

Detection and Characterization of Viruses in the Environment Using Established and Novel Sequencing Approaches

by

Delaney Nash

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Biology

Waterloo, Ontario, Canada, 2025

©Delaney Nash 2025

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Dr. Evelien Adriaenssens Research Leader, Quadram Institute
Supervisor(s)	Dr. Trevor C. Charles Professor
Supervisor(s)	Dr. Jozef I. Nissimov Assistant Professor
Internal Member	Dr. Kirsten M. Müller Professor
Internal-external Member	Dr. Valerie Ward Associate Professor
Other Member	Dr. Josh D. Neufeld Professor

Author Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Delaney Nash was the sole author for chapters 1 and 5, which were written under the supervision of Dr. Jozef I. Nissimov and Dr. Trevor C. Charles and were not written for publication.

This thesis contains three manuscripts written for publication (Chapter 2, 3, and 4). Exceptions to sole authorship are as follows:

Research Presented in Chapter 2:

This research was conducted at the University of Waterloo by Delaney Nash under the supervision of Dr. Jozef I. Nissimov and Dr. Trevor C. Charles. The research project was conceptualized by Delaney Nash. The IAV H3N2 and RSV A sequencing assays were developed by Delaney Nash. Initial assay testing was conducted by Delaney Nash and Ria Menon. RNA extractions from wastewater samples were conducted by Yemurayi Hungwe. Molecular work and preparation of sequencing libraries of wastewater RNA extracts was conducted by Delaney Nash. Samples were loaded and run on an NextSeq 2000 (Illumina) by Samran Prasla and Alyssa K. Overton. Delaney Nash conducted sequence data analysis. Jennifer J. Knapp developed software tool Alvoc, software testing and validation was conducted by Delaney Nash and Jennifer J. Knapp. Delaney Nash wrote the original manuscript. Delaney Nash, Jennifer J. Knapp, Alyssa K. Overton, Ria Menon, Jozef I. Nissimov, and Dr. Trevor C. Charles revised and reviewed the manuscript.

The preprint for this manuscript is available on BioRxiv DOI: 10.1101/2025.07.29.667219.

Research presented in Chapter 3:

This research was conducted at the University of Waterloo by Delaney Nash under the supervision of Dr. Jozef I Nissimov and Dr. Trevor C. Charles. Delaney Nash, Isaac Ellmen, Dr. Trevor C. Charles, Dr. Jozef I. Nissimov, and Dr. JiuJun Cheng conceptualized the project.

ToBRFV-Seq assay was designed by Delaney Nash and Isaac Ellmen. ToBRFV-Seq assay testing and validation was performed by Delaney Nash and Ria Menon. ToBRFV-assay and shotgun sequencing on wastewater samples was conducted by Delaney Nash. Data analysis and interpretation was conducted by Delaney Nash with support from Jennifer J. Knapp. Altob software was developed, tested, and optimized by Delaney Nash, Isaac Ellmen, Jennifer J. Knapp, and Michael D.J. Lynch. The original draft of the manuscript was written by Delaney Nash. Delaney Nash, Jennifer J. Knapp, Alyssa K. Overton, Ria Menon, Dr. Trevor C. Charles, Dr. Jozef I. Nissimov, Dr. Michael D.J. Lynch, and Dr. JiuJun Cheng reviewed and revised the manuscript.

Citation:

Nash D, Ellmen I, Knapp JJ, Menon R, Overton AK, Cheng J, Lynch MDJ, Nissimov JI, Charles TC. A Novel Tiled Amplicon Sequencing Assay Targeting the Tomato Brown Rugose Fruit Virus (ToBRFV) Genome Reveals Widespread Distribution in Municipal Wastewater Treatment Systems in the Province of Ontario, Canada. *Viruses*, March 2024, DOI: 10.3390/v16030460.

Research presented in Chapter 4:

This research was conducted at both the University of Waterloo by Delaney Nash under the supervision of Dr. Jozef I. Nissimov and Dr. Trevor C. Charles, and the University of Toronto Mississauga by Dr. Christine N. Palermo and Dr. Ichiro Inamoto, supervised by Dr. Steven M. Short. Isolation, culturing, DNA extraction, and Illumina sequencing of the CpV-BQ1 virus was performed at the University of Toronto Mississauga. Delaney Nash conducted Nanopore sequencing, targeted amplicon sequencing, genome assembly and annotation, characterized genomic functional potential, plotted data, interpreted results, and wrote the original draft of the manuscript. Funding for this project was obtained by Dr. Trevor C. Charles, Dr. Jozef I. Nissimov, and Dr. Steven M. Short. Assistance with manuscript writing and supervision of the project was done by Dr. Jozef I. Nissimov and Dr. Steven M. Short. Co-authors Dr. Trevor C. Charles, Dr.

Ichiro Inamoto, Dr. Christine N. Palermo, Dr. Jozef I. Nissimov, and Dr. Steven M. Short all helped revise and edit the final manuscript.

Citation:

Nash D, Palermo CN, Inamoto I, Charles TC, Nissimov JI, Short SM. Hybrid sequencing reveals the genome of a *Chrysochromulina parva* virus and highlight its distinct replication strategy. BMC Genomics, May 2025, DOI: 10.1186/s12864-025-11700-z.

Abstract

Viruses play critical roles in both the environment and public health systems. This thesis integrates genomic sequencing methodologies to detect, characterize, and monitor viral diversity, advancing both fundamental ecological research and applied pathogen surveillance. Through three focused research projects, this work demonstrates how targeted and metagenomic sequencing strategies can address key knowledge gaps in environmental virology, plant pathology, and viral public health surveillance.

The first objective (Chapter 2) was to develop and apply novel tiled-amplicon sequencing assays for Influenza A virus (IAV) H3N2 and respiratory syncytial virus (RSV) A in wastewater-based surveillance. These assays successfully recovered near-complete viral genomes from wastewater samples collected during the peak of the 2023/2024 respiratory virus season. Genomic coverage trends mirrored clinical case data within the associated public health region, validating the approach as a complementary tool to traditional clinical surveillance. Variant deconvolution analyses revealed distinct spatiotemporal patterns in variant distribution, demonstrating the capacity of these assays and bioinformatic tools to resolve community-level transmission dynamics and emerging viral lineages.

The second objective (Chapter 3) extended this surveillance framework to the agriculturally significant Tomato brown rugose fruit virus (ToBRFV). A multiplexed tiled-amplicon assay was developed, achieving significant improvements in viral genome recovery compared to metagenomic RNA shotgun sequencing while reducing sequencing costs and resource requirements. ToBRFV clades circulating in Ontario wastewater were identified, including variants later recognized through global surveillance initiatives, underscoring the assay's potential for early detection and the value of wastewater systems as environmental reservoirs for plant viruses.

The third objective (Chapter 4) involved the assembly and characterization of the complete genome of a novel freshwater algal virus, *Chrysochromulina parva* virus BQ1 (CpV-BQ1). Using a hybrid Nanopore and Illumina sequencing strategy, a high-quality 165,454 bp genome was assembled and annotated, revealing hallmark nucleocytoplasmic large DNA virus (NCLDV) genes and diverse functional categories associated with viral replication, host manipulation, and capsid formation. This work offers important insights into viral infection of the ecologically important freshwater algal species *C. parva*. Moreover, it establishes a methodological framework for the complete sequencing, assembly, and annotation of algal virus genomes.

Collectively, this thesis advances environmental virology and viral surveillance systems by developing scalable, sensitive, and cost-effective genomic workflows for virus detection and characterization in freshwater environments. By integrating virus ecology with applied public health and agricultural surveillance within a One Health framework, this work underscores the interconnectedness of environmental, human, and agricultural systems and provides practical guidance for future viral genomics research and pathogen monitoring programs.

Acknowledgements

Thank you to my co-supervisors Dr. Jozef I. Nissimov and Dr. Trevor C. Charles for their supervision, guidance, and mentorship throughout my PhD. Additionally, thank you to my committee members Dr. Kirsten M. Müller and Dr. Josh D. Neufeld for their support and guidance on PhD projects. As well as my external committee members Dr. Evelien Adriaenssens and Dr. Valerie Ward for all of their helpful advice and feedback on my thesis and during my defense.

Thank you to all co-authors of the manuscripts that were included in this thesis, Alyssa K. Overton, Jennifer J. Knapp, Ria Menon, Yemurayi Hungwe, Samran Prasla, Dr. Michael D.J. Lynch, Isaac Ellmen, Dr. Jiujun Cheng, Dr. Christine N. Palermo, Dr. Ichiro Inamoto, Dr. Steven M. Short, Dr. Jozef I. Nissimov, and Dr. Trevor C. Charles.

Thank you to wastewater treatment plant operators and public health units for providing samples for testing.

Thank you to all my family, friends, and especially my partner Max McKee for their constant support and love throughout the pursuit of my PhD. This journey would not have been possible without their endless encouragement and support.

Dedication

To

Lindsey Anne Taylor

My friend, confidant, champion, and bonus mum.

*“May the road rise up to meet you,
may the wind be always at your back,
may the sun shine warm upon your face,
and the rain fall soft upon your fields.*

Until we meet again.”

Table of Contents

List of Figures	xiv
List of Tables	xv
List of Abbreviations	xvi
Chapter 1. An Introduction to Viruses: Replication, Ecological and Host Impacts, and Genome Sequencing Techniques	1
1.1 Defining Attributes of Viruses	1
1.2 The Baltimore Classification System for Viruses	1
1.3 Virus Replication and General Lifecycle	3
1.4 Host Hijacking–Mechanisms to Increase Viral Propagation	4
1.5 The Impact of Viruses on Ecosystems	5
1.6 Impacts of Viruses on Humans—from Pandemics to Food Instability	7
1.7 Genomic Sequencing: a Tool to Study Viruses and their Impacts	9
1.8 Research Focus and Objectives	11
Chapter 2. A Genome Sequence Variant Monitoring Program for Seasonal Influenza A virus H3N2 and Respiratory Syncytial Virus A using Wastewater-Based Surveillance in Ontario, Canada ..	15
2.1. Introduction – Influenza A H3N2 and Respiratory Syncytial Virus A burden, infections, and detection	15
2.2 Materials and Methods	21
2.2.1 Sampling Sites and Wastewater Collection	21
2.2.2 Wastewater and Water Virus Spike-in Samples to Test Primer Specificity ..	21
2.2.3 Wastewater Virus Isolation and RNA Extraction	22
2.2.4 IAV H3N2 and RSV A Tiled-Amplicon Primer Design	23
2.2.5 Primer Pooling	24
2.2.6 cDNA Synthesis and PCR Amplification of RNA Extracts from Seasonal WW Samples	24
2.2.7 PCR Amplification Methods for Virus Spike-in Samples	25
2.2.8 Library Preparation and Sequencing	26
2.2.9 Sequencing Read Quality Control, Alignment, and Calculation of Genome Coverage	27
2.2.10 Variant Calling with Freyja and Alvoc	27
2.2.11 Intraspecific Amplicon Coverage Calculation	28
2.3 Results and Discussion	29
2.3.1 Comparison of Novel and Existing IAV H3N2 and RSV A Targeted Primers	29
2.3.2 Assessment of Genomic Coverage and Comparison to Clinical Percent Positive Tests	31
2.3.3 Evaluation of Synthetic Reads with Freyja and Alvoc	36
2.3.4 Application of Freyja and Alvoc for the Assessment of IAV H3N2 and RSV A Clades in Seasonal Wastewater Samples	41
2.3.5 Intraspecific Amplicon Analysis of Tiled-Amplicon Sequencing Assays ...	45
2.3.6 Future Goals and Assay Optimization	47
2.4 Conclusions	50
Chapter 3. A Novel Tiled-Amplicon Sequencing Assay Targeting the Tomato Brown Rugose Fruit Virus (ToBRFV) Genome	51
3.1 Introduction – Tomato Brown Rugose Fruit Virus Incidence, Genome, and Detection	51
3.2 Materials and Methods	57

3.2.1	<i>Viral RNA Extraction from Wastewater Influent</i>	57
3.2.2	<i>Primer Design and Pooling</i>	58
3.2.3	<i>ToBRFV-Targeted Tiled Amplicon Sequencing Library Preparation</i>	59
3.2.4	<i>RNA Shotgun Sequencing Library Preparation</i>	59
3.2.5	<i>Library Pooling and Illumina NextSeq Sequencing</i>	60
3.2.6	<i>Read Processing, Taxonomic Classification, and ToBRFV Genome Alignment</i>	60
3.2.7	<i>Altob Implementation and Synthetic Read Simulation</i>	61
3.3	<i>Results</i>	63
3.3.1	<i>In Silico Primer Binding Analysis of ToBRFV Genome Sequences of 125 Strains</i>	63
3.3.2	<i>Read Quality Control and Adaptor Trimming</i>	64
3.3.3	<i>Taxonomic Profile of WWI Samples Prepared by RNA Shotgun and ToBRFV-Seq Methods</i>	65
3.3.4	<i>Mapping Reads to a ToBRFV Reference Genome</i>	68
3.3.5	<i>Estimating the Relative Abundance of ToBRFV Clades in Wastewater Influent</i>	70
3.3.6	<i>Assessment of Viral Shotgun Reads at a Species Level</i>	72
3.4	<i>Discussion</i>	75
3.4.1	<i>Evaluation of ToBRFV-Seq Assay Robustness and Specificity</i>	75
3.4.2	<i>Possible Assay Applications to Determine ToBRFV Transmission and Prevalence</i>	76
3.4.3	<i>Effectiveness of and Possible Improvements to Viral Capture and ToBRFV-Seq Procedures</i>	77
3.5	<i>Conclusion</i>	78
<i>Chapter 4. Hybrid Sequencing, Annotation, and Analysis of a Novel Algal Virus Genome—Chrysochromulina parva virus BQ1 (CpV-BQ1)</i>		79
4.1	<i>Introduction – Eukaryotic Algal Viruses of the Phylum Nucleocytoviricota</i>	79
4.2	<i>Materials and Methods</i>	82
4.2.1	<i>Host and Virus Cultivation</i>	82
4.2.2	<i>Nucleic Acid Extraction</i>	82
4.2.3	<i>Illumina and Nanopore library preparation and sequencing</i>	83
4.2.4	<i>Read processing and Genome Assembly</i>	83
4.2.5	<i>Primer Design, PCR Amplification, and Analysis of Random Genomic Regions</i>	85
4.2.6	<i>Open Reading Frame (ORF) Prediction</i>	86
4.2.7	<i>Gene Annotation</i>	86
4.2.8	<i>Taxonomic classification</i>	87
4.3	<i>Results and Discussion</i>	88
4.3.1	<i>Assembly of the CpV-BQ1 Genome</i>	88
4.3.2	<i>Characterization and Description of the CpV-BQ1 Genome</i>	92
4.3.3	<i>Taxonomy</i>	98
4.3.4	<i>DNA and RNA Replication, Recombination, and Repair Enzymes</i>	100
4.3.5	<i>Transcription</i>	100
4.3.6	<i>Nucleotide Transport and Metabolism Proteins</i>	102
4.3.7	<i>DNA Methylation</i>	104
4.3.8	<i>Sugar Manipulation</i>	106
4.3.9	<i>Protein and Lipid Binding, Synthesis, and Modifications</i>	107
4.3.10	<i>DNA Packaging and Genome Completeness</i>	108
4.3.11	<i>Virion Capsid and Associated Structural Proteins</i>	109

4.3.12 Comparison of CpV-BQ1 and CpV-BQ2 Genomes	111
4.4 Conclusions	113
Chapter 5. Summary and Future Perspectives	115
5.1 Overview of Objectives and Conclusions	115
5.2 Summary of Key Findings	117
5.2.1 Wastewater-based Surveillance of Influenza A H3N2 and RSV A	117
5.2.2 Development of a Tiled-Amplicon Sequencing Assay for the ToBRFV ...	118
5.2.3 Assembly and Characterization of the CpV-BQ1 Genome	119
5.3 Integration of Thesis Objectives and Contributions	120
5.4 Final Remarks	121
References	122
Appendices	149
Supplementary Table 2.1. Locations of Wastewater Sample Collection	149
Supplementary Table 2.2. Dates of Wastewater Sample Collection	149
Supplementary Table 2.3. IAV H3N2 Primers	150
Supplementary Table 2.4. RSV A Primers	152
Supplementary Figure 2.1. Location of Public Health Regions in Southern Ontario	154
Supplementary Table 2.5. List of IAV H3N2 Genes	155
Supplementary Table 2.6. List of RSV A Genes	155
Supplementary Table 3.1. List of Samples	156
Supplementary Table 3.2. Accession numbers of 179 ToBRFV Strains Obtained from NCBI	156
Supplementary Table 3.3. List of ToBRFV-Seq Primers	159
Supplementary Figure 3.1. Comparison of ToBRFV Phylogenetic Analysis Resolving Clades	161
Supplementary Table 3.4. Twenty-six Virus Species in Shotgun Samples	162
Supplementary Table 4.1. CpV-BQ1 Genome Specific Primers	163
Supplementary Table 4.2. Assembled PCR Amplified Regions of the CpV-BQ1 Genome	164
Supplementary Table 4.3. CpV-BQ1 Predicted ORFs and Functional Annotation	169
Supplementary Table 4.4. Genes used for Phylogenetic Analysis	180

List of Figures

- Figure 1.1. *Flowchart of Thesis Objectives*
- Figure 2.1. *Comparison of Influenza A H3N2 and RSV A Amplicon Sequencing Assays on Virus Spiked and Seasonal Wastewater Samples*
- Figure 2.2. *Influenza A H3N2 and RSV A Percent of Genomic Coverage Versus Percent of Weekly Positive Tests in Public Health Unit Regions*
- Figure 2.3. *Comparison of Freyja and Alvoc IAV H3N2 Variant Calling using Synthetic Read Data*
- Figure 2.4. *Comparison of Freyja and Alvoc RSV A Variant Calling using Synthetic Data*
- Figure 2.5. *IAV H3N2 Variant Calls with Freyja and Alvoc from Seasonal Wastewater Samples*
- Figure 2.6. *RSV A Variant Calls with Freyja and Alvoc in Wastewater Samples*
- Figure 2.7. *IAV H3N2 and RSV A Amplicons with Coverage in Wastewater Samples*
- Figure 3.1. *ToBRFV-Specific Primer Binding Sites and Resulting Amplicons Mapped to the ToBRFV MK133093 Genome Sequence*
- Figure 3.2. *Taxonomic Classification of RNA Shotgun and ToBRFV-Seq Read Pairs at the Domain and Species-level*
- Figure 3.3. *ToBRFV Clade-Abundance Estimates of Synthetic Read Datasets and WWI Samples with Altob*
- Figure 3.4. *Evaluation of Virus-Species Abundance in WWI Samples using Counts of Viral Reads Taxonomically Classified by Kraken2*
- Figure 4.1. *Phylogenetic Tree of Clustered Assembled Contigs*
- Figure 4.2. *Dotplot Analysis of 16 Assembled Contigs*
- Figure 4.3. *Diagram of CpV-BQ1 Genome Primer Binding Sites*
- Figure 4.4. *ORF Predictions in the CpV-BQ1 Genome*
- Figure 4.5. *Phylogenetic Analysis of Three CpV-BQ1 Genes Universally Found in NCLDV Genomes*
- Figure 4.6. *Biosynthesis of Deoxythymidine Triphosphate (dTTP)*
- Figure 4.7. *Potential Hairpin Structures of CpV-BQ1 genome ends*

List of Tables

Table 1.1. *Defining Features of Baltimore Classification Groups*

Table 3.1. *In silico Analysis of ToBRFV-specific Primer Binding to 137 Tobamovirus Genomes*

Table 3.2. *Total and Filtered Reads from RNA Shotgun and ToBRFV-Seq Prepared Sequencing Libraries*

Table 3.3. *Alignment and Coverage of RNA Shotgun and ToBRFV-Seq Reads Mapped to a ToBRFV Genome*

Table 4.1. *List of All Characterised Genes in the CpV-BQ1 Genome*

List of Abbreviations

Abbreviation	Full Term
Alvoc	abundance learning for variants of concern
AMGs	auxiliary metabolic genes
ASFV	African swine fever virus
bp	base pairs
CbV-PW1	Chrysochromulina brevifilum virus PW1
cDNA	complementary DNA
CDS	coding sequences
ChfuGV	choristoneura fumiferana granulovirus
CpV-BQ1	chrysochromulina parva virus BQ1
CR	clinical RSV
DdRp	DNA dependent RNA polymerase
DNA	Deoxyribose nucleic acid
dTTP	deoxythymidine triphosphate
ER	endoplasmic reticulum
ERV	essential for respiration and vegetative growth
F	fusion protein
GB	Gigabases
Gbp	Gigabase pairs
GT	glycosyltransferase
HA	hemagglutinin
HaV	heterosigma akashiwo virus
HMM	Hidden Markov Model
IAV	Influenza A virus
ICTV	International Committee on Taxonomy of Viruses
kb	kilobases
kbp	kilobase pairs
M	matrix protein
MB	MB-Tuni
Mbp	megabase pairs
MCP	major capsid protein
MERS-CoV	middle East Respiratory Syndrome Coronavirus
MP	movement protein
mRNA	messenger RNA
MSA	multiple sequence alignment
MTases	methyltransferases
NA	neuraminidase
NCLDV	nucleocytoplasmic large DNA viruses

NP	nucleocapsid protein
NS	non-structural proteins
NTPs	nucleoside triphosphates
ORFs	open reading frames
PA	polymerase PA
PB1	polymerase PB1
PB2	polymerase PB2
PBCV-1	paramecium bursaria chlorella virus 1
PEG	polyethylene glycol
PgV-08T	phaeocystis globosa virus 08T
PhyML	phylogenetic maximum likelihood
PLVs	polinton-like viruses
PMMoV	pepper mild mottle virus
polB	DNA polymerase B
RdRp	RNA-dependent RNA polymerase
ReMV	rehmannia mosaic virus
R-M	restriction modification
RNA	Ribonucleic acid
RNA pol	RNA polymerase II
rRNA	ribosomal RNA
RSV	respiratory syncytial virus
RT	reverse transcriptase
RT-qPCR	reverse transcriptase quantitative polymerase chain reaction
SAM	S-adenosyl-L-methionine
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
TA	tiled-amplicon
TCID ₅₀	tissue culture infectious dose 50%
TFs	transcription factors
TMV	tobacco mosaic virus
ToBRFV	tomato brown rugose fruit virus
ToBRFV-Seq	tomato brown rugose fruit virus targeted sequencing assay
ToMMV	tomato mottle mosaic virus
ToMV	tomato mosaic virus
tRNA	transfer RNA
VLTF-3	virus late transcription factor 3
WBS	wastewater-based surveillance
WW	wastewater
WWI	wastewater influent

Chapter 1. An Introduction to Viruses—Replication, Ecological and Host Impacts, and Genome Sequencing Techniques

1.1. Defining Attributes of Viruses

Viruses are abundant and ubiquitous on Earth [1–3]. They can infect Bacteria, Eukarya, and Archaea [4], and are defined as obligate intracellular parasites, meaning they can only replicate within a living host cell [4]. They consist of a nucleic acid genome, made of either single or double stranded RNA or DNA, and are enclosed within a protein capsule, which is referred to as the capsid. In some cases, the capsid is enclosed within a host-derived lipid membrane [4]. A complete infectious virus particle is referred to as a virion [4]. The genome size, structure, and topology, as well as capsid morphology and infection strategy vary greatly between viral species.

1.2. The Baltimore Classification System for Viruses

The diversity of virus genomes necessitates a diverse array of strategies for genome replication, which are reflected in the Baltimore classification system [4–6]. The Baltimore system includes seven categories that group viruses based on the type of their nucleic acid genome and its strandedness (+/-), as well as the mRNA synthesis mechanism (Table 1.1) [5–7]. Group I viruses, such as the algal virus *Chrysochromulina parva* virus BQ1 (Chapter 4), have a dsDNA genome and use an DNA dependent RNA polymerase (DdRp) to transcribe mRNA [6]. Group II viruses have a (+)ssDNA genome, with replication and mRNA synthesis requiring a dsDNA intermediate [6]. Group III viruses use a dsRNA genome, with replication and mRNA synthesis requiring a viral encoded RNA-dependent RNA polymerase (RdRp) [6]. Group IV viruses, such as the tomato brown rugose fruit virus (ToBRFV; Chapter 3), have a (+)ssRNA genome which in itself acts as mRNA template, new genomes are synthesized using an RdRp and (-)ssRNA intermediate [6]. Group V viruses, such as Influenza A virus (IAV) and respiratory syncytial virus (RSV) (Chapter 2), have a (-)ssRNA genome from which complementary mRNA transcripts are synthesized by

an RdRp [6]. Group VI viruses use a (+)ssRNA genome which is reverse transcribed into a DNA intermediate, from which mRNA is synthesized [6]. Lastly, group VII viruses have a dsDNA genome which mRNA is synthesized from, but, use reverse transcriptase (RT) and an RNA intermediate for genome replication [6].

Table 1.1. Defining Features of Baltimore Classification Groups. The seven Baltimore classification groups, their respective type of nucleic acid genome, mRNA synthesis enzyme and polarity, DNA replication enzyme, and corresponding virus realm(s).

Group	Nucleic Acid type	mRNA synthesis and polarity	Genome synthesis	Virus Realms
1	dsDNA	DdRp, mRNA on both strands	DNApol	Varidnaviria, Duplodnaviria, Adnaviria, Monodnaviria
2	(+)ssDNA	DdRp, + strand mRNA polarity	DNApol	Monodnaviria, Varidnaviria
3	dsRNA	RdRp, + strand mRNA polarity	RdRp	Riboviria
4	(+)ssRNA	RdRp, + strand mRNA polarity	RdRp	Riboviria
5	(-)ssRNA	RdRp, + strand mRNA polarity	RdRp	Riboviria, Ribozviria
6	(+)ssRNA	RT, + strand mRNA polarity	RT	Riboviria
7	dsDNA	RT, + strand mRNA polarity	RT	Riboviria

The Baltimore classification system of viruses was devised in 1971 prior to the development of sequencing technology [5,6]. Since, sequence analysis has shown that the Baltimore groups are not monophyletic and that they share complex evolutionary history. However the observations they are based upon still inform our understanding of virus replication and evolution [6]. The International Committee on Taxonomy of Viruses (ICTV) now recognizes seven virus realms, where each includes viruses belonging to one or more of the Baltimore groups (Table 1.1) [6,8,9]. Unlike the three domains of life, viruses do not share a common ancestor; they are a paraphyletic group and lack a single universal gene [6]. Instead, approximately 20 hallmark viral genes are used to define virus realms [6,9]. For example, almost all RNA viruses encode a homologous RdRp or RT gene. As such, these hallmark features unify and distinguish the monophyletic realm *Riboviria* [6,9]. Throughout viral taxonomic levels and groups, monophyletic clades are defined by hallmark genes that are homologous and universally found in viruses of said group. Due to

this, the taxonomic classification of viruses requires specific knowledge of clade evolution and hallmark genes.

1.3. Virus Replication and General Lifecycle

Despite the genomic and structural diversity amongst viruses, the general life cycle of most viruses follows similar steps. This includes: cellular entry, genome uncoating, early mRNA synthesis, genome replication, late gene expression, virion assembly, and host cell exit [4,10].

Viral entry into cells is specially adapted to the host's cellular structure and physiology [4]. In one example, respiratory viruses such as Influenza A have evolved a highly specialized glycoprotein called hemagglutinin (HA) that specifically bind $\alpha 2,6$ -linked sialic acid molecules, which are prevalent on epithelial cells in the upper respiratory tract of humans [11]. The fusion of the HA protein to these surface receptors triggers the formation of an endosomal vesicle around the virion and pulls it inside the cell [11,12]. In another example, plant-infecting viruses, such as the ToBRFV, must by-pass both an outer membrane and a thick protective cell wall [4]. To circumvent these obstacles, the ToBRFV spreads via mechanical contact, for example, insects can damage leaves which creates a cellular opening for virion entry [4,13,14]. Once viruses have entered a host cell, virion unpackaging ensue wherein the capsid is either fully or partially degraded and the viral genome is released [4,10]. Then, expression of early viral mRNA transcripts begins, which typically initiates genome replication and increases the number of genomic templates available for transcription [4]. Viruses either encode their own DNA and/or RNA polymerase or hijack the host's cellular machinery for genome and mRNA synthesis [4].

During the late stages of a virus life cycle, late mRNA transcripts are synthesized. Late transcripts often encode or initiate the transcription of capsid subunits and other virion associated proteins [4]. Capsid structures have a wide range of morphologies and complexities [10]. Some capsids are very simple. For example, the ToBRFV encodes a single capsid protein. Many copies of this

protein are stacked together forming a helical capsid structure [10]. In other cases, the capsid is much more complex [15,16]. For example, phycodnaviruses encode several capsid and scaffolding proteins which form an intricate icosahedral structure [15,16]. Once virions are packaged, they must exit to infect neighbouring cells or a new host. In some cases, such as with IAV and RSV, viruses bud out of the host's cell and become encapsulated in a host-derived lipid membrane [17,18]. Virions then build up in mucus and are transmitted via respiratory droplets when coughing or sneezing occurs [18]. In another example, viruses exit by lysing the host cell, a strategy often used by viruses that infect single-cell organisms [4,19]. In all virus life cycle steps the mechanisms used are diverse and adapted to the host cellular morphology.

1.4. Host Hijacking—Mechanisms to Increase Viral Propagation

Because viruses are reliant on their host for propagation, they have consequently evolved mechanisms to hijack their hosts cellular processes for their own gain. Changes to hosts cellular processes can include, but are not limited to, hijacking nucleic acid replication processes [6,20], ribosomes [21,22], metabolic processes [23–25], and gene expression [26]. These changes facilitate virus resource acquisition, genome replication, mRNA expression, and protein synthesis, all while subverting host defense mechanisms.

The reliance of a virus on its host's replication machinery is largely dependent on the genome nucleic acid type and mRNA synthesis mechanisms, as discussed previously [6]. Although some viruses encode their own polymerases, they do not encode all the components required for replication and rely on certain host factors. For instance, all viruses depend on host nucleoside triphosphates (NTPs) as an energy source to catalyze cellular reactions and for nucleic acid synthesis [20]. Some viruses can increase the availability of NTPs, enabling them to rapidly generate large numbers of mRNA and genomic templates [20,23,25,27]. Moreover, viruses may alter the host RNA polymerase activity to preferentially translate viral mRNAs and/or induce host

mRNA degradation to prevent competition for ribosomes and other resources [21,26,28,29]. Viruses are also dependent on their hosts ribosomal RNA (rRNA) complex as they do not encode their own rRNA [21,22,30,31]. This subversion of host mRNA and protein synthesis can also work to downregulate the expression of host antiviral genes, achieving virus protection and proliferation simultaneously [26,29]. Lastly, viruses can optimize the cellular environment to enhance their replication and propagation by altering their hosts metabolism with viral auxiliary metabolic genes (AMGs) [24,32]. AMGs can be used to influence carbon metabolism, photosynthesis, nutrient acquisition, and the synthesis of nucleotides, amino acids, and lipids [23,33].

1.5. Impact of Viruses on Aquatic Ecosystems

Not only can viruses have huge impacts on their hosts cellular functions and metabolism, but they can also have massive impacts on ecosystems and their host populations. Viruses can play fundamental roles in shaping ecosystems by influencing microbial population and diversity [34,35], driving evolutionary processes [36,37], and regulating biogeochemical cycles [1,38]. In marine environments viral lysis of microbes facilitates the release of organic matter and nutrients back into the environment and preventing the loss of carbon to the oceanic sink [1,39]. This important biogeochemical process regulates global carbon and nutrient cycling and is known as the 'viral shunt' [1,39,40]. In both marine and freshwater ecosystems, microbial lysis redirects organic carbon from higher trophic levels to dissolved organic matter making it available to other microbes, which supports population growth and maintains food web dynamics [1,39–42]. Moreover, viruses contribute to ecosystem stability by controlling harmful freshwater and marine algal blooms that when left unchecked cause hypoxic conditions and produce toxins [35,43–45]. Indeed, by regulating phytoplankton populations, algal viruses can effectively stabilize food webs [41,42], alter community composition [34,35], and mediate the availability of nutrients in aquatic environments [34,40–42].

In freshwater systems worldwide, one of the more prevalent and abundant alga species, in terms of biomass, is the haptophyte *Chrysochromulina parva* [46–49]. This alga plays a substantial role in primary production and can form blooms [47,50,51]. However, it has been understudied and very little is known about its virus-host dynamics [46,47]. Based on our knowledge of virus-host dynamics, viral lysis of *C. parva* should not only control algal biomass but also release dissolved organic carbon and cellular nutrients back into the water column, which has potential implications for microbial dynamics, as well as nutrient and biogeochemical cycling. In 2015, *C. parva* lytic agents were isolated for the first time and were characterized using molecular techniques [47]. One of these *C. parva* lytic agents was the *Chrysochromulina parva* virus BQ1 (CpV-BQ1), a dsDNA virus containing a DNA polymerase B (*polB*) gene with homology to viruses in the *Phycodnaviridae* family [47].

Chapter 4 in this thesis presents the complete genomic assembly and characterization of CpV-BQ1, providing novel insights into the gene content, replication strategies, and potential metabolic impacts of freshwater algal viruses, contributing to a broader understanding of this important alga species and freshwater viral ecology. Furthermore, to recover and characterise this genome a comprehensive approach for both genome assembly and annotation was implemented. This approach used hybrid short- and long-read sequencing, four assembly tools whose outputs were piped into the robust genome consensus tool TryCycler, and a total of 24 gene annotations tools. This detail orientated approach enabled the complete recovery of the CpV-BQ1 genome and sets this work apart from most other studies focusing on algae-infecting viruses. Many studies that characterise cultured algal viruses only use a single genome assembly tool, and thus are unable to properly close genome ends, and underutilize gene prediction tools [52–57]. The most common molecular method to study algal viruses is based on the PCR amplification of specific genes such as the major capsid protein (MCP) and *polB* gene. However, this approach overlooks viruses with divergent sequences limiting the genetic information that can be obtained [52]. Still in other work,

metagenomic shotgun approaches have been used to assemble and characterize novel algal-virus genomes in environmental samples [52]. Although promising, this approach requires significant refinement because it often results in a significant number of partial, chimeric, and poorly characterized algal virus genomes [52,58–62]. With this work we not only provide a detailed analysis of the CpV-BQ1 genome and contribute to our understanding of algal-viruses, but we also established a methodology that we hope will become a gold standard for algal-virus genome assembly and annotation of unknown isolates.

1.6. Impacts of Viruses on Humans—from Pandemics to Food Instability

Beyond their ecological significance, viruses that infect humans, plants, and animals, have profound consequences on public health systems, food security, economic stability, and resource availability [63,64]. Viral pandemics, such as the 1918 influenza pandemic, HIV/AIDS, and the SARS-CoV-2 pandemic, demonstrate how zoonotic spill-over events can alter society by increasing mortality rates, disrupting healthcare systems, destabilizing economies, and changing social behaviours [17,63–66]. These outbreak events have spurred the creation and development of public health infrastructure, including vaccination programs, antiviral treatments, and community-based interventions such as social distancing and masking [67,68].

One technological advancement during the COVID-19 pandemic was the rapid global adoption of wastewater-based surveillance (WBS) as a health monitoring tool [69,70]. WBS monitors viral RNA shed in sewage to assess the prevalence of infections within a community, offering a non-invasive, anonymous, and cost-effective method for early detection of outbreaks [69–71]. Molecular approaches such as RT-qPCR and digital droplet PCR were commonly used to quantify the viral load of SARS-CoV-2 in wastewater [71], whilst genomic sequencing protocols, including tiled-amplicon strategies developed by the ARTIC Network, were used to track and identify SARS-CoV-2 variants [72,73]. During the COVID-19 pandemic, WBS was applied in >70 countries to

monitor the prevalence and/or variants of SARS-CoV-2 [70]. This approach not only provided real-time data on viral circulation but also informed targeted public health responses to communities. The rapid implementation and success of WBS illustrates its scalability and cost-effectiveness, as well as its potential to become a long-term disease surveillance tool for both seasonal and emerging viruses, enhancing pandemic preparedness [71,74–76].

Since the COVID-19 pandemic, the utility of WBS has extended beyond SARS-CoV-2 to include other seasonal pathogens [71,74,75]. Respiratory viruses such as the Influenza A virus (IAV) and respiratory syncytial virus (RSV) have been detected and monitored in wastewater using both molecular [77–79] and sequencing-based techniques [79–81]. Both viruses cause seasonal epidemics and place a significant burden on individuals and healthcare systems annually [82–88]. IAV, particularly the subtype H3N2 is associated with high hospitalization and mortality rates amongst the elderly and other at-risk groups [89–91], whilst RSV poses a severe risk to infants, immunocompromised individuals, and the elderly [83,92–94]. The development of tiled-amplicon sequencing assays for these viruses, presented in Chapter 2 of this thesis, demonstrate the presence and lineage diversity of IAV H3N2 and RSV A in municipal wastewater systems. Thus, these assays offer valuable resources to public health services as a complementary method to clinical testing for respiratory disease outbreaks.

Beyond direct human infections, WBS and environmental surveillance can also be used to detect agriculturally significant plant viruses. The tomato brown rugose fruit virus (ToBRFV) has rapidly emerged as a major threat to global tomato production, with the potential for up to 100% crop loss due to its high transmissibility [95–97]. Thus far, the ToBRFV has been reported in 35 countries, several of which are heavily dependent on tomato exports, such as Mexico, Spain, and Italy. Outbreaks of ToBRFV have now disrupted supply chains and caused significant financial losses for farmers and distributors [95,96,98,99]. Tomatoes are the most globally important vegetable crop and the ToBRFV has had a direct impact on tomato prices and availability, affecting both

consumers and agricultural workers reliant on tomato farming [95,98]. The control of ToBRFV requires labor-intensive mitigation measures, including stringent sanitation protocols, seed testing, and the destruction of crops [96]. These measures increase production costs, place a high burden on farmers, cause economic losses, and affect market stability [95,98,99]. Chapter 3 of this thesis expands the utility of WBS by developing a novel tiled-amplicon sequencing assay for ToBRFV detection in wastewater, illustrating how wastewater systems can be used to monitor plant pathogen transmission and determine the regional distribution of ToBRFV variants.

1.7. Genome Sequencing as a Tool to Study Viruses and their Impacts

As demonstrated by the global deployment of wastewater-based surveillance (WBS) for SARS-CoV-2 and its recent expansion to other pathogens, WBS sequencing methodologies are an essential tool for the detection and monitoring of viral variants and emerging pathogens [71,74]. Moreover, genome sequencing is an essential tool for understanding viral diversity, evolution, disease, and transmission in both clinical samples and the environment [2,3,63,100–102]. It can provide critical insights into the functional potential encoded within the viral genome, inform phylogenetic relationships, and can reveal genomic features linked to life-cycle, host-interactions, and pathogenicity [101,103–106]. Sequencing plays an important role in both novel virus discovery, as well as the genomic surveillance and study of known pathogens [71,77,103,105,107–110]. However, scientists have only sequenced and studied the genomes of a small percentage of viruses in the world [102,106,111]. Thus, there remains a significant amount of work to uncover and understand the vast viral diversity, as well as the impacts these viruses have on their hosts and ecosystems [106,111,112]. In this thesis, sequencing was applied to complete two main objectives: first, to develop targeted sequencing assays for respiratory (Chapter 2) and plant viruses (Chapter 3) in wastewater, and second, to assemble and characterize the genome of the freshwater algal virus *Chrysochromulina parva* virus BQ1 (CpV-BQ1) (Chapter 4).

Two primary approaches employed to sequence viral genomes include shotgun sequencing and targeted sequencing assays. Shotgun sequencing can capture all nucleic acids present in a sample without enriching for or targeting genome sequences from any one species or group of organisms [101,106,113–115]. This is particularly useful when investigating a mixed community of viruses and/or organisms; i.e., metagenomic samples, and can provide information about viral diversity, as well as detect co-infections, quasispecies, and associated host-organisms [101,111,116]. Additionally, shotgun sequencing is highly useful when sequencing novel viruses wherein no prior sequencing information is available, such as CpV-BQ1, as sequence specific targeted approaches cannot be developed or applied [101,113].

However, metagenomic shotgun sequencing also has disadvantages which make it unsuitable for some applications. For example, in environmental samples, such as wastewater, viral loads are typically low while host and/or microbial nucleic acids dominate [101,117–119]. As such, metagenomic approaches necessitate extensive sequencing depth to achieve enough viral reads for subsequent analysis [101,117–119]. Thus, metagenomic shotgun sequencing often lacks the sensitivity and cost-effectiveness required for routine surveillance of viruses in wastewater samples [101,117–119]. To address these short-comings various targeted sequencing approaches, such as bait-capture and PCR amplicon sequencing, have been developed [101,119].

Among targeted approaches, tiled-amplicon sequencing has gained widespread adoption as it can enrich near-complete target viral genomes from samples in an efficient, sensitive, and cost-effective manner, making it highly suitable for wastewater surveillance programs [72–74,101,119,120]. Tiled-amplicon sequencing uses overlapping primers designed from a consensus genome to generate a series of amplicons that cover the target genome [120,121]. Following PCR amplification, the products are sequenced using sequencing platforms, such as Illumina or Oxford Nanopore Technologies, providing high-resolution data suitable for genomic

reconstruction and analysis [72,73]. This technique gained significant popularity during the COVID-19 pandemic as it was applied on a global scale to monitor SARS-CoV-2 variants in wastewater [71,74,122]. Since, a myriad of tiled-amplicon sequencing assays have been developed to target many different virus genomes and have been applied to various sample types [108,123–138]. While highly effective, tiled-amplicon sequencing requires prior knowledge of genome information for primer design, making it unsuitable for discovery of novel viruses [101,120]. Additionally, amplicon dropout can occur if multiple mutations accumulate within the primer binding site, requiring periodic updating of primer schemes [101,120]. Despite these limitations, the high-sensitivity and cost-effectiveness of tiled-amplicon sequencing makes it a powerful tool for viral genomics and environmental surveillance.

1.8. Research Focuses and Objectives

This thesis integrates and applies genome sequencing methodologies to better understand viral diversity and ecology, as well as improve and develop upon currently available public health resources. The overarching objective of this work is to improve and develop methods and workflows for detecting, characterizing, and monitoring viruses in the environment, including algae-, plant-, and human-infecting viruses, using both metagenomic and targeted sequencing methods. Through three focused research projects, this work highlights how genomic sequencing can be used to answer fundamental and applied questions in public health surveillance, plant pathology, and viral ecology.

The first objective (Chapter 2) was to develop and apply targeted tiled-amplicon sequencing assays to monitor two human respiratory viruses, Influenza A virus (IAV) H3N2 and respiratory syncytial virus (RSV) A. Both pathogens contribute significantly to seasonal illness and place an undue burden on healthcare systems [82–84,88]. To monitor these viruses, novel tiled-amplicon sequencing assays targeting nearly complete genomes were developed and validated on

wastewater samples. Then, these assays were applied to wastewater samples collected over a three-month period during the peak of the respiratory virus season. This work demonstrated the successful detection and variant calling of IAV H3N2 and RSV A from wastewater, enabling the identification of circulating lineages and providing critical information for public health decision-making. These assays offer a scalable, cost-effective method to complement clinical surveillance systems, especially during periods of heightened respiratory virus activity.

The second objective (Chapter 3) was to extend this surveillance method to monitor the agriculturally significant *Tomato brown rugose fruit virus* (ToBRFV) in municipal wastewater. ToBRFV is a tobamovirus responsible for severe disease outbreaks in tomato crops worldwide [95–97]. A novel multiplexed tiled-amplicon assay was designed to cover the entire ToBRFV genome, enabling variant detection and clade distribution analysis in wastewater influent samples. This study demonstrated that wastewater systems act as environmental reservoirs for plant viruses and established a proof-of-concept for tracking agricultural pathogens through municipal wastewater networks.

The third objective (Chapter 4) was to characterize the genome of a novel freshwater algal virus, *Chrysochromulina parva* virus BQ1 (CpV-BQ1). This virus was isolated from a Lake Ontario water sample [47] and infects the freshwater haptophyte *Chrysochromulina parva*, an ecologically important algal species. A hybrid metagenomic sequencing strategy was employed using long-read Nanopore and short-read Illumina sequencing platforms to assemble the complete CpV-BQ1 genome. This hybrid approach effectively resolved repetitive and complex genomic regions while maintaining high base-level accuracy. Subsequent genomic annotation uncovered the functional potential encoded within the CpV-BQ1 genome which revealed key viral genes involved in replication, transcription, nucleotide metabolism, capsid formation, and host manipulation. This analysis provided important insights into the taxonomic placement of CpV-BQ1 within the *Phycodnaviridae* family and expanded our understanding of viruses that infect and impact

freshwater phytoplankton. This work now provides a robust model-system which can be used to study the complexities of virus-host interactions of the ecologically significant freshwater alga species *C. parva*.

Together, these research objectives use techniques to sequence and characterise the genomes of viruses in freshwater to address critical knowledge gaps in environmental virology, plant pathogen surveillance, and public health genomics (Figure 1.1). By applying advanced sequencing technologies to freshwater environments, this thesis underscores the interconnectedness of viral ecology and human society through a “One Health” perspective. One Health is an integrated approach that looks to optimize the health of humans, animals, and the environment through the recognition their interconnectedness. Through these three projects, we connect the impacts of viruses to humans, plants, the ecosystems. In chapter two, the direct impact of human-infecting viruses Influenza A H3N2 and RSV A is addressed using a wastewater-based surveillance approach to understand the prevalence and spread of these viruses within municipalities. Chapter three highlights the tomato brown rugose fruit virus and its impact on the greenhouse sector and food supply chains, the genome sequencing tools developed in this work provide a useful resource to track and monitor this viral threat. Chapter four addresses the indirect impact viruses have on humans by exploring a novel algal-infecting virus. Algae are important players in our ecosystems, which impact food webs, biogeochemical cycling, and our atmosphere. By studying the viruses that infect and control algal populations we can better understand ecosystem dynamics and how changes to these systems may impact humans. Furthermore, this work highlights the strengths and limitations of metagenomic and tiled-amplicon sequencing approaches in complex environmental matrices and establishes workflows adaptable for the detection and characterization of both novel and known viral pathogens.

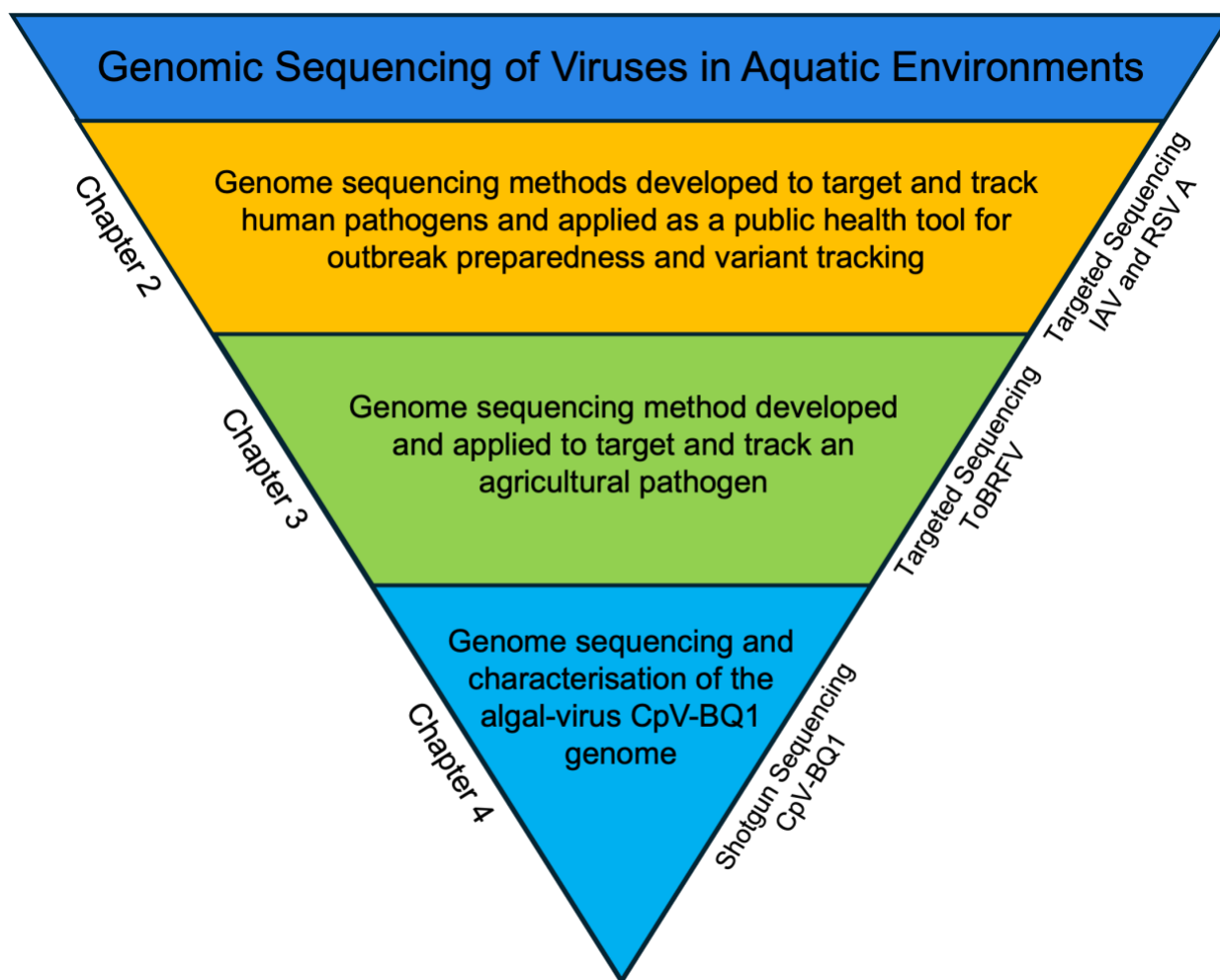


Figure 1.1. Flowchart of Thesis Objectives. The overarching theme of this thesis centers around genomic sequencing of viruses in freshwater. The development and/or application of these strategies were applied in three different objects across chapters 2 to 4 of this thesis, starting with the direct and broad impact of human infecting viruses, followed by the in-direct and more narrow impacts plant- and algal-infecting viruses have on humans.

Chapter 2. A Genome Sequence Variant Monitoring Program for Seasonal Influenza A virus H3N2 and Respiratory Syncytial Virus A using Wastewater-Based Surveillance in Ontario, Canada

2.1 Introduction – Influenza A H3N2 and Respiratory Syncytial Virus A burden, infections, and detection

Seasonal respiratory viruses can be highly contagious and typically circulate in the Northern hemisphere from October to April, and in the Southern hemisphere from February to September [139,140]. Two seasonal respiratory viruses which cause a large number of yearly infections and substantial health care burden are the Influenza A virus (IAV) and Respiratory Syncytial Virus (RSV) [91,92,141,142]. These viruses can lead to serious illness and death particularly in vulnerable populations, including children, elderly adults, immunocompromised individuals, and those with comorbidities [92,94]. Annually, there are an estimated one billion Influenza infections globally, resulting in approximately 290,000 and 650,000 deaths with the vast majority of these in children under five years old [143]. RSV is responsible for an estimated 3.6 million hospitalizations and approximately 100,000 deaths in children under five years old, annually, moreover in 2019 there were an estimated 33 million global RSV-associated acute lower respiratory infection episodes in children under five [144,145]. Broader health consequences following an infection can include chronic illness, increased susceptibility to infections, cardiovascular events, and poor pregnancy outcomes, all of which have negative impacts on quality of life and longevity [83,84,89,90,93]. Infections can also cause financial strains for individuals who are unable to work due to illness or the need to care for family members [85,86]. Additionally, increased hospitalizations and healthcare encounters due to seasonal illness can place a large burden on healthcare and economic systems [82,83,87,88]. In the USA, the healthcare costs attributed to Influenza A and B infections over eight seasons is estimated to be between \$2.0-\$5.8 billion [88],

while estimates of the direct medical cost of RSV infections in the USA range from \$1.52-\$2.99 billion annually for adults over 60 years of age [146]. Routine monitoring along with public awareness campaigns and vaccination of both IAV and RSV is crucial to preventing the spread of these viruses and reducing the burden they impose.

The most common human-infecting IAV subtypes are H3N2 and H1N1(pdm09), with the subtype causing the most infections changing from year-to-year. H3N2-dominated seasons have higher mortality and hospitalization rates in older adults as compared to H1N1(pdm09) or Influenza type B [147]. Additionally, H3N2 is known to have a lower vaccine effectiveness and higher immune evasion [147]. In the 2022-2023 winter season, the subtype H3N2 was found to be the most prevalent IAV subtype in Canada, making up 89% of IAV subtyped laboratory tests [148], whilst in the following 2023-2024 season, the H1N1(pdm09) subtype was the most prevalent IAV in Canada, making up 73% of IAV laboratory tests with subtype information [149]. The IAV H3N2 subtype can be further divided into 35 clades or lineages which are tracked and visualized with NextStrain (<https://nextstrain.org/seasonal-flu/h3n2/ha/12y>).

Two RSV subtypes, A and B, infect humans and co-circulate during infection seasons [150,151]. There are conflicting reports of which subtype causes more severe illness and outcome severity. However, evidence suggests RSV A infections are more prevalent and severe in preterm infants and children with comorbidities [150,151]. Definitions at the subgroup level of RSV A and B have been difficult to discern and have been debated within the scientific community [152–154]. However, recent work by the RSV Genotyping Consensus Consortium has defined 24 distinct lineages for RSV A [154] which can be visualized with NextStrain (<https://nextstrain.org/rsv/a/genome>).

Understanding which viral lineages are circulating is important for understanding the disease severity caused by subtypes/lineages, tracking the spatial and temporal spread of viruses, and

the effectiveness of vaccines and other treatments [74,155]. Traditionally, respiratory virus outbreaks, including those caused by IAV and RSV, have been tracked by testing clinical samples taken from individuals seeking medical treatment. Clinical samples are first tested for viruses using point-of-care diagnostics or reverse-transcription PCR (RT-PCR) analysis, then a subset of positive samples are sequenced to determine the specific subtype or lineage present [148,156–160]. However, since the onset of the SARS-CoV-2 pandemic in 2020, wastewater-based surveillance (WBS) has been used alongside clinical testing to track the spread of respiratory viruses in communities [71,74,77,78,80]. WBS has provided an early warning system for SARS-CoV-2 and other viral outbreaks and has been used to determine which viral subtypes/lineages are circulating in specific areas [69,71,76]. WBS offers an attractive additional tool for use alongside clinical testing because it is cost-effective, does not require individuals to seek medical care or report symptoms, and can capture data from asymptomatic or unreported infections in an unbiased and anonymous manner [75,76,161]. Moreover, WBS can inform preventative public health measures which can reduce the health and economic burden imposed by seasonal respiratory viruses [71,75,76,161,162].

WBS relies on several molecular biology techniques to fully understand the spread and load of viruses in wastewater (WW), including various qPCR and sequencing techniques [71,74,76]. Reverse transcription quantitative PCR (RT-qPCR) methods have been highly successful in tracking the viral load of SARS-CoV-2 and other viruses in wastewater and can determine levels of infection and illness circulating in communities in real time [76,77,162,163]. RT-qPCR can also be used to detect and quantify specific viral variants or subtypes in wastewater if they have distinct mutations that distinguish them from other variants/subtypes. This requires the design and testing of specific primers/probe combinations which target differentiating mutations between subtypes [164–166]. Thus, subtype monitoring requires a suite of primer/probe combinations to capture subtype diversity and also demands the frequent development and testing of new primer/probe

combinations to detect emerging lineages. Furthermore, virus subtypes such as Influenza A virus H3N2 and RSV A can contain many distinct clades which are rarely discernable using an RT-qPCR approach [80,81].

Sequencing viruses in wastewater offers an attractive alternative to RT-qPCR; viral genomes can be readily amplified using a set of PCR primers inclusive to all lineages of a single virus or virus subtype, with the flexibility to modify the primers when multiple mutations arise in primer binding sites that reduce binding efficiency [72,73,120]. Moreover, sequencing can capture the complete genome sequence, which provides more genomic information and insight into the evolution of virus genomes [74,79,167]. A sequencing approach, dubbed tiled-amplicon sequencing or the ARTIC method, has become a popular PCR-based method to specifically enrich the entire genome of a target virus, and has been widely applied in WBS to detect and characterize variants of SARS-CoV-2 [72,73,75,76,120].

Since 2020, WBS using a combination of RT-qPCR analysis and targeted sequencing with the ARTIC method has been a powerful public health tool to monitor the viral load and lineages of SARS-CoV-2 in communities. WBS is now expanding to detect other viruses including IAV and RSV [76–78,168–171]. In the past few years, many studies have been published which monitor the viral load of IAV and RSV in wastewater using RT-qPCR and digital droplet PCR, demonstrating a widespread interest in monitoring these pathogens in wastewater [77,78,163,165,166,168,170–186]. However, relatively few studies have used genomic sequencing to monitor and evaluate the clade-level variation and genomic evolution of these viruses in wastewater [79–81,167,187–191]. Among the studies that have embarked on genomic sequencing, some have only used a small portion of one genome segment, typically the hemagglutinin (HA) segment, thereby limiting the number of clade specific mutations that can be detected and used for genotyping [81,187,188,191]. Other approaches such as a bait-capture [190] and shotgun RNA-Seq [189] did not produce adequate sequencing depth for clade

determination. A few studies have used whole genome IAV and RSV targeted sequencing on wastewater samples; however, these studies employed a genome consensus based strategy using clade reference genomes to determine which clades are present in wastewater [80,167]. Generating a consensus sequence by mapping reads to a reference genome produces a single representative sequence, which includes the most prevalent nucleotide at each position, thus mutations can be removed from data and therefore missed during variant calling, which is particularly problematic for identifying rare and low-abundance variants in circulation. Additionally, the abundance of lineages in wastewater cannot be estimated with this strategy. Comparatively, deconvolution methods, with tools such as Freyja [192], observe the prevalence of mutations at clade-defining nucleotide positions and use statistical models to estimate the frequency of mixed-lineages in a sample [122,193,194]. To expand the utility of WBS, we developed novel tiled-amplicon sequencing assays and used them to monitor Influenza A H3N2 and RSV A in wastewater. Samples were collected from six wastewater treatment/collection sites in Ontario, Canada, on a weekly or biweekly basis from November 2023 to January 2024, which coincides with the annual seasonal respiratory virus period in Canada. We applied deconvolution methods to clades present in samples using Freyja [192], and our in-house tool, Alvoc (abundance learning for variants of concern). Alvoc was adapted for IAV H3N2 and RSV A clade determination [195] from our previously designed tool, Alcov (abundance learning of SARS-CoV-2 variants), which was developed for estimating mixed SARS-CoV-2 lineages in wastewater samples [196]. Alvoc was modified to generate constellation files from NextStrain phylogenetic trees. These files delineate clade-specific mutations which are used to call and estimate clade abundances. This modified function enables user adaptation of Alvoc for any virus that has a NextStrain build [195].

Existing primer sets and assays have been previously developed for the detection of IAV and RSV. The MB-Tuni primers [197] and the clinical tiled amplicon RSV (CR) primers [198], have been successfully applied to clinical IAV [156,157,159,197] and RSV [198,199] samples, as well

as some agricultural samples [200–202]. However, their performance on RNA extracted from wastewater influent has not been thoroughly evaluated, with only one study using MB-Tuni primers on RNA extracted from wastewater [80]. Given the fundamental differences between clinical and environmental wastewater samples, including viral concentration, genome integrity, and the presence of inhibitory substances, the MB-Tuni and CR assays may not be amenable to environmental wastewater samples. Thus, the tiled-amplicon assays designed in this study were compared to these previously designed primer sets using wastewater samples spiked with IAV H3N2 and RSV A virions.

The IAV H3N2 genome consists of eight discrete RNA fragments, called PB2, PB1, PA, HA, NP, NA, MP, and NS, with each segment encoding 1-2 genes [203]. In contrast, the RSV A genome is a single RNA strand approximately 15225 bp in length that encodes 10 genes and 11 proteins [204]. Our assays employ an ARTIC-style primer scheme, wherein overlapping primers are used to generate ~400 bp amplicons across each genome. Our IAV H3N2 assay uses a total of 42 primers, which cover 93.04% of the influenza A H3N2 RefSeq reference genomes NC_007366.1 – NC_007373.1. The RSV A assay uses a total of 52 primers, which covers 98.44% of the Respiratory syncytial virus A reference genome NC_001803.1.

2.2 Materials and Methods

2.2.1 Sampling Sites and Wastewater Collection

The study includes data from seasonal wastewater samples collected from six collection sites between November 6th, 2023, and January 30th, 2024. Either 24 h composite influent wastewater or upstream grab-samples were collected from multiple sites in the Regional Municipalities of York, Peel, Waterloo, and Hamilton (Supplementary Table 2.1) on a weekly basis, however during a few weeks sample collection was paused at certain sites (Supplementary Table 2.2). An on-site autosampler was used to collect either pre-grit or post-grit separation wastewater influent samples at the GE Booth, Kitchener, Cambridge, and Hamilton wastewater treatment plants. Grab fractions taken over a 24 h period were pooled to create a composite sample, which was mixed thoroughly to resuspend fecal solids. 250 mL of the composite sample was poured into a sterile polypropylene bottle and refrigerated at 4°C. Samples taken from the two York Region collection sites were three pooled grab samples taken five minutes apart and thus represent a 10-minute time period. The samples were placed in an insulated container with multiple ice packs and shipped to our laboratory in Waterloo, ON, Canada.

2.2.2 Wastewater and Water Virus Spike-in Samples to Test Primer Specificity

To test primers for specificity and efficiency, wastewater and water samples were spiked with heat-inactivated isolates of IAV H3N2 and RSV A. Upon receiving the wastewater samples, the bottle exteriors were sterilized with 70% ethanol and the bottle was inverted to resuspend solids. Two 20 mL fractions were taken from four wastewater samples and placed in 50 mL falcon tubes. Additionally, four 20 mL volumes of nuclease-free water were placed in falcon tubes to act as spike-in and wastewater controls. One replicate of the four wastewater and two water control samples were spiked with 25 μ L (1.0425×10^4 TCID₅₀) of Influenza H3N2 heat-inactivated culture fluid isolate Hong Kong/2671/19 (*Zeptomatrix*, #0810609CFHI), while the other replicate of four

wastewater samples and two water samples was spiked with 25 μL ($9.7 \times 10^2 \text{ TCID}_{50}$) of RSV-A heat-inactivated culture fluid 2/2015 isolate #2 (*Zeptomatrix*, #0810474CFHI).

2.2.3 Wastewater Virus Isolation and RNA Extraction

Viruses in wastewater and controls were captured and isolated using Ceres Nanotrap Microbiome A beads (*Ceres Nanosciences*, #44202) and RNA was extracted using the QIAGEN RNeasy® Mini Kit (*Qiagen*, #74116). The following amendments to the manufacturer's protocol for the “Nanotrap® Microbiome A; 10mL Manual Protocol with QIAamp® Viral RNA Mini Kit” were made (https://www.ceresnano.com/_files/ugd/f7710c_360d4f6c81094e7e84e161c0c32748b2.pdf). For every 20 mL fraction of wastewater or water, 200 μL of the Ceres Nanotrap® Enhancement Reagent 2 (*Ceres Nanosciences*, #10112-10) and 300 μL of the Microbiome A Particles was used to concentrate the sample. Following a 10 min incubation period, the Microbiome A beads were pelleted using centrifugation at $8,000 \times g$ for 4 min at 4°C . The pellet was resuspended in a small volume of remaining initial sample instead of nuclease-free water and transferred to a 1.5 mL centrifuge tube for ease of use. The supernatant was removed and 700 μL of RLT lysis buffer (*Qiagen*, #74116) with 1% 2-mercaptoethanol was added to the pellet. The sample was vortexed at maximum speed to release the total virus sized particles into solution and then centrifuged at $8,000 \times g$ for 4 min at room temperature. The supernatant was transferred to a fresh 2.0 mL round bottom tube. The supernatant with the viral analytes was loaded onto an automated QIAGEN QIAcube Connect® (*Qiagen*, #9002864) device using the pre-loaded manufacturer's RNeasy® Mini Kit protocol (*Qiagen*, #74116) without on-column DNase digestion. This virus isolation and RNA extraction method was used for both seasonal wastewater samples and samples spiked with heat-inactivated virus particles.

2.2.4 IAV H3N2 and RSV A Tiled-Amplicon Primer Design

To design IAV H3N2 genome specific primers, Influenza A H3N2 genomes were downloaded from the NCBI influenza database, applying the following filters; Type: A, Host: Human, Country: Any, Segment: 1-8, Subtype: H3N2, Collection date: 2020-01-01 to 2022-07-21. All segments were downloaded as individual files. The following number of sequences were used for each genome segment alignment: PB2 – 399, PB1 – 401, PA – 402, HA – 416, NP – 406, NA – 409, M – 407, and NS – 398. Individual multiple sequence alignments (MSA) were generated for each set of the eight segments using the MAFFT-7.0 algorithm. A consensus 90 or consensus 95 sequence was generated from each MSA. Finally, PrimalScheme was used to generate a 400 bp tiled amplicon scheme (Supplementary Table 2.3) from each consensus sequence. The total length of the generated consensus genome was 13629 bp, of which the primer scheme amplifies 12861 bp (i.e., 94.36% of the genome). Primers were also aligned to the IAV H3N2 RefSeq genome segments, accessions NC_007366.1 – NC_007373.1, and covered 93.04% of this reference sequence. Primers in tiled amplicon schemes typically do not target the ends of genomic fragments. The overlapping primer design and uniform size of amplicons necessitated by these assays means a small part of the genome ends, often 20-100 bp, is not amplified. Moreover, the IAV genome consists of eight discrete RNA fragments and thus less genomic coverage of the IAV H3N2 was obtained as compared to the RSV A genome.

To design RSV A genome specific primers, RSV genomes were downloaded from the NCBI (National Center for Biotechnology Information) Virus database which met the following filtering requirements: Virus: Human respiratory syncytial virus A, taxid:208893, Sequence Length: Min - 15200 Max - 15230. A total of 123 genomes were downloaded along with the RSV A RefSeq sequence, NC_001803.1. A MSA was created using the MAFFT-7.0 algorithm and a consensus 85 sequence 15319 bp in length was generated. Finally, PrimalScheme was used to generate a 400 bp tiled amplicon scheme (Supplementary Table 2.4) from the consensus 85 sequence,

which covers 97.84% of the consensus RSV A genome. Primers were also aligned to the RefSeq genome, accession NC_001803.1, and covered 98.44% of this reference sequence.

2.2.5 Primer Pooling

Initially, primers were pooled at equal concentrations, then were used to PCR amplify cDNA synthesized from RNA that was extracted from a water sample that had been spiked with either Influenza A H3N2 or RSV A heat-inactivated fluid cultures. Ceres bead and spin column extraction was used, as described previously. Following amplicon data analysis, as described below, primer concentrations were adjusted to either increase the depth of reads obtained for amplicons with low coverage or to decrease the depth of reads obtained for amplicons with high coverage. The designated primer pool and concentrations for IAV H3N2 primers (Supplementary Table 2.3) and RSV A primer (Supplementary Table 2.4) are listed. These adjusted primer concentrations were used for all described tiled-amplicon PCR assays.

2.2.6 cDNA Synthesis and PCR Amplification of RNA Extracts from Seasonal WW Samples

RNA extract from seasonal wastewater samples were used to synthesize cDNA with the LunaScript RT SuperMix kit (*New England Biolabs*, #E3010L). Briefly, 20 μ L of RNA elution, 8 μ L of LunaScript RT SuperMix, and 12 μ L of sterile nuclease-free were mixed in a PCR tube. Samples were placed in a thermocycler, then incubated at 25°C for 2 min, 65°C for 15 min, 95°C for 1 min, and then held at 12°C. RNA was converted to cDNA which was subsequently amplified using our tiled-amplicon primers. Separate reactions were prepared for IAV H3N2 and RSV A amplification and each primer pool, for a total of four PCRs per sample. Briefly, 12.5 μ L of Q5 High-Fidelity 2x Master Mix (*New England Biolabs*, #M0492L), 6 μ L of cDNA, 1 μ L of either primer pool 1 or 2, and 5.5 μ L of nuclease-free water were mixed in a 0.2 mL PCR

tube. Tubes were incubated on the thermocycler at 98°C for 30 s, followed by 35 cycles of 95°C for 15 s then 63°C for 5 min, afterwards samples were held at 12°C.

2.2.7 PCR Amplification Methods for Virus Spike-in Samples

RNA extracts from IAV H3N2 and RSV A spike in control samples were prepared using both a one-step method, wherein cDNA and PCR amplification occur in a single tube, and a two-step method, wherein cDNA synthesis and PCR amplification occur in two separate tubes. cDNA used in all two-step amplifications was prepared as described above.

One-step Method:

MB-Tuni (MB) Influenza A H3N2 one-step PCR: 1.25 µL One-step enzyme (*Gene Biosystems*, #P611), 12.5 µL HiScript II 2x Mix (*Gene Biosystems*, #P611), 1 µL of 5 µM pooled MB-Tuni forward and reverse primers [197], 5 µL RNA extract, and 5.25 µL nuclease-free water. Thermocycling conditions: 25°C for 2 min, 50°C for 30 min, 95°C for 2 min, 5 cycles of 94°C for 30 s, 45°C for 30 s, and 68°C for 5 min, 31 cycles of 94°C for 30 s, 57°C for 30 s, and 68°C for 5 min, then held at 12°C [197].

Tiled-amplicon (TA) IAV H3N2 one-step PCR: 1.25 µL One-step enzyme (*Gene Biosystems*, #P611), 12.5 µL HiScript II 2x Mix (*Gene Biosystems*, #P611), 1 µL of H3N2 pool 1 or 2 primers, 5 µL RNA extract, and 5.25 µL nuclease-free water. Thermocycling conditions: 25°C for 2 min, 50°C for 30 s, 95°C for 2 min, 35 cycles of 95°C for 30 s, 63°C for 30 s, then 72°C for 3 min and samples held at 12°C.

Clinical RSV (CR) tiled-amplicon one-step PCR: 1.25 µL One-step enzyme (*Gene Biosystems*, #P611), 12.5 µL HiScript II 2x Mix (*Gene Biosystems*, #P611), either 2.55 µL CR primer pool 1 or 2.8 µL CR primer pool 2, 5 µL RNA extract [198], and topped up to 25 µL with nuclease-free

water. Thermocycling conditions: 25°C for 2 min, 50°C for 30 min, 94°C for 3 min, 35 cycles of 98°C for 10 s, 55°C for 10 s, and 72°C for 3 min, then 72°C for 3 min and held at 12°C [198].

Tiled-amplicon RSV A one-step PCR: 1.25 µL One-step enzyme (*Gene Biosystems*, #P611), 12.5 µL HiScript II 2x Mix (*Gene Biosystems*, #P611), 1 µL of RSV A pool 1 or 2 primers, 5 µL RNA extract, and 5.25 µL nuclease-free water. Thermocycling conditions: 25°C for 2 min, 50°C for 30 min, 94°C for 3 min, 35 cycles of 98°C for 10 s, 55°C for 10 s, and 72°C for 3 min, then 72°C for 3 min and held at 12°C.

Two-step method:

MB-Tuni Influenza A H3N2 two-step PCR: 12.5 µL Q5 2x Master Mix (*New England Biolabs*, #M0492L), 1 µL of 5 µM pooled MB-Tuni forward and reverse primers [197], 6 µL cDNA, and 5.5 µL nuclease-free water. Thermocycling conditions: 98°C for 30 s, 5 cycles of 95°C for 30 s, 45°C for 30 s, and 63°C for 5 min, then 30 cycles of 95°C for 30 s, 57°C for 30 s, 63°C for 5 min, then held at 12°C.

Clinical RSV Tiled-amplicon two-step PCR: 12.5 µL Q5 2x Master Mix (*New England Biolabs*, #M0492L), either 2.55 µL CR primer pool 1 or 2.8 µL CR primer pool 2 [198], 6 µL of cDNA, and topped up to 25 µL with nuclease-free water. Thermocycling conditions: 98°C for 30 s, 35 cycles of 98°C for 30 s, 55°C for 10 s, 63°C for 5 min, then held at 12°C.

The previously described PCRs for seasonal wastewater samples were used for the two-step tiled-amplicon IAV H3N2 and RSV A spike-in samples.

2.2.8 Library Preparation and Sequencing

Sequencing libraries were prepared from PCR amplified products using the Illumina DNA prep kit (*Illumina*, #20060059) using half of the manufacturer's recommended volumes. Briefly, PCR

products were tagged using bead-linked transposomes and then indexed using custom Illumina index primers. Samples were pooled and diluted then sequenced on an NextSeq 2000 system (Illumina) using P1 chemistry and 2 x 300 sequencing.

2.2.9 Sequencing Read Quality Control, Alignment, and Calculation of Genome Coverage

Raw reads were trimmed and cleaned up using fastp v0.23.2 with default parameters [205,206]. Quality controlled reads were aligned to the IAV reference genome EPI_ISL_1563628 or RSV A reference genome EPI_ISL_412866, downloaded from GISAID (February 14th, 2024). bwa mem v0.7.12-r118 was used to index reference genomes and align filtered reads [207]. Samtools v1.13 was used to generate sam and bam files from bwa alignments [208]. Genome coverage was calculated from bam files using samtools v1.13 [208]. All raw reads are available on the SRA database under accession PRJNA1288507.

2.2.10 Variant Calling with Freyja and Alvoc

Variant calling was performed with Freyja [192] and Alvoc. Briefly, synthetic read data was generated using DWGSIM v0.1.15 [209]. To generate synthetic reads, one genome listed from each IAV H3N2 clade in the NextStrain build (<https://nextstrain.org/seasonal-flu/h3n2/ha/6y>) and RSV A clade in the NextStrain build (<https://nextstrain.org/rsv/a/genome/6y>) were downloaded from either Gisaide or NCBI virus (March 3rd, 2025). DWGSIM v0.1.15 was used to generate 25,000 synthetic reads from each genome without introducing mutations or off-target reads [209].

To call variants, raw reads from seasonal wastewater sequencing libraries and fastq files containing synthetic reads underwent pre-processing steps, first adaptors were removed using cutadapt 5.0 [210]. Then bam files were generated using minimap2 2.22 [211] and aligned against the HA reference genome NC_007366.1 or RSV A reference genome NC_001803.1. Bam files and filtered reads were then piped into both the Freyja [192] and Alvoc [195,196] tools set to call

IAV H3N2 or RSV A variants. Alvoc only considers mutations at positions that have a read depth greater than 10 bases. For analysis with Freyja, barcode sets specific to each virus were used (<https://github.com/andersen-lab/Freyja-barcodes>). For analysis with Alvoc, constellations were generated by parsing the NextStrain trees for the HA segment of H3N2 (<https://nextstrain.org/seasonal-flu/h3n2/ha/6y>) and the RSV A genome (<https://nextstrain.org/rsv/a/genome/all-time>), mutations present in at least 90% of the nodes belonging to each clade were included in our constellation files. Heatmaps for each virus were generated by both tools for synthetic read data and samples collected at each of the six sampling sites.

2.2.11 Intraspecific Amplicon Coverage Calculation

Individual fasta files were generated for each amplicon using the primer scheme consensus sequences; the binding sites of primer pairs were used as the start and end site for each amplicon sequence. Filtered reads from seasonal samples were aligned to each amplicon, bwa v0.7.12-r118 was used to index amplicon and align reads [207]. Samtools v1.13 was used to generate sam and bam files from bwa alignments, then used to calculate genome coverage from bam files [208].

2.3 Results and Discussion

2.3.1 Comparison of novel and existing IAV H3N2 and RSV A targeted primers

To assess the robustness and performance of our tiled amplicon assays against the previously designed MB-Tuni and clinical RSV (CR) primers, we tested each primer set on water and wastewater samples spiked with heat-inactivated IAV H3N2 and RSV A culture fluid. The MB-Tuni [197] and CR assays [198] use a one-step RT-PCR method, wherein both reactions occur in a single tube. Conversely, our assay conditions employ a two-step method wherein cDNA synthesis is followed by PCR with extended low-temperature extension times. For a comprehensive comparison, both one-step and two-step methods were performed with all four primer sets on virus spiked samples. Assay performance was evaluated using the percentage of genome coverage achieved in each sample, as more efficient and reliable amplification of the target virus genome results in higher genomic coverage (Figure 2.1).

Using a one-step RT-PCR method, the MB-Tuni primers achieved up to 100% genome coverage with virus-spiked water samples. However, they had reduced assay efficiency with virus-spiked wastewater samples, with genome coverage ranging from 19–83% (Figure 2.1a). In contrast, the TA H3N2 assay achieved up to 95% genome coverage in water and maintained comparatively higher coverage in wastewater samples, ranging from 59–84% (Figure 2.1a). The superior performance of the TA H3N2 primers in wastewater is likely due to their design, which generates shorter ~400 bp amplicons. These smaller targets are less susceptible to RNA degradation and inhibitory substances that are common and abundant in wastewater matrices [212]. In comparison, the MB-Tuni primers bind at genome termini and amplify across the genome segments that range from ~800-2300 bp, making them more prone to failure due to RNA segment degradation and the inhibitory PCR conditions [80].

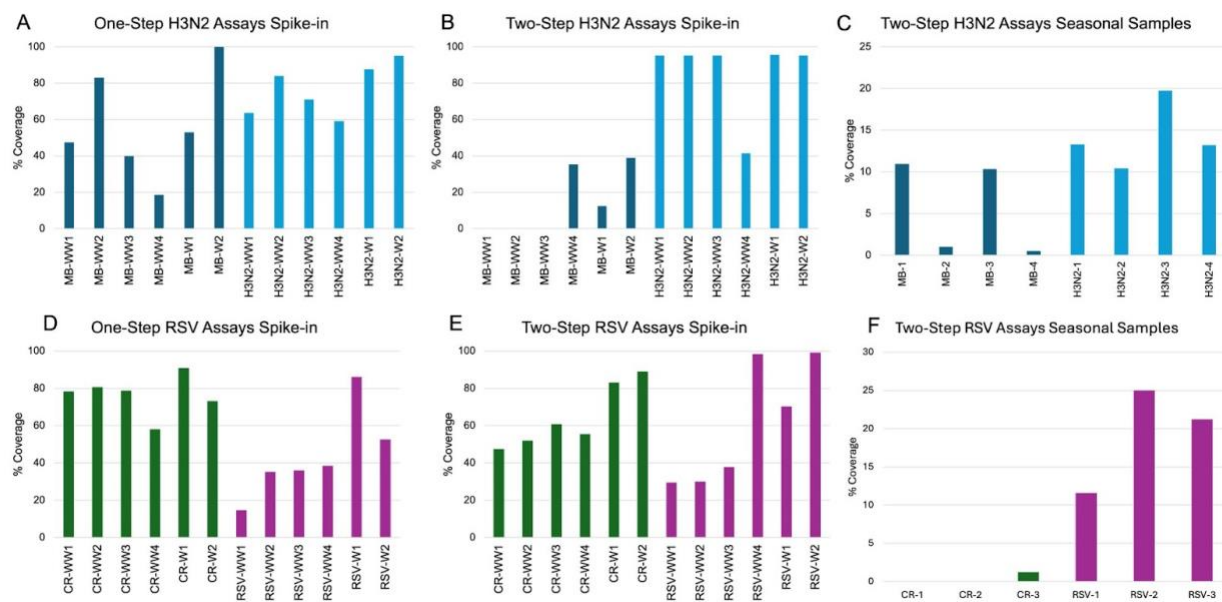


Figure 2.1. Comparison of Influenza A H3N2 and RSV A Amplicon Sequencing Assays on Virus Spiked and Seasonal Wastewater Samples. Comparison of IAV H3N2 primer sets, MB-Tuni (MB) and tiled-amplicon H3N2 (H3N2) (A,B,C). Comparison of RSV primer sets, clinical RSV primers (CR) and tiled-amplicon RSV A (RSV) (D,E,F). One-step (A,D) and two-step (B,E) assay efficiency was compared using four wastewater (WW) and two water (W) virus spike-in samples. Seasonal wastewater samples were used to compare two-step assays targeting IAV H3N2 (C) and RSV A (F).

Further evaluation of these assays across one-step and two-step RT-PCR workflows revealed that the MB-Tuni primers performed better with the one-step method, while the TA H3N2 primers showed improved efficiency with the two-step approach (Figure 2.1a.). This discrepancy may be attributed to differences in polymerase enzymes and thermocycling conditions between the two workflows, which likely affect primer binding and extension efficiency. Notably, while the MB-Tuni one-step assay's conditions closely matched those used in the original publication for MB-Tuni primers [197], the two-step method had a modified extension temperature and time. While this modification is more closely aligned with the ARTIC protocol and the two-step TA H3N2 assay, it potentially compromised the MB-Tuni assay performance. However, regardless of the workflow, the TA H3N2 primers consistently produced higher genome coverage in wastewater samples spiked with IAV virions, than the MB-Tuni primers (Figure 2.1a,b). Additionally, when applied to four unspiked seasonal wastewater samples using the two-step method, the TA H3N2 primers

again demonstrated superior genome amplification efficiency across all four samples compared to the MB-Tuni primers (Figure 2.1c). These findings highlight the TA H3N2 assay's robustness and suitability for amplifying IAV genomes in complex, inhibitor-rich environmental samples like wastewater.

For RSV detection, the clinical RSV (CR) primers demonstrated strong performance in all spike-in samples, regardless of whether a one-step or two-step RT-PCR approach was used and consistently outperformed the TA-RSV A primers under high viral load conditions (Figure 2.1d,e). However, when applied to seasonal wastewater samples without spiked-in virus, the efficiency of the CR assay declined markedly and was surpassed by the TA-RSV A assay (Figure 2.1f). This discrepancy is likely due to the inherently low concentrations and degradation of viral genomes in environmental wastewater samples. Similar to the MB-Tuni primers used for IAV detection, the CR primers are designed to amplify large genomic regions, with some amplicons extending up to 6,000 bp [198]. While effective in controlled spike-in samples and the clinical samples that they were designed for [198], this approach is hindered in real wastewater samples where genomic degradation likely limits the amplification of long fragments [212]. In contrast, the TA-RSV A primers amplify shorter ~400 bp regions, making them less susceptible to the effects of RNA degradation and more reliable for genome recovery in inhibitor-rich environmental wastewater samples. To further validate the assays designed in this study, decreasing concentration of IAV H3N2 and RSV A should be spiked into wastewater to determine the limit of detection.

2.3.2 Assessment of Genomic Coverage and Comparison to Clinical Percent Positive Tests

As infection rates rise within a community, corresponding increases in viral load within wastewater are expected [175–177]. Thus, a robust and efficient PCR assay should yield higher genome coverage and increased sequencing depth as viral load increases in wastewater samples. To assess the performance and epidemiological relevance of our assays, we routinely collected

wastewater from six collection sites (Supplementary Table 2.1) for three months (Supplementary Table 2.2) and used our tiled-amplicon assays on the extracted RNA. We then calculated the percent of genome coverage achieved in each sample and compared these values to the percentage of positive clinical tests reported by the corresponding regional public health units (Figure 2.2). The six sites we monitored are within the catchments of four public health unit regions: the Cambridge and Kitchener treatment plants in the Region of Waterloo public health and emergency services, GE Booth in the Peel public health unit region, York 1 (Leslie Street), and York 2 (Warden Street) in the York Region public health unit , and the Hamilton collection site in the City of Hamilton public health unit region (Supplementary Figure 2.1).

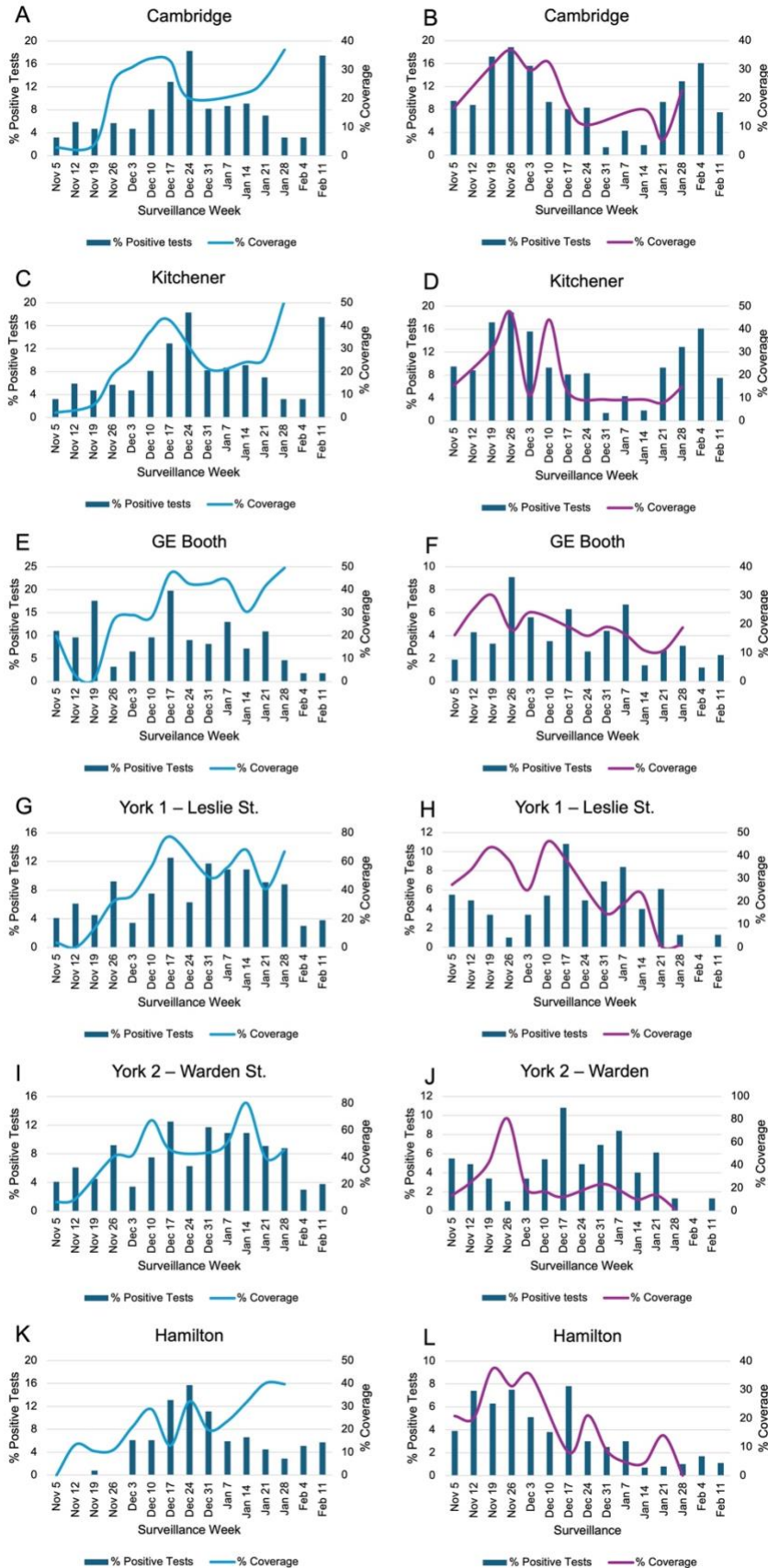


Figure 2.2. Influenza A H3N2 and RSV A Percent of Genomic Coverage Versus Percent of Weekly Positive Tests in Public Health Unit Regions. IAV H3N2 (A,C,E,G,J,K) and RSV A (B,D,F,H,J,L) genomes in wastewater samples from six sites collected on a weekly or bi-weekly basis were PCR amplified with our tiled-amplicon assays. The percent of genomic coverage (solid lines) was determined for each sample and compared to the percent of positive IAV and RSV tests (bars) reported by public health units in each respective area.

Publicly available surveillance data is reported with the Ontario Respiratory Virus Tool (<https://www.publichealthontario.ca/en/Data-and-Analysis/Infectious-Disease/Respiratory-Virus-Tool>) and provides weekly percentages of positive tests and number of cases in each public health region. The percent of positive tests was selected over the absolute number of reported cases, as case counts are directly influenced by testing volume, which can fluctuate weekly. By considering the percentage of positive tests reported in April 2025, we accounted for changes in the level of testing and obtained a more reliable indicator of infection prevalence within each community. This allowed for a meaningful comparison between wastewater-based genome coverage and infection trends in the population serviced by each respective public health unit (Figure 2.2).

Generally, of the six sampling sites, we observed that increases in genome coverage generally corresponded with rising percentages of positive tests within the associated public health regions (Figure 2.2). Notably, peaks in genomic coverage of IAV H3N2 in Kitchener, Cambridge, and York 2, were observed one to three weeks before corresponding increases in clinical positivity rates, suggesting that these assays may provide an early warning of escalating viral transmission within these communities (Figure 2.2a,c,i). For example, when monitoring IAV H3N2 cases in Cambridge and Kitchener, which are both in the Waterloo region, peaks in genomic coverage were observed in the weeks of December 17th, 2023 and January 28th, 2024, one and two weeks before infection levels peaked in this region (Figure 2.2a,c). Additionally, peaks in genomic coverage of RSV A in GE Booth, York 1, York 2, and Hamilton, were observed one to three weeks before peaks in clinical positivity rates (Figure 2.2f,h,j,l). A notable caveat that must be addressed are gaps in clinical reporting data that may affect the comparisons between genomic coverage and clinical positivity rates. In Ontario, clinical testing is only performed on 4 eligible groups, including: symptomatic children under 18 seen in emergency departments, symptomatic

hospitalized patients, symptomatic residents in institutional settings, and the first four symptomatic individuals in an outbreak setting that request testing

(<https://www.publichealthontario.ca/en/Laboratory-Services/Test-Information-Index/Virus-Respiratory>). Thus, clinical testing and positivity rates are biased toward individuals in these groups and do not capture the full scope of infected individuals. Additionally, only a small percentage of IAV and RSV clinical positive tests are subtyped, thus the clinical data includes both IAV subtypes H3N2 and H1N1(pdm09) and RSV A and B subtypes, which could further skew our comparison of clinically positive tests and genomic coverage. Although the current testing data does work as a proxy for the general levels of infection in each public health region, we suspect that if a greater number of individuals were tested and samples were subtyped there may be a better correlation between clinical positivity rates and genomic coverage across all sites and would perhaps alleviate some of the observed discrepancies (Figure 2.2k).

Spearman's rank correlation analysis was used to assess the association between the percent of genomic coverage of virus genomes in wastewater and the percent positivity of clinical tests. No statically significant correlations were observed at any temporal offset examined. Specifically, correlations were non-significant when wastewater genomic coverage was compared to clinical percent positivity with 0-, 1-, 2-, and 3-week lag times (two-tailed $p > 0.01$ for all comparisons). Thus, a consistent monotonic relationship could not be identified in these data sets.

While genome coverage is not a quantitative proxy for viral load or case numbers, these trends across multiple locations demonstrate the utility of our assays for monitoring virus activity in wastewater and supporting public health surveillance. These findings indicate that wastewater-based genomic surveillance can effectively monitor IAV H3N2 and RSV A infections throughout the respiratory virus season. The increases and decreases in genomic coverage mirror patterns of rising and falling clinically reported cases within the respective collection regions. Thus, these

assays offer an accurate real-time assessment of active IAV H3N2 and RSV A cases in municipalities, demonstrating the value and feasibility of monitoring the transmission and prevalence of these seasonal viruses with wastewater-based surveillance. Furthermore, in some instances they have the potential to offer valuable lead-time for public health unit response efforts, however more complete clinical testing of individuals is necessary to properly confirm those trends.

2.3.3 Evaluation of Synthetic Reads with Freyja and Alvoc

To support the identification and tracking of Influenza A H3N2 (IAV H3N2) and respiratory syncytial virus A (RSV A) variants, we developed a lineage deconvolution tool, Alvoc, to estimate abundance of these viruses in wastewater samples. Similar to its predecessor for SARS-CoV-2 variant deconvolution [196], Alvoc uses a normalization approach based on mutation positions characteristic of specific clades and applies a linear regression model via ordinary least squares to estimate variant abundances within a sample [195,196].

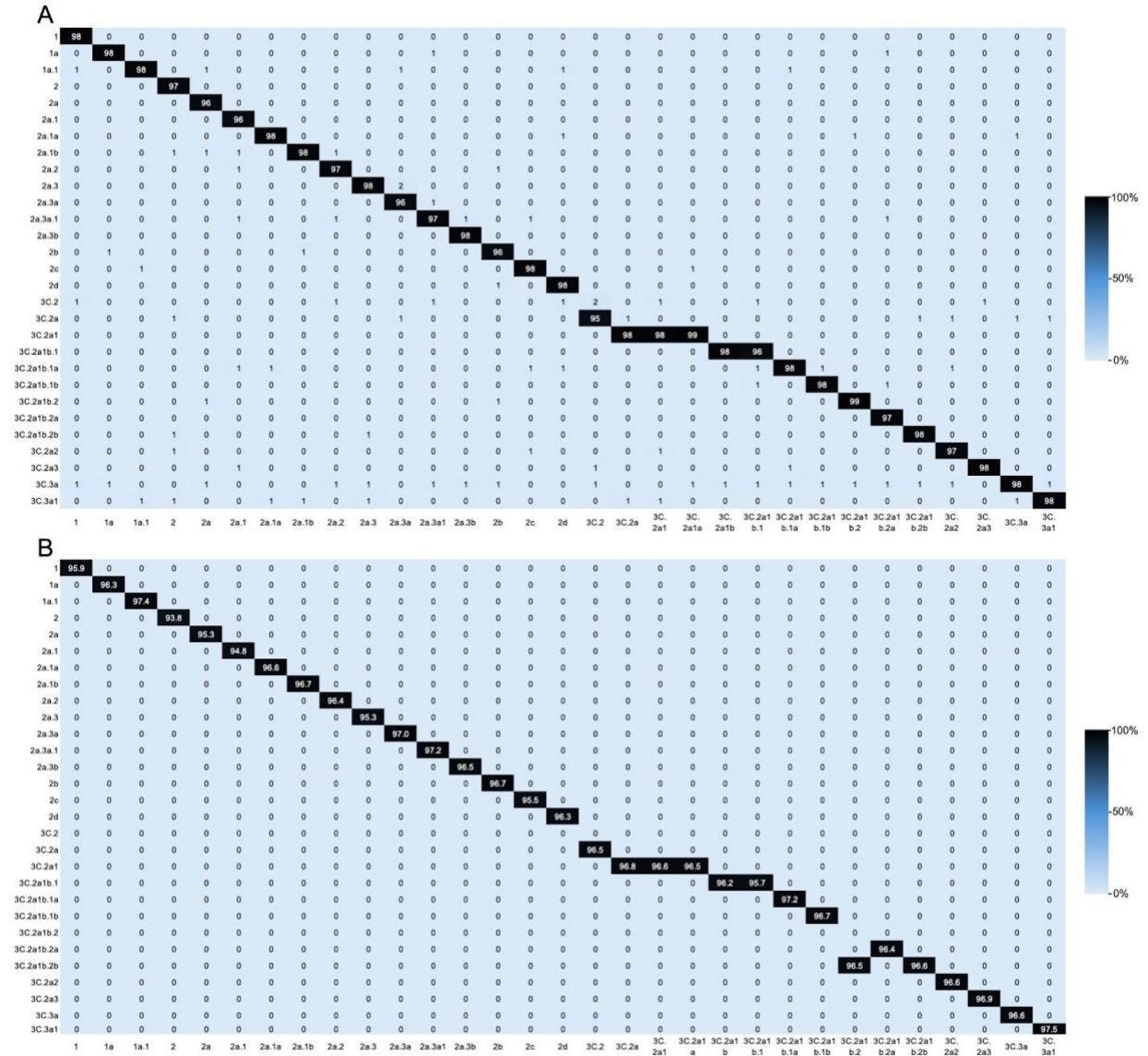
Several other variant-calling tools, such as Freyja, have been developed for deconvoluting mixed viral populations. Freyja similarly relies on lists of clade-defining mutations and their genomic positions but employs a constrained weighted approach using least absolute deviations to estimate variant proportions [192]. Notably, our original SARS-CoV-2 variant deconvolution tool, Alcov, was previously benchmarked against Freyja and seven additional tools, where both Alcov and Freyja demonstrated strong and comparable performance, accurately resolving SARS-CoV-2 variant mixtures in complex samples [194]. Given this, we compared the performance of both Alvoc (i.e., modified Alcov) and Freyja for IAV H3N2 and RSV A detection in wastewater using synthetic read datasets generated from one virus genome per clade.

To evaluate the performance of Alvoc and Freyja for IAV H3N2 and RSV A variant calling, we generated synthetic read data sets from one representative virus genome per clade for each virus

and analyzed these datasets using both tools. Both Alvoc and Freyja used mutations in the hemagglutinin gene (HA) segment to estimate IAV lineage abundances. In these controlled tests on IAV genomes, the two tools showed comparable performance (Figure 2.3). Freyja correctly identified 27 out of 31 clades (Figure 2.3a), while Alvoc identified 26 out of 31 (Figure 2.3b). However, both tools encountered challenges in differentiating certain closely related clades. For example, clade 3C.2 was frequently misclassified as 3C.2a, and subclades such as 3C.2a and 3C.2a1a were called as the closely related clade 3C.2a1. Additionally, Alvoc misclassified clade 3C.2a1b.2 as 3C.2a1b.2b, demonstrating minor difficulties in resolving fine-scale differences between closely related lineages. The sensitivity, defined as the tools ability to correctly call a clade, and specificity, defined as the tools ability to correctly reject a clade, were determined for each tool. Sensitivity was calculated by dividing the number of true positives (correctly called clades, at <2%) divided by the number of false negatives (correct clades that were not called, at >2%) plus true positives. Specificity was calculated by dividing the number of true negatives (incorrect clades are not called, at >2%) by the number of false positives (incorrect clade was called, at <2%) plus the number of true negatives. The sensitivity and specificity of Freyja IAV calls were 87.10% and 99.57%, respectively, and for Alvoc IAV were 83.87% and 99.46%, respectively. Thus, Freyja is slightly better than Alvoc at correctly calling IAV H3N2 clades, while both are able to correctly reject incorrect clades with a similar frequency.

When analyzing RSV A synthetic read datasets, Alvoc outperformed Freyja (Figure 2.4). Alvoc successfully identified all RSV A clades (Figure 2.4b). However, two read data sets were identified as clade mixtures. The A.3.1 read data set was classified as 65.9% A.3.1 and 27% A.D., while A.3.1.1 was classified as 58.7% A.3.1.1 and 32.2% A.D. The misclassifications largely reflected close phylogenetic relationships between these clades that share similar mutations (Figure 2.4b). In contrast, Freyja correctly called only 23 out of 36 RSV A clades, with frequent misclassifications of subclades being grouped under their parent clade (Figure 2.4a). For instance, variants A.D.1.4

through A.D.1.8 were consistently miscalled as the parent clade A.D.1 (Figure 2.4a). This suggests that while both tools can resolve major clades, Alvoc offers greater resolution and specificity when applied to RSV A variant mixtures. The sensitivity and specificity of Freyja RSV A calls were 63.89% and 98.97%, respectively, while for Alvoc RSV A calls were 100% and 99.84%, respectively. Thus, Alvoc was able to correctly identify clades with a much higher frequency than Freyja, while both tools were able to reject incorrect clades with a similar frequency.



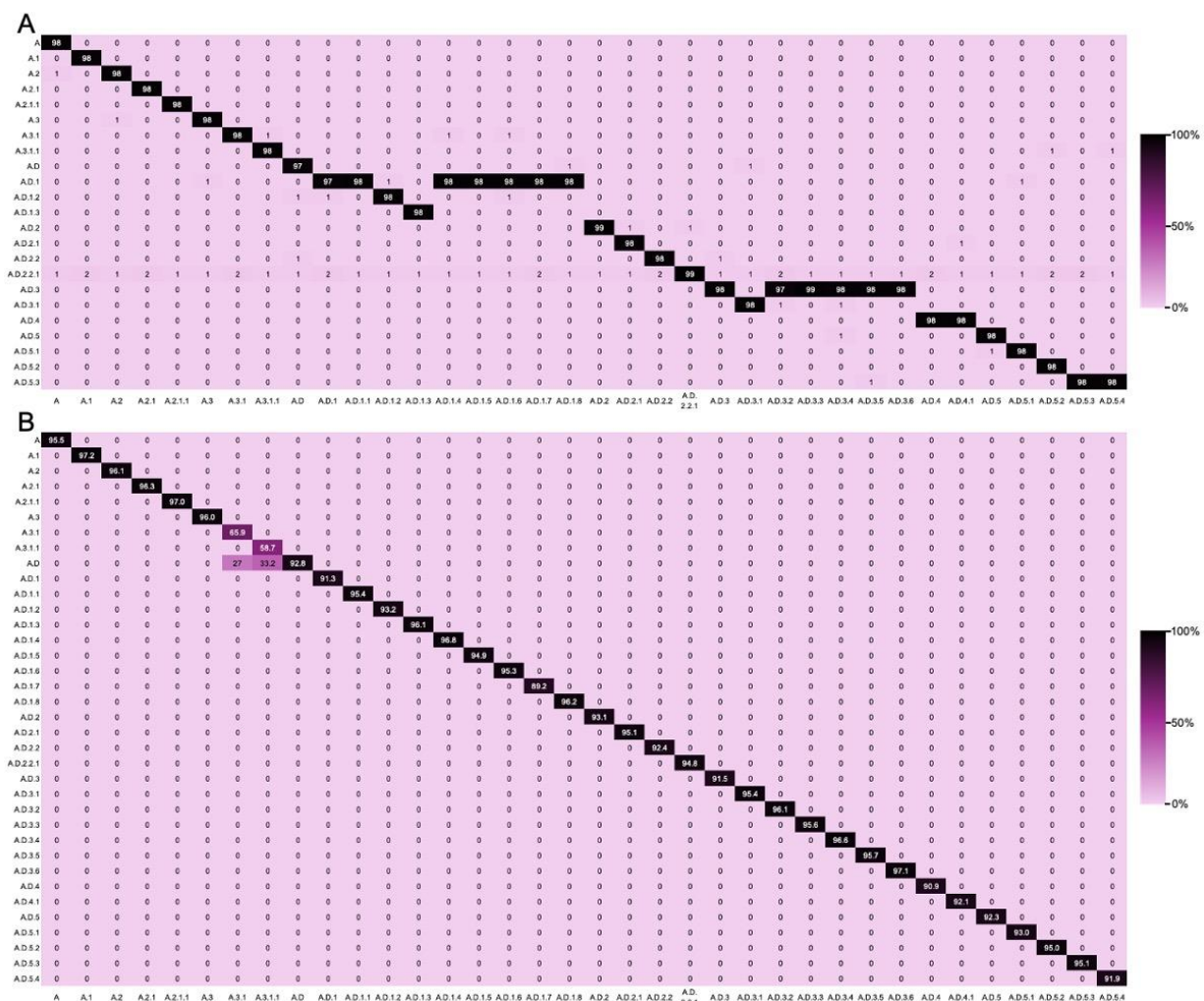


Figure 2.4. Comparison of Freyja and Alvoc RSV A Variant Calling using Synthetic Data. Variant calling on synthetic reads, generated from genomes belonging to RSV A viruses in each clade with DWGSIM, was performed with both Freyja (A) and Alvoc (B). Sample names, i.e. the clade each synthetic read dataset belongs to, is displayed on the horizontal axis, while the clade called for each sample is displayed on the vertical axis. Heatmaps are used to delineate the percentage of reads assigned to each clade. Clades with zero percent calls across all synthetic data sets were excluded from the heatmap.

There are several differences between these tools that likely contribute to the minor discrepancies between variant calls. For one, the statistical method for estimating clade abundances differs, which may result in differing lineage abundance estimates. Most significantly, Freyja and Alvoc use different methods and parameters for generating their constellation files that delineate clade-specific mutations, resulting in differences between clade-defining mutations used by each tool. The mutations used by each tool to define clades plays the largest role in what clades are ultimately called and thus are the most likely source of discrepancies between the variant calling

tool output. On top of utilizing different mutations, Freyja's RSV A constellation files do not define lineages A.D.1.4-A.D.1.8, A.D.3.2-A.D.3.6, or A.D.5.3, which is why these lineages were not correctly called by Freyja (Figure 2.4a). Despite these discrepancies, these results demonstrate that both Alvoc and Freyja are capable of deconvoluting IAV H3N2 variant mixtures in wastewater samples, with Alvoc showing enhanced performance for some closely related RSV A clades. These findings support the feasibility of wastewater-based genomic surveillance for respiratory viruses beyond SARS-CoV-2 and highlight the potential of Alvoc as a valuable tool for public health monitoring of seasonal respiratory virus transmission.

2.3.4 Application of Freyja and Alvoc for the Assessment of IAV H3N2 and RSV A Clades in Seasonal Wastewater Samples

Freyja and Alvoc were applied to sequenced reads from seasonal wastewater samples to estimate H3N2 lineage abundances. Freyja generated lineage calls for more samples than Alvoc, largely due to differences in minimum read depth thresholds. Freyja operates with a default threshold of zero, thus can include low-coverage positions in its abundance estimates, while Alvoc requires a minimum of 10× depth, limiting its ability to analyze samples with sparse coverage. However, this increased detection capacity in Freyja reduced the certainty of some lineage assignments, and often distributed abundance estimates across multiple clades when coverage was low (Figure 2.5).

In samples where both tools produced confident calls, the primary H3N2 lineages detected were 2a, 2a.3a, 2a.3a1, and 3C.2a1b.1b (Figure 2.5). These observations are consistent with global surveillance data. According to the NextStrain H3N2 build, of 143 genomes collected from 42 countries between our sampling period of November 2023 and January 2024, 141 genomes were classified as 2a.3a1, confirming this clade's predominance internationally. Our Ontario wastewater findings similarly showed widespread detection of 2a.3a1 (Figure 2.5). In almost all

samples where 2a.3a1 was called, both Alvoc and Freyja agreed on the clade assignment. In some samples, for example York1 Jan 3rd and York2 Jan 8th, Alvoc called clades 2a, 2a.1, or 2a.3a, instead of the closely related and predominant lineage 2a.3a1. Either, Alvoc had minor difficulties in distinguishing between these closely related clades, or these clades were co-circulating alongside 2a.3a1. The former seems more likely, as 97% of sequenced clinical samples in Canada during this time were assigned to clade 2a.3a1 [213]. When Freyja was used to evaluate these same samples, it was unable to determine the specific clade and instead split the lineage abundance call between 12 possible clades (Figure 2.5). In these samples the HA gene read coverage and/or depth was low, which likely made it difficult for these tools to discern between closely related lineages that share common clade determining mutations. To enhance read recovery, various virion capture and isolation methods, such as PEG precipitation or ultracentrifugation, could be evaluated to determine whether they improve virion isolation, thereby increasing IAV genome concentrations, PCR product yields, and sequencing read counts [168,214,215].

The 2a.3a1 clade (formally 3C.2a1b.2a.2a.3a.1) was initially reported in the United States in 2021, as reported by GISAID, and quickly disseminated worldwide becoming the predominant H3N2 clade in 2023 [216]. This clade carries distinct amino acid changes in the HA and neuraminidase (NA) genes as compared to the 2023/2024 WHO-recommended vaccine strains, which reduced vaccine protection and efficacy during this season [213,217–219]. In Canada, 97% of the sequenced H3N2 infections during the 2023/2024 season belonged to 2a.3a1, and vaccine effectiveness against circulating H3N2 strains was estimated at 40%, indicating moderate protection [213]. In the 2024/25 season, the vaccine against the H3N2 strain was updated using a clade 2a.3a1 strain, which increased the vaccine protection to 54%, and was particularly effective against 2a.3a1 strain infections [220]. The implementation of a genomic surveillance system like this could provide real-time insight into emerging strains, overcoming delays

associated with clinical genotyping of strains. If applied globally, it could help predict which clades are likely to become predominant and better inform vaccine strain selection, improving the vaccines alignment with circulating variants and enhancing their effectiveness [71].

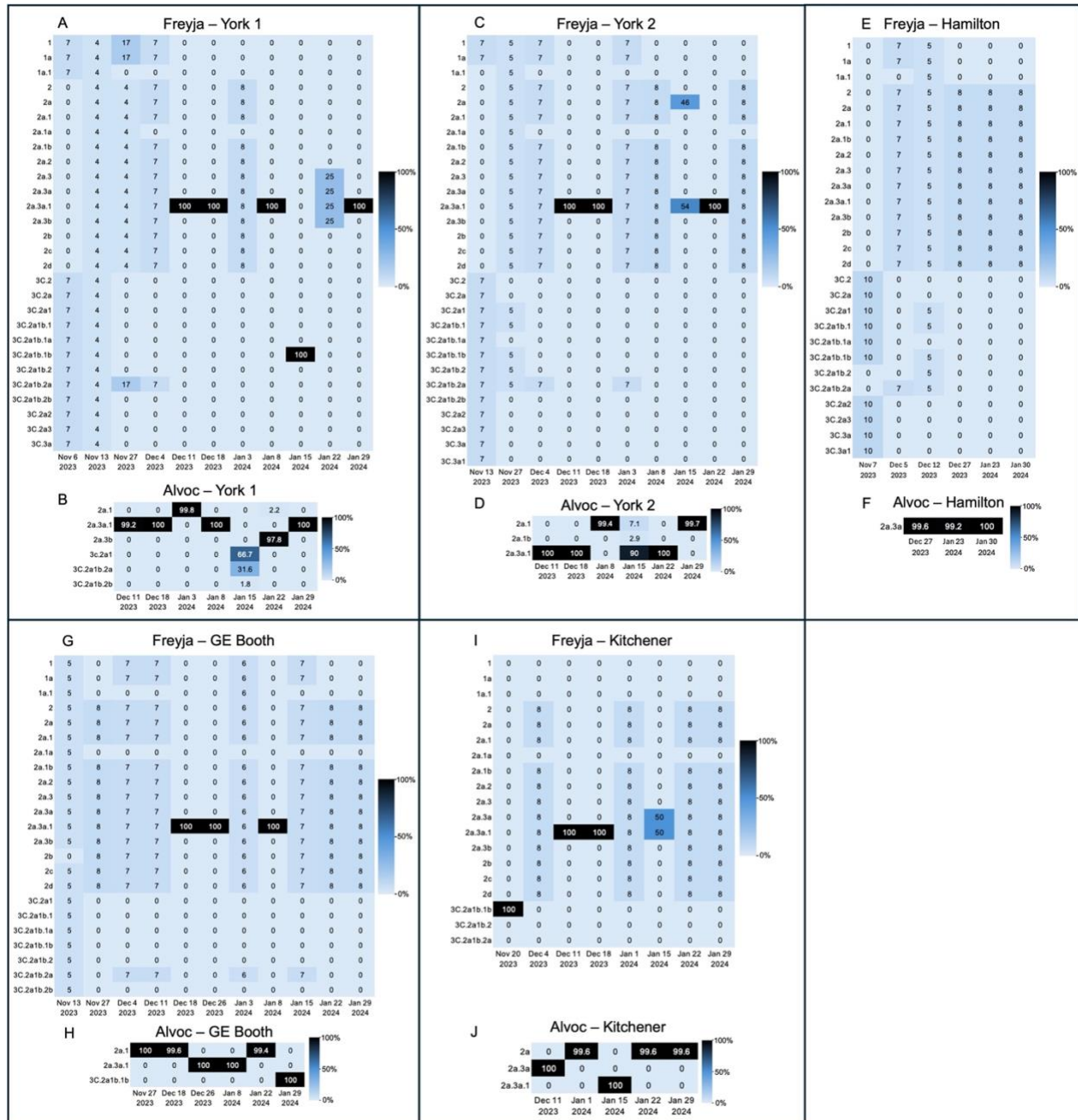


Figure 2.5. IAV H3N2 Variant Calls with Freyja and Alvoc from Seasonal Wastewater Samples. Heatmaps display the lineage abundance calls for IAV H3N2 estimated using Freyja (A,C,E,G,I) and Alvoc (B,D,F,H,J). Includes results from seasonal wastewater samples collected at sites in the York Region (A,B,C,D), Hamilton region (E,F) Peel region (G,H), and the Waterloo Region (I,J).

According to the NextStrain RSV A build, multiple RSV A clades were circulating globally during our sampling period, with A.D.5.2 (24% frequency), A.D.1 (20% frequency), and A.D.3 (13% frequency) representing the three most prevalent clades, several additional clades were also detected at lower frequencies. Overall, variant calls generated by both Freyja and Alvoc were broadly consistent, with minor differences in estimated lineage abundances (Figure 2.6). Discrepancies between tools became more apparent in samples collected toward the end of the RSV season, i.e. the end of December and January, when declining viral RNA concentrations reduced genome recovery, complicating lineage assignment. Notably, in synthetic dataset analyses, Freyja was unable to resolve strains within subclades A.D.3.2–A.D.3.6, instead assigning them to their parent clade A.D.3, whereas Alvoc accurately classified these strains to their respective clades. This pattern was reflected in wastewater samples, where Alvoc detected multiple A.D.3 sub-lineages that Freyja did not resolve (Figure 2.6). Geographically, differences in clade distributions were observed (Supplementary Figure 2.1). In the York (York 1 and York 2) and Peel (GE Booth) Public Health region, which are part of the Greater Toronto Area, A.D.5.2 and A.D. 5.1 emerged as the dominant clade. Whereas samples from the Waterloo (Kitchener and Cambridge) and Hamilton regions, which are located closely together, displayed greater clade diversity with no single dominant lineage (Supplementary Figure 2.1). These findings highlight the regional variation in RSV A transmission dynamics and the ability to capture these dynamics using WBS.

Currently available RSV immunization strategies rely on vaccines incorporating recombinant fusion (F) protein antigens to elicit protective immunity [221], while therapeutic interventions in infected individuals predominantly use monoclonal antibodies targeting the RSV F protein [222,223]. Both approaches are susceptible to the effects of viral evolution, particularly mutations in the F protein that can reduce antibody binding and neutralization [223,224]. Consequently, ongoing genomic surveillance to track the emergence of potential escape variants is critical for

Figure 2 displays 12 heatmaps (A-L) showing the proportion of positive samples for SARS-CoV-2 by date of collection for various locations and time periods. The heatmaps are organized into three rows and four columns. Each panel shows data for a specific location and time period, with a color scale indicating the proportion of positive samples (0% to 100%).

- Row 1: Freyja**
 - A: Freyja – York 1** (Nov 6 to Jan 8, 2023)
 - C: Freyja – York 2** (Nov 13 to Jan 15, 2023)
 - E: Freyja – Hamilton** (Nov 7 to Jan 23, 2023)
- Row 2: Alvoc**
 - B: Alvoc – York 1** (Nov 6 to Jan 8, 2023)
 - D: Alvoc – York 2** (Nov 13 to Jan 15, 2023)
 - F: Alvoc – Hamilton** (Nov 7 to Jan 23, 2023)
- Row 3: Freyja**
 - G: Freyja – GE Booth** (Nov 6 to Jan 8, 2023)
 - I: Freyja – Kitchener** (Nov 6 to Jan 15, 2023)
 - K: Freyja – Cambridge** (Nov 6 to Jan 24, 2023)
- Row 4: Alvoc**
 - H: Alvoc – GE Booth** (Nov 6 to Jan 8, 2023)
 - J: Alvoc – Kitchener** (Nov 6 to Jan 15, 2023)
 - L: Alvoc – Cambridge** (Nov 6 to Jan 24, 2023)

2.3.5 Intraspecific Amplicon Analysis of Tiled-Amplicon Sequencing Assays

45

were covered in 56–57 samples, and NP amplicons 31 and 32 in 53 and 55 samples, respectively. This pattern may reflect PCR primer bias, with certain primers exhibiting greater binding efficiency and amplification success. Alternatively, it may indicate differential RNA degradation across genome segments in wastewater, with PB2 and NP regions being more stable. If so, these segments could serve as more reliable targets for RT-qPCR assays of IAV in wastewater.

For RSV A, amplicon 44, which targets the RNA-directed RNA polymerase (RdRp) gene, displayed the highest coverage, present in 48 of the 67 samples (Figure 2.7b). Some of the currently used RSV A RT-qPCR assays target the N gene [81,163,169,172] and M gene [165]. However, these regions had much lower coverage in the N and M genes compared to the RdRp. Similar to our analysis of IAV H3N2 amplicons, this may suggest greater RNA stability in the RdRp region, making it a preferable target for RT-qPCR assays monitoring RSV in environmental samples.

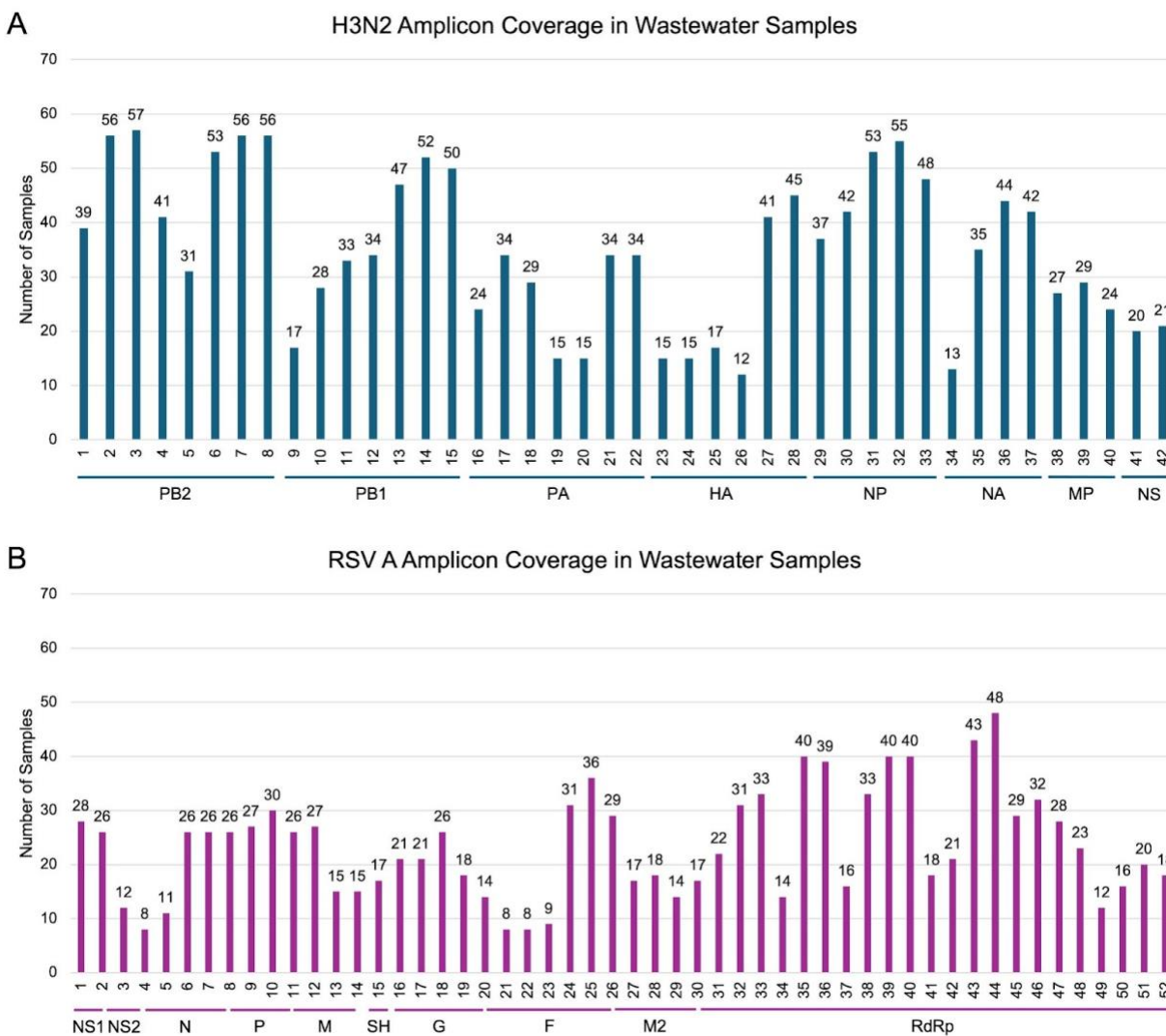


Figure 2.7. IAV H3N2 and RSV A Amplicons with Coverage in Wastewater Samples. The number of samples that contained coverage of IAV H3N2 amplicons (A) and RSV A amplicons (B). A total of 67 samples had genomic coverage of IAV H3N2 and RSV A. The number of samples that had amplicon coverage are labeled above bars. IAV H3N2 assay used 42 amplicons. The genome segment each amplicon bound to is represented by blue bars and segment labels (A) (Supplementary Table 2.5). The RSV A assay used 52 amplicons, the gene each amplicon bound to is represented by purple lines and gene labels (B) (Supplementary Table 2.6).

2.3.6 Future Goals and Assay Optimization

To improve upon our genomic surveillance methods of IAV H3N2 and RSV A, assays and bioinformatic tools could be further optimized. To improve our assays, a variety of experimental modifications regarding virion capture, RNA extraction, and PCR amplification could be evaluated.

As previously discussed, virus concentration and isolation techniques could be compared to increase virion yield. Improving RNA extraction from wastewater could also be explored by testing different kits and including a post-extraction DNase treatment step. Variations of our PCR assays could also be tested to optimize and increase the yield of PCR products prior to library preparation. Such variations could include testing of different thermocycling conditions and increasing the concentration of primers that target genomic regions with low yields. Lastly, the feasibility of pooling IAV H3N2 and RSV A primers to reduce the number of PCR reactions and library preparation samples was not explored, and if successful, could significantly reduce the cost of materials and labour required.

Regarding our bioinformatic approaches, improvements to our IAV H3N2 Alvoc deconvolution tool could also be made. Our RSV A deconvolution tool performed exceedingly well, demonstrating high sensitivity and specificity, and was able to call variants in a greater number of samples than our IAV H3N2 tool. This is likely due to its ability to leverage clade-specific mutations across the entire RSV genome, whereas our H3N2 tool, as well as Freyja, only uses the HA segment to call variants. Thus, a key objective for future surveillance efforts is the development of a robust IAV deconvolution tool capable of leveraging clade-defining mutations distributed across all IAV H3N2 genome segments. This strategy would facilitate accurate clade assignment even in instances of low or absent coverage of the hemagglutinin (HA) gene, by incorporating nucleotide polymorphisms from other genomic regions. Such an approach would enhance the reliability and resolution of lineage determination in metagenomic datasets derived from complex environmental samples, such as wastewater.

Additionally, the scope and effectiveness of an influenza and respiratory virus surveillance program would be strengthened by incorporating a comprehensive, full-genome sequencing assay capable of detecting H1N1(pdm09) viruses. Given that H1N1(pdm09) co-circulates with H3N2 and RSV B co-circulates with RSV A, simultaneous detection and characterization of these

pathogens is crucial for a more complete understanding of community-level viral dynamics. In the present study, the exclusion of H1N1(pdm09) and RSV B detection was a limitation due to resource constraints. Future work should prioritize inclusion of these targets to address current gaps in our WBS methodology. Development of an H1N1(pdm09) targeted assay is of particular importance as this is a previous pandemic swine flu strain that poses a potential future threat, as mutations that arise in this strain could result in increased transmissibility and worsened disease outcomes.

2.4 Conclusions

In summary, this study demonstrates the feasibility and efficacy of wastewater-based genomic surveillance for monitoring the circulation and diversity of respiratory viruses, specifically IAV H3N2 and RSV A. Through comparative analysis of the variant-calling tools Alvoc and Freyja, we show that both are capable of deconvoluting mixed viral populations in environmental samples, with Alvoc exhibiting improved fine-resolution of RSV A lineage assignments. These findings highlight the potential of wastewater genomics to complement traditional clinical surveillance by providing real-time, population-level data on variant prevalence and abundance. To enhance the scope and utility of future surveillance efforts, comprehensive deconvolution tools that use clade-defining mutations from all Influenza A genome segments, along with expanded assay panels to include additional co-circulating H1N1(pdm09) and RSV B genomes should be developed. Furthermore, optimization of virus concentration, isolation, and sequencing workflows will be essential to improve genome recovery and assay sensitivity. Together, these advancements will strengthen the capacity of wastewater-based surveillance systems to inform public health decision-making, guide IAV vaccine strain selection, predict RSV A vaccine escape variants from caused by mutations, and support the early detection of emerging variants with public health significance.

Chapter 3. A Novel Tiled-Amplicon Sequencing Assay Targeting the Tomato Brown Rugose Fruit Virus (ToBRFV) Genome

3.1. Introduction – Tomato Brown Rugose Fruit Virus Incidence, Genome, and Detection

In Canada, fresh market tomatoes have been the most abundantly produced greenhouse vegetable since record keeping began in 1955. In 2022, >293 million kg of fruit with a farm gate value of >CAD 793 million were produced, with >213 million kg of fruit with a farm gate value of >CAD 540 million produced in Ontario [227]. The Tomato Brown Rugose Fruit Virus (ToBRFV) is a devastating plant pathogen that infects tomato and pepper plants and can reach nearly 100% disease incidence, causing massive crop losses and placing a high economic burden on farmers [96,97,228]. Thus, the control or elimination of this virus is a major obstacle in maintaining secure global tomato crop yields.

The first incidence of ToBRFV was reported by Salem and colleagues (2016) following a 2015 outbreak in Jordan. Shortly thereafter, Luria and colleagues (2017) identified ToBRFV as the causative agent of a 2014 tomato disease outbreak in Israel. It has spread to at least 35 countries since, including major tomato-producing countries such as China, Türkiye, and Mexico [96,229–231]. Global transmission and clade-specific mutations of ToBRFV strains are monitored through a curated NextStrain database updated by authorized curators with each updated version referred to as a build [232–234].

The ToBRFV is a member of the Tobamovirus genus, a group of highly virulent plant pathogens that includes 37 virus species with diverse and distinct host ranges [228,235,236]. The ToBRFV genome shares the highest sequence identity with Tobacco Mosaic Virus (TMV) isolate Ohio V at 82.4% shared identity, and fulfills species demarcation criteria [97,237]. All ToBRFV isolates available in NextStrain share greater than 90% sequence identity, meeting the demarcation

criteria. The ToBRFV 2022 NextStrain build (version 3) contains 179 ToBRFV isolates, grouped into eight distinct clades, which represent diverging lineages of ToBRFV strains [232,233].

Tobamovirus virions are rod-shaped particles up to 300 nm long and 20 nm wide and are non-enveloped [235,238]. The genome structure and gene order are conserved amongst all tobamoviruses, consisting of a ~6.4 kbp monopartite single-stranded positive-sense RNA genome containing four open reading frames (ORFs). Replication proteins are encoded by ORF1 and ORF2, whereas virus structural proteins are encoded by ORF3 and ORF4 and expressed through subgenomic RNA [96]. ORF1 encodes a 126 kDa RNA-dependent RNA polymerase (RdRp), while ORF2 encodes a 183 kDa helicase protein through an amber UAG stop codon downstream of ORF1 [97]. ORF3 encodes a 30 kDa movement protein (MP) that facilitates cell-to-cell transmission through the plant plasmodesmata. The MP binds to multiple cellular components that transport the viral RNA through the plasmodesmata. Lastly, ORF4 encodes a coat protein that encapsulates the virus genome forming a thin rod-like structure [97,238]. Changes in the ToBRFV MP sequence break tobamovirus-resistance genes that have been introduced into most commercial crop species to provide protection against tobamoviruses [239]. ToBRFV MP amino acid mutations result in structural changes that prevent MP binding by the plant Tm-2² receptor protein, thereby preventing activation of the hypersensitive response and leading to severe infection symptoms [239]. Recently, ToBRFV resistant tomato seed lines have been commercially developed; however, only a limited number of varieties are currently available (<https://www.vegetables.bayer.com/ca/en-ca/resources/news/bayer-launches-additional-commercial-varieties-with-intermediate.html>, accessed on 16 March 2024; <https://www.syngentavegetables.com/resources/ToBRFV>, accessed on 7 June 2023).

The NextStrain database is a useful resource for scientists and policymakers to understand transmission routes and establish disease control measures [234]. For example, genomic sequence data from commercial seeds analyzed using NextStrain phylogenies alerted Dutch

authorities to the illegal use of a ToBRFV cross-protective species [232]. ToBRFV genome sequences can be submitted and then incorporated into updated builds by authorized curators. This community-based approach has increased the number of genomes available in each version [232,233]. The ToBRFV 2019, 2020, and 2022 NextStrain builds contain 63, 118, and 179 ToBRFV strains, respectively, which can be phylogenetically grouped into three, six, and eight clades, respectively, representing diverging ToBRFV clades [232–234]. However, when Abrahamian and colleagues (2022) analyzed genome sequences of 123 ToBRFV strains retrieved from GenBank as well as another 22 isolated from plant tissues, they identified three distinct clades [237]. Zhang and colleagues (2022) also phylogenetically analyzed 78 ToBRFV genomes and identified three distinct clades [96]. These differences between identified clades are likely a result of the differing phylogenetic analysis tools and genome sequences used by each study. Additionally, the majority of isolated strains originate from the Netherlands and this sampling bias has likely skewed ToBRFV phylogenetic analyses [232,237]. To fully understand ToBRFV phylogeny, an increased number of genome sequences from various regions must be obtained, which can be aided by improved sequencing methods [96,237]. To aid in ToBRFV identification and surveillance efforts, we developed a robust and specific ToBRFV genome tiled amplicon sequencing assay utilizing short-read Illumina sequencing technology.

High-throughput shotgun sequencing technology can produce millions of reads from a single sample [240–242]. However, the low abundance of viral reads and the high abundance of host and background reads in environmental samples can make it challenging to sequence viral genomes without utilizing virus-specific enrichment methods such as PCR [241,243]. PCR enrichment assays are often used to amplify specific DNA sequences; however, they require the development of target-specific primers [120,244,245]. Recently, an open-source collaboration called the ARTIC Network has created adaptable user-friendly tools to develop virus-specific sequencing assays utilizing PCR enrichment (<https://artic.network/>; accessed on 20 February

2024). The ARTIC Network was originally developed for monitoring the transmission and evolution of viruses such as MERS-CoV and Ebola but has since been adapted to streamline the design of species specific tiled amplicon sequencing assays [120,246]. ARTIC primer schemes are designed to target short overlapping fragments, referred to as amplicons, that span a specific genome of interest. PCR-enriched amplicons can be sequenced using either Illumina or Nanopore platforms to obtain complete genome sequences. ARTIC primer schemes have been used to target many viral genomes, including but not limited to Zika, Ebola, Dengue, and SARS-CoV-2 [120,121,246–249]. We used the ARTIC infrastructure to develop a nearly whole genome tiled amplicon sequencing assay that enriches ToBRFV genomes for Illumina sequencing. During our ongoing exploration of SARS-CoV-2 content in wastewater influent (WWI), we observed widespread occurrence of the ToBRFV sequences in samples from multiple treatment plants in Ontario, Canada (data available upon request). Indeed, several recent studies have observed ToBRFV sequences in WWI, indicating this virus is likely widespread and prevalent in wastewater systems in many countries [250–252].

ToBRFV sequences have been observed in wastewater samples from Slovenia and California [250–252]. In 2020, Bačnik and colleagues demonstrated that wastewater influent contains active ToBRFV virions which, when concentrated, can cause asymptomatic infections in tomato plants [250]. Reclaimed wastewater is commonly used to irrigate crops, however several species in the Tobamovirus genus, including Pepper Mild Mottle Virus (PMMoV), remain transmissible in treated wastewater effluent [250]. Thus, ToBRFV could potentially be transmitted to crops via reclaimed wastewater during irrigation. Moreover, in 2023, Mehle and colleagues demonstrated that ToBRFV-infected plants shed active virions through their roots, which can then spread to other plants through shared hydroponic systems that distribute nutrient solution. This suggests that plants exposed to ToBRFV-contaminated wastewater could potentially become infected and cause widespread outbreaks. This could be a challenge for growers who use reclaimed

wastewater to irrigate crops [253,254]. However, it has not yet been shown that ToBRFV virions remain infective after conventional wastewater treatment methods. Thus, the possibility remains that treated wastewater effluent is suitable for irrigation and further work must be performed to determine ToBRFV transmissibility [250,251]. The entry mode of ToBRFV virions into wastewater has not yet been established. One possible explanation for ToBRFV transmission is through environmental sources such as waste from regional greenhouses and fields with ongoing ToBRFV infections [252]. Another possible mode of transmission could be through a human dietary route [252]. If the virus is present in produce and tomato- or pepper-based products, then the excretion of these products by humans could transmit ToBRFV virions into WWI [255]. To test these hypotheses, WWI could be obtained and tested for ToBRFV from regions without ToBRFV infections or where there is no tomato or pepper cultivation to eliminate the possibility of transmission through agricultural waste. Additionally, tomato and pepper products from grocery stores, as well as human feces, could be tested for the ToBRFV to establish a dietary route of transmission [255]. Similarly, a dietary mode of PMMoV transmission to human feces has been established, thus, the human dietary mode of ToBRFV transmission into WWI is the more probable scenario [255].

Current ToBRFV screening procedures, such as RT-qPCR and loop-isothermal amplification assays, can be used to determine the presence or absence of ToBRFV sequences, quantify the viral load, and provide more rapid results than sequencing assays. However, many of these assays cannot detect strain-specific mutations or they require the development of specialized strain-specific primers/probes to track transmission routes and determine sources of infection [96,251,256,257]. Moreover, increased evolutionary pressure from disease eradication efforts could result in the emergence of novel mutations that evade or reduce the effectiveness of current rapid presence/absence detection assays and eradication measures [237]. Thus, sequencing of complete ToBRFV genomes is essential to monitor virus transmission and evolution, predict

emerging threats, and fully understand ToBRFV phylogeny [232]. However, high-throughput shotgun sequencing analysis is not ideal for the recovery of ToBRFV genomes due to the low recovery of viral sequences, high costs, and length of sample processing time [96]. To circumvent this problem, we designed a ToBRFV-specific tiled amplicon assay. Results show our tiled-amplicon assay provides a reliable cost-effective method to sequence ToBRFV genomes as well as aid in global surveillance efforts and uncover novel phylotypes of clades. To aid in ToBRFV surveillance efforts, we used wastewater RNA extracts from five Ontario WWI collection sites to develop and validate a novel ToBRFV tiled amplicon sequencing (ToBRFV-Seq) assay. We prepared RNA sequencing libraries using our ARTIC-style ToBRFV-Seq assay and an RNA shotgun sequencing assay using the NebNext Ultra II RNA Prep kit. By evaluating the recovery of ToBRFV reads by each method, we demonstrate that our assay is a highly specific and efficient method for ToBRFV genome sequencing. Moreover, we adapted our in-house SARS-CoV-2 lineage abundance estimation tool, Alcov [196], to identify and estimate the relative abundance of ToBRFV clades. Our adapted tool, named Altob (<https://github.com/Ellmen/altob>; accessed on 20 February 2024), was able to analyze clade-specific mutations and estimate the abundance of ToBRFV NextStrain clades 1–4 and 6–8 in samples.

3.2. Materials and Methods

3.2.1. Viral RNA Extraction from Wastewater Influent

Wastewater influent (WWI) samples were collected in plastic bottles and shipped in coolers with ice packs from five wastewater treatment plants in Ontario, Canada (Supplementary Table 3.1). Samples from five sites (designated A–E) were processed using virion capture Nanotrap Microbiome A particles (*Ceres Nanosciences*, #44202), then RNA was extracted using the Qiagen RNeasy Mini kit (*Qiagen*, #74104). Briefly, bottles of WWI were inverted to mix and allowed to settle for 1 min at room temperature, then 10 mL of wastewater supernatant containing suspended biosolids was moved to a fresh tube. A negative control sample was prepared alongside extracts using 10 mL of nuclease-free DI water instead of WWI. The supernatant was mixed with 100 µL of ER2 solution by briefly vortexing. Then, 150 µL of Microbiome A particles were added and the samples were inverted to mix. The samples were then incubated at room temperature for 10 min, with three inversions after 5 min of incubation. Samples were placed on a 15 mL tube magnetic stand; the cleared supernatant was discarded, and magnetic beads were retained.

Beads were gently resuspended in 1 mL of nuclease-free water and transferred to a sterile 1.5 mL microfuge tube. Beads were placed on a magnetic stand and incubated for 1 min and the supernatant was discarded without disturbing the bead pellet. Beads were resuspended 700 µL of RLT lysis buffer containing 1% 2-ME (*BioShop*, #MER002.500) by pipetting and then vortexing. Samples were incubated at room temperature for 10 min, then 600 µL of the supernatant from each sample was manually extracted using the Qiagen RNeasy Mini kit (*Qiagen*, #74104). RNA was eluted in 80 µL of RNase-free water and the concentration of DNA and RNA was analyzed on a Qubit 4.0 fluorometer (*ThermoFisher Scientific*, #Q33238). DNA was degraded using the Ambion DNase I kit (*Invitrogen*, #AM2222), samples were incubated with 4 units of DNase I at 37°C for 45 min. RNA was recovered using the RNeasy Mini Kit RNA (*Qiagen*, #74104) clean-up

protocol, DNA and RNA concentration were then determined by Qubit 4.0 analysis to ensure RNA integrity and complete removal of DNA.

3.2.2. Primer Design and Pooling

Primers were designed by first generating a multiple sequence alignment (MSA) from the genomes of 118 ToBRFV strains listed on the 2020 NextStrain Build (version 2), and sequences were obtained from NCBI (Supplementary Table 3.2). A consensus sequence was generated from the MSA and uploaded to the web-based tool PrimalScheme to generate primer sequences targeting overlapping amplicons of approximately 400 base pairs (bp) in length. PrimalScheme generated 20 amplicons for a total of 40 primers that cover 95.7% of the ToBRFV genome NC_028478.1, which was used for downstream read mapping analysis. Primers were synthesized by Integrated DNA Technologies. Primers amplifying odd and even numbered amplicons were diluted and mixed in two separate aliquots (pools 1 and 2, respectively) to avoid mis-priming between overlapping amplicons (Figure 3.1). Each primer was pooled at a concentration of 0.5 μ M in respective pools (Supplementary Table 3.3).

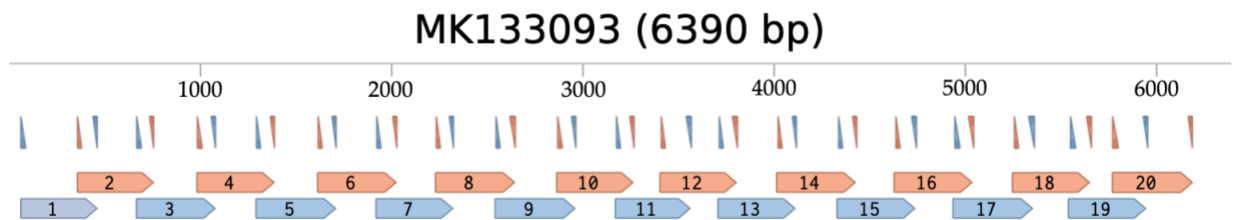


Figure 3.1. *ToBRFV-Specific Primer Binding Sites and Resulting Amplicons Mapped to the ToBRFV MK133093 Genome Sequence.* In silico visualization of primer binding sites; forward and reverse primers are denoted by the upward and downward pointing triangles, respectively, and each resulting amplicon is numbered and denoted by blue and orange flags [48]. Pool 1 primers and amplicons are blue, pool 2 primers and amplicons are orange. Nucleotide positions are labeled at 1000 bp intervals.

3.2.3. ToBRFV-Targeted Tiled-Amplicon Sequencing Library Preparation

RNA extracts from WWI samples were reverse transcribed to synthesize cDNA using the LunaScript SuperMix RT kit (*New England Biolabs*, #E3010L); 10 μ L of each RNA extract was used in a total reaction volume of 20 μ L. First-strand cDNA synthesis products were PCR-amplified in two reaction mixtures containing primer pools 1 or 2 (Figure 3.1). The PCR was performed using Q5 2x Master Mix (*New England Biolabs*, #M0492S), 6 μ L of cDNA synthesis reaction, and 2.5 μ L of one primer pool, adding ddH₂O to a total volume of 25 μ L. Samples were initially denatured at 98°C for 30 s, followed by 35 cycles of 98°C for 15 s, 63°C for 30 s, and 72°C for 30 s, ending with a final extension of 5 min at 72°C. PCR products were prepared for sequencing using the Illumina DNA Prep kit (*Illumina*, #20060060) using half of the manufacturer's recommended volumes. Briefly, PCR products were purified using AMPure XP beads (*Beckman Coulter*, #A63881), then assessed by Qubit 4.0 fluorometer and agarose gel electrophoresis. Purified PCR products were tagmented using on-bead tagmentation and barcoded with custom unique dual indexes using six PCR cycles. Indexed PCR products of ~400 bp were isolated using double-sided magnetic bead purification. The purified fragments were assessed using a Qubit 4.0 fluorometer and agarose gel electrophoresis.

3.2.4. RNA Shotgun Sequencing Library Preparation

RNA extracts from WWI were prepared for RNA sequencing using the NEB Ultra II RNA Sequencing kit (*New England Biolabs*, #E7770S), following the manufacturer's recommended protocol, Section 5. RNA extracts were quantified using a Qubit 4.0 fluorometer, then reverse-transcribed using first- and second-strand synthesis enzymes. AMPure XP beads were used to purify double-stranded cDNA fragments. Fragments were end-prepped by dA-tailing enzymes, then adaptor-ligated. Ligated fragments were purified using AMPure XP beads, PCR-enriched and barcoded with unique dual indexes (*Illumina*, #20027213) using six PCR cycles. PCR

products were purified using AMPure XP beads and then assessed using a Qubit 4.0 fluorometer and agarose gel electrophoresis.

3.2.5. Library Pooling and Illumina NextSeq Sequencing

ToBRFV-Seq and RNA shotgun-prepared libraries were pooled. Samples prepared using the same method, either RNA shotgun or ToBRFV-Seq, were pooled at equal concentrations and shotgun libraries were pooled at 15°C the concentration of ToBRFV-Seq libraries. The pooled libraries were sequenced on an Illumina NextSeq2000 using 2 x 300 sequencing and P1 chemistry. Approximately 65% of the flowcell sequencing capacity was dedicated to these samples, for a total expected yield of 52 GB.

3.2.6. Read Processing, Taxonomic Classification, and ToBRFV Genome Alignment

Sequencing read quality and adaptor content were evaluated using FastQC v0.12.1 [258]. Low-quality reads and adaptor sequences were removed and reads were paired using Trimmomatic v0.39 using flags ILLUMINACLIP:<a>:2:30:7:2:TRUE, LEADING:3, TRAILING:3, SLIDINGWINDOW:5:15, MINLEN:125 [259]. RNA shotgun reads were trimmed using NebNext Illumina adaptor read sequence 1 AGATCGGAAGAGCACACGTCTGAACTCCAGTCA and read sequence 2 AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT while ToBRFV-Seq reads were trimmed using the Illumina DNA Prep adaptor sequence CTGTCTCTTATACACATCT. The overall quality of read trimming was evaluated using FastQC v0.12.1 [258]. Read pairs passing quality control filters were taxonomically classified with Kraken2 v2.0.7-beta, using the flag-paired and the standard database, and reports of read taxonomic rankings were generated for each sample [260]. Quality-controlled read pairs were aligned to the ToBRFV genome NC_028478.1 using bowtie2 v2.3.5 with the flag-local [261]. Alignments were converted into bam files with samtools v1.9 files and anvio-7.1, and then visualized using Tablet 1.21.02.08 [208,262,263].

3.2.7. Altob Implementation and Synthetic Read Simulation

Our in-house SARS-CoV-2 variant of concern lineage abundance estimation tool, Alcov, was adapted to identify and estimate ToBRFV clades in Illumina paired reads [196]. We used Seaview to compute a multiple sequence alignment (MSA) with the MUSCLE algorithm using NC_028478.1 as the root genome and 124 unique genomes listed on NextStrain (Supplementary Table 3.2) [232,264,265]. Then, a phylogenetic Maximum Likelihood tree was generated using a GTR model and 100 bootstrap replicates and compared to the 2022 NextStrain ToBRFV phylogenetic tree (Supplementary Figure 3.1). Phylogenetically grouped isolates were used to generate MSAs for each clade using NC_028478.1 as the root and the MUSCLE algorithm. MSAs were used to generate a list of mutations that appear in at least 50% of clade-specific isolates, which were then used to construct constellations of clade-specific mutations to estimate ToBRFV clade abundances using Altob. Positions in our mutation list are relative to position one of the reference genomes, whereas the mutations reported in the NextStrain phylogeny are relative to position one of their ToBRFV MSAs, which is either 106 or 107 bp upstream of our start site.

Altob was adapted from our in-house tool Alcov by modifying the original mutation constellations and reference genome; all other functions and parameters were maintained. A detailed description of Alcov and its functions can be found in our preprint [196]. Briefly, Altob scans sites with a minimum coverage of 40 reads for the provided mutations and calculates frequencies as the number of mutations over the total reads at that site [196]. Altob computes the clade abundance that best explains the observed pattern of mutations utilizing linear programming, a well-studied optimization technique [196,266]. Synthetic read data used to benchmark Altob were generated using DGWSIM v0.1.13 and genomes of eight randomly selected ToBRFV strains, each belonging to a different NextStrain clade (Supplementary Table 3.2) [209,232]. Three datasets containing either 1000, 10,000, or 100,000 paired 2 x 250 Illumina reads were generated for each of the eight genomes without introducing random mutations or random DNA reads. Files

containing 100,000 paired reads were combined to generate datasets with a mixture of clade-specific mutations. An equal number of reads from eight, four, and two genomes were combined, thus, each clade is represented by 12.5%, 25%, or 50% of reads, respectively. Altob is publicly available on GitHub (<https://github.com/Ellmen/altob>; accessed on 20 February 2024).

3.3. Results

3.3.1. In Silico Primer Binding Analysis of ToBRFV Genome Sequences of 125 Strains

To ensure our 40 ToBRFV-specific primers would inclusively target a wide variety of ToBRFV strains without compromising species-specific stringency, we performed an in silico primer binding analysis using the sequence analysis software Benchling (2023) and 137 tobamovirus genomes [267]. A list of 179 ToBRFV strain accession numbers from the NextStrain ToBRFV build (2022, version 3) was obtained, representing 125 unique sequences, and these genomes were downloaded from NCBI. In addition, 12 tobamovirus Ref-Seq genomes of species closely related to the ToBRFV were obtained from NCBI (accessed on 20 April 2023) and all 137 unique genome sequences were imported into Benchling [236,268]. Benchling was used to probe all 137 genome sequences for potential primer binding sites using the program's Find Primer Binding Sites tool. A successful match required at least 18 matching bases, a maximum of three mismatches, and a maximum of one consecutive nucleotide.

The positions of matching primers were mapped to each genome and visualized using Benchling to ensure the overlapping amplicon primer scheme was maintained. All 40 ToBRFV-specific primers had one matching binding site within each of the 125 ToBRFV-strain genomes (Table 3.1), with between zero and two mismatches observed between primers and ToBRFV-strain genomes. Several binding sites were identified in genomes of tobamovirus species closely related to the ToBRFV; 15, 14, 11, and 9 primers produced partial matches to tobacco mosaic virus (TMV), tomato mottle mosaic virus (ToMMV), rehmannia mosaic virus (ReMV), and tomato mosaic virus (ToMV) genomes, respectively (Table 3.1). Fewer matches, between zero and six, were identified for the other nine tobamovirus genomes, which are more distantly related to the ToBRFV (Table 3.1) [97].

Table 3.1. *In silico* Analysis of ToBRFV-specific Primer Binding to 137 Tobamovirus Genomes. *In silico* primer binding analysis identified one potential binding site in 137 tobamovirus genomes for all 40 ToBRFV-specific primers [267]. A limited number of ToBRFV-specific primers are partially bound to other tobamovirus genomes. A greater number of primer binding sites was observed in genomes of species sharing more recent ancestry with the ToBRFV.

Tobamovirus species	Accession	Primer Count
All 125 unique ToBRFV genomes	Supplementary Table 3.2	40
Tobacco mosaic virus (TMV)	NC_001367.1	15
Tomato mottle mosaic virus (ToMMV)	NC_022230.1	9
Tomato mosaic virus (ToMV)	NC_002692.1	14
Pepper mild mottle virus (PMMoV)	NC_003630.1	3
Bell pepper mottle tobamovirus (BPeMV)	NC_009642.1	6
Obuda pepper virus (ObPV)	NC_003852.1	1
Paprika mild mottle virus (PaMMV)	NC_004106.1	3
Tobacco mild green mosaic virus (TMGMV)	NC_001556.1	2
Cucumber green mottle mosaic virus (CGMMV)	NC_001801.1	0
Cucumber fruit mottle mosaic virus (CFMMV)	NC_002633.1	0
Brugmansia mild mottle virus (BrMMV)	NC_010944.1	6
Zucchini green mottle mosaic virus (ZGMMV)	NC_003878.1	0

Our *in silico* primer binding analysis suggests all 40 primers could amplify the genomic sequences of all 179 ToBRFV strains listed on NextStrain while maintaining an overlapping binding scheme across each genome (Table 3.1, Figure 3.1). *In silico* results demonstrate that our primers are highly specific to ToBRFV genome sequences yet remain non-specific at the strain level. Moreover, limited binding to other tobamovirus genomes shows our primers are unlikely to amplify the complete genome of off-target species; however, partial genome segments could potentially be amplified.

3.3.2. Read Quality Control and Adaptor Trimming

Sequenced RNA shotgun libraries produced on average 4.64 Gbp and 15,627,313 reads per sample, whereas ToBRFV-Seq libraries produced on average 179,820 Mbp and 597,554 reads per sample, and in total 63.7 GB of data were produced (Table 3.2). Due to the low level of viral reads and high level of background eukaryotic and prokaryotic DNA found in environmental samples, shotgun-prepared libraries often require a high sequencing depth to obtain even a

modest number of viral reads [115,241]. Thus, RNA shotgun samples were loaded at a 15 x greater concentration than ToBRFV-Seq prepared samples. The majority of reads were high-quality, with RNA shotgun and ToBRFV-Seq prepared samples containing an average of 94.07% and 83.72% reads passing filters, respectively (Table 3.2). All raw reads are available through the SRA database under BioProject accession PRJNA1030775.

Table 3.2. *Total and Filtered Reads from RNA Shotgun and ToBRFV-Seq Prepared Sequencing Libraries.* The total number of reads (Pairs), number of filtered paired reads (QF Pairs), and percent of quality filtered (%PF) reads obtained from WWI samples and a nuclease-free water blank prepared with either RNA shotgun or ToBRFV-Seq. Samples A–E were averaged, excluding the NTC (no-template control) blank.

Sample	RNA Shotgun				ToBRFV-Seq			
	Total Gbp	Total Read Pairs	QF Read Pairs	% PF	Total Gbp	Total Read Pairs	QF Read Pairs	% PF
Average	4.640000	15,627,313	14,631,264	94.07	0.179820	597,554	502,007	83.72
A	5.500000	18,528,523	17,138,798	92.50	0.168800	560,908	451,928	80.57
B	4.900000	16,434,198	15,693,587	95.49	0.210300	698,848	609,572	87.23
C	3.600000	12,035,815	11,788,803	97.95	0.158900	527,913	445,326	84.36
D	5.400000	18,266,390	16,248,968	88.96	0.205400	682,673	586,488	85.91
E	3.800000	12,871,637	12,286,164	95.45	0.155700	517,430	416,720	80.54
Blank	0.039600	131,614	123,528	93.86	0.000057	190	108	56.84

3.3.3. Taxonomic Profile of WWI Samples Prepared by RNA Shotgun and ToBRFV-Seq Methods

Domain-level taxonomic classification of read pairs by Kraken2 against the standard database was computed to compare the recovery of viral reads obtained with each library prep method (Figure 3.2a,b) [260]. RNA shotgun reads were mostly bacterial, representing on average 89.40% of sample reads; 0.06% were viral, while the remaining 10.53% of reads were eukaryotic, archaeal, or other. Comparatively, on average, 97.38% of ToBRFV-Seq reads were classified as viral while the remaining 2.62% of reads were eukaryotic, viral, archaeal, or other (Figure 3.2c). The high recovery of bacterial reads in cDNA shotgun libraries is likely a result of residual bacterial

genomic DNA, rRNA, and mRNA sequences in prepared samples and is a common limitation of viral metagenomic sequencing [114,269]. DNA contamination could be prevented by including thorough quality checks to confirm DNA has been completely removed from the sample prior to library preparation steps [269]. Notably, DNA contamination did not impact the recovery of viral reads using our ToBRFV-Seq, even though the same extracts were used for each library prep method. Moreover, we have found DNase treatment unnecessary to obtain a high number of specific reads with our ToBRFV-Seq assay, as this step was not used during our library preparation and its absence did not impact the ability to recover ToBRFV reads.

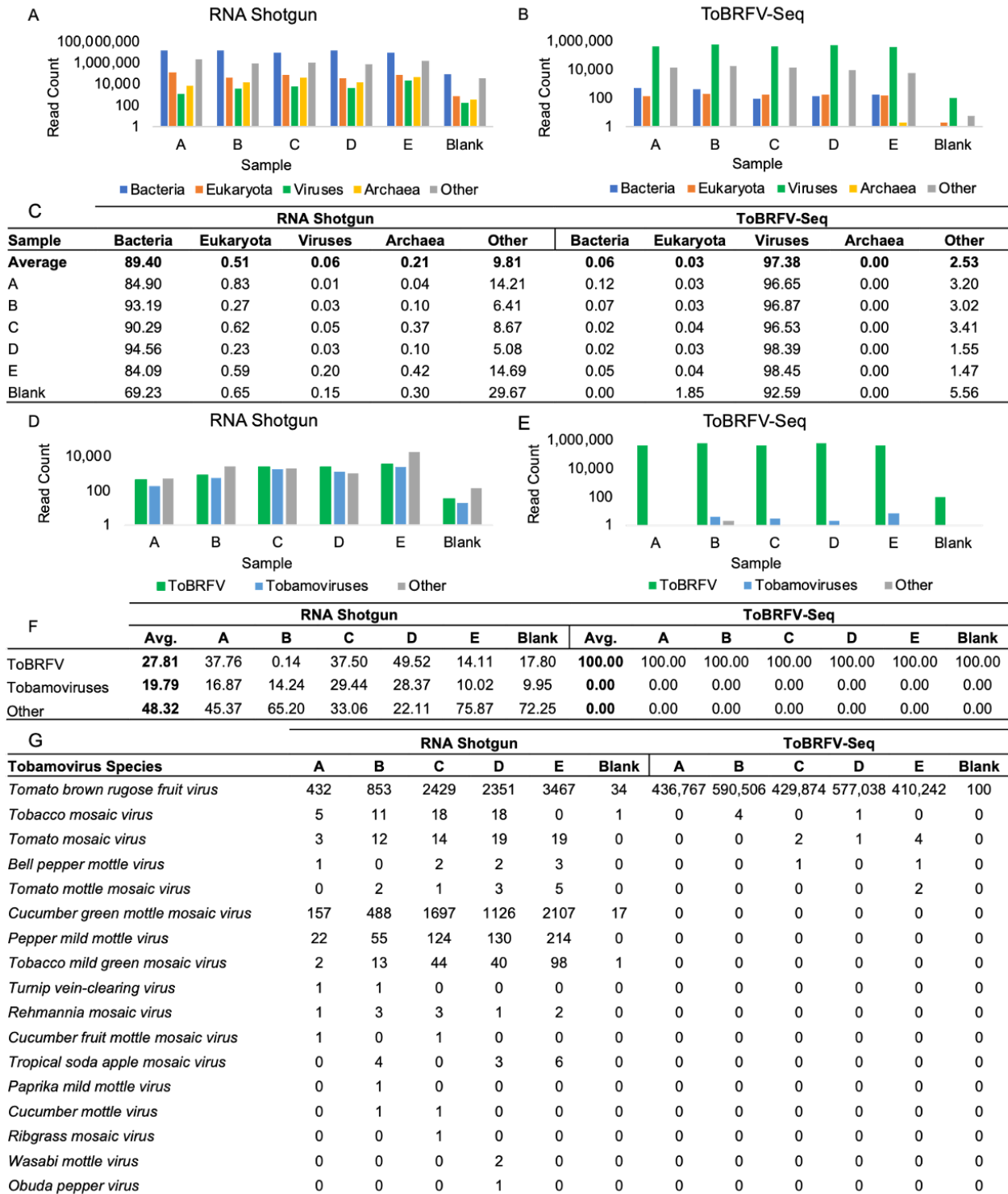


Figure 3.2. Taxonomic Classification of RNA Shotgun and ToBRFV-Seq Read Pairs at the Domain and Species-level. Kraken2 classified read pair counts at the domain level for (A) RNA shotgun library prepared samples and (B) ToBRFV-Seq prepared samples [39]. At the domain level, categories included are viral, archaea, bacterial, eukaryotic, or other reads, where other includes unclassified reads, plasmids, adapters, linkers, and/or primers. (C) The percentage of reads represented by each domain for both prep methods.

Samples A–E were used to calculate the average, the blank was excluded. (D,E) Percentage of total reads classified as ToBRFV (T), other tobamoviruses (OT), and the remainder of virus classified reads (RR) (F). Counts of tobamovirus classified read pairs at the species level (G). (A,B,D,E) Reads are represented on a log scale to magnify and visualize low abundance reads. Two-tailed tests were used to compare the mean read counts of RNA shotgun and ToBRFV-Seq classified reads, respectively. (D) RNA shotgun samples, $p < 0.05$, $n = 5$, $df = 4$, t-values, (T-OT) 3.569, (T-RR)-1.015, and (RR-OT)-1.181, critical t-value ∓ 2.776 . (E) ToBRFV-Seq samples, $p < 0.05$, $n = 5$, $df = 4$, t-values, (T-OT) 12.523, (T-RR) 12.523, and (RR-OT) 2.449, critical t-value ∓ 2.776 . Thus, there is a statistically significant difference between the number of ToBRFV and other tobamovirus reads produced by RNA shotgun sequencing, whereas there is a statistically significant difference between the number of ToBRFV reads produced compared to both other tobamovirus reads and the read remainder.

Viral read pairs were further classified at the species level to evaluate the recovery of ToBRFV, other tobamovirus, and all other virus species by each library preparation method (Figure 3.2d,e). As expected, RNA shotgun sequencing yielded a mixture of tobamovirus reads with a high proportion specific to the ToBRFV, as well as other virus species (Figure 3.2d,f). On the other hand, our ToBRFV-Seq assay almost exclusively produced ToBRFV reads (Figure 3.2e,f). Although other tobamoviruses were observed in all WWI samples, they were not significantly enriched within ToBRFV-Seq samples (Figure 3.2g). This demonstrates that our primers do not amplify these closely related off-target sequences when present in low abundance and are therefore highly specific to ToBRFV genomes. Of note, the ToBRFV-Seq blank sample, prepared using nuclease-free water, contained 100 ToBRFV read pairs (Figure 3.2g). This is likely due to sample contamination prior to or during library preparation and is a commonly encountered issue when preparing tiled amplicon sequencing assays [134]. Thus, extra precautions should be taken to reduce sample contamination during RNA extraction and PCR setup [270].

3.3.4. Mapping Reads to a ToBRFV Reference Genome

The quality of ToBRFV genome recovery was assessed by mapping reads from samples to a ToBRFV genome using bowtie2, and assessing coverage, completeness, and depth of read mapping to the reference [261]. Our ToBRFV-specific primers target 95.7% of the ToBRFV genome sequence, forgoing 64 bp and 210 bp on the 5' and 3' genome ends, respectively. These

regions are less than 400 bp and therefore too small to be included in our tiled amplicon primer scheme. Thus, the percentage of genome coverage was calculated for both the total and PCR-targeted genome regions. Both library preparation methods were able to obtain nearly complete or complete coverage of the total and PCR-targeted genome; however, RNA shotgun samples contained far fewer reads and had a lower depth of coverage than ToBRFV-Seq samples (Table 3.3). Moreover, RNA shotgun samples on average contained 15,627,313 read pairs, yet only 0.02% of reads aligned to the ToBRFV reference genome (Tables 3.2 and 3.3). Conversely, ToBRFV-Seq samples contained on average 597,554 paired reads with 99.86% of reads aligning to the reference genome (Tables 3.2 and 3.3). RNA shotgun samples used a much greater sequencing capacity than our ToBRFV-Seq samples, yet still yielded fewer ToBRFV reads and lower depth of sequencing. Thus, our ToBRFV-Seq assay offers an alternative and more robust approach for sequencing ToBRFV genomes that is highly specific and requires minimal sequencing capacity. On average, ToBRFV-Seq samples covered 99.92% of the PCR-targeted genome sequence with 0–6 bp missing on alignment terminal ends.

Table 3.3. *Alignment and Coverage of RNA Shotgun and ToBRFV-Seq Reads Mapped to a ToBRFV Genome.* Number of ToBRFV reads and percentage of total reads that align to the ToBRFV genome NC_028478.1, % coverage of the total (% Total Cov) and PCR-targeted (% Target Cov) genomic region, the average read depth, and maximum read depth. Samples A–E were included in the average and the blank was excluded.

Sample	RNA Shotgun						ToBRFV-Seq					
	ToBRFV Reads	% Aligned	% Total Cov.	% Target Cov.	Avg. Depth	Max. Depth	ToBRFV Reads	% Aligned	% Total Cov.	% Target Cov.	Avg. Depth	Max. Depth
Average	3,840	0.02	99.06	99.71	155.31	472	977,491	99.86	95.69	99.92	35,694.79	206,304
A	867	0.00	97.45	98.53	29.94	110	873,270	99.78	95.67	99.89	32,084.65	227,756
B	1,732	0.01	99.34	100.00	67.63	184	1,180,695	99.85	95.70	99.85	44,893.64	312,224
C	4,891	0.02	99.5	100.00	211.20	623	859,568	99.90	95.67	99.95	31,268.41	154,096
D	4,742	0.02	99.44	100.00	177.90	556	1,153,735	99.90	95.74	99.95	40,645.85	227,468
E	6,968	0.03	99.58	100.00	289.89	886	820,185	99.85	95.67	99.95	29,581.39	109,976
Blank	70	0.04	72.96	73.36	2.93	16	199	96.60	53.00	56.53	6.33	36

3.3.5. Estimating the Relative Abundance of ToBRFV Clades in Wastewater Influent

We adapted our tool, Alcov, to estimate the abundance of ToBRFV clades in Illumina sequencing reads and named the adapted tool Altob. Clade-specific mutations were defined by constructing a Maximum Likelihood tree from unique ToBRFV isolates listed on the NextStrain 2022 build and clade-specific MSAs aligned to the reference genome NC_028478.1 were used to define mutations [232]. We were able to resolve seven of the eight clades identified on NextStrain; however, clade five isolates did not form a distinct clade and appeared in several divergent branches (Supplementary Table 3.2, Supplementary Figure 3.1). Benchmarking analysis of Altob was performed using synthetic Illumina read datasets of 1000, 10,000, and 100,000 reads using genomes that represent each ToBRFV clade [232]. Altob correctly identified all clades and estimated between 94.2% and 98.7% relative abundance for the respective clades (Figure 3.3a), except for clade five. Clade five contains the wildtype ToBRFV used to root our tree and define mutations in all other clades; therefore, the clade five clade is not called by Altob [232]. Synthetic read datasets containing a mixture of clade-specific reads were also correctly identified using Altob (Figure 3.3b). We expected a 50% abundance call for each clade in files containing two genomes, a 25% call in files containing four genomes, and a 12.5% call in files containing all eight genomes. Slight differences in the estimated and expected read abundance for some clades were observed (Figure 3.3b), which can likely be attributed to the low frequency of mutations in some strains and therefore absence of mutations in some randomly generated reads. Thus, Altob is well suited to identify and distinguish between ToBRFV clades with the exception of clade 5; however, abundance scores should be treated as close estimates and not as exact measurements.

Finally, Altob was used to evaluate the clades present in our WWI samples (Figure 3.3c). In samples A–E, clades four and seven were identified in all RNA Shotgun and ToBRFV-Seq samples, demonstrating the wide-spread transmission of specific ToBRFV-clades in Ontario

WWI. Clade four strains have been identified in North America; as such, the identification of clade four strains in Ontario wastewater influent is unsurprising [232]. Clade seven strains have been identified in the Netherlands and Belgium, suggesting the transmission of these strains to Ontario either through the agricultural industry and/or the import of tomato products [232,252,255]. In each sample, the predicted clade abundance of clade 7 was relatively similar regardless of library preparation method, with an average difference of $\mp 1.86\%$ between RNA shotgun and ToBRFV-Seq estimates for each respective sample. However, clade 4 yielded different abundance calls when samples were prepared by each assay and had an average difference of $\mp 26.92\%$.



Figure 3.3. *ToBRFV Clade-Abundance Estimates of Synthetic Read Datasets and WWI samples with Altob.* Heatmaps display Altob clade abundance estimates of (A) datasets containing 1000, 10,000, or 100,000 synthetic reads from eight ToBRFV genomes representative of each clade. (B) Datasets combining 100,000 synthetic reads from either all eight genomes, four genomes, or two genomes. (C) Reads from WWI samples prepared with RNA shotgun (Shotgun) or our ToBRFV-Seq assay (ToBRFV) (A–E).

3.3.6. Assessment of Viral Shotgun Reads at a Species Level

To investigate the presence and abundance of viruses in WWI, viral reads from RNA shotgun samples were taxonomically classified at the species level and the abundance of species

assigned read pairs was evaluated (Figure 3.4). In all five WWI samples, twenty-six viruses were consistently found, which included plant-infecting viruses, human pathogens, enteric bacteriophages, insect viruses, and amoeba-infecting viruses (Supplementary Table 3.4).

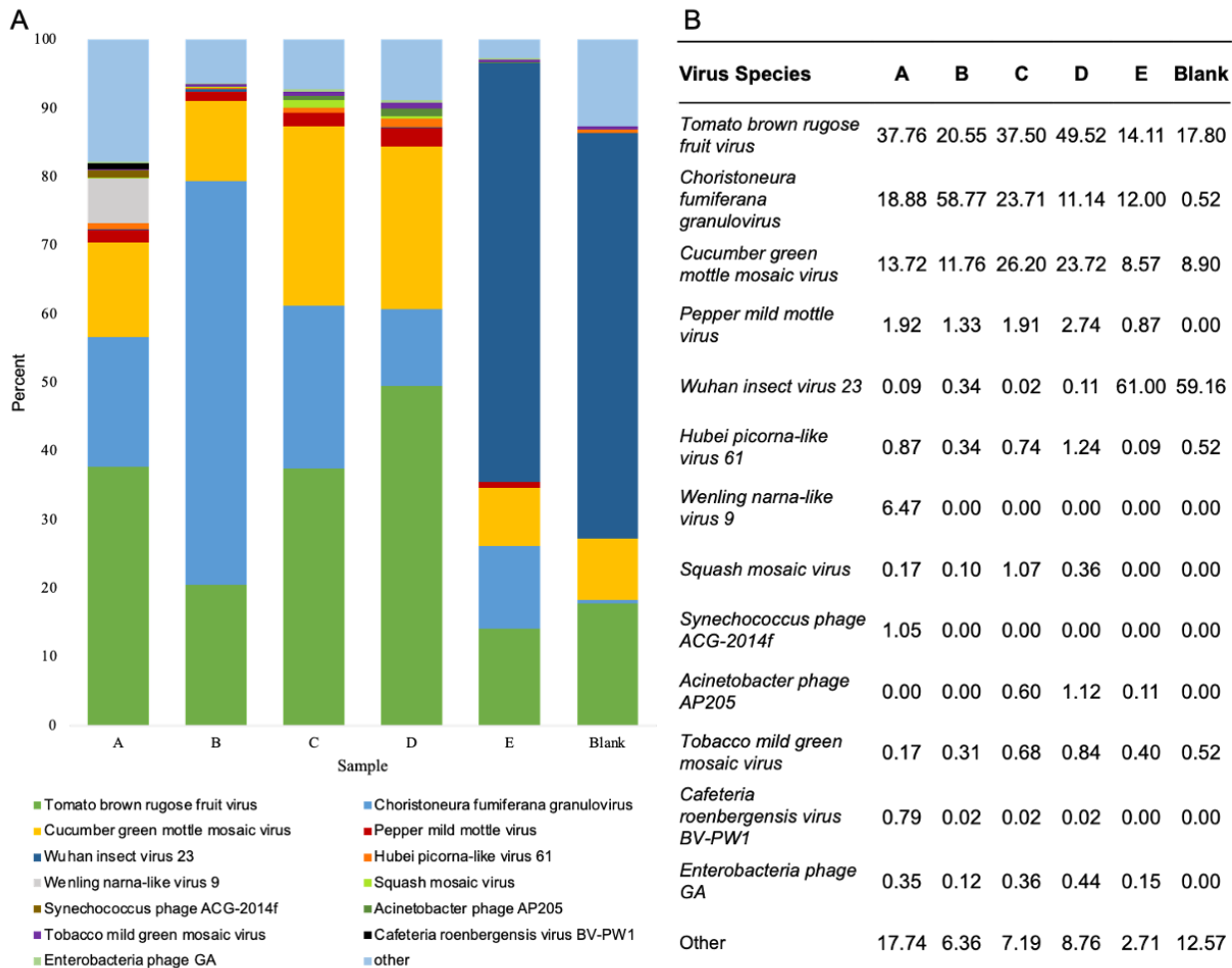


Figure 3.4. Evaluation of Virus-Species Abundance in WWI Samples using Counts of Viral Reads Taxonomically Classified by Kraken2. (A) Percentage of virus paired read counts of the seven most abundant virus species found in each sample and count of all other species. (B) Percentage of virus species classified reads assigned to the most abundant species in each WWI sample prepared by RNA shotgun and a blank. The total number of virus read pairs identified in samples A, B, C, D, E, and Blank were 1 144, 4 150, 6 477, 4 748, 24 578, and 191, respectively.

The counts and percentage of paired reads assigned to the seven most prevalent virus species identified in each sample were compared to ascertain which species were consistently found in high abundance (Figure 3.4). In samples A–D, either ToBRFV or *choristoneura fumiferana* granulovirus (ChfuGV) read pairs were in the highest abundance while cucumber green mottle

mosaic virus (CGMMV) and pepper mild mottle virus (PMMoV) were on average the third and fourth most prevalent virus species, respectively (Figure 3.4b). ToBRFV, CGMMV, and PMMoV are plant pathogens belonging to the *Tobamovirus* genus, while ChfuGV is an insect pathogen that infects budworms [228,271]. In sample E, however, Wuhan insect virus 23, was the most abundant single virus species detected, which has been identified in metagenomic studies and is thought to have an insect host [272]. Preparation or sequencing biases may have resulted in the overestimation of this virus in samples or for some unknown reason this virus may be found at this specific collection site. Replicative data at this site could be used to discern between these and other potential possibilities.

3.4. Discussion

3.4.1 Evaluation of ToBRFV-Seq Assay Robustness and Specificity

ToBRFV-Seq is a whole genome sequencing enrichment assay that uses a tiled amplicon primer scheme for specific PCR amplification. We evaluated our ToBRFV-Seq assay specificity using RNA extracted from five wastewater influent samples. We assessed primer specificity by first performing in silico primer binding analysis to ToBRFV genomes and other closely related tobamovirus genomes (Table 3.1). In silico, all primers were capable of binding the genomes of all 125 ToBRFV strains that were analyzed while relatively few off-target binding sites to other tobamovirus genomes were identified. We then evaluated primer specificity by taxonomically identifying reads obtained from five WWI RNA extracts prepared with our ToBRFV-Seq assay. The vast majority were taxonomically identified as ToBRFV reads and very few reads, 0–0.0017%, were assigned to other tobamoviruses (Figure 3.2g). Furthermore, >99% of ToBRFV-Seq reads mapped to the reference genome NC_028478.1, covering >99% of the target genomic region and >95% of the total genome (Table 3.3). Thus, demonstrating our primers are highly specific to and efficient in enriching ToBRFV genomes. When evaluating our ToBRFV-Seq assay we also prepared each WWI RNA extract for RNA shotgun sequencing. ToBRFV was detected in all RNA shotgun samples; however, a far greater depth of sequencing was required to obtain a modest number of ToBRFV reads when compared to our ToBRFV-Seq assay (Table 3.3). Thus, our assay provides an effective sequence enrichment method that is highly specific to ToBRFV genomes. The high relative abundance of ToBRFV-specific reads in RNA shotgun samples indicates that ToBRFV is widespread and highly prevalent in Ontario WWI (Figure 3.4). Moreover, comparing counts of ToBRFV-specific reads versus all other tobamovirus reads with a one-tailed t-test ($p < 0.05$, $n = 5$, $df = 4$, t-values, t-value = 3.569, critical t-value ∓ 2.776) demonstrates a statistically significant difference in read counts, with ToBRFV reads outnumbering all other tobamovirus reads combined (Figure 3.2d). The higher abundance of ToBRFV reads than

PMMoV and all other tobamoviruses is consistent with the findings of Rothman and Whiteson, 2022, who investigated WWI from Californian treatment plants for the presence of eight tobamoviruses using qPCR and Illumina sequencing. Of the eight tobamoviruses viruses, their analysis identified the ToBRFV as the most abundant in California WWI, even more prevalent than both PMMoV and CGMMV [252].

3.4.2. Possible Assay Applications to Determine ToBRFV Transmission and Prevalence

Wastewater influent has become an important tool in tracking the transmission of specific virus lineages [122,273]. Due to the fact that the ToBRFV is prevalent in wastewater, sequencing of WWI could be used to track and identify currently circulating ToBRFV lineages. To this end, we adapted our SARS-CoV-2 lineage abundance estimation tool, Alcov, to estimate ToBRFV lineages from sequencing reads. Generally, Altob is a robust prediction tool capable of estimating ToBRFV clade abundances in samples with as few as 1000 ToBRFV reads (Figure 3.3). However, some differences in the estimated clades when using differing library preparation methods demonstrate that Altob-predicted clade abundances should be treated as general estimates. The identification of additional ToBRFV strain genome sequences, particularly in clades 6 and 8 which each contain only three known strains, would enable more clade-specific mutations to be identified and added to Altob constellation files, thus improving Altob's ability to classify samples by ToBRFV clade. Together, our ToBRFV-Seq assay and Altob could be a valuable resource to identify clade-specific ToBRFV infections, explore potential transmission routes, and aid in combating transmission events. PMMoV is a commonly used human fecal indicator due to its reliable presence in human feces and has become a popular RT-qPCR standard to normalize cycle threshold (Ct) values from SARS-CoV-2 assays [255,273–275]. Since the SARS-CoV-2 pandemic, a multitude of virus and chemical human fecal indicators, such as PMMoV, caffeine, and nicotine, have been recently evaluated [276–278]. However, a consensus on the best standard has not yet been reached and the reliability of PMMoV and other standards is quite

variable between studies due to factors such as population size, seasonal changes, diet, sample processing, and detection methods [275,278–282]. Although PMMoV was found in all our RNA shotgun samples, it was consistently found in a lower abundance relative to ToBRFV, suggesting ToBRFV may be more readily detectable and thus be a more robust fecal indicator and RT-qPCR standard than PMMoV (Figure 3.4). Indeed, two studies examining human infant oropharyngeal and fecal viromes within the first year of life identified the ToBRFV by metagenomic sequencing in infant fecal [283] and infant oropharyngeal [284] viromes. Moreover, a recent comparative study of PMMoV and ToBRFV as fecal contamination indicators showed that ToBRFV is more prevalent in wastewater and matched the established microbial tracking marker cross-assembly phage (crAssphage) abundance [285]. This suggests that ToBRFV may be a suitable marker replacement for PMMoV in wastewater. ToBRFV has likely been overlooked as a fecal indicator because PMMoV was first introduced as a fecal indicator in 2006, long before the first recorded ToBRFV outbreak in 2014 [97,255,286].

3.4.3. Effectiveness of and Possible Improvements to Viral Capture and ToBRFV-Seq Procedures

Microbiome A particles, manufactured by Ceres Nanosciences, are magnetic capture beads that offer a fast and easy solution to isolate numerous virion species from wastewater, including SARS-CoV-2, influenza, respiratory syncytial virus, PMMoV, norovirus, and others [122,287,288]. We used Microbiome A particles to recover ToBRFV sequences from all of our WWI samples (Figure 3.2) and to our knowledge, this is the first demonstration of ToBRFV capture and concentration using Microbiome A particles. Virus sequence yields in samples concentrated with Microbiome A particles provide a rapid and user-friendly method to isolate ToBRFV and other viruses for targeted sequencing methods. However, Microbiome A beads may not be the best option for virus metagenomic RNA shotgun sequencing, due to the high contamination of bacterial DNA sequences. Thus, when performing viral metagenomic analysis, a combination of ultracentrifugation and filtration or other robust viral isolation methods should be used [289,290].

3.5. Conclusions

To aid in ToBRFV detection and ongoing surveillance efforts, we designed and developed a ToBRFV genome tiled amplicon sequencing assay that is robust, specific, and efficient. Moreover, we designed a computational tool, Altob, that estimates the relative abundance of ToBRFV clades in Illumina paired reads. Additionally, we demonstrate that Microbiome A particles are an effective method for isolating ToBRFV from WWI. Lastly, we demonstrate that ToBRFV is widespread and one of the most abundant RNA virus species in Ontario, Canada wastewater influent. Due to its stability and abundance in WWI, the ToBRFV has the potential to be a useful indicator species in wastewater and/or used as an RT-qPCR standard to quantify viral content in wastewater and more broadly, used in other wastewater epidemiology efforts and human fecal studies; however, this would require further study [285]. Moreover, due to the transmissibility of several tobamovirus species in reclaimed wastewater, including PMMoV [250], the closely related ToBRFV may also be transmissible in reclaimed wastewater used in crop irrigation, which could facilitate transmission of active ToBRFV virions to crops causing viral outbreaks.

Chapter 4. Hybrid Sequencing, Annotation, and Analysis of a Novel Algal Virus

Genome—Chrysochromulina parva virus BQ1 (CpV-BQ1)

4.1. Introduction – Eukaryotic Algal Viruses of the Phylum Nucleocytoviricota

Research into viruses infecting aquatic microbes has increased in recent years. Many aquatic viruses are Nucleocytoplasmic large DNA viruses (NCLDV), a diverse group of eukaryotic viruses [59] that were recently classified under the viral phylum *Nucleocytoviricota*. The *Nucleocytoviricota* includes the class *Megaviricetes* with three orders and numerous families including the *Mesomimiviridae* and the *Phycodnaviridae* [291]. Viruses belonging to the class *Megaviricetes* are characterized by large double-stranded DNA genomes, typically over 100 kb in size [292], an ability to replicate in both the host's nucleus [293,294] or cytoplasm [292,295], and the presence of a core set of shared genes, including those encoding the major capsid protein, DNA polymerase B, and others [292]. Viruses in the *Megaviricetes* are also notable for their complexity and size, with some members known as "giant viruses" having genomes and virions that rival those of small cellular organisms [295–299]. They infect a wide range of eukaryotic hosts, from algae to animals, and are found in various environments, particularly aquatic ecosystems [296]. They encode genes for DNA replication, transcription, and repair [292], have an extensive genomic plasticity and diversity [292], have acquired genes through lateral gene transfer from various sources [296], and have the potential to significantly alter host metabolism during infection [296].

In addition to their unique genomic characteristics, the *Megaviricetes* play significant ecological roles in various ecosystems, particularly in aquatic environments. They can affect the population dynamics of their unicellular hosts, including controlling algal blooms. For example, *Heterosigma akashiwo* virus (HaV) influences seasonal harmful algal blooms in marine environments [45]. Moreover, viruses in the *Megaviricetes* contribute to biological carbon export from marine

surfaces to deep layers through host-cell death [40,300]; can impact nitrogen metabolism and fermentation processes in their hosts [301]; and can alter eukaryotic community structures, particularly in marine environments, by infecting various eukaryotic lineages [300,302–305]. Many of these viruses are involved in metabolic reprogramming by encoding genes involved in nutrient uptake, light harvesting, and central carbon metabolism, allowing them to reprogram host metabolism during infection [296]. The latter includes genes for photosynthesis, diverse substrate transport, and light-driven proton pumps [59]. Indeed, by infecting and lysing their hosts, many *Megaviricetes*, such as viruses that infect the massive bloom-forming alga *Gephyrocapsa huxleyi*, drive the release of organic matter into the environment, contributing to nutrient cycling in ecosystems [40,44,300,306,307]. Collectively, these ecological roles highlight the importance of *Megaviricetes* in shaping ecosystem dynamics, particularly in aquatic environments, and their potential impact on global biogeochemical cycles.

An understudied aquatic microbial eukaryotes that is infected by viruses that are likely to be classified within the *Megaviricetes* is *Chrysochromulina parva* (*C. parva*), a freshwater haptophyte alga [46,308]. It is a small unicellular organism, typically 4-6 µm in size [309] with two flagella (~8 µm long) and a long haptonema (up to 10 times the body length) [309]. *C. parva* has a deep groove running the length of the cell, from which the flagella and haptonema emerge, contains two chloroplasts with internal pyrenoids, each associated with a large lipid body, and has a simple cellular morphology with a eukaryotic nucleus, mitochondria with tubular cristae, and a Golgi apparatus. Unlike most *Chrysochromulina* species, *C. parva* lacks visible scales on its cell surface or within the Golgi apparatus [309].

Viruses capable of infecting and lysing *C. parva* were first isolated in 2015 from a Lake Ontario water sample. This sample, collected from the Bay of Quinte, contained a filterable, heat labile, chloroform sensitive lytic agent that was capable of completely lysing cultures of *C. parva* [47]. PCR primers designed to amplify gene fragments from a range of viruses within the

Phycodnaviridae family were used to amplify DNA polymerase B (*polB*) and major capsid protein (MCP) gene fragments [47]. Sanger sequencing of the resulting amplicons led to the identification of a single DNA *polB* gene fragment and 9 unique MCP fragments. Based on the *polB* phylogeny, the lytic agent was named *Chrysochromulina parva* virus BQ1 (CpV-BQ1) and was putatively classified as a phycodnavirus; the presence of multiple MCP fragments allowed the researchers to speculate that more than one type of virus might have been isolated from this water sample [47]. Further culturing, isolation, and high-throughput sequencing experiments led to the assembly of a giant, 437 kb, virus genome encoding 503 open reading frames (ORFs), as well the ~23 kb genomes of three polinton-like viruses (PLVs), which are viruses similar to virophages and need a helper virus for their replication. However, the CpV-BQ1 *polB* gene fragment initially recovered by PCR was not found within this giant virus genome or the PLV genomes, indicating that an additional virus of *C. parva*, named CpV-BQ2, was present in the Bay of Quinte water sample and had been co-cultured along with CpV-BQ1 [57]. Here we report the second genome sequence of a *C. parva* virus noting that this genome assembly was derived from CpV-BQ1 because it encodes an identical *polB* gene as that which originally led to the classification of CpV-BQ1 as a member of the *Phycodnaviridae*. To study this virus's genome characteristics, we implemented a two-step approach for its sequencing, which combined short read Illumina and long read Nanopore sequencing, coupled with PCR-enabled molecular analysis.

4.2. Materials and Methods

4.2.1. Host and Virus Cultivation

Viruses infecting *Chrysochromulina parva* strain CCMP 291 (National Center for Marine Algae and Microbiota, East Boothbay, Maine, USA) were originally isolated from an embayment of Lake Ontario, Canada, in 2011 [47] and have been regularly propagated in the laboratory since then. The CpV-BQ1 virus was purified from cultures of mixed *C. parva* viruses, as described in Stough et al. (2019), via an end-point dilution approach. Briefly, serial 10-fold dilutions of *C. parva* lysates were inoculated into 96-well microtiter plates with mid-log phase *C. parva* cells at a concentration of approximately 6.0×10^5 cells/mL as determined using a hemocytometer and light microscope. Medium from individual wells that lysed at the highest dilution level (i.e., lowest concentration) of viruses were transferred into 50 mL cultures of mid-log *C. parva* and the resulting lysates were filtered through sterile 0.22 μ m pore-size PVDF Durapore® membrane filters (EMD Millipore, #GVWP00010). This process was repeated three times to ensure that only a single type of CpV was propagated, and the presence of CpV-BQ1 was confirmed throughout this purification process using the qPCR method described in Mirza et al. (2015).

4.2.2. Nucleic Acid Extraction

Following purification of CpV-BQ1, genomic material was prepared for sequencing by extracting nucleic acids from a 600 mL lysate of a *C. parva* culture infected with CpV-BQ1. This lysate was filtered through a 0.22 μ m pore-size Steritop® disposable bottle top filters (MilliporeSigma, #S2GPT10RE) and was concentrated via ultracentrifugation at 31,000 rpm for 1 hour at 20°C in a SW32Ti rotor (Beckman Coulter Life Sciences). The supernatant was decanted and pelleted material was resuspended in 10 mM Tris-Cl, pH 8.5. Nucleic acids were extracted from the concentrated viral lysate using a Maxwell® RSC Viral Total Nucleic Acid extraction kit (Promega, #AS1330) following the manufacturer's protocol for 300 μ L of sample input. The DNA

concentration measured using a Qubit fluorometer with a dsDNA HS kit (*ThermoFisher Scientific*, #Q32851) was 26.2 ng/μL.

4.2.3. Illumina and Nanopore library preparation and sequencing

The DNA was sent to SeqCenter (Pittsburgh, USA) for paired-end whole genome shotgun sequencing. Sample libraries were prepared using the Illumina DNA Prep kit and IDT 10 base pair UDI indices, and were sequenced on an Illumina NovaSeq 6000, producing 2 x 151 base pair reads. Demultiplexing, quality control, and adapter trimming was performed with BCL-Convert v4.0.3, generating 14,757,009 read pairs. The DNA was also prepared for Oxford Nanopore sequencing at the University of Waterloo. Extracted DNA was diluted to a concentration of 5 ng/μL and prepared with the Nanopore Rapid PCR barcoding 24 V14 kit (*Oxford Nanopore*, #SQK-RPB114.24). The prepared DNA library was sequenced on a MinION Mk1B device using an R10.4.1 flow cell (*Oxford Nanopore*, #FLO-MIN114). All raw reads are available on the SRA database under BioProject accession number PRJNA1199504.

4.2.4. Read processing and Genome Assembly

Nanopore sequencing produced a total of 2,325,374 raw reads. First, Filtlong was used to remove reads less than 1000 bp long and 10% of the worst quality reads, resulting in 1,879,244 reads [310]. Kraken2 v2.0.7-beta was then used to filter out classified reads, removing contaminating host, bacterial, and human reads. The standard Kraken2 database v9/26/2022 was used with a 0.001 confidence filter and the unclassified-out flag to generate a separate file of unclassified reads, from which 1,048,312 reads were obtained [260].

Filtered long reads and Illumina reads were then used for assembly with the TryCycler pipeline v0.5.4 [311]. Briefly, Filtlong was used to keep the best 80% of filtered nanopore reads. This final read filtering step yielded 796,152 reads with an N50 of 6,479bp and a mean length of 6,352 bp

[310]. Long reads were then subsampled to create 24 read subsets with approximately 8,645 reads per subset. An estimated genome size of 400,000 bp was used to subset reads, which was based on the size of the CpV-BQ2 virus genome (accession MH918795). Each read subset was assembled with either Flye v2.9.2-b1786, Miniasm & Minipolish v0.1.2, Raven v1.8.1, or Canu v2.2 [312–315]. Of the 24 read subsets, 6 were assembled into contigs by each respective tool. From these 24 genome assemblies, 26 contigs were produced.

Assembled contigs were clustered to determine the MASH distance between assemblies. The MASH distance is based on the Jaccard index and provides a measure of similarity and diversity between each sample. Using MASH distances, a phylogenetic tree was generated during the TryCycler clustering step, of the 26 assembled contigs, 23 clustered closely together. These assembled contigs were reconciled, which ensures contigs are on the same strand and sufficiently similar for downstream assembly. To ensure contigs were sufficiently similar, the TryCycler default thresholds were used, including a 98% sequence identity score and minimum 1-kbp identity score of 25%. Based on these thresholds, 7 of the 23 contigs were discarded. A dotplot was generated to visualize the topology of the remaining 16 contigs.

A multiple sequence alignment was generated from the remaining 16 clustered and reconciled contigs with the MUSCLE algorithm [264]. Long reads were partitioned to each assembly to determine the single best alignment for each read. Then, a linear consensus sequence was generated based on the best read alignments. Long-read correction with the Nanopore tool Medaka v1.11.1 was used to polish the consensus sequence [316,317]. Short Illumina read pairs were quality filtered using Fastp v0.23.4. Then, two iterations of polypolish v0.5.0 were used to correct genome errors using the filtered Illumina reads and generate the final genome assembly [206,318]. Alignment of Illumina reads to the assembled genome produced a mean read depth of 21,688.3x, corrected 54 positions in the genome, and identified a 104 bp region at the C-terminus that had 0% coverage. This 104 bp region was removed from the final assembly and likely arose

from erroneous read extension. The final assembly resulted in a genome length of 165,454 bp and was published under accession number PQ783904.

4.2.5. Primer Design, PCR Amplification, and Analysis of Random Genomic Regions

The CpV-BQ1 assembly was verified by PCR amplification and sequencing of five regions, approximately 9,000 bp in length, across the genome. To design primers, the CpV-BQ1 genome was uploaded to PrimalScheme and an amplicon size of 9,000 bp was selected [120]. PrimalScheme generated 19 potential primer pairs that could be used across the entire genome, of which five were selected to perform random spot checks across the genome (Supplementary Table 4.1). An additional four primers were designed to sequence across the genome ends to investigate the genome topology. PCR amplification was performed with 2x GB-AMP PaCeR HP Master Mix (*GeneBio Systems*, #PCR-002-01), 0.4 μ M of forward and reverse primers, and 1.1 ng of genomic DNA. Thermocycling conditions were: initial denaturation at 95°C for 30 seconds, 35 cycles of 95°C for 15 s, 68°C for 15 s, and 72°C for 10 mins, then a final extension at 72°C for 6.5 min and samples were held at 12°C. PCR products were analyzed by gel electrophoresis on a 1% agarose gel and concentration was evaluated with a Qubit 4.0 fluorometer. PCR fragments were cleaned up with AMPure XP beads (*Beckman-Coulter*, #A63881) to remove primers and small DNA fragments. Recovered products were prepared using the Nanopore Rapid Barcoding kit V14 (*Oxford Nanopore*, #SQK-RBK114.24). The prepared library was loaded onto a MinION R10.4.1 flowcell (*Oxford Nanopore*, #FLO-MIN114) and sequenced on a Nanopore MinION. Reads were analyzed using the Trycycler pipeline as previously described with a few exceptions. Reads were filtered once using filtlong, twelve read subsets were generated using a genome size of 9,000 bp, and only one round of polishing was performed with polypolish. Assembled contigs (Supplementary Table 4.2) were aligned to the CpV-BQ1 genome using the MUSCLE algorithm and visualized with Seaview [264,265]. Raw reads are available on the SRA database under accession PRJNA1199504.

4.2.6. Open Reading Frame (ORF) Prediction

Open reading frames (ORFs) were predicted using a combination of four prediction tools: GeneMarkS, Glimmer3, Prodigal, and FragGeneScan [319–322]. Any ORF predicted by two or more tools was considered a putative gene encoding ORF [323–325]. When differing gene start sites were identified, the positions identified with GeneMarkS were used first, followed by Prodigal, then FragGeneScan [324]. A total of 193 ORFs were predicted, of which GeneMarkS, Prodigal, FragGeneScan, and Glimmer3 predicted 98.45%, 95.85%, 88.08%, and 97.4% of these ORFs, respectively (Supplementary Table 4.3).

4.2.7. Gene Annotation

ORFs were assessed for functional predictions using multiple approaches, including database searches, InterPro analysis, and HMM classification. A BLASTp search against the non-redundant protein sequence database was performed twice, once including all organisms and a second search against only viruses [326]. Additionally, a search was performed against the IMG virus protein BLAST database [327]. InterPro was used to predict protein functionality using protein domain and family predictions [328]. InterPro has multiple tools integrated into its online platform including Panther, NCBIfam, CDD, Cath-Gene3D, SUPERFAMILY, ProSiteProfiles, ProSitePatterns, PRINTS, SMART, Pfam, HHMAP, and FunFam to classify protein families, domains, superfamilies, and amino acid sites. InterPro also identifies signal peptides, transmembrane helices, coils, and disordered regions using TMHMM, Coils, MobliteDB, and Phobius. HMM profile analysis against the PhROGs, efam-xc, RVDB, and VOGDB were performed to classify ORFs [329–332]. MMSeqs was used to perform HMM analysis against the PhROGs database while HMMer3 was used to search all other HMM databases [333,334]. tRNAscan-SE was used to predict tRNA sequences in the assembled CpV-BQ1 genome [335].

4.2.8. Taxonomic classification

Three genes common to all NCLDV were used to taxonomically classify the CpV-BQ1 genome, DNA polymerase type-B, viral A32-like packaging ATPase, and viral late transcription factor 3 (VLTF-3). Protein sequences and accessions for each gene from various NCLDV were downloaded from NCBI (Supplementary Table 4.4). Multiple sequence alignments (MSAs) were generated using the MUSCLE algorithm for each protein using Seaview v5.0.4 [264,265]. MSAs were used to generate phylogenetic trees in Seaview using the phylogenetic maximum likelihood method with the LG model, aLRT (SH-like) branch support, NNI tree searching operation, and a BioNJ starting tree [265,336].

4.3. Results and Discussion

4.3.1. Assembly of the CpV-BQ1 Genome

A hybrid assembly approach with long Nanopore and short Illumina reads was used to determine the CpV-BQ1 genome. Long read assemblies can produce larger contigs and resolve repetitive genome regions with higher accuracy than short Illumina read assemblies. However, Illumina reads have higher sequence accuracy than long reads and can correct for errors introduced through long read sequencing. Thus, using a combination of long and short reads for assembly produces a genome with higher accuracy and instills greater confidence in the resulting assembly [311,337]. Hybrid assembly of the CpV-BQ1 genome was performed using the TryCycler pipeline [311]. TryCycler is a robust tool that uses a unique approach to produce assemblies with a high degree of confidence. Most long-read assembly tools work fairly well, however, they are not perfect and can introduce large- and small-scale errors into an assembled genome that often go undetected [311]. To eliminate these issues, TryCycler uses multiple separate long-read assemblies, generated by a variety of assembly tools, as its input and produces a consensus sequence. By using a variety of assembly tools that utilize different approaches/algorithms, bias and/or errors introduced by any one tool can be detected and eliminated [311].

Cleaned and filtered long reads were divided into 24 subsets. Then, six subsets were assembled with Flye, Minipolish, Raven, and Canu, respectively [312–315]. A total of 26 contigs were assembled, of which 23 formed a single phylogenetic cluster (Figure 4.1). After reconciling these assemblies, 16 contigs showed high sequence similarity. This included five Flye, and Minipolish assemblies, four Raven assemblies, and two Canu assemblies (Figure 4.1). The similarity between the 16 independently generated contigs provided confidence that the assemblies were of very high quality and representative of the true CpV-BQ1 genome. These 16 contigs were used to generate an MSA, reads were partitioned, then a consensus genome was generated and

polished with Illumina reads producing a 165,454 bp genome. Additionally, attempts to circularize the genome during reconciling were unsuccessful. Analysis of these contigs using a dotplot shows the same discrete start and end point in all 16 contigs, which indicates the genome has a linear topology (Figure 4.2). If genomes had a circular topology, we would expect the start and stop sites to vary amongst the 16 assembled genomes, causing the lines within each dotplot to start along the axis edge instead of the corner. Additionally, if some of our assemblies contained gaps or large rearrangements the plots would contain lines with gaps or a discontinuous arrangement. Our dotplot indicates these 16 assemblies are not circular and do not contain gaps or large rearrangements (Figure 4.2).

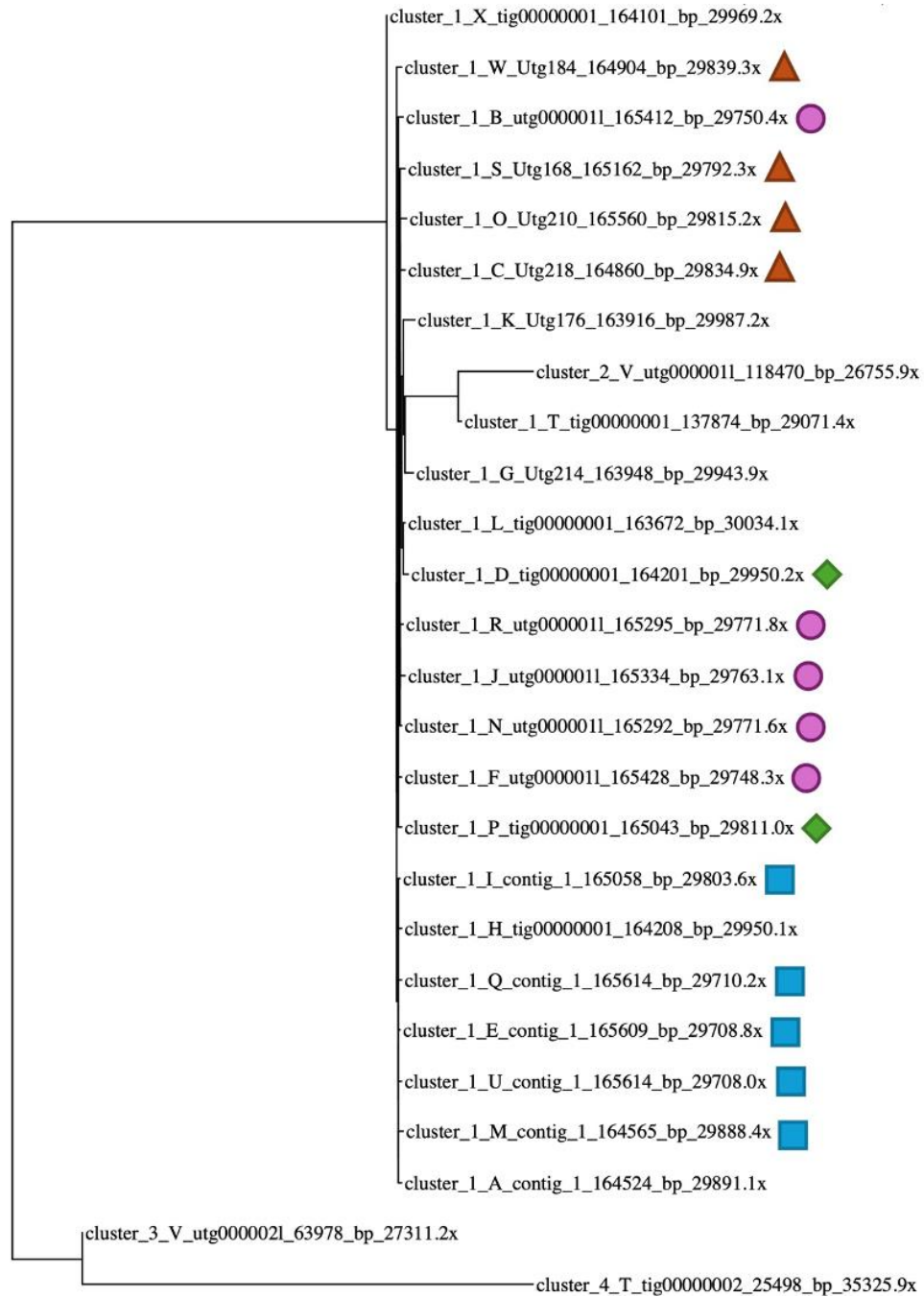


Figure 4.1. Phylogenetic Tree of Clustered Assembled Contigs. The TryCycler clustered assemblies based on the MASH distances between contigs displayed in a phylogenetic tree. Almost all assemblies fall within the cluster 1 branching structure. The 16 assemblies used for reconciling and MSA steps are labeled (e.g., blue squares are Flye assemblies, pink circles are Minipolish assemblies, orange triangles are Raven assemblies, and green diamonds are Canu assemblies).

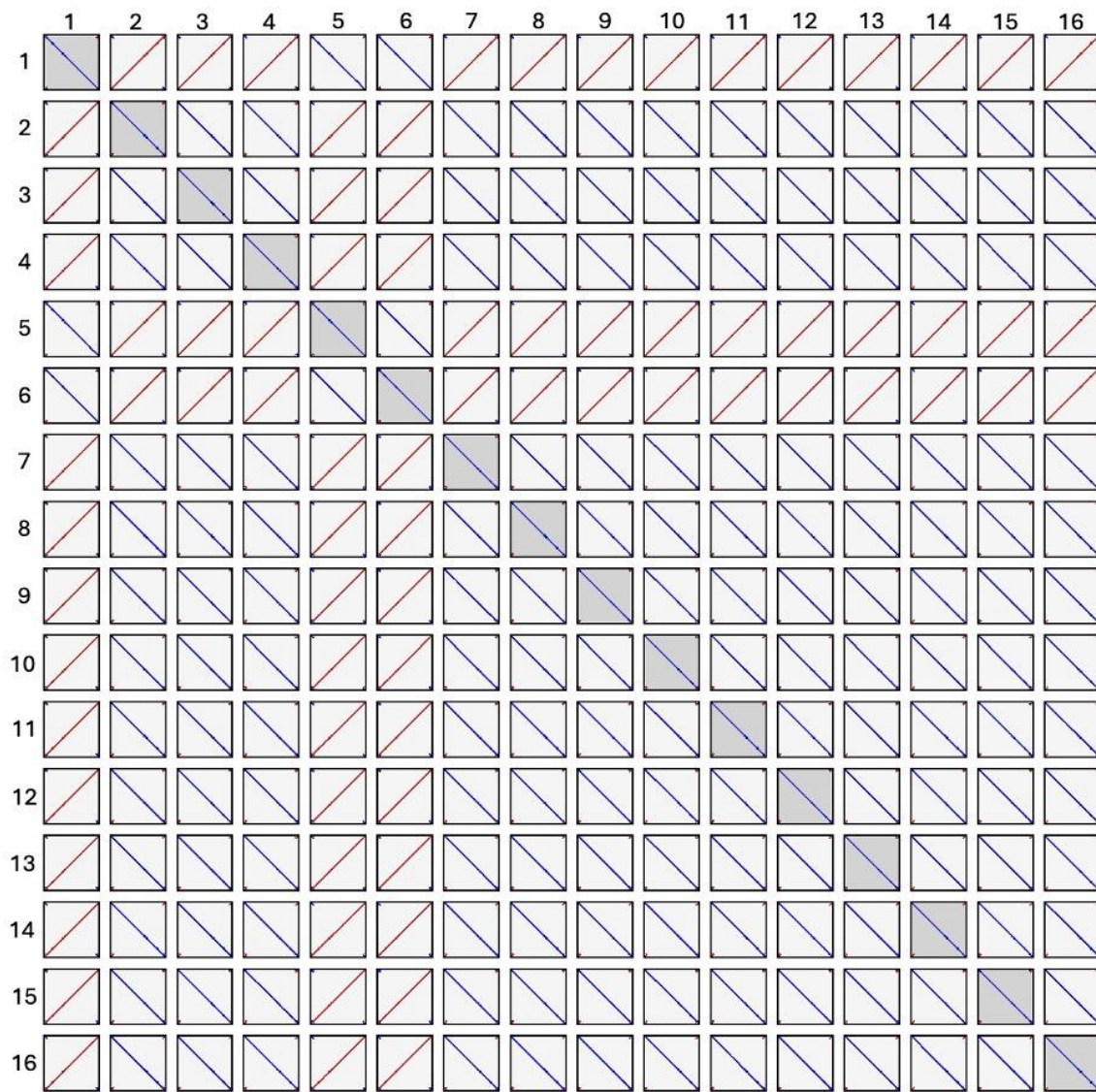


Figure 4.2. *Dotplot Analysis of 16 Assembled Contigs.* Squares are a visual comparison of all pairwise combinations of the 16 assembled contigs. Solid diagonal blue lines indicate the sequences have the same start and end sites, are highly similar, and do not contain any large gaps or rearrangements. The solid diagonal red lines indicate the sequences are the reverse of one another, are highly similar, and do not contain large gaps or rearrangements.

To validate the assembled sequence PCR amplifications, Nanopore sequencing and TryCycler assembly were performed to recover and verify the sequence of five 9,000 bp regions within the genome (Figure 4.3). Additionally, to verify the genomes linear topology, four PCR amplifications across the genome ends were attempted (Figure 4.3).

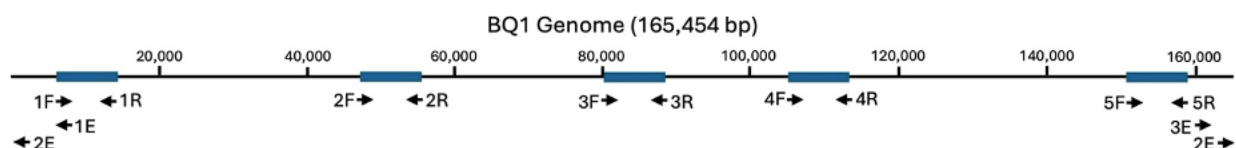


Figure 4.3. *Diagram of CpV-BQ1 Genome Primer Binding Sites.* The complete 165,454 bp CpV-BQ1 genome with the primer binding sites used for genome validation. Primer binding sites are represented by arrows specific to the direction of amplification. Primer pairs 1F&1R, 2F&2R, 3F&3R, 4F&4R, and 5F&5R used to amplify 9,000 bp regions in the genome indicated by blue rectangles. Primer pairs 1E&2E, 1E&3E, 2E&2E, and 2E&3E were used to PCR amplify across the genome ends (Supplementary Table 4.1).

The five ~9,000 bp regions were easily PCR amplified, sequenced, and assembled with TryCycler (Supplementary Table 4.2). Alignment of the assembled contigs to the CpV-BQ1 genome showed almost 100% sequence identity to their respective sites. The very ends of assembled PCR regions # two and # five contained one and 17 bases that differ from the CpV-BQ1 genome, respectively, which can be attributed to low read coverage in these areas. Overall, the similarity of PCR amplified sequences to the assembled genome provides further confidence in the assembly accuracy. Attempts to amplify genome ends to test for circularization were unsuccessful. Of these four attempted PCR amplifications (Figure 4.3), nanopore reads were obtained from three reactions. However, attempts at read assembly with TryCycler resulted in clusters with many spurious, incomplete, and misassembled contigs. Thus, the inability to amplify sequences that span the genome ends provides additional confirmation that the CpV-BQ1 genome has a linear topology.

4.3.2. *Characterization and Description of the CpV-BQ1 Genome*

Assembly of the CpV-BQ1 genome sequence produced a linear genome of 165,454 bp with a GC content of 32.32%. Three transfer RNA (tRNA) sequences and 193 coding sequences (CDSs) were identified in the genome that ranged from 44 to 1693 amino acids (aa) in length, with an average size of 260 aa (Figure 4.4, Supplementary Table 4.3). Coding sequences were found in slightly higher prevalence on the positive strand than the negative strand, occurring at a rate of

58.5% and 41.5%, respectively. Of the 193 identified CDSs, 92 (47.67%) were assigned putative functions based on homology to known genes and protein domains, while the remaining 101 (52.33%) could not be assigned functions and were designated as hypothetical proteins (Supplementary Table 4.3). Functionally annotated CDSs were placed into eight general functional groups (Figure 4.4, Table 4.1). A total of 19 CDSs were involved in DNA replication, recombination, and repair; eight in nucleotide metabolism and DNA packaging; 10 in transcription; three in sugar manipulation; four in DNA methylation; 15 in protein and lipid binding, synthesis, and modification; 12 in virion capsid and associated proteins, and 21 encoded miscellaneous functions (Table 4.1).

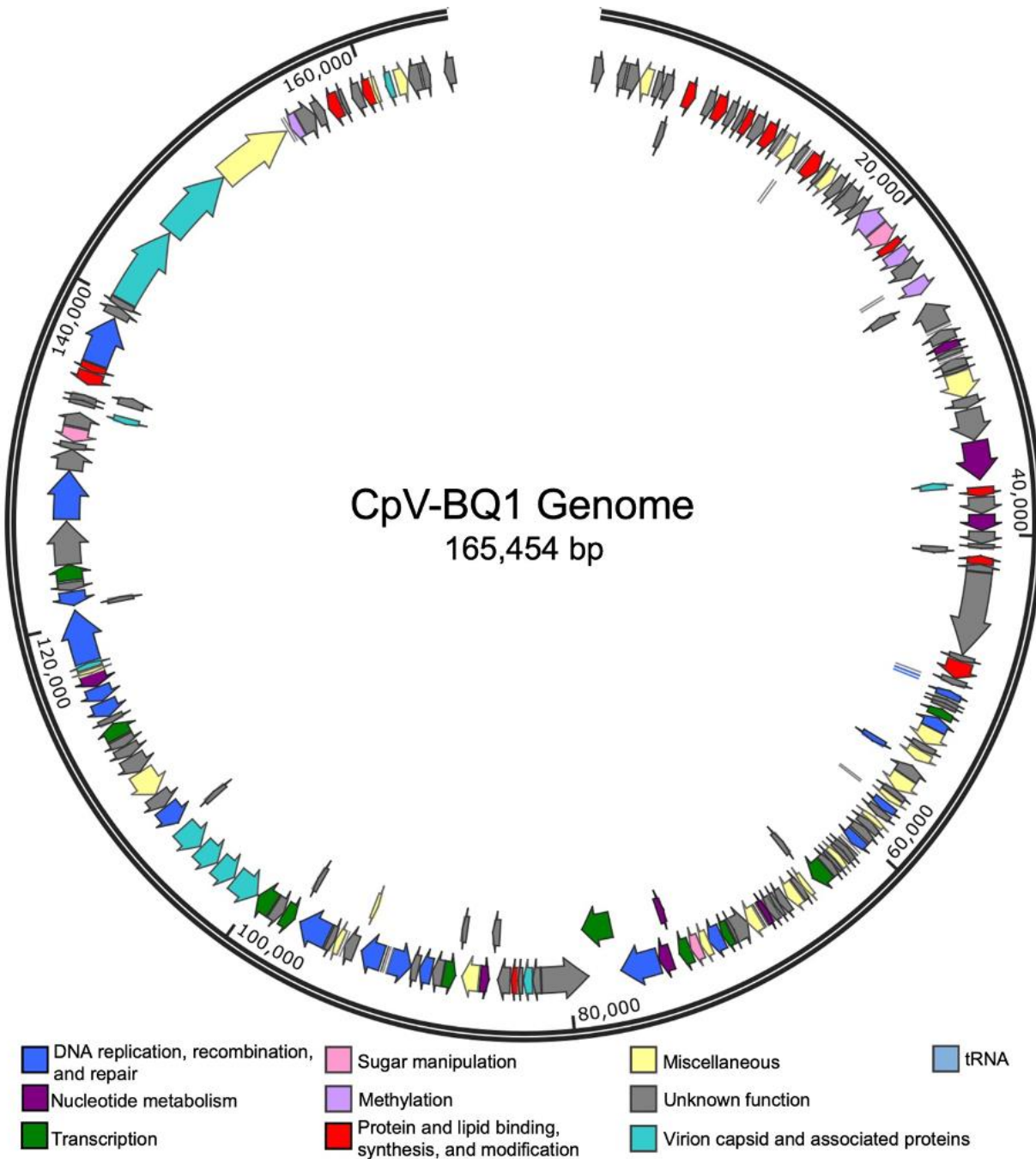


Figure 4.4. ORF Predictions in the CpV-BQ1 Genome. ORFs represented by arrows are aligned to the 165,454 bp CpV-BQ1 genome oriented in their coding direction. Arrows are coloured based on their hypothesized functional group. Blue arrows for DNA replication, recombination, and repair; dark purple for nucleotide metabolism and DNA packaging; green for transcription; pink for sugar manipulation; lavender for methylation; red for protein and lipid binding, synthesis, and modification; aqua for virion capsid and associated proteins; yellow for miscellaneous proteins; light blue for tRNA; and grey for all proteins with unknown functions. The genome is presented as circular for presentation purposes only. Nucleotide positions are denoted at every 20,000 bp. SnapGene Viewer was used to annotate and colour CpV-BQ1 ORFs.

Table 4.1. List of All Characterised Genes in the CpV-BQ1 Genome. All genes in the CpV-BQ1 genome with an identifiable function or domain architecture. Gene name, locus, as well as start and end position within the CpV-BQ1 genome are provided. These protein encoding genes are divided into functional groups including, DNA replication, recombination, and repair; nucleotide metabolism and DNA packaging; transcription; sugar manipulation; DNA methylation; protein and lipid binding, synthesis, and modifications; virion capsid and associated proteins; and miscellaneous.

Protein	Locus	Start	End
DNA and RNA replication, recombination, and repair			
yqaJ-like viral recombinase domain containing like-protein	BQ1_063	49504	49701
pyrimidine dimer DNA glycosylase, endonuclease V	BQ1_064	50107	49682
DNA polymerase family X	BQ1_068	52187	51462
GxxExxY protein	BQ1_072	54374	54754
Holliday junction resolvase, A22	BQ1_078	57463	57972
SWIB/MDM2-domain containing protein	BQ1_083	60100	60714
DNA topoisomerase I	BQ1_104	70386	71225
ATP-dependent DNA ligase	BQ1_110	74487	76832
proliferating cell nuclear antigen, PCNA	BQ1_124	88347	89114
VV A18-like helicase	BQ1_126	90943	89633
NTPase/helicase	BQ1_128	91438	92781
P-loop containing nucleoside triphosphate hydrolase	BQ1_134	94988	96880
putative DNA primase/helicase	BQ1_144	107635	106412
Flap endonuclease 1	BQ1_152	115062	114115
YqaJ-like viral recombinase domain containing protein	BQ1_153	116022	115118
DNA topoisomerase II	BQ1_157	117625	120840
exodeoxyribonuclease III	BQ1_159	121988	121146
Helicase	BQ1_163	126370	129303
DNA polymerase type-B	BQ1_174	135953	138937
Nucleotide transport and metabolism			
thymidine kinase	BQ1_040	28623	28096
ribonucleoside-diphosphate reductase large subunit	BQ1_048	34374	36686
ribonucleoside-diphosphate reductase small subunit	BQ1_052	38805	39767
deoxycytidylate deaminase	BQ1_098	67554	67210
deoxyuridine 5'-triphosphate nucleotidohydrolase	BQ1_108	73618	73169
thymidylate synthase ThyX	BQ1_109	73681	74451
NUDIX hydrolase	BQ1_119	85372	84851
Viral A32 protein	BQ1_154	116887	116066
Transcription			

RNA cap guanine-N2 methyltransferase family protein	BQ1_067	51398	50886
transcription factor IIB	BQ1_090	62855	63874
transcription elongation factor S-II	BQ1_103	69838	70341
ribonuclease III	BQ1_107	72453	73172
mRNA-capping enzyme	BQ1_111	76891	78837
MYM-type Zinc finger with FCS sequence motif-containing protein	BQ1_122	87611	86892
TATA-box binding protein	BQ1_136	97984	97247
late transcription factor VLTF3	BQ1_138	98765	99838
mRNA-capping enzyme	BQ1_150	112713	113603
transcription initiation factor TFIIB	BQ1_161	122685	123626
Sugar manipulation			
glycosyltransferase family 2 protein	BQ1_030	20111	21163
glycosyltransferase 17	BQ1_106	71804	72418
glycosyltransferase	BQ1_166	131902	131078
DNA Methylation			
DNA cytosine-5 methyltransferase	BQ1_029	20038	18737
FkbM family methyltransferase protein	BQ1_032	21810	22706
D12 class N6 adenine-specific DNA methyltransferase protein	BQ1_035	23949	24851
SAM-dependent methyltransferase	BQ1_181	155423	154812
Protein and lipid binding, synthesis, and modifications			
F-box domain containing protein	BQ1_008	5793	6443
F-box domain containing protein	BQ1_010	7749	8501
F-box domain containing protein	BQ1_013	9687	10196
F-box domain containing protein	BQ1_015	11116	11862
RING finger E3 ubiquitin protein ligase	BQ1_022	14225	15103
E3 ubiquitin-protein ligase	BQ1_031	21217	21729
translation initiation factor 4E	BQ1_050	37156	37707
Putative Metallopeptidase WLM domain containing protein	BQ1_056	41801	41286
DNAJ homolog subfamily A member	BQ1_060	47801	48868
E3 ubiquitin-protein ligase	BQ1_092	64318	64644
N1R/p28-like protein	BQ1_116	83104	83520
patatin-like phospholipase family protein	BQ1_172	135387	134623
WLM domain-containing protein	BQ1_173	135893	135411
F-box domain containing protein	BQ1_184	158155	157388
F-box domain containing protein	BQ1_187	160285	159638
Virion Capsid and Associated proteins			

ERV/ALR sulfhydryl oxidase protein	BQ1_049	37107	36658
Ac78 protein	BQ1_099	67798	67643
minor capsid P9 transmembrane helices containing protein	BQ1_114	82228	82785
putative major capsid protein A	BQ1_139	101585	99924
putative major capsid protein B	BQ1_140	102955	101687
putative major capsid protein C	BQ1_141	104372	103038
putative major capsid protein D	BQ1_142	106012	104462
Thioredoxin	BQ1_156	117244	117561
Glutaredoxin-like protein	BQ1_168	133267	132848
tail fiber protein containing WIAG-tail domain	BQ1_177	140154	145088
fibronectin type III domain containing tail fiber protein	BQ1_178	145167	149783
tape measure protein	BQ1_189	161545	161105
Miscellaneous			
RNA binding domain containing protein	BQ1_004	3644	2958
transmembrane protein similar to AMEV204	BQ1_018	12505	13218
Zinc-finger, ring type domain containing protein	BQ1_024	15552	16211
Diverse functionality domain-containing protein	BQ1_045	30081	31643
AAA family ATPase/CfxQ-like protein	BQ1_069	52224	53210
DUF5871 protein	BQ1_071	53711	54340
DUF5764 protein	BQ1_074	55548	56522
DUF5759 protein	BQ1_076	56932	57294
DUF5754 protein	BQ1_080	58923	58603
putative Mn2+ efflux pump MntP	BQ1_087	62104	61712
DUF5871 protein	BQ1_094	65085	65804
2OG-Fe(II) oxygenase	BQ1_100	67910	68626
DUF5872 protein	BQ1_105	71244	71759
DUF5767 protein	BQ1_120	85474	86505
DUF1599 protein	BQ1_129	93038	92778
DUF5760 protein	BQ1_132	94057	94470
IQ motif containing protein	BQ1_146	110246	108597
type II secretory pathway component PulF	BQ1_155	117159	116935
fibronectin type III domain containing transmembrane protein	BQ1_179	149818	154509
high mobility group protein	BQ1_188	160585	160349
RNA binding domain containing protein	BQ1_190	161789	162472

4.3.3. Taxonomy

Viruses within the *Megaviricetes* (previously and informally known as NCLDV) share a few common characteristics, including the presence of a dsDNA genome typically over 100 kb, which encodes a handful of universal NCLDV genes [292]. These genes include the major capsid protein (MCP), a DNA polymerase B (*polB*), the viral A32-like packing ATPase, and the viral late transcription factor 3 (VLTF-3) [292]. Additionally, although the D5-helicase is considered a universal NCLDV gene, it is not found in viruses in the *Phycodnaviridae* family, nor in the CpV-BQ1 genome [292]. Of the universal genes, the MCPs share little sequence identity, and thus cannot be used reliably for phylogenetic analysis. Phylogenetic analysis of the CpV-BQ1 genome was therefore performed using *polB*, viral A32-like packing ATPase, and VLTF-3 (Figure 4.5).

In all three phylogenetic trees, the CpV-BQ1 groups closely with viruses in the *Phycodnaviridae* family (Figure 4.5) such as the Heterosigma akashiwo virus 01, a member of the *Phycodnaviridae* family and the single member of the genus *Raphidovirus*. Within the DNA polymerase B phylogenetic tree CpV-BQ1 groups closest to two viruses within the genus *Prymnesiovirus* [305], Chrysochromulina brevifilum virus PW1 (CbV-PW1) [338] and Phaeocystis globosa virus 08T (PgV-08T) [339]. Molecular analysis of the A32 ATPase and VLTF-3 genes in CbV-PW1 and PgV-08T has not been performed, thus we could not include them in our phylogenetic analysis. Nonetheless, based on the similarity of the *polB* gene and the fact these three viruses all have prymnesiophyte hosts [305] suggests the most appropriate classification for CpV-BQ1 is within the *Prymnesiovirus* genus. Thus, our phylogenetic analysis supports the inclusion of CpV-BQ1 in the NCLDV *Phycodnaviridae* family and suggests it can most appropriately be placed within the *Prymnesiovirus* genus.

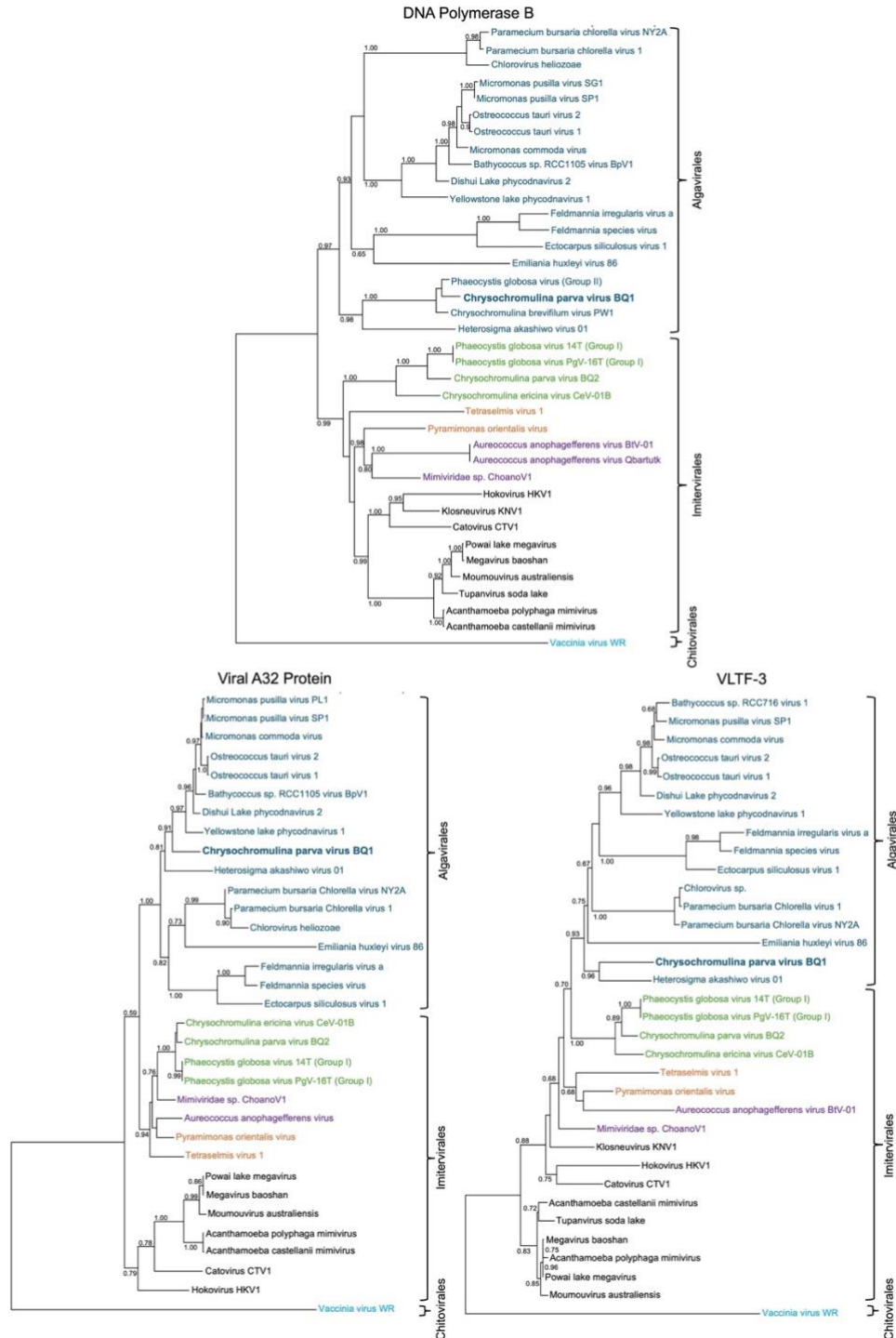


Figure 4.5. Phylogenetic Analysis of Three CpV-BQ1 Genes Universally Found in NCLDV Genomes. Phylogenetic trees generated to determine the taxonomy of the CpV-BQ1 virus using the Phylogenetic maximum likelihood (PhyML) algorithm in Seaview v5.0. Trees were generated using the protein sequences for the universal NCLDV genes *polB*, viral A32 protein, and VLTF-3 (Supplementary Table 4.4). Protein sequences are from viruses belonging to either the *Algavirales*, *Imitevirales*, or *Chitovirales* order. Families within each order are colour coded, *Phycodnaviridae* in blue, *Mesomimiviridae* in green, *Schizomimiviridae* in purple, *Allomimiviridae* in orange, *Mimiviridae* in black, and *Poxviridae* in light blue.

4.3.4. DNA and RNA Replication, Recombination, and Repair Enzymes

NCLDV encode a suite of genes that enable the replication of their own genome with little reliance on their host's genome replication machinery [15,27]. A total of 19 proteins involved in DNA/RNA replication, recombination, and repair were identified in the CpV-BQ1 genome. This consort of enzymes enables CpV-BQ1 to facilitate much of its own DNA replication processes. These replication associated genes include a DNA polymerase type-B (*polB*), a DNA polymerase family X protein, an A22 holliday junction resolvase, a proliferating cell nuclear antigen sliding clamp (PCNA), topoisomerase I and II, four nuclease proteins, four helicase proteins, two Yqaj-like viral recombinases, an ATP-dependent DNA-ligase, a P-loop nucleoside triphosphate hydrolase, and a SWIIB/MDM2-domain protein [15,27]. Some NCLDVs, such as poxviruses [27], mimiviruses [293,295], and phycodnaviruses [294,340] replicate and package their genomes within the cytoplasm of their host cell in subcellular compartments surrounded by endoplasmic reticulum membranes, which are referred to as virus factories [27,293]. However, other NCLDVs, such as members of the family *Phycodnaviridae*, replicate their genome within the nucleus of their host, then package virions in the cytoplasmic virus factories [293,294]. To explore the location of CpV-BQ1 genome replication, proteins involved in DNA replication were analyzed for the presence of nuclear localization signals (NLS) using DeepLoc 2.1 [341]. Indeed, NLS signals were identified in 12 of the 19 predicted DNA replication, recombination, and repair enzymes, including the DNA-dependent DNA replication polymerase, *polB*. This suggests that at least some part of the DNA replication process occurs within the host cell nucleus.

4.3.5. Transcription

Most NCLDVs encode a DNA-dependent RNA polymerase, which allows them to transcribe their viral mRNA within the cytoplasm [342]. However, some members of the *Phycodnaviridae* family, such as chloroviruses and prasinoviruses, do not encode an RNA polymerase gene. Instead they

rely on the host's RNA polymerase and enter the nucleus to initiate viral gene transcription [56]. Similarly, the CpV-BQ1 genome does not encode a recognizable RNA polymerase and therefore its lytic cycle most likely involves a nuclear localization step necessary for gene transcription.

To gain control of the host RNA polymerase II (RNA pol), CpV-BQ1 encodes several eukaryotic-like transcription factors (TFs) that likely interact with the host's RNA pol and alter its activity. CpV-BQ1 encodes two transcription factor IIB (TFIIB) proteins and a TATA box binding protein (TBP), which are similar to general eukaryotic TFs [343]. During eukaryotic transcription initiation, RNA pol is recruited to the DNA promoter by a complex of general TFs including TFIIB and TBP. With virus encoded TFs, viruses can redirect the RNA pol to initiate the transcription of viral genes [343]. CpV-BQ1 also encodes two specific TFs which activate viral genes expressed in the late transcription phase, including virus late transcription factor 3 (VLTF-3) and an MYM-type zinc-finger FCS motif containing protein [344–346].

A handful of other genes important for transcription are encoded by the CpV-BQ1 genome. Transcription elongation factor S-II (TFIIS) stimulates the cleavage activity of RNA pol when the transcription complex becomes stalled, allowing transcription to be restarted at the newly created 3' prime end [313]. Two mRNA-capping enzymes are encoded that ensure the protection and efficient transcription of mRNA. A protein Blast analysis of these two mRNA-capping enzymes show they may differ in their evolutionary origin and possibly function. The first is encoded at locus BQ1_111 and is similar to bacterial and eukaryotic mRNA capping enzymes, while the second locus is similar to archaeal and viral mRNA capping enzymes. Moreover, an RNA cap guanine-N2 methyltransferase family protein likely methylates the 5' prime cap on mRNA, which is suspected to increase the synthesis of viral transcripts [347]. Lastly, the enzyme ribonuclease III (RNase III) is encoded by the CpV-BQ1 genome. RNase III proteins are involved in the processing and maturation of RNA species, including tRNA, rRNA, and mRNA, as well as the

degradation of mRNAs. The specific role of RNase III encoded by NCLDV is not well understood, however it is suspected that it is involved in RNA cleavage and tRNA maturation processes [348].

During the nuclear infection stage, chloroviruses such as *Paramecium bursaria chlorella virus* (PBCV-1) not only redirect host transcriptional machinery, but they also repress the expression of host genes [19]. It is likely that CpV-BQ1 uses mechanisms similar to PBCV-1 to decrease the transcription of host genes and prioritize the expression of its own genes. In the virion, PBCV-1 packages methylation, restriction endonuclease, and chromatin-remodeling enzymes that are released upon nuclear entry of virion particles. In consort, these enzymes methylate, degrade, and remodel host DNA effectively downregulating the production of host transcripts [19]. CpV-BQ1 encodes several methylase enzymes and endonucleases, some of which may target host DNA. Additionally, CpV-BQ1 encodes a SWIB/MDM2-domain containing protein; the SWIB domain, although not fully understood, can function as a transcriptional activator or as a chromatin-remodeling protein [349]. Thus, circumstantial evidence suggests the CpV-BQ1 genome contains the necessary protein apparatus to downregulate host gene transcription.

4.3.6. Nucleotide Transport and Metabolism Proteins

To facilitate rapid genome replication, some NCLDV encode DNA precursor metabolism proteins that generate pools of available deoxythymidine triphosphate (dTTP) [27]. The CpV-BQ1 genome encodes several genes that are important for dTTP synthesis, including deoxycytidylate deaminase (dCD), thymidine kinase (TK), deoxyuridine 5'-triphosphate nucleotidohydrolase (dUTPase), thymidylate synthase ThyX (thyX) and ribonucleoside-diphosphate reductase (RNR) (Figure 4.6) [350].

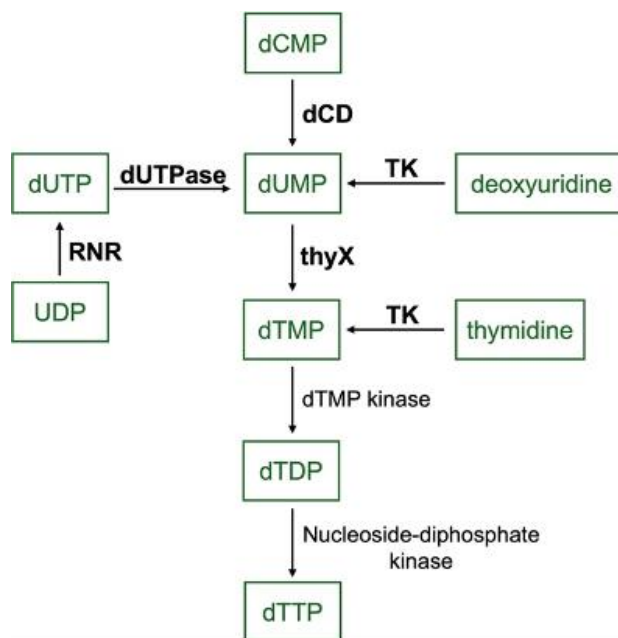


Figure 4.6. *Biosynthesis of Deoxythymidine Triphosphate (dTTP).* KEGG ontology pyrimidine metabolism pathway (<https://www.kegg.jp/pathway/ec00240+3.5.4.12>). Substrates are in green boxes, enzymes written in black text convert substrates following the direction of the arrow heads. Enzyme names that are bolded are encoded by the CpV-BQ1 genome, while the remaining enzymes are not encoded by the CpV-BQ1 genome.

The three enzymes dCD, TK, and dUTPase convert their respective substrates to dUMP, an essential precursor molecule for the production of dTMP by the enzyme thyX (Figure 4.6). Furthermore, TK can also phosphorylate available thymidine to produce dTMP [351]. dTMP can then be phosphorylated by dTMP kinase then nucleoside diphosphate kinase to produce dTTP that can be incorporated into replicating viral DNA, however, the CpV-BQ1 genome does not encode either of these enzymes (Figure 4.6). In addition to contributing to dTTP biosynthesis, RNR can also convert CDP to dCDP with help from the protein glutaredoxin, which is encoded by the CpV-BQ1 genome. dCDP can then be phosphorylated by nucleoside-diphosphate kinase to produce dCTP, which can be used for DNA synthesis. However, it has been shown that these proteins are only essential for maintaining dCTP levels when similar host enzymes have been depleted [352]. Other studies have demonstrated that glutaredoxin is an important component in a redox cascade reaction that is required for virion assembly [353]. Analysis of the CpV-BQ1 encoded glutaredoxin with subcellular prediction tool DeepLoc identified an endoplasmic

reticulum (ER) localization signal, thus this protein is more likely to be located at ER virus factories and involved in virion assembly and packaging.

The CpV-BQ1 genome also encodes an NUDIX hydrolase. NUDIX proteins hydrolyze substrates containing a nucleoside diphosphate linked to some other moiety, increases free nucleotides available for metabolism [354]. It remains unclear what substrate(s) the CpV-BQ1 encoded NUDIX hydrolase has specificity to, however it has been shown that NUDIX enzymes in NCLDV's Vaccinia Virus, African Swine Fever Virus (ASFV), and Mimivirus L375 function as mRNA decapping enzymes. This decapping process is thought to facilitate mRNA turnover and increase the availability of dNTPs [355].

4.3.7. DNA Methylation

Methyltransferases (MTases) are present in all domains of life however they are not highly prevalent amongst viruses. Within viruses, MTases are most often reported in bacteriophages and some members of the *Phycodnaviridae* family [350,356]. *Phycodnaviruses* often use their encoded MTases to methylate their own genome. Moreover, it has been observed that some chloroviruses encode their own restriction-modification (R-M) systems, complete with type II site specific restriction enzymes to degrade host or foreign DNA and MTases to protect their own genome from degradation [356,357]. The CpV-BQ1 genome encodes five methyltransferase enzymes, including a D12 class N6 adenine-specific DNA methyltransferase protein (mA6-MTase), DNA cytosine-5 methyltransferase (mC5-MTase), FbkM methyltransferase, a SAM-dependent methyltransferase, and the previously discussed RNA cap guanine-N2 methyltransferase. The possibility of a CpV-BQ1 encoded R-M system was investigated but is unlikely as we could not identify any type II site-specific restriction enzymes in the CpV-BQ1 genome.

mA6-MTase methylates the amino group at the C6 position of adenine in specific nucleotide motifs, creating an N⁶-methyl-adenine modification [356,358]. N⁶-methyl-adenine modifications are rare amongst eukaryotes but are widespread amongst *Phycodnaviridae* family members [351,356,359,360]. The CpV-BQ1 genome is AT-rich, with a 67.68% AT content, which makes adenine a highly prevalent nucleotide and robust candidate for methylation. Therefore, this enzyme likely specifically methylates the CpV-BQ1 genome to provide protection from degradation by nucleases.

mC5-MTase methylates the 5th carbon on the pyrimidine ring of cytosine in specific nucleotide motifs, creating a 5-methyl-cytosine modification [361,362]. 5-methyl-cytosine modifications are commonly found in eukaryotic genomes and are implicated in various functions such as regulating gene expression, transposon silencing, genomic imprinting, and development [356]. This modification is also found in some virus genomes including *Phycodnaviridae* family members and is thought to provide the viral genome protection from nucleases [350,356,360]. The mC5-MTase encoded by CpV-BQ1 is likely required for methylation of its own genome. However, it also could be involved in manipulating its host's cellular processes, such as downregulating the expression of host genes. To explore the CpV-BQ1 genome methylation landscape single-molecule real-time sequencing could be used to identify both N⁶-methyl-adenine and 5-methyl-cytosine modifications and gain deeper insight into the protection and gene regulation mechanisms provided by these MTases [356,363].

FkbM methyltransferase is a S-adenosyl-L-methionine (SAM) transferase dependent methyltransferase enzyme with O-methylation activity [364]. FkbM was originally isolated from bacteria and has been shown to perform post-modification of the macrocyclic polyketides FK506, FK520, and the antibiotic rapamycin [365]. FkbM has also been found in *Phycodnaviridae* genomes, however, its substrate interactions and function is unknown [350,351,366]. The final methyltransferase encoded by CpV-BQ1 could not be assigned a specific function and is

designated a general SAM-dependent methyltransferase [367]. SAM-MTases are a broad group of methyltransferases found in all domains of life that serve many different biological functions, thus, the functional role of this CpV-BQ1 encoded enzyme is unknown [368].

4.3.8. Sugar Manipulation

Three glycosyltransferase proteins were identified in the CpV-BQ1 genome. Glycosyltransferases attach sugar moieties to proteins, which can constitute important post-translational modifications. Glycosidic bonds are catalyzed using sugar donors that contain either a nucleoside phosphate or a lipid phosphate leaving group [369,370]. Glycosyltransferases are classified into hierarchical groups based on families, clans, and fold structures [371]. A total of 137 glycosyltransferase (GT) families have been classified to date, listed on the carbohydrate-active enzymes (CAZy) database (<http://www.cazy.org/>) [372]. Of the three identified GTs in the CpV-BQ1 genome, two were classified as members of specific families, while the third could not be classified as a specific GT family member.

The GT encoded at locus BQ1_30 belongs to the family GT2. The GT2 family is one of the largest GT groups and contains members with a wide variety of catalytic activities. These enzymatic functions include, but are not limited to, cellulose synthase, chitin synthase, hyaluronan synthase, and β -glucosyltransferase [371]. GT2 family members contain a GT-A fold defined by two closely positioned $\beta/\alpha/\beta$ Rossmann domains, and use an inverting mechanism during catalysis of the donor substrate [370].

The GT encoded at locus BQ1_106 belongs to the family GT17, a small GT group in which all family members exhibit β -1,4-N-Acetylglucosaminyltransferase (GnTIII) activity and use an inverting mechanism during catalysis [371]. In vertebrates, GnTIII catalyzes the formation of bisecting *N-acetylglucosamine* (GlcNAc) residues on *N*-glycans within the Golgi apparatus. This modification inhibits the action of branching enzymes, preventing the formation of highly branched

N-glycan structures [373,374]. In viruses, glycans are commonly *N*-linked to a glycoprotein asparagine residue via GlcNAc [375]. These modifications are prevalent on glycoproteins in virion envelopes and play important roles in viral infection stages, including progeny formation and cellular infection [375].

Moreover, some NCLDV virions have glycans attached to the surface of major capsid proteins via the activity of GTs [376]. Thus, possible functions of the three GTs encoded by CpV-BQ1 may be the synthesis of *N*-glycans on major capsid proteins or glycoproteins in the virion envelope, however, these possibilities have yet to be explored experimentally.

4.3.9. Protein and Lipid Binding, Synthesis, and Modifications

The CpV-BQ1 genome encodes two proteins that likely participate in protein synthesis, the translation initiation factor 4E (eTIF4E) and DNAJ. The eTIF4E facilitates binding of the host's ribosome to the 5' prime cap of mRNA, which initiates protein translation [377]. DNAJ, originally identified in prokaryotes, functions as a co-chaperone protein to Hsp70 and aids in protein folding [378]. However, studies of DNAJ homologs in viruses show this protein has diverse roles and can also be involved in genome replication, transcriptional activation, virion assembly, and cellular transformation [378]. Thus, further study is required to determine the exact functional nature of the DNAJ protein encoded by CpV-BQ1. The CpV-BQ1 genome also encodes three tRNAs, including tRNA-Leu, tRNA-Arg, and tRNA-Ile, which are required for incorporation of these three amino acids into nascent peptide chains [335].

Protein modification enzymes encoded by CpV-BQ1 indicate the host ubiquitination pathway is used by this virus to modulate proteins and perhaps subvert host defense mechanisms. Ubiquitination of proteins can either modulate their activity or signal their degradation [15]. Four E3 ubiquitin ligase proteins are encoded by the CpV-BQ1 genome. Interestingly, one of the E3-ligases encoded by the genome was denoted as a N1r/p28-like protein. The N1r/p28 protein was

the first E3-ligase discovered in poxviruses, it is recruited to virus factories within the cell and is an important virulence factor [15,379]. Thus, the CpV-BQ1 encoded N1R/p28 protein may be involved in the ubiquitination of important proteins within virus factories. Other proteins that may be involved in protein modification include six F-box domain containing proteins, which are thought to be involved in the ubiquitin-ligase complex [380], as well as, two proteins containing a metallopeptidase (WLM) domain, which are also thought to be associated with ubiquitin-signaling pathways [381].

Lastly, the patatin-like phospholipase protein encoded by CpV-BQ1 has also been found in many other NCLDV's however its function is not well understood [382]. In plants, patatin phospholipase catalyzes the cleavage of fatty lipids from membranes [383], while in bacteria, this protein has been implicated in the pathogen-host interaction [384].

4.3.10. DNA Packaging and Genome Completeness

The CpV-BQ1 genome encodes the viral A32 protein found in all NCLDV's that is essential for the packaging of viral DNA into virions [382,385]. Silencing of the viral A32 protein results in virion structures devoid of viral DNA [386]. A32 is thought to form a hexameric ring on the membrane surface of immature virions and pumps complete viral DNA into the virion [27].

In addition to the A32 protein, packaging of many NCLDV genomes, including Poxviridae [387], ASFV [388], phycodnaviruses [389–391], and mimiviruses [297,392], requires the presence of inverted terminal repeats on the distal ends of their genomes [393]. These inverted repeats contain, in order, genes, tandem DNA repeats, and mismatched hairpin ends that interact with packaging enzymes ensuring complete genomes are incorporated into virions [27]. The CpV-BQ1 genome has inverted repeats 3928 bp long at its distal ends that contain identifiable genes and tandem repeats. The possibility of hairpin sequences in the first 30, 40, 50, and 100 bp of the genome was investigated using the DNA secondary structure prediction tool by vector builder

[394] (Figure 4.7). Of the four analyzed sequences, the 50 bp sequence provides the most promising mismatched hairpin structure, as it is similar to the stem-loop organization of the vaccinia virus hairpin motifs [387].

The presence and structure of mismatched hairpin ends cannot be verified by genome assembly alone. However, based on the presence of terminal distal repeats and their well-defined role in other NCLDV [27,389,392,393] as well as preliminary secondary structural analysis, it is likely that the CpV-BQ1 genome forms hairpin structures. Most importantly, as distal inverted repeats are a signature of many NCLDV genome ends, the presence of these inverted repeats on the distal ends of the CpV-BQ1 genome provide strong evidence that the CpV-BQ1 genome has been sequenced completely.

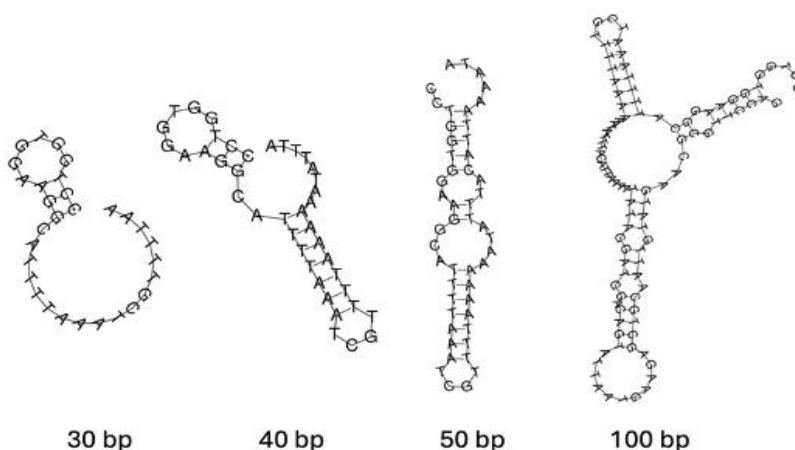


Figure 4.7. *Potential Hairpin Structures of CpV-BQ1 Genome Termini.* Possible secondary DNA hairpin structures formed by the 30, 40, 50, and 100 bp ends of the CpV-BQ1 genome modeled using the DNA secondary structure tool by VectorBuilder [394].

4.3.11. *Virion Capsid and Associated Structural Proteins*

Four putative viral capsid proteins were identified in the CpV-BQ1 genome. InterPro analysis of putative capsid proteins at locus BQ1_139, BQ1_140, and BQ1_141 identified an adenovirus hexon domain in these three proteins. The adenovirus hexon domain contains a jelly-roll fold that is a signature feature of NCLDV capsid proteins [395,396]. Furthermore, these three proteins and

BQ1_142 had a high degree of structural similarity to other NCLDV capsid structures when analyzed with the homology-based protein fold recognition tool Phyre2 [397]. Lastly, these proteins are encoded sequentially on the same strand. This proximity would permit the synthesis of all four genes in tandem, making their synthesis highly efficient. NCLDV capsid proteins have a high degree of sequence variation; a protein Blast analysis of these four proteins did not detect any sequence homology with other proteins [345]. Thus, if these putative capsid proteins are indeed structural components of the virion capsid, this would expand our knowledge of NCLDV capsid sequences and aid in the identification of other capsids. Additionally, two minor capsid proteins were identified in the CpV-BQ1 genome. The minor capsid P9 transmembrane helices containing protein has domains similar to the inner capsid P9 structure of the NCLDV PBCV-1 [396]. The second minor protein is a tape measure protein, which likely spans from one virion icosahedral vertex to another and is thought to play a critical role in mediating the capsid size and maintaining the orientation of capsomers to one another during assembly [398].

The CpV-BQ1 genome also encodes two tail fiber proteins, one of which contains a fibronectin III binding domain. Tail fiber proteins are found in bacteriophages and facilitate the injection of viral DNA into cells. However, they have also been found in some NCLDVs including in chloroviruses and other phycodnaviruses [351,399]. In PBCV-1, tail fiber proteins form a spike on a vertex of the icosahedral capsid structure that interacts with the host cell to facilitate virion entry [399,400]. Similarly, the CpV-BQ1 genome tail fiber proteins may form a spike on the capsid to facilitate cellular entry.

An important post-translational NCLDV capsid protein modification for assembly and stabilization is the formation of disulfide bonds between conserved cysteine residues [15,401]. These bonds are catalyzed by a redox reaction requiring two enzymes, a thioredoxin domain containing protein and an ERV sulfhydryl oxidase [353,401]. The CpV-BQ1 genome encodes an ERV/ALR sulfhydryl oxidase and two thioredoxin domain-containing proteins, a thioredoxin and a glutaredoxin-like

protein. DeepLoc subcellular predictions identified ER signaling domains on the ERV/ALR sulfhydryl oxidase and glutaredoxin-like protein. Together, these two proteins likely complete the redox reaction necessary for assembly and stabilization of capsid structures. DeepLoc predicted the thioredoxin protein is within the cytoplasmic compartment; however, this does not definitively exclude it from involvement in the capsid redox cascade reaction. Lastly, the CpV-BQ1 genome encodes an Ac78 gene that is similar to the baculovirus protein Ac78 [402]. This protein is important for budded virion production, embedding of virions into occlusion bodies, and primary cellular infection. It has been shown to play an important role in virion localization however is not essential for virion assembly or structure [402].

4.3.12. Comparison of CpV-BQ1 and CpV-BQ2 Genomes

With the publication of this work we now have the full genome sequence of two viruses, CpV-BQ1 (this study) and CpV-BQ2 [57], which can infect *C. parva*. CpV-BQ1 and CpV-BQ2 are taxonomically assigned to the NCLDV families *Phycodnaviridae* and *Mesomimiviridae* respectively. One striking difference between CpV-BQ1 and CpV-BQ2 is their genome size. CpV-BQ1 has a 165,454 bp genome with 32.32% GC content that encodes 193 ORFs, whereas CpV-BQ2 has a 437,255 bp genome with 25% GC content that encodes for 503 ORFs [57]. This difference in both size and coding potential indicates that although they share the same host, they likely use remarkably different infection and replication strategies. For example, CpV-BQ2 encodes at least eight restriction-modification (R-M) systems and 13 methyltransferases [57], whereas CpV-BQ1 does not encode any R-M systems and only encodes 4 methyltransferases. This indicates these viruses use very different methods to protect their own DNA and/or degrade host DNA, which can alter cellular processes such as metabolism, transcription, and translation [356,357]. Furthermore, the presence of an RNA polymerase (RNA pol) II gene in CpV-BQ2 and its absence in CpV-BQ1 indicate these viruses use very different infection and replication strategies. RNA pol encoding viruses can transcribe their own DNA, NCLDV that encode their

own RNA pol typically carry out transcription, genome replication, and virion packaging all within the host's cytoplasm [342]. On the other hand, NCLDV's that do not encode an RNA pol require a nuclear infection step to hijack their host's transcriptional machinery that is followed by a cytoplasmic infection wherein virion packaging occurs [56]. Thus, CpV-BQ1 likely uses a two-step infection/replication strategy that includes both cytoplasmic and nuclear infection that requires coordination across subcellular compartments, whereas CpV-BQ2 likely infects *C. parva* through a one-step cytoplasmic approach.

4.4. Conclusions

In this study we have sequenced and annotated the complete linear 165,454 bp genome of *Chrysochromulina parva* virus BQ1 (CpV-BQ1). CpV-BQ1 was originally isolated from a lake in Ontario and is a lytic agent of the haptophyte alga *C. parva*. Taxonomic analysis of *polB*, A32 ATPase, and VLTF-3 protein sequences were used to assign the *Phycodnaviridae* family classification to CpV-BQ1. The genome contains 193 genes, of which 92 could be assigned a known function. CpV-BQ1 has hallmark genes found in many NCLDV necessary for genome replication, virion production, and transcription [292]. Like other phycodnaviruses, CpV-BQ1 most likely has a two-step cellular infection life-cycle, first entering the nucleus for transcription as evidenced by the lack of an RNA polymerase gene, then virion assembly ensues within the cytoplasm [19,56]. Previously, another *C. parva* lytic agent *Chrysochromulina parva* virus BQ2 (CpV-BQ2) was isolated from the same water sample as CpV-BQ1 [47] its genome was sequenced and was taxonomically assigned to the *Mesomimiviridae* family [57]. Thus far, co-infection of a eukaryotic algae with viruses from both *Phycodnaviridae* and *Mesomimiviridae* families has only been observed in the species *Phaeocystis globosa* [52,339]. However, only the complete genome sequence of one *P. globosa* infecting virus, the group II PgV-16T mimivirus, has been sequenced and made available [403]. It is postulated that co-infections of eukaryotic algae with viruses belonging to different NCLDV families is common, however, due to a lack of data this hypothesis cannot be currently supported [52]. Indeed, only ~60 eukaryotic algal viruses have been isolated in culture [52], while thousands of eukaryotic algal species have been isolated and are available in culture collections around the world. This disparity emphasizes the lack of research and knowledge surrounding eukaryotic algal virus diversity, taxonomy, life cycle, and environmental impact. With the work reported here, we have for the first time established an algal-virus system with complete genome sequences for both *Phycodnaviridae* and *Mesomimiviridae* viruses. This system can be used to study the biological, ecological, and environmental

consequences of the coding potential of these viruses, which replicate in the same host. For example, the presence of R-M systems encoded by CpV-BQ2 and their absence in CpV-BQ1 may be relevant to inter-viral competition and suggest that BQ2 may restrict the replication of BQ1, or the PLVs that presumably parasitize one of these viruses. With the genomic information in hand, detailed transcriptional studies can be conducted to further understand the complicated dynamics and relationships between *C. parva* and its viral parasites. In turn, this knowledge will illuminate the potential complexities of algal virus-host interactions and is especially important considering the critical importance of algae in the biosphere and human affairs.

Chapter 5. Conclusions and Future Perspectives

5.1. Overview of Objectives and Contributions

This thesis sought to explore viral diversity and transmission through genomic sequencing techniques, with the dual aims of expanding the known genetic landscape of freshwater algal viruses and developing practical tools for viral surveillance in wastewater systems. Through three major research projects, genomic sequencing was used to detect, characterize, and monitor viruses in aquatic environments. The first two objectives were to develop and apply tiled-amplicon sequencing assays for the detection and lineage characterization of clinically relevant viruses, IAV H3N2 and RSV A, and the agriculturally significant ToBRFV in environmental wastewater samples. The third objective was to characterize the genome of a novel freshwater algal virus using a hybrid long- and short-read sequencing approach.

With a One Health lens, these three research projects collectively examine virus impacts across humans, plants, and ecosystems, illustrating how viral dynamics in diverse biological systems ultimately converge on human health and well-being. Objective one directly addresses human health by developing and applying wastewater-based genomic surveillance approaches for the human-infecting seasonal respiratory viruses IAV H3N2 and RSVA. This work demonstrates that wastewater genomics is an effective, scalable, and complementary public health surveillance tool capable of capturing community-level viral circulation independent of clinical testing biases. By strengthening and supporting existing public health surveillance frameworks, these methods contribute to earlier detection of outbreaks, improved public awareness, and more informed public health decision making. Effective monitoring and control of viral disease outbreaks not only improves individual health outcomes and reduces morbidity and mortality but also reduces pressure on healthcare systems and healthcare professionals. Moreover, these viral genomic surveillance approaches developed in this work enable the detection and tracking of emerging

IAV H3N2 and RSV A and mutations that may comprise or reduce vaccine of anti-viral treatment effectiveness, which supports adaptive public health responses.

Objective two extends the One Health framework to plant health and food product through the development of a robust genomic tool for monitoring the ToBRFV. By enabling analysis and characterization of the ToBRFV genome, clade distributions, and mutations, this approach provides valuable insights for both researchers and growers. Improved understanding of ToBRFV transmission dynamics enhances our capacity to detect outbreaks early, inform management strategies, and mitigate virus spread. Ultimately these efforts can help reduce the frequency and severity of ToBRFV outbreaks, limit crop losses, decrease the financial burden on growers, and support the stability of tomato production. In doing so, this work contributes to plant health, agricultural sustainability, and food security, all of which are essential components of human health within a One Health framework.

Objective three addresses an often-looked ecological dimension of One Health by advancing our understanding of algal viruses through the sequencing and annotation of the novel algal virus CpV-BQ1. This work expands the currently limited genomic knowledge of algal viruses and sheds light on the complex virus-host interactions that regulate aquatic ecosystems. Algal play critical roles in ecosystems through biogeochemical and nutrient cycling, the formation of algal blooms, and the modification of microbial community structures. By influencing algal populations and bloom dynamics, these viruses can indirectly affect the key ecological functions algal play in ecosystems. Through lytic activity, algal-viruses can disrupt harmful algal blooms and modify algal populations, thereby effecting water quality, food web stability, and nutrient cycling, which result in downstream ecosystem effects. Additionally, algal dynamics can influence carbon sequestration, climate and weather processes, and environmental stability. Collectively, these ecological processes underpin the health of the ecosystems upon which human societies depend. By illuminating viral mechanisms operating at the ecosystem level, this objective reinforces the

interconnectedness of environmental, plant, and human health which underscores the importance of viruses within a One Health framework. This work advances the fields of environmental virology, public health surveillance, and viral genomics by contributing a complete genome sequence of a novel algal virus and establishing optimized sequencing assays for the detection of viral variants of public health concern in wastewater. Together, these projects demonstrate how viral genomics can illuminate both fundamental and applied questions in public health surveillance, plant pathology, and viral ecology, using a One Health approach.

5.2. Summary of Key Findings

5.2.1. Wastewater-based Surveillance of Influenza A H3N2 and RSV A

For the first objective of this thesis, two novel tiled-amplicon sequencing assays were developed and applied to wastewater samples to monitor the transmission of IAV H3N2 and RSV A variants. Generally, increases and decreases of genomic coverage obtained with these assays mirrored trends in reported clinical cases within the municipal public health unit region that wastewater samples were taken from, demonstrating the potential of WBS to complement clinical surveillance in tracking community-level transmission of IAV H3N2 and RSV A.

Through the application of variant deconvolution tools, we identified specific IAV H3N2 and RSV A clades and identified distinct spatiotemporal patterns in variant distribution, revealing differences in local transmission dynamics and highlighting the capacity of these tools to resolve community-specific pathogen circulation. Additionally, by comparing the performance of the Freyja tool with our in-house variant deconvolution tool, Alvoc, we showed that Alvoc performs comparably to Freyja for IAV H3N2 and outperforms it in detecting RSV A variants. Collectively, this work demonstrates that our sequencing assays, variant deconvolution tool, and WBS approach offer a valuable complement to clinical respiratory virus surveillance, providing

actionable insights into seasonal virus transmission dynamics, variant emergence, and public health preparedness.

5.2.2. Development of a Tiled-Amplicon Sequencing Assay for the ToBRFV

For the second objective of this thesis, we successfully identified *Tomato brown rugose fruit virus* (ToBRFV) clades present in Ontario municipal wastewater systems using a novel tiled-amplicon sequencing assay. *In silico* analysis of our multiplexed PCR primers demonstrated the ability of these primers to inclusively bind variants in all ToBRFV clades, while excluding closely related off-target viruses within the same *Tobamovirus* genus. ToBRFV genomes were identified in all tested wastewater influent samples, and the virus was identified as the most abundant RNA virus in four out of five samples. When our ToBRFV-Seq assay to RNA shotgun sequencing were compared we observed a statistically significantly increase in the ToBRFV reads that were recovered by our assay.

As expected, when we used our ToBRFV-Seq assay that relies on a PCR-enrichment strategy to recover ToBRFV reads, we achieved a statistically significant increase in both the number of recovered ToBRFV reads and genome depth, while requiring substantially less sequencing capacity, lower cost of materials (\$24 per ToBRFV-Seq samples versus \$36 for RNA shotgun sample), and reduced labour, compared to an RNA shotgun approach.

Lastly, ToBRFV clades present in wastewater were determined using our in-house variant deconvolution tool Altob, which detected variants in clades four and seven. According to global transmission patterns tracked on the NextStrain ToBRFV build, clade four variants were known to circulate in Canada, while clade seven had not yet been reported here. Notably, clade seven variants were detected in Canada in the fourth iteration of the ToBRFV NextStrain build, after the publication of our assay. This highlights the ability of our assay to identify clades in Canada prior to their detection through global genomic surveillance efforts. Furthermore, these findings

demonstrate the value of wastewater-based surveillance for monitoring plant pathogens and capturing regional pathogen diversity, which can be difficult to document by conventional plant disease monitoring that typically relies on symptomatic crops in privately owned agricultural operations.

5.2.3. Assembly and Characterization of the CpV-BQ1 Genome

In the third objective of this thesis, the complete genome of *Chrysochromulina parva* virus BQ1 (CpV-BQ1) was successfully assembled using a hybrid long-read Oxford Nanopore and short-read Illumina sequencing approach. The 165,454 bp genome was thoroughly characterized, it contained 193 ORFs, of which 92 could be assigned a putative function. Putative gene functions were assigned using a comprehensive and extensive analysis of ORFs via a combination of sequence homology, protein domain and family, and hidden Markov model prediction tools. Hallmark NCLDV genes, the DNA polymerase B, A32-like packaging ATPase, and virus late transcription factor 3, were used to confidently assign CpV-BQ1 to the phylum *Nucleocytoviricota*, class *Megaviricetes*, order *Algavirales*, and family *Phycodnaviridae*.

Putative genes were placed in eight distinct categories, and their functional potential was thoroughly investigated using published literature. These functional categories included: DNA/RNA replication, recombination, and repair; nucleotide metabolism and DNA packaging; transcription; sugar manipulation; DNA methylation, protein and lipid binding, synthesis, and modifications; virion capsid and associated protein; and miscellaneous functions.

Genes involved DNA/RNA replication, recombination, and repair were identified, suggesting CpV-BQ1 is not reliant on its hosts cellular machinery for genome replication and the presence of nuclear localization signals on many of these genes indicates replication occurs in the hosts nucleus. Moreover, CpV-BQ1 did not contain an identifiable DNA-dependent RNA polymerase, thus, it is likely dependent on its host transcriptional machinery to generate mRNA transcripts.

From this, we hypothesized that CpV-BQ1 uses a two-step infection process, wherein genome replication and gene transcription occur during a nuclear infection step, followed by virion assembly and packaging in a cytoplasmic infection step. CpV-BQ1 also encoded a comprehensive set of genes to alter host gene expression, transcription, and translation. These extensive host modifications and manipulations by CpV-BQ1 likely create a host environment suited its specific needs and beneficial for its own virion replication processes. Additionally, CpV-BQ1 contains unique major capsids proteins (MCPs) that expand our knowledge of capsid protein sequences and our ability to identify similar proteins with homology-based searches.

With this work we now have a second fully sequenced virus that infects the ecologically significant alga *C. parva*. However, it is the first well characterized genome, with extensive characterization of its genomic content and functional coding-potential. Overall, our knowledge of NCLDV that infect freshwater algal is very limited. Thus, these findings expand our knowledge algal-virus genomics and provides a well-characterized model system for exploring the ecological and metabolic impacts of viruses that infect *C. parva*.

5.3. Integration of Thesis Objectives and Contributions

Across these three objectives, this thesis highlights the capacity of genomic sequencing approaches to address research questions in environmental virology, from the characterization and study of novel freshwater viruses to surveillance of human respiratory and plant viruses in wastewater systems and emphasizes the value of knowledge integration through a One Health framework. The use of hybrid genome assembly with the TryCycler pipeline for the CpV-BQ1 demonstrates its strength in resolving large complex viral genomes. Characterization of the CpV-BQ1 genome expands our understanding of algal-viral host relationships and the functional potential of viruses in freshwater ecosystems. The implementation of tiled-amplicon sequencing for ToBRFV, IAV H3N2, and RSV A demonstrates the sensitivity, cost-effectiveness, scalability,

and adaptability of surveillance in complex wastewater samples where virus titers are often too low to resolve by shotgun sequencing approaches. Moreover, the pairing of this sequencing strategy with variant deconvolution tools advances our capacity to resolve mixed virus lineages in wastewater, offering useful, fine-scale, and real-time resolution of viral transmission dynamics at the community-level.

5.4. Final Remarks and Conclusions

This thesis demonstrates how environmental virology and a One Health approach, powered by genomic sequencing technologies, can advance both applied pathogen surveillance and ecological research. From the application of tiled-amplicon sequencing approaches to detect ToBRFV and respiratory virus in wastewater to complete genomic characterization of a previously undescribed freshwater algal virus, this work contributes to our understanding of viral diversity, ecology, and transmission dynamics in freshwater and human engineered water systems. Collectively, these projects demonstrate the utility and ability of viruses in freshwater to integrate ecological virology and applied surveillance system, highlighting the interconnectedness of environmental, agricultural, and public health systems within a One Health framework. By developing and applying scalable, sensitive, and versatile sequencing approaches, this thesis establishes a practical framework for future studies aimed at both virus discovery and public health preparedness in an era of increased interest in viral research and emerging viral threats.

References

1. Suttle, C.A. Viruses in the Sea. *Nature* **2005**, *437*, 356–361, doi:10.1038/nature04160.
2. Roux, S.; Emerson, J.B. Diversity in the Soil Virosphere: To Infinity and Beyond? *Trends Microbiol.* **2022**, *30*, 1025–1035, doi:10.1016/j.tim.2022.05.003.
3. Wang, X.; Ma, T.; Chen, Z.; Liu, Y.; Wang, K.; Liu, G.; Li, K.; Chen, T.; Zhang, G.; Zhang, W.; et al. Review of Methods for Studying Viruses in the Environment and Organisms. *Viruses* **2025**, *17*, 86, doi:10.3390/v17010086.
4. Acheson, N.H. *Fundamentals of Molecular Virology*; 2nd ed.; John Wiley & Sons, Inc., 2011;
5. Baltimore, D. Viral Genetic Systems. *Trans. N. Y. Acad. Sci.* **1971**, *33*, 327–332, doi:10.1111/j.2164-0947.1971.tb02600.x.
6. Koonin, E.V.; Krupovic, M.; Agol, V.I. The Baltimore Classification of Viruses 50 Years Later: How Does It Stand in the Light of Virus Evolution? *Microbiol. Mol. Biol. Rev.* **2021**, *85*, e00053-21, doi:10.1128/MMBR.00053-21.
7. Baltimore, D. Expression of Animal Virus Genomes. *Bacteriol. Rev.* **1971**, *35*, 235–241, doi:10.1128/br.35.3.235-241.1971.
8. Koonin, E.; Fischer, M.; Yutin, N.; Kuhn, J.; Krupovic, M. *Reorganization of the Realm Varidnaviria*; International Committee on Taxonomy of Viruses, 2024;
9. Simmonds, P.; Adriaenssens, E.M.; Zerbini, F.M.; Abrescia, N.G.A.; Aiewsakun, P.; Alfenas-Zerbini, P.; Bao, Y.; Barylski, J.; Drosten, C.; Duffy, S.; et al. Four Principles to Establish a Universal Virus Taxonomy. *PLOS Biol.* **2023**, *21*, e3001922, doi:10.1371/journal.pbio.3001922.
10. *Structure and Physics of Viruses: An Integrated Guide*; Mateu, M.G., Ed.; Subcellular Biochemistry; Springer Nature Switzerland: Cham, 2024; Vol. 105; ISBN 978-3-031-65186-1.
11. Uyeki, T.M.; Hui, D.S.; Zambon, M.; Wentworth, D.E.; Monto, A.S. Influenza. *The Lancet* **2022**, *400*, 693–706, doi:10.1016/S0140-6736(22)00982-5.
12. Byrd-Leotis, L.; Cummings, R.D.; Steinhauer, D.A. The Interplay between the Host Receptor and Influenza Virus Hemagglutinin and Neuraminidase. *Int. J. Mol. Sci.* **2017**, *18*, 1541, doi:10.3390/ijms18071541.
13. Scholthof, K.-B.G. TOBACCO MOSAIC VIRUS: A Model System for Plant Biology. *Annu. Rev. Phytopathol.* **2004**, *42*, 13–34, doi:10.1146/annurev.phyto.42.040803.140322.
14. Zhang, S.; Griffiths, J.S.; Marchand, G.; Bernards, M.A.; Wang, A. Tomato Brown Rugose Fruit Virus: An Emerging and Rapidly Spreading Plant RNA Virus That Threatens Tomato Production Worldwide. *Mol. Plant Pathol.* **2022**, *23*, 1262–1277, doi:10.1111/mpp.13229.
15. Iyer, L.M.; Balaji, S.; Koonin, E.V.; Aravind, L. Evolutionary Genomics of Nucleo-Cytoplasmic Large DNA Viruses. *Virus Res.* **2006**, *117*, 156–184, doi:10.1016/j.virusres.2006.01.009.
16. Xian, Y.; Xiao, C. Current Capsid Assembly Models of Icosahedral Nucleocytoviricota Viruses. In *Advances in Virus Research*; Elsevier, 2020; Vol. 108, pp. 275–313 ISBN 978-0-12-820761-1.
17. Taubenberger, J.K.; Kash, J.C. Influenza Virus Evolution, Host Adaptation, and Pandemic Formation. *Cell Host Microbe* **2010**, *7*, 440–451, doi:10.1016/j.chom.2010.05.009.
18. Zambon, M.C. Epidemiology and Pathogenesis of Influenza. **1999**.
19. Van Etten, J.L.; Agarkova, I.V.; Dunigan, D.D. Chloroviruses. *Viruses* **2019**, *12*, 20, doi:10.3390/v12010020.
20. Rampersad, S.; Tennant, P. Replication and Expression Strategies of Viruses. In *Viruses*; Elsevier, 2018; pp. 55–82 ISBN 978-0-12-811257-1.

21. Jan, E.; Mohr, I.; Walsh, D. A Cap-to-Tail Guide to mRNA Translation Strategies in Virus-Infected Cells. *Annu. Rev. Virol.* **2016**, *3*, 283–307, doi:10.1146/annurev-virology-100114-055014.
22. Wang, X.; Zhu, J.; Zhang, D.; Liu, G. Ribosomal Control in RNA Virus-Infected Cells. *Front. Microbiol.* **2022**, *13*, 1026887, doi:10.3389/fmicb.2022.1026887.
23. Thaker, S.K.; Ch'ng, J.; Christofk, H.R. Viral Hijacking of Cellular Metabolism. *BMC Biol.* **2019**, *17*, 59, doi:10.1186/s12915-019-0678-9.
24. Hurwitz, B.L.; U'Ren, J.M. Viral Metabolic Reprogramming in Marine Ecosystems. *Curr. Opin. Microbiol.* **2016**, *31*, 161–168, doi:10.1016/j.mib.2016.04.002.
25. Sanchez, E.L.; Lagunoff, M. Viral Activation of Cellular Metabolism. *Virology* **2015**, *479–480*, 609–618, doi:10.1016/j.virol.2015.02.038.
26. Guo, L.; Louis, I.V.-S.; Bohjanen, P.R. Viral Manipulation of Host mRNA Decay. *Future Virol.* **2018**, *13*, 211–223, doi:10.2217/fvl-2017-0106.
27. Greseth, M.D.; Traktman, P. The Life Cycle of the Vaccinia Virus Genome. *Annu. Rev. Virol.* **2022**, *9*, 239–259, doi:10.1146/annurev-virology-091919-104752.
28. Laliberté, J.-F.; Zheng, H. Viral Manipulation of Plant Host Membranes. *Annu. Rev. Virol.* **2014**, *1*, 237–259, doi:10.1146/annurev-virology-031413-085532.
29. Walsh, D.; Mohr, I. Viral Subversion of the Host Protein Synthesis Machinery. *Nat. Rev. Microbiol.* **2011**, *9*, 860–875, doi:10.1038/nrmicro2655.
30. Mizuno, C.M.; Guyomar, C.; Roux, S.; Lavigne, R.; Rodriguez-Valera, F.; Sullivan, M.B.; Gillet, R.; Forterre, P.; Krupovic, M. Numerous Cultivated and Uncultivated Viruses Encode Ribosomal Proteins. *Nat. Commun.* **2019**, *10*, 752, doi:10.1038/s41467-019-08672-6.
31. Walsh, D. Manipulation of the Host Translation Initiation Complex eIF4F by DNA Viruses. *Biochem. Soc. Trans.* **2010**, *38*, 1511–1516, doi:10.1042/BST0381511.
32. Polcicova, K.; Badurova, L.; Tomaskova, J. Metabolic Reprogramming as a Feast for Virus Replication. *Acta Virol.* **2020**, *64*, 201–215, doi:10.4149/av_2020_210.
33. Tian, F.; Wainaina, J.M.; Howard-Varona, C.; Domínguez-Huerta, G.; Bolduc, B.; Gazitúa, M.C.; Smith, G.; Gittrich, M.R.; Zablocki, O.; Cronin, D.R.; et al. Prokaryotic-Virus-Encoded Auxiliary Metabolic Genes throughout the Global Oceans. *Microbiome* **2024**, *12*, 159, doi:10.1186/s40168-024-01876-z.
34. Brussaard, C.P.D. Viral Control of Phytoplankton Populations—a Review. *J. Eukaryot. Microbiol.* **2004**, *51*, 125–138, doi:10.1111/j.1550-7408.2004.tb00537.x.
35. Funaoka, Y.; Hiromoto, H.; Morimoto, D.; Takahashi, M.; Wada, K.; Nagasaki, K. Diversity in Infection Specificity between the Bloom-Forming Microalga *Heterosigma Akashiwo* and Its dsDNA Virus, *Heterosigma Akashiwo Virus*. *Microbes Environ.* **2023**, *38*, n/a, doi:10.1264/jsme2.ME23036.
36. Koonin, E.V.; Dolja, V.V.; Krupovic, M. Origins and Evolution of Viruses of Eukaryotes: The Ultimate Modularity. *Virology* **2015**, *479–480*, 2–25, doi:10.1016/j.virol.2015.02.039.
37. Koonin, E.V. Viruses and Mobile Elements as Drivers of Evolutionary Transitions. *Philos. Trans. R. Soc. B Biol. Sci.* **2016**, *371*, 20150442, doi:10.1098/rstb.2015.0442.
38. Emerson, J.B.; Roux, S.; Brum, J.R.; Bolduc, B.; Woodcroft, B.J.; Jang, H.B.; Singleton, C.M.; Solden, L.M.; Naas, A.E.; Boyd, J.A.; et al. Host-Linked Soil Viral Ecology along a Permafrost Thaw Gradient. *Nat. Microbiol.* **2018**, *3*, 870–880, doi:10.1038/s41564-018-0190-y.
39. Suttle, C.A. Marine Viruses — Major Players in the Global Ecosystem. *Nat. Rev. Microbiol.* **2007**, *5*, 801–812, doi:10.1038/nrmicro1750.
40. Laber, C.P.; Hunter, J.E.; Carvalho, F.; Collins, J.R.; Hunter, E.J.; Schieler, B.M.; Boss, E.; More, K.; Frada, M.; Thamatrakoln, K.; et al. Coccolithovirus Facilitation of Carbon Export in the North Atlantic. *Nat. Microbiol.* **2018**, *3*, 537–547, doi:10.1038/s41564-018-0128-4.

41. Weitz, J.S.; Stock, C.A.; Wilhelm, S.W.; Bourouiba, L.; Coleman, M.L.; Buchan, A.; Follows, M.J.; Fuhrman, J.A.; Jover, L.F.; Lennon, J.T.; et al. A Multitrophic Model to Quantify the Effects of Marine Viruses on Microbial Food Webs and Ecosystem Processes. *ISME J.* **2015**, *9*, 1352–1364, doi:10.1038/ismej.2014.220.
42. Gao, Y.; Lu, Y.; Dungait, J.A.J.; Liu, J.; Lin, S.; Jia, J.; Yu, G. The “Regulator” Function of Viruses on Ecosystem Carbon Cycling in the Anthropocene. *Front. Public Health* **2022**, *10*, 858615, doi:10.3389/fpubh.2022.858615.
43. Tokodi, N.; Łobodzińska, A.; Klimczak, B.; Antosiak, A.; Młynarska, S.; Šulčius, S.; Avrani, S.; Yoshida, T.; Dziga, D. Proliferative and Viability Effects of Two Cyanophages on Freshwater Bloom-Forming Species *Microcystis Aeruginosa* and *Raphidiopsis Raciborskii* Vary between Strains. *Sci. Rep.* **2025**, *15*, 3152, doi:10.1038/s41598-025-87626-z.
44. Wheeler, G.L.; Sturm, D.; Langer, G. *Gephyrocapsa Huxleyi* (*Emiliania Huxleyi*) as a Model System for Coccolithophore Biology. *J. Phycol.* **2023**, *59*, 1123–1129, doi:10.1111/jpy.13404.
45. Hingamp, P.; Grimsley, N.; Acinas, S.G.; Clerissi, C.; Subirana, L.; Poulain, J.; Ferrera, I.; Sarmiento, H.; Villar, E.; Lima-Mendez, G.; et al. Exploring Nucleo-Cytoplasmic Large DNA Viruses in Tara Oceans Microbial Metagenomes. *ISME J.* **2013**, *7*, 1678–1695, doi:10.1038/ismej.2013.59.
46. Nicholls, K.H. Chapter 13 - Haptophyte Algae. In *Freshwater Algae of North America*; Academic Press, 2015; pp. 587–605 ISBN 978-0-12-385876-4.
47. Mirza, S.F.; Staniewski, M.A.; Short, C.M.; Long, A.M.; Chaban, Y.V.; Short, S.M. Isolation and Characterization of a Virus Infecting the Freshwater Algae *Chrysochromulina Parva*. *Virology* **2015**, *486*, 105–115, doi:10.1016/j.virol.2015.09.005.
48. Deodato, C.R.; Barlow, S.B.; Hovde, B.T.; Cattolico, R.A. Naked *Chrysochromulina* (Haptophyta) Isolates from Lake and River Ecosystems: An Electron Microscopic Comparison Including New Observations on the Type Species of This Taxon. *Algal Res.* **2019**, *40*, 101492, doi:10.1016/j.algal.2019.101492.
49. Simon, M.; López-García, P.; Moreira, D.; Jardillier, L. New Haptophyte Lineages and Multiple Independent Colonizations of Freshwater Ecosystems. *Environ. Microbiol. Rep.* **2013**, *5*, 322–332, doi:10.1111/1758-2229.12023.
50. Hansen, L.R.; Kristiansen, J.; Rasmussen, J.V. Potential Toxicity of the Freshwater *Chrysochromulina* Species *C. Parva* (Prymnesiophyceae). **1994**, *287*, 157–159.
51. Shi, X.; Li, S.; Li, H.; Chen, F.; Wu, Q. The Community Structure of Picophytoplankton in Lake Fuxian, a Deep and Oligotrophic Mountain Lake. *Front. Microbiol.* **2019**, *10*, 2016, doi:10.3389/fmicb.2019.02016.
52. Coy, S.R.; Gann, E.R.; Pound, H.L.; Short, S.M.; Wilhelm, S.W. Viruses of Eukaryotic Algae: Diversity, Methods for Detection, and Future Directions. *Viruses* **2018**, *10*, doi:10.3390/v10090487.
53. Derelle, E.; Monier, A.; Cooke, R.; Worden, A.Z.; Grimsley, N.H.; Moreau, H. Diversity of Viruses Infecting the Green Microalga *Ostreococcus Lucimarinus*. *J. Virol.* **2015**, *89*, 5812–5821, doi:10.1128/JVI.00246-15.
54. Kimura, K.; Tomaru, Y. Isolation and Characterization of a Single-Stranded DNA Virus Infecting the Marine Diatom *Chaetoceros* Sp. Strain SS628-11 Isolated from Western JAPAN. *PLoS ONE* **2013**, *8*, e82013, doi:10.1371/journal.pone.0082013.
55. Gallot-Lavallée, L.; Pagarete, A.; Legendre, M.; Santini, S.; Sandaa, R.-A.; Himmelbauer, H.; Ogata, H.; Bratbak, G.; Claverie, J.-M. The 474-Kilobase-Pair Complete Genome Sequence of CeV-01B, a Virus Infecting *Haptolina* (*Chrysochromulina*) *Ericina* (Prymnesiophyceae). *Genome Announc.* **2015**, *3*, e01413-15, doi:10.1128/genomeA.01413-15.

56. Moreau, H.; Piganeau, G.; Desdevises, Y.; Cooke, R.; Derelle, E.; Grimsley, N. Marine Prasinovirus Genomes Show Low Evolutionary Divergence and Acquisition of Protein Metabolism Genes by Horizontal Gene Transfer. *J. Virol.* **2010**, *84*, 12555–12563, doi:10.1128/JVI.01123-10.
57. Stough, J.M.A.; Yutin, N.; Chaban, Y.V.; Moniruzzaman, M.; Gann, E.R.; Pound, H.L.; Steffen, M.M.; Black, J.N.; Koonin, E.V.; Wilhelm, S.W.; et al. Genome and Environmental Activity of a Chrysochromulina Parva Virus and Its Virophages. *Front. Microbiol.* **2019**, *10*, doi:10.3389/fmicb.2019.00703.
58. Andreani, J.; Verneau, J.; Raoult, D.; Levasseur, A.; La Scola, B. Deciphering Viral Presences: Two Novel Partial Giant Viruses Detected in Marine Metagenome and in a Mine Drainage Metagenome. *Virol. J.* **2018**, *15*, 66, doi:10.1186/s12985-018-0976-9.
59. Schulz, F.; Roux, S.; Paez-Espino, D.; Jungbluth, S.; Walsh, D.A.; Denef, V.J.; McMahon, K.D.; Konstantinidis, K.T.; Eloë-Fadrosch, E.A.; Kyripides, N.C.; et al. Giant Virus Diversity and Host Interactions through Global Metagenomics. *Nature* **2020**, *578*, 432–436, doi:10.1038/s41586-020-1957-x.
60. Zhang, W.; Zhou, J.; Liu, T.; Yu, Y.; Pan, Y.; Yan, S.; Wang, Y. Four Novel Algal Virus Genomes Discovered from Yellowstone Lake Metagenomes. *Sci. Rep.* **2015**, *5*, 15131, doi:10.1038/srep15131.
61. Schulz, F.; Andreani, J.; Francis, R.; Boudjemaa, H.; Bou Khalil, J.Y.; Lee, J.; La Scola, B.; Woyke, T. Advantages and Limits of Metagenomic Assembly and Binning of a Giant Virus. *mSystems* **2020**, *5*, 10.1128/msystems.00048-20, doi:10.1128/msystems.00048-20.
62. Moniruzzaman, M.; Wurch, L.L.; Alexander, H.; Dyhrman, S.T.; Gobler, C.J.; Wilhelm, S.W. Virus-Host Relationships of Marine Single-Celled Eukaryotes Resolved from Metatranscriptomics. *Nat. Commun.* **2017**, *8*, 16054, doi:10.1038/ncomms16054.
63. Morens, D.M.; Daszak, P.; Markel, H.; Taubenberger, J.K. Pandemic COVID-19 Joins History's Pandemic Legion. *mBio* **2020**, *11*, e00812-20, doi:10.1128/mBio.00812-20.
64. Morens, D.M.; Daszak, P.; Taubenberger, J.K. Escaping Pandora's Box — Another Novel Coronavirus. *N. Engl. J. Med.* **2020**, *382*, 1293–1295, doi:10.1056/NEJMp2002106.
65. Siff, E.J.; Aghagholi, G.; Gallo Marin, B.; Tobin-Tyler, E.; Poitevien, P. SARS-CoV-2 Transmission: A Sociological Review. *Epidemiol. Infect.* **2020**, *148*, e242, doi:10.1017/S095026882000240X.
66. Mishra, N.P.; Das, S.S.; Yadav, S.; Khan, W.; Afzal, M.; Alarifi, A.; Kenawy, E.-R.; Ansari, M.T.; Hasnain, M.S.; Nayak, A.K. Global Impacts of Pre- and Post-COVID-19 Pandemic: Focus on Socio-Economic Consequences. *Sens. Int.* **2020**, *1*, 100042, doi:10.1016/j.sintl.2020.100042.
67. Bhatia, B.; Sonar, S.; Khan, S.; Bhattacharya, J. Pandemic-Proofing: Intercepting Zoonotic Spillover Events. *Pathogens* **2024**, *13*, 1067, doi:10.3390/pathogens13121067.
68. Ferguson, N.M.; Cummings, D.A.T.; Fraser, C.; Cajka, J.C.; Cooley, P.C.; Burke, D.S. Strategies for Mitigating an Influenza Pandemic. *Nature* **2006**, *442*, 448–452, doi:10.1038/nature04795.
69. Kitajima, M.; Ahmed, W.; Bibby, K.; Carducci, A.; Gerba, C.P.; Hamilton, K.A.; Haramoto, E.; Rose, J.B. SARS-CoV-2 in Wastewater: State of the Knowledge and Research Needs. *Sci. Total Environ.* **2020**, *739*, 139076, doi:10.1016/j.scitotenv.2020.139076.
70. Prado, T.; Rey-Benito, G.; Miagostovich, M.P.; Sato, M.I.Z.; Rajal, V.B.; Filho, C.R.M.; Pereira, A.D.; Barbosa, M.R.F.; Mannarino, C.F.; Da Silva, A.S. Wastewater-Based Epidemiology for Preventing Outbreaks and Epidemics in Latin America – Lessons from the Past and a Look to the Future. *Sci. Total Environ.* **2023**, *865*, 161210, doi:10.1016/j.scitotenv.2022.161210.
71. Parkins, M.D.; Lee, B.E.; Acosta, N.; Bautista, M.; Hubert, C.R.J.; Hruday, S.E.; Frankowski, K.; Pang, X.-L. Wastewater-Based Surveillance as a Tool for Public Health

- Action: SARS-CoV-2 and Beyond. *Clin. Microbiol. Rev.* **2024**, *37*, e00103-22, doi:10.1128/cmr.00103-22.
72. Tyson, J.R.; James, P.; Stoddart, D.; Sparks, N.; Wickenhagen, A.; Hall, G.; Choi, J.H.; Lapointe, H.; Kamelian, K.; Smith, A.D.; et al. Improvements to the ARTIC Multiplex PCR Method for SARS-CoV-2 Genome Sequencing Using Nanopore 2020.
 73. Quick, J.; Lansdowne, L. ARTIC SARS-CoV-2 Sequencing Protocol v4 (LSK114) V.4 2024.
 74. Ladner, J.T.; Sahl, J.W. Towards a Post-Pandemic Future for Global Pathogen Genome Sequencing. *PLOS Biol.* **2023**, *21*, e3002225, doi:10.1371/journal.pbio.3002225.
 75. Berry, I.; Brown, K.A.; Buchan, S.A.; Hohenadel, K.; Kwong, J.C.; Patel, S.; Rosella, L.C.; Mishra, S.; Sander, B. A Better Normal in Canada Will Need a Better Detection System for Emerging and Re-Emerging Respiratory Pathogens. *Can. Med. Assoc. J.* **2022**, *194*, E1250–E1254, doi:10.1503/cmaj.220577.
 76. Keshaviah, A.; Diamond, M.B.; Wade, M.J.; Scarpino, S.V.; Ahmed, W.; Amman, F.; Aruna, O.; Badilla-Aguilar, A.; Bar-Or, I.; Bergthaler, A.; et al. Wastewater Monitoring Can Anchor Global Disease Surveillance Systems. *Lancet Glob. Health* **2023**, *11*, e976–e981, doi:10.1016/S2214-109X(23)00170-5.
 77. 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.; et al. Monitoring Influenza and Respiratory Syncytial Virus in Wastewater. Beyond COVID-19. *Sci. Total Environ.* **2023**, *892*, 164495, doi:10.1016/j.scitotenv.2023.164495.
 78. Hayes, E.K.; Gouthro, M.T.; LeBlanc, J.J.; Gagnon, G.A. Simultaneous Detection of SARS-CoV-2, Influenza A, Respiratory Syncytial Virus, and Measles in Wastewater by Multiplex RT-qPCR. *Sci. Total Environ.* **2023**, *889*, 164261, doi:10.1016/j.scitotenv.2023.164261.
 79. Williams, R.C.; Farkas, K.; Garcia-Delgado, A.; Adwan, L.; Kevill, J.L.; Cross, G.; Weightman, A.J.; Jones, D.L. Simultaneous Detection and Characterization of Common Respiratory Pathogens in Wastewater through Genomic Sequencing. *Water Res.* **2024**, *256*, 121612, doi:10.1016/j.watres.2024.121612.
 80. Lee, A.J.; Carson, S.; Reyne, M.I.; Marshall, A.; Moody, D.; Allen, D.M.; Allingham, P.; Levickas, A.; Fitzgerald, A.; Bell, S.H.; et al. Wastewater Monitoring of Human and Avian Influenza A Viruses in Northern Ireland: A Genomic Surveillance Study. *Lancet Microbe* **2024**, *5*, 100933, doi:10.1016/S2666-5247(24)00175-7.
 81. Allen, D.M.; Reyne, M.I.; Allingham, P.; Levickas, A.; Bell, S.H.; Lock, J.; Coey, J.D.; Carson, S.; Lee, A.J.; McSparron, C.; et al. Genomic Analysis and Surveillance of Respiratory Syncytial Virus Using Wastewater-Based Epidemiology. *J. Infect. Dis.* **2024**, *jiae205*, doi:10.1093/infdis/jiae205.
 82. Cohen, C.; Zar, H.J. Early Life Respiratory Syncytial Virus Disease—a Preventable Burden. *Lancet Infect. Dis.* **2024**, S1473309924002615, doi:10.1016/S1473-3099(24)00261-5.
 83. Averin, A.; Atwood, M.; Sato, R.; Yacisin, K.; Begier, E.; Shea, K.; Curcio, D.; Houde, L.; Weycker, D. Attributable Cost of Adult Respiratory Syncytial Virus Illness Beyond the Acute Phase. *Open Forum Infect. Dis.* **2024**, *11*, ofae097, doi:10.1093/ofid/ofae097.
 84. Macias, A.E.; McElhaney, J.E.; Chaves, S.S.; Nealon, J.; Nunes, M.C.; Samson, S.I.; Seet, B.T.; Weinke, T.; Yu, H. The Disease Burden of Influenza beyond Respiratory Illness. *Vaccine* **2021**, *39*, A6–A14, doi:10.1016/j.vaccine.2020.09.048.
 85. Blanchet Zumofen, M.-H.; Frimpter, J.; Hansen, S.A. Impact of Influenza and Influenza-Like Illness on Work Productivity Outcomes: A Systematic Literature Review. *PharmacoEconomics* **2023**, *41*, 253–273, doi:10.1007/s40273-022-01224-9.
 86. Trautmannsberger, I.; Plagg, B.; Adamek, I.; Mader, S.; De Luca, D.; Esposito, S.; Silfverdal, S.A.; Zimmermann, L.J.I.; Tischer, C.; the ResQ Family study group The

- Multifaceted Burden of Respiratory Syncytial Virus (RSV) Infections in Young Children on the Family: A European Study. *Infect. Dis. Ther.* **2024**, doi:10.1007/s40121-024-00989-0.
87. Tran, P.T.; Nduaguba, S.O.; Wang, Y.; Diaby, V.; Finelli, L.; Choi, Y.; Winterstein, A.G. Economic Burden of Medically Attended Respiratory Syncytial Virus Infections Among Privately Insured Children Under 5 Years of Age in the USA. *Influenza Other Respir. Viruses* **2024**, *18*, e13347, doi:10.1111/irv.13347.
 88. Yan, S.; Weycker, D.; Sokolowski, S. US Healthcare Costs Attributable to Type A and Type B Influenza. *Hum. Vaccines Immunother.* **2017**, *13*, 2041–2047, doi:10.1080/21645515.2017.1345400.
 89. McCarthy, Z.; Xu, S.; Rahman, A.; Bragazzi, N.L.; Corrales-Medina, V.F.; Lee, J.; Seet, B.T.; Neame, D.; Thommes, E.; Heffernan, J.; et al. Modelling the Linkage between Influenza Infection and Cardiovascular Events via Thrombosis. *Sci. Rep.* **2020**, *10*, 14264, doi:10.1038/s41598-020-70753-0.
 90. Xie, Y.; Choi, T.; Al-Aly, Z. Long-Term Outcomes Following Hospital Admission for COVID-19 versus Seasonal Influenza: A Cohort Study. *Lancet Infect. Dis.* **2024**, *24*, 239–255, doi:10.1016/S1473-3099(23)00684-9.
 91. Javanian, M.; Barary, M.; Ghebrehewet, S.; Koppolu, V.; Vasigala, V.; Ebrahimpour, S. A Brief Review of Influenza Virus Infection. *J. Med. Virol.* **2021**, *93*, 4638–4646, doi:10.1002/jmv.26990.
 92. Griffiths, C.; Drews, S.J.; Marchant, D.J. Respiratory Syncytial Virus: Infection, Detection, and New Options for Prevention and Treatment. *Clin. Microbiol. Rev.* **2017**, *30*, 277–319, doi:10.1128/CMR.00010-16.
 93. Blanken, M.O.; Rovers, M.M.; Molenaar, J.M.; Winkler-Seinstra, P.L.; Meijer, A.; Kimpen, J.L.L.; Bont, L. Respiratory Syncytial Virus and Recurrent Wheeze in Healthy Preterm Infants. *N. Engl. J. Med.* **2013**, *368*, 1791–1799, doi:10.1056/NEJMoa1211917.
 94. Ambrosch, A.; Luber, D.; Klawonn, F.; Kabesch, M. Focusing on Severe Infections with the Respiratory Syncytial Virus (RSV) in Adults: Risk Factors, Symptomatology and Clinical Course Compared to Influenza A / B and the Original SARS-CoV-2 Strain. *J. Clin. Virol.* **2023**, *161*, 105399, doi:10.1016/j.jcv.2023.105399.
 95. Salem, N.M.; Jewehan, A.; Aranda, M.A.; Fox, A. Tomato Brown Rugose Fruit Virus Pandemic. *Annu. Rev. Phytopathol.* **2023**, *61*, 137–164, doi:10.1146/annurev-phyto-021622-120703.
 96. Zhang, S.; Griffiths, J.S.; Marchand, G.; Bernards, M.A.; Wang, A. *Tomato Brown Rugose Fruit Virus*: An Emerging and Rapidly Spreading Plant RNA Virus That Threatens Tomato Production Worldwide. *Mol. Plant Pathol.* **2022**, *23*, 1262–1277, doi:10.1111/mpp.13229.
 97. Salem, N.; Mansour, A.; Ciuffo, M.; Falk, B.W.; Turina, M. A New Tobamovirus Infecting Tomato Crops in Jordan. *Arch. Virol.* **2016**, *161*, 503–506, doi:10.1007/s00705-015-2677-7.
 98. García-Estrada, R.S.; Diaz-Lara, A.; Aguilar-Molina, V.H.; Tovar-Pedraza, J.M. Viruses of Economic Impact on Tomato Crops in Mexico: From Diagnosis to Management—A Review. *Viruses* **2022**, *14*, 1251, doi:10.3390/v14061251.
 99. Davino, S.; Caruso, A.G.; Bertacca, S.; Barone, S.; Panno, S. Tomato Brown Rugose Fruit Virus: Seed Transmission Rate and Efficacy of Different Seed Disinfection Treatments. *Plants* **2020**, *9*, 1615, doi:10.3390/plants9111615.
 100. Kibenge, F.S. Emerging Viruses in Aquaculture. *Curr. Opin. Virol.* **2019**, *34*, 97–103, doi:10.1016/j.coviro.2018.12.008.
 101. Houldcroft, C.J.; Beale, M.A.; Breuer, J. Clinical and Biological Insights from Viral Genome Sequencing. *Nat. Rev. Microbiol.* **2017**, *15*, 183–192, doi:10.1038/nrmicro.2016.182.

102. Garza, D.R.; Dutilh, B.E. From Cultured to Uncultured Genome Sequences: Metagenomics and Modeling Microbial Ecosystems. *Cell. Mol. Life Sci.* **2015**, *72*, 4287–4308, doi:10.1007/s00018-015-2004-1.
103. Sijmons, S.; Van Ranst, M.; Maes, P. Genomic and Functional Characteristics of Human Cytomegalovirus Revealed by Next-Generation Sequencing. *Viruses* **2014**, *6*, 1049–1072, doi:10.3390/v6031049.
104. Urban, L.; Perlas, A.; Francino, O.; Martí-Carreras, J.; Muga, B.A.; Mwangi, J.W.; Boykin Okalebo, L.; Stanton, J.L.; Black, A.; Waipara, N.; et al. Real-time Genomics for One Health. *Mol. Syst. Biol.* **2023**, *19*, e11686, doi:10.15252/msb.202311686.
105. Marais, A.; Šafářová, D.; Navrátil, M.; Faure, C.; Cornaggia, D.; Brans, Y.; Suchá, J.; Candresse, T. Complete Genome Sequence of Cherry Virus T, a Novel Cherry-Infecting Tepovirus. *Arch. Virol.* **2020**, *165*, 1711–1714, doi:10.1007/s00705-020-04656-w.
106. Liang, G.; Bushman, F.D. The Human Virome: Assembly, Composition and Host Interactions. *Nat. Rev. Microbiol.* **2021**, *19*, 514–527, doi:10.1038/s41579-021-00536-5.
107. Parameswaran, P.; Charlebois, P.; Tellez, Y.; Nunez, A.; Ryan, E.M.; Malboeuf, C.M.; Levin, J.Z.; Lennon, N.J.; Balmaseda, A.; Harris, E.; et al. Genome-Wide Patterns of Intrahuman Dengue Virus Diversity Reveal Associations with Viral Phylogenetic Clade and Interhost Diversity. *J. Virol.* **2012**, *86*, 8546–8558, doi:10.1128/JVI.00736-12.
108. Stapleton, P.J.; Eshaghi, A.; Seo, C.Y.; Wilson, S.; Harris, T.; Deeks, S.L.; Bolotin, S.; Goneau, L.W.; Gubbay, J.B.; Patel, S.N. Evaluating the Use of Whole Genome Sequencing for the Investigation of a Large Mumps Outbreak in Ontario, Canada. *Sci. Rep.* **2019**, *9*, 12615, doi:10.1038/s41598-019-47740-1.
109. Chen, N.F.G.; Chaguza, C.; Gagne, L.; Doucette, M.; Smole, S.; Buzby, E.; Hall, J.; Ash, S.; Harrington, R.; Cofsky, S.; et al. Development of an Amplicon-Based Sequencing Approach in Response to the Global Emergence of Mpox. *PLOS Biol.* **2023**, *21*, e3002151, doi:10.1371/journal.pbio.3002151.
110. Faria, N.R.; Kraemer, M.U.G.; Hill, S.C.; Goes De Jesus, J.; Aguiar, R.S.; Iani, F.C.M.; Xavier, J.; Quick, J.; Du Plessis, L.; Dellicour, S.; et al. Genomic and Epidemiological Monitoring of Yellow Fever Virus Transmission Potential. *Science* **2018**, *361*, 894–899, doi:10.1126/science.aat7115.
111. Santiago-Rodriguez, T.M.; Hollister, E.B. Unraveling the Viral Dark Matter through Viral Metagenomics. *Front. Immunol.* **2022**, *13*, 1005107, doi:10.3389/fimmu.2022.1005107.
112. Rodrigues, R.A.L.; Andrade, A.C.D.S.P.; Boratto, P.V.D.M.; Trindade, G.D.S.; Kroon, E.G.; Abrahão, J.S. An Anthropocentric View of the Virosphere-Host Relationship. *Front. Microbiol.* **2017**, *8*, 1673, doi:10.3389/fmicb.2017.01673.
113. Jansz, N.; Faulkner, G.J. Viral Genome Sequencing Methods: Benefits and Pitfalls of Current Approaches. *Biochem. Soc. Trans.* **2024**, *52*, 1431–1447, doi:10.1042/BST20231322.
114. Elbehery, A.H.A.; Deng, L. Insights into the Global Freshwater Virome. *Front. Microbiol.* **2022**, *13*, doi:10.3389/fmicb.2022.953500.
115. Pérez-Cataluña, A.; Cuevas-Ferrando, E.; Randazzo, W.; Sánchez, G. Bias of Library Preparation for Virome Characterization in Untreated and Treated Wastewaters. *Sci. Total Environ.* **2021**, *767*, doi:10.1016/j.scitotenv.2020.144589.
116. Rosario, K.; Breitbart, M. Exploring the Viral World through Metagenomics. *Curr. Opin. Virol.* **2011**, *1*, 289–297, doi:10.1016/j.coviro.2011.06.004.
117. Jia, X.; Wang, W.; Wu, M.; Pang, L.; Yang, C.; Ling, Y.; Yi, Z.; Zhang, X. Empirical Assessment of the Enrichment-Based Metagenomic Methods in Identifying Diverse Respiratory Pathogens. *Sci. Rep.* **2024**, *14*, 24493, doi:10.1038/s41598-024-75120-x.
118. Ogunbayo, A.E.; Sabiu, S.; Nyaga, M.M. Evaluation of Extraction and Enrichment Methods for Recovery of Respiratory RNA Viruses in a Metagenomics Approach. *J. Virol. Methods* **2023**, *314*, 114677, doi:10.1016/j.jviromet.2023.114677.

119. Child, H.T.; Airey, G.; Maloney, D.M.; Parker, A.; Wild, J.; McGinley, S.; Evens, N.; Porter, J.; Templeton, K.; Paterson, S.; et al. Comparison of Metagenomic and Targeted Methods for Sequencing Human Pathogenic Viruses from Wastewater. *mBio* **2023**, *14*, e01468-23, doi:10.1128/mbio.01468-23.
120. Quick, J.; Grubaugh, N.D.; Pullan, S.T.; Claro, I.M.; Smith, A.D.; Gangavarapu, K.; Oliveira, G.; Robles-Sikisaka, R.; Rogers, T.F.; Beutler, N.A.; et al. Multiplex PCR Method for MinION and Illumina Sequencing of Zika and Other Virus Genomes Directly from Clinical Samples. *Nat. Protoc.* **2017**, *12*, 1261–1276, doi:10.1038/nprot.2017.066.
121. Quick, J.; Loman, N.J.; Duraffour, S.; Simpson, J.T.; Severi, E.; Cowley, L.; Bore, J.A.; Koundouno, R.; Dudas, G.; Mikhail, A.; et al. Real-Time, Portable Genome Sequencing for Ebola Surveillance. *Nature* **2016**, *530*, 228–232, doi:10.1038/nature16996.
122. Karthikeyan, S.; Levy, J.I.; De Hoff, P.; Humphrey, G.; Birmingham, A.; Jepsen, K.; Farmer, S.; Tubb, H.M.; Valles, T.; Tribelhorn, C.E.; et al. Wastewater Sequencing Reveals Early Cryptic SARS-CoV-2 Variant Transmission. *Nature* **2022**, *609*, 101–108, doi:10.1038/s41586-022-05049-6.
123. Alathari, S.; Chaput, D.L.; Bolaños, L.M.; Joseph, A.; Jackson, V.L.N.; Verner-Jeffreys, D.; Paley, R.; Tyler, C.R.; Temperton, B. A Multiplexed, Tiled PCR Method for Rapid Whole-Genome Sequencing of Infectious Spleen and Kidney Necrosis Virus (ISKNV) in Tilapia. *Viruses* **2023**, *15*, 965, doi:10.3390/v15040965.
124. Nash, D.; Ellmen, I.; Knapp, J.J.; Menon, R.; Overton, A.K.; Cheng, J.; Lynch, M.D.J.; Nissimov, J.I.; Charles, T.C. A Novel Tiled Amplicon Sequencing Assay Targeting the Tomato Brown Rugose Fruit Virus (ToBRFV) Genome Reveals Widespread Distribution in Municipal Wastewater Treatment Systems in the Province of Ontario, Canada. *Viruses* **2024**, *16*, 460, doi:10.3390/v16030460.
125. Nunes-Leite, L.; Liefting, L.W.; Waite, D.W.; Khan, S.; Thompson, J.R. High-Throughput Sequencing Methods for the Detection of Two Strawberry Viruses in Post-Entry Quarantine. *Viruses* **2024**, *16*, 1550, doi:10.3390/v16101550.
126. Huszar, T.I.; Wetton, J.H.; Jobling, M.A. Mitigating the Effects of Reference Sequence Bias in Single-Multiplex Massively Parallel Sequencing of the Mitochondrial DNA Control Region. *Forensic Sci. Int. Genet.* **2019**, *40*, 9–17, doi:10.1016/j.fsigen.2019.01.008.
127. Sikazwe, C.; Neave, M.J.; Michie, A.; Mileto, P.; Wang, J.; Cooper, N.; Levy, A.; Imrie, A.; Baird, R.W.; Currie, B.J.; et al. Molecular Detection and Characterisation of the First Japanese Encephalitis Virus Belonging to Genotype IV Acquired in Australia. *PLoS Negl. Trop. Dis.* **2022**, *16*, e0010754, doi:10.1371/journal.pntd.0010754.
128. Ndiaye, M.; Badji, A.; Dieng, I.; Dolgova, A.S.; Mhamadi, M.; Kirichenko, A.D.; Gladkikh, A.S.; Gaye, A.; Faye, O.; Sall, A.A.; et al. Molecular Detection and Genetic Characterization of Two Dugbe Orthonairovirus Isolates Detected from Ticks in Southern Senegal. *Viruses* **2024**, *16*, 964, doi:10.3390/v16060964.
129. Diagne, M.M.; Ndione, M.H.D.; Mencattelli, G.; Diallo, A.; Ndiaye, E.H.; Di Domenico, M.; Diallo, D.; Kane, M.; Curini, V.; Top, N.M.; et al. Novel Amplicon-Based Sequencing Approach to West Nile Virus. *Viruses* **2023**, *15*, 1261, doi:10.3390/v15061261.
130. Caserta, L.C.; Zhang, J.; Piñeyro, P.; Diel, D.G. Rapid Genotyping of Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) Using MinION Nanopore Sequencing. *PLOS ONE* **2023**, *18*, e0282767, doi:10.1371/journal.pone.0282767.
131. Zhao, K.; Tu, C.; Chen, W.; Liang, H.; Zhang, W.; Wang, Y.; Jin, Y.; Hu, J.; Sun, Y.; Xu, J.; et al. Rapid Identification of Drug-Resistant Tuberculosis Genes Using Direct PCR Amplification and Oxford Nanopore Technology Sequencing. *Can. J. Infect. Dis. Med. Microbiol.* **2022**, *2022*, 1–8, doi:10.1155/2022/7588033.
132. Talts, T.; Moss crop, L.G.; Williams, D.; Tregoning, J.S.; Paulo, W.; Kohli, A.; Williams, T.C.; Hoshler, K.; Ellis, J.; Lusignan, S.D.; et al. Robust and Sensitive Amplicon-Based

- Whole-Genome Sequencing Assay of Respiratory Syncytial Virus Subtype A and B. *Microbiol. Spectr.* **2024**, *12*, e03067-23, doi:10.1128/spectrum.03067-23.
133. Isabel, S.; Eshaghi, A.; Duvvuri, V.R.; Gubbay, J.B.; Cronin, K.; Li, A.; Hasso, M.; Clark, S.T.; Hopkins, J.P.; Patel, S.N.; et al. Targeted Amplification-Based Whole Genome Sequencing of *Monkeypox Virus* in Clinical Specimens. *Microbiol. Spectr.* **2024**, *12*, e02979-23, doi:10.1128/spectrum.02979-23.
 134. Mackie, J.; Kinoti, W.M.; Chahal, S.I.; Lovelock, D.A.; Campbell, P.R.; Tran-Nguyen, L.T.T.; Rodoni, B.C.; Constable, F.E. Targeted Whole Genome Sequencing (TWG-Seq) of Cucumber Green Mottle Mosaic Virus Using Tiled Amplicon Multiplex PCR and Nanopore Sequencing. *Plants* **2022**, *11*, 2716, doi:10.3390/plants11202716.
 135. Cho, A.Y.; Kim, T.-H.; Lee, S.-H.; Lee, H.; Choi, Y.-J.; Seo, Y.-R.; Lee, D.-H.; Hyeon, J.-Y.; Song, C.-S. Whole Genome Sequencing of Avian Metapneumovirus Type B Genomes Directly from Clinical Samples Collected from Chickens in Live Bird Markets Using Multiplex Tiling RT-PCR Method. *Front. Vet. Sci.* **2023**, *10*, 1112552, doi:10.3389/fvets.2023.1112552.
 136. D'Addiego, J.; Curran-French, M.; Smith, J.; Junushov, A.T.; Breininger, I.; Atkinson, B.; Hay, J.; Hewson, R. Whole-Genome Sequencing Surveillance of Siberian Tick-Borne Encephalitis Virus (TBEV) Identifies an Additional Lineage in Kyrgyzstan. *Virus Res.* **2025**, *351*, 199517, doi:10.1016/j.virusres.2024.199517.
 137. Park, K.; Lee, S.-H.; Kim, J.; Lee, J.; Lee, G.-Y.; Cho, S.; Lee, S.H.; Park, K.; No, J.S.; Budhathoki, S.; et al. Multiplex PCR-Based Nanopore Sequencing and Epidemiological Surveillance of Hantaan Orthohantavirus in Apodemus Agrarius, Republic of Korea. *Viruses* **2021**, *13*, 847, doi:10.3390/v13050847.
 138. Sekee, T.R.; Bubuluma, R.; Van Jaarsveldt, D.; Bester, P.A.; Burt, F.J. Multiplex PCR Method for MinION Sequencing of Bagaza Virus Isolated from Wild Caught Mosquitoes in South Africa. *J. Virol. Methods* **2024**, *327*, 114917, doi:10.1016/j.jviromet.2024.114917.
 139. Del Riccio, M.; Caini, S.; Bonaccorsi, G.; Lorini, C.; Paget, J.; Van Der Velden, K.; Meijer, A.; Haag, M.; McGovern, I.; Zanobini, P. Global Analysis of Respiratory Viral Circulation and Timing of Epidemics in the Pre-COVID-19 and COVID-19 Pandemic Eras, Based on Data from the Global Influenza Surveillance and Response System (GISRS). *Int. J. Infect. Dis.* **2024**, *144*, 107052, doi:10.1016/j.ijid.2024.107052.
 140. Bloom-Feshbach, K.; Alonso, W.J.; Charu, V.; Tamerius, J.; Simonsen, L.; Miller, M.A.; Viboud, C. Latitudinal Variations in Seasonal Activity of Influenza and Respiratory Syncytial Virus (RSV): A Global Comparative Review. *PLoS ONE* **2013**, *8*, e54445, doi:10.1371/journal.pone.0054445.
 141. Habbous, S.; Hota, S.; Allen, V.G.; Henry, M.; Hellsten, E. Changes in Hospitalizations and Emergency Department Respiratory Viral Diagnosis Trends before and during the COVID-19 Pandemic in Ontario, Canada. *PLOS ONE* **2023**, *18*, e0287395, doi:10.1371/journal.pone.0287395.
 142. 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.; et al. Population-based Hospitalization Burden Estimates for Respiratory Viruses, 2015–2019. *Influenza Other Respir. Viruses* **2022**, *16*, 1133–1140, doi:10.1111/irv.13040.
 143. *Influenza (Seasonal)*; World Health Organization, 2025;
 144. Li, Y.; Wang, X.; Blau, D.M.; Caballero, M.T.; Feikin, D.R.; Gill, C.J.; Madhi, S.A.; Omer, S.B.; Simões, E.A.F.; Campbell, H.; et al. Global, Regional, and National Disease Burden Estimates of Acute Lower Respiratory Infections Due to Respiratory Syncytial Virus in Children Younger than 5 Years in 2019: A Systematic Analysis. *The Lancet* **2022**, *399*, 2047–2064, doi:10.1016/S0140-6736(22)00478-0.
 145. *Respiratory Syncytial Virus (RSV)*; World Health Organization, 2025;

146. Herring, W.L.; Zhang, Y.; Shinde, V.; Stoddard, J.; Talbird, S.E.; Rosen, B. Clinical and Economic Outcomes Associated with Respiratory Syncytial Virus Vaccination in Older Adults in the United States. *Vaccine* **2022**, *40*, 483–493, doi:10.1016/j.vaccine.2021.12.002.
147. Belongia, E.A.; McLean, H.Q. Influenza Vaccine Effectiveness: Defining the H3N2 Problem. *Clin. Infect. Dis.* **2019**, *69*, 1817–1823, doi:10.1093/cid/ciz411.
148. Schmidt, K.; Ben Moussa, M.; Buckrell, S.; Rahal, A.; Chestley, T.; Bastien, N.; Lee, L. National Influenza Annual Report, Canada, 2022–2023: Canada’s First Fall Epidemic since the 2019–2020 Season. *Can. Commun. Dis. Rep.* **2023**, *49*, 413–424, doi:10.14745/ccdr.v49i10a02.
149. Ben Moussa, M.; Nwosu, A.; Schmidt, K.; Buckrell, S.; Rahal, A.; Lee, L.; Shane, A.; Bastien, N. National Influenza Annual Report 2023–2024: A Focus on Influenza B and Public Health Implications. *Can. Commun. Dis. Rep.* **2024**, *50*, 393–399, doi:10.14745/ccdr.v50i11a03.
150. Bøås, H.; Havdal, L.B.; Størdal, K.; Døllner, H.; Leegaard, T.M.; Bekkevold, T.; Flem, E.; Inchley, C.; Nordbø, S.A.; Rojahn, A.E.; et al. No Association between Disease Severity and Respiratory Syncytial Virus Subtypes RSV-A and RSV-B in Hospitalized Young Children in Norway. *PLOS ONE* **2024**, *19*, e0298104, doi:10.1371/journal.pone.0298104.
151. Ciarlito, C.; Vittucci, A.C.; Antilici, L.; Concato, C.; Di Camillo, C.; Zangari, P.; Villani, A. Respiratory Syncytial Virus A and B: Three Bronchiolitis Seasons in a Third Level Hospital in Italy. *Ital. J. Pediatr.* **2019**, *45*, 115, doi:10.1186/s13052-019-0704-0.
152. Chen, J.; Qiu, X.; Avadhanula, V.; Shepard, S.S.; Kim, D.; Hixson, J.; Piedra, P.A.; Bahl, J. Novel and Extendable Genotyping System for Human Respiratory Syncytial Virus Based on Whole-genome Sequence Analysis. *Influenza Other Respir. Viruses* **2022**, *16*, 492–500, doi:10.1111/irv.12936.
153. Goya, S.; Galiano, M.; Nauwelaers, I.; Trento, A.; Openshaw, P.J.; Mistchenko, A.S.; Zambon, M.; Viegas, M. Toward Unified Molecular Surveillance of RSV: A Proposal for Genotype Definition. *Influenza Other Respir. Viruses* **2020**, *14*, 274–285, doi:10.1111/irv.12715.
154. Goya, S.; Ruis, C.; Neher, R.A.; Meijer, A.; Aziz, A.; Hinrichs, A.S.; Von Gottberg, A.; Roemer, C.; Amoako, D.G.; Acuña, D.; et al. Standardized Phylogenetic Classification of Human Respiratory Syncytial Virus Below the Subgroup Level. *Emerg. Infect. Dis.* **2024**, *30*, doi:10.3201/eid3008.240209.
155. Williams, T.G.S.; Snell, L.B.; Alder, C.; Charalampous, T.; Alcolea-Medina, A.; Sehmi, J.K.; Al-Yaakoubi, N.; Humayun, G.; Miah, S.; Lackenby, A.; et al. Feasibility and Clinical Utility of Local Rapid Nanopore Influenza A Virus Whole Genome Sequencing for Integrated Outbreak Management, Genotypic Resistance Detection and Timely Surveillance. *Microb. Genomics* **2023**, *9*, doi:10.1099/mgen.0.001083.
156. Galli, C.; Ebranati, E.; Pellegrinelli, L.; Airoidi, M.; Veo, C.; Della Ventura, C.; Seiti, A.; Binda, S.; Galli, M.; Zehender, G.; et al. From Clinical Specimen to Whole Genome Sequencing of A(H3N2) Influenza Viruses: A Fast and Reliable High-Throughput Protocol. *Vaccines* **2022**, *10*, 1359, doi:10.3390/vaccines10081359.
157. MacFadden, D.R.; McGeer, A.; Athey, T.; Perusini, S.; Olsha, R.; Li, A.; Eshaghi, A.; Gubbay, J.B.; Hanage, W.P. Use of Genome Sequencing to Define Institutional Influenza Outbreaks, Toronto, Ontario, Canada, 2014–15. *Emerg. Infect. Dis.* **2018**, *24*, 492–497, doi:10.3201/eid2403.171499.
158. Vemula, S.; Zhao, J.; Liu, J.; Wang, X.; Biswas, S.; Hewlett, I. Current Approaches for Diagnosis of Influenza Virus Infections in Humans. *Viruses* **2016**, *8*, 96, doi:10.3390/v8040096.
159. Fall, A.; Han, L.; Yunker, M.; Gong, Y.-N.; Li, T.-J.; Norton, J.M.; Abdullah, O.; Rothman, R.E.; Fenstermacher, K.Z.J.; Morris, C.P.; et al. Evolution of Influenza A(H3N2) Viruses in

- 2 Consecutive Seasons of Genomic Surveillance, 2021–2023. *Open Forum Infect. Dis.* **2023**, *10*, ofad577, doi:10.1093/ofid/ofad577.
160. May, F.; Ginige, S.; Firman, E.; Li, Y.S.; Soonarane, Y.K.; Smoll, N.; Hunter, I.; Pery, B.; Macfarlane, B.; Bladen, T.; et al. Estimating the Incidence of COVID-19, Influenza and Respiratory Syncytial Virus Infection in Three Regions of Queensland, Australia, Winter 2022: Findings from a Novel Longitudinal Testing-Based Sentinel Surveillance Programme. *BMJ Open* **2024**, *14*, e081793, doi:10.1136/bmjopen-2023-081793.
 161. Thampi, N.; Mercier, E.; Paes, B.; Edwards, J.O.; Rodgers-Gray, B.; Delatolla, R. Perspective: The Potential of Wastewater-Based Surveillance as an Economically Feasible Game Changer in Reducing the Global Burden of Pediatric Respiratory Syncytial Virus Infection. *Front. Public Health* **2024**, *11*, 1316531, doi:10.3389/fpubh.2023.1316531.
 162. Mercier, E.; Pisharody, L.; Guy, F.; Wan, S.; Hegazy, N.; D'Aoust, P.M.; Kabir, M.P.; Nguyen, T.B.; Eid, W.; Harvey, B.; et al. Wastewater-Based Surveillance Identifies Start to the Pediatric Respiratory Syncytial Virus Season in Two Cities in Ontario, Canada. *Front. Public Health* **2023**, *11*, 1261165, doi:10.3389/fpubh.2023.1261165.
 163. Zulli, A.; Varkila, M.R.J.; Parsonnet, J.; Wolfe, M.K.; Boehm, A.B. Observations of Respiratory Syncytial Virus (RSV) Nucleic Acids in Wastewater Solids Across the United States in the 2022–2023 Season: Relationships with RSV Infection Positivity and Hospitalization Rates. *ACS EST Water* **2024**, *4*, 1657–1667, doi:10.1021/acsestwater.3c00725.
 164. Nakauchi, M.; Yasui, Y.; Miyoshi, T.; Minagawa, H.; Tanaka, T.; Tashiro, M.; Kageyama, T. One-Step Real-Time Reverse Transcription-PCR Assays for Detecting and Subtyping Pandemic Influenza A/H1N1 2009, Seasonal Influenza A/H1N1, and Seasonal Influenza A/H3N2 Viruses. *J. Virol. Methods* **2011**, *171*, 156–162, doi:10.1016/j.jviromet.2010.10.018.
 165. Zafeiriadou, A.; Kaltsis, L.; Thomaidis, N.S.; Markou, A. Simultaneous Detection of Influenza A, B and Respiratory Syncytial Virus in Wastewater Samples by One-Step Multiplex RT-ddPCR Assay. *Hum. Genomics* **2024**, *18*, 48, doi:10.1186/s40246-024-00614-8.
 166. Malla, B.; Shrestha, S.; Haramoto, E. Optimization of the 5-Plex Digital PCR Workflow for Simultaneous Monitoring of SARS-CoV-2 and Other Pathogenic Viruses in Wastewater. *Sci. Total Environ.* **2024**, *913*, 169746, doi:10.1016/j.scitotenv.2023.169746.
 167. Vo, V.; Harrington, A.; Chang, C.-L.; Baker, H.; Moshi, M.A.; Ghani, N.; Itorralba, J.Y.; Tillett, R.L.; Dahlmann, E.; Basazinew, N.; et al. Identification and Genome Sequencing of an Influenza H3N2 Variant in Wastewater from Elementary Schools during a Surge of Influenza A Cases in Las Vegas, Nevada. *Sci. Total Environ.* **2023**, *872*, 162058, doi:10.1016/j.scitotenv.2023.162058.
 168. Toribio-Avedillo, D.; Gómez-Gómez, C.; Sala-Comorera, L.; Galofré, B.; Muniesa, M. Adapted Methods for Monitoring Influenza Virus and Respiratory Syncytial Virus in Sludge and Wastewater. *Sci. Total Environ.* **2024**, *918*, 170636, doi:10.1016/j.scitotenv.2024.170636.
 169. Hughes, B.; Duong, D.; White, B.J.; Wigginton, K.R.; Chan, E.M.G.; Wolfe, M.K.; Boehm, A.B. Respiratory Syncytial Virus (RSV) RNA in Wastewater Settled Solids Reflects RSV Clinical Positivity Rates. *Environ. Sci. Technol. Lett.* **2022**, *9*, 173–178, doi:10.1021/acs.estlett.1c00963.
 170. Rector, A.; Bloemen, M.; Thijssen, M.; Pussig, B.; Beuselinck, K.; Van Ranst, M.; Wollants, E. Respiratory Viruses in Wastewater Compared with Clinical Samples, Leuven, Belgium. *Emerg. Infect. Dis.* **2024**, *30*, doi:10.3201/eid3001.231011.

171. Li, H.; He, F.; Lv, Z.; Yi, L.; Zhang, Z.; Li, H.; Fu, S. Tailored Wastewater Surveillance Framework Uncovered the Epidemics of Key Pathogens in a Northwestern City of China. *Sci. Total Environ.* **2024**, *926*, 171833, doi:10.1016/j.scitotenv.2024.171833.
172. Boehm, A.B.; Hughes, B.; Duong, D.; Chan-Herur, V.; Buchman, A.; Wolfe, M.K.; White, B.J. Wastewater Concentrations of Human Influenza, Metapneumovirus, Parainfluenza, Respiratory Syncytial Virus, Rhinovirus, and Seasonal Coronavirus Nucleic-Acids during the COVID-19 Pandemic: A Surveillance Study. *Lancet Microbe* **2023**, *4*, e340–e348, doi:10.1016/S2666-5247(22)00386-X.
173. De Melo, T.; Islam, G.; Simmons, D.B.D.; Desaulniers, J.-P.; Kirkwood, A.E. An Alternative Method for Monitoring and Interpreting Influenza A in Communities Using Wastewater Surveillance. *Front. Public Health* **2023**, *11*, 1141136, doi:10.3389/fpubh.2023.1141136.
174. DeJonge, P.M.; Adams, C.; Pray, I.; Schussman, M.K.; Fahney, R.B.; Shafer, M.; Antkiewicz, D.S.; Roguet, A. Wastewater Surveillance Data as a Complement to Emergency Department Visit Data for Tracking Incidence of Influenza A and Respiratory Syncytial Virus — Wisconsin, August 2022–March 2023. *MMWR Morb. Mortal. Wkly. Rep.* **2023**, *72*, 1005–1009, doi:10.15585/mmwr.mm7237a2.
175. Faherty, E.A.G.; Yuce, D.; Korban, C.; Bemis, K.; Kowalski, R.; Gretsche, S.; Ramirez, E.; Poretsky, R.; Packman, A.; Leisman, K.P.; et al. Correlation of Wastewater Surveillance Data with Traditional Influenza Surveillance Measures in Cook County, Illinois, October 2022–April 2023. *Sci. Total Environ.* **2024**, *912*, 169551, doi:10.1016/j.scitotenv.2023.169551.
176. Germano, E.R.; Flores, T.; Freed, G.S.; Kim, K.; Tulinsky, G.H.; Yang, A.; Rose, O.J.; Ray, C.A.; Autry, A.; Catallozzi, M.; et al. Building-Level Wastewater Surveillance Localizes Interseasonal Influenza Variation. *mSphere* **2024**, *9*, e00600-23, doi:10.1128/msphere.00600-23.
177. Girón-Guzmán, I.; Cuevas-Ferrando, E.; Barranquero, R.; Díaz-Reolid, A.; Puchades-Colera, P.; Falcó, I.; Pérez-Cataluña, A.; Sánchez, G. Urban Wastewater-Based Epidemiology for Multi-Viral Pathogen Surveillance in the Valencian Region, Spain. *Water Res.* **2024**, *255*, 121463, doi:10.1016/j.watres.2024.121463.
178. Dlamini, M.; Msolo, L.; Ehi Ebomah, K.; Nontongana, N.; Ifeanyi Okoh, A. A Systematic Review on the Incidence of Influenza Viruses in Wastewater Matrices: Implications for Public Health. *PLOS ONE* **2024**, *19*, e0291900, doi:10.1371/journal.pone.0291900.
179. Ando, H.; Ahmed, W.; Iwamoto, R.; Ando, Y.; Okabe, S.; Kitajima, M. Impact of the COVID-19 Pandemic on the Prevalence of Influenza A and Respiratory Syncytial Viruses Elucidated by Wastewater-Based Epidemiology. *Sci. Total Environ.* **2023**, *880*, 162694, doi:10.1016/j.scitotenv.2023.162694.
180. Boehm, A.B.; Wolfe, M.K.; White, B.J.; Hughes, B.; Duong, D.; Bidwell, A. More than a Tripledemic: Influenza A Virus, Respiratory Syncytial Virus, SARS-CoV-2, and Human Metapneumovirus in Wastewater during Winter 2022–2023. *Environ. Sci. Technol. Lett.* **2023**, *10*, 622–627, doi:10.1021/acs.estlett.3c00385.
181. Lehto, K.-M.; Lämsävaara, A.; Hyder, R.; Luomala, O.; Lipponen, A.; Hokajärvi, A.-M.; Heikinheimo, A.; Pitkänen, T.; Oikarinen, S. Wastewater-Based Surveillance Is an Efficient Monitoring Tool for Tracking Influenza A in the Community. *Water Res.* **2024**, *257*, 121650, doi:10.1016/j.watres.2024.121650.
182. Markt, R.; Stillebacher, F.; Nägele, F.; Kammerer, A.; Peer, N.; Payr, M.; Scheffknecht, C.; Dria, S.; Draxl-Weiskopf, S.; Mayr, M.; et al. Expanding the Pathogen Panel in Wastewater Epidemiology to Influenza and Norovirus. *Viruses* **2023**, *15*, 263, doi:10.3390/v15020263.
183. Mercier, E.; D'Aoust, P.M.; Thakali, O.; Hegazy, N.; Jia, J.-J.; Zhang, Z.; Eid, W.; Plaza-Diaz, J.; Kabir, M.P.; Fang, W.; et al. Municipal and Neighbourhood Level Wastewater

- Surveillance and Subtyping of an Influenza Virus Outbreak. *Sci. Rep.* **2022**, *12*, 15777, doi:10.1038/s41598-022-20076-z.
184. Raya, S.; Malla, B.; Shrestha, S.; Sthapit, N.; Kattel, H.; Sharma, S.T.; Tuladhar, R.; Maharjan, R.; Takeda, T.; Kitajima, M.; et al. Quantification of Multiple Respiratory Viruses in Wastewater in the Kathmandu Valley, Nepal: Potential Implications of Wastewater-Based Epidemiology for Community Disease Surveillance in Developing Countries. *Sci. Total Environ.* **2024**, *920*, 170845, doi:10.1016/j.scitotenv.2024.170845.
 185. Wolken, M.; Sun, T.; McCall, C.; Schneider, R.; Caton, K.; Hundley, C.; Hopkins, L.; Ensor, K.; Domakonda, K.; Kalvapalle, P.; et al. Wastewater Surveillance of SARS-CoV-2 and Influenza in preK-12 Schools Shows School, Community, and Citywide Infections. *Water Res.* **2023**, *231*, 119648, doi:10.1016/j.watres.2023.119648.
 186. Zafeiriadou, A.; Kaltsis, L.; Kostakis, M.; Kapes, V.; Thomaidis, N.S.; Markou, A. Wastewater Surveillance of the Most Common Circulating Respiratory Viruses in Athens: The Impact of COVID-19 on Their Seasonality. *Sci. Total Environ.* **2023**, *900*, 166136, doi:10.1016/j.scitotenv.2023.166136.
 187. Dumke, R.; Geissler, M.; Skupin, A.; Helm, B.; Mayer, R.; Schubert, S.; Oertel, R.; Renner, B.; Dalpke, A.H. Simultaneous Detection of SARS-CoV-2 and Influenza Virus in Wastewater of Two Cities in Southeastern Germany, January to May 2022. *Int. J. Environ. Res. Public Health* **2022**, *19*, 13374, doi:10.3390/ijerph192013374.
 188. Heijnen, L.; Medema, G. Surveillance of Influenza A and the Pandemic Influenza A (H1N1) 2009 in Sewage and Surface Water in the Netherlands. *J. Water Health* **2011**, *9*, 434–442, doi:10.2166/wh.2011.019.
 189. Stockdale, S.R.; Blanchard, A.M.; Nayak, A.; Husain, A.; Nashine, R.; Dudani, H.; McClure, C.P.; Tarr, A.W.; Nag, A.; Meena, E.; et al. RNA-Seq of Untreated Wastewater to Assess COVID-19 and Emerging and Endemic Viruses for Public Health Surveillance. *Lancet Reg. Health - Southeast Asia* **2023**, *14*, 100205, doi:10.1016/j.lansea.2023.100205.
 190. Tisza, M.; Javornik Cregeen, S.; Avadhanula, V.; Zhang, P.; Ayvaz, T.; Feliz, K.; Hoffman, K.L.; Clark, J.R.; Terwilliger, A.; Ross, M.C.; et al. Wastewater Sequencing Reveals Community and Variant Dynamics of the Collective Human Virome. *Nat. Commun.* **2023**, *14*, 6878, doi:10.1038/s41467-023-42064-1.
 191. Wolfe, M.K.; Duong, D.; Bakker, K.M.; Ammerman, M.; Mortenson, L.; Hughes, B.; Arts, P.; Luring, A.S.; Fitzsimmons, W.J.; Bendall, E.; et al. Wastewater-Based Detection of Two Influenza Outbreaks. *Environ. Sci. Technol. Lett.* **2022**, *9*, 687–692, doi:10.1021/acs.estlett.2c00350.
 192. Freyja: Depth-Weighted De-Mixing.
 193. Ferdous, J.; Kunkleman, S.; Taylor, W.; Harris, A.; Gibas, C.J.; Schlueter, J.A. A Gold Standard Dataset and Evaluation of Methods for Lineage Abundance Estimation from Wastewater. *Sci. Total Environ.* **2024**, *948*, 174515, doi:10.1016/j.scitotenv.2024.174515.
 194. Sutcliffe, S.G.; Kraemer, S.A.; Ellmen, I.; Knapp, J.J.; Overton, A.K.; Nash, D.; Nissimov, J.I.; Charles, T.C.; Dreifuss, D.; Topolsky, I.; et al. Tracking SARS-CoV-2 Variants of Concern in Wastewater: An Assessment of Nine Computational Tools Using Simulated Genomic Data. *Microb. Genomics* **2024**, *10*, doi:10.1099/mgen.0.001249.
 195. Alvoc: Abundance Learning for Variants of Concern.
 196. Ellmen, I.; Lynch, M.D.J.; Nash, D.; Cheng, J.; Nissimov, J.I.; Charles, T.C. Alvoc: Estimating Variant of Concern Abundance from SARS-CoV-2 Wastewater Sequencing Data. **2021**, doi:10.1101/2021.06.03.21258306.
 197. Zhou, B.; Donnelly, M.E.; Scholes, D.T.; St. George, K.; Hatta, M.; Kawaoka, Y.; Wentworth, D.E. Single-Reaction Genomic Amplification Accelerates Sequencing and Vaccine Production for Classical and Swine Origin Human Influenza A Viruses. *J. Virol.* **2009**, *83*, 10309–10313, doi:10.1128/JVI.01109-09.

198. Dong, X.; Deng, Y.-M.; Aziz, A.; Whitney, P.; Clark, J.; Harris, P.; Bautista, C.; Costa, A.-M.; Waller, G.; Daley, A.J.; et al. A Simplified, Amplicon-Based Method for Whole Genome Sequencing of Human Respiratory Syncytial Viruses. *J. Clin. Virol.* **2023**, *161*, 105423, doi:10.1016/j.jcv.2023.105423.
199. Köndgen, S.; Oh, D.-Y.; Thürmer, A.; Sedaghatjoo, S.; Patrono, L.V.; Calvignac-Spencer, S.; Biere, B.; Wolff, T.; Dürrwald, R.; Fuchs, S.; et al. A Robust, Scalable, and Cost-Efficient Approach to Whole Genome Sequencing of RSV Directly from Clinical Samples. *J. Clin. Microbiol.* **2024**, *62*, e01111-23, doi:10.1128/jcm.01111-23.
200. Croville, G.; Walch, M.; Sécula, A.; Lèbre, L.; Silva, S.; Filaire, F.; Guérin, J.-L. An Amplicon-Based Nanopore Sequencing Workflow for Rapid Tracking of Avian Influenza Outbreaks, France, 2020-2022. *Front. Cell. Infect. Microbiol.* **2024**, *14*, 1257586, doi:10.3389/fcimb.2024.1257586.
201. Diaz, A.; Enomoto, S.; Romagosa, A.; Sreevatsan, S.; Nelson, M.; Culhane, M.; Torremorell, M. Genome Plasticity of Triple-Reassortant H1N1 Influenza A Virus during Infection of Vaccinated Pigs. *J. Gen. Virol.* **2015**, *96*, 2982–2993, doi:10.1099/jgv.0.000258.
202. Cruz, C.D.; Icochea, M.E.; Espejo, V.; Troncos, G.; Castro-Sanguinetti, G.R.; Schilling, M.A.; Tinoco, Y. Highly Pathogenic Avian Influenza A(H5N1) from Wild Birds, Poultry, and Mammals, Peru. *Emerg. Infect. Dis.* **2023**, *29*, 2572–2576, doi:10.3201/eid2912.230505.
203. Dadonaite, B.; Gilbertson, B.; Knight, M.L.; Trifkovic, S.; Rockman, S.; Laederach, A.; Brown, L.E.; Fodor, E.; Bauer, D.L.V. The Structure of the Influenza A Virus Genome. *Nat. Microbiol.* **2019**, *4*, doi:10.1038/s41564-019-0513-7.
204. Al-leimon, O.; Shihadeh, H.; Yousef, A.A.; Khraim, A.; Siwwad, R. Respiratory Syncytial Virus: A Review of Current Basic and Clinical Knowledge. *Qatar Med. J.* **2024**, *2024*, doi:10.5339/qmj.2024.56.
205. Chen, S. Ultrafast One-pass FASTQ Data Preprocessing, Quality Control, and Deduplication Using Fastp. *iMeta* **2023**, *2*, e107, doi:10.1002/imt2.107.
206. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor. *Bioinformatics* **2018**, *34*, i884–i890, doi:10.1093/bioinformatics/bty560.
207. Li, H.; Durbin, R. Fast and Accurate Long-Read Alignment with Burrows–Wheeler Transform. *Bioinformatics* **2010**, *26*, 589–595, doi:10.1093/bioinformatics/btp698.
208. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M. Twelve Years of SAMtools and BCFtools. *GigaScience* **2021**, *10*, doi:10.1093/gigascience/giab008.
209. Nh13/DWGSIM: Whole Genome Simulator for Next-Generation Sequencing Available online: <https://github.com/nh13/DWGSIM/tree/main> (accessed on 21 February 2024).
210. Martin, M. Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. *EMBnet.journal* **2011**, *17*, 10–12, doi:10.14806/ej.17.1.200.
211. Li, H. New Strategies to Improve Minimap2 Alignment Accuracy. *Bioinformatics* **2021**, *37*, 4572–4574, doi:10.1093/bioinformatics/btab705.
212. Zhang, M.; Roldan-Hernandez, L.; Boehm, A. Persistence of Human Respiratory Viral RNA in Wastewater-Settled Solids. *Appl. Environ. Microbiol.* **2024**, *90*, e02272-23, doi:10.1128/aem.02272-23.
213. Skowronski, D.M.; Zhan, Y.; Kaweski, S.E.; Sabaiduc, S.; Khalid, A.; Olsha, R.; Carazo, S.; Dickinson, J.A.; Mather, R.G.; Charest, H.; et al. 2023/24 Mid-Season Influenza and Omicron XBB.1.5 Vaccine Effectiveness Estimates from the Canadian Sentinel Practitioner Surveillance Network (SPSN). *Eurosurveillance* **2024**, *29*, doi:10.2807/1560-7917.ES.2024.29.7.2400076.
214. Farkas, K.; Kevill, J.L.; Williams, R.C.; Pântea, I.; Ridding, N.; Lambert-Slosarska, K.; Woodhall, N.; Grimsley, J.M.S.; Wade, M.J.; Singer, A.C.; et al. Comparative Assessment

- of Nanotrap and Polyethylene Glycol-Based Virus Concentration in Wastewater Samples. *FEMS Microbes* **2024**, 5, xtae007, doi:10.1093/femsmc/xtae007.
215. Farkas, K.; Pellett, C.; Alex-Sanders, N.; Bridgman, M.T.P.; Corbishley, A.; Grimsley, J.M.S.; Kasprzyk-Hordern, B.; Kevill, J.L.; Pântea, I.; Richardson-O'Neill, I.S.; et al. Comparative Assessment of Filtration- and Precipitation-Based Methods for the Concentration of SARS-CoV-2 and Other Viruses from Wastewater. *Microbiol. Spectr.* **2022**, 10, e01102-22, doi:10.1128/spectrum.01102-22.
 216. Khare, S.; Gurry, C.; Freitas, L.; B Schultz, M.; Bach, G.; Diallo, A.; Akite, N.; Ho, J.; Tc Lee, R.; Yeo, W.; et al. GISAID's Role in Pandemic Response. *China CDC Wkly.* **2021**, 3, 1049–1051, doi:10.46234/ccdcw2021.255.
 217. Zhang, J.; Zhang, L.; Li, J.; Ma, J.; Wang, Y.; Sun, Y.; Ma, C.; Duan, W.; Wang, Q.; Yang, P.; et al. Moderate Effectiveness of Influenza Vaccine in Outpatient Settings: A Test-Negative Study in Beijing, China, 2023/24 Season. *Vaccine* **2025**, 46, 126662, doi:10.1016/j.vaccine.2024.126662.
 218. Martínez-Baz, I.; Navascués, A.; Trobajo-Sanmartín, C.; Pozo, F.; Fernández-Huerta, M.; Olazabal-Arruiz, M.; Argente-Colas, L.; Ezpeleta, G.; Echeverria, A.; Casado, I.; et al. Effectiveness of Influenza Vaccination in Preventing Confirmed Influenza Cases and Hospitalizations in Northern Spain, 2023/24 Season: A Population-Based Test-Negative Case-Control Study. *Int. J. Infect. Dis.* **2025**, 151, 107364, doi:10.1016/j.ijid.2024.107364.
 219. Marron, L.; McKenna, A.; O'Donnell, J.; Joyce, M.; Bennett, C.; Connell, J.; Domegan, L. Influenza Vaccine Effectiveness Against Symptomatic Influenza in Primary Care: A Test Negative Case Control Study Over Two Influenza Seasons 2022/2023 and 2023/2024 in Ireland. *Influenza Other Respir. Viruses* **2024**, 18, e70023, doi:10.1111/irv.70023.
 220. Separovic, L.; Zhan, Y.; Kaweski, S.E.; Sabaiduc, S.; Carazo, S.; Olsha, R.; Mather, R.G.; Dickinson, J.A.; Hasso, M.; Meunier, I.; et al. Interim Estimates of Vaccine Effectiveness against Influenza A(H1N1)Pdm09 and A(H3N2) during a Delayed Influenza Season, Canada, 2024/25. *Eurosurveillance* **2025**, 30, doi:10.2807/1560-7917.ES.2025.30.4.2500059.
 221. Centers of Disease Control and Protection *ACIP Evidence to Recommendations for Use of Protein Subunit RSV Vaccines (GSK Arexvy or Pfizer Abrysvo) in All Adults Aged ≥75 Years and in Adults Aged 60–74 at Increased Risk of Severe RSV Disease*; U.S. Department of Health and Human Services;
 222. Zhu, Q.; Patel, N.K.; McAuliffe, J.M.; Zhu, W.; Wachter, L.; McCarthy, M.P.; Suzich, J.A. Natural Polymorphisms and Resistance-Associated Mutations in the Fusion Protein of Respiratory Syncytial Virus (RSV): Effects on RSV Susceptibility to Palivizumab. *J. Infect. Dis.* **2012**, 205, 635–638, doi:10.1093/infdis/jir790.
 223. Yasui, Y.; Yamaji, Y.; Sawada, A.; Ito, T.; Nakayama, T. Cell Fusion Assay by Expression of Respiratory Syncytial Virus (RSV) Fusion Protein to Analyze the Mutation of Palivizumab-Resistant Strains. *J. Virol. Methods* **2016**, 231, 48–55, doi:10.1016/j.jviromet.2016.01.003.
 224. Holland, L.A.; Holland, S.C.; Smith, M.F.; Leonard, V.R.; Murugan, V.; Nordstrom, L.; Mulrow, M.; Salgado, R.; White, M.; Lim, E.S. Genomic Sequencing Surveillance to Identify Respiratory Syncytial Virus Mutations, Arizona, USA. *Emerg. Infect. Dis.* **2023**, 29, doi:10.3201/eid2911.230836.
 225. Shu, B.; Kirby, M.K.; Davis, W.G.; Warnes, C.; Liddell, J.; Liu, J.; Wu, K.-H.; Hassell, N.; Benitez, A.J.; Wilson, M.M.; et al. Multiplex Real-Time Reverse Transcription PCR for Influenza A Virus, Influenza B Virus, and Severe Acute Respiratory Syndrome Coronavirus 2. *Emerg. Infect. Dis.* **2021**, 27, 1821–1830, doi:10.3201/eid2707.210462.
 226. US Center for Disease Control and Prevention *CDC's Influenza SARS-CoV-2 Multiplex Assay*; US Center for Disease Control and Prevention;

227. Statistics Canada Table 32-10-0456-01 Production and Value of Greenhouse Fruits and Vegetables.
228. Jones, R.A.C. Global Plant Virus Disease Pandemics and Epidemics. *Plants* **2021**, *10*, 233, doi:10.3390/plants10020233.
229. Fresh Tomato Production and Top Producing Countries Available online: <https://www.tridge.com/intelligences/tomato/production> (accessed on 21 February 2024).
230. Fidan, H.; Sarikaya, P.; Yildiz, K.; Topkaya, B.; Erkiş, G.; Calis, O. Robust Molecular Detection of the New Tomato Brown Rugose Fruit Virus in Infected Tomato and Pepper Plants from Turkey. *J. Integr. Agric.* **2021**, *20*, 2170–2179, doi:10.1016/S2095-3119(20)63335-4.
231. Ma, Z.; Zhang, H.; Ding, M.; Zhang, Z.; Yang, X.; Zhou, X. Molecular Characterization and Pathogenicity of an Infectious cDNA Clone of Tomato Brown Rugose Fruit Virus. *Phytopathol. Res.* **2021**, *3*, doi:10.1186/s42483-021-00091-0.
232. Botermans, M.; de Koning, P.P.M.; Oplaat, C.; Fowkes, A.R.; McGreig, S.; Skelton, A.; Adams, I.P.; Fox, A.; De Jonghe, K.; Demers, J.E.; et al. Tomato Brown Rugose Fruit Virus Nextstrain Build Version 3: Rise of a Novel Clade. *PhytoFrontiers™* **2023**, doi:10.1094/phytofr-09-22-0090-a.
233. van de Vossenbergh, B.T.L.H.; Dawood, T.; Woźny, M.; Botermans, M. First Expansion of the Public Tomato Brown Rugose Fruit Virus (ToBRFV) Nextstrain Build; Inclusion of New Genomic and Epidemiological Data. *PhytoFrontiers™* **2021**, *1*, 359–363, doi:10.1094/phytofr-01-21-0005-a.
234. van de Vossenbergh, B.T.L.H.; Visser, M.; Bruinsma, M.; Koenraad, H.M.S.; Westenberg, M.; Botermans, M. Real-Time Tracking of Tomato Brown Rugose Fruit Virus (ToBRFV) Outbreaks in the Netherlands Using Nextstrain. *PLoS ONE* **2020**, *15*, doi:10.1371/journal.pone.0234671.
235. Adams, M.J.; Adkins, S.; Bragard, C.; Gilmer, D.; Li, D.; MacFarlane, S.A.; Wong, S.M.; Melcher, U.; Ratti, C.; Ryu, K.H. ICTV Virus Taxonomy Profile: Virgaviridae. *J. Gen. Virol.* **2017**, *98*, 1999–2000, doi:10.1099/jgv.0.000884.
236. Gibbs, A. Evolution and Origins of Tobamoviruses. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **1999**, *354*, 593–602, doi:10.1098/rstb.1999.0411.
237. Abrahamian, P.; Cai, W.; Nunziata, S.O.; Ling, K.S.; Jaiswal, N.; Mavrodieva, V.A.; Rivera, Y.; Nakhla, M.K. Comparative Analysis of Tomato Brown Rugose Fruit Virus Isolates Shows Limited Genetic Diversity. *Viruses* **2022**, *14*, doi:10.3390/v14122816.
238. Adams, M.J.; Antoniw, J.F.; Kreuze, J. Virgaviridae: A New Family of Rod-Shaped Plant Viruses. *Arch. Virol.* **2009**, *154*, 1967–1972, doi:10.1007/s00705-009-0506-6.
239. Hak, H.; Spiegelman, Z. The Tomato Brown Rugose Fruit Virus Movement Protein Overcomes *Tm-2²* Resistance in Tomato While Attenuating Viral Transport. *Mol. Plant-Microbe Interactions®* **2021**, *34*, 1024–1032, doi:10.1094/MPMI-01-21-0023-R.
240. Bai, G.H.; Lin, S.C.; Hsu, Y.H.; Chen, S.Y. The Human Virome: Viral Metagenomics, Relations with Human Diseases, and Therapeutic Applications. *Viruses* **2022**, *14*, doi:10.3390/v14020278.
241. Pérez-Losada, M.; Arenas, M.; Galán, J.C.; Bracho, M.A.; Hillung, J.; García-González, N.; González-Candelas, F. High-Throughput Sequencing (HTS) for the Analysis of Viral Populations. *Infect. Genet. Evol.* **2020**, *80*, doi:10.1016/j.meegid.2020.104208.
242. Sandybayev, N.; Belousov, V.; Stochkov, V.; Solomadin, M.; Granica, J.; Yegorov, S. Next Generation Sequencing Approaches to Characterize the Respiratory Tract Virome. *Microorganisms* **2022**, *10*, doi:10.3390/microorganisms10122327.
243. Yang, Y.; Walls, S.D.; Gross, S.M.; Schroth, G.P.; Jarman, R.G.; Hang, J. Targeted Sequencing of Respiratory Viruses in Clinical Specimens for Pathogen Identification and Genome-Wide Analysis. In *Methods in Molecular Biology*; Humana Press Inc., 2018; Vol. 1838, pp. 125–140.

244. Kumar, A.; Murthy, S.; Kapoor, A. Evolution of Selective-Sequencing Approaches for Virus Discovery and Virome Analysis. *Virus Res.* **2017**, *239*, 172–179, doi:10.1016/j.virusres.2017.06.005.
245. Lin, X.; Glier, M.; Kuchinski, K.; Ross-Van Mierlo, T.; McVea, D.; Tyson, J.R.; Prystajecky, N.; Ziels, R.M. Assessing Multiplex Tiling PCR Sequencing Approaches for Detecting Genomic Variants of SARS-CoV-2 in Municipal Wastewater. *mSystems* **2021**, *6*, doi:10.1128/msystems.01068-21.
246. ARTIC Network Available online: <https://artic.network/> (accessed on 20 February 2024).
247. Kareinen, L.; Ogola, J.; Kivistö, I.; Smura, T.; Aaltonen, K.; Jääskeläinen, A.J.; Kibiwo, S.; Masika, M.M.; Nyaga, P.; Mwaengo, D.; et al. Range Expansion of Bombali Virus in Mops Condylurus Bats, Kenya, 2019. *Emerg. Infect. Dis.* **2020**, *26*, 3007–3010, doi:10.3201/EID2612.202925.
248. Hill, S.C.; Neto de Vasconcelos, J.; Granja, B.G.; Thézé, J.; Jandondo, D.; Neto, Z.; Mirandela, M.; Sebastião, C. dos S.; Cândido, A.L.M.; Clemente, C.; et al. Early Genomic Detection of Cosmopolitan Genotype of Dengue Virus Serotype 2, Angola, 2018. *Emerg. Infect. Dis.* **2019**, *25*, 784–787, doi:10.3201/eid2504.180958.
249. ARTIC Multiplex PCR (Full Pathogens List) - Laboratory Available online: <https://community.artic.network/t/artic-multiplex-pcr-full-pathogens-list/494> (accessed on 20 February 2024).
250. Bačnik, K.; Kutnjak, D.; Pecman, A.; Mehle, N.; Tušek Žnidarič, M.; Gutiérrez Aguirre, I.; Ravnikar, M. Viromics and Infectivity Analysis Reveal the Release of Infective Plant Viruses from Wastewater into the Environment. *Water Res.* **2020**, *177*, doi:10.1016/j.watres.2020.115628.
251. Mehle, N.; Bačnik, K.; Bajde, I.; Brodarič, J.; Fox, A.; Gutiérrez-Aguirre, I.; Kitek, M.; Kutnjak, D.; Loh, Y.L.; Maksimović Carvalho Ferreira, O.; et al. Tomato Brown Rugose Fruit Virus in Aqueous Environments – Survival and Significance of Water-Mediated Transmission. *Front. Plant Sci.* **2023**, *14*.
252. Rothman, J.A.; Whiteson, K.L. Sequencing and Variant Detection of Eight Abundant Plant-Infecting Tobamoviruses across Southern California Wastewater. *Microbiol. Spectr.* **2022**, doi:10.1128/spectrum.03050-22.
253. Christou, A.; Agüera, A.; Bayona, J.M.; Cytryn, E.; Fotopoulos, V.; Lambropoulou, D.; Manaia, C.M.; Michael, C.; Revitt, M.; Schröder, P.; et al. The Potential Implications of Reclaimed Wastewater Reuse for Irrigation on the Agricultural Environment: The Knowns and Unknowns of the Fate of Antibiotics and Antibiotic Resistant Bacteria and Resistance Genes – A Review. *Water Res.* **2017**, *123*, 448–467, doi:10.1016/j.watres.2017.07.004.
254. Goh, P.S.; Ahmad, N.A.; Lim, J.W.; Liang, Y.Y.; Kang, H.S.; Ismail, A.F.; Arthanareeswaran, G. Microalgae-Enabled Wastewater Remediation and Nutrient Recovery through Membrane Photobioreactors: Recent Achievements and Future Perspective. *Membranes* **2022**, *12*, doi:10.3390/membranes12111094.
255. Zhang, T.; Breitbart, M.; Lee, W.H.; Run, J.Q.; Wei, C.L.; Soh, S.W.L.; Hibberd, M.L.; Liu, E.T.; Rohwer, F.; Ruan, Y. RNA Viral Community in Human Feces: Prevalence of Plant Pathogenic Viruses. *PLoS Biol.* **2006**, *4*, 0108–0118, doi:10.1371/journal.pbio.0040003.
256. Natarajan, A.; Han, A.; Zlitni, S.; Brooks, E.F.; Vance, S.E.; Wolfe, M.; Singh, U.; Jagannathan, P.; Pinsky, B.A.; Boehm, A.; et al. Standardized Preservation, Extraction and Quantification Techniques for Detection of Fecal SARS-CoV-2 RNA. *Nat. Commun.* **2021**, *12*, doi:10.1038/s41467-021-25576-6.
257. Rizzo, D.; Da Lio, D.; Panattoni, A.; Salemi, C.; Cappellini, G.; Bartolini, L.; Parrella, G. Rapid and Sensitive Detection of Tomato Brown Rugose Fruit Virus in Tomato and Pepper Seeds by Reverse Transcription Loop-Mediated Isothermal Amplification Assays (Real Time and Visual) and Comparison With RT-PCR End-Point and RT-qPCR Methods. *Front. Microbiol.* **2021**, *12*, doi:10.3389/fmicb.2021.640932.

258. Babraham Bioinformatics - FastQC A Quality Control Tool for High Throughput Sequence Data Available online: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed on 21 February 2024).
259. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120, doi:10.1093/bioinformatics/btu170.
260. Wood, D.E.; Lu, J.; Langmead, B. Improved Metagenomic Analysis with Kraken 2. *Genome Biol.* **2019**, *20*, doi:10.1186/s13059-019-1891-0.
261. Langmead, B.; Salzberg, S.L. Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359, doi:10.1038/nmeth.1923.
262. Eren, A.M.; Kiehl, E.; Shaiber, A.; Veseli, I.; Miller, S.E.; Schechter, M.S.; Fink, I.; Pan, J.N.; Yousef, M.; Fogarty, E.C.; et al. Community-Led, Integrated, Reproducible Multi-Omics with Anvi'o. *Nat. Microbiol.* **2021**, *6*, 3–6, doi:10.1038/s41564-020-00834-3.
263. Milne, I.; Stephen, G.; Bayer, M.; Cock, P.J.A.; Pritchard, L.; Cardle, L.; Shawand, P.D.; Marshall, D. Using Tablet for Visual Exploration of Second-Generation Sequencing Data. *Brief. Bioinform.* **2013**, *14*, 193–202, doi:10.1093/bib/bbs012.
264. Edgar, R.C. MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797, doi:10.1093/nar/gkh340.
265. Gouy, M.; Tannier, E.; Comte, N.; Parsons, D.P. Seaview Version 5: A Multiplatform Software for Multiple Sequence Alignment, Molecular Phylogenetic Analyses, and Tree Reconciliation. *Methods Mol. Biol. Clifton NJ* **2021**, *2231*, 241–260, doi:10.1007/978-1-0716-1036-7_15.
266. Rencher, A.C.; Christensen, W.F. Methods of Multivariate Analysis | Wiley Series in Probability and Statistics Available online: <https://onlinelibrary.wiley.com/doi/book/10.1002/9781118391686> (accessed on 21 February 2024).
267. Cloud-Based Platform for Biotech R&D | Benchling Available online: <https://www.benchling.com/> (accessed on 21 February 2024).
268. Ochar, K.; Ko, H.C.; Woo, H.J.; Hahn, B.S.; Hur, O. Pepper Mild Mottle Virus: An Infectious Pathogen in Pepper Production and a Potential Indicator of Domestic Water Quality. *Viruses* **2023**, *15*, doi:10.3390/v15020282.
269. Guajardo-Leiva, S.; Chnaiderman, J.; Gaggero, A.; Díe, B. Metagenomic Insights into the Sewage RNA Virosphere of a Large City. *Viruses* **2020**, *12*, doi:10.3390/v12091050.
270. Hirschhorn, J.W.; Avery, A.; Schandl, C.A. Managing a PCR Contamination Event in a Molecular Pathology Laboratory. *Methods Mol. Biol. Clifton NJ* **2023**, *2621*, 15–26, doi:10.1007/978-1-0716-2950-5_2.
271. Rashidan, K.; Guertin, C.; Cabana, J. Granulovirus. In *Encyclopedia of Entomology*; Capinera, J.L., Ed.; Springer Netherlands: Dordrecht, 2008; pp. 1653–1660 ISBN 978-1-4020-6359-6.
272. Shi, M.; Lin, X.-D.; Tian, J.-H.; Chen, L.-J.; Chen, X.; Li, C.-X.; Qin, X.-C.; Li, J.; Cao, J.-P.; Eden, J.-S.; et al. Redefining the Invertebrate RNA Virosphere. *Nature* **2016**, *540*, 539–543, doi:10.1038/nature20167.
273. Karthikeyan, S.; Ronquillo, N.; Belda-Ferre, P.; Alvarado, D.; Javidi, T.; Longhurst, C.A.; Knight, R. High-Throughput Wastewater SARS-CoV-2 Detection Enables Forecasting of Community Infection Dynamics in San Diego County. *mSystems* **2021**, *6*, doi:10.1128/msystems.00045-21.
274. Symonds, E.M.; Sinigalliano, C.; Gidley, M.; Ahmed, W.; McQuaig-Ulrich, S.M.; Breitbart, M. Faecal Pollution along the Southeastern Coast of Florida and Insight into the Use of Pepper Mild Mottle Virus as an Indicator. *J. Appl. Microbiol.* **2016**, *121*, 1469–1481, doi:10.1111/jam.13252.
275. Torii, S.; Oishi, W.; Zhu, Y.; Thakali, O.; Malla, B.; Yu, Z.; Zhao, B.; Arakawa, C.; Kitajima, M.; Hata, A.; et al. Comparison of Five Polyethylene Glycol Precipitation Procedures for

- the RT-qPCR Based Recovery of Murine Hepatitis Virus, Bacteriophage Phi6, and Pepper Mild Mottle Virus as a Surrogate for SARS-CoV-2 from Wastewater. *Sci. Total Environ.* **2022**, *807*, doi:10.1016/j.scitotenv.2021.150722.
276. Maal-Bared, R.; Qiu, Y.; Li, Q.; Gao, T.; Hrudef, S.E.; Bhavanam, S.; Ruecker, N.J.; Ellehoj, E.; Lee, B.E.; Pang, X. Does Normalization of SARS-CoV-2 Concentrations by Pepper Mild Mottle Virus Improve Correlations and Lead Time between Wastewater Surveillance and Clinical Data in Alberta (Canada): Comparing Twelve SARS-CoV-2 Normalization Approaches. *Sci. Total Environ.* **2023**, *856*, doi:10.1016/j.scitotenv.2022.158964.
 277. Ciannella, S.; González-Fernández, C.; Gomez-Pastora, J. Recent Progress on Wastewater-Based Epidemiology for COVID-19 Surveillance: A Systematic Review of Analytical Procedures and Epidemiological Modeling. *Sci. Total Environ.* **2023**, *878*, 162953–162953, doi:10.1016/j.scitotenv.2023.162953.
 278. Oloye, F.F.; Xie, Y.; Challis, J.K.; Femi-Oloye, O.P.; Brinkmann, M.; McPhedran, K.N.; Jones, P.D.; Servos, M.R.; Giesy, J.P. Understanding Common Population Markers for SARS-CoV-2 RNA Normalization in Wastewater – A Review. *Chemosphere* **2023**, *333*, doi:10.1016/j.chemosphere.2023.138682.
 279. Farkas, K.; Walker, D.I.; Adriaenssens, E.M.; McDonald, J.E.; Hillary, L.S.; Malham, S.K.; Jones, D.L. Viral Indicators for Tracking Domestic Wastewater Contamination in the Aquatic Environment. *Water Res.* **2020**, *181*, doi:10.1016/j.watres.2020.115926.
 280. Hsu, S.Y.; Bayati, M.; Li, C.; Hsieh, H.Y.; Belenchia, A.; Klutts, J.; Zemmer, S.A.; Reynolds, M.; Semkiw, E.; Johnson, H.Y.; et al. Biomarkers Selection for Population Normalization in SARS-CoV-2 Wastewater-Based Epidemiology. *Water Res.* **2022**, *223*, doi:10.1016/j.watres.2022.118985.
 281. Kitajima, M.; Sassi, H.P.; Torrey, J.R. Pepper Mild Mottle Virus as a Water Quality Indicator. *Npj Clean Water* **2018**, *1*, doi:10.1038/s41545-018-0019-5.
 282. Rainey, A.L.; Liang, S.; Bisesi, J.H.; Sabo-Attwood, T.; Aurelli, A.T. A Multistate Assessment of Population Normalization Factors for Wastewater-Based Epidemiology of COVID-19. *PLoS One* **2023**, *18*, e0284370–e0284370, doi:10.1371/journal.pone.0284370.
 283. Aguado-García, Y.; Taboada, B.; Morán, P.; Rivera-Gutiérrez, X.; Serrano-Vázquez, A.; Iša, P.; Rojas-Velázquez, L.; Pérez-Juárez, H.; López, S.; Torres, J.; et al. Tobamoviruses Can Be Frequently Present in the Oropharynx and Gut of Infants during Their First Year of Life. *Sci. Rep.* **2020**, *10*, doi:10.1038/s41598-020-70684-w.
 284. Rivera-Gutiérrez, X.; Morán, P.; Taboada, B.; Serrano-Vázquez, A.; Isa, P.; Rojas-Velázquez, L.; Pérez-Juárez, H.; López, S.; Torres, J.; Ximénez, C.; et al. The Fecal and Oropharyngeal Eukaryotic Viromes of Healthy Infants during the First Year of Life Are Personal. *Sci. Rep.* **2023**, *13*, doi:10.1038/s41598-022-26707-9.
 285. Natarajan, A.; Fremin, B.J.; Schmidtke, D.T.; Wolfe, M.K.; Zlitni, S.; Graham, K.E.; Brooks, E.F.; Severyn, C.J.; Sakamoto, K.M.; Lacayo, N.J.; et al. The Tomato Brown Rugose Fruit Virus Movement Protein Gene Is a Novel Microbial Source Tracking Marker. *Appl. Environ. Microbiol.* **2023**, *89*, e0058323, doi:10.1128/aem.00583-23.
 286. Luria, N.; Smith, E.; Reingold, V.; Bekelman, I.; Lapidot, M.; Levin, I.; Elad, N.; Tam, Y.; Sela, N.; Abu-Ras, A.; et al. A New Israeli Tobamovirus Isolate Infects Tomato Plants Harboring Tm-22 Resistance Genes. *PLoS ONE* **2017**, *12*, doi:10.1371/journal.pone.0170429.
 287. Ahmed, W.; Bivins, A.; Korajkic, A.; Metcalfe, S.; Smith, W.J.M.; Simpson, S.L. Comparative Analysis of Adsorption-Extraction (AE) and Nanotrap® Magnetic Virus Particles (NMVP) Workflows for the Recovery of Endogenous Enveloped and Non-Enveloped Viruses in Wastewater. *Sci. Total Environ.* **2023**, *859*, doi:10.1016/j.scitotenv.2022.160072.

288. Nanotrap® Microbiome Particles | Ceres Nanosciences, Inc. | United States Available online: <https://www.ceresnano.com/microbiome> (accessed on 21 February 2024).
289. Fernandez-Cassi, X.; Timoneda, N.; Martínez-Puchol, S.; Rusiñol, M.; Rodriguez-Manzano, J.; Figuerola, N.; Bofill-Mas, S.; Abril, J.F.; Girones, R. Metagenomics for the Study of Viruses in Urban Sewage as a Tool for Public Health Surveillance. *Sci. Total Environ.* **2018**, *618*, doi:10.1016/j.scitotenv.2017.08.249.
290. Hjelmsø, M.H.; Hellmér, M.; Fernandez-Cassi, X.; Timoneda, N.; Lukjancenko, O.; Seidel, M.; Elsässer, D.; Aarestrup, F.M.; Löfström, C.; Bofill-Mas, S.; et al. Evaluation of Methods for the Concentration and Extraction of Viruses from Sewage in the Context of Metagenomic Sequencing. *PLoS ONE* **2017**, *12*, doi:10.1371/journal.pone.0170199.
291. Aylward, F.O.; Abrahão, J.S.; Brussaard, C.P.D.; Fischer, M.G.; Moniruzzaman, M.; Ogata, H.; Suttle, C.A. Taxonomic Update for Giant Viruses in the Order Imitervirales (Phylum Nucleocytoviricota). *Arch. Virol.* **2023**, *168*, 283, doi:10.1007/s00705-023-05906-3.
292. Mönttinen, H.A.M.; Bicep, C.; Williams, T.A.; Hirt, R.P. The Genomes of Nucleocytoplasmic Large DNA Viruses: Viral Evolution Writ Large. *Microb. Genomics* **2021**, *7*, doi:10.1073/pnas.0904716106.
293. Koonin, E.V.; Yutin, N. Origin and Evolution of Eukaryotic Large Nucleo-Cytoplasmic DNA Viruses. *Intervirology* **2010**, *53*, 284–292, doi:10.1159/000312913.
294. Homola, M.; Büttner, C.R.; Füzik, T.; Křepelka, P.; Holbová, R.; Nováček, J.; Chaillet, M.L.; Žák, J.; Grybchuk, D.; Förster, F.; et al. Structure and Replication Cycle of a Virus Infecting Climate--modulating Alga *Emiliania Huxleyi*. *Sci. Adv.* **2024**, *10*.
295. Yoosuf, N.; Yutin, N.; Colson, P.; Shabalina, S.A.; Pagnier, I.; Robert, C.; Azza, S.; Klose, T.; Wong, J.; Rossmann, M.G.; et al. Related Giant Viruses in Distant Locations and Different Habitats: *Acanthamoeba* Polyphaga Moutmouvirus Represents a Third Lineage of the Mimiviridae That Is Close to the Megavirus Lineage. *Genome Biol. Evol.* **2012**, *4*, 1324–1330, doi:10.1093/gbe/evs109.
296. Moniruzzaman, M.; Martinez-Gutierrez, C.A.; Weinheimer, A.R.; Aylward, F.O. Dynamic Genome Evolution and Complex Virocell Metabolism of Globally-Distributed Giant Viruses. *Nat. Commun.* **2020**, *11*, doi:10.1038/s41467-020-15507-2.
297. Raoult, D.; Audic, S.; Robert, C.; Abergel, C.; Renesto, P.; Ogata, H.; La Scola, B.; Suzan, M.; Claverie, J.-M. The 1.2-Megabase Genome Sequence of Mimivirus. *Science* **2004**, *306*, 1344–1350, doi:10.1126/science.1101485.
298. Scola, B.L.; Audic, S.; Robert, C.; Jungang, L.; De Lamballerie, X.; Drancourt, M.; Birtles, R.; Claverie, J.-M.; Raoult, D. A Giant Virus in Amoebae. *Science* **2003**, *299*, 2033–2033, doi:10.1126/science.1081867.
299. Colson, P.; Yutin, N.; Shabalina, S.A.; Robert, C.; Fournous, G.; La Scola, B.; Raoult, D.; Koonin, E.V. Viruses with More Than 1,000 Genes: Mamavirus, a New *Acanthamoeba* Polyphagamimivirus Strain, and Reannotation of Mimivirus Genes. *Genome Biol. Evol.* **2011**, *3*, 737–742, doi:10.1093/gbe/evr048.
300. Kim, K.E.; Joo, H.M.; Lee, T.-K.; Kim, H.-J.; Kim, Y.J.; Kim, B.K.; Ha, S.-Y.; Jung, S.W. Covariance of Marine Nucleocytoplasmic Large DNA Viruses with Eukaryotic Plankton Communities in the Sub-Arctic Kongsfjorden Ecosystem: A Metagenomic Analysis of Marine Microbial Ecosystems. *Microorganisms* **2023**, *11*, doi:10.3390/microorganisms11010169.
301. Liang, J.-L.; Feng, S.; Jia, P.; Lu, J.; Yi, X.; Gao, S.; Wu, Z.; Liao, B.; Shu, W.; Li, J. Unraveling the Habitat Preferences, Ecological Drivers, Potential Hosts, and Auxiliary Metabolism of Soil Giant Viruses across China. *Microbiome* **2024**, *12*, doi:10.1186/s40168-024-01851-8.
302. Endo, H.; Blanc-Mathieu, R.; Li, Y.; Salazar, G.; Henry, N.; Labadie, K.; De Vargas, C.; Sullivan, M.B.; Bowler, C.; Wincker, P.; et al. Biogeography of Marine Giant Viruses

- Reveals Their Interplay with Eukaryotes and Ecological Functions. *Nat. Ecol. Evol.* **2020**, 4, 1639–1649, doi:10.1038/s41559-020-01288-w.
303. Cottrell, M.; Suttle, C. Wide-Spread Occurrence and Clonal Variation in Viruses Which Cause Lysis of a Cosmopolitan, Eukaryotic Marine Phytoplankton *Micromonas Pusilla*. *Mar. Ecol. Prog. Ser.* **1991**, 78, 1–9, doi:10.3354/meps078001.
 304. Tarutani, K.; Nagasaki, K.; Itakura, S.; Yamaguchi, M. Isolation of a Virus Infecting the Novel Shellfish-Killing Dinoflagellate *Heterocapsa Circularisquama*. *Aquat. Microb. Ecol.* **2001**, 23, 103–111, doi:10.3354/ame023103.
 305. Brussaard, C.P.D.; Short, S.M.; Frederickson, C.M.; Suttle, C.A. Isolation and Phylogenetic Analysis of Novel Viruses Infecting the Phytoplankton *Phaeocystis Globosa* (Prymnesiophyceae). *Appl. Environ. Microbiol.* **2004**, 70, 3700–3705, doi:10.1128/AEM.70.6.3700-3705.2004.
 306. Highfield, A.; Evans, C.; Walne, A.; Miller, P.I.; Schroeder, D.C. How Many Coccolithovirus Genotypes Does It Take to Terminate an *Emiliania Huxleyi* Bloom? *Virology* **2014**, 466–467, 138–145, doi:10.1016/j.virol.2014.07.017.
 307. Nissimov, J.I.; Vandzura, R.; Johns, C.T.; Natale, F.; Haramaty, L.; Bidle, K.D. Dynamics of Transparent Exopolymer Particle Production and Aggregation during Viral Infection of the Coccolithophore, *Emiliania Huxleyi*. *Environ. Microbiol.* **2018**, 20, 2880–2897, doi:10.1111/1462-2920.14261.
 308. Parke, M.; Lund, J.W.G.; Manton, I. Observations on the Biology and Fine Structure of the Type Species of Chrysochromulina (C. Parva Lackey) in the English Lake District. *Arch. Für Mikrobiol.* **1962**, 42, 333–352, doi:10.1007/BF00409070.
 309. Hovde, B.T.; Deodato, C.R.; Andersen, R.A.; Starkenburg, S.R.; Barlow, S.B.; Cattolico, R.A. Chrysochromulina: Genomic Assessment and Taxonomic Diagnosis of the Type Species for an Oleaginous Algal Clade. *Algal Res.* **2019**, 37, 307–319, doi:10.1016/j.algal.2018.11.023.
 310. Wick, R.R. Filtlong 2021.
 311. Wick, R.R.; Judd, L.M.; Cerdeira, L.T.; Hawkey, J.; Méric, G.; Vezina, B.; Wyres, K.L.; Holt, K.E. Tricycler: Consensus Long-Read Assemblies for Bacterial Genomes. *Genome Biol.* **2021**, 22, doi:10.1186/s13059-021-02483-z.
 312. Vaser, R.; Šikić, M. Time- and Memory-Efficient Genome Assembly with Raven. *Nat. Comput. Sci.* **2021**, 1, 332–336, doi:10.1038/s43588-021-00073-4.
 313. Wick, R.R.; Holt, K.E. Benchmarking of Long-Read Assemblers for Prokaryote Whole Genome Sequencing. *F1000Research* **2021**, 8, doi:10.12688/f1000research.21782.4.
 314. Kolmogorov, M.; Yuan, J.; Lin, Y.; Pevzner, P.A. Assembly of Long, Error-Prone Reads Using Repeat Graphs. *Nat. Biotechnol.* **2019**, 37, 540–546, doi:10.1038/s41587-019-0072-8.
 315. Koren, S.; Walenz, B.P.; Berlin, K.; Miller, J.R.; Bergman, N.H.; Phillippy, A.M. Canu: Scalable and Accurate Long-Read Assembly via Adaptive *k*-Mer Weighting and Repeat Separation. *Genome Res.* **2017**, 27, 722–736, doi:10.1101/gr.215087.116.
 316. Lee, J.Y.; Kong, M.; Oh, J.; Lim, J.; Chung, S.H.; Kim, J.-M.; Kim, J.-S.; Kim, K.-H.; Yoo, J.-C.; Kwak, W. Comparative Evaluation of Nanopore Polishing Tools for Microbial Genome Assembly and Polishing Strategies for Downstream Analysis. *Sci. Rep.* **2021**, 11, doi:10.1038/s41598-021-00178-w.
 317. Wright, C. Medaka 2023.
 318. Wick, R.R.; Holt, K.E. Polypolish: Short-Read Polishing of Long-Read Bacterial Genome Assemblies. *PLOS Comput. Biol.* **2022**, 18, doi:10.1371/journal.pcbi.1009802.
 319. Besemer, J. GeneMarkS: A Self-Training Method for Prediction of Gene Starts in Microbial Genomes. Implications for Finding Sequence Motifs in Regulatory Regions. *Nucleic Acids Res.* **2001**, 29, 2607–2618, doi:10.1093/nar/29.12.2607.

320. Delcher, A.L.; Bratke, K.A.; Powers, E.C.; Salzberg, S.L. Identifying Bacterial Genes and Endosymbiont DNA with Glimmer. *Bioinformatics* **2007**, *23*, 673–679, doi:10.1093/bioinformatics/btm009.
321. Hyatt, D.; Chen, G.-L.; LoCascio, P.F.; Land, M.L.; Larimer, F.W.; Hauser, L.J. Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification. *BMC Bioinformatics* **2010**, *11*, doi:10.1186/1471-2105-11-119.
322. Rho, M.; Tang, H.; Ye, Y. FragGeneScan: Predicting Genes in Short and Error-Prone Reads. *Nucleic Acids Res.* **2010**, *38*, doi:10.1093/nar/gkq747.
323. Lazeroff, M.; Ryder, G.; Harris, S.L.; Tsourkas, P.K. Phage Commander, an Application for Rapid Gene Identification in Bacteriophage Genomes Using Multiple Programs. *PHAGE* **2021**, *2*, 204–213, doi:10.1089/phage.2020.0044.
324. González-Tortuero, E.; Krishnamurthi, R.; Allison, H.E.; Goodhead, I.B.; James, C.E. Comparative Analysis of Gene Prediction Tools for Viral Genome Annotation 2021.
325. Salisbury, A.; Tsourkas, P.K. A Method for Improving the Accuracy and Efficiency of Bacteriophage Genome Annotation. *Int. J. Mol. Sci.* **2019**, *20*, doi:10.3390/ijms20143391.
326. Altschu, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215*, 403–410, doi:10.1016/S0022-2836(05)80360-2.
327. Chen, I.-M.A.; Chu, K.; Palaniappan, K.; Ratner, A.; Huang, J.; Huntemann, M.; Hajek, P.; Ritter, S.J.; Webb, C.; Wu, D.; et al. The IMG/M Data Management and Analysis System v.7: Content Updates and New Features. *Nucleic Acids Res.* **2023**, *51*, D723–D732, doi:10.1093/nar/gkac976.
328. Paysan-Lafosse, T.; Blum, M.; Chuguransky, S.; Grego, T.; Pinto, B.L.; Salazar, G.A.; Bileschi, M.L.; Bork, P.; Bridge, A.; Colwell, L.; et al. InterPro in 2022. *Nucleic Acids Res.* **2023**, *51*, D418–D427, doi:10.1093/nar/gkac993.
329. Bigot, T.; Temmam, S.; Pérot, P.; Eloit, M. RVDB-Prot, a Reference Viral Protein Database and Its HMM Profiles. *F1000Research* **2019**, *8*, 530, doi:10.12688/f1000research.18776.1.
330. Terzian, P.; Olo Ndela, E.; Galiez, C.; Lossouarn, J.; Pérez Bucio, R.E.; Mom, R.; Toussaint, A.; Petit, M.-A.; Enault, F. PHROG: Families of Prokaryotic Virus Proteins Clustered Using Remote Homology. *NAR Genomics Bioinforma.* **2021**, *3*, doi:10.1093/nargab/lqab067.
331. Trgovec-Greif, L.; Hellinger, H.-J.; Mainguy, J.; Pfundner, A.; Frishman, D.; Kiening, M.; Webster, N.S.; Laffy, P.W.; Feichtinger, M.; Rattei, T. VOGDB—Database of Virus Orthologous Groups. *Viruses* **2024**, *16*, doi:10.3390/v16081191.
332. Zayed, A.A.; Lücking, D.; Mohssen, M.; Cronin, D.; Bolduc, B.; Gregory, A.C.; Hargreaves, K.R.; Piehowski, P.D.; White III, R.A.; Huang, E.L.; et al. Efam: An Expanded, Metaproteome-Supported HMM Profile Database of Viral Protein Families. *Bioinformatics* **2021**, *37*, 4202–4208, doi:10.1093/bioinformatics/btab451.
333. Eddy, S.R. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **2011**, *7*, doi:10.1371/journal.pcbi.1002195.
334. Mirdita, M.; Steinegger, M.; Söding, J. MMseqs2 Desktop and Local Web Server App for Fast, Interactive Sequence Searches. *Bioinformatics* **2019**, *35*, 2856–2858, doi:10.1093/bioinformatics/bty1057.
335. Chan, P.P.; Lin, B.Y.; Mak, A.J.; Lowe, T.M. tRNAscan-SE 2.0: Improved Detection and Functional Classification of Transfer RNA Genes. *Nucleic Acids Res.* **2021**, *49*, 9077–9096, doi:10.1093/nar/gkab688.
336. Guindon, S.; Gascuel, O. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Syst. Biol.* **2003**, *52*, 696–704, doi:10.1080/10635150390235520.
337. Sohn, J.; Nam, J.-W. The Present and Future of de Novo Whole-Genome Assembly. *Brief. Bioinform.* **2016**, *19*, 23–40, doi:10.1093/bib/bbw096.

338. Suttle, C.A.; Chan, A.M. Viruses Infecting the Marine Prymnesiophyte *Chrysochromulina* Spp.: Isolation, Preliminary Characterization and Natural Abundance. *Mar Ecol Prog Ser* **1995**, *118*, 275–282, doi:10.3354/meps118275.
339. Baudoux, A.-C.; Brussaard, C.P.D. Characterization of Different Viruses Infecting the Marine Harmful Algal Bloom Species *Phaeocystis Globosa*. *Virology* **2005**, *341*, 80–90, doi:10.1016/j.virol.2005.07.002.
340. Milrot, E.; Mutsafi, Y.; Fridmann-Sirkis, Y.; Shimoni, E.; Rechav, K.; Gurnon, J.R.; Van Etten, J.L.; Minsky, A. Virus-Host Interactions: Insights from the Replication Cycle of the Large *Paramecium Bursaria Chlorella Virus*: Replication Factories of PBCV-1. *Cell. Microbiol.* **2016**, *18*, 3–16, doi:10.1111/cmi.12486.
341. Ødum, M.T.; Teufel, F.; Thumhuri, V.; Almagro Armenteros, J.J.; Johansen, A.R.; Winther, O.; Nielsen, H. DeepLoc 2.1: Multi-Label Membrane Protein Type Prediction Using Protein Language Models. *Nucleic Acids Res.* **2024**, *52*, W215–W220, doi:10.1093/nar/gkae237.
342. Moniruzzaman, M.; Erazo Garcia, M.P.; Farzad, R.; Ha, A.D.; Jivaji, A.; Karki, S.; Sheyn, U.; Stanton, J.; Minch, B.; Stephens, D.; et al. Virologs, Viral Mimicry, and Virocell Metabolism: The Expanding Scale of Cellular Functions Encoded in the Complex Genomes of Giant Viruses. *FEMS Microbiol. Rev.* **2023**, *47*, doi:10.1093/femsre/fuad053.
343. O'Brien, M.J.; Ansari, A. Critical Involvement of TFIIB in Viral Pathogenesis. *Front. Mol. Biosci.* **2021**, *8*, doi:10.3389/fmolb.2021.669044.
344. Hubbs, A.E.; Wright, C.F. The A2L Intermediate Gene Product Is Required for in Vitro Transcription from a Vaccinia Virus Late Promoter. *J. Virol.* **1996**, *70*, 327–331, doi:10.1128/jvi.70.1.327-331.1996.
345. Iyer, L.M.; Aravind, L.; Koonin, E.V. Common Origin of Four Diverse Families of Large Eukaryotic DNA Viruses. *J. Virol.* **2001**, *75*, 11720–11734, doi:10.1128/JVI.75.23.11720-11734.2001.
346. Keck, J.G. Overexpression, Purification, and Late Transcription Factor Activity of the 17-Kilodalton Protein Encoded by the Vaccinia Virus AIL Gene. *J Virol* **1993**, *67*, 5740–5748.
347. Benarroch, D.; Qiu, Z.R.; Schwer, B.; Shuman, S. Characterization of a Mimivirus RNA Cap Guanine-N2 Methyltransferase. *RNA* **2009**, *15*, 666–674, doi:10.1261/rna.1462109.
348. Zhang, Y.; Calin-Jageman, I.; Gurnon, J.R.; Choi, T.-J.; Adams, B.; Nicholson, A.W.; Van Etten, J.L. Characterization of a Chlorella Virus PBCV-1 Encoded Ribonuclease III. *Virology* **2003**, *317*, 73–83, doi:10.1016/j.virol.2003.08.044.
349. De Souza, R.F.; Iyer, L.M.; Aravind, L. Diversity and Evolution of Chromatin Proteins Encoded by DNA Viruses. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.* **2010**, *1799*, 302–318, doi:10.1016/j.bbagr.2009.10.006.
350. Weynberg, K.; Allen, M.; Wilson, W. Marine Prasinoviruses and Their Tiny Plankton Hosts: A Review. *Viruses* **2017**, *9*, 43, doi:10.3390/v9030043.
351. Weynberg, K.D.; Allen, M.J.; Gilg, I.C.; Scanlan, D.J.; Wilson, W.H. Genome Sequence of *Ostreococcus Tauri* Virus OtV-2 Throws Light on the Role of Picoeukaryote Niche Separation in the Ocean. *J. Virol.* **2011**, *85*, 4520–4529, doi:10.1128/JVI.02131-10.
352. Rajagopal, I.; Ahn, B.-Y.; Moss, B.; Mathews, C.K. Roles of Vaccinia Virus Ribonucleotide Reductase and Glutaredoxin in DNA Precursor Biosynthesis. *J. Biol. Chem.* **1995**, *270*, 27415–27418, doi:10.1074/jbc.270.46.27415.
353. White, C.L.; Senkevich, T.G.; Moss, B. Vaccinia Virus G4L Glutaredoxin Is an Essential Intermediate of a Cytoplasmic Disulfide Bond Pathway Required for Virion Assembly. *J. Virol.* **2002**, *76*, 467–472, doi:10.1128/JVI.76.2.467-472.2002.
354. Bessman, M.J.; Frick, D.N.; O'Handley, S.F. The MutT Proteins or “Nudix” Hydrolases, a Family of Versatile, Widely Distributed, “Housecleaning” Enzymes. *J. Biol. Chem.* **1996**, *271*, 25059–25062, doi:10.1074/jbc.271.41.25059.

355. Kago, G.; Parrish, S. The Mimivirus L375 Nudix Enzyme Hydrolyzes the 5' mRNA Cap. *PLOS ONE* **2021**, *16*, doi:10.1371/journal.pone.0245820.
356. Jeudy, S.; Rigou, S.; Alempic, J.-M.; Claverie, J.-M.; Abergel, C.; Legendre, M. The DNA Methylation Landscape of Giant Viruses. *Nat. Commun.* **2020**, *11*, doi:10.1038/s41467-020-16414-2.
357. Agarkova, I.V.; Dunigan, D.D.; Van Etten, J.L. Virion-Associated Restriction Endonucleases of Chloroviruses. *J. Virol.* **2006**, *80*, 8114–8123, doi:10.1128/JVI.00486-06.
358. Timinskas, A.; Butkus, V.; Janulaitis, A. Sequence Motifs Characteristic for DNA [Cytosine-N4] and DNA [Adenine-N6] Methyltransferases. Classification of All DNA Methyltransferases. *Gene* **1995**, *157*, 3–11, doi:10.1016/0378-1119(94)00783-O.
359. Ratel, D.; Ravanat, J.-L.; Berger, F.; Wion, D. N6-Methyladenine: The Other Methylated Base of DNA. *BioEssays* **2006**, *28*, 309–315, doi:10.1002/bies.20342.
360. Van Etten, J.L.; Dunigan, D.D. Chloroviruses: Not Your Everyday Plant Virus. *Trends Plant Sci.* **2012**, *17*, 1–8, doi:10.1016/j.tplants.2011.10.005.
361. Cheng, X. Structure and Function of DNA Methyltransferases. *Annu. Rev. Biophys.* **1995**, *24*, 293–318, doi:10.1146/annurev.bb.24.060195.001453.
362. Kumar, S.; Cheng, X.; Klimasauskas, S.; Sha, M.; Posfai, J.; Roberts, R.J.; Wilson, G.G. The DNA (Cytosine-5) Methyltransferases. *Nucleic Acids Res.* **1994**, *22*, 1–10, doi:10.1093/nar/22.1.1.
363. Flusberg, B.A.; Webster, D.R.; Lee, J.H.; Travers, K.J.; Olivares, E.C.; Clark, T.A.; Korlach, J.; Turner, S.W. Direct Detection of DNA Methylation during Single-Molecule, Real-Time Sequencing. *Nat. Methods* **2010**, *7*, 461–465, doi:10.1038/nmeth.1459.
364. Ban, Y.H.; Shinde, P.B.; Hwang, J.; Song, M.-C.; Kim, D.H.; Lim, S.-K.; Sohng, J.K.; Yoon, Y.J. Characterization of FK506 Biosynthetic Intermediates Involved in Post-PKS Elaboration. *J. Nat. Prod.* **2013**, *76*, 1091–1098, doi:10.1021/np4001224.
365. Chen, D.; Zhang, L.; Pang, B.; Chen, J.; Xu, Z.; Abe, I.; Liu, W. FK506 Maturation Involves a Cytochrome P450 Protein-Catalyzed Four-Electron C-9 Oxidation in Parallel with a C-31 O-Methylation. *J. Bacteriol.* **2013**, *195*, 1931–1939, doi:10.1128/JB.00033-13.
366. Weynberg, K.D.; Allen, M.J.; Ashelford, K.; Scanlan, D.J.; Wilson, W.H. From Small Hosts Come Big Viruses: The Complete Genome of a Second *Ostreococcus Tauri* Virus, OtV-1. *Environ. Microbiol.* **2009**, *11*, 2821–2839, doi:10.1111/j.1462-2920.2009.01991.x.
367. Schubert, H.L.; Blumenthal, R.M.; Cheng, X. Many Paths to Methyltransfer: A Chronicle of Convergence. *Trends Biochem. Sci.* **2003**, *28*, 329–335, doi:10.1016/S0968-0004(03)00090-2.
368. Martin, J.L.; McMillan, F.M. SAM (Dependent) I AM: The S-Adenosylmethionine-Dependent Methyltransferase Fold. *Curr. Opin. Struct. Biol.* **2003**, *13*.
369. Campbell, J.A.; Davies, G.J.; Bulone, V.; Henrissat, B. A Classification of Nucleotide-Diphospho-Sugar Glycosyltransferases Based on Amino Acid Sequence Similarities. *Biochem. J.* **1997**, *326*, 929–939, doi:10.1042/bj3260929u.
370. Lairson, L.L.; Henrissat, B.; Davies, G.J.; Withers, S.G. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu. Rev. Biochem.* **2008**, *77*, 521–555, doi:10.1146/annurev.biochem.76.061005.092322.
371. Coutinho, P.M.; Deleury, E.; Davies, G.J.; Henrissat, B. An Evolving Hierarchical Family Classification for Glycosyltransferases. *J. Mol. Biol.* **2003**, *328*, 307–317, doi:10.1016/S0022-2836(03)00307-3.
372. Drula, E.; Garron, M.-L.; Dogan, S.; Lombard, V.; Henrissat, B.; Terrapon, N. The Carbohydrate-Active Enzyme Database: Functions and Literature. *Nucleic Acids Res.* **2022**, *50*, D571–D577, doi:10.1093/nar/gkab1045.

373. Brockhausen, I.; Carver, J.P.; Schachter, H. Control of Glycoprotein Synthesis. The Use of Oligosaccharide Substrates and HPLC to Study the Sequential Pathway for *N*-Acetylglucosaminyltransferases I, II, III, IV, V, and VI in the Biosynthesis of Highly Branched *N*-Glycans by Hen Oviduct Membranes. *Biochem. Cell Biol.* **1988**, *66*, 1134–1151, doi:10.1139/o88-131.
374. Narasimhan, S. Control of Glycoprotein Synthesis. UDP-GlcNAc:Glycopeptide Beta 4-N-Acetylglucosaminyltransferase III, an Enzyme in Hen Oviduct Which Adds GlcNAc in Beta 1-4 Linkage to the Beta-Linked Mannose of the Trimannosyl Core of N-Glycosyl Oligosaccharides. *J. Biol. Chem.* **1982**, *257*, 10235–10242, doi:10.1016/S0021-9258(18)34010-9.
375. Bagdonaite, I.; Wandall, H.H. Global Aspects of Viral Glycosylation. *Glycobiology* **2018**, *28*, 443–467, doi:10.1093/glycob/cwy021.
376. Speciale, I.; Laugieri, M.E.; Noel, E.; Lin, S.; Lowary, T.L.; Molinaro, A.; Duncan, G.A.; Agarkova, I.V.; Garozzo, D.; Tonetti, M.G.; et al. Chlorovirus PBCV-1 Protein A064R Has Three of the Transferase Activities Necessary to Synthesize Its Capsid Protein N-Linked Glycans. *Proc. Natl. Acad. Sci.* **2020**, *117*, 28735–28742, doi:10.1073/pnas.2016626117.
377. Poulin, F.; Gingras, A.-C.; Olsen, H.; Chevalier, S.; Sonenberg, N. 4E-BP3, a New Member of the Eukaryotic Initiation Factor 4E-Binding Protein Family. *J. Biol. Chem.* **1998**, *273*, 14002–14007, doi:10.1074/jbc.273.22.14002.
378. Knox, C.; Luke, G.A.; Blatch, G.L.; Pesce, E.-R. Heat Shock Protein 40 (Hsp40) Plays a Key Role in the Virus Life Cycle. *Virus Res.* **2011**, *160*, 15–24, doi:10.1016/j.virusres.2011.06.013.
379. Nerenberg, B.T.H.; Taylor, J.; Bartee, E.; Gouveia, K.; Barry, M.; Früh, K. The Poxviral RING Protein P28 Is a Ubiquitin Ligase That Targets Ubiquitin to Viral Replication Factories. *J. Virol.* **2005**, *79*, 597–601, doi:10.1128/JVI.79.1.597-601.2005.
380. Bai, C.; Sen, P.; Hofmann, K.; Ma, L.; Goebel, M.; Harper, J.W.; Elledge, S.J. SKP1 Connects Cell Cycle Regulators to the Ubiquitin Proteolysis Machinery through a Novel Motif, the F-Box. *Cell* **1996**, *86*, 263–274, doi:10.1016/S0092-8674(00)80098-7.
381. Iyer, L.M.; Koonin, E.V.; Aravind, L. Novel Predicted Peptidases with a Potential Role in the Ubiquitin Signaling Pathway. *Cell Cycle* **2004**, *3*, 1440–1450, doi:10.4161/cc.3.11.1206.
382. Yutin, N.; Wolf, Y.I.; Raoult, D.; Koonin, E.V. Eukaryotic Large Nucleo-Cytoplasmic DNA Viruses: Clusters of Orthologous Genes and Reconstruction of Viral Genome Evolution. *Virol. J.* **2009**, *6*, 223, doi:10.1186/1743-422X-6-223.
383. Mignery, G.A.; Pikaard, C.S.; Park, W.D. Molecular Characterization of the Patatin Multigene Family of Potato. *Gene* **1988**, *62*, 27–44, doi:10.1016/0378-1119(88)90577-X.
384. Banerji, S.; Aurass, P.; Flieger, A. The Manifold Phospholipases A of Legionella Pneumophila – Identification, Export, Regulation, and Their Link to Bacterial Virulence. *Int. J. Med. Microbiol.* **2008**, *298*, 169–181, doi:10.1016/j.ijmm.2007.11.004.
385. Colson, P.; De Lamballerie, X.; Yutin, N.; Asgari, S.; Bigot, Y.; Bideshi, D.K.; Cheng, X.-W.; Federici, B.A.; Van Etten, J.L.; Koonin, E.V.; et al. “Megavirales”, a Proposed New Order for Eukaryotic Nucleocytoplasmic Large DNA Viruses. *Arch. Virol.* **2013**, *158*, 2517–2521, doi:10.1007/s00705-013-1768-6.
386. Cassetti, M.C.; Merchlinsky, M.; Wolffe, E.J.; Weisberg, A.S.; Moss, B. DNA Packaging Mutant: Repression of the Vaccinia Virus A32 Gene Results in Noninfectious, DNA-Deficient, Spherical, Enveloped Particles. *J. Virol.* **1998**, *72*, 5769–5780, doi:10.1128/JVI.72.7.5769-5780.1998.
387. Shenouda, M.M.; Noyce, R.S.; Lee, S.Z.; Wang, J.L.; Lin, Y.-C.; Favis, N.A.; Desaulniers, M.A.; Evans, D.H. The Mismatched Nucleotides Encoded in Vaccinia Virus Flip-and-Flop Hairpin Telomeres Serve an Essential Role in Virion Maturation. *PLOS Pathog.* **2022**, *18*, doi:10.1371/journal.ppat.1010392.

388. De La Vega, I.; González, A.; Blasco, R.; Calvo, V.; Viñuela, E. Nucleotide Sequence and Variability of the Inverted Terminal Repetitions of African Swine Fever Virus DNA. *Virology* **1994**, *201*, 152–156, doi:10.1006/viro.1994.1277.
389. Rey Redondo, E.; Leung, S.K.K.; Yung, C.C.M. Genomic and Biogeographic Characterisation of the Novel Prasinovirus *MANTONIELLA TINHAUANA* Virus 1. *Environ. Microbiol. Rep.* **2024**, *16*, e70020, doi:10.1111/1758-2229.70020.
390. Zhang, Y.; Strasser, P.; Grabherr, R.; Van Etten, J.L. Hairpin Loop Structure at the Termini of the Chlorella Virus PBCV-1 Genome. *Virology* **1994**, *202*, 1079–1082, doi:10.1006/viro.1994.1444.
391. Zhang, W.; Zhou, J.; Liu, T.; Yu, Y.; Pan, Y.; Yan, S.; Wang, Y. Four Novel Algal Virus Genomes Discovered from Yellowstone Lake Metagenomes. *Sci. Rep.* **2015**, *5*, 15131, doi:10.1038/srep15131.
392. Xia, Y.; Cheng, H.; Zhong, J. Hybrid Sequencing Resolved Inverted Terminal Repeats in the Genome of Megavirus Baoshan. *Front. Microbiol.* **2022**, *13*, 831659, doi:10.3389/fmicb.2022.831659.
393. Wilson, W.H.; Van Etten, J.L.; Allen, M.J. The Phycodnaviridae: The Story of How Tiny Giants Rule the World. In *Lesser Known Large dsDNA Viruses*; Van Etten, J.L., Ed.; Current Topics in Microbiology and Immunology; Springer Berlin Heidelberg: Berlin, Heidelberg, 2009; Vol. 328, pp. 1–42 ISBN 978-3-540-68617-0.
394. DNA Secondary Structure.
395. Rux, J.J.; Kuser, P.R.; Burnett, R.M. Structural and Phylogenetic Analysis of Adenovirus Hexons by Use of High-Resolution X-Ray Crystallographic, Molecular Modeling, and Sequence-Based Methods. *J. Virol.* **2003**, *77*, 9553–9566, doi:10.1128/JVI.77.17.9553-9566.2003.
396. Shao, Q.; Agarkova, I.V.; Noel, E.A.; Dunigan, D.D.; Liu, Y.; Wang, A.; Guo, M.; Xie, L.; Zhao, X.; Rossmann, M.G.; et al. Near-Atomic, Non-Icosahedrally Averaged Structure of Giant Virus *Paramecium Bursaria* Chlorella Virus 1. *Nat. Commun.* **2022**, *13*, doi:10.1038/s41467-022-34218-4.
397. Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J.E. The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nat. Protoc.* **2015**, *10*, 845–858, doi:10.1038/nprot.2015.053.
398. Xian, Y.; Avila, R.; Pant, A.; Yang, Z.; Xiao, C. The Role of Tape Measure Protein in Nucleocytoplasmic Large DNA Virus Capsid Assembly. *Viral Immunol.* **2021**, *34*, 41–48, doi:10.1089/vim.2020.0038.
399. Zhang, X.; Xiang, Y.; Dunigan, D.D.; Klose, T.; Chipman, P.R.; Van Etten, J.L.; Rossmann, M.G. Three-Dimensional Structure and Function of the *Paramecium Bursaria* Chlorella Virus Capsid. *Proc. Natl. Acad. Sci.* **2011**, *108*, 14837–14842, doi:10.1073/pnas.1107847108.
400. Agarkova, I.V.; Lane, L.C.; Dunigan, D.D.; Quispe, C.F.; Duncan, G.A.; Milrot, E.; Minsky, A.; Esmael, A.; Ghosh, J.S.; Van Etten, J.L. Identification of a Chlorovirus PBCV-1 Protein Involved in Degrading the Host Cell Wall during Virus Infection. *Viruses* **2021**, *13*, doi:10.3390/v13050782.
401. Hakim, M.; Fass, D. Cytosolic Disulfide Bond Formation in Cells Infected with Large Nucleocytoplasmic DNA Viruses. *Antioxid. Redox Signal.* **2010**, *13*, 1261–1271, doi:10.1089/ars.2010.3128.
402. Li, S.-N.; Wang, J.-Y.; Yuan, M.-J.; Yang, K. Disruption of the Baculovirus Core Gene Ac78 Results in Decreased Production of Multiple Nucleocapsid-Enveloped Occlusion-Derived Virions and the Failure of Primary Infection in Vivo. *Virus Res.* **2014**, *191*, 70–82, doi:10.1016/j.virusres.2014.07.019.
403. Santini, S.; Jeudy, S.; Bartoli, J.; Poirot, O.; Lescot, M.; Abergel, C.; Barbe, V.; Wommack, K.E.; Noordeloos, A.A.M.; Brussaard, C.P.D.; et al. Genome of *Phaeocystis*

Globosa Virus PgV-16T Highlights the Common Ancestry of the Largest Known DNA Viruses Infecting Eukaryotes. *Proc. Natl. Acad. Sci.* **2013**, *110*, 10800–10805, doi:10.1073/pnas.1303251110.

Appendices

Supplementary Table 2.1. *Locations of Wastewater Sample Collection.*

Region	Collection Site	Collection System	Sample Type
Peel	GE Booth	Wastewater Treatment Plant	24 h composite
York	Leslie Street (York 1)	Linear Pipe Collection Site	Grab
York	Warden (York 2)	Pumping Station	Grab
Waterloo	Kitchener	Wastewater Treatment Plant	24 h composite
Waterloo	Cambridge	Wastewater Treatment Plant	24 h composite
Hamilton	Hamilton	Wastewater Treatment Plant	24 h composite

Supplementary Table 2.2. *Dates of Wastewater Sample Collection.* The collection date of samples from six sites and dates of surveillance week.

Surveillance Week	Dates of Surveillance Week	GE Booth	York 1	York 2	Kitchener	Cambridge	Hamilton
45	Nov 5 - Nov 11, 2023	Nov 6, 2023	Nov 6, 2023	Nov 6, 2023	Nov 6, 2023	Nov 6, 2023	Nov 7, 2023
46	Nov 12 - Nov 18, 2023	Nov 13, 2023	Nov 13, 2023	Nov 13, 2023	-	-	Nov 14, 2023
47	Nov 19 - Nov 25, 2023	Nov 20, 2023	Nov 20, 2023	Nov 20, 2023	Nov 20, 2023	Nov 20, 2023	Nov 21, 2023
48	Nov 26 - Dec 2, 2023	Nov 27, 2023	Nov 27, 2023	Nov 27, 2023	Nov 27, 2023	Nov 27, 2023	Nov 28, 2023
49	Dec 3 - Dec 9, 2023	Dec 4, 2023	Dec 4, 2023	Dec 4, 2023	Dec 4, 2023	Dec 4, 2023	Dec 5, 2023
50	Dec 10 - Dec 16, 2023	Dec 11, 2023	Dec 11, 2023	Dec 11, 2023	Dec 11, 2023	Dec 11, 2023	Dec 12, 2023
51	Dec 17 - Dec 23, 2023	Dec 18, 2023	Dec 18, 2023	Dec 18, 2023	Dec 18, 2023	Dec 18, 2023	Dec 15, 2023
52	Dec 24 - Dec 30, 2023	Dec 26, 2023	-	-	-	Dec 25, 2023	Dec 27, 2023
1	Dec 31, 2023 - Jan 6, 2024	Jan 3, 2024	Jan 3, 2024	Jan 3, 2024	Jan 1, 2024	-	Jan 2, 2024
2	Jan 7 - Jan 13, 2024	Jan 8, 2024	Jan 8, 2024	Jan 8, 2024	Jan 15, 2024	-	Jan 9, 2024
3	Jan 14 - Jan 20, 2024	Jan 17, 2024	Jan 15, 2024	Jan 15, 2024	Jan 17, 2024	Jan 15, 2024	Jan 16, 2024
4	Jan 21 - Jan 27, 2024	Jan 22, 2024	Jan 22, 2024	Jan 22, 2024	Jan 22, 2024	Jan 22, 2024	Jan 23, 2024
5	Jan 28 - Feb 3, 2024	Jan 29, 2024	Jan 29, 2024	Jan 29, 2024	Jan 29, 2024	Jan 29, 2024	Jan 30, 2024

Supplementary Table 2.3. IAV H3N2 Primers. List of tiled-amplicon IAV H3N2 primers, including, name, amplicon, sequence, size, %GC, melting temperature, segment name, strand, primer pool, and pool concentration.

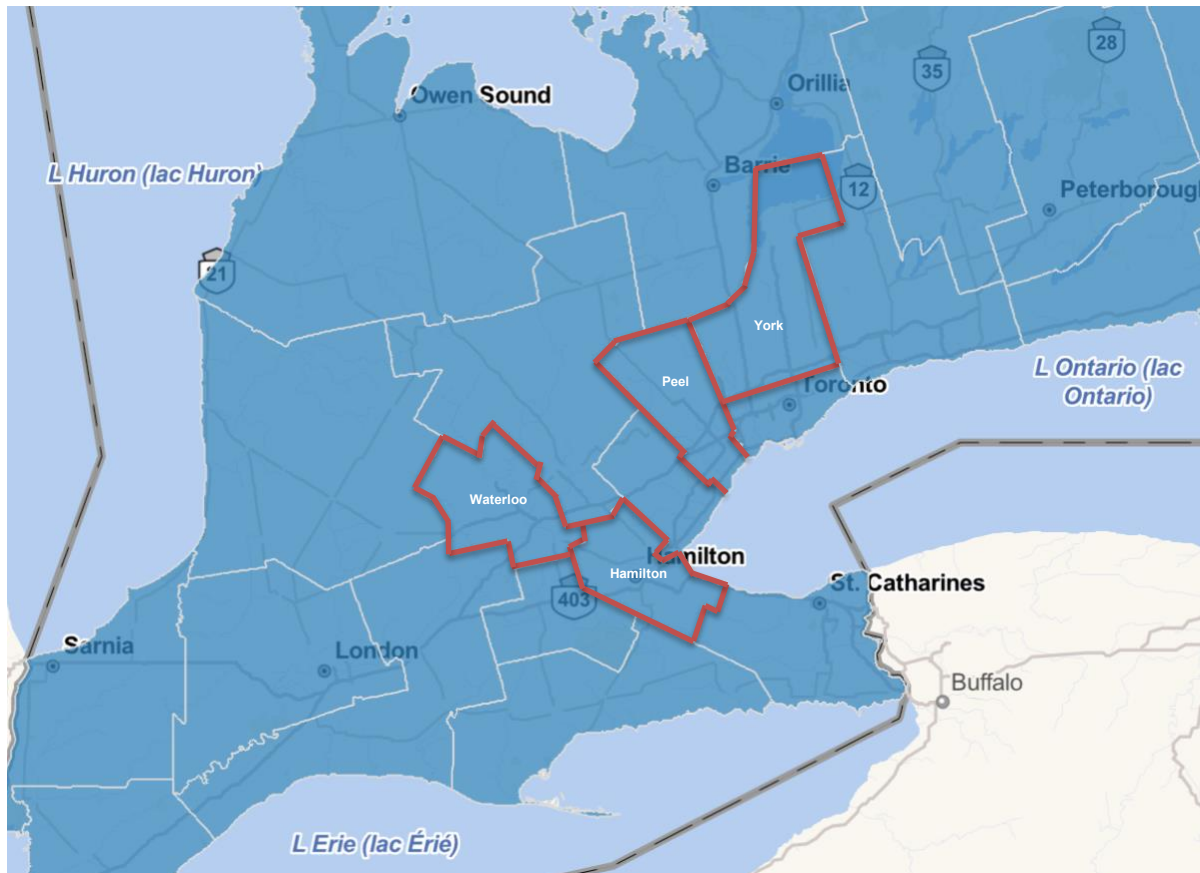
Name	Amplicon	Seq	size	%gc	Tm	Segment Name	Strand	Pool	[pool] (uM)
PB2_1_LEFT	1	AAACCACAGTGGACCATATGGC	22	50	61.07	PB2	+	1	0.4
PB2_1_RIGHT	1	TTGGCACTGAGGTCTGCATGA	21	52.38	62.1	PB2	-	1	0.4
PB2_2_LEFT	2	AACCTTTGGCCCTGTCCATTTT	22	45.45	61.15	PB2	+	2	0.2
PB2_2_RIGHT	2	GGCCGCAATAATTAGGCTTTGG	22	50	60.72	PB2	-	2	0.2
PB2_3_LEFT	3	AGCCAGAATACTAACATCAGAATCACA	27	37.04	60.86	PB2	+	1	0.2
PB2_3_RIGHT	3	AGCTTGTTCTTCAGTCGGGTTC	22	50	60.99	PB2	-	1	0.2
PB2_4_LEFT	4	GCAGTGTGAGCAGATCCACT	20	55	60.05	PB2	+	2	0.4
PB2_4_RIGHT	4	CTTGTGAAAACACCATGGCCAC	22	50	60.98	PB2	-	2	0.4
PB2_5_LEFT	5	ACTAAGAAAAGCAACCAGAAGACTG	25	40	59.85	PB2	+	1	0.4
PB2_5_RIGHT	5	TCTCGAACTCTCAAAAACCGATCA	24	41.67	60.58	PB2	-	1	0.4
PB2_6_LEFT	6	AGGAATAAGAGTCAGCAAAATGGGT	25	40	60.91	PB2	+	2	0.4
PB2_6_RIGHT	6	CCAAGTACGTCTCTCATTTGTTGG	24	45.83	60.16	PB2	-	2	0.4
PB2_7_LEFT	7	AATCTTTAGTCCCCAAGGCCAC	22	50	60.48	PB2	+	1	0.2
PB2_7_RIGHT	7	TTGATGCTTAATGCTGGTCCGT	22	45.45	60.8	PB2	-	1	0.2
PB2_8_LEFT	8	ACTGACTGTGAATGTGAGGGGA	22	50	61.21	PB2	+	2	0.4
PB2_8_RIGHT	8	GAAACAAGGTCGTTTTAACTATTCAGT	29	31.03	60.17	PB2	-	2	0.4
PB1_1_LEFT	9	GTTCCAGCGCAAAATGCCATAA	22	45.45	60.85	PB1	+	1	0.4
PB1_1_RIGHT	9	GACTTCTATGGTGTGGCTAATGC	24	45.83	60.28	PB1	-	1	0.4
PB1_2_LEFT	10	ATGGAAGCCGTTCAACAGACAA	22	45.45	60.93	PB1	+	2	0.4
PB1_2_RIGHT	10	GGGTGTTGCAATAGCCCTTCTT	22	50	61.33	PB1	-	2	0.4
PB1_3_LEFT	11	TGAATAAGAGAGGCTACCTAATAAGAGC	28	39.29	60.61	PB1	+	1	0.4
PB1_3_RIGHT	11	ACATTATTGGTGCGATGCTCAG	22	45.45	59.56	PB1	-	1	0.4
PB1_4_LEFT	12	AAAACCCCCGAATGTTTTGGC	22	45.45	60.93	PB1	+	2	0.2
PB1_4_RIGHT	12	ATGAGGGCAAAATCGTCTGAGG	22	50	60.86	PB1	-	2	0.2
PB1_5_LEFT	13	ACATGCTAAGTACAGTTTtaggagtct	27	37.04	60.54	PB1	+	1	0.3
PB1_5_RIGHT	13	ATTTGGGCTGTTGCTGGTCC	20	55	61.2	PB1	-	1	0.3
PB1_6_LEFT	14	GTCAGCTGACATGAGCATTGGA	22	50	61.12	PB1	+	2	0.2
PB1_6_RIGHT	14	CCATGGGCTGGCATTACTACAG	22	54.55	61.25	PB1	-	2	0.2
PB1_7_LEFT	15	CTTCACATCCCTGAAGTCTGCT	22	50	60.21	PB1	+	1	0.4
PB1_7_RIGHT	15	AATCCTTCCAGACTCGAAGTCAA	23	43.48	59.68	PB1	-	1	0.4
PA_1_LEFT	16	TGGAAGATTTTGTCGACAATGC	23	43.48	61.11	PA	+	1	0.4
PA_1_RIGHT	16	CAAGGTAATATATGTGGACTTCTCCT	28	39.29	60.56	PA	-	1	0.4
PA_2_LEFT	17	GCAACACTACTGGAGCTGGAAA	22	50	60.99	PA	+	2	0.4
PA_2_RIGHT	17	GGCTCTAAAATTCTCAAGGCAGG	23	47.83	59.87	PA	-	2	0.4
PA_3_LEFT	18	AGAAAAATTTGAAATCACAGGAACATATGC	29	31.03	60.47	PA	+	1	0.4
PA_3_RIGHT	18	TCCATGACAGCAGGTAATTTGAATT	25	36	59.67	PA	-	1	0.4
PA_4_LEFT	19	AGTGCATAAAAACATTCTTTGGATGGA	27	33.33	60.75	PA	+	2	0.4
PA_4_RIGHT	19	TTGCAATGTGTTCAATTGGGGC	22	45.45	60.99	PA	-	2	0.4
PA_5_LEFT	20	AACAAGGCCTGTGAGCTAACTG	22	50	60.99	PA	+	1	0.4
PA_5_RIGHT	20	TCTCAAGGACACAATATTTCTCCCA	25	40	60.14	PA	-	1	0.4
PA_6_LEFT	21	GTGGTAACTTTGTGAGCATGGA	23	43.48	59.68	PA	+	2	0.4
PA_6_RIGHT	21	TGCATACAGGCTATTGAATACTGACT	26	38.46	60.63	PA	-	2	0.4
PA_7_LEFT	22	GAGATGAGACGTTGCCTCCTTC	22	54.55	60.92	PA	+	1	0.2
PA_7_RIGHT	22	TGCATGTGTCAGGAAGGAGTTG	22	50	60.99	PA	-	1	0.2
HA_1_LEFT	23	CCATGAAGACTATCATTGCTTTGAGC	26	42.31	61.17	HA	+	1	0.4
HA_1_RIGHT	23	AACTCCAGTGTGCCGGATGA	20	55	61.5	HA	-	1	0.4
HA_2_LEFT	24	TGTTACCCTTATGATGTGCCGG	22	50	60.6	HA	+	2	0.4
HA_2_RIGHT	24	CCAATAGATGCTTATTCTGCTAGGGA	26	42.31	60.74	HA	-	2	0.4
HA_3_LEFT	25	TTCTTTAGTAGATTAATTTGGTTGACCCA	29	31.03	59.96	HA	+	1	0.4

HA_3_RIGHT	25	GGAGTGATGCATTGAGACTTGC	22	50	60.08	HA	-	1	0.4
HA_4_LEFT	26	GGAATCTAATTGCTCCTAGGGGT	23	47.83	59.61	HA	+	2	0.4
HA_4_RIGHT	26	TCAATCGATTGAGCTTCCATTGA	24	41.67	60.95	HA	-	2	0.4
HA_5_LEFT	27	TCATAGAAAATGGTTGGGAGGGAA	24	41.67	60.04	HA	+	1	0.4
HA_5_RIGHT	27	TCACATTTGTGGTATATTTTGAACAACC	29	31.03	60.42	HA	-	1	0.4
HA_6_LEFT	28	GCCCTGGAGAACCAACATACAA	22	50	60.74	HA	+	2	0.2
HA_6_RIGHT	28	CAAGGGTGTTTTAAATTAATGCACTCA	27	33.33	59.83	HA	-	2	0.2
NP_1_LEFT	29	AGCAGGGTTGATAATCACTCACTG	24	45.83	60.95	NP	+	2	0.2
NP_1_RIGHT	29	GCGCCAGATTCGCCTTATTTCT	22	50	61.81	NP	-	2	0.2
NP_2_LEFT	30	TGATGAAAGAAGGAATAAATACCTGGAAG	29	34.48	60.32	NP	+	1	0.4
NP_2_RIGHT	30	AATTTGATCGTTGATCCCCCG	22	50	61.24	NP	-	1	0.4
NP_3_LEFT	31	ATGTGCTCTCTGATGCAGGG	20	55	59.84	NP	+	2	0.2
NP_3_RIGHT	31	TCCCACCAAGGAATATCCCTCT	22	50	60.21	NP	-	2	0.2
NP_4_LEFT	32	TCACAAATCTTGCCTACCTGCC	22	50	61.06	NP	+	1	0.2
NP_4_RIGHT	32	GGAGGCCCTCTGTTGATTAGTG	22	54.55	60.6	NP	-	1	0.2
NP_5_LEFT	33	ACAAATTGCTTCAAATGAGAACATGG	26	34.62	60.01	NP	+	2	0.4
NP_5_RIGHT	33	TCGTA CTCTCTGCATTGTCTCC	23	47.83	60.37	NP	-	2	0.4
NA_1_LEFT	34	AACCAAGTGATGCTGTGTGAAC	22	45.45	59.82	NA	+	1	0.4
NA_1_RIGHT	34	CTGGACCATGCTATGCACACTT	22	50	61.12	NA	-	1	0.4
NA_2_LEFT	35	ATTCGATTAGGCTTTCCGCTGG	22	50	61.25	NA	+	2	0.4
NA_2_RIGHT	35	GTCATTACTACTGTACAAGTTCCATTGA	28	35.71	59.98	NA	-	2	0.4
NA_3_LEFT	36	ACGGGGGATGATAAAATGCAAC	23	43.48	60.12	NA	+	1	0.2
NA_3_RIGHT	36	TTCTGGGTGTGTCTCCAACAAG	22	50	60.6	NA	-	1	0.2
NA_4_LEFT	37	AAAGGATCCAACCGGCCCAT	20	55	61.89	NA	+	2	0.4
NA_4_RIGHT	37	TCAGTTTCCTCTTTTCTTCCCCT	23	43.48	59.54	NA	-	2	0.4
MP_1_LEFT	38	GCAGGTAGATATTGAAAGATGAGCC	25	44	59.97	MP	+	2	0.2
MP_1_RIGHT	38	TATATGAGGCCCATGCAACTGG	22	50	60.41	MP	-	2	0.2
MP_2_LEFT	39	ACATGGACAAAGCAGTTAACTGT	24	37.5	59.68	MP	+	1	0.4
MP_2_RIGHT	39	ACTAGGATGAGTCCCAATGGCT	22	50	60.82	MP	-	1	0.4
MP_3_LEFT	40	AAATGGCTGGATCAAGTGAGCA	22	45.45	60.74	MP	+	2	0.4
MP_3_RIGHT	40	TGCTGACAAAATGACTGTCGTC	22	45.45	59.63	MP	-	2	0.4
NS_1_LEFT	41	GCAGGGTGACAAAGACATAATGG	23	47.83	60.12	NS	+	1	0.2
NS_1_RIGHT	41	TTTCTCCATGATTGCCTGGTCC	22	50	60.81	NS	-	1	0.2
NS_2_LEFT	42	AGAAACTGGTTCATGCTAATGCC	23	43.48	59.81	NS	+	2	0.4
NS_2_RIGHT	42	TCTTCAAACCTCTGACCTAGCTGT	24	41.67	59.98	NS	-	2	0.4

Supplementary Table 2.4. RSV A Primers. List of tiled-amplicon RSV A primers, including, name, amplicon, sequence, size, %GC, melting temperature, strand, primer pool, and pool concentration.

Name	Amp	Seq	size	%gc	Tm	Strand	Pool	[pool] (uM)
RSV_1_LEFT	1	AATGCGTACAACAACTTGCGT	22	40.91	60.4	+	1	0.5
RSV_1_RIGHT	1	GCCATTAGGTTGAGAGCAGTGT	22	50	60.8	-	1	0.5
RSV_2_LEFT	2	TGTTATTACAAGTAGTGATTTGCCCT	28	32.14	59.56	+	2	0.5
RSV_2_RIGHT	2	TCATCAGTCTTTGGTGTGGT	22	45.45	60.01	-	2	0.5
RSV_3_LEFT	3	CAGAAGATAAAAATGGGGCAAATAAATCA	29	31.03	60.12	+	1	0.5
RSV_3_RIGHT	3	TTATGGGTGTGTGCTTGGTAGG	22	50	60.47	-	1	0.5
RSV_4_LEFT	4	AACACAAAATATGGCACTTTCCCT	24	37.5	59.73	+	2	0.5
RSV_4_RIGHT	4	GCCACATAACTTATTAATGTGTTTCTGC	28	35.71	60.55	-	2	0.5
RSV_5_LEFT	5	CCAGCAAATATACCATCCAACGG	23	47.83	60	+	1	0.5
RSV_5_RIGHT	5	GTCATGCCTGTATTCTGGAGCC	22	54.55	61.25	-	1	0.5
RSV_6_LEFT	6	GCAAGCTTAACAACTGAAATTCAAATCA	28	32.14	60.8	+	2	0.5
RSV_6_RIGHT	6	CCCTGCACCATAGGCATTGATA	22	50	60.41	-	2	0.5
RSV_7_LEFT	7	GGTATAGCACAATCTTCTACCAGAGG	26	46.15	60.96	+	1	0.5
RSV_7_RIGHT	7	ACCATTTTCTTTGAGTTGTTTCAGCA	25	36	60.49	-	1	0.5
RSV_8_LEFT	8	TGGCCTAGGCATAATGGGAGAA	22	50	61.16	+	2	0.5
RSV_8_RIGHT	8	TGGTTGAATTTGATGTTATAGGGCTT	26	34.62	59.78	-	2	0.5
RSV_9_LEFT	9	CACATCACCCAAAGATCCCAAGA	23	47.83	60.82	+	1	0.5
RSV_9_RIGHT	9	ATACCATCCCGAGCAGATGTGG	22	54.55	62.12	-	1	0.5
RSV_10_LEFT	10	CAAACGATAATATAACAGCAAGATTAGATAGG	32	31.25	59.98	+	2	0.5
RSV_10_RIGHT	10	GGTGAGTTGGTTTGTGTTGGT	23	43.48	60.56	-	2	0.5
RSV_11_LEFT	11	AGTGACAATGATCTATCACTTGAAGATT	28	32.14	59.56	+	1	0.5
RSV_11_RIGHT	11	CCTTGGGTGTGGATATTTGTTTCAC	25	44	60.95	-	1	0.5
RSV_12_LEFT	12	GTGCCCATGTTCCAATCATCCA	22	50	61.4	+	2	0.5
RSV_12_RIGHT	12	TTTCTGACGCTGATAGATCTTAGGT	25	40	59.79	-	2	0.5
RSV_13_LEFT	13	TGCAGTCTAACATGCCTAAATCAA	25	36	59.67	+	1	0.5
RSV_13_RIGHT	13	TTGCAAATCGTGTAGCTGTGTG	22	45.45	60.21	-	1	0.5
RSV_14_LEFT	14	GCCACAAAGTCAATTCATAGTAGATCT	27	37.04	59.99	+	2	0.5
RSV_14_RIGHT	14	CCATTGGTTGATTTTATCTAGCGTGT	26	38.46	60.62	-	2	0.5
RSV_15_LEFT	15	GTTTCATCAGATCCAGTACTCAAATAAGT	28	35.71	59.82	+	1	0.5
RSV_15_RIGHT	15	GCATGGTGAGATGTTGATGTGG	22	50	60.08	-	1	0.5
RSV_16_LEFT	16	AGCATTCACCAATCTGATAGCTCA	24	41.67	60.47	+	2	0.5
RSV_16_RIGHT	16	CTTTGTGGTTTGCCGAGGCTAT	22	50	61.64	-	2	0.5
RSV_17_LEFT	17	AGGACCTGGGACACTCTCAATC	22	54.55	61.34	+	1	0.5
RSV_17_RIGHT	17	GTTGTTTTGTGGTGGGTTTGCT	22	45.45	60.8	-	1	0.5
RSV_18_LEFT	18	GAGTCAACCCACAAATCCACAA	22	50	60.93	+	2	0.5
RSV_18_RIGHT	18	GTGTTGGAGGTGAGCAGTGTAG	22	54.55	61.05	-	2	0.5
RSV_19_LEFT	19	ACCTCAAACCACAAAACCAAAGG	23	43.48	60.31	+	1	0.5
RSV_19_RIGHT	19	TGAGGATTGGCAACTCCATTGT	22	45.45	60.41	-	1	0.5
RSV_20_LEFT	20	TCCACTCAACCACCTCCGAA	20	55	60.78	+	2	0.5
RSV_20_RIGHT	20	TCTGTACCATTACACTTATTTTCCTTGAT	29	31.03	59.81	-	2	0.5
RSV_21_LEFT	21	GCAAAGGCTATCTTAGTGCTCTAAG	25	44	59.91	+	1	0.5
RSV_21_RIGHT	21	AGCCTTGTTTGTGGATAGTAGAGC	24	45.83	60.95	-	1	0.5
RSV_22_LEFT	22	GGCATTGCCGTATCCAAGGT	20	55	60.77	+	2	0.5
RSV_22_RIGHT	22	TGCTGTCTAACTATTTGAACATTGCT	26	34.62	60.01	-	2	0.5
RSV_23_LEFT	23	ACTACACCTGTAAGCACTTATATGTAAAC	29	34.48	60.12	+	1	0.5
RSV_23_RIGHT	23	TGCAGAGATTTACCTCACTTGGT	23	43.48	59.93	-	1	0.5
RSV_24_LEFT	24	TTTTCCACAAAGCTGAAACATGT	23	39.13	59.8	+	2	0.5
RSV_24_RIGHT	24	TCATCAGAGGGGAACACTAATGGA	24	45.83	61.15	-	2	0.5
RSV_25_LEFT	25	GGGGGTGGATACTGTGTCTGTA	22	54.55	61.08	+	1	0.5
RSV_25_RIGHT	25	GGTGCTATTTTATTTCAGTTACTAAATGCA	30	30	60.2	-	1	0.5

RSV_26_LEFT	26	TATACTGCAAGGCCAGAAGCAC	22	50	60.86	+	2	0.5
RSV_26_RIGHT	26	CATGGGGTGGCCATTCAAAATAA	23	43.48	60.06	-	2	0.5
RSV_27_LEFT	27	GTCACGAAGGAATCCTTGCAAA	22	45.45	59.63	+	1	0.5
RSV_27_RIGHT	27	TCCTGTTGCTTTCAATATATGATATGACA	29	31.03	59.97	-	1	0.5
RSV_28_LEFT	28	AGAGTATGCCCTCGGTGTAGTT	22	50	60.81	+	2	0.5
RSV_28_RIGHT	28	GGAATTTATACTACAAGGATATTTGTGAGGT	31	32.26	60.51	-	2	0.5
RSV_29_LEFT	29	TGGATATCCACAAGAGCATAACCA	24	41.67	59.91	+	1	0.5
RSV_29_RIGHT	29	GGGATCCATTTTGTCCCATAGCT	23	47.83	60.7	-	1	0.5
RSV_30_LEFT	30	AATGAAATCCATTGGACCTCTCAAG	25	40	59.73	+	2	0.5
RSV_30_RIGHT	30	CACCTTTATGATACTTAGATATTAAGGACTGT	32	31.25	60.02	-	2	0.5
RSV_31_LEFT	31	TGGTCCTTATCTCAAAAATGATTATACCA	29	31.03	59.66	+	1	0.5
RSV_31_RIGHT	31	TGTCTGCTTTAAGATGAGATTGATTATCC	29	34.48	60.42	-	1	0.5
RSV_32_LEFT	32	TGGACAAGATGAAGACAACCTCAGT	24	41.67	60.22	+	2	0.5
RSV_32_RIGHT	32	ACTAAGGCTAATATCTTTCCATGTCAAGA	29	34.48	60.99	-	2	0.5
RSV_33_LEFT	33	TGATAGATAATCATACTCTTAGTGATTCCA	31	32.26	60.37	+	1	0.5
RSV_33_RIGHT	33	GGCATCTGTGATGTTGTTGAGC	22	50	60.59	-	1	0.5
RSV_34_LEFT	34	GGTTCTACATAATAAAGAGGTAGAGGGA	29	37.93	60.58	+	2	0.5
RSV_34_RIGHT	34	AGCTCCTCTTAACATACTCAAACCTACT	27	37.04	60.1	-	2	0.5
RSV_35_LEFT	35	AAGACAAGCCATGGATGCTGTT	22	45.45	61	+	1	0.5
RSV_35_RIGHT	35	TCTATATAATTTTGTATGTGTGACGGCA	28	32.14	59.88	-	1	0.5
RSV_36_LEFT	36	TCATAAATGATAAGGCTATATCACCTCCT	29	34.48	60.27	+	2	0.5
RSV_36_RIGHT	36	TTTCTGTAGTTCTAGATCACCATATCTTG	29	34.48	59.51	-	2	0.5
RSV_37_LEFT	37	TGTTCAAGACAAGTTCAAATATTAGCAGAG	29	34.48	60.92	+	1	0.5
RSV_37_RIGHT	37	TCTATATAGTCCACTTTGCTCATCTACA	28	35.71	59.87	-	1	0.5
RSV_38_LEFT	38	TGTCACAATAATATGCACATATAGGCA	27	33.33	59.61	+	2	0.5
RSV_38_RIGHT	38	CCGTTATGTTGGATCGTTTTACTCA	25	40	60.03	-	2	0.5
RSV_39_LEFT	39	AGAGTATGCAGGAATAGGCCAC	22	50	59.81	+	1	0.5
RSV_39_RIGHT	39	GGCAAATTCATATACAATGTTAATGCTGT	29	31.03	60.42	-	1	0.5
RSV_40_LEFT	40	AGAGGTGAAAGTCTATTATGCAGTTTAA	28	32.14	59.5	+	2	0.5
RSV_40_RIGHT	40	GCATTGGGGTTTTTGTCAAACG	22	45.45	59.89	-	2	0.5
RSV_41_LEFT	41	CCATGATTTAAAGATAAACTTCAAGATCTGTC	33	30.3	60.98	+	1	0.5
RSV_41_RIGHT	41	GCAGAAGTCTTTTCCAGTATGTTAGT	26	38.46	59.79	-	1	0.5
RSV_42_LEFT	42	GTTTATGAAAGTTTACCCTTTTATAAAGCAGAG	33	30.3	61.11	+	2	0.5
RSV_42_RIGHT	42	GGTCCTCTCTCACCACGTGTTA	22	54.55	61.58	-	2	0.5
RSV_43_LEFT	43	GGACATAAAATATACAACAAGCACTATAGC	30	33.33	59.96	+	1	0.5
RSV_43_RIGHT	43	CAAAGTGATAATTTGTAGTTCTATAAGCTGGTA	33	30.3	61.17	-	1	0.5
RSV_44_LEFT	44	TGGGTTAACATATGAGAAGGCCAA	24	41.67	60.35	+	2	0.5
RSV_44_RIGHT	44	CATATTGAGTCAAACCTTATTTGTCTGGT	29	31.03	59.62	-	2	0.5
RSV_45_LEFT	45	CCTCCCATATTCACAGGTGATGTT	24	45.83	60.71	+	1	0.5
RSV_45_RIGHT	45	ATCACACTCCAGCTTTGCTCTG	22	50	61.06	-	1	0.5
RSV_46_LEFT	46	GCTGGACATTGGATTCTTATTATACAACCT	29	34.48	60.62	+	2	0.5
RSV_46_RIGHT	46	TGATAATCTATGTTAACAACCCAAGGG	27	37.04	59.72	-	2	0.5
RSV_47_LEFT	47	GGATGTCATAGCTTCAAACCTATGGTT	26	38.46	59.9	+	1	0.5
RSV_47_RIGHT	47	ACCTATACAATAGTCACTCAGTGTCTT	27	37.04	59.88	-	1	0.5
RSV_48_LEFT	48	CACAAGTTCAATGATGAGTTTTTATACTTCT	30	30	59.71	+	2	0.5
RSV_48_RIGHT	48	TGATGAGAAGTAGTAGTGTAAGTTGGT	28	35.71	60.56	-	2	0.5
RSV_49_LEFT	49	TGTATAATTTATTTCTACGGTTGTGATTG	30	30	59.62	+	1	0.5
RSV_49_RIGHT	49	ATGTTGATATGTCCATTGACAGCC	25	40	59.85	-	1	0.5
RSV_50_LEFT	50	TCATAGGTGAAGGAGCAGGGAA	22	50	60.75	+	2	0.5
RSV_50_RIGHT	50	GTCTAATTTGAAATCGATATCATCTTGAGC	30	33.33	60.2	-	2	0.5
RSV_51_LEFT	51	TGGAGCAAGCATGTAAGAAAATGC	24	41.67	60.94	+	1	0.5
RSV_51_RIGHT	51	CGTCCAGCTATAGAATATGATAGTATATCTCC	32	37.5	61.12	-	1	0.5
RSV_52_LEFT	52	TCCCAGTATTTAATGTAGTACAAAATGCT	29	31.03	60.07	+	2	0.5
RSV_52_RIGHT	52	TTTTAAGTTCATTGGTTGTCAAGCTG	26	34.62	59.68	-	2	0.5



Supplementary Figure 2.1. *Location of Public Health Regions in Southern Ontario, Canada.* The boundaries of the Waterloo, Hamilton, Peel, and York public health units outlined in orange.

Supplementary Table 2.5. *List of IAV H3N2 Genes.* Segment, segment name, segment length, gene name, gene abbreviation, and location of genes of in the reference genomes, NC_007366.1 - NC_007373.1.

Segment	Segment Name	Segment Length	Gene Name	Gene abbreviation	Location
1	PB2	2341	Polymerase PB2	PB2	28 - 2307
2	PB1	2341	Polymerase PB1	PB1	25 - 2298
			PB1-F2 protein	PB1-F2	119 -391
3	PA	2233	polymerase PA	PA	25 -2175
			PA-X protein	PA-X	25 - 784
4	HA	1762	Hemagglutinin	HA	30 - 1730
5	NP	1566	Nucleocapsid protein	NP	46 - 1542
6	NA	1467	Neuraminidase	NA	20 - 1429
7	MP	1027	Matrix protein 2	M2	26 - 1007
			Matrix protein 1	M1	26 - 784
8	NS	890	Non-structural protein 2	NS2	27 - 864
			Non-structural protein 1	NS1	27 - 719

Supplementary Table 2.6. *List of RSV A Genes.* Name, abbreviation, and location of genes of in the reference genome, NC_001803.1.

Gene Name	Gene abbreviation	Location
Non-structural protein 1	NS1	45 - 576
Non-structural protein 2	NS2	956 - 1097
Nucleoprotein	N	1125 - 2329
Phosphoprotein	P	2331 - 3220
Matrix protein	M	3224 - 4180
Small hydrophobic protein	SH	4190 - 4599
Attachment Glycoprotein	G	4644 - 5565
Fusion protein	F	5619 - 7521
Matrix (M2/22K)	M2	7567 - 8527
RNA-dependent RNA Polymerase	RdRp	8460 - 15037

Supplementary Table 3.1. List of Samples. Sample names, sites, type of collection system, serviced population size, and raw data accessions.

Sample	Region	Collection Site	Collection System	Population	Sample Type	Collection Date	Sample Preparation Date
A	Kingston	Cataraqui Bay	Wastewater Treatment Plant	52 084	24-h composite	June 28th 2023	July 17th 2023
B	Peel	GE Booth	Wastewater Treatment Plant	1 089 738	24-h composite	July 10th 2023	July 17th 2023
C	Galt	Galt	Wastewater Treatment Plant	89 714	24-h composite	July 10th 2023	July 17th 2023
D	York	Warden - 407	Pumping Station	650 303	Grab	July 10th 2023	July 17th 2023
E	York	Leslie Street	Pumping Station	295 232	Grab	July 10th 2023	July 17th 2023

Supplementary Table 3.2. Accession Numbers of 179 ToBRFV Strains Obtained from NCBI. ToBRFV genomes listed on NextStrain used to design primers and phylogenetic analysis. The strains used for tree building (PhyML Tree), assigned clades (Clade), and genomes used for read simulation (Sim. Reads) are listed. Unique strain genomes used in our PhyML tree are denoted yes (PhyML Tree), identical genomes are denoted no and the representative genome used in the PhyML Tree is listed (Rep. Genome). Strains were assigned to the same clade by both our PhyML tree and the NextStrain build, with the expectation of some clade 5 strains. Clade 5 strains predicted by NextStrain but not our analysis are labeled 5(N).

Accession	PhyML Tree	Rep. Genome	Clade	Sim. Reads
OM515230	yes	-	3	-
OM515231	yes	-	-	-
OM515232	yes	-	-	-
OM515233	yes	-	8	-
OM515234	yes	-	-	-
OM515235	yes	-	8	-
OM515236	yes	-	3	yes
OM515237	yes	-	-	-
OM515238	yes	-	3	-
OM515239	yes	-	1	yes
OM515240	yes	-	3	-
OM515241	yes	-	1	-
OM515242	yes	-	3	-
OM515243	no	MW314135	7	yes
OM515244	yes	-	3	-
OM515245	yes	-	7	-
OM515246	yes	-	7	-
OM515247	yes	-	7	-
OM515248	no	OM515255	7	-
OM515249	no	OM515252	7	-
OM515250	yes	-	-	-
OM515251	no	OM515252	7	-
OM515252	yes	-	7	-
OM515253	no	OM515255	7	-
OM515254	yes	-	7	-
OM515255	yes	-	7	-
OM515256	yes	-	-	-
OM515257	no	OM515266	-	-
OM515258	yes	-	8	yes
OM515259	yes	-	7	-
OM515260	no	OM515245	7	-
OM515261	yes	-	1	-
OM515262	yes	-	-	-
OM515263	yes	-	3	-

OM515264	yes	-	6	-
OM515265	yes	-	7	-
OM515266	yes	-	-	-
OM515267	yes	-	3	-
OM515268	no	OM515255	7	-
OM515269	no	OM515245	7	-
OM515270	yes	-	1	-
OM515271	yes	-	7	-
OM515272	yes	-	7	-
OM718702	yes	-	7	-
OM718703	yes	-	7	-
OM718704	yes	-	7	-
OM718705	no	OM515255	7	-
OM718706	yes	-	1	-
MZ945419	no	MZ945420	1	-
MZ945420	yes	-	1	-
MW349655	yes	-	4	-
MZ323110	yes	-	5	-
MZ438228	yes	-	5	-
MZ202349	yes	-	5 (N)	-
MZ004925	yes	-	5 (N)	-
MT018320	yes	-	5 (N)	-
MW314091	yes	-	5	-
MW314092	yes	-	2	yes
MW314093	no	MN882033	2	-
MW314094	yes	-	2	-
MW314095	no	MW314097	3	-
MW314096	no	MW314097	3	-
MW314097	yes	-	3	-
MW314098	no	MW314097	3	-
MW314099	no	MW314101	3	-
MW314100	no	MW314101	3	-
MW314101	yes	-	3	-
MW314102	no	MN882034	3	-
MW314103	yes	-	3	-
MW314104	no	MW314105	5	-
MW314105	yes	-	5	-
MW314106	yes	-	1	-
MW314107	no	MW314109	5	-
MW314108	yes	-	5	-
MW314109	yes	-	5	-
MW314110	yes	-	5	-
MW314111	yes	-	8	-
MW314112	yes	-	1	-
MW314113	yes	-	-	-
MW314114	yes	-	5	-
MW314115	no	MW314116	-	-
MW314116	yes	-	-	-
MW314117	yes	-	-	-
MW314118	yes	-	1	-
MW314119	no	MW314120	1	-
MW314120	yes	-	1	-
MW314121	yes	-	3	-
MW314122	yes	-	3	-
MW314123	yes	-	6	yes
KT383474	yes	-	5	-
MW314124	no	MW314125	1	-
MW314125	yes	-	1	-
MW314126	yes	-	3	-
MW314127	yes	-	3	-

MW314128	yes	-	1	-
MW314129	no	MW314136	3	-
MW314130	yes	-	1	-
MW314131	yes	-	1	-
MW314132	no	MW314137	1	-
MW314133	no	MW314135	7	-
MW314134	no	MW314135	7	-
MW314135	yes	-	7	-
MW314136	yes	-	3	-
MW314137	yes	-	1	-
MP872414	no	KT383474	5	-
MP872415	no	KX619418	5 (N)	-
MP875816	no	KT383474	5	-
MP875817	no	KX619418	5 (N)	-
MN549394	yes	-	4	yes
MN549395	yes	-	4	-
MN549396	yes	-	4	-
MT002973	yes	-	4	-
MT107885	yes	-	5 (N)	-
MT118666	no	KX619418	5 (N)	-
MN882011	yes	-	3	-
MN882012	yes	-	1	-
MN882013	yes	-	1	-
MN882014	yes	-	1	-
MN882015	yes	-	1	-
MN882016	yes	-	1	-
MN882017	yes	-	1	-
MN882018	yes	-	1	-
MN882019	no	MN882020	3	-
MN882020	yes	-	3	-
MN882021	yes	-	3	-
MN882022	no	MN882034	3	-
MN882023	no	MN882034	3	-
MN882024	no	MN882034	3	-
MN882025	no	MN882064	3	-
MN882026	no	MN882027	1	-
MN882027	yes	-	1	-
MN882028	no	MN882029	3	-
MN882029	yes	-	3	-
MN882030	no	MN882033	2	-
MN882031	no	MN882033	2	-
MN882032	no	MN882033	2	-
MN882033	yes	-	2	-
MN882034	yes	-	3	-
MN882035	no	MN882034	3	-
MN882036	yes	-	3	-
MN882037	no	MN882040	1	-
MN882038	no	MN882040	1	-
MN882039	no	MN882040	1	-
MN882040	yes	-	1	-
MN882041	yes	-	1	-
MN882042	no	MN882043	6	-
MN882043	yes	-	6	-
MN882044	no	MN882045	1	-
MN882045	yes	-	1	-
MN882046	no	MN882047	1	-
MN882047	yes	-	1	-
MN882048	yes	-	3	-
MN882049	no	MN882050	3	-
MN882050	yes	-	3	-

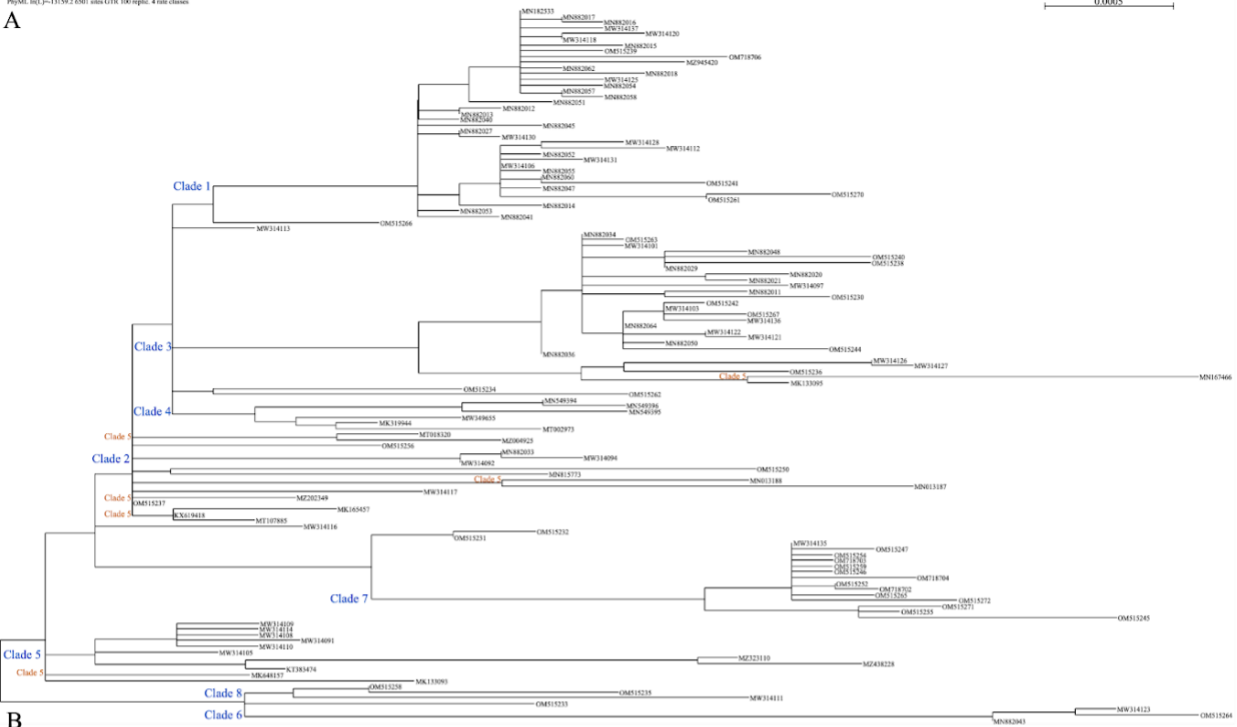
MN882051	yes	-	1	-
MN882052	yes	-	1	-
MN882053	yes	-	1	-
MN882054	yes	-	1	-
MN882055	yes	-	1	-
MN882056	no	MN882015	1	-
MN882057	yes	-	1	-
MN882058	yes	-	1	-
MN882059	no	MN882060	1	-
MN882060	yes	-	1	-
MN882061	no	MN882062	1	-
MN882062	yes	-	1	-
MN882063	no	MN882064	3	-
MN882064	yes	-	3	-
MN815773	yes	-	-	-
MN013187	yes	-	5 (N)	-
MN013188	yes	-	5 (N)	-
MN182533	yes	-	1	-
MK648157	yes	-	5 (N)	-
MN167466	yes	-	3 (5)	-
MK319944	yes	-	4	-
MK165457	yes	-	5 (N)	yes
MK133093	yes	-	5 (N)	-
MK133095	yes	-	3	-
KX619418	yes	-	5 (N)	-

Supplementary Table 3.3. *List of ToBRFV-Seq Primers.* Sequences and concentration of pooled primers.

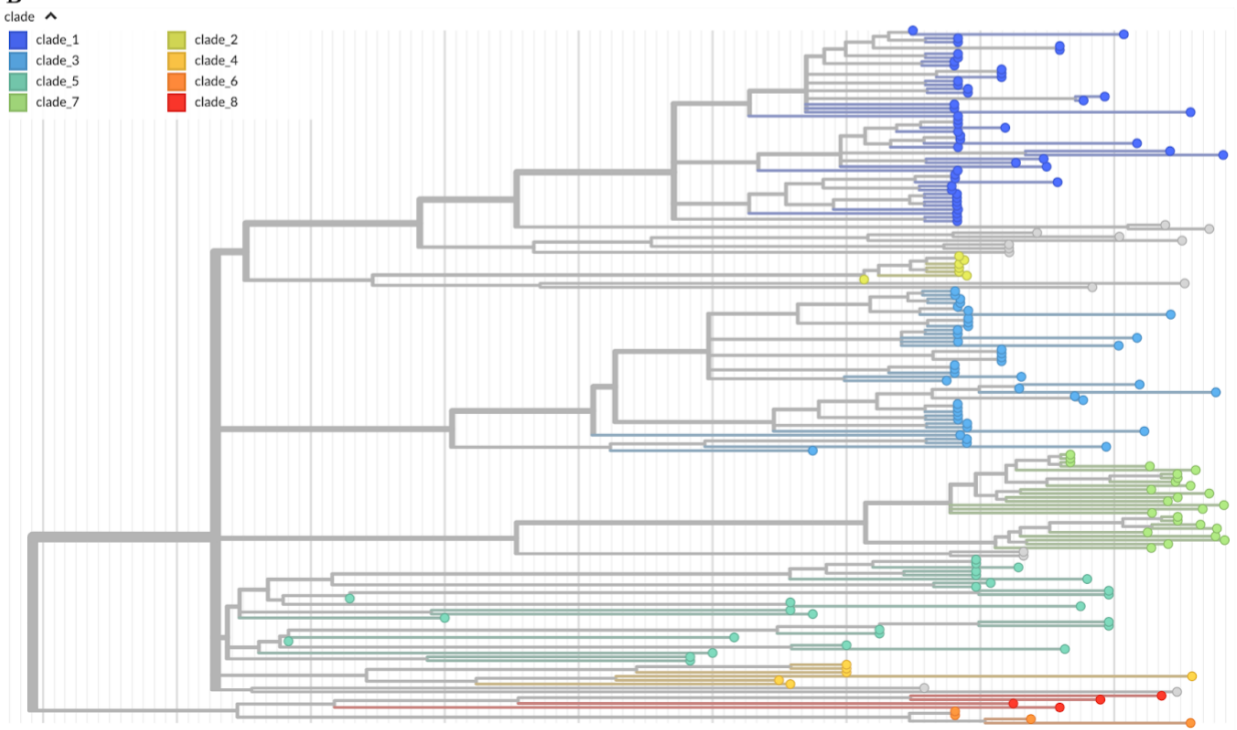
Name	Pool	Sequence	Size	%GC	[Pool]
ToBRFV_1_LEFT	1	TTTACAAC TACAATGGCATACACACA	26	34.62	0.5 μ M
ToBRFV_1_RIGHT	1	ATTGGGCATACAGCAGTGAACA	22	45.45	0.5 μ M
ToBRFV_2_LEFT	2	AATGATGCAGATCCCGTACGGA	22	50	0.5 μ M
ToBRFV_2_RIGHT	2	CGGCGTAACAAACATGGACATT	22	45.45	0.5 μ M
ToBRFV_3_LEFT	1	TGCTATTGCATTGCACAGTATATACG	26	38.46	0.5 μ M
ToBRFV_3_RIGHT	1	TGCGCTGTAAAATTGCTCACTAT	23	39.13	0.5 μ M
ToBRFV_4_LEFT	2	ACACCTGGTTTTGTAAAGTTTTCTAGG	26	38.46	0.5 μ M
ToBRFV_4_RIGHT	2	CTGGCAGTCACTCCGTTGATAA	22	50	0.5 μ M
ToBRFV_5_LEFT	1	GCACTTATCAAGCCAAGGCACT	22	50	0.5 μ M
ToBRFV_5_RIGHT	1	ATCAAGCACTGGCATATCCACC	22	50	0.5 μ M
ToBRFV_6_LEFT	2	CGCATTAGAAATCAGGGTGCCT	22	50	0.5 μ M
ToBRFV_6_RIGHT	2	CCTCTTGCCATTGAACCCCTCA	22	50	0.5 μ M
ToBRFV_7_LEFT	1	TGTCGCATTGGCACTTAAAGATT	23	39.13	0.5 μ M
ToBRFV_7_RIGHT	1	TGGTCGCAACATCTAAGACTCC	22	50	0.5 μ M
ToBRFV_8_LEFT	2	CGGTGTGCAACCTAGTCAAGAT	22	50	0.5 μ M
ToBRFV_8_RIGHT	2	CGGTACTAAGATTAGATCTTCCTCAAAAT	29	34.48	0.5 μ M
ToBRFV_9_LEFT	1	GCACTGCTAAAGTCGTCCTAGT	22	50	0.5 μ M
ToBRFV_9_RIGHT	1	GAGTTTCCACCTCATCAACCTCT	23	47.83	0.5 μ M
ToBRFV_10_LEFT	2	TGGAGACACACAACAAATTCATAC	25	40	0.5 μ M
ToBRFV_10_RIGHT	2	GCGATGATAGATACAGGTGTCGG	23	52.17	0.5 μ M

ToBRFV_11_LEFT	1	ACGGACGTCCATACAGTACATG	22	50	0.5 μ M
ToBRFV_11_RIGHT	1	CAGCATCATAGTTATTTAACATGGTGC	27	37.04	0.5 μ M
ToBRFV_12_LEFT	2	GATGCAGGGACCCAATAGCAAT	22	50	0.5 μ M
ToBRFV_12_RIGHT	2	TCTACTACCAAAGATGCAGTATTCTCA	27	37.04	0.5 μ M
ToBRFV_13_LEFT	1	TTGGAATTTGGTGGCGATGA	22	40.91	0.5 μ M
ToBRFV_13_RIGHT	1	TTTGCCTTGTGAGCTCACTGAA	22	45.45	0.5 μ M
ToBRFV_14_LEFT	2	TTCAAAGCGAATATCCGGCCTT	22	45.45	0.5 μ M
ToBRFV_14_RIGHT	2	TGAATGTTGTAACTCCCACT	22	45.45	0.5 μ M
ToBRFV_15_LEFT	1	ACAAGGCCACAGGAAACTACT	22	45.45	0.5 μ M
ToBRFV_15_RIGHT	1	TGAACTCTTCTAAGTGATCCCAATCC	26	42.31	0.5 μ M
ToBRFV_16_LEFT	2	ACGTGATACATCATGACAGAGGG	23	47.83	0.5 μ M
ToBRFV_16_RIGHT	2	TGAACCAATATCTTATCAACCTTGGAGA	28	35.71	0.5 μ M
ToBRFV_17_LEFT	1	TGAGTTCATAGACTTGTCAAATCAGAA	28	32.14	0.5 μ M
ToBRFV_17_RIGHT	1	CGCAGCAATTTTAACATTTCTAATATTAACGAG	33	30.3	0.5 μ M
ToBRFV_18_LEFT	2	GGTTTCAGTTCAAAGTCGTTCCA	23	43.48	0.5 μ M
ToBRFV_18_RIGHT	2	TCCTTCTCAACCCTAACTTTTAAACC	26	38.46	0.5 μ M
ToBRFV_19_LEFT	1	ACCGGGAAAAAGTTTAGTAGTAAAAGTG	28	35.71	0.5 μ M
ToBRFV_19_RIGHT	1	AGGATCTAGTACCGCATTGTACCT	24	45.83	0.5 μ M
ToBRFV_20_LEFT	2	GGCCGACCCTATAGAATTAATAAATTTATGT	31	32.26	0.5 μ M
ToBRFV_20_RIGHT	2	GGTGCAGAGGACCATTGTAAAC	22	50	0.5 μ M

A



B



Supplementary Figure 3.1. Comparison of ToBRFV Phylogenetic Analysis Resolving Clades. Our ToBRFV phylogenetic tree (A) compared to the NextStrain phylogenetic tree (B). (A) Strains belonging to each clade are downstream of blue labeled nodes. Brown clade 5 labels denote NextStrain assigned Clade 5 strains that did not group with clade 5 in our phylogenetic analysis.

Supplementary Table 3.4. *Twenty-six Virus Species in Shotgun Samples.* Viruses common to all five wastewater influent shotgun sequencing prepared samples.

Virus Species
<i>Acanthamoeba polyphaga moumouvirus</i>
<i>Cactus virus X</i>
<i>Choristoneura fumiferana granulovirus</i>
<i>Cucumber green mottle mosaic virus</i>
<i>Emesvirus japonicum</i>
<i>Emesvirus zinderi</i>
<i>Enterobacteria phage GA</i>
<i>Enterobacteria phage Hgal1</i>
<i>Escherichia phage MS2</i>
<i>Garlic common latent virus</i>
<i>Garlic virus A</i>
<i>Hagavirus psychrophilum</i>
<i>Hubei picorna-like virus 61</i>
<i>Moumouvirus</i>
<i>Norwalk virus</i>
<i>Pepino mosaic virus</i>
<i>Pepper mild mottle virus</i>
<i>Pitaya virus X</i>
<i>Rehmannia mosaic virus</i>
<i>Schlumbergera virus X</i>
<i>Squash mosaic virus</i>
<i>Tobacco mild green mosaic virus</i>
<i>Tobacco mosaic virus</i>
<i>Tomato brown rugose fruit virus</i>
<i>Tomato mosaic virus</i>
<i>Wuhan insect virus 23</i>

Supplementary Table 4.1. *CpV-BQ1 Genome Specific Primers.* Primer name, binding position, sequence, size, GC content (%GC), annealing temperature (T_m), and the amplicon size generated by each primer pair is listed.

Name	Binding Position	Sequence	Size (bp)	%GC	T _m	Amplicon Size
BQ1_1F	5495-5516	GCCTAAGGTGTTGTCCCATGTT	22	50	61	8821
BQ1_1R	14315-14294	AATACCGGATTTGCTCTCGACG	22	50	60.98	
BQ1_2F	46781-46802	ACGATCAGAATGGAGAACCCGA	22	50	61.39	8954
BQ1_2R	55734-55713	GAGGCACCGTCTTGAGTTCTTT	22	50	60.99	
BQ1_3F	79773-79794	TGTGAACTAACTCTCCGCGAGA	22	50	61.31	9023
BQ1_3R	88795-88775	AAGTCGGCTGAAGGCATTGTT	21	47.62	60.92	
BQ1_4F	104825-104846	ATCTGGGGTTGATATTCGCCG	22	50	60.99	8971
BQ1_4R	113795-113774	GAGTCGCAAGCTCAAGATGACT	22	50	60.85	
BQ1_5F	150772-150793	TACTTTGCTGTAGTCCCGTCCT	22	50	61	8971
BQ1_5R	159742-159721	TCGGAGCCATACTCTATTGCCT	22	50	60.94	
BQ1_1E	6167-6146	TACACTAAGGTCCCTCGTTGCT	22	50	61.6	n/a
	2453-2432 &					
BQ1_2E	162977-162998	GGTGTAAGAACTACTGCGCCAA	22	50	61.7	n/a
BQ1_3E	159174-159195	GTGCGCGCTTCCATTCTAAATC	22	50	61.6	n/a

PCR Region	% Id to CpV	Assembled PCR Sequence	Length (bp)
------------	-------------	------------------------	-------------

2 99.98 8930

TGAGAACCCGAGATAGTTTCAACTGAAGTTGAAGTACAAAGATACGATGTACCTACTAATTTTATAGAACCCTAGTAAATTTACAGATTGGCCAAATTATAACTTAGGGATAATATATTTG
AAGCTATGTACAAAACCGCCTTTAACACACTTACAAATTCGGTTAATCCTTTTGATATATACAACTGTTGAACAAATTTGATGCTACATTAGAAGCAAGAAATTTGCTGTAAATAGTAGCT
GGATTTAAAAAAAATAATACAGATTGGAACTATAGGAATTTGGAGATATACAAAAAATCCGGACGTCGATTACACTATAAAAACTATTACTCCTCAACAGATTTTTGGAAAAATTAACAGAA
GCTATAAGCAATAAAGACATATCGGTTTTATCTGAATACTATTTAATGGGTGTAAAGAGCTGAAATAGACAAAAACACACTCTCAGAAGCAAAAGCTATGATAGATTCTTAAAAAACCAT
CTATGAAAACTGAACAGGACGATGTAAGGTTTCAGAAAGAAAGAACAAACGTCGTAAAGCTATGTAATACAAAGAACCCGGGAAACGCTAGAGAGTAAAGAAAGAAATTTTCGTTTATAT
TTAAATTCATTTTCTAATGTATTTTAAATGGCAAGTTCTATCATAGTAGTTTTTCATAAAGATTGTACTCCTCTACAGATCTTGTTTTTCGGCGCTCTCAAAATACAAACCTAATATATCG
AATACATCGATGTTAAACCTGATGAATTCGAATCAGAACTGGTTATAGATGTAGTTCGGTTAAATAGTAGATAACAAAGATATATTTAAAGGTAAAGACCGCTTTTGATGATAAGAA
ATTGGGTAAACTCCTCCTAAAAAGGAAAAACAGCTCTTTAAAAACGATATGGTTAATACATTTATTGAAGATTCTGGTAGTAAAAATCAGAAGATAGATTAGATGCTGGACGAA
AGCGCGCTAATTAATTTACAATTTAAATATTTTATTTTAAATGGATCAGGCCCCAGTTTCCCGACTTTTGCTAAAAATATGAAAATTGCTCAGCAAGTCGCGAGTAAAAATAGATCCACCTC
CCGAACCTAAAAAGCGGAGAGATTAACCGAGGATGAAATGAATAACGCAATAAAGCTCTGTGCTAAATCAGTAACCGATGTTGTTACACCAGATATGTTTGAGATGGGAGAGAGTTCA
AATAAGAAAGAAAAAGGAAAGCAGTTGCCAATAATTAAAGAACCTTCAAAAAATTTCTTTAAAGAAAGAACACCAATTGAAGAAGGTTCAAGTTCAAAAAAACTAAAAAACCCGAGATT
GTAGAAATAGATTCCGATTGTTCCGAAGAATTGGATAACGCGCTTATGAGAACAAAGATATGGCTTTTACATTAAACCGTTAACTAGAAAGAACATATATTAATGGAACATAGAAAAAATTA
GCAATCAGGCGTCAAAAAATTTGATGACAGCAATAAGTGTTCGAAGAAAGAAAAAACTTGCCATAGTCGTTGAACCTGGTATGATGGATAATCAATCTATTAGATTCAACAGAAATGTC
TGACGAGAAAAATAGGTTACGAAACTGGAGATGTAGTTGTACTCTAGACGTAGAAGATCACCCCTTACTTTATGAGAGATGGAATAATTTACTAATGAAAAAGAGATTCTTTGGCTGA
AGCCTATAATCCAATCGTTTTATTGAGCATCTCAACCGTAAAACTTTAAAAATACAGGAGAGCGTTAAACGTTTTTAGTGATGAAGATACAATGTTAAAGAAAGTGCGCTGAGATGGG
AATGCCAAATTTAGGAGCACTAATAAATTTGGAGATCTGTTTTATAGATTAAATGTGTCACACAAAAACAAAAATACCCGAGAAATATAGAGAGTAAATAGTTTATCCACCGCTT
CTTACGATTTCCGGAAGTAAAGAAACTGATATCTTAGAGAAAAATTTGAAATAGTAAAGCAATCGGACCTAGAGTTCTTAGAGTCGGATTCTGATGAGTATGATTTCAGATGAAGATTAT
TCGGATTCCGATTCTGAAGAATTCGGAATAAAGTAAATTTAAAGAGTACTATAAATATAGTAAGATGATTGAAAAATGGATGGGAATTTCTAATTTATAGTCCATTGATGTAATATTATA
CAAACTGTCAAAATAGTATGAAAAAATTTATAATTTGTCTACGAGAAATTTATTAATGAATTTCTATTCCAGAACCATATACATTTTAAAAATGATGATTATTTACTTTTATATTT
CATGGGAGGAGAAATGGACAAATAGAAATTTAGAAAGAAAAATTTGATATAATTTTCTTTAAAGCTGATTTTAAAAACTAGCTATAAAAAATCAGTAATGAAATAGTAATAC
ACATATAATATTGGAGTAGAAATTTATATAAGTCTGGGAGCTATATAGTTTATAGTACATGAAAGAAAGGCTAATAAAATTAACACACGCTAATTTAATTTCTCTGTAATATTGAAAT
AGTCCAAAGTCAATAAGTTCGAATATAAGGAGGTAACTTTTATATATTAGCTATCTATAATTCTATGTAAGCAAGATTGATAAAAAAGTTGTACAGTTATATTTATATAAGAAATTTT
TCAAGATTTCAGACGACATTTATTATAAAAAAATTTATATTTTATTAATAATGTATCACTCAATTTTAATAGTATTCTCTTTTTTATTGAATTACAATATCAAGGTTATTAACTCAAGATA
GAACTAATGAAAAAGCAATTTGAAAAACCGAATATAACATCTCTCAAAAAATTTAAACTAGCAAGAGAACTATCAAAATTTAAAAAGGATCGTTATATTTGCTATTATTACTGAAGATCTG
TGTAATTTTGGTTTTGTACCAATAAGGAATTCGTATTTTTCCATTTAGCGAAATGAGTTTTTTCCGGTGTATAATCTATGATAAGCAATTTATCAATTTTGTGCTACATTTTCCAGG
CATTTGCTGGGTAAACGCTGTTATGTTTCCCTTTGGAATATTACAGGAAATAAGAGCAACACCGCTCTAATTTAATTGACACGCTGTGTATCTTTCCATATCTCTGGGTATTTCTTTAC
ACAAACAAATGAATAATCGTAGAGCCTGAATAATACCGCTAGTAGTTCTTGCTCCATCAGTACATCGGATTATTTTTAAAGCTATTTTATAAATTTTATAGTTGCCACAA
ACATGATGAGCTGACATAACATCTGAGCGTATTCTAATATCAATTTTGACAACATGAGATCAGCTGCAATCTCTGCGCACACATATGGACATATCGATAAAAAAGATATTCATAGTA
TCGATGAATTTGATATATGTAATTCAGTCTTATATATCCATTAAAAATAGCAATTAATAGAGGCATAAATTTAAATTTATTTCTGAGCAGCCCTAACTCTCTATAACGAATCGTTGTAGTTG
TAATTCCCAATCATTTTCAATAACTTCAAATTCAAATCCTATACCTCTTCGCGAAACATATCCATCTCTATTCTAATTGTTCCAGTATCTCTATCGTAAAAATTTAATCTTATAAATGT
GTCCAAATCTAGAAATAAGCAAAATCTGCCACGCTAGCTGTATAGCCATTATTTAATAATTTTAAATTTATTTCTTGAGCAGCCCTGACGCTCTCATACCAAGCAATGTTGT
GTACAGCTCATTTCCAGTAACTTCAAATTCAAACCCCTATGCGCCCCCTTCCCAACCCATCTCTCAGTCTTAGTCGTTCCAAATTTCTTCATCGTAAAAATTTAATCTTATAAAGGCA
CCAGCTCTGAATAATAGTAAAAATCTTGCCAGGTACGTTTATAGCCATTATTAATAATGTATAATTTTAAATGCGATTATATCTGACGGTGGGGTTCGAACCGACGAGG
TTTCCGCGCAGCAGATCTTAAGTCTGCCCTTACAGCCACTCGGGCAACCGTACAGGTATATAATATACCCTCTCTATTTTAAATATTTATTTCTTTATATCTTTTTATAAAAAAGTATAAGT
TTATAGTTATTTTTTAAAAAGTATTAAATGTATAAATTTTAACTCCAGAAATTTTCATTAGTTTGTATATACAAATTTATTTGGAACTTTTAATGCAACAAATCTAGTAAAAATGATAAAT
ATTATCGATTATTTCTAATACATTAAACCATCGAGATTAAATTTAGCTTATCTCTCATCTTTATAGTTTGTATCGGCCCGCGGATCAATAAATACCAAGTCCGTGTAATTTTGTACAT
TTATATCATATTATGAACAGTTATAGTTTATACTTTTGTTGTATTACTTTTTAAAAATTTAAAAATTTTCGGGATCTTTTCAATTTGCTATTACAACCTTAAAACTTTTCAAAAAAGTGTGA
TTTTCTCCTATACAGCGAGTACAGTCACTGATACACTGTCAAAATTTATATATTTTTTATAATATCGTTAATTTGTAATGCTCGAGAAGGTGTTGTATATATATTAGCAATTTGTGAGG
AAGTGTCTAATTTTAAATAGTTAGTATTATATTTTTTAAATGGATTATTAATCTAATCTGCGGCGAGTTTCGCGCATCGCCCTCTCTAGTAAGTTTTTAATACATCAAACTGATCAG
TTTTCTGATGCCCTTTTATCTTTTTTAAATGGTCTAGGTTCCCGGAGTTTATTCATGCTCAATCGTAAATCGTAAATAGCTTTTTGAGCAGGTACGATAAGCGCAAGCTCTGCTGCTCCT
TTTTTTCAGAAAGCATTGTTAGTTAGTACATTTCCATGTATCTATTGACGTGTTATTTACGCGCAGGAGGTGGTTTTCGAACTATTTCAGGAAGTGTCCCAAGGCTCTAGTCTATCTA
TATTTTAAATTTTCGTTTAAATTAATTTGTCGTAATTTGATGAATCAGTTCOCATAGTTGGTGACATAAATCTGCAAAATAGTGCTCTATTGACACGCTGCTGCTCTCTGTGGAAATTT
CGTTGAAATTTTATATTGAGCCACGCGCAGTTGCGTATTCAGGAGGTAAATAAATTTGTAGCCGCTAAGTCCACGGAACGCAATTTCTAATCTGGGCCACTGTTCCGCTGGATAGTAGC
TCCAATTCGGGATCTTCCAGATCCCAACGATATACATTTCTCCAGTGGATGTATTAGTCATATAAACAGGTTGATCATACTGTCTAGACTTTAAATTTGAAATTAATTTATCTGATTTCTCC
TCGACAGGTTGTTCCAAAGTCTAACTTTGTAATAAAGTATAAATTTGGGTCATTACTATATAAATATAATATCTAATAATATAAATAATGAAAGAAACCAATGATGACAGTAAATGTA
ATTCCGATATTACAAATTAATAATTAATACATTAGAAAGTCTAATAGAAATGATAAACGCACTATTTCAGGAAATCGTTACCATCTAAAAAACACCGGAAAGATTTCCATCTAAGATGTACCT
TTTTGCCAGATATTATAGAAATTTAATTGAACATAATATATGATAGGAATGAATAAATTTGAAGACTCAGATAATAGATCAAGTATTATTTTTATACACGATCTCAAGAACTATAATG
TTACATCTGTATTAGAAGGTCCTCCTGGTACAGGTAAACATCTGTGTCTGCAATTTTAGCAAAAAATTTATTCCAAGCTTGAATATTTAAAAAGTAAATTTAAGTAGTAAAGCGTT
CCGATTTAATTTTCAGAATATCTAGGAGGAACGACTATTAACACACTCGAAACCTTTGAATAGGTGTAAAAACCGCGTAATTTTAAATAGATGAAGCTTATTCACTGGGATCTAATGGTTGCTCAG
AAGACATCTACGCCAAAGAAATGTGTAGATACTTTAAATCAATATTATCAGAAAAATGTAGATAAATTTATTGTATTATAGCCGTTATAAAAAAGAGTTGGATTGCGTGTTTTTTTTCATTA
AATCCGAGGCTTAGAAGCAGATTTCCTTGGACTTTTCAAACTCGAAAAATTAATTTCCGGAAGAATTAAGCGATATTTTTTATAAATAAATAAAGAAAGGAATGGGATACAAACATGTGAA
AAAAAGACTTTAACACATTTTAATTAATAAAAAATTTAAATTTATTCGATGGAACCGGTGGAGATATAAATACAATCTGAAAAAGCAATGATAAATAGATTAGAATAAATTTGGCACATG
TAAAAATATACCATAGAAATGAAGATTTTACAGAAGCTTTAGACATTTTTACTACAACCTAAACTAAGCACATATCTTGCCCTCCACCTGGTATGTACTTAATTTTACATATTTACG
CGGTACTTTTCATTAAAAATTTAATAATATTATATTGTAATATGTTTTTCATACAAAAAATAGATTATTAATAAATTTGACCGTAATAAATAAACAACCTTTTATTTAAATGGAGATCTCTAAATCA
AGTCTCCAATTAATAATACGAGTATAACGATGATAAACTTATTCTCAAAATTAACAGTGACGAGATAGTCACAACTATTTTTTAAATTTGTGTCTTATATAGAAAGATTATTTAAAAAT
CGGATAGTAAAAATTTTAAATAAATTTTCAAGTACGAATTAGTATTGACGAATGTAGCAAGTTTACGATTATAACAAGATGAATAGGTAAAAACAGTATAAAAATGGCGGAAAAATTTAT
ATGTTTCATTTAGTTGTAATAAATGGTAGATGTACACTAATAGAATTTACTCAGATTAATAAAAAATTAATTTTTAATTTTCGTTATATTGTAACAAATTAATAAATTTATTTTGAATAAATGACAGA
AATCAAAAAATGAATCCGTAATGAAATCAAAAAACGAATTTAATTTTATGATCCAAATTTAAACTCCTAAAAATATATATTATTCAGAATTTATTCTAATTTCCGAAGAAATTAGTTTTCAAGT
AGCAAAAAATAAAAATCATTTTGTATAAAGTTTAAAGTAAAGCAACACTAATCTAGATGAAAAAGTTTAACTTGTATTAAAAAATTTCCGGATGCGGTTATAGAAATTAACGCTGAAAAA
GTAAAAACATGGTTTGGAAAAAGAAATTAACATAGAGGAATGTAAAGGAATATTCAAAATTAATTTAATAGACAAATAATTTTGTGTTTTTACGATGAAAAATACAGTTTTTTACGATCTTAA
AATACAGAATAGATAAATGCTGTTTACAGATGAATTTATGGAATTCGTTGAATTTAAGTGCAGGATAGTAGTTTATACAAAAAATTTATTTTATTAATGGGAAATAGGTGAGTATAA
AATTAATAAATCAAAATTTAAGTAAATTTAAGTGAATTTTATTCGGAATATAAAAAATTAAGAACTTACCGGAACATCAAAAGATGATGAATAGTTAAAAAACTTGAAGAAATTA
CTCTATTTTAAATTTTAAAGTAAATTAATTTTAACTTTGAGAATGGAACAATTAATAGAATGTTTAAAAAGGTAAGTCTACACTTGGTAACTATTACAAGAAAAATGTTTACCAAGCA
GCTCTGAGCGTTGAACCTAATTTGGCAGGAATTTTATTTCAAAATGAAGTGGTCTTCGATAAAAAATAAAGGTTGTGTGCGTAGGATACGAAAGAGCGGATAGTATTATTTATCAAAA
GAATGTGATCCGGAGTTTATTAATTTGAATTTAAATCTCAAAATACTAAGTTGTCTCTCAAGAAATTAATCAATTTAGAAAGTATATCACTAATCTCAATTTGATAAAATGGAGCTCTTGTA
TTTTTACGAAACGCTAGAAATTTAAGTGTACAGAAAAATCTCATACAAAAATAGATTAACTAAAAATTTGAAGTGATTTTCGTGACATTTCCGCGATATAAAATTAGCGTCAAAATACCA
AAGGTTTATTAAAGCGGAATTTTACTACTTCAAAATTTTTCTTTTATCTTTTGGAACTTTTTCGATTTTTCTAAAAATTTTATATATAGTTTATACATCTCTGAGAGATAGAAATTTAC
AAGTTTAAACCAATGTCTACTGTATTTTTTCTCTGATATTGAAAGTTTTAGTAAATCGTAGTGTGCTGCCCTTGATTTTCTAGTCTATAAATCATGTGTTCTTTTAAATATTTTCTAG
TATGCTCATACAGAACTTTTTTAAATAAACGTTTGTTTTAAATACATAAGAACCTTTCTAGTGTAAAACTATATGAACCAATGTACCTTTTTTTCGCTGAAAAACATGATAGCCATATCTTC
CCCAGAGGCCAACCGTATGCTGTTCTTCTGCTTTTAAACACTGCCATGGATGAGTATGATAAATTAATTTCCATTAGGAGTTCTTACACTCGATTTTTTCCCCCTTACTCTATAATCT
TTCAAAAAAGTTTATACAGAGGTGCTCCGTTTACAAATTTATATCTTAAACAAAAAGAGATCCGGGCTTTTCATGTTTTGCTTTTCCATAACGATGCTTCAATTTACAGACATAAATGTTTATC
TATCATCCACTTGACCTTTTTTCCAGATTAGAACCGGATTTACTTTTAAATTTAGAGCTTACCATTATTATATAAGTTATTTTTTAAAGTTTAAATTAATTAATTTTAAACATGTG
GGAACTTTGAAGTTAATGTTATGTCGGGCAAGGATGAATATCTAGACAATTTACTAAATCAAAATCAACAGAGTATATGATATCACTTAAAGTGTATATCATGATTCTCAGGAG
CAAAACATAGAGCTTAAATTTTCATATTCAAACTTTCAAAAGAACTC

CGGAGGTTAATCGTAAGATCTAAATTTAAGATTTGTTTTATACACATTAAATGATTCATCTTTGTAAAGTGTTTAACTCCTGAAATATCTACTTCTCTTTCAAAATCTTCTCCCTT
AAGTCTCAAAATTTCTCCGGTCTGCTTATTTCTTACCTGATAATCGTGTGTTTTCTTGTAGATAATTTTTAAATACGGTAACGTGTTTCAAAATACCATCTATTAATACCTTTAGCAGTTGCTGC
TTTAGTGTTACTTCCAAAAATTTAATTCAGAAATATGGAATCTTCTGTTCTGTAATTTGAACCTCAAAAAATATCGGATTTAGTTTTATCATCGGCACCGTTTGGAAATCCCTTTAATTAAT
ATTCCTTCATTCTTTCAATAGACTTTTACTTTTAGCTTCGACGATGGATACCTAATACCATACTCGTCGGCAGAACCTAGTGGATCTAAAAATGATAATTTGGATCTGGACATTTACC
TTTCCATGAATACGGGTACAGGATGATCTCTCTGATGAAGTTGTTCCGCTTTTAAATTTCTAGTCTCTTATTTGGTCTCAATGTGCCGGTGGCGATCCAGATTAAGGTTATTTCTCTCCGG
TTTTACTATTTCGCTGTGCAAGGTTTTTTAAGAAAGAAATTTATCGTTTITAGTAAAAAAATCGATAAAAAATCTCTTTAGTAAATTTGAAATAATTAACCTGACAAAGGTGTAAAGATACCC
ACATAAAGATTTCTTTAAATTTTATCTCCACATCTAGACATAGTTAACATAATTTACTCCATATTTGTTTTATTTATCCAGTAATTTTAAATTCGGGGTCCGGAATTTGATACAAATTTGATATA
TCTTTTATCATAGTTTGAATGCGTGACATTTTACTGGAAGAAAAAGACCAATCAATTTTAAATTCCTTTAAAGTTTATGATAGTAATCCATTTATAGTTACTTCAAGCATTTGATGTTTC
CGTGTTTAAACAAATTTCCGATTTGGATCAGTTGGAGCTATTAGATTTTAAATTTTCAAAATTTTATTTCTAAACTCTAGATACACTGAATTTGCCGAGTAAATAGATGTGTATACGATTTAT
AAGGTAAGTATTCAAAATCTGGTAATCCAGTAATTTCTTCAAAATTTTCAGGTGTCTACAGCACCAGATGTTTGATTTTGCAATTAAGAAATCAATTAATAAATTTTAAATTTTCAGATCTT
TAGGAATATTAACTAAATTAATTTAGTCCGTTTTTACTCACTCTTATAGAAGTTTATCAGGGCCGCTCTTATAAAAAATTTATGGCAGTATTTAAAAATTTAGTGTGGCAGTTTTTATGACC
AATTTTTTTGGCCCTCTTTTTAAATTTATCCAGAAATTTGAATTTTTGATAATAAATAGACTGTCTTTCTAAACTAAATTTGTTGTTCTGATAGTTTACTTTCATCTATAATAACAATAGTT
CATCGTTTTAATTCCTCTTGAGTTATTGTTCCATCTATTATTTTTGTTTTAATTCAAAATATCTCCATCATATGAATTTTAAATTAACAATTTCAATTAATTCCTCCAAGTGTAAATATAA
ACTTTTATCTCTCGGATAAGGACACTCATTGTATGAATTTGAGGACCTTTTTATAACAATAACTACAATAAAGTCTTTTTAATTTGGTCCAACCTGGTGGTCTATTATAATTTGAATTA
TGGTTCGACGACCCATGTTCTTTTCTCAGCAAAATTTATCTGTAAAAACCTCTGGAGAATCTTCGGAATCTAATGGAAATCTAGTCTTAAGAAATTTATCATGTCGATAATCTAAATA
CGAATCTAGCTTTGTTGACAAAAATTTAGTGCTATTAAATCTGAGTTCGGAATCCATTTAATTTTATAAATAAATGATTTAAAAAAAATCTATAAATTTAGTAAATGAAAAAAGAAAT
AGAAATATTTAACAACCTCTCGAAGAGTTTTAGAAAAATTTATATCTAAATTTGATACCAACAGAGATAAAACATACAGAAAAATCTGTTATAATGATGAAGAAATGTTCTCCGGAACAA
CCTGTTTAATATTTTGGCAGGTTGTATCAATTTAATAAACGAAATTAAGAGTGAACACGAATCTTTTTTTTGAAGATAGAATAATTAATGAACAAATTTTCCGATTTTATAC
AGAAGACTGTGGATTTGGACGCTTATTTGGAATGATTTGACTCCTATAACTAAAAAGCTATATGGGATACGTTCAATCTCTTTTGTATTGGGAGAAATAGTTGTTAATAATACACAGC
AGTATTGAAAAATATAACGCAATTATATGCAACTGATTATAAAATGAATTTAAAAATTTAAAAACGGAAATTTCAATAAGTTTTTAAATAAATAAATCTTAAATATATAGTAATGACCT
CATATTTGGTAAAGTATATAGCGTATTATTTCTTTTGAAGTTAATCCAATCAGAGGTCGACGCAAAAAATTTAGATATAACGCGATTGACTAGATTAAATATTTTAGTAACCTGCGTA
TCAACACTTTTTTTAAACATTTAAGAAATGATAAGAAATACTACTTGTGCGAATATCTTCACATTACGTTGCGTGGAATATATTTTTATCTTCAACAAAGATTTGGCTGAAGTAAAAAATTT
AAAGTGTAGAAAAAGCATAGAAAAACCGAGTTTGAATTCGCGATGATTAGAAAAATAAAGTTATAGATCAGACAGCGAATAAAAAAGCATGACCAATATAACTTATAAAGGGGA
ATTGATACAGATAATCTTTCTAAATATTTTTATAGAAAAATAGTACTTTCTAAAAATAAGAGGAAGTTTAAATGATGAAAAATATAATATAGCGCTATTATAATATCGTTAGAGGAAA
AGTATCTAACGATATACCTTAATCAGATATAACAATGAATTCAGAGTTAGCGATTATATTAATTTTTAAAAATAAATTTAATATAGTATAAATGGAATTTAATATAGTAAATGAGAA
ATAGATAACAATCTCTTAATAATTTAACTTACTTAGATTAAATCGGTTTGAACAGTTTAAAAAGCCTCCAAATATTGATTATACGATGGCGAGTGTCAATTTAATAATACATAAGCGCCA
TCATTTACAGCTCTTAAATTTGTAACCGATGGTAAAAATATGCGTATTACGGAATACCGGTTACGTTAGTATAGAAAGTTCACAAGTTAAAAATAGTTTTTTCTAAAGAAATTTTAAATATG
ATAGTGTAATAATTAAGTTTAAATAGAACTATTATCAATTTTTACCTTTTAAACGATAATGATTAGTTGGTTACTTTACTGAAACTCGGGATACTCAAGATATCTCAAGATACCAAGAAAA
AGACCAATGAAGAAATTTAGAGAGTTATTTTCAAAATTTTTTGGAGTCTCGGAAAAATTTTGAATCTACTATATATAGTTAGATCTACTCTCAATATTATACGTTTATAGAAAGCAAAAA
AAGTATTATCTGTTGAACAGATTGACAATTTAGAAAAATAGAACTATATACGTCCTGTCCTCAATTTGATGGCAGCAAGTCCGGAATAATTAATAATTTAAATTTGCCCAATATTTTTGTAA
GATTGTTAAAAATCTGGTTGTCAAGCACTTAACACATGTCCTAATTTGTAGAGTAACCAATATAGTGATTAAATCTCTCGATTAAATAAATTTTAAATTAATAATTTTATTTAGTCAAT
ATGAGATCGCTAATTTTATTATTAATAATTTTCAATAGCGCTATTAGGATTTTAAATATAGTGAATAAGACCCGAACCAAAAAAAGTGAATACGCTACTTTAGACTCAACCAACACTTTAG
ACTCAACCAACCTCTCGGATGATGCAATCTGTCGACGACCAAAAAAGTTAAATGTTTAAATGAAATGGCGACGCTTTAAATGAACAAAGTGTCAATATAGAGCTTTTGAAGCTTTGCTAA
GTTGAGATAACTCTGTTAACGTAGGCTTTCCGGCAAGCTTTTTCGAAACCATCAATAAAAAGTAACGGATGAAAAATGGGTTAACTGAAAAATTTAATTTCTTATAAAAACAATTTTCCAT
ATGAGTCAAAATATTAAAAAAGTAGATTCTAACACTAGCAAAAAATTTCCAATTAACAACTCAAGTCTCAATCTTTTGTAAAGTAGTGAAATTTGCCAGAGGGGTAAACAGACTATAT
AGAAGATTTTGATTATTTAAAAATTTCTAAAAATTAGTAAGGAACGCTCTCAGTAAAGACAACTAATTAATCTCCCTCCCATGTAGAATAACAAATCAACAACTAAAGAAATCTAAACACAT
TACCAAGTGAATAGTTTTCAAGACTAAACCTGACGATTTTGAATTTACAGAGCCAAAAATTTAAGAGCTGTAAATGACAAAAACGATATTTTATACGATCTTTAAGAACTTTGAAGA
CGCTAAAAATATATTTAAATTTAATTTAATTTTATATATATAATATATAGTCTTTAGTGGTAGATAAACCGTCCAAAGAAATCAAAATTTAAAAATTTATGAAGCCGATATCT
AAAAATATGAACCAATCAGCTAATAAGATATAGTTAAATTTGTCCTCAAAATTTAGAACCTCCGCGTAAAAATTTAAATATATGACAGCGAACTTAACTAGAAAAATCACAACTCCAAAT
TAGCGATACAAAGACGATATTAATTAAGTTTAAACCAATGAGCAGCAATTTCCGAGATTTTATACATCGAATGTACCCAGATTTTACGATACAAATGACCTCCGCACTGTGGAAGTA
GTTGAAAAAATATTAATAATAGCCCAATATATATAAAGAAATTTGATCCAAAGTTTGTAGAACTTTTGAAGAGTTTGAAGAGGTTAATAAAAAATACGACTTAAATTTTTTACAGACTAGATCTT
ATTAAAAAATAGAGCATCCCAATCCCATGGTTGTACAAGATAGAATTTGATAAGAGATTAATAACTCTAAAAATTCATTAATATAGTTTTAGAGTTTTGTCGAAACATGCTCAATAA
GCTTAATAAAGAAATTTTTGTCGTTTTATGAAATTTAATAGTATCTATAGAAACCCATTTAATGCTGTCTATTCTAGAAATTTTGGATCGGTTTTAATTTTGTATAAAACTTTTATAGTAG
AAATGAATCTATCTCTATAAATCGGACTGAAAGAAATTTTACATAAAATCATATAAATGGGTGACCACTCGGTGTTTTAGAAGTAATACATTTAATTTCTAGATTTAAGAGTATTTTAT
ATAGTTTCTGAATCTGTAATAGCTCCAATCTTTCTCCACGCTCCCTAGATGCTGTTACCTCAGAGTCAAAATTTATCATTTTATTTTCGCACCTACCTCCGCAAAATTTAGCCATTTGTTAT
CGTTATCTTTCCCAATAAGAAAAAGGAGTATTATCTATTGCTCTGCTATAAAGAGAGATACCCGGAAGCGGTATAAAATATTATTTATCTATTATATATTCTATTAAATGATACT
TTTAAATATACTTTTTACATTTAATAATTTATGCGTTTATAAATAAAAAATTTTTTAAACTATTAGATTAAATGAGCGGTTTTGTACCAAAAAATTTAAGGTTGAAGCAGAAAAAGGATTATACA
CCTATCAGTGTAAAAAATCCTATTAAATTAATGGAATTTCTATAGAACACGAGAGTTCGGAAGATAGCGACAGCAGATCGAACCCAGATGAACCTAGTTGTTAAAACTGTGTCGATAGA
AAAACTGTTTTAAATCTGAAAAAAGTTCGCCCCAACAAAGAAAGCAAAATTTTACCCAGAAAGATTACAAAATTTTGTCAATAGTTCTAAAAACAGCTGTAAAAAATACGATTTCGGAA
GAATACGATAGCGATTTCTGAATCTGGTAATTCGGAATAACAGTGAATCTGCTGATTACTCGGATAGTATGTCTGAAAAATCTGAAGACGATAAAAATCGAATGCCAATAAAAAAGAGAA
AAACAAACCCCTCTTTTAAACATCTGATGCTTTAGAAAAAAGGCGTCGAACTAACTAAAAATTTTCAATGAATTCAAAACTATCCGATTTAAAAATTTGAATATGAACCTTCATAAAAATAG
CGCAGAGATTGAAATGAGTATTAATTTACAGCAAAAGATATTAGTTGCGGCAATAAATCGGATTAGAGTTTGAATAAAAAATTTGATCCAATCGCGCGCAAAATAGATGGTTGGTCAGA
ATCTAATAGGACAAATTTAGACGATTACGAATCAGTATTACTAAATACATGAGAAATATAAACACAGGGCAGATTTACCACCAGAAATTAACAATATTGGTTACGTTGGCTGTAGCGC
ATTCGTTTTCATATGACTAAAACTATGTTTGTGTTCAATGTCACCAACAGGAAAAATCTCAATACAGCGACATTATCAAAAAACATGATGCAACAAAAACACAAAGTACTCCAAAAAT
ATGTTTGGAGATATATCTGGACCTCTCCTCAATTTTTACATATTCTTAAAGATGATGACACAATTTCAACTGGTTCCGGTAGAACTCTTAAAGAGGTGACAGTTAATCAGAAAGGAAAG
CGAGCAATTAATCTATAATTTAAATTTAAATTTTTTTATATTATGATTTAATAAATATTCATGGTTTTATATTATAAAGGTTCCATTAACGATGAACCGCTAATTTAAGACGCCCCAC
GAATAGCTCATTTTCTGGTGATAGAATTTTACAAATACAAGAAGTTTATAGTCTCAGCTCGCTGTATATAATAGAAACAACGAGGAAAAAATAGCAGATAGAATTAATATTTAGAAAAAT
CAATTAATGAATATAGAAAAACCTACGAAAGAATTTAGATTTTTTAAACAAAGATTACAATTTGGAAGAAGAAGCTAATATTATTATTATCTCAGATTAAAGGATTATCTATTTACATTTTAT
CTATTATTATAGCAATTTGAATTTAAAAATTTACGGTATTAAATTTAAATGTATAACTTCAACAACCTAGGTATTATAAGTGTATGTATCATTAATAACAGGAATTTCTAAATTTCTCTATTGT
AAGAGTTCCTCCGTTGTTTTAACTTTAAAAAATCTGGAGCTGGAGCAAAATTTAAATTTCCGTCAAAAAGTTTTTATAAACTGACATAATAATAAATCTTATTTTCAAGTAGTTTAT
TTGACAAACAATAAGATTTTGGCAATTAGGAGAACAGAAATTACCAAAAGTTTTATATCTATCTAGGCTTCGCAATATTTAACGGAATGTAAAAAGGTAAGTTAATCAATGGATGATG
ACAATAAATACTAGTTCTATTTTTTTTGAACATAGACTTTTATGAATGTACTATATTTTACGGCTGTCTATTTTAAATTTTGTCTCATCTTCTCATCTATCTATATAAACA
ATTGCTTTTTTGTCTTTTATGAAAAATCAGTGATATTAAAAATAGTCTTTTTCTTTCTTTTAACTTTTTATAGTTAAATTTCCGAATTTTAAATTTATCTCCATTTTCATTGTACGGATCTTTAA
CTATATTAGAAGTATTTTTGTGAGGTTCTCCAAATACAACAGGTTTCGCACACATTTATCGAATAATTTTAAATAGAAGAACTTTCCCATTTTTTTTACGACCTCTTTAGATTTAGGTGGA
ATAATTTCATCTGATCTCATTTATAAATTTTAAAGAACTTTAAATATTAAATTAATAGAGATTTGGGCATTATACACTATAGCAGGATTATTTTCTCTGGATCTTAATTTAAAGAAAT
GCTTAACATTTATACTCTATCAATGTGGTTAAATATAAATCTTTGGAAGATGTTGTAAGAAGAAGAAGATGAATATACATTTGATGTTATAGAAATTTAAACAGACAAACGAAAAAGAACTA
GTCGAGAACTGACTTTAGATGAAGAGACTATAAAATCTATAGAAGAACTGACGATATAATTTACATTATAATTGAGTATATGTTTAAATGAACTTTGATGAAATATATCACTTATAAAAAAG
ATATACAATTTCCGATATATAGGTTTAGAGTTGAACCAACAAGTATTATACCTATCCAGAAAGCATTTTCTTAAATGAAATTAATACACTGATTATGTGACTCCATATCTAGGACCGCTGA
TGTAATTTTTATAGAGATCGCGAAGAACAGCCTTTACTCAAAGATATACTATCGGATCTCTAGATATAAAGATTTCAATTTTGAAGATGGTTTAAATTTTATCTTAACGACACTCC
AATTAGTGGTAAAAATGTTTTATCAAAAAATTTCCATGTGATTAAATTTGGAAGAGGATCGGCGGATAGATCTTAAGGATGACGATAAGTTGGATAATTTAATTTGAAAAAAGTAAATTT
AAAAAAGTAAATTTAAAACTAAACAAATGTCTTCAAAATTTTTGTTAGTTTTACTACGGTTGACAGCAATATAATAAAGAAATTTTATTTGAAGCTTTAAAAAATTTTTTGTGCGATGTAATTT
TACAGTAGATAAATCAGGGATTAATTTAACGACAAATAGACAAATCAAAATTCGCAATAGTAAATCTTAATCTTTATTACAGAAAAATTTGAAGAAATATATATGTGAAAAAGAAATTAATATAG
GTATAAATTTACTTTCAATCTATAAAATACTAAAGGTACAAACACGACAGATGATTTCTTTAGTATTTATGAAAAAGATGCTACATTTTAAACAATTAACCAACCAATAGTAATAAA
AGACAACATATAGAGATGAAAAATTTAAATACTAGACATCGATGAAAAAATATACAATTTCCAGATATTTCATTCGATCTCTATATAACAATG

TAACGTCGAATATTGCTTCTCCCATGAAGAACTCAAACGCGTAGATCCGTAGTGTAAATACCTTCTCCTTGATCCGGAAGAAATTAGCGTCGGAAGCACTGCATCTATTAACTCT
GAAATTAAAGTAGAAGTAGAAGCTGTTATTACCGGTTTCGCCGAATCCGTTGTAAGATGTAGAGCAAACTACTACTTGATCTGTTAAAGCTCCAACATGTAACGAAATAAATATAGAGATA
CATTTAAGATTACAGGATGCAAACTCTAAACATATTCCGAGTAGTACCGGTAGACGATGACACGGGGACGGGAATAGGATCTCTATATATATTAAAAACGCTATTGAGTACTCTATTACAA
TATTTTGCCTTAAAGAAATTTCTTTCAGTGTGTAAATACATGATTTTGAATAACAAGATTAGTTTCTAGCGATGTAAATACACCTGTACCAATGTTATTTAATAGAGTTAAACGTTTAGAT
TATCGTTGGTTAGTCTTGGAATAATATCTATAAGTAACCAACAGTTAAGGCATTAGTAGTAGTCCGATTGTGTAAGAACTCCTTTCATTATTAACCTCTTACCAATAAAGGGTAAAGA
TAGCGAAAAATTTAGAGTTTCTCCACTGGCAGTCTGATGACCAATATTGCTTCCGGTATCAAAATATACAAGTTTATTGTAAATGATTTTCATTTCGAATCAAGCTGCCAGATTCTGAA
TAATTTCTCATATATATATCTCCTGGATATATAGTTTGAATAAAGACCGCTACTTTAAATCTATGAAATCAATAAAATCTAACGGTAATGTTTTTGAATAATATATACCACTACAATC
GTAAAACTAGTCGGAGCTGAATTTTATAGTCCAATTTAAATTCATCTCACTGATTGCGTCTATATGTTTGGTATACGAAAAAGTATCCACATTATTAAGTACCGCTCCATTGCTATGA
GGTAAAGCTGTTAAGAGTACCTTAATTACAGTAGGTCCAGATCCGTTAATACTATTTAATGGCGCTAGATATGAAGTCCGAAGTAAATTTGTCGCAACTATCTGCCGCGTTGGCTCT
AGTTATAGATTGAGACCCAGAGAACTAAAAGTATCTACGGCAATATTGCTTATCCCATTTTATAAATAGATAAATAAATTAATTTAAAAAAAAGATTGTTTTAAAAAAAAGATTTTA
AATATCTAAATACTATAAATGTCTAATTTTGAATGTAACTTAGTGATTAAATAATACCCCGGTTGAGGAACTAAAACCTCAAGCAAGTAGTGTTTTGTCTAAAGAGGACACTCAC
AGATTTCTGAAAAAATACACGAACAGAAATTATTTCAAAATTTTGAATGAAAAAATGGTAAATAATCTATTAATACATTGTGACATTTTATATTACATTGAGAACAGGTAAACATCTT
TATAAGTTTAAATTTACCATCGTTGATAGCATGTACATCGCAACTAAATCTTTGGGTAACCACTTTGAGGTATATTAATAGGAATTATATTATCTGATTTCTTTTCCCGGAC
CACTAAATGAATTTTATTGTTAGTTTATTCTCGAGTCTTTCTAGTAGACCATTAATACTAAAAGACTCTTGAGAACATGATATTTTCTTTTGGTTTTTCCAAATTAATGCAGATAAAA
CACCCGTACTCAACGGAACATAAGAACTTTGATAATCTCTACAACAACCATGAATTCCTGATTGTTGATAGACATTTTGACAAAGTCCGGTTGGAGATTATTTAAAAATATATAGATT
ATTTTGTAAATCCATTTTATTTTGACAACTCTCGATTAGTATAAATAAATACCTGCTTTCCGTTTACTTTGAAATATGCCAAGATCTCTACGTTATAGCCGATTGCGTGATT
TTTAAAAATTTTAAATTTCTAAATAAAGTGGGCTATTTTGAAGACAGATATAAGTTGTAATTTAAACCTTGAGTTTCTCTCGTATCGTTTAAATTAAGCATTTTGTAAATTTTCGATT
TTATCGCTTCTTACACTCGTGCTTTTAACTAAAGCGAGCATGTTATTTGTATAATTTACCATCTTTCTTTATGAAATTTTATCTATGTATACATCTTAAGCAATGTATACACGCTTCG
TAAACACGAGTACTATCAGAAATAGTACACTTATCAGATCCTATTAGTCTAAGTCGTTTTTTGTATACACACTTATCTATTATCTTTCCGAGTCGTTATCGAAATGTTCAACTTTTCC
AAGACTGTTTTAAGCTTACTAATAATTTGCTACGAAATTTAATAGCCGTTTCAACGCTCAACATCAATGAGTAACTGAGTTTCTGTTTATTAAGCTTTTAAACCTGTTTATGCTTTGA
TAATTTCTATGTTTTATCAGGAATAGTAGAAATACATTTGAGATCTTTAATAATATAAATATAAAGATTACTTCTGAAACACTTAAATAAGGTTCTTCATCTATAAATTTCTGAAAG
TACATCAAAATCAATAAAATCTAAAAAATCGGTTTTCTTTCTACCAAGCAATTTTGAATTTTATATTTTATAGCATAAAGTTCTTGGAAATGTGTCGAGTCTTCTGATATAAAGTTT
AATCCCTTCCATAGAATAGTGTAGTAGACGAGTTATCGGTTACGATTTTTCCAGTTGAATAAAACCAATCTTAAGTGGATTATCCATTATTGTTAATTATATATATATATGCTATAT
ATAATTTTAAATAAATTTATCGTTTAAATGTAATTACTGCCACACTTCGCCACTCCATAAATTTGAGGATAAATAATCCGTAATATATCTCGTTCTGCCAAAAATCAACATCTGG
TATTTGTAATCTAATCCGCGCAACTAGTGTCCGCAACACTCGAATCGGTCGACAGTCAATGAGTAACTGCTGAAGCGAGTAAAGCTTGTGCTGCCGCTTGTCCGCAATATAGCTTTTGCT
CTTTTAGCGAGCGCTAAGCGTTCTCGTTGCTGCTGCTGCTTTCTGTTGGGGGCTATCCCATGTAGCGAATAGCATCTGCTGCTGCTCTATCCATCCCGGCTTCTCTATTGCTTC
ATATTTGTACAGACTTTTGTGCGGCTTTGCTGCAACGCCATTTTCTATTTGTTGCAAGCAATTTTGTCTTTTCTCGATTGCGTATGCGGCTATCTAGGATCTTCTGCGGCTTCTGCGGCC
TTACCTCTCCCATAGGTAAGAACTCCAAATTTGGATGTTTTGTAGCTTATCCATTACGAGTTAATCTAGACTTATGTTGAGGTATAGTCAACAAGTTCTTGTGATATCTTTTGTCTTT
TCTCATTTTAAAAATAGAAATTTTATACGTTTTGCTAACTTTTGAAGAAAACTTAGATATATTCTAACTTTTCTTAGCACATTTTCTTGACCAAGTTTGGTTTTCGGCTGCTACTGTTGA
TCCAGGTGGACATCTTTTATTACCTTTACGTTTACATTTTATGTTCTTATTTATCGTATTTTGGCTTATCTCTAAGATATGCTCGCTACCGCCAGGAGGAATATCTAGTTCTACACGAA
TACACGTAAACCAATCATTTGTAAGTATACGCGCACCGCTATAATCTATTATATAATATAAATTTTAAATATTTAATTCATATTTTACGAACTTATCCAGCGGATATCTTT
GAATAGTAGTCCATAGGTTGGTGTCTGCTATGCACTGGGCGGTGGAATAAATATCTCCGCAATTTCTCTACGCATACGGTCTGCGACACGGTTTGGCGGTAGATCCACAT
CTACCGGAGTCGCTTTCAGAAATCTCTACGGATATTTCTTGAATCGGATATGCTTCTGTTGGTTGTAACACTCCGAATGCTGCTCCAGCACTGCTGCTCAAGCACTGATATCTTG
CCACCGCATTTTGTATGACCTTGACTACCTCTCGCGCGCTAGCTTCAAGTATTTACGCCACGATGGCAACGGAACCTTTGTCATCTCGCCACTCTGTCCAACCTTCAGAAGAT
TCGCGCCGAGATTGAGTTCTTCTGCGCGACCTCCACAGTCTCAGCTCGTAATGCTAATGTTGAGGCGTCAAAAGTTGGCGAGGTTGACCGGCTTCTGAGGACGGCTCCGATATC
ATAGCCTGAACCTCGTTATAATCTTCTCAGACATTCATGAAATGGAGGTTCTCGGCAATCTCCTCAAGGCGGACTTGTGCTGCTCCATACCAATTTCACTTTGCGCCAAATCGGCT
CACTTGTTCTATTAACAAAACTTAGGATCTATTTATCCCAACCTGCTCAGGAACCTCGTTATAGGACTTACTATAATCCCTTATCTGCCATATATGACCGGATAGCGTCCATCTTC
TCTTCTCATCTCTTGAGGTGTTTTCTGAAAGTTATAATCTGCTATTTAGATTAAATTTGTTATATAAATAAATACCTTCTCGGAGTAGTCTGCTATTTCCATTTTCACTGATAGCAAA
CCTAATTTCTTCTCCTTCAAGTCTTTCCGCAAAAAATGATATGCTGCATATATAGCTCTTTCTGTTAATCGGCAGACTGCCTAAGCATATCGGTTCCAAATGCACCAACATCATCT
GTAAAGTAAACCTCTTCCGAGTCTTTGCCAGATATAACCGCGTTCCCTTACCAGACCTTACTGCTTCACTCGTATTGGCGTTCCAAATGCTGCTTATACG
CCAACCCAGACTCCCTTCTGATGCTATCTTTGGGCGAGTCTGAGCAGGAGGTGACTCATTTCTAGAAAAAGGGAAGTGTGAAGCCATCTCTCTAAATCTCAGTAACTGGAACGAG
CACCTCACTCCCGGAGCAAGCATACCAAAATTTGATATATTCAATTTATAACTTATTCAGTGTCTGTTAATCTAGACTTTAGTTGAGGTATAGTCAATAAGTCTTTGTTGATCGCTTTT
TATCTTTTCTGGTTTAAAAATGAAATTTCTATTGCGTGTCTAACATTTGAAGAAAACTAAGAACAGATTAACTTTTCTTAGTGCACTTCTTGGACCTGGTTTACGTTTGGTGG
TACTGTATACAGGGGGGCGAGGAGCATCTGGGGATACATCACACTACATTCAGTCTGCTTTAGGAACCTCTTGGTCTTAGGAACCTCTGTTGCTTCCGCGGAACCTTATTTCGGA
ATGGATATTAAACAACTGATTTGACTACCGTATTAAATTTACGCGCTGCTCAATTTTAAATGAAATATTTTAAATTAATTTATATCCAGTTGGATAATTTTCTGAATTTTACCTGT
CGGTGCTCTATCTCGCGGATTTCTCTCGGATGATCGAAGCAATTCATCTCGGGAAGTCTGCGCGGCTGTAATTCGAGGAAAAACCAATTTCTCTTTCGCTCATGCAATTAACACCTG
GTAGGATTGTTGATGTTGATGCTCTGCTACCACTCTCCCTGTTTGGCAGCGGCTCTGCGTACGCTCTCTGCACTCATCTACATCTCGTTGAGCCGCTGCGATTCTAACGAAGCT
TGTGGAACATTCTAATTAGTCTATGCTTCAAGTAAATCTGCGGTGATTAGTAGCTGGAGCAGGAACCTCAATTCTTGGAGTACACGTTTGAAGTAGTCTCACTGGAGCACTAG
CGTTCTGTTTTCACTGCTTCCGCGACCACTTGAGCGTAAGATCTACCTGGCTCAGACGAACTGCGCAATTCGGAACAATTTATTTGTAACATTATCCATGCTCTGTTAATCTAGACTT
TAATTGAGGTATAGTCAATAATTTTGTATATCTTTCTGCTTTTCTCATTTTAAAAATAGAAATTTTATACGTTTGTGTAACCTTTGAAGAAAAATACAAATAGATGTACCAACGA
TGACTTTCTCGCCGAGCTGAGTTTCCGCAAGCCAGATACCTCGTAATCTACCAATAGTTTCACTCGTAGTGCTCTGGTAAAAATGCAATATACCTCCCTAATACAATCTTTGAATA
TTCAATTATAAATCCATAGGAGTAGTCAATCCACTCATAGATGAAGTAAATGAAGATATTATATCTGAAGCAAACTCAGGCGATTTCTCTGCCCAAAAGTTTTCTATTGAGCAATAGTT
CGACCCCTAATTTGGTTCTGTAACCTTTGAAGTAGATTCCCTAGTTGCTTAAATATGCTTCTGAATCTCGTTAGGTGTAGACGATCCACCATATCGAAGCATATTGTGATATTA
TACTTAAAGCTAATTGTACAGTCTAGCATAACTCTCAATTTAATAACCAAAAGATTTTGTATAAATGAAGATTGCGATTATTAATTTTATTAATTTTATGCAATTTATGGCCGCGGT
ATATTCCAAACGCGCAATATGTAAGACTACCGCCCGCAAGAAATTAATATTAGTGAATACACCACTGCTGCTCTACGTAATCTCTGTAAGATCCAAGAGCAATCAATATATTAG
TAAGTTCTTGAAGACGCGGATCAAAAGGTTGCGGATTTGAAATCTCTTAGTACAGATTCCATTAATAAATGTCAAAAACAGGAAAACTAAATTAATAAAGTTTCCGAACATAT
TCTAGCCTTTTCAAGCACTCCTTATCAGCAATTCGAAGATTGAAGTCTTCTGCTGACAATTTGATTTTGAAGAAATCAATCTTTTCCGCTTTTCTGCTGATATTTCAAAAGTTT
CAAAAGTAAACGTAATTTTGTAAATAACCTTACTTTCGAAGACACAGGATCAGTTGCTGCGAATTATGCGATGTCAGACATTTCTGCGTTAATAGTGTCTCTATTGCGTTAATAA
CTGTCTAGCAAGACCAAGGAATATCGTGAATAGCATTAATCTGTTGAACCTTCTGCAACTTCTCTTGTAAATCTTCAATTTCAAGATTCTTTTCTTCTGCTATTGATGATGAGACGCGT
TTTTCAAGGGGATTATTCGGAATTTTGAATATGTTCTCATTTAATATGTAATAATTTTAAATTTTAAAGTGACAACCTGAGAATCTGTAACCAACCTTTAAGCAGCATCTGAGATAA
AACTATTCTTTTCCCTTTCTTTTATAGTAAATCAGAGTCCCATATCAAAATCTAAGTTTGTATATAGAAATAGCATATCAAAAAATTTTATTTCCAAAAACCACTAAAAAATTTAAT
TGCCCAACGGTAGTAAATCAAACTATTTTATCTACATTAAAGCATATCTGCTGCTTTTACAAAAAAGTATTACAATCTATTATAATTCGCTCTGCTCAAAATGATGCGAA
AAATTTTTTGAATACGCTTCAACTGATTTTTATAGTCTAAATAAATATTAATAATAGTTATGTCGCCAGTATTTTGAATTTGGATATATATATATCTTTTGAATAAATTTGTTACT
AACCAGTCTATCAGTCTTAACTAATGATTATTTTGATAAACTATATCTTTTAAATTTCAATTTTATCTTTGTAATAATGAGTAAAAATCCTAAAAATAGTCTCTCTTTAATGTACAGT
TCATCACAAATAGTTATTAATAAATTTTCTTATGTAATTTTATAATAATTTATATAAAGAAACTATACATAAATAAATTAACCTTTATGTTGCAAAATATAGATGATAAATTTAAAAAGA
AAATATATGTTTATTAGTATTGTTGCTAGCAAAATATAGCAATTTTTCCAACACGATATGAGTACTCATTTGAAAGAAAGGACTTATTAACTTAAAAATTTTTATATTATTATT
ACAAAAAATAACAAATATGAAAAAAGGCGTTCTATTTTATTTACAGATAGCAATTTGCGAAAAATAGGCGAGTATTGATGTTAAAAATTTTACTATTTATTATAGATATAAATAG
CAGGAGGATTACTATAAAAAATACCATATTGATGTAACAATATCTGGAGACGATAAATAATATCTATGACACTATATACATTCTGGAATTAATAATTAATATACATTATTGATAG
ATAAGCTGAAGCTGAAAAATTTAAGAAGAAATACAATTACACCGAAGTTTGATATATGCGAATACCTAGAAGAAATTAACAATTTAAATGATACCTCAAAACCGTTTGAAGAAGAAATATT
ATAATCTATAATCTACCTATTAAAAAGGTATTAAACCGGTTGCTTTAAATGGCAACCAATTTGAATTTTATTACTTTAGTTTAAAGTAAATAGAAATGAGACGATTAAATATTATAT
ATTATGTAATTACAAAAAGACTTACCTTTTGGCAAAATACACGATCAGATCCGAGTAAATATATAAACAGATTAATAAATCTAGAAAAATATACAAATGAATGTTATATAGACA
TTTGCTTTAATAAAAAATACAAATTTTACAGTAAAGTAAACCTTTCCAAGTTCTTGTAGACATGTTGAAAAATCTACATATTAATAAATGAAAAATATACTATACAGCAACTGTGAAAT
AAAGAATAATAGTAAATAGAAATACAAAAATAACATAATAGAATAATTTAAATTTTATTATTCTATTA

ATGTCCAAGGTTATGCATAACATCCCGTCTCTACTAATCCAATACAGATTACAGGCACAGTACCAAGTTCTTCAGTATTGGGTATAATGTTTCAACCTCGACGGCGTTCGCCAATTA
TAGTGGTGGGGTCTCAGAGTCTTACTATGGGATCGTTTTCTCCGACTAATAACAATAACAGGTAAGGACGAATAGTATTAATAACTTTGTACCCACACCACTAATTTTACAATGTTGATATT
TCTCTAAACATATCAAGAGCTACTAGTATTACAGCAAAATGTTACTGCTGGTAATTTTACCTTTTCAAAATAGTAATCTTAGAGTTACAATAATAACGGAACTCCGCTGATATAGCATTAT
CTGAGGGGGGAAAAACAGTATAGGAAGGTAACTCTTCACTAATATTGTTTCTATAACAGTTTAAATGGAAATGATTCATAGGTTTACATTAATCTTCTACTGACGGATATAATAGCCACT
TATTTGGGACCAAAAGTTATAGTCGAATACCGGTCACTCACATTCAGTTTGTAGTTTAGCCAAATTTAAATATCGCTATTATAGGAAGCGAAATTAATAGTATAAATCTCTACAAATTCAT
CTGAATCCAAAGATCATGTAATTTTAAATTAATCTTACGGGGGTCTCAAAATAGTATAGGCAATTTTAACTCTAACGTAATCTAGAGCAATAGTTCATGTCCATCTCTGAACCTTA
GTATTTGAAGGAACATTAGAGATTAGGTTGTGCACTAAGCATTATTTTATATTCATTGTATCACAGGTTATAGCAGTTACATATACAGACACAGCAACACATACTATTTTCAATTTACATT
ACCTTTCTACGGCAGCAGTGAACGTAGAATTTCTAGTAATAAGAGCCCTGACATTTTCTTTTACACAATCTGTTTCTAACAGCTGCCAATCCTATAATTAACAGTAATGGGATCCGAGTT
ACTAGAAGTAACAACAATTTTGTGTTAACGTCGCCGATTAAACAGTAAACATCTACTGTTTACTGTAACCTGGCAGTAACAATAACTCTACAGAATTAACCTAAGAAATTTAAATTTAAGCC
CGGCAAAATTTTACATCATCAGTAGATACAGGACAGTTAATCGATTTTGTGTTGACGTTACCAAGTAAACACAGCAGTACGTACTAATATTATTACTCACACACCTAGAACTATTA
TTATAAGATATATCCACTTTATGACAAAATATAACAGAGGTGAGCGAGTTAAGAGCAGAACCTTTTAACTCCATCACAAGCTAAATACACAATAACAGTAACAGCAACGCAAGTAACTTC
AGATAATTGGGAAGTATATAATAGCAGTAAATTCATGAGGTTCTCTACCTACTACAGTTCAAACTCGTTCAAAATAGTCCAGGACATTCTCAGAATAAAGAAATAGACCTCTGCATACT
CTACAATACTACTCCTTACCTAATGGCACATATGCCAAATGATGTCGTAAGGTAACCAATGATCTACCTTAAATATATCCATGACACAGGATACGACCTCAAATAGTCCGACGAAG
AGTTCACATTGCTGACGTAATCAATTTCTCTGGTCTCTCTGTTTACTATATAACTACAAATGCTCAGTCCGGTGAGGCGATTCTTACTAAGACCTGTGGGCAATGTTATGAATGATACTA
CAGTTTATAGGCGAAGTAATGATATCAGAAATATACACTAAGTACGAGCTCATCTTCACTGGTCCATATAGTCACGCACTTTTGGAGAGTTAACTCAACACACGACATATCTTTGTGA
ATTCAATCCGGCTTGTATTTATCGCATCCGATTAATCTATCTGATTTATACATAAGCGCACAAGTCTCGGAATCTTATACGAGCTCTACTGACGAGTCTAATAGTAAATTTAGAAATATATCAG
AACCTGTATATGTCAGGTTTAAACAATACTAAACCTTGGATTTAGAAGTTACAATTTCTCCAAATCAGTCGTCATTAAATGGGACTCAGACATTAATAATATTAGTAAATCTTCTATATA
CCAGCTTCTACTCTCAATTTCAAAATTAATTAAGTATAGTAACTTTAAACAATACATCGTTAAATAGTATATCTTATATTTCCGCGGGTCAAGACATTTTCAATTTATAGAAATTTAGAAAT
TCCAGATAAATATTACGAGCGCGGATTTATAACATTTATCCACCTAATACCTTTTACAGATAATGGATTTTAAACAATCTTTATGAATAGACACTGACCGTAATGCCAGGCGATGTTCT
GAAAGAAATTTTATTAATAATAATAACAACCTACAGTAATCTTACTACACCTAATGCACTAATGCAATATTAAGCAATATTTAGAAATTCAAATATCAGAATAAAGGCTATTAAACAGCTTCC
ACAACTCGGTGGGCTATTATCTTCTTACTAATCAAAATTCGATCAATGTTGACACAGGATTTGACCAACGAAATACATTTCTGTTAAAAATTCGCAATCAAAATTCGAAATCCAAACA
ATAAATTTAAACACGGAAGTACAGAACAGATTTATATTTAGTCTTACTATAATCTTATACAACTTTTACGTACATTAAATGACAAATTTAGTCTGAAATTTAGCTGTTGAGAGGATCTTCTCAT
TTCACTTACTGAGGACTGAATTTAGACAAGGACAAAGATTAAATGGTAGAACTAAAAAGTAGACAGGGTTAAGCGGAACATAAACACTTTTTTATACCAAGTTCTAGGCGAGGCTG
CTCAATATACGACCATTAACACGGGATCGTTTATAAATTTACCAGGAAGTATGCACTGACATACGGGAGGATATAATAGTACGCGAGGTTGCTAATAATATACAGGAACCTGGATTG
ATTGAGGCACTATCTAATGCGTCAATACCTTTGACTTAGGTGTAAGCAGCTTTTGGAGATGATGATCAACAACAGTCCATAGAAAAACATAATAAATGAAGTACCTCCGAGGCTGGT
TCCTCAAAAACTATTATTGTAATAATTCAGATCCGTTAATTTTATTATTTCCGCGACATCTACTAGTACACTAAGTAATTTAGTCTGGAAATTTATCTAGTAAAGAAACCTTGACCTCAA
ATAATTTTAAATGATAGTAACTAAGTAACTTACCTGCTGCGTATGATATTTCTAATAATGGCATATTTGGCTTCAACGCACTAGGAGTTACAGGTTTAAATGTTACGTTATACAGACAGCGGTT
ATTATCTTCCGCGACTCTCGCATTTTATAAGTAATAAATCTCGTTGTGACTATAGTATATACGATAGTTTACGAAATTTGAGTTTGTAGTTTCCAAAACAATATGCGGTAGTATTAA
CTAGCACAAACGTCGCCAGCGGATCGCTATTCTCAGAACGGAAGACCTACTACATATGCAATACGTTACGTTTGAATCTGTCACTGCGCTACAGTTAAATAATGACGAATTTATCTCAATTA
AAAAATTGCTGCCATGACTATATATTATCGCGCGGTAGTAGTAGCTAGTACACGGAATTCGCCGTAGTAAACACTAATCTGTATACACCGTTTATGATGATCAGTGGGATGAATATGTT
ATAATGAAGCGGGCGGAATACCAACACAGGCGGGAATAGAGAAGGATTTAGAAATTTTATGGTCTCATCGGGTAGCAGGGGTTACAACTTAATCATACTGATGATTTTCTCTAA
TTAACAGTAGCTTAATACATAAATTTGAAGTATAACAGATAAAATTTGAATTTAATTTAGTAAATAGTAATTTCAAATTTAATCGGTTATTAATTCAGTTAAATTTTAAATTAACCGC
ATCCACGCGATAGTACGTAAATGGCAAAAAATAATACATTTTAAACCAATTCGACAGGAGTCTGCCGCGTGAAGCAATTTGATAATTTCTTAAAAAAATGGCATTAACATA
GGATGTGGAAGACGATATCACCTGGCATTTATTATTACATTATATAAAAAAATAAATTAATTTTATGATGAGGTATTAATAACACCAACATTTTACATATTTTCTAAAGAAATAAAA
TTAGAAGTTCCGCTTTTAAATCCAAACAAATATAAGTCTTTTGATATAGAATTTAATATGTTGCCCTACTGAAATAAAGTATTTAAATCTAGTACAGCGTTTAAATCTATTTCTGTGAAGA
TTTTTATAATAGTCTACCATATCTACTAAATTCGCGATGGTACCATAGAACATCTCGGACTGGTTCTTCTGTACCGGTTCCGGTCTATTAGTTGAGGCAACAGTAAAAAATAAAC
CATCTGGCTTTAAACATATCGATATCTTTTAAAAAGATCTTTTATAATTCAGGGTCTGTTTCAAAACATCTAGTTGATACGATTTGTCAAAGGTATCATTGTGAATGGTAAATCTTTTGT
TTAGATACAATAGTTACATTCTTCTGCTGTATTACATCACTCCCATTTATAATTTACACTCTTCAATATAGAACAGATTAATTCATTAATATCCCCGCAACCACTGTTTAAATTTGA 8971
ATAAAAAATTTGTAAGTATATTTTAAACAATAGTGTAAATCTCTAGCTTGTGGATGATATTTGTAATTTAATAATAATAATCTTTATATTTATTAAGAAATAGCTCTCTCCATTTT
GAGATCCGACGCAAACTTAATTAATTTGCTCAAGATTACATAGACGCGGGTCTCGATAGATTACCCAGCAGATGAAGTTGTACATTTGATACATTTTGAATTTTGAATTTTATGATAA
GATGATAAGCAAAATCTCAAACTAAACAGGCAATGACGATGTTAAATATAGCCCATATATAGGCCCTTATGACCCAGTTATAGCGGATCTTAAAAAATGAATTTAAAGGATA
AATAGGAATATAAAAAGTCTCCGTTTAAATGTAAGCGCCGAGAAGGATGTTATGCGCGTCTCCGAGTTCTAGCTCCGCAAGAGGCTGTTCTGCCG
GTCTCCGAGTTCAGTCCCAAGAAGGTCGTTCTGCCGCTCTCCGAGTTCAGTCCCAAGAAGGTCGTTCTGCCGCTCTCCGATTTCCGGTGTAGTCTCAAGTAGCGAC
GCAGCTCTCTCAAGTTTCGGAACCCCTAAGTATATAACCGTTAACGCGCGAAGAGCTAACTTTATAGAGGATATACGCGGGGTTTATTATAAAAAACAGCTGGGTAAAAAATTTA
TTTTAAATCGAAAAATGAAAAAATAAATCTTAAATCAGCAGATGGAATCTCGGTACCGGATGGAAAAATTTTAGATTTTTATGAAAAATCAATACGAAAAATACATGGTACAC
TCGCTCTACTAGGTTTAAAGACAGTAAAAATAGGTGACATGCTGGAATTTACTGTTAAACAGACGATTTCTGGAGAAATGTACGATAATATCTACAGATAAATGTTAGTCCAGACGAC
AAATTAGCTGATGATCTAGAAATTCGAAGTTCGTAAGAAAGTAAGACTCCCGCTCTATGAACGTCCTCAAAATTTTCAGCAGTGTCTTCTCGGTCATACAGCTAGAGCGGTTCCGCA
ATTGGTATGTATATAAATCAAAACCGGAATTAATTTAACTGAGGTATAAACTAATATTAATAAATAGTATAATACATTAATTAATGTGCGCATTTAAGTTTATAGTTTACGCGAGTACAA
AAAGAAATAGAAAAATTTAAAAATAAAAAACACATTTATACTTTAATATTTAGAAATTTTTGAAAAATTTAAATTTGAAATATATCACAAAAACAACAAAAATTTTAAATGATTATGATAA
ATACTTAAATCTTTTACACAATTTGAAAAATAATCAGTGTATCTTTTAAACCTTTAATGTTAGATATCTAAAGTTCAAATATAGTCTGTCATTTTAAATAAATTAACCTGACTGA
AGTATTTAAAAATAGAGATAAAACATTTTATATTTTTTCTTATAGATGATGTACTCTTTAATCTCATGATTTTAAATTTAGAAAAAATGAATGTTACTGTTGTAATCCGCTGATGTGAC
AAACGAATGGGTTCGAAGTTTACCTTCACAAATTTTATTATAGAACAACTCGAAATCAGGTTTATGAATGTATTTCGCTACTCTTGGATATAAATACTTGAATAATATACGATACAT
TATTTATTAATTTCCAGTGATATAATTTAAATATATTATATACATACTAGAATAAAAATATTTGTTATTGTAAAAATGTTGATCAAAATTTTATATAATTTATTTATTTATCTC
GGTATCGGCTCGGAATTTTTTAAATTTTAGCATTACAAATAATTTATTTAAATTTAATCTCAAAATAAATTTCTAGTTAGTATAAAAAAGGTTAATTAATCGCCCCAAAGAGCGATTGCGT
GTTAATGCTATTTTTTGGAGTATTAAACCCGCAATTAATTCGCTCGTATAGAGGAAAGACAGACAGTAAGAAAGCGGACTATACACCCCGAAGGTCGTCGCGTAATGCTTTAAT
AAATTAGCTGATGATCTAGAAATTCGAAGTTCGTAAGAAAGTAAGACTCCCGCTCTATGAACGTCCTCAAAATTTTCAGCAGTGTCTTCTCGGTCATACAGCTAGAGCGGTTCCGCA
CTCCTTCTCTATCTCAAGCCACGTCCTGAGGCGCAATGTTACACCTAGAGCATACCGCAGTATAGAGACGTCAGTCTCGGCGAGACATTCAAAATCTCGGCGTCGCGCAATCTAC
CATCGCAATTTGCGAGAGCATCTGAGCAATGCGAACATCCCGCAAGCTCGGTCAAGAAATTTGCGCGCAGTTTCCCAATCGCCGTCAGTAGAACCTCTATGAAGATTGCGCGTGT
TGCGTATTGTACAGACTGGCGCATGTGTAATCGTGAATATGTACATCCGAGAAACGCTGTGAGAGCTAGAAAGCGTCTGTACAGTCTATACCTGCTGCTCGGCGCGTTCGAATAT
TGTAAGCCACATCCGCTGAAGATCGAGCTACAATATCGCGGCAGCTGTGAGCGATTGTACGTTCTGAGTTTCAATGCGGTCCTCTGTAAGAAAGTTCGCTGAGCGGCTTG
CGAAATCCGTCGAGCGACTAGCCACCTCTGTCAGCTTCAGCAGGAACCTCGAAATGCCCTTAGACACCTGCGCGAGTCTACATAGACTTATCGGGTCCAGTCTCTGAAAGCTCTTGACCA
GAAGATCGCATGGAAGAGTATCATATAATCTGGAATTCACGTGACCTTTATGCGCCCGAAGGCGCTGCGCTCATGCTATTTTGGGGGTTTAAATTTTACAAATCAGCTTTCA
CGTTTTAAGATTTAATACCACGACCTCATCCAGCGCGCAAGTCTACGAGCAGCAGCTGTTGCTGCGGCTCTCTGCCACCTTTTGACAACTGGTGGCGCTGTAATCCGCTGCA
CACATTCGTTCTCGCACCTTGGTCTCAATTCAGCTCGACACTCTCTTATCCGATGAAGTTGCGATCCAGAAAAATGAGGCTCTGTGAAGTTGGAACCTCGGACGACAAATGCTACTAG
AGCGAAGAGAGCAAAATCGATTCCACATTCGAATGTCGACTCAATTTAAATTTCAACGACAGTTTATAGATAATAATTAATTTAGCAATTAATACAGCGGCACTTACACACTCTGAAGA
GTGTTGCCGTTTTAATCTAAATGAAAAAGAAAAAATTAACAAATTTCTCCAAATGCTTATAGGCGACTTATTTCTGATGTACATTTGCGGCGACCTACTTCCGCTCTCAAGATTTTGA
TTTGAGATATCAGGTTCTCCAGGTATGCGGTACATTTCTGCTACCTGATTGTGTGCTGGAATAGGAGATTTGATGCTCTGTGACTCGGTCAAGTTTGTGCTGAATGGGCAATTT
CGCACTTCGATTGGAATGAGATCCGATTAACACGAGCTTACAGTCTGTTTCTATTCTAATTTGTGCGCAGGACTTATGCGCTCCCTAAGGAGCGCTGCGGTGAATATTTAAGTTGA
CATGATACGAAGAGTATGTTGTGTAATACTAGATAATTTGTGACTATCAATGCGGTACAGACAGATAAGTGACAAAAATCAGGAACTTCAACAACTTACGCTTACGCT
GCTCACTGAGTTTTTACGCTGCGCAATAGATCTAAGTACTATGAATGTAGAAATGTTTCAAAAGCGCTCTCAACTGTAACTCTTGTGCGGCTTCCATCTCAATCTTCCGCGCA
CATAAACCGAAGCGCTTTGTAACACATCTGCTCGGCTTCCCGACGCGCTCTCCAGCTGTAGTACAAGCTGAATTCGCTAATAAATAGTTCTCATCTGCGGCTATCTGCGCTGACAT
GAGAGCATCTACTGTTGTCCTCTCGGTAGCCATTTTCCGCTTCTCAAAGATAATTTGAGCGCTTTCGAACCTTTTTTTGCTACCTGTTGAGCGCCCAACGAGTCTGCGCTACTTT
TTTGTGTTGAAACGCTCTCGAGTGCCCTGCAGCGAGTGGTGAATAATGTTTCTTGATTTCTACCTCAACAAGCGCTGAACCTCGCGAGTCTTGTCTTGTCTGATCTGGATTTGAGGG
TTGCGATTGCTTTGAGCGCTATTAAATTTGTCAGGTGACCTTATGCGCTCCCTAAGGAGCGCTGCGGTGAATATTACAGAGATCTCAATCTTCGTTTGGAAATCCATGCTTGA
CCACGCGGAATGAGGCAACAAATGACCCACGCCATAGCTATAGGCAATAGAGTG

Supplementary Table 4.3. CpV-BQ1 Predicted ORFs and Functional Annotation. GeneMarkS, GeneFragScan, Prodigal, and Glimmer were used to predict ORFs. Putative ORFs predicted by two or more tools are listed. Each ORF was assigned a locus tag and functionally annotated. ORFs that could not be assigned a functional annotation are labelled as hypothetical proteins. The location of each ORF start and end position and strand they are encoded on are listed. A total of 193 putative ORFs were identified, of these, Prodigal (P) predicted 185, GeneMarkS (GM) predicted 190 ORFs, GeneFragScan (FS) predicted 170 ORFs, and Glimmer (G) predicted 188. For gene annotation, the number of source hits and source names for each protein are listed, collectively, 23 prediction tools were used. Annotation prediction tools included, Blastp analysis against the IMG-VR db, non-redundant (Blastp-NR) db, and non-redundant db with the Virus filter (Blastp-NRV); InterPro, SMART, ProSiteProfiles, ProSitePatterns, SUPERFAMILY, Pfam, Gene3D, Phobius, TMHMM, SignalP, NCBIfam, PRINTS, PANTHER, CDD, PIRSF, FunFam, Phyre2; and HMM databases efam-xc, RVDB, PHROGS, and VOGDB.

Locus tag	Product	Protein accession	Start position	End position	Strand	ORF tools	Number of source hits	Source Hits
BQ1_001	hypothetical protein	XNN68104	55	678	+	P/GM/FS/G	4	Blastp-NR, IMG-VR, RVDB, VOGDB
BQ1_002	hypothetical protein	XNN68105	2018	1569	-	P/GM/FS/G	3	Blastp-NR, IMG-VR, RVDB
BQ1_003	hypothetical protein	XNN68106	2140	2901	+	P/FS/G	2	IMG-VR, VOGDB
BQ1_004	RNA binding domain containing protein	XNN68107	3644	2958	-	GM/FS/G	5	Blastp-NR, Blastp-NRV, IMG-VR, RVDB, VOGDB
BQ1_005	hypothetical protein	XNN68108	3885	4328	+	GM/G	1	IMG-VR
BQ1_006	hypothetical protein	XNN68109	4407	5015	+	GM/FS/G	1	IMG-VR
BQ1_007	hypothetical protein	XNN68110	4999	5400	+	P/GM/FS/G	1	IMG-VR
BQ1_008	F-box domain containing protein	XNN68111	5793	6443	+	P/GM/FS/G	8	IMG-VR, RVDB, VOGDB, SMART, SUPERFAMILY, Pfam, Gene3D, ProSiteProfiles, InterPro
BQ1_009	hypothetical protein	XNN68112	7150	7689	+	P/GM/FS/G	4	Blastp-NR, Blastp-NRV, IMG-VR, PHROG, efam-xc
BQ1_010	F-box domain containing protein	XNN68113	7749	8501	+	P/GM/FS/G	5	IMG-VR, RVDB, ProSiteProfiles, SUPERFAMILY, Pfam, InterPro
BQ1_011	hypothetical protein	XNN68114	8618	9091	+	P/GM/FS/G	1	IMG-VR
BQ1_012	hypothetical protein	XNN68115	9259	9591	+	P/GM/FS/G	1	IMG-VR
BQ1_013	F-box domain containing protein	XNN68116	9687	10196	+	P/GM/FS/G	9	Blastp-NR, IMG-VR, RVDB, VOGDB, SMART, ProSiteProfiles, Pfam, SUPERFAMILY, Gene3D, InterPro
BQ1_014	hypothetical protein	XNN68117	10320	11030	+	P/GM/FS/G	1	IMG-VR
BQ1_015	F-box domain containing protein	XNN68118	11116	11862	+	P/GM/FS/G	7	IMG-VR, SMART, ProSiteProfiles, SUPERFAMILY, Pfam, Gene3D, InterPro
BQ1_016	hypothetical extracellular protein	XNN68119	11937	12236	+	P/GM/FS/G	3	Phobius, SignalP, IMG-VR
BQ1_017	hypothetical protein	XNN68120	12401	12240	-	P/GM/G	2	IMG-VR, RVDB
BQ1_018	transmembrane protein similar to AMEV204	XNN68121	12505	13218	+	P/GM/FS/G	5	IMG-VR, Phobius, TMHMM, RVDB, VOGDB
BQ1_019	hypothetical transmembrane protein	XNN68122	13250	13456	+	P/GM/G	3	IMG-VR, Phobius, TMHMM

BQ1_020	hypothetical transmembrane protein	XNN68123	13854	13453	-	P/GM/FS/G	4	IMG-VR, Phobius, TMHMM, RVDB
BQ1_021	hypothetical transmembrane protein	XNN68124	13908	14087	+	P/GM/FS/G	3	IMG-VR, Phobius, TMHMM
BQ1_022	RING finger E3 ubiquitin protein ligase	XNN68125	14225	15103	+	P/GM/FS/G	11	Blastp-NR, Blastp-NRV, IMG-VR, RVDB, VOGBD, SMART, ProSiteProfiles, SUPERFAMILY, Pfam, Gene3D, InterPro
BQ1_023	hypothetical protein	XNN68126	15211	15507	+	P/GM/FS/G	1	IMG-VR
BQ1_024	Zinc-finger, ring type domain containing protein	XNN68127	15552	16211	+	P/GM/FS/G	10	Blastp-NR, Blastp-NRV, IMG-VR, RVDB, VOGBD, ProSiteProfiles, SUPERFAMILY, Pfam, Gene3D, InterPro
BQ1_025	hypothetical protein	XNN68128	16317	16931	+	P/GM/FS/G	1	IMG-VR
BQ1_026	hypothetical protein	XNN68129	17017	17223	+	GM/FS/G	1	IMG-VR
BQ1_027	hypothetical protein	XNN68130	17268	18155	+	P/GM/FS/G	3	Blastp-NR, IMG-VR, PHROG
BQ1_028	hypothetical transmembrane protein	XNN68131	18183	18731	+	P/GM/FS/G	3	IMG-VR, Phobius, TMHMM
BQ1_029	DNA cytosine-5 methyltransferase	XNN68132	20038	18737	+	P/GM/FS/G	15	Blastp-NR, Blastp-NRV, IMG-VR, PHROG, efam-xc, RVDB, VOGBD, ProSitePatterns, NCBIfam, SUPERFAMILY, Pfam, Gene3D, PRINTS, PANTHER, InterPro
BQ1_030	glycosyltransferase family 2 protein	XNN68133	20111	21163	+	P/GM/FS/G	12	Blastp-NR, Blastp-NRV, IMG-VR, efam-xc, RVDB, VOGBD, PANTHER, SUPERFAMILY, Pfam, CDD, Gene3D, InterPro
BQ1_031	E3 ubiquitin-protein ligase	XNN68134	21217	21729	+	P/GM/FS/G	12	Blastp-NR, Blastp-NRV, IMG-VR, RVDB, VOGBD, SMART, ProSitePatterns, Gene3D, ProSiteProfiles, SUPERFAMILY, Pfam, InterPro
BQ1_032	FkbM family methyltransferase protein	XNN68135	21810	22706	+	P/GM/FS/G	8	Blastp-NR, Blastp-NRV, IMG-VR, efam-xc, RVDB, VOGBD, SUPERFAMILY, InterPro
BQ1_033	hypothetical transmembrane protein	XNN68136	22721	23713	+	P/GM/FS/G	2	IMG-VR, Phobius
BQ1_034	hypothetical transmembrane protein	XNN68137	23883	23710	-	P/GM/G	3	IMG-VR, Phobius, TMHMM
BQ1_035	D12 class N6 adenine-specific DNA methyltransferase protein	XNN68138	23949	24851	+	P/GM/FS/G	17	Blastp-NR, Blastp-NRV, IMG-VR, PHROG, efam-xc, RVDB, VOGBD,

								PANTHER, Pfam, Gene3D, PRINTS, NCBIfam, PIRSF, Gene3D, SUPERFAMILY, ProSitePatterns, InterPro
BQ1_036	hypothetical protein	XNN68139	25422	24841	-	P/GM/G	4	Blastp-NR, Blastp-NRV, IMG-VR, VOGBD
BQ1_037	hypothetical protein	XNN68140	27064	25511	-	P/GM/FS/G	3	Blastp-NR, Blastp-NRV, IMG-VR
BQ1_038	hypothetical protein	XNN68141	27277	27199	-	P/GM/FS/G	1	IMG-VR
BQ1_039	hypothetical transmembrane protein	XNN68142	28064	27318	-	P/GM/FS/G	2	IMG-VR, Phobius
								Blastp-NR, Blastp-NRV, IMG-VR, PHROG, efam-xc, RVDB, VOGBD, Pfam, PIRSF, FunFam, Gene3D, ProSitePatterns, PANTHER, SUPERFAMILY, Gene3D, InterPro
BQ1_040	thymidine kinase	XNN68143	28623	28096	-	P/GM/FS/G	16	IMG-VR, Phobius, TMHMM, efam-xc, RVDB, VOGBD
BQ1_041	hypothetical transmembrane protein	XNN68144	28968	28660	-	P/GM/FS/G	6	Blastp-NR, IMG-VR
BQ1_042	hypothetical protein	XNN68145	29043	29201	+	P/GM/FS/G	2	Blastp-NR, Blastp-NRV, IMG-VR, Phobius, TMHMM, RVDB, VOGBD
BQ1_043	hypothetical transmembrane protein	XNN68146	29797	29204	-	P/GM/FS/G	7	IMG-VR
BQ1_044	hypothetical protein	XNN68147	29848	30051	+	P/GM/FS/G	1	Blastp-NR, IMG-VR, Phobius, TMHMM, RVDB, VOGBD, PANTHER, Pfam, InterPro
BQ1_045	Diverse functionality domain-containing protein	XNN68148	30081	31643	+	P/GM/FS/G	9	IMG-VR, Phobius, TMHMM
BQ1_046	hypothetical transmembrane protein	XNN68149	31671	32288	+	P/GM/FS/G	3	IMG-VR, Phobius, TMHMM
BQ1_047	hypothetical transmembrane protein	XNN68150	32409	34271	+	P/GM/FS/G	3	Blastp-NR, Blastp-NRV, PHROG, efam-xc, RVDB, VOGBD, SUPERFAMILY, Gene3D, CDD, NCBIfam, PRINTS, ProSitePatterns, Pfam, PANTHER, InterPro
BQ1_048	Ribonucleoside-diphosphate reductase large subunit	XNN68151	34374	36686	+	P/GM/FS/G	15	Blastp-NR, Blastp-NRV, IMG-VR, Phobius, TMHMM, RVDB, VOGBD, ProSiteProfiles, SUPERFAMILY, Gene3D, Pfam, InterPro
BQ1_049	ERV/ALR sulfhydryl oxidase protein	XNN68152	37107	36658	-	P/GM/G	12	Blastp-NR, Blastp-NRV, IMG-VR, RVDB, VOGBD, Gene3D, SUPERFAMILY,
BQ1_050	translation initiation factor 4E	XNN68153	37156	37707	+	P/GM/FS/G	10	

BQ1_051	hypothetical protein	XNN68154	37751	38740	+	P/GM/FS/G	4	PANTHER, Pfam, InterPro Blastp-NR, Blastp-NRV, IMG-VR, VOGBD Blastp-NR, Blastp-NRV, IMG-VR, PHROG, Phobius, TMHMM, efam-xc, RVDB, VOGBD, PANTHER, Pfam, Gene3D, CDD, SUPERFAMILY, ProSitePatterns, InterPro
BQ1_052	Ribonuclease-diphosphate reductase small subunit	XNN68155	38805	39767	+	P/GM/FS/G	16	
BQ1_053	hypothetical protein	XNN68156	39822	40547	+	P/GM/FS/G	2	Blastp-NR, IMG-VR
BQ1_054	hypothetical protein	XNN68157	40590	40835	+	P/GM/FS/G	2	Blastp-NR, IMG-VR
BQ1_055	hypothetical protein	XNN68158	40887	41294	+	P/GM/G	2	IMG-VR, RVDB
BQ1_056	Putative Metallopeptidase WLM domain containing protein	XNN68159	41801	41286	-	P/GM/FS/G	4	Blastp-NR, Blastp-NRV, IMG-VR, RVDB
BQ1_057	hypothetical protein	XNN68160	42246	41806	-	P/GM/FS/G	2	Blastp-NR, IMG-VR
BQ1_058	hypothetical protein	XNN68161	42301	47382	+	P/GM/FS/G	2	Blastp-NR, IMG-VR
BQ1_059	hypothetical protein	XNN68162	47424	47768	+	P/GM/FS/G	3	Blastp-NR, IMG-VR, VOGBD Blastp-NR, Blastp-NRV, IMG-VR, PHROG, RVDB, VOGBD, SUPERFAMILY, CDD, SUPERFAMILY, Pfam, Gene3D, PANTHER, InterPro
BQ1_060	DNAJ homolog subfamily A member	XNN68163	47801	48868	+	P/GM/FS/G	13	
BQ1_061	hypothetical protein	XNN68164	48904	49281	+	P/GM/FS/G	2	Blastp-NR, IMG-VR
BQ1_062	hypothetical transmembrane protein	XNN68165	49467	49258	-	P/GM/G	3	IMG-VR, Phobius, TMHMM
BQ1_063	yqaJ-like viral recombinase domain containing protein	XNN68166	49504	49701	+	P/GM/G	3	Blastp-NR, IMG-VR, VOGBD
BQ1_064	pyrimidine dimer DNA glycosylase, endonuclease V	XNN68167	50107	49682	-	P/GM/FS/G	6	Blastp-NR, Blastp-NRV, PHROG, efam-xc, RVDB, VOGBD
BQ1_065	hypothetical protein	XNN68168	50480	50253	-	P/GM/FS/G	1	IMG-VR
BQ1_066	hypothetical protein	XNN68169	50721	50500	-	P/GM/FS/G	1	IMG-VR Blastp-NR, Blastp-NRV, IMG-VR, efam-xc, RVDB, VOGBD, PANTHER, ProSitePatterns, Gene3D, Pfam, SUPERFAMILY, InterPro
BQ1_067	RNA cap guanine-N2 methyltransferase family protein	XNN68170	51398	50886	-	P/GM/FS/G	12	IMG-VR, RVDB, VOGBD, Pfam, Gene3D, SUPERFAMILY, InterPro
BQ1_068	DNA polymerase family X	XNN68171	52187	51462	-	P/GM/FS/G	7	Blastp-NR, Blastp-NRV, IMG-VR, PHROG, RVDB, VOGBD, PRINTS, Pfam, PANTHER, Gene3D, SUPERFAMILY
BQ1_069	AAA family ATPase/CfxQ-like protein	XNN68172	52224	53210	+	P/GM/FS/G	11	

BQ1_070	hypothetical protein	XNN68173	53266	53655	+	P/GM/FS/G	2	IMG-VR, RVDB
BQ1_071	DUF5871 protein	XNN68174	53711	54340	+	P/GM/FS/G	5	Blastp-NRV, IMG-VR, RVDB, VOGBD, Pfam
BQ1_072	GxxExxY protein	XNN68175	54374	54754	+	P/GM/G	8	Blastp-NR, Blastp-NRV, IMG-VR, RVDB, VOGBD, NCBIfam, Pfam, InterPro
BQ1_073	hypothetical protein	XNN68176	55497	54751	-	P/GM/FS/G	5	Blastp-NR, Blastp-NRV, efam-xc, RVDB, VOGBD
BQ1_074	DUF5764 protein	XNN68177	55548	56522	+	P/GM/FS/G	7	Blastp-NR, Blastp-NRV, IMG-VR, RVDB, VOGBD, Pfam, InterPro
BQ1_075	hypothetical transmembrane protein	XNN68178	56880	56530	-	P/GM/FS/G	3	IMG-VR, Phobius, TMHMM
BQ1_076	DUF5759 protein	XNN68179	56932	57294	+	P/GM/FS/G	6	Blastp-NR, Blastp-NRV, IMG-VR, RVDB, VOGBD, Pfam
BQ1_077	hypothetical protein	XNN68180	57294	57428	+	P/GM/G	1	VOGBD
BQ1_078	Holliday junction resolvase, A22	XNN68181	57463	57972	+	P/GM/FS/G	8	Blastp-NR, Blastp-NRV, IMG-VR, RVDB, VOGBD, SUPERFAMILY, Gene3D, Pfam
BQ1_079	hypothetical extracellular protein	XNN68182	58008	58430	+	P/GM/FS/G	3	IMG-VR, RVDB, VOGBD
BQ1_080	DUF5754 protein	XNN68183	58923	58603	-	P/GM/FS/G	7	Blastp-NR, Blastp-NRV, IMG-VR, efam-xc, RVDB, VOGBD, Pfam
BQ1_081	hypothetical transmembrane protein	XNN68184	59428	58952	-	P/GM/FS/G	5	Blastp-NR, Blastp-NRV, Phobius, TMHMM, VOGBD
BQ1_082	hypothetical protein	XNN68185	59986	59450	-	P/GM/FS/G	1	IMG-VR
BQ1_083	SWIB/MDM2-domain containing protein	XNN68186	60100	60714	+	P/GM/FS/G	13	Blastp-NR, Blastp-NRV, IMG-VR, RVDB, VOGBD, Pfam, Gene3D, PANTHER, SUPERFAMILY, CDD, ProSitePatterns, SMART, InterPro
BQ1_084	hypothetical protein	XNN68187	60777	60956	+	P/GM/FS/G	1	IMG-VR
BQ1_085	hypothetical transmembrane protein	XNN68188	61384	61076	-	P/GM/FS/G	5	IMG-VR, Phobius, TMHMM, RVDB, VOGBD
BQ1_086	hypothetical transmembrane protein	XNN68189	61679	61410	-	P/GM/FS/G	4	IMG-VR, Phobius, TMHMM, RVDB
BQ1_087	putative Mn2+ efflux pump MntP	XNN68190	62104	61712	-	P/GM/FS/G	6	Blastp-NR, IMG-VR, Phobius, TMHMM, RVDB, VOGBD
BQ1_088	hypothetical protein	XNN68191	62351	62118	-	P/GM/FS/G	4	Blastp-NR, Blastp-NRV, Phobius, IMG-VR
BQ1_089	hypothetical transmembrane protein	XNN68192	62804	62361	-	P/GM/FS/G	4	Blastp-NR, IMG-VR, Phobius, TMHMM
BQ1_090	transcription initiation factor IIB	XNN68193	62855	63874	+	P/GM/FS/G	5	IMG-VR, efam-xc, RVDB, VOGBD, Gene3D
BQ1_091	hypothetical protein	XNN68194	64236	63871	-	P/GM/G	1	IMG-VR

BQ1_092	E3 ubiquitin-protein ligase	XNN68195	64318	64644	+	P/GM/G	7	Blastp-NR, IMG-VR, Phobius, RVDB, VOGBD, SUPERFAMILY, Gene3D
BQ1_093	hypothetical transmembrane protein	XNN68196	65004	64648	-	P/GM/FS/G	4	IMG-VR, Phobius, TMHMM, efam-xc
BQ1_094	DUF5871 protein	XNN68197	65085	65804	+	P/GM/FS/G	5	Blastp-NR, Blastp-NRV, IMG-VR, RVDB, VOGBD
BQ1_095	hypothetical transmembrane protein	XNN68198	65900	66403	+	P/GM/FS/G	3	IMG-VR, Phobius, TMHMM
BQ1_096	hypothetical transmembrane protein	XNN68199	66427	66687	+	P/GM/FS/G	3	IMG-VR, Phobius, TMHMM
BQ1_097	hypothetical transmembrane protein	XNN68200	66754	67209	+	P/GM/FS/G	4	IMG-VR, Phobius, TMHMM, VOGBD
BQ1_098	deoxycytidylate deaminase	XNN68201	67554	67210	-	P/GM/FS/G	13	Blastp-NR, Blastp-NRV, IMG-VR, PHROG, efam-xc, RVDB, VOGBD, PANTHER, ProSitePatterns, Pfam, Gene3D, SUPERFAMILY, InterPro
BQ1_099	Ac78 protein	XNN68202	67798	67643	-	P/GM/FS/G	5	IMG-VR, Phobius, TMHMM, RVDB, VOGBD
BQ1_100	2OG-Fe(II) oxygenase	XNN68203	67910	68626	+	P/GM/FS/G	8	Blastp-NR, Blastp-NRV, IMG-VR, VOGBD, Pfam, Gene3D, PANTHER, InterPro
BQ1_101	hypothetical transmembrane protein	XNN68204	69527	68631	-	P/GM/FS/G	3	IMG-VR, Phobius, TMHMM
BQ1_102	hypothetical transmembrane protein	XNN68205	69762	69565	-	P/GM/FS/G	3	IMG-VR, Phobius, TMHMM
BQ1_103	transcription elongation factor S-II	XNN68206	69838	70341	+	P/GM/FS/G	13	Blastp-NR, Blastp-NRV, IMG-VR, RVDB, VOGBD, CDD, ProSitePatterns, Gene3D, Pfam, ProSiteProfiles, SMART, SUPERFAMILY, InterPro
BQ1_104	DNA topoisomerase I	XNN68207	70386	71225	+	P/GM/FS/G	10	Blastp-NR, Blastp-NRV, IMG-VR, PHROG, RVDB, VOGBD, Gene3D, SUPERFAMILY, Pfam, InterPro
BQ1_105	DUF5872 protein	XNN68208	71244	71759	+	P/GM/FS/G	9	Blastp-NR, Blastp-NRV, IMG-VR, PHROG, efam-xc, RVDB, VOGBD, Pfam, InterPro
BQ1_106	glycosyltransferase 17	XNN68209	71804	72418	+	P/GM/FS/G	9	Blastp-NR, Blastp-NRV, IMG-VR, efam-xc, RVDB, VOGBD, Pfam, PANTHER, InterPro
BQ1_107	ribonuclease III	XNN68210	72453	73172	+	P/GM/FS/G	17	Blastp-NR, Blastp-NRV, IMG-VR, PHROG, RVDB,

									VOGBD, NCBIfam, Gene3D, SMART, FunFam, SUPERFAMILY, CDD, ProSiteProfiles, PANTHER, Pfam, Hamap, InterPro Blastp-NR, Blastp-NRV, IMG-VR, PHROG, efam-xc, PVDB, VOGBD, PANTHER, Pfam, SUPERFAMILY, Gene3D, InterPro Blastp-NR, Blastp-NRV, IMG-VR, PHROG, efam-xc, RVDB, VOGBD, PANTHER, CDD, ProSiteProfiles, Pfam, NCBIfam, Hamap, SUPERFAMILY, Gene3D, InterPro Blastp-NR, Blastp-NRV, IMG-VR, PHROG, efam-xc, RVDB, VOGBD, Pfam, ProSiteProfiles, SUPERFAMILY, Gene3D, PANTHER, CDD, PRINTS, ProSitePatterns, SMART, InterPro Blastp-NR, IMG-VR, RVDB, VOGBD, Gene3D, SUPERFAMILY, Pfam, PANTHER, InterPro IMG-VR
BQ1_108	deoxyuridine 5'-triphosphate nucleotidohydrolase	XNN68211	73618	73169	-	P/GM/G	12		
BQ1_109	thymidylate synthase ThyX	XNN68212	73681	74451	+	P/GM/FS/G	16		
BQ1_110	DNA ligase, ATP-dependent	XNN68213	74487	76832	+	P/GM/FS/G	17		
BQ1_111	mRNA-capping enzyme	XNN68214	76891	78837	+	P/GM/G	9		
BQ1_112	hypothetical protein	XNN68215	81681	78814	-	P/GM/FS/G	1		
BQ1_113	hypothetical protein	XNN68216	81735	82217	+	P/GM/FS/G	5		
BQ1_114	minor capsid protein P9 transmembrane helices containing protein	XNN68217	82228	82785	+	P/GM/FS/G	7		
BQ1_115	hypothetical protein	XNN68218	82817	83092	+	P/GM/FS/G	1		
BQ1_116	N1R/p28-like protein	XNN68219	83104	83520	+	P/GM/FS/G	13		
BQ1_117	hypothetical extracellular protein	XNN68220	83578	84330	+	P/GM/FS/G	3		
BQ1_118	hypothetical protein	XNN68221	84353	84856	+	P/GM/FS/G	1		
BQ1_119	NUDIX hydrolase	XNN68222	85372	84851	-	P/GM/G	8		

BQ1_120	DUF5767 protein	XNN68223	85474	86505	+	P/GM/FS/G	7	SUPERFAMILY, Gene3D, InterPro Blastp-NR, Blastp-NRV, IMG-VR, RVDB, VOGBD, Pfam, InterPro
BQ1_121	hypothetical transmembrane protein	XNN68224	86549	86908	+	P/GM/G	3	IMG-VR, Phobius, TMHMM
BQ1_122	MYM-type Zinc finger with FCS sequence motif-containing protein	XNN68225	87611	86892	-	P/GM/FS/G	5	Blastp-NR, Blastp-NRV, IMG-VR, RVDB, VOGBD
BQ1_123	hypothetical protein	XNN68226	87656	88303	+	P/GM/FS/G	1	IMG-VR
BQ1_124	proliferating cell nuclear antigen, PCNA	XNN68227	88347	89114	+	P/GM/FS/G	15	Blastp-NR, Blastp-NRV, IMG-VR, RVDB, VOGBD, PANTHER, SUPERFAMILY, CDD, Pfam, NCBIfam, Gene3D, NCBIfam, PRINTS, Hamap, InterPro
BQ1_125	hypothetical transmembrane protein	XNN68228	89158	89631	+	P/GM/FS/G	3	IMG-VR, Phobius, TMHMM
BQ1_126	VV A18-like helicase	XNN68229	90943	89633	-	P/GM/FS/G	15	Blastp-NR, Blastp-NRV, IMG-VR, PHROG, efam-xc, RVDB, VOGBD, Pfam, SMART, CDD, Gene3D, PANTHER, ProSiteProfiles, SUPERFAMILY, InterPro
BQ1_127	hypothetical protein	XNN68230	91149	90976	-	P/GM/FS/G	1	IMG-VR
BQ1_128	NTPase/helicase	XNN68231	91438	92781	+	P/GM/FS/G	10	Blastp-NR, Blastp-NRV, IMG-VR, PHROG, RVDB, VOGBD, ProSiteProfiles, Gene3D, SUPERFAMILY, InterPro
BQ1_129	DUF1599 protein	XNN68232	93038	92778	-	P/GM/FS/G	8	Blastp-NR, Blastp-NRV, IMG-VR, PHROG, efam-xc, VOGBD, Pfam, InterPro
BQ1_130	hypothetical protein	XNN68233	93127	93843	+	P/GM/FS/G	2	IMG-VR, RVDB
BQ1_131	hypothetical transmembrane protein	XNN68234	93864	94028	+	GM/FS/G	3	IMG-VR, Phobius, TMHMM
BQ1_132	DUF5760 protein	XNN68235	94057	94470	+	P/GM/FS/G	6	Blastp-NR, IMG-VR, RVDB, VOGBD, Pfam, InterPro
BQ1_133	hypothetical protein	XNN68236	94889	94503	-	P/GM/FS/G	1	IMG-VR
BQ1_134	P-loop containing nucleoside triphosphate hydrolase protein	XNN68237	94988	96880	+	P/GM/FS/G	9	Blastp-NR, Blastp-NRV, IMG-VR, PHROG, RVDB, VOGBD, Gene3D, SUPERFAMILY, InterPro
BQ1_135	hypothetical transmembrane protein	XNN68238	96926	97255	+	P/GM/FS/G	6	IMG-VR, Phobius, TMHMM, PHROG, RVDB, VOGBD
BQ1_136	TATA-box binding protein	XNN68239	97984	97247	-	P/GM/G	10	Blastp-NR, Blastp-NRV, IMG-VR, RVDB, VOGBD,

BQ1_137	hypothetical protein	XNN68240	98692	98012	-	P/GM/FS/G	2	SUPERFAMILY, Gene3D, Pfam, PANTHER, InterPro
BQ1_138	late transcription factor VLTf3	XNN68241	98765	99838	+	P/GM/FS/G	7	IMG-VR, VOGBD Blastp-NR, Blastp-NRV, IMG-VR, RVDB, VOGBD, Pfam, InterPro
BQ1_139	potential major capsid protein A	XNN68242	101585	99924	-	P/GM/FS/G	5	Phyre2, Gene3D, IMG-VR, RVDB, Blastp-NR
BQ1_140	potential major capsid protein B	XNN68243	102955	101687	-	P/GM/FS/G	4	Phyre2, Gene3D, IMG-VR, Blastp-NR
BQ1_141	potential major capsid protein C	XNN68244	104372	103038	-	P/GM/FS/G	4	Phyre2, Gene3D, IMG-VR, Blastp-NR
BQ1_142	putative major capsid protein D	XNN68245	106012	104462	-	P/GM/FS/G	4	Phyre2, IMG-VR, RVDB, Blastp-NR
BQ1_143	hypothetical transmembrane protein	XNN68246	106093	106446	+	P/GM/G	3	IMG-VR, Phobius, TMHMM
BQ1_144	Putative DNA primase/helicase	XNN68247	107635	106412	-	P/GM/FS/G	5	Blastp-NR, Blastp-NRV, IMG-VR, RVDB, VOGBD
BQ1_145	hypothetical protein	XNN68248	108566	107703	-	P/GM/FS/G	2	Blastp-NR, IMG-VR
BQ1_146	IQ motif containing protein	XNN68249	110246	108597	-	P/GM/FS/G	4	Blastp-NR, IMG-VR, RVDB, ProSiteProfiles
BQ1_147	hypothetical protein	XNN68250	111317	110271	-	P/GM/FS/G	2	Blastp-NR, IMG-VR
BQ1_148	hypothetical protein	XNN68251	112046	111387	-	P/GM/FS/G	1	IMG-VR
BQ1_149	hypothetical protein	XNN68252	112625	112065	-	P/GM/FS/G	5	Blastp-NR, Blastp-NRV, IMG-VR, RVDB, VOGBD
BQ1_150	mRNA capping protein	XNN68253	112713	113603	+	P/GM/FS/G	5	Blastp-NR, Blastp-NRV, Gene3D, RVDB, VOGBD
BQ1_151	hypothetical extracellular protein	XNN68254	114049	113702	-	P/FS/G	2	IMG-VR, Phobius
BQ1_152	Flap endonuclease 1	XNN68255	115062	114115	-	P/GM/FS/G	12	Blastp-NR, Blastp-NRV, IMG-VR, RVDB, VOGBD, SUPERFAMILY, Pfam, Gene3D, SMART, PANTHER, PRINTS, InterPro
BQ1_153	YqaJ-like viral recombinase domain containing protein	XNN68256	116022	115118	-	P/GM/FS/G	14	Blastp-NR, Blastp-NRV, IMG-VR, PHROG, efam-xc, RVDB, VOGBD, SUPERFAMILY, Pfam, Gene3D, PANTHER, NCBIfam, CDD, InterPro
BQ1_154	Viral A32 protein	XNN68257	116887	116066	-	P/GM/FS/G	9	Blastp-NR, Blastp-NRV, IMG-VR, RVDB, VOGBD, Gene3D, SUEPRFAMILY, Pfam, InterPro
BQ1_155	type II secretory pathway component PulF	XNN68258	117159	116935	-	P/GM/FS/G	3	IMG-VR, Phobius, TMHMM
BQ1_156	Thioredoxin	XNN68259	117244	117561	+	P/GM/FS/G	12	Blastp-NR, Blastp-NRV, IMG-VR, PHROG, efam-xc, RVDB, VOGBD, CDD, Gene3D,

BQ1_157	DNA topoisomerase II	XNN68260	117625	120840	+	P/GM/FS/G	16	Pfam, SUPERFAMILY, InterPro Blastp-NR, Blastp- NRV, IMG-VR, PHROG, efam-xc, RVDB, FunFam, Pfam, PRINTS, SUPERFAMILY, SMART, Gene3D, ProSiteProfiles, PANTHER, CDD, InterPro
BQ1_158	hypothetical transmembrane protein	XNN68261	120880	121149	+	P/GM/G	3	IMG-VR, Phobius, TMHMM Blastp-NR, Blastp- NRV, IMG-VR, PHROG, RVDB, VOGBD, NCBIfam,
BQ1_159	exodeoxyribonuclease III	XNN68262	121988	121146	-	P/GM/FS/G	13	Pfam, SUPERFAMILY, ProSiteProfiles, Gene3D, PANTHER, InterPro
BQ1_160	hypothetical protein	XNN68263	122570	122052	-	P/GM/FS/G	2	Blastp-NR, IMG-VR Blastp-NR, Blastp- NRV, IMG-VR, RVDB, VOGBD, Gene3D, CDD, SUPERFAMILY, InterPro
BQ1_161	transcription initiation factor TFIIIB	XNN68264	122685	123626	+	P/GM/FS/G	9	Blastp-NR, IMG- VR, RVDB, VOGBD Blastp-NR, Blastp- NRV, IMG-VR, PHROG, RVDB, VOGBD, Gene3D, SMART, ProSitePatterns, Pfam, SUPERFAMILY, SMART, InterPro
BQ1_162	hypothetical protein	XNN68265	123652	126243	+	P/GM/FS/G	4	IMG-VR Blastp-NR, Blastp- NRV, IMG-VR, PHROG, RVDB, VOGBD, Gene3D, SMART, ProSitePatterns, Pfam, SUPERFAMILY, SMART, InterPro
BQ1_163	Helicase	XNN68266	126370	129303	+	P/GM/FS/G	13	IMG-VR Blastp-NR, Blastp- NRV, IMG-VR, efam-xc, VOGBD
BQ1_164	hypothetical protein	XNN68267	129348	130595	+	P/GM/FS/G	1	Blastp-NR, IMG-VR Blastp-NR, Blastp- NRV, IMG-VR, PHROG, RVDB, VOGBD, Gene3D, SMART, ProSitePatterns, Pfam, SUPERFAMILY, SMART, InterPro
BQ1_165	hypothetical protein	XNN68268	130994	130599	-	P/GM/FS/G	1	IMG-VR Blastp-NR, Blastp- NRV, IMG-VR, efam-xc, VOGBD
BQ1_166	glycosyltransferase	XNN68269	131902	131078	-	P/GM/FS/G	5	Blastp-NR, IMG-VR Blastp-NR, Blastp- NRV, IMG-VR, PHROG, RVDB, VOGBD, Gene3D, SMART, ProSitePatterns, Pfam, SUPERFAMILY, SMART, InterPro
BQ1_167	hypothetical protein	XNN68270	131986	132855	+	P/GM/FS/G	2	IMG-VR Blastp-NR, Blastp- NRV, IMG-VR, efam-xc, VOGBD
BQ1_168	Glutaredoxin-like protein	XNN68271	133267	132848	-	P/GM/G	6	Blastp-NR, IMG-VR Blastp-NR, Blastp- NRV, IMG-VR, PHROG, RVDB, VOGBD, Gene3D, SMART, ProSitePatterns, Pfam, SUPERFAMILY, SMART, InterPro
BQ1_169	hypothetical transmembrane protein	XNN68272	133333	133620	+	P/GM/FS/G	3	IMG-VR, Phobius, TMHMM Blastp-NR, Blastp- NRV, IMG-VR, efam-xc, VOGBD
BQ1_170	hypothetical transmembrane protein	XNN68273	133648	133965	+	P/GM/FS/G	3	IMG-VR, Phobius, TMHMM Blastp-NR, Blastp- NRV, IMG-VR, PHROG, RVDB, VOGBD, Gene3D, SMART, ProSitePatterns, Pfam, SUPERFAMILY, SMART, InterPro
BQ1_171	hypothetical protein	XNN68274	134022	134630	+	P/GM/G	2	IMG-VR Blastp-NR, Blastp- NRV, IMG-VR, PHROG, RVDB, VOGBD, Gene3D, SMART, ProSitePatterns, Pfam, SUPERFAMILY, SMART, InterPro
BQ1_172	patatin-like phospholipase family protein	XNN68275	135387	134623	-	P/GM/FS/G	11	IMG-VR Blastp-NR, Blastp- NRV, IMG-VR, PHROG, RVDB, VOGBD, Gene3D, SMART, ProSitePatterns, Pfam, SUPERFAMILY, SMART, InterPro

BQ1_173	WLM domain-containing protein	XNN68276	135893	135411	-	P/GM/FS/G	5	Blastp-NR, IMG-VR, RVDB, Phobius, VOGBD
BQ1_174	DNA polymerase type-B	XNN68277	135953	138937	+	P/GM/FS/G	15	Blastp-NR, Blastp-NRV, IMG-VR, PHROG, efam-xc, RVDB, VOGBD, Gene3D, PANTHER, ProSitePatterns, PRINTS, Pfam, SUPERFAMILY, SMART, InterPro
BQ1_175	hypothetical protein	XNN68278	139070	139663	+	P/GM/FS/G	5	Blastp-NR, Blastp-NRV, RVDB, IMG-VR, VOGBD
BQ1_176	hypothetical transmembrane protein	XNN68279	140097	139678	-	P/GM/FS/G	3	IMG-VR, Phobius, TMHMM
BQ1_177	tail fiber protein containing WIAG-tail domain	XNN68280	140154	145088	+	P/GM/FS/G	9	Blastp-NR, Blastp-NRV, IMG-VR, PHROG, efam-xc, RVDB, VOGBD, PANTHER, InterPro
BQ1_178	fibronectin type III domain containing tail fiber protein	XNN68281	145167	149783	+	P/GM/FS/G	13	Blastp-NR, Blastp-NRV, IMG-VR, PHROG, VOGBD, CDD, Gene3D, ProSiteProfiles, SUPERFAMILY, SMART, PANTHER, Pfam, InterPro
BQ1_179	fibronectin type III domain containing transmembrane protein	XNN68282	149818	154509	+	P/GM/FS/G	5	IMG-VR, PHROG, Phobius, CDD, SUPERFAMILY
BQ1_180	hypothetical transmembrane protein	XNN68283	154778	154614	-	P/GM/FS/G	3	IMG-VR, Phobius, TMHMM
BQ1_181	SAM-dependent methyltransferase	XNN68284	155423	154812	-	P/GM/FS/G	12	Blastp-NR, Blastp-NRV, IMG-VR, PHROG, efam-xc, RVDB, VOGBD, CDD, Gene3D, Pfam, SUPERFAMILY, InterPro
BQ1_182	hypothetical protein	XNN68285	155469	156491	-	P/GM/FS/G	3	Blastp-NR, IMG-VR, RVDB
BQ1_183	hypothetical protein	XNN68286	156528	157112	+	P/GM/FS/G	1	IMG-VR
BQ1_184	F-box domain-containing protein	XNN68287	158155	157388	-	P/GM/FS/G	9	Blastp-NR, IMG-VR, RVDB, VOGBD, Pfam, ProSiteProfiles, Gene3D, SUPERFAMILY, InterPro
BQ1_185	hypothetical extracellular protein	XNN68288	158515	158261	-	GM/G	3	IMG-VR, Phobius, SignalP
BQ1_186	hypothetical transmembrane protein	XNN68289	159582	158956	-	P/GM/FS/G	3	IMG-VR, Phobius, TMHMM
BQ1_187	F-box domain-containing protein	XNN68290	160285	159638	-	GM/FS/G	9	IMG-VR, RVDB, VOGBD, Gene3D, SMART, SUPERFAMILY, Pfam, ProSitePatterns, InterPro

BQ1_188	high mobility group protein	XNN68291	160585	160349	-	P/GM/FS/G	6	Blastp-NR, IMG-VR, RVDB, Gene3D, SUPERFAMILY, InterPro
BQ1_189	tape measure protein	XNN68292	161545	161105	-	P/GM/FS	2	IMG-VR, efam-xc
BQ1_190	RNA binding domain containing protein	XNN68293	161786	162472	+	GM/FS	5	Blastp-NR, Blastp-NRV, IMG-VR, RVDB, VOGBD
BQ1_191	hypothetical protein	XNN68294	163290	162529	-	P/FS	2	IMG-VR, VOGBD
BQ1_192	hypothetical protein	XNN68295	163412	163861	+	P/GM/FS	3	Blastp-NR, IMG-VR, RVDB
BQ1_193	hypothetical protein	XNN68296	165375	164752	-	P/GM/FS/G	4	Blastp-NR, IMG-VR, RVDB, VOGBD

Supplementary Table 4.4. *Genes used for Phylogenetic Analysis of CpV-BQ1.* The protein accession number and the viruses they are found in, used for the assembly of the polB, A32, and VLTF-3 phylogenetic trees.

Accession	Organism	Phylogenetic Tree
XNN68277	Chrysochromulina parva virus BQ1	PolB
AAB49739.1	Chrysochromulina brevifilum virus PW1	PolB
AAR05089.1	Phaeocystis globosa virus 08T	PolB
BAE06251.1	Heterosigma akashiwo virus 01	PolB
AAC00532.1	Paramecium bursaria Chlorella virus 1	PolB
YP_001427279.1	Chlorovirus heliozoae	PolB
CAY39796.1	Ostreococcus tauri virus 1	PolB
CBI70206.1	Ostreococcus tauri virus 2	PolB
AAB66713.1	Micromonas pusilla virus SP1	PolB
CAI65453.1	Emiliana huxleyi virus 86	PolB
QIG59463.1	Dishui Lake phycodnavirus 2	PolB
XDR79888.1	Micromonas commoda virus	PolB
AAB49746.1	Micromonas pusilla virus SG1	PolB
YP_009174732.1	Yellowstone lake phycodnavirus 1	PolB
AAA88827.1	Paramecium bursaria Chlorella virus NY2A	PolB
YP_004061614.1	Bathycoccus sp. RCC1105 virus BpV1	PolB
NP_077578.1	Ectocarpus siliculosus virus 1	PolB
AAB67116.1	Feldmannia species virus	PolB
YP_009665694.1	Feldmannia irregularis virus a	PolB
AGM15559.1	Phaeocystis globosa virus PgV-16T	PolB
AET73919.1	Phaeocystis globosa virus 14T	PolB
MH918795 region 273836 - 277723	Chrysochromulina parva virus BQ2	PolB
ALH23271.1	Chrysochromulina ericina virus CeV-01B	PolB

ARF11832.1	Klosneuvirus KNV1	PolB
ARF09278.1	Catovirus CTV1	PolB
ARF11096.1	Hokovirus HKV1	PolB
BAV62410.1	Acanthamoeba castellanii mimivirus	PolB
YP_010789826.1	Moumouvirus australiensis	PolB
YP_010788776.1	Megavirus baoshan	PolB
YP_010776439.1	Powai lake megavirus	PolB
YP_010781734.1	Tupanvirus soda lake	PolB
YP_010777236.1	Acanthamoeba polyphaga mimivirus	PolB
UOG94055.1	Aureococcus anophagefferens virus QBARTUTK	PolB
QDY51635.1	Mimiviridae sp. ChoanoV1	PolB
YP_009052217.1	Aureococcus anophagefferens virus BtV-01	PolB
ABU23717.1	Pyramimonas orientalis virus	PolB
YP_010783595.1	Tetraselmis virus 1	PolB
P06856.5	Vaccinia virus WR	PolB
XNN68257	Chrysochromulina parva virus BQ1	A32
YP_009507425.1	Heterosigma akashiwo virus 01	A32
NP_048749.2	Paramecium bursaria Chlorella virus 1	A32
YP_001426918.1	Chlorovirus heliozoae	A32
YP_003212915.1	Ostreococcus tauri virus 1	A32
YP_004063517.1	Ostreococcus tauri virus 2	A32
YP_009664892.1	Micromonas pusilla virus SP1	A32
YP_293826.1	Emiliana huxleyi virus 86	A32
QIG59506.1	Dishui Lake phycodnavirus 2	A32
XDR79698.1	Micromonas commoda virus	A32
AET43649.1	Micromonas pusilla virus PL1	A32
YP_009174839.1	Yellowstone lake phycodnavirus 1	A32
YP_001497732.1	Paramecium bursaria Chlorella virus NY2A	A32
YP_004061509.1	Bathycoccus sp. RCC1105 virus BpV1	A32
NP_077511.1	Ectocarpus siliculosus virus 1	A32
YP_002154709.1	Feldmannia species virus	A32
YP_009665688.1	Feldmannia irregularis virus a	A32
YP_008052384.1	Phaeocystis globosa virus 16T	A32
AET73733.1	Phaeocystis globosa virus 14T	A32
MH918795 region 94067 - 94909	Chrysochromulina parva virus BQ2	A32
ALH23031.1	Chrysochromulina ericina virus CeV-01B	A32
ARF09327.1	Catovirus CTV1	A32
ARF10841.1	Hokovirus HKV1	A32

BAV62527.1	Acanthamoeba castellanii mimivirus	A32
YP_010789698.1	Moumouvirus australiensis	A32
YP_010788906.1	Megavirus baoshan	A32
YP_010776318.1	Powai lake megavirus	A32
YP_010777350.1	Acanthamoeba polyphaga mimivirus	A32
YP_009052240.1	Aureococcus anophagefferens virus	A32
QDY52214.1	Mimiviridae sp. ChoanoV1	A32
QOI90457.1	Pyramimonas orientalis virus	A32
YP_010783224.1	Tetraselmis virus 1	A32
QTC35543.1	Vaccinia virus	A32
XNN68241	Chrysochromulina parva virus BQ1	VLTF-3
YP_009507423.1	Heterosigma akashiwo virus 01	VLTF-3
NP_048850.1	Paramecium bursaria Chlorella virus 1	VLTF-3
WZO77348.1	Chlorovirus sp.	VLTF-3
YP_003212932.1	Ostreococcus tauri virus 1	VLTF-3
YP_004063535.1	Ostreococcus tauri virus 2	VLTF-3
YP_009664910.1	Micromonas pusilla virus SP1	VLTF-3
YP_294161.1	Emiliana huxleyi virus 86	VLTF-3
QIG59621.1	Dishui Lake phycodnavirus 2	VLTF-3
XDR79675.1	Micromonas commoda virus	VLTF-3
YP_009174814.1	Yellowstone lake phycodnavirus 1	VLTF-3
YP_001497859.1	Paramecium bursaria Chlorella virus NY2A	VLTF-3
QOR60143.1	Bathycoccus sp. RCC716 virus 1	VLTF-3
NP_077586.1	Ectocarpus siliculosus virus 1	VLTF-3
YP_002154650.1	Feldmannia species virus	VLTF-3
YP_009665633.1	Feldmannia irregularis virus a	VLTF-3
YP_008052476.1	Phaeocystis globosa virus 16T	VLTF-3
AET73828.1	Phaeocystis globosa virus 14T	VLTF-3
MH918795 region 159579 – 160727	Chrysochromulina parva virus BQ2	VLTF-3
ALH23099.1	Chrysochromulina ericina virus CeV-01B	VLTF-3
ARF11172.1	Klosneuvirus KNV1	VLTF-3
ARF09332.1	Catovirus CTV1	VLTF-3
ARF10614.1	Hokovirus HKV1	VLTF-3
BAV61530.1	Acanthamoeba castellanii mimivirus	VLTF-3
YP_010789706.1	Moumouvirus australiensis	VLTF-3
YP_010788897.1	Megavirus baoshan	VLTF-3
YP_010776327.1	Powai lake megavirus	VLTF-3
YP_010782005.1	Tupanvirus soda lake	VLTF-3

AVG46193.1	Acanthamoeba polyphaga mimivirus	VLTF-3
YP_009052430.1	Aureococcus anophagefferens virus BtV-01	VLTF-3
QDY52221.1	Mimiviridae sp. ChoanoV1	VLTF-3
YP_010804359.1	Pyramimonas orientalis virus	VLTF-3
YP_010783292.1	Tetraselmis virus 1	VLTF-3
AHB35758.1	Vaccinia virus	VLTF-3
