QuantBio- week2

Step 1:

Step 2:

I didn’t write a script.

bwa mem -R "@RG\tID:A01\_35\tSM:A01\_35" -o "A01\_35.sam" sacCer3.fa A01\_35.fastq

Step 3:

samtools sort -o A01\_35.bam -O bam A01\_35.sam

Step 4:

freebayes -f sacCer3.fa --genotype-qualities -p 1 \*.bam >results.vcf

Step 5:

vcffilter -f "QUAL > 20" results.vcf > qualresults.vcf

\*I know this wouldn’t provide a 0.99 probability that the snp is polymorphic, but I’m not totally sure what numbers you would use. I assume the closer to .5 the better because that’s the least likely number to get if you just had single errors that happened to be on the same nucleotide.

Step 6:

vcfallelicprimitives -g -k qualresults.vcf > decompresults.vcf

step 7:

snpeff ann -download R64-1-1.86 -csvStats statsresults.csv decompresults.vcf > annotatedresults.csv

head -n 1000 annotatedresults.csv > HW2.csv

step 8:

see jupyter notebook