Running the pipeline:

1. For FASTQ or FASTQ.gz input files (other kinds of compression are not allowed) (average processing time 1.5hours):

At the command line type:

perl /groups/murray/run\_pipeline/analyseNGS.pl <paired-end indicator> <paired-end extension> <path>

**<paired-end indicator>**: takes on the values 1 or 0. 1: fastq files are paired-end i.e. there are two files for each strain

**>RP: Ask user if it is pair-end?**

**0: single-end**

**1: pair-end**

**Yes-> Ask user for extension. Example “\_R”**

**<paired-end extension>**: e.g. if the paired-end extension is **\_R** the fastq files will be: **teststrain\_R1.fastq & teststrain\_R2.fastq**, if extension is **.** then the fastq files are **teststrain.1.fastq & teststarin.2.fastq**, and so forth.

**<path>**: the folder where the user fastq files were downloaded to

Note all the output will appear in a sub directory called <path>/output/

The TBpredict.R output file is currently stored in this directory and called result.json

The error/log file (has extension error) will be stored in the <path>

1. For VCF input files (processing time 2-3minutes)

At the command line type:

perl /groups/murray/run\_pipeline/analyseVCF.pl <path>

**<path>**: the folder where the user fastq files were downloaded to

Note all the output will appear in a sub directory called <path>/output/ including the result.json file

The error/log file (has extension error) will be stored in the <path>

Failure text 1🡪 "The processing of your file has failed here is out best guess as to why?"

**Mapping** failed if stampy gives an error and/or the vcf and var files contain only headers

if mapping failed

Failure text 2🡪 "The read mapping failed. If your fastq files are paired please verify that they are named following one of these two formats

(1) id\_R1.fastq & id\_R2.fastq

(2) id.1.fastq & id.2.fastq

If the fastq files are single ended please make sure the name contains no periods except at the extension, i.e. id.fastq"

**Variant calling** failed if platypus gives an error and/or the vcf and var files contain only headers

if the variant calling failed

Failure text 2🡪 "The variant calling failed"

**Annotation** failed if flatAnnotatorVCF.pl gives an error and/or the var file contains only headers: as this has been tested many times, this can only happen if the vcf formatting changed (ie if platypus code was changed)

if annotation failed

Failure text 2🡪 "The annotation failed"

**Prediction** failed if TBpredict.R gives an error and/or the result.json file is empty

if prediction failed

Failure text 2🡪 "The resistance prediction failed"

Failure text 3🡪 "We are sorry for any inconvenience, we are looking into the error and will email you about next steps shortly"