# CdiffFragR Manual

*C. difficile* Fluorescent PCR Ribotyping v 1.0.1

CdiffFragR was developed to match query *Clostridioides difficile* chromatograms (.fsa file format) generated on capillary electrophoresis sequencers to a database of know *C. difficile* ribotypes. CdiffFragR is an R tool for DNA fragment analysis developed and maintained by the Walk Lab at MSU Bozeman. CdiffFragR is compatible with both Mac OS X and Windows 10.

## DOWNLOADING SOFTWARE

### 1.) DOWNLOAD R

If you do not already have R installed on your computer, go to: https://cran.r-project.org/bin/windows/base/
and click on Download R 4.1.1 for Windows. Save the file (R-4.1.1-win.exe) to your desktop and run the installation.

If you already have R on your computer, please upgrade to version 4.0 or more recent.

### 2.) DOWNLOAD RTOOLS

If you do not already have Rtools installed on your computer, go to: https://cran.r-project.org/bin/windows/Rtools/
and click on the link for Windows 64-bit: rtools40v2-x86_64.exe. Save the file (rtools40v2x86_64.exe) to your desktop and run the installation.

If you already have Rtools, please upgrade to version 4.0 or more recent.

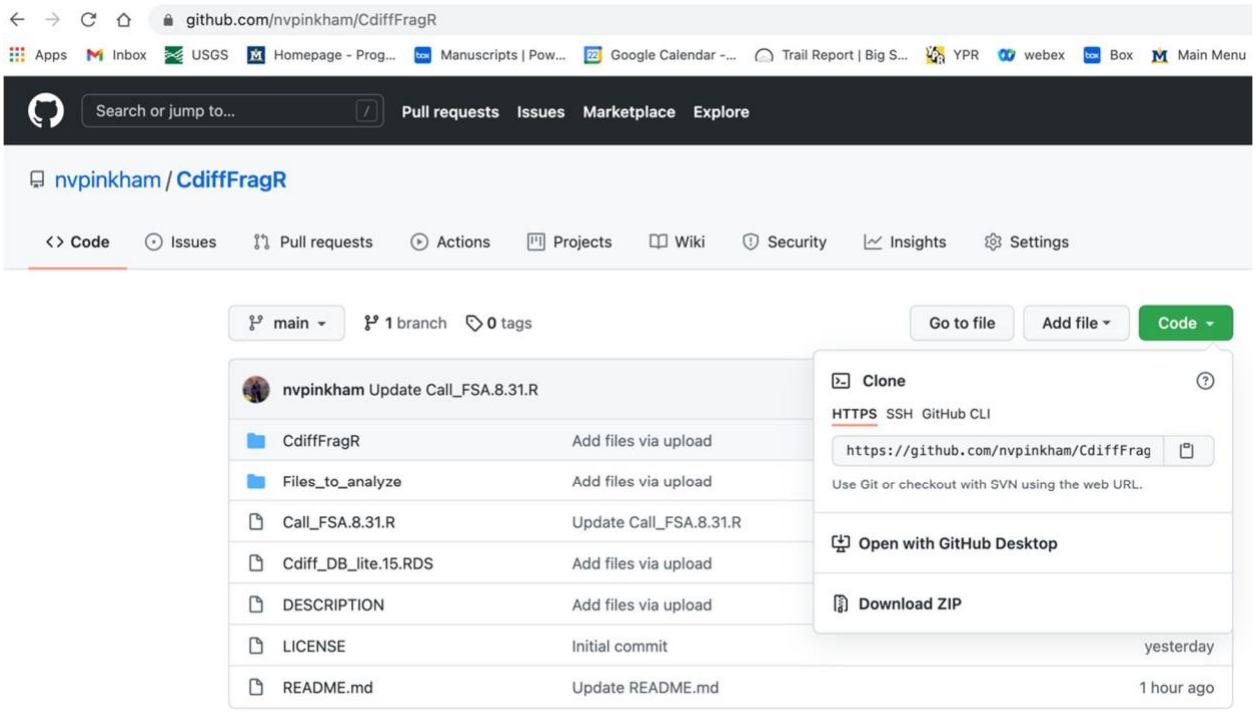
### 3.) DOWNLOAD RSTUDIO

If you do not already have RStudio on your computer, go to: https://www.rstudio.com/products/rstudio/download/
and click the DOWNLOAD button under RStudio Desktop. Save the file (RStudio-1.4.1717.exe) to your desktop and run the installation.

If you already have RStudio, please upgrade to 1.4 or more recent.

### 4.) DOWNLOAD CdiffFragR

Download CdiffFragR-main.zip from github.com/nvpinkham/CdiffFragR by clicking the Code button and the Download Zip option. Save this to your computer.




Following download, extract the contents of the zip file to a directory on your computer.

Windows – Right click on CdiffFragR-main.zip and select Extract All… from the dropdown window. When prompted, click Browse button and select Desktop.

Mac – Uncompress "CdiffFragR-main.zip"

**RUNNING CdiffFragR FOR THE FIRST TIME**

We set up the download process to run a test on previously generated datafiles before running new datafiles. If you look in the directory "Files_to_analyze", you will see 4 .fsa files. The following is a test run for these files:

Open "CdiffFragR-main.Rproj" with Rstudio.

Windows – Right click on the file, select "Open with…", and select Rstudio
Mac – Single click on the file, select "Open With", and Rstudio

Once Rstudio has launched, open the script "Call_FSA.12.06.R" by going to File>Open File…

Once the script is loaded, highlight all the text and click the Run button in the upper right-hand corner of the script window.

After peaks are called and chromatograms are processed, each .fsa file will be moved from the "Files_to_analyze" directory to a new directory called "Files_analyzed". Results will populate a directory called "Results_YYYY-MM-DD_Hour_Min_Sec". The script generates jpeg summaries of chromatograms (filenames begin with "chrom_") and overlay plots comparing each query against the closest (best) match to the database (filenames begin with "hits_"). A "SUMMARY.csv" table is also produced, summarizing the results as follows.

| ***Column Name*** | ***Definition*** |
|---|---|
| query_file | Directory and filename of query file |
| Confidence | Computationally defined match |
| Hit_1 | Directory and filename of the closest match to the database |
| ribo_1 | Ribotype of closest database match |
| Dist_1 | Bray-Curtis dissimilarity between query and database chromatogram |
| hit_1_jpeg | Directory and filename of query-match overlay plot |
| chromatogram_jpeg | Directory and filename of chromatogram |

**INTERPRETING RESULTS**

1) Open the SUMMARY.csv file and look at the Dist_1 column. Numbers in this column refer to the Bray-Curtis dissimilarity between the query and hit chromatograms. The value ranges between 0 and 1, and the closer the number is to 0, the better the match. Based on tens of thousands of analyzed .fsa files, the "Confidence" of matches fall into one of three categories according to the Bray-Curtis distance as shown in the table below.

| **Dist_1** | **Confidence** |
|---|---|
| <0.10 | "good match" |
| 0.10 – 0.20 | "questionable match" |
| >0.20 | "poor match" |

2) Our interpretation:
   Good match – This is a solid match to the database with a very low chance of a false positive.

Poor match – This is not a match to the database. Take a look at the chromatogram and hits images. Either 1) the chromatogram looks fine, and the ribotype is not yet in the Walk Lab database, or 2) something went wrong with the amplification or capillary electrophoresis and the sample should be re-analyzed.

Questionable match – This may or may not be a match to the database. Take a look at the hit image and see whether the peaks line up (peaks within 1-2 base pairs should be considered the same peaks).

Two examples:
From the SUMMARY.csv file, you will see that one of the queries ("i24_20210723163608_20210723131726_Fragment_C8_06.fsa") had a "questionable match". If you open up the corresponding chromatogram ("chrom_i24_20210723163608_20210723131726_Fragment_C8_06.jpeg") you will see that the peaks are quite small compared to those of the size standard. This isolate should be re-analyzed.

From the SUMMARY.csv file, you will see that one of the queries ("i18_20210723155914_20210723131726_Fragment_C2_05.fsa") Had a "poor match". If you open up the corresponding chromatogram ("chrom_i18_20210723155914_20210723131726_Fragment_C2_05.jpeg") you will see that the peaks are nice a tall compared to those of the size standard, indicating that there was plenty of signal. Next, if you open up the corresponding query-match overlay plot ("hit_i18_20210723155914_20210723131726_Fragment_C2_05.jpeg") you will see that the query contains at least two peaks that the best match to the database does not and is missing at least two peaks that the best match to the database has. This is likely a new ribotype.

New ribotypes:
If you believe you have identified a new ribotype, it would be best to re-analyze and confirm the previous result. Next, please contact the Walk Lab to see if it can be added to the database. We would love to continue developing this resource for the scientific community and adding new ribotypes as they are observed would certainly help.

## RUNNING CdiffFragR AGAIN

Each time you run CdiffFragR, the program simply looks for files in the "Files_to_analyze" directory and generates a time-stamped results directory ("Results_YYYY-MM-

DD_Hour_Min_Sec"). To begin a new analysis, simply drop files into "Files_to_analyze" and follow the instructions above.

***From time to time, please check whether a new version of the database (F-Ribotyping_Reps.lite.15) is available. If it is, simply replace these directories to make updated ribotype calls.