

# UF Research Computing: Introduction and Getting Started

Erik Deumens  
Matt Gitzendanner

August 27, 2012

# UF Research Computing



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

# UF Research Computing

- ▶ Funding
  - Faculty participation (i.e. grant money) provides funds for hardware purchases
    - Matching grant program!
- ▶ Any UF Faculty can use
  - Up to 8 cores at a time
  - Investors gain priority and access to additional resources
- ▶ Comprehensive management
  - Hardware maintenance and 24x7 monitoring
  - Relieve researchers of the majority of systems administration tasks

# Matching Program

Consolidating Resources to Improve Efficiency and Capacity



The Research Computing Matching Program pooled \$642k, thereby creating synergies and improving research infrastructure.

# Services

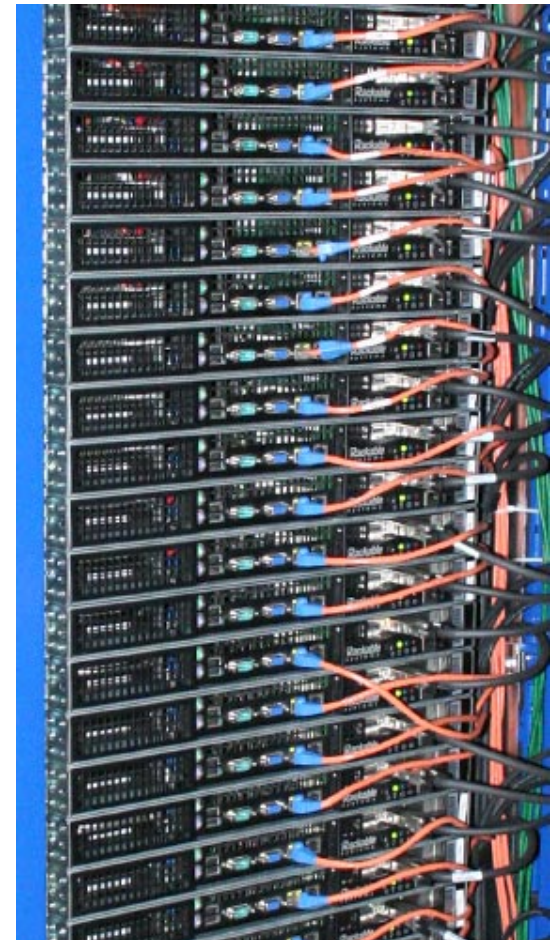
- ▶ NCU Normalized Compute Units
  - 1 core in node with RAM, network, access to 230 TB fast scratch storage
  - Acquisition is for 5 years
  - Cost: \$400 per NCU
- ▶ RSU Replicated long-term Storage Units
  - 1 TB for 1 year
  - Cost: \$250 per RSU
- ▶ LSU Long-term Storage Units
  - 1 TB for 1 year
  - Cost: \$125 per LSU

# UF Research Computing

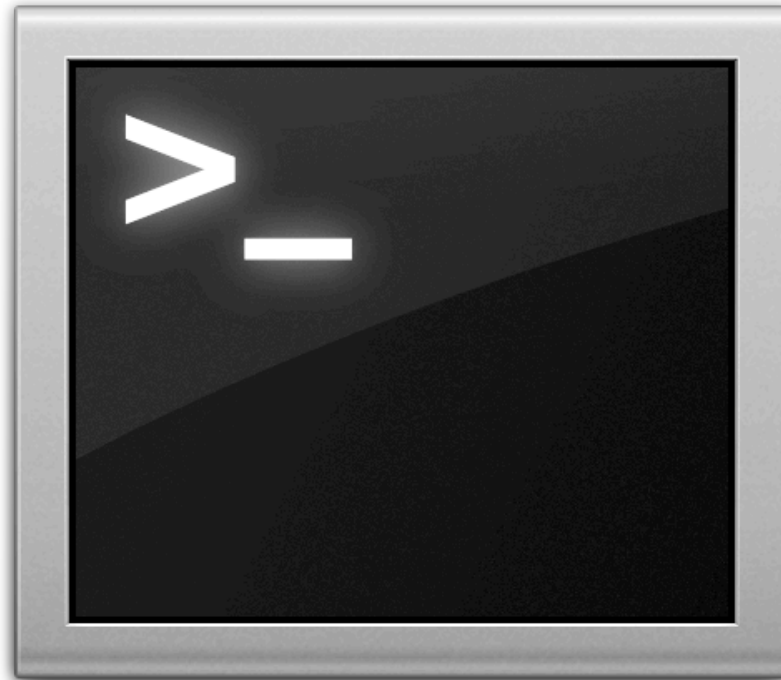


# UF Research Computing

- ▶ Shared Hardware Resources
  - Over **6K cores** AMD and Intel
  - High-speed, low-latency **InfiniBand** interconnects
  - **>1 PB**, high performance Lustre and Nexenta storage
  - **GPUs**—CUDA code, etc.
  - Several large memory (**512GB**) nodes



# UF Research Computing



Where do you start?

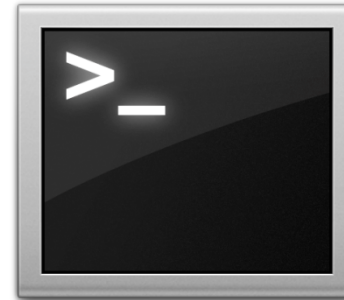


# Job priorities

- ▶ Investors get fast access
  - To the nr of cores they bought
  - Within minutes to at most an hour
    - Check that your job can actually run
    - Also check that others in the group did not use all slots
  - Can **burst** up to 10 times that nr if cores are idle
    - Not continuous use of 10 times the nr of cores
- ▶ Non-investors can use
  - Up to 8 cores simultaneously
  - Can burst to 10 times that
  - Lower priority than investors

# What can you run?

▶ Galaxy



▶ Linux

▶ Generally command line driven applications

▶ Graphical apps can be setup

◦ SAS



◦ BEAUti



# UF Research Computing

- ▶ User Accounts
  - Qualifications:
    - Current UF faculty, UF graduate student, and researchers
  - Request at: <http://www.hpc.ufl.edu/support/>
  - Requirements:
    - GatorLink Authentication
    - Faculty sponsorship for graduate students and researchers



The screenshot shows the 'HPC Center Account Request' form on the University of Florida website. The page header includes the University of Florida logo and the text 'The Foundation for The Gator Nation'. The main title is 'HPC Center Account Request'. Below the title is a navigation menu with links: Home, About, Users, Facilities, Support, Queues, Proposals, Publications, Contact, Frequently Asked Questions, Account Requests, Reset Password, Bug Reports, Wiki, and Test Nodes. The form fields are: First Name, Last Name, Organization, Email Address, Gatorlink Username, Sponsor, Sponsor Email, Type of Research, and Comments. A 'Submit Request' button is located at the bottom right of the form.

# Galaxy

The screenshot shows the Galaxy web interface. At the top, there is a navigation bar with the following items: **Galaxy / UF HPC**, **Analyze Data**, **Workflow**, **Shared Data**, **Visualization**, **Admin**, **Help**, and **User**. Below the navigation bar is a sidebar on the left containing a list of tools under the heading "Tools". The main content area features the University of Florida logo and the text "UNIVERSITY of FLORIDA". Below the logo, there are sections for "UFL HPC Galaxy Docs:", "UFL HPC Galaxy News:", and a list of news items. The right sidebar contains a "History" section with a list of recent jobs, including "7: UCSC Main on Human:", "6: megablast on db", "5: blastn on db", "4: blastn on db", "3: blastn on db", "2: RBCL blastn on nt", and "1: RBCL".

**Tools** Options ▾

- Get Data
- Send Data
- ENCODE Tools
- Lift-Over
- Text Manipulation
- Filter and Sort
- Join, Subtract and Group
- Convert Formats
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Regional Variation
- Multiple regression
- Multivariate Analysis
- Evolution
- Motif Tools
- Multiple Alignments
- Metagenomic analyses
- FASTA manipulation
- NCBI BLAST+
- NGS: QC and manipulation
- NGS: Picard (beta)
- NGS: Mapping
- NGS: Indel Analysis
- NGS: RNA Analysis
- NGS: SAM Tools
- NGS: GATK Tools
- NGS: Peak Calling
- NGS: Simulation
- SNP/WGA: Data; Filters
- SNP/WGA: QC; LD; Plots
- SNP/WGA: Statistical Models
- Human Genome Variation
- Genome Diversity
- VCF Tools
- PacBio/Illumina Assembly
- Workflows

**UFL HPC Galaxy Docs:**

[UFL HPC Galaxy docs in the HPC Wiki.](#)

**UFL HPC Galaxy News:**

- **2001-10-03: Beta testing**  
As all UF HPC users can log into UF Galaxy already we're looking for people to run actual data analyses and report any encountered problems, so we could fix them before Galaxy is opened up for unrestricted public use. Please use the "Help>Email comments, bug reports, or suggestions" Galaxy menu to email us reports and suggestions or file a report for Galaxy in the software section of the [HPC Support Website](#). Please share the history that shows the encountered problem with the Galaxy user (galaxy@hpc.ufl.edu) when you send a report.
- **2001-09-29: Research Computing Day demo**  
UF HPC Galaxy demo at the First Annual UF Research Computing Day. Galaxy beta announcement.
- **2001-09-15: MACS workshop for GMS 6001**  
GMS 6001 class had a hands-on workshop analyzing CHIP-Seq data using HPC Galaxy. Shared Data Library used for this class is available as "GMS 6001 MACS Exercise" in the Galaxy.
- **2001-08-09: Prototype Galaxy Instance**  
An instance of [Galaxy Platform](#) for Biological Research Computing was brought online at the University of Florida [High-Performance Computing Center](#) for testing and demonstration purposes. This instance is not available for public use, yet. However, you can email [HPC](#) or the [biological applications support](#) directly to request to be notified of its general availability.

The Galaxy project is supported in part by [NSF](#), [NHGRI](#), and the [Huck Institutes of the Life Sciences](#).

**History** Options ▾

UFGI Grad Demo 2.5 Mb

- 7: UCSC Main on Human: [eye](#) [edit](#) [delete](#)
- 6: megablast on db [eye](#) [edit](#) [delete](#)
- 5: blastn on db [eye](#) [edit](#) [delete](#)
- 4: blastn on db [eye](#) [edit](#) [delete](#)
- 3: blastn on db [eye](#) [edit](#) [delete](#)
- 2: RBCL blastn on nt [eye](#) [edit](#) [delete](#)
- 1: RBCL [eye](#) [edit](#) [delete](#)

# 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

Tools Options ▾

- [Get Data](#)
- [Send Data](#)
- [ENCODE Tools](#)
- [Lift-Over](#)
- [Text Manipulation](#)
- [Filter and Sort](#)
- [Join, Subtract and Group](#)
- [Convert Formats](#)
- [Extract Features](#)
- [Fetch Sequences](#)
- [Fetch Alignments](#)
- [Get Genomic Scores](#)
- [Operate on Genomic Intervals](#)
- [Statistics](#)
- [Graph/Display Data](#)
- [Regional Variation](#)
- [Multiple regression](#)
- [Multivariate Analysis](#)
- [Evolution](#)
- [Motif Tools](#)
- [Multiple Alignments](#)

- [Multivariate Analysis](#)
- [Evolution](#)
- [Motif Tools](#)
- [Multiple Alignments](#)
- [Metagenomic analyses](#)
- [FASTA manipulation](#)
- [NCBI BLAST+](#)
- [NGS: QC and manipulation](#)
- [NGS: Picard \(beta\)](#)
- [NGS: Mapping](#)
- [NGS: Indel Analysis](#)
- [NGS: RNA Analysis](#)
- [NGS: SAM Tools](#)
- [NGS: GATK Tools](#)
- [NGS: Peak Calling](#)
- [NGS: Simulation](#)
- [SNP/WGA: Data; Filters](#)
- [SNP/WGA: QC; LD; Plots](#)
- [SNP/WGA: Statistical Models](#)
- [Human Genome Variation](#)
- [Genome Diversity](#)
- [VCF Tools](#)
- [PacBio/Illumina Assembly](#)

- NGS: Mapping**
- [Lastz](#) map short reads against reference sequence
  - [Lastz paired reads](#) map short paired reads against reference sequence
  - [Map with Bowtie](#) for Illumina
  - [Map with Bowtie](#) for SOLiD
  - [Map with BWA](#) for Illumina
  - [Map with BWA](#) for SOLiD
  - [Map with BFAST](#)
  - [Megablast](#) compare short reads against htgs, nt, and wgs databases
  - [Parse blast XML output](#)
  - [Map with PerM](#) for SOLiD and Illumina
  - [Re-align with SRMA](#)
  - [Map with Mosaik](#)

# Galaxy

## Map with Bowtie for Illumina

Will you select a reference genome from your history or use a built-in index?:

Use a built-in index

Built-ins were indexed using default options

Select a reference genome:

if your genome of interest is not listed – contact Galaxy team

Is this library mate-paired?:

Single-end

FASTQ file:

Must have ASCII encoded quality scores

Bowtie settings to use:

Commonly used

For most mapping needs use Commonly used settings. If you want full control use Full parameter list

Suppress the header in the output SAM file:

Bowtie produces SAM with several lines of header information by default

Execute

# Galaxy

The screenshot displays the Galaxy web interface at <http://galaxy.hpc.ufl.edu/root>. The main content area shows a BLASTN search result for a query sequence: `gi|344217682|dbj|AB665989.1| Dendropanax trifidus rbcL gene for ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds`. The search parameters include a database of all GenBank+EMBL+DDBJ+PDB sequences (excluding EST, STS, GSS, environmental samples, and phase 0, 1, or 2 HTGS sequences) and a total of 14,487,257 sequences. The query length is 744. The results section lists sequences producing significant alignments, with the top hit being `dbj|AB665989.1| Dendropanax trifidus rbcL gene for ribulose-1,5-...`. The interface also features a navigation menu on the left with categories like 'Tools', 'Get Data', and 'NCBI BLAST+', and a history panel on the right showing previous jobs such as 'UCSC Main on Human' and 'megablast on db'.



# What can you run?

▶ Galaxy



▶ Linux

▶ Generally command line driven applications

▶ Graphical apps can be setup

◦ SAS



◦ BEAUti



# 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](mailto:support@hpc.ufl.edu)



# 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

