Targeted enrichment and viral metagenomics in the detection of livestock and wildlife viruses

By

Amber Papineau

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Biological Sciences

University of Manitoba

Winnipeg

Abstract:

This thesis focuses on metagenomic viral detection based on high-throughput sequencing. Metagenomics sequencing results in a mixture of sequences, including viral, host and other non-target sequences. Metagenomic viral detection can lack sensitivity due to the low incidence of viral sequences in these mixtures. For this reason, enrichment methods are often employed in metagenomic viral detection. Targeted enrichment is a method based on the hybridization of probes designed to enrich the desired target from a mixture. ViroCap is a method of targeted enrichment designed for vertebrate viruses. It has previously been validated to aid in the detection of viruses from human samples.

The second chapter of this thesis describes the evaluation of ViroCap in 10 wildlife and livestock hosts. Enrichment was calculated on a blinded panel of 24 animal and zoonotic viral species. The average enrichment of percent-viral-reads across the 22 viral families tested was 182-fold. Depth of coverage increased 123 times and breadth of coverage increased by 43%. In addition, 6 unexpected viruses were detected. Chapter 3 of this thesis describes the results of the use of viral metagenomics and ViroCap enrichment in a diagnostic disease investigation. ViroCap enrichment and metagenomic high-throughput sequencing revealed the presence of 4 viruses in a mass die-off of Canada and Snow geese, including avian metapneumovirus and avian adeno-associated virus. In addition, the entire genome sequence of a novel species of *Gammacoronavirus* is described. The first sequence information of goose adenovirus in Canada is also described. Chapter 4 describes the fecal virome of arctic and red foxes. 6 viral families were detected with the use of ViroCap targeted enrichment. A diversity of divergent circoviruses and parvoviruses are described. In addition, numerous avian influenza and canine kobuvirus subtypes were also detected in the feces of several arctic foxes. In conjunction with the description of wildlife viromes, this thesis establishes the utility of targeted enrichment in the detection of livestock and wildlife viruses.

Table of contents:

List of Figures:

stem loop-like motif and octamer motif are contained within the 3' UTR. Genome organization Figure was constructed using Geneious™ (Biomatters, v 9.1.8). * indicate ACoV 4b homologues. Proteins are named numerically from the 5' end of the genome, with the exception of the structural genes, which are denoted by their common names.

Figure 7: The phylogeny of gammacoronavirus spike and nucleocapsid proteins. A maximum likelihood tree built, using the amino acid sequences of the spike protein (A) and nucleocapsid protein (B) domains aligned with ClustalW (Larkin et al. 2007), in MEGA X using the Jones-Taylor-Thornton (JTT) substitution model and 1000 bootstraps (Kumar et al. 2018). IBV Infectious Bronchitis virus, TCoV Turkey Coronavavirus, PCoV Pigeon Coronavirus, DCoV Duck Coronavirus.

Figure 8: The phylogeny of Goose Coronavirus CB17. A maximum likelihood tree built, using the amino acid sequences of the replicase and helicase protein domains aligned with ClustalW (Larkin et al. 2007), in MEGA X using the Jones-Taylor-Thornton (JTT) substitution model and 1000 bootstraps (Kumar et al. 2018). Numbers at nodes indicate the bootstrap value.

Figure 9: Maximum likelihood tree constructed using the nucleotide sequence of the glycoproteins of avian metapneumovirus. The tree was constructed using IQ-TREE (Nguyen et al. 2015), with the substitution model TIM2+F+I+G4 and 1000 ultra-fast bootstraps.

Figure 10: Maximum likelihood tree constructed using the nucleotide sequence of the capsid proteins of dependoparvovirus genus. The tree was constructed using IQ-TREE (Nguyen et al. 2015), with the substitution model TIM3e+G4 and 1000 ultra fast bootstraps.

Figure 11: Avian influenza hemagglutinin type distribution in 4 arctic fox fecal samples, estimated with IRMA (Shepard et al. 2016).

List of Tables:

Acknowledgements:

This project was funded by the Canadian Food Inspection Agency (CFIA) project WIN-A-1408 and Canadian Safety and Security Program project TI-2222 for this project. I would like to thank Dr. Samuel Sharpe for his work on the geese die-off. The arctic fox samples used in this thesis were provided by Dr. James Roth and Chloe Rodriguez, thank you for your hard work. Thank you to Dr. Ed Taboada and Dr. Steve Whyard for their guidance and advice.

I would like to thank the staff of the National Centre for Foreign Animal Disease for their assistance and guidance throughout this project. Thank you to Cody Buchanan, Katherine Handel, Mathew Fisher, Michelle Nebroski and Peter Kruczkiewicz for all your guidance and help, but especially for your friendship. I would like to thank Dr. Oliver Lung for his leadership and for giving me the opportunity to work on this project. I've been very lucky to have so many wonderful scientists as mentors. I loved working with all of you.

I'm very been very lucky to work on the things that fascinate me most in life, so thank you to all the researchers that came before me and made this work possible.

A specail thank you to all my friends and family, I wouldn't have been able to do this without all the cheerleaders in my life.

Dedication:

I would like to dedicate this thesis to my parents and my sisters. You've supported me through this whole process and beyond. Thank you for giving me the space and encouragement to be my nerdy self. Thank you for everything you've done for me.

Chapter 1: Literature review

Here a virus, there a virus, everywhere a virus

Viruses are infectious agents which consist of a RNA or DNA genome encased by a protein shell (or capsid). They are cellular parasites, meaning they are dependant on a host cell to replicate. This fact has led to a long debate over whether they are alive or inanimate (Koonin & Starokadomskyy 2016). While many believe they do not constitute a form of life (Moreira & López-García 2009), more recently it has been proposed life can be classified into ribosome containing (the 3 current recognized domains of life) and non-ribosome containing (viruses) (Raoult & Forterre 2008). Regardless of their classification, they are ubiquitous. They are the most abundant biological entity on earth. Estimates of viral abundance in environmental samples boggle the mind; 10^10 in 1 litre of seawater (Bergh et al. 1989), and 10^9 in a g of soil (Williamson et al. 2005).

With this abundance in number, comes a huge diversity in shape (Rossmann 2013), size (Chapman & Rossmann 1993, Colson et al. 2017), genome size (Arslan et al. 2011) (Ellis 2014) and host specificity (Bekliz et al. 2016; Cauldwell et al. 2014). Estimate are that there is a minimum of 320,000 viral species in mammalian hosts, most of which have yet to be described (Anthony et al. 2013). Many groups have taken to the task of identifying this undescribed viral diversity (Radford et al. 2012). This age of viral discovery, enabled by new technologies, has led to a huge strain on viral taxonomy systems (Simmonds 2018).

Viral taxonomy is a difficult business, to begin with. This is due to the lack of a shared common gene both amongst all viral families and with other forms of life. For this reason, they are excluded from the universal tree of life (Brüssow 2009). Attempts are still made to impose order on the chaos of the viral world. In 1971, David Baltimore proposed 7 groups to

classify viruses based on their genome structure; I Double-stranded DNA viruses, II Single-stranded DNA viruses, III Double-stranded RNA viruses, IV Single stranded positive-sense RNA viruses, V Single stranded negative-sense RNA viruses, VI Positive-sense single-stranded RNA viruses that replicate through a DNA intermediate and VII: dsDNA with an ssRNA replication intermediate (Baltimore 1971). While this classification helps to understand the variety of viral genomic structures, it fails to accurately describe viral evolutionary history. For example, double-stranded DNA viruses are believed to be polyphyletic in nature (Koonin et al. 2015). The international committee on viral taxonomy (ICTV) maintains a finer classification for viruses. In 1971 the ICTV released their first report, listing 290 viral species organized in 2 families. The most recent report lists 14 orders, 150 families, 79 subfamilies, 1021 genus and 5560 species (Lefkowitz et al. 2018).

The World Organisation for Animal Health (OIE) lists diseases caused by 50 of these viral species as reporTable. The OIE is an intergovernmental organization that coordinates international efforts to control and prevent the spread of animal diseases (Brückner 2009). The control of animal diseases is important for our economic well being (Knight-Jones & Rushton 2013). It is also very important for our health, as many of the emerging diseases in humans originate in domestic animals (Zhou et al. 2017). In turn, many viruses that impact livestock originate in wildlife (Miller et al. 2013). This has led to an increased interest in describing the virome (entire viral content) of known wildlife and domestic vectors of disease (Wu et al. 2016). As most viral detection methods are targeted, traditional viral detection methods don't provide the necessary diversity to analyse the breadth of possible viral targets in virome analysis.

Overview of viral detection methods

Viral detection began with the discovery of a non-bacterial disease-causing agent. Dmitri Ivanovsky discovered that sap from a tobacco plant which had been filtered of bacteria remained infectious. (Ivanovsky 1892.) Martinus Beijerinck called this new infectious agent a virus and showed it was dependent on a host to replicate (Beijerinck *1898*). Soon after viruses were found to be causative agents of disease in animals. Foot-and-mouth was the first animal virus discovered (Loeffler et al. 1898). Viruses were then found to be causative agents of human disease as well, beginning with yellow fever (Reed et al. 2001). It's since become apparent viruses infect all domains of life (d'Herelle 1917) (Prangishvili et al. 2006). Electron microscopy gave shape to these invisible pathogens and therefore a new system of classification (Roingeard 2008). This lead to the use of electron microscopy in the diagnosis of viral disease (Nagler & Rake 1948). Transmission electron microscopy (TEM) became important not only in the discovery of viruses but also in their classification, as some of the first viral classification systems were based on morphology (Roingeard 2008). TEM was then replaced with antigenic methods (Bryan 1987) and molecular-based testing (Jackson 1990) in large part due to their superior sensitivity and rapid turnaround time.

One of these molecular methods is Sanger sequencing, a method that would become and still is important in viral diagnostic programs. Frederick Sanger invented Sanger sequencing in 1977. Sanger sequencing is often referred to as sequencing by termination, as it employs di-deoxynucleotides which halt DNA synthesis. Each of the four dideoxynucleotides is labelled with different fluorescent dyes, allowing nucleotides at the terminal position to be identified when visualized on an acrylamide gel (Sanger et al. 1977). Sequencing has allowed more in-depth characterization of viruses. For example, genotyping HIV to detect drug-resistant mutants (Hirsch et al. 2000).

The Age of Viral Metagenomics

One trait these traditional tests have in common is the need for prior knowledge of a virus in order to detect it. PCR and Sanger sequencing require primers homologous to the target virus's nucleic acid (Rodríguez et al. 2015) and serological assays require high-quality antibodies (Tabll et al. 2015). High-throughput sequencing (HTS) doesn't require *a priori* knowledge of a pathogen in order to study it; this enables a broad-range approach which makes it particularly well suited for viral detection and characterization. HTS is a group of technologies that have succeeded first generation sequencing (Sanger) due to their higher throughput. The high throughput of HTS is achieved through millions of sequencing reactions occurring in parallel (Buermans & den Dunnen 2014). The high throughput of HTS makes it unnecessary to target a single microbe in a test, allowing a metagenomic approach to virus detection and characterization. Metagenomics is a method of study in which the complete nucleic acid content of a sample is analyzed. Metagenomic analysis is often used in the study of environmental microbial populations (Zeigler Allen et al. 2017), but also in virus discovery in vertebrates (Palinski et al. 2016) and insects (Ergünay et al. 2017).

In the case of viral detection and characterization, a metagenomic approach, facilitated by HTS, provides an enhanced ability to detect unexpected, divergent or novel pathogens. The recent discovery of Schmallenberg virus in European cattle is an excellent example of metagenomic HTS being used to detect and characterize novel viruses in livestock. Using a 454 sequencer, Hoffman et al. (2009) were able to identify *Orthobunyavirus* in sick cattle from just seven sequencing reads. Further sequencing exposed a novel virus, subsequently, named Schmallenberg virus. An RT-PCR assay was designed using the sequence data acquired from HTS, which was then used to track the outbreak and study infected cattle. This case demonstrates that metagenomic HTS also provides the advantage of more specific outbreak management and response (Hoffmann et al. 2012).

As mentioned above, HTS technology is able to overcome some of the limitations of traditional diagnostic tools. This is apparent in cases like the Uganda Ebola virus outbreak in 2008. Towner et al. (2008) used HTS to identify Ebola virus as the cause of disease in patients who had been previously diagnosed as negative for Ebola virus by PCR. This study demonstrates that very sensitive diagnostic tests, such as RT-PCR, can fail if highly divergent pathogens arise (Towner et al. 2008). This fact is echoed by other cases, such as a study by Yozwiak et al. (2012). Using HTS, Yozwiak et al. investigated febrile illness of unknown etiology in children. HTS revealed Dengue virus as the cause (Yozwiak et al. 2012).

Another advantage HTS has over traditional laboratory testing is its ability to detect multiple pathogens in a single sample. This allows HTS to correctly identify co-infections that might have been missed by tests that are targeted to a single pathogen, or a set of pathogens. For instance, HTS has been used to diagnose co-infections of Zika virus and Chikungunya virus (Sardi et al. 2016). This advantage is also particularly apparent in the study of multifactorial diseases such as postweaning multisystemic wasting syndrome, a disease that results from co-infection with multiple viruses (Blomström et al. 2016).

As outlined above, viral detection and characterization using HTS has many advantages. However, it isn't without its challenges. A major challenge to viral detection by HTS is the typically low abundance of viral nucleic acid in comparison to other nucleic acids, such as the host's genome. This can lead to poor sequencing depth, poor breadth of coverage and ultimately failure to detect viruses (Daly et al. 2011). Also, it results in inefficient use of resources, as the majority of sequencing is done on non-target host material and other contaminating nucleic acid. This challenge has led to a variety of sample preparation techniques intended to enrich viral nucleic acid from a heterogeneous sample.

When a known virus is being sequenced, its genome can be amplified using PCR before sequence library preparation (Baillie et al. 2012). However, this technique can fail if an isolate is significantly divergent (Towner et al. 2008) and it cannot be applied to an unknown or unexpected pathogen. For viral metagenomic sequencing, there are a variety of other sample enrichment methods. Generally, metagenomic viral HTS involves a pre-treatment of the sample for viral nucleic acid enrichment, followed by cDNA synthesis, library preparation and sequencing. A review of the literature will show that low-speed centrifugation, syringe based filtration and nuclease pre-treatments are some of the most commonly used enrichment methods (Montmayeur et al. 2017; Zou et al. 2017; Theuns et al. 2016; Luk et al. 2015). Nuclease treatments are thought to work to increase the proportion of viral nucleic acid relative to host because viral nucleic acid is protected from nucleases by the viral capsid, while host nucleic acid is sensitive to degradation by nucleases. In some cases, DNase treatments occur post extraction of nucleic acid, so that viruses with RNA genomes are enriched for sequencing. Centrifugation and filtration both work to separate virus and host material through physical means due to differences in the density and size of viral particles and host material. Filtration allows small viral particles to pass through a filter while host material, bacteria and fungi cannot. Similarly, low-speed centrifugation pellets host material from solution while viral particles remain in the supernatant (Hall et al. 2014).

Nuclease treatments can consist of DNase alone (Logan et al. 2014) or a combination of DNase and RNase. Rossel et al. (2015) found that DNase/RNase treatments prior to nucleic acid extraction did not always improve viral sequencing. Specifically, it worked well in serum but not tissue samples. This is likely due to the higher number of host cells in tissue relative to serum. They did find that nuclease treatments in combination with centrifugation works well to increase the proportion of viral reads in both serum and tissue (Rosseel et al. 2015). Other publications have found combinations of these pre-treatments to be effective in viral nucleic

acid enrichment. For example, a similar study by Hall et al. (2014) found that a 3-step method consisting of centrifugation, filtration and nuclease pre-treatments worked best to increase the percentage of viral reads. The authors were able to show a 10 fold increase in influenza viral reads (from 0.001% to 0.01%) and a 20 fold increase for enterovirus reads (from 0.16% to 4.74%) in artificial samples(Hall et al. 2014). Ribosomal RNA (rRNA) removal has also been used to enhance viral metagenomic HTS. Rossel et al. (2015) found that rRNA removal, performed with magnetic beads, improved the percentage of viral reads in tissue but not serum.

Technological advancements have enabled the development of a more targeted enrichment method known as sequence capture. Sequence capture is a technique in which a group of desired nucleic acids are removed from a larger population. This is achieved through hybridization of targeted DNA to homologous capture probes. The probes are biotinylated in order to bind streptavidin-coated magnetic beads, allowing unbound DNA to be washed away. This technique is being used to enrich for genes of interest in areas of human health research, such as cancer research (Hagemann et al. 2013) and mutation studies (Wang et al. 2016). Wylie et al. (2015) developed a sequence capture panel called ViroCap. ViroCap is designed to capture genetic material from all known viral species that infect vertebrates, with the aim of viral nucleic acid enrichment prior to HTS. Wylie et al. (2015) designed ViroCap using sequence data from 337 viral species from 190 genera, making up almost a billion base pairs of sequence data. The resulting 100 bp probes were screened against the human genome and any probes that shared over 75% sequence identity with the human genome were removed. Using Virocap on human clinical samples they detected 32 viruses post-capture, as opposed to 21 viruses pre-capture. They also note an increase in genomic coverage (2.0% to 83.2%). A median fold increase of the percentage of viral reads of 674 and 296, in comparison to 10 and 20 fold increases of the percentage of viral reads using centrifugation, filtration and DNase

reported by Hall et al. (2014). They also show that Virocap can enrich highly divergent viruses, such as *Anelloviruses* (Wylie et al. 2015). Briese et al. (2015) created a targeted sequence capture panel similar to Wylie et al's. They use 2 million probes designed from vertebrate virus sequence. They report fold increases of viral reads of 100 to 1000. Briese et al. demonstrate that sequence capture performed in this manner can capture partial viral genomes of novel viruses (Briese et al. 2015).

Another concern to consider in viral metagenomic HTS is the synthesis of cDNA. RNA viruses make up a significant proportion of viral pathogens. Bacterial metagenomic studies can take advantage of conserved sequences that span all bacterial taxons, such as 16S ribosomal RNA genes (Yergeau et al. 2017). Viral metagenomic approaches do not have this benefit. Often random priming is the strategy adopted for RNA viral cDNA synthesis (van Gurp et al. 2013). Sequence-independent, single-primer amplification (SISPA) is a method used for cDNA synthesis based on random priming. SISPA uses a random primer tagged with a conserved sequence for cDNA synthesis; the conserved sequence can then be used for subsequent amplification to achieve the required input for HTS library construction(Reyes & Kim 1991). Rosseel et al. (2012) used SISPA in combination with a DNase enrichment to detect Schmallenberg virus in sheep and cattle.

Objectives

The main goal of this thesis research project is to evaluate and validate new methods for the broad range detection and characterization of unknown and unexpected viruses from animals. The primary focus of this project is the application of broad-range sequence capture to enrich viral nucleic acid for detection and characterization by HTS. Metagenomic enrichment methods, such as the 3-step method proposed by Hall et al. (2014), are very non-specific and therefore produce limited viral enrichment and detection. Species-specific

methods, such as the conserved cDNA synthesis method developed by Logan et al.'s (2015), are difficult to apply when the identity of the virus is unknown. Here I have evaluated and validate broad range sequence capture for use in veterinary diagnostics. This goal was achieved through 3 research objectives:

1. Evaluate ViroCap targeted enrichment in the use of veterinary diagnostics, by testing a blinded panel of animal viruses consisting of multiple hosts and sample types.

2. Demonstrate the utility of ViroCap targeted enrichment in veterinary diagnostic disease investigations with a case study of a large die-off of Canada and Snow Geese.

3. Demonstrate the utility of ViroCap targeted enrichment in the exploration of wildlife viromes.

References

Anthony SJ, Epstein JH, Murray KA, Navarrete-Macias I, Zambrana-Torrelio CM, et al. 2013. A strategy to estimate unknown viral diversity in mammals. *MBio*. 4(5):e00598-13

Arslan D, Legendre M, Seltzer V, Abergel C, Claverie J-M. 2011. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. *Proc Natl Acad Sci USA*. 108(42):17486–91

Baillie GJ, Galiano M, Agapow P-M, Myers R, Chiam R, et al. 2012. Evolutionary dynamics of local pandemic H1N1/2009 influenza virus lineages revealed by whole-genome analysis. *J. Virol.* 86(1):11–18

Baltimore D. 1971. Expression of animal virus genomes. *Bacteriol. Rev.* 35(3):235–41

Beijerinck MW (1898) Über ein *contagium vivum fluidum* als Ursache der Fleckenkrankheit der Tabaksblätter. Verh Kon Akad Wetensch 65 (Sectie 2, Deel 6): 3–21. (Eng. translation: Concerning a contagium vivum fluidum as cause of the spot disease of tobacco leaves. Phytopath Classics 7 (1942): 33–54)

Bekliz M, Colson P, La Scola B. 2016. The expanding family of virophages. *Viruses*. 8(11):

Bergh O, Børsheim KY, Bratbak G, Heldal M. 1989. High abundance of viruses found in aquatic environments. *Nature*. 340(6233):467–68

Blomström A-L, Fossum C, Wallgren P, Berg M. 2016. Viral Metagenomic Analysis Displays the Co-Infection Situation in Healthy and PMWS Affected Pigs. *PLoS ONE*. 11(12):e0166863

Briese T, Kapoor A, Mishra N, Jain K, Kumar A, et al. 2015. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *MBio*. 6(5):e01491-15

Brückner GK. 2009. The role of the World Organisation for Animal Health (OIE) to facilitate the international trade in animals and animal products. *Onderstepoort J. Vet. Res.* 76(1):141–46

Brüssow H. 2009. The not so universal tree of life or the place of viruses in the living world. *Philos Trans R Soc Lond, B, Biol Sci*. 364(1527):2263–74

Bryan JA. 1987. The serologic diagnosis of viral infection. An update. *Arch. Pathol. Lab. Med.* 111(11):1015–23

Buermans HPJ, den Dunnen JT. 2014. Next generation sequencing technology: Advances and applications. *Biochim. Biophys. Acta*. 1842(10):1932–41

Cauldwell AV, Long JS, Moncorgé O, Barclay WS. 2014. Viral determinants of influenza A virus host range. *J. Gen. Virol.* 95(Pt 6):1193–1210

Chapman MS, Rossmann MG. 1993. Structure, sequence, and function correlations among parvoviruses. *Virology*. 194(2):491–508

Colson P, La Scola B, Raoult D. 2017. Giant viruses of amoebae: A journey through innovative research and paradigm changes. *Annu. Rev. Virol.* 4(1):61–85

d'Herelle, F., 1917. Sur un microbe invisible antagoniste des bacilles dysentériques. *CR Acad. Sci. Paris*, *165*, pp.373-375.

Daly GM, Bexfield N, Heaney J, Stubbs S, Mayer AP, et al. 2011. A viral discovery methodology for clinical biopsy samples utilising massively parallel next generation sequencing. *PLoS ONE*. 6(12):e28879

Ellis J. 2014. Porcine circovirus: a historical perspective. *Vet. Pathol.* 51(2):315–27

Ergünay K, Brinkmann A, Litzba N, Günay F, Kar S, et al. 2017. A novel rhabdovirus, related to Merida virus, in field-collected mosquitoes from Anatolia and Thrace. *Arch. Virol.* 162(7):1903–11

Hagemann IS, Cottrell CE, Lockwood CM. 2013. Design of targeted, capture-based, next generation sequencing tests for precision cancer therapy. *Cancer Genet.* 206(12):420–31

Hall RJ, Wang J, Todd AK, Bissielo AB, Yen S, et al. 2014. Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. *J. Virol. Methods*. 195:194–204

Hirsch MS, Brun-Vézinet F, D'Aquila RT, Hammer SM, Johnson VA, et al. 2000. Antiretroviral drug resistance testing in adult HIV-1 infection: recommendations of an

International AIDS Society-USA Panel. *JAMA*. 283(18):2417–26

Hoffmann B, Scheuch M, Höper D, Jungblut R, Holsteg M, et al. 2012. Novel orthobunyavirus in Cattle, Europe, 2011. *Emerging Infect. Dis.* 18(3):469–72

Ivanovsky D (1892) Über die Mosaikkrankheit der Tabakspflanze. Bull Acad Imper Sci St. Petersburg 35: 67–70; id. (1893) Beih Bot Centralbl 3: 266–268; Eng translation: (1942) Concerning the mosaic disease of tobacco. Phytopath Classics 7: 25–30

Jackson JB. 1990. The polymerase chain reaction in transfusion medicine. *Transfusion*. 30(1):51–57

Kleiner M, Hooper LV, Duerkop BA. 2015. Evaluation of methods to purify virus-like particles for metagenomic sequencing of intestinal viromes. *BMC Genomics*. 16:7

Knight-Jones TJD, Rushton J. 2013. The economic impacts of foot and mouth disease - what are they, how big are they and where do they occur? *Prev. Vet. Med.* 112(3–4):161–73

Koonin EV, Dolja VV, Krupovic M. 2015. Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology*. 479–480:2–25

Koonin EV, Starokadomskyy P. 2016. Are viruses alive? The replicator paradigm sheds decisive light on an old but misguided question. *Stud. Hist. Philos. Biol. Biomed. Sci.* 59:125–34

Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, Siddell SG, Smith DB. 2018. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res.* 46(D1):D708–17

Loeffler, F. and Frosch, P., 1898. Report of the commission for research on foot-and-mouth disease. *Zentrabl. Bacteriol. Parastenkunde Infektionkrankh*, *23*, pp.371-391

Logan G, Freimanis GL, King DJ, Valdazo-González B, Bachanek-Bankowska K, et al. 2014. A universal protocol to generate consensus level genome sequences for foot-and-mouth disease virus and other positive-sense polyadenylated RNA viruses using the Illumina MiSeq. *BMC Genomics*. 15:828

Luk K-C, Berg MG, Naccache SN, Kabre B, Federman S, et al. 2015. Utility of Metagenomic Next-Generation Sequencing for Characterization of HIV and Human Pegivirus Diversity. *PLoS ONE*. 10(11):e0141723

Miller RS, Farnsworth ML, Malmberg JL. 2013. Diseases at the livestock-wildlife interface: status, challenges, and opportunities in the United States. *Prev. Vet. Med.* 110(2):119–32

Montmayeur AM, Ng TFF, Schmidt A, Zhao K, Magaña L, et al. 2017. High-Throughput Next-Generation Sequencing of Polioviruses. *J. Clin. Microbiol.* 55(2):606–15

Moreira D, López-García P. 2009. Ten reasons to exclude viruses from the tree of life. *Nat. Rev. Microbiol.* 7(4):306–11

Nagler FP, Rake G. 1948. The use of the electron microscope in diagnosis of variola, vaccinia, and varicella. *J. Bacteriol.* 55(1):45–51

Palinski RM, Mitra N, Hause BM. 2016. Discovery of a novel Parvovirinae virus, porcine parvovirus 7, by metagenomic sequencing of porcine rectal swabs. *Virus Genes*. 52(4):564–67

Prangishvili D, Garrett RA, Koonin EV. 2006. Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. *Virus Res.* 117(1):52–67

Radford AD, Chapman D, Dixon L, Chantrey J, Darby AC, Hall N. 2012. Application of next-generation sequencing technologies in virology. *J. Gen. Virol.* 93(Pt 9):1853–68

Raoult D, Forterre P. 2008. Redefining viruses: lessons from Mimivirus. *Nat. Rev. Microbiol.* 6(4):315–19

Reed W, Carroll J, Agramonte A. 2001. The etiology of yellow fever: an additional note . 1901. *Mil. Med.* 166(9 Suppl):44–53

Reyes GR, Kim JP. 1991. Sequence-independent, single-primer amplification (SISPA) of complex DNA populations. *Mol. Cell. Probes*. 5(6):473–81

Rodríguez A, Rodríguez M, Córdoba JJ, Andrade MJ. 2015. Design of primers and probes for quantitative real-time PCR methods. *Methods Mol. Biol.* 1275:31–56

Roingeard P. 2008. Viral detection by electron microscopy: past, present and future. *Biol. Cell*. 100(8):491–501

Rosseel T, Ozhelvaci O, Freimanis G, Van Borm S. 2015. Evaluation of convenient pretreatment protocols for RNA virus metagenomics in serum and tissue samples. *J. Virol. Methods*. 222:72–80

Rossmann MG. 2013. Structure of viruses: a short history. *Q. Rev. Biophys.* 46(2):133–80

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA*. 74(12):5463–67

Sardi SI, Somasekar S, Naccache SN, Bandeira AC, Tauro LB, et al. 2016. Coinfections of Zika and Chikungunya Viruses in Bahia, Brazil, Identified by Metagenomic Next-Generation Sequencing. *J. Clin. Microbiol.* 54(9):2348–53

Simmonds P. 2018. A clash of ideas - the varying uses of the “species” term in virology and their utility for classifying viruses in metagenomic datasets. *J. Gen. Virol.*

Tabll A, Abbas AT, El-Kafrawy S, Wahid A. 2015. Monoclonal antibodies: Principles and

applications of immmunodiagnosis and immunotherapy for hepatitis C virus. *World J. Hepatol.* 7(22):2369–83

Theuns S, Conceição-Neto N, Zeller M, Heylen E, Roukaerts IDM, et al. 2016. Characterization of a genetically heterogeneous porcine rotavirus C, and other viruses present in the fecal virome of a non-diarrheic Belgian piglet. *Infect. Genet. Evol.* 43:135–45

Towner JS, Sealy TK, Khristova ML, Albariño CG, Conlan S, et al. 2008. Newly discovered ebola virus associated with hemorrhagic fever outbreak in Uganda. *PLoS Pathog.* 4(11):e1000212

van Gurp TP, McIntyre LM, Verhoeven KJF. 2013. Consistent errors in first strand cDNA due to random hexamer mispriming. *PLoS ONE*. 8(12):e85583

Wang J, Yu H, Zhang VW, Tian X, Feng Y, et al. 2016. Capture-based high-coverage HTS: a powerful tool to uncover a wide spectrum of mutation types. *Genet. Med.* 18(5):513–21

Williamson KE, Radosevich M, Wommack KE. 2005. Abundance and diversity of viruses in six Delaware soils. *Appl. Environ. Microbiol.* 71(6):3119–25

Wu Z, Yang L, Ren X, He G, Zhang J, et al. 2016. Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases. *ISME J.* 10(3):609–20

Wylie TN, Wylie KM, Herter BN, Storch GA. 2015. Enhanced virome sequencing using targeted sequence capture. *Genome Res.* 25(12):1910–20

Yergeau E, Michel C, Tremblay J, Niemi A, King TL, et al. 2017. Metagenomic survey of the taxonomic and functional microbial communities of seawater and sea ice from the

Canadian Arctic. *Sci. Rep.* 7:42242

Yozwiak NL, Skewes-Cox P, Stenglein MD, Balmaseda A, Harris E, DeRisi JL. 2012. Virus identification in unknown tropical febrile illness cases using deep sequencing. *PLoS Negl. Trop. Dis.* 6(2):e1485

Zeigler Allen L, McCrow JP, Ininbergs K, Dupont CL, Badger JH, et al. 2017. The baltic sea virome: diversity and transcriptional activity of DNA and RNA viruses. *mSystems*. 2(1):

Zhou L, Ren R, Yang L, Bao C, Wu J, et al. 2017. Sudden increase in human infection with avian influenza A(H7N9) virus in China, September-December 2016. *Western Pac. Surveill. Response J.* 8(1):6–14

Zou X, Tang G, Zhao X, Huang Y, Chen T, et al. 2017. Simultaneous virus identification and characterization of severe unexplained pneumonia cases using a metagenomics sequencing technique. *Sci. China Life Sci.* 60(3):279–86

Chapter 2: Evaluation of a Broad Range Targeted Enrichment Method for the Detection of Wildlife and Livestock Viruses

Foreword

The mandate of the Canadian Food Inspection Agency's National Centre for Foreign Animal Disease (NCFAD) includes diagnostics, surveillance and response to multiple foreign animal diseases (FAD), including foot-and-mouth disease virus (FMDV) and classical swine fever virus (CSFV). The control of FADs is consequential for the health (Zhou et al. 2017) and economic wellbeing of the public (Knight-Jones & Rushton 2013). The Genomics Unit of the NCFAD provides diagnostic support and is actively researching new diagnostic methods for the detection and characterization of FADs. The primary purpose of this thesis project is to evaluate high-throughput sequencing (HTS) based methods to improve the identification and characterization of unknown and unexpected viral pathogens at the NCFAD.

Abstract

The availability of high-throughput sequencing (HTS) technology has improved the ability to detect viral pathogens without *a priori* knowledge. However, the abundance of host and environmental nucleic acids in samples can impact the success of accurately identifying low abundance viral nucleic acids. This challenge has led to a variety of sample preparation techniques intended to enrich viral nucleic acids from complex samples, including targeted enrichment. ViroCap is a targeted enrichment panel which consists of capture probes designed to enrich sequences of viral species that infect vertebrates. This study reports an evaluation of ViroCap against a broad range of animal and zoonotic viruses representing 22 families, 27 genera, using a panel of blinded nucleic acid of viral isolates and clinical samples from wildlife and livestock. In addition to the accurate identification of the expected panel viruses, six unexpected viruses were detected, including several viruses with sequence divergence above

50% from known viruses. Also, a disease investigation case study is presented to demonstrate the utility of ViroCap targeted enrichment in true diagnostic veterinary scenarios. This work shows ViroCap targeted enrichment improves the detection and characterization of viruses in domestic species, such as sheep and pigs, as well as wildlife, and it can be an instrumental tool in veterinary disease investigations.

Introduction

Increasingly metagenomic high-throughput sequencing (HTS) is being used in the discovery of novel viruses (Duraisamy et al. 2018, Vanmechelen et al. 2018). It provides the ability to detect viral pathogens without prior knowledge of sequence information. This ability can overcome the challenges of traditional testing used in diagnostic laboratories such as PCR. Traditional testing can fail to detect highly divergent viruses, including novel subtypes of known species (Towner et al. 2008). In fact, metagenomic HTS has had such a large impact on viral detection that it is straining the current viral taxonomic system, due to the high number of new species being discovered (Simmonds et al. 2017). In addition to detection, HTS reveals high-resolution sequence information important for understanding the atielogy and spread of a viral disease outbreak (Zhou et al. 2018). However, the sensitivity of viral discovery using metagenomic HTS can be challenging in complex samples, such as tissue, due to the low abundance of viral nucleic acid in comparison to the host's nucleic acid and other environmental nucleic acids (Rosseel et al. 2012). This challenge can lead to poor sequencing depth, poor breadth of coverage of viral genomes and ultimately failure to detect viruses (Daly et al. 2011).

A variety of enrichment techniques are employed to increase the quantity of viral nucleic acid relative to the background, including filtration and centrifugation (Montmayeur et al. 2017, Zou et al. 2017). Nuclease treatments are also commonly used to enrich viral

sequence in clinical samples (Rosseel et al. 2015). The work of Hall et al. demonstrates employing nuclease treatments in combination with filtration and centrifugation will enrich sequencing reads of both RNA and DNA viruses (Hall et al. 2014). Targeted enrichment differs from previous enrichment methods, as it is based on using biotinylated oligonucleotide capture probes that will hybridize to the target of interest for pull-down enrichment using streptavidin-coated beads. This technique is being used to enrich for genes of interest in areas of human health, such as cancer research (Hagemann et al., 2013).

ViroCap is a sequence targeted enrichment method designed to enrich vertebrate viruses. Wylie et al. (2015) designed ViroCap using sequence data from 337 viral species from 190 genera, making up almost a billion base pairs of sequence data. The resulting 100 bp probes were screened against the human genome and any probes that shared over 75% sequence identity with the human genome were removed. The focus of previous studies using ViroCap were viruses that affect human health (Wylie et al. 2015, 2018). In this study, we have expanded the laboratory evaluation of ViroCap to include veterinary and zoonotic viruses. This study also includes a case study to demonstrate the utility of ViroCap enrichment in veterinary diagnostic disease investigations.

Methods

Previously extracted total nucleic acids from tissue culture (n=3) and field/clinical (n=19) samples from different sources containing viruses were blinded by staff not involved in this study prior to being used for the evaluation of the Virocap method in multiple batches. For batch 1, no cDNA synthesis was performed. Libraries were sheared to 500 bp with the Covaris™ M220 Focused-ultrasonicator™ (ThermoFisher). Library construction was performed on extracted total nucleic acids with the KAPA HyperPlus library kit (Roche), according to Nimblegen's SeqCap EZ HyperCap Workflow User's Guide V1.

For the 2 remaining batches, cDNA synthesis was performed using Superscript IV First-Strand Synthesis System (SSIV) (ThermoFisher). 11 uL of extracted total nucleic acid was mixed with dNTPS (10 mM) and a tagged random nonamer primer (40 uM) (GTT TCC CAG TCA CGA TAN NNN NNN NN). Samples were incubated at 65°C for 5 minutes, then placed on ice for 1 minute. A reagent mixture of 5x SSIV Buffer, Ribonuclease Inhibitor (40 U/μL), DTT (100 mM) and SuperScript™ IV Reverse Transcriptase was then added. The samples were incubated for 10 minutes at 23°C, 10 minutes at 50°C and 10 minutes at 80°C.

Second strand synthesis was performed using Sequenase Version 2.0 DNA Polymerase (ThermoFisher). The first strand synthesis product was incubated with 10 uL of Sequenase Version 2.0 DNA Polymerase diluted in 5x reaction buffer and nuclease free water. Samples were then heated to 37°C over 5 minutes and incubated at 37°C for 12 minutes, followed by 2 minutes at 95°C. Samples were cooled to 10°C and 1.2 uL of Sequenase DNA polymerase in dilution buffer was added. Samples were again ramped to 37°C over five minutes and incubated at 37°C for 12 minutes, followed by 8 minutes at 95°C.

A total of 6 uL of the second strand synthesis product was used as template for amplification. AccuPrime™ *Taq* DNA Polymerase (Thermofisher) was mixed with 10X AccuPrime™ PCR Buffer I, nuclease free water and a primer for the nonomer's tag (100 uM). 30 cycles of PCR were then performed with the following parameters: 30 secs at 94°C, 30 secs at 40°C, 30 sec at 50°C and 1 minute at 72°C.

cDNA/DNA mixtures were then cleaned with Genomic DNA Clean & Concentrator columns (Zymo Research) and eluted in 20 mM Tris (ThermoFisher). Batches 2 and 3 underwent library preparation with the KAPA HyperPrep library kit (Roche). Sequence library construction and capture were carried out according to Nimblegen's SeqCap EZ HyperCap Workflow User's Guide V1. Samples were pooled in equal amounts by weight

prior to capture.

Sequencing was performed on an Illumina MiSeq instrument in the National Centre for Foreign Animal biocontainment level 3 sequencing facility. Batch 1 was sequenced on a V3 flow cell with a 600 cycle cartridge. Batch 2 was sequenced using 500 cycle cartridges and on a V2 Micro. The same libraries used then sequenced on a V3 flow cell with a 600 cycle cartridge to allow comparison between capture and uncaptured sequencing statistics. Batch 3 was sequenced on V2 flow cells with 300 cycle cartridges. The same libraries used in capture, but unenriched were then sequenced on a V2 Micro flow cell with a 500 cycle cartridge. Flowcells were chosen with aim of obtaining at least 1-2 million reads per sample. A total of 18 pmoles of library was loaded on V3 flow cells and 8-9 pm loaded on V2 flow cells.

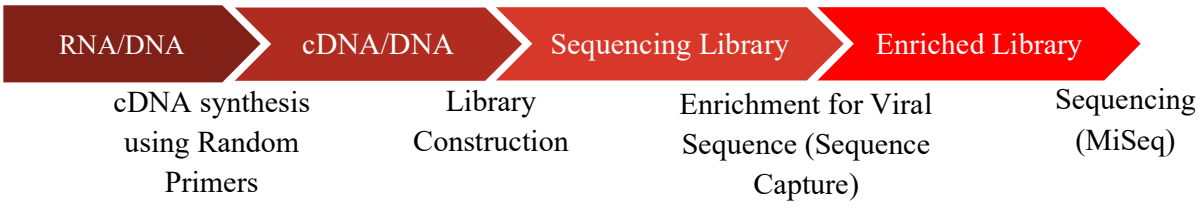

Figure 1: Laboratory methods for ViroCap targeted enrichment.

To calculate enrichment, the same libraries were sequenced before and after capture, with the exception of batch 1. For batch 1, uncaptured stats were calculated based on a library that was prepared in the following way. The samples were sheared for 550 base-pair (bp) insert size using Covaris M220 ultrasonicator. Libraries were then constructed on a Illumina NeoPrep instrument according to instructions in the TruSeq Nano DNA Library Prep for NeoPrep Reference Guide. The quantified and normalized libraries were pooled in equimolar amounts and sequenced on a MiSeq instrument with a V3 flow cell and a 600-cycle reagent cartridge

Read quality was assessed using FastQC and trimming was performed using Trimmomatic (Version 0.36), according to read quality (Bolger et al. 2014). *de novo* assembly was performed using SPAdes (v3.12.0) (Bankevich et al. 2012). SPAdes contigs were classified using Blastn (Altschul et al. 1990). The most appropriate references for reference assemblies were chosen based on sequence similarity to the SPAdes contigs. Reference assemblies were then performed using bwa mem (Version: 0.7.17-r1188) (Li & Durbin 2010). Duplicate sequences were removed with samtools rmdup (Li et al. 2009). Enrichment was calculated by dividing the percentage of viral reads in captured samples by the percent of viral reads in uncaptured samples. For viruses with segmented genomes, segments were concatenated to calculate sequencing statistics. For porcine astrovirus 1, no suiTable reference could be found in NCBI due to the high sequence divergence, therefore, contigs from the de novo assembly were used to calculate sequencing statistics. Alignments and percent pairwise identities were performed and calculated using ClustalW (Thompson et al. 1994). Highly divergent sequences were assembled by first sorting reads using Rambo-k (Tausch et al. 2015).

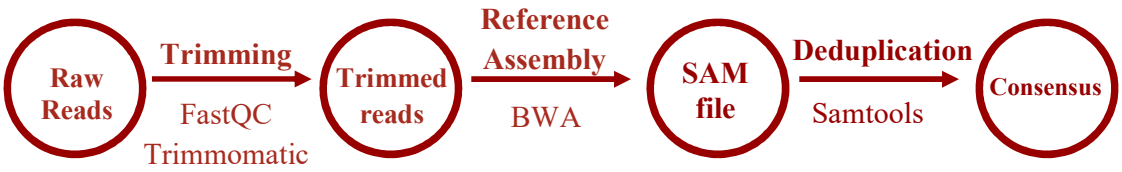

Figure 2: Bioiformatic analysis pipeline for Virocap targeted enrichment.

Results

Blinded panel

A blinded panel of 22 clinical or cell culture amplified samples were processed in 3

batches (n=5, 6, 11). The panel consisted of 19 clinical samples from animals and humans and 3 viral isolates grown in cell culture. A total of 24 viral species were sequenced from the panel (Table 1). These species represent all 7 Baltimore classes and include representatives from 22 viral families and 27 viral genera. Among these viruses are 11 viruses that cause OIE (World Organization for Animal Health) reportable diseases, such as African swine fever virus. 3 biosafety level 3 and 4 zoonotic viruses were also sequenced, such as Nipah virus. The panel, as shown in Table 1, consisted of 11 sample types from 10 host species, including domestic animal and wildlife species. Nine of the viruses in the panel had DNA genomes: six belong to Baltimore class I (double-stranded DNA viruses), two belong to Baltimore class II (single-stranded DNA viruses) and one virus belonged to Baltimore class VII (double-stranded DNA retroviruses). The rest of the viruses in the panel were RNA viruses (15 of 24).

The effect of ViroCap enrichment on the blinded panel sequencing is demonstrated in Table 1. Increase in the breadth of coverage was observed in all samples, with the exception of samples such as a Lassa mammarenavirus, Senecavirus A and Sheeppox virus, which were fully sequenced in the unenriched sample and therefore could not be increased. The average increase in the breadth of coverage was 43%. The depth of coverage of the panel increased by an average of 123. Percent viral reads (PVR) enrichment ranged from 1 to 1965 fold, with an average of 182-fold.

PVRs with and without enrichment are demonstrated in Figure 3. Viral titer in clinical samples can vary drastically (Olesen et al. 2017). This is reflected in the PVR of the blinded panel. Samples such as Seneca Valley virus contained high PVR in unenriched samples (20 and 40% viral reads) leaving little room for enrichment and thus resulting in small fold enrichment of PVR (e.g. 1). Similarly, Figure 4 demonstrates that increases in breadth of coverage due to ViroCap enrichment are drastic in cases with low PVR prior to enrichment (African Swine Fever virus clinical sample = 0.00140% PVR unenriched), while in other

cases with high unenriched PVR ViroCap targeted enrichment enabled the closing of coverage gaps (Porcine epidemic diarrhea virus sample = 0.4 PVR unenriched).

Veterinary Diagnostic Case Study: Geese Die-off

A total of 6,476,274 reads for 8 pooled samples were sequenced with ViroCap enrichment. 23,450,882 reads were sequenced using the same sequence libraries without ViroCap enrichment, meaning on average four times more reads were sequenced for the unenriched samples. Four viruses were detected (Table 2). The increase of PVR with ViroCap enrichment ranged from 19 to 524. The average breadth of coverage increased by 20% and the depth of coverage increased by an average of 47x. Only one virus was completely sequenced, for the remaing viruses genome coverage was 40-45%. Coverage is fragmented for the remaining three viruses, which could be due to the advanced decomposition of the samples (Figure 5).

Table 1: Enrichment of Viral Genomes with ViroCap targeted enrichment

Species	**Host**	**Sample type**	**% nt identity to VCap Design**	**Fold Increase**		
				PVR	**Breadth(%)**	**Depth (x)**
Ambystoma tigrinum virus[A]	western tiger salamander	Skin	99	535	43.1	113.2
Ambystoma tigrinum virus[A]	western tiger salamander	Skin	97.4	297	89.6	5.5
African Swine Fever virus[A]	Cell culture	N/A	94.7	77	0.6	227.5
African Swine Fever virus[B]	Domestic pig	Serum	99.5	169	87.8	26.5
Avian avulavirus[A, D]	Domestic sheep	Scab	99.8	N/A	97.1	4.2
Bovine viral diarrhea virus 1[B, D]	Northern Long Eared Bat	Brain Tissue	94.3	N/A	69.7	4
Epizootic hemorrhagic disease virus[C]	Cell Culture	N/A	97.4	16	80.1	20.8
Foot-and-mouth disease virus[B]	Domestic sheep	Oral swab	99.5	1965	17.3	19.8
Frog virus 3[C]	Snapping Turtle	Liver tissue	99.5	8	17.2	124
Frog virus 3[C]	Snapping Turtle	Liver tissue	99.6	2	67.4	14
Hepatitis B virus[C]	Human	Serum	98.7	282	92	62.3
Lassa mammarenavirus[C]	Cell Culture	N/A	59.7	4	0	283.6

Porcine astrovirus 1 [B, D]	Domestic pig	Serum	65.1	7	84.3	6.1
Nipah virus[B]	Domestic pig	Brain tissue	99.9	96	84.3	134.9
Orf virus[A]	Domestic sheep	Scab	98.6	763	14.7	136.3
Porcine Cytomegalovirus[B, D]	Domestic pig	Serum	98.9	69	40.1	2.3
Porcine epidemic diarrhea virus[B]	Domestic pig	Rectal swab	99.9	50	3.5	226.5
Porcine parvovirus 2[B, D]	Domestic pig	Serum	44.5	5	16.9	35.9
Porcine reproductive and respiratory syndrome virus 2[C]	Domestic pig	Rectal swab	88.2	31	75.4	276.1
Porcine Rotavirus [B, D]	Domestic pig	Serum	88.6	N/A	44.2	9.9
Rabbit hemorrhagic disease virus[C]	Rabbit	Liver tissue	84.7	2	0.1	243.6
Rabies virus[B]	Northern Long Eared Bat	Brain tissue	96.3	100	82.3	258.1
Rift Valley Fever phlebovirus[B]	Domestic Goat	Serum	99	12	1.7	119.1
Senecavirus[AC]	Domestic pig	Tonsil tissue	96.6	1	0	242.9
Senecavirus[AC]	Domestic pig	Oral Swab	95.6	1	0.3	271.8

Sheeppox virus[A]	Cell culture	Ovine testetells	99.1	9	0	208.4
Simian immunodeficiency virus[C]	Macaques	Blood	98.4	14	1.5	406.5
Simian immunodeficiency virus[C]	Macaques	Blood	98.2	20	67.1	7.7
Torque teno sus virus 1[B, D]	Domestic pig	Serum	92.2	204	77.2	67.4

A=Batch 1, B=Batch 2, C=Batch 3, D=Unexpected
*segmented viral genomes were concatenated for calculation
PVR = Percent Viral Reads, VCap = ViroCap

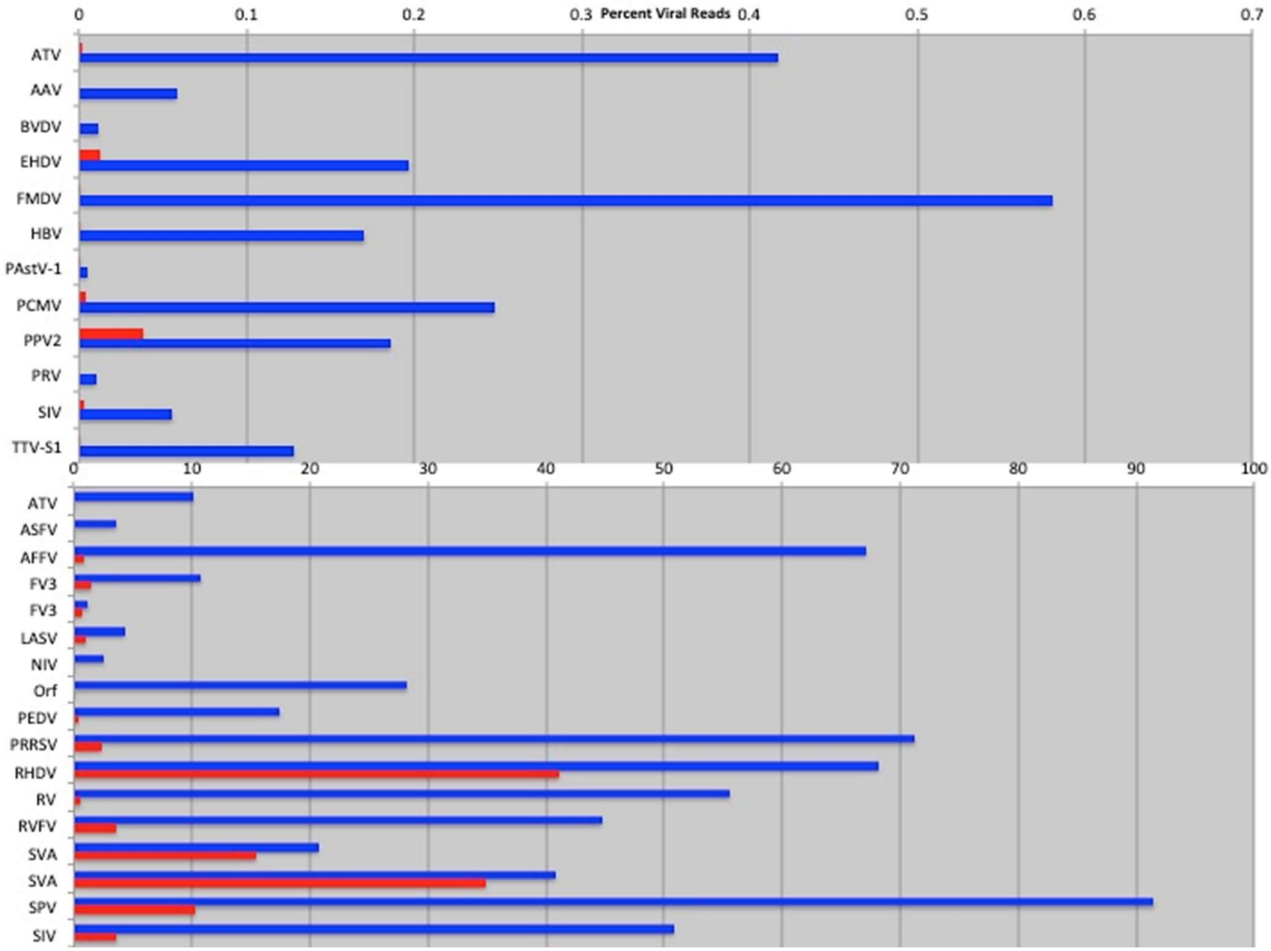

Figure 3: Percent viral reads of various viruses with and without ViroCap enrichment. Red indicates PVR before enrichment and blue indicates PVR with ViroCap targeted enrichment. ATV = Ambystoma tigrinum virus, AAV = Avian avulavirus, ASFV = African Swine Fever virus, BVDV = Bovine viral diarrhea virus, EHDV = Epizootic hemorrhagic disease virus, FMDV = Foot-and-mouth disease virus, FV3 = Frog virus 3, HBV = Hepatitis B virus, LASV = Lassa mammarenavirus, NIV = Nipah virus, PAstV-1 = Porcine astrovirus 1, PCMV = Porcine Cytomegalovirus, PEDV = Porcine epidemic diarrhea virus, PPV2 = Porcine parvovirus 2, PRRSV = Porcine reproductive and respiratory syndrome virus, PRV = Porcine Rotavirus, RHDV = Rabbit hemorrhagic disease virus, RV = Rabies virus, RVFV = Rift Valley Fever virus, SVA = Senecavirus A, SPV = Sheeppox virus, SIV = Simian immunodeficiency virus, TTV-S1 = Torque teno sus virus 1.

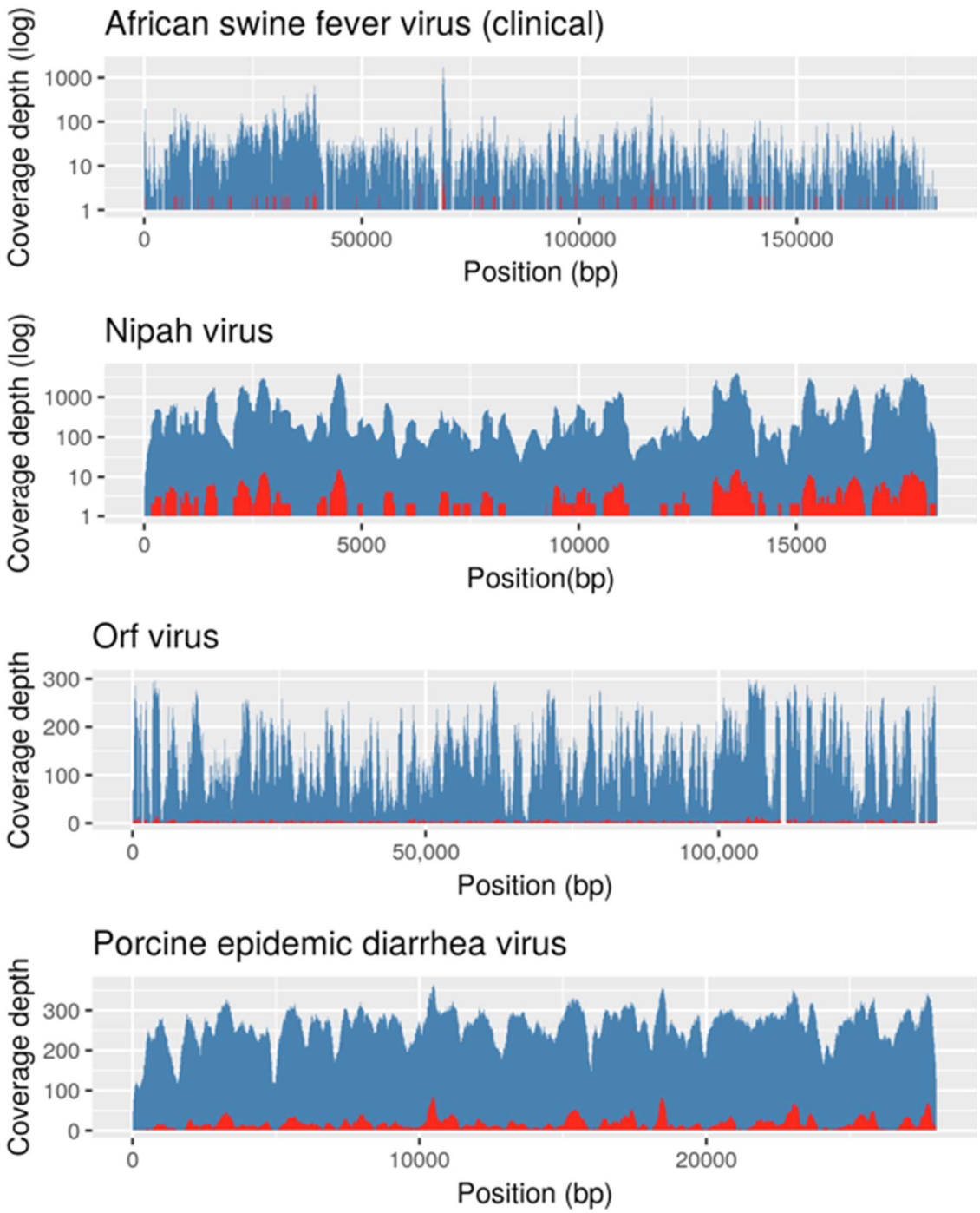

Figure 4: Sequencing Coverage plots with and without ViroCap targeted enrichment for selected viruses. Red indicates before enrichment and blue indicates PVR with ViroCap targeted enrichment.

Table 2: High-throughput sequencing and ViroCap enrichment statistics of 4 viruses detected in a large die-off of Canada and Snow Geese.

Virus	Host - Sample type	Millions of reads		Breadth of coverage		Mean Depth of Coverage		PVR		Enric-hment	% identity to ViroCap Reference
		w VCap	w/o VCap	w VCap	w/o VCap	w VCap	w/o VCap	w VCap	w/o VCap		
Adeno-associated virus	SG - CS	1.53	7.56	42	19	4.6	0.7	0.06	0.0003	213	71
Goose Adeno-virus 4	SG - CS	1.53	7.56	45	12	10.2	0.3	0.4	0.0008	476	96
Avian Meta-Pneum-ovirus	CG - PS	2.05	9.14	40	12	265	26	4.6	0.0014	3420	98
Goose Corona-virus CB17	CG - CS	2.90	6.75	98	100	125	190	34	1.8	19	61

SG = Snow goose, CG = Canada goose, CS = Cloacal swab, PS = Pharyngeal swab, Vcap = Virocop, w = with and w/o = without

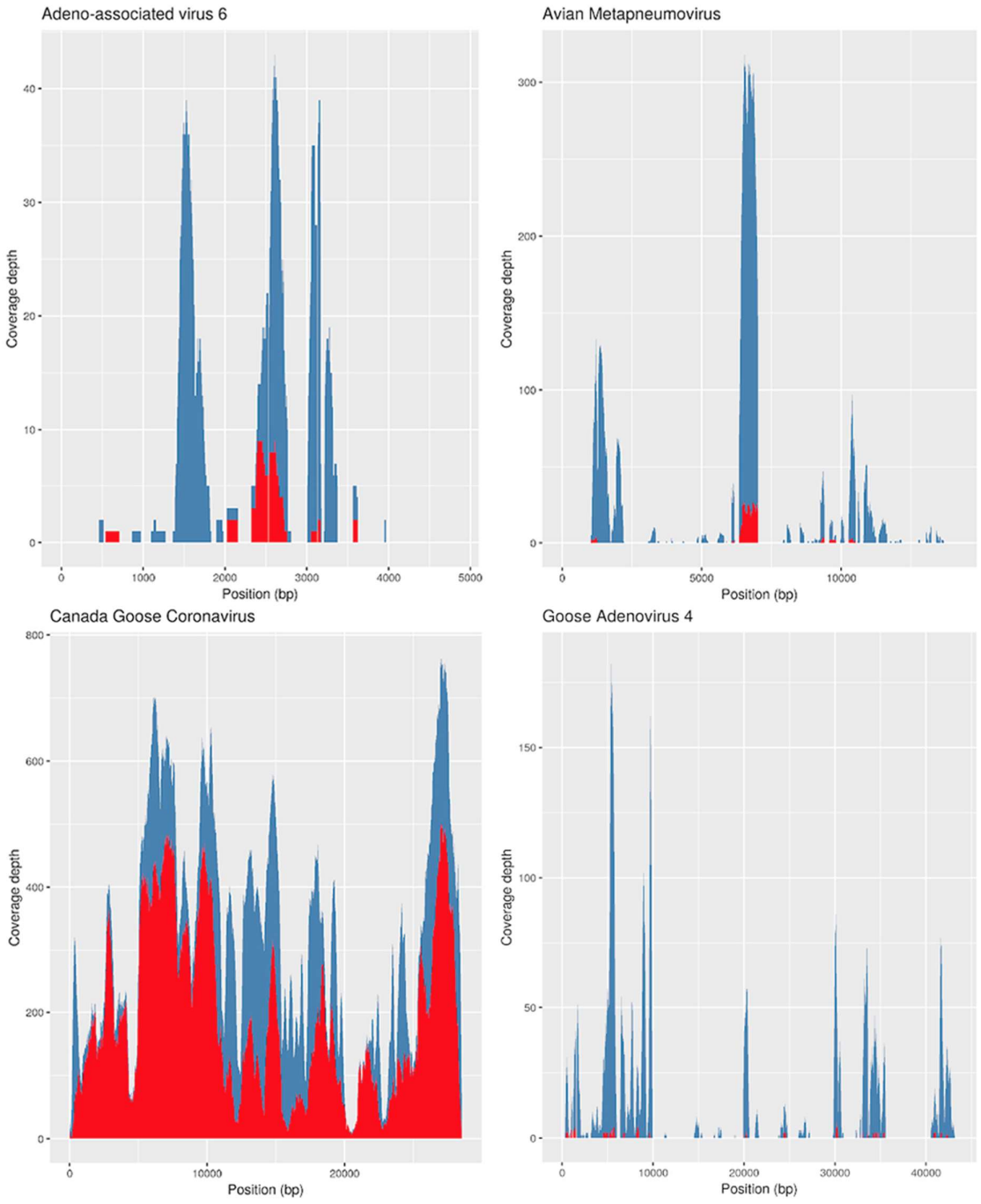

Figure 5: Coverage plots of four viruses detected in a mass die-off of Canada and Snow geese. Red indicates coverage before ViroCap enrichment and blue indicates coverage with ViroCap enrichment.

Discussion

Metagenomic viral detection is currently being used in human medicine to identify previously elusive viral pathogens, such as the cause of febrile illness in children. However, low sensitivity can make cell culture amplification of viral isolates necessary for complete genetic characterization (Haddad-Boubaker et al. 2019). Improvements in sensitivity would therefore increase the utility of viral metagenomics in clinical diagnostics. To facilitate this, sequence capture panels have been designed for groups of viruses associated with specific disease scenarios, such as a panel designed for 34 viruses which can cause respiratory illness (Yang et al. 2018). In contrast, several broader panels encompassing all known vertebrate viruses have been designed (Briese et al. 2015, Wylie et al. 2015). These panels have been designed for and validated on human samples. Previously, a panel of veterinary viral sequence capture probes was designed for felid species. Specifically, Lee et al. designed a capture for the use in domestic cats, bobcats and mountain lions, which detects 7 viral pathogens and a bacterial pathogen (Lee et al. 2017). This study, to the authors' knowledge, demonstrates the first evaluation of a broad range targeted enrichment method applied to numerous animal hosts. This design of ViroCap incorporated several steps with the objective of improving the panel for use in human diagnostics, e.g. probes were screened against the human genome and probes with over 75 % identity were eliminated. Despite this, the mean enrichment seen in this study (182) is comparable to that of the original ViroCap study for human viruses in two experiments (674 and 296) (Wylie et al. 2015). An increase in breadth and depth of coverage was observed in viral genomes tested from a wide range of host backgrounds, demonstrating the potential of ViroCap targeted enrichment for use in veterinary diagnostics.

In addition to correctly identifying all viruses in the blinded panel, 6 unexpected viruses were also detected. These viruses were not known to be present in the panel prior to

sequencing. The majority of these cases were from mixtures of viral isolates or reagent contaminants. As is the case of bovine viral diarrhea virus (BVDV) sequenced from bat brain. This sample was processed for virus isolation in media containing fetal bovine serum, in which BVDV is a common contaminant. One case arose from a probable co-infection. African swine fever virus, porcine astrovirus 1, porcine cytomegalovirus, porcine parvovirus 2, and torque teno sus virus were all sequenced from one swine serum sample. Despite the read number between enriched and unenriched samples being comparable (1-2 million reads), porcine rotavirus was not detected in the sequencing run without ViroCap. The ability to detect a broad spectrum of viruses from many different taxonomic groups in one reaction is a major improvement over traditional testing that targets a single pathogen or a limited number of pathogens. Multiplexing strategies, such as multiplex PCR, are rarely extended beyond 20 targets (Hanson & Couturier 2016). These unexpected viruses would likely go undetected if traditional tests, such as PCR, had been used for detection. This is particularly important as some diseases are multifactorial in nature, such as postweaning multisystem wasting syndrome (Blomström et al. 2016) and viral co-infection can impact disease severity (Scotta et al. 2016). The ability to detect multiple viruses is therefore particularly important in disease investigations.

Validation of tests in vitro on known and well-described viruses doesn't reflect the true nature of diagnostics. In the field, animal carcases degrade and are subject to predation. Transportation can introduce more degradation, especially from remote locations. The blinded panel evaluation was effective in demonstrating the range of viruses and host backgrounds in whichViroCap enrichment is able to improve viral characterization. The second part of this study describes the use of ViroCap in a true diagnostic disease scenario. A disease investigation case study was presented to demonstrate the utility of ViroCap targeted enrichment using an actual diagnostic submission from a die-off of Canada and snow geese.

In August 2017, dozens of deceased juvenile Canada and snow geese were discovered on a beach in Cambridge Bay, Nunavut. Following the failure of toxicological and microbiological testing to detect a cause for the die-off, the samples then became an "unknown", making them an ideal candidate for ViroCap enrichment and HTS. 8 samples from the geese die-off were subjected to ViroCap enrichment and HTS. Cloacal and pharygynal swabs were collected from 3 birds (2 Canada geese and one snow goose). Lung tissue was only available from 2 Canada geese, as many internal organs were missing due to predation.

This investigation was included to demonstrate the utility of ViroCap enrichment in veterinary diagnostics. As mentioned above, the difficulty of characterizing viral genomes in clinical samples often means isolation is required before viruses can be fully genetically characterized, due to the fact that a complete genome is difficult to sequence without amplification to increase PVR (Jaune et al. 2018). Viral isolation was attempted using samples from the Canada and snow geese die-off by passaging twice in embryonated chicken eggs, but infectious virus could not be isolated. This negative result is most likely due to the poor quality and poor state of the samples by the time it reached the laboratory. Poor quality and/or limited sample are common scenarios in viral diagnostics (Jonassen et al. 2005). Howevewr, HTS combined with ViroCap targeted enrichment revealed 4 viruses in these samples. Furthermore, the near complete genome of a novel Goose Coronavirus CB17, a novel *Gammacoronavirus,* was detected in samples from all three birds submitted for testing (Papineau et al. 2019). The novel coronavirus was present in high enough PVR (1.3%) to be nearly completely genetically characterized without capture, if sufficient sequencing was used on the samples. Still, the PVR was 20x higher with ViroCap enrichment, despite a very low nucleotide pairwise identity to known coronaviruses (61.2%). This demonstrates that ViroCap can be useful in the discovery and genetic characterization of novel viruses. This is also evident in the case of porcine parvovirus 2 (PPV2) sequenced from a swine serum sample

included in the blinded panel. PPV2 showed 5 times PVR enrichment, despite only a 45% pairwise nucleotide identity to the closest virus sequence used in the design of ViroCap.

In addition to the novel *Gammacoronavirus*, 3 other viruses were detected in the geese die-off samples. Avian metapneumovirus (AMPV) was detected in the pharyngeal swab of one Canada goose. Despite obtaining 4 times more sequencing reads for this sample in the unenriched sequencing run, the breadth of coverage increased from 11% to 44% with ViroCap enrichment. This increase in coverage enabled sequencing of the complete glycoprotein ORF, a gene important for the determination of host range and virulence (Bennett et al. 2005). Thus, ViroCap enrichment enabled more detailed genetic characterization with lower sequencing throughput. Similar results were observed for goose adenovirus 4 and an adeno-associated virus.

This increase in depth of genetic characterization can play an important role in diagnostic scenarios. Traditional nucleic acid detection tests are based on the detection of small genomic targets through hybridization with primers and probes. In addition to being sensitive to mutations that can result in false negative results (Towner et al. 2008), these PCR-based tests also provide no sequence information. The ability to detect and sequence the genome of a virus can improve outbreak response, e.g. improving biosecurity procedures by increasing the understanding of the spread of disease (Houlihan et al. 2018). Therefore, enrichment methods like ViroCap can have an impact on not just the ability to detect viral pathogens but also on the management of the disease these pathogens cause.

Examples of other methods available for enrichment of virus from metagenomic samples for sequencing include nuclease treatments, centrifugation and filtration. The efficiency of these methods can vary depending on the sample type. For example, Rossel et al (2015) demonstrated DNase treatment resulted in a maximum viral enrichment of 83x in

serum and 32x in tissue samples (Rossel et al. 2015). This method was targeted for RNA viruses and would reduce the effectiveness of acquiring sequence information from viruses with DNA genomes. The maximum viral enrichment achieved in this study in tissue was 535x for DNA viruses and 100x for RNA viruses. It should be noted that there are many factors that differ between this study and Rossel et al.'s (2015), ranging from sample composition to the sequencing platform used. Hall et al (2014) evaluated physical enrichment methods on both RNA and DNA viruses. They also demonstrated the vartiability of the effect of DNase treatments and recommend that this technique be combined with centrifugation and filtration (Hall et al, 2014). Hall et al. (2014) used artificial samples constructed from a mixture of human cells, *E. coli* cells and virus. Neither of these studies evaluated swab samples, an important sample type in veterinary diagnostics. Our work demonstrates that in true diagnostic veterinary samples representing a range of sample types, ViroCap enrichment has a positive impact on the ability to detect both RNA and DNA viruses. As ViroCap is based on known viral sequence, there is a possibility that novel viruses can be missed, something that physical based enrichment methods would be advantageous for in theory. However, this work also demonstrates that the ViroCap targeted enrichment successfully sequenced viruses with high sequence divergence from the viral sequences used in the capture probe design.

PCR is commonly used (including in this study) to generate sufficient input material for library construction from low-input samples, such as field collected or clinical material. PCR is thought to contribute to artificial read duplicates in HTS data (Li et al. 2009). Thus, in this study, coverage was calculated after removal of PCR duplicates in order to prevent PCR duplicates from affecting sequencing statistics. As low input library construction methods become available and compatible with the capture process, future experiments can decrease cycle numbers required. In addition to the PCR duplicates, PCR can also lead to increase

index jumping when samples are multiplexed. While some adapter jumping is common in Illumina sequencing (Griffiths et al. 2018), capture can contribute to increased index jumping. Index jumping was observed in our study as multiple samples were batched in the same run to reduce the cost of sequencing. This problem can potentially be reduced in future studies with unique dual indexing strategies (MacConaill et al. 2018).

To summarize, the data presented here shows ViroCap targeted enrichment can be used to increase the sensitivity of viral detection and characterization in animal studies. The detection of unexpected and highly divergent viruses shows it is suitable for the detection of viruses in veterinary diagnostics. This study highlights the benefits of viral enrichment through broad range sequence capture, including, higher sensitivity for viral pathogens, ability to identify mixed infections, highly divergent and unknown/unexpected pathogens.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215(3):403–10

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19(5):455–77

Bennett RS, LaRue R, Shaw D, Yu Q, Nagaraja KV, et al. 2005. A wild goose metapneumovirus containing a large attachment glycoprotein is avirulent but immunoprotective in domestic turkeys. *J. Virol.* 79(23):14834–42

Blomström A-L, Fossum C, Wallgren P, Berg M. 2016. Viral Metagenomic Analysis

Displays the Co-Infection Situation in Healthy and PMWS Affected Pigs. *PLoS ONE*. 11(12):e0166863

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30(15):2114–20

Briese T, Kapoor A, Mishra N, Jain K, Kumar A, et al. 2015. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *MBio*. 6(5):e01491-15

Daly GM, Bexfield N, Heaney J, Stubbs S, Mayer AP, et al. 2011. A viral discovery methodology for clinical biopsy samples utilising massively parallel next generation sequencing. *PLoS ONE*. 6(12):e28879

Duraisamy R, Akiana J, Davoust B, Mediannikov O, Michelle C, et al. 2018. Detection of novel RNA viruses from free-living gorillas, Republic of the Congo: genetic diversity of picobirnaviruses. *Virus Genes*. 54(2):256–71

Griffiths JA, Richard AC, Bach K, Lun ATL, Marioni JC. 2018. Detection and removal of barcode swapping in single-cell RNA-seq data. *Nat. Commun.* 9(1):2667

Haddad-Boubaker S, Joffret M-L, Pérot P, Bessaud M, Meddeb Z, et al. 2019. Metagenomic analysis identifies human adenovirus 31 in children with acute flaccid paralysis in Tunisia. *Arch. Virol.* 164(3):1–9

Hall RJ, Wang J, Todd AK, Bissielo AB, Yen S, et al. 2014. Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. *J. Virol. Methods*. 195:194–204

Hanson KE, Couturier MR. 2016. Multiplexed molecular diagnostics for respiratory,

gastrointestinal, and central nervous system infections. *Clin. Infect. Dis.* 63(10):1361–67

Houlihan CF, Frampton D, Ferns RB, Raffle J, Grant P, et al. 2018. Use of Whole-Genome Sequencing in the Investigation of a Nosocomial Influenza Virus Outbreak. *J. Infect. Dis.* 218(9):1485–89

Jaune FW, Taques IIGG, Dos Santos Costa J, Araújo JP, Catroxo MHB, et al. 2018. Isolation and genome characterization of canine parvovirus type 2c in Brazil. *Braz. J. Microbiol.*

Jonassen CM, Kofstad T, Larsen I-L, Løvland A, Handeland K, et al. 2005. Molecular identification and characterization of novel coronaviruses infecting graylag geese (Anser anser), feral pigeons (Columbia livia) and mallards (Anas platyrhynchos). *J. Gen. Virol.* 86(Pt 6):1597–1607

Knight-Jones TJD, Rushton J. 2013. The economic impacts of foot and mouth disease - what are they, how big are they and where do they occur? *Prev. Vet. Med.* 112(3–4):161–73

Lee JS, Mackie RS, Harrison T, Shariat B, Kind T, et al. 2017. Targeted enrichment for pathogen detection and characterization in three felid species. *J. Clin. Microbiol.* 55(6):1658–70

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 26(5):589–95

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25(16):2078–79

MacConaill LE, Burns RT, Nag A, Coleman HA, Slevin MK, et al. 2018. Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. *BMC Genomics*.

19(1):30

Montmayeur AM, Ng TFF, Schmidt A, Zhao K, Magaña L, et al. 2017. High-Throughput Next-Generation Sequencing of Polioviruses. *J. Clin. Microbiol.* 55(2):606–15

Olesen AS, Lohse L, Boklund A, Halasa T, Gallardo C, et al. 2017. Transmission of African swine fever virus from infected pigs by direct contact and aerosol routes. *Vet. Microbiol.* 211:92–102

Papineau A, Berhane Y, Wylie TN, Wylie KM, Sharpe S, Lung O. 2019. Genome Organization of Canada Goose Coronavirus, A Novel Species Identified in a Mass Die-off of Canada Geese. *Sci. Rep.* 9(1):5954

Rosseel T, Ozhelvaci O, Freimanis G, Van Borm S. 2015. Evaluation of convenient pretreatment protocols for RNA virus metagenomics in serum and tissue samples. *J. Virol. Methods*. 222:72–80

Rosseel T, Scheuch M, Höper D, De Regge N, Caij AB, et al. 2012. DNase SISPA-next generation sequencing confirms Schmallenberg virus in Belgian field samples and identifies genetic variation in Europe. *PLoS ONE*. 7(7):e41967

Scotta MC, Chakr VCBG, de Moura A, Becker RG, de Souza APD, et al. 2016. Respiratory viral coinfection and disease severity in children: A systematic review and meta-analysis. *J. Clin. Virol.* 80:45–56

Simmonds P, Adams MJ, Benkő M, Breitbart M, Brister JR, et al. 2017. Consensus statement: Virus taxonomy in the age of metagenomics. *Nat. Rev. Microbiol.* 15(3):161–68

Tausch SH, Renard BY, Nitsche A, Dabrowski PW. 2015. RAMBO-K: Rapid and Sensitive Removal of Background Sequences from Next Generation Sequencing Data. *PLoS ONE*.

10(9):e0137896

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22(22):4673–80

Towner JS, Sealy TK, Khristova ML, Albariño CG, Conlan S, et al. 2008. Newly discovered ebola virus associated with hemorrhagic fever outbreak in Uganda. *PLoS Pathog.* 4(11):e1000212

Vanmechelen B, Bletsa M, Laenen L, Lopes AR, Vergote V, et al. 2018. Discovery and genome characterization of three new Jeilongviruses, a lineage of paramyxoviruses characterized by their unique membrane proteins. *BMC Genomics*. 19(1):617

Wylie KM, Wylie TN, Buller R, Herter B, Cannella MT, Storch GA. 2018. Detection of viruses in clinical samples by use of metagenomic sequencing and targeted sequence capture. *J. Clin. Microbiol.* 56(12):

Wylie TN, Wylie KM, Herter BN, Storch GA. 2015. Enhanced virome sequencing using targeted sequence capture. *Genome Res.* 25(12):1910–20

Yang Y, Walls SD, Gross SM, Schroth GP, Jarman RG, Hang J. 2018. Targeted Sequencing of Respiratory Viruses in Clinical Specimens for Pathogen Identification and Genome-Wide Analysis. *Methods Mol. Biol.* 1838:125–40

Zhou L, Ren R, Yang L, Bao C, Wu J, et al. 2017. Sudden increase in human infection with avian influenza A(H7N9) virus in China, September-December 2016. *Western Pac. Surveill. Response J.* 8(1):6–14

Zhou P, Fan H, Lan T, Yang X-L, Shi W-F, et al. 2018. Fatal swine acute diarrhoea syndrome

caused by an HKU2-related coronavirus of bat origin. *Nature*. 556(7700):255–58

Zou X, Tang G, Zhao X, Huang Y, Chen T, et al. 2017. Simultaneous virus identification and characterization of severe unexplained pneumonia cases using a metagenomics sequencing technique. *Sci. China Life Sci.* 60(3):279–86

Chapter 3: Viruses Detected in a Mass Die-off of Canada and Snow Geese

Abstract

A mass die-off of Canada and snow geese occurred in Nunavut Canada in the fall of 2017. Following the failure of toxicological and microbiological testing to detect pathogens, ViroCap targeted enrichment and high-throughput sequencing revealed 4 viruses. The complete genome of a novel coronavirus was detected in the cloacal swab of one Canada goose. Phylogenetics and genetic analysis reveal it is a new species of the genus gammacoronavirus, which was given the designation Canada goose coronavirus. The genome of Canada goose coronavirus includes 6 novel ORFs, a partial duplication of the 4 gene and a presumptive change in the proteolytic processing of polyproteins 1a and 1ab. Three partial genomes were also detected, including that of an avian metapneumovirus sharing high nucleotide similarity to the Colorado strain. In addition, the first genetic information of Goose adenovirus 4 and Avian dependoparvovirus 1 in wild Canadian geese is presented. Metagenomic investigation revealed the presence of several viruses with the ability to infect domestic avian species and a novel coronavirus of unknown pathogenicity, indicating the need for increased surveillance and additional studies to determine pathogenicity of Canada goose coronavirus.

Introduction

Given the role of migratory birds in the transmission and maintenance of livestock and human pathogens (Endo & Nishiura 2018), it is important to monitor the viral community of migratory birds. A large die-off of juvenile Canada and snow geese occurred in Cambridge Bay Nunavut in August of 2017. Samples were provided from three birds (1 snow goose and 2 Canada geese). All geese were in poor body condition; subcutaneous and intracoelomic adipose tissue was absent and pectoral mass was reduced. The carcasses were heavily predated and decomposing. Most internal organs were missing and in one case (the snow goose) the head was entirely missing. Despite the poor condition, samples were tested at the University of Calgary. Mineral and toxicological testing were unremarkable. Tests for duck viral enteritis virus, fowl cholera (*Pasteurella multocidasepticemia*), avian influenza, Newcastle disease and Duck Viral Enteritis were negative. The pathology report produced by the University of Calgary and Samuel Sharpe indicated the possibility that salt intoxication and/or poor body condition may have contributed to the large die-off.

The viruses detected in the 2017 mass die-off of Canada and snow geese were described in Chapter 2, as part of a discussion of the effectiveness of ViroCap in veterinary diagnostic disease investigations. This chapter will describe the genetics and phylogeny of the 4 viruses detected in the mass die-off of Canada and Snow geese. The 4 viruses detected in the mass die-off are listed in Table 3.

Table 3: Viruses detected in a mass die-off of Canada and Snow Geese. CG = Canada Goose, SG = Snow Goose, CS = Cloacal Swab, PS = Pharyngynal swab.

Virus	**Host -**	**Breadth of coverage**	**Mean Depth of Coverage**

	Sample type		
Goose Coronavirus CB17	CG – CS	98	125
Avian Meta-pneumovirus	CG – PS	40	265
Goose Adenovirus 4	SG – CS	45	10.2
Adeno-associated virus	SG – CS	42	4.6

Methods

Source of samples

A large die-off of Canada and snow geese occurred in the fall of 2017 near the Arctic in Cambridge Bay, Nunavut, Canada. Due to poor carcass quality and remote location, samples were only collected from two dead Canada geese and one Snow goose, all of which had undergone predation and decomposition. Cloacal and pharyngeal swabs were collected from all three birds, lung tissue was collected from only the Canada geese. Other organs were not present or were in extremely poor condition. Detection of both common avian pathogens, such as avian influenza and avian paramyxovirus by the National Reference Laboratory, by routine laboratory testing gave negative results. Virus isolation was performed by two serial passages in SPF chicken eggs using protocols prescribed by the World Organization for Animal Health (OIE) for the most closely related gammacoronavirus, infectious bronchitis virus (IBV). Samples were then subjected to targeted sequence enrichment (Wylie et al. 2015) and high-throughput sequencing on an Illumina MiSeq platform.

Sample pre-treatment

Tissues were homogenized using a Precellys Evolution homogenizer (Bertin Instruments) according to the manufacturer's instructions. Following a clarification by centrifugation at 3000 rpm for 10 minutes, nucleic acids were extracted using the MagMAX Pathogen RNA/DNA Kit (Ambion) according to the manufacturer's instructions.

cDNA synthesis was then performed using SuperScript™ IV First-Strand Synthesis System (SSIV) (ThermoFisher) according to the manufacturer's recommendation. A total of 11 uL of extracted total nucleic acid was mixed with dNTPS (10 mM) and a tagged random nonamer primer (40 uM) (GTT TCC CAG TCA CGA TAN NNN NNN NN). Samples were incubated at 65°C for 5 minutes, and then placed on ice for 1 minute. A reagent mixture of 5x SSIV Buffer, Ribonuclease Inhibitor (40 U/μL), DTT (100 mM) and SuperScript™ IV Reverse Transcriptase was then added. The samples were incubated for 10 minutes at 23°C, 10 minutes at 50°C and 10 minutes at 80°C.

Second strand synthesis was performed using Sequenase Version 2.0 DNA Polymerase (ThermoFisher) according to the manufacturer's recommendation. The first strand synthesis product was incubated with 10 uL of Sequenase Version 2.0 DNA Polymerase diluted in 5x reaction buffer and nuclease free water. Samples were then heated to 37°C over five minutes and incubated at 37°C for 12 minutes, followed by 2 minutes at 95°C. Samples were then cooled to 10°C and 1.2 uL of Sequenase DNA polymerase in dilution buffer was added. Samples were again ramped to 37°C over five minutes and incubated at 37°C for 12 minutes, followed by 8 minutes at 95°C. A total of 6 uL of the second strand synthesis product was then used as template for amplification. AccuPrime™ *Taq* DNA Polymerase (Thermofisher) was mixed with 10X AccuPrime™ PCR Buffer I, nuclease free water and a primer for the nonomer's tag (100 uM). 30 cycles of PCR were then performed with the following parameters: 30 seconds at 94°C, 30 seconds at 40°C, 30 seconds at 50°C and 1 minute at

72°C. cDNA/DNA mixtures were then cleaned with Genomic DNA Clean & Concentrator columns (Zymo Research) and eluted in 20 mM Tris (ThermoFisher).

Library preparation and sequencing

Sequence libraries were prepared with the KAPA HyperPlus library kit (Roche). Sequence library construction and capture were carried out according to Nimblegen's SeqCap EZ HyperCap Workflow User's Guide V1. Samples were pooled in equal amounts by weight prior to capture. Sequencing was performed on an Illumina Miseq instrument in the National Centre for Foreign Animal Disease biocontainment level 3 sequencing facility. A V2 flow cell was used with a 500 cycle reagent cartridge (Illumina).

5' Race and Sanger sequencing

5' RACE was used to obtain the missing leader sequence (52 bp). The SMARTer 5' RACE and 3' RACE kit (Takarabio) was used according to the kit instructions. The gene specific primer used for 5' RACE was TCAGCTACAGTAGAGGGAGATGTCATAGGTGC. For Sanger sequencing, amplicons was performed using KAPA HiFi HotStart ReadyMixPCR Kit (KAPABiosystems). The primers CTAAAGAGAAGGTGGACACTGGT and CTAAGAATGCGAACTTCACAGAGC were used to amplify the gene 4b homologue region. The primers GTTGTTGTGTTACAAGGCAAGGG and GGATTATGATCAAACCATGAACCTGG were used to amplify the NSP 10/12 region. Cycling conditions used to generate amplicon for Sanger sequencing were: 1 cycle: 95°C for 3 minutes, 40 cycles: 98°C for 20 seconds, 65°C for 15 seconds, 72°C for 2.5 minutes, and 1 cycle: 72°C for 3 minutes. Amplicons were cleaned using AMPure XP beads (Beckman Coulter) according to the manufacturer's directions. Sanger sequencing was performed on the

ABI Genetic Analyzer 3130XL platform using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) according to the user manual

Bioinformatics

Read quality was assessed using FastQC and trimmed using Trimmamatic (Bolger et al. 2014) (Version 0.36). Host reads were then filtered with RAMBO- K, using the only complete genome of a goose species (swan goose, *Anser cygnoides*) currently available and DCoV (Tausch et al. 2015). The near complete genome sequence of Goose Coronavirus CB17 (GCoV-CB17) was assembled from HTS derived sequences from a cloacal swab of one Canada goose using SPAdes (Bankevich et al. 2012). Sanger reads were aligned to the draft genome in Geneious™ (Biomatters, v 9.1.8). Annotations were performed using Geneious and protein domains were identified using PFAM (El-Gebali et al. 2019). The Canada goose coronavirus genome is available under accession number MK359255 on NCBI.

The Genome Orginization of Canada Goose Coronavirus

Viruses belonging to the *Coronaviridae* family have a single stranded positive sense RNA genome of 26-31 kb. Members of this family include both human pathogens, such as severe acute respiratory syndrome virus (SARS-CoV) (Tsang et al. 2003), and animal pathogens, such as porcine epidemic diarrhea virus (Wood 1977). Currently, the International Committee on the Taxonomy of Viruses (ICTV) recognizes four genera in the *Coronaviridae* family: *Alphacoronavirus, Betacoronavirus, Gammacoronavirus and Deltacoronavirus*. While the reservoirs of the *Alphacoronavirus* and *Betacoronavirus* genera are believed to be bats, the *Gammacoronavirus* and *Deltacoronavirus* genera have been shown to spread primarily through birds[3]. The first three species of the *Deltacoronavirus* genus were

discovered in 2009 (Woo et al. 2009) and recent work has vastly expanded the *Deltacoronavirus* genus, adding seven additional species (Woo et al. 2012).

By contrast, relatively few species within the *Gammacoronavirus* genus have been identified. There are currently two recognized species in the *Gammacoronavirus* genus: avian coronavirus (ACoV) and beluga whale coronavirus SW1 (SW1). ACoVs infect multiple avian hosts and include several important poultry pathogens, such as infectious bronchitis virus (IBV) and turkey coronavirus (TCoV) (de Groot et al. 2008). IBV was first described in the United States (Fabricant 1998) but has since been described around the globe (Bande et al. 2017). Turkey Coronavirus is the cause of acute enteritis in domestic turkeys (Lin et al. 2002). The second species in the *Gammacornavirus* genus, SW1, was first discovered in beluga whales (Mihindukulasuriya et al. 2008) but has since been detected in other cetaceans, such as Indo-Pacific bottlenose dolphins (Woo et al. 2014). Despite IBV being the first discovered coronavirus and the impact it has on the poultry industry (Jones 2010), the number of identified species within the *Gammacoronavirus* genus remains small in comparison to the other coronavirus genera. Coronaviruses from several other avian hosts for which partial sequences are available suggest relatedness to IBV and TCoV. These viruses, which include goose coronavirus (GCoV), were tentatively classified as part of the ACoV species. An approximately 3 kb region, including the nucleocapsid gene and several accessory genes, of GCoV were previously sequenced from a greylag goose in Norway (Jonassen et al. 2005). Here we present the full genome of Goose coronavirus CB17 (GcoV-CB17) sequenced directly from the cloacal swab of a Canada goose, which expired in a mass die-off in a remote region near the arctic in Nunavut, Canada. Our analyses demonstrate that it should be classified as a novel species in the *Gammacoronavirus* genus.

Due to the remote location of the die off, samples from the dead birds were not collected immediately and sent to a diagnostic laboratory until severe predation and

decomposition had occurred. The poor sample quality, in addition to the difficulty of coronavirus isolation, led to the failure to isolate infectious virus using standard methods. However, the nearcomplete genome of a novel gammacoronavirus was assembled from high throughput sequencing reads derived from the cloacal swab of a single Canada goose. The assembled genome of the novel Goose Coronavirus (GCoV-CB17) is 28,539 nts in length (excluding the poly(A) tail) and has 38.4% GC-content. GCoV-CB17 is approximately 1000 nts longer than the reference genomes for ACoV available in GenBank. The genome organization of GCoV-CB17 is presented in figure 6. The 5' UTR of CGCoV is 553 nt in length and contains a higher GC content (48.3%) relative to the genome as a whole. The 5' UTR of CGCoV shares only 68% pairwise identity with that of duck coronavirus (DCoV) and 47.5% pairwise identity to that of SW1. Like all coronavirus genomes reported to date, GCoV-CB17's genome is dominated by the coding regions for the large polyproteins 1a and 1ab, followed by the structural and accessory genes. The heptanucleotide slippery sequence UUUAAAC, associated with the ribosomal slippage that produces polyprotein 1ab, was present at nt positon 11995. GCoV-CB17's genome contains genes for all four structural proteins common to coronaviruses; spike (S), envelope (E), membrane (M) and nucleocapsid (N). In addition, GCoV-CB17 contains 10 open reading frames (ORFs) predicted to encode accessory proteins. The order of the structural and accessory protein-coding ORFs in GCoV-CB17 resembles that of ACoV, but there are noTable differences. The general genome organization of ACoV is 1ab-S-3a-3b-E-M-4b-4c-5a-5b-N-6b (Cao et al. 2008). However, there is some variance in the genome organization within the ACoV species. For example, Australian IBV strains lack ORFs 4a, 4b and 5b (Mardani et al. 2008). Overall, GCoV-CB17 contains a larger number (n=14) of ORFs coding for predicted accessory and structural proteins downstream of the polyprotein 1ab coding region. Two additional ORFs (7a and 7b) are found between the GCoV-CB17 M and N ORFs. There are also two additional ORFs (10

and 11) following the N gene. While some ACoVs do have ORFs following the N gene, ORFs 10 and 11 in GCoV-CB17 do not share obvious homology to those of IBV and TCoV. The 3' UTR of GCoV-CB17 is 301 nucleotides in length and contains the stem loop-like motif 113 bp upstream from the poly(A) tail. This stem loop-like motif was first identified in astroviruses (Jonassen et al. 1998) but is also present in ACoVs and SARS-CoV (Cao et al. 2008). Further downstream in the 3' UTR, the octanucleotide motif (GGAAGAGC) is found 71 bp upstream of the poly(A) tail. The 3' UTR of GCoV-CB17 shares 98% pairwise identity to the partially sequenced GCoV and 84 % pairwise identity to IBV.

A trait suggesting common ancestry between GCoV-CB17 and ACoV is the canonical ACoV transcription regulatory sequence (TRS) found at the end of the leader sequence in GCoV-CB17. The TRS of GCoV-CB17 is identical to that identified by Cao et al (2008) as the TRS of TCoV (CTTAACAAA). Body TRS's regulate viral gene expression by forming a complex with the leader TRS, causing discontinuous transcription of mRNA (Dufour et al. 2011). Ten putative body TRSs were found in the 3' end of the GCoV-CB17 genome (figure 6). Four of the ten putative TRSs (4, 6, 8, 9) were exact matches to the canonical leader TRS. Three TRSs (2, 7, 11) contained one mismatch and the remaining three TRSs (3, 5, 10) contained two mismatches to the leader TRS. The functionality of these TRSs would need to be experimentally determined; however, previous studies have shown that TRSs of ACoVs are subject to some variation (Bentley et al. 2013, Cao et al. 2008). GCoV-CB17 contains twice the number of TRS's as ACoVs and a similar number compared to the nine contained in SW1[9]. Table 4 demonstrates the nucleotide distances between the TRS and the start codon of ORFs found in GCoV-CB17's, which are comparable to those of TCoV (Cao et al. 2008).

Table 4: Putative viral proteins of Canada goose coronavirus

Protein	Top Match in NCBI	Top match - aa % identity*	Size (aa)	Distance between TRS and start codon (nt)
1a	1a-Infectious bronchitis virus strain B1648	43	3825	480
1ab	1ab-Infectious bronchitis virus strain ck/CH/LJL/05I	57	6510	480
Spike	Spike-Infectious bronchitis virus strain N2-75	53	1184	82
3	n/a	n/a	53	0
4a	n/a	n/a	55	3
Envelope	Envelope-Infectious bronchitis virus strain IS-1494	69	100	n/a
Membrane	Membrane-Duck Coronavirus isolate DK/GD/2014	72	235	74
5b	4b-Infectious bronchitis virus strain Georgia 1998 Vaccine	41	88	n/a
6	n/a	n/a	63	5
7a	4b-Duck Coronavirus isolate DK/GD/2014	23	92	3

7b	n/a	n/a	69	n/a
8a	5a-Duck Coronavirus isolate DK/GD/2014	37	65	4
8b	5b-Duck Coronavirus isolate DK/GD/2014	46	85	n/a
Nucleocapsid	Nucleocapsid-Goose Coronavirus	94	414	94
10	ORFxg-Goose Coronavirus	92	97	0
11	ORFyg-Goose Coronavirus	81	180	91

*Matches below 20% coverage not shown.

The start codon of GCoV-CB17's polyprotein 1ab is located 567 nucleotides downstream of the leader TRS. The coronavirus polyprotein 1ab is cleaved into 15-16 non-structural proteins (NSPs) by two viral proteases (Ziebuhr et al. 2000). Putative cleavage sites for these proteases are present in GCoV-CB17's 1a and 1ab polyproteins, with the exception of the NSP 10/11 (polyprotein 1a) and NSP 10/12 (polyprotein 1ab) cleavage sites. The missing cleavage site would be located near the end of polyprotein 1a, producing the NSPs 10 and 11, and also in the alternatively transcribed polyprotein 1ab, producing NSPs 10 and 12. The absence of the NSP10/11 and 10/12 protease recognition site was confirmed with Sanger sequencing. With the exception of the missing cleavage sites, the putative cleavage sites would produce NSPs of sizes congruent with other *Gammacoronavirus* species (table 5). No

Gammacoronavirus species to date, including GCoV-CB17, have a papain-like protease cleavage site between NSP 1-2 (Kint et al. 2016).

Table 5: Non-structural proteins size and cleavage site of gammacoronaviruses.

	GCoV-CB17		**TcoV**		**IBV**		**SW1**	
Protein	**Cleavage site**	**Size aa**	**Cleavage Site**	**Size aa**	**Cleavage site**	**Size aa**	**Cleavage site**	**Size aa**
NSP1/2	AG^GH	609	AG^GK	673	AG^GK	673	VD^GD	636
NSP3	AG^GV	1532	AG^GV	1594	AG^GI	1592	LG^GV	1586
NSP4	LQ^AG	503	LQ^AG	514	LQ^SG	514	LQ^AG	537
NSP5	LQ^SN	307	LQ^SS	307	LQ^SS	307	LQ^SN	303
NSP6	VQ^SK	295	VQ^SK	297	VQ^AK	293	VQ^SK	303
NSP7	LQ^AV	83	LQ^SV	83	LQ^SV	83	LQ^AV	83
NSP8	LQ^NN	212	LQ^NN	210	LQ^NN	210	LQ^NN	198
NSP9	LQ^GK	111	LQ^SK	111	LQ^SK	111	LQ^HG	112
NSP10	**SRFV***	**173**	**VQ^SA**	**145**	**VQ^SV**	**145**	**LQ^SV**	**189**
NSP11	**-**	**-**	**-**	**23**	**-**	**23**	**-**	**17**
NSP12	**SRFV***	**1101**	**VQ^SA**	**941**	**VQ^SV**	**940**	**LQ^SV**	**926**
NSP13	LQ^SC	599	LQ^SC	601	LQ^SC	600	LQ^AS	601
NSP14	LQ^SN	522	LQ^GT	521	LQ^GT	514	LQ^SQ	528
NSP15	LQ^SI	338	LQ^SI	338	LQ^SI	338	LQ^SL	349
NSP16	LQ^SG	298	LQ^SA	302	LQ^SA	302	LQ^SD	312

*amino acids present in GCoV-CB17 where putative protease cleavage sites were observed in TCoV, IBV and SW1

While the genome structure of GCoV-CB17 resembles that of ACoV, there are some noTable differences. For example, there are no homologues to ACoV's 3a or 3b accessory proteins in GCoV-CB17, a trait shared with SW1. Furthermore, GCoV-CB17 has a number of ORFs that do not appear to have homologues in other sequenced *Gammacoronavirus* species, such as the ORFs for putative proteins 3 and 4a (Figure 6). These two ORFs are found in GCoV-CB17 in the corresponding location of ACoV's 3a and 3b ORFs (between the S and E ORFs) and are also similar in size to ACoV's 3a and 3b proteins. However, they share no obvious sequence similarity with any 3a or 3b gene, or any other entry in NCBI (table 3). ACoV's 3a and 3b proteins have been shown to be unnecessary for replication (Hodgson et al. 2006), however knock-out mutants with these accessory genes deleted are attenuated (Laconi et al. 2018, Liu et al. 1991). The IBV's 3 gene is functionally tricistronic, meaning the 3a, 3b and E proteins are under the control of a single TRS (Brooks et al. 2004). This is not the case in GCoV-CB17, as the E ORF of GCoV-CB17 shares a TRS with only the 4a ORF in GCoV-CB17 and 3 ORF is preceded by a separate TRS (Figure 6).

An additional TRS is also found in between GCoV-CB17's M and N ORFs, preceding the proteins 7a and 7b (Figure 6). Commonly, ACoVs have two ORFs between the M and 5 genes, coding for the 4b and 4c accessory proteins. GCoV-CB17 contains 4 ORFs between the M and 8 gene (AcoV 5 gene homologue). Two of these ORFs (5b and 7a) are ACoV 4b homologues, likely the result of gene duplication. This region in IBV has been identified as a hotspot for recombination (Woo et al. 2012). The region between the ACoV M and 5 gene was formally called the intergenic region because of the lack of a TRS. However, it was later shown that gene 4 is expressed using an alternative TRS in IBV (Bentley et al. 2013). Notably, one of the 4b homologs (i.e. 5b) in GCoV-CB17 does have a TRS (figure 6). The use of template switching at TRSs is thought to lend to recombination in coronaviruses

(Simon-Loriere & Holmes 2011). The two GCoV-CB17 4b homologs are not identical to each other (table 1). Amino acid sequence identity to other 4b proteins is low for both GCoV-CB17 4b homologues, 41% to IBV and 23% to DCoV respectively. The gene 4 duplication was also confirmed by Sanger sequencing of the genomic region between the M ORF to the 8 gene.

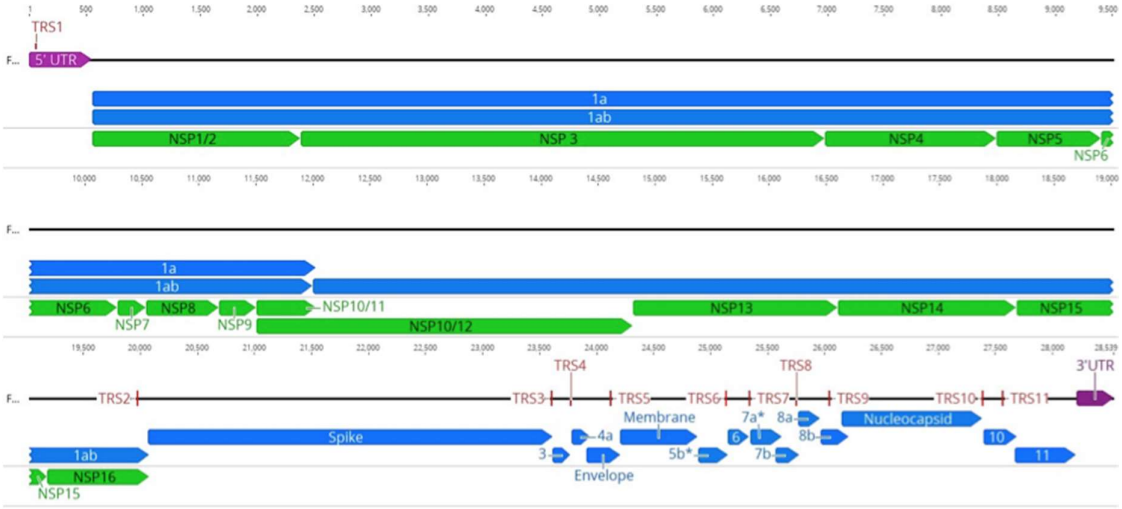

Figure 6: Genome organization of Canada goose coronavirus. Purple indicates untranslated regions, blue indicates putative proteins, green indicates coding region of mature non-structural proteins (NSP) and red indicates transcription regulatory sequences (TRS). The stem loop-like motif and octamer motif are contained within the 3' UTR. Genome organization Figure was constructed using Geneious™ (Biomatters, v 9.1.8). * indicate ACoV 4b homologues. Proteins are named numerically from the 5' end of the genome, with the exception of the structural genes, which are denoted by their common names.

The ACoV 5a and 5b accessory proteins (8a and 8b in GCoV-CB17) appear to be the only accessory proteins conserved in all 3 *Gammacoronavirus* species, although gene order

differs. ORFs encoding putitive proteins 5a and 5b belong to the bicistronic gene 5 of ACoVs and are also unnecessary for replication (Laconi et al. 2018). To date, all publically available sequence information suggest that *Gammacoronavirus* species have lost the NSP1 cleavage site. The function of NSP1 in alphacoronaviruses and betacoronaviruses is the inhibition of host protein production. Accessory protein 5a is shown to have adopted this function in place of NSP1 in IBV (Kint et al. 2016).

The majority of structural proteins of GCoV-CB17 also share low amino acid sequence identity (53-72%) with IBV and DCoV. Phylogenetic analysis of the spike gene show that the GCoV-CB17 spike gene clusters with the IBV spike gene, separate from the TCoV cluster (Figure 7). Figure 7 also demonstrates the nucleocapsid gene of GCoV-CB17 is distantly related to those of ACoVs. However the GCoV-CB17 nucleocapid protein does share 94% amino acid sequence identity with the nucleocapsid protein encoded in the partially sequenced graylag GCoV genome (Cao et al. 2008). In addition, ORFs 10 and 11, which are preceded by the nucleocapsid gene, also share high amino acid identity with graylag GCoV proteins, 92% and 81% respectively. It should be noted that, among full and partial genomes of gammacoronaviruses sequenced to date, ORFs 10 and 11 seem to be unique to GCoV-CB17 and GCoV and are both preceded by a TRS, suggesting that these ORFs are very likely expressed. The fact that some GCoV-CB17 proteins share higher amino acid sequence similarity with the partial GCoV sequences available suggest these two viruses are more closely related to each other than to other gammacoronaviruses known to date.

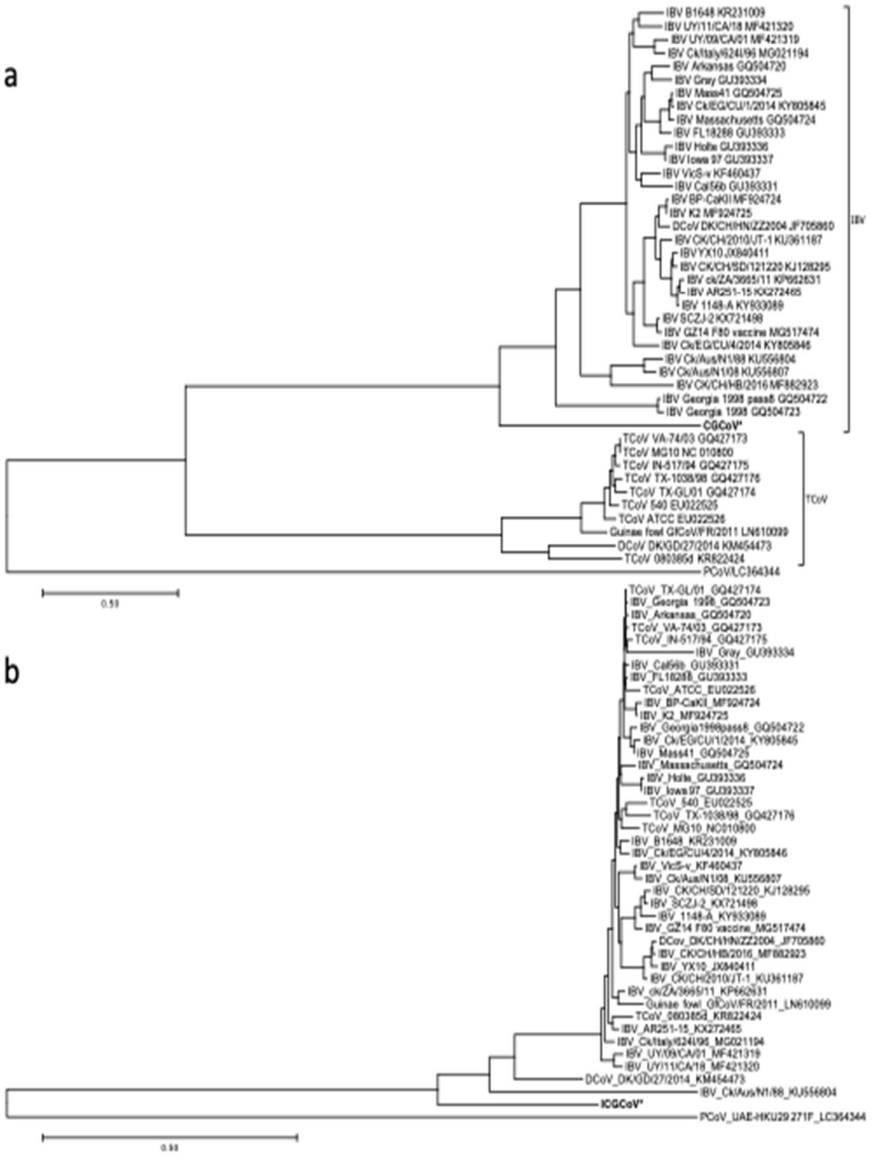

Figure 7: The phylogeny of gammacoronavirus spike and nucleocapsid proteins. A maximum likelihood tree built, using the amino acid sequences of the spike protein (A) and nucleocapsid protein (B) domains aligned with ClustalW (Larkin et al. 2007), in MEGA X using the Jones-Taylor-Thornton (JTT) substitution model and 1000 bootstraps (Kumar et al. 2018). IBV Infectious Bronchitis virus, TCoV Turkey Coronavavirus, PCoV Pigeon Coronavirus, DCoV Duck Coronavirus.

The phylogenetic tree built using the coding regions for the conserved replicase and helicase domains demonstrates that GCoV-CB17 clusters with gammacoronaviruses and shares a more recent common ancestor with ACoV than with the cetacean gammacoronaviruses (Figure 8). Further comparisons suggest that GCoV-CB17 is a separate species from ACoV. Current taxonomy of *Coronaviridae* is determined using pairwise comparisons of the amino acid sequence of seven conserved domains in the 1ab polyprotein. Members of the same species share over 90% amino acid identity in these seven conserved domains (De Groot et al. 2008). Percent identity of GCoV-CB17 falls well below the 90% threshold set by ICTV with ACoV and SW1, suggesting GCoV-CB17 is a separate species (Table 6). Within *Coronaviridae*, GCoV-CB17 shares the highest homology (68%) in the 7 conserved domains to the gammacoronaviruses TCoV and DCoV.

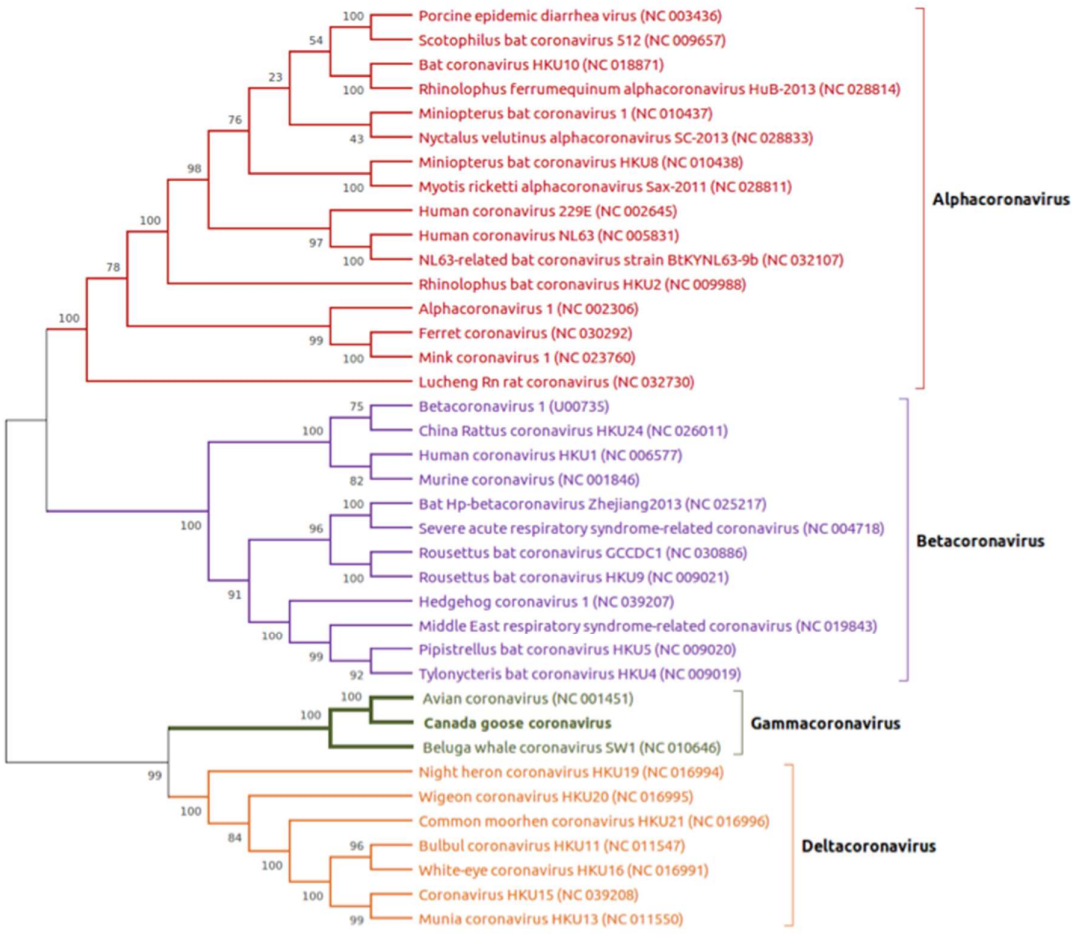

Figure 8: The phylogeny of Canada goose coronavirus. A maximum likelihood tree built, using the amino acid sequences of the replicase and helicase protein domains aligned with ClustalW (Larkin et al. 2007), in MEGA X using the Jones-Taylor-Thornton (JTT) substitution model and 1000 bootstraps (Kumar et al. 2018). Numbers at nodes indicate the bootstrap values.

Table 7: Comparison of the amino acid pairwise identity of 7 conserved coronavirus domains in the poly1ab protein of Canada goose coronavirus to other gammacoronaviruses.

Domain	aa % identity to IBV	aa % identity to TcoV	aa % identity to DcoV	aa % Identity to SW1

ADP-ribose-1″ -phosphatase	42	43	38	23
3C-like Protease	56	58	57	49
RdRp	80	80	83	69
Helicase 1	89	90	92	78
Exonuclease	78	72	77	56
Endoribonuclease	53	53	54	41
Ribose-2'-O methyltransferase	74	77	76	65
Average	67	68	68	54

As the full genome was sequenced from only the cloacal swab of a single Canada goose, a screening PCR was designed based on the 4b duplication region unique to GCoV-CB17 and performed on all samples. The Sanger sequencing primers of the region between the M and 8 gene were used, as this area of the genome is specific to GCoV-CB17. All samples were found to be positive, with the exception of the pharyngeal swab of the snow goose and the lung tissue of the second Canada goose which could not be tested as the sample was exhausted. Amplicons were Sanger sequenced and confirmed to match the GCoV-CB17 genome. High throughput sequencing conducted on RNA extracted from cloacal swabs from the second Canada goose and the snow goose also resulted in partial (64 and 18%) genomes of the GCoV-CB17. While this does confirm the virus's presence in all animals that perished in the die-off, this shows GCoV-CB17 was present in all birds that were available for testing. Further studies will require the availability of an infectious virus to determine the

pathogenicity of GCoV-CB17 and its ability to cause mortality in Canada geese and snow geese.

To summarize, the complete genome of GCoV-CB17, a novel *Gammacoronavirus* species was sequenced directly from the cloacal swab of a Canada goose associated with a mass die-off. The GCoV-CB17 genome was also detected in samples derived from a second Canada goose and a snow goose that perished in the die-off, using PCR, Sanger and high throughput sequencing. Comparative genomics and phylogenetic analysis indicate GCoV-CB17 clusters with ACoV but is a distinct *Gammacoronavirus* species. Interesting features of this new species include the presence of two 4b homologues, a putative change in the proteolytic processing of the polyproteins 1a and 1ab, and six novel accessory genes.

Partial Avian Metapneumovirus genome

In addition to the complete genome of Canada Goose Coronavirus, 3 partial viral genomes were also detected, including avian metapneumovirus (AMPV). AMPV belongs to the family *Pneumoviridae*, now recognized as its own family but formerly a subfamily of *Paramyxoviridae*. *Pneumoviridae* contains the genus *Metapneumovirus*, whose members infects both avian and mammalian hosts. The genus contains only two species; human metapneumovirus and AMPV. AMPV infects a wide range of domestic and wild avian species (Gough et al. 1988). In domestic avian species, AMPV is associated with reduced egg-laying (Sugiyama et al. 2006), swollen-head syndrome in domestic chickens (Maharaj et al. 1994) and rhinotracheitis in domestic turkeys (Lupini et al. 2011). There are four subtypes of AMPV determined by antigenic difference and sequence divergence of the glycoprotein. Subtypes A and B are prevalent in Asia and Europe (Banet-Noach et al. 2005; Tucciarone et al. 2017, 2018). Subtype C is the only subtype present in North America (Jardine et al. 2018, Seal 1998). 42% of an AMPV genome was sequenced from a pharyngeal swab of a Canada

goose in this study. It shares 98% nucleotide sequence identity to the Colorado strain of AMPV C. The Colorado strain was the first strain of AMPV isolated from sick turkeys in the United States (Seal 1998).

The partial genome sequenced from the Cambridge-Bay die-off included the glycoprotein gene, which was used to construct the tree in Figure 9. This tree demonstrates the AMPV detected in this study belongs in the subtype C clade (in red). AMPV has previously been isolated from Canada geese (Bennett et al. 2005). While the genome of AMPV sequenced in this study shares 98% nucleotide identity with other AMPV C strains, the glycoprotein gene is significantly divergent. The glycoprotein of the AMPV from this study shares 81% amino acid identity to that of other Canada Goose AMPV and 92% to that of the Colorado strain of AMPV (Table 7).

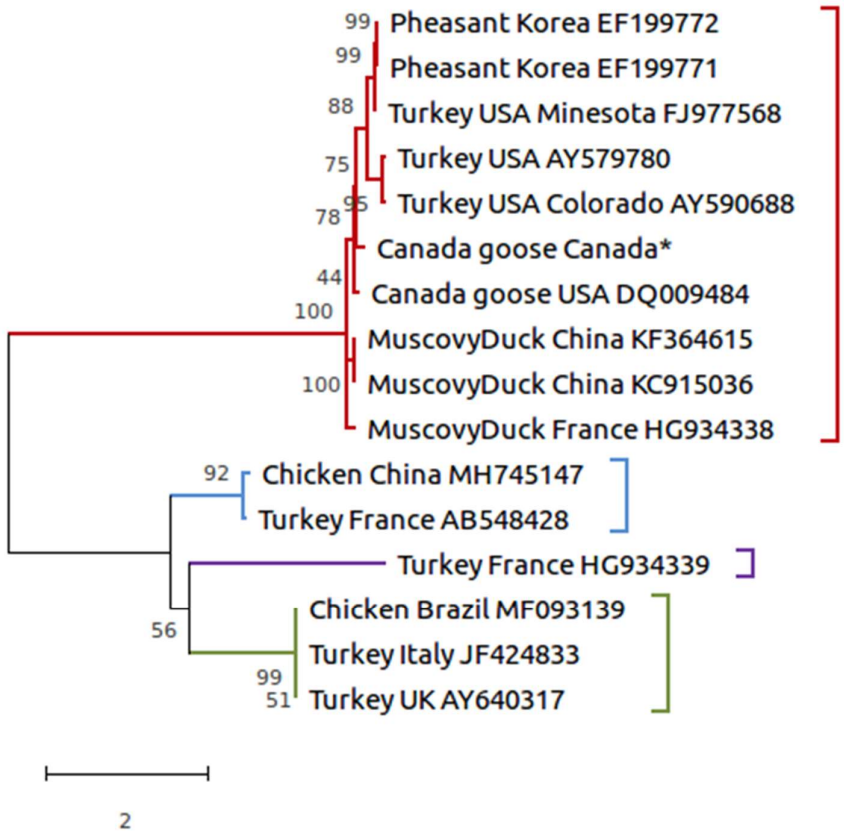

Figure 7: Maximum likelihood constructed using the nucleotide sequence of the coding sequences of the glycoproteins of avian metapneumovirus. The tree was constructed using IQ-TREE (Nguyen et al. 2015), with the substitution model TIM2+F+I+G4 and 1000 ultra-fast bootstraps.

Table 7: Pairwise amino acid identities of glycoproteins of Avian metapneumoviruses

Domestic strains of AMPV	**Canada goose/Cambridge Bay 2017**	**Canada goose/USA/DQ009484**
Chicken/Brazil/MF093139	11	9

Chicken/China/MH745147	13	10
Canada goose/USA/DQ009484	81	100
MuscovyDuck/China/KC915036	69	80
MuscovyDuck/China/KF364615	69	80
MuscovyDuck/France/HG934338	69	78
Pheasant/Korea/EF199771	84	80
Pheasant/Korea/EF199772	84	80
Turkey/France/AB548428	13	11
Turkey/France/HG934339	13	10
Turkey/Italy/JF424833	11	10
Turkey/UK/AY640317	11	10
Turkey/USA/Colorado/AY579780	92	89
Turkey/USA/Minesota/FJ977568	81	78

Partial Adeno-Associated virus genome

42% of an adeno-associated virus was also sequenced from the cloacal swab of the sampled snow goose. Small contigs (600 bp) matching this virus were also observed in the cloacal swab of one sampled Canada goose. Adeno-associated viruses belong to the *Dependoparvirus* genus in the *Parvoviridae* family. The name adeno-associated originates from their discovery as contaminants in adenovirus cultures (Atchison et al. 1965). The linear single-stranded genomes of the *Dependoparvovirus* genus are approximately 4.7 kb in length and contain 2 ORFs, encoding replication and capsid proteins (Cotmore et al. 2019). Mammalian dependoparvoviruses are best known for their ability to integrate into the host

genome and therefore their applicability in gene editing (Naso et al. 2017). Avian dependoparvoviruses, on the other hand, are best known for the diseases they cause in domestic waterfowl.

ICTV currently recognizes two species of avian *Dependoparvovirus*; Avian dependoparvovirus 1 and Anseriform dependoparvovirus 1. The Avian dependoparvovirus 1 species contains the helper-dependent non-pathogenic members of the *Dependoparvovirus* genus (true adeno-associated viruses) (Kapgate et al. 2018). Anseriform dependoparvovirus 1 is a species of waterfowl pathogens. There are two main subtypes within the Ansiform dependoparvovirus 1 species. Goose parvovirus, the causative agent of Derzsy's disease, was first identified in the 1960s (Derzsy 1967). There are two clinical manifestations of anseriform dependoparvovirus in geese; one which involves ascites, hydropericardium, hepatitis and myocarditis and a second clinical manifestation of enteric disease. The second subtype of anseriform dependoparvirus 1 was isolated from Muscovy ducks, in which it causes a range of clinical disorders ranging from muscular and myocardial disorders, hepatitis and neurological disorders (Glávits et al. 2005). Both subtypes are capable of infecting both geese and Muscovy ducks. While severe disease is observed in Muscovy ducks infected by both subtypes, only virus of goose origin causes severe disease in gosliHTS (Glávits et al. 2005).

The partial dependoparvovirus genome identified in this study shares 90% nucleotide pairwise identity to a muscovy duck dependoparvovirus isolated in China. To further complicate the situation, Chinese muscovy duck dependoparvovirus was not classified in the ansiform dependoparvovirus 1 species (the pathogenic waterfowl group), but rather in the avian dependoparvovirus 1 species (the true adeno-associated viruses). This new Muscovy duck dependoparvovirus was isolated from a co-infection with an adenovirus (Su et al. 2017). A phylogenetic tree was constructed using the complete capsid gene that could be recovered

from the partial genome of the dependoparvovirus sequenced from the Cambridge Bay die-off (Figure 10). The avian helper-dependant dependoparvoviruses are grouped in red. The tree demonstrates, that like the recently isolated Chinese muscovy duck avian dependoparvovirus, the partial dependoparvovirus from this study also beloHTS to the avian dependovirus grouping (in red) and not the pathogenic dependoparvovirus grouping (in purple). The identity of this virus as a helper-dependent virus, rather than a pathogenic dependoparvovirus, is further supported by the presence of an adenovirus in the Cambridge Bay die-off.

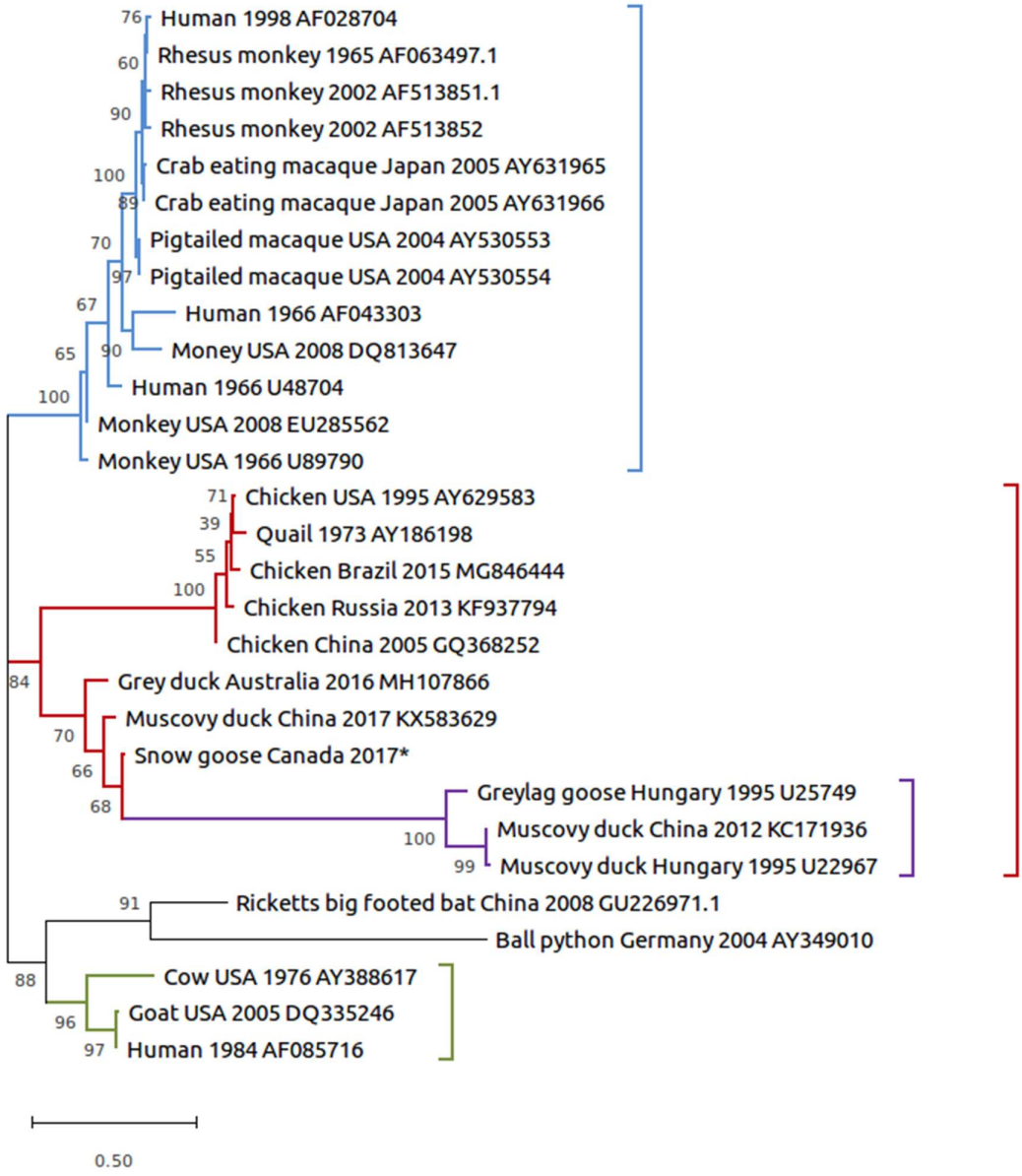

Figure 10: Maximum likelihood constructed using the nucleotide sequence of the capsid proteins of dependoparvovirus genus. The tree was constructed using IQ-TREE (Nguyen et al. 2015), with the substitution model TIM3e+G4 and 1000 ultra fast bootstraps. Blue = primate group, red = avian group, purple = pathogenic avian subgroup, green = livestock group.

Partial Goose Adenovirus genome

45% of an adenovirus was recovered from the cloacal swab of the snow goose tested. The *Adenoviridae* family contains the genus *Aviadenovirus*, a genus which infects only birds. Aviadenoviruses have increasingly become a concern for the poultry industry (Schachner et al. 2018). *Aviadenovirus* genomes are double-stranded DNA and range in size from 43 to 45 kb. Species demarcation in the *Aviadenovirus* genus is determined by over 95% amino acid pairwise identity in the polymerase gene. There are 3 different disease scenarios caused by aviadenoviruses; adenoviral gizzard erosion, hydropericardium-hepatitis syndrome (HHS) and inclusion body hepatitis (Hess 2017). Goose adenovirus has been associated with the HHS syndrome in juvenile domestic geese. This species has caused large-scale die-offs in Hungarian farms. The isolate from the Hungarian farm goose die-off was designated Goose adenovirus 4 and is the only complete goose adenovirus genome available on NCBI (Ivanics et al. 2010). Unfortunately, due to predation and decomposition, there was no liver or heart tissue to be examined for pathology or metagenomics in the case of the Cambridge bay die-off. The partial adenovirus genome obtained from the snow goose cloacal swab (including the polymerase gene), shared 96% nucleotide identity with goose adenovirus (Table 8), meaning this genome falls just?above the demarcation line to be included in the Goose adenovirus species.

Table 8: Pairwise distances of Snow goose adenovirus detected in a mass die-off Canada and Snow geese to known aviadenoviruses.

Accession number	**Adenovirus species**	**nt % identity to snow goose adenovirus**
NC_000899.1	Fowl adenovirus D, complete genome	46%
NC_001720.1	Fowl adenovirus A, complete genome	50%

NC_014564.2	Turkey adenovirus 1, complete genome	44%
NC_014969.1	Fowl adenovirus E, complete genome	46%
NC_015323.1	Fowl adenovirus C, complete genome	49%
NC_017979.1	Goose adenovirus 4, complete genome	96%
NC_021221.1	Fowl adenovirus 5 strain 340, complete genome	46%
NC_022612.1	Turkey adenovirus 4 isolate TNI1, complete genome	47%
NC_022613.1	Turkey adenovirus 5 isolate 1277BT, complete genome	49%
NC_024474.1	Pigeon adenovirus 1 complete genome, strain IDA4	47%
NC_024486.1	Duck adenovirus 2 strain GR, complete genome	55%
NC_031503.1	Pigeon adenovirus 2 isolate YPDS-Y-V1.A19.11-2013, complete genome	50%
NC_038332.1	Fowl adenovirus 6 strain CR119, complete genome	45%
NC_039032.1	Psittacine aviadenovirus B isolate CS15-4016, complete genome	41%

Discussion

Mass die-offs of wild birds aren't entirely uncommon. Many causes have been identified, including industrial waste (Patton et al. 2017), parasites (Patton et al. 2017), bacteria (Wobeser et al. 1997), and viruses (Krone et al. 2018). The Cambridge Bay die-off occurred in the arctic breeding ground, where large masses of birds congregate. It has been

demonstrated that migratory birds can experience large die-offs caused by microbes during large gatheriHTS, like those that occur during migration (Bi et al. 2015). It's also been demonstrated arctic breeding grounds can serve as a breeding ground for new viral subtypes as well the birds that carry them (Global Consortium for H5N8 and Related Influenza Viruses 2016). Given the location of the 2017 Cambridge Bay die-off, infectious causes for the die-off are of particular interest.

Migratory birds, waterfowl in particular, are susceptible to a number of viruses. Influenza is the most infamous of viruses carried by wild waterfowl, because of its impact on human (Guo et al. 2018) and livestock health (Pasick et al. 2015). Wild waterfowl do carry other viruses of importance to human and domestic animal health. Specifically, Canada geese have been identified as a natural reservoir of AMPV, a pathogen of domestic chicken and turkeys. Both influenza (Berhane et al. 2014) and AMPV (Turpin et al. 2008) can be carried subclinically. There are several known viral pathogens that cause illness and death in Canada geese, such as bornaviruses (Murray et al. 2017) and high path influenza (Berhane et al. 2014).

While we have demonstrated the presence of 4 viruses in geese that were involved in the Cambridge Bay die-off, it can't be determined which virus, if any, caused the die-off from this data alone. Two viruses at least can be established as unlikely pathogens. AMPV has been detected in Canada geese in numerous surveys from animals without clinical disease (Bennett et al. 2002, Turpin et al. 2008). Similarly, adeno-associated viruses belonging to the Avian dependoparvovirus 1 species are not pathogenic (Su et al. 2017, Wang et al. 2011). Despite being an unlikely cause of the outbreak, the data presented here still offers valuable new information regarding viruses present in wild birds in Canada. The presence of adeno-associated viruses in wild waterfowl in Canada with high similarity to those circulating in China has been established. This study has also widened the known viral diversity of AMPV

in Canada geese. As mentioned above, AMPV has been isolated from Canada geese by Bennett et al (2005). This previously isolated Canada geese AMPV varied in the glycoprotein sequence with AMPV circulating in domestic turkeys in the US. Bennet et al demonstrated that experimental infections of domestic turkeys with Canada goose AMPV did not produce clinical disease, though replication was demonstrable. The observation that the wild goose AMPV sequenced from Cambridge Bay shares more identity with a domestic Turkey AMPV than with other wild geese APMV, helps solidify the role of migrating Canada geese as the source of seasonal outbreaks of AMPV in domestic fowl (Bennett et al. 2005, Shin et al. 2000).

The pathogenicity of the remaining 2 viruses cannot be established from the data presented here alone. One virus can be confirmed as present in all 3 tested viruses. GCoV-CB17 was detected in the cloacal swabs of all three birds. While the full genome was assembled and analyzed from a single Canada Goose cloacal swab, 63.7% of GCoV-CB17's genome was also detected in the second Canada goose's cloacal swab and 18% was of GCoV-CB17's genome was detected in the snow goose's cloacal swab. This data links one virus to all three birds. However, it should be cautioned that viral metagenomics should not be used for determining the absence of a virus. The factors that affect the efficiency of detection are so numerous and diverse, it's questionable if it's possible to determine a lower limit of detection that would be applicable to more than one particular viral infection and sample type. For example, it's already been established sample type effects the ability to enrich viral DNA and therefore the sensitivity of viral metagenomic detection (Rosseel et al. 2015). In the case of sequence-dependent enrichment, such as sequence capture, the factors affecting the efficiency of detection become even more complicated. For these reasons, viral metagenomic results cannot be treated as proof of the absence of a virus. In the case of the Cambridge Bay die-off, this is particularly true, as the degraded nature of the samples is reflecting in the

fractured nature of the genomes recovered. This is a fact that should be taken into account when considering the adenovirus detected in the Cambridge Bay die-off. Goose adenovirus could not be detected in any sample but the cloacal swab of the single snow goose tested. However, the adeno-associated virus was detected in two samples. This indicates a possibility of adenovirus that was either beyond the lower limit of detection or too degraded to sequence, i.e. in the Canada geese as well as the Snow geese.

What can be established with this data, is the presence of several viruses in two wild geese species that have not been reported before in Canada. First, a new species of coronavirus has been discovered. It has been accepted that wild gammacoronaviruses are the source of domestic gammacoronaviruses pathogens, such as IBV (Woo et al. 2014). However, this work is the first detailed description of a gammacoronavirus from a wild bird which is related to the pathogens of domestic avian coronaviruses. Most importantly, a method (the 4 gene duplication PCR) has been established for the detection and differentiation of this new species. Previous surveys have established that gamacoronaviruses can be detected in healthy geese species (Jonassen et al. 2005) and from healthy geese in the arctic (Muradrasoli et al. 2010). The pathogenicity of this new species in Canada and snow geese would have to be determined experimentally. Similarly, pathogenicity would have also need to be established for goose adenovirus in these species. The adenovirus family have been associated with large die-off in wild birds (Hollmén et al. 2003a,b). This is the first confirmed report of Goose adenovirus in Canada and in these species. While there is evidence of aviadenoviruseses in Canada (Brochu et al. 2019), there have been no reports of goose adenovirus in Canada confirmed with sequence data. There is a report of a goose adenovirus causing disease in domestic geese in Saskatchewan, but no sequence was uploaded to NCBI for comparison. 25% of gosliHTS on a farm succumbed to myocarditis and hepatitis, from which viral inclusion bodies characteristic of adenoviruses were observed (Riddell 1984). As the geese

involve in the Cambridge Bay die-off were juveniles, the adenovirus is also of interest in relation to the mortality of the geese. Given the poor sample quality prevented isolation and full sequencing, further studies are needed. However, this investigation is an important first step.

References

Anthony SJ, Epstein JH, Murray KA, Navarrete-Macias I, Zambrana-Torrelio CM, et al. 2013. A strategy to estimate unknown viral diversity in mammals. *MBio*. 4(5):e00598-13

Atchison RW, Casto BC, Hammon WM. 1965. ADENOVIRUS-ASSOCIATED DEFECTIVE VIRUS PARTICLES. *Science*. 149(3685):754–56

Bande F, Arshad SS, Omar AR, Hair-Bejo M, Mahmuda A, Nair V. 2017. Global distributions and strain diversity of avian infectious bronchitis virus: a review. *Anim. Health Res. Rev.* 18(1):70–83

Banet-Noach C, Simanov L, Perk S. 2005. Characterization of Israeli avian metapneumovirus strains in turkeys and chickens. *Avian Pathol.* 34(3):220–26

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19(5):455–77

Bennett RS, LaRue R, Shaw D, Yu Q, Nagaraja KV, et al. 2005. A wild goose metapneumovirus containing a large attachment glycoprotein is avirulent but immunoprotective in domestic turkeys. *J. Virol.* 79(23):14834–42

Bennett RS, McComb B, Shin HJ, Njenga MK, Nagaraja KV, Halvorson DA. 2002. Detection of avian pneumovirus in wild Canada (Branta canadensis) and blue-winged teal (Anas

discors) geese. *Avian Dis.* 46(4):1025–29

Bentley K, Keep SM, Armesto M, Britton P. 2013. Identification of a noncanonically transcribed subgenomic mRNA of infectious bronchitis virus and other gammacoronaviruses. *J. Virol.* 87(4):2128–36

Berhane Y, Embury-Hyatt C, Leith M, Kehler H, Suderman M, Pasick J. 2014. Pre-exposing Canada Geese (Branta canadensis) to a low-pathogenic H1N1 avian influenza virus protects them against H5N1 HPAI virus challenge. *J. Wildl. Dis.* 50(1):84–97

Bi Y, Zhang Z, Liu W, Yin Y, Hong J, et al. 2015. Highly pathogenic avian influenza A(H5N1) virus struck migratory birds in china in 2015. *Sci. Rep.* 5:12986

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30(15):2114–20

Brochu NM, Guerin MT, Varga C, Lillie BN, Brash ML, Susta L. 2019. A two-year prospective study of small poultry flocks in Ontario, Canada, part 1: prevalence of viral and bacterial pathogens. *J. Vet. Diagn. Invest.*, p. 1040638719843577

Brooks JE, Rainer AC, Parr RL, Woolcock P, Hoerr F, Collisson EW. 2004. Comparisons of envelope through 5B sequences of infectious bronchitis coronaviruses indicates recombination occurs in the envelope and membrane genes. *Virus Res.* 100(2):191–98

Cao J, Wu C-C, Lin TL. 2008. Complete nucleotide sequence of polyprotein gene 1 and genome organization of turkey coronavirus. *Virus Res.* 136(1–2):43–49

Cotmore SF, Agbandje-McKenna M, Canuti M, Chiorini JA, Eis-Hubinger A-M, et al. 2019. ICTV virus taxonomy profile: parvoviridae. *J. Gen. Virol.* 100(3):367–68

Derzsy D. 1967. A viral disease of gosliHTS. I. Epidemiological, clinical, pathological and

aetiological studies. *Acta Vet. Acad. Sci. Hung.* 17(4):443–48

de Groot, R.J 2008. Revision of the family Coronaviridae. *Taxonomic proposal to the ICTV Executive Committee*:1-37.

Dufour D, Mateos-Gomez PA, Enjuanes L, Gallego J, Sola I. 2011. Structure and functional relevance of a transcription-regulating sequence involved in coronavirus discontinuous RNA synthesis. *J. Virol.* 85(10):4963–73

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, et al. 2019. The Pfam protein families database in 2019. *Nucleic Acids Res.* 47(D1):D427–32

Endo A, Nishiura H. 2018. The Role of Migration in Maintaining the Transmission of Avian Influenza in Waterfowl: A Multisite Multispecies Transmission Model along East Asian-Australian Flyway. *Can. J. Infect. Dis. Med. Microbiol.* 2018:3420535

Fabricant J. 1998. The early history of infectious bronchitis. *Avian Dis.* 42(4):648–50

Glávits R, Zolnai A, Szabó E, Ivanics E, Zarka P, et al. 2005. Comparative pathological studies on domestic geese (Anser anser domestica) and Muscovy ducks (Cairina moschata) experimentally infected with parvovirus strains of goose and Muscovy duck origin. *Acta Vet. Hung.* 53(1):73–89

Global Consortium for H5N8 and Related Influenza Viruses. 2016. Role for migratory wild birds in the global spread of avian influenza H5N8. *Science*. 354(6309):213–17

Gough RE, Collins MS, Cox WJ, Chettle NJ. 1988. Experimental infection of turkeys, chickens, ducks, geese, guinea fowl, pheasants and pigeons with turkey rhinotracheitis virus. *Vet. Rec.* 123(2):58–59

Guo L, Hou M, Ning R, Li W, Yang Z, et al. 2018. A family cluster of two fatal cases infected

with influenza A (H7N9) virus in Kunming China, 2017. *Infect. Genet. Evol.* 66:152–58

Hess M. 2017. Commensal or pathogen - a challenge to fulfil Koch's Postulates. *Br. Poult. Sci.* 58(1):1–12

Hodgson T, Britton P, Cavanagh D. 2006. Neither the RNA nor the proteins of open reading frames 3a and 3b of the coronavirus infectious bronchitis virus are essential for replication. *J. Virol.* 80(1):296–305

Hollmén TE, Franson JC, Flint PL, Grand JB, Lanctot RB, et al. 2003a. An adenovirus linked to mortality and disease in long-tailed ducks (Clangula hyemalis) in Alaska. *Avian Dis.* 47(4):1434–40

Hollmén TE, Franson JC, Kilpi M, Docherty DE, Myllys V. 2003b. An adenovirus associated with intestinal impaction and mortality of male common eiders (Somateria mollissima) in the Baltic Sea. *J. Wildl. Dis.* 39(1):114–20

Ivanics E, Palya V, Markos B, Dán A, Ursu K, et al. 2010. Hepatitis and hydropericardium syndrome associated with adenovirus infection in gosliHTS. *Acta Vet. Hung.* 58(1):47–58

Jardine CM, Parmley EJ, Buchanan T, Nituch L, Ojkic D. 2018. Avian metapneumovirus subtype C in Wild Waterfowl in Ontario, Canada. *Transbound. Emerg. Dis.* 65(4):1098–1102

Jonassen CM, Jonassen TO, Grinde B. 1998. A common RNA motif in the 3' end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus. *J. Gen. Virol.* 79 (Pt 4):715–18

Jonassen CM, Kofstad T, Larsen I-L, Løvland A, Handeland K, et al. 2005. Molecular

identification and characterization of novel coronaviruses infecting graylag geese (Anser anser), feral pigeons (Columbia livia) and mallards (Anas platyrhynchos). *J. Gen. Virol.* 86(Pt 6):1597–1607

Jones RC. 2010. Viral respiratory diseases (ILT, aMPV infections, IB): are they ever under control? *Br. Poult. Sci.* 51(1):1–11

Kapgate SS, Kumanan K, Vijayarani K, Barbuddhe SB. 2018. Avian parvovirus: classification, phylogeny, pathogenesis and diagnosis. *Avian Pathol.* 47(6):536–45

Kint J, Langereis MA, Maier HJ, Britton P, van Kuppeveld FJ, et al. 2016. Infectious bronchitis coronavirus limits interferon production by inducing a host shutoff that requires accessory protein 5b. *J. Virol.* 90(16):7519–28

Knight-Jones TJD, Rushton J. 2013. The economic impacts of foot and mouth disease - what are they, how big are they and where do they occur? *Prev. Vet. Med.* 112(3–4):161–73

Krone O, Globig A, Ulrich R, Harder T, Schinköthe J, et al. 2018. White-Tailed Sea Eagle (Haliaeetus albicilla) Die-Off Due to Infection with Highly Pathogenic Avian Influenza Virus, Subtype H5N8, in Germany. *Viruses*. 10(9):

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* 35(6):1547–49

Laconi A, van Beurden SJ, Berends AJ, Krämer-Kühl A, Jansen CA, et al. 2018. Deletion of accessory genes 3a, 3b, 5a or 5b from avian coronavirus infectious bronchitis virus induces an attenuated phenotype both in vitro and in vivo. *J. Gen. Virol.* 99(10):1381–90

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*. 23(21):2947–48

Lin TL, Loa CC, Tsai SC, Wu CC, Bryan TA, et al. 2002. Characterization of turkey coronavirus from turkey poults with acute enteritis. *Vet. Microbiol.* 84(1–2):179–86

Liu DX, Cavanagh D, Green P, Inglis SC. 1991. A polycistronic mRNA specified by the coronavirus infectious bronchitis virus. *Virology*. 184(2):531–44

Lupini C, Cecchinato M, Ricchizzi E, Naylor CJ, Catelli E. 2011. A turkey rhinotracheitis outbreak caused by the environmental spread of a vaccine-derived avian metapneumovirus. *Avian Pathol.* 40(5):525–30

Maharaj SB, Thomson DK, da Graca JV. 1994. Isolation of an avian pneumovirus-like agent from broiler breeder chickens in South Africa. *Vet. Rec.* 134(20):525–26

Mardani K, Noormohammadi AH, Hooper P, Ignjatovic J, Browning GF. 2008. Infectious bronchitis viruses with a novel genomic organization. *J. Virol.* 82(4):2013–24

Mihindukulasuriya KA, Wu G, St Leger J, Nordhausen RW, Wang D. 2008. Identification of a novel coronavirus from a beluga whale by using a panviral microarray. *J. Virol.* 82(10):5084–88

Muradrasoli S, Bálint A, Wahlgren J, Waldenström J, Belák S, et al. 2010. Prevalence and phylogeny of coronaviruses in wild birds from the Bering Strait area (Beringia). *PLoS ONE*. 5(10):e13640

Murray M, Guo J, Tizard I, JenniHTS S, Shivaprasad HL, et al. 2017. Aquatic Bird Bornavirus-Associated Disease in Free-Living Canada Geese (Branta canadensis) in the Northeastern USA. *J. Wildl. Dis.* 53(3):607–11

Naso MF, Tomkowicz B, Perry WL, Strohl WR. 2017. Adeno-Associated Virus (AAV) as a Vector for Gene Therapy. *BioDrugs*. 31(4):317–34

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32(1):268–74

Pasick J, Berhane Y, Joseph T, Bowes V, Hisanaga T, et al. 2015. Reassortant highly pathogenic influenza A H5N2 virus containing gene segments related to Eurasian H5N8 in British Columbia, Canada, 2014. *Sci. Rep.* 5:9484

Patton RT, Goodenough KS, De La Cruz SEW, Nevins H, Cole R, et al. 2017. Mass Mortality Attributed to Acanthocephaliasis at a Gull-billed Tern (Gelochelidon nilotica) Colony in Coastal California, USA. *J. Wildl. Dis.* 53(4):885–90

Riddell C. 1984. Viral hepatitis in domestic geese in Saskatchewan. *Avian Dis.* 28(3):774–82

Rosseel T, Ozhelvaci O, Freimanis G, Van Borm S. 2015. Evaluation of convenient pretreatment protocols for RNA virus metagenomics in serum and tissue samples. *J. Virol. Methods*. 222:72–80

Schachner A, Matos M, Grafl B, Hess M. 2018. Fowl adenovirus-induced diseases and strategies for their control - a review on the current global situation. *Avian Pathol.* 47(2):111–26

Seal BS. 1998. Matrix protein gene nucleotide and predicted amino acid sequence demonstrate that the first US avian pneumovirus isolate is distinct from European strains. *Virus Res.* 58(1–2):45–52

Shin HJ, Njenga MK, McComb B, Halvorson DA, Nagaraja KV. 2000. Avian pneumovirus (APV) RNA from wild and sentinel birds in the United States has genetic homology with RNA from APV isolates from domestic turkeys. *J. Clin. Microbiol.* 38(11):4282–84

Simon-Loriere E, Holmes EC. 2011. Why do RNA viruses recombine? *Nat. Rev. Microbiol.* 9(8):617–26

Su XN, Liu JJ, Zhou QF, Zhang XH, Zhao LC, et al. 2017. Isolation and genetic characterization of a novel adeno-associated virus from Muscovy ducks in China. *Poult. Sci.* 96(11):3867–71

Sugiyama M, Koimaru H, Shiba M, Ono E, Nagata T, Ito T. 2006. Drop of egg production in chickens by experimental infection with an avian metapneumovirus strain PLE8T1 derived from swollen head syndrome and the application to evaluate vaccine. *J. Vet. Med. Sci.* 68(8):783–87

Tausch SH, Renard BY, Nitsche A, Dabrowski PW. 2015. RAMBO-K: Rapid and Sensitive Removal of Background Sequences from Next Generation Sequencing Data. *PLoS ONE*. 10(9):e0137896

Tsang KW, Ho PL, Ooi GC, Yee WK, Wang T, et al. 2003. A cluster of cases of severe acute respiratory syndrome in Hong Kong. *N. Engl. J. Med.* 348(20):1977–85

Tucciarone CM, Andreopoulou M, Franzo G, Prentza Z, Chaligiannis I, Cecchinato M. 2017. First Identification and Molecular Characterization of Avian metapneumovirus Subtype B from Chickens in Greece. *Avian Dis.* 61(3):409–13

Tucciarone CM, Franzo G, Lupini C, Alejo CT, Listorti V, et al. 2018. Avian Metapneumovirus circulation in Italian broiler farms. *Poult. Sci.* 97(2):503–9

Turpin EA, Stallknecht DE, Slemons RD, Zsak L, Swayne DE. 2008. Evidence of avian metapneumovirus subtype C infection of wild birds in Georgia, South Carolina, Arkansas and Ohio, USA. *Avian Pathol.* 37(3):343–51

Wang J, Zhu L, Zhu J, Sun H, Zhu G. 2011. Molecular characterization and phylogenetic analysis of an avian adeno-associated virus originating from a chicken in China. *Arch. Virol.* 156(1):71–77

Wobeser G, Baptiste K, Clark EG, Deyo AW. 1997. Type C botulism in cattle in association with a botulism die-off in waterfowl in Saskatchewan. *Can. Vet. J.* 38(12):782

Woo PCY, Lau SKP, Lam CSF, Lai KKY, Huang Y, et al. 2009. Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus. *J. Virol.* 83(2):908–17

Woo PCY, Lau SKP, Lam CSF, Lau CCY, Tsang AKL, et al. 2012. Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. *J. Virol.* 86(7):3995–4008

Woo PCY, Lau SKP, Lam CSF, Tsang AKL, Hui S-W, et al. 2014. Discovery of a novel bottlenose dolphin coronavirus reveals a distinct species of marine mammal coronavirus in Gammacoronavirus. *J. Virol.* 88(2):1318–31

Wood EN. 1977. An apparently new syndrome of porcine epidemic diarrhoea. *Vet. Rec.* 100(12):243–44

Wylie TN, Wylie KM, Herter BN, Storch GA. 2015. Enhanced virome sequencing using targeted sequence capture. *Genome Res.* 25(12):1910–20

Ziebuhr J, Snijder EJ, Gorbalenya AE. 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J. Gen. Virol.* 81(Pt 4):853–79

Chapter 4: Fecal Virome of Arctic and Red Foxes

Abstract

Metagenomic virome analysis has vastly expanded the known diversity of viral species. This is especially true for viruses that infect wildlife species, for which we have only begun to understand the estimated viral diversity. This study is the first report on the fecal virome of arctic foxes. 11 swabs and fecal samples belonging to arctic and red foxes underwent targeted enrichment and high-throughput sequencing. Sequences belonging to 7 viral families were detected. Homology of the majority of viruses detected suggested these viruses originated from the diet of the foxes, e.g. avian influenza. However, at least one virus (canine kobuvirus) indicates the possibility of transmission of viruses between arctic foxes and other canid species, such as domestic dogs. This study lays the foundation for understanding the diversity of the fecal virome in arctic and red foxes in general and in particular in the Churchill region of Manitoba.

Introduction

The following chapter will describe preliminary results from a project investigating the virome of arctic and red foxes from the Churchill area of northern Manitoba. The previous chapter demonstrates that viral metagenomics and targeted enrichment have a place in modern veterinary diagnostics. This chapter will demonstrate viral metagenomics can also play an important role in basic explorations of viral diversity. These studies fall into a subcategory of metagenomics, referred to as virome analyses. Virome analyses intend to describe the entire viral content of a sample and are often performed on healthy individuals.

Not all viral infections are clinically significant, meaning investigating viral content of healthy animals can reveal diverse viral communities (Ling et al. 2019). In particular, it has been routinely demonstrated that wild animals can carry human and livestock pathogens with little harm to their own health (Jones et al. 2019). Recent outbreaks hightlight the role of wildlife as a reservoir of viral pathogens of humans (Baudel et al. 2019), and of domestic animals (Jori & Etter 2016). Describing the virome of wildlife can therefore help understand and control future outbreaks. In addition to revealing information about potential sources of zoonotic illnesses, virome studies of wildlife also reveal basic information about viral ecology and evolution (Yinda et al. 2019).

The fecal virome of red foxes in Europe have been described in several studies (Lojkić et al. 2016; Bodewes et al. 2013b, 2014b). These studies have revealed the presence of numerous viral families in fox fecal viromes, including but not limited to *Circoviridae, Parvoviridae, Pircornavirales* and *Astroviridae*. The authors are unaware of a previous study describing the fecal virome of red foxes in the arctic or describing the fecal virome of arctic foxes. The following chapter describes the fecal virome of arctic and red foxes from the Churchill area of northern Manitoba. This study was performed in collaboration with Dr. James Roth and Chloes Rodrigues Jr.

Methods

Fecal matter was collected from the carcasses of trapped foxes, donated by hunters. Fecal swabs were collected from live caught foxes in the Churchill area. Samples are described in Table 9. Swabs were stored in viral transport media, fecal matter was stored at -20°C. Total nucleic acid was extracted using DNeasy® Blood & Tissue Kit (Qiagen), following the user made protocol: Purification of viral DNA from animal stool using the

DNeasy® Blood & Tissue Kit (https://www.qiagen.com/ca/resources/resourcedetail?id=5cdd289f-65c5-4228-8852-fe962e2ca3bf&lang=en). In brief**,** swabs were homogenized prior to extraction. Fecal matter was suspended in saline (0.89% NaCl) and vortexed to homogenization. Centrifugation at 10,000 RPM for 10 minutes was used to clarify samples. Samples were then syringe-filtered using a 0.4 micron filter. The DNeasy Blood & Tissue Kit extraction kit (Qiagen) was then used according to the manufacturer instructions. cDNA synthesis, library sequence preparation and sequence capture were then performed as described in chapter 3. The samples were sequenced on a V2 flow cell and a 300 cycle paired-end cartridge. Read quality was assessed using FastQC and trimming was performed using Trimmomatic (Version 0.36), according to read quality (Bolger et al. 2014). Kraken was then used to filter bacteria reads (Wood & Salzberg 2014). *de novo* assembly was performed using SPAdes (v3.12.0) (Bankevich et al. 2012). SPAdes contigs were classified using Blastn (Altschul et al. 1990) and DIAMOND (Buchfink et al. 2015). Contigs identified as viral were then used in combination with reference sequences to sort viral reads with RAMBO-K (Tausch et al. 2015). The domestic dog genome was used as a host reference as there was no fox genome available at the time this work was carried out. Filtered viral reads were then reassembled with SPAdes and reassessed with BLASTn and DIAMOND. Viral contigs larger than 1 kb were then used as a reference for assembly in Geneious using the lowest sensitivity and 10 iterations. Assemblies were visually inspected and annotated in Geneious. IRMA was used to calculated hemagglutinin and neuraminidase avian influenza percentages (Shepard et al. 2016).

Table 9: Fecal swabs and fecal sample from red and arctic in the Churchill Area fox summary

Sample number	**Sample Type**	**Species**	**Sex**	**Number of reads**
Fox 07	Feces	Arctic Fox	N/A	1.93E+06
Fox 11	Swab suspension	Red fox	Male	3.37E+06
Fox 17	Feces	Arctic Fox	N/A	1.01E+06
Fox 18	Feces	Arctic Fox	N/A	8.58E+05
Fox 29	Feces	Arctic Fox	N/A	1.65E+06
Fox 53	Swab suspension	Arctic Fox	Male	1.41E+06
Fox 54	Swab suspension	Arctic Fox	Female	1.70E+06
Fox 55	Swab suspension	Arctic Fox	Female	5.97E+06
Fox 63	Feces	Arctic Fox	N/A	1.13E+06
Fox 76	Swab suspension	Arctic Fox	Female	3.89E+06
Fox 96	Swab suspension	Arctic Fox	Male	1.48E+06

Overview of viruses detected

Sequence belonging to 7 viral families were detected (Table 10). The most abundant viral species detected was avian influenza (AIV). AIV was found in 4 of 12 foxes, a range of AIV subtypes were detected in each AIV positive sample. Circovirus was also detected in four arctic foxes. Parvovirus was detected in 3 arctic foxes. All of the circovirus and parvovirus sequence detected shared low homology to recognized species. One full genome of canine kobuvirus (*Picornaviridae*) was detected in an arctic fox. A second partial genome of a kobuvirus was detected in a separate arctic fox. Trace amounts of picornavirales, annelovirus and astrovirus were also detected. However,

contigs were small and depth of coverage was low. Viruses with significant contigs/genomes are discussed in depth below.

Table 10: Viruses detected in Arctic and Red fox fecal virome.

Fox	Family	Genus	Contig length	Depth of Coverage	top Blastx hit (NCBI Accession)	% identity (query cover)	Complete Genome
Fox 07	Circoviridae	UC	2079	25	replication-associated protein [Golden silk orbweaver associated circular virus 1] (AXL65901.1)	60(43)	Yes
Fox 07	Circoviridae	UC	1773	10	Rep [Rodent circovirus] (ATP66719.1)	73(49)	Yes
Fox 07	Parvoviridae	Dependo-parvovirus	2923	131	VP1 [Murine adeno-associated virus 1] (AWB14638.1)	61(68)	No
Fox 07	Picornavirales	unclassified	3212	31	nonstructural polyprotein [Fesa-like virus] (AWU65874.1)	55(49)	No
Fox 07	Picornavirales	unclassified	2764	9	nonstructural polyprotein [Fesa-like virus] (AWU65874.1)	38(85)	No

Fox 17	Parvoviridae	Protoparvo-virus	2783	117	NS1 [Tusavirus] (AIT18928.1)	50(60)	No
Fox 17	Astroviridae	unclassified	1200	5.8	capsid protein [Marmot astrovirus 2] (AVX29489.1)	48(77)	No
Fox 17	Circoviridae	unclassified	4130	197	Rep [Rodent circovirus] (ATP66707.1)	48(21)	No
Fox 17	Anelloviridae	unclassified	1903	4.2	ORF1 [Torque teno canis virus] (ASV72278.1)	55(67)	No
Fox 29	Parvoviridae	Protoparvo-virus	2743	35	NS1 [Tusavirus] (AIT18928.1)	51(59)	No
Fox 29	Parvoviridae	Protoparvo-virus	1633	8	capsid protein [Parvovirus fur seal/ATROP40/BR/2012] (AKI82154.1)	48(84)	No
Fox 29	Picornaviridae	Kobuvirus	8358	2972	polyprotein [Canine kobuvirus US-PC0082] AEO19724.1	99(88)	Yes
Fox 55	Circoviridae	unclassified	1187	5	Rep [Bat circovirus] (AIF76268.)	88(39)	Yes
Fox 63	Picornaviridae	Kobuvirus	3664	119	polyprotein [Canine kobuvirus US-PC0082] (AEO19724.1)	89(81)	No

Fox 63	Picornaviridae	Kobuvirus	4250	126	polyprotein [Canine kobuvirus] (AGH29114)	96(99)	No

UC = Unclassified

Influenza

Influenza viruses belong to the family *Orthomyxoviridae*. Their genomes are negative-sense RNA and consist of 8 segments. There are four 4 genera of influenza viruses: *Influenzavirus A, Influenzavirus B, Influenzavirus C and Influenzavirus D.* The first three genera can cause infections in humans. Influenzavirus A and B cause most human influenza infections, while influenzavirus C infections in humans are rare and mild (Webster & Govorkova 2014). Influenza B viruses are only known to infect humans and seals (Bodewes et al. 2013a). The host range and dynamics of Avian Influenza virus (AIV) are more complicated. The hemagglutinin (H) and neuraminidase (N) proteins of influenza viruses are used to subtype the virus, as they are important for cell attachment and therefore host range. AIV has 16 H subtypes, which originated in avian host species. Three H subtypes have caused pandemics in humans; HINI was the cause of the Spanish flu pandemic in 1918-1920, H2N2 was the cause of the Asian flu epidemic in 1957-1958, H3N2 caused the Hong Kong flu pandemic in 1968-1969 and H1N1 caused the recent swine flu pandemic in 2009-2010. The Spanish flu was the largest of these pandemics, causing 50 million deaths. Five other HA types (H5, H6, H7, H9, and H10) are able to infect humans. Two HA subtypes (H3 and H7) can infect horses and five H subtypes infect swine (H1, H2, H3, H4, H5, H7, and N9) (Webster & Govorkova 2014). Canines are also susceptible to several AIV subtypes. H3N8 is known to have been

transmitted to canines from horses (Crawford et al. 2005). Canines have been shown to be susceptible to H5N1 (Maas et al. 2007) and H3N2 (Voorhees et al. 2018).

While it is apparent the potential host range of AIV is large, for the vast majority of virus subtypes waterfowl are the natural hosts (Horimoto & Kawaoka 2001). It is therefore not surprising that arctic foxes which feed on waterfowl would contain AIV RNA in their feces. All fox samples in which AIV was detected were mixtures of many subtypes. This is a particularly challenging situation to deal with bioinformatically. Most HTS assemblers were designed for organisms with higher fidelity polymerases. RNA viruses, AIV in particular, introduce mutations in every replication cycle leading to high viral population diversity. In addition, the segmented genomes of AIV lends themselves to reassortment events (Westgeest et al. 2014). These issues, in addition to samples which contain multiple viral subtypes, can lead to many errors in assemblies such as chimeric contigs. The fox AIV sequence data is particularly difficult because it isn't a case of coinfection of two virus types, but rather a sampling of many viruses from the foxes' diet. SPAdes failed to produce accurate assemblies (segments of the correct length and containing intact ORFs), even following read filtering. For this reason, IRMA was used to assess the influenza content of the arctic fox fecal virome. IRMA was designed to analyze viral HTS data. There is an IRMA module available for influenza data (Shepard et al. 2016). This module was used to estimate the H and N type diversity in 4 arctic fox fecal viromes. IRMA outputs read counts by H and N types, these were used to construct Figures 11 and 12.

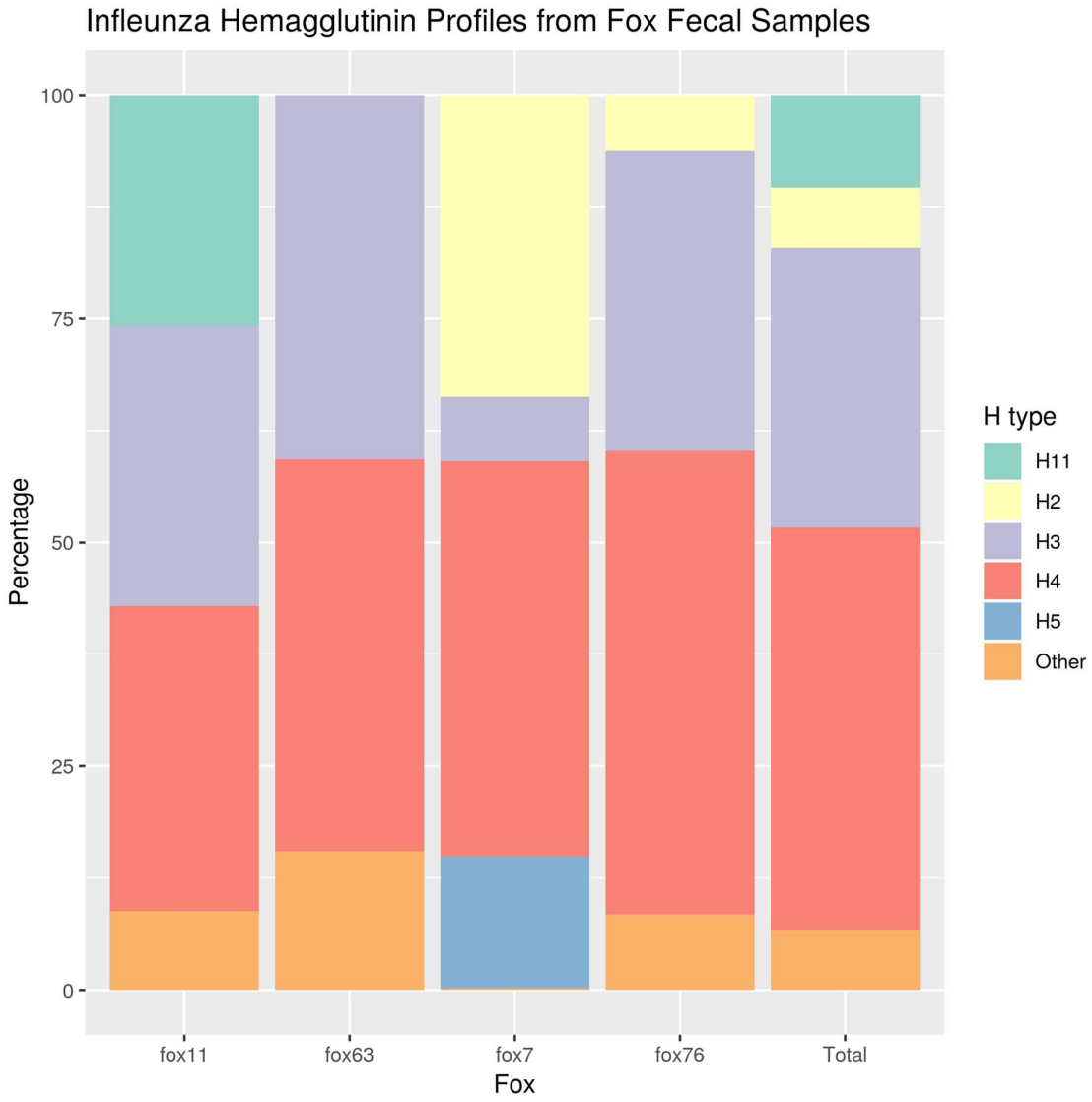

Figure 11: Avian influenza hemagglutinin type distribution in 4 arctic fox fecal samples, estimated with IRMA (Shepard et al. 2016).

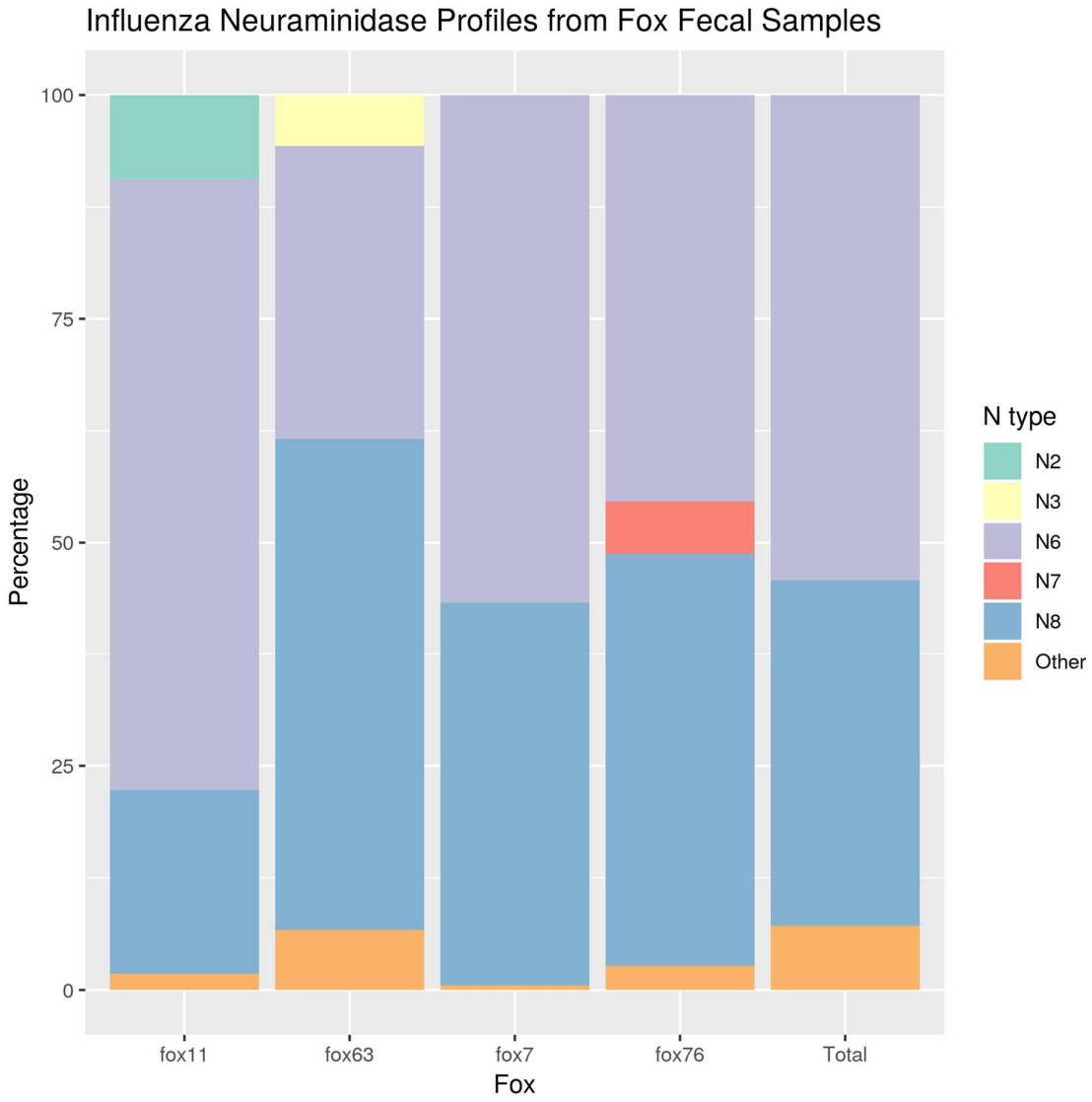

Figure 12: Avian influenza neuraminidase type distribution in 4 arctic fox fecal samples, estimated with IRMA (Shepard et al. 2016).

More variance was observed in the H types than N types. Interestingly, all samples contain H3 and N8, one of the known influenza subtypes which can infect domestic dogs. To the authors knowledge, it has not been demonstrated that red or arctic foxes are susceptible to this strain of canine influenza. Seroprevalence studies in the past have failed to demonstrate canine influenza infection in free ranging foxes (DiGeronimo et al. 2019) (Lempp et al. 2017).

A single fox contained reads assigned as H5. The H5N1 AIV subtype has been shown to be highly pathogenic in a number of species belonging to the Order Carnivora. Carnivores can contract H5N1 through feeding on infected carcases (Keawcharoen et al. 2004), and possibly by transmission from one carnivore to another (Thanawongnuwech et al. 2005). Transmission of H5N1 from prey to carnivores has been shown to occur in the wild (Klopfleisch et al. 2007). In particular, studies have shown red foxes are susceptible to H5N1 and can contract the virus by feeding on infected carcases (Reperant et al. 2008). However, no N1 type was detected.

The most likely scenario is that all AIV detected in this study is from the foxes' diet. All fecal samples contained feathers and observations from researchers who collected these samples confirm the foxes were consuming a large number of birds at the time the samples were collected. The diet of arctic foxes in northern Manitoba can consist heavily of migratory birds during seasonal migration (Roth 2002). The detection of viruses originating from diet is common in fecal virome studies (Li et al. 2011c; Donaldson et al. 2010). Previous studies of avian influenza in the Canadian Arctic have shown H3 to be a common subtype in arctic geese (Liberda et al. 2017). While the AIV detected in this study has likely originated from diet, we have demonstrated the foxes are coming into contact with numerous AIV subtypes that could infect carnivore and canid species. Seroprevalence studies in arctic foxes for antibodies against AIV could be performed out to determine if the foxes are being actively infected by AIV. As well, the fact that foxes are coming into contact with several AIV subtypes should be taken into consideration for the biosecurity of both hunters and researchers (Root et al. 2015).

Kobuviruses

Kobuvirus is a relatively new genus in the large viral family *Picornaviridae*. Currently this genus contains six species, which infect a range of mammalian species. Three *Kobuvirus* species are known to infect domestic animals; Aichivirus B infects cattle and sheep, while Aichivirus C infects pigs (Li et al. 2011a) and Aichivirus D has also been isolated from cattle. Aichivirus E was isolated from rabbits (Pankovics et al. 2016). Aichivirus F, the latest discovery in the genus, was discovered in bats (Wu et al. 2016). Aichivirus A was the first kobuvirus species discovered, and contains three subtypes. Aichi Virus 1, the first subtype, was isolated from humans with gastroenteritis in Japan (Yamashita et al. 1993). It has since been studied in other parts of Asia (Yamashita et al. 1995), as well as Europe and South America (Oh et al. 2006). Canine kobuvirus 1 and murine kobuvirus 1 are the other two subtypes in the Aichivirus A species.

Members of the *Kobuvirus* genus contain positive sense RNA genomes of approximately 8.2 to 8.4 kb in length. The genome is dominated by a single polyprotein of approximately 2400 amino acids in length. This polyprotein is cleaved by viral proteases into the virus's 3 structural proteins and 8 nonstructural proteins (Reuter et al. 2011). The full genome of a kobuvirus was assembled from the reads of arctic fox29. Coverage of the genome was very high (2972x). This genome was 8285 bp in length and contained an ORF 2445 amino acids in length, which encoded the viral polyprotein. Fox63 also contained SPAdes contigs identified by BLAST as belonging to the *Kobuvirus* genus. However, the complete genome could not be assembled. Two contigs were assembled by SPAdes; one contig was 3664 bp long and had a sequencing coverage of 119x and a second contig of 4250 bp and coverage of 126 times. When aligned to the most closely related full genome from NCBI these two contigs reveal three gaps in coverage across the genome. 34 bp was missing from the 5' end of the genome, a gap of

approximately 210 bp was observed starting at position 3628, and 280 bp were also missing preceding the poly(A) tail. Therefore these contigs represent approximately 94% of the kobuvirus genome.

The kobuviruses detected in this study both share highest homology (>85% pairwise nucleotide identity) to canine kobuviruses, suggesting both viruses belong to the canine kobuvirus genotype of Aichivirus 1. This similarity falls well above the threshold set by the ICTV for species demarcation (Adams et al. 2013). The tree in Figure 13 demonstrates that the kobuvirus from fox29 clusters reliably with a canine kobuvirus isolated from a domestic dog in the United States. The kobuvirus from fox 63 is an outlier from other known canine kobuviruses. This is interesting as it indicates there are multiple subtypes of canine kobuvirus circulating in arctic foxes, some of which share high homology (94% nucleotide identity) to viruses isolated from domestic dogs in North America and a second virus that may represent a unique subtype.

This isn't the first report of canine kobuvirus in a fox species, although it is the first report in arctic foxes. A fecal virome study of red foxes from Spain revealed a small portion (506 bp) of a canine kobuvirus. It shared a high similarity (97% nt) to kobuviruses from diarrhetic domestic dogs in Italy (Bodewes et al. 2014b). This 506 bp sequence shared 91% similarity to fox63's kobuvirus and and 95% identity to fox29's kobuvirus. Canine kobuvirus has also been detected in several wild canine species, including spotted hyenas (Olarte-Castillo et al. 2015) and wolves (Melegari et al. 2018). The high similarity of fox29's kobuvirus to american domestic dog kobuviruses, indicates the possibility of cross species transmission. Both kobuviruses were detected in arctic foxes, but previous studies, showing red foxes are susceptible as well, raise the question of red fox to arctic fox kobuvirus transmission. These two species co-exist and have been shown to exchange viruses belonging to other families (Balboni et al. 2019).

To date, disease associated with kobuvirus in the wild carnivores hasn't been reported. In both cases of kobuvirus detection in red foxes, the foxes were healthy (Di Martino et al. 2014; Olarte-Castillo et al. 2015). Canine kobuviruses has been associated with diarrhea in domestic dogs, but has also been isolated from healthy dogs (Li et al. 2011a). In fact, there is no significant correlation between diarrhetic domestic dogs and kobuvirus detection, meaning kobuviruses are detected equally in diarrhetic and healthy dogs (Soma et al. 2016). Seroprevalence studies could be undertaken to examine the prevalence of kobuvirus in these two species as well as domestic dogs in the arctic. Further studies could investigate the possibility of these wild species as reservoirs for domestic dog pathogens, and vice versa.

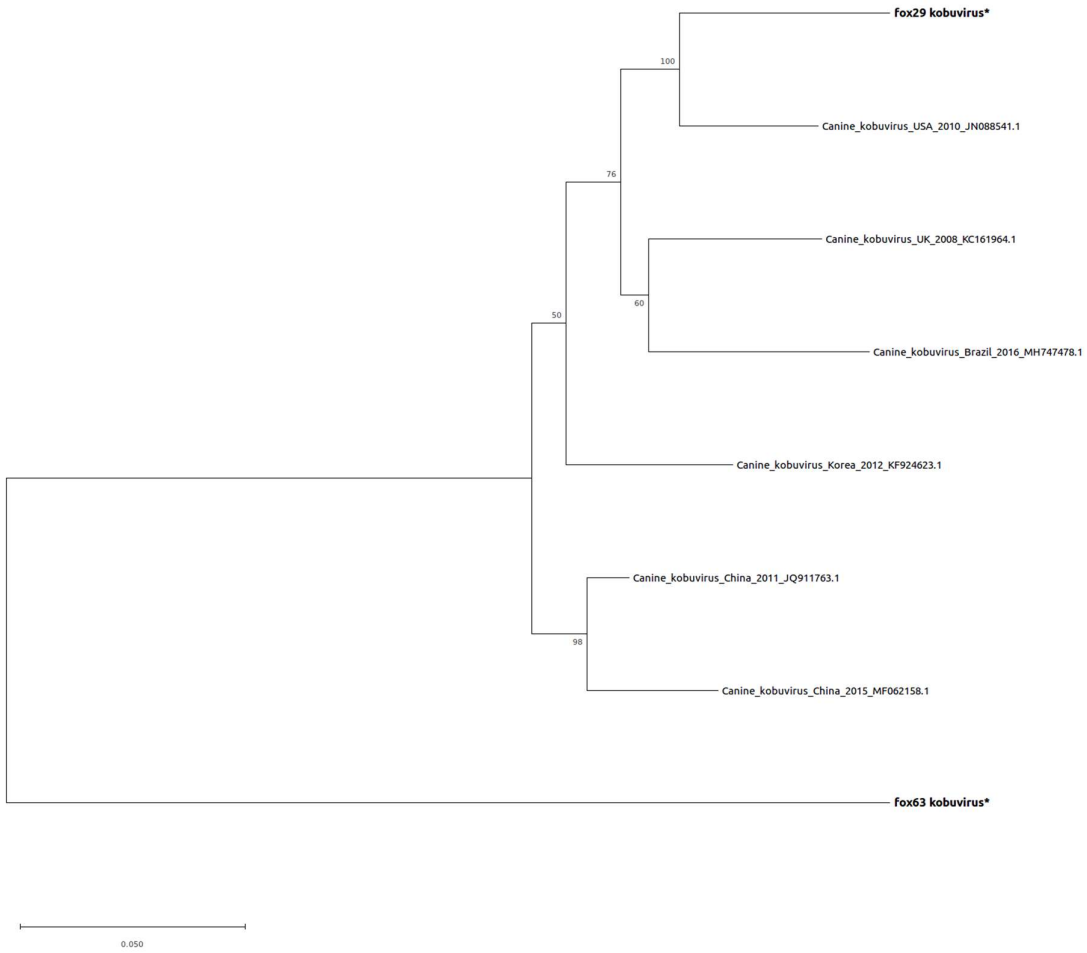

Figure 13: Maximum likelihood tree constructed using the full genomes of canine kobuviruses. The tree was constructed using IQ-TREE (Nguyen et al. 2015), with the substitution model GTR+F+I+G4 and 1000 ultra fast bootstraps.

Circoviruses

Circoviridae genomes range from 1.8 kb to 2.1 kb in length. Their genomes are circular single-stranded DNA. *Circoviridae* genomes are identifiable by the presence of two proteins, a replication-associated protein (often referred to as the replicase) and a capsid protein. This family contains notable pathogens of livestock, such as beak and feather disease virus (Hakimuddin et al. 2016) and porcine circovirus 2 (Xu et al. 2019).

Circoviridae have recently undergone a large taxonomic change. This is largely due to the number of *Circoviridae* species discovered with the advent of metagenomic sequencing. Metagenomic investigations have found these viruses to be particularly ubiquitous in fecal viromes (Shulman & Davidson 2017). It's therefore unsurprising, previous virome studies in red foxes have revealed circoviruses (Bodewes et al. 2013b).

As mentioned above, the taxonomy of *Circoviridae* has been revisited since the advent of metagenomics. *Circoviridae* species are classified into two genera: *Circovirus* and *Cyclovirus,* the former being a new addition to the family (Rosario et al. 2017). The *Cyclovirus* genus was discovered in various hosts, such as humans, chimpanzees (Li et al. 2010), bats (Ge et al. 2011), cows, goats (Li et al. 2011b), horses (Li et al. 2015), squirrels (Sato et al. 2015). One major distinction between cycloviruses and classic circoviruses (which are now classified in the *Circovirus* genus) is the discovery of cyclovirus in insect hosts, such as dragonflies (Dayaram et al. 2013) and cockroaches (Padilla-Rodriguez et al. 2013). While a good majority of cycloviruses have been detected in feces (Li et al. 2010), some cycloviruses have also been detected in cerebrospinal fluid (Tan et al. 2013) and respiratory secretions (Phan et al. 2014a).

Circoviridae and *Circoviridae*-like contigs were detected in 4 foxes in this study (fox07, fox17, fox29 and fox55). Contigs in this study were identified as circovirus-like due to the fact they contain replicases similar to those of the *Circoviridae* family. Figure 14 presents a maximum likelihood tree built using the amino acid sequence of these replicase proteins. With the exception of the fox7 circovirus 2, all circoviruses detected in this share little to no homology to classified circoviruses at the nucleotide level. It is only at the amino acid level that contigs could be identified as *Circoviridae*-like due to the presence of replicases. This is why DIAMOND is needed in addition to Blastn to identify viral contigs.

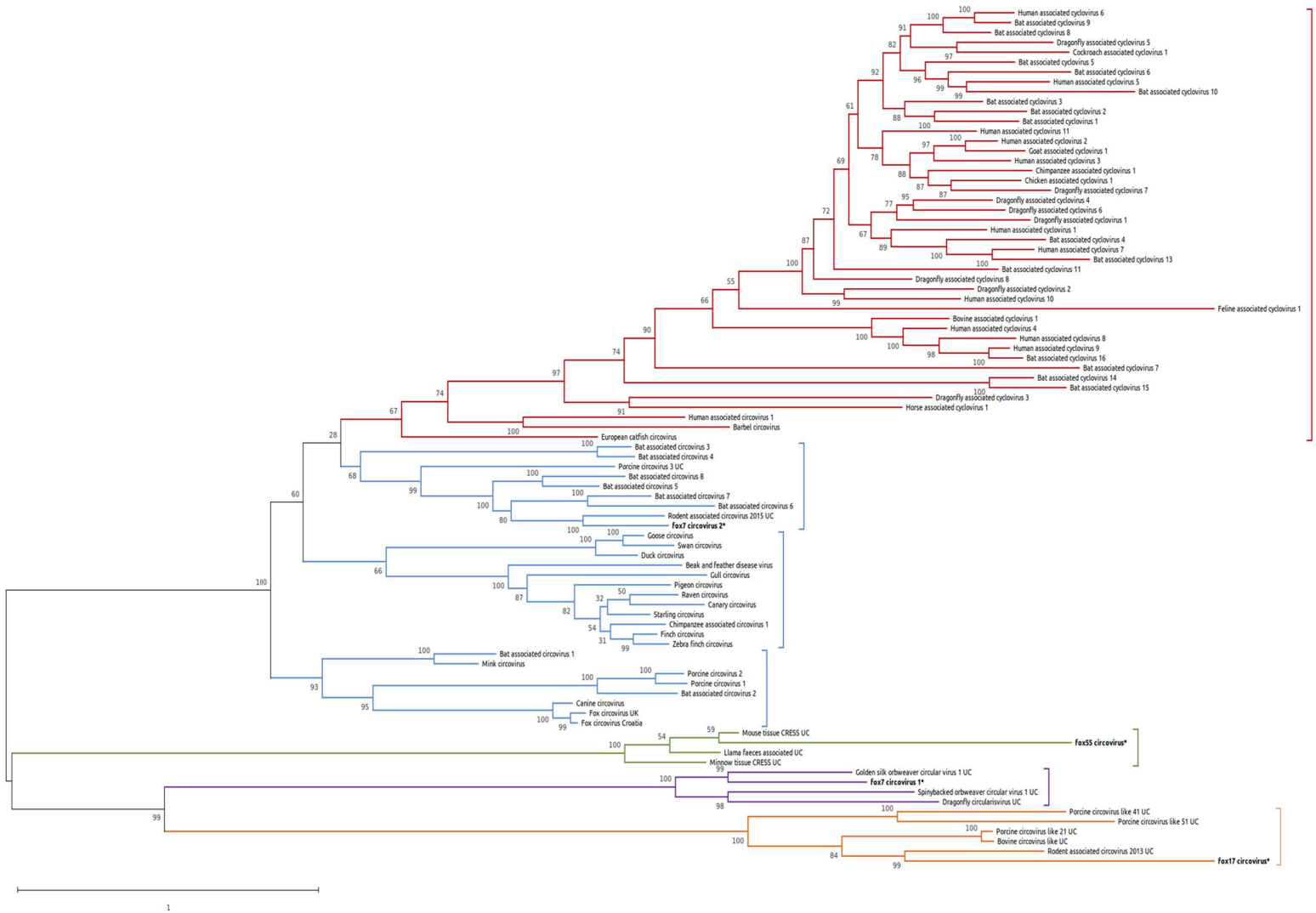
Human associated cyclovirus 6
Bat associated cyclovirus 9
Bat associated cyclovirus 8
Dragonfly associated cyclovirus 5
Cockroach associated cyclovirus 1
Bat associated cyclovirus 5
Bat associated cyclovirus 6
Human associated cyclovirus 5
Bat associated cyclovirus 10
Bat associated cyclovirus 3
Bat associated cyclovirus 2
Bat associated cyclovirus 1
Human associated cyclovirus 11
Human associated cyclovirus 2
Goat associated cyclovirus 1
Human associated cyclovirus 3
Chimpanzee associated cyclovirus 1
Chicken associated cyclovirus 1
Dragonfly associated cyclovirus 7
Dragonfly associated cyclovirus 4
Dragonfly associated cyclovirus 6
Dragonfly associated cyclovirus 1
Human associated cyclovirus 1
Bat associated cyclovirus 4
Human associated cyclovirus 7
Bat associated cyclovirus 13
Bat associated cyclovirus 11
Dragonfly associated cyclovirus 8
Dragonfly associated cyclovirus 2
Human associated cyclovirus 10
Feline associated cyclovirus 1
Bovine associated cyclovirus 1
Human associated cyclovirus 4
Human associated cyclovirus 8
Human associated cyclovirus 9
Bat associated cyclovirus 16
Bat associated cyclovirus 7
Bat associated cyclovirus 14
Bat associated cyclovirus 15
Dragonfly associated cyclovirus 3
Horse associated cyclovirus 1
Human associated circovirus 1
Barbel circovirus
European catfish circovirus
Bat associated circovirus 3
Bat associated circovirus 4
Porcine circovirus 3 UC
Bat associated circovirus 8
Bat associated circovirus 5
Bat associated circovirus 7
Bat associated circovirus 6
Rodent associated circovirus 2015 UC
fox7 circovirus 2*
Goose circovirus
Swan circovirus
Duck circovirus
Beak and feather disease virus
Gull circovirus
Pigeon circovirus
Raven circovirus
Canary circovirus
Starling circovirus
Chimpanzee associated circovirus 1
Finch circovirus
Zebra finch circovirus
Bat associated circovirus 1
Mink circovirus
Porcine circovirus 2
Porcine circovirus 1
Bat associated circovirus 2
Canine circovirus
Fox circovirus UK
Fox circovirus Croatia
Mouse tissue CRESS UC
fox55 circovirus*
Llama faeces associated UC
Minnow tissue CRESS UC
Golden silk orbweaver circular virus 1 UC
fox7 circovirus 1*
Spinybacked orbweaver circular virus 1 UC
Dragonfly circulariscirvirus UC
Porcine circovirus like 41 UC
Porcine circovirus like 51 UC
Porcine circovirus like 21 UC
Bovine circovirus like UC
Rodent associated circovirus 2013 UC
fox17 circovirus*
1

Figure 14: Maximum likelihood tree constructed using the amino acid sequence of the replicase proteins of classified and unclassified circoviruses. The tree was constructed using IQ-TREE (Nguyen et al. 2015), with the substitution model LG+R5 and 1000 ultra fast bootstraps. UC demarcates unclassified species.

Only one of the 5 *Circoviridae*-like contigs detected in arctic foxes clusters with classified *Circoviridae* species. Circovirus 2 from fox7 clusters together with rodent circoviruses in the *Circovirus* genus. The replication protein is congruent in size (~300 amino acids) with other replicases of the *Circovirus* genus. The replicase of this virus shares 72% amino acid pairwise identity to the replicase of a rodent circovirus. This indicates that this virus might have originated from rodents in the fox's diet. The entire genome shares 59% nucleotide pairwise identity to a rodent circovirus. This falls far below the new species demarcation rule of 80%, but above the threshold of 55% to be considered a member of the family. The second criteria for taxonomic assignment in *Circoviridae* of genome organization could not be determined for circovirus 2 from fox7. While a complete replicase for this virus could be identified, a capsid-like protein could not be (Figure 15). It is therefore difficult to determine if this should be classified as a member of the Circoviridae family, given that inclusion of members of this family are also based on gene orientation and spacing (Rosario et al. 2017). The inability to elucidate a capsid protein in fox7 circovirus 2, could be due to low amino acid similarity or do to misassembly due to the low coverage of this genome (10x).

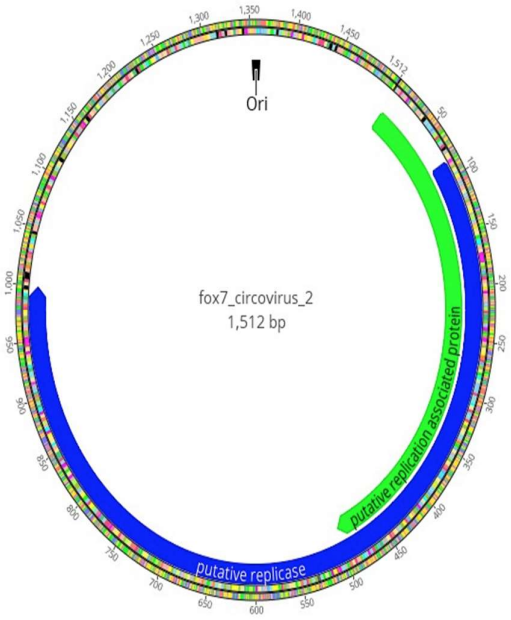

Figure 15: Genome organization of a circovirus detected in the fecal matter of an arctic fox.

The remaining 4 circovirus-like viruses cluster with unclassified circoviruses. These unclassified viruses have been given the designation of CRESS (Circular Rep-encoding ssDNA viruses). Rosario et al have proposed that viruses sharing similarity to *Circoviridae* but sharing less than 55% genome-wide pairwise identity should be classified as CRESS. As these viruses are studied further they may require the designation of a new family or genus (Rosario et al. 2017). Still, the phylogeny of these unclassified viruses is interesting, as they form three distinct clades.

The clade in green, in Figure 14, not only cluster together but share similar genome organization (Figure 16). Circoviridae viruses are ambisense, meaning the capsid and replicase ORFs code in opposite orientations. Genomes in the green clade, which contains arctic fox55 circovirus, code replicase and capsid ORFs in the same orientation. This fact reinforces that they likely do not belong to the Circoviridae family. Fox55 circovirus genome is small (1187 bp) and sequencing coverage was low (5x).

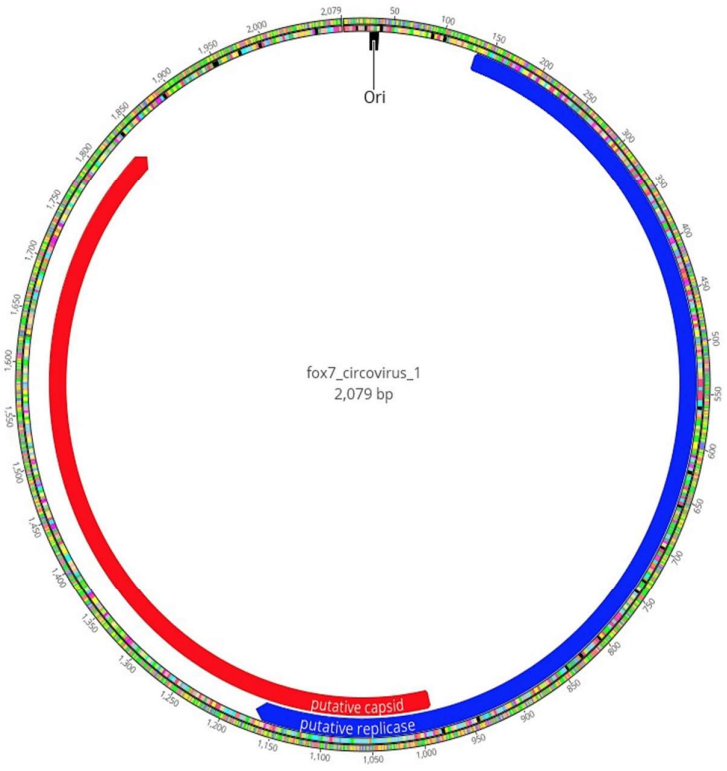

Figure 16: Genome organization of a circovirus detected in the fecal matter of arctic fox55.

The clade in orange of Figure 14 is made entirely of insect associated CRESS. One of the arctic fox7 circoviruses clusters in this group. The full genome of fox7 circovirus 1 was assembled, at moderate coverage (25x). The genome is 2079 bp in length and

contained a complete replicase and capsid gene. As arctic foxes are hosts to a number of insect parasites, it is possible that this virus originates from a parasite (Skírnisson et al. 1993, Stien et al. 2010) . This clades also lacks the ambisense genome organization of true circoviruses (Figure 17). A third CRESS clade is formed with porcine and bovine like-circoviruses (in purple). Fox17 circovirus clusters in this clade. A complete genome could not be recovered for this virus, despite relatively high coverage (197).

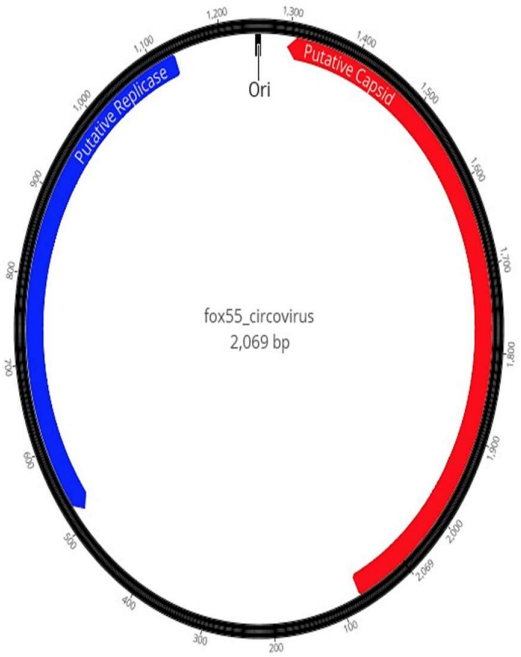

Figure 17: Genome organization of a circovirus detected in the fecal matter of a arctic fox17.

It isn't currently possible to determine a host for CRESS viruses, meaning these viruses could infect a parasite, a prey item or the animal whose feces they have been detected. Previous studies have speculated that some of these viruses infect parasitic protozoans (Shan et al. 2011). Others have attempted to infer pathogenicity through presence in diseased animals and absence in healthy animals (Guo et al. 2018). Further studies are needed to determine characteristics beyond sequence information for these unclassified viruses, specifically isolation of virus. For the true circoviridae species detected in this study, it is possible to infer possible host identity through similarity to known viruses. There are known circovirus pathogens of fox species (Bexton et al. 2015) and these pathogens can be transmitted among carnivore species (Yang et al. 2018). However, this is another case where the more likely explanation is that the virus is of prey origin, given the similarity of fox 7 circovirus 2 to rodent circoviruses and the diet of foxes, which can consists heavily of lemmings.

Parvoviruses

Viruses belonging to the *Parvoviridae* family have single-stranded DNA genomes. Parvoviridae genomes are linear and 4-6 kb in length. Similar to *Circoviridae*, the genomes code a replicase and capsid (or virion) protein. In addition, they encode genus-specific accessory proteins (Cotmore et al. 2019). The family contains two subfamilies; *Parvovirinae*, which infect vertebrate hosts, and *Densovirinae*, which infect invertebrates. There are currently eight genera in the *Parvovirina* subfamily (Cotmore et al. 2014). In this study, contigs identified belonged to the protoparvovirus and dependoparvovirus genera.

The *Protoparvirus* genus includes the species Carnivore protoparvirus 1. Carnivore protoparvovirus 1 is the species designation given to a group of viruses including, feline panleukopenia virus (FPV), mink enteritis virus (MEV), canine parvovirus (CPV) and

racoon parvovirus (Cotmore et al. 2019). FPV virus was the first discovered virus in this species. It causes a highly contagious and often lethal disease in felids. While the symptoms of FPV related illness can be devastating (anorexia, vomiting, diarrhea, neutropenia and lymphopenia), FPV can also be isolated from healthy cats (Stuetzer & Hartmann 2014). MEV emerged in minks in 1952, originally it was indistinguishable from FPV. MEV causes highly contagious enteritis in minks (Wills 1952). CPV emerged in the 1970s, causing highly contagious enteritis in domestic dogs (Appel et al. 1979). Due to their genetic similarity, these carnivore parvoviruses have been classified into the species designation Carnviore protoparvovirus 1 (Cotmore et al. 2019). Carnivore protoparvovirus 1 has also been demonstrated to infect other carnivore species, including skunks (Barker et al. 1983), arctic foxes and red foxes (Tryland et al, 2018). The *Protoparvovirus* genus also includes several clades of human parvoviruses. Bufaviruses were discovered in 2012 (Phan et al. 2012) and Tusaviruses in 2014 (Phan et al. 2014b). Both species were isolated from children with diarrhea in Africa. Both viruses share high homology with animal parvoviruses (Väisänen et al. 2017).

Two of the parvovirus contigs detected in the arctic fox fecal virome in this study belonged to the *Protoparvovirus* genus. Fox17 contained a single *Protoparvovirus* contig with sufficient coverage (117x). This contig contained a complete polymerase ORF. Two *Protoparvovirus* contigs were assembled from the reads of fox29, one of which contained a partial polymerase gene and the other contained a partial capsid gene. The contigs do not overlap. The protoparvoviruses from fox29 and fox7 are more similar (98% nucleotide similarity) to each other than to classified species. Despite evidence that Carnivore protoparvirus 1 infects wild foxes (Truyen et al. 1998), neither *Protoparvovirus* detected in this study do not belong to the Carnivore protoparvirus 1 species. Figure 18 is a maximum likelihood tree constructed from an alignment of

Protoparvovirus polymerase proteins. 4 distinct clades are formed, the largest of which includes FPV, MEV, and CPV. 2 clades are formed by human protoparvoviruses, bufavirus and tusavirus. The tree demonstrates that protoparvoviruses detected in this cluster with unclassified parvoviruses detected in seals. This is not too surprising considering Arctic foxes do feed on seal carcases from polar bear kills (Roth 2003). However, the seal viruses clustering with the arctic fox protoparvoviruses from this study were detected in fur seals from Antarctica. This grouping holds true for the capsid proteins as well (Figure 19). There are known parvoviruses which seals are susceptible to, but these infections are caused by viruses belonging to a different genus (Bodewes et al. 2014a). Limited virome analysis has been performed on seals (Kluge et al. 2016). There is also evidence in the literature that seal viruses can be transmitted long distances (Kennedy et al. 1989) and that viruses can be transmitted between dogs and seals (Barrett et al. 1992). Though the amino acid similarity is very low, this is still a very interesting finding, as it could indicate a marine group of protoparvoviruses.

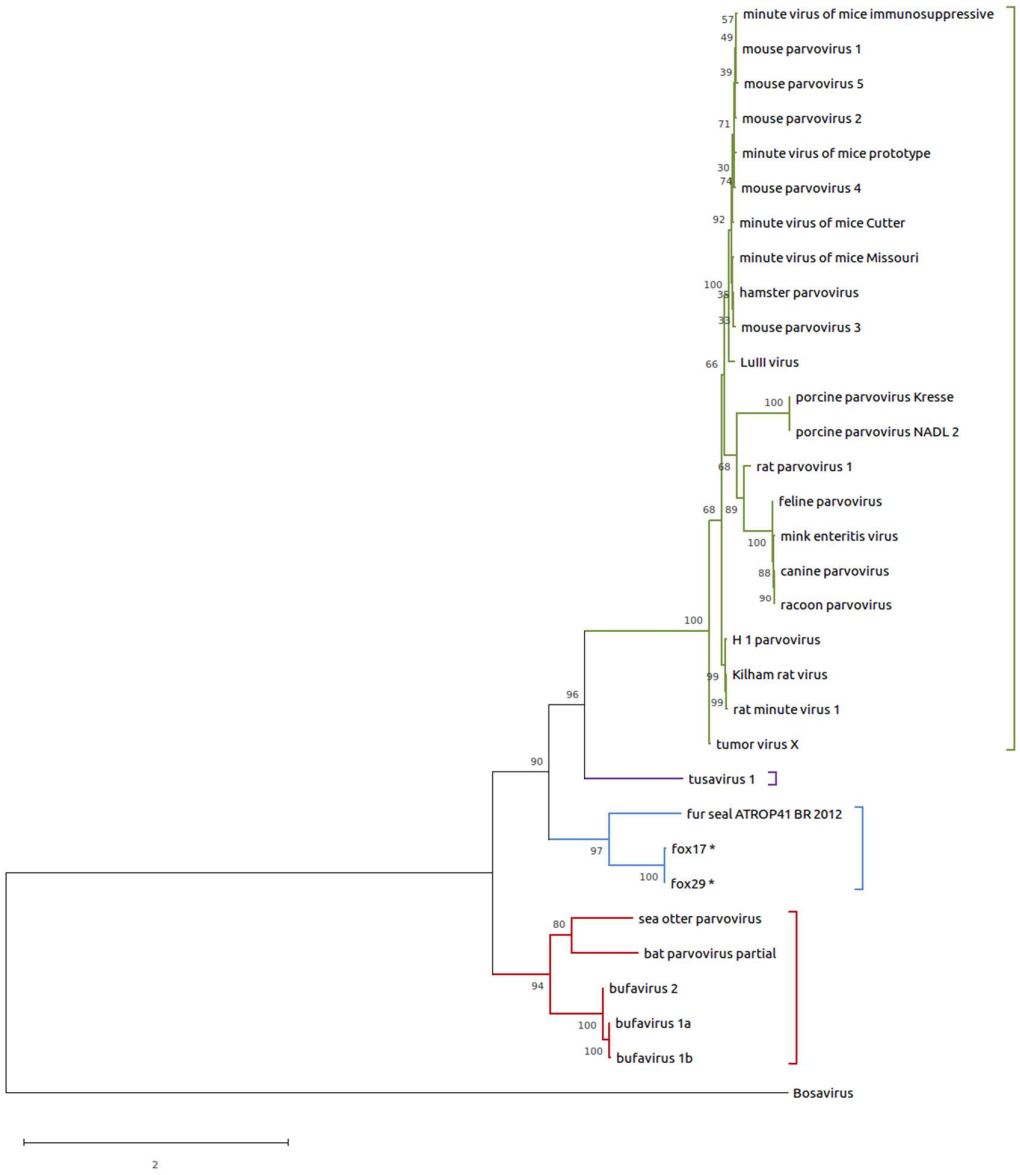

Figure 18: Maximum likelihood tree constructed using the amino acid sequence of the replicase proteins of protoparvirus genus. The tree was constructed using IQ-TREE (Nguyen et al. 2015), with the substitution model LG+F+G4 and 1000 ultra fast bootstraps. UC demarcates unclassified species.

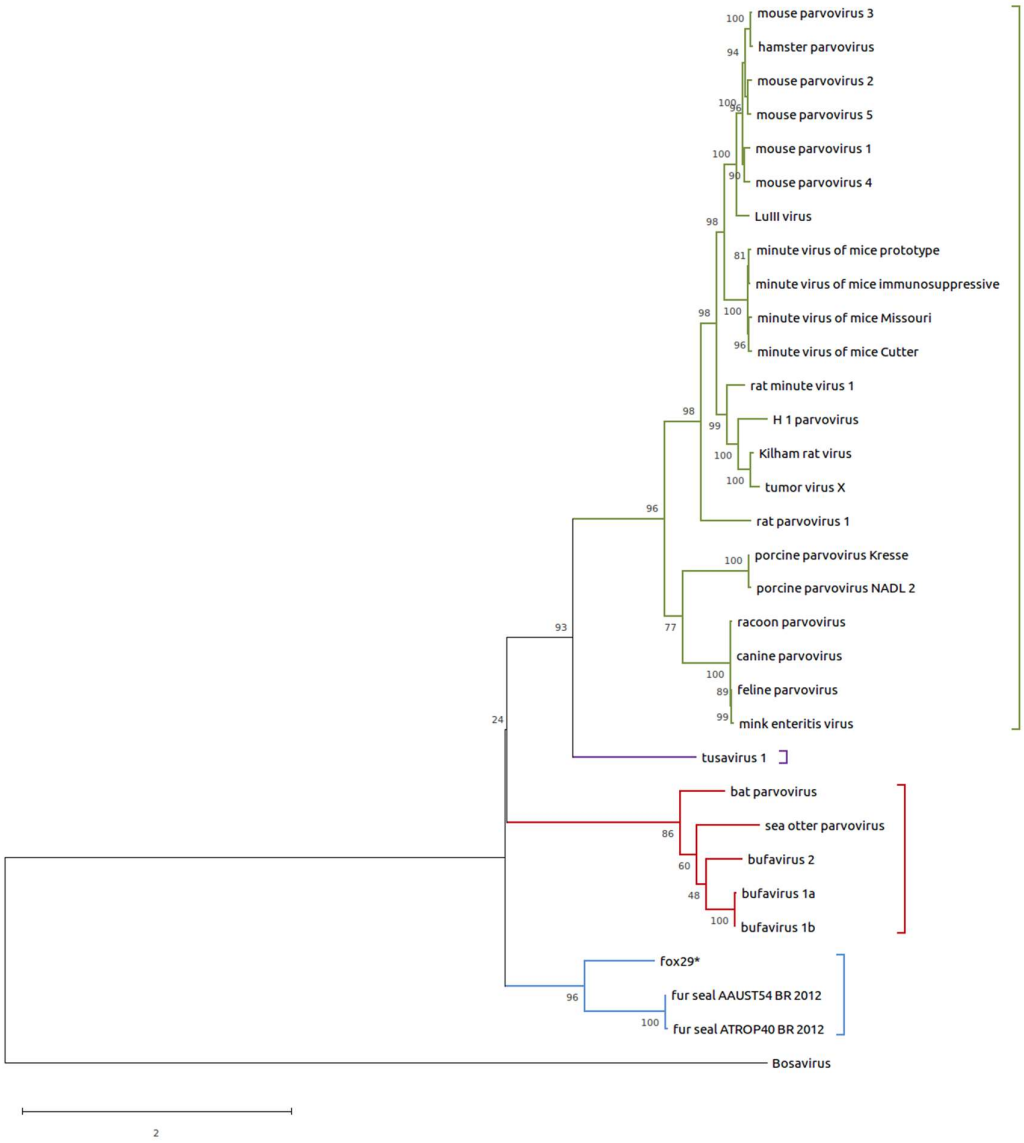

Figure 19: Maximum likelihood tree constructed using the amino acid sequence of the capsid proteins of protoparvirus genus. The tree was constructed using IQ-TREE (Nguyen et al. 2015), with the substitution model LG+F+G4 and 1000 ultra fast bootstraps. UC demarcates unclassified species.

The second parvovirus identified in this study belonged to the dependoparvovirus genus. The majority of viruses in this genus are dependant on a helper virus, meaning

they require co-infection to replicate. They are capable of integrating into the host genome and therefore there are a wide variety of species with endogenous viruses with homology to this genus (Kapoor et al. 2010). Currently, they have not been associated with disease (Zinn & Vandenberghe 2014). The exception to this rule is the ansiform dependoparvovirus group. Viruses in this genus can cause disease in waterfowl with mortality rates as high as 80% (Glávits et al. 2005).

Fox7 contained a partial parvovirus genome with homology to viruses in the *Dependoparvovirus* genus. The contig was 2923 bp in length and the sequencing coverage was 131x. A partial capsid and polymerase ORF could be annotated from either end of the contig. ICTV has declared that species share over 85% similarity with each other, indicating that this partial genome would represent a new species. This is again a situation were homology indicates the possibility of the virus originating from a prey item. The *Dependoparvovirus* contig from fox7 shared low homology to a murine dependoparvovirus . Previous studies have shown that rodents can shed high numbers of parvovirus (Nobach et al. 2015). Bodewes et al did detect dependoparvovivurs in fecal virome of red foxes, however, due to the low number of reads sequences were not uploaded to NCBI so cannot be compared (Bodewes et al. 2013b).

Concluding Remarks and Future Directions

This study has demonstrated a diversity of viral sequences are presented in the feces of arctic and red foxes, originating from a wide range of possible sources. Congruent with other virome analysis, viruses from the diet were shown to be detectable in the fecal virome (Zhang et al. 2006). This gives a new perspective on the food web. Especially, considering that several of the viruses discussed here can be contracted from

feeding on infected prey. In addition to the possibility of pathogen transmission from prey, this data also raises the question of the possibility of virus transmission between wild canids and domestic dogs. The data presented here shows kobuviruses, which share high homology to those detected in domestic dogs, are present in the feces of arctic foxes. Previous studies have shown that domestic dogs are capable of introducing pathogens into wildlife (Canuti et al. 2017). In some areas, vulnerable fox populations have been impacted by dog viruses (Timm et al. 2009). More studies are needed to understand if the exchange of kobuviruses is occurring between canid species in the Churchill area, especially in a changing climate.

Arctic foxes in particular are predicted to be largely impacted by climate change (Fuglei & Ims 2008). For example, the encroachment of red foxes into arctic fox territory is already occurring (Fuglei & Ims 2008). Animal movement into virgin territory means species will come into contact with the pathogens of encroaching species. Viruses introduced into naive populations can have devastating impacts (Price et al. 2014). This study is important groundwork to understand a baseline of viral diversity in these species to compare sick animals or even outbreaks to (Zhang et al. 2017).

Estimates of huge viral diversity indicate there are large numbers of undiscovered viruses in eukaryotes (Anthony et al. 2013). Other groups have begun to reveal some of this viral diversity in the Antarctic (Yang et al. 2019, Zablocki et al. 2014). This study has found a surprising link between the viruses of the arctic and antarctic, in the form of a group of parvoviruses. As the climate of our planet undergoes changes that can have an impact on the spread of disease, it is particularly important to studyviral diversity of changing ecosystems, like the arctic (McIntyre et al. 2017). Like previous virome studies, this study has led to many more questions than answers, but to know what to ask is the first step.

References

Adams MJ, King AMQ, Carstens EB. 2013. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2013). *Arch. Virol.* 158(9):2023–30

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215(3):403–10

Anthony SJ, Epstein JH, Murray KA, Navarrete-Macias I, Zambrana-Torrelio CM, et al. 2013. A strategy to estimate unknown viral diversity in mammals. *MBio*. 4(5):e00598-13

Appel MJ, Cooper BJ, Greisen H, Scott F, Carmichael LE. 1979. Canine viral enteritis. I. Status report on corona- and parvo-like viral enteritides. *Cornell Vet.* 69(3):123–33

Balboni A, Tryland M, Mørk T, Killengreen ST, Fuglei E, Battilani M. 2019. Unique genetic features of canine adenovirus type 1 (CAdV-1) infecting red foxes (Vulpes vulpes) in northern Norway and arctic foxes (Vulpes lagopus) in Svalbard. *Vet. Res. Commun.*

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19(5):455–77

Barker IK, Povey RC, Voigt DR. 1983. Response of mink, skunk, red fox and raccoon to inoculation with mink virus enteritis, feline panleukopenia and canine parvovirus and prevalence of antibody to parvovirus in wild carnivores in Ontario. *Can. J. Comp. Med.* 47(2):188–97

Barrett T, Crowther J, Osterhaus AD, Subbarao SM, Groen J, et al. 1992. Molecular and serological studies on the recent seal virus epizootics in Europe and Siberia. *Sci. Total*

Environ. 115(1–2):117–32

Baudel H, De Nys H, Mpoudi Ngole E, Peeters M, Desclaux A. 2019. Understanding Ebola virus and other zoonotic transmission risks through human-bat contacts: Exploratory study on knowledge, attitudes and practices in Southern Cameroon. *Zoonoses Public Health*. 66(3):288–95

Bexton S, Wiersma LC, Getu S, van Run PR, Verjans GMGM, et al. 2015. Detection of Circovirus in Foxes with Meningoencephalitis, United Kingdom, 2009-2013. *Emerging Infect. Dis.* 21(7):1205–8

Bodewes R, Hapsari R, Rubio García A, Sánchez Contreras GJ, van de Bildt MWG, et al. 2014a. Molecular epidemiology of seal parvovirus, 1988-2014. *PLoS ONE*. 9(11):e112129

Bodewes R, Morick D, de Mutsert G, Osinga N, Bestebroer T, et al. 2013a. Recurring influenza B virus infections in seals. *Emerging Infect. Dis.* 19(3):511–12

Bodewes R, Ruiz-Gonzalez A, Schapendonk CME, van den Brand JMA, Osterhaus ADME, Smits SL. 2014b. Viral metagenomic analysis of feces of wild small carnivores. *Virol. J.* 11:89

Bodewes R, van der Giessen J, Haagmans BL, Osterhaus ADME, Smits SL. 2013b. Identification of multiple novel viruses, including a parvovirus and a hepevirus, in feces of red foxes. *J. Virol.* 87(13):7758–64

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30(15):2114–20

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND.

Nat. Methods. 12(1):59–60

Canuti M, Rodrigues B, Whitney HG, Lang AS. 2017. Introduction of canine parvovirus 2 into wildlife on the Island of Newfoundland, Canada. *Infect. Genet. Evol.* 55:205–8

Cotmore SF, Agbandje-McKenna M, Canuti M, Chiorini JA, Eis-Hubinger A-M, et al. 2019. ICTV virus taxonomy profile: parvoviridae. *J. Gen. Virol.* 100(3):367–68

Cotmore SF, Agbandje-McKenna M, Chiorini JA, Mukha DV, Pintel DJ, et al. 2014. The family Parvoviridae. *Arch. Virol.* 159(5):1239–47

Crawford PC, Dubovi EJ, Castleman WL, Stephenson I, Gibbs EPJ, et al. 2005. Transmission of equine influenza virus to dogs. *Science*. 310(5747):482–85

Dayaram A, Potter KA, Moline AB, Rosenstein DD, Marinov M, et al. 2013. High global diversity of cycloviruses amoHTSt dragonflies. *J. Gen. Virol.* 94(Pt 8):1827–40

Di Martino B, Di Profio F, Melegari I, Robetto S, Di Felice E, et al. 2014. Molecular evidence of kobuviruses in free-ranging red foxes (Vulpes vulpes). *Arch. Virol.* 159(7):1803–6

DiGeronimo PM, Van Why K, Glass H, Dubovi EJ, Latney LV. 2019. Serosurvey for Influenza Virus Subtypes H3N8 and H3N2 Antibodies in Free-Ranging Canids in Pennsylvania, USA. *J. Wildl. Dis.* 55(1):227–30

Donaldson EF, Haskew AN, Gates JE, Huynh J, Moore CJ, Frieman MB. 2010. Metagenomic analysis of the viromes of three North American bat species: viral diversity among different bat species that share a common habitat. *J. Virol.* 84(24):13004–18

Fuglei E, Ims RA. 2008. Global warming and effects on the Arctic fox. *Sci. Prog.* 91(Pt 2):175–91

Ge X, Li J, Peng C, Wu L, Yang X, et al. 2011. Genetic diversity of novel circular ssDNA

viruses in bats in China. *J. Gen. Virol.* 92(Pt 11):2646–53

Glávits R, Zolnai A, Szabó E, Ivanics E, Zarka P, et al. 2005. Comparative pathological studies on domestic geese (Anser anser domestica) and Muscovy ducks (Cairina moschata) experimentally infected with parvovirus strains of goose and Muscovy duck origin. *Acta Vet. Hung.* 53(1):73–89

Guo Z, He Q, Tang C, Zhang B, Yue H. 2018. Identification and genomic characterization of a novel CRESS DNA virus from a calf with severe hemorrhagic enteritis in China. *Virus Res.* 255:141–46

Hakimuddin F, Abidi F, Jafer O, Li C, Wernery U, et al. 2016. Incidence and detection of beak and feather disease virus in psittacine birds in the UAE. *Biomolecular Detection and Quantification*. 6:27–32

Horimoto T, Kawaoka Y. 2001. Pandemic threat posed by avian influenza A viruses. *Clin. Microbiol. Rev.* 14(1):129–49

Jones MEB, Amman BR, Sealy TK, Uebelhoer LS, Schuh AJ, et al. 2019. Clinical, Histopathologic, and Immunohistochemical Characterization of Experimental Marburg Virus Infection in A Natural Reservoir Host, the Egyptian Rousette Bat (Rousettus aegyptiacus). *Viruses*. 11(3):

Jori F, Etter E. 2016. Transmission of foot and mouth disease at the wildlife/livestock interface of the Kruger National Park, South Africa: Can the risk be mitigated? *Prev. Vet. Med.* 126:19–29

Kapoor A, Simmonds P, Lipkin WI. 2010. Discovery and characterization of mammalian endogenous parvoviruses. *J. Virol.* 84(24):12628–35

Keawcharoen J, Oraveerakul K, Kuiken T, Fouchier RAM, Amonsin A, et al. 2004. Avian influenza H5N1 in tigers and leopards. *Emerging Infect. Dis.* 10(12):2189–91

Kennedy S, Smyth JA, Cush PF, McCullough SJ, Allan GM, McQuaid S. 1989. Transatlantic spread of seal virus. *Nature*. 337(6209):695

Klopfleisch R, Wolf PU, Wolf C, Harder T, Starick E, et al. 2007. Encephalitis in a stone marten (Martes foina) after natural infection with highly pathogenic avian influenza virus subtype H5N1. *J. Comp. Pathol.* 137(2–3):155–59

Kluge M, Campos FS, Tavares M, de Amorim DB, Valdez FP, et al. 2016. Metagenomic Survey of Viral Diversity Obtained from Feces of Subantarctic and South American Fur Seals. *PLoS ONE*. 11(3):e0151921

Lempp C, Jungwirth N, Grilo ML, Reckendorf A, Ulrich A, et al. 2017. Pathological findiHTS in the red fox (Vulpes vulpes), stone marten (Martes foina) and raccoon dog (Nyctereutes procyonoides), with special emphasis on infectious and zoonotic agents in Northern Germany. *PLoS ONE*. 12(4):e0175469

Li L, Giannitti F, Low J, Keyes C, Ullmann LS, et al. 2015. Exploring the virome of diseased horses. *J. Gen. Virol.* 96(9):2721–33

Li L, Kapoor A, Slikas B, Bamidele OS, Wang C, et al. 2010. Multiple diverse circoviruses infect farm animals and are commonly found in human and chimpanzee feces. *J. Virol.* 84(4):1674–82

Li L, Pesavento PA, Shan T, Leutenegger CM, Wang C, Delwart E. 2011a. Viruses in diarrhoeic dogs include novel kobuviruses and sapoviruses. *J. Gen. Virol.* 92(Pt 11):2534–41

Li L, Shan T, Soji OB, Alam MM, Kunz TH, et al. 2011b. Possible cross-species transmission of circoviruses and cycloviruses among farm animals. *J. Gen. Virol.* 92(Pt 4):768–72

Li L, Shan T, Wang C, Côté C, Kolman J, et al. 2011c. The fecal viral flora of California sea lions. *J. Virol.* 85(19):9909–17

Liberda EN, Meldrum R, Charania NA, Davey R, Tsuji LJ. 2017. Avian influenza prevalence among hunter-harvested birds in a remote Canadian First Nation community. *Rural Remote Health*. 17(1):3864

Ling Y, Zhang X, Qi G, Yang S, Jingjiao L, et al. 2019. Viral metagenomics reveals significant viruses in the genital tract of apparently healthy dairy cows. *Arch. Virol.* 164(4):1059–67

Lojkić I, Biđin M, Prpić J, Šimić I, Krešić N, Bedeković T. 2016. Faecal virome of red foxes from peri-urban areas. *Comp. Immunol. Microbiol. Infect. Dis.* 45:10–15

Maas R, Tacken M, Ruuls L, Koch G, van Rooij E, Stockhofe-Zurwieden N. 2007. Avian influenza (H5N1) susceptibility and receptors in dogs. *Emerging Infect. Dis.* 13(8):1219–21

McIntyre KM, Setzkorn C, Hepworth PJ, Morand S, Morse AP, Baylis M. 2017. Systematic assessment of the climate sensitivity of important human and domestic animals pathogens in europe. *Sci. Rep.* 7(1):7134

Melegari I, Sarchese V, Di Profio F, Robetto S, Carella E, et al. 2018. First molecular identification of kobuviruses in wolves (Canis lupus) in Italy. *Arch. Virol.* 163(2):509–13

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.*

32(1):268–74

Nobach D, Bourg M, Herzog S, Lange-Herbst H, Encarnação JA, et al. 2015. Shedding of Infectious Borna Disease Virus-1 in Living Bicolored White-Toothed Shrews. *PLoS ONE*. 10(8):e0137018

Oh DY, Silva PA, Hauroeder B, Diedrich S, Cardoso DDP, Schreier E. 2006. Molecular characterization of the first Aichi viruses isolated in Europe and in South America. *Arch. Virol.* 151(6):1199–1206

Olarte-Castillo XA, Heeger F, Mazzoni CJ, Greenwood AD, Fyumagwa R, et al. 2015. Molecular characterization of canine kobuvirus in wild carnivores and the domestic dog in Africa. *Virology*. 477:89–97

Padilla-Rodriguez M, Rosario K, Breitbart M. 2013. Novel cyclovirus discovered in the Florida woods cockroach Eurycotis floridana (Walker). *Arch. Virol.* 158(6):1389–92

Pankovics P, Boros Á, Bíró H, Horváth KB, Phan TG, et al. 2016. Novel picornavirus in domestic rabbits (Oryctolagus cuniculus var. domestica). *Infect. Genet. Evol.* 37:117–22

Phan TG, Luchsinger V, Avendaño LF, Deng X, Delwart E. 2014a. Cyclovirus in nasopharyngeal aspirates of Chilean children with respiratory infections. *J. Gen. Virol.* 95(Pt 4):922–27

Phan TG, Sdiri-Loulizi K, Aouni M, Ambert-Balay K, Pothier P, et al. 2014b. New parvovirus in child with unexplained diarrhea, Tunisia. *Emerging Infect. Dis.* 20(11):1911–13

Phan TG, Vo NP, Bonkoungou IJO, Kapoor A, Barro N, et al. 2012. Acute diarrhea in West African children: diverse enteric viruses and a novel parvovirus genus. *J. Virol.* 86(20):11024–30

Price SJ, Garner TWJ, Nichols RA, Balloux F, Ayres C, et al. 2014. Collapse of amphibian communities due to an introduced Ranavirus. *Curr. Biol.* 24(21):2586–91

Reperant LA, van Amerongen G, van de Bildt MWG, Rimmelzwaan GF, Dobson AP, et al. 2008. Highly pathogenic avian influenza virus (H5N1) infection in red foxes fed infected bird carcasses. *Emerging Infect. Dis.* 14(12):1835–41

Reuter G, Boros A, Pankovics P. 2011. Kobuviruses - a comprehensive review. *Rev. Med. Virol.* 21(1):32–41

Root JJ, Shriner SA, Ellis JW, VanDalen KK, Sullivan HJ, Franklin AB. 2015. When fur and feather occur together: interclass transmission of avian influenza A virus from mammals to birds through common resources. *Sci. Rep.* 5:14354

Rosario K, Breitbart M, Harrach B, Segalés J, Delwart E, et al. 2017. Revisiting the taxonomy of the family Circoviridae: establishment of the genus Cyclovirus and removal of the genus Gyrovirus. *Arch. Virol.* 162(5):1447–63

Roth JD. 2002. Temporal varitability in arctic fox diet as reflected in sTable-carbon isotopes; the importance of sea ice. *Oecologia*. 133(1):70–77

Roth JD. 2003. Varitability in marine resources affects arctic fox population dynamics. *J. Anim. Ecol.* 72(4):668–76

Sato G, Kawashima T, Kiuchi M, Tohya Y. 2015. Novel cyclovirus detected in the intestinal contents of Taiwan squirrels (Callosciurus erythraeus thaiwanensis). *Virus Genes*. 51(1):148–51

Shan T, Li L, Simmonds P, Wang C, Moeser A, Delwart E. 2011. The fecal virome of pigs on a high-density farm. *J. Virol.* 85(22):11697–708

Shepard SS, Meno S, Bahl J, Wilson MM, Barnes J, Neuhaus E. 2016. Viral deep sequencing needs an adaptive approach: IRMA, the iterative refinement meta-assembler. *BMC Genomics*. 17:708

Shulman LM, Davidson I. 2017. Viruses with Circular Single-Stranded DNA Genomes Are Everywhere! *Annu. Rev. Virol.* 4(1):159–80

Skírnisson K, Eydal M, Gunnarsson E, Hersteinsson P. 1993. Parasites of the arctic fox (Alopex lagopus) in Iceland. *J. Wildl. Dis.* 29(3):440–46

Soma T, Matsubayashi M, Sasai K. 2016. Detection of kobuvirus RNA in Japanese domestic dogs. *J. Vet. Med. Sci.* 78(11):1731–35

Stien A, Voutilainen L, Haukisalmi V, Fuglei E, Mørk T, et al. 2010. Intestinal parasites of the Arctic fox in relation to the abundance and distribution of intermediate hosts. *Parasitology*. 137(1):149–57

Stuetzer B, Hartmann K. 2014. Feline parvovirus infection and associated diseases. *Vet. J.* 201(2):150–55

Tan LV, van Doorn HR, Nghia HDT, Chau TTH, Tu LTP, et al. 2013. Identification of a new cyclovirus in cerebrospinal fluid of patients with acute central nervous system infections. *MBio*. 4(3):e00231-13

Tausch SH, Renard BY, Nitsche A, Dabrowski PW. 2015. RAMBO-K: Rapid and Sensitive Removal of Background Sequences from Next Generation Sequencing Data. *PLoS ONE*. 10(9):e0137896

Thanawongnuwech R, Amonsin A, Tantilertcharoen R, Damrongwatanapokin S, Theamboonlers A, et al. 2005. Probable tiger-to-tiger transmission of avian influenza

H5N1. *Emerging Infect. Dis.* 11(5):699–701

Timm SF, Munson L, Summers BA, Terio KA, Dubovi EJ, et al. 2009. A suspected canine distemper epidemic as the cause of a catastrophic decline in Santa Catalina Island foxes (Urocyon littoralis catalinae). *J. Wildl. Dis.* 45(2):333–43

Truyen U, Müller T, Heidrich R, Tackmann K, Carmichael LE. 1998. Survey on viral pathogens in wild red foxes (Vulpes vulpes) in Germany with emphasis on parvoviruses and analysis of a DNA sequence from a red fox parvovirus. *Epidemiol. Infect.* 121(2):433–40

Väisänen E, Fu Y, Hedman K, Söderlund-Venermo M. 2017. Human Protoparvoviruses. *Viruses*. 9(11):

Voorhees IEH, Dalziel BD, Glaser A, Dubovi EJ, Murcia PR, et al. 2018. Multiple Incursions and Recurrent Epidemic Fade-Out of H3N2 Canine Influenza A Virus in the United States. *J. Virol.* 92(16):

Webster RG, Govorkova EA. 2014. Continuing challenges in influenza. *Ann. N. Y. Acad. Sci.* 1323:115–39

Westgeest KB, Russell CA, Lin X, Spronken MIJ, Bestebroer TM, et al. 2014. Genomewide analysis of reassortment and evolution of human influenza A(H3N2) viruses circulating between 1968 and 2011. *J. Virol.* 88(5):2844–57

Wills CG. 1952. Notes on infectious enteritis of mink and its relationship to feline enteritis. *Can. J. Comp. Med. Vet. Sci.* 16(12):419–20

Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 15(3):R46

Wu Z, Yang L, Ren X, He G, Zhang J, et al. 2016. Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases. *ISME J.* 10(3):609–20

Xu P-L, Zhao Y, Zheng H-H, Tian R-B, Han H-Y, et al. 2019. Analysis of genetic variation of porcine circovirus type 2 within pig populations in central China. *Arch. Virol.* 164(5):1445–51

Yamashita T, Sakae K, Ishihara Y, Isomura S, Utagawa E. 1993. Prevalence of newly isolated, cytopathic small round virus (Aichi strain) in Japan. *J. Clin. Microbiol.* 31(11):2938–43

Yamashita T, Sakae K, Kobayashi S, Ishihara Y, Miyake T, et al. 1995. Isolation of cytopathic small round virus (Aichi virus) from Pakistani children and Japanese travelers from Southeast Asia. *Microbiol. Immunol.* 39(6):433–35

Yang Q, Gao C, Jiang Y, Wang M, Zhou X, et al. 2019. Metagenomic characterization of the viral community of the south scotia ridge. *Viruses*. 11(2):

Yang Y, Cheng Y, Li N, Cheng S, Guo L, et al. 2018. Mink circovirus can infect minks, foxes and raccoon dogs. *Virol. Sin.* 33(6):561–64

Yinda CK, Vanhulle E, Conceição-Neto N, Beller L, Deboutte W, et al. 2019. Gut virome analysis of cameroonians reveals high diversity of enteric viruses, including potential interspecies transmitted viruses. *mSphere*. 4(1):

Zablocki O, van Zyl L, Adriaenssens EM, Rubagotti E, Tuffin M, et al. 2014. Niche-dependent genetic diversity in Antarctic metaviromes. *Bacteriophage*. 4(4):e980125

Zhang T, Breitbart M, Lee WH, Run J-Q, Wei CL, et al. 2006. RNA viral community in

human feces: prevalence of plant pathogenic viruses. *PLoS Biol.* 4(1):e3

Zhang W, Yang S, Shan T, Hou R, Liu Z, et al. 2017. Virome comparisons in wild-diseased and healthy captive giant pandas. *Microbiome*. 5(1):90

Zinn E, Vandenberghe LH. 2014. Adeno-associated virus: fit to serve. *Curr. Opin. Virol.* 8:90–97

Chapter 5: General conclusions and future directions

The impact of metagenomics on the field of virology is clear and it can be established with this work that enrichment can play a vital role in viral metagenomics. Specficially, chapter 2 describes the impact of targeted enrichment on metagenomic viral detection in vetereinary diagnostics. The validation performed in this chapter demonstrates the breadth of sample types, host species and viral diversity this method is applicable to. This is especially important as this test is especially well suited to diagnostic laboratories, such as the NCFAD. In terms of unknowns, unexpected and coinfecions, this type of testing provides an unparalleled opportunity for viral detection and characterization. That being said, there are several ways in which the potential impact of ViroCap targeted enrichment and metagenomics could be improved further. Given that this test is meant to be used diagnostically, shortening incubation times would improve turnaround times and therefore improve it's use in diagnostics. Likewise, adaption of the ViroCap method to new low-input libraries would reduce amplification cycles and therefore PCR duplication and polymerase induced errors.

In addition to evaluating the use of targeted enrichment in animal disease investigations and virome analysis, this thesis describes the viromes of several wildlife species. The first being in a die-off of Canada and Snow Goose and the second being the fecal virome of arctic and red foxes. Several times I have discussed wildlife as reservoirs for human domestic animal pathogens. This point is underscored by the results of the mass die-off of geese. Previous studies have assumed the source of fowl and cetacean *Gammacoronavirus* pathogens are wild birds, I have fully described the genome of the first species of gammacoronavirus detected in wild birds. In addition, I further the known diversity of AMPV in Canada geese, the reservoir for this fowl pathogen. These findings leave several unaswered questions for future researchers to pursue, including including the pathogenicity and natural resevoir of Goose Coronavirus CB-17 and the pathogenicity and full identity of goose

adenoviruses in Canada.

The work presented here also underscores the fact that most viral diversity is yet to be described, this is especially true pertaining to wildlife. This fact is demonstrated by the indication of a novel group of marine parvoviruses linking the Arctic and Antarctic revealed in Chapter 4 of this thesis. Establishing the presence of unclassified circoviruses and parvoviruses in the arctic adds to our understanding of the ecology of these viruses, and in turn the hosts that carry them. This chapter described the first evaluation of the fecal virome of arctic foxes and consequently raises several new lines of research. Were the kobuviruses detected in this study introduced to the arctic foxes by encroaching red foxes or domestic dogs?

While the virome investigations presented here make it clear that viral metagenomics cannot answer all questions, we've demonstrated that it is an invaluable first step in the understanding the diversity of the viral world. While the investigations presented here have a;; lead to more questions about pathogenicity or host range of the detected viruses, we now know what viruses to ask these questions about.