fragmentR *Clostridium difficile* Tutorial 1.0.1

**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. 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.

Included in the package is the database F-Ribotyping\_Reps.lite.15.zip”, that query FSAs are going to be match against. It must be uncompressed before it is used. There is also an accompanying file; “Cdiff\_DB\_lite.15.RDS”, summarizing the database is also needed.

On a mac:

Rscript setup\_fragmentR.R

On windows:

FragmentR-main/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/ FragmentR-main/Call\_FSA.R

On windows:

Desktop/FragmentR-main/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