Part 2: Extend your slurm script to run fastqc on the downloaded file

Author: Zach Maas

The goal of this exercise is to extend the script you just wrote to run fastqc, a quality control tool, on the fastq file that you're downloading. This requires that you use the module system available on the server to load fastqc and then to download the results from the server to your local machine so that you can look at them.

Things you need to do:

1. Write or extend your script to run fastqc on the fastq file that you downloaded. You will need to use the module system to load fastqc first inside of your script, using fastqc version 0.11.5. Put the output of this fastqc run into your results folder.
2. Use rsync to copy the output files from your fastqc run to your local machine and open them in a browser. How do they look?

Some useful things to know:

- The module command allows you to load different programs for use in your script
- module spider will find a program if available
- module load will load a program to be run. You may need to specify the version
- The rsync command allows you to not just move files around on a single machine, but between multiple machines
- The -e flag allows you to specify that you're using the ssh protocol
- The -P flag will show you progress
- The -r flag will copy directories and the files in them recursively
- Make sure you're on the machine that you want to download to. Why can't you send a file from the server to your machine from the server while logged into the server? Why do you have to download the file using your local machine instead?
- You can view some documentation on fastqc here: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Often project pages like that will have helpful documentation for explaining what a tool does, although sometimes it's better to read the original paper instead
- It's essential to keep track of the specific versions of tools that you use so that your code and analysis can be reproduced by others