Class 15: Lab 2, Using Remote AWS Instances

Kira Jung

My **Instance ID**: i-02a406fbed4ef34ee

When uncompressing the FASTA protein files with `gunzip *.faa.gz`: **Q1**: What does the star character accomplish here? *Answer*: The asterisk allows all three of the .faa.gz files to be uncompressed at once without individually typing and entering the three files.

Q2: How many sequences are in the `mouse.1.protein.faa` file? *Answer*: Using `grep -c ">"` `mouse.1.protein.faa` there are 36872 sequences in that file.

Q3: What happens if you run the above command without the `> mm-first.fa` part? *Answer*: Without the output redirection or making a new file called `mm-first.fa`, we would not be able to move the first two sequences into the newly created file.

Q4: What happens if you were to use two '>' symbols (i.e. » mm-first.fa)? *Answer*: The '»' symbol implies that you are moving the sequences to a pre-existing file. Since we didn't previously create a new file called `mm-first.fa` the command would not be successful.

Q5: How would you determine how many sequences are in the mm-second.fa file? *Answer*: There are 87 sequences in the `mm-second.fa` files. This was done by entering `grep -c ">"` `mm-second.fa` into the command line.

```
scp -i ~/Downloads/bioinf_kira.pem -r ubuntu@ip-172-31-14-243:~/work/* .
```

Q6: Note the addition of the -r option here: What is it's purpose? Also what about the , *what is it's purpose here?* Answer*: The -r option means read permissions and allows the file to be opened and viewed.