**SUPPLEMENTAL**

**Supplemental 1A. VCF file production from fastq files with Orca1 server.**

#!/bin/bash

#This is assuming that you transferred the wanted fastq files onto #the server

mkdir MICB421

cd ~/MICB421

mkdir concat

cd ~/MICB421/concat/

cat \*.fastq>compiled\_final.fastq

minimap2 -x map-ont -a Reference.fasta compiled\_final.fastq | samtools view -bS | samtools sort > aln\_final.sorted.bam

echo SORTING COMPLETE

samtools mpileup -u -f Reference.fasta aln\_final.sorted.bam -I>aln\_final.bcf

bcftools call --ploidy 1 -mv aln\_final.bcf> aln\_final.vcf

**Supplemental 1B. VCF Annotation with snpEff and snpSift on Local Drive.**

#!/bin/bash

#transferring files from Orca1 server

scp cwielunski\_mb19@orca1.bcgsc.ca:/home/cwielunski\_mb19/MICB421/concat/Final.vcf ~/Documents/Assignments-uni/5th\_Year\_Courses/MICB421/Project\_2/snpEFF/snpEff

#working directory

cd [your working directory]

#executing snpEff for annotation

java -jar snpEff.jar -v Escherichia\_coli\_str\_k\_12\_substr\_mg1655 Final.vcf >Final\_annotated.vcf

#sifting through annotated vcf file for important information

java -jar SnpSift.jar extractFields -s "," -e "." Final\_annotated.vcf CHROM POS REF ALT QUAL "EFF[\*].EFFECT" "EFF[\*].AA" "EFF[\*].GENE">Final\_annotated\_sifted.vcf

**Supplemental 1C. Annotated VCF Sorting with Rstudio.**

library(data.table)

library(tidyverse)

library(reshape2)

#directory and opening snp\_EFF

setwd([your working directory])

snpEff\_genes\_final=fread("snpEFF/snpEff/genes.txt")

snpEff\_genes\_filtered\_final=filter(snpEff\_genes\_final,variants\_effect\_missense\_variant!=0)

amino\_vcf=fread([wherever your Final\_annotated\_sifted.vcf file is])

Filtered\_amino\_1 <- separate(amino\_vcf,"EFF[\*].EFFECT",into=c("Primary\_Effect"),sep=",") %>%

 separate("EFF[\*].GENE",into=c("Primary\_Gene"),sep=",") %>%

 filter(Primary\_Effect=="missense\_variant")%>%

 separate("EFF[\*].AA",into=c("AA\_Substitution"),sep=",") %>%

 separate("AA\_Substitution",into=c("p.","AA\_Substitution")) %>%

 select(-"p.")

Merged\_amino <- merge(Filtered\_amino\_1, snpEff\_genes\_filtered\_final, by.x="Primary\_Gene", by.y="#GeneName")

write\_csv(Merged\_amino, "Final\_excel.csv")