UF Research Computing
An Introduction

Matt Gitzendanner
Assoc. Sci., Biology/FLMNH
UF HPC User Support
magitz@ufl.edu

Oleksandr "Alex" Moskalenko
UF HPC Bio User Support
om@hpc.ufl.edu

UF Research Computing

UNIVERSITY of FLORIDA

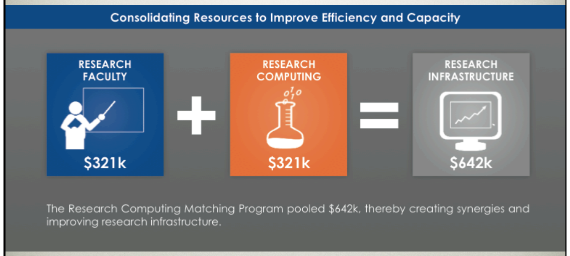
✦ Mission
- Improve opportunities for research and scholarship
- Improve competitiveness in securing external funding
- Provide high-performance computing resources **and support** to UF researchers

UF Research Computing

✦ Funding
- Faculty participation (i.e. grant money) provides funds for hardware purchases
  - Matching grant program!

✦ Comprehensive management
- Hardware maintenance and 24x7 monitoring
- Relieve researchers of the majority of systems administration tasks

Matching Program



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

UF Research Computing

✦ Shared Hardware Resources

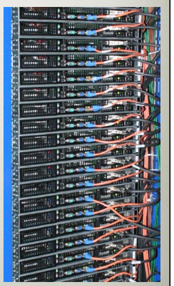
Bio Cluster

- Over 6K cores AMD and Intel
- InfiniBand interconnects
- >1 PB, high performance Lustre and Nexenta storage
- NVidia Tesla (C1060) GPUs
- Several large memory (512GB) nodes

UF Research Computing

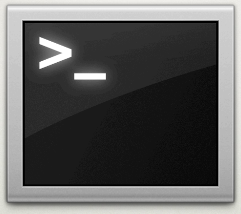
Machine room at Larson Hall

UF Research Computing



- ✦ Large resources available
- ✦ Staff to help you succeed

UF Research Computing



- ✦ Where do you start?

What can you run?

- ✦ Galaxy
- ✦ Linux
- ✦ Generally command line driven applications
- ✦ Though can setup graphical apps
  - SAS
  - BEAUti

UF Research Computing

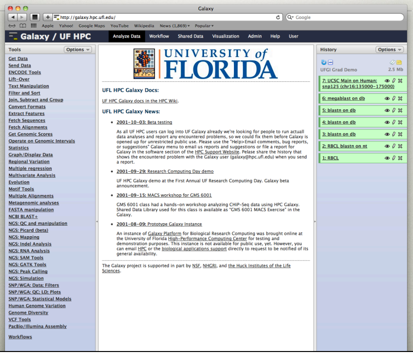
- ✦ User Accounts
  - Qualifications:
    - Current UF faculty, UF graduate student, and researchers
  - Request at: http://www.hpc.ufl.edu/support/
  - Requirements:
    - GatorLink Authentication
    - Faculty sponsorship for graduate students and researchers

UF Research Computing

- ✦ Account Policies
  - Personal activities are strictly prohibited on HPC Center systems
  - Class accounts deleted at end of semester
  - **Data are not backed up!**
  - **Home** directories must not be used for I/O
    - Use /scratch/hpc/
  - Storage systems may not be used to archive data from other systems
  - Passwords expire every 6 months

Making the leap

Galaxy



Galaxy

http://galaxy.hpc.ufl.edu/

UFL HPC Galaxy Docs:

UF 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 actuall 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. Pelase 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

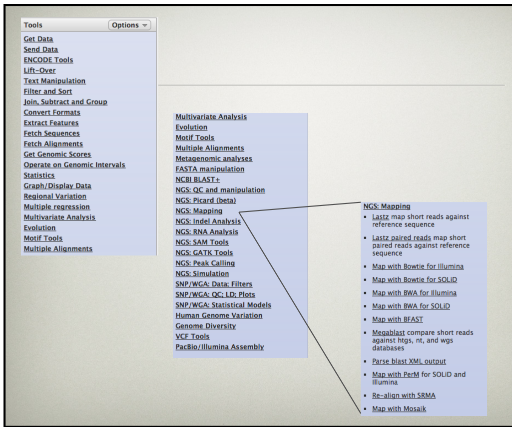
  GMS 6001 class had a hands-on workshop analyzing ChIP-Seq data using HPC Galaxy. Shared Data Library used for this class is available as "GMS 6001 MACS Exercise" in the Galaxy.
- 2001-08-09: Prototype Galaxy Instance

  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.

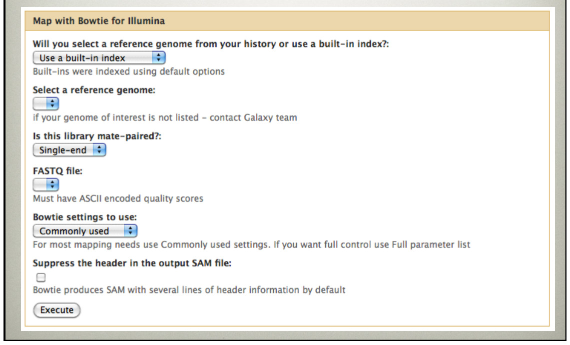
Galaxy:
Data intensive biology *for everyone*

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



- 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

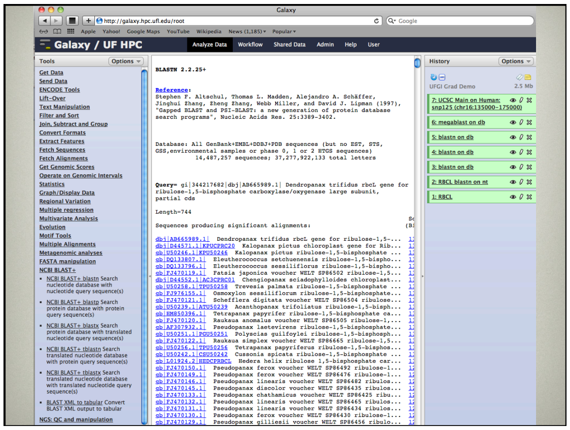
Select a reference genome:
If your genome of interest is not listed - contact Galaxy team

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



Galaxy

http://galaxy.hpc.ufl.edu/root

- NCBI BLAST+ blastn Search nucleotide database with nucleotide query sequence(s)
- NCBI BLAST+ blastp Search protein database with protein query sequence(s)
- NCBI BLAST+ blastx Search protein database with translated nucleotide query sequence(s)
- NCBI BLAST+ tblastn Search translated nucleotide database with protein query sequence(s)
- NCBI BLAST+ tblastx Search translated nucleotide database with translated nucleotide query sequence(s)
- BLAST XML to tabular Convert BLAST XML output to tabular

```
BLASTN 2.2.25+


Reference:
Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),
"Gapped BLAST and PSI-BLAST: a new generation of protein database
search programs", Nucleic Acids Res. 25:3389-3402.



Database: All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS,
GSS,environmental samples or phase 0, 1 or 2 HTGS sequences)
           14,487,257 sequences; 37,277,922,133 total letters



Query= gi|344217682|dbj|AB665989.1| Dendropanax trifidus rbcL gene for
ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit,
partial cds

Length=744
                                                                      Sc
Sequences producing significant alignments:                          (Bi

dbj|AB665989.1|  Dendropanax trifidus rbcL gene for ribulose-1,5-...
dbj|D44571.1|KPUCPRC20  Kalopanax pictus chloroplast gene for Rib...
gb|U50246.1|KPU50246  Kalopanax pictus ribulose-1,5-bisphosphate ...
gb|DQ133807.1|  Eleutherococcus setchuenensis ribulose-1,5-bispho...
gb|DQ133796.1|  Eleutherococcus sessiliflorus ribulose-1,5-bispho...
gb|FJ470119.1|  Fatsia japonica voucher WELT SP86502 ribulose-1,5...
dbj|D44552.1|AC3CPRC01  Chengiopanax sciadophylloides chloroplast...
gb|U50258.1|TPU50258  Trevesia palmata ribulose-1,5-bisphosphate ...
gb|FJ976155.1|  Osmoxylon sessiliflorum ribulose-1,5-bisphosphate...
gb|FJ470121.1|  Schefflera digitata voucher WELT SP86504 ribulose...
gb|U50239.1|ATU50239  Acanthopanax trifoliatus ribulose-1,5-bisph...
gb|HM850396.1|  Tetrapanax papyrifer ribulose-1,5-bisphosphate ca...
gb|FJ470120.1|  Raukaua anomalus voucher WELT SP86505 ribulose-1,...
gb|AF307932.1|  Pseudopanax laetevirens ribulose-1,5-bisphosphate...
gb|U50251.1|PGU50251  Polyscias guilfoylei ribulose-1,5-bisphosph...
gb|FJ470122.1|  Raukaua simplex voucher WELT SP86665 ribulose-1,5...
gb|U50256.1|TPU50256  Tetrapanax papyriferus ribulose-1,5-bisphos...
gb|U50242.1|CSU50242  Cussonia spicata ribulose-1,5-bisphosphate ...
gb|L01924.2|HEDCPRBCL  Hedera helix ribulose 1,5-bisphosphate car...
gb|FJ470150.1|  Pseudopanax ferox voucher WELT SP86492 ribulose-1...
gb|FJ470149.1|  Pseudopanax ferox voucher WELT SP86476 ribulose-1...
gb|FJ470146.1|  Pseudopanax linearis voucher WELT SP86482 ribulos...
gb|FJ470145.1|  Pseudopanax discolor voucher WELT SP86435 ribulos...
gb|FJ470133.1|  Pseudopanax chathamicus voucher WELT SP86425 ribu...
gb|FJ470132.1|  Pseudopanax linearis voucher WELT SP86465 ribulos...
gb|FJ470131.1|  Pseudopanax linearis voucher WELT SP86434 ribulos...
gb|FJ470130.1|  Pseudopanax ferox voucher WELT SP86430 ribulose-1...
gb|FJ470129.1|  Pseudopanax gilliesii voucher WELT SP86456 ribulo...
```

Making the leap

Compute resources

Cluster basics

User interaction

Galaxy

Login node (Head node)

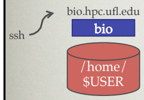
Scheduler

Tell the scheduler what you want to do

Compute resources

Your job runs on the cluster

Cluster login



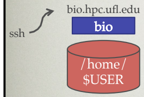
ssh <user>@bio.hpc.ufl.edu

Windows: PuTTY
Mac/Linux: Terminal

Head node(s)

Login to head node

Cluster login

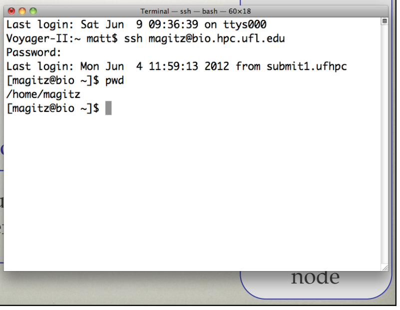


ssh <user>@bi

Windows: Pu
Mac/Linux: Te

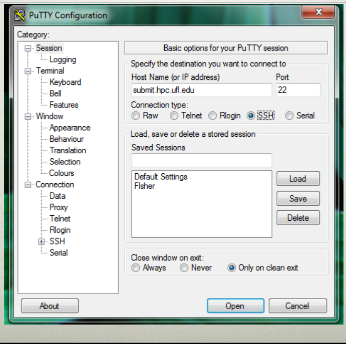
Head node(s)

node



```
Last login: Sat Jun  9 09:36:39 on ttys000
Voyager-II:~ matt$ ssh magitz@bio.hpc.ufl.edu
Password:
Last login: Mon Jun  4 11:59:13 2012 from submit1.ufhpc
[magitz@bio ~]$ pwd
/home/magitz
[magitz@bio ~]$ 
```

Logging in



Specify the destination you want to connect to

Host Name (or IP address): submit.hpc.ufl.edu
Port: 22

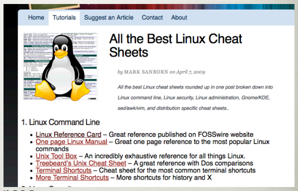
Connection type: ○ Raw ○ Telnet ○ Rlogin ● SSH ○ Serial

Load, save or delete a stored session

Saved Sessions: Default Settings, Flsher

Close window on exit: ○ Always ○ Never ● Only on clean exit

Linux Command Line



All the Best Linux Cheat Sheets

by MARK SANBORN *on April 7, 2009*

All the best Linux cheat sheets rounded up in one post broken down into Linux command line, Linux security, Linux administration, Gnome/KDE, sed/awk/vim, and distribution specific cheat sheets..

1. Linux Command Line

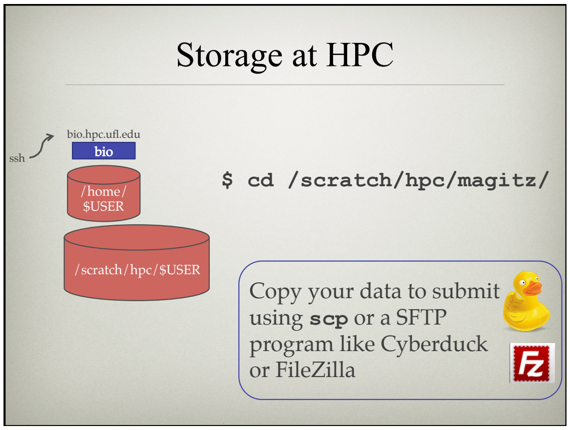
- Linux Reference Card – Great reference published on FOSSwire website
- One page Linux Manual – Great one page reference to the most popular Linux commands
- Unix Tool Box – An incredibly exhaustive reference for all things Linux.
- Treebeard's Unix Cheat Sheet – A great reference with Dos comparisons
- Terminal Shortcuts – Cheat sheet for the most common terminal shortcuts
- More Terminal Shortcuts – More shortcuts for history and X

- ✦ Lots of online resources
  - Google: linux cheat sheet
- ✦ Training sessions: TBA
- ✦ User manuals for applications

UF Research Computing

- ✦ Storage
  - Home Area: /home/$USER
    - For code compilation and user file management only, do not write job output here
  - On UF-HPC: Lustre File System
    - /scratch/hpc/$USER, 460 TB

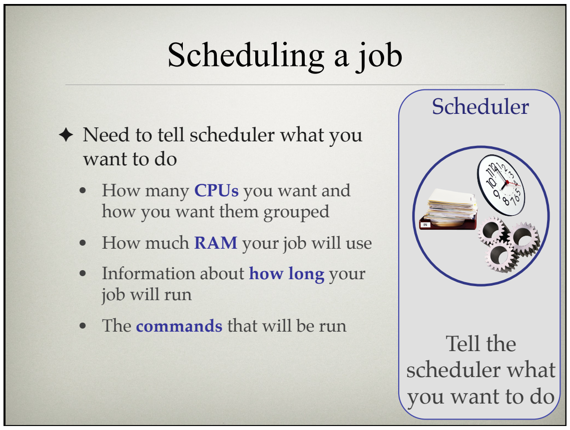
Must be used for all file I/O



Storage at HPC

```
$ cd /scratch/hpc/magitz/
```

Copy your data to submit using **scp** or a SFTP program like Cyberduck or FileZilla



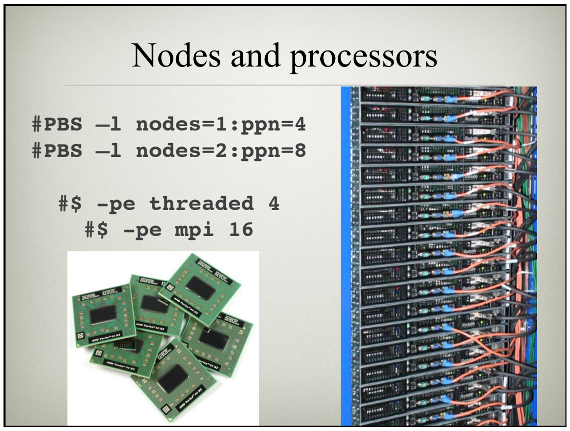
Scheduling a job

✦ Need to tell scheduler what you want to do

- How many **CPUs** you want and how you want them grouped
- How much **RAM** your job will use
- Information about **how long** your job will run
- The **commands** that will be run

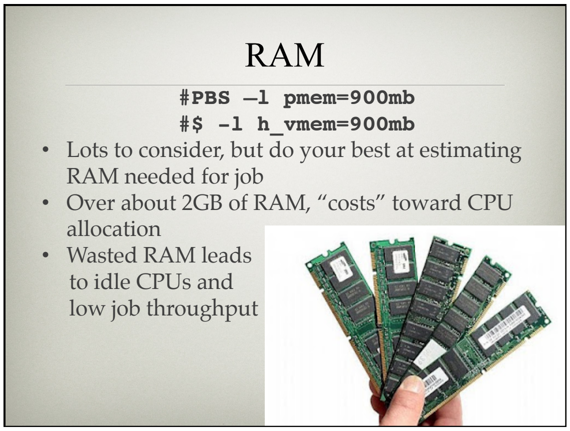
Scheduler

Tell the scheduler what you want to do



Nodes and processors

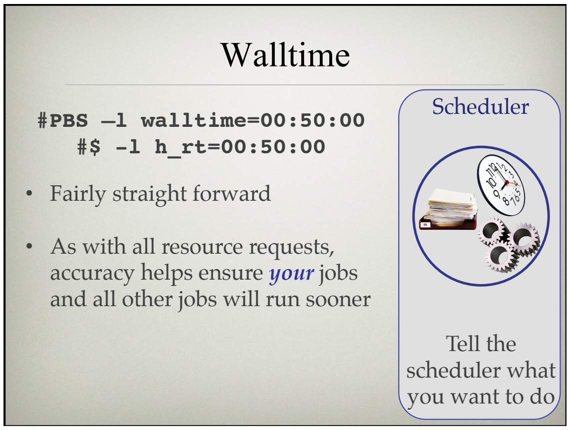
```
#PBS –l nodes=1:ppn=4
#PBS –l nodes=2:ppn=8

  #$ -pe threaded 4
    #$ -pe mpi 16
```



RAM

```
#PBS –l pmem=900mb
#$ -l h_vmem=900mb
```

- Lots to consider, but do your best at estimating RAM needed for job
- Over about 2GB of RAM, “costs” toward CPU allocation
- Wasted RAM leads to idle CPUs and low job throughput



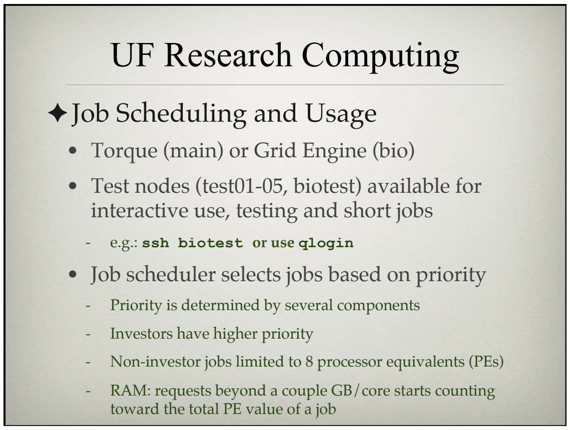
Walltime

```
#PBS –l walltime=00:50:00
   #$ -l h_rt=00:50:00
```

- Fairly straight forward
- As with all resource requests, accuracy helps ensure ***your*** jobs and all other jobs will run sooner

Scheduler

Tell the scheduler what you want to do



UF Research Computing

✦ Job Scheduling and Usage

- Torque (main) or Grid Engine (bio)
- Test nodes (test01-05, biotest) available for interactive use, testing and short jobs
  - e.g.: `ssh biotest` or use `qlogin`
- Job scheduler selects jobs based on priority
  - Priority is determined by several components
  - Investors have higher priority
  - Non-investor jobs limited to 8 processor equivalents (PEs)
  - RAM: requests beyond a couple GB/core starts counting toward the total PE value of a job

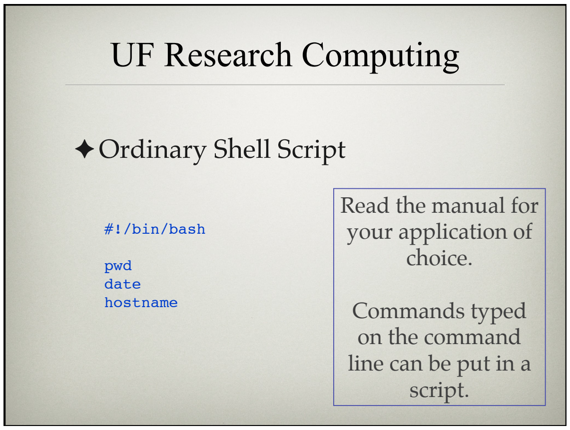
UF Research Computing

✦Ordinary Shell Script

```
#!/bin/bash

pwd
date
hostname
```

Read the manual for your application of choice.

Commands typed on the command line can be put in a script.



UF Research Computing

✦Submission Script

```
#!/bin/bash
#
#PBS -N My_Job_Name
#PBS -M Joe_Shmoe@ufl.edu
#PBS -m abe
#PBS -o My_Job_Name.log
#PBS -j oe
#PBS -l nodes=1:ppn=1
#PBS -l walltime=00:05:00
#PBS –l pmem=900mb

cd $PBS_O_WORKDIR
pwd
date
hostname
```

Tell the scheduler what you want to do

UF Research Computing

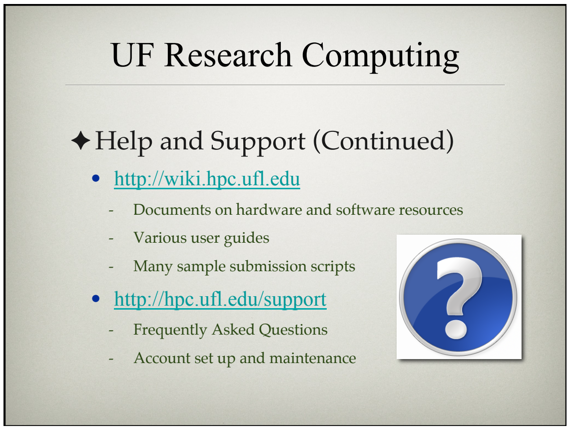
✦Job Management

- qsub <file_name>: job submission
- qstat –u <user>: check queue status
- qdel <JOB_ID>: job deletion

UF Research Computing

✦Help and Support

- Help Request Tickets
  - https://support.hpc.ufl.edu
  - Not just for "bugs" but for any kind of question or help requests
  - Searchable database of solutions
- We are here to help!
  - 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