Basic Assembly QC Exercises

# Quast

Run Quast on spades assemblies found in the folder precomputed\_data/spades\_assemblies/spades\_renamed/

Make sure to include all contigs (by default Quast filters out contigs <500 bp)

Use environment env\_BTG\_Quast

Use quast -h for input options

Open the report.html file produced by Quast.

Have a look at the top table with basic assembly stats. Which assemblies would you flag for potential contamination or poor quality?

Also have a look at the other output files produced by Quast to get an idea of what’s in them.

# Kraken2

Use environment BTG\_kraken2

Generate a kraken classification file and a kraken report file for one of the assemblies in

/home/gebt/BTG\_2024/precomputed\_data/spades\_assemblies/spades\_filtered/

The folder

/home/gebt/BTG\_2024/precomputed\_data/precomputed\_data/kraken\_spades/ contains kraken output and kraken reports for all the Listeria assemblies

Have a look at the kraken.report output from an assembly you believe to be a good, clean assembly.
Then have a look at the reports from assemblies you flagged in the quast exercise. Can you find an explanation for the poor assembly quality in the kraken results?

CHEAT SHEET

# Quast

conda activate BTG\_env\_quast

quast -h

quast -m 0 -o output\_dir \*.fasta

# Kraken2

To run kraken on reads use

kraken2 --db minikraken2\_v2\_8GB\_201904\_UPDATE --gzip-compressed --paired --output [output.kraken.txt] --report [output.kraken.report.txt] [input.fastq.gz]

To run kraken on an assembly

kraken2 --db minikraken2\_v2\_8GB\_201904\_UPDATE --output test.kraken --report [input.fasta]

# In silico MLST (Multilocus sequence typing)

Use environment BTG\_env\_mlst

To run MLST on assemblies simply use

mlst /path/to/assemblies/\*.fasta > mlst.txt

# Bakta

The newest version of Bakta is installed in environment BTG\_env\_bakta\_1.9.2

However…

One of Baktas dependencies has a bug in it’s most recent version.
After activating the environment, downgrade Diamond to version 2.1.8 using

micromamba install bioconda::diamond=2.1.8

Now you’re ready to run bakta. Look at the options with

bakta -h

Try to run bakta on one of the assemblies in precomputed\_data/spades\_assemblies/spades\_filtered

Hints

Kraken2 is installed in environment BTG\_kraken2

## Setup a kraken2 database

In case you need to set up a kraken2 database

You can find a list of various indexed kraken2 databases at <https://benlangmead.github.io/aws-indexes/k2>

For now, we will work with the minikraken database (a bit outdated, but we want a small DB that can be set up in a reasonable timeframe:

Use:
wget <https://genome-idx.s3.amazonaws.com/kraken/minikraken2_v2_8GB_201904.tgz>

Will take ~15 minutes

Unzip the tar-ball

tar -xvf minikraken2\_v2\_8GB\_201904.tgz

will take a few minutes

SRR27240828 Contaminated with L. welshiment

SRR27240830 Contaminated (Shigella ~8%)

SRR27240832 Mixed strain, ST224 / ST7 / STEC

SRR27240829 Low coverage

SRR27240831 High quality R1 but low quality R2