

Micro 612 – Bacterial Genomics Pre-course Homework

Instructor – Evan Snitkin (esnitkin@umich.edu)

Course assistant – Ali Pirani (apirani@umich.edu)

Course assistant – Stephanie Thiede (thiede@umich.edu)

Course assistant – Zena Lapp (zenalapp@umich.edu)

Review you Unix and R

While the focus of the workshop is on analyzing bacterial genomes, you will also be learning about doing computational research using a combination of Unix and R. Unix and R form the foundation for not just microbial genomics research, but bioinformatics research in general, which is why we had you first take the Software Carpentry workshop before this one. If you took an earlier SWC workshop, I strongly recommend reviewing the SWC materials on Unix (<https://swcarpentry.github.io/shell-novice/>) and R (<http://swcarpentry.github.io/r-novice-inflammation/>). The Unix material is especially important, as many of the tools we will be using are only available on a Unix platform. If you would like additional practice with Unix, check out command line boot camp: http://rik.smith-unna.com/command_line_bootcamp/?id=9xnbkx6eaof

Getting setup on Flux

Next, we want to get you setup on the Flux compute cluster, which is the platform we will be working on during the workshop. This is a computational system available to researchers at the University of Michigan, and in addition to facilitating big data analysis; it also acts as a pre-configured compute environment for doing a lot of common bioinformatics analyses. First, to use Flux you have to request an account (<https://arc-ts.umich.edu/fluxform>) and activate Duo authentication (<http://its.umich.edu/accounts-access/uniqnames-passwords/two-factor-authentication/enroll-in-duo>).

Once you have gotten your Flux account activated, try to login. To do this, open a Unix session on your computer (terminal for Macs and Gitbash for Windows) and establish an ssh (secure shell) connection to flux by typing the following command:

```
ssh username@flux-login.arc-ts.umich.edu
```

You will then be prompted for your level 1 password, followed by your preferred method for Duo authentication. If you have successfully logged in, you should see some welcome text and usage agreement information. **Note that if you are off campus, you will have to be on the VPN to login to Flux.**

Getting oriented on Flux

When you log in to a remote system, you are typically placed in your home directory. Use the “pwd” command (present working directory) to see where your home directory is on flux.

For the purposes of the class we created an alternative home directory inside the class folder.

To verify that your home directory exists:

- i) Go to “/scratch/micro612w18_fluxod/” (**type “cd /scratch/micro612w18_fluxod/”**)
- ii) List the directories and verify that a directory exists that corresponds to your username (**use ls command**)
- iii) Go into your user directory (**use cd command**)
- iv) Type “pwd” again to see what the path is to your class home directory

Genomics tools for the workshop

In addition to the work we will be doing on Flux and R, we will use several graphical tools for the purposes of data visualization.

Please download and install the following on your computer.

- i) IGV - <https://www.broadinstitute.org/software/igv/log-in>
- ii) Artemis - <http://www.sanger.ac.uk/science/tools/artemis>
- iii) ACT - <http://www.sanger.ac.uk/science/tools/artemis-comparison-tool-act>
- iv) Mauve - <http://darlinglab.org/mauve/download.html>
- v) Seaview - <http://doua.prabi.fr/software/seaview>

Additional background reading

- i) Bacterial genomics review
<http://www.nature.com/nrmicro/journal/v13/n12/full/nrmicro3565.html>
- ii) Review on the use of genomics to study epidemiology
<http://www.nature.com/nrg/journal/v13/n9/full/nrg3226.html>
<http://www.sciencedirect.com/science/article/pii/S1369527414001635>
- iii) Review on the use of genomics to study bacterial pathogenesis
<http://www.sciencedirect.com/science/article/pii/S1931312816301512>