Chapter 1

Salil jindal

2024-10-04

## To login in the OSCER

ssh biol726306@schooner.oscer.ou.edu

## To access software through the environment: (This needs to be done only once)

module load Mamba  
  
mamba init

## To activate the environment: (Everytime you start wokring)

mamba activate /home/mbtoomey/.conda/envs/BIOL7263\_Genomics

## To retrieve sequence data:

### We can use wget command to retrieve data from NCBI, EBI or any other databse. Here we will download raw data of ecoli:

wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR857/SRR857279/SRR857279\_1.fastq.gz  
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR857/SRR857279/SRR857279\_2.fastq.gz

## To prevent accidental deletion of files, we can make the files read-only using:

chmod 444 \*.gz

This command changes the permissions of all files in the current directory that have a .gz extension (which typically indicates Gzip-compressed files) to read-only for everyone.

In similar fashion, sequence data of pseudomonas was retrieved.

wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR491/SRR491287/SRR491287\_1.fastq.gz  
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR491/SRR491287/SRR491287\_2.fastq.gz

## To get reference data:

Again, using the ‘wget’:mag: command we can download files directly from the web to our local dicrectory. The files are ‘gzipped’, this means they are compressed to save space, it also allows us to make sure the data has not been corrupted during the transfer. We will also need to unzip them with the program ‘gunzip’:mag

wget ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/845/GCF\_000005845.2\_ASM584v2/GCF\_000005845.2\_ASM584v2\_genomic.fna.gz  
wget ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/845/GCF\_000005845.2\_ASM584v2/GCF\_000005845.2\_ASM584v2\_genomic.gff.gz

## To unzip .gz files:

gunzip ecoli/\*.gz

# Change write permissions, so that we can’t edit them by accident

chmod -R 444 ecoli/*.fna chmod -R 444 ecoli/*.gff

## To create database (pfam):

We need to download the HMMs and .dat files for Pfam-A

wget http://ftp.ebi.ac.uk/pub/databases/Pfam/current\_release/Pfam-A.hmm.gz  
wget http://ftp.ebi.ac.uk/pub/databases/Pfam/current\_release/Pfam-A.hmm.dat.gz  
wget http://ftp.ebi.ac.uk/pub/databases/Pfam/current\_release/active\_site.dat.gz

Once downloaded, we can unzip the files using gunzip command.

## For more information:

<https://github.com/mbtoomey/genomics_adventure/blob/release/home.md>