TITLE

Phylogenomic insights to the origin and spread of Phocine Distemper Virus in European harbour seals in 1988 and 2002

AUTHORS

Iben Stokholm[1*], Tero Härkönen[2], Karin Harding[3], Ursula Siebert[4], Kristina Lehnert[4], Rune Dietz[5], Jonas Teilmann[5], Anders Galatius[5], Linnea Worsøe Havmøller[5], Emma Carroll[6], Ailsa Hall[6], and Morten Tange Olsen[1*]

AFFILIATIONS

[1]Section for Evolutionary Genomics, Natural History Museum of Denmark, University of Copenhagen, DK-1350 Copenhagen K, Denmark

[2]Department of Environmental Research and Monitoring, Swedish Museum of Natural History, Box 50007, S-10405 Stockholm, Sweden

[3]Department of Biological and Environmental Sciences, University of Gothenburg, Box 463 SE-405 30, Sweden

[4]Institute for Terrestrial and Aquatic Wildlife Research, University of Veterinary Medicine Hannover, 25761 Büsum, Germany,

[5]Marine Mammal Research, Department of Bioscience, Aarhus University, DK-4000 Roskilde, Denmark

[6]Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews, St Andrews, Fife, KY16 8LB

CORRESPONDING AUTHORS

*Iben Stokholm (ibenstokholm@gmail.com)

*Morten Tange Olsen (morten.olsen@snm.ku.dk)

RUNNING HEAD

Phocine Distemper Virus in harbour seals

ABSTRACT

The 1988 and 2002 Phocine Distemper Virus (PDV) outbreaks in European harbour seals (*Phoca vitulina*) are among the largest mass mortality events recorded in marine mammals. Despite its large impact on harbour seal population numbers, and three decades of studies, many questions regarding the spread and temporal origin of PDV remain unanswered. Here, we sequenced and analysed 7,123 base pairs of the PDV genome, including the coding and non-coding regions of the entire P, M, F and H genes in tissues from 44 harbour seals to shed new light on the origin and spread of PDV in 1988 and 2002. The phylogenetic analyses trace the origin of the PDV strain causing the 1988 outbreak to between May 1987 and April 1988, while the origin of the strain causing the 2002 outbreak can be traced back to between June 2001 and May 2002. The analyses further point to several independent introductions of PDV in 1988, possibly linked to a southward mass immigration of harp seals in the winter and spring of 1987-1988. The vector for the 2002 outbreak is unknown, but the epidemiological analyses suggest that the subsequent spread of PDV from the epicentre in Kattegat, Denmark, to haul-out sites in the North Sea through several independent introductions.

KEYWORDS

Morbillivirus; epidemic; viral phylogeny; virus evolution, wildlife pathogen

INTRODUCTION

Over the past few decades, several epidemics have been recorded in marine mammal populations around the globe (Jo *et al.,* 2018; Duignan *et al.,* 2014). The pathogens responsible for these epidemics have mainly been identified as Influenza A (Krog *et al.,* 2015) or viruses belonging to the morbillivirus genus such as various cetacean morbilliviruses (Kennedy *et al.,* 1989; Domingo *et al.,* 1990; Aguilar & Antonio Raga 1993; Birkun *et al.,* 1999; Lipscomb *et al.,* 1996), Canine Distemper Virus (CDV) (Mamaev *et al.,* 1995) and Phocine Distemper Virus (PDV) (Dietz *et al*., 1989a: Curran *et al.,* 1990; Harding *et al.,* 2002; Härkönen *et al.,* 2006). The morbillivirus genus is a group of highly infectious viruses from the *Paramyxoviridae* family. The genus also includes the Measles Virus (MeV) and the now eradicated Rinderpest (RPV) that have caused severe disease outbreaks among humans and cattle respectively, resulting in suffering and economic losses in affected populations (Furuse *et al.,* 2010; Roeder *et al.,* 2013). Morbillivirus infection occurs through the respiratory route where the virus replicates in macrophages and dendritic cells (Rijks *et al.,* 2012). Members of the genus have frequently caused epidemic disease in previously unexposed populations, and during the past century interspecies infections have been reported on several occasions (Mamaev *et al.*, 1995; Goldstein *et al.*, 2009; Roelke-Parker *et al.*, 1996). Thus, new zoonotic diseases could potentially emerge from wildlife reservoirs of morbilliviruses.

The PDV epidemics in 1988 and 2002 are among the largest mass mortality events ever recorded in marine mammals, resulting in the death of over 23,000 harbour seals in 1988 and approximately 30,000 harbour seals in 2002 (Dietz *et al*., 1989a: Harding *et al.,* 2002; Härkönen *et al.,* 2006). The two outbreaks share similarities with respect to the timing, duration and location of the first cases, occurring early in spring on the island of Anholt in central Kattegat, Denmark. From there, the virus seems to have spread to harbour seal colonies across northern Europe both to neighbouring haul-out

sites and via large jumps in location (Härkönen *et al.,* 2006). The sudden appearance and spread of the previously undescribed morbillivirus among European harbour seals in 1988 immediately sparked investigations into the virus' origin and mechanisms of transmission and spread. Antibodies against morbillivirus have subsequently been found in samples collected from harp seals (*Pagophilus groenlandicus*) and ringed seals (*Pusa hispida*) from Greenland and Svalbard waters prior to the European outbreak in 1988, suggesting that the virus is circulating in Arctic seals (Dietz *et al.,* 1989b). In contrast, the virus seems not to be circulating in European harbour seals. PDV antibodies were not detected in harbour seals in these waters prior to the 1988 outbreak, and the proportion of the European population carrying antibodies declined steadily after each outbreak, dropping from more than 50% in 2003 to less than 10% in 2007 (Bodewes *et al.,* 2013; Ludes-Wehrmeister *et al.,* 2016; Pomeroy *et al.,* 2005). It has been hypothesized that the unusual mass migration of harp seals in the winter and spring 1987/1988 lead to the introduction of the virus to the European harbour seals in 1988 (Haug & Nilssen, 1995; Nilssen *et al.,* 1998), but the exact timing of these events have never been investigated, and a vector for the 2002 epidemic has not been proposed. Moreover, while the long-distance transmission events of PDV among harbour seal colonies was initially suggested to be aided by the less susceptible and more mobile grey seal (*Haliocherus grypus*) (Härkönen *et al.,* 2006; McConnell *et al.,* 1999) more recent tagging data indicate that while harbour seals are, in general, philopatric, individual seals occasionally undertake long-distance movements (Dietz *et al*., 2013; Reijnders *et al.,* 2010; Tougaard *et al.* 2008; Aarts *et al.,* 2016). Therefore, dispersing harbour seals could have served as a vector for spreading the virus across Europe.

Thus, despite nearly three decades of research on PDV, a number of fundamental questions remain to be addressed regarding the origin and evolution of PDV, the diversity and relationship of

97 circulating PDV strains, the source and spread of PDV, and the mechanisms of virus transmission.
98 Here, we sequenced a continuous 7,123 bp sequence of the PDV genome, including the complete
99 coding and non-coding regions of the viral P, M, F and H genes directly from harbour seal samples
100 collected throughout the geographic and temporal progression of the 1988 and 2002 outbreaks to
101 provide the first detailed epidemiological assessment of the emergence and spread of PDV among
102 harbour seals in northern Europe.
103
104 **MATERIALS & METHODS**
105 *Sampling*
106 Tissue samples from harbour seals were collected at haul-out sites along the coasts of Denmark,
107 Sweden, Germany, the Netherlands, England and Scotland during the 1988 and 2002 outbreaks
108 (Supplementary Table 1). Unfortunately high quality samples (e.g. lung tissue stored in RNAlater at
109 -80°C) were only available animals that died in Sweden during the 2002 outbreak. Thus we tested
110 32 liver and muscle tissue samples collected at locations in Denmark and Germany in 1988, and 67
111 spleen, lung and muscle tissue samples from 2002, covering haul-out sites in most Northern Europe.
112 To increase the likelihood of detecting and sequencing the PDV virus, preference was given to
113 samples collected from carcasses that were fresh at the time of collection. Furthermore, we
114 attempted to increase the phylogenomic inference by selecting samples representing as broad a
115 geographical and temporal range as possible, prioritizing samples from early, mid and late stages of
116 each outbreak.
117
118 *Virus screening and sequencing*
119 Viral RNA was extracted from spleen, lung, liver or muscle tissue using the Thermo Scientific
120 KingFisher^{TM} Cell and Tissue DNA Kit. The extractions were first tested for the presence of viral

RNA by RT-PCR runs with the H5 (5'-AGATGATATCTTTCCTCC-3') and H6 (5'-ATCCATATGAGTTGCTCC-3') primers. Different sequencing methods were attempted to generate full genomes, but ultimately only primer walking followed by standard Sanger sequencing of four genes was possible. Primer pairs designed for sequencing the PDV genome (de Vries *et al.* 2013) were used to amplify and sequence 13 separate, but overlapping segments of 7,123 bp of the PDV genome, covering the complete sequence from position 1,744 to 8,915 (excluding stop codons), including the phosphoprotein (P), the unglycosylated matrix protein (M), the fusion protein (F), and the attachment protein hemagglutinin (H) (Supplementary Table 2 and Supplementary Figure 1). RT-PCRs were performed with the QIAGEN OneStep RT-PCR kit with cycling conditions of 30 min at 50 °C, 10 min at 95 °C, 40-38 cycles of 1 min at 95 °C, 1 min at 55-65 °C and 1 min at 72 °C (Supplementary Table 3) and a final step of elongation for 10 min at 72 °C followed by 4 °C. The total volume for each reaction was 25 µl and consisted of 5 µl QIAGEN OneStep RT-PCR Buffer, 1µl dNTP (10 mM), 1.5 µl of each primer (10 µM), 1 µl QIAGEN OneStep RT-PCR Enzyme Mix, 12.5 µl H_2O and 2.5 µl RNA template. PCR products were purified and sequenced by Sanger sequencing at Macrogen Europe (Amsterdam, the Netherlands).

Sequence preparation

The raw sequences were assembled and examined for errors in Geneious version 9.1.2 (Kearse *et al.*, 2012) (Supplementary Table 4). All ambiguous base calls and singletons (i.e. mutations only occurring in a single sample) were checked by re-sequencing. Low signal regions at the beginning and end of each sequence were removed, and two bases (TT) were deleted before the M gene (position 1,651-1,652) and the H gene (position 5,302-5,303) to place the coding regions in the correct reading frame. The resulting sequences were mapped to the PDV reference genome from 1988 (NCBI GenBank accession no. KC802221; de Vries *et al.*, 2013) using MUSCLE (Edgar,

2004) to obtain a continuous 7,123 bp sequence, spanning the coding and non-coding regions of the viral P, M, F and H genes. Three different alignments were generated; one including sequences from the 1988 outbreak (eight sequences), one including the 2002 outbreak sequences (36 sequences) and one of all sequences combined from the two outbreaks (44 sequences).

Substitution saturation and recombination analyses

The full data set was tested for the absence of substitution saturation using the entropy-based index in DAMBE (Xia et al. 2003; Xia, 2013) (Supplementary Table 5). The results showed no indication of substitution saturation among the sequences. Likewise, no evidence of recombination was found using GARD recombination analysis under the HKY85 nucleotide evolution model implemented in Datamonkey (www.datamonkey.org) (Hasegawa *et al.,* 1985; Pond *et al.*, 2006; Delport *et al.*, 2010).

Phylogenetic analyses

Phylogenetic trees and divergence times were estimated by using the Bayesian Markov chain Monte Carlo (MCMC) method in the BEAST 2.4.6 software package (Bouckaert *et al.,* 2014). Sequence alignments were partitioned into coding regions, 1st, 2nd and 3rd codon positions and non-coding intergenic regions to account for the expected variation in the substitution rates among in this region of the sequences (Lio & Goldman, 1998). The data were analysed using the HKY substitution model (Hasegawa *et al.,* 1985) determined as the best fit model based on the final weight of three criteria; AIC (Akaike, 1973), BIC (Akaike, 1973) and DT (Minin *et al.,* 2003) estimated in JModeltest (Posada, 2008) (Supplementary Table 6). The evolution of the partitions was described by the Gamma site model thereby allowing the substitution rates of each site to correspond to a continuous distribution of different rates among sites (Haug, 1990). Runs were carried out under the

169 assumption of a coalescent constant population growth model for the dataset, and a coalescent
170 exponential population growth model for the separate 1988 and 2002 datasets, respectively, using a
171 relaxed uncorrelated lognormal molecular clock in all analyses (Drummond *et al.,* 2006)
172 (Supplementary Table 7 and 8). The tree was calibrated by the inclusion of sampling dates (tip
173 dates), specified in years. Four independent runs were conducted with a chain length of 20,000,000
174 iterations, with sampling of the tree parameters for every 2,000th iteration and a burn-in of 10 %.
175 The results of each individual run were visually inspected in Tracer v1.6.0 for convergence and to
176 check that the effective sample size (ESS) values were > 200. Subsequently, the files produced from
177 independent runs of the MCMC yielding similar results were merged into one tree and log file using
178 LogCombiner v2.4.6 (Bouckaert *et al.,* 2014) with 50% sampling frequency and a burn-in of 10%.
179 The tree files were summarized in maximum clade credibility trees in TreeAnnotator and viewed in
180 Figtree v1.4.3 (Rambaut 2016). The time to the most recent common ancestor (TMRCA) of the
181 combined 1988 and 2002 dataset was estimated as a mean value with its 95% high-probability
182 density (HPD) interval. To account for uncertainties in branch lengths and topology the final tree
183 was constructed using Densitree (Bouckaert & Heled, 2014).
184
185 **RESULTS AND DISCUSSION**
186 *Sequencing PDV from the 1988 and 2002 outbreaks*
187 A total of 99 harbour seal samples were screened from the outbreaks in 1988 (n = 32) and 2002 (n =
188 67) from which 7,123 base pairs of the PDV genome (position 1,744-8,915, excluding stop codons)
189 was successfully amplified and sequenced from 44 samples; eight from 1988 and 36 from 2002
190 (Figure 1; Supplementary Table 1; Supplementary Table 4). In total 112 polymorphic sites were
191 found in the combined 1988 and 2002 dataset, 16 polymorphic sites (including 11 singletons)
192 occurred in the sequences from the 1988 samples, and 57 polymorphic sites (32 singletons)

193 occurred in the sequences from the 2002 samples (Supplementary Tables 9 and 10). The
194 amplification and sequencing success was higher for samples from 2002 than 1988, and within the
195 2002 samples success was higher for samples that had been stored in RNAlater at -80°C compared
196 to storage in ethanol or frozen dry at -20°C. Nevertheless, our results illustrate that PDV virus can
197 be extracted and sequenced from infected seals even after three decades of suboptimal sample
198 storage.
199
200 *A winter origin of the 1988 and 2002 PDV strains*
201 The origin of the virus sequenced from the 1988 epidemic was estimated to be mid-November
202 1987, with a 95% probability interval ranging from mid-May 1987 to early April 1988 (Table 1).
203 This estimate correlates well with the timing of observed mass migrations of harp seals from the
204 Barents Sea south along the coasts of Norway and into the North Sea in the winter of 1987-1988
205 (Haug & Nilssen, 1995). The migration event is widely believed to have introduced PDV to the
206 previously unexposed European harbour seals (Haug & Nilssen, 1995; Nilssen *et al.*, 1998; Dietz *et*
207 *al.*, 1989a,b; Härkönen *et al.,* 2006), and was presumably caused by the collapse of the Barents Sea
208 capelin (*Mallotus villosus*) and herring (*Clupea harengus*) stocks (Haug & Nilssen, 1995).
209 Although the morbillivirus itself has not been detected in Arctic seals (Kreutzer *et al.,* 2008)
210 morbillivirus specific antibodies have been found in Greenlandic harp seals and ringed seals
211 sampled between 1984 and 1987 (Dietz *et al.,* 1989b), as well as in Barents Sea harp seal samples
212 collected between 1987 and 1989 (Markussen & Have, 1992). Markussen and Have (1992)
213 documented that the percentage of harp seals with antibodies increased from 1987 to 1989,
214 indicating that a large proportion of the Barents Sea seals carried a morbillivirus before and during
215 the European epidemic in 1988. Harp seals are migratory and even discrete populations have
216 overlapping ranges during part of the year (Øien & Øritsland, 1995; Folkow *et al.,* 2004),

supporting the possibility of periodical encounters and intraspecific and interspecific disease transmissions across the Arctic.

The origin of the virus sequenced from the 2002 epidemic was estimated to be mid-January 2002, with a 95% probability interval ranging from late June 2001 to early May 2002 (Table 1). Thus, like the 1988 outbreak, the 2002 PDV strain can be traced back to the preceding winter. However, unlike the 1988 outbreak, there was no obvious candidate for the introduction of PDV in 2002. The phylogenetic analyses presented here strongly suggest that the 2002 strain did not originate directly from the 1988 strain (Figure 2B). This is also supported by previous analyses of shorter PDV sequence fragments (Nielsen *et al.,* 2009; Earle *et al.,* 2011), as well as an absence of virus in European harbour seals screened between the two outbreaks (Ludes-Wehrmeister *et al.,* 2016; Bodewes *et al.,* 2013). PDV neutralizing antibodies have been detected in harp seals, ringed seals, hooded seals (*Cystophora cristata*) and polar bears (*Ursus maritimus*) sampled between 1988 and 1996 in the waters surrounding Greenland, the Canadian Arctic and North America (Nielsen *et al.,* 1997; Cattet *et al.,* 2004). PDV positive harbour seals, grey seals and harp seals were also detected along the North Atlantic coast of North America between 2004-2007 (Siembieda *et al.,* 2017). Thus there are several potential candidate species that could have been reservoir hosts or carriers. Similar to 1988, the Barents Sea capelin and herring stock experienced declines in the years just before the 2002 epidemic (Loeng & Drinkwater 2007); however, this collapse did not result in a new mass migration of harp seals as occurred in 1998. Nevertheless, the virus could have been carried and transmitted by single migrating individuals as harp seals tagged in Canada were observed to visit the North Sea between 2001 and 2002 (Kreutzer *et al*., 2008).

241 *Multiple introductions of PDV in 1988*

242 In addition to supporting the hypothesised introduction of PDV by harp seals in the winter of 1987-
243 88, the phylogenetic analyses of the 1988 data provide new insight to the emergence and
244 subsequent pattern of spread among harbour seals. In the beginning of March 1988, the viral strains
245 split into two clades; one clade containing sequences from central Kattegat and a second clade
246 containing sequences from the Wadden Sea and the Limfjord. The analysis of the 1988 data in
247 isolation – as well as the combined 1988-2002 data – places a sample (#42524) collected in the
248 Limfjord on 22 July 1988 at the base of the tree (Supplementary Figure 2). Interestingly, while the
249 viral sequence from sample #42524 contains mutations that are unique to the 1988 outbreak, it
250 diverges from the rest of the 1988 sequences by six mutations (Supplementary Table 9). The
251 presence of these mutations was confirmed by forward and reverse sequencing and Phred scores of
252 > 30, and they do not appear to be artefacts caused by DNA damage and deamination as only three
253 of the six substitutions are C-T or G-A. Moreover, another sample from the Limfjord (#42523) –
254 which was omitted from the full phylogenetic analyses because full sequences only could be
255 obtained for positions 3,910-4,535, 5,027-6,204 and 6,715-8,916, respectively – also contained four
256 of the six unique mutations found in sample #42524. The basal phylogenetic placement of the
257 Limfjord sequence indicates that the 1988 outbreak among European harbour seals consisted of not
258 just a single, but several introductions of related PDV strains. The observed rise in the number of
259 harbour seal deaths along the Danish west coast by February 1988 (Danielsen *et al.,* 1988; Dietz *et*
260 *al.,* 1989a) might have been caused by such an introduction, independent of the full-scale outbreak
261 two months later in Kattegat. This peculiar Limfjord strain could represent an early and partly
262 unsuccessful variant of the virus, later enhanced in the form of the second, more virulent and
263 widespread strain. Further clarification on the role of harp seals as vectors of the 1988 outbreak, as
264 well as the potential introduction, evolution, pathogenicity and circulation of several PDV strains

265 throughout this outbreak, would require the successful retrieval and comparison of viral RNA from
266 European harbour seals and Barents Sea harp seals collected before, during and after the 1988
267 outbreak.
268
269 *Spread of PDV in 2002 through multiple transmission events*
270 In 2002, the first harbour seal mortalities were observed on the island of Anholt in Kattegat on May
271 4th, and by May 30th, sick seals had been reported along the Danish and Swedish coasts of Kattegat
272 and Skagerrak (Härkönen *et al.,* 2006). On June 16th the first cases were reported from Vlieland in
273 the Dutch Wadden Sea (Härkönen *et al.,* 2006), and by July 18th the index case of England was
274 recorded in the Wash area (Lawson & Jepson, 2003), 800-1000 km from Kattegat. Intriguingly, sick
275 seals were not observed in the geographically intermediate German and Danish Wadden Sea and
276 the Limfjord before mid-September (Härkönen *et al.,* 2006). Thus, based on the observed sequence
277 of mortalities, the predominant hypothesis has been that the virus made a long-distance jump from
278 the Kattegat epicentre to localities in the Dutch Wadden Sea from where it spread in two waves:
279 west to England, Belgium and France, and east to the German and Danish Wadden Sea before
280 going north along the coast to the Limfjord (Härkönen *et al.,* 2006).
281
282 The phylogenetic analysis presented here supports a Kattegat epicentre of the PDV outbreak in
283 2002, but offers new insights into the subsequent long-distance spread among European harbour
284 seal colonies (Figure 2A; Supplementary Figure 3). The most basal sample of the phylogenetic tree
285 is from Kattegat, which by early May splits into branches containing samples from Kattegat,
286 Skagerrak, Øresund and southwestern Baltic, in agreement with the observed spatio-temporal
287 pattern of emergence. In contrast, the phylogenetic analysis does not agree with the presumed
288 scenario in which harbour seal colonies outside of the epicentre were infected by one long-distance

289 transmission event from Kattegat to the Dutch Wadden Sea and further to other North Sea colonies.
290 Rather, based on the phylogenetic analyses we hypothesise that PDV arrived in southern England,
291 the Wadden Sea and the Limfjord, respectively, through at least three independent long-distance
292 transmission events from the Kattegat epicentre (Figure 2C). The integrity of each of these
293 subclades is supported by posterior values above 0.8, but the positioning of samples within
294 subclades, as well as the relationship among subclades is poorly resolved (Figure 2A;
295 Supplementary Figure 3).
296
297 The finding of multiple transmission events poses the question of who the likely carriers of PDV
298 among harbour seal colonies are. As harbour seals show strong site fidelity throughout the summer
299 (Dietz *et al.,* 2013) and exhibit fine-scale population genetic structuring (Olsen *et al.,* 2014), it has
300 previously been assumed that the more mobile and less susceptible grey seal acted as carrier of
301 PDV between geographically distant haul-out sites (Hammond *et al.,* 2005; Pomeroy *et al.,* 2005;
302 Härkönen *et al.,* 2006). However, while grey seals did haul-out on Anholt in 2002, the numbers
303 were likely no more than 5-10 animals, several orders of magnitude lower than the 800-1000
304 harbour seals hauling out. Thus, although it only takes one infected animal to start an epidemic, the
305 likelihood of transmission from harbour seal to harbour seal is substantially greater than from grey
306 seal to harbour seal. Moreover, recent tagging and genetic studies indicate that especially juvenile
307 harbour seals undertake movements across much greater distances than previously reported, and
308 thus potentially could have played a larger role in transporting PDV longer distances. For instance,
309 Wadden Sea harbour seals have been shown to undertake foraging trips of a couple of hundred
310 kilometres into the North Sea (Dietz *et al*., 2013; Tougaard *et al.,* 2008; Aarts *et al.,* 2016), and
311 recent population genetic analyses suggested movements among the seals in southern England,
312 France and the Dutch Wadden Sea (Olsen *et al.,* 2017). Likewise, tagging of harbour seals in the

fall and early winter of 2016 provided the first documented movements from a haul-out in southern Kattegat to the central Limfjord (Rune Dietz and collaborators, unpublished) by an animal that has now been genetically determined to originate from the Limfjord (Morten Tange Olsen, unpublished). Thus, while grey seals cannot be ruled out, we hypothesise that both short- and long-distance PDV transmissions in 2002 could have been facilitated by harbour seal movements. Although beyond the scope of the present analysis, additional insights into the dispersal of seals and transmission of PDV could be obtained by integration of genetic, epidemiological and tagging data (Swinton *et al.,* 1998; Hall *et al.,* 2006).

New insights to PDV virus evolution

The mean substitution rate of the full data set – including coding and non-coding regions – was found to be 6.73×10^{-4} sub/site/yr, while the substitution rate estimates for the sequences of each outbreak were 2.28×10^{-3} sub/site/yr and 1.58×10^{-3} sub/site/yr for 1988 and 2002, respectively. The results correspond well with the rates of 10^{-3} to 10^{-4} sub/site/yr reported for other morbilliviruses (Furuse *et al.,* 2010; Hanada *et al.,* 2004; Xu *et al.,* 2014). The substitution rates estimated at gene level did not differ much and were found to be between 8.03×10^{-4} to 2.2×10^{-4} for the full data set, 5.75×10^{-3} to 4.62×10^{-3} for the 1988 sequences and 5.72×10^{-3} to 1.53×10^{-3} for the 2002 sequences (Table 2). In the full dataset, the F, M and H genes had similar substitution rates with overlapping HPD intervals, whereas the substitution rate of the P gene (2.2×10^{-4}) was substantially lower. Interestingly, while the P gene had the lowest substitution rate, it had the highest proportion of non-synonymous substitutions (64%), followed by the F gene (50%) and H gene (48%), and finally the M gene (23%). The P gene encodes for three different proteins; P (1,524 bp), V (899 bp) and C (525 bp), which use different reading frames of the same nucleotide sequence, and high sequence variability has previously been documented in MeV (Bankamp *et al.,* 2008; Beaty & Lee, 2016;

337 Devaux & Cattaneo, 2004). Nine amino acid changes were found in the reading frame of the P
338 protein, which is a part of the polymerase complex and is essential for viral RNA replication
339 (Bankamp *et al.,* 2011). Six amino acid changes were found in the reading frame of the V protein
340 and two amino acid changes were found in the reading frame of the C protein. The V and C proteins
341 both play important roles in the inhibition of interferons (Nanda & Baron 2006; Chinnakannan *et*
342 *al.,* 2014) and influence the antiviral host response. Additionally, the C protein is believed to affect
343 the transcription of mRNA of the viral genes (Baron & Barrett 2000) and this could explain why the
344 percentage of changes is lower than the V protein. H and F genes code for the hemagglutinin
345 glycoprotein and the fusion protein, respectively, allowing the virus to attach and fuse with the host
346 cell membrane (Wild *et al.,* 1991). The two genes are known targets for the virus-neutralizing
347 antibodies (Xu *et al.,* 2014), and would be expected to contain a high amount of genetic variation,
348 which our study also suggests. The substitution rate of the H gene for MeV have previously been
349 estimated to be 6.44×10^{-4} sub/site/yr (Furuse *et al.,* 2010) – an estimate which corresponds well to
350 our estimate of 5.6×10^{-4} for PDV. The M gene codes for the matrix protein and is considered to be
351 relatively conserved (Beaty & Lee, 2016), which the relatively low proportion of non-synonymous
352 substitutions estimated for PDV also suggests.
353
354 *Perspectives*
355 Wildlife pathogens, such as PDV and other morbilliviruses, represent a major risk for -epidemic
356 outbreaks resulting in extensive die offs in naïve populations and species. The past centuries'
357 human activities and extensive utilization of natural resources in terrestrial and marine
358 environments have led to substantial habitat alteration, loss of biodiversity and climatic changes
359 (Jones *et al.,* 2008). As a consequence, wildlife species shift their geographical ranges, increasing
360 the likelihood of disease transmissions and outbreaks within and among naïve species and

populations (Karesh *et al.*, 2012). Indeed, the mass migration of harp seals hypothesised to have brought PDV to the naïve harbour seal population in 1988 was caused by a collapse of capelin and herring stocks following intense human exploitation (Haug & Nilssen, 1995). Such outbreaks could lead to ecological or economic losses and have huge consequences for already exposed or endangered species (Osterhaus *et al.,* 1997; Thorne & Williams, 1988). Further work should focus on understanding the dynamics of pathogen emergence and spread, including the co-evolutionary dynamics of intra- and interspecific pathogen transmissions.

ACKNOWLEDGEMENTS

The authors would like to thank Professor Tom Gilbert for encouraging the initiation of the project, Dr. Line Nielsen for initial advice on RT-PCR amplifications, and not least the field personnel that collected tissue samples from thousands of dead harbour seals in 1988 and 2002. The study was supported by the Villum Foundation, the Danish Ministry of the Environment, the Volkswagen Foundation (Az.: 89911), and by the BONUS program BaltHealth, which has received funding from BONUS (Art. 185), funded jointly by the EU, Innovation Fund Denmark (grants 6180-00001B and 6180-00002B), Forschungszentrum Jülich GmbH, German Federal Ministry of Education and Research (grant FKZ 03F0767A), Academy of Finland (grant 311966) and Swedish Foundation for Strategic Environmental Research (MISTRA).

AUTHOR CONTRIBUTIONS

IS and MTO conceived and designed the study; MTO, TH, KH, US, RD, JT and AH provided funding; TH, KH, US, KL, RD, JT, AG, LWH, EC and AH collected and curated the samples; IS conducted the lab work and phylogenetic analyses; IS and MTO drafted the manuscript. All authors provided input to the manuscript and approved the final version.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA ACCESSIBILITY

The edited PDV sequences will be made available on NCBI GenBank upon acceptance of the manuscript.

REFERENCES

Aguilar, A. & Antonio Raga, J., 1993. The Striped Dolphin Epizootic in the Mediterranean Sea. *Royal Swedish Academy of Sciences*, 22(8), pp.524–528.

Akaike, H., 1973. Information Theory and an Extension of the Maximum Likelihood Principle. *In B. N. Petrov, & F. Csaki (Eds.), Proceedings of the 2nd International Symposium on Information Theory*, pp.267–281.

Bankamp, B. et al., 2011. Genetic characterization of measles vaccine strains. *Journal of Infectious Diseases*, 204(SUPPL. 1), pp.533–548.

Bankamp, B. et al., 2008. Genetic variability and mRNA editing frequencies of the phosphoprotein genes of wild-type measles viruses. *Virus Research*, 135(2), pp.298–306.

Baron, M.D. & Barrett, T., 2000. Rinderpest viruses lacking the C and V proteins show specific defects in growth and transcription of viral RNAs. *Journal of virology*, 74(6), pp.2603–11. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10684274%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC111748.

Beaty, S.M. & Lee, B., 2016. Constraints on the genetic and antigenic variability of measles virus.

400 *Viruses*, 8(4), pp.1–20.
401 Birkun, A. et al., 1999. Epizootic of morbilliviral disease in common dolphins (Deiphinus deiphis
402 ponticus) from the Black Sea. *The Veternary Record*, 144, pp.85–92.
403 Bodewes, R. et al., 2013. Prevalence of phocine distemper virus specific antibodies: bracing for the
404 next seal epizootic in north-western Europe. *Emerging Microbes & Infections*, 2(1), p.e3.
405 Available at: http://www.nature.com/doifinder/10.1038/emi.2013.2.
406 Bouckaert, R. et al., 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis.
407 *PLoS Computational Biology*, 10(4), pp.1–6.
408 Bouckaert, R.R. & Heled, J., 2014. DensiTree 2: Seeing trees through the forest. *bioRxiv - Preprint*
409 *Server*, pp.1–11. Available at:
410 http://biorxiv.org/content/early/2014/12/08/012401.abstract%0Ahttp://biorxiv.org/lookup/doi/1
411 0.1101/012401.
412 Cattet, M.R.L. et al., 2004. Antibodies to Canine Distemper and Phocine Distemper Viruses in
413 Polar Bears from the Canadian Arctic. *Journal of Wildlife Diseases*, 40(2), pp.338–342.
414 Chinnakannan, S.K. et al., 2014. Different functions of the common P/V/W and V-specific domains
415 of rinderpest virus V protein in blocking IFN signalling. *Journal of General Virology*,
416 95(PART 1), pp.44–51.
417 Curran, M.D. et al., 1990. Nucleotide sequence analysis of phocine distemper virus reveals its
418 distinctness from canine distemper virus. *Veterinary Record*, 127(17), pp.430–431.
419 Danielsen, F., Durinck, J. & Skov, H., 1988. Ilanddrevne sæler langs udvalgte danske kyster juni
420 1988. *Ornis Consult*, p.Vesterbrogade 140, DK-1620 København V, Denmark (i.
421 Delport, W. et al., 2010. Datamonkey 2010: a suit of phylogenetic analysis tools for evolutionary
422 biology. *Bioinformatics*, 26, pp.2455–2457.
423 Dietz, R. et al., 1989. Clue to seal epizootic. *Nature*, 338(627).

424 Dietz, R. et al., 2013. Movement and site fidelity of harbour seals (Phoca vitulina) in Kattegat,
425 Denmark, with implications for the epidemiology of the phocine distemper virus. *ICES*
426 *Journal of Marine Science*, 70, pp.186–195.
427 Dietz, R., Heide-Jørgensen, M.P. & Härkönen, T., 1989. Mass deaths of Harbor seals (Phoca
428 vitulina) in Europe. *AMBIO*, 18(5), pp.258–264.
429 Domingo, M. et al., 1990. Morbillivirus in dophins. *Nature*, 348, p.21.
430 Drummond, A.J. et al., 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology*,
431 4(5), pp.699–710.
432 Duffy, S., Shackelton, L.A. & Holmes, E.C., 2008. Rates of evolutionary change in viruses: patterns
433 and determinants. *Nature reviews. Genetics*, 9(4), pp.267–76. Available at:
434 http://dx.doi.org/10.1038/nrg2323.
435 Duignan, P.J. et al., 2014. Phocine Distemper Virus: Current Knowledge and Future Directions.
436 *Viruses*, 6, pp.5093–5134. Available at: www.mdpi.com/journal/viruses.
437 Earle, J.P. et al., 2011. Phocine Distemper Virus in Seals, East Coast, United States, 2006.
438 *Emerging Infectious Diseases*, 17(2), pp.215–220.
439 Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high
440 throughput. *Nucleic Acids Research*, 32(5), pp.1792–1797.
441 Folkow, L.P., Nordøy, E.S. & Blix, A.S., 2004. Distribution and diving behaviour of harp seals
442 (Pagophilus groenlandicus) from the Greenland Sea stock. *Polar Biology*, 27(5), pp.281–298.
443 Furuse, Y., Suzuki, A. & Oshitani, H., 2010. Origin of measles virus: divergence from rinderpest
444 virus between the 11th and 12th centuries. *Virology journal*, 7, p.52.
445 Goldstein, T. et al., 2009. Phocine distemper virus in northern sea otters in the Pacific Ocean,
446 Alaska, USA. *Emerging Infectious Diseases*, 15(6), pp.925–927.
447 Hall, A.J. et al., 2006. Phocine distemper virus in the North and European Seas - Data and models,

448 nature and nurture. *Biological Conservation*, 131(2), pp.221–229.
449 Hammond, J.A. et al., 2005. Identification and real-time PCR quantification of Phocine distemper
450 virus from two colonies of Scottish grey seals in 2002. *Journal of General Virology*, 86(9),
451 pp.2563–2567.
452 Hanada, K., Suzuki, Y. & Gojobori, T., 2004. A large variation in the rates of synonymous
453 substitution for RNA viruses and its relationship to a diversity of viral infection and
454 transmission modes. *Molecular Biology and Evolution*, 21(6), pp.1074–1080.
455 Hasegawa, M., Kishino, H. & Yano, T. aki, 1985. Dating of the human-ape splitting by a molecular
456 clock of mitochondrial DNA. *Journal of Molecular Evolution*, 22(2), pp.160–174.
457 Haug, T., 1990. This report not to be cited without prior reference to authors International Council
458 for Exploration of the Sea C.M. 1990.
459 Haug, T. & Nilssen, K.T., 1995. Ecological implications of harp seal Phoca groenlandica invasions
460 in northern Norway. *Whales, seals, fish and man: Proceedings of the International Symposium*
461 *on the biology of Marine Mammals in the North East Atlantic, Tromsø, Norway, 29 November-*
462 *1 December 1994*, pp.545–556.
463 Härkönen, T., Dietz, R. & Reijnders, P., 2006a. A review of the 1988 and 2002 phocine distemper
464 virus epidemics in European harbour seals. *Dis Aquat Org*, 68, pp.115–130.
465 Härkönen, T., Dietz, R. & Reijnders, P., 2006b. A review of the 1988 and 2002 phocine distemper
466 virus epidemics in European harbour seals. *Dis Aquat Org*, 68, pp.115–130.
467 Jo, W.K., Osterhaus, A.D. & Ludlow, M., 2018. Transmission of morbilliviruses within and among
468 marine mammal species. *Current Opinion in Virology*, 28, pp.133–141. Available at:
469 http://dx.doi.org/10.1016/j.coviro.2017.12.005.
470 Jones, K.E. et al., 2008. Global trends in emerging infectious diseases. *Nature*, 451(7181), pp.990-
471 U4.

472 Karesh, W.B. et al., 2012. Ecology of zoonoses: Natural and unnatural histories. *The Lancet*,
473 380(9857), pp.1936–1945. Available at: http://dx.doi.org/10.1016/S0140-6736(12)61678-X.
474 Kearse, M. et al., 2012. Geneious Basic: An integrated and extendable desktop software platform
475 for the organization and analysis of sequence data. *Bioinformatics*, 28(12), pp.1647–1649.
476 Kennedy et al., 1989. Viral distemper now found in porpoises. *Nature*, 336(21), pp.97–103.
477 Kreutzer, M. et al., 2008. Brief Report In search of virus carriers of the 1988 and 2002 phocine
478 distemper virus outbreaks in European harbour seals. *Archives of Virology*, 153, pp.187–192.
479 Krog, J.S. et al., 2015. Influenza A (H10N7) virus in dead harbor seals, Denmark. *Emerging*
480 *Infectious Diseases*, 21(4), pp.684–687.
481 Lawson, B. & Jepson, P.D., 2003. UK Phocine Distemper Virus Epizootic Investigation Report
482 2002 / 2003.
483 Lio, P. & Goldman, N., 1998. Models of molecular evolution and phylogeny. *Genome Res*, 8(12),
484 pp.1233–1244.
485 Lipscomb, T.P. et al., 1996. Morbillivirus epizootic in bottlenose dolphins of the Gulf of Mexico.
486 *Journal of Veterinary Diagnostic Investigation*, 8(3), pp.283–290.
487 Loeng, H. & Drinkwater, K., 2007. An overview of the ecosystems of the Barents and Norwegian
488 Seas and their response to climate variability. *Deep-Sea Research II*, 54, pp.2478–2500.
489 Ludes-Wehrmeister, E. et al., 2016. Phocine distemper virus (PDV) seroprevalence as predictor for
490 future outbreaks in harbour seals. *Veterinary Microbiology*, 183, pp.43–49.
491 Mamaev, L. V. et al., 1995. Characterisation of morbilliviruses isolated from Lake Baikal seals
492 (Phoca sibirica). *Veterinary Microbiology*, 44(2–4), pp.251–259.
493 Markussen, N.H. & Have, P., 1992. Phocine distemper virus infection in harp seals (Phoca
494 groenlandica). *Marine Mammal Science*, 8(1), pp.19–26. Available at:
495 http://doi.wiley.com/10.1111/j.1748-7692.1992.tb00121.x.

McConnell, B.J. et al., 1999. Movements and foraging of grey seals in the North sea. *The Journal of Applied Ecology*, 36, pp.573–590.

Minin, V. et al., 2003. Performance-Based Selection of Likelihood Models for Phylogeny Estimation. *Systematic Biology*, 52(5), pp.674–683. Available at: https://academic.oup.com/sysbio/article-lookup/doi/10.1080/10635150390235494.

Nanda, S.K. & Baron, M.D., 2006. Rinderpest Virus Blocks Type I and Type II Interferon Action: Role of Structural and Nonstructural Proteins. *Journal of Virology*, 80(15), pp.7555–7568. Available at: http://jvi.asm.org/cgi/doi/10.1128/JVI.02720-05.

Nielsen, L. et al., 2009. Genetic diversity and phylogenetic analysis of the attachment glycoprotein of phocine distemper viruses of the 2002 and 1988 epizootics. *Virus Research*, 144(1–2), pp.323–328.

Nielsen, O. et al., 1997. Epizootiology of Morbillivirus infection in harp, hooded, and and ringed seals from the Canatian Arctic and Western Atlantic. *Journal of Wildlife Diseases*, 33(1), pp.7–19.

Nilssen, K.T. et al., 1998. Invasions of harp seals Phoca groenlandica ERXLEBEN to coastal waters of Norway in 1995: Ecological and demographic implications. *Sarsia*, 83(4), p.345.

Olsen, M.T. et al., 2017. Genetic population structure of harbour seals in the United Kingdom and neighbouring waters. *Aquatic Conservation: Marine and Freshwater Ecosystems*, (May 2016), pp.1–7.

Olsen, M.T. et al., 2014. Integrating genetic data and population viability analyses for the identification of harbour seal (Phoca vitulina) populations and management units. *Molecular Ecology*, 23(4), pp.815–831.

Osterhaus, A. et al., 1997. Morbillivirus in monk seal mass mortality. *Nature*, 388(August), pp.834–838.

520 Pomeroy, P.P. et al., 2005. Morbillivirus neutralising antibodies in Scottish grey seals Halichoerus
521 grypus: Assessing the effects of the 1988 and 2002 PDV epizootics. *Marine Ecology Progress*
522 *Series*, 287(Cdv), pp.241–250.
523 Pond, S.L.K. et al., 2006. Automated phylogenetic detection of recombination using a genetic
524 algorithm. *Molecular Biology and Evolution*, 23(10), pp.1891–1901.
525 Posada, D., 2008. jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution*,
526 25(7), pp.1253–1256.
527 Rambaut, A., 2016. http://tree.bio.ed.ac.uk/software/figtree/.
528 Reijnders, P. et al., 2010. Population development and status of harbour seals (*Phoca vitulina*) in
529 the Wadden Sea. *NAMMCO Scientific Publications*, 8, pp.95–106. Available at:
530 http://septentrio.uit.no/index.php/NAMMCOSP/article/view/2677.
531 Reineking, B., 2003. *Seal epidemic 2002: Information on Dead Seals in the North Sea, Wadden Sea*
532 *and the Kattegat/Skagerrak Area in 2002*,
533 Rijks, J.M., Osterhaus, A.D.M.. & Kuiken, T., 2012. Morbillivirus infections. In *Infectious*
534 *Diseases of Wild Mammals and Birds in Europe*. pp. 99–118.
535 Roeder, P., Mariner, J. & Kock, R., 2013. Rinderpest: the veterinary perspective on eradication. *The*
536 *Royal Society*, 368(1623).
537 Roelke-Parker, M.E. et al., 1996. A canine distemper virus epidemic in Serengeti lions (Panthera
538 leo). *Nature*, 379(6564), pp.441–445. Available at:
539 http://www.ncbi.nlm.nih.gov/pubmed/8559247%5Cnhttp://www.nature.com/nature/journal/v3
540 79/n6564/pdf/379441a0.pdf.
541 Siembieda, J.L. et al., 2017. Epidemiology of a Phocine Distemper Virus Outbreak Along the North
542 Atlantic Coast of the United States. *Aquatic Mammals*, 43(3), pp.254–263. Available at:
543 http://www.aquaticmammalsjournal.org/index.php?option=com_content&view=article&id=16

65:epidemiology-of-a-phocine-distemper-virus-outbreak-along-the-north-atlantic-coast-of-the-united-states&catid=160&Itemid=326.

Swinton, J. et al., 1998. Persistence thresholds for phocine distemper virus infection in harbour seal Phoca vitulina metapopulations. *Journal of Animal Ecology*, 67(1), pp.54–68.

Thorne, E.T. & Williams, E.S., 1988. Disease and Endangered Species: The Black-Footed Ferret as a Recent Example. *Conservation Biology*, 2(1), pp.66–74.

Tougaard, J., Teilmann, J. & Tougaard, S., 2008. Harbour seal spatial distribution estimated from Argos satellite telemetry: Overcoming positioning errors. *Endangered Species Research*, 4(1–2), pp.113–122.

Vandamme, A.-M., Bertazzoni, U. & Salemi, M., 2000. Evolutionary strategies of human T-cell lymphotropic virus type II. *Gene*, 261(1), pp.171–180. Available at: http://www.sciencedirect.com/science/article/pii/S037811190000473X.

de Vries, R.D. et al., 2013. Complete Genome Sequence of Phocine Distemper Virus Isolated from a Harbor Seal (Phoca vitulina) during the 1988 North Sea Epidemic. *Genome Announcements*, 1(3), pp.6–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3695424&tool=pmcentrez&rendertype=abstract.

Wild, T.F., Malvoisin, E. & Buckland, R., 1991. Measles virus: Both the haemagglutinin and fusion glycoproteins are required for fusion. *Journal of General Virology*, 72(2), pp.439–442.

Xia, X. et al., 2003. An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution*, 26(1), pp.1–7.

Xia, X., 2013. DAMBE5: A comprehensive software package for data analysis in molecular biology and evolution. *Molecular Biology and Evolution*, 30(7), pp.1720–1728.

Xu, S. et al., 2014. Evolutionary genetics of genotype H1 measles viruses in China from 1993 to

2012. *Journal of General Virology*, 95(PART 9), pp.1892–1899.

Øien, N. & Øritsland, T., 1995. Use of mark-recapture experiments to monitor seal populations subject to catching. *Developments in Marine Biology*, 4(C), pp.35–45.

Aarts, G. et al., 2016. Spatial distribution and habitat preference of harbour seals (Phoca vitulina) in the Dutch North Sea. *Wageningen University & Research Report C118/16*, (November).

FIGURES AND TABLES

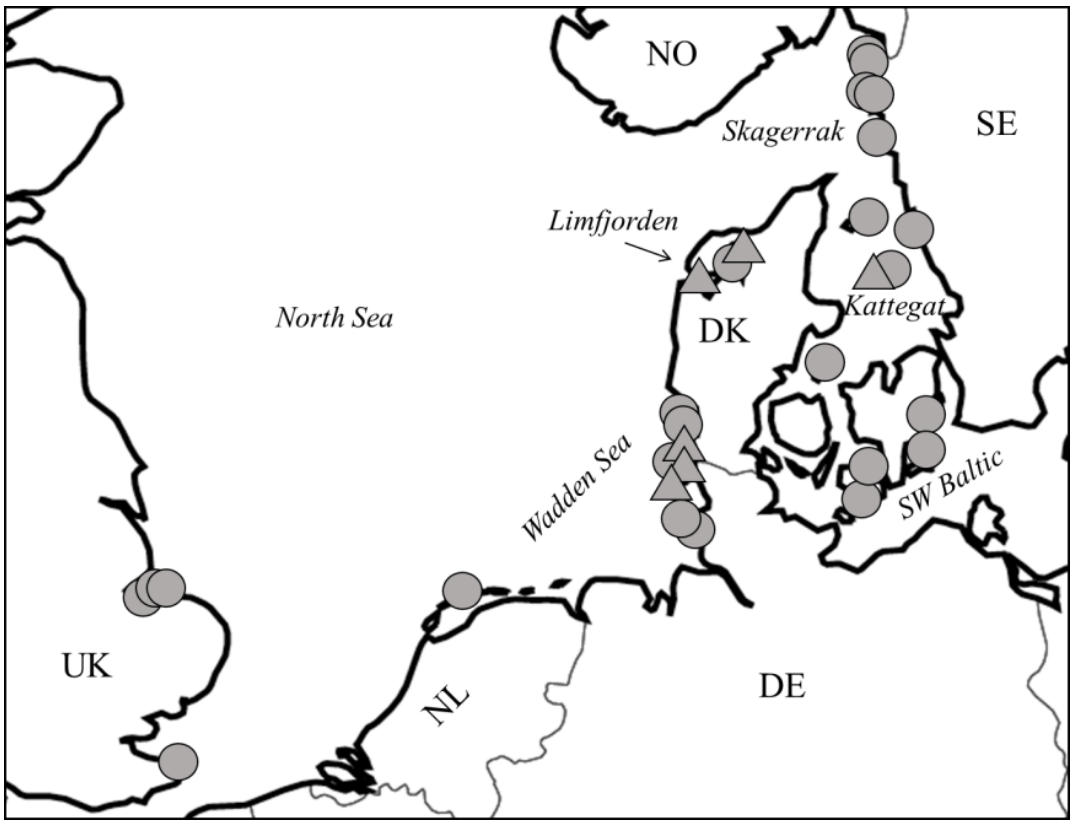


Figure 1 Spatial and temporal origin of the 44 harbour seals and PDV sequences used in the phylogenetic analysis. Triangles = Samples from 1988 collected in Kattegat (N=2), Limfjorden (N=2) and the German Wadden Sea (N=4). Circles = Samples from 2002 collected in Kattegat (N=11), Skagerrak (N=6), the Baltic (N=3), Limfjorden (N=3), the Danish Wadden Sea (N=4), the German Wadden Sea (N=4), the Dutch Wadden Sea (N=1) and the southeastern coast of England (N=5).

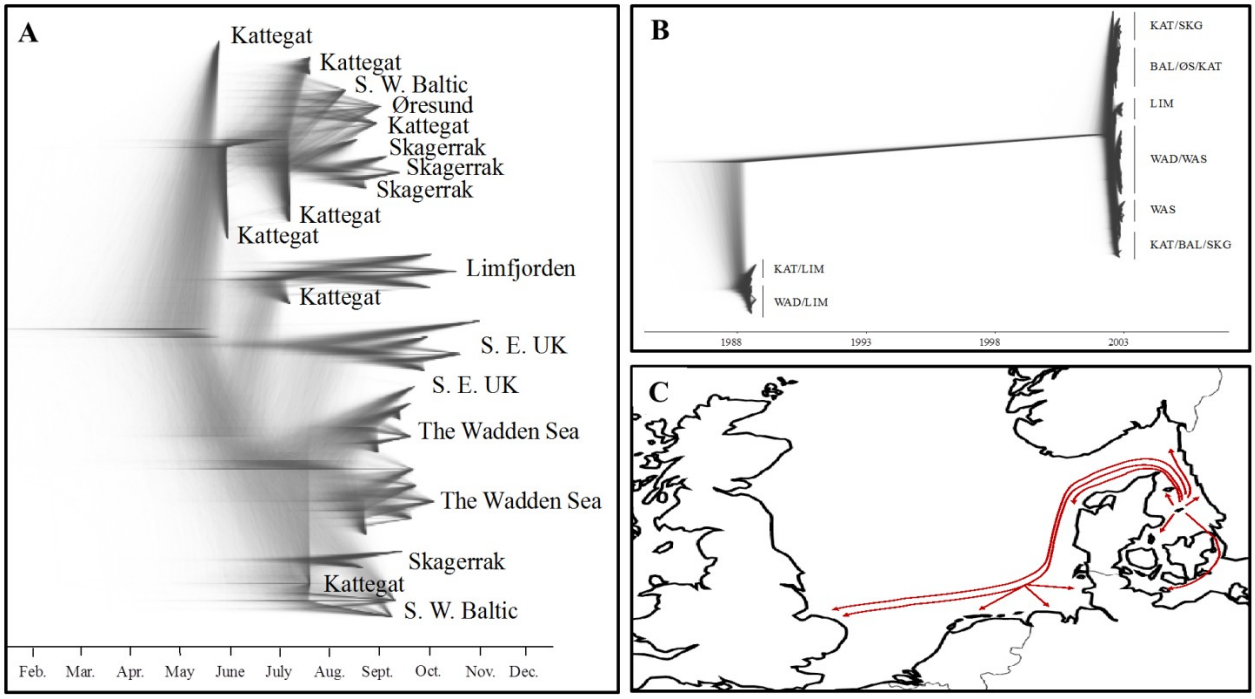


Figure 2 Phylogenetic relationship and hypothesised spread of PDV among European harbour seals. A) Bayesian phylogenetic analysis of 36 PDV sequences obtained from harbour seals sampled during the 2002 outbreak and sequenced in 7,123bp (position 1,744-8,915) of the viral P, M, F and H genes. The analysis provides strong support for Kattegat as the origin of the outbreak, with subsequent independent pulses of spread to harbour seal localities throughout northern Europe. Uncertainties in the tree topology are visualized as a smear around the tree nodes and branches. B) Phylogenetic analysis of the combined data set of 44 samples from 1988 and 2002 supporting the independent origin of the strain leading to the 2002 outbreak. C) A novel hypothesis for the spread of PDV during 2002 from the Kattegat epicentre to other European harbour seal localities through a series of independent transmission events.

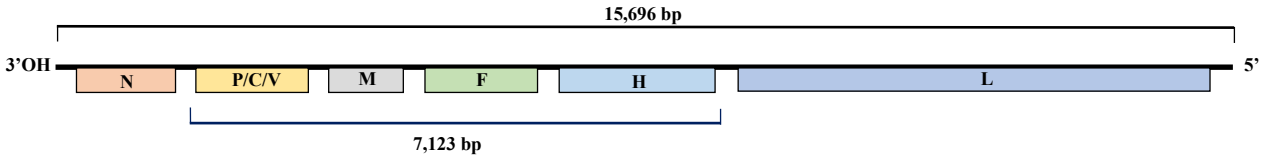
Table 1 Summary statistics for the Bayesian phylogenetic analyses of the sequences from the PDV outbreaks in European harbour seal populations

Data set	ucld.stdev	Coefficient of variance	Rate of evolution sub/site/yr	Tree height	Growth rate	TMRCA date [95 % HPD] 1988	TMRCA date [95 % HPD] 2002
PDVall	0.71	0.75	6.73*10-4	15.54 = 15 y 198 d		25/11-1987	7/1-2002
(constant growth)	[0.41, 1.04]	[0.4, 1.15]	[5.04*10-4. 8.41*10-4]	[14.62, 17.02]	-	[10/6-87, 3/4-88]	[28/7-01, 30/4-02]
	3.74*10-3	4.09*10-3	2.42*10-6	0.01			
PDV1988	0.20	0.19	2.28*10-3	0.38	12.97		
(exp. growth)	[6.52*10-10, 0.69]	[5.61*10-10, 0.64]	[4.54*10-4, 4.22*10-3]	[9.19, 0.73]	[1.76, 23.91]	-	-
	2.55*10-3	2.31*10-3	1.55*10-5	3.17*10-3	0.09		
PDV2002	0.23	0.23	1.58*10-3	0.56 = 203,16 d	8.48		
(exp. growth)	[8.15*10-9, 0.62]	[8.12*10-9, 0.63]	[1.03*10-3, 2.18*10-3]	[0.46, 0.68]	[4.84, 11.96]	-	-
	2.40*10-3	2.47*10-3	5.49*10-6	1,18*10-3	0.03		

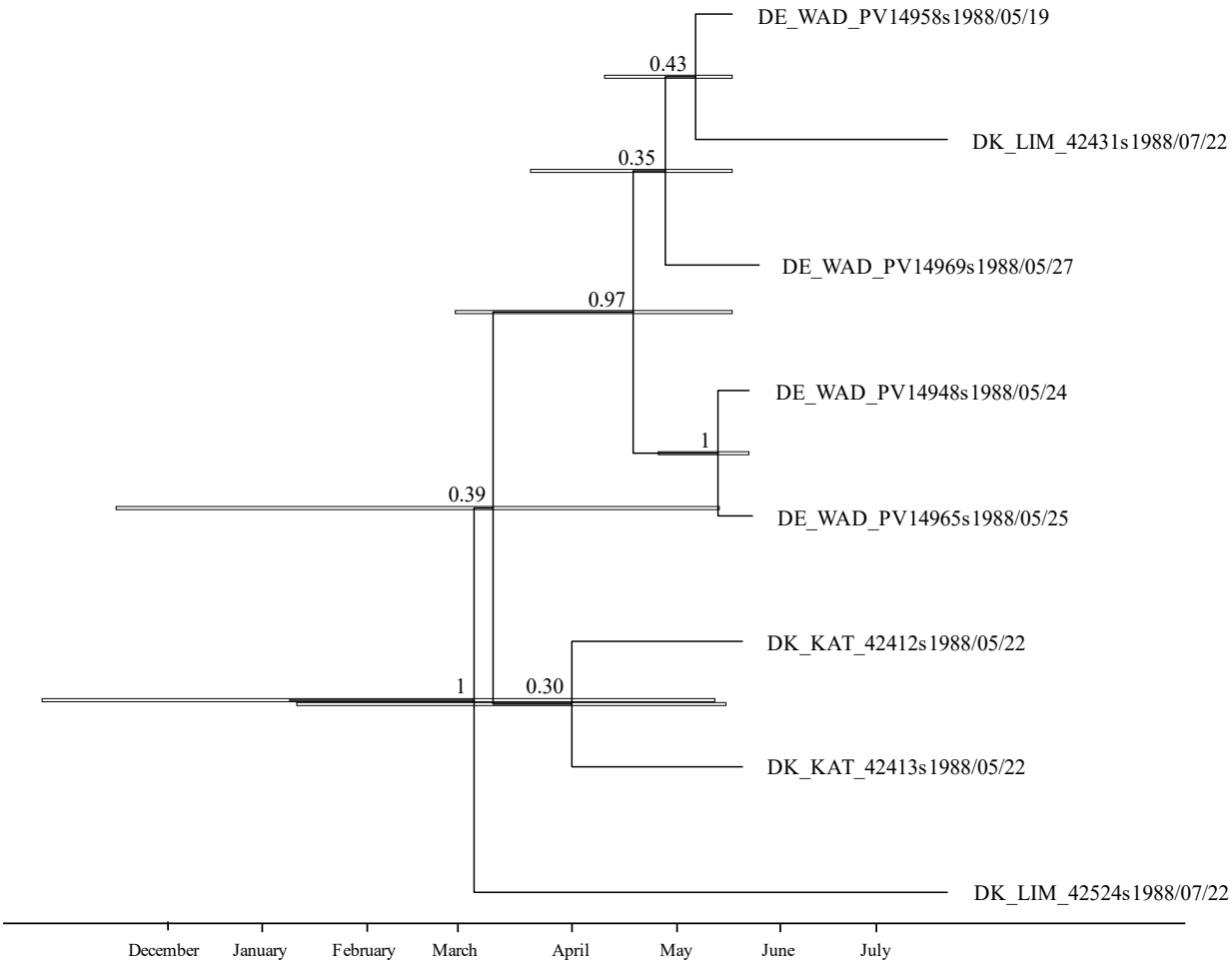
Table 2 The number of genetic differences, amino acid changes and the substitution rates for the F, H, M and P genes calculated using an uncorrelated relaxed clock. The M gene from the 1988 data set contained no substitutions and thus no substitution rate was calculated.

Gene	Lenght	Number of mutations			Amino acid changes			Substitution rate		
		PDVall	PDV1988	PDV2002	PDVall	PDV1988	PDV2002	PDVall	PDV1988	PDV2002
F	1,893 (2,206)	28	4	15	0.50	0	0.66	8.03*10-4 [4.44*10-4, 1.16*10-3] 4.08*10-4	4.62*10-3 [4.52*10-4, 9.82*10-3] 3.4*10-5	3.96*10-3 [1.3*10-3, 7.08*10-3] 5.35*10-5
H	1,821 (1,857)	31	5	12	0.48	0.40	0.50	5.6*10-4 [2.56*10-4, 8.96*10-4] 5.11*10-6	4.75*10-3 [2.44*10-4, 0.01] 4.76*10-5	2.19*10-3 [4.89*10-4, 4.37*10-3] 3.88*10-5
M	1,005 (1,447)	13	0	10	0.23	0	0.20	5.73*10-4 [1.9*10-4, 1.01E*10-3] 6.74*10-6	-	5.72*10-3 [1.57*10-3, 0.01] 6.70*10-5
P	1,521 (1,653)	14	4	8	0.64	0.75	0.75	2.2*10-4 [6.26*10-5, 4.24*10-4] 2.51*10-6	5.71*10-3 [1.89*10-4, 0.01] 5.97*10-5	1.53*10-3 [3.79*10-4, 2.94*10-3] 1.25*10-5

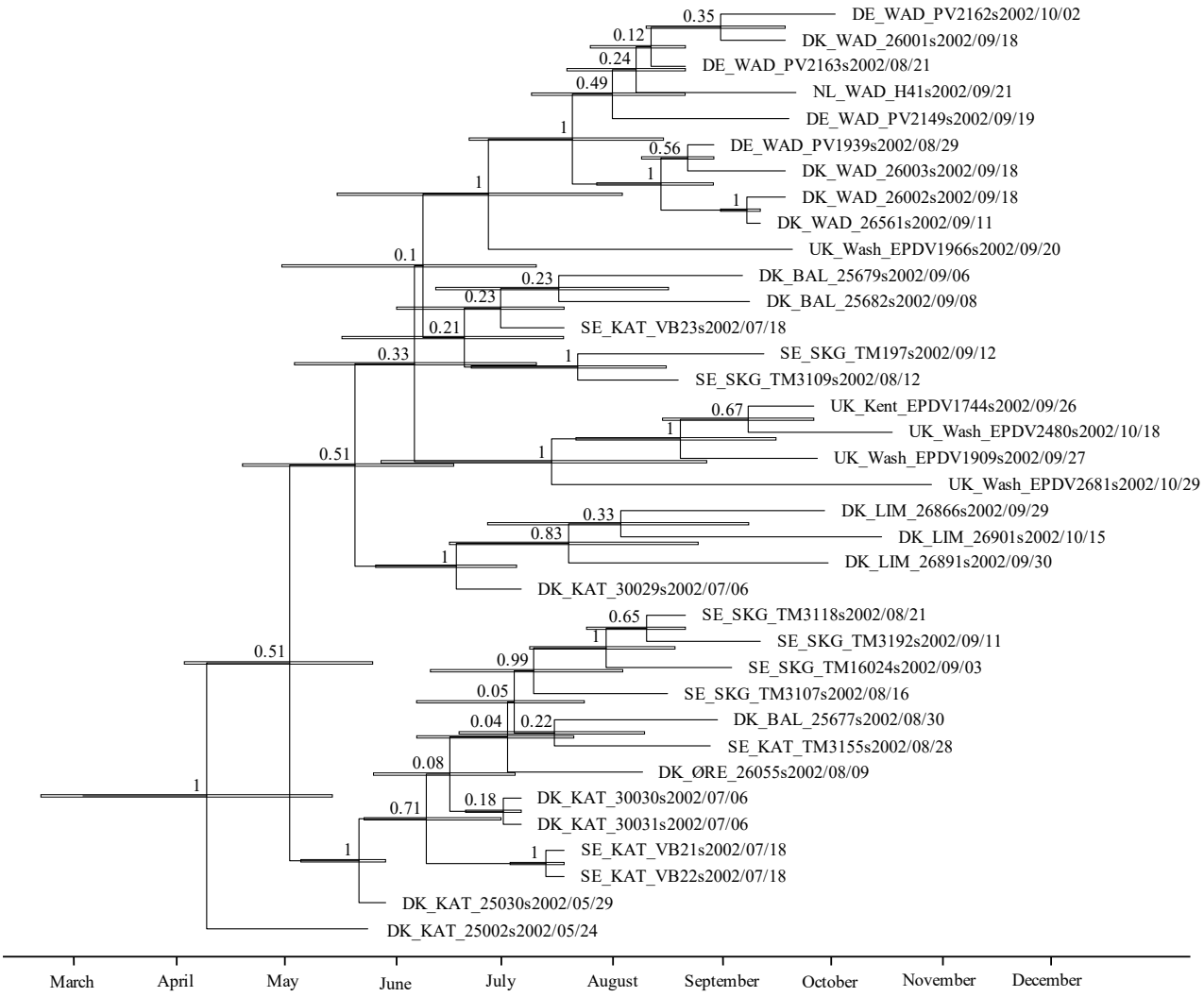
SUPPLEMENTARY MATERIAL



Supplementary Figure 1 A schematic drawing of the genome of Phocine Distemper Virus (PDV) including the six non-overlapping genes; N (position 108-1,679), P (position 1,801-3,324), M (position 3,432-4,439), F (position 5,028-6,923), H (position 7,079-8,902) and L (position 9,036-15,590) coding for at least eight proteins as depicted (de Vries *et al.,* 2013) sequenced in this study (position 1,744-8,915).



Supplementary Figure 2 Phylogenetic relationship among eight PDV strains sequenced from harbour seals infected during the 1988 epidemic. The analysis suggests at least two independent introductions of the PDV strain to Limfjorden (DK_LIM_42524; 1988/07/22) and Kattegat (DK_KAT_42412; 1988/05/22), respectively. Node bars are shown as the 95 % interval of the node heights thereby indicating their uncertainty.



Supplementary Figure 3 Phylogenetic relationship among PDV strains sequenced from harbour seals infected during the 2002 epidemic. The analysis provides strong support for Kattegat as the origin of the outbreak (DK_KAT_25002; 2002/05/24), with subsequent pulses of spread to harbour seal localities in Skagerrak (SKG), SW Baltic (BAL), the Danish German and Dutch Wadden Sea (WAD), Limfjorden (LIM) and the Wash (UK_Wash). Node bars are shown as the 95 % interval of the node heights thereby indicating their uncertainty.

Supplementary Table 1 Summary of tissue type, storage conditions and the number of samples extracted and included in the analyses.

Country	Tissue type		Storage temperature		Sample storage		Number of samples		Number of samples included in the data set	
	1988	2002	1988	2002	1988	2002	1988	2002	1988	2002
Denmark	Liver	Spleen/lung	-20 °C	-20 °C	Organs in bags	Organs in bags	27	28	4	16
England	-	Spleen/lung	-	-20 °C/-80 °C	-	-	-	17		5
Germany	Muscle	Spleen/lung	-20 °C	-20 °C	Organs in bags	Organs in bags	5	5	4	4
The Netherlands	-	Muscle	-	-20 °C	-	-	-	2		1
Scotland	-	Spleen/lung	-	-20 °C	-	-	-	4		0
Sweden	-	Spleen	-	-20 °C/-80 °C	-	Tubes RNAlater	-	11		10
Total							32	67	8	36

Supplementary Table 2 Table of primers used to amplify the fragments sequenced. The primers were compiled from several previous studies and combined to construct the complete genome sequence of PDV (de Vries et al., 2013).

Primer	Sequence	Position start	Position end	Coverage base pairs	Gene
4	5'-TCAGCTAAGTTCTAAAAGACATTGC-3' 3'-CAGTTAGATGAAGCATTTCCTTCGG-5'	1,672	2,303	631	P
5	5'-GAAGGAAGAGAGGGAAGCCTTGATG-3' 3'-GTTTTTTGATCGAGTCAATCTCTCC-5'	2,224	2,854	630	P
6	5'-TGAGGATAATCAATCAATACTTAGC-3' 3'-TGAGTCCTAAGTTTTTTGTAATTGC-5'	2,778	3,407	629	P
7	5'-ACCAATGCCTTTGTTAGTAATCAGG-3' 3'-TCGGAAATCAAACATTCCGCGAGGG-5'	3,333	3,974	641	P/M
8	5'-GTAGTGTACATGAGTATTACTAGAC-3' 3'-AACAGACTAAAGCTAATTAATGCTG-5'	3,9	4,549	649	M
9	5'-TAATGATATAGACTGAAGCACTCAC-3' 3'-TTCAGGGTTTAGAAATGGTGAATCG-5'	4,464	5,108	644	M/F
10	5'-ATGACTCGAGTCAAGAAACTCCCAG-3' 3'-CAGCAGTTGCTACACCTAGAGCTGC-5'	5,028	5,655	627	F
11	5'-GGGATCAGGTAGAAGACAGAGGCG-3' 3'-GGGACAGTAGTGTACCATTCCTGTG-5'	5,582	6,217	635	F
12	5'-GGGTTGTCGTACATAGATTGGAGGC-3' 3'-ATACCGGGAGTGCAATGAGGGTGCC-5'	6,148	6,777	629	F
13	5'-AATCCTCTGATCAAATCCTTGACAC-3' 3'-ATCCTTGACTTGATGATGCACTGCC-5'	6,703	7,348	645	F/H
14	5'-AAAAAGCAATTTGGAGTTCAATAAG-3' 3'-GATAATCCTGTAGTTTGTTGTTTGG-5'	7,27	7,9	630	H
H5/H6	5'-AGATGATATCTTTCCTCC-3' 3'-ATCCATATGAGTTGCTCC-5'	7,615	8,065	451	H
15	5'-GATTCGTGTCTTTGAGATTGGATTC-3' 3'-CCATTTTTAGGTGGAATTGTTAGCC-5'	7,819	8,448	629	H
16	5'-ATAATATTTAACGGTGATGGGATGG-3' 3'-AAACAAGGGAAAGAAACACTTACCG-5'	8,366	8,972	606	H

Supplementary Table 3 RT-PCR conditions for the 14 different primer sets used in this study.

Cycles	1	1	40	/	38	x	(1	1	1)	1	1
Time (min)	30	15					1	1	1	10	∞
Primer 4 (ºC)	50	95	x				94	64	72	72	4
Primer 5 (ºC)	50	95	x				94	64	72	72	4
Primer 6 (ºC)	50	95	x				94	63	72	72	4
Primer 7 (ºC)	50	95			x		94	65	72	72	4
Primer 8 (ºC)	50	95	x				94	63	72	72	4
Primer 9 (ºC)	50	96	x				94	62	72	72	4
Primer 10 (ºC)	50	95	x				94	62	72	72	4
Primer 11 (ºC)	50	95	x				94	62	72	72	4
Primer 12 (ºC)	50	95	x				94	62	72	72	4
Primer 13 (ºC)	50	95	x				94	62	72	72	4
Primer 14 (ºC)	50	95	x				94	58	72	72	4
(ºC)	50	95	x				94	55	72	72	4
Primer 15 (ºC)	50	95	x				94	58	72	72	4
Primer 16 (ºC)	50	95			x		94	64	72	72	4

Supplementary Table 4 The final sample set of PDV sequences covering the coding and non-coding regions of the P, M, F and H genes of samples collected in 2002 and 1988. Numbers 1-4 indicate identical sequences. * Note that viral RNA was not successfully sequenced with primer 6 for the samples 42412 and 42413 and a fragment from a German sample was used to fill the gap (see Table 9).

Sample	Species	Location	Area	Date	Tissue
25002	Harbour seal	Læsø	DK-KAT	24. May 2002	(Liver)
25030 [4]	Harbour seal	Anholt	DK-KAT	29. May 2002	Spleen
30029	Harbour seal	Samsø	DK-KAT	6. July 2002	Spleen
30030 [4]	Harbour seal	Samsø	DK-KAT	6. July 2002	Spleen
30031 [4]	Harbour seal	Samsø	DK-KAT	6. July 2002	Spleen
VB21 [3]	Harbour seal	Prästaskär	SE-KAT	18. July 2002	Spleen
VB22 [3]	Harbour seal	Flatbåden	SE-KAT	18. July 2002	Spleen
VB23	Harbour seal	Prästaskär	SE-KAT	18. July 2002	Spleen
TM3155	Harbour seal	Luseskär	SE-KAT	28. August 2002	Spleen
TM197	Harbour seal	Väderöarna	SE-SKG	12. September 2002	Spleen
TM3107	Harbour seal	Sälvik	SE-SKG	16. August 2002	Spleen
TM3109	Harbour seal	Koster	SE-SKG	19. August 2002	Spleen
TM3118	Harbour seal	Lilla Sockna	SE-SKG	21. August 2002	Spleen
TM3192	Harbour seal	Tjärnö	SE-SKG	11. September 2002	Spleen
TM16024	Harbour seal	Ramsökalven	SE-SKG	3. September 2002	Spleen
26055	Harbour seal	Karlslunde	DK-ØRE	9. August 2002	Spleen
25679	Harbour seal	Møn	DK-BAL	6. September 2002	Spleen
25682	Harbour seal	Rødsand	DK-BAL	8. September 2002	Spleen
25677	Harbour seal	Kippinge	DK-BAL	30. August 2002	Spleen
H41	Harbour seal	Ameland	NL-WAD	9. September 2002	Muscle
Pv 1939	Harbour seal	Eiderstedt	DE-WAD	29. August 2002	Spleen/lung
Pv 2149	Harbour seal	g	DE-WAD	19. September 2002	Spleen/lung
Pv 2162	Harbour seal	Sylt	DE-WAD	2. October 2002	Spleen
Pv 2163 [2]	Harbour seal	Eiderstedt	DE-WAD	21. August 2002	Spleen
26001 [2]	Harbour seal	(North)	DK-WAD	18. September 2002	Spleen
26002 [1]	Harbour seal	(North)	DK-WAD	18. September 2002	Spleen
26003	Harbour seal	(North)	DK-WAD	18. September 2002	Spleen
26561 [1]	Harbour seal	Koresand	DK-WAD	11. September 2002	Spleen
26866	Harbour seal	Blinde Røn	DK-LIM	29. September 2002	Spleen
26891	Harbour seal	Blinde Røn	DK-LIM	30. September 2002	Spleen
26901	Harbour seal	Blinde Røn	DK-LIM	15. October 2002	Spleen

Supplementary Table 4 (continued)

Sample	Species	Location	Area	Date	Tissue
EPDV2681	Harbour seal	Wash	UK-Wash	29. September 2002	Spleen/lung
EPDV2480	Harbour seal	Wash	UK-Wash	18. October 2002	Spleen/lung
EPDV1966	Harbour seal	Wash	UK-Wash	20. September 2002	Spleen/lung
EPDV1909	Harbour seal	Wash	UK-Wash	27. September 2002	Spleen/lung
EPDV1744	Harbour seal	Kent	UK-Wash	26. September 2002	Spleen/lung
PV14965	Harbour seal	Wadden Sea	DE-WAD	25. May 1988	Muscle/blubber
PV14958	Harbour seal	Wadden Sea	DE-WAD	19. May1988	Muscle/blubber
PV14948	Harbour seal	Wadden Sea	DE-WAD	24. May 1988	Muscle/blubber
PV14969	Harbour seal	Wadden Sea	DE-WAD	27. May 1988	Muscle/blubber
42524	Harbour seal	Limfjorden	DK-LIM	22. July 1988	Liver
42531	Harbour seal	Limfjorden	DK-LIM	22. July 1988	Liver
42412 *	Harbour seal	Anholt	DK-KAT	22. May 1988	Liver
42413 *	Harbour seal	Anholt	DK-KAT	22. May 1988	Liver

Supplementary Table 5 Results of the substitution saturation tests of the alignment of PDV sequences from 1988 and 2002. No substitution saturation is detected if the Iss is significantly lower than Iss.c (Xia et al. 2003).

Codon(s)	Iss	Iss.c
Codon position 1 and 2	0.0117	0.8026
Codon position 3	0.0027	0.7909

Supplementary Table 6 Results of the substitution model tests conducted in JModeltest (Posada 2008). The score is the weights of AIC, AICc, BIC and DT and the model of the highest weight is the model which is best fitted for the given alignment.

Alignment	AIC		AICc		BIC		DT	
	Model	Score	Model	Score	Model	Score	Model	Score
PDVall	TVM + I	0.164	TPM1uf +I	0.165	HKY + I	0.307	HKY + I	0.016
PDV1988	TPM1uf	0.302	TPM1uf	0.304	HKY	0.487	HKY	0.057
PDV2002	TPM1uf +I	0.137	TPM1uf +I	0.138	HKY	0.731	HKY	0.031

Supplementary Table 7 General priors applied to the parameters in BEAUTi.

Parameter	Distribution	Initial value	Interval
gammaShape.s:1stpos	Exponential	1	[-∞, ∞]
gammasShape.s:2ndpos	Exponential	1	[-∞, ∞]
gammasShape.s:3rdpos	Exponential	1	[-∞, ∞]
gammasShape.s:coding	Exponential	1	[-∞, ∞]
gammasShape.s:noncoding	Exponential	1	[-∞, ∞]
Kappa.s:1stpos	Log normal	2	[0,0, ∞]
Kappa.s:2ndpos	Log normal	2	[0,0, ∞]
Kappa.s:3rdpos	Log normal	2	[0,0, ∞]
Kappa.s:coding	Log normal	2	[0,0, ∞]
Kappa.s:noncoding	Log normal	2	[0,0, ∞]

Supplementary Table 8 Priors applied for the different tests of the coalescence constant and exponential population growth.

Parameter	Distribution	Initial value	Interval
ePopSize.t.tree	Gamma	1	[0,001, 1000]
ClockRate.c:clock	Uniform	1	[-∞, ∞]
ucldMean.c:clock (PDVall)	Uniform	0.0014	[0, 0.1]
ucldMean.c:clock (1988 and 2002)	Uniform	4.329E-4	[0, 0.1]
GrowthRate.t.tree	Laplace Distribution	3.0E-4	[-∞, ∞]

Supplementary Table 9 Mutations found in the sequences from samples collected in 1988. The parentheses indicate the positions where a fragment from a German sample was used to fill gaps. Results from an additional sample (42523) which was not fully sequenced was included and non-sequenced areas marked with -.

Samples	Locality	1,868 (Q = 52)	2,251 (Q = 52)	2,481 (Q = 55)	3,131 (Q = 55)	3,357 (Q = 52)	4,562 (Q = 52)
PV14965	Germany	C	A	C	G	G	G
PV14958	Germany	C	A	C	G	G	G
PV14948	Germany	C	A	C	G	**A**	G
PV14969	Germany	C	**G**	C	G	G	**A**
42524	Limfjorden	C	A	**T**	**A**	G	G
42523	Limfjorden	-	-	-	-	-	-
42531	Limfjorden	C	A	C	G	G	G
42412	Anholt	**A**	A	C	G	(G)	(G)
42413	Anholt	C	A	C	G	(G)	(G)

Samples	Locality	4,869 (Q = 54)	4,978 (Q = 59)	5,126 (Q = 30)	5,202	5,816	5,891
PV14965	Germany	A	C	C	C	T	**C**
PV14958	Germany	A	C	C	C	T	**C**
PV14948	Germany	A	C	C	C	T	**C**
PV14969	Germany	A	C	C	C	T	**C**
42524	Limfjorden	A	C	C	**T**	**G**	T
42523	Limfjorden	-	-	C	**T**	**G**	T
42531	Limfjorden	**C**	C	**T**	C	T	**C**
42412	Anholt	A	C	C	C	T	T
42413	Anholt	A	**T**	C	C	T	T

Samples	Locality	7,414	7,444	7,602 (Q = 52)	8,023	8,201 (Q = 52)
PV14965	Germany	T	A	T	**C**	A
PV14958	Germany	T	A	T	A	A
PV14948	Germany	T	A	T	**C**	A
PV14969	Germany	T	A	T	A	A
42524	Limfjorden	**C**	**G**	T	A	A
42523	Limfjorden	**C**	**G**	T	A	A
42531	Limfjorden	T	A	T	A	A
42412	Anholt	T	A	**G**	A	A
42413	Anholt	T	A	T	A	**G**

Supplementary Table 10 The mutations (≥ 2) of the PDV2002 data set. Samples are listed with a number indicating the clade to which they belong to in the tree.

Sample	Area	1,792 (P gene)	1,795 (P gene)	2,152 (P gene)	2,473 (P gene)
0_25002	DK-KAT	T	T	G	T
1_25030 [4]	DK-KAT	T	T	G	C
1_30030 [4]	DK-KAT	T	T	G	C
1_30031 [4]	DK-KAT	T	T	G	C
1_VB21 [3]	SE-KAT	T	T	G	C
1_VB22 [3]	SE-KAT	T	T	G	C
1_TM3155	SE-KAT	T	T	G	C
1_TM3107	SE-SKG	T	C	G	C
1_TM3118	SE-SKG	T	C	G	C
1_TM3192	SE-SKG	T	C	G	C
1_TM16024	SE-SKG	T	C	G	C
1_26055	DK-ØRE	T	T	G	C
1_25677	DK-BAL	T	T	G	C
2_26866	DK-LIM	T	T	G	T
2_26891	DK-LIM	T	T	G	T
2_26901	DK-LIM	T	T	G	T
2_30029	DK-KAT	T	T	G	T
3_EPDV2681	UK-Wash	T	T	G	T
3_EPDV2480	UK-Wash	T	T	G	T
3_EPDV1909	UK-Wash	T	T	G	T
3_EPDV1744	UK-Wash	T	T	G	T
4_EPDV1966	UK-Wash	T	T	G	T
4_H41	NL-WAD	T	T	G	T
4_Pv 1939	DE-WAD	C	T	G	T
4_Pv 2149	DE-WAD	T	T	G	T
4_Pv 2162	DE-WAD	T	T	G	T
4_ Pv 2163 [2]	DE-WAD	T	T	G	T
4_ 26001 [2]	DK-WAD	T	T	G	T
4_ 26002 [1]	DK-WAD	C	T	A	T
4_26003	DK-WAD	C	T	G	T
4_26561 [1]	DK-WAD	C	T	A	T
5_VB23	SE-KAT	T	T	G	T
5_TM197	SE-SKG	T	T	G	T
5_TM3109	SE-SKG	T	T	G	T
5_25679	DK-BAL	T	T	G	T
5_25682	DK-BAL	T	T	G	T

Sample	Area	3,614 (M gene)	4,010 (M gene)	4,698 (Non c. M gene)	4,747 (Non c. M gene)
0_25002	DK-KAT	T	C	T	A
1_25030 [4]	DK-KAT	T	C	T	A
1_30030 [4]	DK-KAT	T	C	T	A
1_30031 [4]	DK-KAT	T	C	T	A
1_VB21 [3]	SE-KAT	T	C	**C**	A
1_VB22 [3]	SE-KAT	T	C	T	A
1_TM3155	SE-KAT	T	C	**C**	A
1_TM3107	SE-SKG	T	C	T	A
1_TM3118	SE-SKG	T	C	T	A
1_TM3192	SE-SKG	T	C	T	A
1_TM16024	SE-SKG	T	C	T	A
1_26055	DK-ØRE	T	C	T	A
1_25677	DK-BAL	T	C	T	A
2_26866	DK-LIM	T	C	T	A
2_26891	DK-LIM	T	C	T	A
2_26901	DK-LIM	T	C	T	A
2_30029	DK-KAT	T	C	T	A
3_EPDV2681	UK-Wash	**C**	C	T	A
3_EPDV2480	UK-Wash	**C**	C	T	A
3_EPDV1909	UK-Wash	**C**	C	T	A
3_EPDV1744	UK-Wash	**C**	C	T	A
4_EPDV1966	UK-Wash	T	**T**	T	**G**
4_H41	NL-WAD	T	**T**	T	A
4_Pv 1939	DE-WAD	T	**T**	T	A
4_Pv 2149	DE-WAD	T	**T**	T	**G**
4_Pv 2162	DE-WAD	T	**T**	T	A
4_ Pv 2163 [2]	DE-WAD	T	**T**	T	A
4_ 26001 [2]	DK-WAD	T	**T**	T	A
4_ 26002 [1]	DK-WAD	T	**T**	T	A
4_26003	DK-WAD	T	**T**	T	A
4_26561 [1]	DK-WAD	T	**T**	T	A
5_VB23	SE-KAT	T	C	T	A
5_TM197	SE-SKG	T	C	T	A
5_TM3109	SE-SKG	T	C	T	A
5_25679	DK-BAL	T	C	T	A
5_25682	DK-BAL	T	C	T	A

Sample	Area	5,064 (F gene)	5,266 (F gene)	5,306 (F gene)	6,893 (F gene)
0_25002	DK-KAT	C	T	A	G
1_25030 [4]	DK-KAT	C	T	A	G
1_30030 [4]	DK-KAT	C	T	A	G
1_30031 [4]	DK-KAT	C	T	A	G
1_VB21 [3]	SE-KAT	C	T	A	G
1_VB22 [3]	SE-KAT	C	T	A	G
1_TM3155	SE-KAT	C	T	A	G
1_TM3107	SE-SKG	C	T	A	G
1_TM3118	SE-SKG	**T**	T	A	G
1_TM3192	SE-SKG	**T**	T	A	G
1_TM16024	SE-SKG	**T**	T	A	G
1_26055	DK-ØRE	C	T	A	G
1_25677	DK-BAL	C	T	A	G
2_26866	DK-LIM	C	T	A	G
2_26891	DK-LIM	C	T	A	G
2_26901	DK-LIM	C	T	A	G
2_30029	DK-KAT	C	T	A	G
3_EPDV2681	UK-Wash	C	T	A	G
3_EPDV2480	UK-Wash	C	T	A	**A**
3_EPDV1909	UK-Wash	C	T	A	**A**
3_EPDV1744	UK-Wash	C	T	A	**A**
4_EPDV1966	UK-Wash	C	T	A	G
4_H41	NL-WAD	C	T	A	G
4_Pv 1939	DE-WAD	C	**G**	A	G
4_Pv 2149	DE-WAD	C	T	A	G
4_Pv 2162	DE-WAD	C	T	A	G
4_ Pv 2163 [2]	DE-WAD	C	T	A	G
4_ 26001 [2]	DK-WAD	C	T	A	G
4_ 26002 [1]	DK-WAD	C	**G**	**G**	G
4_26003	DK-WAD	C	**G**	A	G
4_26561 [1]	DK-WAD	C	**G**	**G**	G
5_VB23	SE-KAT	C	T	A	G
5_TM197	SE-SKG	**T**	T	A	G
5_TM3109	SE-SKG	C	T	A	G
5_25679	DK-BAL	C	T	A	G
5_25682	DK-BAL	C	T	A	G