**Phylogeny Exercise**

**Exercise Files**

* Sequenced genomes of seven (7) bacterial isolates have been provided to you as fasta (**fastas** folder) files. Please visit the Google drive link and **download these files to your local machine**.
* A metadata sheet containing epidemiological information of all the seven (7) strains has also been provided in the Google drive folder. **Download the ‘metadata.xlsx’ file**.
* A reference genome has also been selected for you for the exercise. **Download the file named ‘reference.fas**’.
* Replace FigTree by iTOL

# Software/Tools

You would need to use the following software during the exercise:

* CSIPhylogeny (https://cge.food.dtu.dk/services/CSIPhylogeny/): The CSIPhylogeny workflow is a free tool developed and maintained by the Center for Genomic Epidemiology in Denmark is used to construct phylogenies from whole genome raw reads (fastq format) or assemblies (fasta format). This is a web-based tool and does not need to be installed locally.
* iTOL (Interactive tree of life)
* Microreact (https://microreact.org): Microreact is a useful tool for the visualisation of constructed phylogenies with contextual metadata. It is also a web-based tool and **does not** need to be installed locally.

# Exercise Instructions Phylogeny construction

* Go to the CSIPhylogeny webpage at https://cge.food.dtu.dk/services/CSIPhylogeny/
* In the section ‘**Upload reference genome (fasta format)**’, click on **Choose file** and navigate to the folder where the reference genome file is stored on your local machine. Click ‘**Open**’. Afterwards, check/mark the box below it to **Include reference in final phylogeny** (this is optional).
* Leave all the other parameters as default and scroll to the section ‘**Upload read files and/or assembled genomes (fasta or fastq format)**’. We would be using the fasta assemblies you have already downloaded from the shared Google drive for this exercise.
* Click on ‘**Isolate File**’ and navigate to the folder where the fasta files are stored on your local machine. Hold **Control** on your keyboard (or **Command** for Mac users) and select all 7 fasta files. Click **Open**.
* Scroll down and click ‘**Upload**’. **Note:** This may take a while, depending on your internet speed. **Please do not refresh the page or you would have to start over!**
* Once it is done uploading, the page would automatically refresh and redirect to the jobs page.
* On the Jobs page, enter your email address and click submit. This allows you to be notified by email when the run is finished.

# Accessing and manipulating the tree file

* Follow interactive use of Microreact
* **Visualisation (Microreact)**
* Visit https://microreact.org
* Click on the **Instructions** tab to see specific information on how to prepare your metadata file (you must have downloaded this already) and your tree file for upload and visualization.

# Please note that you should add other information gleaned from previous exercises (including MLST, species identification, AMR prediction, etc.) to your metadata file to be able to make concrete and informed inferences.

**Exercise Question**

Narrate an epidemiological story for the strains in your collection.

**Hints**: Was there an outbreak going on at the time of isolation? What strains were involved, if any? What was the source? What pieces of information on the visualization made you infer this?