# For the user to fill in after running MetaCHIP and prodigal separately (ie not using

# MetaCHIP prodigal calls)

conda\_source="/monfs00/scratch/lwoo0007/WoodsL/miniconda/bin/activate"

conda\_env="/monfs00/scratch/lwoo0007/WoodsL/miniconda/conda/envs/seqkit\_env"

metachip\_all\_HGT\_faa="Mackay\_MAGs\_noblast\_MetaCHIP\_wd/Mackay\_MAGs\_noblast\_combined\_pcofg\_HGTs\_ip90\_al200bp\_c75\_ei80\_f10kbp/Mackay\_MAGs\_noblast\_pcofg\_detected\_HGTs\_recipient\_genes.faa"

prodigal\_folder="../../prodigal-M/Mackay\_MAGs/"

op\_dir="./"

subfolder\_prefix=$(basename $metachip\_all\_HGT\_faa ".faa")

metachip\_ids\_op\_folder="${op\_dir}/${subfolder\_prefix}\_by\_sample\_metachip\_ids/"

prodigal\_ids\_op\_folder="${op\_dir}/${subfolder\_prefix}\_by\_sample\_prodigal\_ids/"

echo $metachip\_ids\_op\_folder

# Output folders

if [ ! -d $op\_dir ]

then

mkdir $op\_dir

fi

if [ ! -d $metachip\_ids\_op\_folder ]

then

mkdir $metachip\_ids\_op\_folder

fi

if [ ! -d $prodigal\_ids\_op\_folder ]

then

mkdir $prodigal\_ids\_op\_folder

fi

source $conda\_source base

conda activate $conda\_env

for item in ${prodigal\_folder}/\*.faa

do

sample\_name=$(basename $item ".faa")

grep $sample\_name $metachip\_all\_HGT\_faa | sed 's/^.//' > ${op\_dir}/temp\_${sample\_name}\_ids.txt

seqkit grep -f ${op\_dir}/temp\_${sample\_name}\_ids.txt $metachip\_all\_HGT\_faa | seqkit mutate -p 1:M > ${op\_dir}/temp\_${sample\_name}\_mutated.faa

seqkit common -s ${op\_dir}/temp\_${sample\_name}\_mutated.faa ${prodigal\_folder}/${sample\_name}.faa | seqkit sort -s > ${metachip\_ids\_op\_folder}/${sample\_name}.faa

seqkit common -s ${prodigal\_folder}/${sample\_name}.faa ${op\_dir}/temp\_${sample\_name}\_mutated.faa | seqkit sort -s > ${prodigal\_ids\_op\_folder}/${sample\_name}.faa

rm ${op\_dir}/temp\_${sample\_name}\_ids.txt

rm ${op\_dir}/temp\_${sample\_name}\_mutated.faa

done