Command overview

# Tool linked commands

## CheckM

checkm taxonomy\_wf genus Campylobacter -x .fasta /home/dvannassauw/QUADRAM\_FASTA\_INPUT /home/dvannassauw/CheckM\_QUADRAM > /home/dvannassauw/CheckM\_QUADRAM/checkmoutput.tsv

## Blast

Consisted of two steps:

1. Create a database of the blast fragments
2. Nucleotide blast against the databases
3. makeblastdb -in MapA\_gene\_sequence.fasta -parse\_seqids -dbtype nucl -out MapA\_gene\_blastdb

makeblastdb -in CeuE\_gene\_sequence.fasta -parse\_seqids -dbtype nucl -out CeuE\_gene\_blastdb

1. find \* | grep fasta$ | while read fasta ; do blastn -task blastn -query $fasta -db ../Blast/CeuE\_gene\_blastdb -outfmt 6 | while read hit ; do echo $fasta $hit ; done ; done > ../Blast/Blast\_output/CeuE\_gene\_quadram/Output\_CeuE\_gene.tsv

find \* | grep fasta$ | while read fasta ; do blastn -task blastn -query $fasta -db ../Blast/MapA\_gene\_blastdb -outfmt 6 | while read hit ; do echo $fasta $hit ; done ; done > ../Blast/Blast\_output/MapA\_gene\_quadram/Output\_MapA\_gene.tsv

## Parsnp

**FOR COLI**

parsnp -g ~/ParSNP/GCF\_002024185.1\_ASM202418v1\_genomic.gbff -d ../ -c -p 16 -o ./

**FOR JEJUNI**

parsnp -g ~/ParSNP/C\_jejuniGCF\_000009085.1\_ASM908v1\_genomic.gbff -d ~/DEPICT/Clustered\_DEPICT/Jejuni\_2/Jejuni\_2-4 -c -p 16 -o ~/DEPICT/Clustered\_DEPICT/Jejuni\_2/Jejuni\_2-4/ParSNP/ &

**VISUALISATION**

gingr Output\_Quadram/parsnp.ggr

figtree parsnp.tree

## Harvesttools

harvesttools -i parsnp.ggr -M parsnp.harvesttools.fasta

## Gubbins

run\_gubbins2.py ../parsnp\_without\_ref.fasta --tree\_builder fasttree --threads 16 &

**VISUALISATION**

gubbins\_drawer -t ./parsnp\_without\_ref.final\_tree.tre -o recombination\_Coli1-1.pdf ./parsnp\_without\_ref.recombination\_predictions.embl

Dendroscope parsnp.harvesttools.node\_labelled.final\_tree.tre

**RECOMBINATION STATISTICS**

seqstat parsnp.harvesttools.filtered\_polymorphic\_sites.fasta

## Masking

This part is divided in multiple steps:

1. Use Excel to retrieve column 4 and 5 from the recombination.predictions.gff file
2. Change the two columns into a comma separated file, while the areas are written as 2 minus separated values in one row
   1. sed 's/\t/-/' Jejuni\_2-4\_column\_4\_5.txt > minus-seperated\_columns4\_5.txt
   2. cat minus-seperated\_columns4\_5.txt | sed -n -e 'H;${x;s/\n/,/g;s/^,//;p;}' > comma\_minus\_separated\_C4\_5.txt
3. Addition of the recombination areas to the Masking script

#!/bin/bash

ls \*.fasta | while read file ; do echo "maskseq -sequence $file -regions 829494-830767,705976-706974,299657-301010,708818-710675,854694-855522….**etcetera** --outseq masked.$file -maskchar 'N'" ; done

1. Split all the contigs in the parsnp.harvestools file in separate files

seqretsplit parsnp.harvesttools.fasta

1. Run script

Chmod +x Masking$

./Masking\_C3-3.sh | parallel -j 16 &

1. Concatenation of masked files back to parsnp.harvestool construction and sequence statistics

* cat masked.\* > core\_genome\_without\_recombination.fasta
* seqtk comp C2\_core\_genome\_without\_recombination.fasta > C2\_seqtk\_statistics.txt

## BEAST

After the XML files were created in BEAUti and ran in BEAST using their interface. Tracer was used for visualisation, TreeAnnotator for creating the target tree and Figtree for visualisation of the tree.

* tracer Coli\_1\_\_10x.log.txt
* ~/bin/BEASTv1.10.4/bin/treeannotator -burnin 1000000 -heights keep Coli\_1\_\_10x.trees.txt > Coli\_1\_\_10x.mcc.tree
* figtree Coli\_1\_\_10x.mcc.tree

## Plasmid blast against close related strains

makeblastdb -in 103292-003-009.fasta -parse\_seqids -dbtype nucl -out 103292-003-009\_blastdb

find 103292-003-009/\* | grep fasta$ | while read fasta ; do blastn -task blastn -query $fasta -db 103292-003-009/103292-003-009\_blastdb -outfmt "6 std qcovs" -dust no -soft\_masking false | while read hit ; do echo $fasta $hit ; done ; done > 103292-003-009/Output\_103292-003-009.tsv

# Non-tool linked commands

## Removal of strains in alignment file

**FOR COLI**

cat parsnp.harvesttools.fasta | ~/bin/fasta2tab.sh | grep -v GCF\_0020 | ~/bin/tab2fasta.sh > ./parsnp\_without\_ref.fasta

**FOR JEJUNI**

cat parsnp.harvesttools.fasta | ~/bin/fasta2tab.sh | grep -v C\_jejuniGCF | ~/bin/tab2fasta.sh > ./parsnp\_without\_ref.fasta

## Add .fasta to every value in column 1 in a textfile

awk -F '\t' '{ $1 = $1".fasta" }1' < Filenames\_Data\_extracted.txt > fastafilenames\_data\_extracted.txt

## Add something to the end of each line in a text file

sed 's/$/<@string>/' txt.txt > txt2.txt

## Move files which filename occurs in textfile and move it to desired directory

for i in ~/DEPICT/Clustered\_DEPICT/Coli\_1/\*.fasta; do

if grep -Fq -f ~/DEPICT/Clustered\_DEPICT/Coli\_1/Coli\_1-1.txt <<< "$i"; then mv -t ~/DEPICT/Clustered\_DEPICT/Coli\_1/Coli\_1-1 "$i" ; fi ; done

## Create a directory for every file in a directory and copy the file in it.

for x in ./\*.txt; do

mkdir "${x%.\*}" && mv "$x" "${x%.\*}" ; done