

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



◆ 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

Consolidating Resources to Improve Efficiency and Capacity



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

UF Research Computing

◆ Shared Hardware Resources

- **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

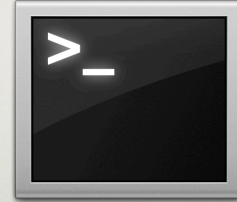


UF Research Computing

- ◆ 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



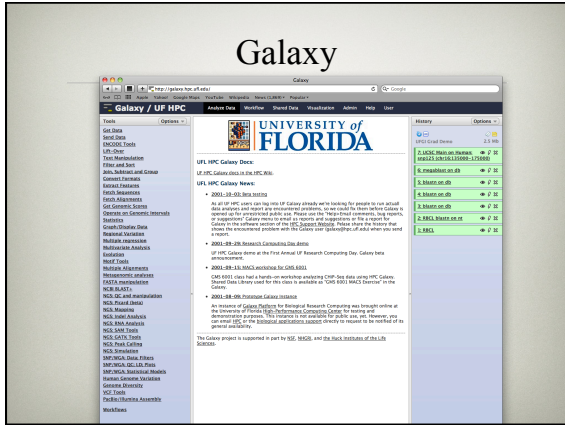
UF Research Computing

◆ 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

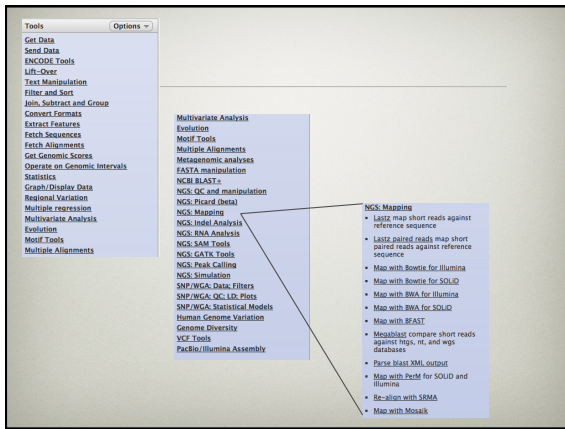




Galaxy:

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

Map with Bowtie for Illumina

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

Built-ins were indexed using default options

Select a reference genome:

FASTQ file:

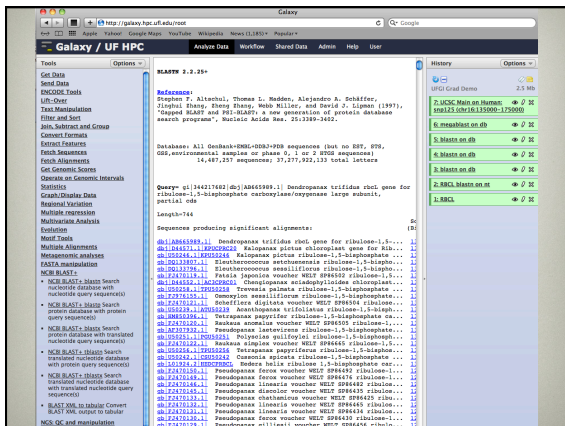
Must have ASCII encoded quality scores

Bowtie settings to use:

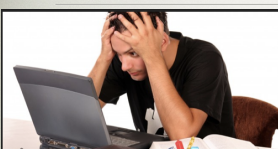
For most mapping needs use Commonly used settings. If you want full control use Full parameter list


Suppress the header in the output SAM file:

Bowtie produces SAM with several lines of header information by default

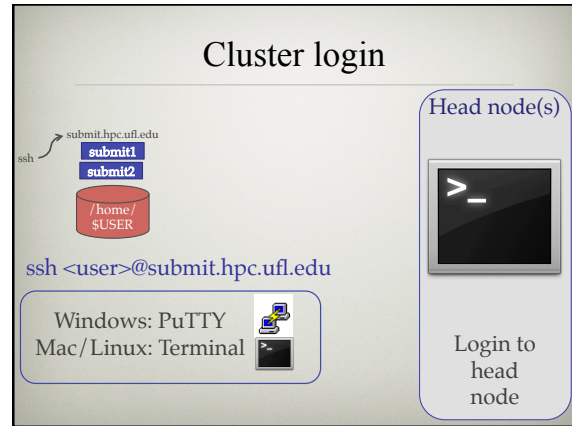
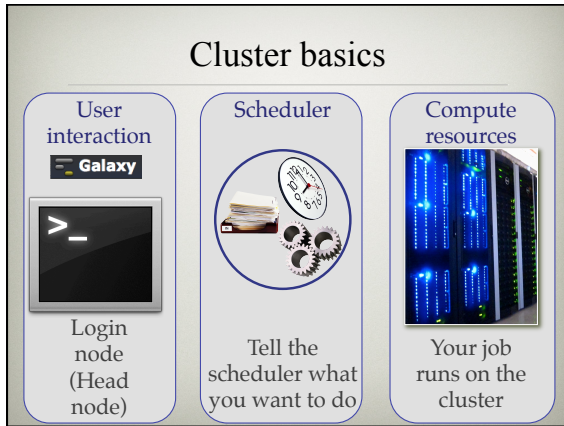


Making the leap





Compute resources



Cluster login

```

submit.hpc.ufl.edu
ssh submit.hpc.ufl.edu
Last login: Mon Jun 11 21:49:41 on ttys000
Voyager-II:~ matt$ ssh magitz@submit.hpc.ufl.edu
magitz@submit.hpc.ufl.edu's password:
Last login: Tue Jun 12 16:01:13 2012 from submit.hpc.ufl.edu
Welcome to the UF HPC Center.

Do not run interactive jobs on the login nodes. If you need to
run an interactive job, there are interactive/test nodes for that.
UF HPC Center Account Policies can be found here:
http://www.hpc.ufl.edu/users/accounts.php

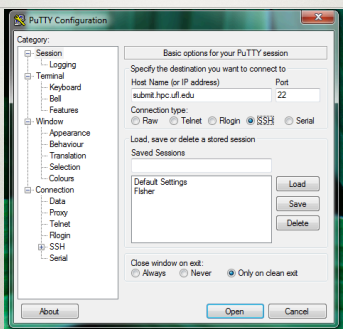
[magitz@submit1 ~]$ pwd
/home/magitz
[magitz@submit1 ~]$
                    
```

Head node(s)

Windows: PuTTY


Mac/Linux: Terminal


Logging in



The screenshot shows the PuTTY Configuration dialog box. The 'Session' category is selected, showing fields for Host Name (submit.hpc.ufl.edu) and Port (22). The 'Connection type' is set to SSH. The 'Close window on exit' options are set to 'Always'.

Linux Command Line





- ◆ 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


ssh → submit.hpc.ufl.edu



submit1
submit2

`$ cd /scratch/hpc/magitz/`

/home/\$USER

/scratch/hpc/\$USER

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
```



RAM

```
#PBS -l pmem=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
```

- 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

- Test nodes (test01-05) available for interactive use, testing and short jobs
 - e.g.: `ssh test01`
- 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_Shmoefl.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
```

Scheduler



Tell the scheduler what you want to do



UF Research Computing

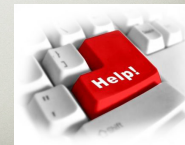
◆ Job Management

- `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

