fragmentR *Clostridium difficile* Tutorial 1.0.0

**Running from the command line:**

What you will need:

R version 4 +

One does not need any R experience to run fragmentR! It can be run directly from the command line. All you need is to download the *C. difficile* database and the FSA files from the walk lab website (https://thewalklab.com/tools). We find it is best to create a folder specifically for running fragmentR. Place the “F-RibotypingFiles” in the directory you are running fragmentR from. For example, FP503 is a ribotype, within the folder 7958.D02.FP503.fsa, and 7959.D02.FP503.fsa are the representative chromatogram files within that ribotype.

Table

Description automatically generated with medium confidence

From the terminal navigate to the directory you plan on running fragmentR from. To download the required packages run the “setup\_fragmentR.R” script from the terminal. This will also create a “Files\_to\_analyze” folder. A file summarizing the database is also needed for the F-RibotypingFiles database use “Cdiff\_DB\_list.5.4.rds”, this file must be in same directory as F-RibotypingFiles.

On a mac:

Rscript Desktop/Run\_Fragment\_Analysis/setup\_fragmentR.R

On windows:

Desktop/Run\_Fragment\_Analysis/setup\_fragmentR.R

Drag and drop files you wish to match to a ribotyping database into the “Files\_to\_analyze” folder.

On a mac:

Rscript Desktop/Run\_Fragment\_Analysis/Call\_FSA.R

On windows:

Desktop/Run\_Fragment\_Analysis/Call\_FSA.R

The results will populate a folder named “Results YYYY-MM-DD Hour/Min/Sec”. As the script runs it will populate the results folder with jpegs of chromatograms and plots comparing the query and best hits in the database. At the end a SUMMARY.csv table is also produced, summarizing all the files that were in “file\_to\_analyze” folder.

Graphical user interface, text, application

Description automatically generated

Sample chromatogram (chrom\_ 002-g01-34814.jpeg) visualizes the raw data from the machine and the peaks that were called in the query and ladder channels. All channels are plotted. The cutoff is only revised for the ladder channel.

chrom\_ 002-g01-34814.jpeg

Chart, histogram

Description automatically generated

Sample hit plot shows (hit\_ 002-g01-34814.jpeg) the normalized peak intensity plotted against fragment size (base pairs) of the query and the closest match in the database. Black is the query and red represents the hit in the database. The summary of each fsa files is calculated by measuring the Bray-Curtis (BC) distance between each entry in the database and the query. The lower the BC distance between a query and hit the better match it is. chromatograms are classified as either a good match (<0.10 distance), questionable match (0.10 - 0.20 distance), or poor match (>.20 distance). If a match is questionable, consider visually inspecting the chromatograms.

hit\_ 002-g01-34814.jpeg

Chart, histogram

Description automatically generated