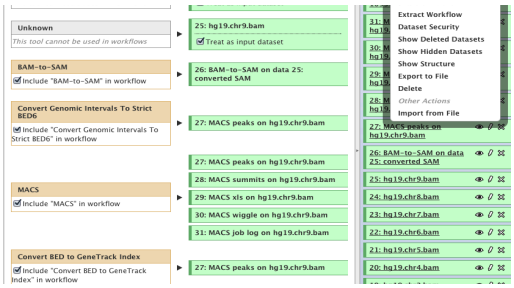
Galaxy Tool Suites

- 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

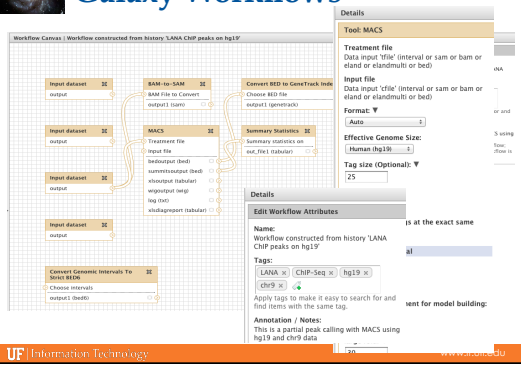
A galaxy of tools



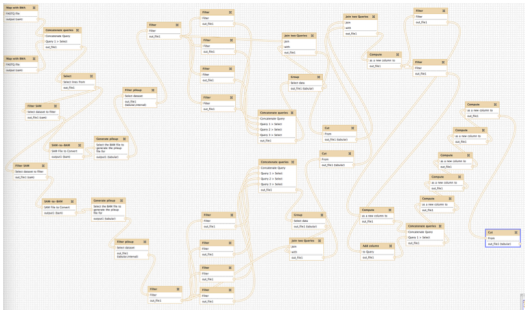
Galaxy Workflows



Galaxy Workflows

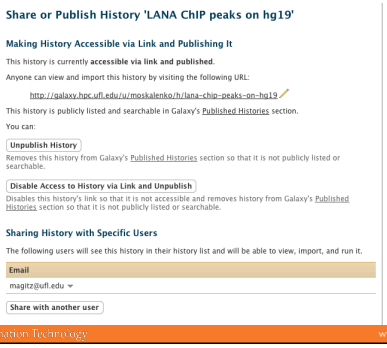


Galaxy Workflows



Sharing and publishing

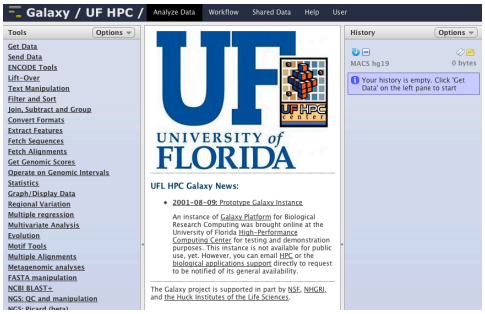
Sharing and publishing



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