**SUPPORTING INFORMATION**

**Longitudinal analysis of pinnipeds in the Northwest Atlantic provides insights on endemic circulation of Phocine distemper virus**

Wendy Puryear, Kaitlin Sawatzki, Andrea Bogomolni, Nichola Hill, Alexa Foss, Iben Stokholm, Morten Tange Olsen, Ole Nielsen, Thomas Waltzek, Tracey Goldstein, Kuttichantran Subramaniam, Thais Carneiro Santos Rodrigues, Manjunatha Belaganahalli, Lynda Doughty, Lisa Becker, Ashley Stokes, Misty Niemeyer, Allison Tuttle, Tracy Romano, Mainity Batista Linhares, Deborah Fauquier, Jonathan Runstadler

Supplemental Methods

**Sequence generation**

## Complete PDV genomes

**US/Hg/IFAW18-001/2018 and US/Pv/NEAQ18-044/2018** (GenBank: MW642077-MW642078): Libraries were generated with the Kapa HyperPrep Library kit as per manufacturer protocol (Roche, Basel, CH). Sequencing was performed on Illumina MiSeq using 250bp PE chemistry at University of California Davis Genome Centre. Geneious R11 software was used to assemble contigs and unique singletons with US/Pv/2006 as a reference.

**CA/Pg/WVL181140/2017** (GenBank: MW504062): Libraries were generated with the NEBNext Ultra II RNA Library Prep kit as per manufacturer protocol (New England Biolabs, Ipswich, MA). Sequencing was performed on Illumina HiSeq using 150bp PE chemistry at the University of Florida Interdisciplinary Center for Biotechnology Research. CLC Genomic Workbench v12.0 was used to filter low-quality reads and quality trimming, prior to assembly with NL/Pv/Wad/1988 as a reference.

**US/Pv/MME-287/2018 and US/Pv/MME-343/2018** (GenBank: MW581015-MW581016): Libraries were generated with the NEBNext Ultra II RNA Library Prep kit as per manufacturer protocol. Sequencing was performed on Illumina HiSeq using 125bp PE chemistry at the Icahn School of Medicine at Mt. Sinai Genetics and Genomics Core. Reads were processed with Trimmomatic (v0.27) and Ngs crumbs and assembled with IDBA-hybrid (v1.1.3) using NL/Pv/Wad/1988 as a reference.

**US/Pv/MME-430/2018** (GenBank accession: MW581017): NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs) was used to select mRNA and cDNA generated using NEBNext Ultra II RNA Library Prep kit. Libraries were generated with Oxford Nanopore Technologies (ONT) (Oxford, UK) 1D-ligation kit (SQK-LSK109) with the PCR barcoding expansion kit (EXP-PBC001). Sequencing was performed on ONT MinION with basecalling in MinKNOW (v2.0). Reads were processed with Porechop (v0.2.4) and FiltLong (v0.2.0) and assembled using the ONT pomoxis pipeline (v0.2.0) using NL/Pv/Wad/1988 as a reference.

**Hemagglutinin genes**

RT-PCR was performed using SuperScript III Platinum One-Step qRT-PCR Kit (ThermoFisher) and forward (5’- GGGCCCAGGTAGTTCAAC-3’) and reverse (5’- CTCAACCTCAGTGGGTAC-3’) oligonucleotides as per manufacturer instructions. Libraries were generated with the ONT 1D-ligation kit (SQK-LSK109) with native barcoding expansion kit (EXP-NBD104) and sequenced and analyzed on ONT MinION as with US/Pv/MME-420/2018. GenBank accession numbers MW581018- MW581026.

**Time-scaled phylogenetic model and clock analysis**

We used the full N=61 dataset inclusive of all H gene sequences to investigate the temporal signal of the H gene using TempEst v.1.5.3, and determined that PDV evolves according to a molecular clock (diverge over time, R-squared: 0.86, **Figure S1**) [1]. Time-scaled phylogenetic models were tested using BEAST (v1.10.4) [2]. Six parameterizations were compared using path and stepping-stone sampling (**Table S2**) [3]. Using Bayes factors to guide model selection, an uncorrelated relaxed clock and HKY γ substitution model was selected with Coalescent: constant growth prior. Input and output files are available in the electronic supplementary data.

**Maximum-likelihood phylogenetic analysis.** RAxML was used to generate maximum-likelihood trees for comparative analysis of mammalian *Morbillivirus* host receptors (CD150, N=117 and Nectin-4, N=124). Representative mammalian amino acid sequences were accessed via NCBI Orthologs, aligned using ClustalΩ and phylogenetic trees were generated using RAxML under the JTT+γ model with bootstrapping (x 5000).

Electronic supplementary data including sequence alignments, phylogenetic output and assay output can be found at https://github.com/ksawatzki/Supp\_data/

Supplemental References

1. Rambaut A., Lam T.T., Carvalho L.M., Pybus O.G. 2016 Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* **2**(1). (doi:10.1093/ve/vew007).

2. Drummond A.J., Rambaut A. 2007 BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **7**, 214. (doi:10.1186/1471-2148-7-214).

3. Baele G., Lemey P., Bedford T., Rambaut A., Suchard M.A., Alekseyenko A.V. 2012 Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol* **29**(9), 2157-2167. (doi:10.1093/molbev/mss084).

Supplemental Table 1

Number of animals sampled each year, reported for sample type and species.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **2010** | **2011** | **2012** | **2013** | **2014** | **2015** | **2016** | **2017** | **2018** | **2019** | **2020** | **Total** |
| **RT-PCR Swabs** |  |  |  |  |  |  |  |  |  |  |  |  |
| Wild Grey Seal Pups | ND | ND | ND | ND | ND | ND | 127 | 121 | 96 | 65 | 83 | 492 |
|   |  |  |  |  |  |  |   |   |   |   |   |   |
| **RT-PCR Swabs** |  |  |  |  |  |  |  |  |  |  |  |  |
| Stranded Harbor Seals | ND | ND | ND | ND | ND | ND | ND | ND | 102 | 96 | 49 | 247 |
| Stranded Grey Seals | ND | ND | ND | ND | ND | ND | ND | ND | 39 | 51 | 35 | 125 |
| Stranded Harp Seals | ND | ND | ND | ND | ND | ND | ND | ND | 8 | 70 | 14 | 92 |
|   |  |  |  |  |  |  |   |   |   |   |   |   |
| **RT-PCR Tissues** |  |  |  |  |  |  |  |  |  |  |  |  |
| Stranded Harbor Seals | ND | 1 | 2 | 2 | 0 | 2 | ND | ND | 5 | ND | ND | 12 |
| Stranded Grey Seals | ND | 1 | 2 | 1 | 1 | 3 | ND | ND | 1 | ND | ND | 9 |
|   |  |  |  |  |  |  |   |   |   |   |   |   |
| **Serum for Serology** |  |  |  |  |  |  |  |  |  |  |  |  |
| Wild Grey Seals | 0 | 0 | 0 | 8 | ND | ND | ND | ND | ND | ND | ND | 8 |
| Stranded Harbor Seals | 76 | 69 | 65 | 6 | ND | ND | ND | ND | ND | ND | ND | 216 |
| Stranded Grey Seals | 3 | 9 | 5 | 12 | ND | ND | ND | ND | ND | ND | ND | 29 |
| Stranded Harp Seals | 15 | 12 | 0 | 3 | ND | ND | ND | ND | ND | ND | ND | 30 |

 ND = not done

Supplemental Table 2

Tree and clock model fit comparison using BEAST

|  |  |  |  |
| --- | --- | --- | --- |
| Models | Clocks | PS\* | SS\* |
| Constant size | Strict | -24128.1 | -24128.1 |
| Exponential growth | Strict | -24130.4 | -24130.2 |
| Bayesian skyride | Strict | -24141.8 | -24141.8 |
| Constant size | Relaxed | -24123.9 | -24123.8 |
| Exponential growth | Relaxed | -24131.6 | -24132 |
| Bayesian skyride | Relaxed | -24140.3 | -24140.3 |

\* Path sampling (PS) and stepping-stone sampling (SS) log marginal likelihood. Higher values indicate a better model to describe the data.

Supplemental Table 3

PDV RT-PCR results for 15 stranded animals from 2011-2015.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Animal ID | Species | Brain | Kidney | Liver | Lung | Spleen |
| Pv/IFAW303/2011 | Harbor | + | + | - | - | - |
| Pv/IFAW258/2012 | Harbor | + | - | + | ND | - |
| Pv/IFAW278/2012 | Harbor | + | - | - | + | - |
| Pv/IFAW009/2013 | Harbor |  +\* | - | - | - | - |
| Pv/IFAW177/2013 | Harbor | - | - | - | + | - |
| Pv/IFAW227/2015 | Harbor | + | - | - | - | - |
| Pv/NMLC025/2015 | Harbor | + | ND | ND | ND | ND |
| Hg/IFAW237/2011 | Grey |  +\* | + | - | - | - |
| Hg/IFAW265/2012 | Grey | + | - | + | - | + |
| Hg/IFAW267/2012 | Grey |  +\* | - | - | - | - |
| Hg/IFAW049/2013 | Grey | + | - | + | - | - |
| Hg/IFAW140/2014 | Grey |  +\* | - | - | - | - |
| Hg/IFAW087/2015 | Grey | + | - | - | - | - |
| Hg/IFAW098/2015 | Grey | + | - | - | + | - |
| Hg/IFAW145/2015 | Grey |  +\* | + | - | - | - |

Asterisks indicate samples where H gene was sequenced. ND is not done due to tissue/RNA unavailability.

Supplemental Table 4

Nucleotide identity of newly sequenced Phocine Distemper Virus full genomes to those already published. New sequences denoted with asterisks.

|  |  |  |
| --- | --- | --- |
|   | Wad/88(western Europe) | US/06(eastern N.Am) |
| NL/Pv/Wad/1988 | 100.0000 | 99.3126 |
| US/Pv/2006 | 99.3126 | 100.0000 |
| \*CA/Pg/WVL18-1140/2017 | 99.0115 | 99.1270 |
| \*US/Pv/MME-287/2018 | 99.0364 | 99.1520 |
| \*US/Pv/MME-343/2018 | 98.9207 | 99.8330 |
| \*US/Pv/MME-430/2018 | 99.0364 | 99.1392 |
| \*US/Hg/IFAW18-001/2018 | 99.0364 | 99.1649 |
| \*US/Pv/NEAQ-044/2018 | 99.0299 | 99.1456 |



Supplemental Figure 1. Root-to-tip divergence analysis to evaluate the temporal signal of the hemagglutinin gene of PDV. Clock-like evolution of the 61 taxa was indicated by the strong linear correlation between divergence and time (R-squared = 0.8599, correlation co-efficient = 0.9273) using TempEst v 1.5.3.



Supplemental Figure 2. PDV Hemagglutinin receptor binding site (RBS) domains are identical between virus sequenced from different seals. This alignment illustrates amino acids, where the top sequence, NL/Pv/NC028249/1988 is a reference and a dot indicates identity to the reference. Background color denotes virus derived from harbor (white), grey (light grey) or harp (dark grey) seals. New sequences denoted with asterisks.



Supplemental Figure 3. Comparative phylogenetics of CD150/SLAMF1 among mammals. CD150 is the probable primary host receptor for PDV entry and infection. One hundred-seventeen full length CD150 coding region sequences representing a diverse range of mammals were compared using a maximum-likelihood approach in RAxML. Grey and Harbor seal are marked in blue, additional marine mammals are marked in orange. The scale bar corresponds to substitution rate.



Supplemental Figure 4. Comparative phylogenetics of Nectin-4/PVRL4 among mammals. Nectin-4 is a probable secondary host receptor for PDV entry and infection. One hundred-twenty-four full length Nectin-4 coding region sequences representing a diverse range of mammals were compared using a maximum-likelihood approach in RAxML. Grey and Harbor seal are marked in blue, additional marine mammals are marked in orange. The scale bar corresponds to substitution rate.



Supplemental Figure 5. Maximum clade credibility phylogenetic tree using only Phosphoprotein-Matrix-Fusion-Hemagglutinin genes.