




OPEN

Impact of viral features, host jumps and phylogeography on the rapid evolution of Aleutian mink disease virus (AMDV)

Giovanni Franzo , Matteo Legnardi, Laura Grassi, Giorgia Dotto, Michele Drigo, Mattia Cecchinato & Claudia Maria Tucciarone

Aleutian mink disease virus (AMDV) is one of the most relevant pathogens of domestic mink, where it can cause significant economic losses, and wild species, which are considered a threat to mink farms. Despite their relevance, many aspects of the origin, evolution, and geographic and host spreading patterns of AMDV have never been investigated on a global scale using a comprehensive biostatistical approach. The present study, benefitting from a large dataset of sequences collected worldwide and several phylodynamic-based approaches, demonstrates the ancient origin of AMDV and its broad, unconstrained circulation from the initial intercontinental spread to the massive among-country circulation, especially within Europe, combined with local persistence and evolution. Clear expansion of the viral population size occurred over time until more effective control measures started to be applied. The role of frequent changes in epidemiological niches, including different hosts, in driving the high nucleotide and amino acid evolutionary rates was also explored by comparing the strengths of selective pressures acting on different populations. The obtained results suggest that the viral passage among locations and between wild and domesticated animals poses a double threat to farm profitability and animal welfare and health, which is particularly relevant for endangered species. Therefore, further efforts must be made to limit viral circulation and to refine our knowledge of factors enhancing AMDV spread, particularly at the wild-domestic interface.

Aleutian disease is probably the most relevant cause of economic losses in global mink farming, in terms of both direct and control-associated costs¹. The disease is caused by a virus classified as the species *Carnivore amdoparvovirus 1* in genus *Amdoparvovirus* and typically features a progressive immune-complex-related syndrome leading to glomerulonephritis and arteritis, although the clinical outcome is affected by the interactions between Aleutian mink disease virus (AMDV) strain virulence and host-related factors^{2,3}. An acute syndrome can also occur in newborn minks and is characterized by often fatal interstitial pneumonia⁴. Due to the North American origin of the farmed mink species, AMDV was thought to have originated from this area before or after domestication, later spreading to other countries because of local and international animal trades⁵. However, definitive and robust evidence on this topic is not available, and the origin of AMDV from another area and/or feral mustelid species cannot be excluded⁶.

Similar to other members of the *Parvoviridae* family, AMDV is characterized by single-stranded DNA (ssDNA) ~ 4.8 kb in size encoding 3 nonstructural (NS1-3) and 2 capsid structural proteins (VP1-2)⁷. In most viruses, higher phenotypic variability and genotypic variability affect structural proteins that evolve under the pressure of escaping the host immune response. However, likely because of the antibody-mediated enhancement that takes part in the viral infectious cycle, such pressure is not present. The primary replication site for AMDV is in circulating macrophages, and viral entry is mediated by cellular Fc receptors recognizing antibody-coated viral particles⁸. Therefore, escape from the host immune response could paradoxically be detrimental to viral fitness⁹. On the other hand, NS proteins, particularly NS1, which is essential for viral replication, are characterized by marked genetic variability and have thus been widely studied and sequenced for epidemiological purposes^{5,10}, allowing the identification of different clusters. The causes of NS protein variability are still uncertain, although a combination of the high evolutionary rate typical of ssDNA viruses and immunity-induced selective pressures

Department of Animal Medicine, Production and Health (MAPS), University of Padua, 35020 Legnaro, PD, Italy.
✉ email: giovanni.franzo@unipd.it

can be hypothesized. Except for local aggregates, clear geographical grouping does not seem to occur¹¹. Several small-scale studies reported noteworthy diversity in the same region or even within farms, suggesting that long- and short-distance animal trades are likely to play a relevant role in shaping the wide and largely unconstrained viral circulation of AMDV and its epidemiological patterns^{5,12}.

In addition to American minks (*Neovison vison*), AMDV is known to infect other mustelids, including European minks (*Mustela lutreola*), ferrets (*Mustela putorius furo*), short-tailed weasels (*Mustela erminea*), European pine martens (*Martes martes*), stone martens (*Martes foina*), European otters (*Lutra lutra*), North American river otters (*Lontra canadensis*) and Eurasian badgers (*Meles meles*), and carnivorous species, including striped skunks (*Mephitis mephitis*), raccoons (*Procyon lotor*), bobcats (*Lynx rufus*) and common genets (*Genetta genetta*)¹¹.

The scenario is further complicated by the presence of feral minks, which have become established across Europe and America as a result of escape and deliberate releases by animal rights activists⁵. Additionally, in this case, whether AMDV is native to wild minks and was transmitted to captive mink populations or the virus originated in mink farms and wild animals were infected through accidental escapes or deliberate releases of infected captive animals is still a matter of discussion⁵. The respective roles of wild and domestic populations in infection maintenance and the prevalent direction of viral flux are similarly unknown.

Finally, the features of wild and domestic populations (e.g., effective contacts, population size and density, turnover, immunity, health status, and genetic background) could affect both viral population size and selective pressures acting on the pathogen, thus differentially contributing to AMDV evolution and new variant emergence.

The aim of the present work is to provide a comprehensive depiction of epidemiological AMDV patterns at the global scale and investigate how viral mutation, population size, the interaction with the host and geographical spreading have shaped AMDV evolution over time at the genomic level, using different biostatistic and phylodynamic approaches.

Results

Datasets. Based on the selected inclusion criteria and after the removal of poor-quality sequences (i.e., poorly aligned ones and those with frameshift mutations or premature stop codons), a dataset of partial NS1 (286 bases long) sequences, including 927 sequences with known collection dates (spanning the period 1963–2019) and countries (19 countries) and 689 sequences with known collection dates and hosts (5 hosts), was obtained. In the latter case, the status of the mink host (farmed or wild) was available for 623 out of 675 strains. The list of used sequences is provided in the Appendix.

Additionally, 94 complete NS1 sequences with known sampling hosts were included in the study. The phylogenetic signal was adequate for all datasets, including the randomly generated ones. The temporal signal, evaluated using the same databases, demonstrated a constantly positive correlation between root-to-tip distance and collection date, ranging between 0.27 and 0.38, and the estimated evolutionary rate was approximately 10^{-4} substitutions/site/year. Therefore, despite not being strong, the temporal signal was present and considered adequate for further analysis.

Viral population parameters and phylogeographic analysis. The AMDV MRCA was estimated to 1885.57 (median = 1895.32; 95 HPD = 1799.12–1933.67). Fully comparable values were obtained based on the ten randomly generated datasets (Fig. 1). Similarly, the overall analysis and the analyses based on random databases produced analogous evolutionary rate estimates (mean = 2.23×10^{-3} ; median = 2.23×10^{-3} ; 95 HPD = 1.4×10^{-3} – 3.7×10^{-3}) (Fig. 1). Marginal likelihood estimation suggested a relaxed molecular clock rather than a strict clock. The analysis of viral population dynamics revealed a consistent pattern, independent of the randomly included sequences, featuring a progressive increase in viral relative genetic diversity from the AMDV origin to approximately the 1990s, when a stationary phase began, followed by a slow decline at the beginning of the new millennium and a more abrupt decline in the last years of the study (Fig. 2).

The phylogeographic model evaluation did not support an asymmetric migration model over the symmetric model based on Bayesian Factor (BF) calculation, independent of the considered random dataset. Despite a certain degree of variability among runs in the among-country analysis of inferred significant migration rates, some consistent features could be identified (Fig. 3 and Supplementary Figure 1). With the exception of only Run 1, where the AMDV origin was estimated to be in North America, in all instances, an initial European source was predicted, involving Northern European countries in particular. Nevertheless, the posterior probability associated with the country of origin was constantly low (essentially between 0.5 and 0.8) (Supplementary Figure 2), demonstrating the uncertainty of this prediction. Several well-supported (BF > 10) migration rates were detected between countries and involved long-distance connections between North America and Europe and between Europe and China, as well as several within-Europe links. In all runs, independent connections of Europe with the USA and Canada were identified, linked particularly to Sweden, from which many additional significant migration paths with other European countries were proven, although a certain degree of variability in European connections occurred depending on the specific dataset (Fig. 3).

The overall geographic pattern featured different clusters of strains collected from the same country (Supplementary Figure 1), which is suggestive of multiple strain introductions in the same area, followed by local persistence, evolution and co-circulation.

Structured coalescent. The application of a structured coalescent model, considering the hosts of origin as different *demes*, provided concordant estimates compared to those of the previously described serial coalescent method in terms of tMRCA and evolutionary rates. The host of origin was predicted to be “Other wild animals” with a high posterior probability (i.e., > 0.9) (Supplementary Figure 3). Thereafter, several host switches/jumps

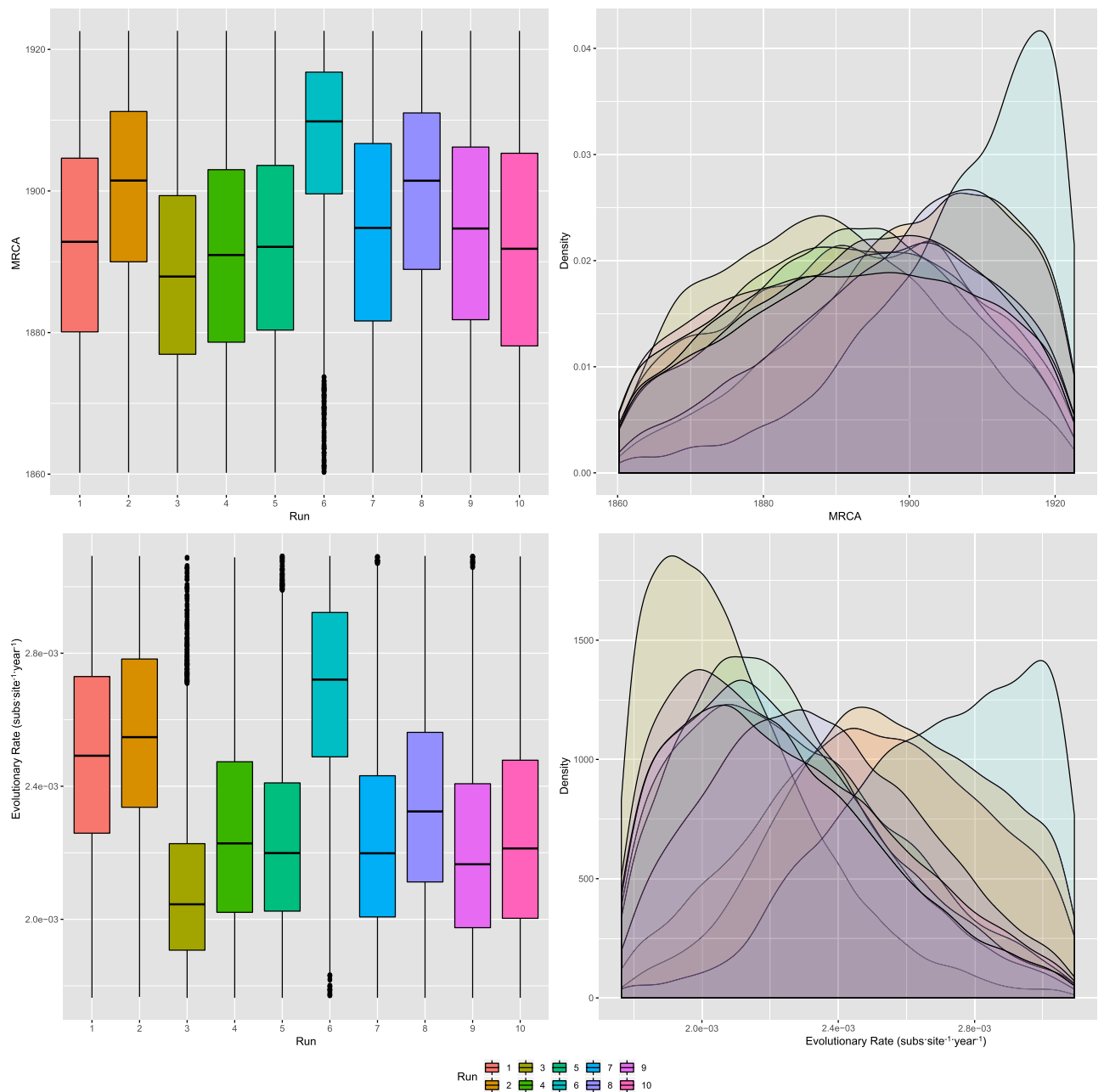


Figure 1. AMDV MRCA and evolutionary rate. Upper figure: Boxplot (left) and Densityplot (right) of the MRCA posterior probability. Lower figure: Boxplot (left) and Densityplot (right) of the mean evolutionary rate (expressed in base-10 logarithm) posterior probability. Results have been estimated performing ten independent runs based on randomly sampled sequences. The 95HPD intervals are reported for both figures.

were inferred, particularly between domestic minks and wild animals. More specifically, the estimated migration rates were as follows: farmed mink to wild mink = 2.88×10^{-2} , farmed mink to other wild animals = 1.71×10^{-2} , wild mink to other wild animals = 1.48×10^{-3} , other wild animals to farmed mink = 1.54×10^{-2} , other wild animals to farmed mink = 1.75×10^{-3} and other wild animals to wild mink = 2.44×10^{-3} .

Selective pressure analysis. The analysis of pervasive selective pressure revealed the presence of different sites under diversifying selection detected as significant by at least two implemented methods (Supplementary Figure 4 and Supplementary Table 1). Similarly, several codons were demonstrated to be under episodic diversifying selection by MEME (Supplementary Table 2). In both instances, homology modelling of the NS1 protein allowed the location of sites under diversifying selection on the protein surface (Fig. 4).

The comparison of selective pressure patterns acting on the strains collected from farmed and wild/feral minks demonstrated a significant difference in the proportion of sites under selection ($p < 0.001$) but not in the overall selection strength ($p = 0.854$) or selective regime ($p = 0.853$). Accordingly, the site-by-site comparison performed

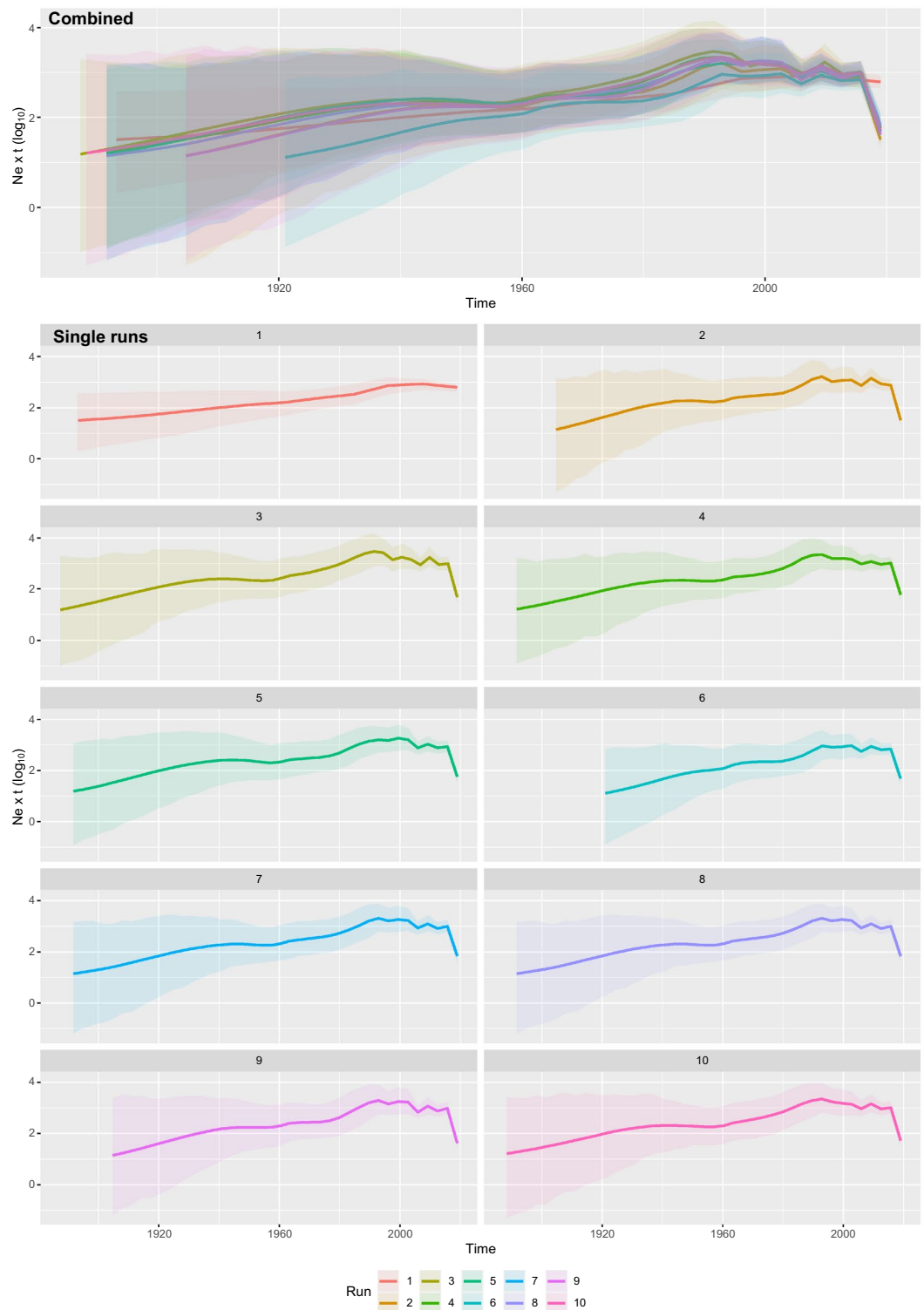


Figure 2. AMDV population dynamics. Mean relative genetic diversity over time estimated performing ten independent runs (colour-coded) based on randomly sampled sequences. Results have been reported both superimposed (upper figure) and individually (lower figures) for comparison purpose. The lines represent the mean value while the shaded areas depict the 95HPD ranges.

on the alignment of strains collected from domestic and wild/feral minks revealed 12 sites (i.e., codons 12, 27, 83, 120, 153, 211, 317, 345, 353, 393, 450 and 484) with differential selective pressure and exposed on the NS1

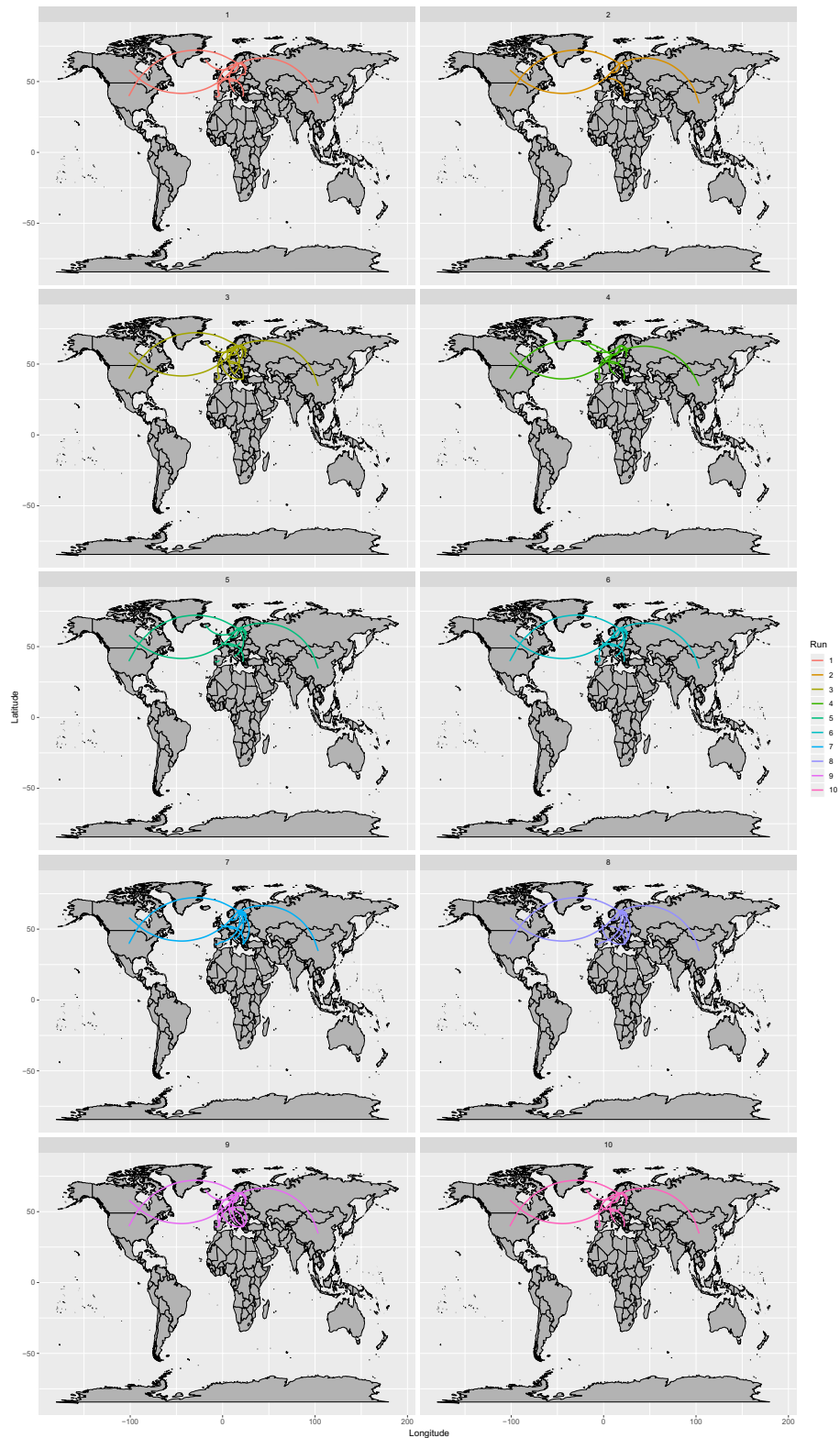


Figure 3. AMDV migration paths. Well supported migration paths (i.e. $BF > 10$) among countries are depicted. Different frames report the results of independent runs. Maps were generated using the ggplot⁵⁴ library in R.

surface (Fig. 4). For 8 of these codons, i.e., 16, 27, 83, 120, 211, 353, 393 and 484, the dN/dS ratio was higher in domestic animals. Finally, DEPS analysis highlighted 33 sites under episodic directional selection between wild

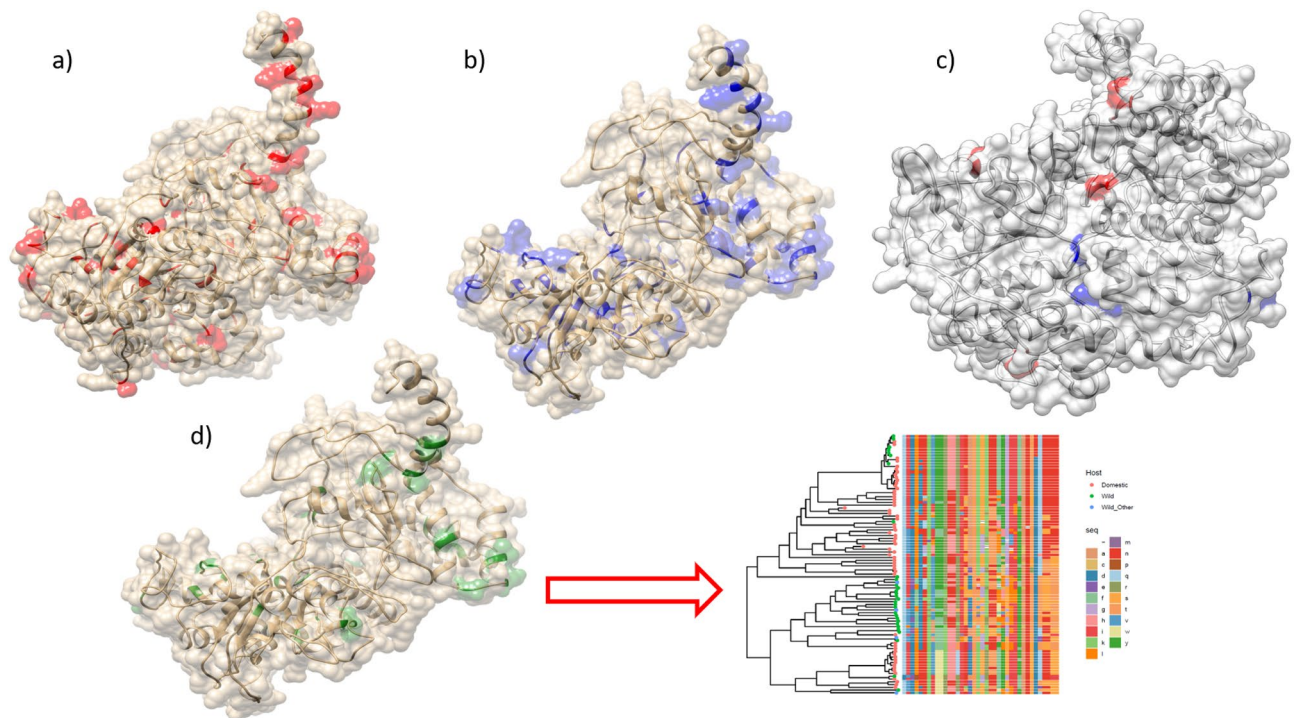


Figure 4. Tertiary structure of the NS1 protein estimated through homology modelling. Figures were generated using the Chimera software⁵³. Amino acids ribbon and relative surface (when exposed) have been colour-coded to depict the results of different selective pressures analyses: (a) sites detected under pervasive diversifying selection are highlighted in red; (b) sites detected under episodic diversifying selection are highlighted in blue; (c) sites detected under a significantly higher diversifying selection in domestic and wild minks are highlighted in red and blue, respectively; (d) (left figure) sites detected under episodic directional selection (wild subjects collected sequences have been used as foreground), (right figure) plot reporting the alignment of the NS1 protein amino acids under episodic directional selection with respect to their position in the phylogenetic tree (strains have been colour-coded according to the collection host with the highest posterior probability). To facilitate the evaluation of specific patterns, the latter plot is provided as Supplementary Figure 5 also.

and domestic minks (Supplementary Table 3). Additionally, in this case, the involved amino acids were located on the predicted surface of the NS1 protein (Fig. 4).

Discussion

AMDV is one of the most relevant pathogens of minks and mustelids worldwide. It can cause severe economic losses in commercially raised minks, especially since no vaccine is currently available and control must be based on eradication programs^{13,14}. However, viral features, including persistent infection and environmental resistance^{2,15}, make this approach challenging. The risk of new introductions from other farms or even wild animals further complicates the scenario. Conversely, viral spread from domestic to wild animals can pose a threat to these species, particularly endangered ones⁵. Despite their relevance, many of the aspects of AMDV epidemiology have not been quantitatively evaluated on a global scale. In contrast to several previous studies limited to single countries, locations or even individual farms^{6,15,16}, the present study attempts to complement their results by providing an overall depiction of AMDV molecular epidemiology and evolution from the AMDV origin to the present, benefitting from the coalescent theory robust framework, which allowed us to statistically support the performed analysis rather than rely on subjective evaluation of strain genetic distances and/or phylogenetic analysis. The viral MRCA was estimated to the end of the XIX century. Although previous studies have reported the absence of temporal signal in AMDV evolution⁶, when preliminarily evaluated using randomly generated datasets, a positive correlation was constantly detected between sampling date and root-to-tip distance, which suggests a certain temporal structure. Additionally, the relaxed molecular clock model was preferred over the strict model based on BF calculations. It was thus possible to conclude that AMDV evolution follows a temporal structure, compliant with a molecular-clock evolutionary model, although with significant variations over time and/or among lineages. Based on these considerations and the high repeatability of the results obtained using the randomly generated datasets and different analytical models, we considered our AMDV origin estimates reliable. It can therefore be concluded that AMDV originated well before its first identification as a clinical problem in fur minks. The actual host of origin is still largely unknown and a matter of discussion, with transmission from wild minks to domestic minks or domestic emergence and subsequent “escape” to the wild being the two debated scenarios^{5,6,17}. The present study estimation reveals a high probability of a wild origin, albeit in non-mink species. While unexpected, the scenario could be supported by the markedly high AMDV circulation in several wild

species¹⁷. Mink farming began in the late XIX century in North America⁵, approximately in the upper range of the tMRCA estimated by our analyses. Initial breeding stocks were acquired from locally caught wild minks; thus, clear compartmentalization between wild and domestic animals did not occur in the early phases, likely facilitating strain exchange in a “homogeneous” and connected wild/domestic environment. This environment, at the same time, might have created favourable circumstances for viral expansion due to nascent farming conditions. The viral population size showed a progressive increase over time until recent decades, which conforms to the proposed scenario. The subsequent stabilization and decrease in viral expansion could be attributed to the implementation of more effective control measures. For example, AMDV in Denmark has been subjected to a control campaign since 1976, but eradication has been supported by legislation since 1999¹³, when an actual effect on AMDV relative genetic diversity became clear (Fig. 2). The heterogeneity of control programmes around the world and continuous viral circulation in feral minks and wild animals can, on the other hand, explain the limited overall reduction compared to expectations based on Danish epidemiological data.

Of note, AMDV demonstrated a high evolutionary rate, in the upper range typical of ssDNA viruses¹⁸. From this perspective, AMDV appears more similar to other parvoviruses that underwent a host jump, such as canine parvovirus, than to others, such as feline parvovirus, that are well adapted to a specific host^{19,20}. This evidence could indicate progressive adaptation to new hosts or at least ecological niches, which could be consistent with our non-mink origin hypothesis.

However, other non-conflicting processes could have at least partially contributed to the observed high evolutionary rate, including the host population size, density in the farming setting and the enhanced transmission rate.

The analysis of selective pressures demonstrated the presence of a significant number of NS1 protein amino acids under pervasive and, even more, episodic diversifying selection, in agreement with the findings of other studies^{10,21}.

Homology modelling revealed that these amino acids were mostly exposed on the protein surface. Even if the absence of an experimentally determined tertiary structure demands caution, the observed scenario suggests a relevant role of immune-induced pressure in shaping NS1 evolution. Interestingly, the comparison of strains collected from wild and domestic animals yielded a significant difference in the selective force distribution, and the dN/dS ratio was proven to be different at several codon positions, being higher in domestic animals in 8 out of 12 instances. Evidence of directional selective pressures was also uncovered. Taken together, these results could be indicative of adaptation to new environmental conditions. Of note, also in these cases, the involved sites were exposed on the NS1 surface, which could reflect the need for a differential interaction with host proteins and/or the immune system. Variations in population immunity ascribable to differences in prevalence, host genetic background or immune system functionality (pathophysiologic) might be only some of the potentially involved factors.

However, the limited number of available complete NS1 sequences, especially from wild animals, does not allow us to confidently exclude the presence of a “founder” effect (i.e., when foreground and background strains in the phylogenetic tree (see “Material and methods”) are separated by a single or few branches) rather than actual directional selection²². In fact, when such a scenario occurs, it is almost impossible to discriminate whether the observed amino acid profiles are due to random and casual selection on an ancestor carrying a certain phenotype or to causal forces favouring the specific amino acid variant. Further investigations would help solve this issue when more sequences are available.

Nevertheless, after the revelation of a likely wild origin, the analysis of viral flux highlighted a preeminent directionality from domestic to wild populations, which fits well with the frequent introduction of domestic animals to the wild due to escapes or deliberate releases by animal rights movements^{23–25}. The presence of large wild-only clusters (Supplementary Figure 3) confirms the capacity of AMDV to persist and evolve over time in this environment independently of new introductions¹¹. Wild species should thus be considered more threatened by domestic species rather than a threat to those species. Nevertheless, as the viral migration rate from wild to farmed animals is not zero, it is clear that biosecurity measure improvement is still warranted and could contribute to the prevention of new outbreaks⁵.

Nevertheless, the present study and other studies demonstrated that the main risk is represented by the introduction of new strains from other farms¹³. An intense transmission network was reconstructed over long distances (i.e., different countries and continents), implying artificial movements of animals or other contamination sources such as fomites²⁶.

In particular, the link between North America and Northern European countries has already been proposed^{1,5}. Our estimates support a European origin, which poorly fits with historic epidemiological data: the mink was initially domesticated in North America and thereafter extensively exported to Northern Europe²⁷. Therefore, a North American origin of AMDV, followed by exportation to Europe, would appear more likely. The limited number of old sequences jeopardizes the ability to estimate such old spreading events. Long branches and the lack of historical data largely prevent inference of the spatial history of older viral lineages with high confidence. Additional spatial movements between multiple locations are likely concealed in the “long tree branches” during this time frame²⁸. Accordingly, also from a statistical perspective, the symmetric migration model was preferred over the asymmetric model, highlighting the lack of evidence supporting significant directionality of the viral flux between country pairs. Above all, the disproportion of sequences collected from Europe could have biased the results, despite our attempts to create more balanced datasets by random sampling. Therefore, based on these considerations and with caution, migration patterns should be evaluated in terms of “contact” among countries, avoiding overinterpretation of their directionality.

European countries were also involved in most of the mink trades and AMDV spreading events, as previously suggested by other authors^{6,11,15}. Interestingly, the mink farms in some of these countries were fully integrated into the Soviet Union’s mink-farming industry before independence, which could have contributed to eastward

viral dispersal¹¹. In China, mink farming started in the 1950s, likely with imports from the Soviet Union²⁹, which fits with our estimate of AMDV introduction to this country in the late 1950s. Unfortunately, Asian sequences with adequate features for inclusion in our dataset were essentially limited to China, preventing higher resolution of the viral spread to and among oriental countries. In contrast, the abundance of European sequences allowed us to demonstrate an intense diffusion of AMDV within this area. Frequent introduction events in different countries were identified and followed by local persistence and evolution, which suggests the limited efficacy of biosecurity measures both among and within countries, similar to what has been described for several other livestock infections^{30–32}. Likely, a comparable scenario could be true for other regions, including Asia.

The present study provides several details regarding AMDV epidemiology and evolution, which appear to occur in multiple environments linked by sporadic contact. AMDV seems to persist and evolve in different wild or domestic hosts, even for a noteworthy time, sporadically jumping to other hosts and beginning a new pathway¹¹. Similarly, despite frequent and multiple introductions to new geographic areas that characterize AMDV epidemiology, local persistence and evolution also occur. These conditions likely explain some of the peculiarities of the viral strains, as well as the high but heterogeneous evolutionary rate of AMDV, which periodically switches between stable and variable environmental conditions. From a practical perspective, the limited efficacy of currently applied control measures in preventing viral passage among countries and farms and contact between wild and domesticated animals were also demonstrated. Such conditions pose a double threat to farm profitability along with animal welfare and health, which is particularly relevant for endangered species. Further efforts must thus be made to limit such viral spreading, and dedicated studies should be performed to refine our knowledge of AMDV spreading risk/enhancing factors, particularly at the wild-domestic interface.

Material and methods

Sequence dataset preparation. Freely available NS1 sequences were considered in the present study, aiming to reconstruct AMDV population dynamics, phylogeography and among-host transmission.

Only a limited number of complete NS1 sequences, properly annotated with metadata of interest (i.e. collection date, sampling host and country) were available. To deal with this data shortage and maximize available information, different datasets were created according to the main purpose of the specific analysis, and the partial NS1 region with the highest coverage of available sequences was selected.

Therefore, only sequences with available collection date were selected to estimate evolution rate and population dynamics. Similarly, phylogeographic and host transmission analyses were performed separately on sequences with known sampling date, collection country and host, respectively.

For each database, recombinant sequences were identified using RDP4³³ and excluded from further analysis. The RDP4 settings for each method were adjusted to account for the dataset features according to the RDP manual recommendations. In particular, RDP, GENECONV, Chimaera and 3Seq were used in a primary scan while the full set of available methods was used for the analysis refinement. Only recombination events detected by more than 2 methods with a significance value lower than 10^{-5} (p-value $< 10^{-5}$) and Bonferroni correction were accepted. The absence of residual recombination signal was assessed by GARD³⁴. The presence of an adequate phylogenetic signal despite the relatively short selected sequence region was assessed using the likelihood mapping approach implemented in IqTree³⁵. A maximum likelihood (ML) phylogenetic tree was reconstructed using the same software selecting as substitution model the one with the lowest Akaike information criterion (AIC) and used to preliminarily evaluate the temporal signal using TempEst³⁶.

A dataset of complete NS1 sequences was also downloaded and processed as previously described.

Sequence analysis. *Viral population parameters and phylogeographic analysis.* The evolutionary rate, time to the most recent common ancestor (tMRCA) and population dynamics were jointly estimated in a Bayesian fashion using the serial coalescent approach in BEAST 1.8.4³⁷. The best substitution model (i.e. GTR + G) was selected based on the Bayesian Information Criterion (BIC) calculated using Jmodeltest 2³⁸, while the relaxed lognormal molecular clock³⁹ was preferred over the strict one based on the BF value, calculated through the marginal likelihood estimation using the Path Sampling and Stepping Stone methods, as described by Baele et al.⁴⁰.

The non-parametric skygrid⁴¹ model was selected to account for and reconstruct the viral population dynamics.

A comparable approach was implemented on the dataset of the sequences with available collection country to reconstruct the viral spreading over time using the discrete state phylogeographic approach described by Lemey et al.⁴². The symmetric migration rate was preferred over the asymmetric one based on the marginal likelihood calculation of the two models⁴⁰. The Bayesian Stochastic Search Variable Selection (BSSVS) was also implemented, allowing the definition of the most parsimonious viral dispersal path and the identification of the statistically supported migration rates between countries through BF testing⁴². The sequence availability differed significantly among countries and time periods, potentially biasing the analysis results. For this reason, attempting to control and limit this phenomenon, 10 more balanced dataset were created by randomly selecting without replacement up to a maximum of 10 sequences for each country-year pair.

For each dataset, 200 million generations-long runs were performed, sampling trees and parameters every 20 thousand generations. Run results were accepted only if Estimate Sample Size (ESS) exceeded 200 and mixing and convergence, evaluated by visually inspecting the trace plots using Tracer 1.6, were adequate. Population parameters were summarized as mean and 95% Highest Posterior Density (95HPD) after discarding the first 20% of the run as burn-in. The annotated maximum clade credibility tree was generated using Treeannotator, after discarding the first 20% of the posterior trees.

Overall, a $BF > 10$ was considered adequate to prefer the more complex model over the simpler ones. Similarly, migration rates among countries were considered non-zero (i.e. well supported) when the BF, calculated using SPREAD3⁴³ was greater than 10.

Structured coalescent. The connections and viral flux featuring wild and domestic animals were evaluated using a structured coalescent analysis. Viral strains were categorized as originating from “Farmed mink”, “Wild mink” and “Other wild animals” based on information present in Genbank or literature search.

Briefly, according to structured coalescent, each of these groups was treated as a separate *deme* with a certain size, connected to other *demes* through strain migration occurring at a certain rate. The advantage of this model over discrete trait analysis is that migration events affecting lineages are explicitly parameterized and estimated, avoiding the potentially biasing assumption that sampling intensity is proportional to subpopulation size^{44,45}. Such an assumption can be particularly strong when dealing with wild animals, whose sampling is especially challenging and economical relevance is typically marginal, further reducing the related research activities.

The partial NS1 sequences, for which the collection host category was available were analyzed using the structured coalescent approach as implemented in the package BASTA⁴⁵ of BEAST2.5⁴⁶. The substitution and clock models were respectively selected based on the BIC and BF values (calculated through estimation of the marginal likelihood using the Path Sampling and Stepping Stone methods). A MCMC run of 200 million generations was performed sampling trees and parameters every 20 thousand generations. Run results were accepted only if ESS exceeded 200 and mixing and convergence, evaluated by visually inspecting the trace plots using Tracer 1.6, were adequate. Population parameters were summarized as mean and 95HPD after discarding the first 20% of the run as burn-in. The annotated maximum clade credibility tree was generated using Treeannotator, after discarding the first 20% of the posterior trees.

Selective pressure analysis. The dataset of complete NS1 sequences was used to estimate sites under different evolution pressures applying different methods, all based on the comparison of relative rates of non-synonymous (dN) to synonymous (dS) substitutions. The SLAC, FEL and FUBAR methods^{47,48} were selected to estimate codons under pervasive purifying or diversifying selection, while the MEME method⁴⁹ was used to evaluate the occurrence of episodic diversifying selection. All methods are implemented in the HYPHY software⁵⁰. The significance level was set to $p < 0.05$ for SLAC, FEL and MEME, and to a posterior probability higher than 0.9 for FUBAR.

The comparison of selective forces acting on wild and domestic animals was performed with different statistical approaches. The presence of an overall different selection strength, proportion of sites under selection and selective regime (both strength and proportion) between alignments of strains collected from farmed and wild minks was assessed using the *dNdSDistributionComparison.bf* function of the HYPHY package. Differences in the site by site-selective pressure strength between the two alignments were also evaluated using the *Compare-SelectivePressure.bf* function of the same package. Finally, the occurrence of episodic directional selection was tested with MEDS⁵¹, considering phylogenetic tree branches leading to strains collected from wild-animals as foreground.

To tentatively evaluate the spatial distribution of domains under selective pressures, the tertiary structure of the NS1 was estimated through homology modelling, using Phyre2⁵². The obtained model was edited and visualized using Chimera⁵³.

Data availability

Accession number and relevant metadata of uses sequences are available in Appendix.

Received: 27 March 2021; Accepted: 3 August 2021

Published online: 12 August 2021

References

- Ryt-Hansen, P. *et al.* Global phylogenetic analysis of contemporary aleutian mink disease viruses (AMDVs). *Viol. J.* **14**, 1–6 (2017).
- Best, S. M. & Bloom, M. E. Pathogenesis of Aleutian mink disease parvovirus and similarities to B19 infection. *J. Vet. Med. Ser. B* **52**, 331–334 (2005).
- Porter, D. D., Larsen, A. E. & Porter, H. G. Aleutian disease of mink. *Adv. Immunol.* **29**, 261–286 (1980).
- Alexandersen, S. *et al.* Acute interstitial pneumonia in mink kits inoculated with defined isolates of Aleutian mink disease parvovirus. *Vet. Pathol.* **31**, 216–228 (1994).
- Nituch, L. A., Bowman, J., Wilson, P. & Schulte-Hostedde, A. I. Molecular epidemiology of Aleutian disease virus in free-ranging domestic, hybrid, and wild mink. *Evol. Appl.* **5**, 330–340 (2012).
- Knuutila, A., Uzcátegui, N., Kankkonen, J., Vapalahti, O. & Kinnunen, P. Molecular epidemiology of Aleutian mink disease virus in Finland. *Vet. Microbiol.* **133**, 229–238 (2009).
- Canuti, M., Whitney, H. G. & Lang, A. S. Amdoparvoviruses in small mammals: Expanding our understanding of parvovirus diversity, distribution, and pathology. *Front. Microbiol.* **6**, 1–9 (2015).
- Kanno, H., Wolfenbarger, J. B. & Bloom, M. E. Aleutian mink disease parvovirus infection of mink macrophages and human macrophage cell line U937: Demonstration of antibody-dependent enhancement of infection. *J. Virol.* **67**, 7017–7024 (1993).
- Canuti, M. *et al.* Driving forces behind the evolution of the Aleutian mink disease parvovirus in the context of intensive farming. *Virus Evol.* **2**, 1–17 (2016).
- Olofsson, A. *et al.* Unusual, high genetic diversity of Aleutian mink disease virus. *J. Clin. Microbiol.* **37**, 4145–4149 (1999).
- Leimann, A., Knuutila, A., Maran, T., Vapalahti, O. & Saarma, U. Molecular epidemiology of Aleutian mink disease virus (AMDV) in Estonia, and a global phylogeny of AMDV. *Virus Res.* **199**, 56–61 (2015).
- Ryt-Hansen, P. *et al.* Outbreak tracking of Aleutian mink disease virus (AMDV) using partial NS1 gene sequencing. *Viol. J.* **14**, 119 (2017).

13. Christensen, L. S., Gram-Hansen, L., Chriél, M. & Jensen, T. H. Diversity and stability of Aleutian mink disease virus during bottleneck transitions resulting from eradication in domestic mink in Denmark. *Vet. Microbiol.* **149**, 64–71 (2011).
14. Cho, H. J. & Greenfield, J. Eradication of Aleutian disease of mink by eliminating positive counterimmunoelectrophoresis test reactors. *J. Clin. Microbiol.* **7**, 18–22 (1978).
15. Prieto, A. *et al.* Molecular epidemiology of Aleutian mink disease virus causing outbreaks in mink farms from Southwestern Europe: A retrospective study from 2012 to 2019. *J. Vet. Sci.* **21**, 1–13 (2020).
16. Hagberg, E. E., Pedersen, A. G., Larsen, L. E. & Krarup, A. Evolutionary analysis of whole-genome sequences confirms inter-farm transmission of Aleutian mink disease virus. *J. Gen. Virol.* **98**, 1360–1371 (2017).
17. Farid, A. H. Aleutian mink disease virus in furbearing mammals in Nova Scotia, Canada. *Acta Vet. Scand.* **55**, 10 (2013).
18. Duffy, S., Shackelton, L. A. & Holmes, E. C. Rates of evolutionary change in viruses: Patterns and determinants. *Nat. Rev. Genet.* **9**, 267–276 (2008).
19. Tucciarone, C. M. *et al.* Molecular insight into Italian canine parvovirus heterogeneity and comparison with the worldwide scenario. *Infect. Genet. Evol.* **66**, 171–179 (2018).
20. Franzo, G., Tucciarone, C. M., Cecchinato, M. & Drigo, M. Canine parvovirus type 2 (CPV-2) and Feline panleukopenia virus (FPV) codon bias analysis reveals a progressive adaptation to the new niche after the host jump. *Mol. Phylogenet. Evol.* **114**, 82–92 (2017).
21. Gottschalck, E., Alexandersen, S., Storgaard, T., Bloom, M. E. & Aasted, B. Sequence comparison of the non-structural genes of four different types of Aleutian mink disease parvovirus indicates an unusual degree of variability. *Adv. Virol.* **138**, 213–231 (1994).
22. Kosakovsky Pond, S. L., Poon, A. F. Y., Leigh Brown, A. J. & Frost, S. D. W. A maximum likelihood method for detecting directional evolution in protein sequences and its application to influenza A virus. *Mol. Biol. Evol.* **25**, 1809–1824 (2008).
23. Bonesi, L. & Palazon, S. The American mink in Europe: Status, impacts, and control. *Biol. Conserv.* **134**, 470–483 (2007).
24. Wildhagen, A. Present distribution of North American mink in Norway. *J. Mammal.* **37**, 116 (1956).
25. Gerell, R. Dispersal and acclimatization of the mink (*Mustela vison* Schreb.) in Sweden. (1967).
26. Prieto, A. *et al.* Distribution of Aleutian mink disease virus contamination in the environment of infected mink farms. *Vet. Microbiol.* **204**, 59–63 (2017).
27. Mañas, S. *et al.* Aleutian mink disease parvovirus in wild riparian carnivores in Spain. *J. Wildl. Dis.* **37**, 138–144 (2001).
28. Nelson, M. I. *et al.* Global migration of influenza A viruses in swine. *Nat. Commun.* **6**, 6696 (2015).
29. Leng, X. *et al.* Genetic diversity and phylogenetic analysis of Aleutian mink disease virus isolates in north-east China. *Adv. Virol.* **163**, 1241–1251 (2018).
30. Franzo, G. *et al.* Free to circulate: An update on the epidemiological dynamics of porcine circovirus 2 (PCV-2) in Italy reveals the role of local spreading, wild populations, and Foreign countries. *Pathogens* **9**, 221 (2020).
31. Franzo, G. *et al.* Think globally, act locally: Phylogenetic reconstruction of infectious bronchitis virus (IBV) QX genotype (GI-19 lineage) reveals different population dynamics and spreading patterns when evaluated on different epidemiological scales. *PLoS ONE* **12**, e0184401 (2017).
32. Franzo, G., Cortey, M., Segalés, J., Hughes, J. & Drigo, M. Phylogenetic analysis of porcine circovirus type 2 reveals global waves of emerging genotypes and the circulation of recombinant forms. *Mol. Phylogenet. Evol.* **100**, 269–280 (2016).
33. Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1**, 1–5 (2015).
34. Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H. & Frost, S. D. W. GARD: A genetic algorithm for recombination detection. *Bioinformatics* **22**, 3096–3098 (2006).
35. Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44**, W232–W235 (2016).
36. Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).
37. Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
38. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. JModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* **9**, 772 (2012).
39. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).
40. Baele, G. *et al.* Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol. Biol. Evol.* **29**, 2157–2167 (2012).
41. Gill, M. S. *et al.* Improving Bayesian population dynamics inference: A coalescent-based model for multiple loci. *Mol. Biol. Evol.* **30**, 713–724 (2013).
42. Lemey, P., Rambaut, A., Drummond, A. J. & Suchard, M. A. Bayesian phylogeography finds its roots. *PLoS Comput. Biol.* **5**, e1000520 (2009).
43. Bielejec, F. *et al.* Sprea D3: Interactive visualization of spatiotemporal history and trait evolutionary processes. *Mol. Biol. Evol.* **33**, 2167–2169 (2016).
44. Vaughan, T. G., Kühnert, D., Poppinga, A., Welch, D. & Drummond, A. J. Efficient Bayesian inference under the structured coalescent. *Bioinformatics* **30**, 2272–2279 (2014).
45. De Maio, N., Wu, C.-H., O'Reilly, K. M. & Wilson, D. New routes to phylogeography: A Bayesian structured coalescent approximation. *PLoS Genet.* **11**, e1005421 (2015).
46. Bouckaert, R. *et al.* BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).
47. Kosakovsky Pond, S. L. & Frost, S. D. W. Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* **22**, 1208–1222 (2005).
48. Murrell, B. *et al.* FUBAR: A fast, unconstrained bayesian AppRoximation for inferring selection. *Mol. Biol. Evol.* **30**, 1196–1205 (2013).
49. Murrell, B. *et al.* Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **8**, e1002764 (2012).
50. Kosakovsky Pond, S. L., Frost, S. D. W. & Muse, S. V. HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679 (2005).
51. Murrell, B., de Oliveira, T., Seebregts, C., Kosakovsky Pond, S. L. & Scheffler, K. Modeling HIV-1 drug resistance as episodic directional selection. *PLoS Comput. Biol.* **8**, e1002507 (2012).
52. Kelley, L. A. & Sternberg, M. J. E. Protein structure prediction on the web: A case study using the phyre server. *Nat. Protoc.* **4**, 363–373 (2009).
53. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
54. Ginestet, C. ggplot2: Elegant graphics for data analysis. *J. R. Stat. Soc. A. Stat. Soc.* **174**, 245–246 (2011).

Author contributions

G.F. wrote the majority of the manuscript and communicated with the coauthors to coordinate the document editing. G.F., C.M.T., M.L., G.D. and L.G. prepared the datasets and performed the literature search. G.F. performed the statistical and phylogenetic analysis. G.F. designed the outline of the study. M.C. and M.D. supervised the study and participated to result interpretation. All the authors revised and edited the manuscript and approved the final version for its publication.

Funding

The authors received no specific funding for this work.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-96025-z>.

Correspondence and requests for materials should be addressed to G.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021