# Quality control

mkdir checkm\_output

conda activate checkm

checkm -h

checkm taxonomy\_wf -h

bunzip2 mags/\*

checkm taxonomy\_wf domain Bacteria mags checkm\_output -t 4

less -S checkm\_output/storage/bin\_stats\_ext.tsv

conda deactivate

# Taxonomic assignment

#cd to the data directory

mkdir phylophlan\_output

conda activate ppa

phylophlan\_metagenomic -h

phylophlan\_metagenomic --database\_list

phylophlan\_metagenomic -i mags -o phylophlan\_output/ppa\_m --nproc 4 -n 1 -d CMG2425 --database\_folder ppa\_db --verbose

less -S phylophlan\_output/ppa\_m.tsv

conda deactivate

# Genome annotation

#!/bin/bash

CONDA\_BASE=$(conda info --base)

source $CONDA\_BASE/etc/profile.d/conda.sh

conda activate prokka

prokka -h

mkdir prokka\_output

for f in mags/\*; do

mag=$(basename $f .fna)

echo $mag

echo $f

mkdir prokka\_output/${mag}

prokka mags/${mag}.fna --outdir prokka\_output/${mag} --prefix ${mag} –compliant

done

conda deactivate

#To check some info

cat prokka\_output/${mag}/${mag}.txt

cat prokka\_output/${mag}/${mag}.tsv | grep 'hypothetical'

cat prokka\_output/${mag}/${mag}.tsv | grep -v 'hypothetical' | grep 'CDS' | wc -l

# Pangenome Analysis

conda activate roary

roary -h

roary prokka\_output/\*/\*.gff -f roary\_output -i 95 -cd 90 -p 4

cd roary\_output conda deactivate && conda activate roary\_plots

curl https://raw.githubusercontent.com/sanger-pathogens/Roary/master/bin/create\_pan\_genome\_plots.R \ -o create\_pan\_genome\_plots.R

chmod +x create\_pan\_genome\_plots.R

conda deactivate && conda activate roary\_plots

Rscript create\_pan\_genome\_plots.R

curl https://raw.githubusercontent.com/sanger-pathogens/Roary/master/contrib/roary\_plots/roary\_plots.py \ -o roary\_plots.py

python roary\_plots.py -h

python roary\_plots.py accessory\_binary\_genes.fa.newick gene\_presence\_absence.csv

# Phylogenetic analysis

conda activate roary

roary -h

roary prokka\_output/\*/\*.gff -f roary\_output\_w\_aln -cd 90 -p 4 -e -n

FastTreeMP -pseudo -spr 4 -mlacc 2 -slownni -fastest -no2nd -mlnni 4 -gtr -nt -out roary\_output\_w\_aln/core\_gene\_phylogeny.nwk roary\_output\_w\_aln/core\_gene\_alignment.aln

conda deactivate