

E. coli Subtyping Exercise

Introduction

In this module we would be using two subtyping approaches to subtype the *Escherichia coli* strains. We would first perform MLST analysis using the command line [MLST tool](#) and then serotype the strains using the [ECTyper tool](#).

MLST analysis

- Open the terminal on your virtual machine and print your current working directory
`pwd`
- Make a directory called “subtyping” and navigate into the directory
`mkdir subtyping`
`cd subtyping`
- Make a directory called “fastas”
`mkdir fastas`
- Link the fasta input files from the Phylogeny directory into it.
`ln -s ~/Phylogeny/*.fasta fastas/`
- Create another directory called “mlst” within the subtyping directory and navigate into that directory
`mkdir mlst`
`cd mlst`
- Let’s ensure the mlst tool is properly installed and check which schemes are available
`mlst --list`
- Now run the analysis
`mlst --legacy --nopath --quiet --scheme ecoli_achtman_4`
`../fastas/*.fasta > mlst_output.tsv`
- View the output
`less mlst_output.tsv`

Serotyping analysis

- Navigate back to the “subtyping” directory,
- Create a directory called “serotyping”, and navigate into the “serotyping” directory
- Check that the ECTyper tool is installed

```
ectyper --help
```

- Now run the analysis

```
ectyper -i ../fastas/ -o ectyper_output --verify
```

- View the output

```
cut -f1-8 ectyper_output/output.tsv | column -ts $'\t'
```

Questions

1. How many different *E. coli* sequence types were detected?
2. What are the serotypes of the *E. coli* strains?
3. What do we know about these STs and serotypes? Hint: Pathogenicity? Epidemiology?
4. From the available data (serotype and ST), can we tell if there is an outbreak? If yes, which STs/serotypes are part of the outbreak? If no, why?