

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

UF Research Computing

◆ Shared Hardware Resources

- **Over 5K cores** AMD and Intel
- InfiniBand interconnects
- **>1 PB**, high performance Lustre and Nexenta storage



- NVidia Tesla (C1060) GPUs (8)
- Several large memory (**512GB**) nodes

◆ Fisher Cluster (for UFGI affiliates)

- 136 cores, 332 GB RAM, 48TB storage

UF Research Computing

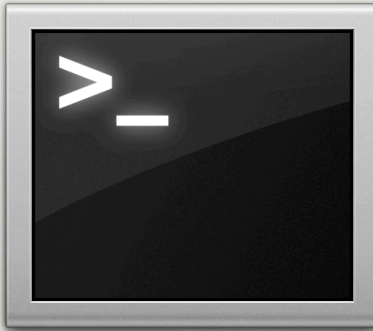
Machine room at Larson Hall



UF Research Computing

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

UF Research Computing



◆ Where do you start?

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. The form includes the following fields: First Name, Last Name, Organization, Email Address, GatorLink Username, Sponsor, Sponsor Email, Type of Research, and Comments. There is a 'Submit Request' button at the bottom of the form. The page header includes 'HPC Center of the University of Florida' and 'UF FLORIDA'.

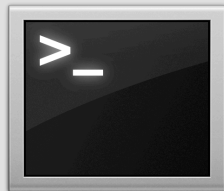
UF Research Computing

◆ 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

Cluster basics

Head node(s)



Login to
head
node

Scheduler



Tell the
scheduler what
you want to do

Compute resources



Your job
runs on the
cluster

Cluster login

ssh → submit.hpc.ufl.edu

submit1

submit2

/home/
\$USER

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

Windows: PuTTY

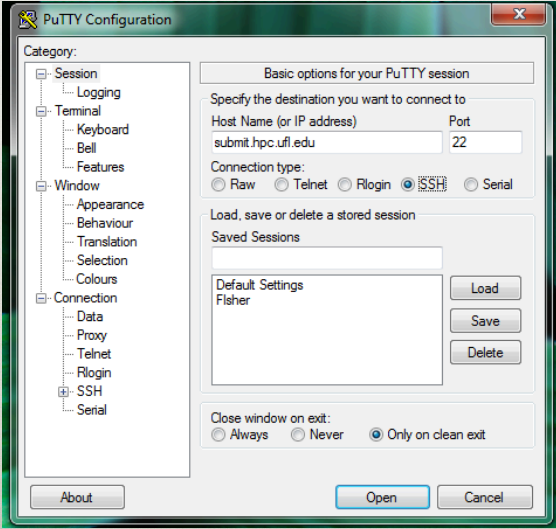
Mac/Linux: Terminal

Head node(s)

>_

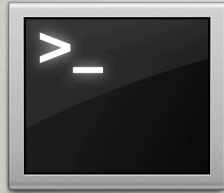
Login to head node

Logging in



The screenshot shows the PuTTY Configuration dialog box. The 'Category' list on the left has 'SSH' selected. The 'Basic options for your PuTTY session' section is visible, showing the 'Host Name (or IP address)' field set to 'submit.hpc.ufl.edu' and the 'Port' field set to '22'. The 'Connection type' section has 'SSH' selected with a radio button. Below this, there are 'Load', 'Save', and 'Delete' buttons for saved sessions. At the bottom, the 'Close window on exit' section has 'Only on clean exit' selected.

Linux Command Line



- ◆ Lots of online resources
 - Google: linux cheat sheet
- ◆ Training sessions: Feb 9th
- ◆ 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, 230 TB Must be used for all file I/O

Storage at HPC

ssh → submit.hpc.ufl.edu



submit1
submit2

/home/
\$USER

/scratch/hpc/\$USER

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

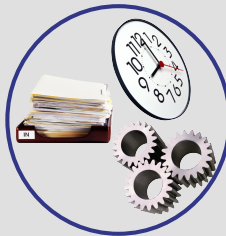
Copy your data to submit using **scp** or a SFTP program like Cyberduck or FileZilla

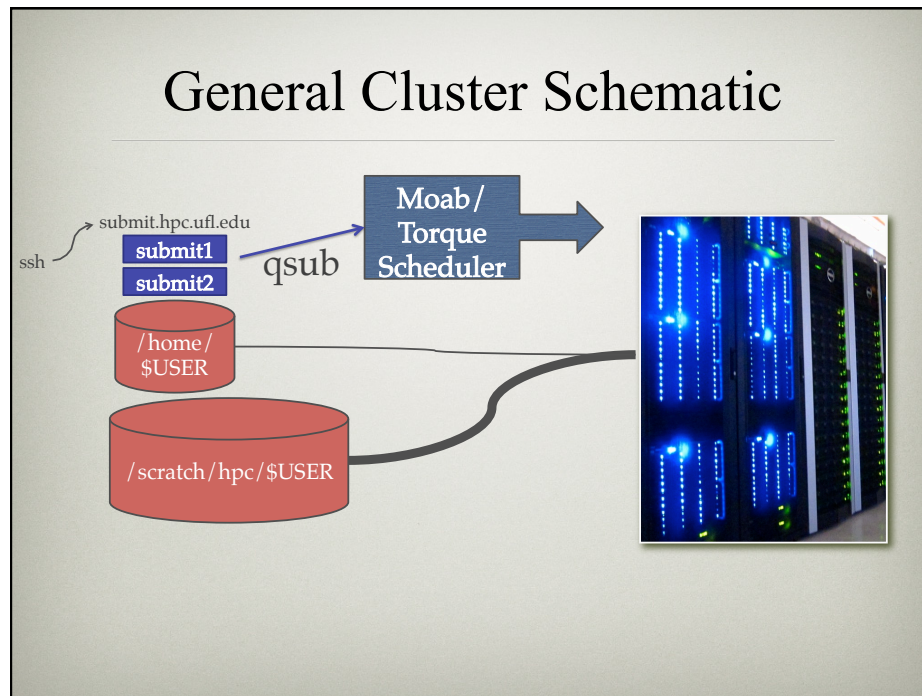
Scheduling a job

- ◆ Need to tell scheduler what you want to do
 - Information about **how long** your job will run
 - How many **CPUs** you want and how you want them grouped
 - How much **RAM** your job will use
 - The **commands** that will be run

Scheduler



Tell the scheduler what you want to do



UF Research Computing

◆ Job Scheduling and Usage

- PBS/Torque batch system
- Test nodes (test01, 04, 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

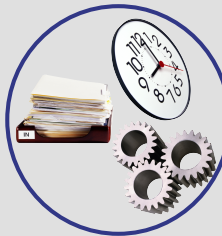
◆ Submission Script

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

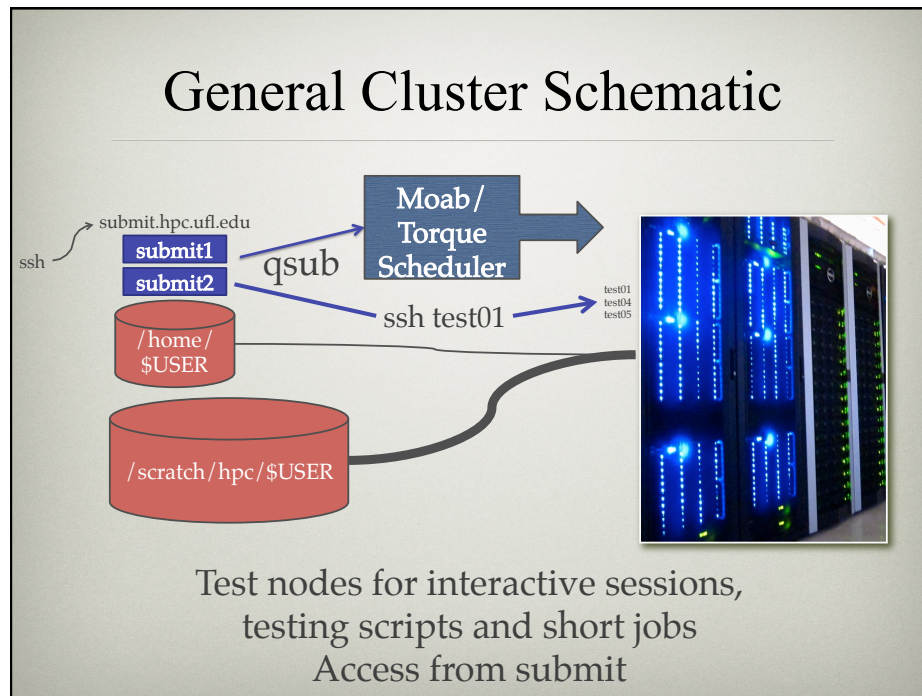
cd $PBS_O_WORKDIR
pwd
date
hostname
```



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
 - `qdelmine` : to delete all of your jobs

UF Research Computing

◆ Current Job Queues @ UFHPC

- **submit**: job queue for general use
 - **investor**: dedicated queue for investors
 - **other**: job queue for non-investors
- **testq**: test queue for small and short jobs
- **tesla**: queue for jobs utilizing GPUs: #PBS -q tesla
- **bigmem**: high memory nodes (256, 512 GB of RAM)

Galaxy

The screenshot shows the Galaxy web interface. The main content area features the University of Florida logo and the following text:

UFL HPC Galaxy Docs:
 UFL HPC Galaxy docs in the HPC Wiki.

UFL HPC Galaxy News:

- **2001-10-03: Beta testing**
 As all UFL HPC users can log into UFL 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**
 UFL 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 itpc@hpc.ufl.edu or the biological applications support directly to request to be notified of its general availability.

The Galaxy project is supported in part by NSF, NH&R, and the Huck Institutes of the Life Sciences.

The History sidebar on the right shows the following jobs:

Job ID	Name	Size
7	UCSC Main on Human snp125 (chr16:135000-175000)	2.5 Mb
6	megalblast on db	
5	blastn on db	
4	blastn on db	
3	blastn on db	
2	RBCL blastn on mt	
1	RBCL	

Galaxy

◆ galaxy.hpc.ufl.edu

- Local instance of Galaxy
 - Faster access to storage, easier upload
 - Local compute resources
 - Local control

◆ Training session: Feb 2nd

The screenshot shows the Galaxy web interface with a 'Tools' sidebar on the left and a main content area. A callout box points to the 'NGS: Mapping' category in the main content area.

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?:

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?:

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

The screenshot shows the Galaxy web interface for the BLASTN 2.2.25+ tool. The main content area displays the following information:

BLASTN 2.2.25+

Reference:
 Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zhong 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

Sequences producing significant alignments:

Accession	Species	Gene	Length	Score
dbj AB665989.1	Dendropanax trifidus	rbcL gene for ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds	744	86.0
dbi D44521.1 KP028728	Kalopanax pictus	chloroplast gene for RbcL	744	85.0
gb U50246.1 KP050246	Kalopanax pictus	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds	744	85.0
gb DQ133807.1	Eleutherococcus setchuensis	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds	744	85.0
gb DQ133796.1	Eleutherococcus sessiliflorus	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds	744	85.0
gb F470119.1	Fatsia japonica	voucher WELT SP86592	744	85.0
dbi D44552.1 AC3CP801	Chengiopanax soiadophylloides	chloroplast gene for RbcL	744	85.0
gb U50258.1 TP050258	Trevesia palmata	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds	744	85.0
gb F376155.1	Osmoxylon sessiliflorum	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds	744	85.0
gb F470121.1	Schefflera digitata	voucher WELT SP86504	744	85.0
gb U50239.1 AY050239	Acanthopanax trifoliatum	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds	744	85.0
gb U50236.1	Tetrapanax papyrifer	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds	744	85.0
gb F470120.1	Raukias anomalus	voucher WELT SP86505	744	85.0
gb AF307932.1	Pseudopanax laetevirens	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds	744	85.0
gb U50251.1 PG050251	Polystichum guilfoylei	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds	744	85.0
gb F470122.1	Raukias simplex	voucher WELT SP86465	744	85.0
gb U50256.1 TP050256	Tetrapanax papyrifer	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds	744	85.0
gb U50242.1 LC050242	Cussonia spicata	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds	744	85.0
gb F470124.1 HRC050242	Hedera helix	ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, partial cds	744	85.0
gb F470150.1	Pseudopanax ferox	voucher WELT SP86492	744	85.0
gb F470149.1	Pseudopanax ferox	voucher WELT SP86476	744	85.0
gb F470146.1	Pseudopanax linearis	voucher WELT SP86482	744	85.0
gb F470145.1	Pseudopanax discolor	voucher WELT SP86435	744	85.0
gb F470133.1	Pseudopanax chathamense	voucher WELT SP86425	744	85.0
gb F470132.1	Pseudopanax linearis	voucher WELT SP86465	744	85.0
gb F470131.1	Pseudopanax linearis	voucher WELT SP86434	744	85.0
gb F470130.1	Pseudopanax ferox	voucher WELT SP86430	744	85.0
gb F470129.1	Pseudopanax gilliesii	voucher WELT SP86456	744	85.0

The interface also shows a left-hand navigation menu with options like 'Tools', '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', 'Meta-genomic analyses', 'FASTA manipulation', 'NCBI BLAST+', 'NCBI BLAST+ blastn Search', 'NCBI BLAST+ blastx Search', 'NCBI BLAST+ tblastn Search', 'NCBI BLAST+ tblastx Search', 'BLAST XML to tabular', and 'NGS, QC and manipulation'. The right-hand panel shows a 'History' section with a list of previous jobs, including 'UFGI Grad Demo' and several 'blastn on db' jobs.

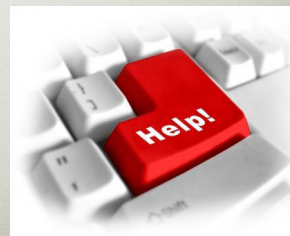
Training Schedule

- ✓ Jan 26: Intro to UFHPC, getting started
- ◆ Feb 2: Galaxy overview and basic usage
- ◆ Feb 9: Intro to Linux shell
- ◆ Feb 16: No session
- ◆ Feb 23: Modules and basic submission scripts
- ◆ Mar 1: Working with NGS data
- ◆ Mar 8: No session – Spring break
- ◆ Mar 15: Perl intro
- ◆ Mar 22: Perl in practice with Gordon Burleigh
- ◆ Mar 29: Statistical analyses at HPC intro
- ◆ Apr 5: R and SAS in practice with TBA
- ◆ Apr 12: TBA
- ◆ Apr 19: TBA

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

