Monitoring SARS-CoV-2 Evolution Through Targeted Next-gen Sequencing of Wastewater Extracts in Kent County

Professional Science Masters in Cell and Molecular Biology

Dr. Tsou & Dr. Blackman, MoM Lab, GVSU

Fall, 2023

Table of Contents

Learning and Internship Objectives

This internship, which was started during the summer 2023 semester, had four primary objectives. First, I was to be trained in the process by which we sequence and track variants of SARS-CoV-2 in the Kent County wastewater. Second, I was to refine the variant sequencing process by creating a duplex PCR for two genetic regions of interest, allowing for increased variant tracking efficiency. Thirdly, I was to synthesize graphics displaying the change in wastewater variant composition throughout Kent County over time. Fourth, I was to work alongside [redacted] a fellow graduate student studying computer science, to develop and refine a new variant identification python program.

Introduction

First appearing in late 2019, SARS-CoV-2 quickly swept the world. As clinical tests for SARS-CoV-2 infection became available, millions were administered nationwide, and were used not only to detect the presence of SARS-CoV-2, but also to detect and track variants of the virus. As viruses spread and multiply, they tend to accrue mutations, a rare few of which will result in an increase in that particular virus's evolutionary fitness. If a particular viral variant, distinguishable by its unique mutations, increases its fitness enough, it will become dominant within a population, outcompeting other variants. As cases of Covid-19 fall and clinical testing rates decline, fewer and fewer samples are available from which to track variant composition. Wastewater testing has risen to fill this gap in testing, allowing health departments across the United States to track not only the prevalence of SARS-CoV-2 in various communities, but also to determine which variants of said virus are present therein.

SARS-CoV-2 infects human cells using the "spike" protein which binds the human ACE2 receptor. This protein is under great selective pressure, as any mutations which increase the binding efficiency of this protein would make said virus more infectious, and thus increase the fitness of the virus. For this reason, the receptor binding domain of this protein, which interacts with the ACE2 receptor directly, tends to contain numerous clinically relevant mutations. While there are many methods by which variants can be identified, the MoM lab used next-gen amplicon sequencing of just the receptor binding domain of the spike protein to identify variants following a protocol developed by the Johnson Lab at the University of Missouri. Sequencing only this portion of the SARS-CoV-2 genome allows us to identify most clinically relevant variants while keeping costs comparatively low and reporting times expedient.

Description of Work

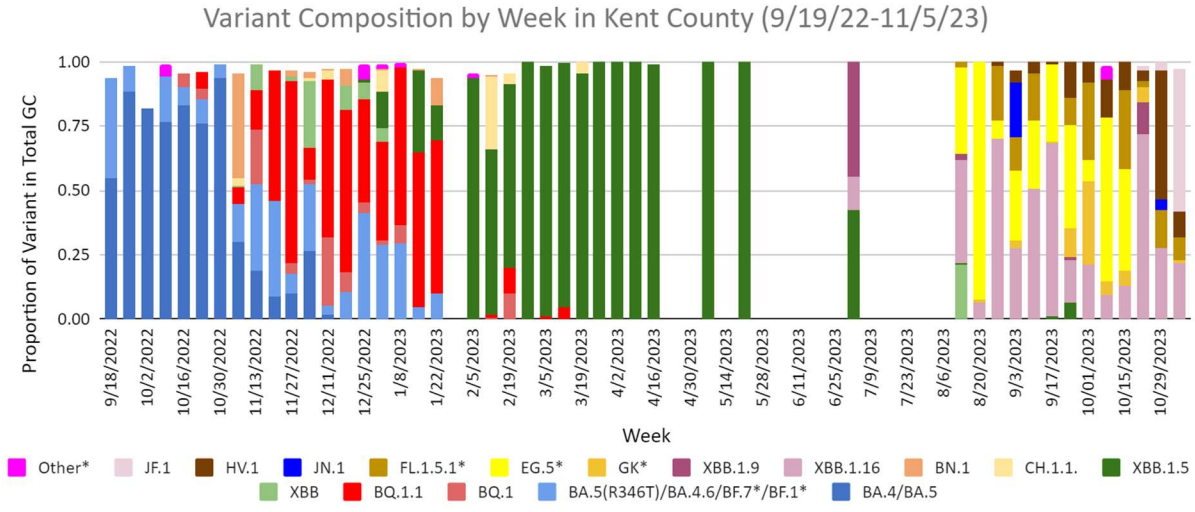
The bulk of my internship involved performing variant sequencing on samples of SARS-CoV-2 RNA collected weekly from wastewater monitoring sites throughout Kent County. This weekly process began with RNA extraction and purification performed by other members of the MoM lab team. When we received samples with concentrations of SARS-CoV-2 RNA over a set threshold, I then processed the samples using a two-step, or nested RT-PCR process. The first step in this process involved reverse transcribing the receptor binding domain RNA into DNA. The second step involved making billions of copies of this DNA while adding adaptor sequences to their ends which facilitate sequencing. Successful reverse transcription and amplification was determined by gel electrophoresis, and thereafter samples of amplified DNA were sent to our partners at the University of Missouri for sequencing. After sequencing was complete, our partners would send us the results, detailing both the mutations of each variant in every sample as well as the proportion of the sample those mutations made up. I then interpreted these mutations to determine the proportions of each variant in our samples employing variant identification programs developed by [redacted].

My work with [redacted] involved creating and refining a new variant identification program. The original program I had used was developed by Ian Staves and compared the lists of mutations to a hand-curated variant dictionary. This system required analysts to continually update this variant dictionary as new variants emerged. By contrast, [redacted]'s program queried the online database GISAID – which contains SARS-CoV-2 genome date collected worldwide – eliminating the need for a hand-curated dictionary.

One limitation of this process is its focus on the receptor binding domain, because variants with identical mutations within this region cannot be differentiated by our process. In

particular, the variants XBB.1.5 and XBB.1.9 could not be distinguished within the receptor binding domain alone. To remedy this issue, we began sequencing the "N-gene", a section of the nucleocapsid gene which can be used to differentiate between the two variants; however, sequencing two regions instead of one doubled the amount of work required for each sample. To increase our process efficiency, we began investigating the possibility of duplex PCR, which would allow us to amplify both sections of DNA in a single reaction instead of two. This process is still in its early stages, so its success cannot yet be determined.

As part of our work monitoring wastewater throughout Kent County for SARS-CoV-2, we report our findings to the Kent County health department and other stakeholders. To facilitate the clear and efficient conveyance of this information, I created a graph displaying the change in variant composition throughout Kent County over time. This graph contained data from all wastewater treatment plants within Kent County and was normalized for flow rate (see below).



Internship Discussion

Of my four original objectives, two were fully completed and two are works in progress. I was trained on all aspects of the variant testing and data management process and continue to perform these roles through the present. I successfully created a graphic displaying the change in wastewater variant composition overtime in Kent County and have continued updating the graph as new data becomes available. While I began developing a duplex PCR method to increase the efficiency of our process, the method has not been finalized as of yet. I have been using [redacted]'s variant identification program for roughly a month and a half, and while it is fully functional, we continue to refine its outputs.

This internship taught me not only how to perform variant sequencing, but also transferable skills like accurate data management and concise scientific communication. Although the PSM coursework did not greatly prepare me for the specific scientific portions of my internship, I feel they instilled in me the professionalism and organizational skills necessary to be successful in a research environment. This project often required me to process large numbers of samples, which resulted in large quantities of data. Throughout the course of this internship, I not only learned to process and organize these large sums of data, but also to communicate this data to stakeholders of diverse educational backgrounds. I expect that my ability to manage and communicate complex data will serve me well in the workplace. Overall, I feel this internship has instilled in me many soft skills necessary to succeed outside of academia.