 doi:10.20944/preprints201805.0184.v1

Communication

Virological sampling of inaccessible wildlife with drones

Jemma L. Geoghegan*[1,a], Vanessa Pirotta[1,a], Erin Harvey[2,a], Alastair Smith[3], Jan P. Buchmann[2], Martin Ostrowski[4], John-Sebastian Eden[2,5], Robert Harcourt[1], Edward C. Holmes[2]

[1] Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia; jemma.geoghegan@mq.edu.au; vanessa.pirotta@hdr.mq.edu.au; robert.harcourt@mq.edu.au
[2] Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Life and Environmental Sciences and Sydney Medical School, The University of Sydney, Sydney, NSW 2006, Australia; ehar6487@uni.sydney.edu.au; jan.buchmann@sydney.edu.au; js.eden@sydney.edu.au; edward.holmes@sydney.edu.au
[3] Heliguy Scientific Pty Ltd, Sydney, NSW 2204, Australia; alastair@heliguy.tv
[4] Department of Molecular Sciences, Macquarie University, Sydney, NSW 2109, Australia; martin.ostrowski@mq.edu.au
[5] Westmead Institute for Medical Research, Centre for Virus Research, Westmead, NSW 2145, Australia; js.eden@sydney.edu.au
[a] Authors contributed equally.
* Correspondence: jemma.geoghegan@mq.edu.au; Tel.: +61-2-9850-8204

Abstract: There is growing interest in characterizing the viromes of diverse mammalian species, particularly in the context of disease emergence. However, little is known about virome diversity in aquatic mammals, in part due to difficulties in sampling. We characterized the respiratory virome of the Eastern Australian humpback whale (*Megaptera novaeangliae*). To achieve an unbiased survey of virome diversity a meta-transcriptomic analysis was performed on 19 pooled whale blow samples collected via a purpose-built Unmanned Aerial Vehicle (UAV, or drone) approximately 3km off the coast of Sydney, Australia during the 2017 winter annual northward migration from Antarctica to northern Australia. Despite the relatively small number of animals surveyed, we identified six novel virus species from five viral families. This work demonstrates the potential of UAVs in studies of virus disease, diversity, and evolution.

Keywords: whale; virome; drone; mammalian host; virosphere.

There is a growing interest in understanding the diversity, evolution and disease associations of viruses in natural populations [1]. Although sampling of many terrestrial species is relatively straightforward, there may be serious logistical challenges for animals that live in inaccessible habitats. Marine environments are one such habitat [2-4]. It has recently been shown that wild populations can be sampled using Unmanned Aerial Vehicles (UAVs) [5, 6]. UAVs are rapidly transforming wildlife science, allowing sampling from dangerous and inaccessible environments to address questions previously only approached by theory. Here, for the first time to our knowledge, we use UAVs to sample viruses, which may ultimately enable a better understanding of the patterns and drivers of disease emergence in wild populations.

There is evidence that marine mammal health is deteriorating as anthropogenic stressors on the world's oceans increase [9]. However, contemporary assessments of marine mammal health are strongly biased towards animals whose health is already compromised, such as stranded animals, which in part reflects the difficulties in sampling aquatic environments. Sampling from free-ranging marine mammals is therefore critical to assess whether healthy animal populations are potential reservoirs of viruses and other transmittable agents.

 doi:10.20944/preprints201805.0184.v1

Utilizing UAV technology for sampling, we used a meta-transcriptomic approach [7, 8] to characterize the respiratory virome of an important marine mammal, the Eastern Australian humpback whale (*Megaptera novaeangliae*), which serves as a model for work in this area. Recent analyses of whale breath, or 'blow', have revealed an extraordinary diversity and abundance of microbiota. Importantly, the microbial communities observed were divergent from those present in the surrounding seawater such that they could be considered as distinctly whale blow-associated [5, 6]. To date, however, these studies have not included virus sampling, and little is known about the diversity of the whale virome and whether this differs fundamentally from that seen in terrestrial mammals.

We collected whale blow samples from 19 humpbacks during the 2017 annual northward migration from Antarctica to northern Australia (Figure 1a; see Video S1). RNA was extracted using RNeasy Plus Universal mini kit (Qiagen). Due to low RNA concentration, all 19 samples were pooled and concentrated using a NucleoSpin RNA Clean-up XS kit (Macherey-Nagel). A single library was produced for RNA sequencing using the Low Input SMARTer Stranded Total RNA Sample Prep Kit with Mammalian rRNA depletion (Clontech), with 1ng of the pooled whale blow RNA as input. Paired-end (100 bp) sequencing of the RNA library was performed on the HiSeq 2500 platform (Illumina) at the Australian Genome Research Facility.

RNA sequencing of the rRNA-depleted library resulted in 19,389,378 paired reads (100 nt in length) that were assembled *de novo* into 107,681 contigs. Sequencing reads were first quality trimmed then assembled using Trinity [10]. The assembled transcriptome was annotated based on similarity searches against the NCBI nucleotide (nt) and non-redundant protein (nr) databases using BLASTn and Diamond (BLASTX) [11], and an e-value threshold of $1x10^{-5}$. Transcript abundance was estimated using RSEM [12] implemented within Trinity.

Our transcriptome data revealed that the humpback whale respiratory virome contains a wide diversity of DNA and RNA viruses. BLAST analysis revealed the relative abundance of taxonomic classes present in the non-rRNA transcriptome data, of which bacteria occupied ~45%, while ciliates were the second-most abundant source at ~29%. Baleen whale species contributed 0.9% of the transcriptome data and were the most abundant source of mammalian RNA. Viruses occupied just ~0.01% of the non-rRNA transcriptome. Despite this relatively low abundance, the viral contigs observed fell into 42 classified viral families, including 29 families of bacteriophage (Figure 1b). Among the most abundant viral families that are known to infect eukaryotes were small single-stranded (ss) DNA viruses, specifically the *Circoviridae* (and *Circoviridae*-like viruses) (6.5% of all viruses), as well as members of the *Parvoviridae* (2.4%) and an RNA virus family, the *Tombusviridae* (0.9%).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 May 2018 doi:10.20944/preprints201805.0184.v1

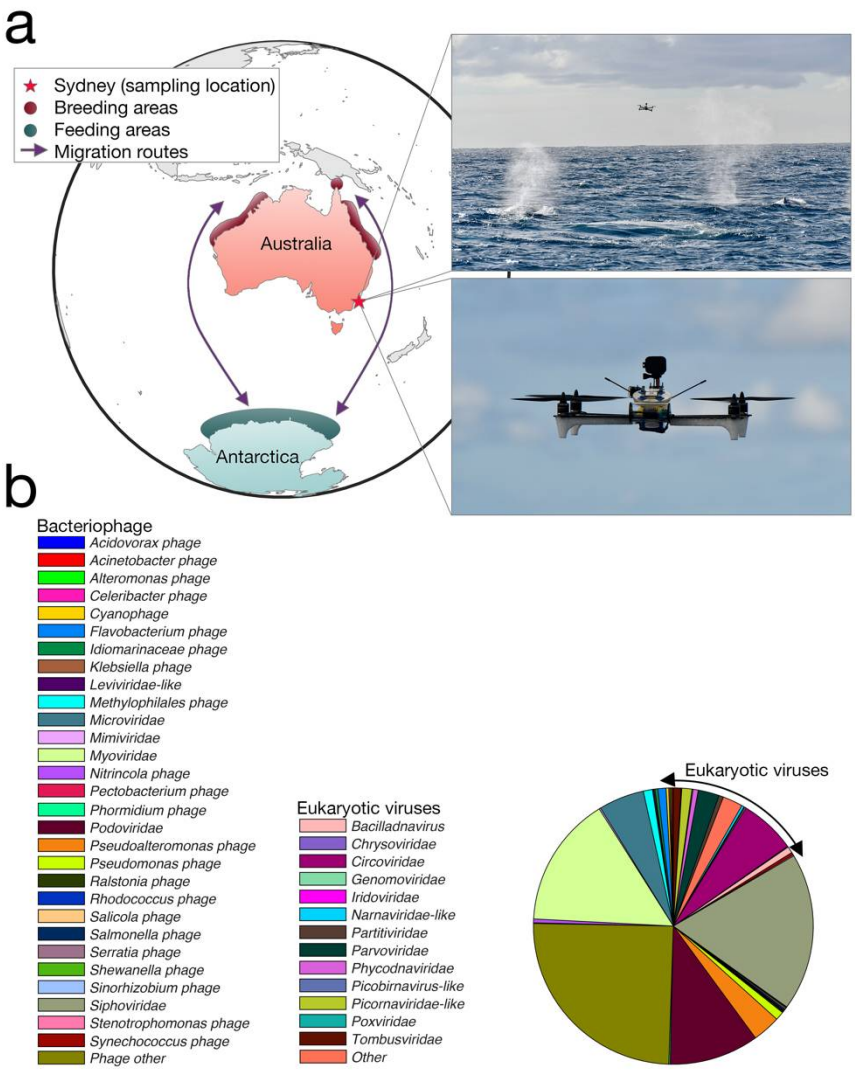


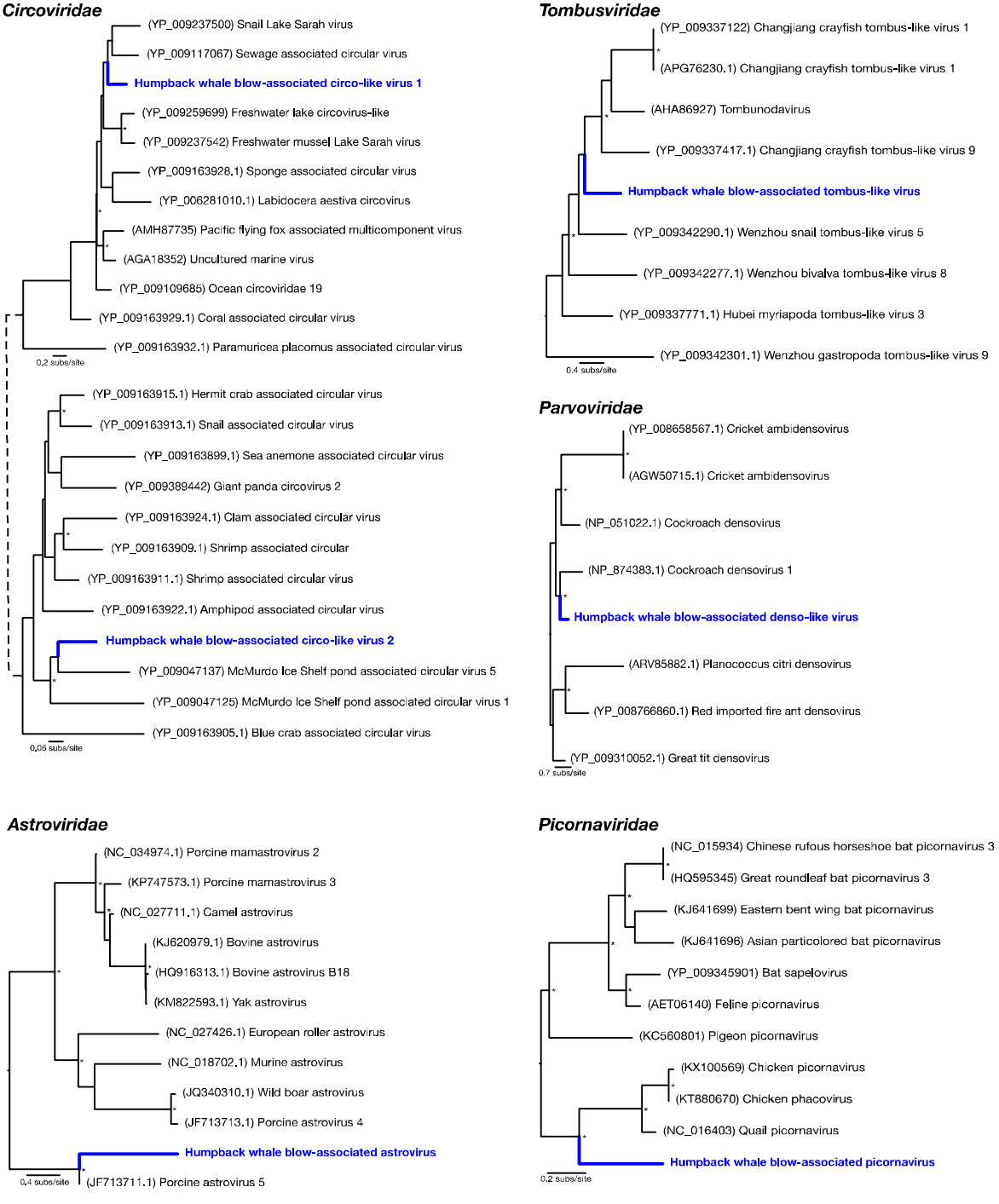
Figure 1. (a) Map showing the humpback whale sampling location (red star), approximately 3km off the coast of Sydney, New South Wales, Australia. Purple arrows indicate the typical seasonal migratory routes of the humpback whale from their likely feeding ground in Antarctica (dark green) to their breeding areas around northern Australia (dark red). Photographs demonstrate the UAV in action. (b) Relative abundance of viruses and their taxonomic families. Taxonomy was based on both protein and nucleotide BLAST search results, taking the best e-value for each (for those with identical e-values we used the taxa with the closest percentage identity). This included 44 viral families, including 30 families of bacteriophage.

We inferred the evolutionary (phylogenetic) relationships of the viruses contained in whale blow with their closest genetic relatives. Of the most abundant viruses, two novel (as determined by phylogenetic analysis) circular Rep-encoding ssDNA viruses (CRESS-DNA viruses) *Circoviridae*-like viruses were identified, denoted here as humpback whale blow-associated circo-like virus 1 and 2 (Table 1; Figure 2). Related viruses have previously been identified in many aquatic systems where marine invertebrates are thought to be a primary host [13]. Humpback whale blow-associated circo-like virus-1 shared 51% amino acid identity to replication-associated protein (Rep) of its closest genetic relative, sewage-associated circular DNA virus-29, and 46% amino acid identity to the Rep of Lake Sarah-associated circular virus-32. Humpback whale blow-associated circo-like virus-2 shared

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 May 2018 doi:10.20944/preprints201805.0184.v1

46% amino acid identity to the Rep of McMurdo Ice Shelf virus-5, isolated from a freshwater pond in Antarctica [14]. As these ssDNA viruses appear to be major virome components in many aquatic environments [13], they are likely associated with aquatic ecosystems in general and are not specifically whale-associated.

Another relatively abundant viral contig was a partial genome of a novel densovirus (family *Parvoviridae*). The most similar amino acid sequence to this new virus, denoted here as humpback whale blow-associated denso-like virus, was a densovirus isolated from a *Periplaneta fuliginosa* (i.e. a cockroach), sharing only 47% sequence similarity to the non-structural protein (Table 1; Figure 2). Similarly, a novel tombus-like viral partial genome, falling into the *Tombusviridae*, was identified and was closely related to Changjiang tombus-like virus-9 isolated from crayfish, with 41% sequence similarity to the RdRp. We denote this virus humpback whale blow-associated tombus-like virus (Table 1; Figure 2).



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 May 2018 doi:10.20944/preprints201805.0184.v1

Figure 2. Phylogenetic relationships, inferred using PhyML v3.1, of the viruses discovered from assembled contigs along with their closest genetic relatives obtained from GenBank. Sequences were aligned using MAFFT v.3.4, employing the E-INS-I algorithm and poorly aligned regions then removed using trimAl v.1.2. We selected the optimal amino acid substitution model identified using the Bayesian Information Criterion as implemented in Modelgenerator v0.85. Families described here are: *Circoviridae*, *Parvoviridae*, *Tombusviridae*, *Picornaviridae* and *Astroviridae*. The maximum likelihood phylogenetic trees show the topological position of the newly discovered viruses (blue). Asterisks indicate branch support >70%, based on 1000 bootstrap replicates. All branches are scaled per the number of amino acid substitutions per site. Trees were mid-point rooted for clarity only.

Virus family	**Virus species**	**Contig length (nt)**	**% Relative abundance in library**	**% Amino acid identify**	**Closest match (GenBank accession number)**
Circoviridae	Humpback whale blow-associated circo-like virus 1	702	0.000115%	51%	Sewage-associated circular DNA virus-29 (YP_009117067)
Circoviridae	Humpback whale blow-associated circo-like virus 2	909	0.000164%	46%	McMurdo Ice Shelf pond-associated circular DNA virus-5 (YP_009047137)
Parvoviridae	Humpback whale blow-associated denso-like virus	315	0.000143%	47%	*Periplaneta fuliginosa* densovirus (NP_051022.1)
Tombusviridae	Humpback whale blow-associated tombus-like virus	279	0.000164%	41%	Changjiang tombus-like virus-9 (YP_009337417.1)
Picornaviridae	Humpback whale blow-associated picornavirus	255	(N/A – assembled contigs from raw reads)	61%	Quail picornavirus (NC_016403)
Astroviridae	Humpback whale blow-associated astrovirus	130	(N/A – assembled contigs from raw reads)	76%	Porcine astrovirus 5 (YP_009010969)

Table 1. Amino acid identity, contig length and relative frequency of the viruses identified in this study.

To reveal viruses at very low abundance a Diamond BLAST (11) analysis was performed against the raw 100bp sequencing reads. This process identified several raw sequencing reads, later assembled into small contigs, including two potentially new RNA viruses that fell into two virus families: the *Picornaviridae* and the *Astroviridae*. Humpback whale blow-associated picornavirus shared 61% amino acid similarity to the RdRp of the most closely related *Coturnix coturnix* (quail) picornavirus (Table 1; Figure 2). Similarly, humpback whale blow-associated astrovirus shared 76%

amino acid identity with the non-structural protein 1a of porcine astrovirus-5 (Figure 2). Both picornaviruses and astroviruses are single-stranded, positive-sense RNA viruses with small icosahedral capsids and no external envelope, which may aid their preservation in harsh environments. As only short fragments of these viruses genomes were identified in our data, set their phylogenetic position cannot be confirmed. This is likely due to the low quantity of RNA isolated from the whale blow samples and the pooling of individual samples that adds complexity to the transcriptome.

Little is known about the transmission of whale viruses. Analyses of whale influenza viruses suggest that they likely originated from gulls and that feeding activities of gulls and whales often place them in close contact, such that oral-fecal transmission through seawater is a likely route [15] and which might explain our observation of viruses associated with aquatic ecosystems. In addition, given the vast aerosol produced by whales, and their close contact within migrating pods as well as at feeding and breeding grounds, respiratory transmission may also play an important role in the movement of viruses in whales.

In sum, this study shows that drone-based virological surveys of previously inaccessible wildlife populations has the potential to help reveal the diversity of the virosphere, facilitating the detection of viruses infecting wildlife and aiding evaluation of their pathogenic and zoonotic potential.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Video S1: GoPro footage from UAV demonstrates whale blow sampling.

Author Contributions: Conceptualization, J.L.G, V.P., R.H, E.C.H.; Methodology, J.L.G, V.P., E.H., A.S., R.H, E.C.H; Validation, J.L.G., J.B., J.S.E, E.H., E.C.H.; Formal Analysis, J.L.G., E.H., J.S.E, J.B.; Investigation, J.L.G, V.P., E.H., A.S., M.O., R.H, E.C.H; Resources, V.P., A.S., R.H, E.C.H.; Writing-Original Draft Preparation, J.L.G, V.P., R.H, E.C.H.; Writing-Review & Editing, J.L.G, V.P., E.H., M.O., R.H, E.C.H; Visualization, J.L.G; Supervision, J.L.G., R.H., E.C.H.; Project Administration, J.L.G.; Funding Acquisition, R.H., E.C.H, A.S.

Funding: VP is supported by a Macquarie Research Excellence scholarship. ECH is funded by an ARC Australian Laureate Fellowship (FL170100022).

Acknowledgments: This research was conducted under the Macquarie University Animal Ethics Committee Animal Research Authority 2016/010 and the NSW Office of Environment and Heritage Scientific Research permit SL101743. We thank Heliguy and Guy Alexander for providing drone manufacturing and testing support, CASA certification and operational insurances, and NSW Maritime and Oliver Masens for their support and use of vessels for sampling. We also thank the Cape Solander Whale Migration Study, in particularly Wayne Reynolds, Sue Rennie Wright and Mark Mcgeachie, and Sally Browning and Air Pig Productions for their donation of drone batteries.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Geoghegan JL, Holmes EC. Predicting virus emergence amid evolutionary noise. *Open Biol*. **2017**, 7, 170189.
2. Suttle CA. Viruses in the sea. *Nature* **2005**, 437, 356-61.
3. Culley AI, Lang AS, Suttle CA. Metagenomic analysis of coastal RNA virus communities. *Science* **2006**, 312, 1795-8.
4. Bogomolni AL, Gast RJ, Ellis JC, Dennett M, Pugliares KR, Lentell BJ, Moore MJ. Victims or vectors: a survey of marine vertebrate zoonoses from coastal waters of the Northwest Atlantic. *Dis Aquat Organ*. **2008**, 81, 13-38.
5. Pirotta V, Smith A, Ostrowski M, Russell D, Jonsen ID, Grech A, Harcourt R. An economical custom-built drone for assessing whale health. *Frontiers Marine Sci*. **2017**, 4, 425.
6. Apprill A, Miller CA, Moore MJ, Durban JW, Fearnbach H, Barrett-Lennard LG. Extensive core microbiome in drone-captured whale blow supports a framework for health monitoring. *mSystems* **2017**, 2, e00119-17.
7. Li C-X, Shi M, Tian J-H, Lin X-D, Kang Y-J, et al. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *eLife* **2015**, 4:e05378.
8. Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, Qin XC, Li J, Cao JP, Eden JS, Buchmann J, Wang W, Xu J, Holmes EC, Zhang YZ. Redefining the invertebrate RNA virosphere. *Nature* **2016**, 540, 539-43.
9. Gulland FMD, Hall AJ. Is marine mammal health deteriorating? Trends in the global reporting of marine mammal disease. *EcoHealth* **2007**, 4, 135-50.
10. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, et al. *De novo* transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nature Protocol*. **2013**, 8, 1494–1512.
11. Buchfink B, Xie C, Huson DH.. Fast and sensitive protein alignment using DIAMOND. *Nature Meth*. **2014**, 12, 59.
12. Li B, Dewey CN.. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **2011**, 12, 323.
13. Rosario K, Schenck RO, Harbeitner RC, Lawler SN, Breitbart M. Novel circular single-stranded DNA viruses identified in marine invertebrates reveal high sequence diversity and consistent predicted intrinsic disorder patterns within putative structural proteins. *Frontiers Microbiol*. **2015**, 6, 696.
14. Zawar-Reza P, Arguello-Astorga GR, Kraberger S, Julian L, Stainton D, Broady PA, Varsani A.. Diverse small circular single-stranded DNA viruses identified in a freshwater pond on the McMurdo Ice Shelf (Antarctica). *Infect Genet Evol* **2014**, 26, 132-8.
15. Hinshaw VS, Bean WJ, Geraci J, Fiorelli P, Early G, Webster RG. Characterization of two influenza A viruses from a pilot whale. *J Virol*. **1986**, 58, 655-6.