UF Research Computing

Phylogenetics applications at UF HPC

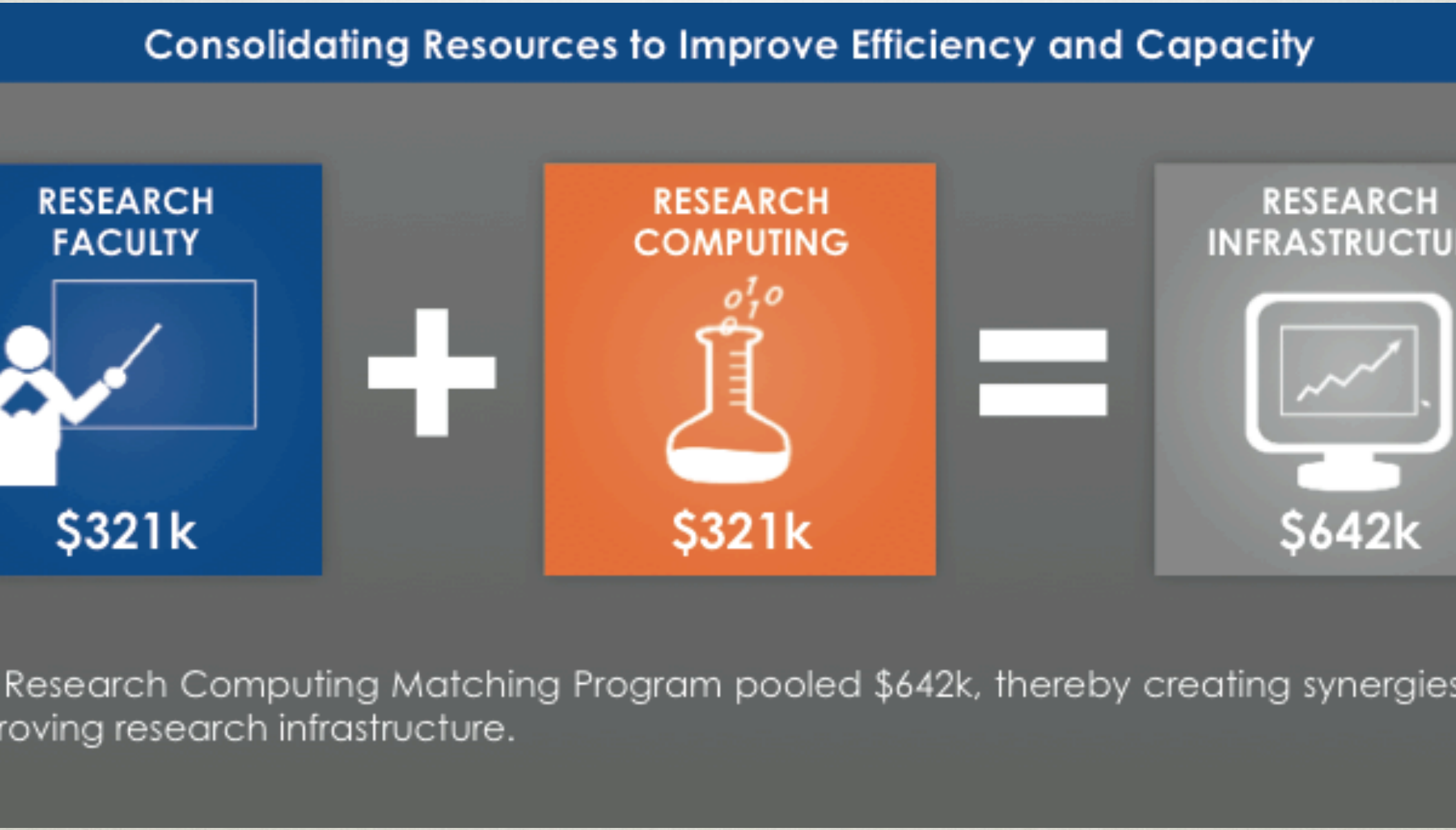
Matt Gitzendanner
Assoc. Sci., Biology / FLMNH
UF HPC User Support
magitz@ufl.edu

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

Matching Program



UF Research Computing

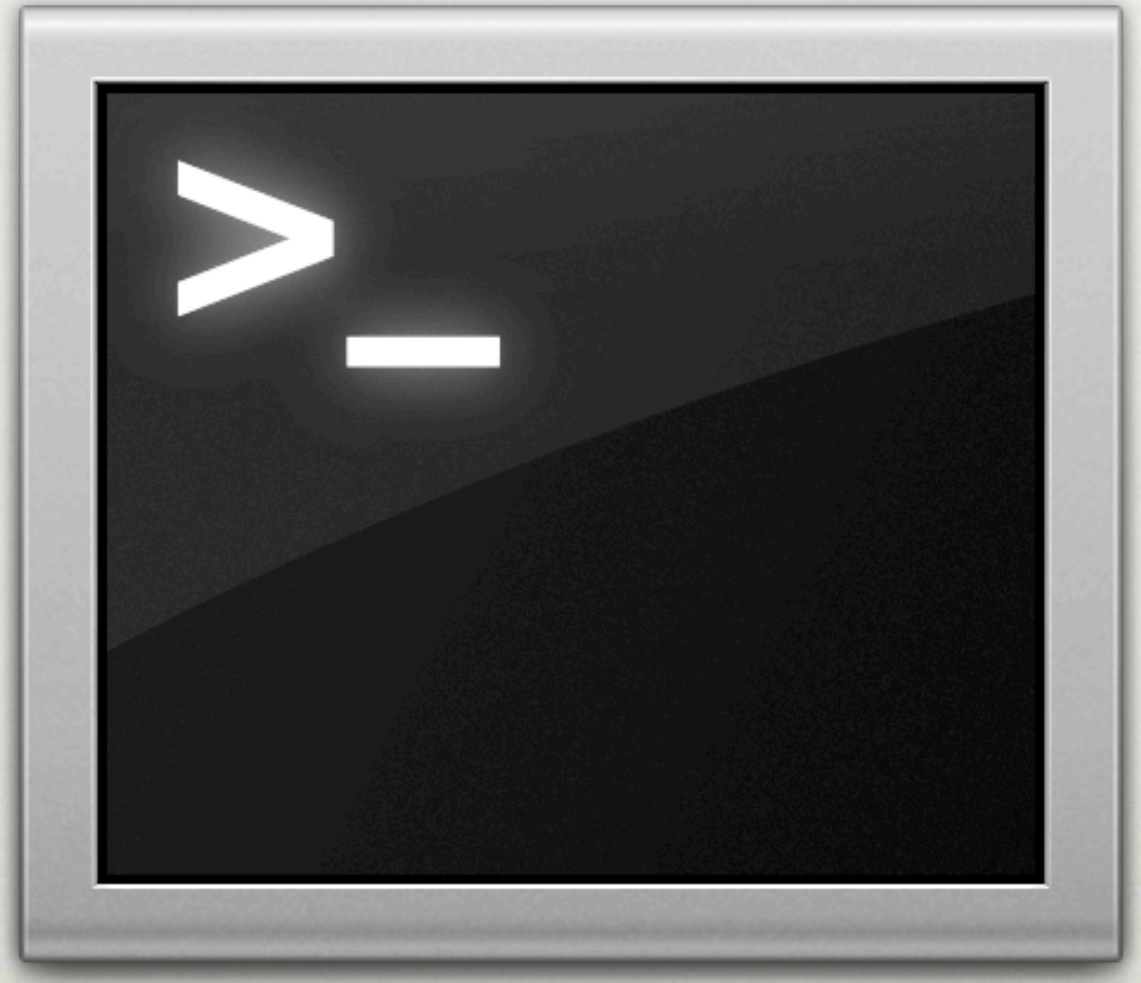
✦Shared Hardware Resources

- 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

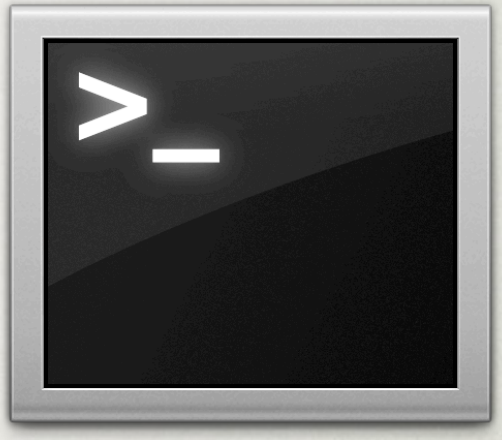
UF Research Computing



- Where do you start?

What can you run?

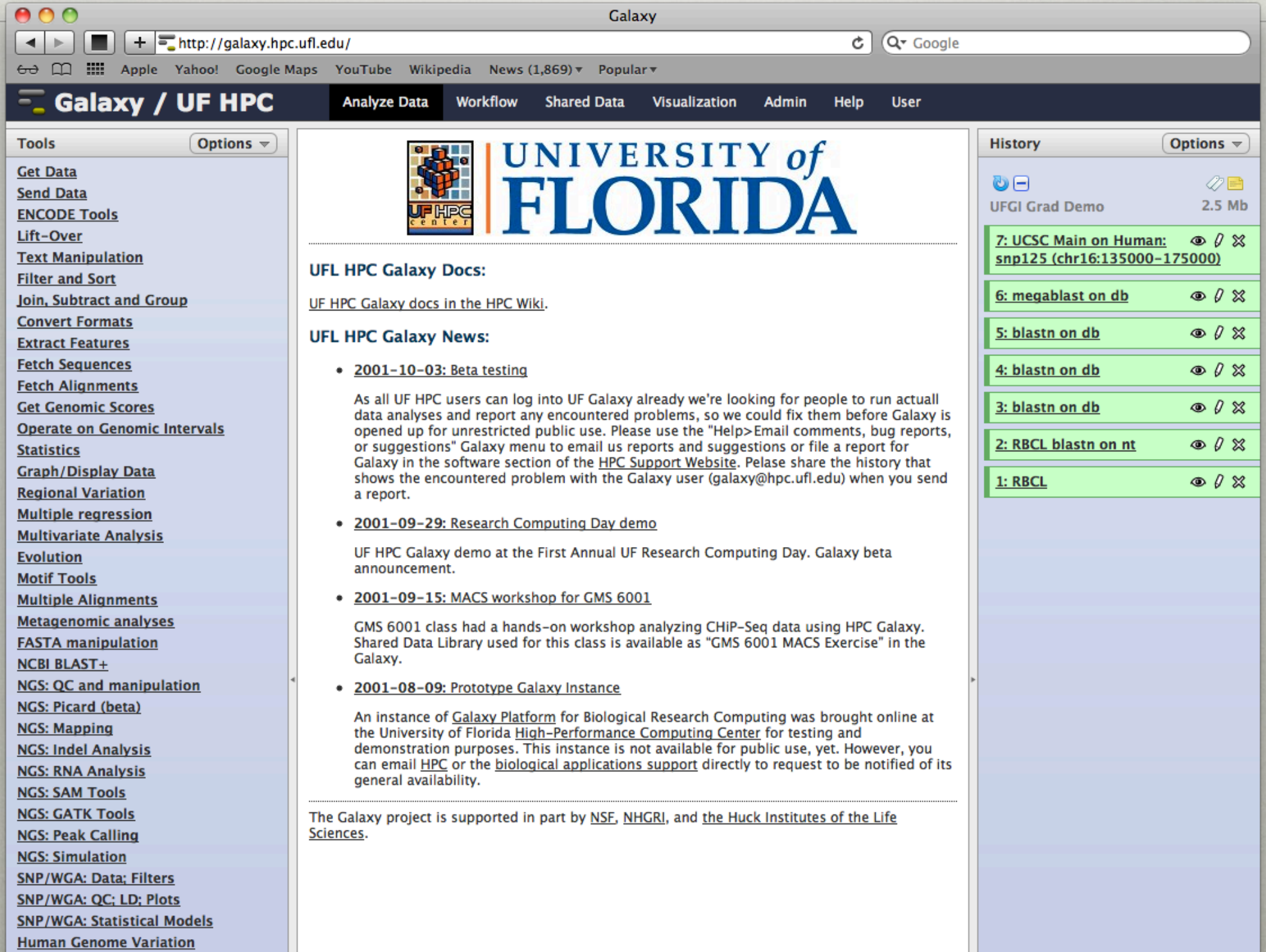
- Galaxy



- Linux
- Generally command line driven applications
- Graphical apps can be setup
  - SAS

  - BEAUti

Galaxy


Galaxy
http://galaxy.hpc.ufl.edu/
Google
Apple
Yahoo!
Google Maps
YouTube
Wikipedia
News (1,869)
Popular
Galaxy / UF HPC
Analyze Data
Workflow
Shared Data
Visualization
Admin
Help
User
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
UF HPC center
UNIVERSITY of FLORIDA
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.
History
Options
UFGI Grad Demo
2.5 Mb
7: UCSC Main on Human: snp125 (chr16:135000-175000)
6: megablast on db
5: blastn on db
4: blastn on db
3: blastn on db
2: RBCL blastn on nt
1: RBCL

Data intensive biology *for everyon*

- ✦ *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

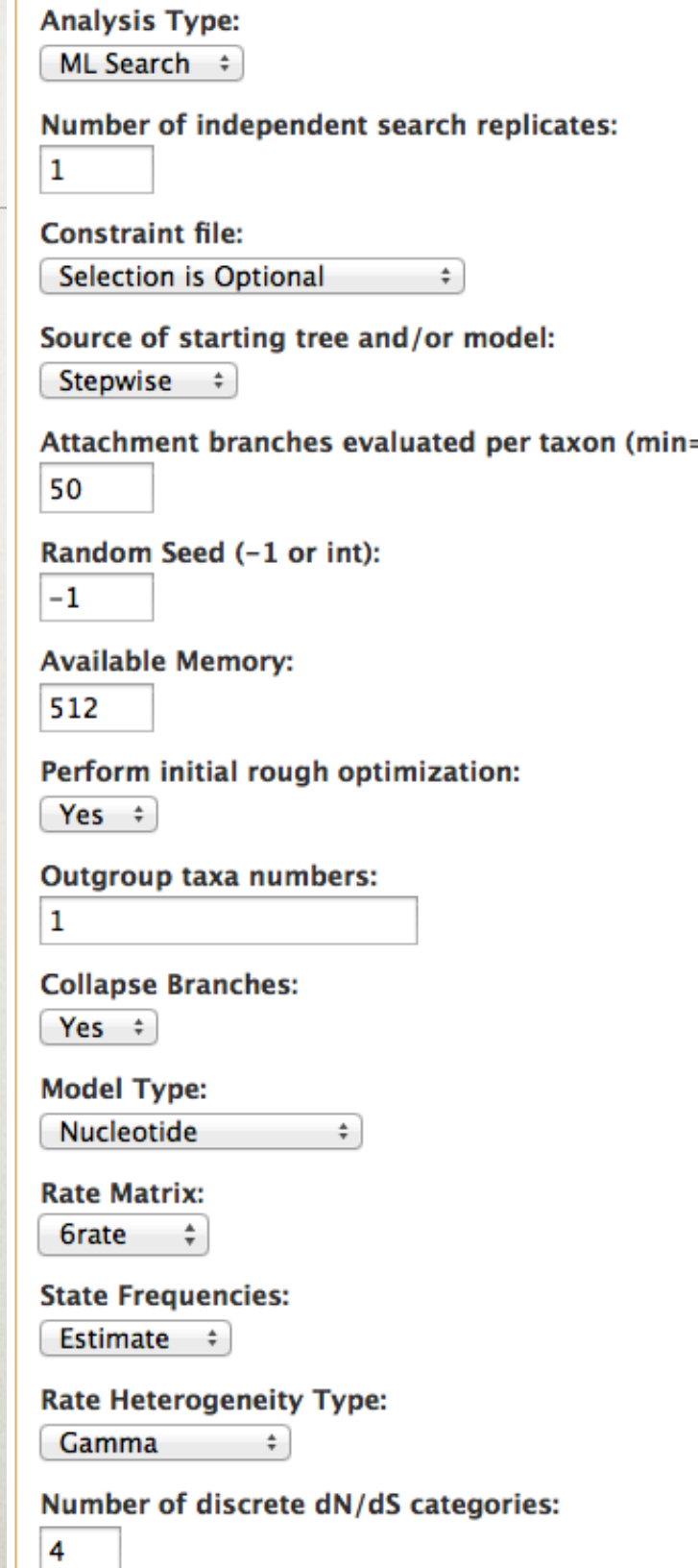
Phylogenetics

- Garli phylogenetic inference using the maximum-likelihood
- Beast Bayesian MCMC analysis of molecular sequences.
- TreeAnnotator BEAST tree annotator.

Garli 2.0

east and TreeAnnotator

RAxML in development



Cluster basics

User teraction

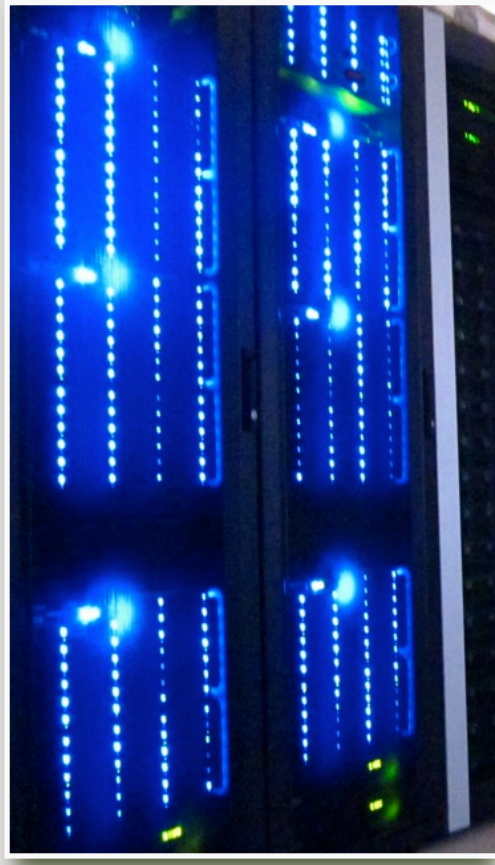
Galaxy

Login node (Head

Scheduler

Tell the scheduler what

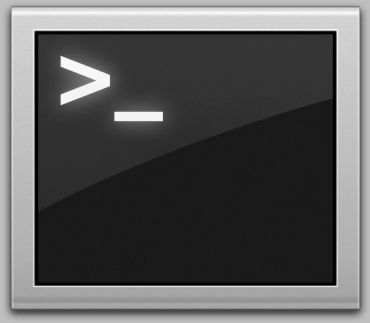
Comput resource



Your job runs on t

Tools

h client to connect to
submit.hpc.ufl.edu

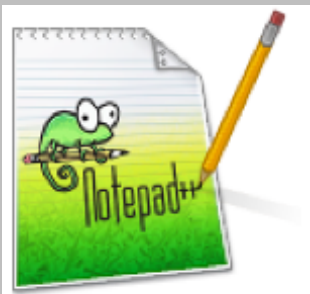


e.g.: Terminal, PuTTY

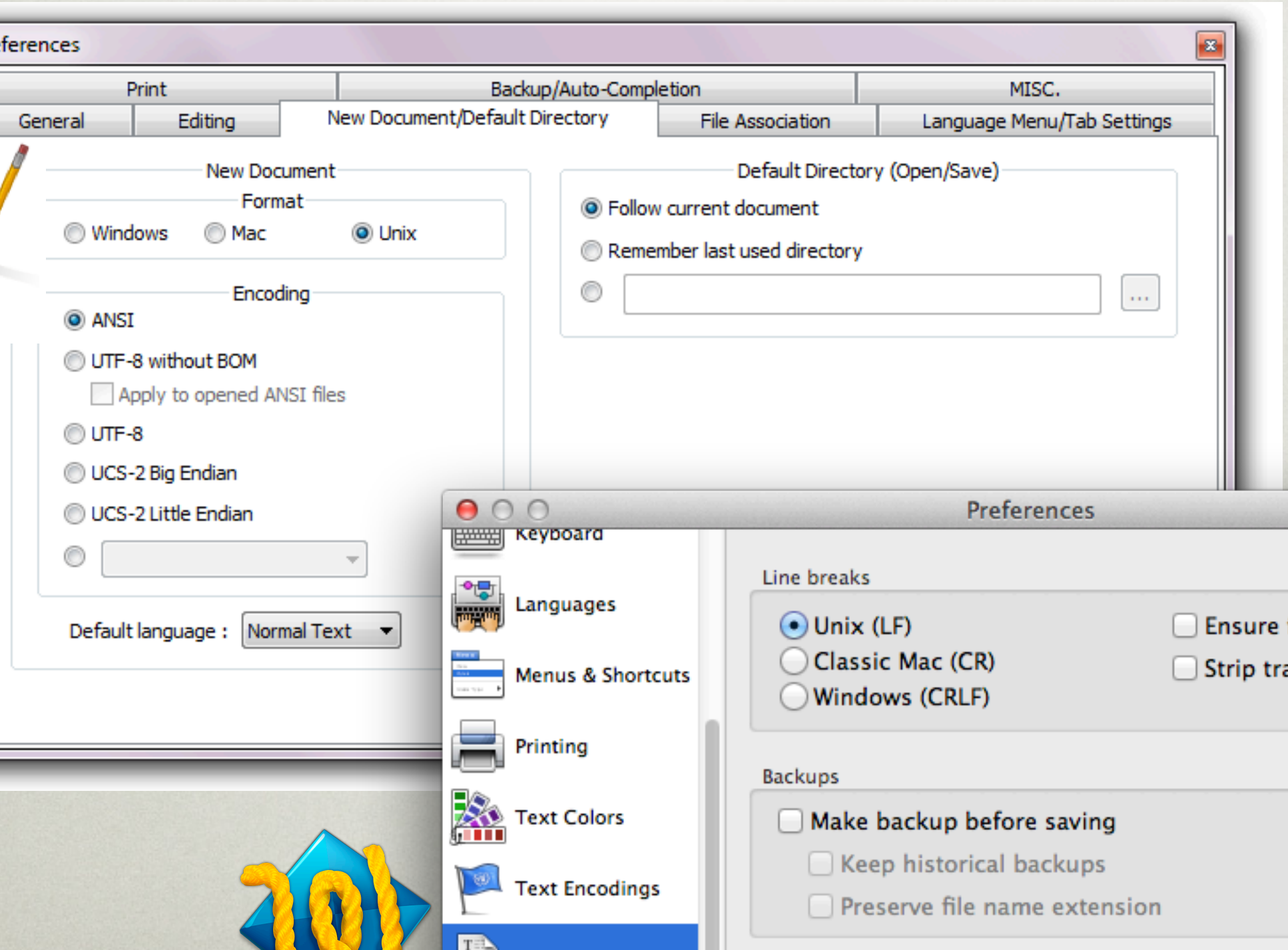
TP client to move files
/ from your computer

e.g.: Cyberduck, FileZilla

t editor to prepare files
cially on Windows, be sure to convert
line breaks to Unix, and *don't use Word*



Unix line breaks



Cluster basics

User teraction

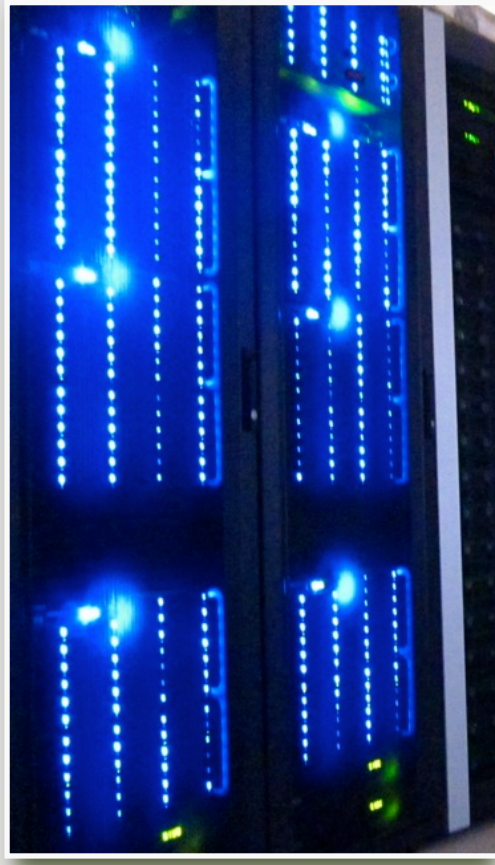
Galaxy

Login node (Head

Scheduler

Tell the scheduler what

Comput resource



Your job runs on t

Cluster login

ubmit.hpc.ufl.edu

submit1

submit2

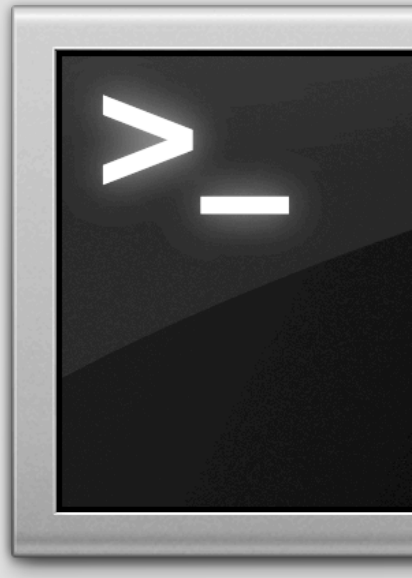
/home/
$USER

user>@submit.hpc.ufl.edu

indows: PuTTY

/Linux: Terminal

Head no



Login
head

Cluster login

Head no

ubmit.hpc.

submi

submi

/home
$USE

user>

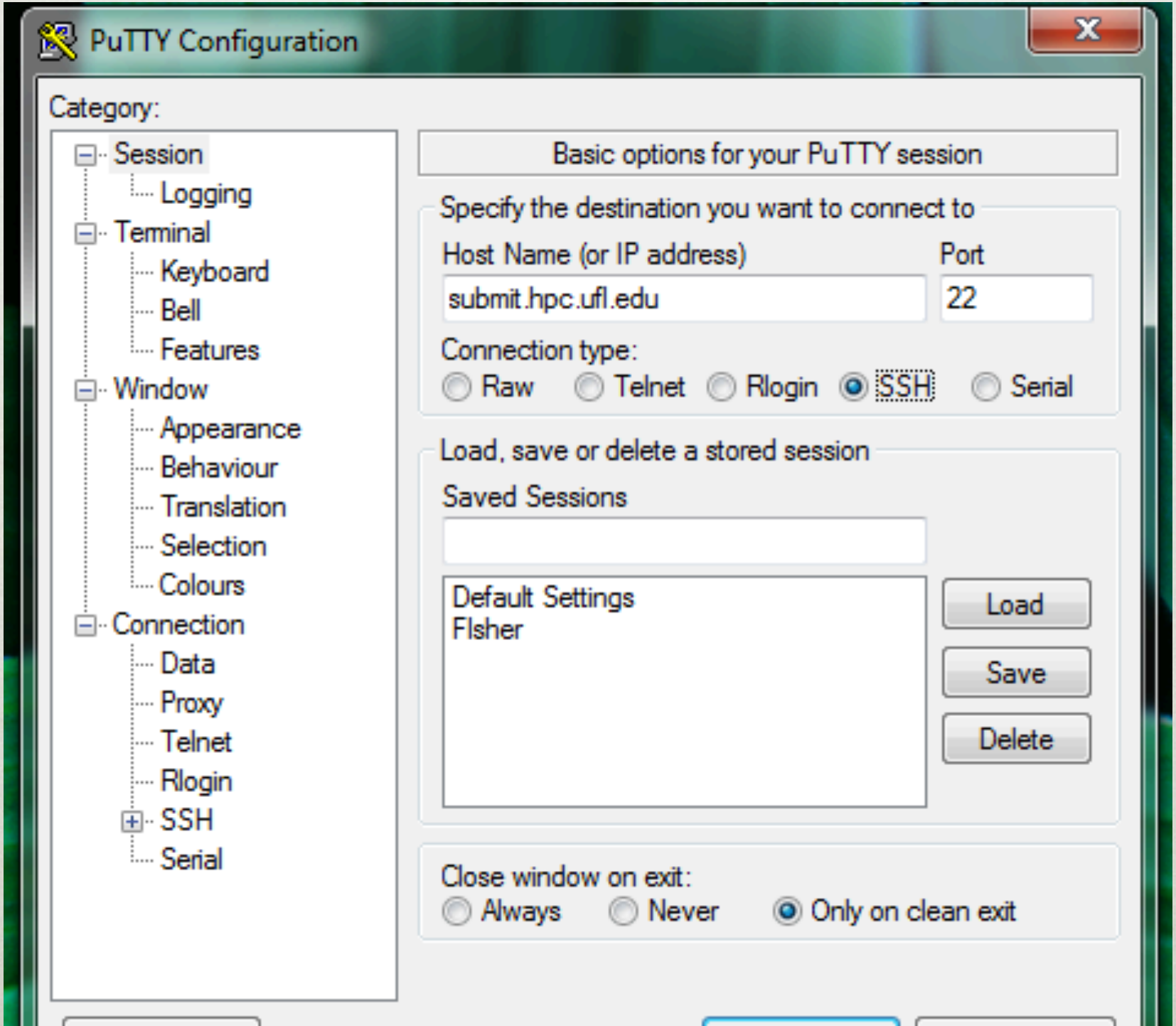
indov

/Lin

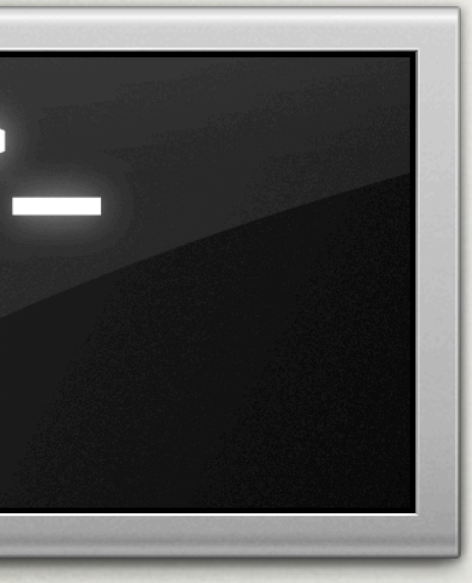
magitz@submit1:~ — ssh — bash — 67×17

```
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.e

Welcome to the UF HPC Center.

Do not run interactive jobs on the login nodes.  If you ne
run an interactive job, there are interactive/test nodes f

UF HPC Center Account Policies can be found here:

http://www.hpc.ufl.edu/users/accounts.php

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

Logging in


PuTTY Configuration
Category:
Session
Logging
Terminal
Keyboard
Bell
Features
Window
Appearance
Behaviour
Translation
Selection
Colours
Connection
Data
Proxy
Telnet
Rlogin
SSH
Serial
Basic options for your PuTTY session
Specify the destination you want to connect to
Host Name (or IP address)
Port
submit.hpc.ufl.edu
22
Connection type:
Raw
Telnet
Rlogin
SSH
Serial
Load, save or delete a stored session
Saved Sessions
Default Settings
FIsher
Load
Save
Delete
Close window on exit:
Always
Never
Only on clean exit

Linux Command Line



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

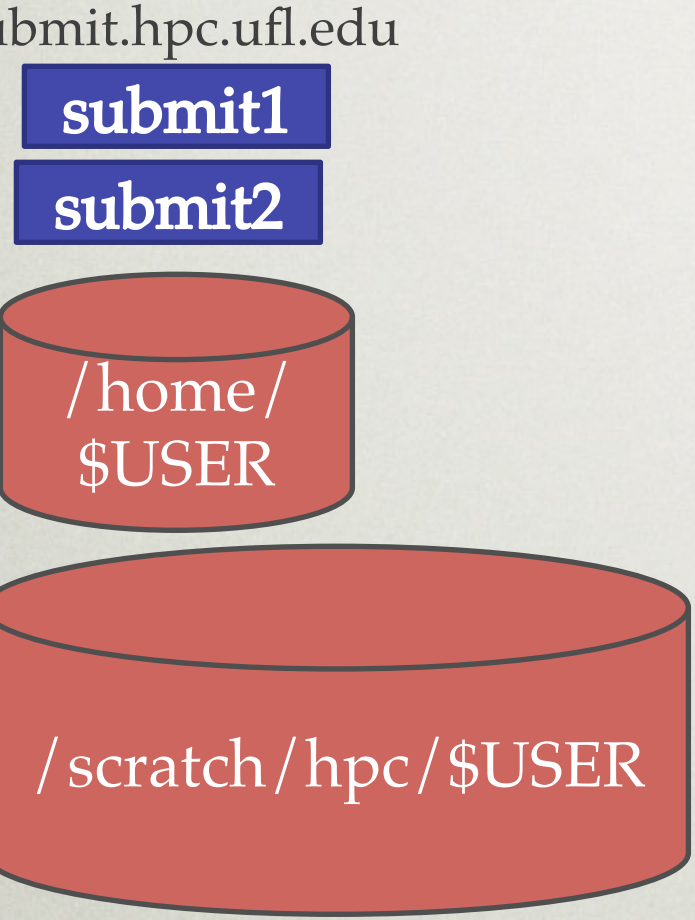
UF Research Computing

✦Storage

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

Must be us
all file I

Storage at HPC




```
$ cd /scratch/hpc/mag
```

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

aking life easier: Module syste

- Paths, libraries
- Compilers, MPI implementations, software versions
- All seamlessly taken care of for you
- Also allows for discovery
  - module spider
- **module load raxml**

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

Schedu

Tell th
scheduler

Nodes and processors

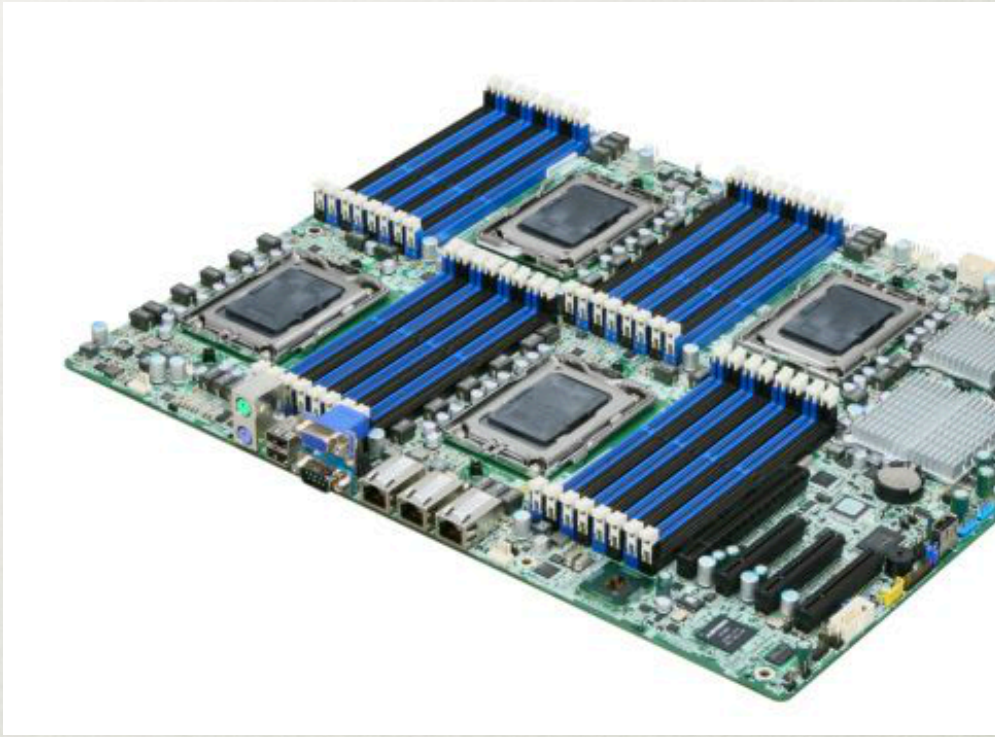
```
#S –l nodes=1:ppn=4
#S –l nodes=2:ppn=8
```

Parallelism

- **Message Passing Interface (MPI)**

- Messages passed among different nodes
- Can be slowed by time

- **Threaded, PTHREADS, OpenMP**



- All messages stay within a node
- Limited by CPUs and

RAM

```
#PBS –l pmem=900mb
```

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

Walltime

```
BS -l walltime=00:50:00
```

airly straight forward

s with all resource requests,
ccuracy helps ensure *your* jobs
nd all other jobs will run sooner

Schedu

Tell th
scheduler

RAxML

- ✦ raxml-SSE3
  - Single threaded
- ✦ raxml-PTHREADS-SSE3
  - Multi-threaded, all on one node
- ✦ raxml-HYBRID-SSE3
  - MPI and multi-threaded, span multiple nodes

MrBayes

- intel/11.1 mrbayes
  - mb –single threaded
- intel/11.1 openmpi mrbayes
  - mb –MPI version,
  - Can span multiple nodes
    - But doesn't need to: nodes=1:ppn=8 is much preferred to nodes=8:ppn=1
      - Faster for your job, fewer points of failure, doesn't partially occupy lots of nodes

GARLI

- For single ML search
  - Single threaded
  - Multi-threaded, probably not worth it
- For bootstrap
  - MPI, splits each replicate onto a processor

Others

- BayesPhylogenies
- BEAST
- PhyML
- PhyloBayes
- PAML
- SATé
- RAxML-Light

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

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