PHA4GE-GHRU BIOINFORMATICS WORKSHOP – MODULE 1

Phylogeny Exercise

Please ensure you have downloaded and read the '**General Instructions on Exercises.pdf**' file from the shared Google drive folder before starting this exercise.

Exercise Files

- Sequenced genomes of seven (7) bacterial isolates have been provided to you as both fasta (**fasta** folder) and paired-end fastq files (**fastq** folder'). Please visit the Google drive link and **download these files to your local machine**. (https://drive.google.com/drive/folders/1Lay00lIavQjOPzohroBRxCx9wvPQTWX3?usp=sharing).
- A metadata sheet containing epidemiological information of all the seven (7) strains has also been provided in the Google drive folder. See the **metadata** folder and **download the 'metadata.xlsx' file** as well.
- A reference genome has also been selected for you for the exercise. This is in the '**references**' folder. Please download this file as well.
- The FigTree software files "**FigTree.v1.4.4.zip**" (for Windows users) and **FigTree.v1.4.4.dmg** (for Mac OSX users) have been provided for you in the Google drive. Please download the appropriate file for your operating system.

Software/Tools

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

- CSIPhylogeny (https://cge.cbs.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.
- FigTree (https://github.com/rambaut/figtree/releases): FigTree would be used for visualization and manipulation of your tree output from CSIPhylogeny. This tool needs to be installed on your local machine. **Please see the installation instructions in the Installation section of this document**.
- 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.

Installation

FigTree requires Java 1.6 and above to be installed, which you should have installed as part of the QC exercise on Day 2. If not, visit https://java.com/en/download to download and install Java.

Windows Users

- Right-click on the "**FigTree.v1.4.4.zip**" file and click "**Extract All**", then click "**Extract**".

- Open the extracted folder and open the file named "**FigTree.v1.4.4**". That's it! Nothing else needs to be done.

Mac Users

- Simply double-click on the **FigTree.v1.4.4.dmg** file once downloaded and follow the prompts.

Exercise Instructions

Phylogeny construction

- Go to the CSIPhylogeny webpage at https://cge.cbs.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 submit so you can get notified by email when the run is finished.

Accessing and manipulating the tree file

- Once the job is finished (may take some time depending on the queue), you would get an email notification with a link to the results page.
- Scroll to the part of the page that says '**Download phylogeny as:**' and click on **Newick**. This would download the tree file in Newick format.
- Once downloaded, open the FigTree software you already installed. On the FigTree interface, click on **File > Open** and select the file you just downloaded. Click **Open.**
- On the left side pane, select **Trees** > **Root tree**, click the dropdown button and select **Midpoint**. Next, check/mark the **Order nodes** box. You should have a fairly nice-looking tree now.
- To save the tree with this view, select File > Export Trees and click the dropdown to change the format from **NEXUS** to **Newick**. Check/mark the box that says "**save as currently displayed**" and click **OK**. Type in the desired name for your tree file in the **Save As** box and add the extension "**.nwk**" at the end of the name (e.g.,

mytreefile.nwk). **Note: it is good practice to avoid using spaces, upper case letters and special characters like @ or : in your file names.**

- Select the destination folder and click **Save**.

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 visualisation.
- **Please note that you would need to add other information gleaned from previous exercises (including MLST, species identification, as well as virulence, plasmid and AMR prediction) 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 visualisation made you infer this?