

Phylocluster Access and Responsible Use Policies

Overview:

The Phylocluster is a shared resource for researchers affiliated with the Center for Computational Biology and Bioinformatics (CCBB). On any given day, the Phylocluster will be used by many individual users for their work. Technical problems with this resource will usually affect several users at once. Following the guidelines laid out in this document will help keep the Phylocluster up and running properly for all users.

As a phylocluster user, you have two main areas of access. The first is your home directory. This is from where you launch jobs and store files directly relating to the jobs you are currently running. You should also receive a space on your PI's share space. This space is for storing files medium- and long-term. Shared space can also be used to store files for collaborations between lab mates. Though these spaces belong to you and your lab, actions that you take within your space can affect the functionality for other users.

Guidelines:

- If you are running many jobs at once, package jobs into arrays. Arrays limit you from taking over the whole phylocluster at once by limiting the number of slots you run on.

Example: If you are bootstrapping a phylogenetic tree, each bootstrap replicate can take up space to run. If you package these replicates into arrays, you free up more space for other users to run jobs.

- If you are not using a file or group of files on a regular basis, save these files as in .tar format and put them into your lab's space on the share drive. Your home directory is not appropriate for long-term storage. If you will not be using a file or group of files in a given week, you should label them appropriately and store them as a .tar file. If you will not be using them in the course of a month or longer, they need to go into your lab's share space.

Example: You recently published a paper using some analyses you ran from your home directory on the Phylocluster. Now, you have all kinds of data in your home directory. Your home directory being overly full decreases the amount of space available for other users to store data and launch jobs. You should package these data into folders, make .tar files and move these into your lab's shared space.

- Work with your lab PI to decide what data need to be saved and for how long. Develop reasonable expectations of how much data you will be generating and inform your PI of this, so that appropriate storage can be allocated for you.

Example: Your PI may have a protocol for storing and archiving data from publications to ensure that the data are available to be accessed by other researchers on request.

- Make sure your scripts clean your workspace on the node after a job is completed.

Example: You are running a program that generates a log file, an output file with your answers, and a summary file that explains different parameter estimates related to results. You might copy the parameter estimates and the results back to your home directory for use in your analysis. But if you don't delete the other files left on the node, they will fill the node and limit the space other

users have to perform their own analysis. If you include a line in your run scripts to delete these extraneous files, everyone will have more room for their work.

- If you will be leaving UT for a long period (moving on to a postdoc, going into the field, etc), make arrangements for the long-term storage of your data. At minimum, this should involve tarring all your files for easy backing up and moving nonessential files into your PI's share space. If you are leaving UT Austin, this will likely involve tarring all your files and storing them in your PI's share space.

Example: If you leave for the semester (e.g., for field work), appropriately storing data will ensure not just ample space for everyone to work, but it will make backups of the system run more efficiently.

- Run only the smallest of scripts in your home directory. Running large computation in your home directory will take up memory on the head node, slowing down the cluster for everyone else using it.

Example: A UNIX command to pull a value from many text files is fine. Running software, performing Approximate Bayesian Computation or running an MCMC chain in R will slow down the Phylocluster for other users.

Policy violations:

If your jobs are violating the guidelines listed above, your actions may negatively impact other cluster users. If this is the case, your access to the Phylocluster may be revoked temporarily, and your PI may be contacted about your behavior.

How to get help:

If you aren't sure about how to do something, ask. You may know other students, postdocs or researchers who have used the cluster extensively and can help you. The CCBB senior systems administrator, Ming Cheng, can be reached at <mingcheng@utexas.edu>. We also have a wiki at <https://wikis.utexas.edu/display/CCBB/Home>