UNIVERSITY OF CALIFORNIA
SANTA CRUZ

PROBE-FREE QUANTIFICATION OF VIRAL POPULATIONS IN WASTEWATER ON THE MINION NANOPORE

A thesis submitted in partial satisfaction of the
requirements for the degree of

MASTER OF SCIENCE

in

MICROBIOLOGY AND ENVIRONMENTAL TOXICOLOGY

by

Ryan M. Modlin

June 2021

The Thesis of Ryan M. Modlin
is approved:

Professor David Bernick, Chair

Professor Karen Ottemann

Professor Benedict Paten

Professor Chad Saltikov

Quentin Williams
Acting Vice Provost and Dean of Graduate Studies

Ryan M. Modlin

2021

Table of Contents

List of Figures

List of Tables

Abstract

PROBE-FREE QUANTIFICATION OF VIRAL POPULATIONS IN WASTEWATER ON THE MINION NANOPORE

by

Ryan M. Modlin

Recently, wastewater surveillance of viral populations, especially SARS-CoV-2, has become a popular choice to actively monitor community spread of disease. While current methods like digital and quantitative PCR are popular choices for wastewater surveillance, their multiplexing capabilities are limited to the number of color channels available. Additionally, the use of probes requires prior knowledge of a target sequence making identification of an unknown agent impossible. The ability to identify, monitor, and track the spread of viral outbreaks in early stages are essential to preventing the spread of disease and potential pandemics. In this paper, we define a method to identify and quantify viral populations from wastewater samples on the MinION Nanopore. The MinION Nanopore, a Next Generation Sequencing device, is able to sequence nucleic acids without prior knowledge of its sequence and as a result, can distinguish between all unique species in a sample. Furthermore, by determining the number of molecules of a specific species and calculating its starting concentration from mass and read length information, we are able to quantify every unique sequence in a sample.

Two nanopore experiments were run to test this model. In the first experiment, we quantified several gene blocks of a known starting concentration to test the precision and accuracy of the device. In the second experiment, we quantified viral RNA from wastewater and aligned the reads to the NCBI viral database to determine the identity of each molecule. Quantification of this run was directly compared to digital PCR results for SCV2 and Pepper Mild Mottle Virus amplification. Although we are only comparing the results of two viruses, depending on the results of this experiment, theoretically, this device has the potential to quantify every single unique organism or pathogen present in a sample.

Acknowledgments

I want to thank my friend and mentor, David Bernick, who has supported me in countless ways over the past few years. Thank you to everyone in the Bernick lab who has supported myself, my project, and my mental health through a crazy two years; James Hahn, Melody Azimi, Torey Brownell, Kyra Eyerman, Stephen Hwang, Allie Jorgenson, Matthew Kozubov, Rachel Mace, and Sophia Sneddon. I would also like to thank my advising committee, Karen Ottemann, Benedict Paten, and Chad Saltikov. And finally, for everyone else who has assisted in supporting this project or providing me with advice along the way; Bryan Condy, David Ghilarducci, Gail Newel, Eric Gross, Combinati, Hugh Olsen, Miten Jain, George Spix, John MacDonald, and Scott Brown.

Chapter 1

Introduction

Over the past year, the SARS-CoV-2 (SCV2) pandemic has made it clear how important it is that we monitor, study, and respond to pathogenic viral spread. There are an estimated 164 million SCV2 cases globally which has resulted in over 3 million deaths since the beginning of the pandemic in January 2020. (CDC 2021) Locally, Santa Cruz county estimates there are 16,000 cases and 250 cases here at UCSC. (Santa Cruz Health 2020, UCSC 2020) In our perceptually narrow view of time, pandemics may seem quite rare, but their effects are painfully apparent for years to come. It is essential once the world tackles the SCV2 pandemic we do not become complacent and prepare for future epidemics.

The most common method of detecting and tracking community spread of SCV2 currently comes in the form of nasopharyngeal swabs. While nasopha-

ryngeal swabs are ideal for determining the presence of SCV2 in an individual, data stemming from this method are difficult to use when assessing SCV2 prevalence because of incomplete representation from testing bias towards symptomatic individuals, uneven test accessibility, and stigma associated with being tested. Recently, wastewater surveillance of viral populations, especially SCV2, has become a popular choice to actively monitor community spread of disease. SCV2, specifically, is found in human fecal matter and quantifying the amount of viral RNA present in a composite sample can help indicate current community spread. (Mallapaty 2020) Wastewater surveillance of viruses is advantageous to traditional nasopharyngeal swabs for collecting prevalence data because it is non-invasive and unbiased towards individuals who choose to be tested. Additionally, wastewater surveillance can also be used as an early warning system for viral spread by setting up response clusters in specific communities.

Regardless, nasopharyngeal swabs and wastewater surveillance systems are both important detection systems currently in place to combat the SCV2 pandemic. These methods, in combination with public health policies, should be replicated in the chance of a future viral outbreak, but it is not enough to predict and catch a pandemic before it becomes widespread. Nasopharyngeal swabs and wastewater surveillance have different collection, extraction, and concentration protocols, but share the same quantification methods. Digital PCR (dPCR) and

quantitative PCR (qPCR) are staples in biology and biochemistry because they provide high sensitivity and specificity for the detection of nucleic acids. The specificity offered by both techniques is highly valued, but the reliance on fluorescent and dye based probes makes the multiplexing capabilities limited. Additionally, the use of probes requires prior knowledge of a target sequence. It is imperative that we take measures to monitor all pathogenic organisms, known or unknown, to more rapidly predict and prevent future pandemics.

1.1 Significance and Goals

The MinION Nanopore has recently seen an increase in popularity as a Next Generation Sequencing (NGS) device. Nanopore strand sequencing is possible by recording the current differential as a molecule passes through the pore and decoding these reads into nucleobases using an algorithm, Bonito, Guppy, etc. (Dreamer 2016) We want to capitalize on these MinION capabilities, incorporate a method of digital quantification, and create a quantification device with a limitless multiplexing ability.

Digital quantification of nucleic acids and limited dilution is currently being used in dPCR. dPCR works by partitioning the sample into droplets or wells and amplifying them with a fluorescent dye, creating a set of fluorescent and non-fluorescent partitions. The starting concentration of the sample is then calculated

using limited dilution. If the molecule of interest is randomly distributed in the sample, such that 20-80% of the partitions are empty, the starting concentration can be determined by counting the number of empty partitions and modeling this value to the probability mass function (PMF) of the Poisson distribution. (Morley 2014) Unlike dPCR however, the MinION does not require the use of probes and is therefore not limited to the number of color channels available on a machine or pathogens whose prior sequence is known. The MinION is able to distinguish between different molecules because its primary function is as a sequencer.

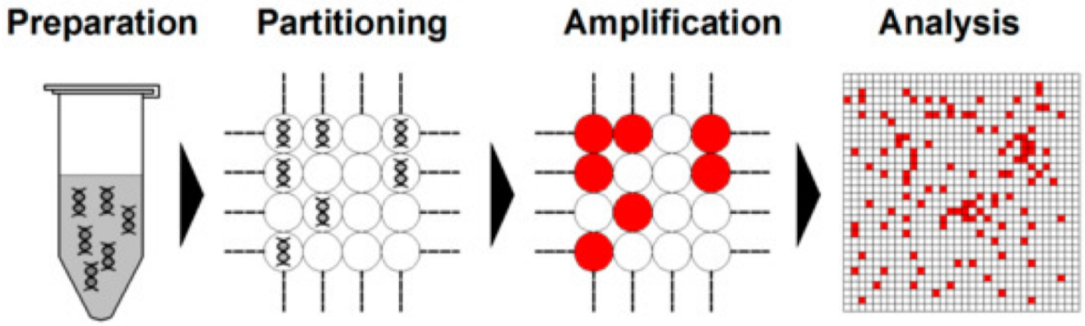

Figure 1.1: Digital PCR Workflow - The figure above shows the general workflow of dPCR and how the user is able to generate quantification data from a sample.

Instead of partitioning the sample in droplets or wells, our goal was to create a method to partition the sample digitally, based on time. As each molecule passes through the pore, the MinION records a timestamp. Depending on the total number of reads for a particular model and the length of the run, we could select time windows such that 20-80% of the partitions will be empty. We would then

count the number of instances that the molecule of interest does not pass through the pore in a window and model this value to the Poisson distribution. The goal was to generate a total count a particular molecule passed through the pore. With this value we can determine the starting concentration of the original molecule by calculating the number of molecules that did not pass through the flow cell based on the total mass and the total number of bases read. This new method of quantitation holds several key advantages of dPCR and other quantitation devices. First, this method of partitioning is more flexible and can be changed based on the duration or mass of a Nanopore run. Secondly, this method of quantitation does not rely on the use of fluorescent probes or prior knowledge of the sequence, resulting in an effectively limitless multiplexing capability. Lastly, unlike other quantitation devices, this method also provides sequence data for each molecule, reducing the risk of a false positive and contributing to genomic studies.

This leads to three main goals for this project. First, as part of our lab's response to the SCV2 outbreak, we will amplify and sequence SCV2 from wastewater samples as a way of collecting variant data for the Watsonville community. This portion of the project is not meant to quantify viruses in wastewater and instead provide robust genomic information through the ARTIC sequencing protocol. Second, we will test the precision and accuracy of the MinION's quantitative ability by preparing a library of several gene blocks of known concentrations. Finally,

we will extract and concentrate viral RNA from wastewater and quantify viral populations in the Watsonville community.

Chapter 2

Research Plan and Methods

2.1 Quantifying MinION Nanopore Reads

Considering digital quantification of nucleic acids is currently being used in dPCR, our first goal was to demonstrate that we could model MinION reads to the Poisson distribution. Similar to partitioning based on droplets or wells, we can model the Probability Mass Function (PMF) by counting the number of time windows where the molecule of interest did not pass through the pore. The equation for the PMF is defined as $P(k|\lambda) = \lambda^k e^{-\lambda} / k!$, where λ is the average number of samples per partition and k is the number of occurrences in a partition. According to the PMF, when k equals 0, the equation becomes $P(0|\lambda) = e^{-\lambda}$. In this instance, if we are able to count the number of time windows where the

molecule of interest did not pass through the pore, we can calculate the average number of occurrences of a particular molecule in solution. Additionally, since our partition is digital, we are able to adjust this value based on the amount of template loaded and length of the run.

$$P(k \mid \lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{2.1}$$

For this method to work, cDNA needs to be randomly distributed inside the sample. We are able to partition on time because only one strand can pass through the pore at any given moment. I created a file reader that interprets fast5 files, nanopore file format, using the ONT Fast5 api and records the number of reads of a particular molecule, its timestamp of when it entered the pore, and the total length of the run. (Nanoporetech 2017) These values are then fed into the model. The number of time windows is set equal to 22.5% of the total number of reads of the molecule of interest which helps ensure that 20% - 80% of the partitions are empty. Afterwards, the number of reads in each respective time window is counted, along with the number of instances where the time window is empty, and the data is modeled to the PMF of the Poisson distribution. Due to limited access to data, I used publicly available nanopore reads provided by Monash University. (Wick 2019) I selected two gram negative organisms, *Klebsiella pneumoniae* and *Stenotrophomonas pavanii*, and the modeling results of each can be seen below

in Figure 2.1. In both instances, each separate MinION run followed the fit of the Poisson distribution. This indicated that we can successfully partition our sample based on flexible time windows and are able to quantify different nucleic acid sequences on the MinION nanopore.

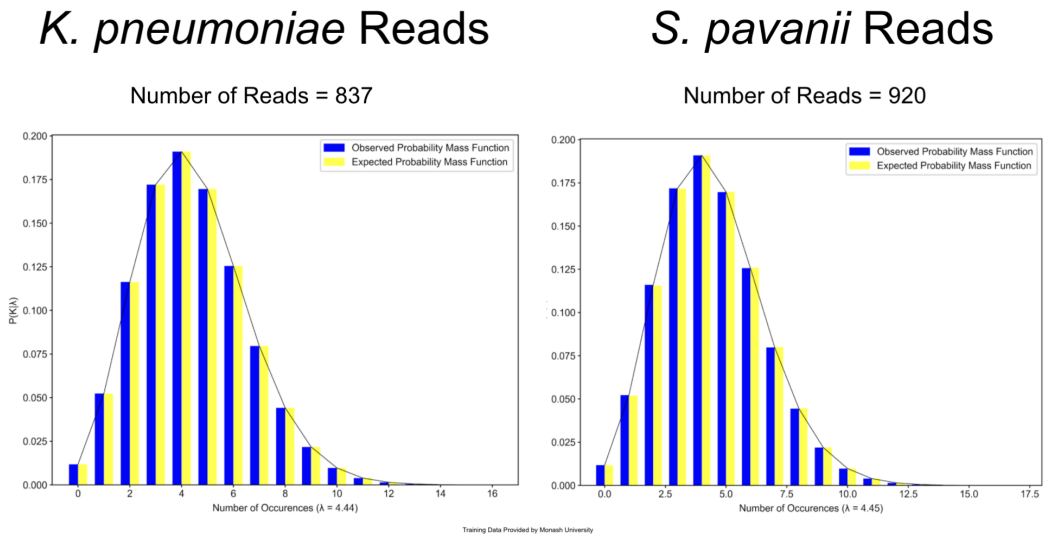

Figure 2.1: Poisson Distribution of *K. pneumoniae* (left) and *S. pavanii* (right) Reads - The two graphs above show two separate MinION runs modeled to the Poisson distribution. The graph on the left shows the distribution of *K. pneumoniae* reads and the graph on the right shows the distribution of *S. pavanii* reads. The y-axis represents the frequency of reads that pass through the pore given the average number of occurrences and the x-axis represents the number of reads that pass through the pore in any given window. The expected number of occurrences for each graph are 4.44 and 4.45 respectively.

Additionally, we also need to account for any nucleic acids that are lost on the flow cell. A majority of nucleic acids loaded onto the flow cell are lost during sequencing because they more often than not get stuck on the walls of the flow cell. For DNA, roughly 1 in 100,000 molecules that are loaded actually get sequenced.

In order to translate these counts to their starting concentrations, we need to calculate the number of molecules that get stuck in the flow cell and are not read by the MinION. Below are the two equations we use to calculate the starting number of molecules in the sample from the MinION counts. The first equation, Equation 2.2, calculates our copy number ratio. In this case we define the copy number ratio as the ratio of the number of molecules that pass through the flow cell compared to the total number of molecules that are loaded onto the flow cell, EB_T. The copy number ratio is calculated by the total mass loaded onto the flow cell divided by the number of bases recorded on the MinION at a particular point in time. The constant converts our results from nanograms to daltons. In the second equation, Equation 2.3, we determine the total number of molecules, in the library by multiplying the copy number ratio to the total number of reads, R, for a particular molecule. We can subsequently apply this to each unique molecule recorded in the MinION run to calculate the starting concentration of each species.

$$CNR = \frac{TotalMassLoaded}{EB_T * 1.08E - 10} \tag{2.2}$$

$$\#Molecules = R_S * CNR \tag{2.3}$$

2.2 Digital PCR Quantification

We decided to use dPCR as a control to verify our MinION quantification results. Unlike qPCR, dPCR also uses the limited dilution principle for quantification and is theoretically more accurate since it is not reliant on a standard. (Morley 2014) Wastewater samples were analyzed in a 4-plex assay. Two of our channels, FAM and HEX, were used to quantify two different positions on the N-Gene of the SCV2 genome. This allowed us to compare these channels against each other and have more confidence in our SCV2 measurements. One issue when quantifying SCV2 from wastewater comes from viral loss during RNA extraction and concentration. In order to quantify SCV2 loss, we used a normalizing constant in our third channel, CY5, which binds the Replication Associated Protein (RAP) of Pepper Mild Mottle Virus (PMMV), to estimate viral loss in each individual run. PMMV is a pathogenic plant virus that is considered to be one of the most abundant viruses found in human fecal matter. It is estimated that PMMV is found in concentrations of up to 109 virions per gram of dry fecal matter and is present in 66.7% of individuals in North America. (Zhang 2005) The last channel measured an optional Bovine Coronavirus (BCoV) spike in and which was used as quality control for RNA extraction and concentration for multiple runs. 5 μL of template DNA was mixed with the 4-plex primer probe mix, and a master mix containing enzymes for cDNA generation and run on Combanati's dPCR machine.

The sequences for the primers and probes are listed in supplemental Table A.1.

2.3 Extraction and Concentration of Wastewater RNA

In conjunction with this project, our lab is performing wastewater analysis for the City of Watsonville. We used several RNA extraction and concentration techniques on wastewater samples including Trizol and chloroform, ultracentrifugation and magbead cleanup, the Berkeley 4S (Sewage, Salt, Silica, SARS-CoV-2) protocol, and Nanotrap Magnetic Virus Particles. Initial wastewater samples were spiked with Newcastle Disease Virus (NDV) which served as a positive control. The first round of experiments adapted Thermo Fisher's Trizol cleanup protocol. (Thermo Fisher N/A) While we believe this method did yield traceable amounts of RNA, we deemed it ineffective because of consistently high phenol and guanidinium salt contamination resulting in unreliable PCR results. It also did not scale in a cost effective manner compared to the other three methods. Next, we adapted the EPA's ultracentrifugation protocol for concentration and the UCSC MDL's SCV2 extraction method used for nasopharyngeal swabs. (CDC 2020) Here, we were successfully able to detect our NDV positive control but were unable to quantify SCV2. While no salt or phenol contamination was present, we

were not able to extract SCV2 RNA in high enough yields. The magbeads bind both DNA and RNA and have a maximum binding capacity of 8 μg. Since we concentrated and processed 40 mL of effluent material, the columns were overloaded with Human and *E. coli* nucleic acids.

We found success with the remaining two methods of wastewater RNA extraction. The first method, the 4S protocol, stands for sewage, salt, silica, and SARS-CoV-2. This technique lyses all virus and live cells in solution and collects RNA from the sample non specifically. The second method employs the use of Nanotrap Magnetic Virus Particles which are able to bind intact virus and allow for the selection and concentration of viral RNA specifically.

2.3.1 RNA Selection using the 4S Protocol (Sewage, Salt, Silica, SARS-CoV-2)

The 4S protocol, created by UC Berkeley, showed the most success on our initial samples. This method uses 4M salt to lyse 40 mL of effluent material and separates most of the larger molecules in a 5 micron filter. The filtered effluent is then concentrated on a silica column and eluted with 100 μL of H2O. The outline for this protocol can be seen in Figure 2.2. As opposed to the magbeads, the silica columns have a binding capacity of 300 μg, which is more than sufficient for our current sample volume. (Whitney 2020) On average this protocol has yielded

about 50 μg of nucleic acids or 500 $\frac{ng}{\mu l}$.

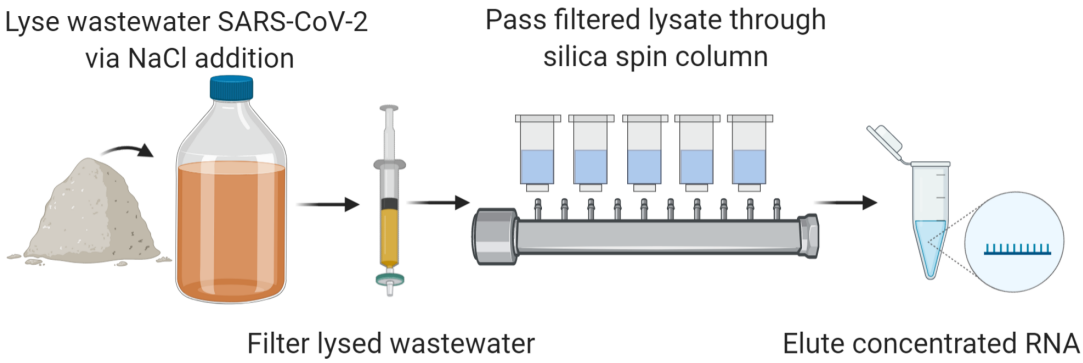

Figure 2.2: RNA Concentration and Extraction in the 4S Protocol - The image above shows the general step by step process for concentrating and extracting all RNA from wastewater in 4S protocol. (Figure provided by Whitney 2020)

2.3.2 Viral RNA Selection using Nanotrap Magnetic Virus Particles

While the 4S protocol worked very well for quantification using primers and probes, we later realized that an overabundance of non viral RNA led to problems downstream for cDNA generation from hexamers for nanopore purposes. Although this protocol is a bit more time and resource intensive, we used Nanotrap Magnetic Viral Particles which are able to bind and capture intact virions. (CERES Nanosciences 2021) Using a magnetic separation rack, we can then separate these beads from the rest of the sample. This method would allow us to specifically target viral RNA and avoid the collection of free floating or lysed RNA

from prokaryotes and eukaryotes in solution. The workflow of the protocol can be seen in Figure 2.3 below. After the separation of viral RNA from solution, the downstream process of concentration was accomplished with a commercial kit. In this case, we resuspended intact virus in Phosphate Buffered Saline (PBS) and then lysed, washed, and concentrated our sample to 50 μL using Qiagen's Viral Mini Kit. Viral RNA generated from this method was used for viral quantification on the MinION Nanopore and compared to the dPCR 4-plex assay.

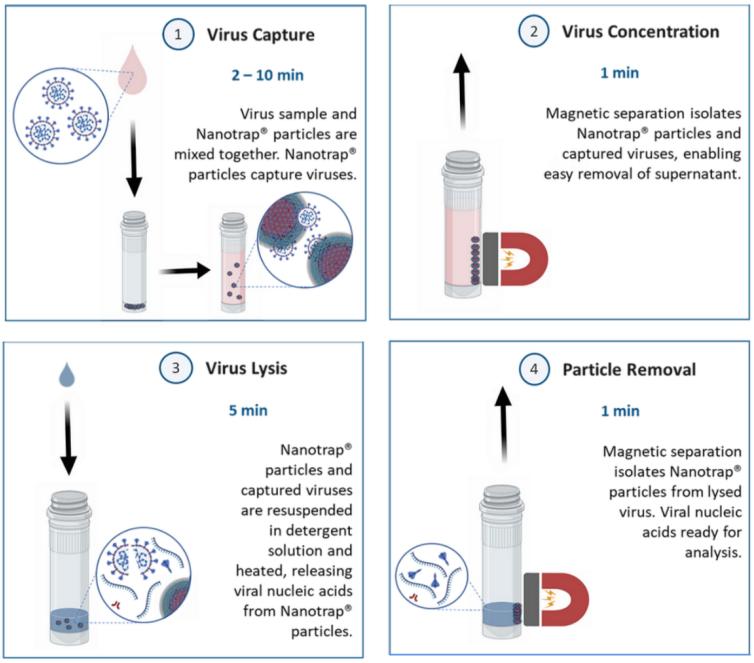

Figure 2.3: Viral RNA Extraction using Nanotrap Particles - The panels above show the general protocol on how to capture, concentrate, lyse, and extract virus using Nanotrap Magnetic Viral Particles. (Figure sourced and modified from CERES Nanosciences)

2.4 SARS-CoV-2 ARTIC Sequencing

One potential concern with quantifying SCV2 from wastewater samples is ensuring the molecule of interest is well above the minimum threshold of detection. Originally, we were going to quantify SCV2 RNA to reduce the number of preliminary steps and modification to the sample, but after learning the sensitivity of the MinION is 100x higher for cDNA and ssDNA compared to RNA, we felt it was best to change course and use cDNA. (Dreamer 2016) While protocols for detecting wastewater sampled virus on the MinION are scarce, researchers at the University of British Columbia established a method of amplifying SCV2 from nasopharyngeal swabs for the nanopore. (Tyson 2020) The ARTIC protocol provides full genome coverage and high sensitivity through direct amplification of SCV2 using tiled, multiplex primers. 109 primers sets are split between two pools and generate 400 bp amplicons that span the full genome. Pools are designed with alternating primer sets so there is no overlap in order to reduce the potential for primer mismatch during amplification.

RNA was extracted from wastewater using the 4S protocol and quantified through dPCR to confirm the presence of SCV2. We eventually modified the original ARTIC protocol, to use Combanati's enzyme with ARTIC primers in order to generate SCV2 cDNA specifically instead of Lunascript RT Supermix. PCR reactions were run on a gel after amplification to confirm the presence of

amplicons. For reactions that did not show up on the gel, 5 μL of the PCR reaction mix was used as template for a new reaction and amplified again to increase the concentration of amplicons, sequential PCR. The sequential PCR mix was rerun on a gel to confirm the presence of amplicons.

Had both primer pool 1 and primer pool 2 been present on the gel, the PCR reactions would have been cleaned using the monarch clean and concentrate kit. The SCV2 amplicons would then be prepared for nanopore sequencing following Josh Quick's ARTIC library preparation protocol. (Quick 2020)

2.5 Library Preparation, Barcoding, and Adapter Ligation

Both the quantification control and viral wastewater experiment used a modified version of the Oxford Nanopore Technologies (ONT) Ligation Sequencing Kit 109 (LSK-109) library preparation protocol. Initial template concentration was selected to maximize nanopore reads. For the 9.4.1 flow cells, this meant aiming for a final library of 5-50 fmol. (Nanopore Community 2021) 20-50 μL of template DNA was mixed with NEBNext Ultra II End Prep Enzyme and Reaction buffer. For the quantification control experiment, each end prep-reaction was then ligated with a NBXX barcode and mixed together after incubation. This allowed

us to save time in having to perform two separate wash steps. From this point, the end-prepped cDNA from the viral wastewater experiment and the barcoded DNA from the quantitation control experiment were then concentrated using AM-Pure XP beads on a magnetic rack and bathed twice with 70% ethanol. Samples were eluted with 61 μL of MilliQ. Adapter ligation and sample loading onto the MinION followed the recommended LSK-109 protocol from this point on. (ONT 2021)

Chapter 3

Results and Discussion

3.1 ARTIC Sequencing

Our first goal was to sequence SCV2 and look for variants on the MinION Nanopore. Initial results with the ARTIC method showed no amplification with either primer pool. The protocol was adjusted for touchdown PCR and increased template but amplification was never observed. Because of this outcome, we adjusted the ARTIC protocol and used enzymes provided by Combanati as opposed to NEBs Lunascript. Lunascript contains random hexamers which we theorized were binding Human and *E. coli* RNA instead. cDNA generation with Combanati's enzyme and ARTIC primers in a two step PCR protocol resulted in amplicon generation in primer pool 2. The gel in Figure 3.1 below displays the PCR results

of a semi successful ARTIC experiment, with template RNA from January 2021 when SCV2 prevalence levels were peaking in the Watsonville area. Amplicons are present around 400 bp for pool 2 PCR and sequential PCR reactions, V2 and V2S. Unfortunately, we never managed to achieve amplification with the pool 1 primers, regardless of the sample. Consistent dPCR results around 70 $\frac{copies}{\mu L}$ for this month, Supplimental Figure A.1, and amplification in the pool 2 primers support the idea that a lack of SCV2 present was not an issue for amplification. For this reason, we believe amplicon generation issues in pool 1 most likely stems from an issue with the primer pool we were given from IDT. Additionally, we seem to see additional bands in both pool 1 and pool 2 around 1500 bp. In this case, we are unsure what caused amplification in this region but think it's most likely due to primer mismatching.

Since amplicon generation was not achieved in both pool 1 and pool 2, variant sequencing of SCV2 from wastewater on the nanopore was set aside to focus on the two quantification experiments.

3.2 MinION Quantification Control Test

Before we attempted to quantify viral populations on the MinION, we needed to test how precise the device was under controlled conditions. Three gene fragments synthesized from IDT, 7DH, AdhE2, and AtoB, were selected to serve

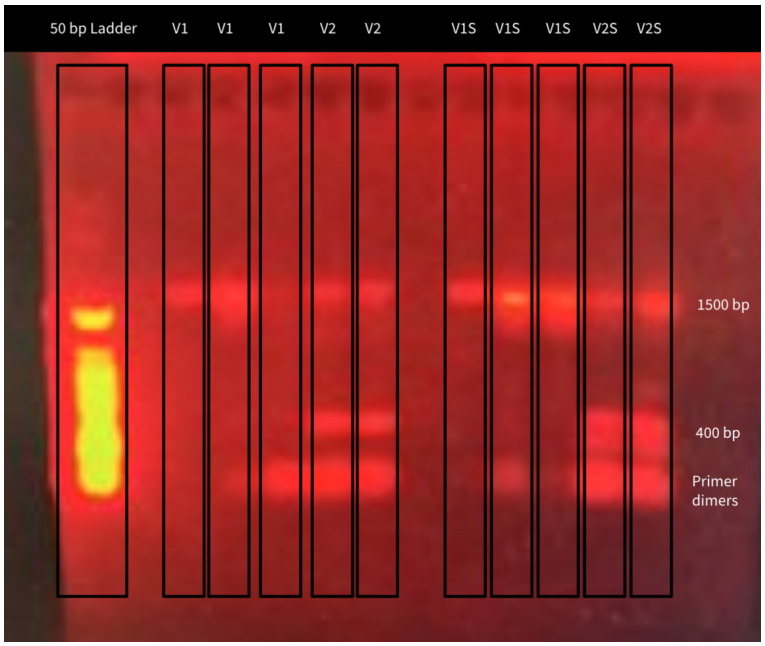

Figure 3.1: Gel Electrophoresis of Wastewater Samples Amplified with ARTIC Primer Pools - The above figure shows the PCR and sequential PCR results of multiple samples. V1 and V2 denotes which ARTIC primer pool was used and samples with S are sequential PCR samples. Bands can be seen at the 1500 and 400 bp marks.

as quantification controls. Gene blocks were selected over other template types because this gave us complete control over starting concentration and fragment length. Depending on the fragment length, 90 - 120 ng of each gene was used in our sequencing library. Gene information and expected quantification results from the MinION can be seen below in Figure 3.2.

The library was run on the MinION Nanopore for 24 hours. Statistics for the run can be seen in Figure 3.3 and Figure 3,4 below. Figure 3.3 shows the total number of bases read during the duration of the run. The information is useful

Gene	Sequence Length	Mass Loaded	Expected Copy Number
7DH	1119	30	2.48 E+10
AdhE2	944	30	2.94 E+10
AtoB	1189	30	2.34 E+10

Figure 3.2: Nanopore Quantification Input Controls - The table above shows the input of the three gene controls loaded onto the flow cell.

because it allows us to quantify the sample at any particular time point we choose. Figure 3.4 shows the average read length of our library over the 24 hour period. This is important to consider because the MinION has a tendency to sequence smaller molecules towards the beginning of the run, as seen below. It isn't until about 12 hours into the run where the average read length begins to peak and steady off around 1070 bp. If quantification of the sample takes place too early, at 6 hours for example, the MinION will be more likely to overrepresent smaller molecules and underrepresent larger molecules.

Using this information, we decided to pick 4 time intervals, 6 hours, 12 hours, 18 hours, and 24 hours, to quantify our genes. Read counts were directly taken from MinION barcoding counts in this instance and multiplied by each respective copy number ratio. The molecule counts from this experiment can be seen in Figure 3.5 and the percent error of each quantification can be seen in Figure 3.6.

For the estimated molecule counts at 6 hours, we see the problem of over and underrepresenting certain fragments described above. AdHE2, our smallest

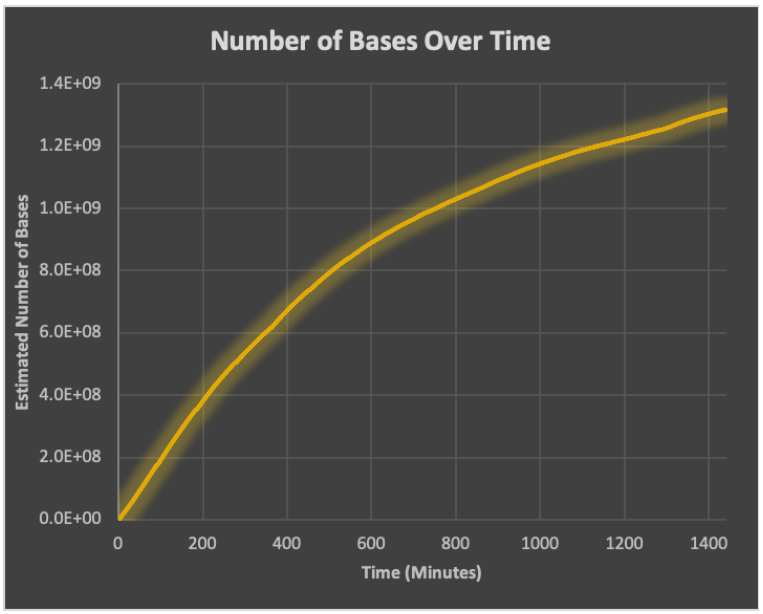

Figure 3.3: Total Count of Bases Sequenced - The above figure shows the number of bases sequenced by the MinION over the course of the 24 hour run. The x-axis shows the duration of the run, time, in minutes and the y-axis shows the estimated number of bases.

gene block, and AtoB, our largest gene block, each have a relatively high error at 15%. 7DH happens to be relatively close to the actual count, but this is more a result of the presence and absence of the smaller and larger molecules. The other three time periods showed consistent results for measuring all three genes with a maximum percent error for AdhE2 at 18 hours with 9.91%. From this experiment, we concluded that we can quantify molecules on the MinION nanopore with a relatively high level of accuracy and led us to believe that results from any further wastewater quantification experiments on the nanopore would be feasible.

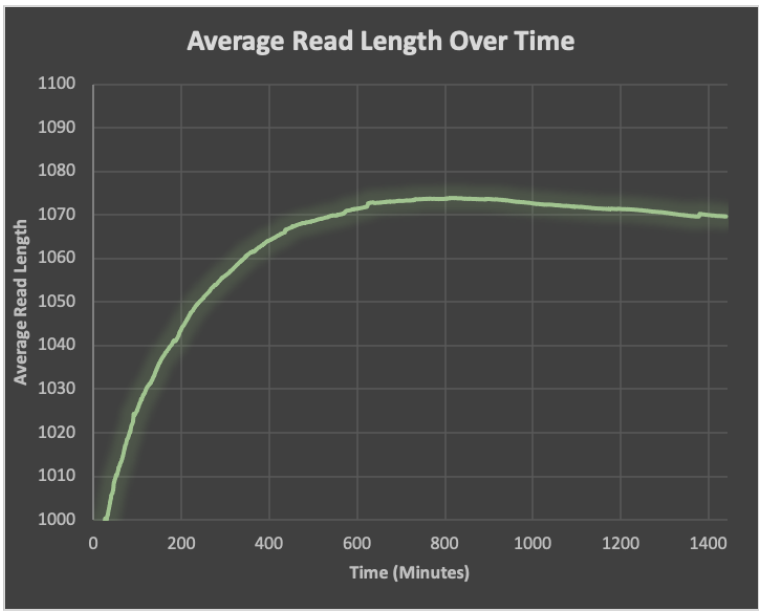

Figure 3.4: Average Read Length Over Time - The above figure shows the average read length from the flow cell over the duration of the Nanopore Run. The x-axis shows the duration of the run, time, in minutes and the y-axis shows the average read length.

3.3 Viral Quantification from Wastewater

Viral cDNA was generated from Nanotrap Particles and Lunacript RT Supermix. The resulting sample was prepped and loaded onto the MinION nanopore. Two unique samples were run. The first sample, taken on May 5th, 2021, was run a week after collection. dPCR confirmed the presence of SCV2 at a concentration of 0.5 $\frac{copies}{\mu L}$, well below the level of detection on the MinION. Regardless, reads were basecalled and aligned to the NCBI database. Overall viral counts were quite low and we also measured an extremely high concentration of human and bacterial

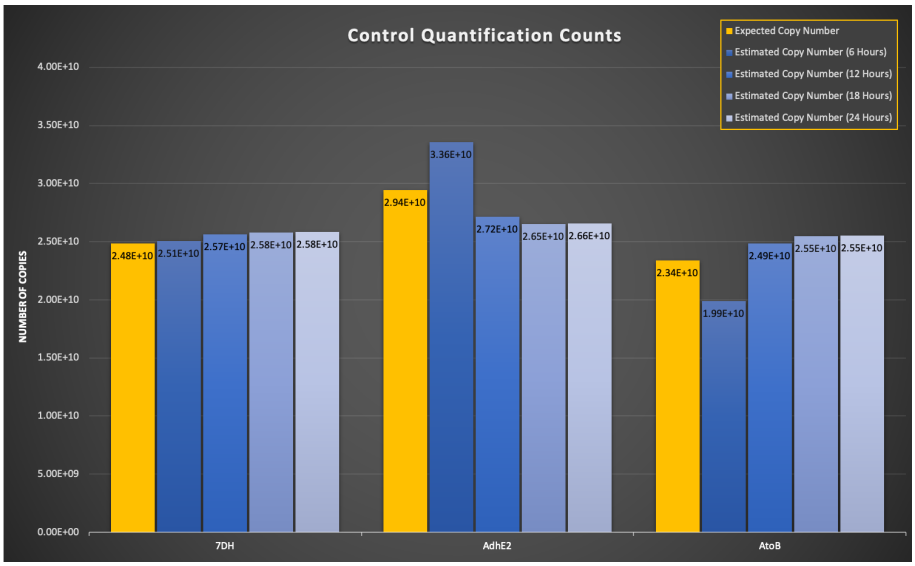

Figure 3.5: Nanopore Control Quantification Counts - The graph above shows the calculated counts of our 3 different control genes compared to the expected at each 6 hour time interval. The expected copy number is displayed in yellow.

Sample Identity	Percent Error (6 hrs)	Percent Error (12 hrs)	Percent Error (18 hrs)	Percent Error (24 hrs)
7DH	0.86%	3.28%	3.78%	4.02%
AdhE2	14.10%	7.77%	9.91%	9.71%
AtoB	14.88%	6.41%	8.97%	9.22%

Figure 3.6: Percent Error of Quantification Measurements - The table above shows the percent error of the nanopore quantification measurements for each gene at 6 hour intervals. The highest percent error collectively is at 6 hours and the lowest is at 12 hours.

DNA, suggesting an error in the extraction and concentration steps. Additionally, since viral counts were quite low, we theorized that this could be due to the fact that the sample has been sitting at 4 °C for an extended period of time, resulting in the loss of a majority of intact virus. The second sample was taken on May

19th, 2021, and was run the following morning. dPCR once again confirmed the presence of SCV2 at a slightly higher concentration of 2 $\frac{copies}{\mu L}$, which is consistent with the concentrations we have been seeing recently with the 4S protocol. The sample was run on the MinION and again, reads were basecalled and aligned to the NCBI database. The 20 most abundant viral populations can be seen below in Figure 3.7.

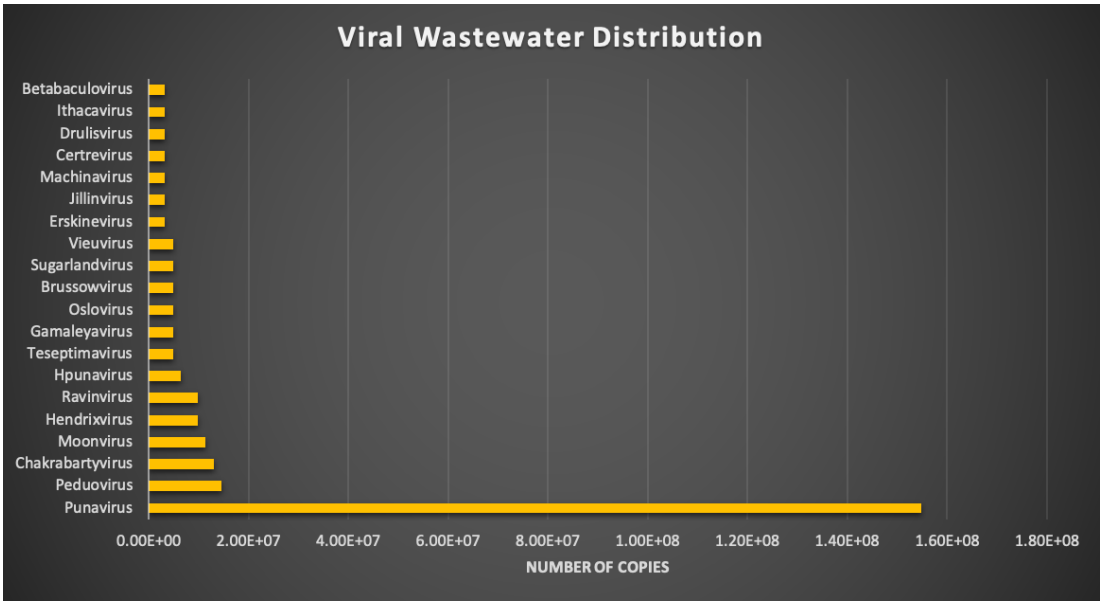

Figure 3.7: Viral Species Counts in Wastewater - The graph above shows the calculated concentration of the 20 most abundant viruses in wastewater. Each species is displayed on the y-axis and the number of copies of each individual species is displayed on the x-axis.

The most abundant virus, by far, in our sample was Punavirus. This virus was almost 10x more abundant compared to any other virus. The full list of viruses can be found in supplemental information. Unfortunately, SCV2 was not

one of the viruses detected. Considering dPCR results estimated that a total of 100 SCV2 molecules were present in our sample before library prep, SCV2 was well below the limit of detection. It's possible we would have been able to detect SCV2 during peak prevalence months in the Watsonville community instead of May. Another explanation, although less likely and has other consequences, is that our method of viral extraction specifically binds and concentrates intact virus. Nanotrap Particles are not able to bind dead virus or floating viral RNA and as such, our quantification results will favor more durable viruses in solution. Although the viral population was overall higher in our newer sample, we still had a large abundance of bacterial and human DNA in our sample. This accounted for roughly 80% of our sample once again suggesting that our viral extraction protocol did not work as intended. Currently, we believe that we are picking up unwanted human and bacterial nucleic acids that are stuck to the side of our falcon tube when resuspending the Nanotrap Particles in PBS. One possible change that might improve this protocol is the introduction of lo-bind tubes in its place to minimize that amount of human and bacterial nucleic acids collected. Although bacterial quantification in wastewater was not the goal of this project, this data is still interesting and worth presenting. The 20 most abundant bacterial populations can be seen below in Figure 3.8.

Human DNA was not included on the graph above but made up a majority of

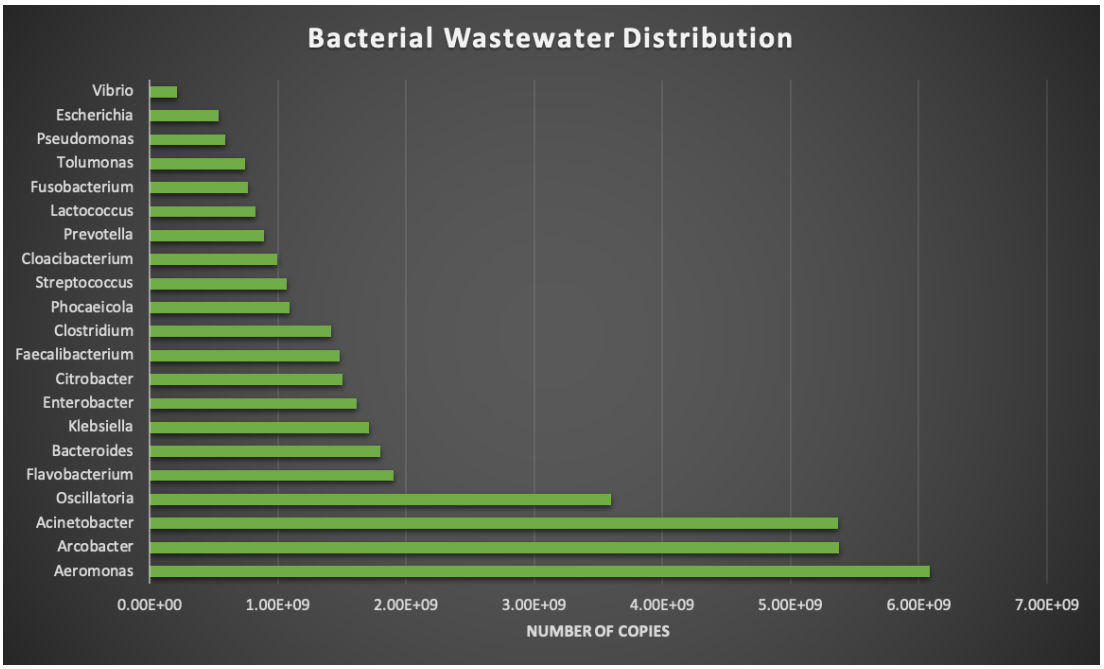

Figure 3.8: The graph above shows the calculated concentration of the 20 most abundant species of bacteria in wastewater. Each species is displayed on the y-axis and the number of copies of each individual species is displayed on the x-axis.

the wastewater sample, roughly 25%. The graph above shows specifically the 20 most abundant bacterial species in the May 19th wastewater composite sample. In this case, *Aeromonas* was the most abundant bacterial species followed by *Arcobacter*, and *Acinetobacter*. Additionally, each bacterial species listed above is more abundant than our most abundant virus, Punaviurs.

Chapter 4

Conclusion and Future Work

While the overall goal of this project to quantify viral populations from wastewater on the MinION was moderately successful, there is still plenty of room for near term improvement. In our first experiment, quantification of gene block controls on the MinION confirmed its ability to act as a limitless multiplex quantification device. This experiment helped validate the quantification results of viral and bacterial species generated on the MinION in the second experiment. Here, we accomplished two major goals of this project despite current issues with viral RNA extraction issues. Unfortunately, our reads do not include SCV2 because the estimated levels are currently well below the limit of detection, especially with our sample not being viral specific. Additionally, due to issues with time and PCR amplification problems, variation detection via ARTIC sequencing was

not completed.

Three suggested future goals for this project include troubleshooting the Viral RNA specific extraction method, introducing a viral-multiplexing method through non-specific amplification, and quantifying viral communities from different sample types. Troubleshooting our extraction method will allow us to redo our viral wastewater nanopore quantification experiment which is currently picking up unwanted human and bacterial nucleic acids. Our current theory is that the introduction of lo-bind tubes will help decrease the amount of human and bacterial nucleic acids significantly. The introduction of a non-specific amplification method for viral multiplexing could also allow us to detect molecules well below the limit of detection, such as SCV2, without compromising quantification accuracy due to PCR prone error. Non-specific amplification of cDNA before library preparation, assuming the sample is limited to viral RNA, would be possible with random hexamers in this case since the sample does not include unwanted nucleic acids. Since cDNA generation and amplification would be uniform and measurable, we will still be able to quantify viral populations with a reasonable level of accuracy. Additionally, while we are testing and introducing this method of quantification for viral populations in wastewater, it would theoretically be possible to quantify viral communities in human samples such as blood or saliva as well. Quantifying viral populations in these samples could have a major impact on the way we di-

agnose and treat viral disease in patients. Considering we have a semi-working method in wastewater, adapting the protocol for different sample types would not be a huge leap either.

Biblography

1. CDC (2020). COVID-19 Cases, Deaths, and Trends in the US — CDC COVID Data Tracker. Centers for Disease Control and Prevention. Available at: https://covid.cdc.gov/covid-data-tracker [Accessed November 27, 2020].

2. CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel Available at: https://www.fda.gov/media/134922/download [Accessed November 27, 2020].

3. CERES Nanosciences (2020). Rapid SARS-CoV-2 Viral Isolation from Wastewater. Available at: https://www.ceresnano.com/post/app-note-rapid-sars-cov-2-viral-isolation-from-wastewater [Accessed April 2, 2021].

4. Community Nanopore (2020). Assessing input DNA (MinION). Available at: https://community.nanoporetech.com/protocols/input- dna-rna-qc/v/idi_s1006_v1_revb_18apr2016/assessing-input-dna [Accessed January 2nd, 2021]

5. de Lannoy, C., de Ridder, D., and Risse, J. (2017). A sequencer coming of age: De novo genome assembly using MinION reads. F1000Res 6, 1083. doi:10.12688/f1000research.12012.1

6. Deamer, D., Akeson, M., and Branton, D. (2016). Three decades of nanopore sequencing. Nat Biotechnol 34, 518–524. doi:10.1038/nbt.3423.

7. Kim, D., Lee, J.-Y., Yang, J.-S., Kim, J. W., Kim, V. N., and Chang, H. (2020). The Architecture of SARS-CoV-2 Transcriptome. Cell 181, 914-921.e10. doi:10.1016/j.cell.2020.04.011.

8. Mallapaty, S. (2020). How sewage could reveal true scale of coronavirus outbreak. Nature 580, 176–177. doi:10.1038/d41586-020-00973-x.

9. Morley, A. A. (2014). Digital PCR: A brief history. Biomolecular Detection and Quantification 1, 1–2. doi:10.1016/j.bdq.2014.06.001.

10. Nanoporetech/bonito (2020). Oxford Nanopore Technologies Available at: https://github.com/nanoporetech/bonito [Accessed November 27, 2020].

11. Nanoporetech/ont_fast5_api (2020). Oxford Nanopore Technologies Available at: https://github.com/nanoporetech/ont_fast5_api [Accessed November 27, 2020].

12. Quick, J. (2020). nCoV-2019 sequencing protocol v3 (LoCost). Avail-

able at: https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye [Accessed May 24, 2021].

13. Rozevsky, Y., Gilboa, T., van Kooten, X. F., Kobelt, D., Huttner, D., Stein, U., et al. (2020). Quantification of mRNA Expression Using Single-Molecule Nanopore Sensing. ACS Nano 14, 13964–13974. doi:10.1021/acsnano.0c06375.

14. Santa Cruz Health., COVID-19 Forecast Models Available at: https://www.santacruzhealth.org/HSAHome/HSADivisions/PublicHealth/CommunicableDiseaseControl/CoronavirusHome/LocalCOVID-19ForecastModels.aspx [Accessed November 27, 2020a].

15. Thermofisher.,Trizol Reagent Guide and RNA Extraction: Available at: http://tools.thermofisher.com/content/sfs/manuals/trizol_reagent.pdf.

16. Tyson, J. R., James, P., Stoddart, D., Sparks, N., Wickenhagen, A., Hall, G., et al. (2020). Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. bioRxiv, 2020.09.04.283077. doi:10.1101/2020.09.04.283077.

17. UCSC., Tracking COVID-19 Roadmap to Recovery. Available at: https://recovery.ucsc.edu/reporting-covid/covid-tracking/.

18. Whitney, O. (2020). V.3 - Direct wastewater RNA capture and purification via the 34;Sewage, Salt, Silica and SARS-CoV-2 (4S)" method. doi:10.17504/protocols.io.bngsmbwe.

19. Wick, R. (2019). Training data. 118132963659 Bytes. doi:10.26180/5C5A5F5FF20ED.

20. Zhang, T., Breitbart, M., Lee, W. H., Run, J.-Q., Wei, C. L., Soh, S. W. L., et al. (2005). RNA Viral Community in Human Feces: Prevalence of Plant Pathogenic Viruses. PLOS Biology 4, e3. doi:10.1371/journal.pbio.0040003.

Appendix A

Supplementary Tables and Figures

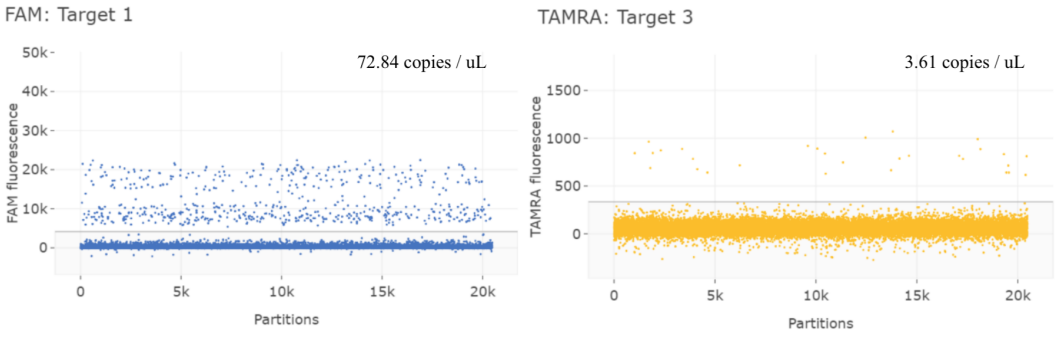

Figure A.1: January Digital PCR Wastewater Surveillance Results - The panels show dPCR results from a wastewater sample collected in January 2021. The graph on the left shows the results for the N gene of SCV2 with an estimated concentration of 72.84 $\frac{copies}{\mu L}$. The graph on the right shows the results for the RAP region of PMMV with an estimated concentration of 3.61 $\frac{copies}{\mu L}$. In both graphs, the X-axis represents the partition number on the plate and on the Y-axis, the fluorescence of the partition.

Usage	Oligo Name	Sequence
SCV2 N Gene (5' end)	$nCOV_{N1}$ Forward	GACCCCAAAATC AGCGAAAT
	$nCOV_{N1}$ Forward	TCTGGTTACTGCC AGTTGAATCTG
	$nCOV_{N1}$ Probe	FAM-ACCCCGCATT ACGTTTGGTGGACC
SCV2 N Gene (3' end)	$nCOV_{N2}$ Forward	TTACAAACATTGG CCGCAAA
	$nCOV_{N2}$ Reverse	GCGCGACATTCCG AAGAA
	$nCOV_{N2}$ Probe	HEX-ACAATTTGCCC CCAGCGCTTCAG
PMMV RAP	PMMV Forward	GAGTGGTTTGACCT TAACGTTGA
	PMMV Reverse	TTGTCGGTTGCAAT GCAAGT
	PMMV Probe	Cy5-CCTACCGAAGCA AAT
BCoV Spike In	BCoV Forward	CTGGAAGTTGGTG GAGTT
	BCoV Reverse	ATTATCGGCCTAAC ATACATC
	BCoV Probe	TAM-CCTTCATATCTA TACACATCAAGTT

Table A.1: The table above shows the sequences and names of the different primer and probe combinations used in the dPCR wastewater quantification assay.

Species	# Molecules	Species	# Molecules	Species	# Molecules	Species	# Molecules
Punavirus	1.55E+08	Certrevirus	4.08E+06	Slopekvirus	2.28E+06	Moineauvirus	1.63E+06
Peduovirus	1.47E+07	Drulisvirus	3.59E+06	Tequatrovirus	2.20E+06	Nickievirus	1.63E+06
Chakrabartyvirus	1.30E+07	Ithacavirus	3.42E+06	Tulanevirus	2.20E+06	Pepyhexavirus	1.63E+06
Moonvirus	1.14E+07	Betabaculovirus	3.26E+06	Winklervirus	1.96E+06	Septimatrevirus	1.63E+06
Hendrixvirus	9.94E+06	Eneladusvirus	3.18E+06	Friunavirus	1.87E+06	Sextaecvirus	1.63E+06
Ravinvirus	9.78E+06	Fletchervirus	3.10E+06	Przondovirus	1.79E+06	Yuavirus	1.63E+06
Hpunavirus	6.52E+06	Flaumdravirus	3.02E+06	Enquatrovirus	1.79E+06	Cyprinivirus	1.63E+06
Teseptimavirus	6.19E+06	Metrivirus	2.93E+06	Ledergergvirus	1.79E+06	Muromegalovirus	1.63E+06
Gamaleyavirus	6.03E+06	Mieseafarmvirus	2.85E+06	Uetakevirus	1.79E+06	Ranavirus	1.63E+06
Oslovirus	5.71E+06	Felixounavirus	2.77E+06	Bronvirus	1.79E+06	Marseillevirus	1.63E+06
Brussowvirus	5.54E+06	Phikzvirus	2.69E+06	Ceetrepovirus	1.71E+06	Cafeteriavirus	1.63E+06
Sugarlandvirus	5.22E+06	Popoffvirus	2.61E+06	Cetovirus	1.71E+06	Mimivirus	1.63E+06
Vieuvirus	4.89E+06	Risingsunvirus	2.53E+06	Eyrevirus	1.71E+06	Prasinovirus	1.63E+06
Erskinevirus	4.56E+06	Tegunavirus	2.45E+06	Jerseyvirus	1.71E+06	Cervidpoxvirus	1.63E+06
Jillinvirus	4.40E+06	Gelderlandvirus	2.28E+06	Kojivirus	1.63E+06	Alphaentomopoxvirus	1.63E+06
Machinavirus	4.24E+06	Karamvirus	2.28E+06	Lamdavirus	1.63E+06	Betatectivirus	1.63E+06

Figure A.2: The table above shows the full list of viruses identified and quantified during the viral wastewater quantification experiment. Viral species with a molecule count around 163 E+06 were just above the minimum limit of detection of the MinION.