Fill in with a list of commands

1. Login to cluster, Download data to the cluster, perform fastqc on the data

wget https://www.dropbox.com/s/z7lxix6o4m3ijcy/AE3002P_S95_L007_R1.fastq.gz?dl=0

mv 'AE3002P_S95_L007_R1.fastq.gz?dl=0' AE3002P_S95_L007_R1.fastq.gz

Make a sbatch script then run fastqc put screen shot result in this document.

Use module avail to find fastqc.

Then module load in your sbatch script to run fastqc.

fastqc AE3002P_S95_L007_R1.fastq.gz (is the command for the sbatch script)

Download the html result and place a single plot from fastqc here. Also, provide a screen shot of using ls -thor that the files have been converted.

2. Oh NO! You have phiX in your data! Figure out how much is in the file: AE3002P_S95_L007_R1.fastq.gz

Download the phix174 genome from dropbox

wget https://www.dropbox.com/s/6idv927seno7kul/phix174.fa?dl=0

mv ‘phix174.fa?dl=0’ phix174.fa

Use bowtie2 to map the phix genome. Make an sbatch script and run it.

Use module avail to find bowtie2

Then module load in your sbatch script to run bowtie2 commands:

module load BOWTIE (find it in module avail)

#make db for phix174

bowtie2-build phix174.fa phixDB

#map reads to phixDB

bowtie2 -p 28 -x phixDB -q AE3002P_S95_L007_R1.fastq.gz -S AE3002P_local.sam –very-sensitive-local

When completed put the number of reads mapped output here. Also, provide a screen shot of using ls -thor that the files have been converted.

3. Okay! Now remove the phiX reads from the sam file and convert back to a fastq

Use module avail to find samtools, then use module load in you sbatch script.

Commands:

#convert sam to bam

samtools view -bS AE3002P_local.sam > AE3002P_local.bam

#convert bam to unmapped bam

samtools view -b -f 4 AE3002P_local.bam >AE3002P_unmapped-local.bam

#convert unmapped bam to fastq

samtools fastq AE3002P_unmapped-local.bam >AE3002P_unmapped-local.fastq

Use either fastqc or other to count the number of sequences. Give the numbers here (original vs. removed vs. final reads). Also, provide a screen shot of using ls -thor that the files have been converted.

4. Provide an updated github with a repo for the course. With your materials and notes. Within the BINF2111 repo (use bin, data, src, doc, results). Also, update to add your photo and your profile. Provide links to your github here with a screen shot.

5. Write a bash script that tells me my username, current directory, the location of our desktop directory, and the date/time with the cluster! Also, provide a screen shot of using ls -thor that the files have been converted. And, the output of the script here.

BONUS 1 (3pts):

Write a bash script that prints all the lengths of the sequences (MultiN.fastq)
For example, seq 1 – 101 etc. Place code here and make an executable script.

EXTRA BONUS (5 pts):

Convert all of example2.fasta into it's amino acids using any command(s) you want?

- Script here