# MONITORING THE SPREAD OF HUMAN RESPIRATORY VIRUSES: INTEGRATING WASTEWATER-BASED EPIDEMIOLOGY AND TARGET CAPTURE SEQUENCING

by

Lauren Roppolo Brazell

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Bioinformatics and Computational Biology

Charlotte

2024

Approved By:

_____
Dr. Cynthia Gibas

_____
Dr. Jessica Schlueter

_____
Dr. Anthony Fodor

_____
Dr. Alex Dornburg

_____
Dr. Alan Dow

# ABSTRACT

LAUREN ROPPOLO BRAZELL. Monitoring the Spread of Human Respiratory Viruses: Integrating Wastewater-Based Epidemiology and Target Capture Sequencing.
(Under the direction of DR. CYNTHIA GIBAS)

Wastewater-based epidemiology (WBE) has emerged as a valuable tool for monitoring the spread of human respiratory viruses, particularly in the context of the COVID-19 pandemic. By bypassing traditional clinical testing, WBE can serve as an early indicator for viral outbreaks, enabling communities to make informed public health decisions. While WBE has been primarily used for SARS-CoV-2, its potential extends to other human respiratory viruses, including influenza A and B, and respiratory syncytial virus. In this dissertation, we implemented a next-generation sequencing protocol to assess human respiratory virus RNA in both wastewater and nasopharyngeal swabs that PCR tested negative for SARS-CoV-2. Control mixtures containing synthetic human respiratory virus RNA were spiked into wastewater and nuclease-free water to evaluate any matrix effects on sequencing outcomes. Bioinformatics analyses used taxonomic classification and direct alignment methods to compare the accuracy of human respiratory virus identification in wastewater and clinical samples. Despite the potential of commercial next generation sequencing-based target-capture assays to detect viral genera, sequencing results from both wastewater and clinical samples demonstrated low depth and breadth of coverage, with discordant outputs from different bioinformatics pipelines. These findings highlight the need for rigorous benchmarking of laboratory and computational methods to ensure accurate human respiratory detection in wastewater and suggest that current sequencing approaches may fall short in providing the strain-specific information required for detailed public health surveillance.

# DEDICATION

To my parents, for encouraging me to chase every single one of my dreams. And to my husband Paul, for loving me while I chased this one.

# ACKNOWLEDGEMENTS

I would like to express my most sincere gratitude to my committee members: Dr. Cynthia Gibas, Dr. Jessica Schlueter, Dr. Anthony Fodor, Dr. Alex Dornburg, and Dr. Alan Dow. Without their support, this dissertation would not have been possible. I am grateful for my coworkers-turned-friends from the COVID-19 Wastewater Lab for their help with any and everything. April Harris, Isaiah Young, and Gari New, so much of this work could not have happened without your help in sample processing and sequencing. Samuel Kunkleman and Kevin Lambirth, the bioinformatics guidance and technical reviews you've provided me with have been instrumental in my success. Lolo Aboufoul, I am so thankful for your help with the Tripledemic project both on and off the bench. I am so proud of everything you have accomplished! Dr. Jannatul Ferdous, seeing you finish your Ph.D. with grace and humility was so inspiring and gave me the strength to continue with mine. You are the most radiant, wonderful person, and I am so lucky to have gotten to experience this chapter of my academic life as your friend! To my Lab Dad, William Taylor, your friendship and mentorship has been the biggest gift of all. I will forever be grateful to you for guiding me through wins, losses, and everything in between- with a joke and a smile. I would not be where I am today without you, and a thanks will never be enough. Last but not least, my support system: my family, friends, and my precious Tina, I love you all beyond measure. Thank you for everything.

TABLE OF CONTENTS

LIST OF FIGURES

# LIST OF ABBREVIATIONS

HRV: human respiratory viruses.
SARS-CoV-2: severe acute respiratory syndrome coronavirus 2.
RI: respiratory illness.
DNA: deoxyribonucleic acid.
RNA: ribonucleic acid.
mRNA: messenger ribonucleic acid.
PCR: polymerase chain reaction.
RT-qPCR: reverse transcriptase quantitative polymerase chain reaction.
ddPCR: droplet digital polymerase chain reaction.
WBE: wastewater-based epidemiology.
COVID-19: coronavirus disease 2019.
bp: base pairs.
kb: kilobase pairs.
VOC: variants of concern.
HRnV: human rhinovirus.
HAdV: human adenovirus.
HPIV: human parainfluenza virus.
hMPV: human metapneumovirus.
RSV: respiratory syncytial virus.
vRNP: viral ribonucleoprotein.
ARDS: acute respiratory distress syndrome.
HCoV: human coronavirus.
NP: nasopharyngeal.
NAAT: nucleic acid amplification test.
NGS: next-generation sequencing.
dNTP: deoxynucleoside triphosphate.
ONT: Oxford Nanopore Technologies.
PreMiEr: Precision Microbe Engineering.
Mpox: monkeypox virus.
CDC: Centers for Disease Control.
NWSS: National Wastewater Surveillance System.
LoD: limit of detection.
rRNA: ribosomal ribonucleic acid.
WMGS: whole-metagenome shotgun sequencing.
RVOP: Respiratory Viral Oligonucleotide Panel (Illumina, Inc.).
ng: nanograms.
cDNA: complementary deoxyribonucleic acid.
dA-tailing: polyA-tailing.
ssRNA: single-stranded ribonucleic acid.
dsDNA: double-stranded deoxyribonucleic acid.
HS: high-sensitivity.
pM: picoMolar.
QC: quality control.
cp: copies.

µL: microliter.
mL: milliliter.
Q30: quality score 30.
SISPA: sequence independent single-primer amplification.
Gbp: giga base pairs.
WW: wastewater.
CN: clinical negative.
MERS: Middle East respiratory syndrome.
SARS: severe acute respiratory syndrome.
NFW: nuclease-free water.
NCDHHS: North Carolina Department of Health and Human Services.
O/E: observed vs. expected.

# CHAPTER 1: INTRODUCTION

Even though SARS-CoV-2 is still the dominant pathogen of concern worldwide, a variety of other human respiratory viruses continue to pose a threat to the general public. Since respiratory viruses are transmitted easily from person to person and cause varying degrees of respiratory illness, finding a way to effectively monitor the spread of these viruses and alert the public to take precautions for their overall health is critical. Molecular testing by PCR-based methods can reliably be used to detect specific viruses in clinical samples and to monitor levels of a few of the most common respiratory viruses in wastewater, but it does not scale to more than a few specific targets per reaction. Next-generation sequencing methods have the potential to allow testing for the presence of a wide variety of respiratory pathogens at once in a way that is simple and minimally invasive. However, shotgun metagenomic sequencing of samples produces large volumes of data that, while interesting, may not be useful for public health systems. Intermediate between these two strategies are methods that capture a fraction of interest from the original sample, whether by filtration, sequestration with magnetic beads, or hybridization-based target capture. We tested the potential of target capture-based sequencing methods to identify prime viral suspects of respiratory illness at the community level via wastewater, and at the individual level in COVID-19 negative clinical tests. While hybridization-capture protocols are promising for identification of respiratory virus targets at a species level, they are not capable of providing the variant-level identification that can be obtained from targeted sequencing protocols such as the ARTIC protocol which was widely used in SARS-CoV-2 genomic surveillance.

## 1.1 Overview and Objectives

While recent human respiratory virus (HRV) research has placed a strong emphasis on monitoring and slowing the spread of SARS-CoV-2, HRVs of many different kinds are still circulating and causing widespread respiratory illnesses (RIs). Infectious HRV particles are highly communicable from person to person through exhaled aerosolized droplets (Wang et al., 2021). Individuals who contract RIs often present with symptoms like coughing and sneezing that facilitate transmission (Subbarao & Mahanty, 2020), however, depending on the virus, an asymptomatic individual can still exhale infectious viral particles via talking and breathing (Scheuch, 2020). Given the nature of their spread, HRVs can quickly spread through communities and cause varying degrees of sickness, hospitalizations, and in some cases, death (Tregoning & Schwarze, 2010).

HRVs responsible for RIs span six major families, including adenoviruses, influenzaviruses, pneumoviruses, paramyxoviruses, enteroviruses, and coronaviruses (Van Elden et al., 2002). These families can then be further characterized by their genome type. Viruses within the adenovirus family have non-enveloped, double-stranded DNA genomes ranging from 26-45 kb in length (Kajon et al., 2019). The remaining viral families have single-stranded RNA genomes, which can be further subgrouped by the type of RNA that serves as the genome (Payne, 2017). Enteroviruses and coronaviruses both consist of positive-stranded RNA genomes that form functional mRNAs, though enteroviruses differ from coronaviruses in that they lack enveloped capsids (Modrow et al., 2013). Influenzaviruses, pneumoviruses, and paramyxoviruses consist of negative-stranded RNA genomes encapsulated by a nucleoprotein (Arbeitskreis Blut, 2009). Single-stranded RNA viruses are very diverse in size, ranging from 4.5-30 kb in length (Lindenbach, 2022).

Though their genetic makeup is diverse, most HRV infections present with similar symptoms and often cannot be differentiated empirically in the clinic (Leung, 2021). Multiplex molecular testing methods including RT-qPCR and ddPCR are commonly used to detect specific pathogen(s) in nasopharyngeal swab samples, which is important for health care providers choosing among available treatments (Jiang et al., 2022). Clinical tests are often sought only if an individual is symptomatic, and not all viruses are targets for testing and antiviral treatment. Clinical testing can therefore provide only an incomplete picture of viral prevalence.

Early in the SARS-CoV-2 pandemic, it was determined that it was possible to to quantify viral material being shed from a population by testing community wastewater, using similar molecular detection methods as used in clinical testing (Kwong et al., 2015). Prior to the SARS-CoV-2 pandemic wastewater surveillance had been used in a limited way to track other viral infections such as Zika virus and polio (Ryerson et al., 2022; Wong et al., 2024). The key observation that led to widespread adoption of WBE for SARS-CoV-2 surveillance was that detectable copies in community wastewater was a leading indicator of waves of infection, and could therefore inform public health officials of increasing incidence of infection faster than clinical testing or metrics such as emergency room visits. . In addition to community-level surveillance, WBE was widely adopted during the COVID-19 pandemic for localized populations living in communal situations such as hospitals, jails, schools, and universities. The WBE approach used at the University of North Carolina at Charlotte in 2020 showed that building-level early detection/alert systems are effective as a first step in identifying infected students on campus in order to prevent further spread of SARS-CoV-2 (Gibas et al., 2021).

A public health surveillance, rather than clinical diagnostic, approach to respiratory samples provides information about circulating viruses that are driving waves of respiratory

infection, and both discarded clinical samples and wastewater samples can be used as material for a public health surveillance approach. While PCR-based testing provides accurate detection of viral copies, target multiplexing in ddPCR and qPCR is limited to 4-6 targets, and is limited in its ability to distinguish circulating viral strains or detect unknowns.

Next-generation sequencing has been widely used during the COVID-19 pandemic to provide strain level information about circulating SARS-CoV-2, which has continued to mutate rapidly since its emergence in 2019. The virus tends to mutate in ways that increase its transmissibility and partially evade immunity previously acquired through vaccination or infection. It is desirable to monitor the details of changes in the viral genome sequence to identify new emerging strains that may require updates to existing vaccines or public health strategies. Influenza has been similarly closely tracked, with a large collection of sequence data collected annually to inform updates to the annual influenza vaccine. However, systematic sequencing of other viral genomes is not widespread (Dahui, 2019).

Second-generation sequencing technologies like those from Illumina Inc. (San Diego, CA) have been trusted by scientists and clinicians for over twenty years. Illumina's technologies have made marked improvements on first-generation sequencing strategies, generating high-accuracy whole-genome sequence information quickly and efficiently (Kwong et al., 2015). The Illumina sequencing technology is constrained by short read lengths between 50-300 bp at a maximum (Profaizer et al., 2015). These short reads are often unable to span any repetitive structures or GC-rich regions in a genome that extend beyond the upper limits of the read length, resulting in a fragmented genome assembly with "holes" in the areas that it cannot resolve by sequencing (Z. Chen et al., 2020). In order to overcome this obstacle, Oxford Nanopore Technologies (Oxford, UK), developed a third-generation sequencing method that allowed for

increased read lengths in excess of 10 kb (Pollard et al., 2018). With the improvements in read length, though, came an increase in error rate. In comparison to Illumina's < 0.5%, Oxford Nanopore's reads have error rates between 10% and 15% (H. Lu et al., 2016). Both Illumina and Oxford Nanopore sequencing methods are viable options that can be used to monitor HRV's in both clinical and wastewater samples.

UNC Charlotte's SARS-CoV-2 surveillance effort used a tiling amplicon sequencing protocol to track SARS-CoV-2 and its variants on campus and within the Charlotte community (Lambisia et al., 2022). However, implementation of tiling amplicon genomic surveillance is not implemented for the majority of viruses that cause human respiratory illness, and such sequencing is challenging to implement for multiple targets in parallel. More research is needed to explore the possibility of agnostic or broad-range sequencing assays that will expand the reach of surveilled HRVs beyond SARS-CoV-2 and influenza. Shotgun microbiome sequencing, while it is a completely agnostic strategy, generates large volumes of data which may be tangential to the goals of public health agencies. In this study, we investigate the practicality of using a more focused, but still broad assay. Target capture protocols use hybridization-based pulldown to capture viral genomic material of interest, making it possible to detect and sequence dozens or hundreds of common circulating HRVs. The Twist assay explored in this study is designed to capture 41 common circulating HRVs. The overall objective of this project is to keep the public healthy and informed about infection risk from HRVs using a broad-based viral assay that can be applied to wastewater, thus minimizing negative societal impacts. This study is designed to achieve this, broadly, by the following:

1) Test and develop SOP's for viral target capture-based sequencing from clinical and environmental samples using Twist Bioscience's Respiratory Viral Panel.

2) Benchmark the performance of the Twist assay in wastewater using known virome spike-in mixtures.
3) Develop an applied bioinformatics pipeline for analysis of virome data generated using both types of sequencing procedures.

## 1.2 Background

Respiratory infections (RIs) are among the most common diseases of the respiratory tract globally (Derbyshire & Calder, 2021). Most often, they affect the mucosa of the upper airway in a self-limiting manner and resolve without medical intervention (Van Doorn & Yu, 2020). These infections can sometimes remain entirely asymptomatic, or present with various symptoms like sore throat, runny nose, coughing and sneezing, which can easily be addressed with over-the-counter remedies (National Health Services, n.d.; Van Doorn & Yu, 2020). However, some infections descend into the lower airway where they pose a larger threat of acute illness and impair normal respiration (Van Doorn & Yu, 2020). These acute infections will sometimes develop into severe syndromes including bronchiolitis, asthma, and pneumonia, and require medical intervention to treat (Pavia, 2011).

Of the possible causative agents of RI, human respiratory viruses (HRVs) pose the most critical threat as their modes of transmission make them especially communicable from person to person. Respiratory viruses are transmitted directly via physical contact between an infected individual and a susceptible individual, indirectly via contact with contaminated surfaces or objects, or directly through the air from one respiratory tract to another via large respiratory droplets or via fine respiratory aerosols (Leung, 2021). Many precautions have been suggested to prevent the spread of HRVs from person to person including frequent hand-washing, surface sanitization, isolation of likely cases, and wearable barriers (gloves, masks, etm.) (Jefferson et al., 2020). These options are effective in the case of symptomatic RIs, but an additional layer of complexity is added when a person is infected by a HRV and shows no symptoms. Wastewater based surveillance samples from populations regardless of whether or not they actively seek medical care, and can provide a window into the circulation of viral families that are

asymptomatic or transmissible prior to the emergence of symptoms, as well as viruses that commonly cause mild to moderate illness but do not necessarily result in a visit to a healthcare provider. Of interest are six major families of viruses: enteroviruses, adenoviruses, pneumoviruses, paramyxoviruses, influenzaviruses, and coronaviruses (Van Elden et al., 2002).

Enteroviruses are members of the family *Picornaviridae*. They are small, positive-stranded RNA viruses with genomes that range from 7.2-8.5 kb in length (Chien et al., 2019). This family includes the human rhinoviruses (HRnVs), of which subspecies A and B account for more than 50% of "common cold" cases (Blaas & Fuchs, 2016). HRnVs A and B are structurally similar, consisting of a non-enveloped protein capsid containing their RNA genome that enters the host cell by way of receptor-mediated endocytosis (Fuchs & Blaas, 2010). Following entry into the cell, the virus releases its genetic material into the cytoplasm and begins translation, replication, and assembly of new virions for release from the host cell (Kerr et al., 2021). If not entirely asymptomatic, symptoms generally remain mild and common cold-like for most individuals, however, some infections can exacerbate existing respiratory ailments and progress into more severe RIs that require medical attention (Turner & Lee, 2009).

Adenoviruses are members of the family *Adenoviridae*. They are medium-sized, non-enveloped, linear, double-stranded DNA viruses with genomes that range from 26-45 kb in length (Davison et al., 2003). This family includes the mastadenovirus genus that comprises the mammalian adenoviruses, within which there are seven species (A-G) and over 60 serotypes of human adenoviruses (HAdVs) (Hoeben & Uil, 2013). The general structure of these viruses consist of the icosahedral nucleocapsid shell and the inner, genome-containing core (Shieh, 2022). The nucleocapsid's composition allows HAdVs to bind to epithelial cell surface receptors of the airway, triggering internalization by endocytosis, virion escape, and finally, delivery of the

genome to the nucleus via the nuclear pore (Nestić et al., 2021). Following nuclear delivery, viral mRNA is transcribed, genomic material replicated, and progeny virion particles assembled just before the infected cell lyses to release them and proliferate infection (Shieh, 2022). Upon infection, individuals will either be asymptomatic or present with a range of common cold-like respiratory symptoms including cough, fever, runny nose, and sore throat (Khanal et al., 2018). Occasionally, some infections produce symptoms that manifest as conjunctivitis or gastrointestinal upset (Rajaiya et al., 2021). Most HAdV infections are mild enough to avoid significant disruption to an individual's daily life and, as such, have little difficulty spreading quickly from person to person in closed/crowded settings (Binder et al., 2017).

The *Paramyxoviridae* family is inclusive of both the *Paramyxovirus* and the *Pneumovirus* genera. Both genera within this family are enveloped, linear, negative-stranded RNA viruses with genomes of about 13-15 kb for pneumoviruses, and 15-19 kb for paramyxoviruses (Cifuentes-Muñoz & Ellis Dutch, 2019; Nayak et al., 2008). These viruses share similarities in their general structure, consisting of an outer attachment protein-bearing envelope, and an inner RNA genome-containing core (Farrukee et al., 2020). These viruses bind to the surface of epithelial cells that line the airway, where they trigger internalization primarily by direct fusion to the plasma membrane, though some forms of this virus are able to enter the host cell via an endocytic mechanism (Cox et al., 2015; Smith et al., 2009). Upon entry into the cell, the nucleocapsid delivers the RNA genome into the cytoplasm, where it is then transcribed, assembled into new virions, and exported from the host cell to proliferate infection (El Najjar et al., 2014). One of the most common human pathogens from the *paramyxovirus* genera is the parainfluenza virus (HPIV), types of which are known to cause respiratory symptoms including fever, sore throat, and cough (Branche & Falsey, 2016). In adults, these symptoms are generally

mild, however, in infants and children, symptoms can manifest as a severe barking cough known as croup (Branche & Falsey, 2016). In the *pneumovirus* genera, common pathogens of note are the human metapneumovirus (hMPV) and respiratory syncytial virus (RSV). Particularly, RSV has gained global recognition as of late, spreading like wildfire among children and sometimes causing severe respiratory illness (Hernández-Rivas et al., 2023). Generally, RSV manifests with classic common cold-like symptoms, but can progress quickly to medical emergencies including wheezing, pneumonia, and acute respiratory failure (Hanish Jain et al., 2023). RSV, coupled with flu and COVID-19 infections, begets what is known as a "triple epidemic" of RIs in the United States (Tanne, 2022).

Influenzaviruses are members of the family *Orthomyxoviridae*. They are small, segmented, negative-stranded RNA viruses with genomes around 13.5 kb in length (Bouvier & Palese, 2008; Talha N. Jilani et al., 2024). Within this family are two genera clinically relevant to humans, A and B (Arbeitskreis Blut, 2009). Like many other viruses, the influenzavirus' genome is contained within a protein-based nucleocapsid, allowing it to bind to the surface of the respiratory epithelium (Sanders et al., 2011; Suzanne Clancy, 2008). Within this nucleocapsid are viral ribonucleoprotein (vRNP) complexes that contain "packages" of each of the 8 RNA segments and an RNA-dependent RNA polymerase required for transcription (Zhu et al., 2023). Similar to other viruses, infection begins when the virus is bound to a cell-surface receptor and enters the cell via an endocytic event. The vRNPs are then released into the cytoplasm of the host cell and delivered to the nucleus, where they begin the replication process (Dou et al., 2018). Viral mRNA is then used to make viral proteins which are then assembled into new viral particles and exported from the cell, thus proliferating the infection further (*Biology for Majors II: Steps of Virus Infections*, n.d.). Upon infection, the hallmark symptoms of an influenza A or B

infection generally present as a fever, but soon give rise to head and body aches, fatigue, and cough (Arbeitskreis Blut, 2009). Infections cause symptoms that range anywhere from mild to moderate, but sometimes progress into the severe category when normal respiration is compromised. Influenza-related pneumonia, acute respiratory distress syndrome (ARDS), and secondary bacterial infections can become emergencies that require immediate medical attention, especially for children and the elderly (Kalil & Thomas, 2019). Akin to other RIs, influenzaviruses spread notoriously quickly among groups of people that spend significant amounts of time together. Additionally, their propensity for mutation makes them likely suspects for pandemic-level spread.

Mutation rates are extremely high for the influenza A virus, and slightly less for the influenza B virus (Nobusawa & Sato, 2006). Thus, exploring mechanisms of mutation in influenza A offers some insight into exactly how and why pandemic-level spreads of this strain have happened throughout the years. Influenza A differs from influenza B in that it is zoonotic in nature. The natural reservoir for this strain exists in aquatic avian species, but this virus has adapted to infect several other types of animals including swine, domestic poultry, dogs, horses, and many others (Harrington et al., 2021). It is likely to have bridged the gap between humans and animals via direct contact with them, and, as such, has been documented as the likely cause of the “Spanish Flu” (Mostafa et al., 2018). In 1918, an outbreak of poultry-borne influenza A virus was responsible for a deadly pandemic that affected approximately 40% of the human population, touting a death rate just above 2% (Carter & Sanford, 2012). This particular strain of virus was named “H1N1” for the hemagglutinin and neuraminidase antigen proteins that line its nucleocapsid, and this nomenclature remained in effect to describe subsequent influenza strains (Pineo, 2021). While time and the advent of mainstream vaccines nearly two decades after the

Spanish Flu pandemic caused a slow in spread and mortality of the original virus, it didn't take long for the virus to "shift" and "drift", accumulating mutations that allowed it to reach pandemic-level infection rates in humans once again (Barberis et al., 2016). The next influenza A pandemic, H2N2, emerged in the late 1950s as a result of a "shift" of the H and N genes from an avian reservoir, and an accumulation of minor "drifting" point mutations that allowed it to easily evade immune responses of people worldwide (Kim et al., 2018). Again, the passing of time and new technologies applied to vaccine development helped resolve the pandemic, but it was not long before H2N2 shifted into H3N2, causing another influenza pandemic over a decade later in late 1960 (Flahault & Zylberman, 2010). The most recent influenza pandemic of note was in 2009, caused by a quadruple reassortment of the swine-originating H1N1 (Al Hajjar & McIntosh, 2010). Owing to extensive research and improvements in preventative measures and medical treatments, these flu pandemics have been less extreme over time, but still caused mild to moderate illness in a large number of people, and severe illness in some. At present, seasonal influenzavirus infections continue to burden people at varying levels worldwide, but a much greater emphasis has been placed on the ongoing coronavirus pandemic.

Coronaviruses are members of the family *Coronaviridae*. They are large, enveloped, positive-stranded RNA viruses with genomes that range from 25-32 kb in length (Payne, 2017). Within this family exists the *Coronavirinae* subfamily, comprising the human coronaviruses (HCoVs), which are notable pathogens in RIs worldwide. The seven identified HCoVs include HCoV-229E, HCoV-OC43, HCoV-NL63, HCoV-HKU1, SARS-CoV, MERS-CoV, and SARS-CoV-2. HCoV-229E, HCoV-OC43, HCoV-NL63, and HCoV-HKU1 are known to cause the common cold, whereas the severe acute respiratory syndrome coronavirus (SARS-CoV) or the Middle East respiratory syndrome coronavirus (MERS-CoV) cause more severe respiratory

disturbances with relatively high mortality rates (B. Chen et al., 2020). All seven of these HCoVs have zoonotic origins, primarily from bats, but also include mice and other domestic animals as well (Z.-W. Ye et al., 2020). Structurally, HCoVs are made up of an outer envelope laden with spike-shaped surface proteins, and their inner core consists of their nucleocapsid-bound genome (D. X. Liu et al., 2021). These viruses use their spike-shaped surface proteins to bind to and enter the epithelial cells of the airway, upon which they release their RNA genome into the cytoplasm (V'kovski et al., 2021). The RNA is then translated and transcribed, assembled into virions, and released from the host cell to continue infection (Brant et al., 2021). All HCoV infections proliferate quickly throughout the body and cause varying degrees of respiratory symptoms that range from mild to moderate common-cold like symptoms. On the moderate to severe end of the spectrum, and of particular concern at the moment, of course, is SARS-CoV-2.

Since its initial discovery in late 2019, SARS-CoV-2, the virus responsible for the ongoing COVID-19 pandemic, continues to affect the lives of people across the globe. The World Health Organization's COVID-19 dashboard reports nearly 650,000,000 cumulative cases, and 6,645,812 deaths worldwide as of December 2022 (*WHO COVID-19 Dashboard*, 2024). Its spread to pandemic-level status is to multiple factors, including its novelty, which meant that much of the world's population was effectively immunologically naive when the virus emerged, the ease of transmission via respiratory droplets and the high average number of secondary infections caused by a single infected individual, $R_0$ (D'Arienzo & Coniglio, 2020). Like other HCoV infections, COVID-19 is sometimes asymptomatic, but most infections present with mild symptoms including fever, headache, shortness of breath, cough, and fatigue and are often resolved without medical intervention (Yuen et al., 2020). However, some severe infections do occur and give rise to pneumonia, acute respiratory distress syndrome (ARDS), and respiratory

failure (Budinger et al., 2021). Frequent mutations allow the virus to evade recently-acquired immune responses.The spike protein in particular has undergone individual point mutations at its receptor binding domains, allowing the virus to partially escape immune response (Cosar et al., 2022). Since 2019, there have been about 4,000 mutations in the spike protein gene, several of which have generated VoCs that are more easily transmissible than the original strain (Cosar et al., 2022).

Molecular testing for SARS-CoV-2 infection became widespread during the early years of the pandemic, during which time other common respiratory viruses were suppressed along with COVID-19 by public health measures. After the end of mitigations such as social distancing and masking, viruses such as influenza and RSV returned to more normal levels, and clinical tests which target SARS-CoV-2, influenza, and RSV in parallel are increasingly common.. A diagnostic test for a RI usually involves the collection of a nasopharyngeal (NP) swab that will then be subjected to a nucleic acid amplification test (NAAT) for the detection of one or more respiratory pathogens (Park, 2021). While generally NAATs are an accurate and cost-effective approach to individual testing, this cost is not sustainable for widespread asymptomatic surveillance of populations. Their performance can degrade if a patient's viral load is below the detectable limit, or viral sequence variants differ from the assay's primers and probes (Ginocchio & McAdam, 2011). In these cases, a patient who is actually "positive" for a RI from a certain virus could receive a false "negative" diagnosis. Asymptomatic patients may not seek NAATs at all, continuing to unknowingly shed viral particles that can infect others they come into contact with as they continue their daily lives.

Respiratory viruses are known to be detectable via extrapulmonary specimens, including wastewater (Sears et al., 1986; Zafeiriadou et al., 2023). In 2020, researchers turned to

wastewater as a potential medium for monitoring of SARS-CoV-2 prevalence in communities. Wastewater democratizes testing and captures material from both symptomatic and asymptomatic patients, without the need for the patient to voluntarily interact with a healthcare provider (Köndgen et al., 2010; Y. Ye et al., 2016; Zhu et al., 2015). While wastewater surveillance has been widely used during the COVID-19 pandemic, it is still subject to the molecular limitations of NAATs. A possible alternative to NAATs in clinical and wastewater samples is via next-generation sequencing methods.

Since they gather information on a nucleotide by nucleotide basis, next-generation sequencing (NGS) technologies enable the accurate identification of respiratory pathogens regardless of sequence mutations and variants. Depending on the ultimate goal of the sequence analysis, there are many different NGS platforms available to date that offer fast, accurate, and cost-effective solutions. Illumina, Inc. (San Diego, CA) sequencing instruments perform short-read sequencing by synthesis. An appropriately prepared library of genomic material is introduced to a flow cell, where the library fragments then hybridize to the surface and undergo bridge PCR amplification, forming a mono-template clonal cluster (Aigrain, 2021). These clusters then undergo sequencing by synthesis, where a single fluorescent dye-labeled dNTP is added one at a time, fluorescence is measured by optical equipment, and the next dNTP is added so that the next round of sequencing can be performed ("Principle and Workflow of Illumina Next-Generation Sequencing," 2018). This process continues for a set number of sequencing cycles depending on the needs of the experiment. Following sequencing, the fluorescent signals are converted into base calls, which can then be used in bioinformatics analyses downstream. This sequencing method is fast and effective with low error rates, though it is limited in the length of fragments that it can sequence, capping out around 2x 300 bp at a maximum (*Maximum*

*Read Length for Illumina Sequencing Platforms*, 2024). Conversely, Oxford Nanopore Technologies (Oxford, UK) sequencing instruments are capable of sequencing longer fragments. Library prep for ONT sequencing follows similar methods as Illumina, but the difference lies in the sequencing itself- the fragment is passed through the proprietary nanopore, and each base passing through it is recorded as an electrical signal. This electrical signal is converted into a base call, which is then used in bioinformatics analyses downstream. With these long sequencing reads, though, comes an increase in error rate (Zhang et al., 2020). There are advantages and drawbacks to each sequencing method, depending on which is used, and both are excellent options in surveilling HRVs in wastewater and clinical samples when NAATs are not enough.

The work in this dissertation explores these NGS methods, building on UNC Charlotte's COVID-19 genomic surveillance work by expanding the reach of surveilled HRVs beyond SARS-CoV-2. The main objective of this work is to keep people healthy and informed when it comes to HRVs in a way that is quick, efficient, and as least disruptive as possible.

# CHAPTER 2: LIMITATIONS OF TARGET-CAPTURE SEQUENCING FOR IDENTIFYING HUMAN RESPIRATORY VIRUSES IN WASTEWATER AND CLINICAL SAMPLES

The rise of the COVID-19 pandemic demonstrated the need for wastewater-based epidemiology (WBE) methods that bypass traditional clinical testing for the virus. Increasing abundance of SARS-CoV-2 viral RNA in wastewater was demonstrated to be a leading indicator for clinical case peaks, which can alert communities of potential outbreaks and allow them to make informed public health decisions in the absence of widespread individual testing. WBE has been increasingly used to screen wastewater for other human respiratory viruses, primarily influenza A and B, respiratory syncytial virus (RSV), and SARS-CoV-2. Here we describe the use of a next-generation sequencing (NGS) protocol that extracts viral RNA from wastewater and nasopharyngeal (NP) swabs that PCR tested negative for the presence of SARS-CoV-2 for use in a human respiratory virus target capture/enrichment sequencing assay. Following downstream bioinformatics analyses, the wastewater sequencing results were compared to those found clinically in the NP swabs. Genomes from both wastewater and clinical samples showed low average sequencing depths and breadth of coverage despite non-specific amplification of viral RNA in library preparation, and different bioinformatics pipelines used to determine relative abundance produced discordant results. Our results suggest that while target-capture sequencing may provide broad information at the genus level about the presence and abundance of human respiratory viruses in wastewater samples, this approach is not likely to provide the information that public health entities desire from sequencing; namely, sensitive and specific identification of circulating viral strains.

## 2.1 Introduction

Respiratory illnesses are among the most common bacterial and viral infections globally (Derbyshire & Calder, 2021). Human respiratory viruses pose a critical threat as their modes of transmission make them especially communicable from person to person. Individuals who contract respiratory illnesses often present with symptoms like coughing and sneezing (Subbarao & Mahanty, 2020); however, an asymptomatic individual may still exhale infectious viral particles via talking and exhaling (Scheuch, 2020). Given the potential ease of transmission, Human respiratory viruses can quickly spread within communities with outcomes potentially including significant morbidity and mortality (Tregoning & Schwarze, 2010). While the primary focus of current research has been on monitoring and slowing the spread of SARS-CoV-2, respiratory diseases caused by various other pathogens such as respiratory syncytial virus (RSV), influenza virus, parainfluenza viruses, human metapneumoviruses and many others remain a global health concern (Tiwari et al., 2024).

Most human respiratory viral infections present with similar symptoms and often cannot be easily differentiated from one another in the clinic (Leung, 2021). Multiplex molecular testing methods including RT-qPCR and ddPCR are commonly used to identify infectious pathogens from nasopharyngeal (NP) swab samples, offering healthcare providers better insight into appropriate treatment options for their patients (Jiang et al., 2022). However, clinical tests are often sought only if an individual voluntarily presents to a healthcare provider with symptoms. Such tests require the collection of a swab sample for testing and are not scalable to widespread routine surveillance of the asymptomatic population, except in extraordinary situations.

As an alternative to invasive NP swabs for routine surveillance, researchers began to explore the possibility of using wastewater to monitor the prevalence of SARS-CoV-2 infection

in communities, using similar molecular detection methods to those used for clinical samples (Kwong et al., 2015). These efforts have since extended to influenza, RSV and Mpox (*Beyond COVID-19: Influenza, RSV, Mpox, Etc.*, n.d.). The WBE approach is able to identify the presence of virus without individual opt-in to health care services, as wastewater encompasses shed viral material from people experiencing any degree of symptoms (Kumar et al., 2021). In the USA wastewater-shed SARS-CoV-2 levels are monitored at varying degrees via partnerships at the state, tribal, local, and territorial level, and data is aggregated at the national level via the CDC's National Wastewater Surveillance System (NWSS) (Adams et al., 2024). WBE methods were also adopted during the COVID-19 pandemic for populations living in high-density situations including hospitals, nursing homes, jails, schools, and dormitories (*Tracking COVID-19 Through Wastewater*, 2023). These public health surveillance efforts were effectively used to alert decisionmakers to rising case numbers and to anticipate increased staffing needs for medical care.

In most scenarios, molecular testing methods provide reliable results from both clinical and wastewater samples. However, if any limitations of the assay arise the test may be inconclusive or deemed a false negative. Molecular-based tests are subject to an assay's limit of detection (LoD), below which a sample would test negative regardless of initial matrix. This is a well-documented challenge in the case of clinically tested SARS-CoV-2 infections as a patient's viral load is dependent on the state of their infection (Takahashi et al., 2022). Other regularly monitored human respiratory viruses including Influenza and RSV are similar to SARS-CoV-2 with respect to illness duration and viral load (Bagga et al., 2013; Puhach et al., 2022). A similar logic can be applied to wastewater: while there may be some individuals experiencing an infection, in the context of a large composite volume their viral load is diluted by many orders of

magnitude and may fall below a molecular assay's LoD (Ahmed et al., 2022). Research into the shedding rates and viral loads for many different human respiratory viruses in wastewater has been conducted (Lowry et al., 2023). Further, wastewater is a complex matrix. Many biological and chemical compounds, including those known to inhibit PCR, exist in wastewater at variable concentrations. These types of compound matrices may lead to poor viral recovery, detection, and reproducibility in PCR-based assays (Kumblathan et al., 2022). There has been an extensive amount of research addressing these LoD constraints, including methods for wastewater concentration, rRNA depletion, and viral particle enrichment (Crits-Christoph et al., 2021; LaTurner et al., 2021).

In this study, we address another challenge of molecular-based testing methods. RT-qPCR and ddPCR-based assays rely on fluorescent probes and allow only limited multiplexing of molecular targets. There are many respiratory viruses for which routine clinical and wastewater molecular testing is not widely available. Human respiratory viruses whose identification would result in no significant difference in treatment or outcome in the clinic and emerging viral pathogens may go undetected in these scenarios. In such cases, molecular sequencing has the potential to provide simultaneous detection of a wider variety of viruses than PCR-based detection alone..

NGS methods used to identify human respiratory viruses from wastewater and clinical samples fall into three main categories: whole metagenome shotgun sequencing (WMGS), amplicon-based sequencing, and hybridization-based sequencing. WMGS is the least biased approach, shearing the entirety of the genomic material into small fragments that are then randomly primed and sequenced one at a time. A drawback of using this method to sequence human respiratory viruses is that without specific selection and amplification of viral material,

sequences present in very low abundance are likely to be missed unless each sample is sequenced very deeply. Amplicon-based sequencing methods can be used to target low-abundance viral material, but these approaches have the same problem as PCR-based methods in that they limit multiplexing of targets. However, the amplicons generated are sequenced deeply enough to allow for identification of specific genetic variants, which permits identification and monitoring of specific viral strains. The middle ground between these two approaches is found in assays that capture a variety of viral targets by hybridization prior to sequencing. Such assays can target dozens or even hundreds of viral species simultaneously. The draw of this assay strategy is that, if viral material of interest is present in a sample, it will be captured, and off-target sequences will simply be washed away. This step saves sequencing resources for only the targets of interest. Illumina Inc.'s Respiratory Virus Enrichment Kit (formerly RVOP) is one such assay. It targets a total of 40 different human respiratory viruses and has been validated with various sample types including both clinical and wastewater (*Detection and Characterization of Respiratory Viruses, Including SARS-CoV-2, Using Illumina RNA Prep with Enrichment*, 2020; *Surveillance of Infectious Disease through Wastewater Sequencing*, 2022). Another target capture assay is offered by Twist Biosciences. The Twist Respiratory Viral Research Panel targets 29 common respiratory viruses and allows for modifications in cDNA generation and sequencing method, making it an attractive choice for labs that need some flexibility. This method has not been tested on wastewater samples, but its related assay with a pan-viral scope, the Twist Comprehensive Viral Research Panel, has (Tisza et al., 2023).

In the fall of 2022 and early 2023, health officials warned of a "tripledemic" in the United States: Influenza, SARS-CoV-2, and RSV (Apoorva Mandavilli, 2022). This simultaneous surge

in viral infections from all three viruses was predicted to place an additional strain on hospitals that were already overwhelmed with COVID-19 cases, and it did. Widespread mask use and social distancing had curbed flu rates in 2020 and 2021, but many Americans abandoned those precautions as COVID-19 restrictions began to ease, with predictable results (Furlow, 2023). RT-PCR based wastewater surveillance data targeting  influenza A, RSV, human metapneumovirus and SARS-CoV-2 was collected in the Greater San Francisco Bay Area during the tripledemic period. Wastewater levels of all four viruses correlated well with clinical detections, suggesting that the potential for community-wide viral surveillance as a proxy for clinical testing was not limited to SARS-CoV-2 (Boehm et al., 2023). Many challenges remain in the interpretation of wastewater viral surveillance data, such as lack of clinical testing availability for validation, variation in population sizes, dilution of viral material, and unexplored variability in viral shedding rates (Tiwari et al., 2024), but it seems likely that applications of WBE will continue to expand, and that one way to expand them might be viral enrichment strategies.

Here, we explore the use of the Twist viral enrichment protocol to sequence clinical and wastewater samples collected from on-campus locations on the UNC Charlotte campus during fall and winter of 2022. Our original aim was to determine whether clinical trends in the student population were reflected in contemporaneous wastewater samples collected from residence halls. However, what we found were issues both in the generated data and in the recommended bioinformatics strategy for data analysis that make obtaining clear, specific and actionable results with this platform challenging. Our study addresses these limitations and highlights alternative interpretations, noting that the lack of viral specificity generated from this NGS method may limit its use for connecting wastewater and clinical data as originally intended.

## 2.2 Methods

### 2.2.1. Sample Collection

#### 2.2.1.1. Clinical swabs

Clinical nasopharyngeal swabs were collected from the Student Health Center at UNC Charlotte, where they were tested for the presence of SARS-CoV-2, Flu A and Flu B via Cepheid's *GeneXpert*® Xpress SARS-CoV-2/Flu Plus Diagnostic Assay test. The nucleocapsid (N) and envelope (E) and RNA-dependent RNA polymerase (RdRP) genes of the SARS-CoV-2 virus genome, influenza A matrix (M), influenza A basic polymerase (PB2), influenza A acidic protein (PA), influenza B matrix (M), influenza B non-structural protein (NS), and the RSV A and RSV B nucleocapsid primer sequences are targeted in this qPCR-based test (*Xpert® Xpress CoV-2/Flu/RSV Plus*, 2023). If any one (or more) of the aforementioned targets are detected, a sample is considered "positive" for that target, and "negative" for that target if it is undetected- though at the time of testing, the SHC did not include RSV in the panel of results. Samples that tested negative for the presence of SARS-CoV-2 targets and were either positive or negative for the presence of Flu A/B targets were retained from the testing center, deidentified, and stored in tubes containing small amounts of viral transport medium (VTM) at -80°C prior to their arrival to the lab for processing.

#### 2.2.1.2. Wastewater

Wastewater samples were collected from residence hall sites on the UNC Charlotte campus and were concentrated as part of an ongoing COVID-19

wastewater monitoring project (Gibas et al., 2021). Briefly, up to 10L composite wastewater samples were auto sampled every 30 minutes over a 24 hour period and collected into a bottle. Samples were aliquot to a volume of 45 mL and spiked with Bovine Coronavirus (BCoV) (BOVILIS® Coronavirus, Merck Animal Health, NE, USA) as a process control for downstream PCR. Each sample was sonicated for one minute, centrifuged at 10,000xg to remove solids, and the supernatant was then subject to filtration through 0.05 um PS Hollow Fiber Filter Tips (Innovaprep) using the automatic CP Select (Innovaprep). Viral particles were then eluted using 0.075% Tween-20/Tris WetFoam elution fluid (Innovaprep) into a final volume ranging from 250 µL to 500 µL (Juel et al., 2021).

2.2.2. RNA extraction; wastewater ddPCR

RNA was extracted from both sample types using the QIAamp viral mini kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions with minor modifications. RNA was treated with a one-step PCR inhibitor removal kit (Zymo, Irvine, CA, USA) to remove any impurities that might remain in wastewater samples after RNA extraction. All wastewater samples were ddPCR tested for the presence of the SARS-CoV-2 N2 target using Bio-Rad's One-Step RT-ddPCR Advanced Kit for Probes (Bio-Rad, Hercules, CA, USA) as part of the ongoing wastewater monitoring project at UNCC. Samples that tested negative for the presence of the target were used in this study. A small selection of clinical samples were extracted using the Ceres $Nanotrap^{®}$ Microbiome A Automated Protocol (Ceres Nanosciences, Manassas, VA) with $MagMAX^{™}$ kit and the $KingFisher^{™}$ Apex (Thermo Fisher Scientific, Waltham, MA)

according to the manufacturer's instructions when the laboratory switched to an automated nucleic acid extraction protocol. Each sample was quantified to determine the presence of viral RNA using the Qubit Broad-Range RNA assay (Thermo Fisher, Waltham, MA, USA) and then stored at -80°C until further processing.

2.2.3. Sequence-Independent, Single Primer Amplification of RNA and Library Preparation

Sequence independent, single-primer amplification (SISPA) and subsequent library preparation was performed using the Roppolo Brazell et al. protocol (Roppolo Brazell et al., 2023). Most RNA concentrations for both clinical and wastewater SARS-CoV-2 negative samples were below the library preparation protocol-recommended 50 ng of input material, so RNA was randomly primed for non-specific PCR amplification to increase the amount of viral material going into library preparation. DNA fragmentation, end repair, and dA-tailing followed Twist Bioscience's cDNA Library Prep Kit for ssRNA Virus Detection (Twist Bioscience, San Francisco, CA, USA). The dual-indexed, purified libraries were then prepared for hybridization with Twist's Respiratory Virus Research Panel following the Twist Bioscience's Target Enrichment Protocol (Twist Bioscience, San Francisco, CA, USA). Following the final PCR amplification step in this protocol, the libraries were quantified again using the Qubit 1X dsDNA HS assay to ensure input genomic material was appropriate for sequencing on an Illumina platform.

2.2.4. Sequencing

The DNA pools were then diluted to a final loading concentration of 750 pM and prepared for sequencing on an Illumina NextSeq 2000 instrument, according to the

protocol described in the Illumina DNA Prep Reference Guide (Illumina Inc., San Diego, CA). Sequencing generated paired-end reads of 150 bp in length, and those that passed QC constraints were stored in Illumina's BaseSpace Sequence Hub where further analyses could be performed within their cloud-based environment using their built-in analysis tool, DRAGEN Bio-IT.

2.2.5. Bioinformatics

The raw sequence reads that were retrieved off the Illumina BaseSpace Sequence Hub were stored as zipped .fastq files and saved to UNCC's high-powered computing cluster. They were trimmed for technical sequence/adapter content using Trimmomatic v0.39 (ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36), and then underwent a final quality check using FastQC v0.11.9's default parameters (Bolger et al., 2014; S Andrews, 2010).

A standard taxonomic classification against the NCBI viral database (built and accessed on July 15, 2024) was performed using default parameters for Kraken2 v2.1.3 followed by the use of its sister tool, Bracken v2.9, to re-estimate relative abundances of taxa in each sample (J. Lu et al., 2017; Wood et al., 2019). Default parameters were used with the exception of specifying a read length of 150, a classification at the species level, and a threshold of 10 for the minimum number of reads required for classification at this level (-r150 -t10 -lS).

Read mapping was also performed by aligning reads to a multifasta file that contained the reference genomes of Twist's Respiratory Viral Research Panel via Bowtie2 v2.4.1, allowing one mismatch in a very sensitive local alignment (--very-sensitive-local -N 1) (Langmead & Salzberg, 2012; Supplementary Table 1). Reads that mapped to the

reference genomes were then sorted by genome, and average depth/breadth of coverage computed. Relative abundance was then calculated by dividing the number of reads mapped to the reference genomes by the total number of reads, multiplied by 100.

Depth for each genomic position within each sample was computed following analysis with SAMtools depth v1.10 with the -a parameter specified to include positions with zero coverage (Li et al., 2009). Average depth per genome was calculated by dividing the sum of the coverage at each position by the number of positions in the genome, and breadth of coverage was calculated by dividing the number of covered positions by the number of positions in the genome. MosDepth was used to calculate the median coverage of 500-bp chunks of each genome at different thresholds (--use-median –by 500 –thresholds 1,10,20,30) (Pedersen & Quinlan, 2018).

TBlastX v2.11.0 was used to examine sequences with coverage at 10X or greater (McGinnis & Madden, 2004). Each 500-bp chunk of the genome at this coverage level was given a unique identifier corresponding to its positional range. That range of the genome was then extracted from the reference multifasta file via SeqKit v0.16.1 and used as input for a translated nucleotide search using TBlastX against the NCBI's viral nt database (built and accessed on August 13, 2024) (Shen et al., 2016). The output format selected for the TBlastX was inclusive of the query sequence IDs, query/subject taxID, %ID, length, mismatch tolerance, gapopen, query/subject sequence start/end, e-value, bitscore, and query/subject scientific/common names (-outfmt 6). The accession numbers of each query sequence ID were merged with their TaxID from genbank's Accession2TaxID map, and the original subject TaxIDs assigned by the Blast database

were matched and replaced with those from genbank's Accession2TaxID map for congruence between databases.

The Accession2TaxID map's nodes.dmp and names.dmp files were downloaded and parsed to identify hierarchical taxonomic relationships at the genus and species level. These relationships were traversed and logged for each query/subject sequence in the TBlastX output file, allowing for post-processing steps to remove alignments where the subject taxa was equivalent to the query taxa at the desired levels. Should the desired level not have information that could be traversed in the Accession2TaxID map, it was logged as "Unknown".

All statistical analyses and figures were created using RStudio v4.3.2.

## 2.3 Results and Discussion

In total, 84 building wastewater samples and 86 clinical samples that dd/qPCR tested negative for SARS-CoV-2 were collected and sequenced during the tripledemic period of Fall 2022 - Winter 2023. In order to be considered negative for SARS-CoV-2, the concentration of the detected gene in each sample must fall below each assay's LoD. For ddPCR tested wastewater, the LoD for the assay detailed in the methods was 330.011 cp/L for N2 (Barua et al., 2022). For qPCR-tested clinical samples, the LoD for the assay detailed in the methods was 70 cp/mL for the E-gene, 403 cp/mL for the N2-gene, and 200 cp/mL for RdRp (Noble et al., 2022). Each wastewater sample is representative of a collective of individuals who visited a specific wastewater testing site. Wastewater sample collection sites used in this study were from a designated set of 6 locations that embodied the north, south, and central parts of campus. Because these samples were collected as part of a routine wastewater monitoring program, many sites were collected on the same date, of which 27 unique dates span from August 2, 2022 to February 21, 2023. Where wastewater is considered a composite sample, each clinical sample is representative of an individual who sought testing at the SHC. While there are some dates where more than one individual was tested, the clinical samples spanned 52 unique dates from September 6, 2022 to February 17, 2023. The two different sample types underwent RNA extraction, SISPA, library prep, and sequencing on an Illumina NextSeq 2000 instrument, yielding a 90.5% average Q30 score, with 78.98% of reads passing internal filters. The total read yield was 40.44 Gbp.

2.3.1. Sequencing outcomes show low average depth and breadth of coverage of wastewater & clinical SARS-CoV-2 negative samples using the target enrichment protocol

Following sequencing, trimming and QC, reads generated from wastewater samples collected on the same date were combined to represent a larger composite wastewater sample. The average number of reads per combined wastewater sample was 1973713, whereas the average number of reads per clinical sample was 188282.

Breadth of coverage at 1X and average sequencing depth were calculated for all genomes in both types of sample [Figures 2.1, 2.2, Supplementary Table 2]. While there are no published studies that validate this exact methodology(i.e., SISPA amplification on RNA from concentrated wastewater) with modifications to accommodate for low viral load on wastewater or clinical samples, our breadth of genome coverage and average sequencing depth for this sample type were purely exploratory. Many samples have a low breadth of genome coverage regardless of sample type (clinical vs wastewater). A one-way ANOVA revealed a significant effect of sample type on the breadth of genome coverage, where $F(1,758) = [7.27]$, $p = 0.007$. The same test revealed that there was no significant effect of sample type on the average sequencing depth, where $F(1,758) = [0.20]$, $p = 0.657$. There is variability in both categories between samples of the same type, which could be due in part to the lack of normalization of RNA input concentration prior to SISPA-mediated cDNA conversion. Viral material was quantified following cDNA conversion, but no normalization was necessary until after adapter ligation and amplification, prior to pooling libraries for hybridization.

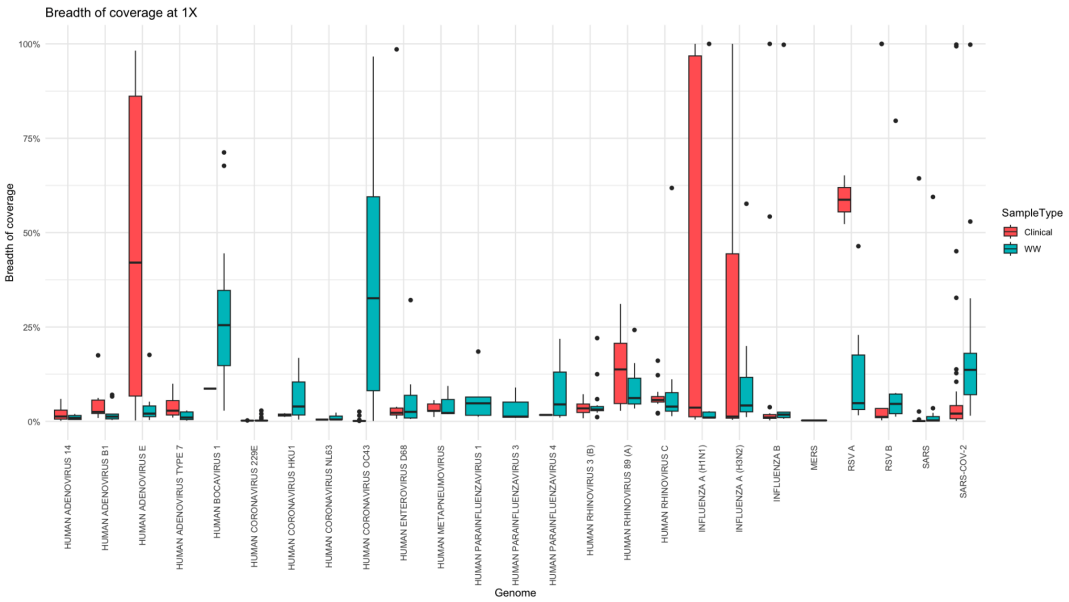


Figure 2.1. Distribution of breadth of genome coverage at 1X across all target genomes present in clinical (red) and wastewater (teal) samples. Of the 29 target genomes, only four (rubella, mumps, measles, and human rubulavirus 2) are not represented in this data. Breadth of genome coverage is computed as a percentage of the length of each genome.

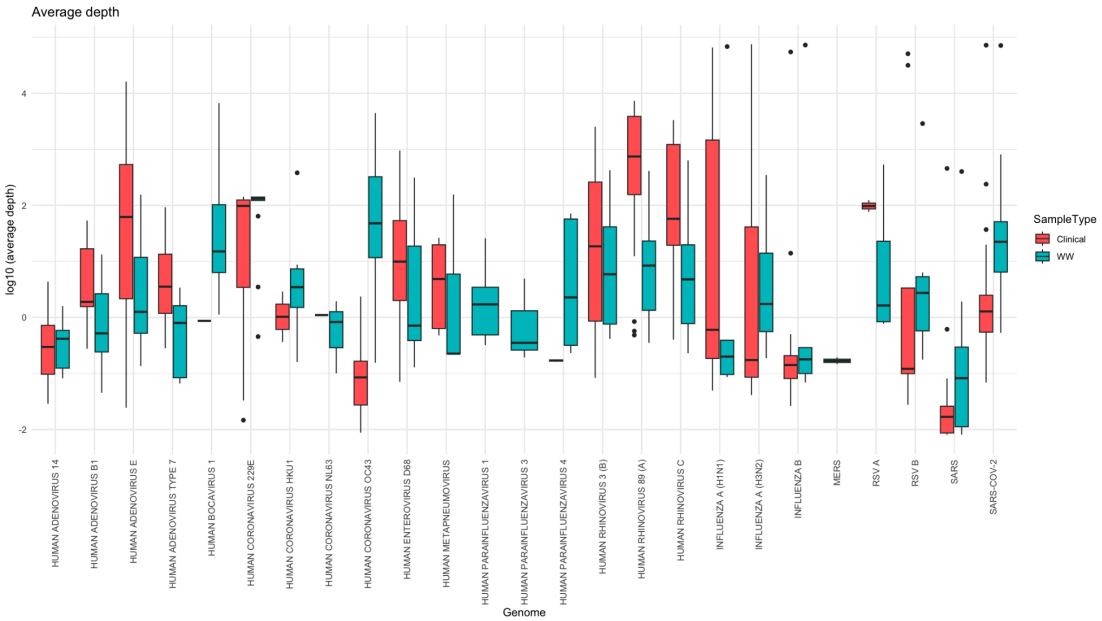


Figure 2.2. Distribution of the average sequencing depth at each position in each target genome present in clinical (red) and wastewater (teal) samples. Of the 29 target genomes, only four (rubella, mumps, measles, and human rubulavirus 2) are not represented in this data. Average sequencing depth is reported on a logarithmic scale (base 10).

A 2023 study that analyzed the performance of the Twist assay on clinical samples, following the standard manufacturer protocol for the assay, found that the majority of the patient samples (62%) sequenced did not produce concordant results with a multiplex qPCR assay of the same samples. Detection criteria used in that study were implemented in the bioinformatics pipeline One Codex, which is recommended by Twist for analysis of data generated using this assay (Kapel et al., 2023). The bioinformatics strategy behind One Codex is an exact alignment to the One Codex Database via Minimap2 using k-mers (k=31) (Denise Lynch, 2023). Any viral hit with 100 reads or greater is able to have secondary statistics generated- average depth, coverage, and sequence identity- all of which are used to determine whether or not a virus is considered "determinate" (at least 20% of the genome covered at an average depth of 10x across entire genome) or "indeterminate" (at least 5% of the genome covered at an average depth of 10x across entire genome) (Christopher Smith, 2024). With this in mind, we generated reports from One Codex and then applied similar criteria to our own analysis of the sequence data that we generated. Only one of the clinical samples in this study was positive for any of the Cepheid targets, so it was difficult to test for concordance with qPCR on behalf of positive results. We did observe (97.7%) concordance with regard to SARS-CoV-2 negatives, with only 2 of 88 samples meeting One Codex's criteria for a positive SARS-CoV-2 result as it relates to average sequencing depth and breadth of coverage. The sample that qPCR tested positive for influenza A did not meet the One Codex criteria for a determinate or indeterminate positive for this genome. Further, there was noticeable variability in the detection statuses between samples and genomes identified as "determinate" and "indeterminate" in each sample [Supplementary Table 2].

2.3.2. Differences in viral abundance between sample types and classification methods

Bearing in mind that a k-mer based classification was under the hood of One Codex, Kraken2, a common k-mer based metagenomic classification tool, was used for an initial analysis of our data [Supplementary Table 3]. Briefly, Kraken2 works by splitting sequence reads into k-mers, comparing them to reference genomes in the specified databases, and then assigning them to their lowest common ancestor taxonomic classifications (Wood et al., 2019). Sequence hits are then interpreted and refined with integration into Kraken2's companion software, Bracken, along with k-mer distributions for the reference and query sequences to reestimate the relative abundances of each taxonomic group per sample. 63/88 (71.6%) of the clinical samples that qPCR tested negative for SARS-CoV-2 were assigned a partial or entire abundance taxonomic classification of SARS-CoV-2 using this analysis method. Counting SARS-CoV-2, only nine viruses of the 29-virus research panel were detected using this method, including adenoviruses B and E, rhinovirus 89 (A), enterovirus D 68, rhinovirus 3 (B), RSV, influenza A and influenza B [Fig. 2.3]. The sample that qPCR tested positive for influenza A showed no relative abundance of this virus using the Kraken2 classification method; instead, the most abundant virus in this sample was SARS-CoV-2 at 60%.

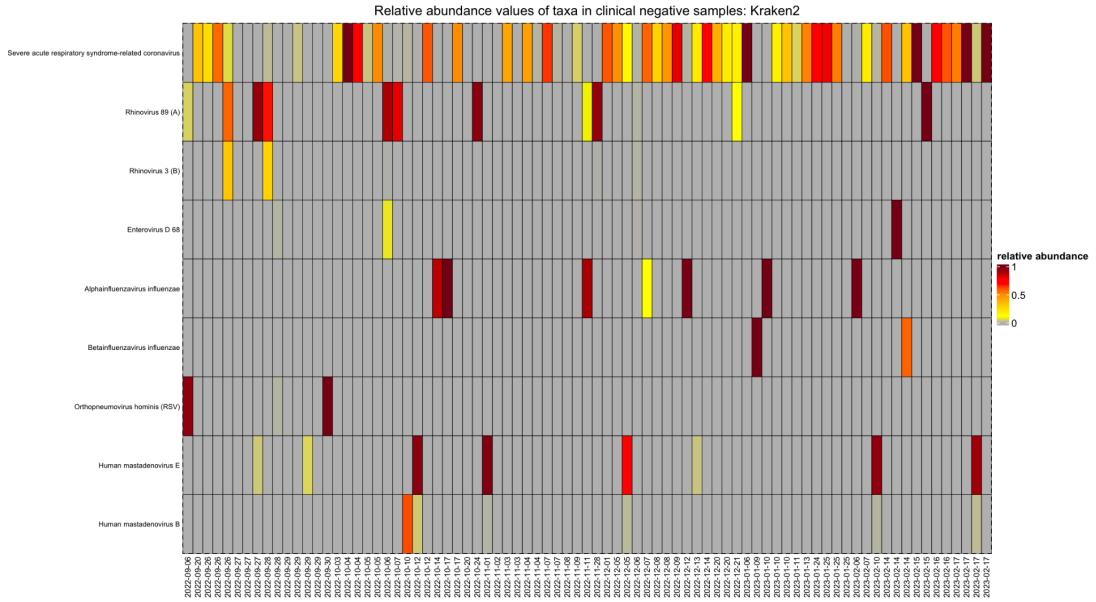


Figure 2.3. Relative abundance of each on-target viral genome in each clinical sample as identified by Kraken2. Samples are organized in chronological order.

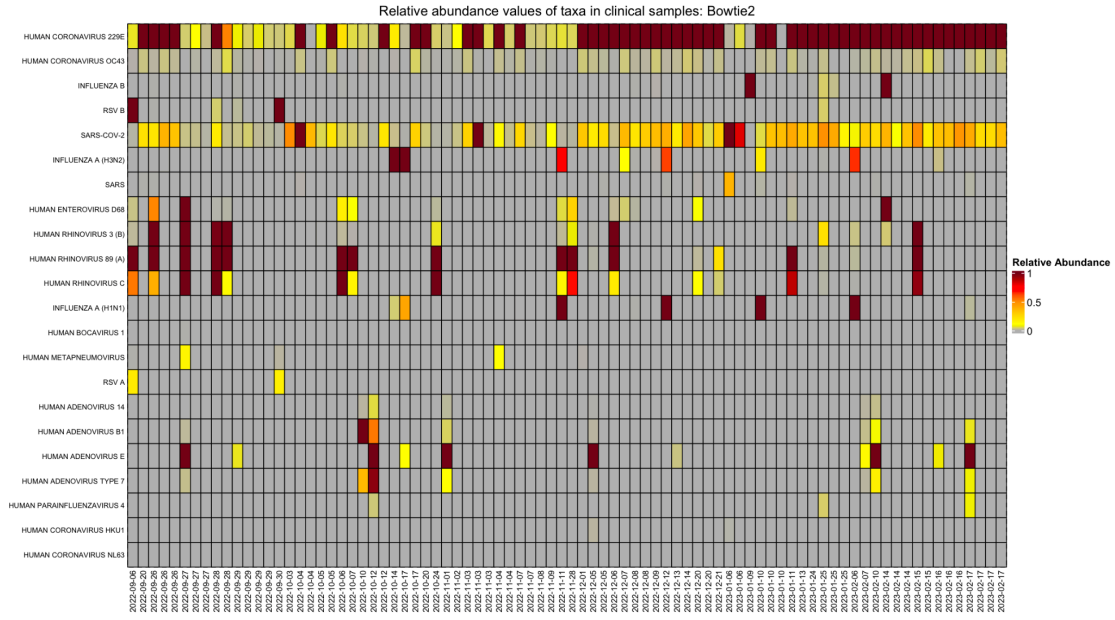


Figure 2.4. Relative abundance of each on-target viral genome in each clinical sample as identified by Bowtie2. Samples are organized in chronological order.

Because Kraken2 classifications only identified six pathogens beyond SARS-CoV-2, we chose to also pursue the assignment of taxonomy based on mapping to the reference genomes [Supplementary Table 4]. These were selected based upon the NCBI accession numbers provided by Twist as the reference genomes their panel was designed from. Given this, 84 of the same samples were assigned a partial abundance taxonomic classification of SARS-CoV-2 by mapping reads to the reference genome [Figure 2.4]. We expected that the overwhelming majority of clinical negative samples would represent infections with influenza A/B, and RSV A/B. Sequencing results using Twist's Respiratory Viral Research Panel detected a much wider variety of these on-target hits than Kraken2 did; 21 of the 29 respiratory viruses targets were seen in clinical samples, excluding only human parainfluenzaviruses 1, 2, 3, measles, MERS, mumps, and rubella.

Wastewater SARS-CoV-2 negative samples were subject to the same analyses. Similar to clinical PCR-based tests, wastewater samples that have a viral load less than the assay's LoD will go undetected. SARS-CoV-2 viral RNA in wastewater is reflective of the viral load being shed in feces regardless of whether an individual is showing symptoms, and the fecal viral load can sometimes be higher than the viral load in a NP swab (Daou et al., 2022). We expected that, once reads from wastewater samples collected on the same date were combined, we would see presence of SARS-CoV-2 in all samples. Kraken2 assigned 26 of 27 wastewater samples a partial abundance of SARS-CoV-2, where the read-mapping method assigned all samples a partial abundance of SARS-CoV-2. Only 13 of the 29-virus research panel targets were detected using Kraken2 [Figure 2.5].

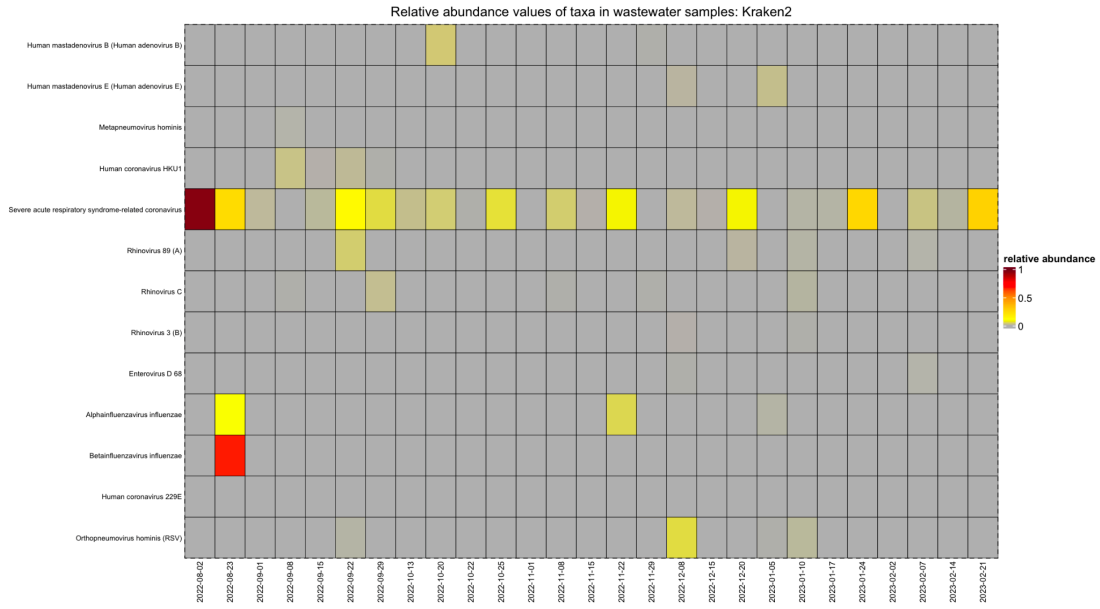


Figure 2.5. Relative abundance of each on-target viral genome in each composite wastewater sample as identified by Kraken2. Samples are organized in chronological order.

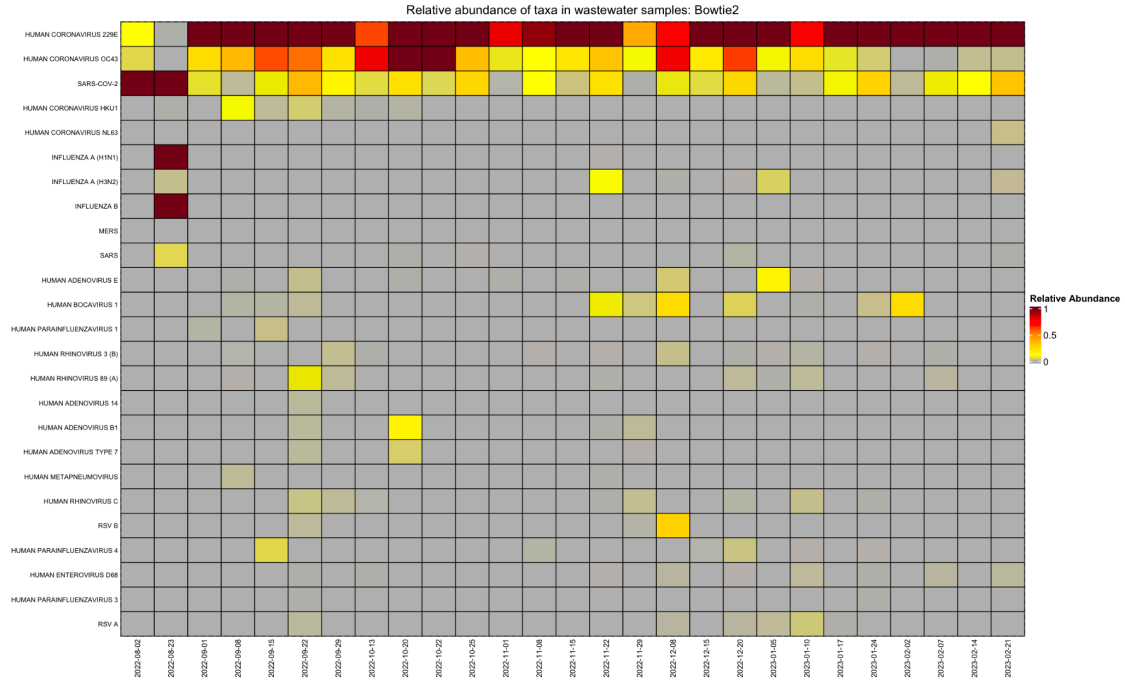


Figure 2.6. Relative abundance of each on-target viral genome in each composite wastewater sample as identified by Bowtie2. Samples are organized in chronological order.

Wastewater samples boasted a slightly more diverse taxonomic profile than clinical negative samples with respect to on-target hits. 25 of the 29 viruses targeted by the panel were detected using the direct alignment method, excluding only parainfluenzavirus 2, measles, mumps, and rubella [Figure 2.6].

### 2.3.2. Evaluating metagenomic analysis approaches for target-capture sequencing

While microbiome studies are often the standout metagenomic study types, virome studies have increased in popularity. In particular, the COVID-19 pandemic demonstrated the utility of using WBE to monitor the spread of the SARS-CoV-2 and its many variants throughout communities. Metagenomics bioinformatics strategies are often employed to deconvolute complex microbiomes in NGS samples. Outside of qPCR and shotgun NGS-based metagenomics, there are limited options available for bioinformatics methods that are designed to consider only members of the virome in a targeted hybridization, such as Twist Bioscience's Respiratory Viral Research Panel. This type of NGS method begs the question, is a virome sample truly a "metagenomic" sample, and can it be treated as such bioinformatically? Targeted, hybridized samples are not true metagenomic samples and the Kraken pipeline may not be appropriate as it may interpret false matches as part of a taxonomic cluster at a genus level or above.

In the previous section, we demonstrated that Kraken2 was used for an initial analysis of our data. Taken together with breadth/depth "determinate" calls and relative abundance results from both methods, we suspected that Kraken2 was overrepresenting sequences that were much less abundant based on direct alignment and mapping using Bowtie2. This behavior has been observed in shotgun metagenomics datasets and

discussed in several recent publications- Johnson et al. demonstrated in 2022 that Kraken2 reports “phantom taxa” that reflect classification errors of high-abundance taxa to lower abundance ones in simulated datasets (Johnson et al., 2022). Bioinformatic contamination is not unprecedented in these types of studies and it is important to take this into account when selecting an appropriate analysis pipeline.

Important to note is that, when compared to the most recent iteration of the Kraken2 Viral database and NCBI’s reference genomes of the viruses included in the panel, many off-target hits are being called at the species level [Figure 2.7, Figure 2.8]. Hybridization capture approaches like the methods used in this study generally work well but can sometimes suffer from off-target captures that affect the interpretation of results (Head et al., 2014). Kraken2’s Viral database reported the species of these off-target hits [Supplementary Files 1 & 2]. Kraken2 identified 60 viral taxa among clinical samples, and of these, only 7 (11.7%) were considered on-target. Viral species that showed relative abundance of 50% or greater that were off-target included *Feravirus neuropterus*, BeAn virus 58058, *Betabaculovirus chofumiferanae*, and human mastadenoviruses E & B. *Feravirus neuropterus* is part of the Phasmaviridae family of viruses with negative stranded RNA genomes (Kuhn & Hughes, 2024). BeAn virus 58058 is a member of the double-stranded DNA virus family Poxviridae. Interestingly, this specific virus is a zoonotic orthopoxvirus capable of infecting a wide range of animal and human hosts- and has been previously identified in postmortem COVID-19 patients as a frequently nonpathogenic detected species (Ferravante et al., 2022). *Betabaculovirus chofumiferana*e is a member of the double-stranded DNA virus family Baculoviridae. Baculoviridae are insect-specific viruses, though they have been found in humans and

other non-target organisms, they have no negative effect on their health [Davis, 2019]. Human mastadenoviruses E & B are part of the Adenoviridae family of viruses, of which four of the expected targets (human adenoviruses 14, B1, E, and Type 7) belong to. Adenoviruses cause acute respiratory disease and occasionally pneumonia, manifesting symptoms including fever, upper respiratory tract infection and conjunctivitis (Walter Doerfler, 1996). Given these clinical presentations, these off-target assignments are understandable in clinical samples. In wastewater samples, Kraken2 identified 345 viral taxa, of which only 9 (2.6%) were considered on-target. The only virus species that showed a relative abundance of 50% or greater aside from on-target hits was *Gihfavirus pelohabitans*. This species belongs to the Steitzviridae family of RNA viruses that infect prokaryotes, and is considered a Leviviricetes phage (Neri et al., 2022). It is no surprise to see so many off-target hits in wastewater as the matrix is full of bacteria that can be infected by phage, however, there were far less on-target hits to the range of viruses that were expected to be in the panel in this sample type.

For a point of comparison to Kraken2's off-target hits, we assigned the proportion of reads that did not align to the reference genomes with Bowtie2 as off-target hits. The distribution of off-target vs. on-target reads as identified by Bowtie2 alignments did not dramatically differ from what was seen in the Kraken2 results. Clinical samples showed the most similar distributions of on vs. on-target read distributions between both methods, while there was a marked difference between on vs. off-target reads in wastewater samples.

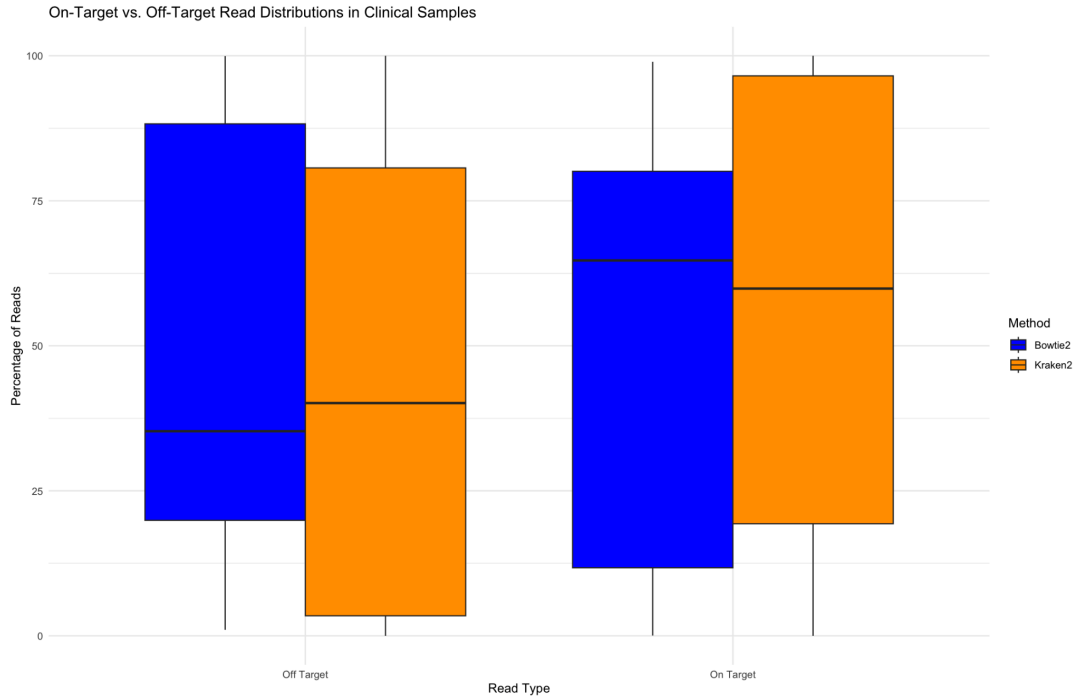


Figure 2.7. On-target vs. off-target distribution of taxonomies in clinical samples. These distributions represent the portion of total reads that are classified as on or off target as identified by Bowtie2 and Kraken2.

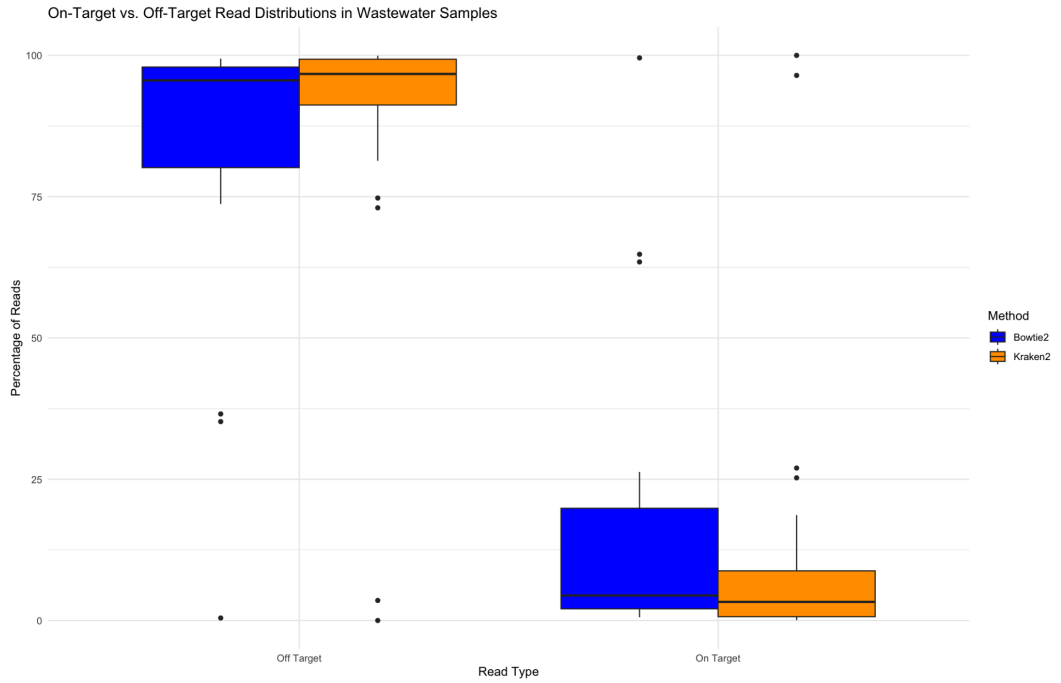


Figure 2.8. On-target vs. off-target distribution of taxonomies in wastewater samples. These distributions represent the portion of total reads that are classified as on or off target as identified by Bowtie2 and Kraken2.

2.3.4. No correlation between high coverage depth and cross-reactivity at genus and species levels

With low coverage common and no dramatic difference between on vs off-target hits abundant, we took 500 base-pair genomic intervals covered at 10X or higher for each virus and examined them more closely [Supplementary Tables 5, 6, & 7]. The main goal of this portion of the study is to determine whether or not there was any consistency in taxa between areas that sequenced with greater coverage. Any of these genomic intervals that showed any cross-reaction with other taxa at the genus and species level suggest that there may be a lack of specificity in this assay. Our bioinformatic approach to assign taxonomy to these highly covered regions is detailed in the methods section.

At the genus level, the large majority of hits outside of the query sequence taxa were assigned to an "unknown" genus for both wastewater and clinical samples, and were further examined at the species level [Supplementary Figures 3 & 4]. In wastewater samples, human coronavirus HKU1 and human parainfluenzavirus 4 were exceptions. *Human parainfluenzavirus 4* showed two regions of the genome that showed cross reactions at the genus level with hits to Respirovirus, and a small number of cross reactions at the genus level with hits to Jeilongvirus. Human coronavirus HKU1 showed a single region of the genome with cross reacts at the genus level with hits to the following genera: *Alpha/Betabaculovirus* (a member of which we saw in the clinical samples assigned by Kraken2), *Begomovirus, Biavirus, Carvajevirus, Catovirus, Dabirmavirus, Dongdastvirus, Hokovirus, Igirivirus, Kayvirus, Lymphocryptovirus,*

*Megavirus, Mimivirus, Nonavirus, Pelagivirus, Pourcelvirus, Rheavirus, Sashavirus, Rotavirus, Tequintavirus, Theiavirus, Tupanvirus,* and *Whopevirus.*

For both wastewater and clinical samples that contained genus level cross-reactions from the same genomes, direct comparisons could be made. While they classified exclusively as unknown, clinical samples showed more genomic chunks covered than those seen in wastewater samples for human adenoviruses B1 and E. SARS showed similar results, with some cross-reacts with the *Alphacoronavirus* genus. Both sample types showed hits for a cross-react in the same area of the SARS-CoV-2 genome with the *Morbillivirus* genus. Cross-reacts at the genus level for *Influenzavirus B* in both sample types showed an overwhelming majority of the *Influenzavirus A* genus. A scatter plot was generated to visualize the average sequencing depth of coverage vs. the number of off-target hits at the genus level for both wastewater and clinical samples. In order to determine if there were any visual relationships between depth and off target hits, a loess curve was fitted to the data, revealing no significant trends.

In order to resolve the areas where genus level could not be discerned, the same bioinformatic methods were used to explore cross-reacts at the species level in both wastewater and clinical samples [Supplementary Figures 5 & 6]. Full resolution at the species level was achieved, though some cross-reacts were still observed. In clinical samples, *Human adenovirus E* showed the most cross-reacts over nearly the entire genome. Cross-reacts between other similar species of *Human adenovirus*, as well as those from primates were observed. A small number of hits were observed in clinical samples that cross reacted with the *Enterovirus* and *Adenovirus* species, and *Human rhinovirus 89 (A)* had a small number of cross-reacts with both the *Enterovirus* and

*Rhinovirus C* species. Similar results were found in wastewater samples. In the case of *Rhinovirus 89 (A)*, wastewater samples showed the same distribution of hits outside of the query species taxa, and for *Human adenovirus B1*, there was one fewer cross-react species than there was in the clinical samples (unidentified Adenovirus). For *Human adenovirus E*, wastewater samples covered much less of the genome with cross-reacts and had fewer species in their distribution of cross-reacts- though they were all similar to clinical samples in that they were either human or primate species-derived. In the case of SARS, clinical and wastewater samples generally agreed in their distribution of cross-reacts, though clinical samples still covered more of the genome and had more hits. The species-level cross-reacts for SARS are all different coronaviruses and sarbecoviruses. For SARS-CoV-2, wastewater and clinical samples agreed exactly with their species-level cross-reacts.

Where *Human parainfluenzavirus 4* had some unresolved taxonomies that cross-reacted at the genus level in wastewater, they were resolved at the species level to hits of various species of *Orthorubulavirus, Paramyxovirus, Parainfluenzavirus, Respirovirus*, and *Jeilongvirus*. Human coronavirus OC43 showed an almost entire genome coverage of cross-reacts with various species of *coronavirus*. Perhaps the most interesting of the wastewater genomes that showed cross-reacts was that of *Human coronavirus HKU1*. While the span of the genome that was covered with cross-reacts was small, the cross reactions were the most diverse. In these areas, we observed the expected cross-reactions with other coronavirus species, but we also saw the presence of phage species that infect various bacteria. We also observed specific species that followed the known genera that were assigned in the previous step resolve specific species. A scatter

plot was generated to visualize the average sequencing depth of coverage vs. the number of off-target hits at the genus and species level for both wastewater and clinical samples. In order to determine if there were any visual relationships between depth and off target hits, a loess curve was fitted to the data, revealing no significant trends [Supplementary Figures 7 & 8].

With the genus and species-level cross-reactions examined, our results establish that there are some problems with genus and species level specificity. Some solutions to this may lie in masking areas of the genome that have large numbers of known cross-reactions to the reference genomes.

## 2.4 Conclusions and Future Work

Target capture-based sequencing methods like the Twist Respiratory Virus Research panel have utility in clinical diagnostics. However, few studies have forayed beyond testing the capabilities of this assay in synthetic RNA controls and patient samples. Our study offers a comprehensive trial and review of the Twist Respiratory Viral Research Panel's capabilities in both clinical and wastewater sample types. This study overcomes the limitations of traditional PCR-based assays by employing NGS methods that enhance multiplex capacity beyond a few targets and identify whole viral genomes at the nucleotide level. While we demonstrate that it is feasible to non-specifically amplify viral RNA in wastewater and use it for input in this assay, our results show that the average depth and breadth of coverage for each targeted genome is not sufficient in wastewater and clinical sample types. Further, a lack of specificity at the genus and species level suggests that assay and bioinformatic modifications may be considered when selecting this assay over others.

In addition to piloting a use-case for this assay with wastewater, this study originally sought to use the results of this NGS assay to draw conclusions about the "Tripledemic", determining how well wastewater could be used as a proxy to clinical tests in predicting respiratory viral trends on the university campus. As part of an existing infrastructure that was established in the Fall of 2020 to monitor SARS-CoV-2 levels on campus, a demonstrated success in the use of this assay would open the possibility of surveillance for a wider variety of respiratory viruses outside of SARS-CoV-2 to keep students, faculty, and staff informed of potential infection risks. Our pilot study falls short of being able to confidently provide this information, however, improvements could be

made on future iterations of this experiment that may help resolve these problems. A larger sample size would increase statistical power. Wastewater and clinical samples alike would benefit from exact viral load quantification confirmed by an input RNA Qubit and again by way of RT-PCR prior to library preparation with this assay. Such information would impart additional context that could be used to produce statistically significant results. Further, normalizing the RNA input prior to cDNA synthesis with SISPA may ensure that there is no sequence bias, allowing for overrepresentation of some sequences over others. While this does not directly address the bioinformatic overrepresentation issues that we experienced using a k-mer based classification method, it would improve accuracy in other bioinformatic analysis methods- many of which we did not attempt in this study.

Addressing the low breadth of genome coverage and sequencing depth is paramount for increased confidence in the results achieved using a hybridization-based sequencing assay. A potential solution to this problem could be to identify and remove identified areas that are prone to cross-reactions for a more accurate taxonomic identification. Moreover, better genome coverage sequenced at greater depths could improve the ability to call variant genomes that deviate from the references used in this panel, which could be useful in identifying variants of concern for existing viral pathogens. Should sequence reads achieve sufficient depth and breadth of coverage needed for a particular virus, it would be possible to use these metrics to confidently evaluate other bioinformatics methods for precision and accuracy. For example, we hoped to observe strong concordance in viral detection across samples analyzed with different bioinformatics methods, but we did not in this study [Supplementary Table 8].

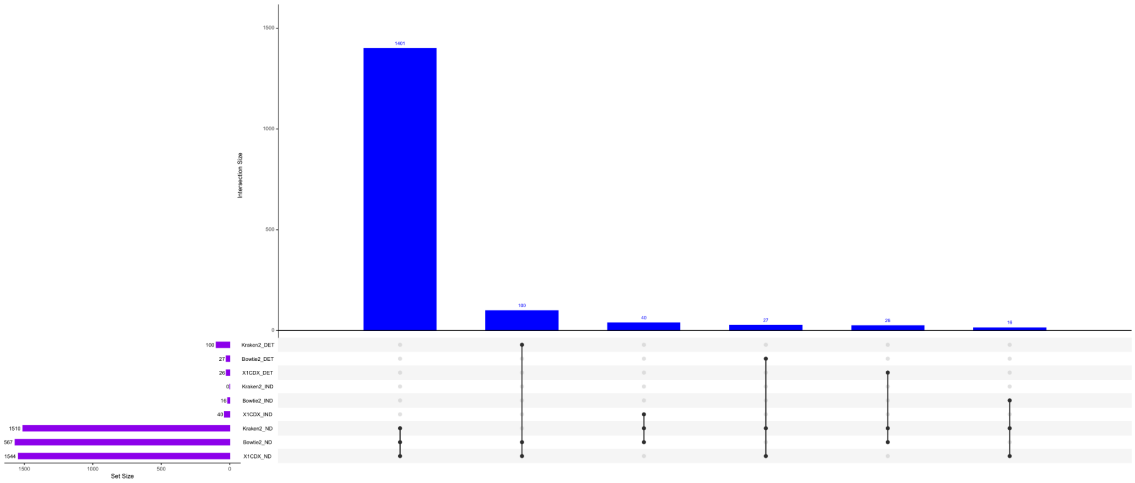

Figure 2.9. Clinical upset plot. The data used to generate this plot is based on the Kraken2 detections of viruses as well as Bowtie2/One Codex detections that were considered “determinate” or “indeterminate”. This information was extrapolated into a binary matrix used to create the upset plot. The purple bars represent the counts of each method/detection combination. The three vertical dots represent the combination of methods and determinations that were seen, and the blue bars represent the number of times that particular combination was seen.

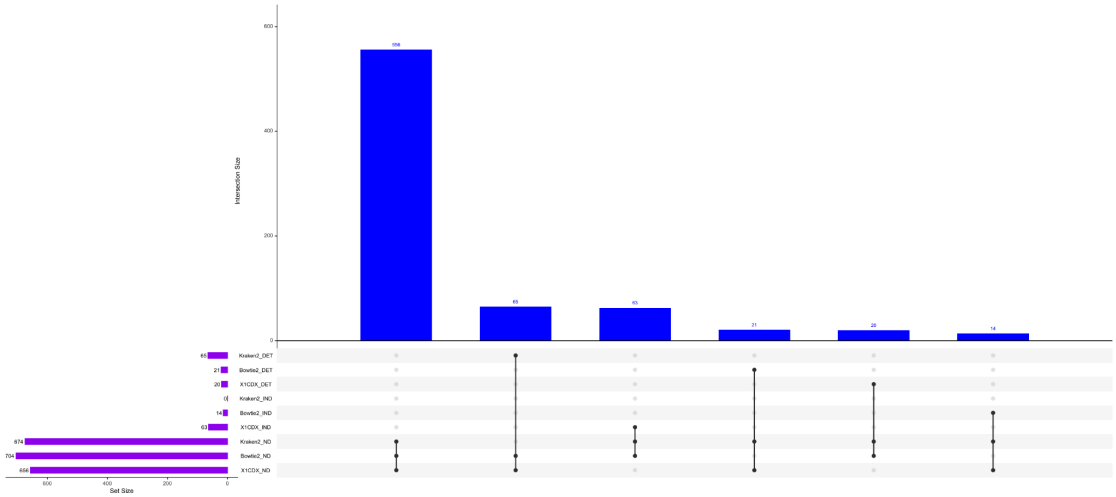

Figure 2.10. Wastewater upset plot. The data used to generate this plot is based on the Kraken2 detections of viruses as well as Bowtie2/One Codex detections that were considered “determinate” or “indeterminate”. This information was extrapolated into a binary matrix used to create the upset plot. The purple bars represent the counts of each method/detection combination. The three vertical dots represent the combination of methods and determinations that were seen, and the blue bars represent the number of times that particular combination was seen.

The overwhelming majority of clinical and wastewater samples had viruses that were considered “not detected” across all three methods, and there are no instances where each individual sample for a particular virus agreed on a “determinate” call across all three methods [Figures 2.9 & 2.10]. With low breadth of coverage and low average sequencing depths common in both sample types, a non-detected status is reasonable- though it is not particularly informative. Detections of viruses called with Kraken2 are the most frequent outside of the non-detects across both methods. This supports the idea that the k-mer based alignment approach is slightly more permissive than a direct alignment approach using the parameters that we specified in our Bowtie2 alignments.

Taken together, this study concludes that while the Twist Respiratory Viral Research Panel is a useful tool for surveilling a large number of respiratory viruses at once, it fell short of our expectations. While this study highlights a low-cost, multi-target alternative to PCR-based testing that is adaptable for various sample types, our results suggest further refinement is needed to confidently connect wastewater and clinical data for predicting viral trends.

# CHAPTER 3: MATRIX EFFECTS AND TARGET-CAPTURE SEQUENCING ACCURACY IN WASTEWATER SURVEILLANCE OF HUMAN RESPIRATORY VIRUSES

While our study shows the Twist Respiratory Viral Research Panel's capacity in analyzing clinical and wastewater samples, the complex nature of wastewater introduces unique challenges that may impact assay accuracy. The complex microbial matrix of wastewater can sometimes hinder the effectiveness of certain sequence-based molecular detection techniques. To address this issue, we evaluate how wastewater matrices influence sequencing outcomes by testing known RNA mixtures in both nuclease-free water and wastewater samples. We used taxonomic classification software and direct alignment methods to determine how well the sequencing assay reflected the contents of the RNA mixtures. The rate of recovery of expected viral taxa was higher in the nuclease-free water spike-ins than the wastewater spike-ins, suggesting that there may be a matrix effect of wastewater on the efficacy of this assay. These results highlight the importance of benchmarking laboratory and computational tools to ensure the most accurate information is being reported from community-level wastewater monitoring to broader public health surveillance.

## 3.1 Introduction

Human respiratory viruses are significant contributors to respiratory infections, which pose a significant public health concern worldwide. Viral respiratory infections place a consequential burden on healthcare resources, leading to illnesses of varying severity - from mild to severe - affecting people of all ages. Their ease of spreading from person to person regardless of symptoms makes them prime candidates for widespread outbreaks of disease. A timely example of this is the recent pandemic caused by the SARS-CoV-2 virus. Researchers and healthcare professionals needed time to determine the best ways to treat COVID-19 and track its spread in communities. Initially, this was accomplished by way of clinical testing results from PCR-based molecular detection methods. Positivity rates were then reported to local, city, state, and national health officials before being communicated to the public. Clinical test results were informative but did not provide a complete picture of the occurrence rates of SARS-CoV-2 because they only included information from symptomatic individuals who sought a clinical test. In an effort to address this bias towards symptomatic individuals, researchers began testing wastewater for the presence of SARS-CoV-2 RNA. Wastewater testing for SARS-CoV-2 offered a clearer picture of infection rates by measuring RNA concentrations in specific communities, regardless of whether individuals were showing symptoms. Building on its success, researchers applied WBE to detect other fecal-shed respiratory viruses of public health concern to begin communicating similar trends to the public. Examples of this can be seen in various contexts. Respiratory syncytial virus (RSV) and influenza viruses have been popular targets for wastewater monitoring projects and have shown favorable success when correlated with clinical data

(Toribio-Avedillo et al., 2023). Recently, advancements have been made in WBE that surveil for viruses outside of the usual suspects, including those that cause the common cold like adenoviruses, parainfluenzaviruses, metapneumoviruses, and more (Rector et al., 2024).

Although WBE is effective for monitoring human respiratory viruses, wastewater remains a challenging sample type to work with for several reasons. Wastewater is a complex matrix with a distinct composition that can hinder the effectiveness of common molecular detection methods. Sewersheds collect wastewater from sources that extend beyond the toilet, including residential/industrial greywater, agricultural leachate, and stormwater runoff (Henze et al., 2019; Rachana Dubey & Subhasis Swain, 2019). Wastewater not only contains fecal-shed viral RNA but also includes various biological and chemical compounds that can inhibit PCR, leading to challenges in recovering, detecting, and reproducing viral material (Kumblathan et al., 2022). Extensive efforts have been made to address these wastewater-specific inhibitory issues in PCR-based assays and many commercial solutions have been developed in response to this widespread problem. Filtration and concentration of wastewater before RNA extraction have greatly improved the detection of viral material in molecular tests. A 2021 publication evaluated several of these methods ahead of testing for SARS-CoV-2 and concluded that wastewater concentration via hollow or ultrafiltration concentrating pipette tips improved the sensitivity of viral detection in downstream qPCR assays (Juel et al., 2021). This method is among one of the official recommendations for SARS-CoV-2 viral RNA concentration from wastewater by the Association of Public Health Laboratories (*SARS-CoV-2 Wastewater Surveillance Testing Guide for Public*

*Health Laboratories*, 2022). Further, RNA extraction methods have been optimized to isolate as much viral material as possible. A recent study evaluated several of these modifications in the context of SARS-CoV-2 detection via PCR-based assay (Zafeiriadou et al., 2024). Many commercial solutions are available to address the specific issues in wastewater that inhibit detection, enabling more accurate PCR-based molecular testing.

PCR-based methods are effective for monitoring human respiratory viruses in wastewater, but analyzing the viral material at a whole-genome level can provide additional valuable insights. Following viral material enhancement and RNA extraction, next-generation sequencing (NGS) of wastewater-originating RNA provides an alternative to standard PCR-based molecular detection methods. Similar to PCR-based methods, NGS methods can be customized during library preparation and beyond to increase the amount of viral material detected. Of note are two key methods: amplicon-based and hybridization/target capture-based NGS. Amplicon-based NGS is a viable option for high-throughput sequencing of specific pathogens that mutate frequently, giving rise to genetic variation. Briefly, short sequence amplicons are generated from overlapping sets of primer pairs that span the entire genome of a single pathogen. In this way, these amplicons capture the highly variant regions of the target genome, allowing for the identification of variants. In the case of SARS-CoV-2 wastewater monitoring, this was an attractive option to identify and monitor variants of concern and it is the standard NGS protocol used by the CDC for this purpose (Rasmussen et al., 2022). The amplicon-based NGS approach generates enough sequencing depth for reliable results, but it requires prior knowledge of the target pathogen to select the appropriate primer schemes for generating amplicons. Further, it is

limited to a single target genome. There is some debate on whether the use of multi-variant primers in a PCR-based molecular detection assay vs. an amplicon-based NGS assay is most appropriate for these cases. The amplicon-based NGS method fails to outperform the PCR-based methods in scenarios where the pathogen spans multiple taxa. Prior knowledge of the expected taxa is necessary to select primers that will amplify in these assays, and each assay is limited to the number of fluorescent probes that the qPCR/ddPCR instrument can detect simultaneously. It is here where the hybridization/target-capture based NGS methods can truly demonstrate their utility in wastewater virome studies.

Hybridization/target-capture based NGS assays work by combining a panel of probes that cover target sequences across multiple genomes. If these sequences are present in a sample after library preparation, they will bind to the biotinylated probes in the next step. This allows them to be separated from off-target sequences using magnetic bead selection. This crucial step allows for off-target sequences to be washed away, ensuring that only the fragments of interest move forward to be sequenced. This approach is cost-effective because it avoids wasting sequencing resources on irrelevant information and focuses on deep sequencing the important data that is present. Though hybridization/target-capture based assays are effective ways to specify only target sequences for downstream NGS, they are not without flaws. Hybridization-based assays sometimes fall victim to a lack of specificity, noticed first in the early days of gene expression studies using microarray technology. Issues with oligonucleotide microarray hybridizations include low accuracy, precision, and specificity due largely in part to non-specific binding in cross-hybridization (Draghici et al., 2006; Jaksik et al., 2015).

Factors that influence the efficiency and specificity of a hybridization reaction include the probe sequence and the hybridization temperature (Halling & Wendel, 2009). Some similarity between on and off-target sequences presents an opportunity for a probe to bind to a sequence that is not its exact complement, resulting in the capture of information that does not accurately represent the sequence of origin (Reilly et al., 2006). This issue is common in studies where probe sets target genome sequences with varying levels of homology, making it harder to interpret the results of a hybridization assay (Evertsz et al., 2001). While these challenges with hybridization-based assays do exist, many strategies have been developed and tested to improve their efficacy and accuracy.

Hybridization and target capture-based NGS methods are popular in virome studies because they can detect a wide variety of targets. There are several commercial choices for probe sets and panels that target hundreds of viruses from many different families, genera, and species. Twist Bioscience offers a Comprehensive Viral Research Panel that covers reference sequences for 3153 different DNA/RNA viruses across a set of ~1 million unique, 120-basepair probes (*Novel Virus Detection Using the Twist Comprehensive Viral Research Panel*, 2023). Recent studies detailed the use of this particular pan-viral hybridization panel in wastewater (Tisza et al., 2023). Although the probes in this panel are unique and can detect novel and existing human viral pathogens, it is important to consider the previously mentioned specificity issues, particularly when testing in a wastewater matrix. In particular, recent studies highlighted that broader panels tend to have lower sensitivity for low-abundance viruses. They also include more probes that may capture background sequences similar to the targets found in wastewater (Kantor & Jiang, 2024). In light of this, several hybridization and target capture-based

panels have been developed to focus on a smaller group of viral species, prioritizing key pathogens that cause respiratory infections. Several comparative studies have been carried out to evaluate and contrast these respiratory viral panels as they relate to patient samples (Kantor & Jiang, 2024; Kapel et al., 2023; Rehn et al., 2021).

Outside of testing control mixtures of synthetic RNA spiked into a sterile background, there are no studies that offer insight into the associated matrix effect of wastewater on the target capture-based sequencing method (*NGS Target Enrichment of Viral Pathogens Using Twist Respiratory Virus Research Panel: Application Note*, 2021). Here, we demonstrate the use of the Twist Respiratory Virus Research Panel in wastewater respiratory viral detection. Our study offers a unique look into the viral recovery efficiency of this assay in both wastewater and sterile backgrounds by way of various bioinformatics methods. Additionally, this study aims to determine if mixtures of human respiratory viruses from different groups cross-react with each other, which could obscure the expected results from the spike-ins. Taken together, the results of this study also present an opportunity to discuss appropriate use cases and downstream bioinformatics analysis for this method of sequencing.

## 3.2 Methods

### 3.2.1. Creation of control mixtures

Nine synthetic RNAs (Twist Bioscience, San Francisco, CA, USA) were selected to reflect ssRNA segments targeted by Twist Bioscience's Respiratory Viral Panel. The stock solutions were diluted to a concentration of $1x10^4$ gene copies (cp)/μL, various volumes of which were used to create mixtures representing various combinations of viral taxa [Supplementary Tables 1 & 2]. These synthetic RNA mixtures were then spiked into both nuclease-free water and composite wastewater matrices.

#### 3.2.1.1. Synthetic RNA spiked into nuclease-free water

Synthetic RNA mixtures were spiked into a nuclease-free water (NFW) matrix according to the volumes detailed in Table 1. Due to the sterility of the matrix, samples spiked into nuclease-free water required the addition of Universal Human Reference RNA (ThermoFisher, Waltham, MA) to ensure appropriate material for hybridization before loading onto a 96-well plate for subsequent library preparation.

#### 3.2.1.2. Synthetic RNA spiked into wastewater matrix

Synthetic RNA mixtures were spiked into a composite wastewater mixture and loaded onto a 96-well plate for subsequent library preparation.

##### 3.2.1.2.1. Wastewater sample collection and concentration

As part of an ongoing WBE project funded by NCDHHS, wastewater samples were collected on various dates from wastewater treatment plants across several counties in North Carolina. Briefly, up to 10L composite wastewater samples were auto sampled every 30 minutes

over a 24-hour period and collected into a bottle. 60mL of each sample was partitioned and centrifuged at 10,000 x *g* to remove solids, and the supernatant was then subject to filtration through 0.05 μm PS Hollow Fiber Filter Tips (Innovaprep) using the automatic CP Select (Innovaprep). Viral particles were then eluted using 0.075% Tween-20/Tris WetFoam elution fluid (Innovaprep) into a final volume ranging from 250 μL to 500 μL and frozen for storage until they were sent to the lab at UNC-Charlotte for RNA extraction and quantification (Juel et al., 2021).

3.2.1.2.2. RNA extraction and SARS-CoV-2 detection

RNA extraction of wastewater samples was performed using the Ceres *Nanotrap*® Microbiome A Automated Protocol (Ceres Nanosciences, Manassas, VA) with *MagMAX*™ kit and the *KingFisher*™ Flex Purification System (Thermo Fisher Scientific, Waltham, MA) according to the manufacturer's instructions. 20 μL of extracted RNA was then used in downstream SARS-CoV-2 detection using a droplet-digital PCR assay targeting the N1/N2 genes.

3.2.1.2.3. Selection of representative wastewater samples

11 of the extracted RNA samples that were designated as "negative" following a concentration of 0.00 cp/μL for the N1/N2 gene in the ddPCR assay were selected for use as part of a representative wastewater matrix [Supplementary Table 3]. 20 μL of each sample were pooled together to create the wastewater matrix used for the RNA mixture spike-ins for a total volume of 220 μL.

3.2.2. Library preparation

DNA fragmentation, end repair, and dA-tailing were conducted following the Twist Bioscience cDNA Library Prep Kit for ssRNA Virus Detection (Twist Bioscience, San Francisco, CA, USA). A Qubit 1X dsDNA HS assay (Thermo Fisher, Waltham, MA, USA) was used to quantify the samples, ensuring sufficient genomic material was present for hybridization. After preparing the dual-indexed and purified libraries, hybridization was performed following Twist Bioscience's Target Enrichment Protocol. Post-PCR amplification, the libraries were pooled together and re-quantified with a Qubit 1X dsDNA HS assay as well as a High Sensitivity D1000 ScreenTape Assay for TapeStation (Agilent Technologies, Inc., Santa Clara, CA, USA) to confirm adequate genomic input and fragment size for sequencing on an Illumina platform.

3.2.3. Sequencing on an Illumina NextSeq 2000 platform

The DNA libraries were diluted to a final concentration of 750 pM for sequencing on the Illumina NextSeq 2000, following the procedures outlined in the Illumina DNA Prep Reference Guide (Illumina Inc., San Diego, CA). Paired-end 150 bp reads were generated, and quality-filtered reads were uploaded to the Illumina BaseSpace Sequence Hub. Here, additional analyses were carried out using DRAGEN Bio-IT, a cloud-based tool provided within the platform.

3.2.4. Bioinformatics

The raw sequence data generated from the Illumina BaseSpace Sequence Hub was downloaded as compressed .fastq files and stored on UNC-Charlotte's high-performance computing cluster. Trimmomatic v0.39 was used to remove adapter sequences and low-quality bases (ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3

TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36), followed by a quality check with FastQC v0.11.9 using default settings (Bolger et al., 2014; S Andrews, 2010).

Taxonomic classification was conducted using Kraken2 v2.1.3 against the NCBI viral database (as of July 15, 2024) and relative abundance re-estimated via Bracken v2.9, specifying 150 bp reads and a minimum of 10 reads for classification at the species level (-r150 -t10 -lS) (J. Lu et al., 2017; Wood et al., 2019).

Reads were aligned to reference genomes from Twist's Respiratory Viral Research Panel using Bowtie2 v2.4.1 with sensitive local alignment parameters, allowing one mismatch (--very-sensitive-local -N 1) (Langmead & Salzberg, 2012). The mapped reads were sorted by genome, and the average depth and breadth of coverage were calculated. Relative abundance was determined by dividing the mapped reads by the total reads, multiplied by 100.

Per-genome depth across positions was calculated using SAMtools v1.10, including zero-coverage positions (-a), while average depth was computed by dividing the total coverage by the number of positions in the genome (Li et al., 2009). Breadth was calculated as the ratio of covered positions to total positions. MosDepth was used to compute median coverage across 500-bp genome chunks at varying thresholds (--use-median --by 500 --thresholds 1,10,20,30) (Pedersen & Quinlan, 2018).

All statistical analyses and figures were created using RStudio v4.3.2.

## 3.3 Results and Discussion

18 mixtures of 10 different synthetic control RNA targeted by the viral panel were spiked into both nuclease-free water and composite wastewater backgrounds [Supplementary Table 1]. The wastewater background samples had ddPCR tested negative for the SARS-CoV-2 N1/N2 gene. The composite wastewater background included samples from nine unique sites collected on 10 different dates between spring 2023 and spring 2024. Each mixture created in nuclease-free water was duplicated in the wastewater background for coherence throughout the study.

The samples and mixtures detailed above underwent the same library preparation detailed in the methods, as well as sequencing on an Illumina NextSeq 2000 instrument, producing 39.76 Gbp of data with an 85.23% average of reads at Q30, and 77.38% of reads passing Illumina's internal quality filter.

### 3.3.1. Sequencing performance across background type

Wastewater samples averaged 501683 reads per sample, whereas nuclease-free water samples averaged 1522212 reads per sample (a nearly 33% increase from wastewater samples). Breadth of genome coverage at 1X and average sequencing depth were calculated for all genomes in both types of sample [Figures 3.1, 3.2, Supplementary Table 5]. A one-way ANOVA revealed a marginally significant effect of sample type on the breadth of genome coverage, where $F(1,373) = [2.85]$, $p = 0.0924$. The same test revealed that there was a significant effect of sample type on the average sequencing depth, where $F(1,373) = [27.2]$, $p = 3.05e\text{-}07$.

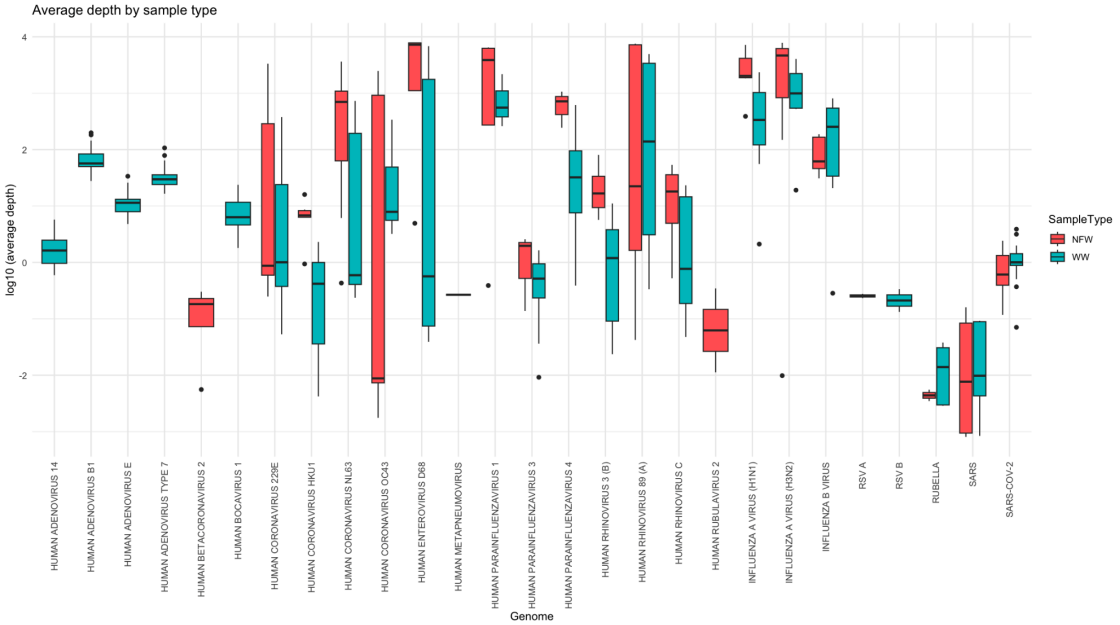


Figure 3.1. Average depth of genome coverage across all target genomes present in nuclease-free water (red) and wastewater (teal) samples. Average sequencing depth is reported on a logarithmic scale (base 10).

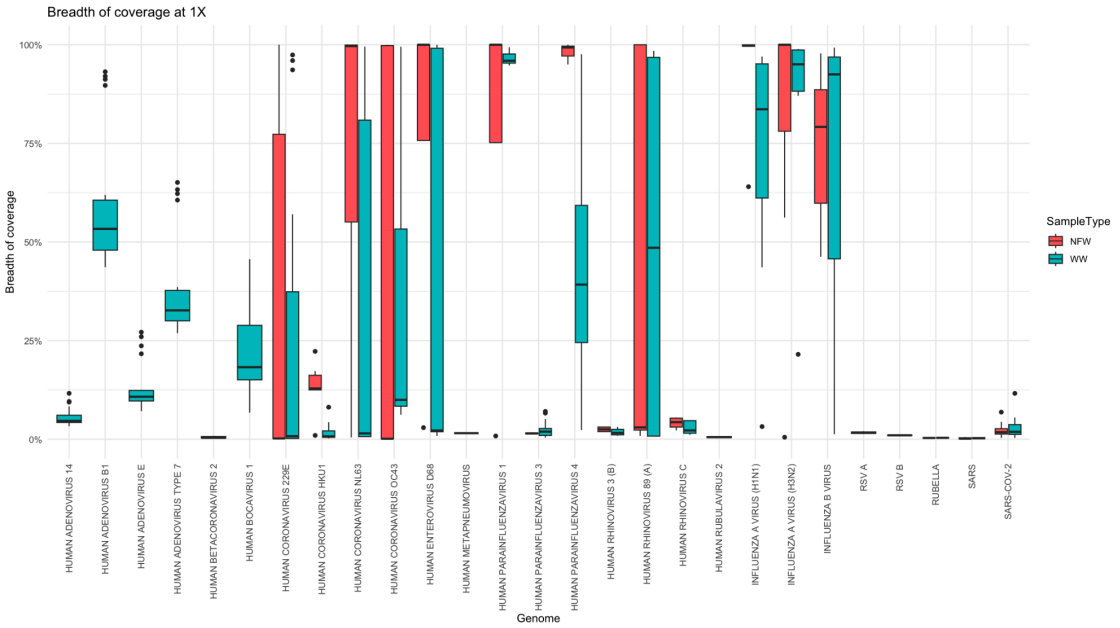


Figure 3.2. Distribution of breadth of genome coverage at 1X across all target genomes present in nuclease-free water (red) and wastewater (teal) samples.

In wastewater samples, certain viruses were detected exclusively, including RSV B, human bocavirus 1, and multiple human adenoviruses (types 14, B1, E, and 7). In contrast, nuclease-free water samples exclusively detected human beta coronavirus 2 (MERS) and human rubulavirus 2. Generally, nuclease-free water samples exhibited higher average depths for viruses such as human coronaviruses 229E, HKU1, NL63, and OC43, as well as enterovirus D68, human parainfluenza viruses 3 and 4, human rhinoviruses (types A, B, and C), and both strains of influenza A. On the other hand, wastewater samples showed higher average depths for influenza B, SARS, rubella, and SARS-CoV-2. Distributions were the most similar between wastewater and nuclease-free water samples for the human rhinovirus A genome. These results suggest that there are differences in either viral abundance or capture efficiency for some viruses in these different sample matrices.

In the case of breadth of genome coverage, nuclease-free water samples demonstrated broader coverage across most detected genomes as compared to wastewater samples. The exceptions were SARS-CoV-2, influenza B, and parainfluenza virus 3, where wastewater samples achieved better breadth of genome coverage. This variation highlights how the sample matrix can influence the capture of different viral genomes.

### 3.3.2. Stronger alignment of observed and expected taxa abundances in nuclease-free water spike-ins

The observed vs. expected (O/E) ratios of taxa spiked into each sample type were calculated based on the relative abundance that was reported using Bowtie2 alignments. A stark difference in the O/E ratio is visible between the nuclease-free water and

wastewater spike-ins, with nuclease-free water spike-ins outperforming their wastewater counterpart for all viruses except influenza B [Figure 3.3].

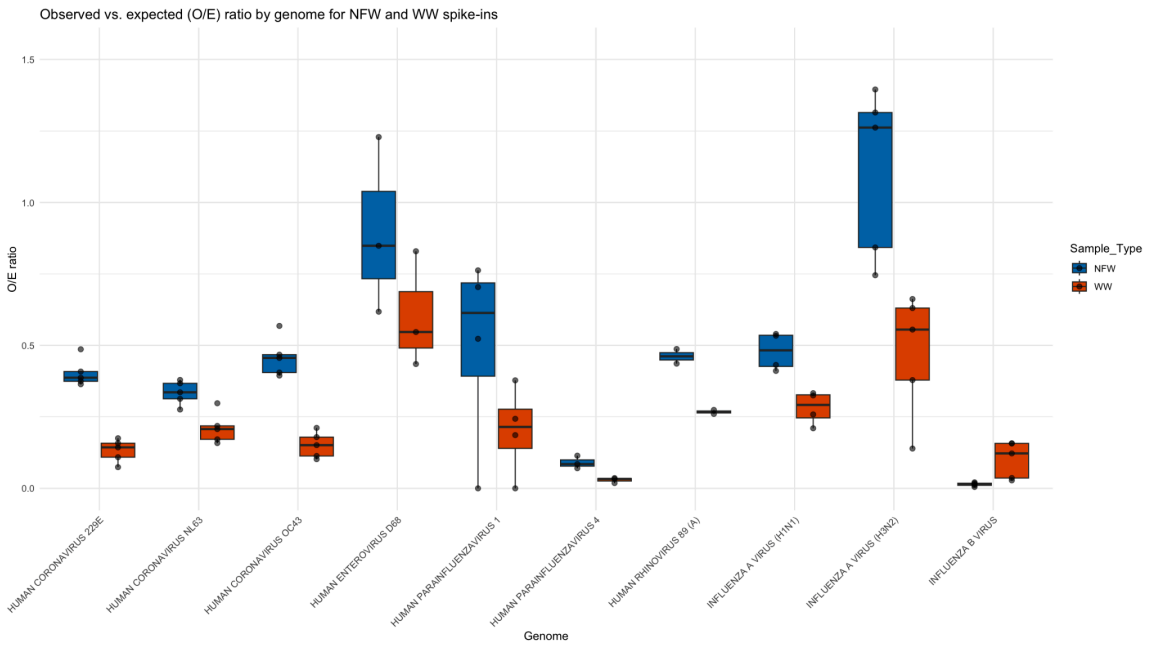


Figure 3.3. Observed vs. expected ratios for each viral genome spiked into both types of control mixture. Blue box and whisker plots represent the distribution of O/E ratios in nuclease-free water samples, and orange box and whisker plots represent the distribution of O/E ratios in wastewater samples.

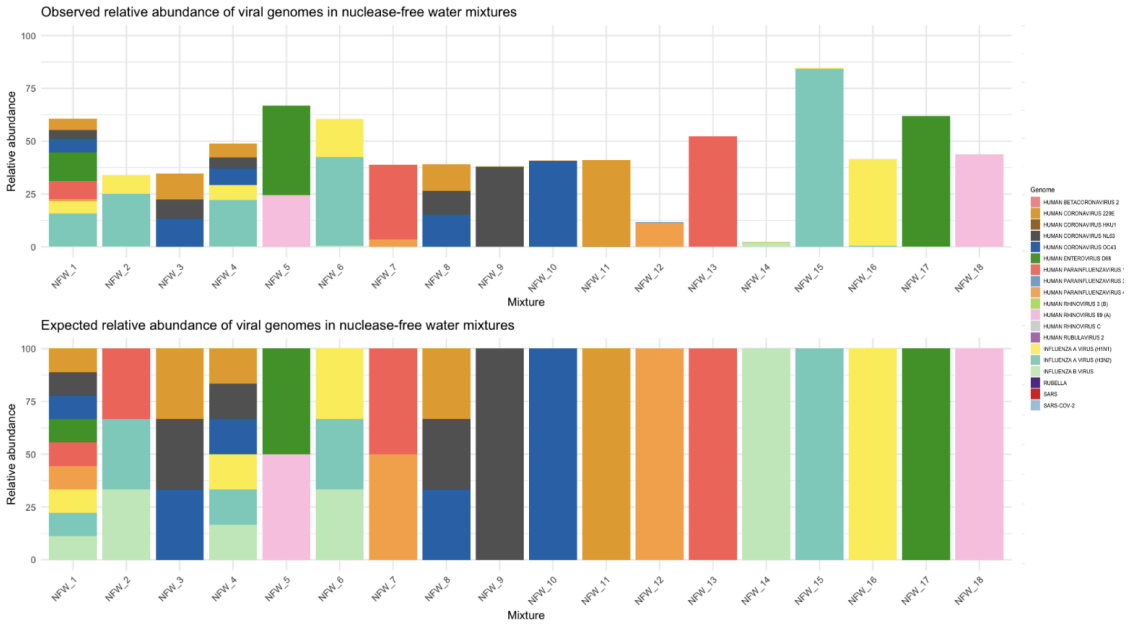


Figure 3.4. Observed vs. expected relative abundances in nuclease-free water mixtures. Observed relative abundances reported based on presence identified with Bowtie2 alignments Bowtie2 alignments to the reference genomes included in the Twist Respiratory Viral Research Panel.

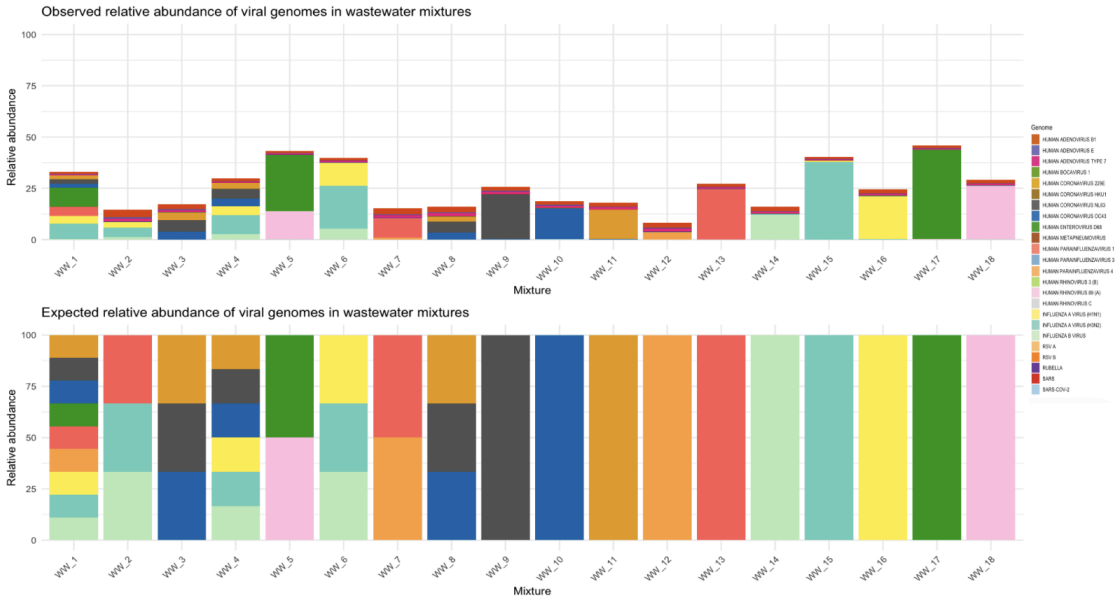


Figure 3.5. Observed vs. expected relative abundances in wastewater mixtures. Observed relative abundances reported based on presence identified with Bowtie2 alignments to the reference genomes included in the Twist Respiratory Viral Research Panel.

The O/E ratio provides strong evidence of a wastewater matrix effect at play in this kind of hybridization and target-capture based sequencing assay. To investigate this further, the individual O/E abundances of taxa within each mixture were examined. Using bar plots for each mixture, we assessed how much of each genomic component was recovered across sample types [Figures 3.4 & 3.5]. In both sample types, the observed relative abundance of respiratory virus target taxa never reached 100%. While this outcome is common in wastewater spike-ins for some other NGS assays, it is important to note that the respiratory viral panel was expected to select for and sequence only the specific targets included in the probe set. With this in mind, the entirety of the reads generated from this experiment should represent genomes from within the probe set, but they do not. Perhaps this is a more reasonable outcome to expect in a wastewater spike-in, but observing this in a nuclease-free water spike-in raises a few questions. A recent study demonstrated a similar pattern of frequent off-target hits in wastewater samples when using the Twist Respiratory Viral Research Panel and alignment-based bioinformatics tools, prompting further investigation into their origin [Roppolo Brazell et al., 2024]. Nuclease-free water samples recovered a far less relative abundance of genomes that were spiked into them, though this can be explained in part due to the addition of Universal Human Reference RNA to this type of spike-in [Supplementary File 1]. One of the most dramatic differences was noticed in the case of influenza B: in every expected mixture that included this virus, recovery was observed to be markedly decreased. Similar results were noticed in human parainfluenzavirus 4. Nuclease-free water samples also display low relative abundances of reads that aligned to SARS-CoV-2 and human coronavirus 229E, except for where human coronavirus 229E was expected.

Finally, nuclease-free water spike-ins did demonstrate some cross-reactions of closely related genomes. This can be seen in the cases of members of the *Orthocoronavirinae* family in NFW9, 10 & 11, as well as in NFW15 & 16 in the *Alphainfluenzavirus* genus/species. While they are members of the same viral family, human coronaviruses 229E and OC43 belong to two different genera; *Alphacoronavirus* and *Betacoronavirus*, respectively. Minor sequence differences between these two species occur in the surface spike protein-coding gene that interacts with the host receptor (Kistler & Bedford, 2021). Similarly, there are sequence differences in the host-receptor binding sites between closely related influenzavirus A subtypes H1N1 and H3N2 (Y. Liu et al., 2023). Considering the stringency of the Bowtie2 parameters, this is not unprecedented: our bioinformatics analysis allowed for a single mismatch in a very sensitive local alignment.

In wastewater samples, similar results were observed concerning the recovery of a lower relative abundance of viral genomes as well as the cross-reactions of members of the *Orthocoronavirinae* family and *Alphainfluenzavirus* genus/species. Wastewater samples did demonstrate the inclusion of targets that were outside of the spike-ins, but this was to be expected in a wastewater background where the composition was unknown outside of SARS-CoV-2. In all wastewater samples, reads that aligned to the SARS-CoV-2, human adenoviruses B1, E, and type 7, human bocavirus 1, and human coronaviruses 229E and OC43 genomes were detected at low relative abundances except for where they were expected in a spike-in. It is reasonable to conclude that the wastewater background used for the spike-ins consisted of low levels of these genomes. Further examination of the remaining reads that did not align with the spike-in reference genomes revealed off-target hits. A brief Kraken2 analysis showed these reads were often

classified as bacteria, fungi, and archaea, highlighting the lack of specificity of the probes in this panel [Supplementary File 2].

3.3.3. Potential kit-associated microbial background

Three no-template controls underwent library preparation alongside the nuclease-free water and wastewater spike-in samples as an internal control. Interestingly, these samples showed an overwhelming majority of reads mapping to SARS-CoV-2- the highest of which showed a relative abundance of 96% [Supplementary Figure 1]. Reads with marginal relative abundance (> 0.01%, > 5.0%) mapped to human coronavirus 229E and SARS in all three samples, respectively.

While this result does not align with what we would expect to see in a negative control made exclusively of sterile molecular-biology grade water, it is not unprecedented: not only did we see low levels of SARS-CoV-2 relative abundance in all nuclease-free water and wastewater spike-ins where we didn't expect to see its presence at all, we also experienced similar results in previous experiments. A prior iteration of this experiment was performed ~eight months before the one detailed in this publication. This experiment differed from the one presently described in that it included synthetic SARS-CoV-2 variant RNA in its control mixtures. What was found in this experiment was that there was an overwhelming majority of reads in the nuclease-free water negative control mapped to the SARS-CoV-2 reference genome using k-mer based classification software Kraken2 (Wood et al., 2019; Supplementary Table 6). This discovery prompted a second attempt at this exact experiment, this time utilizing triplicates of the no-template controls, new reagents, thorough decontamination measures, different laboratory personnel for library preparation, and eliminating all SARS-CoV-2 variants from the

control RNA mixtures to ensure this type of contamination would not be a reason for failure. Despite improvements in the experimental design, similar outcomes were observed, suggesting the possibility of underlying contamination mechanisms or the presence of a kit-derived background microbiome influencing the assay results.

One major difference between the no-template controls and the nuclease-free water spike-ins, aside from containing no viral RNA mixtures, was that they did not contain any Universal Human Reference RNA. Perhaps the potential for the large relative abundance of these taxa is due in part to the absence of this RNA as its purpose is to improve capture efficiency of hybridization-based assays, but for a truly negative control, we would not expect to see the presence of any viral material.

Interestingly, Twist Bioscience released an application note that suggested the preparation of a negative control for library prep outside of the preparation of experimental samples (*NGS Target Enrichment of Viral Pathogens Using Twist Respiratory Virus Research Panel: Application Note*, 2021). In short, the note suggests that negative control data shows that even small cross-contaminations can be detected when using this hybridization-based panel. They recommended preparing viral control and negative control libraries separately, either at different times or in physically separate locations, or by leaving blank wells between samples to prevent cross-contamination risks.

## 3.4 Conclusions and Future Work

This study investigates the effectiveness of a hybridization and target-capture based NGS assay specific to respiratory viruses, focusing on the potential impact of wastewater matrix on capture efficiency. These findings indicate that there is a slight matrix effect that should be taken into account when employing this NGS method in wastewater analysis. Our methods indicate that concentrating wastewater and removing inhibitors are effective for successfully applying this assay; however, further exploration and optimization of additional viral material enrichment techniques alongside this assay are recommended. Enhancements in upstream viral material enrichment before library preparation also have the potential to improve the breadth of genome coverage in wastewater samples. Although the assay demonstrates high capture efficiency in wastewater, it may still introduce some errors that could impact the interpretation of the results. In future experiments, sequencing the wastewater background used for spike-ins will provide insights into its contents, aiding in the differentiation between spike-in components, true background signal, and potential off-target hits. Accurate identification of off-target hits could facilitate the creation of a bioinformatics tool designed to filter out common contaminants that may interfere with result interpretation.

# CHAPTER 4: NEXT-GENERATION SEQUENCING FOR VIROME SURVEILLANCE: ASSESSING THE EFFICACY OF ILLUMINA'S RESPIRATORY VIRAL OLIGONUCLEOTIDE PANEL IN WASTEWATER

While our findings show promising capture efficiency of target capture-based NGS assays in wastewater, the matrix effect remains a variable that could influence accuracy. To further evaluate the efficiency of capture-based technologies in wastewater, we evaluate the Illumina Respiratory Virus Oligo Panel alongside the Twist panel, comparing their performance and specificity in wastewater samples to better understand their applications for respiratory virus monitoring. This study marks the first direct evaluation of these panels in wastewater samples. Although the Illumina panel produced limited genome coverage and low average sequence depth for targeted viral pathogens, it also offered insights into cross-reactivity with homologous sequences and off-target taxonomic classifications. In the tested samples, there were no off-target alignment hits detected at the genus level, and few cross-reactions observed at the species level. This observation positions the Illumina panel above the Twist Biosciences panel, though more 1:1 parallel testing is necessary to determine which panel outperforms the other. The overall findings of this study highlight the potential utility of these types of panels for wastewater applications, with important considerations for coverage and specificity that could inform broader public health monitoring efforts.

## 4.1 Introduction

Human respiratory viruses are quickly transmitted from person to person, where symptoms can range in severity from mild to acute, and place a large burden on healthcare systems that treat these illnesses in high-risk individuals (Kutter et al., 2018; Zimmerman et al., 2022). Viral detection and monitoring is an important component of public health. Point-of-care tests for specific human respiratory viruses are commonplace in clinical settings but require some degree of prior knowledge of a specific pathogen to test for it, and are often limited in scale by testing one or a few pathogens and not a whole array of respiratory viruses. A more passive approach,  wastewater-based epidemiology (WBE), has developed on a larger scale following the pandemic to allow for composite testing of populations as a whole instead of relying on individual tests. Initially used in the 1930s, WBE surged in popularity during teh SARS-CoV-2 pandemic when it was shown to be a sentinel for predicting outbreaks in the population being monitored (Paul et al., 1939; Singer et al., 2023). WBE as an early alert warning system for SARS-CoV-2 allowed officials to monitor the presence and quantity of the virus in wastewater regardless of whether community members were symptomatic, guiding their decisions on important public health policies (Ahmed et al., 2021). This monitoring capability, in a post-pandemic world, is being extended to surveil other fecal-shed human respiratory viruses to assist in keeping the public safe and informed of their risk (Ahmed et al., 2023)

Molecular methods are commonly used to test for the presence of human respiratory viruses in clinical and wastewater contexts, and PCR-based assays are the standard for viral detection in both cases. For example, the US Centers for Disease Control and Prevention (CDC) recommend SARS-CoV-2 detection via quantitative reverse-transcriptase PCR (qRT-PCR), though many testing centers use quantitative reverse-transcriptase droplet digital qPCR

(qRT-ddPCR) to glean a more sensitive measurement of viral material (Vasudevan et al., 2021). Some of these molecular methods allow for the multiplexing of several gene targets, the number of which is limited to the fluorescent capacity of the assay and the instrument being used. While these methods are effective in the detection of human respiratory viruses in clinical and wastewater samples, wastewater remains challenging sample typeas it is a complex matrix of organic and inorganic matter that has inhibitors that can affect the performance of the assay if not handled appropriately. These targeted assays can detect variants, but require a priori information to design and test probes for detection, making novel variant identification, new mutations, and other nucleotide-level changes challenging to detect early. Next-generation sequencing (NGS) methods overcome these target level challenges, but are not with tradeoffs in the bioinformatics required to interpret the results. Commonly used NGS methods for human respiratory viral detection include whole genome shotgun (WGS) sequencing, amplicon-based sequencing, and hybridization-based sequencing. WGS sequencing allows for the full genome sequence to be fragmented and sequenced, however, the drawback of this method is the difficult bioinformatic analysis downstream of sequencing in a complex sample such as wastewater encompassing population level data. Amplicon-based sequencing allows for the generation of short amplicons that tile a viral genome, providing greater resolution of the genome in areas that are prone to variation, but only targets one genome of interest, not the whole population of viral particles in a sample. Multiplexing for multiple organisms that are not closely related is generally not supported. The final method, hybridization-based sequencing, is an attractive alternative that incorporates the positives from both of the aforementioned methods.

The target capture/hybridization-based NGS approach hinges on the use of a specific panel of probes to hybridize only with the target sequences of interest. All unbound sequence

fragments are washed away during library prep, ensuring that only the intended genomic material is sequenced and analyzed. These hybridization panels excel in multiplex capabilities and can span multiple genomes over hundreds, sometimes thousands, of probes. With this said, there are some drawbacks to the target capture/hybridization NGS method.

Several target capture/hybridization-based sequencing panels have been developed for use in the detection of human respiratory viruses. Twist Biosciences has developed a Respiratory Viral Research Panel of probes for this purpose. This panel, along with their library preparation and hybridization assay kits, is capable of detecting 29 different respiratory viruses using ~41000 probes at viral titers as low as 100 copies of viral material per virus (*Twist Respiratory Virus Research Panel: Twist NGS Target Enrichment Panel for Accurate Identification of Respiratory Viruses beyond SARS-CoV-2*, 2020). While it has not exclusively been tested for use in wastewater outside of our lab at the time of this publication, we demonstrated that it is possible to use this panel to enrich respiratory virus sequences from wastewater samples. Another commercial solution for target capture/hybridization of human respiratory viruses has been developed by Illumina, Inc., in their Respiratory Viral Oligonucleotide Panel (RVOP). This panel targets 41 different respiratory viruses using ~7800 probes, and has been validated for nearly whole-genome coverage of SARS-CoV-2 at as low as 1000 copies of viral material per virus (*Detection and Characterization of Respiratory Viruses, Including SARS-CoV-2, Using Illumina RNA Prep with Enrichment*, 2020). In contrast, the Illumina RVOP has been tested in several studies using wastewater, showing moderate success. The RVOP was used in wastewater testing in 2021 (though it placed an emphasis on SARS-CoV-2) and demonstrated that sufficient quality reads were generated from this sequencing experiment, leading to the characterization of SARS-CoV-2 variant identification as well as the identification of relative abundance of other

respiratory viruses targeted by the panel (Crits-Christoph et al., 2021). Another 2021 SARS-CoV-2 focused study used the RVOP and found that detection of a variety of on-target human respiratory viruses post enrichment was possible from wastewater samples (Rothman et al., 2021). The most recent published study in 2023 determined that while the target capture/hybridization NGS methods can be used for the detection of human respiratory viruses in wastewater, the method best suited for the job with respect to optimal genome coverage and variant identification is amplicon-based sequencing (Child et al., 2023).

In summary, this study leverages the availability of target capture and hybridization-based sequencing panels for the detection of human respiratory viruses, aiming to validate the Illumina Respiratory Virus Oligo Panel (RVOP) on our wastewater samples. By comparing its performance with that of the Twist Respiratory Virus Research Panel, we address a notable gap in research, as no direct comparison of these two panels in wastewater samples has been conducted. This approach not only validates our bioinformatics pipeline but also allows us to assess the suitability of each method for wastewater sequencing, with the potential to offer reliable, actionable public health information. Achieving a high standard for virus detection in wastewater could allow these methods to provide insights similar to those from our established COVID-19 wastewater testing program.

## 4.2 Methods

### 4.2.1. Wastewater Sample Collection and Concentration

As part of an ongoing WBE project funded by NCDHHS, wastewater samples were collected on various dates from wastewater treatment plants across several counties in North Carolina, as well as several sites on campus at UNC-Charlotte. Briefly, up to 10L composite wastewater samples were auto sampled every 30 minutes over a 24 hour period and collected into a bottle. 60mL of each sample was centrifuged at 10,000xg to remove solids, and the supernatant was then subject to filtration through 0.05 um PS Hollow Fiber Filter Tips (Innovaprep) using the automatic CP Select (Innovaprep). Viral particles were then eluted using 0.075% Tween-20/Tris WetFoam elution fluid (Innovaprep) into a final volume ranging from 250uL to 500uL and immediately proceeded to RNA extraction (Juel et al., 2021).

### 4.2.2. RNA Extraction and SARS-CoV-2 Detection

RNA extraction of wastewater samples was performed using the Ceres *Nanotrap*® Microbiome A Automated Protocol (Ceres Nanosciences, Manassas, VA) with *MagMAX*™ kit and the *KingFisher*™ Apex (Thermo Fisher Scientific, Waltham, MA) according to the manufacturer's instructions. Finally, the RNA was treated with a one-step PCR inhibitor removal kit (Zymo, Irvine, CA, USA) to remove any impurities that might remain in wastewater samples after RNA extraction.

### 4.2.3. Library preparation

Extracted RNA was used as input for Illumina's RNA Prep with Tagmentation (L) and Enrichment library preparation kit (Illumina Inc., San Diego, CA). In brief, RNA is denatured, and cDNA libraries are generated and tagmented to enable multiplexing. After

a bead-based cleanup, probes that target the viruses in Illumina's Respiratory Virus Oligo Panel are hybridized to the sequence fragments, which are then magnetically captured and purified. Only the hybridized fragments of interest are then amplified and pooled together for sequencing.

4.2.4. Sequencing

The DNA pools were diluted to a final loading concentration of 650 pM and prepared for sequencing on an Illumina NextSeq 2000 instrument, according to the protocol described in the Illumina DNA Prep Reference Guide (Illumina Inc., San Diego, CA). Sequencing generated paired-end reads of 76 bp in length, and those that passed QC constraints were stored in Illumina's BaseSpace Sequence Hub where further analyses could be performed within their cloud-based environment via their built-in analysis tool, DRAGEN Bio-IT.

4.2.5. Bioinformatics

The raw sequence reads that were retrieved off the Illumina BaseSpace Sequence Hub were stored as zipped .fastq files and saved to UNC-Charlotte's high-powered computing cluster. They were trimmed for technical sequence/adapter content using Trimmomatic v0.39 (ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36), and then underwent a final quality check using FastQC v0.11.9's default parameters (Bolger et al., 2014; S Andrews, 2010).

A standard taxonomic classification against the NCBI viral database (built and accessed on July 15, 2024) was performed using default parameters for Kraken2 v2.1.3 followed by the use of its sister tool, Bracken v2.9, to re-estimate relative abundances of

taxa in each sample (J. Lu et al., 2017; Wood et al., 2019). Default parameters were used with the exception of specifying a read length of 150, a classification at the species level, and a threshold of 10 for the minimum number of reads required for classification at this level (-r150 -t10 -lS).

Read mapping was also performed by aligning reads to a multifasta file that contained the reference genomes of Illumina's Respiratory Virus Oligos Panel via Bowtie2 v2.4.1, allowing one mismatch in a very sensitive local alignment (--very-sensitive-local -N 1) (Langmead & Salzberg, 2012); Supplementary Table 1). Reads that mapped to the reference genomes were then sorted by genome, and the average depth/breadth of coverage was computed. Relative abundance was then calculated by dividing the number of reads mapped to the reference genomes by the total number of reads, multiplied by 100.

Depth for each genomic position within each sample was computed following analysis with SAMtools depth v1.10 [Li et al., 2009] with the -a parameter specified to include positions with zero coverage. Average depth per genome was calculated by dividing the sum of the coverage at each position by the number of positions in the genome, and breadth of coverage was calculated by dividing the number of covered positions by the number of positions in the genome. MosDepth was used to calculate the median coverage of 500-bp chunks of each genome at different thresholds (--use-median –by 500 –thresholds 1,10,20,30) (Pedersen & Quinlan, 2018).

TBlastX v2.11.0 was used to examine sequences with coverage at 10X or greater (McGinnis & Madden, 2004). Each 500-bp chunk of the genome at this coverage level was given a unique identifier corresponding to its positional range. That range of the

genome was then extracted from the reference multifasta file via SeqKit v0.16.1 and used as input for a translated nucleotide search using TBlastX against the NCBI's viral nt database (built and accessed on August 13, 2024) (Shen et al., 2016). The output format selected for the TBlastX was inclusive of the query sequence IDs, query/subject taxID, %ID, length, mismatch tolerance, gapopen, query/subject sequence start/end, e-value, bitscore, and query/subject scientific/common names (-outfmt 6). The accession numbers of each query sequence ID were merged with their TaxID from genbank's Accession2TaxID map, and the original subject TaxIDs assigned by the Blast database were matched and replaced with those from genbank's Accession2TaxID map for congruence between databases.

The Accession2TaxID map's nodes.dmp and names.dmp files were downloaded and parsed to identify hierarchical taxonomic relationships at the genus and species level. These relationships were traversed and logged for each query/subject sequence in the TBlastX output file, allowing for post-processing steps to remove alignments where the subject taxa was equivalent to the query taxa at the desired levels. Should the desired level not have information that could be traversed in the Accession2TaxID map, it was logged as "Unknown".

All statistical analyses and figures were created using RStudio v4.3.2.

## 4.3 Results and Discussion

In brief, 32 wastewater samples were sequenced. 16 NC county samples were used, representing eight different sites across four different dates in the spring of 2023. Of these, two samples were wastewater "blanks", process control samples for the WWTP that should be negative for SARS-CoV-2. In addition to these 16 samples were 16 unique campus sites from UNC-Charlotte.

The samples detailed above underwent the same library preparation detailed in the methods, as well as sequencing on an Illumina NextSeq 2000 instrument, producing 22.44 Gbp of data with an 94.86% average of reads at Q30, and 133.46M reads passing Illumina's internal quality filter.

### 4.3.1. Sequencing outcomes similar to competitor product

Breadth of genome coverage at 1X and average sequencing depth were calculated for those of the 11/41 reference genomes that were present in each sample [Figures 4.1 & 4.2]. Some outliers were observed in both categories, but the distributions remained consistently low throughout for genomes detected by the panel. While these results are less than ideal, this experiment was again exploratory to test the efficacy of this assay in wastewater samples. Our lab observed comparable findings in two recent studies using the Twist Respiratory Virus Research Panel on wastewater. The first study was a pilot project designed to evaluate whether this panel serves as a reliable indicator of clinical outcomes through WBE, while the second study investigated the potential matrix effects of wastewater on the effectiveness of target capture and hybridization in the assay [Roppolo Brazell et al., 2024]. Breadth of genome coverage detected in wastewater

across both studies was also low- in the pilot study, most wastewater samples did not achieve greater than 20% genome coverage, with a slightly higher breadth of coverage for genomes detected in wastewater samples from the matrix effect study. It is important to note that in the latter study, there was an entirely different cDNA synthesis protocol followed, so it is possible that variation between both breadth of coverage and average sequencing depth could be due to this. The average sequencing depth in genomes detected in wastewater samples in the aforementioned studies is closely aligned with the results seen in this study as well.

Even across entirely different library preparation methods, this trend appears common in wastewater. These sequencing outcomes suggest that further improvements could be made both up and downstream of library preparation.

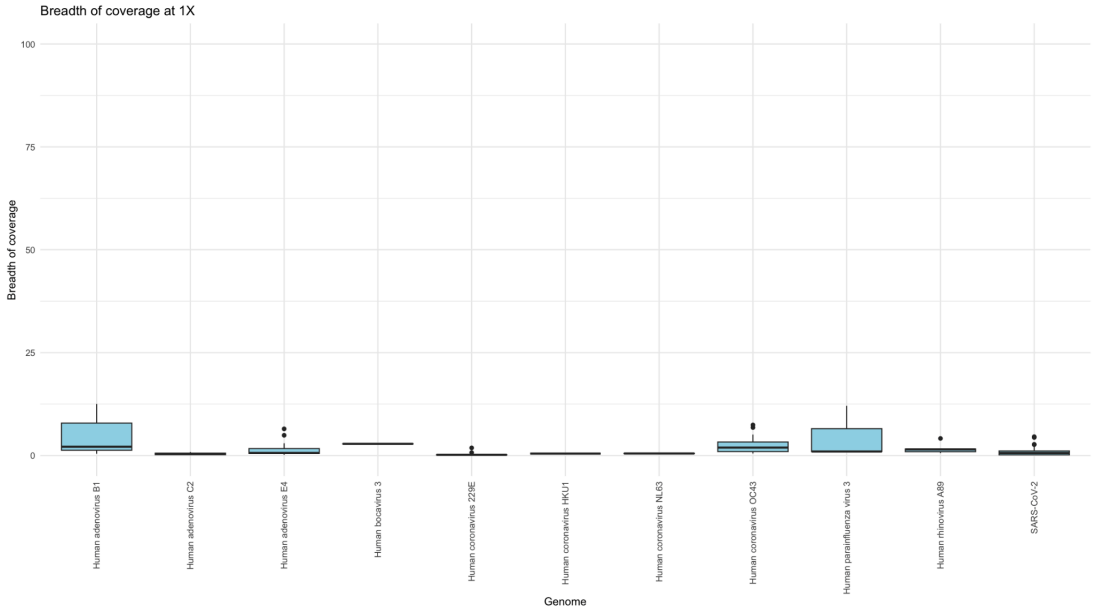


Figure 4.1. Distribution of breadth of genome coverage at 1X across all target genomes present in wastewater samples. Of the 41 target genomes, 30 are not represented in this data. The breadth of genome coverage is computed as a percentage of the length of each genome.

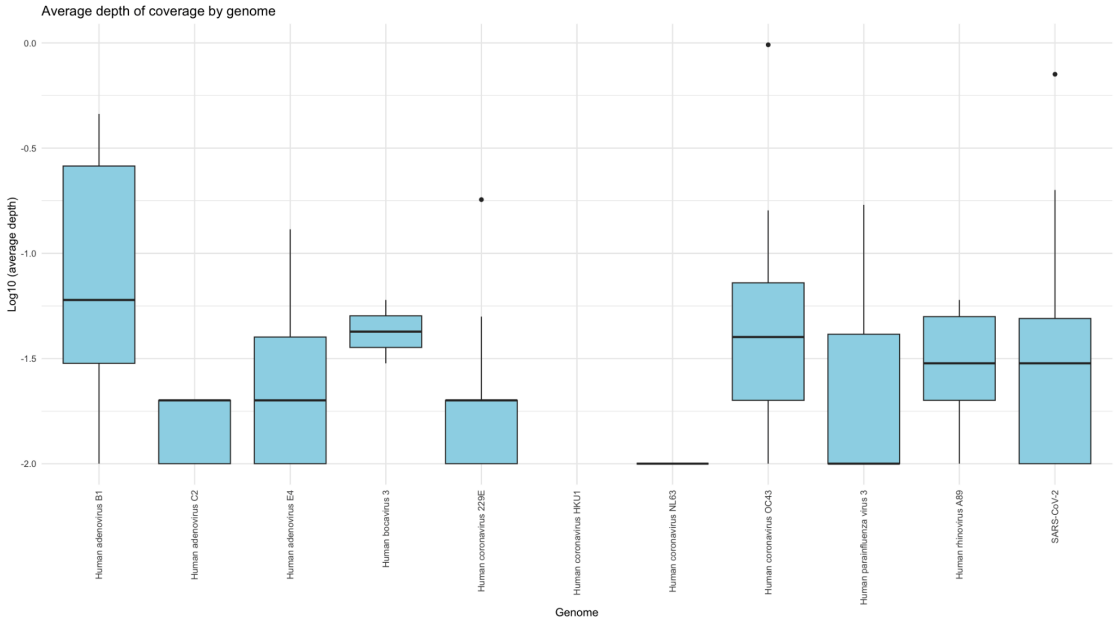


Figure 4.2. Average depth of genome coverage across all target genomes present in wastewater samples. Average sequencing depth is reported on a logarithmic scale (base 10).

4.3.2. Relative abundance comparisons between common taxonomic identification methods

Relative abundances of the detected viral species were computed using both k-mer based and direct-alignment based classification methods. The first method, Kraken2, is frequently used in metagenomics studies to deconvolute complex microbial communities like the ones found in wastewater. This program operates by dividing sequence reads into k-mers and comparing them against reference genomes in the designated databases. It subsequently assigns these k-mers to their lowest common ancestor taxonomic classifications (Wood et al., 2019). The sequence hits are further analyzed and refined through integration with Kraken2's companion software, Bracken, which utilizes k-mer distributions from both the reference and query sequences to reestimate the relative abundances of each taxonomic group in each sample. Of the 41

reference genomes contained in the Illumina Respiratory Viral Oligo Panel, only three (Human mastadenovirus B, E, and SARS-CoV-2) were detected at very low abundances [Figure 4.3; Supplementary Table 3]. Perhaps this is not unexpected given the low breadth of genome coverage and average sequencing depth for these genomes in the above analysis. Further, the low detection of targets mirrored what we saw in our pilot study using the Twist Respiratory Virus Research Panel.

In contrast to the Kraken2 k-mer based classification method, the direct alignment-based approach using Bowtie2 alignment of reads to the reference genomes yielded a larger distribution of detected target genomes at 11/41 [Figure 4.4; Supplementary Table 2]. The three viruses that Kraken2 detected were present in the Bowtie2 analysis, though their relative abundances differed greatly from one another. Again, this is in line with what we observed in our pilot study- the slightly permissive Bowtie2 parameters that were used allowed for the inclusion of hits to other respiratory virus genomes. Omnipresent in nearly all of the samples identified using this method were members of the *Coronaviridae* family, SARS-CoV-2, Human coronavirus 229 and OC43. This trend was also observed exactly in the pilot study using the Twist Respiratory Virus Research Panel in wastewater.

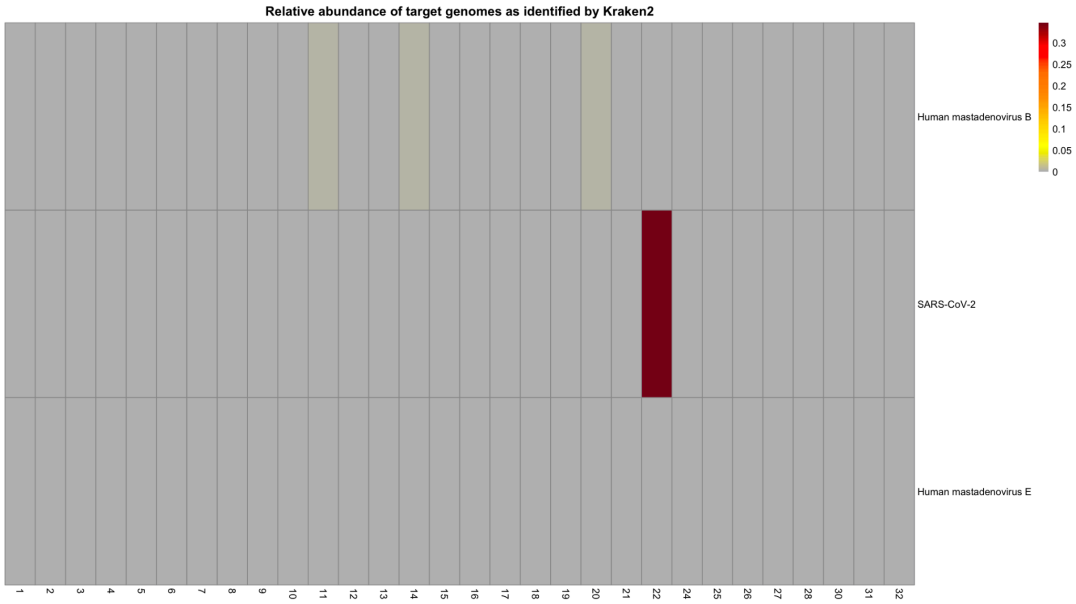


Figure 4.3. Relative abundance of each on-target viral genome in each wastewater sample as identified by Kraken2. Relative abundances are reported on a logarithmic scale (base 10).

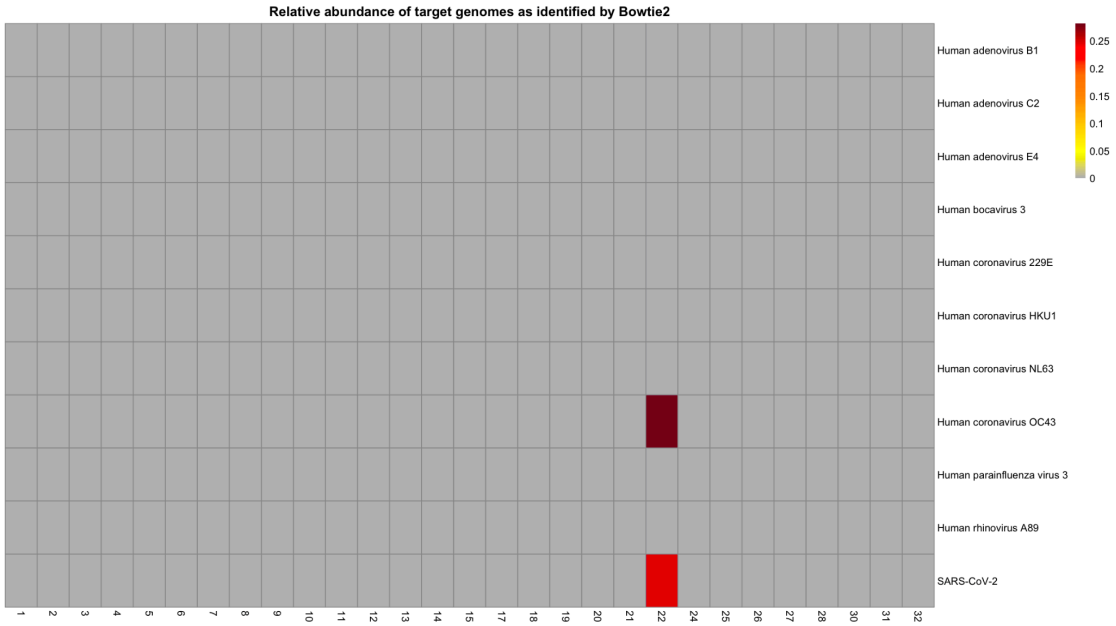


Figure 4.4. Relative abundance of each on-target viral genome in each wastewater sample as identified by Bowtie2. Relative abundances are reported on a logarithmic scale (base 10).

The low relative abundances of the on-target viruses included in the Illumina RVOP warranted further exploration. The Kraken2 analysis compared these reads to the most recent version of the viral database to make its taxonomic classifications, reporting 102 viral species outside of the expected targets at varying relative abundances within each sample [Supplementary Table 3; Supplementary Figures 1 & 2]. The Kraken2 relative abundances of off vs. on-target hits were compared to those observed by Bowtie2 and both showed that the majority of their composition dramatically favored off-target hits (Kraken2: Mean = 98.7, Median = 99.9, SD = 6.29; Bowtie2: Mean = 100.00, Median = 100.00, Standard Deviation= 0.0979).

4.3.3. No cross-reactions observed at genus level; significant homology at species level

Given the presence of off-target hits, it was crucial to determine whether or not the on-target hits demonstrated sequence specificity. In the pilot study using the Twist Respiratory Viral Research Panel, our lab determined that there was some degree of cross-reaction between the targets and homologous sequences. We observed a similar pattern using the Illumina RVOP. Cross-reactions were observed at the species level for the three most abundant viral taxa across all samples [Figures 4.5, 4.6, & 4.7].

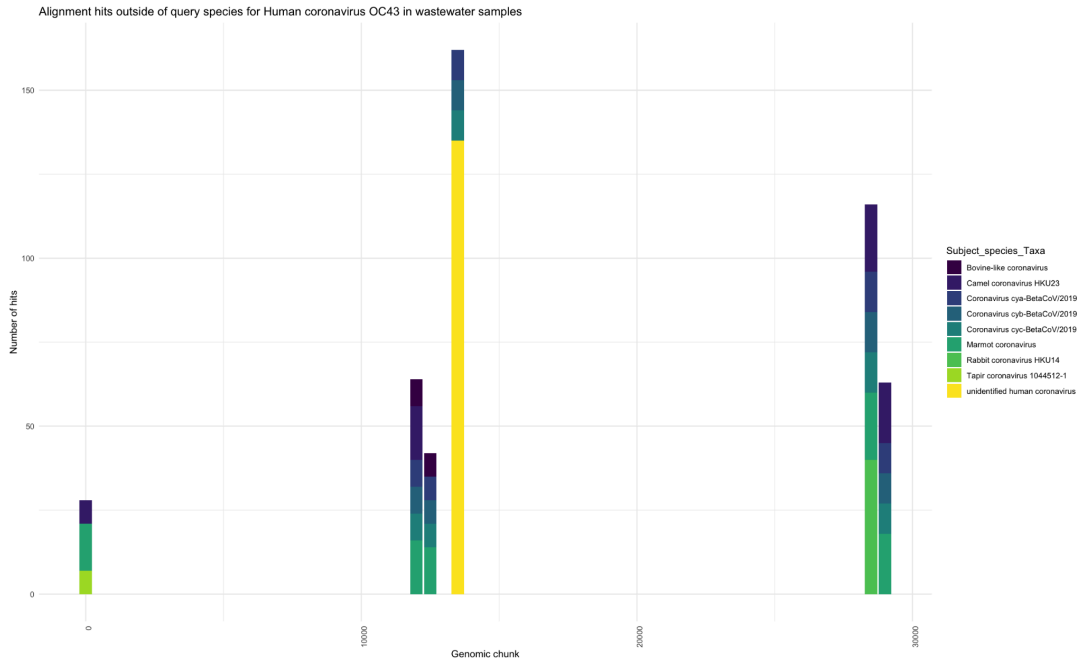


Figure 4.5. Bar plot distributions of the number of alignment hits outside of the query species for Human coronavirus OC43.

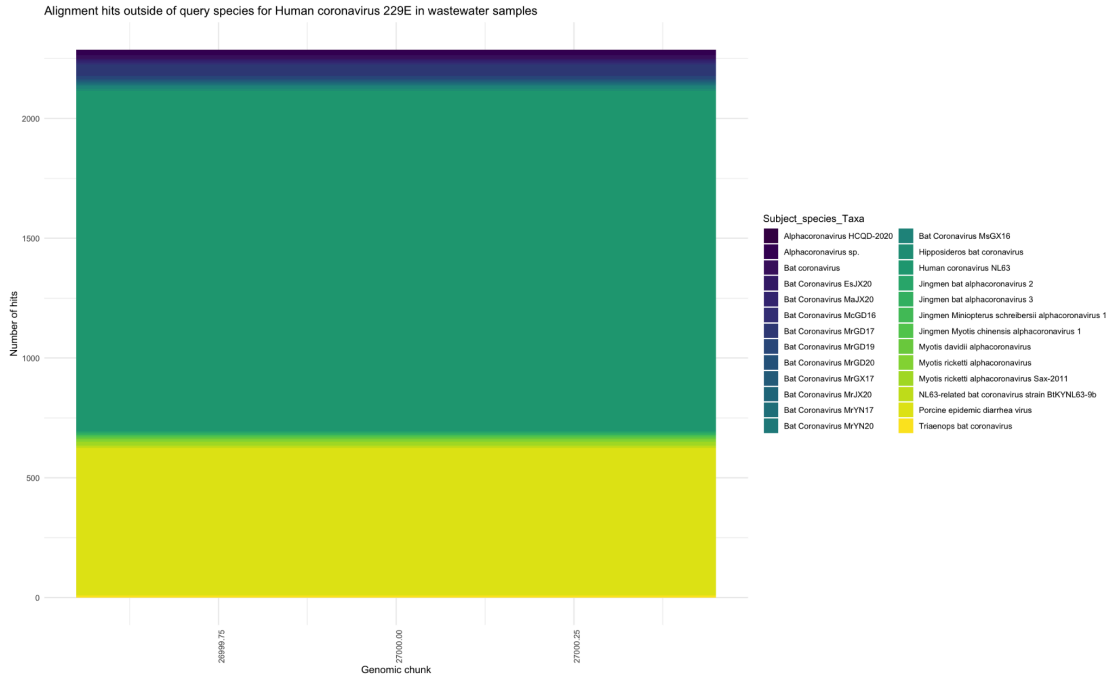


Figure 4.6. Bar plot distributions of the number of alignment hits outside of the query species for Human coronavirus 229E. There was only one location in the genome where any cross-reacts occurred, denoted by the single solid bar.

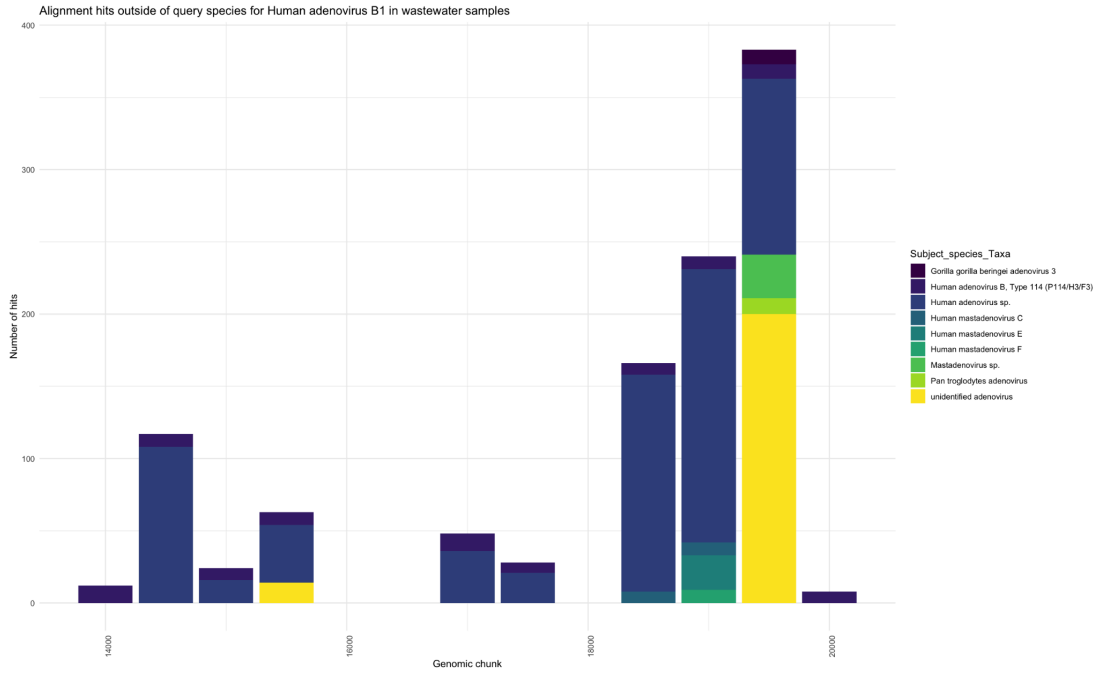


Figure 4.7. Bar plot distributions of the number of alignment hits outside of the query species for Human adenovirus B1.

The cross-reactions noted in these three species are not unprecedented. For human coronavirus OC43, cross-reactions were observed in only six unique 500-bp regions across the genome. Of these, all hits were coronaviruses spanning multiple host species. With this said, there was a far smaller distribution of species that were considered off-target to the query sequence than there were for those in the Twist Respiratory Virus Research Panel [Supplementary Table 4]. For human coronavirus 229E, the only off-target cross-reactions were seen in a single area of the genome. Similar to human coronavirus OC43, all these alignment hits were coronaviruses spanning multiple host species. Finally, for human adenovirus B1, a similar trend was observed where all off-target alignment hits were adenoviruses that affect various host species. At the genus level, no cross-reactions were detected in any of the tested samples. This finding suggests that, in these specific wastewater samples, this panel is capable of precise hybridization and target capture at the genus level. Even considering all off-target cross reactions

at the species level, these results demonstrate that this panel can confidently identify pathogens of interest if not exactly, it gets very close. Taken together with the low sequencing depth and breadth of coverage of these genomes, it is possible that better resolution could be achieved if higher quality sequences were obtained.

## 4.4 Conclusions and Future Work

The Illumina Respiratory Virus Oligo Panel (RVOP) shows considerable promise for applications in both clinical diagnostics and wastewater monitoring. This study validates the assay's effectiveness in detecting human respiratory viruses in wastewater while providing a comparison with a similar commercial product used in the same context. Both investigations highlight significant challenges related to the target capture and hybridization of respiratory viral sequences from wastewater samples, evident in the limited genome coverage and low average sequencing depth observed for the identified genomes. The findings from our study do not meet the necessary quality standards for variant identification, as noted in previous studies employing the same methodology. These issues are particularly pronounced in this analysis, seen further in low relative abundances for genomes when assessed using both k-mer and direct alignment detection methods. Finally, our bioinformatics analysis of cross-reactions favored the use of this panel as it identified all detected viral genera at 100% accuracy with no cross-reactions, and came very close with a few exceptions at the species level. Overall, our research indicates that the Illumina RVOP detection panel demonstrates that it does work in the context of wastewater, but its performance should be considered and optimized in future studies to ensure its success.

It is crucial to recognize the many differences in viral RNA extraction and library preparation between these two methods, as these variations may significantly impact their performance outcomes. To draw more robust conclusions about their efficacy, a direct one-to-one comparison of both kits using identical wastewater samples would be essential. Additionally, incorporating known control mixtures prepared alongside these samples would be beneficial. Given that wastewater samples represent a unique microbiome, the specific microbial constituents remain largely unidentified without whole-genome sequencing for reference. Our

laboratory recently conducted a series of studies aimed at evaluating the matrix effects of wastewater on the recovery of known viral spike-ins. We have gathered data from these experiments using the Twist Respiratory Viral Research Panel, as well as from a separate amplicon-based SARS-CoV-2 study. The results suggested that there is a slight matrix effect from wastewater on these samples, even when care is taken to remove known inhibitors. In this way, establishing a reference point for the RVOP would enhance the future interpretation of results when applying this assay to wastewater samples.

# DISSERTATION SUMMARY AND CONCLUSIONS

Even as the urgency of the SARS-CoV-2 begins to fade, it is important to remember that other human respiratory viruses still pose a significant public health problem worldwide. Insights from the SARS-CoV-2 pandemic demonstrated that wastewater-based epidemiology is a viable method to monitor the spread of fecal-shed respiratory viruses at the community level in order to inform the public of their risk. A shift from limited molecular-based detection methods towards NGS for wastewater surveillance of human respiratory viruses outside of SARS-CoV-2 is underway to provide the most accurate information to the public. Target capture and hybridization NGS methods allow for the identification of more viral targets than standard multiplex PCR assays, providing a more comprehensive snapshot of the virome through wastewater. The results of this dissertation demonstrate that while there is a discernible effect of the wastewater matrix on the performance of the hybridization, enrichment of viral material from wastewater prior to library preparation and target capture sequencing is possible. This method preserves sequencing and bioinformatics resources to only the targets of interest, though our results suggest that cross-reactions do occur and there is a lack of strain level specificity in these assays. Both commercial target capture panels demonstrated in all components of this study that breadth of genome coverage and average sequencing depth of the targets are not sufficient for variant identification from wastewater. Taken together, the results of this dissertation raise the question: what is the best use case for this type of NGS assay? Is it the best way to track circulating respiratory viruses from wastewater?

Careful consideration is needed to decide if a target-capture based assay is the right choice for wastewater surveillance of human respiratory viruses. A broad surveillance of

respiratory viruses where lack of specificity does not matter may have utility in determining the presence or absence of a specific virus within a community.

# REFERENCES


Adams, C., Bias, M., Welsh, R. M., Webb, J., Reese, H., Delgado, S., Person, J., West, R., Shin, S., & Kirby, A. (2024). The National Wastewater Surveillance System (NWSS): From inception to widespread coverage, 2020–2022, United States. *Science of The Total Environment*, *924*, 171566. https://doi.org/10.1016/j.scitotenv.2024.171566

Ahmed, W., Bivins, A., Metcalfe, S., Smith, W. J. M., Verbyla, M. E., Symonds, E. M., & Simpson, S. L. (2022). Evaluation of process limit of detection and quantification variation of SARS-CoV-2 RT-qPCR and RT-dPCR assays for wastewater surveillance. *Water Research*, *213*, 118132. https://doi.org/10.1016/j.watres.2022.118132

Ahmed, W., Bivins, A., Stephens, M., Metcalfe, S., Smith, W. J. M., Sirikanchana, K., Kitajima, M., & Simpson, S. L. (2023). Occurrence of multiple respiratory viruses in wastewater in Queensland, Australia: Potential for community disease surveillance. *Science of The Total Environment*, *864*, 161023. https://doi.org/10.1016/j.scitotenv.2022.161023

Ahmed, W., Tscharke, B., Bertsch, P. M., Bibby, K., Bivins, A., Choi, P., Clarke, L., Dwyer, J., Edson, J., Nguyen, T. M. H., O'Brien, J. W., Simpson, S. L., Sherman, P., Thomas, K. V., Verhagen, R., Zaugg, J., & Mueller, J. F. (2021). SARS-CoV-2 RNA monitoring in wastewater as a potential early warning system for COVID-19 transmission in the community: A temporal case study. *Science of The Total Environment*, *761*, 144216. https://doi.org/10.1016/j.scitotenv.2020.144216

Aigrain, L. (2021). Beginner's guide to next-generation sequencing. *The Biochemist*, *43*(6), 58–64. https://doi.org/10.1042/bio_2021_135

Al Hajjar, S., & McIntosh, K. (2010). The first influenza pandemic of the 21st century. *Annals of Saudi Medicine*, *30*(1), 1–10. https://doi.org/10.4103/0256-4947.59365

Apoorva Mandavilli. (2022, October 27). A 'Tripledemic'? Flu, R.S.V. and Covid May Collide This Winter, Experts Say. *The New York Times*. https://www.nytimes.com/2022/10/23/health/flu-covid-risk.html

Arbeitskreis Blut. (2009). Influenza Virus. *Transfusion Medicine and Hemotherapy*, *36*(1), 32–39. https://doi.org/10.1159/000197314

Bagga, B., Woods, C. W., Veldman, T. H., Gilbert, A., Mann, A., Balaratnam, G., Lambkin-Williams, R., Oxford, J. S., McClain, M. T., Wilkinson, T., Nicholson, B. P., Ginsburg, G. S., & DeVincenzo, J. P. (2013). Comparing Influenza and Rsv Viral and Disease Dynamics in Experimentally Infected Adults Predicts Clinical Effectiveness of Rsv Antivirals. *Antiviral Therapy*, *18*(6), 785–792. https://doi.org/10.3851/IMP2629

Barberis, I., Myles, P., Ault, S. K., Bragazzi, N. L., & Martini, M. (2016). History and evolution of influenza control through vaccination: From the first monovalent vaccine to universal vaccines. *Journal of Preventive Medicine and Hygiene*, *57*(3), E115–E120.

Barua, V. B., Juel, M. A. I., Blackwood, A. D., Clerkin, T., Ciesielski, M., Sorinolu, A. J., Holcomb, D. A., Young, I., Kimble, G., Sypolt, S., Engel, L. S., Noble, R. T., & Munir, M. (2022). Tracking the temporal variation of COVID-19 surges through wastewater-based epidemiology during the peak of the pandemic: A six-month long study in Charlotte, North Carolina. *Science of The Total Environment*, *814*, 152503. https://doi.org/10.1016/j.scitotenv.2021.152503

Binder, A. M., Biggs, H. M., Haynes, A. K., Chommanard, C., Lu, X., Erdman, D. D., Watson, J. T., & Gerber, S. I. (2017). Human Adenovirus Surveillance—United States, 2003–2016.

*MMWR. Morbidity and Mortality Weekly Report*, *66*(39), 1039–1042. https://doi.org/10.15585/mmwr.mm6639a2

*Biology for Majors II: Steps of Virus Infections*. (n.d.). Lumen Learning. Retrieved October 30, 2024, from https://courses.lumenlearning.com/wm-biology2/steps-of-virus-infections-text-version/

Blaas, D., & Fuchs, R. (2016). Mechanism of human rhinovirus infections. *Molecular and Cellular Pediatrics*, *3*(1), 21. https://doi.org/10.1186/s40348-016-0049-3

Boehm, A. B., Wolfe, M. K., White, B. J., Hughes, B., Duong, D., & Bidwell, A. (2023). More than a Tripledemic: Influenza A Virus, Respiratory Syncytial Virus, SARS-CoV-2, and Human Metapneumovirus in Wastewater during Winter 2022–2023. *Environmental Science & Technology Letters*, *10*(8), 622–627. https://doi.org/10.1021/acs.estlett.3c00385

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bouvier, N. M., & Palese, P. (2008). The biology of influenza viruses. *Vaccine*, *26*, D49–D53. https://doi.org/10.1016/j.vaccine.2008.07.039

Branche, A., & Falsey, A. (2016). Parainfluenza Virus Infection. *Seminars in Respiratory and Critical Care Medicine*, *37*(04), 538–554. https://doi.org/10.1055/s-0036-1584798

Brant, A. C., Tian, W., Majerciak, V., Yang, W., & Zheng, Z.-M. (2021). SARS-CoV-2: From its discovery to genome structure, transcription, and replication. *Cell & Bioscience*, *11*(1), 136. https://doi.org/10.1186/s13578-021-00643-z

Budinger, G. R. S., Misharin, A. V., Ridge, K. M., Singer, B. D., & Wunderink, R. G. (2021). Distinctive features of severe SARS-CoV-2 pneumonia. *Journal of Clinical Investigation*, *131*(14), e149412. https://doi.org/10.1172/JCI149412

*California Department of Public Health Wastewater Surveillance Beyond COVID-19: Influenza, RSV, Mpox, etc.* (n.d.). California Department of Public Health. Retrieved October 30, 2024, from https://www.cdph.ca.gov/Programs/CID/DCDC/CDPH%20Document%20Library/COVID-19/WWSBeyondSC2.pdf

Carter, R. W., & Sanford, J. C. (2012). A new look at an old virus: Patterns of mutation accumulation in the human H1N1 influenza virus since 1918. *Theoretical Biology and Medical Modelling*, *9*(1), 42. https://doi.org/10.1186/1742-4682-9-42

Chen, B., Tian, E.-K., He, B., Tian, L., Han, R., Wang, S., Xiang, Q., Zhang, S., El Arnaout, T., & Cheng, W. (2020). Overview of lethal human coronaviruses. *Signal Transduction and Targeted Therapy*, *5*(1), 89. https://doi.org/10.1038/s41392-020-0190-2

Chen, Z., Erickson, D. L., & Meng, J. (2020). Benchmarking Long-Read Assemblers for Genomic Analyses of Bacterial Pathogens Using Oxford Nanopore Sequencing. *International Journal of Molecular Sciences*, *21*(23), 9161. https://doi.org/10.3390/ijms21239161

Chien, Y.-S., Luo, S.-T., Tsao, K.-C., Huang, Y.-C., Chung, W.-Y., Liao, Y.-C., Tan, Y., Das, S. R., & Lee, M.-S. (2019). Genomic analysis of serologically untypable human enteroviruses in Taiwan. *Journal of Biomedical Science*, *26*(1), 49. https://doi.org/10.1186/s12929-019-0541-x

Child, H. T., Airey, G., Maloney, D. M., Parker, A., Wild, J., McGinley, S., Evens, N., Porter, J., Templeton, K., Paterson, S., Van Aerle, R., Wade, M. J., Jeffries, A. R., & Bassano, I.

(2023). Comparison of metagenomic and targeted methods for sequencing human pathogenic viruses from wastewater. *mBio*, *14*(6), e01468-23. https://doi.org/10.1128/mbio.01468-23

Christopher Smith. (2024, July). *Bioinformatics Details of the Twist Viral Panels*. One Codex. https://docs.onecodex.com/en/articles/4648676-bioinformatics-details-of-the-twist-viral-panels

Cifuentes-Muñoz, N., & Ellis Dutch, R. (2019). To assemble or not to assemble: The changing rules of pneumovirus transmission. *Virus Research*, *265*, 68–73. https://doi.org/10.1016/j.virusres.2019.03.002

Cosar, B., Karagulleoglu, Z. Y., Unal, S., Ince, A. T., Uncuoglu, D. B., Tuncer, G., Kilinc, B. R., Ozkan, Y. E., Ozkoc, H. C., Demir, I. N., Eker, A., Karagoz, F., Simsek, S. Y., Yasar, B., Pala, M., Demir, A., Atak, I. N., Mendi, A. H., Bengi, V. U., … Demir-Dora, D. (2022). SARS-CoV-2 Mutations and their Viral Variants. *Cytokine & Growth Factor Reviews*, *63*, 10–22. https://doi.org/10.1016/j.cytogfr.2021.06.001

Cox, R. G., Mainou, B. A., Johnson, M., Hastings, A. K., Schuster, J. E., Dermody, T. S., & Williams, J. V. (2015). Human Metapneumovirus Is Capable of Entering Cells by Fusion with Endosomal Membranes. *PLOS Pathogens*, *11*(12), e1005303. https://doi.org/10.1371/journal.ppat.1005303

Crits-Christoph, A., Kantor, R. S., Olm, M. R., Whitney, O. N., Al-Shayeb, B., Lou, Y. C., Flamholz, A., Kennedy, L. C., Greenwald, H., Hinkle, A., Hetzel, J., Spitzer, S., Koble, J., Tan, A., Hyde, F., Schroth, G., Kuersten, S., Banfield, J. F., & Nelson, K. L. (2021). Genome Sequencing of Sewage Detects Regionally Prevalent SARS-CoV-2 Variants. *mBio*, *12*(1), e02703-20. https://doi.org/10.1128/mBio.02703-20

Dahui, Q. (2019). Next-generation sequencing and its clinical application. *Cancer Biology & Medicine*, *16*(1), 4–10. https://doi.org/10.20892/j.issn.2095-3941.2018.0055

Daou, M., Kannout, H., Khalili, M., Almarei, M., Alhashami, M., Alhalwachi, Z., Alshamsi, F., Tahseen Al Bataineh, M., Azzam Kayasseh, M., Al Khajeh, A., Hasan, S. W., Tay, G. K., Feng, S. F., Ruta, D., Yousef, A. F., Alsafar, H. S., & on behalf of the UAE COVID-19 Collaborative Partnership. (2022). Analysis of SARS-CoV-2 viral loads in stool samples and nasopharyngeal swabs from COVID-19 patients in the United Arab Emirates. *PLOS ONE*, *17*(9), e0274961. https://doi.org/10.1371/journal.pone.0274961

D'Arienzo, M., & Coniglio, A. (2020). Assessment of the SARS-CoV-2 basic reproduction number, R0, based on the early phase of COVID-19 outbreak in Italy. *Biosafety and Health*, *2*(2), 57–59. https://doi.org/10.1016/j.bsheal.2020.03.004

Davison, A. J., Benkő, M., & Harrach, B. (2003). Genetic content and evolution of adenoviruses. *Journal of General Virology*, *84*(11), 2895–2908. https://doi.org/10.1099/vir.0.19497-0

Denise Lynch. (2023, July). *One Codex Database*. One Codex. https://docs.onecodex.com/en/articles/3761205-one-codex-database

Derbyshire, E. J., & Calder, P. C. (2021). Respiratory Tract Infections and Antibiotic Resistance: A Protective Role for Vitamin D? *Frontiers in Nutrition*, *8*, 652469. https://doi.org/10.3389/fnut.2021.652469

*Detection and characterization of respiratory viruses, including SARS-CoV-2, using Illumina RNA Prep with Enrichment*. (2020). Illumina, Inc. https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-rna-enrichment-coronavirus-app-note-470-2020-005/illumina-rna-enrichment-coronavirus-app-note-470-2020-005.pdf

Dou, D., Revol, R., Östbye, H., Wang, H., & Daniels, R. (2018). Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement. *Frontiers in Immunology*, *9*, 1581. https://doi.org/10.3389/fimmu.2018.01581

Draghici, S., Khatri, P., Eklund, A., & Szallasi, Z. (2006). Reliability and reproducibility issues in DNA microarray measurements. *Trends in Genetics*, *22*(2), 101–109. https://doi.org/10.1016/j.tig.2005.12.005

El Najjar, F., Schmitt, A., & Dutch, R. (2014). Paramyxovirus Glycoprotein Incorporation, Assembly and Budding: A Three Way Dance for Infectious Particle Production. *Viruses*, *6*(8), 3019–3054. https://doi.org/10.3390/v6083019

Evertsz, E. M., Au-Young, J., Ruvolo, M. V., Lim, A. C., & Reynolds, M. A. (2001). Hybridization Cross-Reactivity within Homologous Gene Families on Glass cDNA Microarrays. *BioTechniques*, *31*(5), 1182–1192. https://doi.org/10.2144/01315dd03

Farrukee, R., Ait-Goughoulte, M., Saunders, P. M., Londrigan, S. L., & Reading, P. C. (2020). Host Cell Restriction Factors of Paramyxoviruses and Pneumoviruses. *Viruses*, *12*(12), 1381. https://doi.org/10.3390/v12121381

Ferravante, C., Sanna, G., Melone, V., Fromentier, A., Rocco, T., D'Agostino, Y., Lamberti, J., Alexandrova, E., Pecoraro, G., Pagliano, P., Astorri, R., Manzin, A., Weisz, A., Giurato, G., Galdiero, M., Rizzo, F., & Franci, G. (2022). Nasopharyngeal virome analysis of COVID‐19 patients during three different waves in Campania region of Italy. *Journal of Medical Virology*, *94*(5), 2275–2283. https://doi.org/10.1002/jmv.27571

Flahault, A., & Zylberman, P. (2010). Influenza pandemics: Past, present and future challenges. *Public Health Reviews*, *32*(1), 319–340. https://doi.org/10.1007/BF03391605

Fuchs, R., & Blaas, D. (2010). Uncoating of human rhinoviruses. *Reviews in Medical Virology*, *20*(5), 281–297. https://doi.org/10.1002/rmv.654

Furlow, B. (2023). Triple-demic overwhelms paediatric units in US hospitals. *The Lancet Child & Adolescent Health*, *7*(2), 86. https://doi.org/10.1016/S2352-4642(22)00372-8

Gibas, C., Lambirth, K., Mittal, N., Juel, M. A. I., Barua, V. B., Roppolo Brazell, L., Hinton, K., Lontai, J., Stark, N., Young, I., Quach, C., Russ, M., Kauer, J., Nicolosi, B., Chen, D., Akella, S., Tang, W., Schlueter, J., & Munir, M. (2021). Implementing building-level SARS-CoV-2 wastewater surveillance on a university campus. *Science of The Total Environment*, *782*, 146749. https://doi.org/10.1016/j.scitotenv.2021.146749

Ginocchio, C. C., & McAdam, A. J. (2011). Current Best Practices for Respiratory Virus Testing. *Journal of Clinical Microbiology*, *49*(9_Supplement). https://doi.org/10.1128/JCM.00698-11

Halling, K. C., & Wendel, A. J. (2009). In Situ Hybridization: Principles and Applications. In T. C. Allen & P. T. Cagle (Eds.), *Basic Concepts of Molecular Pathology* (Vol. 2, pp. 109–118). Springer US. https://doi.org/10.1007/978-0-387-89626-7_12

Hanish Jain, John W. Schweitzer, & Nathaniel A. Justice. (2023). Respiratory Syncytial Virus Infection in Children. In *StatPearls [Internet]*. StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK459215/

Harrington, W. N., Kackos, C. M., & Webby, R. J. (2021). The evolution and future of influenza pandemic preparedness. *Experimental & Molecular Medicine*, *53*(5), 737–749. https://doi.org/10.1038/s12276-021-00603-0

Head, S. R., Komori, H. K., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D. R., & Ordoukhanian, P. (2014). Library construction for next-generation sequencing: Overviews and challenges. *BioTechniques*, *56*(2), 61–77.

https://doi.org/10.2144/000114133
Henze, M., Van Loosdrecht, M. C. M., Ekama, G. A., & Brdjanovic, D. (2019). *Biological Wastewater Treatment: Principles, Modeling and Design*. IWA Publishing. https://doi.org/10.2166/9781780408613
Hernández-Rivas, L., Pedraz, T., Calvo, C., San Juan, I., Mellado, M. [a] J., & Robustillo, A. (2023). Respiratory syncytial virus outbreak during the COVID-19 pandemic. How has it changed? *Enfermedades Infecciosas y Microbiología Clínica*, *41*(6), 352–355. https://doi.org/10.1016/j.eimc.2021.12.003
Hoeben, R. C., & Uil, T. G. (2013). Adenovirus DNA Replication. *Cold Spring Harbor Perspectives in Biology*, *5*(3), a013003–a013003. https://doi.org/10.1101/cshperspect.a013003
Jaksik, R., Iwanaszko, M., Rzeszowska-Wolny, J., & Kimmel, M. (2015). Microarray experiments and factors which affect their reliability. *Biology Direct*, *10*(1), 46. https://doi.org/10.1186/s13062-015-0077-2
Jefferson, T., Del Mar, C. B., Dooley, L., Ferroni, E., Al-Ansary, L. A., Bawazeer, G. A., Van Driel, M. L., Jones, M. A., Thorning, S., Beller, E. M., Clark, J., Hoffmann, T. C., Glasziou, P. P., & Conly, J. M. (2020). Physical interventions to interrupt or reduce the spread of respiratory viruses. *Cochrane Database of Systematic Reviews*, *2020*(11). https://doi.org/10.1002/14651858.CD006207.pub5
Jiang, X.-W., Huang, T.-S., Xie, L., Chen, S.-Z., Wang, S.-D., Huang, Z.-W., Li, X.-Y., & Ling, W.-P. (2022). Development of a diagnostic assay by three-tube multiplex real-time PCR for simultaneous detection of nine microorganisms causing acute respiratory infections. *Scientific Reports*, *12*(1), 13306. https://doi.org/10.1038/s41598-022-15543-6
Johnson, J., Sun, S., & Fodor, A. A. (2022). *Systematic classification error profoundly impacts inference in high-depth Whole Genome Shotgun Sequencing datasets*. https://doi.org/10.1101/2022.04.04.487034
Juel, M. A. I., Stark, N., Nicolosi, B., Lontai, J., Lambirth, K., Schlueter, J., Gibas, C., & Munir, M. (2021). Performance evaluation of virus concentration methods for implementing SARS-CoV-2 wastewater based epidemiology emphasizing quick data turnaround. *Science of The Total Environment*, *801*, 149656. https://doi.org/10.1016/j.scitotenv.2021.149656
Kajon, A. E., Weinberg, J. B., & Spindler, K. R. (2019). Adenoviruses. In *Reference Module in Biomedical Sciences* (p. B9780128012383000866). Elsevier. https://doi.org/10.1016/B978-0-12-801238-3.00086-6
Kalil, A. C., & Thomas, P. G. (2019). Influenza virus-related critical illness: Pathophysiology and epidemiology. *Critical Care*, *23*(1), 258. https://doi.org/10.1186/s13054-019-2539-x
Kantor, R. S., & Jiang, M. (2024). Considerations and Opportunities for Probe Capture Enrichment Sequencing of Emerging Viruses from Wastewater. *Environmental Science & Technology*, *58*(19), 8161–8168. https://doi.org/10.1021/acs.est.4c02638
Kapel, N., Kalimeris, E., Lumley, S., Decano, A., Rodger, G., Lopes Alves, M., Dingle, K., Oakley, S., Barrett, L., Barnett, S., Crook, D., Eyre, D. W., Matthews, P. C., Street, T., & Stoesser, N. (2023). Evaluation of sequence hybridization for respiratory viruses using the Twist Bioscience Respiratory Virus Research panel and the OneCodex Respiratory Virus sequence analysis workflow. *Microbial Genomics*, *9*(9). https://doi.org/10.1099/mgen.0.001103
Kerr, S.-L., Mathew, C., & Ghildyal, R. (2021). Rhinovirus and Cell Death. *Viruses*, *13*(4), 629.

https://doi.org/10.3390/v13040629
Khanal, S., Ghimire, P., & Dhamoon, A. (2018). The Repertoire of Adenovirus in Human Disease: The Innocuous to the Deadly. *Biomedicines*, *6*(1), 30. https://doi.org/10.3390/biomedicines6010030
Kim, H., Webster, R. G., & Webby, R. J. (2018). Influenza Virus: Dealing with a Drifting and Shifting Pathogen. *Viral Immunology*, *31*(2), 174–183. https://doi.org/10.1089/vim.2017.0141
Kistler, K. E., & Bedford, T. (2021). Evidence for adaptive evolution in the receptor-binding domain of seasonal coronaviruses OC43 and 229e. *eLife*, *10*, e64509. https://doi.org/10.7554/eLife.64509
Köndgen, S., Schenk, S., Pauli, G., Boesch, C., & Leendertz, F. H. (2010). Noninvasive Monitoring of Respiratory Viruses in Wild Chimpanzees. *EcoHealth*, *7*(3), 332–341. https://doi.org/10.1007/s10393-010-0340-z
Kuhn, J. H., & Hughes, H. R. (2024). ICTV Virus Taxonomy Profile: Phasmaviridae 2024: This article is part of the ICTV Virus Taxonomy Profiles collection. *Journal of General Virology*, *105*(7). https://doi.org/10.1099/jgv.0.002002
Kumar, M., Joshi, M., Patel, A. K., & Joshi, C. G. (2021). Unravelling the early warning capability of wastewater surveillance for COVID-19: A temporal study on SARS-CoV-2 RNA detection and need for the escalation. *Environmental Research*, *196*, 110946. https://doi.org/10.1016/j.envres.2021.110946
Kumblathan, T., Piroddi, N., Hrudey, S. E., & Li, X.-F. (2022). Wastewater Based Surveillance of SARS-CoV-2: Challenges and Perspective from a Canadian Inter-laboratory Study. *Journal of Environmental Sciences*, *116*, 229–232. https://doi.org/10.1016/j.jes.2022.01.039
Kutter, J. S., Spronken, M. I., Fraaij, P. L., Fouchier, R. A., & Herfst, S. (2018). Transmission routes of respiratory viruses among humans. *Current Opinion in Virology*, *28*, 142–151. https://doi.org/10.1016/j.coviro.2018.01.001
Kwong, J. C., Mccallum, N., Sintchenko, V., & Howden, B. P. (2015). Whole genome sequencing in clinical and public health microbiology. *Pathology*, *47*(3), 199–210. https://doi.org/10.1097/PAT.0000000000000235
Lambisia, A. W., Mohammed, K. S., Makori, T. O., Ndwiga, L., Mburu, M. W., Morobe, J. M., Moraa, E. O., Musyoki, J., Murunga, N., Mwangi, J. N., Nokes, D. J., Agoti, C. N., Ochola-Oyier, L. I., & Githinji, G. (2022). Optimization of the SARS-CoV-2 ARTIC Network V4 Primers and Whole Genome Sequencing Protocol. *Frontiers in Medicine*, *9*, 836728. https://doi.org/10.3389/fmed.2022.836728
Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923
LaTurner, Z. W., Zong, D. M., Kalvapalle, P., Gamas, K. R., Terwilliger, A., Crosby, T., Ali, P., Avadhanula, V., Santos, H. H., Weesner, K., Hopkins, L., Piedra, P. A., Maresso, A. W., & Stadler, L. B. (2021). Evaluating recovery, cost, and throughput of different concentration methods for SARS-CoV-2 wastewater-based epidemiology. *Water Research*, *197*, 117043. https://doi.org/10.1016/j.watres.2021.117043
Leung, N. H. L. (2021). Transmissibility and transmission of respiratory viruses. *Nature Reviews Microbiology*, *19*(8), 528–545. https://doi.org/10.1038/s41579-021-00535-6
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence

Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lindenbach, B. D. (2022). Reinventing positive-strand RNA virus reverse genetics. In *Advances in Virus Research* (Vol. 112, pp. 1–29). Elsevier. https://doi.org/10.1016/bs.aivir.2022.03.001

Liu, D. X., Liang, J. Q., & Fung, T. S. (2021). Human Coronavirus-229E, -OC43, -NL63, and -HKU1 (Coronaviridae). In *Encyclopedia of Virology* (pp. 428–440). Elsevier. https://doi.org/10.1016/B978-0-12-809633-8.21501-X

Liu, Y., Wang, Y., Wang, Y., Mai, H., Chen, Y., Zhang, Y., Ji, Y., Cong, X., & Gao, Y. (2023). Phylogenetic analysis of HA and NA genes of influenza A viruses in immunosuppressed inpatients in Beijing during the 2018–2020 influenza seasons. *Virology Journal*, *20*(1), 101. https://doi.org/10.1186/s12985-023-02067-2

Lowry, S. A., Wolfe, M. K., & Boehm, A. B. (2023). Respiratory virus concentrations in human excretions that contribute to wastewater: A systematic review and meta-analysis. *Journal of Water and Health*, *21*(6), 831–848. https://doi.org/10.2166/wh.2023.057

Lu, H., Giordano, F., & Ning, Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics & Bioinformatics*, *14*(5), 265–279. https://doi.org/10.1016/j.gpb.2016.05.004

Lu, J., Breitwieser, F. P., Thielen, P., & Salzberg, S. L. (2017). Bracken: Estimating species abundance in metagenomics data. *PeerJ Computer Science*, *3*, e104. https://doi.org/10.7717/peerj-cs.104

*Maximum read length for Illumina sequencing platforms*. (2024, August 30). Illumina. Maximum read length for Illumina sequencing platforms

McGinnis, S., & Madden, T. L. (2004). BLAST: At the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research*, *32*(Web Server), W20–W25. https://doi.org/10.1093/nar/gkh435

Modrow, S., Falke, D., Truyen, U., & Schätzl, H. (2013). Viruses with Single-Stranded, Positive-Sense RNA Genomes. In S. Modrow, D. Falke, U. Truyen, & H. Schätzl, *Molecular Virology* (pp. 185–349). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-20718-1_14

Mostafa, A., Abdelwhab, E. M., Mettenleiter, T. C., & Pleschka, S. (2018). Zoonotic Potential of Influenza A Viruses: A Comprehensive Overview. *Viruses*, *10*(9), 497. https://doi.org/10.3390/v10090497

National Health Services. (n.d.). *Respiratory Tract Infections (RTIs)*. National Health Services. https://www.nhs.uk/conditions/respiratory-tract-infection/#:~:text=Respiratory%20tract%20infections%20(RTIs)%20are,need%20to%20see%20a%20GP.

Nayak, B., Kumar, S., Collins, P. L., & Samal, S. K. (2008). Molecular characterization and complete genome sequence of avian paramyxovirus type 4 prototype strain duck/Hong Kong/D3/75. *Virology Journal*, *5*(1), 124. https://doi.org/10.1186/1743-422X-5-124

Neri, U., Wolf, Y. I., Roux, S., Camargo, A. P., Lee, B., Kazlauskas, D., Chen, I. M., Ivanova, N., Zeigler Allen, L., Paez-Espino, D., Bryant, D. A., Bhaya, D., Krupovic, M., Dolja, V. V., Kyrpides, N. C., Koonin, E. V., Gophna, U., Narrowe, A. B., Probst, A. J., … Wang, Z. (2022). Expansion of the global RNA virome reveals diverse clades of bacteriophages. *Cell*, *185*(21), 4023-4037.e18. https://doi.org/10.1016/j.cell.2022.08.023

Nestić, D., Božinović, K., Pehar, I., Wallace, R., Parker, A. L., & Majhen, D. (2021). The Revolving Door of Adenovirus Cell Entry: Not All Pathways Are Equal. *Pharmaceutics*,

*13*(10), 1585. https://doi.org/10.3390/pharmaceutics13101585

*NGS Target Enrichment of Viral Pathogens using Twist Respiratory Virus Research Panel: Application Note*. (2021, July 12). Twist Bioscience. https://www.twistbioscience.com/sites/default/files/resources/2021-07/AppNote_NGS_RespiratoryPanel_TEofViralPathogensusingRespiratoryVirusResearchPanel_12JUL21_Rev2.0.pdf

Noble, L. D., Scott, L. E., Munir, R., Du Plessis, M., Steegen, K., Hans, L., Marokane, P., Da Silva, P., & Stevens, W. S. (2022). Rapid Evaluation of the Xpert® Xpress CoV-2 plus and Xpert® Xpress CoV-2/Flu/RSV plus Tests. *Diagnostics (Basel, Switzerland)*, *13*(1), 34. https://doi.org/10.3390/diagnostics13010034

Nobusawa, E., & Sato, K. (2006). Comparison of the Mutation Rates of Human Influenza A and B Viruses. *Journal of Virology*, *80*(7), 3675–3678. https://doi.org/10.1128/JVI.80.7.3675-3678.2006

*Novel Virus Detection Using the Twist Comprehensive Viral Research Panel*. (2023, February). Twist Bioscience. https://www.twistbioscience.com/sites/default/files/resources/2023-02/DOC-001207_AppNote_CompViralResPanel-REV2-singles.pdf

Park, C.-E. (2021). Diagnostic Methods of Respiratory Virus Infections and Infection Control. *The Korean Journal of Clinical Laboratory Science*, *53*(1), 11–18. https://doi.org/10.15324/kjcls.2021.53.1.11

Paul, J. R., Trask, J. D., & Culotta, C. S. (1939). Poliomyelitic Virus in Sewage. *Science*, *90*(2333), 258–259. https://doi.org/10.1126/science.90.2333.258

Pavia, A. T. (2011). Viral Infections of the Lower Respiratory Tract: Old Viruses, New Viruses, and the Role of Diagnosis. *Clinical Infectious Diseases*, *52*(Supplement 4), S284–S289. https://doi.org/10.1093/cid/cir043

Payne, S. (2017). Family Coronaviridae. In *Viruses* (pp. 149–158). Elsevier. https://doi.org/10.1016/B978-0-12-803109-4.00017-9

Pedersen, B. S., & Quinlan, A. R. (2018). Mosdepth: Quick coverage calculation for genomes and exomes. *Bioinformatics*, *34*(5), 867–868. https://doi.org/10.1093/bioinformatics/btx699

Pineo, R. (2021). Four Flu Pandemics: Lessons that Need to Be Learned. *Journal of Developing Societies*, *37*(4), 398–448. https://doi.org/10.1177/0169796X211047221

Pollard, M. O., Gurdasani, D., Mentzer, A. J., Porter, T., & Sandhu, M. S. (2018). Long reads: Their purpose and place. *Human Molecular Genetics*, *27*(R2), R234–R241. https://doi.org/10.1093/hmg/ddy177

Principle and Workflow of Illumina Next-generation Sequencing. (2018, October 17). *CD Genomics*. https://www.cd-genomics.com/blog/principle-and-workflow-of-illumina-next-generation-sequencing/

Profaizer, T., Coonrod, E. M., Delgado, J. C., & Kumánovics, A. (2015). Report on the effects of fragment size, indexing, and read length on HLA sequencing on the Illumina MiSeq. *Human Immunology*, *76*(12), 897–902. https://doi.org/10.1016/j.humimm.2015.08.002

Puhach, O., Meyer, B., & Eckerle, I. (2022). SARS-CoV-2 viral load and shedding kinetics. *Nature Reviews Microbiology*. https://doi.org/10.1038/s41579-022-00822-w

Rachana Dubey & Subhasis Swain. (2019, October). Wastewater use in agriculture: Issues and prospects. *Agriculture & Food: E-Newsletter*, *1*(10), 254–257.

Rajaiya, J., Saha, A., Ismail, A. M., Zhou, X., Su, T., & Chodosh, J. (2021). Adenovirus and the Cornea: More Than Meets the Eye. *Viruses*, *13*(2), 293. https://doi.org/10.3390/v13020293

Rasmussen, L. D., Richter, S. R., Midgley, S. E., & Franck, K. T. (2022). Detecting SARS-CoV-2 Omicron B.1.1.529 Variant in Wastewater Samples by Using Nanopore Sequencing. *Emerging Infectious Diseases*, *28*(6). https://doi.org/10.3201/eid2806.220194

Rector, A., Bloemen, M., Thijssen, M., Pussig, B., Beuselinck, K., Van Ranst, M., & Wollants, E. (2024). Respiratory Viruses in Wastewater Compared with Clinical Samples, Leuven, Belgium. *Emerging Infectious Diseases*, *30*(1). https://doi.org/10.3201/eid3001.231011

Rehn, A., Braun, P., Knüpfer, M., Wölfel, R., Antwerpen, M. H., & Walter, M. C. (2021). Catching SARS-CoV-2 by Sequence Hybridization: A Comparative Analysis. *mSystems*, *6*(4), 10.1128/msystems.00392-21. https://doi.org/10.1128/msystems.00392-21

Reilly, C., Raghavan, A., & Bohjanen, P. (2006). Global assessment of cross-hybridization for oligonucleotide arrays. *Journal of Biomolecular Techniques: JBT*, *17*(2), 163–172.

Roppolo Brazell, L., Taylor, W., Aboufoul, L., Ferdous, J., A Schlueter, J., & Gibas, C. (2023). *Effective Target Capture/Enrichment of Respiratory Viruses from Wastewater v1*. https://doi.org/10.17504/protocols.io.8epv5xrjng1b/v1

Rothman, J. A., Loveless, T. B., Kapcia, J., Adams, E. D., Steele, J. A., Zimmer-Faust, A. G., Langlois, K., Wanless, D., Griffith, M., Mao, L., Chokry, J., Griffith, J. F., & Whiteson, K. L. (2021). RNA Viromics of Southern California Wastewater and Detection of SARS-CoV-2 Single-Nucleotide Variants. *Applied and Environmental Microbiology*, *87*(23), e01448-21. https://doi.org/10.1128/AEM.01448-21

Ryerson, A. B., Lang, D., Alazawi, M. A., Neyra, M., Hill, D. T., St. George, K., Fuschino, M., Lutterloh, E., Backenson, B., Rulli, S., Ruppert, P. S., Lawler, J., McGraw, N., Knecht, A., Gelman, I., Zucker, J. R., Omoregie, E., Kidd, S., Sugerman, D. E., … Yee, E. (2022). Wastewater Testing and Detection of Poliovirus Type 2 Genetically Linked to Virus Isolated from a Paralytic Polio Case—New York, March 9–October 11, 2022. *MMWR. Morbidity and Mortality Weekly Report*, *71*(44), 1418–1424. https://doi.org/10.15585/mmwr.mm7144e2

S Andrews. (2010, April 26). *FastQC: A quality control tool for high throughput sequence data*. Babraham Bioinformatics. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Sanders, C. J., Doherty, P. C., & Thomas, P. G. (2011). Respiratory epithelial cells in innate immunity to influenza virus infection. *Cell and Tissue Research*, *343*(1), 13–21. https://doi.org/10.1007/s00441-010-1043-z

*SARS-CoV-2 Wastewater Surveillance Testing Guide for Public Health Laboratories*. (2022, March). Association of Public Health Laboratories. https://www.aphl.org/aboutAPHL/publications/Documents/EH-2022-SARSCoV2-Wastewater-Surveillance-Testing-Guide.pdf

Scheuch, G. (2020). Breathing Is Enough: For the Spread of Influenza Virus and SARS-CoV-2 by Breathing Only. *Journal of Aerosol Medicine and Pulmonary Drug Delivery*, *33*(4), 230–234. https://doi.org/10.1089/jamp.2020.1616

Sears, S. D., Ferreccio, C., & Levine, M. M. (1986). Sensitivity of Moore sewer swabs for isolating Salmonella typhi. *Applied and Environmental Microbiology*, *51*(2), 425–426. https://doi.org/10.1128/aem.51.2.425-426.1986

Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for

FASTA/Q File Manipulation. *PLOS ONE*, *11*(10), e0163962. https://doi.org/10.1371/journal.pone.0163962

Shieh, W.-J. (2022). Human adenovirus infections in pediatric population—An update on clinico–pathologic correlation. *Biomedical Journal*, *45*(1), 38–49. https://doi.org/10.1016/j.bj.2021.08.009

Singer, A. C., Thompson, J. R., Filho, C. R. M., Street, R., Li, X., Castiglioni, S., & Thomas, K. V. (2023). A world of wastewater-based epidemiology. *Nature Water*, *1*(5), 408–415. https://doi.org/10.1038/s44221-023-00083-8

Smith, E. C., Popa, A., Chang, A., Masante, C., & Dutch, R. E. (2009). Viral entry mechanisms: The increasing diversity of paramyxovirus entry. *The FEBS Journal*, *276*(24), 7217–7227. https://doi.org/10.1111/j.1742-4658.2009.07401.x

Subbarao, K., & Mahanty, S. (2020). Respiratory Virus Infections: Understanding COVID-19. *Immunity*, *52*(6), 905–909. https://doi.org/10.1016/j.immuni.2020.05.004

*Surveillance of infectious disease through wastewater sequencing*. (2022). Illumina, Inc. https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/covidseq-wastewater-epidemiology-app-note-m-gl-00429/covidseq-wastewater-epidemiology-app-note-m-gl-00429.pdf

Suzanne Clancy. (2008). Genetics of the Influenza Virus. *Nature Education*, *1*(1), 83.

Takahashi, H., Ichinose, N., & Okada, Y. (2022). False-negative rate of SARS-CoV-2 RT-PCR tests and its relationship to test timing and illness severity: A case series. *IDCases*, *28*, e01496. https://doi.org/10.1016/j.idcr.2022.e01496

Talha N. Jilani, Radia T. Jamil, Andrew D. Nguyen, & Abdul H. Siddiqui. (2024). H1N1 Influenza. In *StatPearls [Internet]*. StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK513241/

Tanne, J. H. (2022). US faces triple epidemic of flu, RSV, and covid. *BMJ*, o2681. https://doi.org/10.1136/bmj.o2681

Tisza, M., Javornik Cregeen, S., Avadhanula, V., Zhang, P., Ayvaz, T., Feliz, K., Hoffman, K. L., Clark, J. R., Terwilliger, A., Ross, M. C., Cormier, J., Moreno, H., Wang, L., Payne, K., Henke, D., Troisi, C., Wu, F., Rios, J., Deegan, J., … Maresso, A. W. (2023). Wastewater sequencing reveals community and variant dynamics of the collective human virome. *Nature Communications*, *14*(1), 6878. https://doi.org/10.1038/s41467-023-42064-1

Tiwari, A., Radu, E., Kreuzinger, N., Ahmed, W., & Pitkänen, T. (2024). Key considerations for pathogen surveillance in wastewater. *Science of The Total Environment*, *945*, 173862. https://doi.org/10.1016/j.scitotenv.2024.173862

Toribio-Avedillo, D., Gómez-Gómez, C., Sala-Comorera, L., Rodríguez-Rubio, L., Carcereny, A., García-Pedemonte, D., Pintó, R. M., Guix, S., Galofré, B., Bosch, A., Merino, S., & Muniesa, M. (2023). Monitoring influenza and respiratory syncytial virus in wastewater. Beyond COVID-19. *Science of The Total Environment*, *892*, 164495. https://doi.org/10.1016/j.scitotenv.2023.164495

*Tracking COVID-19 Through Wastewater*. (2023, March 23). National Institutes of Health COVID-19 Research. https://covid19.nih.gov/news-and-stories/tracking-covid-19-through-wastewater-testing

Tregoning, J. S., & Schwarze, J. (2010). Respiratory viral infections in infants: Causes, clinical symptoms, virology, and immunology. *Clinical Microbiology Reviews*, *23*(1), 74–98. https://doi.org/10.1128/CMR.00032-09

Turner, R. B., & Lee, W.-M. (2009). Rhinovirus. In D. D. Richman, R. J. Whitley, & F. G.

Hayden (Eds.), *Clinical Virology* (1st ed., pp. 1063–1082). Wiley. https://doi.org/10.1128/9781555815981.ch46

*Twist Respiratory Virus Research Panel: Twist NGS target enrichment panel for accurate identification of respiratory viruses beyond SARS-CoV-2*. (2020, July 2). Twist Bioscience. https://www.twistbioscience.com/sites/default/files/resources/2020-07/ProductSheet_NGS_RespiratoryVirusResearchPanel_2JUL20_Rev1.1.pdf

Van Doorn, H. R., & Yu, H. (2020). Viral Respiratory Infections. In *Hunter's Tropical Medicine and Emerging Infectious Diseases* (pp. 284–288). Elsevier. https://doi.org/10.1016/B978-0-323-55512-8.00033-8

Van Elden, L. J. R., Van Kraaij, M. G. J., Nijhuis, M., Hendriksen, K. A. W., Dekker, A. W., Rozenberg‐Arska, M., & Van Loon, A. M. (2002). Polymerase Chain Reaction Is More Sensitive than Viral Culture and Antigen Testing for the Detection of Respiratory Viruses in Adults with Hematological Cancer and Pneumonia. *Clinical Infectious Diseases*, *34*(2), 177–183. https://doi.org/10.1086/338238

Vasudevan, H. N., Xu, P., Servellita, V., Miller, S., Liu, L., Gopez, A., Chiu, C. Y., & Abate, A. R. (2021). Digital droplet PCR accurately quantifies SARS-CoV-2 viral load from crude lysate without nucleic acid purification. *Scientific Reports*, *11*(1), 780. https://doi.org/10.1038/s41598-020-80715-1

V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. (2021). Coronavirus biology and replication: Implications for SARS-CoV-2. *Nature Reviews Microbiology*, *19*(3), 155–170. https://doi.org/10.1038/s41579-020-00468-6

Walter Doerfler. (1996). Chapter 67 Adenoviruses. In Baron S (Ed.), *Medical Microbiology* (4th ed.). https://www.ncbi.nlm.nih.gov/books/NBK8503/

Wang, C. C., Prather, K. A., Sznitman, J., Jimenez, J. L., Lakdawala, S. S., Tufekci, Z., & Marr, L. C. (2021). Airborne transmission of respiratory viruses. *Science*, *373*(6558), eabd9149. https://doi.org/10.1126/science.abd9149

*WHO Health Emergencies Programme: WHO COVID-19 dashboard*. (2024, October 30). World Health Organization Data. https://data.who.int/dashboards/covid19/cases?n=c

Wong, J. C. C., Tay, M., Hapuarachchi, H. C., Lee, B., Yeo, G., Maliki, D., Lee, W., Mohamed Suhaimi, N.-A., Chio, K., Tan, W. C. H., & Ng, L. C. (2024). Case report: Zika surveillance complemented with wastewater and mosquito testing. *eBioMedicine*, *101*, 105020. https://doi.org/10.1016/j.ebiom.2024.105020

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, *20*(1), 257. https://doi.org/10.1186/s13059-019-1891-0

*Xpert® Xpress CoV-2/Flu/RSV plus Instructions for Use*. (2023, July). Cepheid Innovation. https://www.fda.gov/media/152163/download

Ye, Y., Ellenberg, R. M., Graham, K. E., & Wigginton, K. R. (2016). Survivability, Partitioning, and Recovery of Enveloped Viruses in Untreated Municipal Wastewater. *Environmental Science & Technology*, *50*(10), 5077–5085. https://doi.org/10.1021/acs.est.6b00876

Ye, Z.-W., Yuan, S., Yuen, K.-S., Fung, S.-Y., Chan, C.-P., & Jin, D.-Y. (2020). Zoonotic origins of human coronaviruses. *International Journal of Biological Sciences*, *16*(10), 1686–1697. https://doi.org/10.7150/ijbs.45472

Yuen, K.-S., Ye, Z.-W., Fung, S.-Y., Chan, C.-P., & Jin, D.-Y. (2020). SARS-CoV-2 and COVID-19: The most important research questions. *Cell & Bioscience*, *10*(1), 40. https://doi.org/10.1186/s13578-020-00404-4

Zafeiriadou, A., Kaltsis, L., Kostakis, M., Kapes, V., Thomaidis, N. S., & Markou, A. (2023). Wastewater surveillance of the most common circulating respiratory viruses in Athens: The impact of COVID-19 on their seasonality. *Science of The Total Environment*, *900*, 166136. https://doi.org/10.1016/j.scitotenv.2023.166136

Zafeiriadou, A., Nano, K., Thomaidis, N. S., & Markou, A. (2024). Evaluation of PCR-enhancing approaches to reduce inhibition in wastewater samples and enhance viral load measurements. *Science of The Total Environment*, *955*, 176768. https://doi.org/10.1016/j.scitotenv.2024.176768

Zhang, H., Jain, C., & Aluru, S. (2020). A comprehensive evaluation of long read error correction methods. *BMC Genomics*, *21*(S6), 889. https://doi.org/10.1186/s12864-020-07227-0

Zhu, Z., Fodor, E., & Keown, J. R. (2023). A structural understanding of influenza virus genome replication. *Trends in Microbiology*, *31*(3), 308–319. https://doi.org/10.1016/j.tim.2022.09.015

Zhu, Z., Liu, Y., Xu, L., Guan, W., Zhang, X., Qi, T., Shi, B., Song, Z., Liu, X., Wan, Y., Tian, D., He, J., Zhang, X., Wu, M., Lu, H., Lu, S., Zhang, Z., Yuan, Z., & Hu, Y. (2015). Extra-pulmonary viral shedding in H7N9 Avian Influenza patients. *Journal of Clinical Virology*, *69*, 30–32. https://doi.org/10.1016/j.jcv.2015.05.013

Zimmerman, R. K., Balasubramani, G. K., D'Agostino, H. E. A., Clarke, L., Yassin, M., Middleton, D. B., Silveira, F. P., Wheeler, N. D., Landis, J., Peterson, A., Suyama, J., Weissman, A., & Nowalk, M. P. (2022). Population‐based hospitalization burden estimates for respiratory viruses, 2015–2019. *Influenza and Other Respiratory Viruses*, *16*(6), 1133–1140. https://doi.org/10.1111/irv.13040

# PUBLICATIONS

First author publications:

Roppolo Brazell, L., Taylor, W., Aboufoul, L., Ferdous, J., Harris, A., Schlueter, J., Gibas, C. “Effective target capture/enrichment of respiratory viruses from wastewater”. Protocols.io, 4 October 2023, dx.doi.org/10.17504/protocols.io.8epv5xrjng1b/v1.

Roppolo Brazell, L., Stetz, S., Hipp, A., Taylor, S., Stark, N., Jensen, K., Juel, Md. A. I., Deegan, P., Munir, M., Schlueter, J., Weller, J., Gibas, C. “Environmental screening for surface SARS-COV-2 contamination in urban high-touch areas.” *medRxiv*, 4 May 2021, https://doi.org/10.1101/2021.05.04.21256107.

Other publications:

Khanal N, Roppolo Brazell, L, Juel MAI, Gibas C, Schlueter J, Munir M. Investigation of SARS-CoV-2 Contamination of Indoor Air and Highly Touched Surfaces On-Campus Buildings. *Applied Microbiology*. 2024; 4(3):1384-1395. https://doi.org/10.3390/applmicrobiol4030095.

Gibas, C., Lambirth, K., Mittal, N., Juel, M. A. I., Barua, V. B., Roppolo Brazell, L., Hinton, K., Lontai, J., Stark, N., Young, I., Quach, C., Russ, M., Kauer, J., Nicolosi, B., Chen, D., Akella, S., Tang, W., Schlueter, J., & Munir, M. (2021). Implementing building-level SARS-CoV-2 wastewater surveillance on a university campus. *The Science of the total environment*, *782*, 146749. https://doi.org/10.1016/j.scitotenv.2021.146749.

# SUPPLEMENTARY INFORMATION

## CHAPTER 2

Supplementary Table 1. Reference genomes for the Twist Respiratory Virus Research Panel
Supplementary Table 2. Breadth and depth of coverage; One Codex detections
Supplementary Table 3. Kraken2 classifications; on and off target
Supplementary Figures 1 & 2. Kraken2 heatmaps for on and off target hits
Supplementary Table 4. Bowtie2 relative abundances
Supplementary Tables 5 & 6. Raw TblastX results for clinical and wastewater samples
Supplementary Table 7. Unique 500bp chunks identified by TblastX
Supplementary Figures 3 & 4. TblastX off-target alignments at genera level for clinical and wastewater samples
Supplementary Figures 5 & 6. TblastX off-target alignments at species level and sequence depth for clinical and wastewater samples
Supplementary Figures 7 & 8. Depth vs. off-target alignment hits at genus and species level
Supplementary Table 8. One Codex vs. Bowtie2 vs. Kraken2 concordance

## CHAPTER 3

Supplementary Table 1. NFW and wastewater spike-in compositions
Supplementary Table 2. Control dilutions
Supplementary Table 3. Negative wastewater sample metadata
Supplementary Table 4. Reference genomes for the Twist Respiratory Virus Research Panel
Supplementary Table 5. Bowtie2 breadth of coverage and average sequencing depth
Supplementary Table 6. Previous experiment's Kraken2 abundances
Supplementary Files 1 & 2. NFW and WW spike-in Krona plots
Supplementary Figure 1. NTC contamination

## CHAPTER 4

Supplementary Table 1. Illumina RVOP Reference Genomes
Supplementary Table 2. Bowtie2 relative abundances
Supplementary Table 3. Kraken2 relative abundances
Supplementary Figure 1. On vs. off-target distributions
Supplementary Figure 2. Off-target Kraken2 viral abundances
Supplementary Table 4. TblastX off-target alignments