


**UF HPC Training**  
Working with NGS data  
March 11, 2013

### Fastq filtering example

1. Copy the file sample submission script to your scratch directory
  - a. Login: `ssh <user>@submit.hpc.ufl.edu`
  - b. cd to /scratch: `cd /scratch/hpc/<user>/`
  - c. Copy file: `cp /project/bio/training/2013-03-11/sickle_demo.pbs .`
2. Edit sickle\_demo.pbs to add you e-mail
  - a. `nano sickle_demo.pbs`
3. Submit this file and check out the results

### Fastq filtering demo from Galaxy dataset

1. The example script is in the training directory, called sickle\_demo\_galaxy.pbs
2. To get the path of a Galaxy dataset, click the  icon.

3. **Tool: FASTQ Groomer**

Name:	FASTQ Groomer on data 2
Created:	Oct 08, 2012
Filesize:	376.9 MB
Dbkey:	?
Format:	fastqsanger
Tool Version:	
Tool Standard Output:	<a href="#">stdout</a>
Tool Standard Error:	<a href="#">stderr</a>
Full Path:	<a href="#">/galaxy/run/prod/database/files/025/dataset_25369.dat</a>

See [http://wiki.hpc.ufl.edu/doc/Galaxy\\_Data\\_Import](http://wiki.hpc.ufl.edu/doc/Galaxy_Data_Import) for information on getting data back into Galaxy. This is the same method used for scp/sftp upload.

### Things to remember:

- Don't store data in /home/<user>--10GB limit
- Don't read and write data to /home/<user>
- /scratch/hpc/<user> is the current location for most users, /bio/ available for Bio investors, /project/ for longer term storage. New cluster will have new storage.
  - None of the storage is backed up
  - None of the storage is intended for long-term storage of data.
- Some tasks are quick and can be run interactively **on a test node**—`ssh test01` (We have six test nodes: test01–test06)
  - QC checking, filtering, data manipulation, etc.
  - **Do not run interactive jobs on submit nodes**
  - Even zipping of large files should be done on test node