Viruses in marine animals: Discovery, detection, and characterization

Elizabeth Fahsbender
University of South Florida, efahsben@mail.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
ACKNOWLEDGMENTS

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 (opportunist 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
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 californianus*) 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 to 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


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

This appendix was accepted to be published in Virus Evolution.

Diverse and highly recombinant anelloviruses associated with Weddell seals in Antarctica

Elizabeth Fahsbender¹, Jennifer M. Burns²*, Stacy Kim³, Simona Kraberger⁴,⁵, Greg Frankfurter⁶, Alice A. Eilers⁷, Michelle R. Shero², Roxanne Beltran²,⁸, Amy Kirkham²,⁹, Robert McCorkell¹⁰, Rachel Berngartt¹¹, Maketalena F. Male⁵,¹², Grant Ballard¹³, David G. Ainley¹⁴, Mya Breitbart¹, Arvind Varsani⁴,⁵,¹⁵*

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

Point Blue Conservation Science, Petaluma, CA 94954, USA

HT Harvey and Associates, Los Gatos, CA 95032, USA

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 marten, 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 MgSO₄) 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 × 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/](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**


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.,


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


<table>
<thead>
<tr>
<th>Order</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artiodactyla</td>
<td>Sus scrofa</td>
</tr>
<tr>
<td>Carnivora</td>
<td>Canis lupus</td>
</tr>
<tr>
<td>Carnivora</td>
<td>Leptonychotes weddellii</td>
</tr>
<tr>
<td>Carnivora</td>
<td>Phoca vitulina</td>
</tr>
<tr>
<td>Carnivora</td>
<td>Zalophus californianus</td>
</tr>
<tr>
<td>Carnivora</td>
<td>Martes martes</td>
</tr>
<tr>
<td>Carnivora</td>
<td>Felis catus</td>
</tr>
<tr>
<td>Chiroptera</td>
<td>Tadarida brasiliensis</td>
</tr>
<tr>
<td>Diptera</td>
<td>Anopheles sp.</td>
</tr>
<tr>
<td>Equine</td>
<td>Equus caballus</td>
</tr>
<tr>
<td>Primate</td>
<td>Aotus sp.</td>
</tr>
<tr>
<td>Primate</td>
<td>Chlorocebus sabaesus</td>
</tr>
<tr>
<td>Primate</td>
<td>Gorilla gorilla</td>
</tr>
<tr>
<td>Primate</td>
<td>Homo sapiens</td>
</tr>
<tr>
<td>Primate</td>
<td>Pan troglodytes</td>
</tr>
<tr>
<td>Primate</td>
<td>Saguinus midas</td>
</tr>
<tr>
<td>Primate</td>
<td>Macaca fuscata</td>
</tr>
<tr>
<td>Rodentia</td>
<td>Microtus agrestis</td>
</tr>
<tr>
<td>Rodentia</td>
<td>Myodes glareolus</td>
</tr>
<tr>
<td>Rodentia</td>
<td>Rattus norvegicus</td>
</tr>
<tr>
<td>Rodentia</td>
<td>Apodemus sylvaticus</td>
</tr>
</tbody>
</table>

0.3 nucleotide substitutions per site
Species

Leptonychotes weddellii
Phoca vitulina
Zalophus californianus

Branch support
- 60
- 62
- 64
- 66
- 68
- 70
- 72
- 74
- 76
- 78
- 80
- 82
- 84
- 86
- 88
- 90
- 92
- 94
- 96
- 98
- 100

Sample type
SKP
WSK
WSP
WSV
WSN
Order | Species
--- | ---
Artiodactyla | Sus scrofa
Carnivora | Canis lupus
Carnivora | Leptonychotes weddellii
Carnivora | Phoca vitulina
Carnivora | Zalophus californianus
Carnivora | Martes martes
Chiroptera | Tadarida brasiliensis
Equine | Equus caballus
Carnivora | Felis catus
Primate | Aotus sp.
Primate | Chlorocebus sabaeus
Primate | Gorilla gorilla
Primate | Homo sapiens
Primate | Pan troglodytes
Primate | Saginus midas
Primate | Macaca fuscata
Rodentia | Microtus agrestis
Rodentia | Myodes glareolus
Rodentia | Rattus norvegicus
Rodentia | Apodemus sylvaticus

0.3 nucleotide substitutions per site
<table>
<thead>
<tr>
<th>Nasal</th>
<th>Nasal (resampled)</th>
<th>Vaginal</th>
<th>Vaginal (resampled)</th>
<th>Faeces</th>
<th>Kidney</th>
</tr>
</thead>
<tbody>
<tr>
<td>WSN04; WS14-04</td>
<td>WSV04; WS14-04</td>
<td>WSN06; WS14-06</td>
<td>WSV06; WS14-06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSN07; WS14-07</td>
<td>WSN39; WS14-39</td>
<td>WSV07; WS14-07</td>
<td>WSV39; WS14-39</td>
<td>WSP46; WS14-08</td>
<td></td>
</tr>
<tr>
<td>WSN08; WS14-08</td>
<td>WSN41; WS14-41</td>
<td>WSV08; WS14-08</td>
<td>WSV41; WS14-41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSN09; WS14-09</td>
<td>WSV09; WS14-09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSN10; WS14-10</td>
<td>WSN33; WS14-33</td>
<td>WSV10; WS14-10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSN11; WS14-11</td>
<td>WSN27; WS14-27</td>
<td>WSV11; WS14-11</td>
<td>WSV27; WS14-27</td>
<td>WSP47; WS14-11</td>
<td></td>
</tr>
<tr>
<td>WSN13; WS14-13</td>
<td>WSV13; WS14-13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSN15; WS14-15</td>
<td>WSN45; WS14-45</td>
<td>WSV15; WS14-15</td>
<td>WSV45; WS14-45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSN21; WS14-21</td>
<td>WSN36; WS14-36</td>
<td>WSV21; WS14-21</td>
<td>WSV36; WS14-36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSN22; WS14-22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSN23; WS14-23</td>
<td>WSN40; WS14-40</td>
<td>WSV23; WS14-23</td>
<td>WSV40; WS14-40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSN24; WS14-24</td>
<td>WSN26; WS14-26</td>
<td>WSV24; WS14-26</td>
<td>WSV26; WS14-26</td>
<td>WSP48; WS14-14</td>
<td></td>
</tr>
<tr>
<td>WSN25; WS14-25</td>
<td>WSN14; WS14-14</td>
<td>WSV25; WS14-25</td>
<td>WSP49; WS14-14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSN26; WS14-26</td>
<td>WSV26; WS14-26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSN29; WS14-29</td>
<td>WSV29; WS14-29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSN30; WS14-30</td>
<td>WSV30; WS14-30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSN31; WS14-31</td>
<td>WSV31; WS14-31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSN32; WS14-32</td>
<td>WSV32; WS14-32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSN34; WS14-34</td>
<td>WSV34; WS14-34</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSN35; WS14-35</td>
<td>WSV35; WS14-35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSN37; WS14-37</td>
<td>WSV37; WS14-37</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSN38; WS14-38</td>
<td>WSV38; WS14-38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSN42; WS14-42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSN43; WS14-43</td>
<td>WSV43; WS14-43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSN44; WS14-44</td>
<td>WSV05; WS14-05</td>
<td>WSV44; WS14-44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WSN46; WS14-46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

WS01, SK3
WS02, SK15
WS03, SK20
WS04, SK21
WS05, SK24
WS09, SK25
WS09, SK26
WS10, SK27
WS11, SK28
WS12, SK29
WS13, SK30
WS14, SK32
WS15, SK34
WS16, SK35
WS17, SK36
WS18, SK37
WS19, SK38
WS20, SK39
WS21, BR1
WS22, BR2
WS23, BR3
WS24, BR4
WS25, HC1
WS26, HC2
WS27, HC3
WS28, HC4
WS29, HC5
WS30, HC6
WS31, HC7
WS32, HC8
WS33, HC9
WS34, HC10
WS35, HC11
WS36, HC12
WS37, HC13
WS38, TR1
WS39, TR2
WS40, TR3
WS41, TR4
WS42, TR5
WS43, RB5
WS44, RB6
WS45, HC8

WSK
<table>
<thead>
<tr>
<th>Primer</th>
<th>TTLwV type</th>
<th>Sequence (5’→3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>anello1_F</td>
<td>1</td>
<td>AGAAGAAAGAGAATCTCCGTTMAGGGTTGGG</td>
</tr>
<tr>
<td>anello1_R</td>
<td></td>
<td>CCGTGAGGGGTTATATCTTTACTGTAGC</td>
</tr>
<tr>
<td>anello2_F</td>
<td>1</td>
<td>GAAAAATCTGATTGGCTGCCGGGGATG</td>
</tr>
<tr>
<td>anello2_R</td>
<td></td>
<td>TCCGCTTTTTAACAACCTTTTTTCGAGGTAC</td>
</tr>
<tr>
<td>12092anello_F</td>
<td>1</td>
<td>AAAAACCTCCTCTATAGAGCTCAGTACTAC</td>
</tr>
<tr>
<td>12092anello_R</td>
<td></td>
<td>GAAGGTAAGAAATGGTGACCCAGG</td>
</tr>
<tr>
<td>12124anello_F</td>
<td>1</td>
<td>CCATAATTATCCATTGGGTCTGAGAAG</td>
</tr>
<tr>
<td>12124anello_R</td>
<td></td>
<td>CAGATGGGGACCCCTCTGCTCATATG</td>
</tr>
<tr>
<td>12421anello_F</td>
<td>2</td>
<td>TGGGAAGGACTCTTTAACATCACCAGC</td>
</tr>
<tr>
<td>12421anello_R</td>
<td></td>
<td>GCTAGCAGGGGGTTCTACGTATAC</td>
</tr>
</tbody>
</table>
APPENDIX 2: DISCOVERY OF A NOVEL CIRCULAR DNA VIRUS IN THE FORBES SEA STAR, ASTERIAS FORBESI

This appendix was published in:

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

Elizabeth Fahsbender1 · Ian Hewson2 · Karyna Rosario1 · Allison D. Tuttle3 · Arvind Varsani4,5,6 · Mya Breitbart1

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.

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 Circovirus [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.

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 μl 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 to 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
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 Fahsbender¹, Makenzie Burrows¹, Cheryl Morrison², Mya Breitbart¹

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 EIIIs 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


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 same region of the capsid gene.

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

Table 1. Sample summary.
Table 2. Primer sequences and amplification success in each sample.

<table>
<thead>
<tr>
<th>Primer designed by</th>
<th>Primer set</th>
<th>Primer Name</th>
<th>Forward sequence 5'→3'</th>
<th>Reverse sequence 5'→3'</th>
<th>Green</th>
<th>Logerhead (1)</th>
<th>Loggerhead (2)</th>
<th>Kemp's ridley (1)</th>
<th>Kemp's ridley (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>USGS</td>
<td>1</td>
<td>Repeat</td>
<td>GCCACCTAACCACCTCCCTC</td>
<td>ACCAGGAGCTACACACGCGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>2</td>
<td>Repeat</td>
<td>TGAGGCTGAGGACCAACGG</td>
<td>TGGCGAGAGAGAGACTTGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>3</td>
<td>Repeat</td>
<td>ATTCATGTCATTCACGCTTC</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>4</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>5</td>
<td>Repeat</td>
<td>TGTTGCTACATAGCTGAGC</td>
<td>TGTTGCTACATAGCTGAGC</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>USGS</td>
<td>6</td>
<td>Repeat</td>
<td>GGAGAATCGACAGTACGAGC</td>
<td>GGAGAATCGACAGTACGAGC</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>USGS</td>
<td>7</td>
<td>Repeat</td>
<td>ATTCATGTCATTCACGCTTC</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>8</td>
<td>Repeat</td>
<td>TGTTGCTACATAGCTGAGC</td>
<td>TGTTGCTACATAGCTGAGC</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>USGS</td>
<td>9</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>10</td>
<td>Repeat</td>
<td>TGTTGCTACATAGCTGAGC</td>
<td>TGTTGCTACATAGCTGAGC</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>USGS</td>
<td>11</td>
<td>Repeat</td>
<td>TGTTGCTACATAGCTGAGC</td>
<td>TGTTGCTACATAGCTGAGC</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
</tr>
<tr>
<td>USGS</td>
<td>12</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>13</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>14</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>15</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>16</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>17</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>18</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>19</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>20</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>21</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>22</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>23</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>24</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>25</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>26</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>27</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>28</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>29</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>30</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>31</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>32</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>33</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>34</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>35</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>36</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>37</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>38</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>39</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>40</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>41</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>42</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>43</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>44</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>45</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>46</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>47</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>48</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>49</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>USGS</td>
<td>50</td>
<td>Repeat</td>
<td>TTCTTCTCTGCTATGACGTT</td>
<td>TTATATAGTACTGACGGG</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>
APPENDIX 4: DEVELOPMENT OF A SEROLOGICAL ASSAY FOR THE NOVEL SEA LION (ZALOPHUS CALIFORNIAS) ANELLOVIRUS, ZCAV

This appendix was published in:

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

Elizabeth Fahsbender¹, Karyna Rosario¹, John P. Cannon², Frances Gulland³, Larry J. Dishaw² & Mya Breitbart¹,²

¹University of South Florida, College of Marine Science, 140 7th Avenue South, Saint Petersburg, FL 33701 USA, ²University of South Florida, College of Medicine, Department of Pediatrics, 801 6th Street South, Saint Petersburg, FL 33701 USA, ³The 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 increasing¹⁴, 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 limitations⁵. 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 animals⁶–⁹, 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 primates¹⁴, domestic animals¹⁵, Pacific harbor seals⁶, 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 individuals⁷. 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, 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 pigs11,17-20. 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 seroconverion 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 strandings 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 genome22-24. RCA products were then used as the DNA templates for the PCR assay using primers ZcAV1056F (5‘-9637 | DOI: 10.1038/srep09637)

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 Na₂CO₃, 2.93 g of NaHCO₃, 0.2 g of NaN₃, and dH₂O 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.

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 lion samples tested positive for ZcAV by both ELISA and PCR, and 48% of sea lion samples 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 115,26. 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
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 DNA11. 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 sampling 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 inoculation18, although once the antibodies are produced they persist in serum for long periods19–20. 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 variability9,31,32, which can pose problems for both DNA detection and antibody detection. For certain TTV genotypes, it has been shown that multiple serotypes may exist in circulation20. 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.

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).

<table>
<thead>
<tr>
<th>Lung PCR positive n (%)</th>
<th>Lung PCR negative n (%)</th>
<th>ELISA totals n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serum ELISA positive n (%)</td>
<td>Serum ELISA negative n (%)</td>
<td>PCR totals n (%)</td>
</tr>
<tr>
<td>Lung PCR positive n (%)</td>
<td>Lung PCR negative n (%)</td>
<td>ELISA totals n (%)</td>
</tr>
<tr>
<td>11 (11%)</td>
<td>22 (23%)</td>
<td>33 (34%)</td>
</tr>
<tr>
<td>17 (18%)</td>
<td>46 (48%)</td>
<td>63 (66%)</td>
</tr>
<tr>
<td>28 (29%)</td>
<td>68 (71%)</td>
<td>96 (100%)</td>
</tr>
</tbody>
</table>

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.
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).