

INPUT FASTQ Data

QC and Host Filtering

Resistome Analysis

Alignment to MEGARes

Resistome  
Analyzer

Rarefaction  
Analyzer

SNPfinder/  
freebayes

\*Count de-duplication

OUTPUT  
Resistome

OUTPUT  
Rarefaction

OUTPUT  
SNPs

\*extract genes requiring  
"SNPConfirmation"

Resistance  
Gene  
Identifier

"SNP confirmed"  
OUTPUT resistome

Microbiome Analysis

*k*-mer match to  
krakenDB

Taxonomic naming

OUTPUT Microbiome

INPUT FASTQ Data

Base programs:  
Nextflow C and C++  
Python  
Numpy

## QC and Host Filtering

Remove Low-Quality Reads (Trimmomatic and Trimmomatic\_stats.py) - - - - - *RunQC*

Identify host reads (BWA) - - - - - *BuildHostIndex*  
- - - - - *AlignReadsToHost*

Remove host reads (Samtools and samtoolsidx\_stats.py) - - - - - *RemoveHostDNA*

Convert file types (Bedtools) - - - - - *NonHostReads*

## Resistome Analysis

Identify AMR reads (BWA)  
- - - - - *AlignToAMR*

Apply gene fraction (ResistomeAnalyzer)  
- - - - - *RunResistome*

Build count matrix (amr\_long\_to\_wide.py)  
- - - - - *ResistomeResults*

Extract reads mapped to SNP genes (AWK)  
- - - - - *ExtractSNP*

Confirm SNPs required for resistance (RGI)  
- - - - - *RunRGI*

Combine AMR counts (amr\_long\_to\_wide.py)  
- - - - - *Confirmed\_ResistomeResults*

Rarefy (RarefactionAnalyzer)  
- - - - - *RunRarefaction*

Identify SNPs (SNPFinder, Freebayes)  
- - - - - *RunSNPFinder*,  
*RunFreebayes*

OUTPUT resistome count matrix

## Microbiome Analysis

Build kraken *k*-mer DB  
(download is already built)

Build *k*-mer index for each sample, match to DB (kraken)  
- - - - - *RunKraken*

Build count matrix  
(kraken\_long\_to\_wide.py)  
- - - - - *KrakenResults*  
*FilteredKrakenResults*

OUTPUT microbiome count matrix

Underline indicates a program/script needed to run Nextflow pipeline  
*Italics indicates an output directory from Nextflow pipeline. Resistome analysis performs the same steps for de-duplicated alignments.*