**connect to the lab computers via ssh**

ssh epp622@160.36.207.XX

**make your own project folder and copy the qc pipeline directory into it**

cd Documents/EPP622/projects

mkdir <netid> (make a directory for yourself using your netid)

cd <netid>

cp -r ../../qc\_pipeline ./ (copy the folder 'qc\_pipeline' into your personal project directory)

**inital qc (this takes a few minutes)**

cd qc\_pipeline/reads/

bash ./fastx\_qc.sh (this script calls fastx toolkit and creates a quality distribution for reads in the 'qc' folder)

**trim buffer sequence (first 6 positions)**

python3 ../tools/trimmer.py Beauregard\_R1.fastq 6 0

python3 ../tools/trimmer.py Beauregard\_R2.fastq 6 0

optional: use sed to look at the difference in sequence length

sed -n '2,2p' trimmed\_Beauregard\_R1.fastq

sed -n '2,2p' Beauregard\_R1.fastq

**demultiplex and remove barcodes**

cd demultiplex

bash ./dm\_forward.sh (this script calls sabre to demultiplex your samples into 24 smaller fastq files using the provided barcodes)

bash ./dm\_reverse.sh (this script calls sabre again using 16 reverse barcodes on the 24 fastq files, resulting in 384 demultiplexed sets of reads for the forward and reverse reads, each)

**isolate reads that possess the expected restriction sites**

bash ./overhang.sh (this script only keeps reads that possess "TCC" or "TCT" after the barcode region, reads are concatenated based on length and fastx toolkit is called again to give us quality scores in the 'qc' folder)

**trim the low-quality ends of the variable-length reads**

bash ./trimmer.sh

**remove reads that do not have a Phred score > 30 across ALL reads (very stringent)**

bash ./filter.sh

cd ../filtered

fastx\_quality\_stats -Q33 -i final\_Beauregard\_R1.fastq -o stats\_R1.txt

fastx\_quality\_stats -Q33 -i final\_Beauregard\_R2.fastq -o stats\_R2.txt

fastq\_quality\_boxplot\_graph.sh -i stats\_R1.txt -t Quality-Chart -o qual\_stats\_R1.png

fastq\_quality\_boxplot\_graph.sh -i stats\_R2.txt -t Quality-Chart -o qual\_stats\_R2.png