DIGITAL COMMONS
@ UNIVERSITY OF SOUTH FLORIDA

University of South Florida
Digital Commons @ University of South Florida

USF Tampa Graduate Theses and Dissertations
USF Graduate Theses and Dissertations

July 2017

Viruses in marine animals: Discovery, detection, and characterizarion

Elizabeth Fahsbender
University of South Florida, efahsben@mail.usf.edu

Follow this and additional works at: https://digitalcommons.usf.edu/etd

Part of the Other Oceanography and Atmospheric Sciences and Meteorology Commons

Scholar Commons Citation
Fahsbender, Elizabeth, "Viruses in marine animals: Discovery, detection, and characterizarion" (2017). *USF Tampa Graduate Theses and Dissertations.*
https://digitalcommons.usf.edu/etd/6832

This Dissertation is brought to you for free and open access by the USF Graduate Theses and Dissertations at Digital Commons @ University of South Florida. It has been accepted for inclusion in USF Tampa Graduate Theses and Dissertations by an authorized administrator of Digital Commons @ University of South Florida. For more information, please contact digitalcommons@usf.edu.

Viruses in Marine Animals: Discovery, Detection, and Characterization

by

Elizabeth Fahsbender

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
College of Marine Science
University of South Florida

Major Professor: Mya Breitbart, Ph.D.
John Paul, Ph.D.
Larry Dishaw, Ph.D.
Peter Medveczky, Ph.D.
Thierry Work, DVM

Date of Approval:
June 6, 2017

Keywords: Animal virus, Viral metagenomics, Whole-genome sequencing, Serology

Copyright © 2017, Elizabeth Fahsbender

ACKKNOWLEDGMENTS

This dissertation was made possible through the guidance and encouragement of Dr. Mya Breitbart. I would also like to thank my committee members Dr. John Paul, Dr. Larry Dishaw, Dr. Peter Medveczky, and Dr. Thierry Work for their insight throughout my graduate work. Thank you to Dr. Karyna Rosario for her support. I am also grateful for the help and friendship of my lab mates. I would like to thank my family for their unwavering support.

Table of Contents

ABSTRACT……………………………………………………………………………………ii

INTRODUCTION……………………………………………………………………………...1

DISCUSSION…………………………………………………………………………………11

REFERENCES………………………………………………………………………………..20

APPENDIX 1: DIVERSE AND HIGHLY RECOMBINANT ANELLOVIRUSES ASSOCIATED WITH WEDDELL SEALS IN ANTARCTICA……………………………….29

APPENDIX 2: DISCOVERY OF A NOVEL CIRCULAR DNA VIRUS IN THE FORBES SEA STAR, *ASTERIAS FORBESI*……………………………………………………………………30

APPENDIX 3: SEQUENCING OF THE CHELONID FIBROPAPILLOMATOSISASSOCIATED HERPESVIRUS GENOME FROM THREE FLORIDA SEA TURTLE SPECIES………………………………………………………………………………...31

APPENDIX 4: DEVELOPMENT OF A SEROLOGICAL ASSAY FOR THE NOVEL SEA LION (*ZALOPHUS CALIFORNIAS*) ANELLOVIRUS, ZCAV……………………………….32

ABSTRACT

Diseases in marine animals are emerging at an increasing rate. Disease forecasting enabled by virus surveillance presents a proactive solution for managing emerging diseases. Broad viral surveys aid in disease forecasting by providing baseline data on viral diversity associated with various hosts, including many that are not associated with disease. However, these viruses can become pathogens due to expansion in host or geographic range, as well as when changing conditions shift the balance between commensal viruses and the host immune system. Therefore, it is extremely valuable to identify and characterize viruses present in many different hosts in a variety of environments, regardless of whether the hosts are symptomatic or not.

The lack of a universal gene shared by all viruses makes virus surveillance difficult, because no single assay exists that can detect the enormous diversity of viruses. Viral metagenomics circumvents this issue by purifying viral particles directly from host tissues and sequencing the nucleic acids, allowing for virus identification. However, virus identification is only the first step, which should ideally be followed by complete sequencing of the viral genome to identify genes of interest and develop assays to reveal viral prevalence, tropism, ecology, and pathogenicity. This dissertation focuses on the discovery of novel viruses in marine animals, characterization of complete viral genomes, and the development of subsequent diagnostic assays for further analysis of virus ecology.

First, viral metagenomics was used to explore the viruses present in the healthy Weddell seal (*Leptonychotes weddellii*) population in Antarctica, which led to the discovery of highly prevalent small, circular single-stranded DNA (ssDNA) viruses. The lack of knowledge regarding the viruses of Antarctic wildlife warrants this study to determine baseline viral communities in healthy animals that can be used to survey changes over time. From the healthy Weddell seals, viral metagenomics led to the discovery of 152 novel anellovirus genomes, encompassing two anellovirus species. Characterizing these viruses is important for understanding the prevalence and diversity of ssDNA viruses, which have only recently been described in marine animals. Furthermore, since emerging diseases can be caused by changing conditions affecting host susceptibility to a virus that was previously not related to disease (opportunistic pathogen), having baseline data allows for quick identification of the pathogen.

In addition to determining baseline data, viral metagenomics can explore the role of viruses in disease. A novel virus, *Asterias forbesi*-associated circular virus (AfaCV), was discovered in the Atlantic sea star *Asterias forbesi* displaying symptoms of sea star wasting disease (SSWD). AfaCV was the first circular replicase-encoding ssDNA (CRESS-DNA) virus discovered in echinoderms, but it was only present in 10% of SSWD sea stars indicating it is not involved in the development of the disease.

This dissertation also focuses on elucidating the role of two previously characterized viruses, chelonid fibropapillomatosis-associated herpesvirus (CHHV5; Chelonid herpesvirus 5, ChHV5) and *Zalophus californianus* anellovirus (ZcAV), in animal health. PCR amplicon sequencing was used to obtain large portions of the 132 kb genome of ChHV5, the putative etiological agent of the neoplastic sea turtle disease, fibropapillomatosis. Obtaining the genome of ChHV5 from Florida green, Kemp's ridley, and loggerhead sea turtles provides data for

phylogenetic analysis across geographic locations and sea turtle species, as well as a reference for designing downstream molecular assays to examine viral latency.

ZcAV was first described from the lungs of captive sea lions involved in a mortality event. PCR could not detect ZcAV in the blood of infected animals, and since sea lions are a protected species, it is not possible to obtain lung biopsies from live sea lions to determine ZcAV prevalence or its role in sea lion health. To answer these important questions, an enzyme-linked immunosorbent assay (ELISA) was developed to detect antibodies to ZcAV in serum from wild sea lion populations. This newly developed ELISA showed that sea lions mount an immune response to ZcAV, and was used to determine the prevalence of ZcAV among wild sea lion populations.

This dissertation makes an important contribution to marine science through discovery and characterization of viruses present in healthy and diseased marine animals. Several different methods were used for virus whole-genome sequencing including viral metagenomics, PCR amplicon sequencing, and target enrichment. These findings were expanded upon by developing and using PCR assays and a serological assay to screen for virus prevalence. These methods have implications for viral surveillance and understanding the role of novel viruses in animal health.

CHAPTER ONE:

Introduction

Viruses are intracellular parasites capable of causing disease in humans, animals, and plants. Viral epidemics are typically the result of a newly emerged virus, a jump from one species to the next, or the re-emergence of a virus and its expansion to a new area (Morse 1991, Brown 2004). While the genetic plasticity of viruses is a major factor in disease emergence, the propensity for the virus to cause an epidemic lies in the delicate balance between host, pathogen, and the environment. Environmental conditions such as population density, climate, and water quality are all factors that contribute to pathogen transmission. Consequently, environmental degradation can impart stressors on a population, leaving it susceptible to disease (Harb et al. 1993, Appawu et al. 2001, Patz et al. 2005).

Diseases are emerging at an increasing rate, and many are a direct result of environmental changes (Harvell et al. 1999, Dobson and Foufopoulos 2001, Lederberg et al. 2003, Wilcox and Gubler 2005). Human pathogens include viruses, bacteria, protozoa, fungi, and helminths, with viruses representing the majority (44%) of emerging pathogens (Fig. 1) (Taylor et al. 2001). Although virus research has largely been anthropocentric, human disease is inextricably linked to animal health and therefore understanding disease in wildlife is an important component in understanding disease emergence in humans. In fact, about 75% of emerging human pathogens are zoonotic, meaning they are passed from animals to humans, with forty-one percent of the

emerging zoonotic diseases caused by viruses (Fig. 1). This underscores the significance of animals as viral reservoirs.

Although zoonotic events are an important reason for surveying animal viruses, instances of viruses being passed from humans to animals (reverse zoonosis) are increasing and should not be ignored (Dobson and Foufopoulos 2001, Messenger et al. 2014). Humans impact the outcome of wildlife disease especially through environmental degradation and the disease itself can become a driver of ecosystem changes as the balance between trophic levels is altered (Epstein et al. 2003). Evaluating ecosystem health through sentinel species can provide insight into animal health at a population level, as well as human and environmental health (Bossart 2006).

Surveying viruses in wildlife populations is important for disease forecasting, and exploring viruses in the marine realm is no exception; with over half of the population living near coastlines, human impact on marine life is significant. However, the knowledge of marine animal viruses is largely limited to those that affect fisheries and aquaculture, and those that cause mass mortality events. Significantly less is known about modes of virus transmission and virus reservoirs. Unfortunately, marine diseases are especially difficult to contain due to the lack of management techniques that are effective for the prevention and containment of terrestrial diseases, such as vaccines and quarantining diseased individuals (McCallum et al. 2004). Complex life cycles and migration patterns of marine animals combined with the general lack of understanding of virus transmission processes makes it difficult to contain disease upon detection. Once a disease becomes an epidemic, it can be detrimental to ecosystem community structures by disrupting trophic interactions, resulting in further environmental decay (Epstein et

al. 2003). Therefore, rapid detection of emerging disease through viral surveillance is imperative for implementing management techniques to control virus spread in the marine environment.

Viral surveillance is not without obstacles, especially when the intention is to identify emerging viruses. Unlike bacteria, viruses do not contain universal gene markers that can be amplified and sequenced for identification. Current diagnostic methods, such as polymerase chain reaction (PCR), microarray, and serology, are lacking in that they can only detect known viruses and their close relatives (Wang et al. 2002, Symonds et al. 2009). Transmission electron microscopy (TEM) is also used to identify the presence of viruses, but many viruses are indistinguishable based on appearance, or undetectable by TEM. There are roughly 2.2 million animal species in the ocean (Mora et al. 2011), with each species potentially harboring multiple viruses, suggesting there is a wealth of viral diversity that these traditional methods cannot efficiently detect. However, viral metagenomics circumvents these problems and has emerged as an effective method for describing novel viruses (Edwards and Rohwer 2005, Delwart 2007). Viral metagenomics (sequencing the nucleic acids of purified viral particles) was originally used to describe uncultured viruses in the marine environment by virus purification followed by shotgun sequencing (Breitbart et al. 2002). Viral metagenomics has since been used to explore viruses of mucus, blood, feces, cell culture and tissue samples from animals, plants, and insects (Ng et al. 2009, Svraka et al. 2010, Ng et al. 2011, Ng et al. 2011, Rosario and Breitbart 2011). Originally, Sanger sequencing was used, but advances in sequencing technology has led the field to move towards the use of next-generation (NGS) sequencing, including 454 pyrosequencing, Illumina (appendices 1, 2, and 3), and PacBio (Appendix 3).

Despite the advances in virus discovery due to metagenomics and NGS technology, determining the etiological agent of disease by fulfilling Koch's postulates is an onerous and

oftentimes impossible task (Fredricks and Relman 1996). Koch's postulates state (i) the pathogen occurs in every case of disease, (ii) the pathogen can be isolated from the host and grown in culture, (iii) a healthy host inoculated with the pathogen from pure culture induces the disease, (iv) the same pathogen can be isolated from the newly infected host (Koch and Saure 1890). However, there are often factors that confound etiology of infectious disease, including differences in disease manifestation based on genetic susceptibility, environmental conditions, coinfections that can increase pathogenicity, and temporal delays between exposure and disease (Evans 1976). Viruses that are unculturable, associated with chronic illnesses or multifactor diseases, and those that infect protected species that cannot be experimentally infected must be classified as causative agents through other methods. Viruses have been implicated as the causative agent of neoplasms such as the Epstein-Barr virus in Burkitt's lymphoma, and human papillomavirus genotype 16 in cervical cancer using methods such as seroepidemiology, gene expression or viral replicative form analysis, and presence of intact viral particles and viral genome in affected tissues (Zur Hausen et al. 1970, Evans 1976, Fredricks and Relman 1996, Walboomers et al. 1999).

Fortunately, the generation of genomic sequences is beneficial for the creation of downstream diagnostic assays for specific viruses. Demonstrating an adaptive immune response that is specific to the virus, proving of viral replication, measuring viral titers, and performing epidemiology studies are necessary to infer disease causality (Lipkin 2010). With the development of downstream applications in mind, characterizing whole genomes instead of focusing on specific genes is necessary for understanding the ecological importance of the virus (Houldcroft et al. 2016). Whole genome sequencing (WGS) can be used to track virus outbreaks when certain marker genes with low genetic variability are not useful for transmission studies;

for example, transmission of the recent outbreak of the Ebola virus was tracked using WGS (Quick et al. 2016).

Currently, the three most common NGS approaches used for WGS of viruses are viral metagenomics, PCR amplicon sequencing, and target enrichment (Houldcroft et al. 2017). It should be noted that there are numerous techniques used for WGS including single virus genomics, the assembly of RNA viruses through transcriptomics, the assembly of virus-derived small silencing RNAs, and dsRNA isolation followed by NGS (Wu et al. 2010, Allen et al. 2011, Shi et al. 2016, Urayama et al. 2016); however, this dissertation will focus on viral metagenomics, amplicon sequencing, and target enrichment. While viral metagenomic approaches are useful for the identification of novel viruses and characterization of viral diversity, target enrichment sequencing and PCR amplicon sequencing are effective methods for characterizing genomes of known viruses or their close relatives. PCR amplicon sequencing, using complimentary primers to a known sequence, is the most common approach for sequencing small viral genomes such as norovirus (Cotten et al. 2014), but this method can become labor intensive when amplifying larger genomes. Larger genomes require more PCR primers which involves optimizing each PCR assay and results in the need for large sample amounts, which may be a limiting factor (Renzette et al. 2011). PCR amplicon sequencing has been more successful for WGS than metagenomics when virus concentrations are low, although target enrichment may be a more efficient method if the genome is large (Thomson et al. 2016).

Target enrichment is used to sequence genomes directly from samples without the need for virus amplification using PCR or culturing prior to the enrichment, thus providing sequences with fewer mutations than in amplified templates (Depledge et al. 2014, Tsangaras et al. 2014, Wylie et al. 2015). The success of target enrichment sequencing is directly related to the

availability of a suitable reference genome of the target virus from which capture probes can be designed. A higher number of reference genomes will result in an efficient capture of target sequences because the specificity of the probes increases when they are designed against a panel of reference genomes that can capture genetic diversity.

Once the viral genome is completed, it can be used for the creation of downstream assays to determine the role of the virus in animal health. One of the first steps in elucidating the effects of a virus on animal health is to determine its prevalence among healthy and diseased animals and determine its tropism. To screen for prevalence, newly developed assays based on the genomic information such as PCR are useful and cost effective, but not always feasible. In cases where the animal is an endangered or protected species, tissue samples are difficult to obtain and if the virus does not circulate in the blood, serological assays are useful to test for exposure to the virus. In addition, serological assays provide valuable information about the adaptive immune response of the host and therefore clues about virus-host interactions.

Broad viral surveys are critical for characterizing pathogens and baseline viral communities. Environmental impacts have contributed to complex, multifactor diseases and the rise of opportunistic pathogens (Lee Miller et al. 2001, Bossart et al. 2002, King et al. 2002, Ylitalo et al. 2005). Therefore, establishing baseline data is necessary for forecasting emerging diseases and understanding virus-host dynamics. Reducing the time between a disease outbreak and detection of the responsible virus will impact the outcome for disease management and containment.

This dissertation focuses on the discovery of novel viruses in marine animals, characterization of complete viral genomes, and the development of diagnostic assays to determine the ecological role of these viruses.

- **Diverse and highly recombinant anelloviruses associated with Weddell seals in Antarctica** (Appendix 1). The viruses circulating among Antarctic wildlife remain largely unknown. In an effort to identify viruses associated with Weddell seals (*Leptonychotes weddellii*) inhabiting the Ross sea, vaginal and nasal swabs and fecal samples were collected for metagenomic analysis. In addition, a Weddell seal kidney and South Polar skua feces were collected. Based on Illumina sequencing results, abutting primers were designed and used for the recovery of 152 novel anellovirus genomes that share 63-70% genome-wide identities with other pinniped anelloviruses. Phylogenetic analysis revealed two anellovirus species, torque teno *Leptonychotes weddellii* virus (TTLwV) -1 and -2. TTLwV-1 is highly recombinant and subdivided into 40 genotypes. TTLwV-2 is subdivided into three genotypes and is much less recombinant, with only 26% of the sequences having recombination events, compared to the 89% of TTLwV-1 sequences. TTLwV is diverse and ubiquitous among Weddell seals with multiple genotypes infecting individuals, but the role these viruses play in seal health remains unknown.
- **Discovery of a novel circular DNA virus in the forbes sea star, *Asterias forbesi*** (Appendix 2). Sea stars are keystone predators in the marine environment, essential for structuring intertidal invertebrate communities that are critical for ecosystem health. An ongoing massive sea star die off in both the Atlantic and Pacific oceans is due to sea star wasting disease (SSWD). Symptoms include white lesions on the epidermis, deflation, and eventual death of the sea star. Although the disease is decimating the populations of multiple species of sea stars, the etiological agent causing this disease was unknown at the time of this study. This study explored the role of viruses in SSWD in *Asterias*

forbesi from the Atlantic. Viral metagenomic analysis (virus purification followed by Illumina DNA sequencing) was used to identify a single-stranded DNA (ssDNA) virus, Asterias forbesi-associated circular virus (AfaCV). The AfaCV genome organization is typical of circular Rep-encoding ssDNA (CRESS-DNA) viruses belonging to the *Circoviridae* family. PCR-based assays designed to amplify the newly characterized genome detected AfaCV in approximately 10% of *Asterias forbesi* with SSWD, demonstrating the lack of a clear association between this virus and the disease. However, AfaCV represents the first CRESS-DNA virus detected in echinoderms, adding to the growing diversity of these viruses recently recovered from invertebrates.

- **Sequencing of the chelonid fibropapillomatosis-associated herpesvirus genome from three Florida sea turtle species** (Appendix 3). Fibropapillomatosis (FP) is a debilitating disease of panzootic proportions that has been documented in all hard-shelled sea turtles, but primarily affects green sea turtles *(Chelonia mydas*) (Barragan and Sarti 1994, Herbst 1994, Lackovich et al. 1999, Herbst et al. 2004, Ene et al. 2005, Williams and Bunkley-Williams 2006, Work et al. 2009, Alfaro-Núñez and Gilbert 2014). FP is the most significant neoplastic disease in reptiles due to its rapid increase in prevalence and geographic distribution (Herbst 1994). FP was first discovered in the late 1930's by Smith and Coates (1938) in a Florida green turtle and has since spread worldwide with a circumtropical distribution (Aguirre and Lutz 2004). Despite the large proportion of sea turtles afflicted with FP, the etiological agent remains unconfirmed. However, there is overwhelming evidence suggesting the disease is caused by a virus, now known as the chelonid fibropapillomatosis-associated herpesvirus (CHHV5). Unfortunately, all attempts to culture this virus, a precursor needed to fulfill Koch's postulates (Koch and

Saure 1890), have failed, hindering efforts to understand the role of ChHV5 in FP. Currently, the only genome available of ChHV5 was sequenced from the glottis tumor of a Hawaiian green sea turtle using the bacterial artificial clone (BAC) system (Ackermann et al. 2012). To sequence ChHV5 from tumors of Florida green, loggerhead, and Kemp's ridley sea turtles, the Hawaiian genome has been used as a scaffold for long-range PCR followed by Illumina sequencing of amplicons; however, long regions of repeat sections (>10 kb) have impeded genome completion. To obtain the repeat sections of the genome, SureSelect Target enrichment is currently being used to enrich for the herpesvirus and PacBio technology was used to obtain long sequence reads from the genome.

- **Development of a serological assay for the novel sea lion *(Zalophus* californias) anellovirus, ZcAV** (Appendix 4). In response to a mortality event of several captive California sea lions (*Zalophus californianus*), viral metagenomics was previously used to identify a novel anellovirus (ZcAV), which was subsequently determined to actively replicate in the lungs of 100% of the sea lions that died in the mortality event (Ng et al. 2009). The identification of the virus is an important first step, but subsequent studies are needed to determine the role of the virus in the disease. Sea lions are a protected species; thus, it is not possible to obtain lung samples from live sea lions. Furthermore, ZcAV was never detected in the blood of infected animals, and as a result, blood samples from live sea lions could not be screened for ZcAV. This study developed an enzyme-linked immunosorbent assay (ELISA) for the detection of antibodies to ZcAV in sea lion serum. The ELISA was created based on regions predicted to be immunogenic in the putative capsid protein of ZcAV. Ninety-six paired lung and serum samples were tested for ZcAV using PCR and ELISA tested, respectively. 48% of samples were both ELISA and

PCR negative, 11% were positive for both ELISA and PCR, 18% were PCR positive, but ELISA negative, and 23% were PCR negative, but ELISA positive. These results indicate that using an ELISA along with PCR provides robust results of ZcAV prevalence.

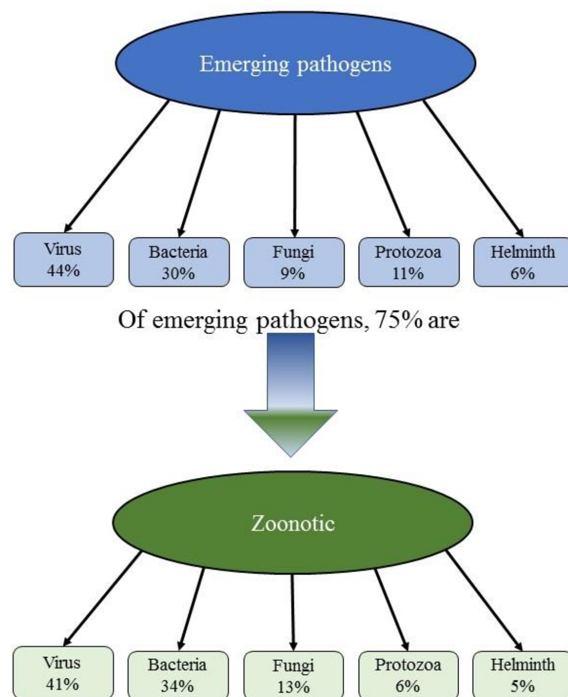


Figure 1. The breakdown (by %) of emerging and zoonotic pathogens. 41% of viruses are the etiological agent of emerging zoonotic disease.

CHAPTER 2:

Discussion

Traditional laboratory methods used to detect viruses, including both molecular and immunological assays, rely on previous knowledge of the sequence or antigenic properties of the virus of interest. While these techniques are useful for the detection of known viruses and their close relatives, they remain ineffective for the detection of genetically divergent viruses, rendering them inefficient for viral surveillance (Wang et al. 2002, Symonds et al. 2009). The relatively new method of viral metagenomics is a sequence-independent tool that circumvents the issues associated with traditional laboratory techniques. However, technical and ethical complications often arise when trying to link virus to disease. The coupling of traditional diagnostic tools with metagenomics is imperative for uncovering the role of viruses in health. This dissertation explores viruses of marine animals and moves beyond viral discovery by developing subsequent diagnostic methods.

The lack of long-term disease surveillance in Antarctica has left many questions about disease risk and prevalence. Moreover, it is unknown if Antarctic wildlife harbor unique microbes because of the extreme climate and geographic isolation. Viral metagenomics analysis of Weddell seals (*Leptonychotes weddellii*) revealed the presence of anelloviruses, which are small, single-stranded DNA viruses (Appendix 1). Multiple PCR assays were designed using these sequences, resulting in the discovery of 152 anellovirus genomes encompassing two

species, TTLwV-1 and TTLwV-2. Exploring the viruses associated with Weddell seals is a step towards establishing baseline data that can monitored for changes over time.

The TTLwVs were ubiquitous among the apparently healthy Weddell seal population. This is not unusual as anelloviruses have been found in up to 100% of human populations (Ninomiya et al. 2008), with multiple genotypes co-occurring in the body without evidence of specific tropism. While they have been incidentally associated with many diseases, no direct causal evidence exists between TTV and a specific disease (Okamoto 2009, Spandole et al. 2015). However, despite the lack of correlation between anelloviruses and any specific disease, they are indicative of immune system health in humans by acting as a biomarker for immune suppression, with high titers of TTVs indicating immunosuppression (Eloit et al. 2015). It is not known if this correlation is true for Weddell seals or other marine mammals in which anelloviruses have been described. Since Weddell seals have relatively naïve immune systems, it would be interesting for future studies to explore the potential link between immune system strength to anellovirus quantity in this species (Yochem et al. 2009).

Virus research has typically focused on pathogens, but recent advances in virus discovery have introduced the emerging paradigm that viromes are more than a collection of pathogens and that viruses not correlated with disease may still have implications for health. Original studies of the microbiome were limited to bacteria and the symbiosis between microbe and host, but recent studies have indicated the virome ranges from pathogens, to innocuous viruses that may be harmful in a small fraction of the population, to commensal and even symbiotic viruses (Griffiths 1999, Bernardin et al. 2010, Haynes and Rohwer 2011, Virgin 2014, Cadwell 2015, Moustafa et al. 2017). For example, it has been shown that certain eukaryotic viruses in the human gastrointestinal tract may have parallel functions to beneficial bacteria living there (Kernbauer et

al. 2014, Cadwell 2015). Eukaryotic viruses have the capacity to support intestinal homeostasis and shape mucosal linings (Kernbauer et al. 2014), therefore the potential contribution of viruses to health should not be ignored, even if there is no apparent disease association.

Metagenomics is also a powerful tool for exploring viruses associated with disease, with the caveat that the presence of a virus does not necessarily mean it is associated with disease. From the leg of a sea star displaying symptoms of SSWD, a novel circular virus, Asterias forbesi-associated circular virus (AfaCV), was discovered using viral metagenomics (Appendix 2). The genome organization is typical of a circular rep-encoding ssDNA (CRESS-DNA) virus, but it is a genetically divergent virus from previously characterized genomes. Although AfaCV is not associated with SSWD, this was the first CRESS-DNA virus discovered in echinoderms, and is therefore an important discovery. Since then, CRESS-DNA viruses have been discovered in other echinoderms including the green sea urchin (*Strongylocentrotus droebachiensis*) and giant California sea cucumber (*Parastichopus californicus*) (Jackson et al. 2016). Metagenomic studies reveal that CRESS-DNA viruses may be more prevalent in marine systems than originally thought (Rosario et al. 2012, Rosario et al. 2015). Although AfaCV was not involved in the development of SSWD, a metagenomics study investigating SSWD in asteroids from the Northeast Pacific Ocean implicated a densovirus (sea star-associated densovirus; SSaDV) as the most likely cause of SSWD in the region (Hewson et al. 2014).

Viral metagenomics is a critical first step in disease surveillance, as identification of a viral sequence allows the virus to be further investigated with diagnostic tests such as PCR, in situ hybridization, serology, and analysis of replication. For example, AfaCV was originally identified in a sea star displaying symptoms of SSWD, but when a PCR assay was designed to screen for prevalence, it was only found in 10% of samples, indicating it is not associated with

the disease, despite its initial disease-related discovery. PCR was used to explore SSaDV in SSWD in East coast Forbes sea star. Twenty three of the 48 samples (48%) were positive for SSaDV; however, this may be an underestimation of true SSaDV prevalence due the use of standard PCR, which is less sensitive than quantitative PCR using an internal TaqMan probe. Nevertheless, SSaDV appears to be associated with SSWD in Atlantic Forbes sea stars, similar to the findings of Hewson et al. (2014) in the Pacific.

Unfortunately, viral metagenomic sequencing is not an effective method for capturing all viruses, including ChHV5 (Appendix 3). To limit host nucleic acids from tissue samples for sequencing viral nucleic acids, a combination of filtering, chloroform and nuclease treatment is used (Ng et al. 2009, Ng et al. 2009). For enveloped viruses such as ChHV5, chloroform will destroy the enveloped membrane, possibly rendering the viral DNA susceptible to nuclease treatment (Allander et al. 2001). It is hypothesized that ChHV5 is in the episomal form (Alfaro-Núñez and Gilbert 2014, Work et al. 2014), which would also be destroyed during the nuclease step. PCR amplicon enrichment and target enrichment followed by next-generation sequencing are effective, sequence-dependent methods useful for the recovery of viral genomes from directly from tissue, without the requirement of culturing (Cotten et al. 2014, Depledge et al. 2014, Tsangaras et al. 2014, Gardy et al. 2015, Wylie et al. 2015, Quick et al. 2016). Using primers based on a reference ChHV5 genome sequenced from the glottis tumor of a Hawaiian green sea turtle using BAC cloning (Ackermann et al. 2012), long-range PCR followed by Illumina sequencing of amplicons was used to sequence the ChHV5 genomes from several species of Florida sea turtles. Since the disease manifests differently between Hawaii and Florida it was important to determine if genetic differences of the virus contributed to those differences.

Generally, whole genomes are not obtained through metagenomic sequencing. Whole genome sequencing (WGS) is necessary both for novel and previously described viruses. Through WGS, molecular epidemiology, evolutionary genomics, and the identification of interesting genes and recombination events is more effective (Houldcroft et al. 2017). This dissertation highlights the importance of WGS through the development of back-to-back primers based on metagenomic sequences, which led to the discovery of two anellovirus species encompassing 43 genotypes associated with Weddell seals. This diversity of the TTLwVs would not have been captured if the genomes were not sequenced.

It is necessary to pair WGS methods with traditional diagnostic methods to elucidate the role of the virus in disease, but this can often pose ethical and technical challenges, especially when the virus is unculturable. Obtaining tissue samples from wildlife is difficult, especially if the species is protected. Creating serology-based assays is an option to screen live animals for exposure to the virus and to determine whether the animal is mounting an immune response to the virus. The ELISA developed to test for ZcAV exposure is an example of how a newly discovered virus is a jumping off point for creating new diagnostic assays (Appendix 4). ZcAV was originally found in a protected species and was not detectable in the blood. Since there was no way to screen for ZcAV in live sea lions, an ELISA was developed to screen for antibodies to ZcAV. Furthermore, ZcAV was identified from a mixed community of viruses and has not been . This ELISA provides a way to screen for exposure to this virus for prevalence and epidemiological studies. Since many newly discovered viruses are unculturable and/or have difficulties associated with sampling the host, moving toward developing an ELISA is one step towards linking virus to disease. Developing an ELISA using synthesized peptides predicted from next-generation sequencing data is currently used for exploring plant viruses, but as more

unculturable viruses are discovered, this may become more common for animal virus (Boonham et al. 2014).

This dissertation explores the strengths and limitation of viral metagenomics for characterizing viruses and their role in health. Viral metagenomics identified sequences of novel viruses (appendices 1 and 2) and expanded the known host range of CRESS-DNA viruses. This method is useful for recovering viruses from healthy and diseased animals, but is limited in its ability to elucidate virus-host interactions. Therefore, traditional diagnostic tools developed from the genetic information obtained through metagenomics are required to examine the link between virus and host (Appendix 4). Another limitation is the ability to detect all viruses (Appendix 3). Viruses at low copy numbers, and those in the latent stage, cannot be purified using typical virus isolation methods, and consequently will not be detected. Although attempted, ChHV5 was not recovered using viral metagenomic techniques. Therefore, other WGS methods are being used to complete the genome.

The appropriate WGS technique should be determined based on whether a targeted or a non-targeted approach is needed to achieve the goal of the study. Metagenomics is most efficiently used as a non-targeted approach. If metagenomics is used while looking for a specific virus, generally very few reads are recovered from the target virus. For example, metagenomic analysis resulted in only 0.008% EBV reads in the blood of a healthy adult (Allen et al. 2016). Samples can be enriched for viruses, which increases read coverage and proportion, but the proportion of reads are still relatively low when compared to the amount of host-derived sequences (Matranga et al. 2014, Calvet et al. 2016). The advantage of metagenomics is that it is effective for identifying viral diversity and emerging viruses. If a specific virus is being

investigated, then a targeted approach like PCR amplicon sequencing or target enrichment may be effective (Thomson et al. 2016).

In the past, diseases caused significant mortality or morbidity before the etiological was discovered. The metagenomic approach provides the potential for proactive pathogen surveillance, instead of a reactive approach that is typically used during an epidemic. Benefits of virus surveillance include (i) identifying genetic clues that explain the jump of a virus from one host to the next, (ii) understanding drivers of disease emergence by tracking virus diversity and evolution during land-use changes, (iii) deciphering genetic changes that increase virulence, and (iv) providing time for the development of vaccines.

No disease has been successfully predicted before its outbreak in humans, which is partially due to the lack of knowledge of viral diversity in wildlife (Jones et al. 2008, Morse et al. 2012). Assessing the viromes associated with animals that frequently carry zoonotic viruses such as bats, rodents, blood-sucking arthropods, and animals that come in close contact with humans or are consumed by humans is one way to increase surveillance efficiency. Another way to increase the efficiency of disease forecasting is to focus on the interface between wildlife and humans. Detecting a pathogen before the human encroachment of wildlife, allows management to take prophylactic measures against disease outbreak. Vector-enabled metagenomics, the targeting of disease vectors for metagenomic analysis, is productive for characterizing viromes within an ecosystem (Ng et al. 2011, Brinkmann et al. 2016). PCR assays can be subsequently used to target specific viruses of interest and identify their hosts.

Understudied ecosystems and the animals that inhabit them should not be ignored. Although zoonotic events are often short term, it is not known where many well-established human viruses originated. Many animals in the marine environment remain unexplored, despite

the increase in disease emergence and significant mortality events that occasionally follow (Harvell et al. 1999, Gulland and Hall 2007, Anthony et al. 2012, Hewson et al. 2014). As more viruses are being characterized from the marine environment, the link between terrestrial and marine disease is becoming evident. The viral disease distemper and caliciviruses can move between marine and terrestrial environments (Smith et al. 1998, Philippa et al. 2004). A novel picornavirus recently discovered in seals is related to human hepatitis A (Anthony et al. 2015), indicating that well established viruses in humans may originate from the ocean. Viruses of terrestrial animals may have originated from the marine environment and therefore marine animals should not be ignored as a reservoir of potential emerging diseases.

Advances in sequencing technology have led to an influx of newly identified viruses (Massart et al. 2017, Moustafa et al. 2017), with the overwhelming majority of them lacking a culture system (Fredricks and Relman 1996). This has led to proposed revisions of Koch's postulates to better suit viruses and technological advances (Rivers 1937, Falkow 1988, Fredricks and Relman 1996, Mokili et al. 2012). While Koch's postulates remain the gold standard for determining etiology, molecular-based methods have been used to bypass fulfilling the postulates. Merkel cell polyomavirus (MCV), human papillomavirus (HPV), and human immunodeficiency virus (HIV) were identified as the causative agents of Merkel's cell carcinoma, cervical cancer, AIDS, respectively, without satisfying Koch's postulates (Montagnier et al. 1987, Walboomers et al. 1999, Feng et al. 2008).

Metagenomic studies are limited by the ability for taxonomic identification of sequences based on comparisons against public databases using similarity-based searches, such as BLAST (Altschul et al. 1990, Huson et al. 2007). Unfortunately, most sequences obtained through viral metagenomics have no sequence homology to any sequences in the database, which is

incomplete and does not represent total viral diversity (Edwards and Rohwer 2005, McHardy et al. 2007). In cases where similarity-based sequence identification fails, composition-based identification methods are used based on GC content and tetranucleotide signatures (Teeling et al. 2004). Composition-based identification has mostly been used for identifying bacteria, but it remains a promising area for viral identification (Fancello et al. 2012). *In silico* advancements are needed to identify unknown reads as viruses, thus contributing to a more robust database of viral sequences. Quick identification of emerging viruses and their hosts will result in a rapid response to enhance the safety of public health. This dissertation describes three new viruses that are now in the NCBI database to aid with viral identification in future studies.

Viral metagenomics is a salient technology that will enhance the understanding of viral diversity, transmission, and emergence. By continuing to describe viruses from healthy and diseased animals, the role of viruses in animal health will become apparent and contribute to the paradigm shift from "all viruses are pathogens" to the idea of a normal, healthy virome. As virus discovery continues, research needs to move towards a holistic approach. The outcome of viral infection is largely due to context, meaning that host genotype, coinfections, presence of commensal bacteria, and environmental factors all contribute to the phenotypic outcome of the virus. As viral diversity and dynamics are better characterized, disease prediction and forecasting should eventually be possible (Anthony et al. 2015).

REFERENCES

Ackermann, M., M. Koriabine, F. Hartmann-Fritsch, P. J. de Jong, T. D. Lewis, N. Schetle, T. M. Work, J. Dagenais, G. H. Balazs and J.-A. C. Leong (2012). "The genome of chelonid herpesvirus 5 harbors atypical genes." PLOS ONE **7**(10): e46623.

Aguirre, A. A. and P. Lutz (2004). "Marine turtles as sentinels of ecosystem health: Is fibropapillomatosis an indicator?" EcoHealth **1**(3): 275-283.

Alfaro-Núñez, A. and M. T. P. Gilbert (2014). "Validation of a sensitive pcr assay for the detection of chelonid fibropapilloma-associated herpesvirus in latent turtle infections." Journal of Virological Methods. 206 (2014): 38-41.

Allander, T., S. U. Emerson, R. E. Engle, R. H. Purcell and J. Bukh (2001). "A virus discovery method incorporating dnase treatment and its application to the identification of two bovine parvovirus species." Proceedings of the National Academy of Sciences **98**(20): 11609-11614.

Allen, L. Z., T. Ishoey, M. A. Novotny, J. S. McLean, R. S. Lasken and S. J. Williamson (2011). "Single virus genomics: A new tool for virus discovery." PLOS ONE **6**(3): e17722.

Allen, U. D., P. Hu, S. L. Pereira, J. L. Robinson, T. A. Paton, J. Beyene, N. Khodai-Booran, A. Dipchand, D. Hébert, V. Ng, T. Nalpathamkalam and S. Read (2016). "The genetic diversity of Epstein–Barr virus in the setting of transplantation relative to non-transplant settings: A feasibility study." Pediatric Transplantation **20**(1): 124-129.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman (1990). "Basic local alignment search tool." Journal of Molecular Biology **215**(3): 403-410.

Anthony, S. J., A. Islam, C. Johnson, I. Navarrete-Macias, E. Liang, K. Jain, P. L. Hitchens, X. Che, A. Soloyvov, A. L. Hicks, R. Ojeda-Flores, C. Zambrana-Torrelio, W. Ulrich, M. K. Rostal, A. Petrosov, J. Garcia, N. Haider, N. Wolfe, T. Goldstein, S. S. Morse, M. Rahman, J. H. Epstein, J. K. Mazet, P. Daszak and W. I. Lipkin (2015). "Non-random patterns in viral diversity." Nature Communications **6**: 8147.

Anthony, S. J., J. A. St. Leger, E. Liang, A. L. Hicks, M. D. Sanchez-Leon, K. Jain, J. H. Lefkowitch, I. Navarrete-Macias, N. Knowles, T. Goldstein, K. Pugliares, H. S. Ip, T. Rowles and W. I. Lipkin (2015). "Discovery of a novel hepatovirus (phopivirus of seals) related to human hepatitis A virus." mBio **6**(4).

Anthony, S. J., J. A. St. Leger, K. Pugliares, H. S. Ip, J. M. Chan, Z. W. Carpenter, I. Navarrete-Macias, M. Sanchez-Leon, J. T. Saliki, J. Pedersen, W. Karesh, P. Daszak, R. Rabadan, T. Rowles and W. I. Lipkin (2012). "Emergence of fatal avian influenza in new england harbor seals." mBio **3**(4).

Appawu, M. A., S. K. Dadzie, A. Baffoe-Wilmot and M. D. Wilson (2001). "Lymphatic filariasis in ghana: Entomological investigation of transmission dynamics and intensity in communities served by irrigation systems in the upper east region of ghana." Tropical Medicine & International Health **6**(7): 511-516.

Barragan, A. R. and L. Sarti (1994). "A possible case of fibropapilloma in Kemp's ridley turtle (*Lepidochelys kempii*)." Marine Turtle Newsletter **67**(28).

Bernardin, F., E. Operskalski, M. Busch and E. Delwart (2010). "Transfusion transmission of highly prevalent commensal human viruses." Transfusion **50**(11): 2474-2483.

Boonham, N., J. Kreuze, S. Winter, R. van der Vlugt, J. Bergervoet, J. Tomlinson and R. Mumford (2014). "Methods in virus diagnostics: From ELISA to next generation sequencing." Virus Research **186**: 20-31.

Bossart, G. D. (2006). "Marine mammals as sentinel species for oceans and human health." Oceanography **19**(2): 134-137.

Bossart, G. D., R. Y. Ewing, M. Lowe, M. Sweat, S. J. Decker, C. J. Walsh, S.-j. Ghim and A. B. Jenson (2002). "Viral papillomatosis in Florida manatees (*Trichechus manatus latirostris*)." Experimental and Molecular Pathology **72**(1): 37-48.

Breitbart, M., P. Salamon, B. Andresen, J. M. Mahaffy, A. M. Segall, D. Mead, F. Azam and F. Rohwer (2002). "Genomic analysis of uncultured marine viral communities." Proceedings of the National Academy of Sciences **99**(22): 14250-14255.

Brinkmann, A., A. Nitsche and C. Kohl (2016). "Viral metagenomics on blood-feeding arthropods as a tool for human disease surveillance." International Journal of Molecular Sciences **17**(10): 1743.

Brown, C. (2004). "Emerging zoonoses and pathogens of public health significance--an overview." Revue scientifique et technique-office international des epizooties **23**(2): 435-442.

Cadwell, K. (2015). "Expanding the role of the virome: Commensalism in the gut." Journal of Virology **89**(4): 1951-1953.

Cadwell, K. (2015). "The virome in host health and disease." Immunity **42**(5): 805-813.

Calvet, G., R. S. Aguiar, A. S. O. Melo, S. A. Sampaio, I. de Filippis, A. Fabri, E. S. M. Araujo, P. C. de Sequeira, M. C. L. de Mendonça, L. de Oliveira, D. A. Tschoeke, C. G. Schrago, F. L. Thompson, P. Brasil, F. B. dos Santos, R. M. R. Nogueira, A. Tanuri and A. M. B. de Filippis (2016). "Detection and sequencing of zika virus from amniotic fluid of fetuses with microcephaly in Brazil: A case study." The Lancet Infectious Diseases **16**(6): 653-660.

Cotten, M., V. Petrova, M. V. T. Phan, M. A. Rabaa, S. J. Watson, S. H. Ong, P. Kellam and S. Baker (2014). "Deep sequencing of norovirus genomes defines evolutionary patterns in an urban tropical setting." Journal of Virology **88**(19): 11056-11069.

Delwart, E. L. (2007). "Viral metagenomics." Reviews in Medical Virology **17**(2): 115-131.

Depledge, D. P., S. Kundu, N. J. Jensen, E. R. Gray, M. Jones, S. Steinberg, A. Gershon, P. R. Kinchington, D. S. Schmid, F. Balloux, R. A. Nichols and J. Breuer (2014). "Deep sequencing of viral genomes provides insight into the evolution and pathogenesis of varicella zoster virus and its vaccine in humans." Molecular Biology and Evolution **31**(2): 397-409.

Dobson, A. and J. Foufopoulos (2001). "Emerging infectious pathogens of wildlife." Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **356**(1411): 1001-1012.

Edwards, R. A. and F. Rohwer (2005). "Viral metagenomics." Nat Rev Micro **3**(6): 504-510.

Eloit, M., J. Cheval, C. Hebert and M. Lecuit (2015). Anellovirus genome quantification as a biomarker of immune suppression, Google Patents.

Ene, A., M. Su, S. Lemaire, C. Rose, S. Schaff, R. Moretti, J. Lenz and L. H. Herbst (2005). "Distribution of chelonid fibropapillomatosis-associated herpesvirus variants in Florida: Molecular genetic evidence for infection of turtles following recruitment to neritic developmental habitats." J Wildl Dis **41**(3): 489-497.

Epstein, P. R., E. Chivian and K. Frith (2003). "Emerging diseases threaten conservation." Environmental Health Perspectives **111**(10): A506-A507.

Evans, A. S. (1976). "Causation and disease: The Henle-Koch postulates revisited." The Yale Journal of Biology and Medicine **49**(2): 175-195.

Falkow, S. (1988). "Molecular Koch's postulates applied to microbial pathogenicity." Review of Infectious Diseases **10**(Supplement 2): S274-S276.

Fancello, L., D. Raoult and C. Desnues (2012). "Computational tools for viral metagenomics and their application in clinical research." Virology **434**(2): 162-174.

Feng, H., M. Shuda, Y. Chang and P. S. Moore (2008). "Clonal integration of a polyomavirus in human merkel cell carcinoma." Science **319**(5866): 1096-1100.

Fredricks, D. N. and D. A. Relman (1996). "Sequence-based identification of microbial pathogens: A reconsideration of Koch's postulates." Clin Microbiol Rev **9**(1): 18-33.

Gardy, J. L., M. Naus, A. Amlani, W. Chung, H. Kim, M. Tan, A. Severini, M. Krajden, D. Puddicombe, V. Sahni, A. S. Hayden, R. Gustafson, B. Henry and P. Tang (2015). "Whole-genome sequencing of measles virus genotypes H1 and D8 during outbreaks of infection following the 2010 olympic winter games reveals viral transmission routes." The Journal of Infectious Diseases **212**(10): 1574-1578.

Griffiths, P. (1999). "Time to consider the concept of a commensal virus?" Reviews in Medical Virology **9**(2): 73-74.

Gulland, F. M. D. and A. J. Hall (2007). "Is marine mammal health deteriorating? Trends in the global reporting of marine mammal disease." EcoHealth **4**(2): 135-150.

Harb, M., R. Faris, A. M. Gad, O. N. Hafez, R. Ramzy and A. A. Buck (1993). "The resurgence of lymphatic filariasis in the nile delta." Bulletin of the World Health Organization **71**(1): 49-54.

Harvell, C. D., K. Kim, J. M. Burkholder, R. R. Colwell, P. R. Epstein, D. J. Grimes, E. E. Hofmann, E. K. Lipp, A. D. M. E. Osterhaus, R. M. Overstreet, J. W. Porter, G. W. Smith and G. R. Vasta (1999). "Emerging marine diseases--climate links and anthropogenic factors." Science **285**(5433): 1505.

Haynes, M. and F. Rohwer (2011). The human virome. Metagenomics of the human body. K. E. Nelson. New York, NY, Springer New York**:** 63-77.

Herbst, L., A. Ene, M. Su, R. Desalle and J. Lenz (2004). "Tumor outbreaks in marine turtles are not due to recent herpesvirus mutations." Curr Biol **14**(17): R697-699.

Herbst, L. H. (1994). "Fibropapillomatosis of marine turtles." Annual Review of Fish Diseases **4**(0): 389-425.

Hewson, I., J. B. Button, B. M. Gudenkauf, B. Miner, A. L. Newton, J. K. Gaydos, J. Wynne, C. L. Groves, G. Hendler, M. Murray, S. Fradkin, M. Breitbart, E. Fahsbender, K. D. Lafferty, A. M. Kilpatrick, C. M. Miner, P. Raimondi, L. Lahner, C. S. Friedman, S. Daniels, M. Haulena, J. Marliave, C. A. Burge, M. E. Eisenlord and C. D. Harvell (2014). "Densovirus associated with sea-star wasting disease and mass mortality." Proceedings of the National Academy of Sciences **111**(48): 17278-17283.

Houldcroft, C. J., M. A. Beale and J. Breuer (2017). "Clinical and biological insights from viral genome sequencing." Nat Rev Micro **15**(3): 183-192.

Houldcroft, C. J., J. M. Bryant, D. P. Depledge, B. K. Margetts, J. Simmonds, S. Nicolaou, H. J. Tutill, R. Williams, A. J. J. Worth, S. D. Marks, P. Veys, E. Whittaker and J. Breuer (2016). "Detection of low frequency multi-drug resistance and novel putative maribavir resistance in immunocompromised pediatric patients with cytomegalovirus." Frontiers in Microbiology **7**: 1317.

Huson, D. H., A. F. Auch, J. Qi and S. C. Schuster (2007). "Megan analysis of metagenomic data." Genome Research **17**(3): 377-386.

Jackson, E. W., K. S. I. Bistolas, J. B. Button and I. Hewson (2016). "Novel circular single-stranded DNA viruses among an Asteroid, Echinoid and Holothurian (phylum: Echinodermata)." PLOS ONE **11**(11): e0166093.

Jones, K. E., N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman and P. Daszak (2008). "Global trends in emerging infectious diseases." Nature **451**(7181): 990-993.

Kernbauer, E., Y. Ding and K. Cadwell (2014). "An enteric virus can replace the beneficial function of commensal bacteria." Nature **516**(7529): 94-98.

King, D. P., M. C. Hure, T. Goldstein, B. M. Aldridge, F. M. D. Gulland, J. T. Saliki, E. L. Buckles, L. J. Lowenstine and J. L. Stott (2002). "Otarine herpesvirus-1: A novel gammaherpesvirus associated with urogenital carcinoma in california sea lions (*Zalophus californianus*)." Veterinary Microbiology **86**(1–2): 131-137.

Koch, R. and T. Saure (1890). Aetiology of tuberculosis, William R. Jenkins.

Lackovich, J. K., D. R. Brown, B. L. Homer, R. L. Garber, D. R. Mader, R. H. Moretti, A. D. Patterson, L. H. Herbst, J. Oros, E. R. Jacobson, S. S. Curry and P. A. Klein (1999). "Association of herpesvirus with fibropapillomatosis of the green turtle *chelonia mydas* and the loggerhead turtle caretta caretta in florida." Dis Aquat Organ **37**(2): 89-97.

Lederberg, J., M. A. Hamburg and M. S. Smolinski (2003). Microbial threats to health: Emergence, detection, and response, National Academies Press.

Lee Miller, D., R. Ewing and G. Bossart (2001). Emerging and resurging diseases. Crc handbook of marine mammal medicine, CRC Press**:** 15-30.

Lipkin, W. I. (2010). "Microbe hunting." Microbiology and Molecular Biology Reviews **74**(3): 363-377.

Massart, S., T. Candresse, J. Gil, C. Lacomme, L. Predajna, M. Ravnikar, J.-S. Reynard, A. Rumbou, P. Saldarelli, D. Škorić, E. J. Vainio, J. P. T. Valkonen, H. Vanderschuren, C. Varveri and T. Wetzel (2017). "A framework for the evaluation of biosecurity, commercial, regulatory, and scientific impacts of plant viruses and viroids identified by ngs technologies." Frontiers in Microbiology **8**: 45.

Matranga, C. B., K. G. Andersen, S. Winnicki, M. Busby, A. D. Gladden, R. Tewhey, M. Stremlau, A. Berlin, S. K. Gire, E. England, L. M. Moses, T. S. Mikkelsen, I. Odia, P. E. Ehiane, O. Folarin, A. Goba, S. H. Kahn, D. S. Grant, A. Honko, L. Hensley, C. Happi, R. F. Garry, C. M. Malboeuf, B. W. Birren, A. Gnirke, J. Z. Levin and P. C. Sabeti (2014). "Enhanced methods for unbiased deep sequencing of lassa and ebola rna viruses from clinical and biological samples." Genome Biology **15**(11): 519.

McCallum, H. I., A. Kuris, C. D. Harvell, K. D. Lafferty, G. W. Smith and J. Porter (2004). "Does terrestrial epidemiology apply to marine systems?" Trends in Ecology & Evolution **19**(11): 585-591.

McHardy, A. C., H. G. Martin, A. Tsirigos, P. Hugenholtz and I. Rigoutsos (2007). "Accurate phylogenetic classification of variable-length DNA fragments." Nature methods **4**(1): 63-72.

Messenger, A. M., A. N. Barnes and G. C. Gray (2014). "Reverse zoonotic disease transmission (zooanthroponosis): A systematic review of seldom-documented human biological threats to animals." PLOS ONE **9**(2): e89055.

Mokili, J. L., F. Rohwer and B. E. Dutilh (2012). "Metagenomics and future perspectives in virus discovery." Current Opinion in Virology **2**(1): 63-77.

Montagnier, L., J. C. Chermann, F. Barre-Sinoussi, F. Brun-Vezinet, C. Rouzioux, W. Rozenbaum, C. Dauguet, J. Gruest, M. T. Nugeyre and F. Rey (1987). Human immunodeficiency viruses associated with acquired immune deficiency syndrome (AIDS), a diagnostic method for aids and pre-aids, and a kit therefor, Google Patents.

Mora, C., D. P. Tittensor, S. Adl, A. G. B. Simpson and B. Worm (2011). "How many species are there on earth and in the ocean?" PLOS Biology **9**(8): e1001127.

Morse, S. S. (1991). "Emerging viruses: Defining the rules for viral traffic." Perspect Biol Med **34**(3): 387-409.

Morse, S. S., J. A. K. Mazet, M. Woolhouse, C. R. Parrish, D. Carroll, W. B. Karesh, C. Zambrana-Torrelio, W. I. Lipkin and P. Daszak (2012). "Prediction and prevention of the next pandemic zoonosis." The Lancet **380**(9857): 1956-1965.

Moustafa, A., C. Xie, E. Kirkness, W. Biggs, E. Wong, Y. Turpaz, K. Bloom, E. Delwart, K. E. Nelson, J. C. Venter and A. Telenti (2017). "The blood DNA virome in 8,000 humans." PLOS Pathogens **13**(3): e1006292.

Ng, T. F. F., C. Manire, K. Borrowman, T. Langer, L. Ehrhart and M. Breitbart (2009). "Discovery of a novel single-stranded DNA virus from a sea turtle fibropapilloma by using viral metagenomics." Journal of Virology **83**(6): 2500-2509.

Ng, T. F. F., W. K. Suedmeyer, E. Wheeler, F. Gulland and M. Breitbart (2009). "Novel anellovirus discovered from a mortality event of captive california sea lions." Journal of General Virology **90**(5): 1256-1261.

Ng, T. F. F., E. Wheeler, D. Greig, T. B. Waltzek, F. Gulland and M. Breitbart (2011). "Metagenomic identification of a novel anellovirus in pacific harbor seal (*Phoca vitulina richardsii*) lung samples and its detection in samples from multiple years." Journal of General Virology **92**(6): 1318-1323.

Ng, T. F. F., D. L. Willner, Y. W. Lim, R. Schmieder, B. Chau, C. Nilsson, S. Anthony, Y. Ruan, F. Rohwer and M. Breitbart (2011). "Broad surveys of DNA viral diversity obtained through viral metagenomics of mosquitoes." PLOS ONE **6**(6): e20579.

Ninomiya, M., M. Takahashi, T. Nishizawa, T. Shimosegawa and H. Okamoto (2008). "Development of PCR assays with nested primers specific for differential detection of three human anelloviruses and early acquisition of dual or triple infection during infancy." Journal of Clinical Microbiology **46**(2): 507-514.

Okamoto, H. (2009). "History of discoveries and pathogenicity of tt viruses." Curr Top Microbiol Immunol **331**: 1-20.

Patz, J. A., U. Confalonieri, F. Amerasinghe, K. Chua, P. Daszak, A. Hyatt, D. Molyneux, M. Thomson, L. Yameogo and M. Lazaro (2005). "Human health: Ecosystem regulation of infectious diseases." Ecosystems and Human Well-Being: Current State and Trends: Findings of the Condition and Trends Working Group of the Millennium Ecosystem Assessment: 391-415.

Philippa, J. D., F. A. Leighton, P. Y. Daoust, O. Nielsen, M. Pagliarulo, H. Schwantje, T. Shury, R. Van Herwijnen, B. E. Martina, T. Kuiken, M. W. Van de Bildt and A. D. Osterhaus (2004). "Antibodies to selected pathogens in free-ranging terrestrial carnivores and marine mammals in canada." The Veterinary record **155**(5): 135-140.

Quick, J., N. J. Loman, S. Duraffour, J. T. Simpson, E. Severi, L. Cowley, J. A. Bore, R. Koundouno, G. Dudas, A. Mikhail, N. Ouédraogo, B. Afrough, A. Bah, J. H. J. Baum, B. Becker-Ziaja, J. P. Boettcher, M. Cabeza-Cabrerizo, Á. Camino-Sánchez, L. L. Carter, J. Doerrbecker, T. Enkirch, I. G. Dorival, N. Hetzelt, J. Hinzmann, T. Holm, L. E. Kafetzopoulou, M. Koropogui, A. Kosgey, E. Kuisma, C. H. Logue, A. Mazzarelli, S. Meisel, M. Mertens, J. Michel, D. Ngabo, K. Nitzsche, E. Pallasch, L. V. Patrono, J. Portmann, J. G. Repits, N. Y. Rickett, A. Sachse, K. Singethan, I. Vitoriano, R. L. Yemanaberhan, E. G. Zekeng, T. Racine, A. Bello, A. A. Sall, O. Faye, O. Faye, N. F. Magassouba, C. V. Williams, V. Amburgey, L. Winona, E. Davis, J. Gerlach, F. Washington, V. Monteil, M. Jourdain, M. Bererd, A. Camara, H. Somlare, A. Camara, M. Gerard, G. Bado, B. Baillet, D. Delaune, K. Y. Nebie, A. Diarra, Y. Savane, R. B. Pallawo, G. J. Gutierrez, N. Milhano, I. Roger, C. J. Williams, F. Yattara, K. Lewandowski, J. Taylor, P. Rachwal, D. J. Turner, G. Pollakis, J. A. Hiscox, D. A. Matthews, M. K. O. Shea, A. M. Johnston, D. Wilson, E. Hutley, E. Smit, A. Di Caro, R. Wölfel, K. Stoecker, E. Fleischmann, M. Gabriel, S. A. Weller, L. Koivogui, B. Diallo, S. Keïta, A. Rambaut, P. Formenty, S. Günther and M. W. Carroll (2016). "Real-time, portable genome sequencing for ebola surveillance." Nature **530**(7589): 228-232.

Renzette, N., B. Bhattacharjee, J. D. Jensen, L. Gibson and T. F. Kowalik (2011). "Extensive genome-wide variability of human cytomegalovirus in congenitally infected infants." PLOS Pathogens **7**(5): e1001344.

Rivers, T. M. (1937). "Viruses and Koch's postulates." Journal of Bacteriology **33**(1): 1-12.

Rosario, K. and M. Breitbart (2011). "Exploring the viral world through metagenomics." Current Opinion in Virology **1**(4): 289-297.

Rosario, K., S. Duffy and M. Breitbart (2012). "A field guide to eukaryotic circular single-stranded DNA viruses: Insights gained from metagenomics." Archives of Virology **157**(10): 1851-1871.

Rosario, K., R. O. Schenck, R. C. Harbeitner, S. N. Lawler and M. Breitbart (2015). "Novel circular single-stranded DNA viruses identified in marine invertebrates reveal high sequence diversity and consistent predicted intrinsic disorder patterns within putative structural proteins." Frontiers in Microbiology **6**(696).

Shi, M., X.-D. Lin, J.-H. Tian, L.-J. Chen, X. Chen, C.-X. Li, X.-C. Qin, J. Li, J.-P. Cao, J.-S. Eden, J. Buchmann, W. Wang, J. Xu, E. C. Holmes and Y.-Z. Zhang (2016). "Redefining the invertebrate RNA virosphere." Nature **540**(7634): 539-543.

Smith, A. W., D. E. Skilling, N. Cherry, J. H. Mead and D. O. Matson (1998). "Calicivirus emergence from ocean reservoirs: Zoonotic and interspecies movements." Emerging Infectious Diseases **4**(1): 13-20.

Spandole, S., D. Cimponeriu, L. M. Berca and G. Mihăescu (2015). "Human anelloviruses: An update of molecular, epidemiological and clinical aspects." Archives of Virology **160**(4): 893-908.

Svraka, S., K. Rosario, E. Duizer, H. van der Avoort, M. Breitbart and M. Koopmans (2010). "Metagenomic sequencing for virus identification in a public-health setting." Journal of General Virology **91**(11): 2846-2856.

Symonds, E. M., D. W. Griffin and M. Breitbart (2009). "Eukaryotic viruses in wastewater samples from the united states." Applied and Environmental Microbiology **75**(5): 1402-1409.

Taylor, L. H., S. M. Latham and M. E. Woolhouse (2001). "Risk factors for human disease emergence." Philos Trans R Soc Lond B Biol Sci **356**(1411): 983-989.

Teeling, H., A. Meyerdierks, M. Bauer, R. Amann and F. O. Glöckner (2004). "Application of tetranucleotide frequencies for the assignment of genomic fragments." Environmental Microbiology **6**(9): 938-947.

Thomson, E., C. L. C. Ip, A. Badhan, M. T. Christiansen, W. Adamson, M. A. Ansari, D. Bibby, J. Breuer, A. Brown, R. Bowden, J. Bryant, D. Bonsall, A. Da Silva Filipe, C. Hinds, E. Hudson, P. Klenerman, K. Lythgow, J. L. Mbisa, J. McLauchlan, R. Myers, P. Piazza, S. Roy, A. Trebes, S. B. Vattipally, J. Witteveldt, S.-H. consortium, E. Barnes and P. Simmonds (2016). "Comparison of next generation sequencing technologies for the comprehensive assessment of full-length hepatitis c viral genomes." Journal of Clinical Microbiology.

Tsangaras, K., N. Wales, T. Sicheritz-Pontén, S. Rasmussen, J. Michaux, Y. Ishida, S. Morand, M.-L. Kampmann, M. T. P. Gilbert and A. D. Greenwood (2014). "Hybridization capture using short PCR products enriches small genomes by capturing flanking sequences (capflank)." PLOS ONE **9**(10): e109101.

Urayama, S.-i., Y. Takaki and T. Nunoura (2016). "Flds: A comprehensive dsRNA sequencing method for intracellular RNA virus surveillance." Microbes and Environments **31**(1): 33-40.

Virgin, Herbert W. (2014). "The virome in mammalian physiology and disease." Cell **157**(1): 142-150.

Walboomers, J. M., M. V. Jacobs, M. M. Manos, F. X. Bosch, J. A. Kummer, K. V. Shah, P. J. Snijders, J. Peto, C. Meijer and N. Munoz (1999). "Human papillomavirus is a necessary cause of invasive cervical cancer worldwide." The Journal of pathology **189**(1): 12-19.

Wang, D., L. Coscoy, M. Zylberberg, P. C. Avila, H. A. Boushey, D. Ganem and J. L. DeRisi (2002). "Microarray-based detection and genotyping of viral pathogens." Proceedings of the National Academy of Sciences **99**(24): 15687-15692.

Wilcox, B. A. and D. J. Gubler (2005). "Disease ecology and the global emergence of zoonotic pathogens." Environmental Health and Preventive Medicine **10**(5): 263-272.

Williams, E. H. and L. Bunkley-Williams (2006). "Early fibropapillomas in Hawaii and occurrences in all sea turtle species: The panzootic, associated leeches wide-ranging on sea turtles, and species of study leeches should be identified." Journal of virology **80**(9): 4643-4644.

Work, T. M., J. Dagenais, G. H. Balazs, N. Schettle and M. Ackermann (2014). "Dynamics of virus shedding and in situ confirmation of chelonid herpesvirus 5 in Hawaiian green turtles with fibropapillomatosis." Veterinary Pathology **52**(6): 1195-1201.

Work, T. M., J. Dagenais, G. H. Balazs, J. Schumacher, T. D. Lewis, J. A. Leong, R. N. Casey and J. W. Casey (2009). "In vitro biology of fibropapilloma-associated turtle herpesvirus and host cells in Hawaiian green turtles (*chelonia mydas*)." J Gen Virol **90**(Pt 8): 1943-1950.

Wu, Q., Y. Luo, R. Lu, N. Lau, E. C. Lai, W.-X. Li and S.-W. Ding (2010). "Virus discovery by deep sequencing and assembly of virus-derived small silencing rnas." Proceedings of the National Academy of Sciences **107**(4): 1606-1611.

Wylie, T. N., K. M. Wylie, B. N. Herter and G. A. Storch (2015). "Enhanced virome sequencing using targeted sequence capture." Genome Research **25**(12): 1910-1920.

Ylitalo, G. M., J. E. Stein, T. Hom, L. L. Johnson, K. L. Tilbury, A. J. Hall, T. Rowles, D. Greig, L. J. Lowenstine and F. M. D. Gulland (2005). "The role of organochlorines in cancer-associated mortality in california sea lions (*Zalophus californianus*)." Marine Pollution Bulletin **50**(1): 30-39.

Yochem, P. K., B. S. Stewart, T. S. Gelatt and D. B. Siniff (2009). Health assessment of weddell seals, *Leptonychotes weddellii*, in Mcmurdo sound, Antarctica. Health of antarctic wildlife: A challenge for science and policy. K. R. Kerry and M. Riddle. Berlin, Heidelberg, Springer Berlin Heidelberg**:** 123-138.

Zur Hausen, H., H. Schulte-Holthausen, G. Klein, G. Henle, W. Henle, P. Clifford and L. Santesson (1970). "Epstein-Barr virus in Burkitt's lymphoma and nasopharyngeal carcinoma. [ii] ebv DNA in biopsies of burkitt tumours and anaplastic carcinomas of the nasopharynx." Nature **228**: 1056-1058.

APPENDIX 1: DIVERSE AND HIGHLY RECOMBINANT ANELLOVIRUSES ASSOCIATED WITH WEDDELL SEALS IN ANTARCTICA

This appendix was accepted to be published in Virus Evolution.

Fahsbender, E., J.M. Burns, S. Kim, S. Kraberger, G. Frankfurter, A.A. Eilers, M.R. Shero, R. Beltran, A. Kirkham, R. McCorkell, R. Berngartt, M.F. Male, G. Ballard, D.G. Ainley, M. Breitbart, A. Varsani. Diverse and highly recombinant anelloviruses associated with Weddell seals in Antarctica.

Diverse and highly recombinant anelloviruses associated with Weddell seals in Antarctica

Elizabeth Fahsbender[1], Jennifer M. Burns[2*], Stacy Kim[3], Simona Kraberger[4,5], Greg Frankfurter[6], Alice A. Eilers[7], Michelle R. Shero[2], Roxanne Beltran[2,8], Amy Kirkham[2,9], Robert McCorkell[10], Rachel Berngartt[11], Maketalena F. Male[5,12], Grant Ballard[13], David G. Ainley[14], Mya Breitbart[1], Arvind Varsani[4, 5, 15*]

[1] College of Marine Science, University of South Florida, Saint Petersburg, FL 33701, USA

[2] Department of Biological Sciences, University of Alaska Anchorage, 3211 Providence Drive, Anchorage, AK 99508, USA

[3] Moss Landing Marine Laboratories, Moss Landing, CA 95039, USA

[4] The Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine, School of Life sciences, Arizona State University, Tempe, AZ 85287-5001, USA

[5] School of Biological Sciences, University of Canterbury, Christchurch 8140, New Zealand

[6] Wildlife Health Center, School of Veterinary Medicine, University of California Davis, Davis, CA 95616, USA

[7] Pink Palace Museum, Memphis, TN 38111, USA

[8] Department of Biology and Wildlife, University of Alaska Fairbanks, PO Box 756100, Fairbanks, AK 99775, USA

[9] College of Fisheries and Ocean Sciences, University of Alaska Fairbanks, 17101 Point Lena Loop Rd. Juneau, Alaska 99801, USA

[10] Faculty of Veterinary Medicine, University of Calgary, Calgary, Alberta, Canada

[11] Bridge Veterinary Services, LLC, Juneau, AK 99801, USA

[12] School of Environmental and Life Sciences, The University of Newcastle, Callaghan NSW 2308, Australia

[13] Point Blue Conservation Science, Petaluma, CA 94954, USA

[14] HT Harvey and Associates, Los Gatos, CA 95032, USA

[15] Structural Biology Research Unit, Department of Clinical Laboratory Sciences, University of Cape Town, Observatory, Cape Town, South Africa

* Corresponding authors:

Jennifer M Burns: jmburns@alaska.edu

Arvind Varsani: Arvind.varsani@asu.edu

Keywords: *Anelloviridae*, Weddell seal, South polar skua, Ross Sea, Antarctica

Abstract: 146 words

Text: 3810 words

Figures: 5

Supplementary tables: 3

Supplementary figure: 2

Supplementary Data: 1

GenBank Accession #s: KY246476 - KY246627

Abstract

The viruses circulating among Antarctic wildlife remain largely unknown. In an effort to identify viruses associated with Weddell seals (*Leptonychotes weddellii*) inhabiting the Ross Sea, vaginal and nasal swabs, and faecal samples were collected between November 2014 and February 2015. In addition, a Weddell seal kidney and South Polar skua (*Stercorarius maccormicki*) faeces were opportunistically sampled. Using high throughput sequencing, we identified and recovered 152 anellovirus genomes that share 63-70% genome-wide identities with other pinniped anelloviruses. Genome-wide pairwise comparisons coupled with phylogenetic analysis revealed two novel anellovirus species, tentatively named torque teno Leptonychotes weddellii virus (TTLwV) -1 and -2. TTLwV-1 (n=133, genomes encompassing 40 genotypes) is highly recombinant, whereas TTLwV-2 (n=19, genomes encompassing three genotypes) is relatively less recombinant. This study documents ubiquitous TTLwVs among Weddell seals in Antarctica with frequent co-infection by multiple genotypes, however, the role these anelloviruses play in seal health remains unknown.

Introduction

Infectious diseases are the leading causes of mass mortality in wildlife and the global uptick of emerging viral disease makes pathogen surveillance crucial for the protection of animal health (Blomström, 2011; Dobson and Foufopoulos, 2001; Morner et al., 2002). However, identifying viruses is difficult because traditional molecular techniques such as polymerase chain reaction (PCR) amplification and serology-based assays are only useful for the detection of known viruses and their close relatives (Symonds et al., 2009; Wang et al., 2002). Metagenomic approaches that enrich for viruses and use high throughput sequencing platforms are powerful approaches for revealing viral communities, including novel viruses, in animal populations (Delwart, 2007; Edwards and Rohwer, 2005; Rosario and Breitbart, 2011).

Parts of Antarctica, a continent characterized by extreme climate and isolation, are inhabited by unique wildlife. The dense breeding grounds of some animals creates an increased risk of infectious diseases spreading amongst the population at an epidemic scale (Kerry et al., 1999). Similarly, increased human activity on the continent through tourism and research bases is thought to have exposed wildlife to diseases previously attributed to domestic animals. Such findings have led to concerns about pathogen introduction associated with anthropogenic activities (Austin and Webster, 1993; Gardner et al., 1997; Olsen et al., 1996; Retamal et al., 2000; Torres, 2000). Unfortunately, little is known about the pathogens associated with Antarctic animals.

Within the context of viral pathogens identified in Antarctic wildlife, the use of sequencing approaches has led to the identification of some viruses in Antarctic penguins, i.e. Adélie penguin (*Pygoscelis adeliae*), Chinstrap penguins (*Pygoscelis antarctica*) and Gentoo penguins (*Pygoscelis papua*), including an adenovirus, paramyxoviruses, orthomyxoviruses, a polyomavirus and a papillomavirus (Hurt et al., 2016; Hurt et al., 2014; Lee et al., 2014; Lee et al., 2016; Thomazelli et al., 2010; Varsani et al., 2014; Varsani et al., 2015). Similarly, a polyomavirus has been identified in sharp-spined notothen (*Trematomus pennelii*), an Antarctic fish (Buck et al., 2016), a parapoxvirus (Tryland et al., 2005) and a polyomavirus (Varsani et al., 2017) in Weddell seals (*Leptonychotes weddellii*), and an adenovirus has been identified in South Polar skua (*Stercorarius maccormicki*) (Park et al., 2012). In addition, serology-based assays have enabled the detection of

a putative birnavirus and flavivirus (Gardner et al., 1997; Morgan and Westbury, 1981; Morgan et al., 1985) in penguins (Adélie penguin, Blue penguin; *Eudyptula minor* and Emperor penguin; *Aptenodytes forsteri*), and a putative herpesvirus in Antarctic seals, namely Weddell seals and crabeater seals (*Lobodon carcinophaga*) (Harder et al., 1991; Stenvers et al., 1992).

A recent health assessment testing Weddell seals for antibodies to specific known bacterial and viral pathogens indicated that this population remains relatively naïve, leaving them potentially vulnerable to mass die-offs due to their close living proximity and lack of herd immunity (Yochem et al., 2009). Due to this vulnerability, it is important to identify viruses associated with these populations. Viral surveys will provide insight into the viral diversity that is currently associated with these seals, and will provide the genetic information necessary to develop new molecular assays to assess the prevalence of identified viruses and begin to elucidate their impact on animal health.

As part of an ongoing study on Weddell seals in the Ross Sea, we opportunistically sampled faeces, and took nasal and vaginal swabs to identify viruses associated with these animals. In these samples, as well as in a kidney sample from a deceased Weddell seal and a faecal sample from a South Polar skua (a bird that scavenges placenta and carcasses of seals), we identified a diversity of anelloviruses. Anelloviruses are non-enveloped, circular, negative sense, single-stranded DNA viruses that belong to the family *Anelloviridae* (Biagini, 2009; Okamoto et al., 1998b). The first anellovirus, human torque teno virus (TTV), was discovered in a Japanese patient with post transfusion hepatitis of unknown etiology (Nishizawa et al., 1997). Since then, numerous anelloviruses have been characterized and grouped into 12 different genera, which have been found in a variety of hosts including pigs, wild boar, dogs, seals, sea lions, pine marteen, bats, horses, cats, sea turtles and a range of primates (Abe et al., 2000; Al-Moslih et al., 2007; Fahsbender et al., 2015; King, 2011; Martinez et al., 2006; Ng et al., 2009a; Ng et al., 2009b; Ng et al., 2011; Nishiyama et al., 2014; Okamoto et al., 2001; Romeo et al., 2000). Despite their ubiquity and ability to cause persistent infections, the etiology of anelloviruses remains unknown.

Most of what is known about anelloviruses is based on human TTV, which has a prevalence as high as 100% in some human populations (Ninomiya et al., 2008). Individuals frequently harbor

multiple TTV genotypes (Niel et al., 2000; Nishiyama et al., 2014) and these have been identified throughout the body including in cervical secretions, nasal secretions, the umbilical cord, kidneys, blood, gastric tissue, and sweat (Spandole et al., 2015). There is no indication of tropism, but patterns of genotype compartmentalization, similar to human immunodeficiency virus, have been documented within the host (Maggi et al., 1999).

Anelloviruses have been found to be highly diverse, even at the amino acid level of the coding open reading frames (ORFs), yet the genome organization remains relatively similar with at least two ORFs and a conserved untranslated region (UTR). ORF1 is the largest ORF and is predicted to encode the capsid protein, however, this has not been definitively confirmed (Kamahora et al., 2000; Okamoto et al., 2000). ORF2 encodes proteins thought to be involved in regulation of the innate and adaptive immune system, but the lack of an appropriate culture system has hindered the ability to determine the functionality of these proteins and TTV pathogenicity (Huang et al., 2012; Kakkola et al., 2009; Kakkola et al., 2007; Yu et al., 2007).

Here we analyze the genomes of the anelloviruses recovered from faeces, kidney, and vaginal and nasal swabs of Weddell seals and a faecal sample of a South Polar skua. Sequence analysis revealed two phylogenetically distinct anellovirus species that are prevalent in all sample types, indicating that anelloviruses are ubiquitous among Weddell seals.

Methods

Sample collection

As part of an ongoing diet study in the Ross Sea on Weddell seals, 42 Weddell seal faecal samples were opportunistically collected on the fast ice of McMurdo Sound (Antarctic) during the 2014/2015 field season. Even though there were tagged animals in the vicinity of the faeces, it was not possible to associate the sampled faeces to a particular tagged animal. In addition to these samples, we collected paired nasal and vaginal swabs from Nov 2014 to Feb 2015 25 adult female Weddell seals; for two additional animals, only nasal swabs were taken, and for three animals we also collected faecal samples (Supplementary Table 1). A subset of females sampled in Nov / Dec were resampled approximately 60 days later in Jan/Feb, providing an additional eight nasal and nine vaginal samples. The nasal and vaginal swabs were refrigerated and stored in UTM™ Viral

Transport Media (Copan, USA). During the field season, a kidney was sampled from the carcass of a 14-year-old female (Flipper Tag# 8714A; specimen ID 17461) and frozen at -80°C prior to analysis, see Varsani et al. (2017) for necropsy details. Finally, a South Polar skua faecal sample was collected off fresh snow at Cape Crozier, Ross Island.

All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. Weddell seal samples were collected under National Marine Fisheries Service Marine Mammal permit #17411, Antarctic Conservation Act permit #2014-003, and University of Alaska Anchorage's Institutional Animal Care and Use Committee approval #419971, with funding from the National Science Foundation grant ANT-1246463 to Jennifer M Burns. The skua faecal sample at Cape Crozier was collected under Animal Care and Use Permit #4130 through Oregon State University, Corvallis, OR, USA, and Antarctic Conservation Act Permit #2006-010 from NSF through H.T. Harvey & Associates.

Sample processing

For each sample, ~5 g of the faecal sample or tissue samples (in the case of the kidney) was resuspended in 20 ml of SM buffer (0.1 M NaCl, 50 mM Tris/HCl – pH 7.4, 10 mM $MgSO_4$) and homogenized by vortexing for 30 sec. The suspension was centrifuged at 10000 × g for 10 mins. Following this, the supernatant was sequentially filtered through 0.45 µm and 0.2 µm (pore size) syringe filters. Three grams of PEG 8000 (Sigma, USA) was added to each of the filtrates and the solution was mixed gently to resuspend the PEG. The resulting suspension was incubated overnight at 4°C to precipitate virions. The solution was centrifuged at 10000 *x g* for 20 mins and the resulting pellet was resuspended in 2 ml of SM buffer.

Viral DNA was extracted using the High Pure Viral Nucleic Acid Kit (Roche Diagnostics, USA) from the resuspended virions (200 µl) from the faecal and kidney samples, and 200 µl of the UTM™ Viral Transport Media in which the swabs were stored. We used rolling-circle amplification (RCA) using the TempliPhi™ kit (GE Healthcare, USA) to randomly amplify nucleic acids.

High throughput sequencing and sequence analysis

A 5 µl aliquot of the randomly-amplified DNA from each of the Weddell seals faecal samples, nasal swabs and vaginal swabs was taken, pooled and labeled as WSP, WSN and WSV, respectively. The enriched DNA from the Weddell seal kidney sample was labelled as WSK and the faecal sample from the South Polar skua as SKP. The DNA samples WSP, WSN, WSV, WSK and SKP were then processed to generate ~100 bp paired-end libraries for multiplex Illumina sequencing and sequenced on an Illumina 2500 (Illumina, USA) platform at Macrogen Inc. (Korea). The resulting paired-end reads were *de novo* assembled using ABySS v1.9 (Simpson et al., 2009) with a k-mer of 64. Contigs of >750 nts were analyzed for viral-like sequences using BLASTx (Altschul et al., 1990) against a local viral sequence database.

In all of the WSP, WSN, WSV, WSK and SKP *de novo* assembled contigs, we identified sequences with similarities to anelloviruses. Based on these sequences we designed four pairs of abutting primers (Supplementary Table 2) to screen and recover the complete anellovirus genomes from each individual sample. The RCA product from each sample was used as a template for PCR amplification using Kapa HiFi Hotstart DNA polymerase with the following thermal cycling conditions: (95°C for 3 mins; 25 cycles of 98°C for 20 sec, 60°C for 15 sec, 72°C for 2 mins and a final extension of 72°C for 3 min). The amplicons were resolved on a 0.7% agarose gel stained with SYBR Safe (ThermoFisher, USA) and ~2 kb size fragments were excised, gel purified and cloned into pJET1.2 plasmid vector (ThermoFisher, USA). The resulting recombinant plasmids (five from each positive sample type) were Sanger sequenced by primer walking at Macrogen Inc. (South Korea).

To investigate the anellovirus diversity detected in seals, the pairwise identities of the anellovirus genomes and ORF1 sequences were determined using SDT v1.2 (Muhire et al., 2014). All anellovirus sequences with a detectable complete ORF1 were downloaded from GenBank (on the 18th of March 2017). 727 ORF1 sequences (including 152 from this study) were translated, aligned using MUSCLE (Edgar, 2004) and then back translated. The resulting alignment was used to infer a Maximum likelihood phylogenetic tree using IQ-TREE (Nguyen et al., 2015) with GTR+I+G4 substitution model selected using ModelFinder (Kalyaanamoorthy et al., 2017). Branches with less than 60% bootstrap support (1000 bootstrap iterations) were collapsed using TreeGraph2 (Stover and Muller, 2010).

Evidence of recombination in the anelloviruses identified in this study was determined using RDP 4.58 (Martin et al., 2015) with default settings. Sequences were auto-masked for optimal recombination detection and only events detected with more than three different methods implemented in RDP 4.58 coupled with phylogenetic support for recombination and a p-value of <0.05 were considered credible.

Results

Anellovirus identification and genome characterization

Analysis of the contigs from the *de novo* assembled reads from Illumina sequencing of WSP, WSN, WSV, WSK and SKP samples revealed high abundance of anellovirus-like sequences, polymavirus-like sequences from WSK which is reported in Varsani et al. (2017), circular replication-associated protein encoding viral-like sequences and various DNA bacteriophages. Given the high abundance of anellovirus-like sequences, we decided to focus on these for this report. Since anelloviruses have small circular genomes, PCRs using abutting primers were performed to recover 152 anellovirus genomes from the various samples, ranging in size from 2105 to 2212 nts. Analysis of the genome-wide pairwise identity of these revealed that the seal associated anellovirus genomes share more than 64% pairwise identity (Figure 1). Of the 152 genomes identified, 74 genomes sharing more than 64% identity were recovered from seal faeces, 37 genomes sharing more than 70% identity from nasal swabs, 34 genomes sharing more than 72% identity from vaginal swabs, and four genomes sharing more than 73% identity from the kidney. In addition, three anellovirus genomes sharing more than 75% identity were recovered from South Polar skua faeces. These Antarctic anelloviruses share ~63-70% genome-wide identities (Supplementary Data 1) with other pinniped anelloviruses from Pacific harbor seals (*Phoca vitulina*) sampled in USA (Pacific coast) and the Netherlands (HQ287751, KF373758, KF373760, KM262781, KM262785).

Based on species demarcation criteria of 35% divergence of the ORF1 amino acid sequences endorsed by the International Committee for the Taxonomy of Viruses (ICTV) (King, 2011), the anelloviruses identified in this study represent two species for which we propose the name torque teno Leptonychotes weddellii virus (TTLwV) -1 and -2. This species demarcation is also supported

by the phylogenetic analysis of the ORF1 protein sequences of the TTLwVs (Figure 2). The genome organization of representatives from the two phylogenetically distinct anellovirus species, TTLwV-1 and TTLwV-2, are illustrated in Figure 1. Both genomes have the same organization, with three open reading frames and a hypothetical protein, and are approximately the same genome size, ~2.1 kb. However, there are differences in ORF size and position, specifically for ORF3.

A Maximum likelihood phylogenetic analysis of all available complete ORF1 nucleotide sequences from GenBank (n=727) show that TTLwVs are related to other anelloviruses discovered in pinnipeds (Figure 2), and most closely related to the Pacific harbor seal anelloviruses and torque teno Zalophus virus (Figures 2 and 3). In general, the ORF1 phylogenetic analysis (Supplementary Figure 1) shows some level of host specificity and within lineages there appears to be a significant level of concordance between the phylogenies of the anelloviruses and their hosts (host phylogeny inferred with TimeTree; http://www.timetree.org/; (Hedges et al., 2015; Kumar et al., 2017)). Furthermore, it is clear that there are two lineages of porcine-associated, pinniped-associated and rodent-associated anelloviruses. Thus, it is highly likely that there were multiple diverse anelloviruses that were circulating amongst the most recent common ancestor (MRCA) of mammals. Within the hominoid-associated anelloviruses, those from chimpanzees (n=10) and gorillas (n=3) appear to be interspersed with those from humans (Figure 2 and Supplementary figure 1). Within the primate lineage, given the depth of sampling of non-human primates, it is difficult to test for a coevolution hypothesis or infer any cross-species transmission events. The mosquito-associated anelloviruses are almost certainly derived from a vertebrate blood meal (see taxa marked with * in Supplementary Figure 2).

TTLwV-1 (n=133) was identified in the South Polar skua faeces, as well as the kidney, nasal and vaginal swabs, and faeces from Weddell seals, while TTLwV-2 (n=19) was exclusively found in seal faeces (Figure 3). All of the genomes identified belong to the pinniped clade, indicating that the seal itself is the most probable host of all of TTLwVs.

Based on the distribution of the pairwise identities of TTLwV-1 and TTLwV-2 ORF1 nucleotide sequences (Figure 4), we established that genomes with greater or equal to 87% identity should be

grouped into the same genotype. Accordingly, TTLwV-1 sequences were subdivided into a total of 40 genotypes, while TTLwV-2 was subdivided into three genotypes (Figure 4).

Evidence of recombination

Evidence of recombination within the genomes was detected by analyzing each species individually with RDP 4.58 (Martin et al., 2015). TTLwV-1 and TTLwV-2 genomes have differing recombination patterns. 89% of TTLwV-1 sequences had at least one recombination event, while only 26% of TTLwV-2 sequences had a recombination event (Figure 5). The majority of the recombination events in TTLwV-1 were located in the highly conserved translated region (UTR), which is similar to the recombination hotspots found in human TTV and in a global anellovirus analysis (Lefeuvre et al., 2009; Worobey, 2000). In contrast with the cold spots in ORF1 of the TTLwV-1 genome, TTLwV-2 has recombination hotspots in the coding regions, ORF1 and ORF3. Recombination may be a driving force of anellovirus diversity, especially in TTLwV-1 (Figure 5).

Prevalence of TTLwV in Weddell seals

Paired nasal and vaginal swabs were collected from 25 seals and 76% (19/25) of these tested positive for TTLwV-1 in at least one of the paired samples. When parsed individually, 64% of the nasal swabs and 72% of the vaginal swabs were TTLwV-1 positive. None of these samples tested positive for TTLwV-2. Of the 45 faecal samples collected, all were TTLwV positive, with 98% positive for TTLwV-1 and 40% for TTLwV-2.

The number of genotypes per type of sample, and the prevalence of each genotype, illustrated in Figure 4 and Supplementary Figure 2, show the most common genotypes of TTLwV-1 are 1, 8, and 11, while genotype 2 completely dominates TTLwV-2. Notably TTLwV-1 genotype 1 was also recovered from the South Polar skua faeces. The seal faeces, vaginal, and nasal swabs had comparable diversity with 25, 24, and 21 characterized genotypes, respectively. Although a single kidney sample was analyzed, four genotypes were identified in this organ. Additionally, while most seals harbored one or two TTLwV-1 genotypes, a few seals harbored many more (maximum n = 9 genotypes seal ID 16603; supplementary Figure 2).

There is no clear distribution pattern of the 40 genotypes of TTLwV-1, yet TTLwV-2 was only found in the seal faeces (40% of the faecal samples). Since TTLwV-2 is related to other anelloviruses isolated from pinnipeds, it is likely to also be infecting seals, but may have a different and more specific tropism than TTLwV-1, which seems to have pan-tropism. The overlapping genotypes of TTLwV-1 found in the faeces, nasal, and vaginal swabs suggest possible fecal-oral transmission of this virus, and indicates that future studies may be able to test the faeces alone to capture anellovirus diversity in seals.

TTLwV-1 identification in South Polar skua faeces

TTLwV-1 was recovered from South Polar skua faeces that was opportunistically sampled. South Polar Skua faeces contained three TTLwV-1 genotypes (1, 13, 17; Figures 3 and 4), which were also identified in the seal kidney, faeces, nasal, and vaginal swabs. The presence of various TTLwV-1 genotypes in South Polar skua faeces may reflect viruses that are dietary in origin since these predatory birds feed on Weddell seal placenta and scavenge seal carcasses, which may contain TTLwV, hence the likely explanation for presence of this in its faeces.

Conclusion

The advent of NGS technology has proven to be a powerful tool for virus discovery and has changed the field of virology. Here we took advantage of this technology to investigate viral presence and diversity in Antarctic wildlife, which remains largely unknown. Previous studies of Weddell seals have focused on wildlife exposure to known viruses, limiting our understanding of the myriad of viruses present in this unique environment.

Samples tested from Weddell seals from the Ross Sea led to the discovery of 152 anellovirus genomes. Sequencing complete genomes enabled the recognition of two new species, TTLwV-1 and TTLwV-2 and the role recombination plays in driving TTLwV diversity. TTLwV is phylogenetically related to the other pinniped-associated anelloviruses that were recovered from the brain and lungs of the Pacific harbor seal, the lungs, liver, lymph nodes and tonsils of a California sea lion (*Zalophus californianus*), and the faeces from subantarctic fur seals (*Arctocephalus tropicalis*) and South American fur seals (*Arctocephalus australis*) (Bodewes et al., 2013; Kluge et al., 2016; Ng et al., 2009b; Ng et al., 2011). However, this is the first time

anelloviruses have been characterized in Weddell seals and the first time they have been described in Antarctic vertebrates.

Although anelloviruses cause persistent infections and are ubiquitous among humans and various animal species, their etiology remains a mystery (Spandole et al., 2015). This was the first time a prevalence study showed anelloviruses to be ubiquitous within a pinniped population, with TTLwVs present in 100% of the seal fecal samples. TTLwV-1 genomes from the seal faeces were related to those recovered from the vaginal and nasal swabs and kidney, indicating that these were shedding from the seal itself and not infecting seal food sources. The fact that the same TTLwV-1 genotypes were found within the seal tissues and faeces may also suggest that this anellovirus species is transmitted through the faecal-oral route, which is a hypothesis proposed for the transmission of TTV in humans (Okamoto et al., 1998a; Ukita et al., 1999). Although TTLwV-2 was only detected in faecal samples, phylogenetic analysis indicates that this species may also infect seals. Failure to detect TTLwV-2 in the seal tissues tested here suggests that this species has more specific tropism than TTLwV-1.

Nevertheless, sampling of Weddell seal faeces alone could give a broad perspective of anellovirus diversity circulating within this Antarctic pinniped population. Sampling faeces may therefore be a valuable, non-invasive sampling tool for capturing the diversity and prevalence of pinniped anelloviruses. Previous prevalence studies of pinniped-associated anelloviruses screened the serum and lungs, which may have greatly underestimated the prevalence of anelloviruses in pinniped populations (Fahsbender et al., 2015; Ng et al., 2011).

Future studies focusing on the anelloviruses in Antarctica will provide insight into their presence in other species and how they are transmitted through the food web. The South Polar skua faeces derived TTLwVs are most likely acquired from scavenging Weddell seal placenta or carcasses. Therefore, detection of TTLwV in South Polar skua faeces and Weddell seal samples may provide an example of a situation in which viruses could be used as proxies for trophic interactions (Dayaram et al., 2016; Godinho et al., 2017). Additionally, it remains to be determined whether TTLwV is present in Antarctic surface waters, as has been shown for some anelloviruses in Italy and Japan (Haramoto et al., 2005; Verani et al., 2006).

More work needs to be done to determine the role of TTLwVs in Weddell seal health. TTLwVs are diverse and pervasive in this population, with individuals infected with multiple genotypes. The health effects of infection by specific genotypes or co-infection with different genotypes remain unknown. With the exception of the dead Weddell seal from which the kidney sample infected with a polyomavirus (Varsani et al., 2017) was obtained, all other animals were in apparent good health, with most sighted months and years following handling. Determining the viral load of TTLwV may prove to be important for providing clues as to the strength of seal immune systems since anellovirus loads are thought to be good indicators of immunosuppression (Hofer, 2014).

Acknowledgements

David G Ainley and Grant Ballard are supported by the US National Science Foundation (NSF; ANT-0944411). All the samples were collected with logistics supplied from the US Antarctic Program. Stacy Kim was supported by NSF ANT-0944747. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Elizabeth Fahsbender and Mya Breitbart were supported by the US National Science Foundation's Assembling the Tree of Life Program Grant DEB-1239976. Amy Kirkham and Roxanne Beltran were supported by Institutional Development Awards (IDeA) Networks of Biomedical Research Excellence Assistantships (grant number P20GM103395) from the National Institute of General Medical Sciences of the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily reflect the official views of the NIH. All the wet bench molecular work and sequencing was supported by personal fund of Arvind Varsani.

Data availability: All sequence data reported in this study has been deposited in GenBank (https://www.ncbi.nlm.nih.gov/genbank/) under accession #s KY246476 - KY246627.

Figure legends and table text

Figure 1: (A) Distribution of genome-wide pairwise identities of TTLwVs from this study. (B) Schematic genome organizations of representatives from TTLwV-1 and TTLwV-2 which highlights three open reading frames and a hypothetical protein.

Figure 2: Maximum-likelihood phylogenetic tree inferred from aligned ORF1 sequences of all publicly available anellovirus sequences together with those from this study. Branches with less than 60% bootstrap support have been collapsed.

Figure 3: Maximum-likelihood phylogenetic tree inferred from aligned ORF1 sequences of TTLwVs and other closely related anelloviruses. Branches with less than 60% bootstrap support have been collapsed. TTLwV sequences from South Polar skua faeces (SKP) are shown with dotted branches.

Figure 4: (A) Distribution of the genome-wide pairwise identities of TTLwV-1 and TTLwV-2 supporting 87% genotype demarcation for TTLwV-1 genome sequences. (B) Summary of genotypes found in different sample types. Each bar represents one genome. Total number of genotypes of TTLvW-1 and TTLvW-2 from each sample type are provided at the bottom.

Figure 5: Summary of recombinant regions identified in (A) TTLwV-1 (light green bars) and (B) TTLwV-2 (dark green bars). Recombination breakpoint distribution plots for (C) TTLwV-1 and (D) TTLwV-2, with breakpoint hot-spots indicated in red and cold-spots in blue. The dark and light grey areas of the plots indicate 95 and 99% confidence intervals, respectively. Detectable breakpoint positions are indicated by vertical lines at the top of the graphs. The thick black line represents the plot of the number of breakpoints detected within the 200 nt window region (window was moved along each of the represented alignments 1 nt at a time).

Supplementary Table 1: Summary of samples used in this study.

Supplementary Table 2: Details of abutting primers used to recover the full genomes of TTLwV.

Supplementary Table 3: Summary of the recombination events detected in TTLwV-1 and -2.

Supplementary Figure 1: Maximum-likelihood phylogenetic tree inferred from aligned ORF1 sequences of anelloviruses and a host phylogeny inferred with TimeTree (http://www.timetree.org/) (Hedges et al., 2015; Kumar et al., 2017). * marks sequences derived from mosquitoes and thus likely of vertebrate origin sampled during a blood meal by the insects.

Supplementary Figure 2: Summary of TTLwV genotypes recovered from various sample types and individual samples. Unless stated, each colour coded square represents one genome.

Supplementary Data 1: Genome-wide pairwise identities of Antarctic anelloviruses and other closely related pinniped anelloviruses from *Phoca vitulina* sampled in USA and the Netherlands (HQ287751, KF373758, KF373760, KM262781, KM262785).

References

Abe, K., Inami, T., Ishikawa, K., Nakamura, S., Goto, S., 2000. TT virus infection in nonhuman primates and characterization of the viral genome: identification of simian TT virus isolates. Journal of virology 74, 1549-1553.

Al-Moslih, M.I., Perkins, H., Hu, Y.W., 2007. Genetic relationship of Torque Teno virus (TTV) between humans and camels in United Arab Emirates (UAE). Journal of medical virology 79, 188-191.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of molecular biology 215, 403-410.

Austin, F.J., Webster, R.G., 1993. Evidence of ortho- and paramoxyoviruses in fauana from Antarctica. Journal of Wildlife Diseases 29, 568-571.

Biagini, P., 2009. Classification of TTV and related viruses (anelloviruses), in: Villiers, E.-M., Hausen, H. (Eds.), TT Viruses. Springer Berlin Heidelberg, pp. 21-33.

Blomström, A.-L., 2011. Viral metagenomics as an emerging and powerful tool in veterinary medicine. Veterinary Quarterly 31, 107-114.

Bodewes, R., Rubio Garcia, A., Wiersma, L.C., Getu, S., Beukers, M., Schapendonk, C.M., van Run, P.R., van de Bildt, M.W., Poen, M.J., Osinga, N., Sanchez Contreras, G.J., Kuiken, T.,

Smits, S.L., Osterhaus, A.D., 2013. Novel B19-like parvovirus in the brain of a harbor seal. PloS one 8, e79259.

Buck, C.B., Van Doorslaer, K., Peretti, A., Geoghegan, E.M., Tisza, M.J., An, P., Katz, J.P., Pipas, J.M., McBride, A.A., Camus, A.C., McDermott, A.J., Dill, J.A., Delwart, E., Ng, T.F., Farkas, K., Austin, C., Kraberger, S., Davison, W., Pastrana, D.V., Varsani, A., 2016. The Ancient Evolutionary History of Polyomaviruses. PLoS pathogens 12, e1005574.

Dayaram, A., Galatowitsch, M.L., Arguello-Astorga, G.R., van Bysterveldt, K., Kraberger, S., Stainton, D., Harding, J.S., Roumagnac, P., Martin, D.P., Lefeuvre, P., Varsani, A., 2016. Diverse circular replication-associated protein encoding viruses circulating in invertebrates within a lake ecosystem. Infect Genet Evol 39, 304-316.

Delwart, E.L., 2007. Viral metagenomics. Reviews in Medical Virology 17, 115-131.

Dobson, A., Foufopoulos, J., 2001. Emerging infectious pathogens of wildlife. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 356, 1001-1012.

Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC bioinformatics 5, 113.

Edwards, R.A., Rohwer, F., 2005. Viral metagenomics. Nat Rev Micro 3, 504-510.

Fahsbender, E., Rosario, K., Cannon, J.P., Gulland, F., Dishaw, L.J., Breitbart, M., 2015. Development of a Serological Assay for the Sea Lion (Zalophus californianus) Anellovirus, ZcAV. Scientific reports 5, 9637.

Gardner, H., Kerry, K., Riddle, M., Brouwer, S., Gleeson, L., 1997. Poultry virus infection in Antarctic penguins. Nature 387, 245.

Godinho, M.T., Paula, D.P., Varsani, A., Ribeiro, S.G., 2017. Genome Sequence of Cauliflower Mosaic Virus Identified in Earwigs (Doru luteipes) through a Metagenomic Approach. Genome Announc 5.

Haramoto, E., Katayama, H., Oguma, K., Yamashita, H., Nakajima, E., Ohgaki, S., 2005. One-year monthly monitoring of Torque teno virus (TTV) in wastewater treatment plants in Japan. Water research 39, 2008-2013.

Harder, T.C., Plötz, J., Liess, B., 1991. Antibodies against european phocine herpesvirus isolates detected in sera of Antarctic seals. Polar Biology 11, 509-512.

Hedges, S.B., Marin, J., Suleski, M., Paymer, M., Kumar, S., 2015. Tree of life reveals clock-like speciation and diversification. Mol Biol Evol 32, 835-845.

Hofer, U., 2014. Microbiome: anelloviridae go viral. Nature reviews. Microbiology 12, 4-5.

Huang, Y.W., Patterson, A.R., Opriessnig, T., Dryman, B.A., Gallei, A., Harrall, K.K., Vaughn, E.M., Roof, M.B., Meng, X.J., 2012. Rescue of a porcine anellovirus (torque teno sus virus 2) from cloned genomic DNA in pigs. Journal of virology 86, 6042-6054.

Hurt, A.C., Su, Y.C., Aban, M., Peck, H., Lau, H., Baas, C., Deng, Y.M., Spirason, N., Ellstrom, P., Hernandez, J., Olsen, B., Barr, I.G., Vijaykrishna, D., Gonzalez-Acuna, D., 2016. Evidence for the Introduction, Reassortment, and Persistence of Diverse Influenza A Viruses in Antarctica. Journal of virology 90, 9674-9682.

Hurt, A.C., Vijaykrishna, D., Butler, J., Baas, C., Maurer-Stroh, S., Silva-de-la-Fuente, M.C., Medina-Vogel, G., Olsen, B., Kelso, A., Barr, I.G., Gonzalez-Acuna, D., 2014. Detection of evolutionarily distinct avian influenza a viruses in antarctica. mBio 5, e01098-01014.

Kakkola, L., Hedman, K., Qiu, J., Pintel, D., S”derlund-Venermo, M., 2009. Replication of and Protein Synthesis by TT Viruses, in: de Villiers, E.-M., Hausen, H.z. (Eds.), TT Viruses: The Still Elusive Human Pathogens. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 53-64.

Kakkola, L., Tommiska, J., Boele, L.C., Miettinen, S., Blom, T., Kekarainen, T., Qiu, J., Pintel, D., Hoeben, R.C., Hedman, K., Soderlund-Venermo, M., 2007. Construction and biological activity of a full-length molecular clone of human Torque teno virus (TTV) genotype 6. The FEBS journal 274, 4719-4730.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14, 587-589.

Kamahora, T., Hino, S., Miyata, H., 2000. Three spliced mRNAs of TT virus transcribed from a plasmid containing the entire genome in COS1 cells. Journal of virology 74, 9980-9986.

Kerry, K., Riddle, M., Clarke, J., 1999. Diseases of Antarctic wildlife. A report for SCAR and COMNAP. SCAR.

King, A.M., 2011. Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier.

Kluge, M., Campos, F.S., Tavares, M., de Amorim, D.B., Valdez, F.P., Giongo, A., Roehe, P.M., Franco, A.C., 2016. Metagenomic Survey of Viral Diversity Obtained from Feces of Subantarctic and South American Fur Seals. PloS one 11, e0151921.

Kumar, S., Stecher, G., Suleski, M., Hedges, S.B., 2017. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Mol Biol Evol 34, 1812-1819.

Lee, S.Y., Kim, J.H., Park, Y.M., Shin, O.S., Kim, H., Choi, H.G., Song, J.W., 2014. A novel adenovirus in Chinstrap penguins (Pygoscelis antarctica) in Antarctica. Viruses 6, 2052-2061.

Lee, S.Y., Kim, J.H., Seo, T.K., No, J.S., Kim, H., Kim, W.K., Choi, H.G., Kang, S.H., Song, J.W., 2016. Genetic and Molecular Epidemiological Characterization of a Novel Adenovirus in Antarctic Penguins Collected between 2008 and 2013. PloS one 11, e0157032.

Lefeuvre, P., Lett, J.M., Varsani, A., Martin, D.P., 2009. Widely conserved recombination patterns among single-stranded DNA viruses. Journal of virology 83, 2697-2707.

Maggi, F., Fornai, C., Morrica, A., Casula, F., Vatteroni, M.L., Marchi, S., Ciccorossi, P., Riente, L., Pistello, M., Bendinelli, M., 1999. High prevalence of TT virus viremia in italian patients, regardless of age, clinical diagnosis, and previous interferon treatment. The Journal of infectious diseases 180, 838-842.

Martin, D.P., Murrell, B., Golden, M., Khoosal, A., Muhire, B., 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus evolution 1, vev003.

Martinez, L., Kekarainen, T., Sibila, M., Ruiz-Fons, F., Vidal, D., Gortazar, C., Segales, J., 2006. Torque teno virus (TTV) is highly prevalent in the European wild boar (Sus scrofa). Veterinary microbiology 118, 223-229.

Morgan, I.R., Westbury, H.A., 1981. Virological studies of Adelie Penguins (Pygoscelis adeliae) in Antarctica. Avian diseases 25, 1019-1026.

Morgan, I.R., Westbury, H.A., Campbell, J., 1985. Viral infections of little blue penguins (Eudyptula minor) along the southern coast of Australia. J Wildl Dis 21, 193-198.

Morner, T., Obendorf, D., Artois, M., Woodford, M., 2002. Surveillance and monitoring of wildlife diseases. Revue Scientifique et Technique-Office International des Epizooties 21, 67-76.

Muhire, B.M., Varsani, A., Martin, D.P., 2014. SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. PloS one 9, e108277.

Ng, T.F., Manire, C., Borrowman, K., Langer, T., Ehrhart, L., Breitbart, M., 2009a. Discovery of a novel single-stranded DNA virus from a sea turtle fibropapilloma by using viral metagenomics. Journal of virology 83, 2500-2509.

Ng, T.F., Suedmeyer, W.K., Wheeler, E., Gulland, F., Breitbart, M., 2009b. Novel anellovirus discovered from a mortality event of captive California sea lions. The Journal of general virology 90, 1256-1261.

Ng, T.F., Wheeler, E., Greig, D., Waltzek, T.B., Gulland, F., Breitbart, M., 2011. Metagenomic identification of a novel anellovirus in Pacific harbor seal (Phoca vitulina richardsii) lung samples and its detection in samples from multiple years. The Journal of general virology 92, 1318-1323.

Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32, 268-274.

Niel, C., Saback, F.L., Lampe, E., 2000. Coinfection with multiple TT virus strains belonging to different genotypes is a common event in healthy Brazilian adults. Journal of clinical microbiology 38, 1926-1930.

Ninomiya, M., Takahashi, M., Nishizawa, T., Shimosegawa, T., Okamoto, H., 2008. Development of PCR assays with nested primers specific for differential detection of three human anelloviruses and early acquisition of dual or triple infection during infancy. Journal of clinical microbiology 46, 507-514.

Nishiyama, S., Dutia, B.M., Stewart, J.P., Meredith, A.L., Shaw, D.J., Simmonds, P., Sharp, C.P., 2014. Identification of novel anelloviruses with broad diversity in UK rodents. The Journal of general virology 95, 1544-1553.

Nishizawa, T., Okamoto, H., Konishi, K., Yoshizawa, H., Miyakawa, Y., Mayumi, M., 1997. A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. Biochemical and biophysical research communications 241, 92-97.

Okamoto, H., Akahane, Y., Ukita, M., Fukuda, M., Tsuda, F., Miyakawa, Y., Mayumi, M., 1998a. Fecal excretion of a nonenveloped DNA virus (TTV) associated with posttransfusion non-A-G hepatitis. Journal of medical virology 56, 128-132.

Okamoto, H., Nishizawa, T., Kato, N., Ukita, M., Ikeda, H., Iizuka, H., Miyakawa, Y., Mayumi, M., 1998b. Molecular cloning and characterization of a novel DNA virus (TTV) associated with posttransfusion hepatitis of unknown etiology. Hepatol Res 10, 1-16.

Okamoto, H., Nishizawa, T., Takahashi, M., Tawara, A., Peng, Y., Kishimoto, J., Wang, Y., 2001. Genomic and evolutionary characterization of TT virus (TTV) in tupaias and comparison with species-specific TTVs in humans and non-human primates. The Journal of general virology 82, 2041-2050.

Okamoto, H., Nishizawa, T., Tawara, A., Takahashi, M., Kishimoto, J., Sai, T., Sugai, Y., 2000. TT virus mRNAs detected in the bone marrow cells from an infected individual. Biochemical and biophysical research communications 279, 700-707.

Olsen, B., Bergstrom, S., McCafferty, D.J., Sellin, M., Wistrom, G., 1996. Salmonella enteritidis in Antarctica: zoonosis in man or humanosis in penguins? Lancet (London, England) 348, 1319-1320.

Park, Y.M., Kim, J.H., Gu, S.H., Lee, S.Y., Lee, M.G., Kang, Y.K., Kang, S.H., Kim, H.J., Song, J.W., 2012. Full genome analysis of a novel adenovirus from the South Polar skua (Catharacta maccormicki) in Antarctica. Virology 422, 144-150.

Retamal, P., Blank, O., Abalos, P., Torres, D., 2000. Detection of anti-Brucella antibodies in pinnipeds from the Antarctic territory. Vet Rec 146, 166-167.

Romeo, R., Hegerich, P., Emerson, S.U., Colombo, M., Purcell, R.H., Bukh, J., 2000. High prevalence of TT virus (TTV) in naive chimpanzees and in hepatitis C virus-infected humans: frequent mixed infections and identification of new TTV genotypes in chimpanzees. The Journal of general virology 81, 1001-1007.

Rosario, K., Breitbart, M., 2011. Exploring the viral world through metagenomics. Current Opinion in Virology 1, 289-297.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J., Birol, I., 2009. ABySS: a parallel assembler for short read sequence data. Genome research 19, 1117-1123.

Spandole, S., Cimponeriu, D., Berca, L.M., Mihaescu, G., 2015. Human anelloviruses: an update of molecular, epidemiological and clinical aspects. Archives of virology 160, 893-908.

Stenvers, O., Plotz, J., Ludwig, H., 1992. Antarctic seals carry antibodies against seal herpesvirus. Archives of virology 123, 421-424.

Stover, B.C., Muller, K.F., 2010. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC bioinformatics 11, 7.

Symonds, E.M., Griffin, D.W., Breitbart, M., 2009. Eukaryotic Viruses in Wastewater Samples from the United States. Applied and Environmental Microbiology 75, 1402-1409.

Thomazelli, L.M., Araujo, J., Oliveira, D.B., Sanfilippo, L., Ferreira, C.S., Brentano, L., Pelizari, V.H., Nakayama, C., Duarte, R., Hurtado, R., Branco, J.O., Walker, D., Durigon, E.L., 2010. Newcastle disease virus in penguins from King George Island on the Antarctic region. Veterinary microbiology 146, 155-160.

Torres, D., 2000. Antarctic territory. The Veterinary Record 146, 166-167.

Tryland, M., Klein, J., Nordoy, E.S., Blix, A.S., 2005. Isolation and partial characterization of a parapoxvirus isolated from a skin lesion of a Weddell seal. Virus research 108, 83-87.

Ukita, M., Okamoto, H., Kato, N., Miyakawa, Y., Mayumi, M., 1999. Excretion into bile of a novel unenveloped DNA virus (TT virus) associated with acute and chronic non-A-G hepatitis. The Journal of infectious diseases 179, 1245-1248.

Varsani, A., Frankfurter, G., Stainton, D., Male, M.F., Kraberger, S., Burns, J.M., 2017. Identification of a polyomavirus in Weddell seal (Leptonychotes weddellii) from the Ross Sea (Antarctica). Archives of virology 162, 1403-1407.

Varsani, A., Kraberger, S., Jennings, S., Porzig, E.L., Julian, L., Massaro, M., Pollard, A., Ballard, G., Ainley, D.G., 2014. A novel papillomavirus in Adelie penguin (Pygoscelis adeliae) faeces sampled at the Cape Crozier colony, Antarctica. The Journal of general virology 95, 1352-1365.

Varsani, A., Porzig, E.L., Jennings, S., Kraberger, S., Farkas, K., Julian, L., Massaro, M., Ballard, G., Ainley, D.G., 2015. Identification of an avian polyomavirus associated with Adelie penguins (Pygoscelis adeliae). The Journal of general virology 96, 851-857.

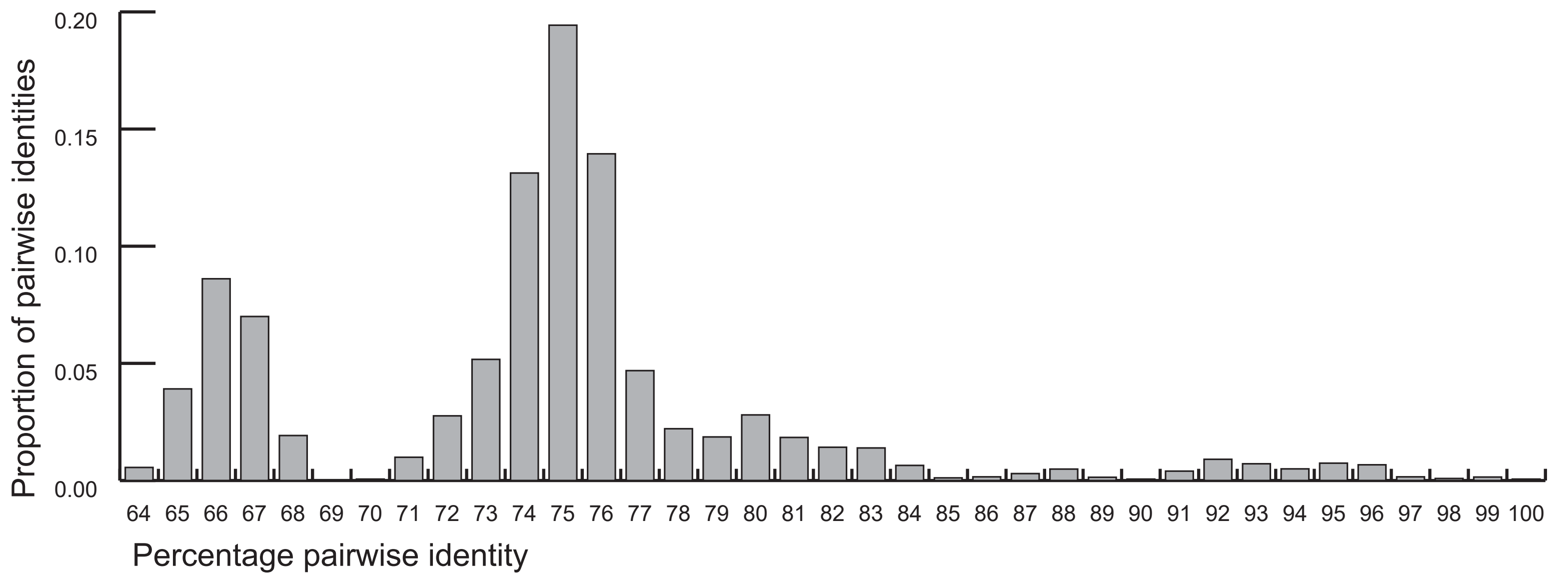
Verani, M., Casini, B., Battistini, R., Pizzi, F., Rovini, E., Carducci, A., 2006. One-year monthly monitoring of Torque teno virus (TTV) in river water in Italy. Water Science and Technology 54, 191.

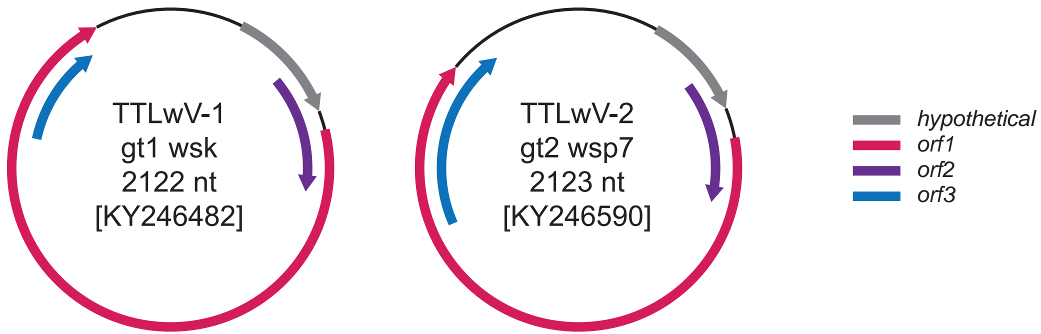
Wang, D., Coscoy, L., Zylberberg, M., Avila, P.C., Boushey, H.A., Ganem, D., DeRisi, J.L., 2002. Microarray-based detection and genotyping of viral pathogens. Proceedings of the National Academy of Sciences 99, 15687-15692.

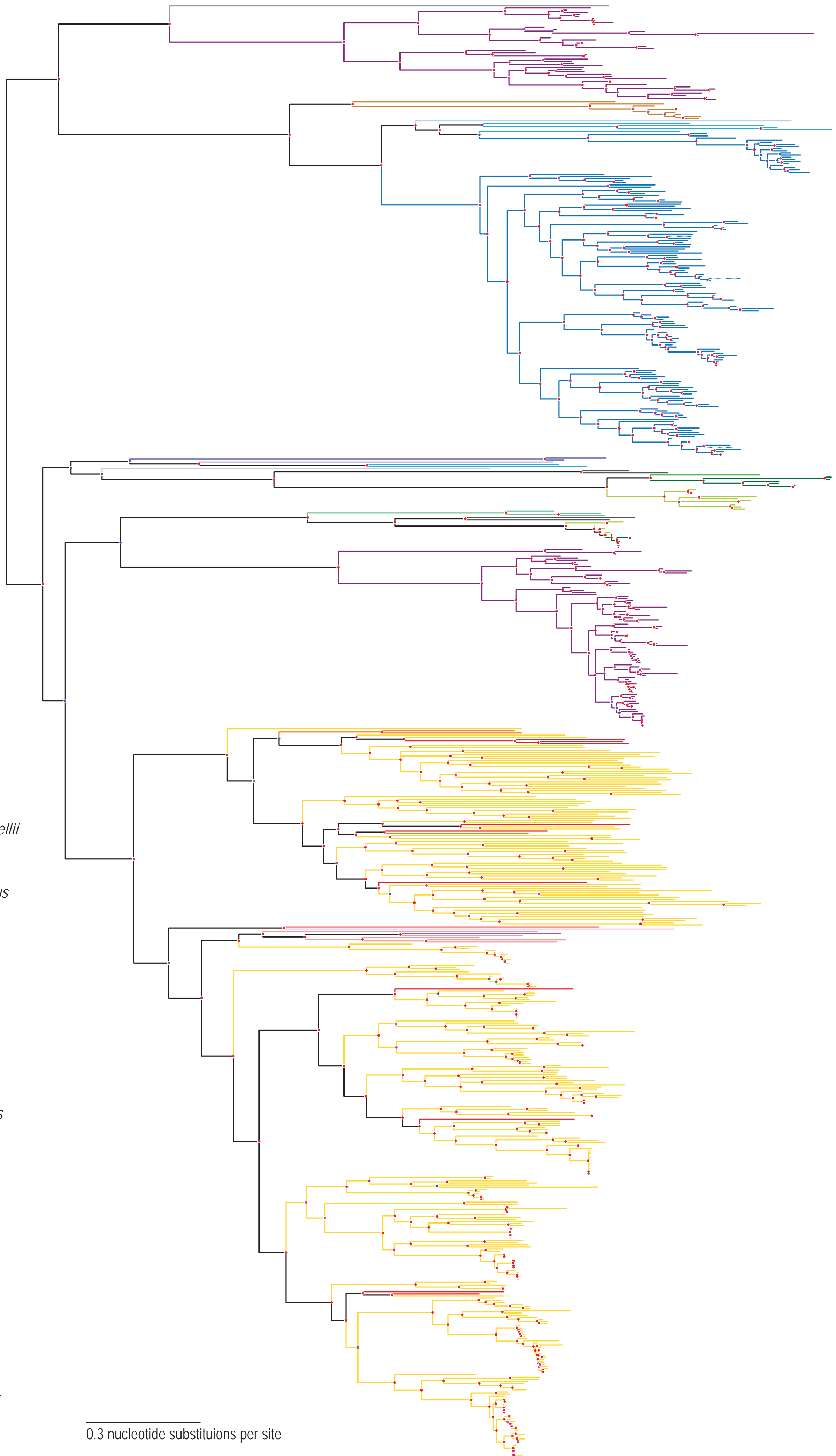
Worobey, M., 2000. Extensive homologous recombination among widely divergent TT viruses. Journal of virology 74, 7666-7670.

Yochem, P.K., Stewart, B.S., Gelatt, T.S., Siniff, D.B., 2009. Health Assessment of Weddell Seals, Leptonychotes weddellii, in McMurdo Sound, Antarctica, in: Kerry, K.R., Riddle, M. (Eds.), Health of Antarctic Wildlife: A Challenge for Science and Policy. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 123-138.

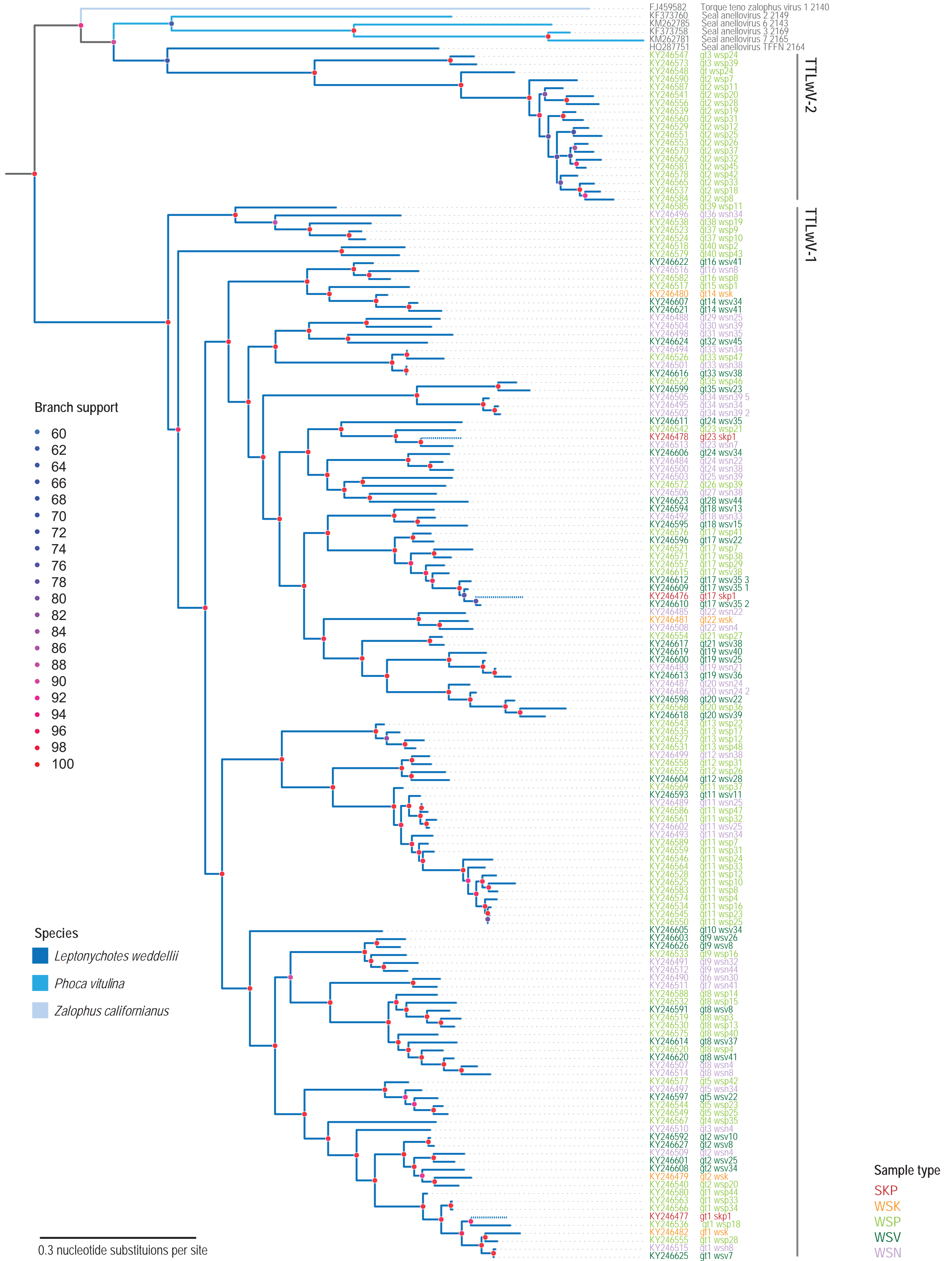
Yu, X., Zhu, C., Zheng, X., He, S., Liu, X., 2007. Genome analysis and epidemiological investigation of goose circovirus detected in eastern China. Virus Genes 35, 605-609.


A
Proportion of pairwise identities
0.20
0.15
0.10
0.05
0.00
64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
Percentage pairwise identity


B
TTLwV-1
gt1 wsk
2122 nt
[KY246482]
TTLwV-2
gt2 wsp7
2123 nt
[KY246590]
hypothetical
orf1
orf2
orf3



Branch support
60
62
64
66
68
70
72
74
76
78
80
82
84
86
88
90
92
94
96
98
100
Order
Species
Artiodactyla
Sus scrofa
Carnivora
Canis lupus
Carnivora
Leptonychotes weddellii
Carnivora
Phoca vitulina
Carnivora
Zalophus californianus
Carnivora
Martes martes
Carnivora
Felis catus
Chiroptera
Tadarida brasiliensis
Diptera
Anopheles sp.
Equine
Equus caballus
Primate
Aotus sp.
Primate
Chlorocebus sabaeus
Primate
Gorilla gorilla
Primate
Homo sapiens
Primate
Pan troglodytes
Primate
Saguinus midas
Primate
Macaca fuscata
Rodentia
Microtus agrestis
Rodentia
Myodes glareolus
Rodentia
Rattus norvegicus
Rodentia
Apodemus sylvaticus
0.3 nucleotide substituions per site



FJ459582 Torque teno zalophus virus 1 2140
KF373760 Seal anellovirus 2 2149
KM262785 Seal anellovirus 6 2143
KF373758 Seal anellovirus 3 2169
KM262781 Seal anellovirus 7 2165
HQ287751 Seal anellovirus TFFN 2164
KY246547 gt3 wsp24
KY246573 gt3 wsp39
KY246548 gt wsp24
KY246590 gt2 wsp7
KY246587 gt2 wsp11
KY246541 gt2 wsp20
KY246556 gt2 wsp28
KY246539 gt2 wsp19
KY246560 gt2 wsp31
KY246529 gt2 wsp12
KY246551 gt2 wsp25
KY246553 gt2 wsp26
KY246570 gt2 wsp37
KY246562 gt2 wsp32
KY246581 gt2 wsp45
KY246578 gt2 wsp42
KY246565 gt2 wsp33
KY246537 gt2 wsp18
KY246584 gt2 wsp8
KY246585 gt39 wsp11
KY246496 gt36 wsn34
KY246538 gt38 wsp19
KY246523 gt37 wsp9
KY246524 gt37 wsp10
KY246518 gt40 wsp2
KY246579 gt40 wsp43
KY246622 gt16 wsv41
KY246516 gt16 wsn8
KY246582 gt16 wsp8
KY246517 gt15 wsp1
KY246480 gt14 wsk
KY246607 gt14 wsv34
KY246621 gt14 wsv41
KY246488 gt29 wsn25
KY246504 gt30 wsn39
KY246498 gt31 wsn35
KY246624 gt32 wsv45
KY246494 gt33 wsn34
KY246526 gt33 wsp47
KY246501 gt33 wsn38
KY246616 gt33 wsv38
KY246522 gt35 wsp46
KY246599 gt35 wsv23
KY246505 gt34 wsn39 5
KY246495 gt34 wsn34
KY246502 gt34 wsn39 2
KY246611 gt24 wsv35
KY246542 gt23 wsp21
KY246478 gt23 skp1
KY246513 gt23 wsn7
KY246606 gt24 wsv34
KY246484 gt24 wsn22
KY246500 gt24 wsn38
KY246503 gt25 wsn39
KY246572 gt26 wsp39
KY246506 gt27 wsn38
KY246623 gt28 wsv44
KY246594 gt18 wsv13
KY246492 gt18 wsn33
KY246595 gt18 wsv15
KY246576 gt17 wsp41
KY246596 gt17 wsv22
KY246521 gt17 wsp7
KY246571 gt17 wsp38
KY246557 gt17 wsp29
KY246615 gt17 wsv38
KY246612 gt17 wsv35 3
KY246609 gt17 wsv35 1
KY246476 gt17 skp1
KY246610 gt17 wsv35 2
KY246485 gt22 wsn22
KY246481 gt22 wsk
KY246508 gt22 wsn4
KY246554 gt21 wsp27
KY246617 gt21 wsv38
KY246619 gt19 wsv40
KY246600 gt19 wsv25
KY246483 gt19 wsn21
KY246613 gt19 wsv36
KY246487 gt20 wsn24
KY246486 gt20 wsn24 2
KY246598 gt20 wsv22
KY246568 gt20 wsp36
KY246618 gt20 wsv39
KY246543 gt13 wsp22
KY246535 gt13 wsp17
KY246527 gt13 wsp12
KY246531 gt13 wsp48
KY246499 gt12 wsn38
KY246558 gt12 wsp31
KY246552 gt12 wsp26
KY246604 gt12 wsv28
KY246569 gt11 wsp37
KY246593 gt11 wsv11
KY246489 gt11 wsn25
KY246586 gt11 wsp47
KY246561 gt11 wsp32
KY246602 gt11 wsv25
KY246493 gt11 wsn34
KY246589 gt11 wsp7
KY246559 gt11 wsp31
KY246546 gt11 wsp24
KY246564 gt11 wsp33
KY246528 gt11 wsp12
KY246525 gt11 wsp10
KY246583 gt11 wsp8
KY246574 gt11 wsp4
KY246534 gt11 wsp16
KY246545 gt11 wsp23
KY246550 gt11 wsp25
KY246605 gt10 wsv34
KY246603 gt9 wsv26
KY246626 gt9 wsv8
KY246533 gt9 wsp16
KY246491 gt9 wsn32
KY246512 gt9 wsn44
KY246490 gt6 wsn30
KY246511 gt7 wsn41
KY246588 gt8 wsp14
KY246532 gt8 wsp15
KY246591 gt8 wsv8
KY246519 gt8 wsp3
KY246530 gt8 wsp13
KY246575 gt8 wsp40
KY246614 gt8 wsv37
KY246520 gt8 wsp4
KY246620 gt8 wsv41
KY246507 gt8 wsn4
KY246514 gt8 wsn8
KY246577 gt5 wsp42
KY246497 gt5 wsn34
KY246597 gt5 wsv22
KY246544 gt5 wsp23
KY246549 gt5 wsp25
KY246567 gt4 wsp35
KY246510 gt3 wsn4
KY246592 gt2 wsv10
KY246627 gt2 wsv8
KY246509 gt2 wsn4
KY246601 gt2 wsv25
KY246608 gt2 wsv34
KY246479 gt2 wsk
KY246540 gt2 wsp20
KY246580 gt1 wsp44
KY246563 gt1 wsp33
KY246566 gt1 wsp34
KY246477 gt1 skp1
KY246536 gt1 wsp18
KY246482 gt1 wsk
KY246555 gt1 wsp28
KY246515 gt1 wsn8
KY246625 gt1 wsv7
TTLwV-2
TTLwV-1
Branch support
60
62
64
66
68
70
72
74
76
78
80
82
84
86
88
90
92
94
96
98
100
Species
Leptonychotes weddellii
Phoca vitulina
Zalophus californianus
Sample type
SKP
WSK
WSP
WSV
WSN
0.3 nucleotide substituions per site

A

TTLwV-1

TTLwV-2

Percentage pairwise identities

100 99 98 97 96 95 94 93 92 91 90 89 88 **87** 86 85 84 83 82 81 80 79 78 77 76 75 74 73 72 71 70 69 68 67

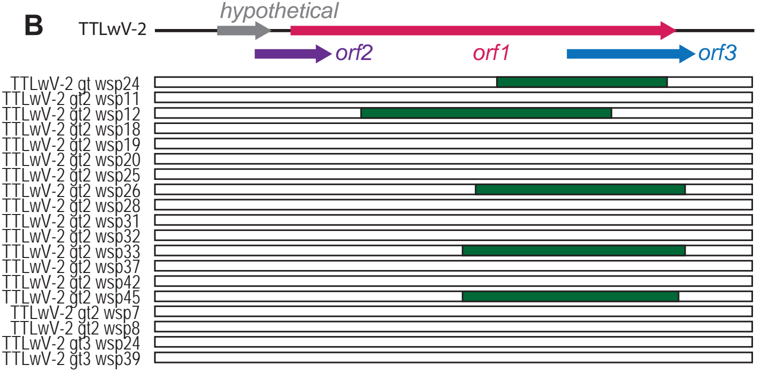
0.25 0.20 0.15 0.10 0.05 0.00 0.00 0.05 0.10 0.15 0.20

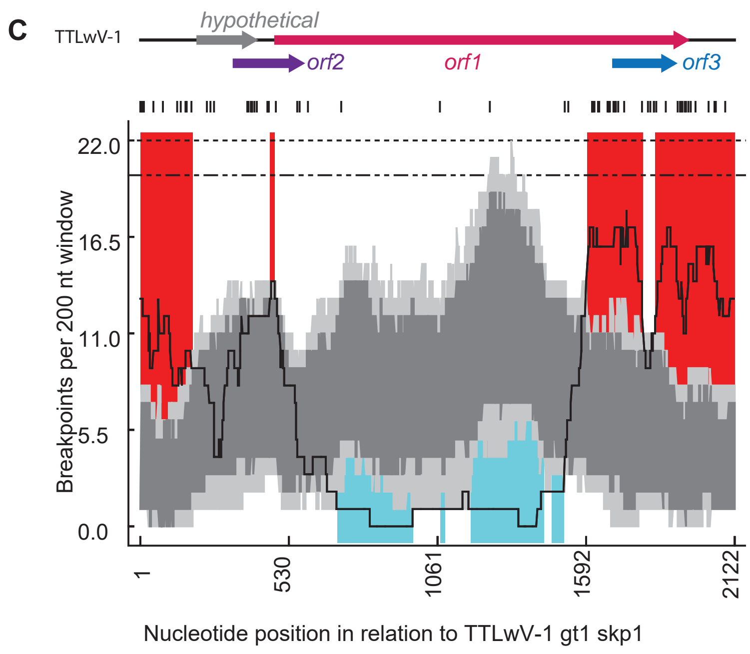
Proportion of pairwise identities

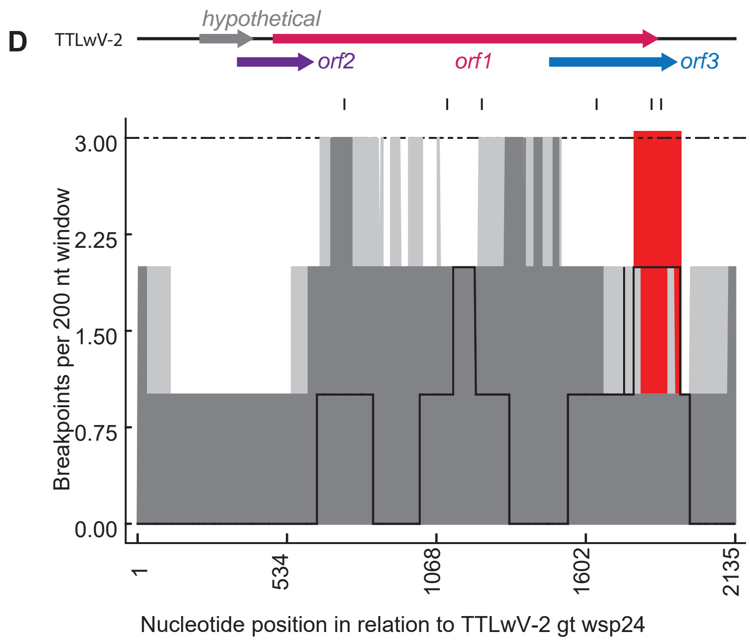
B

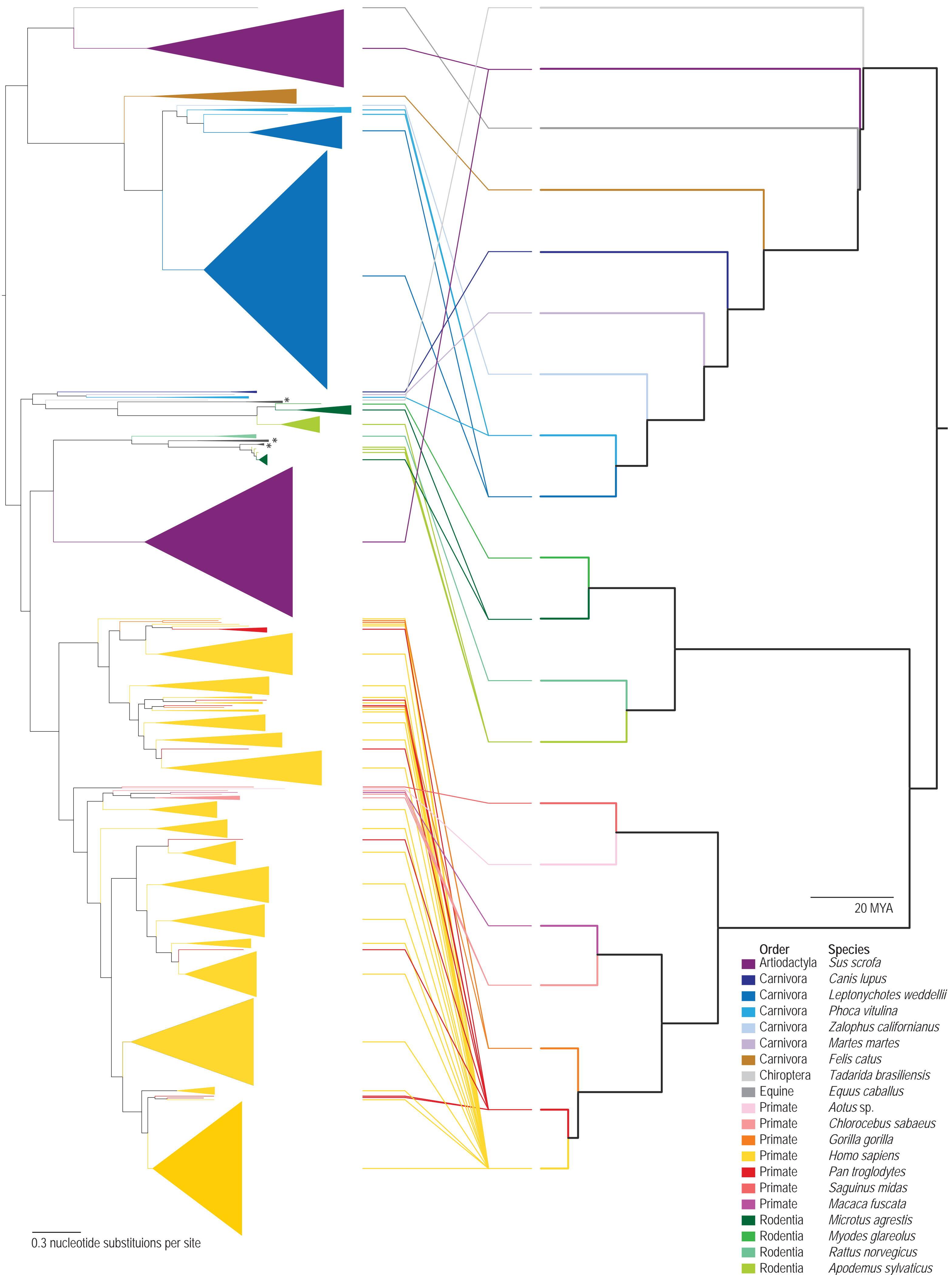
	Genotypes	SKP	WSK	WSP	WSV	WSN	Genomes
TTLwV-1	1						10
	2						7
	3						1
	4						1
	5						5
	6						1
	7						1
	8						10
	9						5
	10						1
	11						18
	12						4
	13						5
	14						3
	15						1
	16						3
	17						9
	18						4
	19						4
	20						4
	21						2
	22						3
	23						2
	24						5
	25						1
	26						1
	27						1
	28						1
	29						1
	30						1
	31						1
	32						1
	33						4
	34						2
	35						2
	36						1
	37						2
	38						1
	39						1
	40						2
TTLwV-2	1						1
	2						17
	3						2
		3	4	25	24	21	

A
TTLwV-1
hypothetical
orf2
orf1
orf3
TTLwV-1 gt1 skp1
TTLwV-1 gt1 wsk
TTLwV-1 gt1 wsn8
TTLwV-1 gt1 wsp18
TTLwV-1 gt1 wsp28
TTLwV-1 gt1 wsp33
TTLwV-1 gt1 wsp34
TTLwV-1 gt1 wsp44
TTLwV-1 gt1 wsv7
TTLwV-1 gt10 wsv34
TTLwV-1 gt11 wsn25
TTLwV-1 gt11 wsn34
TTLwV-1 gt11 wsp10
TTLwV-1 gt11 wsp12
TTLwV-1 gt11 wsp16
TTLwV-1 gt11 wsp23
TTLwV-1 gt11 wsp24
TTLwV-1 gt11 wsp25
TTLwV-1 gt11 wsp31
TTLwV-1 gt11 wsp32
TTLwV-1 gt11 wsp33
TTLwV-1 gt11 wsp37
TTLwV-1 gt11 wsp4
TTLwV-1 gt11 wsp47
TTLwV-1 gt11 wsp7
TTLwV-1 gt11 wsp8
TTLwV-1 gt11 wsv11
TTLwV-1 gt11 wsv25
TTLwV-1 gt12 wsn38
TTLwV-1 gt12 wsp26
TTLwV-1 gt12 wsp31
TTLwV-1 gt12 wsv28
TTLwV-1 gt13 wsp12
TTLwV-1 gt13 wsp17
TTLwV-1 gt13 wsp22
TTLwV-1 gt13 wsp48
TTLwV-1 gt14 ws
TTLwV-1 gt14 wsv34
TTLwV-1 gt14 wsv41
TTLwV-1 gt15 wsp1
TTLwV-1 gt16 wsn8
TTLwV-1 gt16 wsp8
TTLwV-1 gt16 wsv41
TTLwV-1 gt17 skp1
TTLwV-1 gt17 wsp29
TTLwV-1 gt17 wsp38
TTLwV-1 gt17 wsp41
TTLwV-1 gt17 wsp7
TTLwV-1 gt17 wsv22
TTLwV-1 gt17 wsv35 1
TTLwV-1 gt17 wsv35 2
TTLwV-1 gt17 wsv35 3
TTLwV-1 gt17 wsv38
TTLwV-1 gt18 wsn33
TTLwV-1 gt18 wsv13
TTLwV-1 gt18 wsv15
TTLwV-1 gt19 wsn21
TTLwV-1 gt19 wsv25
TTLwV-1 gt19 wsv36
TTLwV-1 gt19 wsv40
TTLwV-1 gt2 wsk
TTLwV-1 gt2 wsn4
TTLwV-1 gt2 wsp20
TTLwV-1 gt2 wsv10
TTLwV-1 gt2 wsv25
TTLwV-1 gt2 wsv34
TTLwV-1 gt2 wsv8
TTLwV-1 gt20 wsn24
TTLwV-1 gt20 wsn24 2
TTLwV-1 gt20 wsp36
TTLwV-1 gt20 wsv22
TTLwV-1 gt20 wsv39
TTLwV-1 gt21 wsp27
TTLwV-1 gt21 wsv38
TTLwV-1 gt22 ws22
TTLwV-1 gt22 wsk
TTLwV-1 gt22 wsn4
TTLwV-1 gt23 skp1
TTLwV-1 gt23 wsn7
TTLwV-1 gt23 wsp21
TTLwV-1 gt24 wsn22
TTLwV-1 gt24 wsn38
TTLwV-1 gt24 wsv34
TTLwV-1 gt24 wsv35
TTLwV-1 gt25 wsn39
TTLwV-1 gt26 wsp39
TTLwV-1 gt27 wsn38
TTLwV-1 gt28 wsv44
TTLwV-1 gt29 wsn25
TTLwV-1 gt3 wsn4
TTLwV-1 gt30 wsn39
TTLwV-1 gt31 wsn35
TTLwV-1 gt33 wsv34
TTLwV-1 gt33 wsv38
TTLwV-1 gt33 wsn38
TTLwV-1 gt33 wsp47
TTLwV-1 gt33 wsv38
TTLwV-1 gt34 wsn34
TTLwV-1 gt34 wsn39 2
TTLwV-1 gt34 wsn39 5
TTLwV-1 gt35 wsp46
TTLwV-1 gt35 wsv23
TTLwV-1 gt36 wsn34
TTLwV-1 gt37 wsp10
TTLwV-1 gt37 wsp9
TTLwV-1 gt38 wsp19
TTLwV-1 gt39 wsp11
TTLwV-1 gt4 wsp35
TTLwV-1 gt40 wsp2
TTLwV-1 gt40 wsp43
TTLwV-1 gt5 wsn34
TTLwV-1 gt5 wsp23
TTLwV-1 gt5 wsp25
TTLwV-1 gt5 wsp42
TTLwV-1 gt5 wsv22
TTLwV-1 gt6 wsn30
TTLwV-1 gt7 wsn41
TTLwV-1 gt8 wsn4
TTLwV-1 gt8 wsn8
TTLwV-1 gt8 wsp13
TTLwV-1 gt8 wsp14
TTLwV-1 gt8 wsp15
TTLwV-1 gt8 wsp3
TTLwV-1 gt8 wsp4
TTLwV-1 gt8 wsp40
TTLwV-1 gt8 wsv37
TTLwV-1 gt8 wsv41
TTLwV-1 gt8 wsv8
TTLwV-1 gt9 wsn32
TTLwV-1 gt9 wsn44
TTLwV-1 gt9 wsp16
TTLwV-1 gt9 wsv26
TTLwV-1 gt9 wsv8


B
TTLwV-2
hypothetical
orf2
orf1
orf3
TTLwV-2 gt wsp24
TTLwV-2 gt2 wsp11
TTLwV-2 gt2 wsp12
TTLwV-2 gt2 wsp18
TTLwV-2 gt2 wsp19
TTLwV-2 gt2 wsp20
TTLwV-2 gt2 wsp25
TTLwV-2 gt2 wsp26
TTLwV-2 gt2 wsp28
TTLwV-2 gt2 wsp31
TTLwV-2 gt2 wsp32
TTLwV-2 gt2 wsp33
TTLwV-2 gt2 wsp37
TTLwV-2 gt2 wsp42
TTLwV-2 gt2 wsp45
TTLwV-2 gt2 wsp7
TTLwV-2 gt2 wsp8
TTLwV-2 gt3 wsp24
TTLwV-2 gt3 wsp39


C
TTLwV-1
hypothetical
orf2
orf1
orf3
22.0
16.5
11.0
5.5
0.0
Breakpoints per 200 nt window
1
530
1061
1592
2122
Nucleotide position in relation to TTLwV-1 gt1 skp1


D
TTLwV-2
hypothetical
orf2
orf1
orf3
3.00
2.25
1.50
0.75
0.00
Breakpoints per 200 nt window
1
534
1068
1602
2135
Nucleotide position in relation to TTLwV-2 gt wsp24
Breakpoint number
Local 95% confidence interval
Local 99% confidence interval
Global 95% confidence limit
Global 99% confidence limit
Recombination hotspot
Recombination coldspot



20 MYA
0.3 nucleotide substituions per site
Order
Species
Artiodactyla
Sus scrofa
Carnivora
Canis lupus
Carnivora
Leptonychotes weddellii
Carnivora
Phoca vitulina
Carnivora
Zalophus californianus
Carnivora
Martes martes
Carnivora
Felis catus
Chiroptera
Tadarida brasiliensis
Equine
Equus caballus
Primate
Aotus sp.
Primate
Chlorocebus sabaeus
Primate
Gorilla gorilla
Primate
Homo sapiens
Primate
Pan troglodytes
Primate
Saguinus midas
Primate
Macaca fuscata
Rodentia
Microtus agrestis
Rodentia
Myodes glareolus
Rodentia
Rattus norvegicus
Rodentia
Apodemus sylvaticus

Nasal	Nasal (resampled)	Vaginal	Vaginal (resampled)	Faeces	Kidney
WSN04; WS14-04		WSV04; WS14-04			
WSN06; WS14-06		WSV06; WS14-06			
WSN07; WS14-07	WSN39; WS14-39	WSV07; WS14-07	WSV39; WS14-39		
WSN08; WS14-08	WSN41; WS14-41	WSV08; WS14-08	WSV41; WS14-41	WSP46; WS14-08	
WSN09; WS14-09		WSV09; WS14-09			
WSN10; WS14-10	WSN33; WS14-33	WSV10; WS14-10			
WSN11; WS14-11	WSN27; WS14-27	WSV11; WS14-11	WSV27; WS14-27	WSP47; WS14-11	
WSN13; WS14-13		WSV13; WS14-13			
WSN15; WS14-15	WSN45; WS14-45	WSV15; WS14-15	WSV45; WS14-45		
WSN21; WS14-21	WSN36; WS14-36	WSV21; WS14-21	WSV36; WS14-36		
WSN22; WS14-22		WSV22; WS14-22			
WSN23; WS14-23	WSN40; WS14-40	WSV23; WS14-23	WSV40; WS14-40		
WSN24; WS14-24	WSN26; WS14-26	WSV24; WS14-24	WSV26; WS14-26		
WSN25; WS14-25		WSN14; WS14-14	WSV25; WS14-25	WSP48; WS14-14	
WSN28; WS14-28		WSV28; WS14-28			
WSN29; WS14-29		WSV29; WS14-29			
WSN30; WS14-30		WSV30; WS14-30			
WSN31; WS14-31		WSV31; WS14-31			
WSN32; WS14-32		WSV32; WS14-32			
WSN34; WS14-34		WSV34; WS14-34			
WSN35; WS14-35		WSV35; WS14-35			
WSN37; WS14-37		WSV37; WS14-37			
WSN38; WS14-38		WSV38; WS14-38			
WSN42; WS14-42					
WSN43; WS14-43		WSV43; WS14-43			
WSN44; WS14-44		WSV05; WS14-05	WSV44; WS14-44		
WSN46; WS14-46					
				WSP01; SK3	
				WSP02; SK15	
				WSP03; SK20	
				WSP04; SK21	
				WSP07; SK24	
				WSP08; SK25	
				WSP09; SK26	
				WSP10; SK27	
				WSP11; SK28	
				WSP12; SK29	
				WSP13; SK30	
				WSP14; SK32	
				WSP15; SK34	
				WSP16; SK35	
				WSP17; SK36	
				WSP18; SK37	
				WSP19; SK38	
				WSP20; SK39	
				WSP21; BR1	
				WSP22; BR2	
				WSP23; BR3	
				WSP24; BR4	
				WSP25; HC1	
				WSP26; HC2	
				WSP27; HC3	
				WSP28; HC4	
				WSP29; HC5	
				WSP30; HC6	
				WSP31; HC7	
				WSP32; HC8	
				WSP33; HC9	
				WSP34; HC10	
				WSP35; HC11	
				WSP36; HC12	
				WSP37; HC13	
				WSP38; TR1	
				WSP39; TR2	
				WSP40; TR3	
				WSP41; TR4	
				WSP42; TR5	
				WSP43; RB5	
				WSP44; RB6	
				WSP45; HC6	
					WSK

Primer	TTLwV type	Sequence (5'→3')
anello1_F	1	AGAAGAAAGAGAATCTCCGTTMGAGGTTGGG
anello1_R		CCGTGAGGGGTAATATCTTACTGTAGC
anello2_F	1	GAAAATCTGATTGGCTGCCCGGGATG
anello2_R		TTCCGCTTTTTAAAACCCACTTTTCGAGGTAC
12092anello_F	1	AAAAACCTCCTCTATAGAGCTCAGTACTAC
12092anello_R		GAAGGTAAAGAAATGGTGACCCCAGG
12124anello_F	1	CCATAATTATCCATTGGGTCTGAGAAG
12124anello_R		CAGATGGGGACCCTTCTGTCCATATG
12421anello_F	2	TGGGAAGGACTCTTTAATCTACCATCAGC
12421anello_R		GCTAGCAGGGGGTCTCACGTATACT

APPENDIX 2: DISCOVERY OF A NOVEL CIRCULAR DNA VIRUS IN THE FORBES SEA STAR, *ASTERIAS FORBESI*

This appendix was published in:

Fahsbender, E., I. Hewson, K. Rosario, A. Tuttle, A. Varsani, M. Breitbart. Discovery of a novel circular DNA virus in the Forbes sea star, Asterias forbesi. Archives of Virology. 160:2349-2351

Arch Virol (2015) 160:2349–2351
DOI 10.1007/s00705-015-2503-2

CrossMark

BRIEF REPORT

Discovery of a novel circular DNA virus in the Forbes sea star, *Asterias forbesi*

Elizabeth Fahsbender[1] · Ian Hewson[2] · Karyna Rosario[1] · Allison D. Tuttle[3] · Arvind Varsani[4,5,6] · Mya Breitbart[1]

Received: 22 April 2015 / Accepted: 15 June 2015 / Published online: 26 June 2015
© Springer-Verlag Wien 2015

Abstract A single-stranded DNA (ssDNA) virus, Asterias forbesi-associated circular virus (AfaCV), was discovered in a Forbes sea star displaying symptoms of sea star wasting disease (SSWD). The AfaCV genome organization is typical of circular Rep-encoding ssDNA (CRESS-DNA) viruses and is similar to that of members of the family *Circoviridae*. PCR-based surveys indicate that AfaCV is not clearly associated with SSWD, whereas the sea star-associated densovirus (SSaDV), recently implicated in SSWD in the Pacific, was prevalent in symptomatic specimens. AfaCV represents the first CRESS-DNA virus detected in echinoderms, adding to the growing diversity of these viruses recently recovered from invertebrates.

✉ Mya Breitbart
mya@usf.edu

1 College of Marine Science, University of South Florida, 140 7th Avenue South, Saint Petersburg, FL 33701, USA

2 Department of Microbiology, Cornell University, Wing Hall, 123 Wing Drive, Ithaca, NY 14853-8101, USA

3 Mystic Aquarium, 55 Coogan Boulevard, Mystic, CT 06355, USA

4 School of Biological Sciences and Biomolecular Interaction Centre, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand

5 Electron Microscope Unit, Division of Medical Biochemistry, Department of Clinical Laboratory Sciences, University of Cape Town, Rondebosch, Cape Town 7701, South Africa

6 Department of Plant Pathology and Emerging Pathogens Institute, University of Florida, Gainesville, FL 32611, USA

Sea stars (Asterozoa) are keystone predators in the marine environment, essential for structuring intertidal invertebrate communities that are critical for ecosystem health [15]. Sea star wasting disease (SSWD), which is characterized by white epidermal lesions, deflation, limb autotomy, and in most cases, eventual death ("melting"), is currently affecting multiple species of asteroids in both the Atlantic and Pacific oceans. A study investigating SSWD in asteroids from the Northeast Pacific Ocean implicated a densovirus (sea star-associated densovirus; SSaDV) as the most likely cause of SSWD in the region [9]. To evaluate if other viruses could be associated with this disease in the Atlantic Ocean, viral metagenomics was performed on Forbes sea star (*Asterias forbesi*) specimens from the Northwest Atlantic coast that were exhibiting SSWD symptoms.

Visually healthy *Asterias forbesi* collected from Point Judith Pond, Rhode Island, in June of 2011 later developed signs of SSWD in captivity, and two individuals were sent to the University of South Florida for viral metagenomic analysis. Upon receipt, samples were flash frozen in liquid nitrogen, and gut contents were removed before crushing each sample inside a sterile Whirl-Pak®. Asteroid tissue was homogenized in sterile 1X phosphate-buffered saline (PBS) solution using a bead-beater (BioSpec, USA) with 1.0-mm glass beads (Research Products International, USA) for 1 min. Virus particles were purified from homogenates by pelleting the cells by centrifugation at 6,000 × *g* for 6 min and filtering the supernatant through a 0.22-µm Sterivex filter (Millipore, USA). The filtrate was treated with 0.2 volumes of chloroform, followed by 2.5 U of DNase I per µl at 37 °C for 3 hours to degrade non-encapsidated nucleic acids [13, 14]. EDTA was then added at a final concentration of 20 mM to inactive the nuclease activity. Viral DNA was extracted from the purified viral

particles using a QIAamp MinElute Virus Spin Kit (QIAGEN, USA) and then amplified and fragmented using a GenomePlex Whole Genome Amplification (WGA) Kit (Sigma-Aldrich) according to the manufacturers' protocols. The resulting fragmented DNA was purified using a UltraClean® PCR Clean-Up Kit (MO BIO, USA) and sequenced at a next-generation sequencing service provider using an Illumina HiSeq 2000 (Illumina, USA) instrument.

The metagenomic single-end reads were *de novo* assembled using Assembly By Short Sequences (ABySS) [19], and viral contigs were identified through BLAST [1] analysis. Back-to-back primers (Seastarcirco_Fout 5'-TCT CAA TGG TGG GAT GGA TA-3' and Seastarcirco_Rout 5'-TCT CTG CTT CCA GTA GGC GT-3') were then designed based on a contig exhibiting significant similarities to circular single-stranded DNA (ssDNA) viruses and utilized for inverse PCR to complete and verify the genome sequence. Viral DNA was extracted using a QIAamp MinElute Virus Spin Kit (QIAGEN, USA) and amplified through rolling-circle amplification (RCA), which is known to enrich for small circular templates [10, 11], using the illustra GenomiPhi V2 Kit (GE Healthcare, USA). This amplified DNA was used as the template for the inverse PCR assay using a Herculase II Fusion Polymerase Kit (Agilent Technologies, USA) under the following conditions: 95 °C for 4 minutes, 45 cycles of 95 °C for 20 seconds, 53 °C for 20 seconds with a touchdown of -0.1 °C per cycle, and 72 °C for 2 minutes, followed by a final extension at 72 °C for 5 minutes. The PCR product was cloned using a TOPO TA Cloning Kit (Invitrogen, USA) and commercially Sanger sequenced by primer walking.

The metagenomics approach resulted in the discovery of a novel circular ssDNA genome, named *Asterias forbesi*-associated circular virus (AfaCV). The AfaCV genome (GenBank accession number KR186219) is 2414 nt in length with two major open reading frames (ORFs) encoding a replication initiator protein (Rep) and a putative capsid protein (Cap) in an ambisense orientation (Fig. 1). The genome also exhibits a putative origin of replication marked by a conserved nonanucleotide motif (TAGTATTAC) at the apex of a hairpin structure. This genome organization is typical of circular Rep-encoding ssDNA (CRESS-DNA) viruses of the family *Circoviridae*, specifically members of the proposed genus *Cyclovirus* [12, 18]. The AfaCV Rep shares 52 % identity with that of a CRESS-DNA virus discovered in an estuarine mollusk (GenBank accession number AJP36419.1) [4], whereas the putative Cap shares 40 % identity with a CRESS-DNA virus discovered in dragonfly larvae (GenBank accession number YP_009001744) [3]. Similar to members of the *Circoviridae*, the AfaCV putative Cap contains an arginine-rich region within the first 50 amino acids.

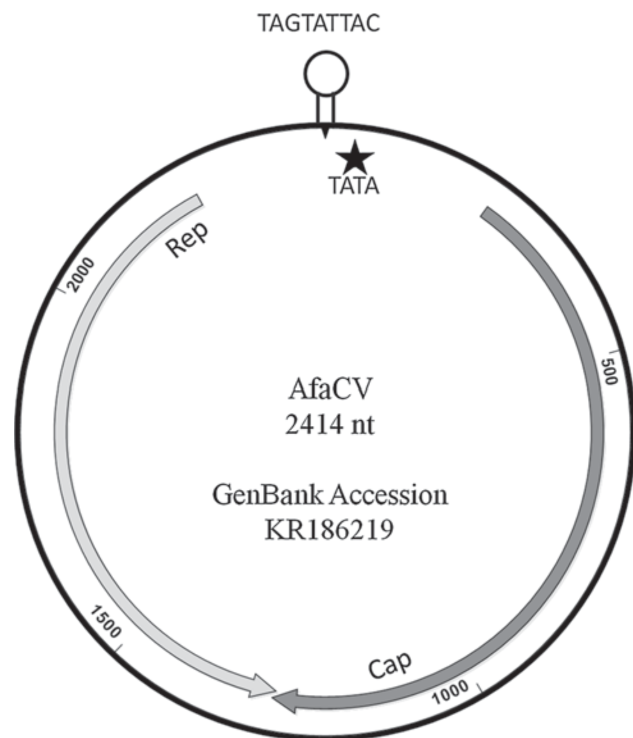


Fig. 1 Schematic genome organization of AfaCV. The figure highlights a replication-initiator ORF (Rep; light gray), a putative capsid ORF (Cap; dark gray), and a stem-loop structure containing the conserved nonanucleotide motif (TAGTATTAC). The black star represents the TATA box

To explore the association of the novel AfaCV with SSWD, 48 SSWD-affected *A. forbesi* sea stars collected from Westerly, Rhode Island, in February of 2013 were tested for both AfaCV and the previously described SSaDV by PCR. DNA was extracted from sea star arm tissue using a ZR Tissue and Insect DNA MiniPrep Kit (Zymo, USA) and enriched by RCA using an illustra TempliPhi Kit (GE Healthcare, USA) for use as the template for both PCR assays. Primers AfaCV_2283F (5'-CTG TCC TCC TTT GGT TTC CA-3') and AfaCV_65R (5'-CTG GAT TTT CCT GTT CCT GTT G-3') were designed to target a 197-nt region of AfaCV. The PCR (containing 1 μM each primer, 200 μM dNTPs, 1 U of Red *Taq* [Sigma-Aldrich], 1X Red *Taq* reaction buffer and 1 μl of DNA template in a 50-μl reaction) proceeded as follows: 95 °C for 5 minutes, 45 cycles of 94 °C for 1 minute, 54 °C for 1 minute with a touchdown of -0.1 °C per cycle, and 72 °C for 30 seconds, followed by a final extension at 72 °C for 10 minutes. PCR products of the correct size were cloned using a TOPO TA Cloning Kit (Invitrogen) and Sanger sequenced to confirm their identity as AfaCV and verify the specificity of the assay through comparisons against GenBank. Only five of the 48 samples tested (10 %) were positive for AfaCV, showing that AfaCV is unlikely to be associated with

SSWD in the Atlantic Ocean. Previously published primers for the VP4 region of SSaDV [9] were used to screen these same samples for SSaDV. The PCR (containing 1 µM each primer, 200 µM dNTPs, 1 U of Red *Taq* [Sigma-Aldrich], 1X Red *Taq* reaction buffer and 1 µl of DNA template in a 50-µl reaction) proceeded as follows: 95 °C for 5 minutes, 45 cycles of 94 °C for 1 minute, 56 °C for 1 minute with a touchdown of -0.1 °C per cycle, and 72 °C for 30 seconds, followed by a final extension at 72 °C for 10 minutes. Twenty-three of the 48 samples (48 %) were positive for SSaDV; however, this may be an underestimation of the true SSaDV prevalence due the use of standard PCR, which is less sensitive than quantitative PCR using an internal TaqMan probe. Nevertheless, SSaDV appears to be associated with SSWD in Atlantic Forbes sea stars, similar to the findings of Hewson et al. [9] in the Pacific.

Recent metagenomic studies have revealed that CRESS-DNA viruses similar to members of the *Circoviridae* are more widespread than previously thought and may thrive in a wide variety of hosts, including invertebrates [18]. Specifically, these viruses have been identified in insects [2, 5, 16, 17], copepods [6], mollusks [4] and water fleas [8]. Although recent studies have identified linear ssDNA densoviruses in echinoderms [7, 9], AfaCV is the first CRESS-DNA virus identified in echinoderms and provides further support that these viruses may infect a wider host range than previously thought. Therefore, although AfaCV has no association with SSWD, future studies are needed to determine the prevalence of AfaCV and other CRESS-DNA viruses amongst various sea star species and elucidate their roles in the ecology of these keystone predators.

Acknowledgments This work was funded through grant DEB-1239976 from the National Science Foundation's Assembling the Tree of Life Program to K.R and M.B. EF is funded by the Sanibel-Captiva Shell Club and Mary and Al Bridell Memorial Fellowship. This manuscript constitutes Sea Research Foundation publication #256.

References

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
2. Dayaram A, Potter KA, Moline AB, Rosenstein DD, Marinov M, Thomas JE, Breitbart M, Rosario K, Argüello-Astorga GR, Varsani A (2013) High global diversity of cycloviruses amongst dragonflies. J Gen Virol 94:1827–1840
3. Dayaram A, Galatowitsch M, Harding JS, Argüello-Astorga GR, Varsani A (2014) Novel circular DNA viruses identified in *Procordulia grayi* and *Xanthocnemis zealandica* larvae using metagenomic approaches. Infect Genet Evol 22:134–141
4. Dayaram A, Goldstien S, Argüello-Astorga GR, Zawar-Reza P, Gomez C, Harding JS, Varsani A (2015) Diverse small circular DNA viruses circulating amongst estuarine molluscs. Infect Genet Evol 31:284–295
5. Dayaram A, Potter KA, Pailes R, Marinov M, Rosenstein DD, Varsani A (2015) Identification of diverse circular single-stranded DNA viruses in adult dragonflies and damselflies (Insecta: Odonata) of Arizona and Oklahoma, USA. Infect Genet Evol 30:278–287
6. Dunlap DS, Ng TFF, Rosario K, Barbosa JG, Greco AM, Breitbart M, Hewson I (2013) Molecular and microscopic evidence of viruses in marine copepods. Proc Natl Acad Sci USA 110:1375–1380
7. Gudenkauf BM, Eaglesham JB, Aragundi WM, Hewson I (2014) Discovery of urchin-associated densoviruses (family *Parvoviridae*) in coastal waters of the Big Island, Hawaii. J Gen Virol 95:652–658
8. Hewson I, Ng G, Li W, LaBarre BA, Aguirre I, Barbosa JG, Breitbart M, Greco AW, Kearns CM, Looi A, Schaffner LR, Thompson PD, Hairston NG (2013) Metagenomic identification, seasonal dynamics, and potential transmission mechanisms of a *Daphnia*-associated single-stranded DNA virus in two temperate lakes. Limnol Oceanogr 58:1605–1620
9. Hewson I, Button JB, Gudenkauf BM, Miner B, Newton AL, Gaydos JK, Wynne J, Groves CL, Hendler G, Murray M, Fradkin S, Breitbart M, Fahsbender E, Lafferty KD, Kilpatrick AM, Miner CM, Raimondi P, Lahner L, Friedman CS, Daniels S, Haulena M, Marliave J, Burge CA, Eisenlord ME, Harvell CD (2014) Densovirus associated with sea-star wasting disease and mass mortality. Proc Natl Acad Sci USA 111:17278–17283
10. Kim K-H, Chang H-W, Nam Y-D, Roh SW, Kim M-S, Sung Y, Jeon CO, Oh H-M, Bae J-W (2008) Amplification of uncultured single-stranded DNA viruses from rice paddy soil. Appl Environ Microbiol 74:5975–5985
11. Kim K-H, Bae J-W (2011) Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. Appl Environ Microbiol 77:7663–7668
12. Li L, Kapoor A, Slikas B, Bamidele OS, Wang C, Shaukat S, Masroor MA, Wilson ML, Ndjango J-BN, Peeters M, Gross-Camp ND, Muller MN, Hahn BH, Wolfe ND, Triki H, Bartkus J, Zaidi SZ, Delwart E (2010) Multiple diverse circoviruses infect farm animals and are commonly found in human and chimpanzee feces. J Virol 84:1674–1682
13. Ng TFF, Manire C, Borrowman K, Langer T, Ehrhart L, Breitbart M (2009) Discovery of a novel single-stranded DNA virus from a sea turtle fibropapilloma by using viral metagenomics. J Virol 83:2500–2509
14. Ng TFF, Suedmeyer WK, Wheeler E, Gulland F, Breitbart M (2009) Novel anellovirus discovered from a mortality event of captive California sea lions. J Gen Virol 90:1256–1261
15. Paine RT (1969) The *Pisaster-Tegula* interaction: prey patches, predator food preference, and intertidal community structure. Ecology 50:950–961
16. Rosario K, Marinov M, Stainton D, Kraberger S, Wiltshire EJ, Collings DA, Walters M, Martin DP, Breitbart M, Varsani A (2011) Dragonfly cyclovirus, a novel single-stranded DNA virus discovered in dragonflies (Odonata: Anisoptera). J Gen Virol 92:1302–1308
17. Rosario K, Dayaram A, Marinov M, Ware J, Kraberger S, Stainton D, Breitbart M, Varsani A (2012) Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). J Gen Virol 93:2668–2681
18. Rosario K, Duffy S, Breitbart M (2012) A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. Arch Virol 157:1851–1871
19. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol İ (2009) ABySS: A parallel assembler for short read sequence data. Genome Res 19:1117–1123

APPENDIX 3: SEQUENCING OF THE CHELONID FIBROPAPILLOMATOSIS-ASSOCIATED HERPESVIRUS GENOME FROM THREE FLORIDA SEA TURTLE SPECIES

Sequencing of the chelonid fibropapillomatosis-associated herpesvirus, ChHV5 genome from three Florida sea turtle species

Elizabeth Fahsbender[1], Makenzie Burrows[1], Cheryl Morrison[2], Mya Breitbart[1]

[1] College of Marine Science, University of South Florida, Saint Petersburg, FL 33701, USA
[2]Leetown Science Center, U.S. Geological Survey, Kearneysville, West Virginia, USA

Abstract

Fibropapillomatosis (FP) is a debilitating disease of panzootic proportions affecting all hard shelled sea turtles, including green *(Chelonia mydas*), loggerhead *(Caretta caretta*), and Kemp's ridley *(Lepidochelys kempii*). There is overwhelming evidence suggesting this neoplastic disease is caused by the chelonid fibropapillomatosis-associated herpesvirus (CFPHV; Chelonid herpesvirus 5, ChHV5). Unfortunately, all attempts to culture this virus, a precursor needed to fulfill Koch's postulates, have failed. The lack of a culture system has hindered efforts to understand the role of ChHV5 in FP, highlighting the need for molecular studies. Currently, the only genome available of ChHV5 was sequenced from the glottis tumor of a Hawaiian green sea turtle using the bacterial artificial clone (BAC) system. To sequence ChHV5 from tumors of Florida green, loggerhead, and Kemp’s ridley sea turtles, this study used the Hawaiian genome as a scaffold for long-range PCR followed by Illumina sequencing of amplicons; however, long regions of repeat sections (> 9 kb) have impeded genome completion. Preliminary data was used to create an updated ChHV5 phylogeny using a 4,152 bp region of the capsid gene. This new phylogeny contradicts previous phylogenies constructed with other genes, indicating that the topology is dependent on the gene analyzed. Currently, efforts are under way to recover full genomes through target enrichment and PacBio sequencing, in order to enable genome-wide phylogenetic comparisons and provide insight into ChHV5 evolution.

Introduction

Chelonid fibropapillomatosis-associated herpesvirus (CFPHV; Chelonid herpesvirus 5, ChHV5) is most likely the etiological agent of fibropapillomatosis (FP) (Lackovich, et al. 1999; Lu, et al. 2000a; Lu, et al. 2000b; Quackenbush, et al. 1998), which is considered the most significant neoplastic disease in reptiles due to its rapid increase in sea turtle populations and worldwide prevalence (Herbst 1994). This debilitating disease has reached panzootic proportions and while it primarily affects green sea turtles (*Chelonia mydas*), it has been documented in all hard shelled sea turtles, including loggerheads (*Caretta caretta*) and Kemp's ridley (*Lepidochelys kempii)* (Alfaro-Núñez and Gilbert 2014). FP manifests as cutaneous fibropapillomas or fibromas and in severe cases, internal fibromas, myxofibromas, or fibrosarcomas (Herbst 1994; Work, et al. 2004). These tumors can interfere with feeding, predator avoidance, buoyancy, hydrodynamics, and vision, ultimately contributing to mortality and stranding events (Brooks, et al. 1994; Jacobson, et al. 1989; Lucke 1938; Smith and Coates 1938). In especially severe cases, FP may lead to immunosuppression, secondary bacterial infections, emaciation and death (Work and Balazs 1999; Work, et al. 2003).

FP has been documented in Florida's waters since the late 1930's (Smith and Coates 1938), but has been reported in increasing numbers over the past two decades (Foley, et al. 2005; Herbst 1994). In some regions of Florida (including the Indian River Lagoon, the Florida Keys, and the Gulf coast), FP affects >50% of the green turtles (Coberley 2002; Foley, et al. 2005; Greenblatt, et al. 2005a). Spontaneous tumor regression in sea turtles that are not severely afflicted with FP has been documented since the 1990's, but the mechanism of regression is unknown (Aguirre, et al. 1999; Bennett, et al. 1999; Ehrhart 1991). Although FP may not be lethal to all sea turtles,

this disease has severely affected Florida's sea turtles and due to its prevalence in juvenile animals, it poses a threat to these endangered animals.

Numerous attempts to isolate and culture ChHV5 have failed, leaving Koch's postulates unfulfilled and limiting our ability to definitively conclude the role of this virus in the disease. Nevertheless, there is strong evidence that ChHV5 plays an important role in the development of FP. Transmission studies implicate an enveloped virus as the etiological agent of FP (Herbst, et al. 1995; Herbst, et al. 1996) and transmission electron microscopy (TEM) of tumors revealed herpesvirus-like particles, which correlate in size and chloroform sensitivity to the transmission experiments (Herbst, et al. 1996). Additionally, herpesviral DNA sequences have been detected by PCR in the tumors of FP symptomatic turtles from around the world. Quantitative PCR of the DNA polymerase gene shows that ChHV5 concentrations are 2.5 to 4.5 logs higher and gene expression is over 200-fold higher in tumors than in unaffected tissues of turtles with fibropapillomatosis (Lu, et al. 2000a; Quackenbush, et al. 2001). In situ hybridization localized herpesviral DNA to the nuclei of clusters of acanthotic epithelial cells of tumors, and mRNA transcripts to the nuclei of clusters of epithelial cells of sea turtles from Puerto Rico (Kang, et al. 2008). More recently, herpesvirus-like particles were also found by TEM in areas of ballooning degeneration with eosinophilic epidermal intranuclear inclusions (EIIs), indicating viral replication in the epidermal cells of tumors (Work, et al. 2014). These collective data provide strong evidence that ChHV5 is the causative agent of FP.

Without an established culture system, molecular biology methods are the best way to study the role of ChHV5 in the development, progression, and regression of FP. Currently, 132,233 base

pairs of ChHV5 DNA have been sequenced from the glottis tumor of a Hawaiian green sea turtle (Ackermann, et al. 2012). Multiple sequencing studies have suggested that ChHV5 is an alphaherpesvirus (Ackermann, et al. 2012; Greenblatt, et al. 2005a; Herbst, et al. 2004; Nigro, et al. 2004), determined by sequence homology, as well as its gene order and orientation. Previous phylogenetic analyses have focused on short segments of the polymerase gene or concatenated regions of multiple ChHV5 genes, revealing the existence of geographic ChHV5 variants (Ene, et al. 2005; Greenblatt, et al. 2005a; Greenblatt, et al. 2005b; Herbst, et al. 2004). This has provided insight into the evolution of these viruses, indicating that ChHV5 has co-evolved with its sea turtle host over millions of years, and suggesting that environmental changes have likely affected virus transmission and disease expression from an already established viral infection (Ene, et al. 2005; Greenblatt, et al. 2005a; Herbst, et al. 2004). The geographically driven genetic differences of ChHV5 is concordant with differences in disease manifestation by location. For example, oral tumors are commonly found on Hawaiian turtles while they are rarely found in Florida turtles (Aguirre, et al. 2002), analysis of EIIs using immunohistochemistry revealed FP turtles from Florida, but not Hawaii, developed antibodies to ChHV5 antigens (Herbst, et al. 1998), and unlike Florida, FP is declining in Hawaii (Chaloupka, et al. 2009; Chaloupka, et al. 2008; Hirama and Ehrhart 2007).

Most of the genome information originates from Hawaiian sea turtles, highlighting the need for virus genome information from Florida, especially given the geographic variability in ChHV5 genetics. However, the latent stage of ChHV5 has impeded the use of typical laboratory methods used to prepare viral templates for sequencing, such as cell culture and virion purification (Greenblatt, et al. 2004; Lu, et al. 1999; Work, et al. 2009). Attempts to isolate viral

particles (as described in appendices 1 and 2) and episomal DNA failed. Furthermore, relatively low ChHV5 copy numbers compared to host DNA precludes sequencing the virus directly from the tumor (Quackenbush, et al. 2001). The recently sequenced ChHV5 genome from the glottis tumor of a Hawaiian green was completed by screening over 10,000 BAC clones (Ackermann, et al. 2012), making this a fairly inefficient approach. However, sequencing the BAC clone has provided a way to circumvent the methodological limitations of sequencing ChHV5 by providing a template for LR-PCR. Next-generation sequencing of LR-PCR amplicons was used to characterize the partial genome of ChHV5 from three different Florida sea turtle species, including loggerhead *(Caretta caretta)*, green *(Chelonia mydas)*, and Kemp's ridley (*Lepidochelys kempii*).

Methods

Tumor collection and DNA extraction

Tumors were collected from Florida loggerhead (n=2), Kemp's ridley (n=2), and green (n=1) sea turtles in accordance with MTP# 15-00223 (Table 1). Tumors were archived at -80°C. One tumor from each sea turtle was dissected (>25mg) for DNA extraction using the DNeasy Blood and Tissue Kit (Qiagen, Valencia, California).

PCR amplification and Illumina sequencing

PCR primers were designed to amplify ~5000 bp regions of the ChHV5 genome using the Hawaiian green sea turtle ChHV5 genome as a reference genome (Table 2). Long range polymerase chain reactions (LR-PCR) were performed using the TaKaRa LA polymerase kit (ClonTech, Palo Alto, California), which is optimized for longer amplicons. The standard

protocol containing 0.2 µM of each primer, 200 µM of dNTPs, 1.25 U/50 µl of PrimeSTAR GXL buffer, 1.25 U/50 µl of PrimeSTAR GXL DNA Polymerase, and 2.5 µl of DNA template in a 25 µl reaction was performed on all samples as follows: 93°C for 3 min, followed by 30 cycles of [98°C for 10 seconds, 60°C for 15 seconds, and 68°C for 5 minutes]. If the standard protocol was unsuccessful, the reaction was optimized according to the manufacturer's troubleshooting manual. LR-PCR products were verified by gel electrophoresis and cleaned using the DNA Clean and Concentrator™ kit (Zymo, Irvine, California). If there was non-specific primer binding resulting in multiple bands, the band of the expected size band was purified directly from the gel using the UltraClean® GelSpin® DNA extraction kit (MoBio, Carlsbad, California). Cleaned amplicons were quantified using the Qubit dsDNA HS assay kit on a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, Massachusetts) and normalized to 1 ng/µl using sterile water. Samples were pooled and libraries were prepped using the Nextera XT DNA Library Prep Kit according to the manufacturer's manual following suggested modification for larger insert size (Illumina, San Diego, California). Libraries were validated for size using the Agilent BioAnalyzer High Sensitivity DNA Kit (Agilent Technologies, Palo Alto, California), quantified using the Qubit 2.0 Fluorometer with the Qubit dsDNA HS Assay Kit (Invitrogen, Carlsbad, California) and normalized to 4 nM using 10 mM Tris, pH 8.5. The pooled libraries were run at 15 pM with a 5% PhiX174 spike on the Illumina MiSeq with a 600 cycle v3 reagent kit (Illumina, San Diego, California).

Data analysis

Machine demultiplexed fastq files were trimmed and imported onto BaseSpace. Reads were checked for quality using FastQC. Geneious v. 7.0 was used to assemble reads to the ChHV5

genome sequence (accession HQ878327) and to generate consensus sequences. For each sample, consensus sequences were aligned using Geneious. A Region of the capsid gene with full coverage for all samples were removed and aligned using MUSCLE. This 4,152 bp region of the capsid gene was compared against the GenBank non-redundant database using BLASTn. ChHV5 sequences with 100% query coverage were included into the alignment for phylogenetic analysis. Maximum-Likelihood trees were created in MEGA with bootstrap replicates of 1000 (Tamura, et al. 2007).

Results

The primer success rates were different for each sample, resulting in different coverage of the reference genome for each sample (n=5) (Fig. 1). The loggerhead consensus sequences covered 89.5% and 83.4 % of the reference genome, while the Kemp's ridley consensus sequences had 87.9% and 45.7% coverage (Table 1). The low coverage Kemp's ridley consensus sequence (Kemp's ridley 2) was excluded from the phylogenetic analysis. The repeat sections of each ChHV5 genome are missing due to low primer success rates and sequencing difficulties. Phylogenetic analysis of a 4,152 bp region of the capsid gene region shows that two of the clades group by location, and one clade groups by species (Fig. 2A). A pairwise identity analysis is concordant with the phylogenetic tree, with the green sea turtles form HI (AF035003.3 and AY644454.1) being the most divergent from the other samples (Fig. 2B). Although the loggerhead (#2) sea turtle is in a clade by itself it is 96-100% identical to the other samples.

Discussion

The incidence of fibropapillomatosis among Florida's sea turtles has increased significantly over the past two decades (Herbst 1994; Hirama and Ehrhart 1999; Hirama and Ehrhart 2007) and has significantly contributed to stranding events of Florida sea turtles. Although fibropapillomatosis is a threat to sea turtle populations, the development and progression of FP is not understood, partially due to the technical difficulties of researching this virus. There are currently 125 partial ChHV5 sequences on NCBI, ranging from 181-22,947 bp, with only one genome. The lack of sequence data makes it difficult to understand ChHV5 evolution, thereby impeding our understanding of the contribution of ChHV5 genetics to disease development and regression.

Long-range PCR paired with next-generation sequencing has proven useful for characterizing the majority of the ChHV5 genome, but large gaps still remain. The high GC content, large genome size and the repetitive nature of the genome create PCR difficulties. The large genome size also makes PCR amplicon sequencing inefficient for large sample sizes, especially with the primers not working for all samples. It is also a possibility that sequence regions that did not amplify using the primers designed from the reference genome are divergent. Furthermore, even if the PCR was successful, sequencing depth is also decreased in cases of high GC content and repetitive sequences (Jia, et al. 2014).

Previous phylogenetic analyses focusing on specific genes or concatenated regions of genes such as the DNA polymerase gene and glycoprotein B gene have revealed the existence of geographic ChHV5 variants by (Ene, et al. 2005; Greenblatt, et al. 2005a; Greenblatt, et al. 2005b; Herbst, et al. 2004; Patrício, et al. 2012). These phylogenetic analyses have proven to be informative and have provided clues to how ChHV5 has spread. However, the new phylogenetic analysis of part

of the capsid region performed here, is different from the previous analyses that have shown ChHV5 to clade by location. The topology of the capsid tree shows clades based on species, as well as location. While evolutionary relationships are typically based on comparisons of a subset of genes, this may not be the best way to capture ChHV5 evolution, since phylogenetic trees comparing different genes may result in different topology, as shown here (Nei 1987; Zimmerer 1991). Phylogenetic analysis of a specific gene, or a subset of genes, is informative for understanding gene evolution, but to elucidate the evolutionary history of ChHV5, genome-wide comparisons are necessary. Unlike gene comparisons, genome-wide comparisons can detect lateral gene transfer, transposons, and recombination events.

In the available ChHV5 genome, repeat sections contain genes homologous to human latency modulating genes, including ICP0, ICP4, and the latency-associated nuclear antigen (LANA) (Ackermann, et al. 2012). In the human Kaposi's sarcoma herpesvirus (KSHV), the LANA also inhibits the host tumor suppressor gene p53, preventing apoptosis and thereby allowing for the development of tumors (Friborg, et al. 1999). Therefore, despite the sequence difficulties associated with the repeat sections, they contain important genetic information and should not be ignored. Additionally, the identification of ChHV5 oncogenes, as well as genes that are involved in modulating latency may have important implications for management and treatment of FP.

Ongoing studies are focusing on characterizing the genome of ChHV5 using target enrichment to capture ChHV5 DNA directly from tumor tissue, paired with PacBio, which provides long reads that can be assembled back to the reference genome (Rhoads and Au 2015). The newly sequenced genomes can then be used to create an updated phylogeny comparing ChHV5 across

multiple locations and species, evaluating the selective pressure on each gene and providing insight into the role genetics plays in the manifestation of FP.

References

Ackermann, Mathias, et al. 2012. The genome of chelonid herpesvirus 5 harbors atypical genes. PLOS ONE 7(10):e46623.

Aguirre, A. Alonso, et al. 2002. Pathology of oropharyngeal fibropapillomatosis in green turtles *Chelonia mydas*. Journal of Aquatic Animal Health 14(4):298-304.

Aguirre, AA, et al. 1999. Survey of fibropapillomatosis and other potential diseases in marine turtles from Moreton Bay, Queensland, Australia. Proceedings of the Nineteenth Annual Symposium on Sea Turtle Conservation and Biology. H. Kalb and T. Wibbels. South Padre Island, Texas, 1999.

Alfaro-Núñez, Alonzo, and M Thomas P Gilbert 2014. Validation of a sensitive PCR assay for the detection of chelonid fibropapilloma-associated herpesvirus in latent turtle infections. Journal of Virological Methods (206):38-41.

Bennett, P, U Keuper-Bennett, and GH Balazs 1999. Photographic evidence for the regression of fibropapillomas afflicting green turtles at Honokowai, Maui, in the Hawaiian Islands. Proc. 19th Ann. Symp. Sea Turtle Biol Cons, 1999, pp. 37-39.

Brooks, D. E., et al. 1994. Ocular fibropapillomas of green turtles (*Chelonia mydas*). Veterinary Pathology Online 31(3):335-339.

Chaloupka, Milani, George H. Balazs, and Thierry M. Work 2009. Rise and fall over 26 Years of a marine epizootic in Hawaiian green sea turtles. Journal of Wildlife Diseases 45(4):1138-1142.

Chaloupka, Milani, et al. 2008. Cause-specific temporal and spatial trends in green sea turtle strandings in the Hawaiian Archipelago (1982–2003). Marine Biology 154(5):887-898.

Coberley, Sadie Shea 2002. The role of herpesviruses in marine turtle diseases, University of Florida.

Ehrhart, LM 1991. Fibropapillomas in green turtles of the Indian River Lagoon, Florida: distribution over time and area. Research Plan for Marine Turtle Fibropapilloma. NOAA Tech. Memo. NMFS-SEFSC-156. US Dept. Commerce:59-61.

Ene, A., et al. 2005. Distribution of chelonid fibropapillomatosis-associated herpesvirus variants in Florida: molecular genetic evidence for infection of turtles following recruitment to neritic developmental habitats. J Wildl Dis 41(3):489-97.

Foley, Allen M., et al. 2005. Fibropapillomatosis in stranded green turtles *(Chelonia mydas)* from the Eastern United States (1980-98): Trends and associations with environmental factors. Journal of Wildlife Diseases 41(1):29-41.

Friborg, Jacques, et al. 1999. p53 inhibition by the LANA protein of KSHV protects against cell death. Nature 402(6764):889-894.

Greenblatt, Rebecca J., et al. 2005a. Genomic variation of the fibropapilloma-associated marine turtle herpesvirus across seven geographic areas and three host species. Journal of Virology 79(2):1125-1132.

Greenblatt, Rebecca J., et al. 2004. The Ozobranchus leech is a candidate mechanical vector for the fibropapilloma-associated turtle herpesvirus found latently infecting skin tumors on Hawaiian green turtles (*Chelonia mydas*). Virology 321(1):101-110.

Greenblatt, Rebecca J., et al. 2005b. Geograhic variation in marine turtle fibropapillomatosis. Journal of Zoo and Wildlife Medicine 36(3):527-530.

Herbst, L., et al. 2004. Tumor outbreaks in marine turtles are not due to recent herpesvirus mutations. Curr Biol 14(17):R697-9.

Herbst, Lawrence H. 1994. Fibropapillomatosis of marine turtles. Annual Review of Fish Diseases 4(0):389-425.

Herbst, Lawrence H., et al. 1998. Serological association between spirorchidiasis, herpesvirus infection, and fibropapillomatosis in green turtles from Florida. Journal of Wildlife Diseases 34(3):496-507.
Herbst, LH, et al. 1995. Experimental transmission of green turtle fibropapillomatosis using cell-free tumor extracts. Diseases of Aquatic Organisms 22(1):1-12.

Herbst, LH, et al. 1996. Sensitivity of the transmissible green turtle fibropapillomatosis agent to chloroform and ultracentrifugation conditions. Diseases of aquatic organisms 25(3):225-228.

Hirama, S, and LM Ehrhart 1999. Prevalence and severity of green turtle fibropapillomatosis in the Indian River lagoon. Florida Scientist 62(1):35.

Hirama, Shigetomo, and Llewellyn M. Ehrhart 2007. Description, prevalence, and severity of green turtle fibropapillomatosis in three developmental habitats on the East coast of Flotida. Florida Scientist 70(4):435-448.

Jacobson, E. R., et al. 1989. Cutaneous fibropapillomas of green turtles (*Chelonia mydas*). Journal of Comparative Pathology 101(1):39-52.

Jia, Haiying, et al. 2014. Long-range PCR in next-generation sequencing: comparison of six enzymes and evaluation on the MiSeq sequencer. Scientific Reports 4:5737.

Kang, K. I., et al. 2008. Localization of fibropapilloma-associated turtle herpesvirus in green turtles (*Chelonia mydas*) by in-situ hybridization. Journal of Comparative Pathology 139(4):218-225.

Lackovich, J. K., et al. 1999. Association of herpesvirus with fibropapillomatosis of the green turtle *Chelonia mydas* and the loggerhead turtle Caretta caretta in Florida. Dis Aquat Organ 37(2):89-97.

Lu, Y., et al. 2000a. Detection of herpesviral sequences in tissues of green turtles with fibropapilloma by polymerase chain reaction. Archives of Virology 145(9):1885-1893.

Lu, Yuanan, et al. 1999. Establishment and characterization of 13 cell lines from a green turtle (Chelonia mydas) with fibropapillomas. In Vitro Cellular & Developmental Biology - Animal 35(7):389-393.

Lu, Yuanan, et al. 2000b. Detection of green turtle herpesviral sequence in saddleback wrasse *Thalassoma duperrey*: A possible mode of transmission of green turtle fibropapilloma. Journal of Aquatic Animal Health 12(1):58-63.

Lucke, Balduin 1938. Studies on tumors in cold-blooded vertebrates. Annual Report of the Tortugas Laboratory of the Carnegie Institute of Washington 38:92-94.

Nei, Masatoshi 1987. Molecular evolutionary genetics: Columbia university press.

Nigro, O., et al. 2004. Sequencing and characterization of the full-length gene encoding the single-stranded DNA binding protein of a novel Chelonian herpesvirus. Archives of Virology 149(2):337-347.

Patrício, A. R., et al. 2012. Global phylogeography and evolution of chelonid fibropapilloma-associated herpesvirus. Journal of General Virology 93(5):1035-1045.

Quackenbush, Sandra L., et al. 2001. Quantitative analysis of herpesvirus sequences from normal tissue and fibropapillomas of marine turtles with real-time PCR. Virology 287(1):105-111.
Quackenbush, Sandra L., et al. 1998. Three closely related herpesviruses are associated with fibropapillomatosis in marine turtles. Virology 246(2):392-399.

Rhoads, Anthony, and Kin Fai Au 2015. PacBio sequencing and its applications. Genomics, Proteomics & Bioinformatics 13(5):278-289.

Smith, GM, and CW Coates 1938. Fibro-epithelial growths of the skin in large marine turtles, *Chelonia mydas* (Linnaeus). Zoologica 23:93-98.

Tamura, Koichiro, et al. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software Version 4.0. Molecular Biology and Evolution 24(8):1596-1599.

Work, T.M., Work, et al. 2004. Retrospective pathology survey of green turtles Chelonia mydas with fibropapillomatosis in the Hawaiian Islands, 1993–2003. Diseases of Aquatic Organisms 62(1-2):163-176.

Work, T. M., et al. 2014. Dynamics of virus shedding and in situ confirmation of chelonid herpesvirus 5 in Hawaiian green turtles with fibropapillomatosis. Veterinary Pathology.

Work, T. M., et al. 2009. In vitro biology of fibropapilloma-associated turtle herpesvirus and host cells in Hawaiian green turtles (*Chelonia mydas*). J Gen Virol 90(Pt 8):1943-50.

Work, T. M., and George H. Balazs 1999. Relating tumor score to hematology in green turtles with fibropapillomatosis in Hawaii. Journal of Wildlife Diseases 35(4):804-807.

Work, T. M., et al. 2003. Bacteraemia in free-ranging Hawaiian green turtles *Chelonia mydas* with fibropapillomatosis. Diseases of aquatic organisms 53(1):41-46.

Zimmerer, Edmund J 1991. Fundamentals of molecular evolution. BioScience 41(7):513-515.

Figures

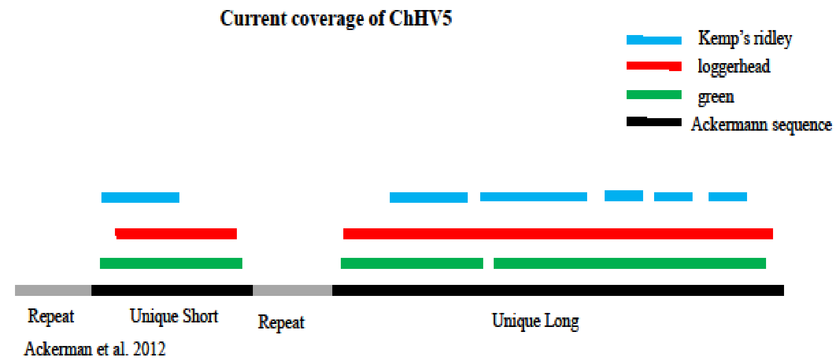


Figure 1. Current coverage of the reference genome by LR-PCR and Illumina sequencing of ChHV5 from Florida green, loggerhead, and Kemp's ridley sea turtles.

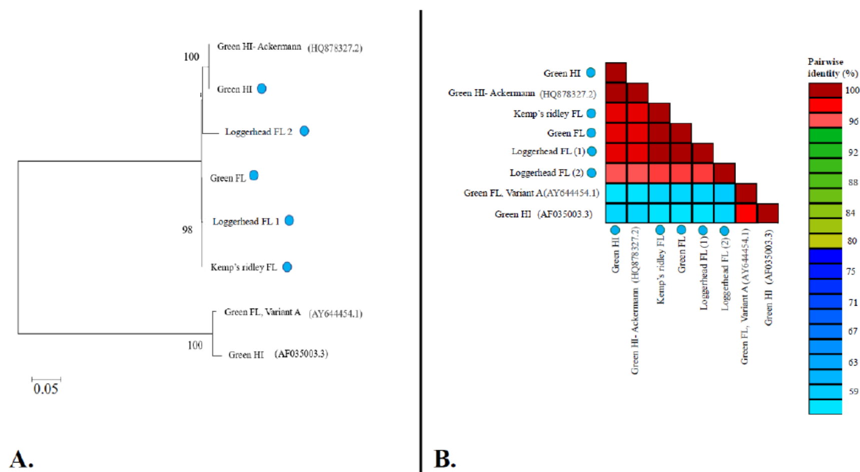


Figure 2. (A) Maximum-Likelihood phylogenetic tree inferred from a MUSCLE alignment of a 4,152 nt region of the capsid gene. (B) Pairwise-identities of the of the same region of the capsid gene.

Species	Collected by	Collection date	Coverage (%)
Loggerhead (1)	Mote	3/1/2008	89.5
Loggerhead (2)	UGA	unknown	83.4
Kemp's ridley (1)	FWC	1/17/2015	87.9
Kemp's ridley (2)	FWC	10/7/2014	45.7
Green	Mote	8/16/2007	68.2

Table 1. Sample summary.

Primer designed by	Primer set	Region	Primer Name	Forward sequence 5'-3'	Primer name	Reverse sequence 5'-3'	Green	Loggerhead (1)	Loggerhead (2)	Kemp's ridley (2)	Kemp's ridley (1)
USGS	1	Repeat	5F	CCCACTAAACCGCTCTCCTC	4654R	ACGAGGAGTCAAGCAAAGGG		N			
USGS	2	Repeat	4110F	TGATGAAGAACCCGAACCCG	9076R	TGACGGGAAAAGAGGGCTTC	N	N	N	N	N
USGS	3	Repeat	7756F	AAAAGGCTTTCTCGGAGGGG	13013R	GCTGCAAATCTCTGCTCGTG	P	N	N	N	P
USF	4	Repeat	8835F	GAACGTTACGGAGGCGAAT	13911R	TTCAGGAGGAGGAGTTTAAAGA			P	P	P
USGS	5	US	12112F	TCGCGTTATCTGACGAGCTC	16987R	GAGATCCTTACTGCCGGACG	P	P	N	N	P
USF	6	US	15558F	CCCTCCCCTGCAATATAAGC	20559R	ATAGCGTGACGTTTCTGCCT			N	N	P
USGS	7	US	15665F	AGTAGAGGTTCGGCTTTGGC	21211R	CGGACAGCATTTCATCTGCG	P	P	N		P
USF	8	US	18796F	GTGCCGCAGTTATTTTTGGT	23827R	GGACGGATGACCGAAGCGT			P	P	P
USGS	9	US	20658F	CGTGCTATTTGAATGCCCCG	26071R	ATCATCGCCACCGGATGTAC	N	P	N		P
USF	10	US	20749F	ATTCGGGATTGTACGTGCTC	25765R	TGTGGGGAGTTTAGGAACCC		P	N	N	N
USGS	11	Repeat	25448F	GGTCCTCTTCTTTGGGACGG	31059R	ATTTCCGCCCCGTACTTCTG		P			
USF	12	Repeat	26013F	GCTTCGTTTTTGGGACCTTT	31017R	CCGTACTTCTGCCAGTTTGG		P	N	P	P
USGS	13	Repeat	30044F	TTCGGAGGAGGAGGAAGAGG	35335R	TTCCGGAGCGAGAAACTCAC		P			
USF	14	Repeat	31321F	AGGGGACGCTCGATCTGTAT	36260R	CGTGTCGTACAGCCGATAAG		P	P	N	P
USGS	15	UL	34335F	CCATAATCGGACGGTCCTGG	39730R	TGCCTGTGTTTCCTCTCGTC	P	P	N		P
USF	16	UL	36311F	GCAGATCACGGACACCATT	41036R	GAGACCTGGCGTTCCATATC			P	P	P
USGS	17	UL	39688F	TCTTGATCAGGTTGGGCGTC	45529R	CCAAAGGGGTGGATGTCTCC	P	P	N		P
USF	18	UL	39697F	ATCAGGTTGGGCGTCCCCGA	45445R	TACAGGCAGCGAGGTTTTTC	P	N	N	N	N
USF	19	UL	41039F	TCAGCGTCGACCTGGTTAC	46068R	TTTCAACGCGGTCATGACC			P	N	P
USGS	20	UL	45288F	TTTACGGAGCACACCTCACC	50531R	GGAGACGATCCTTCAGCGAG	P	N	P	N	P
USGS	21	UL	50001F	AAGTCCGCGAACGTTTTTGG	55036R	AAAGACCCGAAGCCTTGAGG	P	P	P	N	P
USGS	22	UL	58560F	TGGGCTCGTAAACTTCGGAC	63624R	CGGCGAAAGATAACCGCTTG	P	P	P	N	P
USGS	23	UL	62964F	GAAGTCTCCGCGTACCTAGC	68428R	CGATTCGGCTCTTTCGCAAG	P	P	P		P
USF	24	UL	63561F	TCGTTATCTCTCACGGGTACG	68603R	CTTACGTTTCGGGCACCTG			P	P	P
USGS	25	UL	67715F	TTTTCGACAAGCACGGCATG	72955R	GCGATTGATCCAACGTGAGC	P	P	P	P	P
USF	26	UL	67723F	AAGCACGGCATGTCGAGCT	72728R	TATTCGGTTGAGGGCCGGTT			P	P	P
USGS	27	UL	72610F	TTCCTCTTCGGGGGAGAGAG	77568R	ATGAACGAAGACCTGCTGGG	P	P	P	N	P
USF	28	UL	77121F	GGGCGTCAAAGCCTGATAC	81869R	CGCACGGTTACCTGCTGAA			P	P	P
USGS	29	UL	77394F	TCTGCCACAGCTTCAGAGTG	82866R	AACGCCAAGACTGAAGAGGG	P	P	P		P
USGS	30	UL	82303F	AGACCGGTAAAAGGCTTGGG	87753R	GCGTACCGCGTGTTATTTCC	P	P	P		
USF	31	UL	82426F	TATTCCTTGCACGAAGCGCTC	87471F	TTCGGGCTCTATCATAACGA			P	P	P
USGS	32	UL	87591F	CGGAACCCGCTCACAAAAAG	92786R	GTCTTATTGGACGCGCGTTC	P	N	N	N	
USF	33	UL	87768F	ACAACCTGCCTCTCAATTCC	91830R	GTCCCAGAACAAGCGATTTC	P	P	N	N	P
USGS	34	UL	91526F	AATCGTCGACCACACACCTC	96800R	GGCGGTTTTATCGGTGATGC	N	P	P	N	N
USF	35	UL	91530F	TCTCAATCCGGGACTACGAA	96486R	TTTCGAGGTAGAGCTGGCC		P	P	P	P
USF	36	UL	92554F	AAAGATGCGTGACCTCGTG	97740R	GAACGACAACCTGCCTCACT		P	P	N	P
USGS	37	UL	96260F	CCAATTCGATCGCGTTCCAC	101663R	TTTTCAGCTGAAACGCCTGC	P	P	P		
USF	38	UL	97953F	GCTTTAGCGGTTCTTCATCG	102994R	ATACACGGAAAGCGTTGGAGG		P	P	P	P
USGS	39	UL	101478F	TACATTCGACTCCGGCATCG	106607R	ACGGGTGAGTCTTCGGTTTC	P	P	P	N	P
USF	40	UL	106256F	TTCACTATCGCCGTGCAAC	111292R	TCCGCTAGAGATTCACGGTCT			P	N	P
USGS	41	UL	106424F	CGTACGCCGTCAACGAAATC	111576	CGTGTTTTGCTTGGTGGGAG	P	P	P	N	N
USGS	42	UL	111236F	GTGGACGCAAAAGGGTTTCC	116514R	TTTGCGAGCACAACTTGACG	N	P	N	N	N
USF	43	UL	111403F	CCTGACCAGACCGTGAATCT	112772R	GCCACCGAATACCTTTTGAA	P	P	P	N	P
USF	44	UL	111564F	CGTGCGTGTCTCCATCTAAA	116453R	CCAACTTTTCAAGCGAGGAA		P	N	P	P
USGS	45	UL	115877F	TCGCTATCCGGTGCATAACC	121021R	TTGCTTTTGGACAATGGCGG	P	P	P	N	P
USF	46	UL	120801F	TCATGCGGCAAATGTAGAGGT	125708R	CGGATGGGTCACCTCTTATA			P	N	P
USGS	47	UL	120901F	AAACGTGTTCCTTCGCTTGC	125756R	TCCAGTCGAAACTCCGTTGG	P	P	N	N	N
USGS	48	UL	125270F	TTTCACACCCTCTTCGAGCC	130355R	GCAGATGTACGAGTACGGGG	P	P	N	N	P
USGS	49	UL	128058F	CCAAGTGACTGTTTTGCCGG	132233R	GAAATCCATCCCCGGGCC		P			
USF	50	UL	130278F	TTGACCAGGCACAGCATCTT	131974R	TCTGCCAAGTAAGCATTCTGC		P	P	N	P

Table 2. Primer sequences and amplification success in each sample.

APPENDIX 4: DEVELOPMENT OF A SEROLOGICAL ASSAY FOR THE NOVEL SEA LION (*ZALOPHUS CALIFORNIAS*) ANELLOVIRUS, ZCAV

This appendix was published in:

Fahsbender, E., K, Rosario, J.P. Cannon, F. Gulland, L.J. Dishaw, M. Breitbart. Development of a serological assay for the sea lion (*Zalophus californianus*) Anellovirus, ZcAV, Scientific Reports. 5:9637

SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
VIRAL EPIDEMIOLOGY
METAGENOMICS

Received
10 December 2014

Accepted
13 March 2015

Published
12 May 2015

Correspondence and requests for materials should be addressed to M.B. (mya@usf.edu)

Development of a Serological Assay for the Sea Lion (*Zalophus californianus*) Anellovirus, ZcAV

Elizabeth Fahsbender[1], Karyna Rosario[1], John P. Cannon[2], Frances Gulland[3], Larry J. Dishaw[2] & Mya Breitbart[1,2]

[1]University of South Florida, College of Marine Science, 140 7th Avenue South, Saint Petersburg, FL 33701 USA, [2]University of South Florida, College of Medicine, Department of Pediatrics, 801 6th Street South, Saint Petersburg, FL 33701 USA, [3]The Marine Mammal Center, 2000 Bunker Road, Sausalito, CA 94965 USA.

New diseases in marine animals are emerging at an increasing rate, yet methodological limitations hinder characterization of viral infections. Viral metagenomics is an effective method for identifying novel viruses in diseased animals; however, determining virus pathogenesis remains a challenge. A novel anellovirus (*Zalophus californianus* anellovirus, ZcAV) was recently reported in the lungs of captive California sea lions involved in a mortality event. ZcAV was not detected by PCR in the blood of these animals, creating the inability to assess the prevalence of ZcAV in live sea lions. This study developed an enzyme-linked immunosorbent assay (ELISA) to detect antibodies to ZcAV in sea lion serum. To assess ZcAV prevalence, paired serum and lung samples (n = 96) from wild sea lions that stranded along the California coast were tested through ELISA and PCR, respectively. Over 50% of the samples tested positive for ZcAV by ELISA (34%), PCR (29%), or both (11%) assays. ZcAV is prevalent in stranded wild sea lion populations and results suggest that PCR assays alone may grossly underestimate ZcAV exposure. This ELISA provides a tool for testing live sea lions for ZcAV exposure and is valuable for subsequent studies evaluating the potential pathogenicity of this anellovirus.

The rate of emergence of new diseases in marine animals is increasing[1–4], generating a need for surveillance of potential pathogens to protect marine mammals against epidemics. However, it remains difficult to characterize and diagnose viral infections because of methodological limitations[5]. Current detection methods such as degenerate PCR and pan-viral microarrays can detect close relatives of previously described viruses, but are limited for detecting novel viruses. Viral metagenomics (virus particle purification followed by shotgun sequencing) is an effective method for identifying viruses involved in mortality events in marine animals[6–9], yet it remains difficult to establish a connection between a novel virus and disease due to limitations of culturing the viruses as well as difficulties in obtaining fresh diagnostic tissues from wild marine mammals.

Viral metagenomics performed on lung tissue of several necropsied captive California sea lions (*Zalophus californianus*) with signs of respiratory disease that were involved in a mortality event of unknown cause revealed the presence of a novel anellovirus, *Zalophus californianus* anellovirus (ZcAV)[7]. ZcAV was found by specific PCR in the lungs of all three of the sea lions that died in the mortality event, but it was not found in sea lions from the same zoo that died of unrelated causes. In addition, ZcAV was found to be actively replicating in lung tissue of a sea lion from the mortality event, further suggesting an association of ZcAV with the death of these animals. In addition to the captive sea lions, 11% of lung samples from wild sea lions stranded off the California coast tested positive for ZcAV by PCR, indicating that this anellovirus is present in wild populations.

The initial ZcAV discovery and prevalence studies have raised many questions about the potential pathogenicity of this virus. Anelloviruses have been extensively studied in humans, where these viruses can be highly prevalent (infecting up to 100% of the population)[10–13], yet they have not been linked to human disease. Although anelloviruses have also been found in a wide range of mammals including non-human primates[14], domestic animals[15], Pacific harbor seals[9], and Risso's dolphins (E.M. Fahsbender et al., unpublished data), anellovirus-associated pathology remains unknown. Initial evidence suggested that ZcAV may be linked to the mortality event of captive California sea lions; however, studies investigating the pathogenicity of this virus are difficult since ZcAV has only been detected by PCR in tissues of necropsied animals, and cannot be detected by PCR in the blood of infected individuals[7]. The inability to detect ZcAV in blood samples by PCR severely limits further

testing for this virus, since obtaining lung biopsies from live sea lions is not possible. Hence, there is a critical need for an assay to detect ZcAV exposure in blood samples to investigate the epidemiology of this virus, understand its association with disease, and preemptively develop management strategies that can prevent the spread of this virus in captive and rehabilitation animals. To overcome these technical limitations of studying the role of ZcAV in disease, here we describe the development of an enzyme-linked immunosorbent assay (ELISA) for ZcAV, and demonstrate that sea lions mount an immune response against ZcAV.

Similar to other anelloviruses[16], ZcAV contains a small (2140 nucleotide (nt)), negative sense, single-stranded DNA circular genome that encodes three major open reading frames (ORFs). Based on similarities to other anelloviruses, ORF 1 is believed to encode the capsid protein, although this has not been experimentally demonstrated for any anellovirus. For other anelloviruses, the ORF 1 gene product has been predicted to be antigenic due to the presence of major hydrophilic regions, and ORF 1 has been successfully used in seroprevalence studies in humans and pigs[11,17–20]. Here we developed an ELISA based on hydrophilic regions of the ORF 1 gene product of the ZcAV genome and demonstrated that this assay is capable of detecting anti-ZcAV antibodies in sea lion serum. This ELISA provides a tool for studying ZcAV epidemiology and identifying seroconversion upon symptom development in captive sea lions, enabling future research investigating the pathogenesis of ZcAV. Finally, the creation of this assay lays the groundwork for bridging the gap between genome discovery via viral metagenomics and assessing the epidemiology of novel viruses and their significance for wild populations.

Methods

Sample collection. Paired serum and lung samples were collected from 100 California sea lions involved in stranding events along the coast of northern California. All samples were collected by The Marine Mammal Center located in Sausalito, California in accordance with Marine Mammal Protection Act permit no 932-1905/MA-009526 to Gulland. Serum samples were collected during routine clinical examinations of live stranded sea lions as described by Bossart et al.[21] and archived at -70°C. Lung samples were collected from the same animals upon necropsy, with animals stored at 4°C between time of death and necropsy (within 24 hours of death). All protocols were approved under the Marine Mammal Protection Act. Paired lung and serum samples were shipped to the University of South Florida and stored at -80°C for processing using PCR (for lung samples) and ELISA (for serum samples).

PCR testing of sea lion lung samples. Lung samples were tested with a ZcAV-specific PCR assay. For this purpose, DNA was extracted from approximately 25 mg of sea lion lung tissue using the DNeasy Blood and Tissue Kit (Qiagen). DNA was amplified through rolling circle amplification (RCA) using the Illustra Templiphi Kit (GE Healthcare), which is known to enrich for small circular templates such as the ZcAV genome[22,23]. RCA products were then used as the DNA templates for the PCR assay using primers ZcAV1056F (5′-AGG CAC TCA CAT AAT CTA TTC AA-3′) and ZcAV1369R (5′-CCC AGG CAT TAC AGG CTT TA-3′) designed to target a 314 nt region from ORF 1 of ZcAV. The PCR [containing 1 μM of each primer, 200 μM of dNTPs, 1 U of Red *Taq* (Sigma-Aldrich), 1X Red *Taq* reaction buffer and 2.5 μl of DNA template in a 25 μl reaction] proceeded as follows: 95°C for 5 minutes, 45 cycles of [94°C for 1 minute, 54°C for 1 minute, and 72°C for 1 minute], followed by a final extension at 72°C for 10 minutes. PCR products of the correct size were verified by gel electrophoresis. Four samples were cloned using a TOPO TA Cloning Kit (Invitrogen) and commercially sequenced. Sequences were then compared against the GenBank database in order to verify the specificity of the assay. To determine assay sensitivity, a positive control was prepared as previously described. The PCR product was cleaned using the UltraClean PCR Clean-Up Kit (Mo Bio Laboratories) and quantified using a NanoDrop ND-1000 apparatus (NanoDrop Technologies). The number of targets per microliter was back-calculated and the positive control was serially diluted to determine the sensitivity of the assay, which was 10 targets.

Optimization of ELISA parameters. An ELISA was developed to screen sea lion serum for antibodies to ZcAV. Since the virus was identified from a mixed community using metagenomics, there were no purified virus particles that could be used as an antigen for the ELISA. Therefore, the ELISA was developed based on regions predicted to be immunogenic based on hydrophilic sections found in the predicted protein sequence of ORF 1. Four peptides (on average, 25 amino acids in length) from different regions of ORF 1 were synthesized at 70% purity by Pacific Immunology. Each of these peptides was scrambled, synthesized and used as a control peptide. Sea lion sera were used as primary antibodies, while alkaline phosphatase (AP)-conjugates of protein A (Roche) and protein L (Thermo Scientific) were used as detection reagents. Protein A is a bacterially-derived molecule that binds to the Fc region of IgG of many mammalian species; in contrast, protein L, also derived from bacteria, binds to immunoglobulin light chains and thus has the ability to detect a wider range of serum antibodies, including IgG, IgM, IgA and IgE[24–26]. Similar reagents have been used previously in sea lion serological testing[26]. Checkerboard titrations were performed to find optimal concentrations of peptide, serum, and proteins. After initial screening, peptide 4T (amino acid sequence: GMENTPPKRVRFRQSDVLRKHKHRI), along with its scrambled control peptide 4C (amino acid sequence: QRLHPKKHIRSETKFVRVDRNPGMR) were chosen as primary targets due to their low background. Protein L was selected as the detection reagent since it is able to bind more classes of immunoglobulins than protein A.

ELISA testing of sea lion serum samples. High binding 96-well microtiter plate wells were coated with a final concentration of 1 μg/ml peptide diluted in 0.5 M carbonate buffer (pH 9.6) (1.59 g of Na_2CO_3, 2.93 g of $NaHCO_3$, 0.2 g of NaN_3, and dH_2O to 1 liter) and the plate was incubated overnight at 4°C. All wells were washed three times with a phosphate-buffered saline (PBS) solution containing 0.1% Tween-20 (PBS-tween) using an automated microtiter plate washer, and then blocked with 1% bovine serum albumin (BSA) diluted 1 : 5 in PBS-tween for 1 hour. Sea lion serum samples were diluted 1 : 50 in PBS-tween containing 1% BSA, added to each well in triplicates, and left to incubate overnight at 4°C. The plate was then washed and blocked as previously described before secondary reagent protein L was added in PBS-tween containing 1% BSA at a 1 : 10,000 dilution and left to incubate at 37°C for 1 hour. Each well was washed as previously described. ELISA Blue (SureBlue Reserve) was added and incubated at room temperature for 10–20 minutes. Plates were read on a PerkinElmer Enspire machine to measure optical density (OD) at 650 nm.

Statistical Analysis. The statistical software package GraphPad Instat (GraphPad Software, Inc.) was used to perform an analysis of variance (ANOVA) on averages of the triplicate sea lion serum samples. Samples with p-values <0.05 were considered statistically significant. Samples were considered positive if serum incubated with the test peptide (4T) had a statistically significant higher OD value than the serum incubated with the control peptide (4C).

Results

Development of the ELISA using serum samples from sea lions whose lung samples were PCR-positive for ZcAV demonstrated that sea lions mount an immune response to ZcAV infection. Paired serum and lung samples from 100 stranded California sea lions were then tested for the presence of ZcAV nucleic acids and anti-ZcAV antibodies through PCR and the newly developed ELISA, respectively. Four samples were removed from the data set due to high background with the ELISA. Of the 96 samples tested, 50 (52%) tested positive for ZcAV exposure (ELISA; serum samples; $n=33$ [34%]) and/or DNA presence (PCR; lung samples; $n=28$ [29%]) (Table 1). Eleven percent of the sea lions tested positive for ZcAV by both ELISA and PCR, and 48% of sea lions tested negative by both methods. Of the 28 sea lions with PCR-positive lung samples, 39% were also serum-positive by ELISA. However, 32% of the PCR-negative sea lions ($n=68$) also tested positive for ZcAV exposure via ELISA. The Kappa value of 0.066, calculated using GraphPad, indicates poor correlation between the PCR and ELISA test results.

Discussion

The ELISA developed in this study provides the first and only method to test for ZcAV exposure in live sea lions. Here, this assay was used in combination with PCR to assess the prevalence of ZcAV in the wild, stranded sea lion population along the northern coast of California. Protein-L-based detection of ZcAV-specific antibodies suggests the exposed sea lions are able to recognize and mount immune responses to the virus. These data, which are the first obtained for marine mammal anelloviruses, support prior studies showing that human anelloviruses are immunogenic, with both IgG and IgM class antibodies produced upon exposure to ORF 1[27,28]. The combined ELISA and PCR results demonstrate that ZcAV infections are prevalent and persistent in wild California sea lions.

The development of this ELISA was unconventional due to the many difficulties of working with marine mammals. Gaining access to 100 paired serum and lung samples from a protected species is both arduous and time consuming. Furthermore, the sea lion

Table 1 | Prevalence of ZcAV in paired serum and lung samples from a stranded wild population of California sea lions (*Zalophus californianus*; n=96) assessed by ELISA and PCR, respectively

	Serum ELISA positive n= (%)	Serum ELISA negative n= (%)	PCR totals n= (%)
Lung PCR positive n= (%)	11 (11%)	17 (18%)	28 (29%)
Lung PCR negative n= (%)	22 (23%)	46 (48%)	68 (71%)
ELISA totals n= (%)	33 (34%)	63 (66%)	

immune system is not frequently studied; therefore, there are no commercial antibodies for sea lions. In addition to the many ambiguities of the sea lion immune system, ZcAV is not a well-understood virus. ZcAV was discovered from a mixed community and attempts to isolate ZcAV virions in culture have been unsuccessful, limiting sensitivity testing of this ELISA. In addition, without more knowledge of the types of related anelloviruses found in marine mammals, the specificity of this assay cannot be determined.

Although the ELISA and PCR results obtained in this study were not correlated, the ZcAV ELISA provides complementary results to the PCR data, providing a greater scope to measure virus prevalence, as use of the ZcAV ELISA increased the detection of exposure to ZcAV. There are multiple explanations for the incongruences between the ELISA and PCR data. ELISA-positive, but PCR-negative samples (23%) are likely due to previous ZcAV infections in which the viral DNA is no longer present but the ELISA can detect the presence of (long-lived) antibodies, suggesting a prior exposure. Similar results have been seen for the human anellovirus Torque teno virus (TTV), where antibodies were detected despite the absence of TTV DNA[11]. In addition, if the ZcAV infection occurred in another sea lion organ instead of the lungs, we could expect a PCR-negative but ELISA-positive result. Conversely, sea lion samples may have tested positive by PCR but negative by ELISA (18%) if the sea lions were actively infected upon stranding and died before a detectable antibody response was mounted. Although the timing of the response of the sea lion immune system to ZcAV is unknown, development of the human antibody response to TTV is slow and known to take up to 21 weeks after virus inoculation[29], although once the antibodies are produced they persist in serum for long periods[28,30]. Finally, slight sequence mutations or antigenic diversity of anelloviruses may also contribute to data incongruency. Anelloviruses are known to demonstrate a high degree of genetic variability[19,31,32] which can pose problems for both DNA detection by PCR primers and antibody detection. For certain TTV genotypes, it has been shown that multiple serotypes may exist in circulation[20]. Future work should compare the diversity of anelloviruses that can be detected by the PCR and ELISA utilized in this study, as the discrepancy between the two assay types may reflect differences in the diversity of genotypes and serotypes recovered.

Through a combination of PCR- and ELISA-based testing, this study demonstrated that sea lions mount an immune response to the anellovirus ZcAV and revealed the high prevalence of ZcAV in a stranded wild sea lion population. The ELISA created here will enable future research on the epidemiology of ZcAV in live sea lions and allow further study of ZcAV pathogenesis by measuring seroconversion in infected captive sea lions. Finally, this assay has direct implications for protecting sea lion health since it is the only method available for the detection of ZcAV in live sea lions. The ZcAV ELISA enables the screening of rehabilitation animals before admitting them into facilities with healthy animals, which can prevent the spread of this virus to other captive sea lions and provide a valuable management step if further work links ZcAV with disease.

1. Harvell, C. D. *et al*. Emerging marine diseases--climate links and anthropogenic factors. *Science* **285**, 1505–1510 (1999).
2. Gulland, F. D. & Hall, A. Is marine mammal health deteriorating? Trends in the global reporting of marine mammal disease. *EcoHealth* **4**, 135–150 (2007).
3. Jones, K. E. *et al*. Global trends in emerging infectious diseases. *Nature* **451**, 990–993 (2008).
4. Keesing, F. *et al*. Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature* **468**, 647–652 (2010).
5. Delwart, E. L. Viral metagenomics. *Rev Med Virol* **17**, 115–131 (2007).
6. Ng, T. F. F. *et al*. Discovery of a novel single-stranded DNA virus from a sea turtle fibropapilloma by using viral metagenomics. *J Virol* **83**, 2500–2509 (2009).
7. Ng, T. F. F., Suedmeyer, W. K., Wheeler, E., Gulland, F. & Breitbart, M. Novel anellovirus discovered from a mortality event of captive California sea lions. *J Gen Virol* **90**, 1256–1261 (2009).
8. Li, L. *et al*. The fecal viral flora of California sea lions. *J Virol* **85**, 9909–9917 (2011).
9. Ng, T. F. F. *et al*. Metagenomic identification of a novel anellovirus in Pacific harbor seal (*Phoca vitulina richardsii*) lung samples and its detection in samples from multiple years. *J Gen Virol* **92**, 1318–1323 (2011).
10. Prescott, L. E., Simmonds, P. & Collaborators & International. Global distribution of transfusion-transmitted virus. *New Engl J Med* **339**, 776–777 (1998).
11. Handa, A., Dickstein, B., Young, N. S. & Brown, K. E. Prevalence of the newly described human circovirus, TTV, in United States blood donors. *Transfusion* **40**, 245–251 (2000).
12. Biagini, P. *et al*. Distribution and genetic analysis of TTV and TTMV major phylogenetic groups in French blood donors. *J Med Virol* **78**, 298–304 (2006).
13. Okamoto, H. in *TT Viruses* Vol. 331 *Current Top Microbiol* (eds Ethel-Michele de Villiers & Haraldzur Hausen) Ch. 1, 1–20 (Springer Berlin Heidelberg, 2009).
14. Romeo, R. *et al*. High prevalence of TT virus (TTV) in naive chimpanzees and in hepatitis C virus-infected humans: frequent mixed infections and identification of new TTV genotypes in chimpanzees. *J Gen Virol* **81**, 1001–1007 (2000).
15. Biagini, P. *et al*. Circular genomes related to anelloviruses identified in human and animal samples by using a combined rolling-circle amplification/sequence-independent single primer amplification approach. *J Gen Virol* **88**, 2696–2701 (2007).
16. Biagini, P. in *TT Viruses* Vol. 331 *Current Top in Microbiol* (eds Ethel-Michele Villiers & Haraldzur Hausen) Ch. 2, 21–33 (Springer Berlin Heidelberg, 2009).
17. Huang, Y. W. *et al*. Expression of the putative ORF1 capsid protein of Torque teno sus virus 2 (TTSuV2) and development of western blot and ELISA serodiagnostic assays: Correlation between TTSuV2 viral load and IgG antibody level in pigs. *Virus Res* **158**, 79–88 (2011).
18. Huang, Y.-W. *et al*. Serological profile of torque teno sus virus species 1 (TTSuV1) in pigs and antigenic relationships between two TTSuV1 genotypes (1a and 1b), between two species (TTSuV1 and -2), and between porcine and human anelloviruses. *J Virol* **86**, 10628–10639 (2012).
19. Okamoto, H. *et al*. The entire nucleotide sequence of a TT virus isolate from the United States (TUS01): Comparison with reported isolates and phylogenetic analysis. *Virology* **259**, 437–448 (1999).
20. Ott, C. *et al*. Use of a TT virus ORF1 recombinant protein to detect anti-TT virus antibodies in human sera. *J Gen Virol* **81**, 2949–2958 (2000).
21. Bossart, G. D., Reidarson, T. H., Dierauf, L. A. & Duffield, D. A. in *CRC Handbook of Marine Mammal Medicine* 2nd edn (eds Dierauf, L.A. and Gulland, F.M.D.) (CRC Press 2001).
22. Kim, K.-H. *et al*. Amplification of uncultured single-stranded DNA viruses from rice paddy soil. *Appl Environ Microbiol* **74**, 5975–5985 (2008).
23. Kim, K.-H. & Bae, J.-W. Amplification methods bias metagenomic libraries of uncultured single-stranded and double-stranded DNA viruses. *Appl Environ Microbiol* **77**, 7663–7668 (2011).
24. Kessler, S. W. Rapid isolation of antigens from cells with a staphylococcal protein A-antibody adsorbent: Parameters of the interaction of antibody-antigen complexes with protein A. *J Immunol* **115**, 1617–1624 (1975).
25. Nilson, B. H., Solomon, A., Björck, L. & Akerström, B. Protein L from *Peptostreptococcus magnus* binds to the kappa light chain variable domain. *J Biol Chem* **267**, 2234–2239 (1992).
26. De Chateau, M. *et al*. On the interaction between protein L and immunoglobulins of various mammalian species. *Scand J Immunol* **37**, 399–405 (1993).
27. Tsuda, F. *et al*. Determination of antibodies to TT virus (TTV) and application to blood donors and patients with post-transfusion non-A to G hepatitis in Japan. *J Virol Methods* **77**, 199–206 (1999).
28. Tsuda, F. *et al*. IgM-class antibodies to TT virus (TTV) in patients with acute TTV infection. *Hepatol Res* **19**, 1–11 (2001).
29. Tawara, A. *et al*. Transmission of human TT virus of genotype 1a to chimpanzees with fecal supernatant or serum from patients with acute TTV infection. *Biochem Biophys Res Commun* **278**, 470–476 (2000).

30. Maggi, F. & Bendinelli, M. in *TT Viruses* Vol. 331 *Current Top Microbiol* (eds Ethel-Michele de Villiers & Haraldzur Hausen) Ch. 5, 65–90 (Springer Berlin Heidelberg, 2009).
31. Okamoto, H. *et al*. Molecular cloning and characterization of a novel DNA virus (TTV) associated with posttransfusion hepatitis of unknown etiology. *Hepatol Res* **10**, 1–16 (1998).
32. Hijikata, M., Takahashi, K. & Mishiro, S. Complete circular DNA genome of a TT virus variant (isolate name SANBAN) and 44 partial ORF2 sequences implicating a great degree of diversity beyond genotypes. *Virology* **260**, 17–22 (1999).

Acknowledgments

This project was funded through an Interdisciplinary Research Development Grant from the University of South Florida's College of Marine Science. The authors thank Dr. Michael Shamblott, Marci O'Driscoll, and Elizabeth Wheeler for guidance with experiments, use of equipment, and feedback on the manuscript, and the staff and volunteers of The Marine Mammal Center for care and sampling of the stranded sea lions.

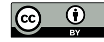
Author contributions

E.F. developed the ELISA, tested samples, and wrote the manuscript. F.G. collected the sea lion samples. K.R., M.B., J.C. and L.D. advised and aided with the development of the ELISA and data analysis. All authors reviewed the manuscript.

Additional information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Fahsbender, E. *et al*. Development of a Serological Assay for the Sea Lion (*Zalophus californianus*) Anellovirus, ZcAV. *Sci. Rep*. **5**, 9637; DOI:10.1038/srep09637 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/