MADA Data Analysis Project Proposal

Monica Chan

2021-09-17

Applying MADA learned techniques to a RT-PCR dataset to determine the effectiveness of sample enhancement strategies for environmental surface sampling of low bio-burden viruses.

# 1 Data Description

*What the data is*

Compilation of RT-PCR results for a couple of probes (N1 & N2) to detect recovered heat-inactivated SARS-CoV-2 samples that have undergone sample enhancement procedures.

*How was it collected*

Surface sampling, sample processing, RT-PCR run results, and result compilation were collected and performed by myself.

*How many observations*

The current set has 379 observations and 23 variables. Additional data will be added into this set as experiments are completed throughout the course.

*Additional information?*

Additional variables (Environmental and introduced) will be tested for effects.

# 2 Data

library(readr)  
Data<-read\_csv("../../data/raw\_data/2021.08.18\_Spike.csv")

## Rows: 379 Columns: 23

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (11): EXP, Extraction, PCR, EXP2, Tool, Method, Type, Dilution.Tube, Sam...  
## dbl (11): Dilution.Factor, Tool.Vol, Input.Vol, Vol.Adjustment, FIN.Vol, PER...  
## lgl (1): PCR.REP

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

str(Data)

## spec\_tbl\_df [379 x 23] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ EXP : chr [1:379] "SpikeEluent" "SpikeEluent" "SpikeEluent" "SpikeEluent" ...  
## $ Extraction : chr [1:379] "2021.05.10" "2021.05.10" "2021.05.10" "2021.05.10" ...  
## $ PCR : chr [1:379] "2021.05.12" "2021.05.12" "2021.05.12" "2021.05.12" ...  
## $ PCR.REP : logi [1:379] NA NA NA NA NA NA ...  
## $ EXP2 : chr [1:379] "SpikeEluent" "SpikeEluent" "SpikeEluent" "SpikeEluent" ...  
## $ Tool : chr [1:379] "Sanigen" "Sanigen" "Sanigen" "Sanigen" ...  
## $ Method : chr [1:379] "CDC" "CDC" "CDC" "CDC" ...  
## $ Type : chr [1:379] "Sample" "Sample" "Sample" "Sample" ...  
## $ Dilution.Tube : chr [1:379] "I" "I" "II" "II" ...  
## $ Dilution.Factor: num [1:379] -2 -2 -3 -3 -4 -4 -2 -2 -3 -3 ...  
## $ Tool.Vol : num [1:379] 10000 10000 10000 10000 10000 10000 10000 10000 10000 10000 ...  
## $ Input.Vol : num [1:379] 120 120 120 120 120 120 120 120 120 120 ...  
## $ Vol.Adjustment : num [1:379] 0 0 0 0 0 0 0 0 0 0 ...  
## $ FIN.Vol : num [1:379] 120 120 120 120 120 120 120 120 120 120 ...  
## $ PER.SAMP : num [1:379] 0.012 0.012 0.012 0.012 0.012 0.012 0.012 0.012 0.012 0.012 ...  
## $ IDEAL.Copies : num [1:379] 22 22 2 2 0.224 0.224 22 22 2 2 ...  
## $ Copy.Adjustment: num [1:379] 0 0 0 0 0 0 0 0 0 0 ...  
## $ REAL.Copies : num [1:379] 22 22 2 2 0.224 0.224 22 22 2 2 ...  
## $ Sample.Name : chr [1:379] "CDC 120 I" "CDC 120 I" "CDC 120 II" "CDC 120 II" ...  
## $ Target : chr [1:379] "N1 Probe" "N1 Probe" "N1 Probe" "N1 Probe" ...  
## $ Ct : chr [1:379] "32.42782593" "31.51000595" "34.5826683" "35.53193665" ...  
## $ Column3 : num [1:379] NA NA NA NA NA NA NA NA NA NA ...  
## $ Column4 : num [1:379] NA NA NA NA NA NA NA NA NA NA ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. EXP = col\_character(),  
## .. Extraction = col\_character(),  
## .. PCR = col\_character(),  
## .. PCR.REP = col\_logical(),  
## .. EXP2 = col\_character(),  
## .. Tool = col\_character(),  
## .. Method = col\_character(),  
## .. Type = col\_character(),  
## .. Dilution.Tube = col\_character(),  
## .. Dilution.Factor = col\_double(),  
## .. Tool.Vol = col\_double(),  
## .. Input.Vol = col\_double(),  
## .. Vol.Adjustment = col\_double(),  
## .. FIN.Vol = col\_double(),  
## .. PER.SAMP = col\_double(),  
## .. IDEAL.Copies = col\_double(),  
## .. Copy.Adjustment = col\_double(),  
## .. REAL.Copies = col\_double(),  
## .. Sample.Name = col\_character(),  
## .. Target = col\_character(),  
## .. Ct = col\_character(),  
## .. Column3 = col\_double(),  
## .. Column4 = col\_double()  
## .. )  
## - attr(\*, "problems")=<externalptr>

summary(Data)

## EXP Extraction PCR PCR.REP   
## Length:379 Length:379 Length:379 Mode:logical   
## Class :character Class :character Class :character NA's:379   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## EXP2 Tool Method Type   
## Length:379 Length:379 Length:379 Length:379   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## Dilution.Tube Dilution.Factor Tool.Vol Input.Vol   
## Length:379 Min. :-5.000 Min. :10000 Min. : 120   
## Class :character 1st Qu.:-4.000 1st Qu.:10000 1st Qu.: 120   
## Mode :character Median :-3.000 Median :10000 Median : 300   
## Mean :-3.103 Mean :10000 Mean :1629   
## 3rd Qu.:-2.000 3rd Qu.:10000 3rd Qu.:1000   
## Max. :-2.000 Max. :10000 Max. :8400   
##   
## Vol.Adjustment FIN.Vol PER.SAMP IDEAL.Copies   
## Min. :-2220.0 Min. : 120 Min. :0.0120 Min. : 0.22   
## 1st Qu.: 0.0 1st Qu.: 120 1st Qu.:0.0120 1st Qu.: 2.00   
## Median : 0.0 Median : 300 Median :0.0300 Median : 22.00   
## Mean : -351.5 Mean :1278 Mean :0.1278 Mean : 4831.44   
## 3rd Qu.: 0.0 3rd Qu.:1000 3rd Qu.:0.1000 3rd Qu.: 187.00   
## Max. : 0.0 Max. :6180 Max. :0.6180 Max. :84000.00   
##   
## Copy.Adjustment REAL.Copies Sample.Name Target   
## Min. : 0.00 Min. : 0.22 Length:379 Length:379   
## 1st Qu.: 0.00 1st Qu.: 2.00 Class :character Class :character   
## Median : 0.00 Median : 22.00 Mode :character Mode :character   
## Mean : 24.22 Mean : 4807.21   
## 3rd Qu.: 0.00 3rd Qu.: 187.00   
## Max. :414.00 Max. :84000.00   
##   
## Ct Column3 Column4   
## Length:379 Min. :20.18 Min. :0.0048   
## Class :character 1st Qu.:29.42 1st Qu.:0.0507   
## Mode :character Median :32.51 Median :0.1883   
## Mean :32.01 Mean :0.3781   
## 3rd Qu.:35.40 3rd Qu.:0.3788   
## Max. :39.80 Max. :1.8626   
## NA's :318 NA's :341

knitr::kable(Data[1:5,6:22] ,caption = 'First 5 Rows of data and a selection of variables table.')

Table 2.1: First 5 Rows of data and a selection of variables table.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Tool | Method | Type | Dilution.Tube | Dilution.Factor | Tool.Vol | Input.Vol | Vol.Adjustment | FIN.Vol | PER.SAMP | IDEAL.Copies | Copy.Adjustment | REAL.Copies | Sample.Name | Target | Ct | Column3 |
| Sanigen | CDC | Sample | I | -2 | 10000 | 120 | 0 | 120 | 0.012 | 22.000 | 0 | 22.000 | CDC 120 I | N1 Probe | 32.42782593 | NA |
| Sanigen | CDC | Sample | I | -2 | 10000 | 120 | 0 | 120 | 0.012 | 22.000 | 0 | 22.000 | CDC 120 I | N1 Probe | 31.51000595 | NA |
| Sanigen | CDC | Sample | II | -3 | 10000 | 120 | 0 | 120 | 0.012 | 2.000 | 0 | 2.000 | CDC 120 II | N1 Probe | 34.5826683 | NA |
| Sanigen | CDC | Sample | II | -3 | 10000 | 120 | 0 | 120 | 0.012 | 2.000 | 0 | 2.000 | CDC 120 II | N1 Probe | 35.53193665 | NA |
| Sanigen | CDC | Sample | III | -4 | 10000 | 120 | 0 | 120 | 0.012 | 0.224 | 0 | 0.224 | CDC 120 III | N1 Probe | Undetermined | NA |

## 2.1 Questions to answer

There are 3 major aims of this:

1. Do the sample enhancement strategies work?
2. How well do they work compared to each other?
3. Are there any environmental variables recorded that have a positive impact on detection?

Testing several sample enhancement strategies (Nanoparticles and Bait & Capture) for molecular detection of heat-inactivated SARS-CoV-2 from stainless-steel surfaces. The primary aim of this work is to determine if any of the enhancement strategies increase the concentration of RNA copies in a sample that will express itself as a lower Ct value from the Rt-PCR output. Lower Ct values equate to more copies being present in the sample meaning less time needed to confirm presence.

In the case the sample enhancements work, we would then compare how well they worked against an automated extraction method. The degree of “wellness” will be determined by how low the Ct value results are produced after a sample has going through a specific method.

In addition to the comparison of enhancement strategies the effects of the tool used for sampling (macrofoam or knit-fiber), time held, and other environmental variables will be explored any significant effects, with interest in positive impacts to detection.

### 2.1.1 Proposed Analysis

1. General data wrangling and cleaning
2. Summarize data for overview.
3. Compare Standard Curve Data for run differences
4. Compare Ct values between Methods, Tools, and Inputs (boxplots)
5. Adjust for difference if there were significant changed (ln?) between Serial dilutions on each PCR run.
6. Investigate any effects on the Ct recovery
7. Investigate loss between spiking experiment and sampling experiments

### 2.1.2 Background Basics

Outbreak surveillance is contingent on the samples recovered from the environment and then identified. The introduction of molecular methods has helped in the rapid identification of species causing outbreaks, but these methods are only as good as the sample collected and used.

To insure a good sample can be taken, culture methods are generally required to grow all things present in the sample. Culture based methods are the gold standard in identification, but require resources and time to culture and a multitude of other factors may cause targeted organisms to die off or out competed during culturing.

Enter sample enhancement and a direct to molecular method workflow for quick ready to sequence samples.

Low-quality samples often are confounded by contamination or just low presence. Here I plan to utilize two types of sample enhancement strategies to see if these technologies work and how well they work against an automated sample extractor (Promega Maxwell 48). The first to be tested are Nanotrap Particles (Ceres Nanoscience), a hydrogel particle enhanced with specific affinity dyes that attract a target organism. The second strategy to be tested will be capture Baits (Arbor Sciences)- customized complementary baits for a target that can be used to select specific targets.

Though this project was created with Antimicrobial resistant organisms in mind, the 2020 pandemic halted research on non mission critical (Covid related) projects. Explaining the use of the SARS-CoV-2 virus in this study.