How to run the Bioinformatics Pipeline

### Summary

* Download data from operator or SEPA lab
* Run Qiime2 based taxonomic assignment script
* Assess quality, export and format the taxa data
* Calculate the IQI score based on the bacteria taxa
* Upload the bacteria based IQI scores to the Access database
* Calculate the Area footprint in the Spotfire area calculator tool
* Report results to Compliance team

### Installation

To run the [qiime2](https://qiime2.org/) based taxa assignment, [install](https://docs.qiime2.org/2022.8/install/) qiime2 and conda software libraries. These run on MacOS or Linux. For Windows users, recommend using Windows Subsystem for Linux (WSL).

### Saving files

The Operator or SEPA lab will provide a link to where sequenced data is stored.

1. Check these files will be in .fastq files.
2. Save the files to your One Drive within the ‘dna-data’ folder.
3. Name the folder with the unique survey name.

### Run Assignment

1. On Windows, open WSL ubuntu 20.04, and ‘mount’ (mnt), the Windows folder containing the sequence data files. For example:

* cd /mnt/c/Users/Tim.Foster/OneDrive\ -\ Scottish\ Environment\ Protection\ Agency/Reports/Fish\ Farm/operating\_iqi\_surveys/ScreeningTool\_SequenceData/dna-data/
* NOTE: For reference the `qiime2-taxa-assign.txt` file can be found here:
* [aquaman/inst/extdat/qiime2-bash-script at main · aquaMetrics/aquaman (github.com)](https://github.com/aquaMetrics/aquaman/tree/main/inst/extdat/qiime2-bash-script)
* This file is also stored in the /dna-files folder.

1. Activate conda python environment required for Qiime2.

* conda activate qiime2-2022.2

1. Run the bash script and provide a relative path to sequenced files.

* bash /qiime2-taxa-assign.txt directory/containing/files

1. This produces a number of output files which will be stored in the sub-directory with dna-files.

* NOTE: If changes made to qiime2-taxa-assign.txt file, then may require cleaning the file of carriage returns introduced by editing.
* tr -d "/r" < qiime2-taxa–assign.txt > qiime2-taxa-assign-cleaned.txt

### Quality Control

Refer to the DNA SOP for current requirements on number of reads required. Assess the ‘…demuxed\_paired\_end.qzv’ files using [Qiime2 Viewer](https://view.qiime2.cn/) where the number of reads can be verified.

The classify-consensus-blast Qiime2 command used in the script assign taxa uses the default settings to select taxa matches. The fixed percent identity to a taxa is 0.80.

### Export Taxa

Taxa information must be exported to .CSV for later processing in R.

1. Drag and drop filenames containing ‘…S16….qzv’ to [Qiime2 Viewer](https://view.qiime2.cn/).
2. Select Taxonomic Level 7 from drop-down and click the ‘Download as CSV’ button.
3. Save files along side the .qzv files in the relevant dna-files/ directory.

### Format Assigned Taxa

Prior to running the IQI assessment of the taxonomic data, the .CSV data is converted to the correct input format.

1. Open RStudio

library(aquaman)  
taxa <- aquaman:::import\_qiime2(path = "C:\\Users\\Tim.Foster\\OneDrive - Scottish Environment Protection Agency\\Reports\\Fish Farm\\operating\_iqi\_surveys\\ScreeningTool\_SequenceData\\dna-data\\K\\taxa.csv")

### Calculate IQI

USe the iqi() function to run the IQI model and calculate IQI.

iqi\_score <- iqi(taxa)

(Check if set.seed will influence this?) May need to store output as static file. Not reproducible.

### Format for Access

Outputs may need to be format to fit with existing internal Access database. Copy output into Access database:

* C:\Users\Tim.Foster\OneDrive - Scottish Environment Protection Agency\FishFarmDatabaseMaster\MPFFSelfMonDatabase.accdb
* This requires Access software limited to some virtual desktop machines.

### Run Spotfire Area Tool

Calculate area in Spotfire tool (maybe need data refresh overnight to get latest data from Access)?

### Compliance

Submit results to Compliance team via email to complete compliance review.