

Class 13: Unix for Bioinformatics

Katherine Wong (A16162648)

11/9/2021

Q1. What does the star character accomplish here? Ask Barry, or your class neighbor, if you are not sure!

The star character is the wildcard. This means that if we use the command `'gunzip *.faa.gz'`, this will uncompress all files that end with `.faa.gz`.

Q2. How many sequences are in this `mouse.1.protein.faa` file? Hint: Try using `grep` to figure this out...

67876 protein sequences. Using `"grep -c">" mouse.1.protein.faa`.

Q3. What happens if you run the above command without the `> mm-first.fa` part?

Prints to stdout the first two FASTA protein sequences.

Q4. What happens if you were to use two `'>'` symbols (i.e. `» mm-first.fa`)?

It redirects the output by appending the two protein sequences to the end of the existing `mm-first.fa` file

Q5. How would you determine how many sequences are in the `mm-second.fa` file?

`grep -c ">" mm-second.fa`. There are 83 sequences in the file.

Q6. Note the addition of the `-r` option here: What is its purpose? Also what about the `*`, what is its purpose here?

The purpose of `-r` here is that it recursively copies the files and its sub-directories. So everything in the directory should be copied over, including subdirectories and the files within those subdirectories, etc. The purpose of the `*` or the wildcard is that it copies all the files and directories in specified directory, in this case `home/work/` from the remote machine.