

Tutorial 07 - Github Review and installing tools remotely

How to do an assignment from now on

- 1) fork the TAs repository

Do this on Github in a web browser

- 2) clone the github repo to your local machine

Copy the url on Github, then `git clone <paste url>`

- 3) do your work locally; periodically store a version of your work locally and remotely

`git add -A`, `git commit -m "message"`, and `git push`

- 4) do one last add, commit, and push to record all of your changes

`git add -A`, `git commit -m "message"`, and `git push`

- 5) submit a pull request by 10:25 on Fridays to “turn in” your answers

Do this on Github in a web browser

Make a file here and teleport there

Assuming we have a text file called `data.txt` in our **local** present working directory, we can copy it to the `Private` directory in our home directory on the **remote machine** called `crcfe01.crc.nd.edu` with the following syntax **on our local machine**:

```
scp source destination
```

```
scp local_file remote_address:remote_path
```

```
scp ./data.txt user_name@crcfe01.crc.nd.edu:~/Private
```

Challenge - Make a file there and teleport it here

Assuming we have a text file called `data.txt` in the `Private` directory in our home directory on the **remote machine** called `crcfe01.crc.nd.edu`, we can copy it to our **local** present working directory with the following syntax **on our local machine**:

```
scp source destination
```

```
scp remote_address:remote_file local_path
```

```
scp user_name@crcfe01.crc.nd.edu:~/Private/data.txt .
```

Why Muscle and hmmer?

When used together these tools represent a simple, but common, workflow in bioinformatics.

- ▶ We have some information in the form of reference gene sequences
- ▶ We have some additional data from other sources
- ▶ We want to find gene sequences like our reference sequences in the additional data

This allows us to answer biological questions like - Does this organism have the genes to do this biological function? or How abundant are these genes compared to those genes in an environment?

Muscle

Recall that Muscle is designed to generate multiple sequence alignments. These alignments attempt to identify and “line up” conserved residues in sequences. Once an alignment is generated it can be used to infer evolutionary relationships as the basis of a molecular-based phylogenetic tree or create profile hidden markov models (HMMs). The basic syntax for generating an alignment with Muscle is:

```
./muscle -in refsequencesfile -out filenameforresults
```

hmmer

Remember that `hmmer` is used to find sequences that match a particular pattern of interest. We use `hmmer` to describe a pattern of interest with a profile HMM. A common application of `hmmer` is to generate a profile HMM for a particular protein family, domain, or active site sequence. The two key `hmmer` tools we'll need are `hmmbuild`, to build a profile HMM from a sequence alignment (the output from `muscle`) and `hmmsearch`, to search a sequence database with a profile HMM (the output from `hmmbuild`). The basic syntax for these tools are:

```
./hmmbuild resultsfile alignmentfilefrommuscle
```

```
./hmmsearch --tblout resultsfile hmmbuildresults filetosearch
```

Installation is hard

It is common to run into issues with tool installation, so don't be discouraged!

Issues arise because the installation process is slightly different for each tool and for each system you try to install it on. In part, this is why we are using the CRC for this portion of the class.

Using the slides from Wednesday's lecture, try to get the two tools installed!

Walkthrough is for you to practice for the Bioinformatics Project!

There is no exercise assigned today. Instead we will assign the Bioinformatics Project in lecture on Monday.

To get some practice with these tools before starting the Bioinformatics project, go to Stuart's repo (link on Sakai), then
fork -> clone -> go to work