

# Driving forces behind the evolution of the Aleutian mink disease parvovirus in the context of intensive farming

Marta Canuti,<sup>1,\*</sup>† Kimberly E. O’Leary,<sup>2</sup> Bruce D. Hunter,<sup>3,‡</sup> Grant Spearman,<sup>4</sup> Davor Ojkic,<sup>5</sup> Hugh G. Whitney,<sup>2</sup> and Andrew S. Lang<sup>1,\*</sup>

<sup>1</sup>Department of Biology, Memorial University of Newfoundland, 232 Elizabeth Ave., St. John’s, Newfoundland and Labrador, A1B 3X9, Canada, <sup>2</sup>Animal Health Division, Forestry and Agrifoods Agency, P.O. Box 7400, St. John’s, Newfoundland and Labrador, A1E 3Y5, Canada, <sup>3</sup>Department of Pathobiology, Ontario Veterinary College, University of Guelph, Guelph, Ontario, N1G 2W1, Canada, <sup>4</sup>Department of Agriculture, Animal Health Laboratory, 65 River Rd., Truro, Nova Scotia, B2N 5E3, Canada and <sup>5</sup>Animal Health Laboratory, 419 Gordon Street, University of Guelph, Guelph, Ontario, N1G 2W1, Canada

\*Corresponding authors: E-mail: marta.canuti@gmail.com; aslang@mun.ca

†<http://orcid.org/0000-0002-9959-128X>

‡Deceased.

## Abstract

Aleutian mink disease virus (AMDV) causes plasmacytosis, an immune complex-associated syndrome that affects wild and farmed mink. The virus can also infect other small mammals (e.g., ferrets, skunks, ermines, and raccoons), but the disease in these hosts has been studied less. In 2007, a mink plasmacytosis outbreak began on the Island of Newfoundland, and the virus has been endemic in farms since then. In this study, we evaluated the molecular epidemiology of AMDV in farmed and wild animals of Newfoundland since before the beginning of the outbreak and investigated the epidemic in a global context by studying AMDV worldwide, thereby examining its diffusion and phylogeography. Furthermore, AMDV evolution was examined in the context of intensive farming, where host population dynamics strongly influence viral evolution. Partial NS1 sequences and several complete genomes were obtained from Newfoundland viruses and analyzed along with numerous sequences from other locations worldwide that were either obtained as part of this study or from public databases. We observed very high viral diversity within Newfoundland and within single farms, where high rates of co-infection, recombinant viruses and polymorphisms were observed within single infected individuals. Worldwide, we documented a partial geographic distribution of strains, where viruses from different countries co-exist within clades but form country-specific subclades. Finally, we observed the occurrence of recombination and the predominance of negative selection pressure on AMDV proteins. A surprisingly low number of immunoeptopic sites were under diversifying pressure, possibly because AMDV gains no benefit by escaping the immune response as viral entry into target cells is mediated through interactions with antibodies, which therefore contribute to cell infection. In conclusion, the high prevalence of AMDV in farms facilitates the establishment of co-infections that can favor the occurrence of recombination and enhance viral diversity. Viruses are then exchanged between different farms and countries and can be introduced into the wild, with the rapidly evolving viruses producing many parallel lineages.

**Key words:** Aleutian mink disease virus; amdoparvoviruses; parvoviruses; viral evolution; viral recombination; AMDV.

## 1. Introduction

Aleutian mink disease virus (AMDV), also known as *Carnivore amdoparvovirus 1*, is the first-discovered member of the genus *Amdoparvovirus* within the family *Parvoviridae* and subfamily *Parvovirinae*. This genus includes a group of viruses able to infect various terrestrial carnivores (Canuti, Whitney, and Lang 2015). The primary hosts of AMDV are American mink (*Neovison vison*) and European mink (*Mustela lutreola*) but other mustelids and small animals are also frequently infected (Mañas et al. 2001; Farid et al. 2012; Knuutila et al. 2015). Three other species within the genus *Amdoparvovirus* have been discovered recently and have so far only been identified in foxes and/or raccoon dogs; these are the gray fox amdovirus (GFAV) (Li et al. 2011), the raccoon dog and fox amdoparvovirus (RFAV) (Shao et al. 2014), and the red fox fecal amdovirus (Bodewes et al. 2014). However, a wider host range for these viruses is plausible and other AMDV-related but not yet discovered species likely exist (Canuti, Whitney, and Lang 2015).

Parvovirus virions are composed of a protein capsid encompassing a monosense, single-stranded DNA molecule. The genome contains two main open reading frames (ORFs) flanked by terminal untranslated regions that can fold into hairpins and mediate viral DNA replication (Cotmore and Tattersall 2014). The ORF located at the 5'-side of the genome encodes the major non-structural protein NS1 and two other smaller proteins, NS2 and NS3, which are generated by alternative splicing (Huang et al. 2014). The 3'-ORF encodes the two capsid proteins, VP1 and VP2, which share partial amino acid (aa) sequence but VP1 has approximately forty additional N-terminal residues, the VP1 unique region (VP1u) (Bloom et al. 1997).

In adult mink, AMDV causes an immune complex-associated progressive syndrome, called plasmacytosis or Aleutian disease, characterized by weight loss, hypergammaglobulinemia, necrotizing arteritis, and kidney complications (Porter, Larsen, and Porter 1973). Virus-antibody complexes allow viral particles to penetrate into their target cells, circulating macrophages, facilitating viral replication (Kanno, Wolfenbarger, and Bloom 1993), and deposit in tissues leading to arteritis and glomerulonephritis (Porter, Larsen, and Porter 1973). In kits, AMDV infects cells in the lungs causing a fulminant interstitial pneumonia (Alexandersen 1986).

The infection is widespread in wild as well as in farmed animals (Canuti, Whitney, and Lang 2015). In farmed mink, AMDV infection is associated with high mortality, reduced pregnancy rates, decreased litter size, and abortion (Broll and Alexandersen 1996) and results in severe economic consequences. Vaccination as a preventive measure is not feasible because of the peculiar pathogenic mechanism of AMDV, and the only possible approach to eliminate the virus from an affected farm is to implement eradication measures, which consist of identifying infected animals and culling them (Cho and Greenfield 1978; Christensen et al. 2011). However, eradication is difficult because bodily fluids from infected animals contain viral particles that can resist inactivation and persist in the environment (Canuti, Whitney, and Lang 2015). Furthermore, maintaining a disease-free status can also be challenging because the reintroduction of the virus from the outside, via commercial routes or from the wild where the virus is not eradicable, can quickly reestablish epidemics (Gunnarsson 2001). At the same time, farms are also a source of viruses for wild animals, after accidental escape or deliberate release of infected animals, where AMDV has the potential for detrimental

effects on wild animal populations (Mañas et al. 2001; Fournier-Chambrillon et al. 2004). Finally, as a consequence of animal trading within the fur industry, the virus, which presumably originated in North America, is currently distributed worldwide (Canuti, Whitney, and Lang 2015).

Not much is known about the evolutionary dynamics driving genetic changes in amdoparvoviruses, but presumably some of the same processes characterizing the evolution of other parvoviruses are also involved. Single-stranded DNA viruses are characterized by high levels of genetic diversity and evolve at rates approaching those observed in RNA viruses (Duffy, Shackelton, and Holmes 2008), and for parvoviruses, this has been estimated to be approximately  $10^{-4}$  substitutions/site/year (Shackelton et al. 2005; Shackelton and Holmes 2006; Zehender et al. 2010). Another mechanism that increases genetic variation in parvoviruses is recombination, whereby chimeric genomes are generated after the simultaneous infection of the same cell by two different strains (Shackelton et al. 2007; Ohshima and Mochizuki 2009; Tyumentsev et al. 2014). This evolutionary potential combined with an already existing high viral diversity are likely linked to the emergence of strains with novel characteristics (e.g., increased pathogenicity or virulence) and the ability to replicate in novel hosts (Moya, Holmes, and González-Candelas 2004; Duffy, Shackelton, and Holmes 2008), as already documented for the emergence of canine parvovirus (Shackelton et al. 2005).

Viral evolution is also influenced by host population dynamics. Intensive farming practices, where a large number of animals live in very close contact and in fairly restricted areas, create favorable conditions for disease spread, with high local host density offering the perfect opportunity for efficient host-to-host transmission (Mennerat et al. 2010). Additionally, the high turnover and the shorter lifespan of hosts in these settings favor further viral transmission due to the continual availability of new susceptible individuals. These dynamics are likely to impact the efficiency of viral spread in terms of basic reproductive number  $R_0$  (average number of secondary infections stemming from a single infected host). An  $R_0$  above 1 is needed for substantial transmission and high host densities facilitate the selection for a higher  $R_0$ . In these conditions, higher virulence and faster within-host growth rate (i.e., shorter generation time), and therefore a more rapid diversification, are also favored (Elena and Sanjuán 2005; Duffy, Shackelton, and Holmes 2008; Mennerat et al. 2010). It is therefore important to study and understand the basic mechanisms of viral evolutionary change in these circumstances.

An outbreak of Aleutian disease involving ten mink farms started in 2007 on the Island of Newfoundland, a large island off the east coast of Canada, and offered an opportunity to study AMDV evolutionary dynamics in the context of intensive farming. We have studied the molecular epidemiology and genetic features of AMDV from farmed and wild animals of Newfoundland over a period of 11 years, from before the beginning of the outbreak until 2014. Furthermore, by comparing these viral sequences to those identified worldwide, we have analyzed the global AMDV epidemic and studied the underlying viral evolutionary dynamics.

## 2. Materials and Methods

### 2.1 Sample collections

The sequences obtained in this study came from viruses circulating between 2004 and 2014 at various sites on the Island of Newfoundland, other parts of North America, and Europe.

Carcasses from suspected AMDV-positive mink from Newfoundland were obtained from ten different farms in 2007–9 and again in 2014 from one of these farms. Samples from farmed mink in Nova Scotia, Ontario, Wisconsin, and Denmark were also included. Samples were collected in 2014 from various wild animals: ten American mink, sixteen ermine (*Mustela erminea*), two Newfoundland lynx (*Lynx canadensis subsolanus*), nineteen red foxes (*Vulpes vulpes deletrix*), and forty coyotes (*Canis latrans*). Additionally, sequences obtained from AMDV-positive wild mink collected in 2004 and in 2007–8 were also included. These sequences originated from two pilot studies showing AMDV-positive levels of 14 per cent (9/64) and 45.2 per cent (51/93) for samples collected from wild mink across Newfoundland before and after the beginning of the farm epidemic, respectively (Bradley 2005). During the 2007–8 study samples from six ermine, twenty-nine Newfoundland pine marten (*Martes americana atrata*), and six American red squirrel (*Tamiasciurus hudsonicus*) tested AMDV-negative (H.G. Whitney, unpublished data).

Spleens were used as the source of material for virus characterization. Each spleen collected between 2007–9 was stabbed in several locations with cotton-tipped swabs that were then vortexed in 500 µl tryptose phosphate broth (tryptose 20 g/l, dextrose 5 g/l, NaCl 5 g/l, disodium phosphate 2.5 g/l; pH = 7.3). For the 2014 samples, a 10 mg piece was taken from each spleen.

## 2.2 Screening and sequencing

Swab suspensions or spleen tissues were used for DNA isolation and molecular screening following two different procedures. DNA was extracted from 200-µl aliquots of tissue swab suspensions by using the MagNA Pure LC instrument (Roche Diagnostics) and the MagNA Pure LC total Nucleic Acid Isolation Kit (Roche Diagnostics) and screened for AMDV by polymerase chain reaction (PCR) with the primer pair ADV-1207..1239 (5'-KTTGGTTGCTTTACTCC-3')/ADV-1707..1688 (5'-RTCTACTTTTATCATCACCAC-3'). DNA sequences of PCR products were determined at the University of Guelph Laboratory Services sequencing facility by Sanger sequencing with an Applied Biosystems 3730 DNA analyzer.

DNA was extracted from the spleen tissues with the DNeasy Blood and Tissue Kit (Qiagen) and screened for AMDV by PCR with two primer pairs. Three novel members of the genus *Amdoparvovirus* were described between 2011 and 2014 (Li et al. 2011; Bodewes et al. 2014; Shao et al. 2014), so the primer sequences were updated to detect all known amdoparvoviruses. Those primers, GF-F (5'-GACAACRAACCAACCAAAG-3')/GF-R (5'-CCHAMSMACAGTGAATATG-3') and ScF (5'-TGTTGCTTTACTCCAGAAG-3')/ScR (5'-WCCWCCTCCAGTAATRGC-3'), were designed to amplify two different portions of the NS1 gene (sequence positions 329–597 and 1207–1690, respectively, of the strain AMDV-G, accession number JN040434). Both primer pairs ADV-1207..1239/ADV-1707..1688 and ScF/ScR are able to bind to all known AMDVs. Amplified products from positive ScF-ScR PCRs were purified with AMPure Beads (Beckman Coulter), cloned into pGEM<sup>®</sup>-T Easy Vector (Promega), and up to ten clones for each positive PCR were sequenced at The Center for Applied Genomics (Toronto, Canada) by Sanger sequencing using an Applied Biosystems 3730XL DNA analyzer.

Coding sequences from a selection of strains containing multiple viral variants (two samples containing putative recombinant viruses and a third randomly chosen sample containing a mixed infection of viruses from the major identified clades) were amplified, cloned, and sequenced in the same way.

Complete genome sequencing for a selection of strains (three randomly selected viruses from single infections identified in farmed mink representing every major group and one virus identified in one wild mink) was achieved by amplifying the entire viral coding sequence with two overlapping hemi-nested PCRs that were subjected to direct sequencing using internal primers. Sequences of all primers used in this study are available in the [Supplementary Table S1](#).

## 2.3 Phylogenetic analyses

The dataset of sequences obtained for this study included 174 partial NS1 sequences from viruses detected between 2004 and 2014 in Newfoundland, other parts of North America (Nova Scotia, Ontario and Wisconsin), and Europe (Denmark). Sequences from Newfoundland originated from both wild mink and from ten different farms. A subset of 131 sequences was selected by excluding sequences that presented observable double peaks on chromatograms, recombinant sequences (identified using RDP software, see below), and all identical sequences derived from the same year and the same geographical area (same farm for Newfoundland) and used for phylogenetic inference. For each sample collected in 2014, only one representative clonal sequence for every strain was included. Reference sequences included only viruses for which the complete coding sequence of one or both ORFs were available and AMDV-like viruses (RFAV and GFADV) identified in foxes and raccoon dogs (Bloom et al. 1988; Gottschalck et al. 1991, 1994; Oie et al. 1996; Schuierer et al. 1997; Li et al. 2011, 2012; Huang et al. 2012; Shao et al. 2014).

Other sequence datasets were built with sequences downloaded from GenBank ([www.ncbi.nlm.nih.gov/genbank](http://www.ncbi.nlm.nih.gov/genbank)) by mapping all available AMDV sequences ( $N = 456$ ) to a reference complete genome to identify the genomic regions most frequently used for molecular epidemiological studies (Olofsson et al. 1999; Dyer, Ching, and Bloom 2000; Mañas et al. 2001; Knuutila et al. 2009, 2015; Jahns et al. 2010; Christensen et al. 2011; Jensen et al. 2012; Leimann et al. 2015). Genomic areas represented by the majority of the sequences were selected, and sequences that mapped to these areas were extracted and used to perform phylogenetic analyses.

Sequences were aligned with ClustalX 2.1 (Larkin et al. 2007), alignments were manually edited when needed with BioEdit 7.0.5.3 (Hall 1999) and then used for phylogenetic inference. A model selection was performed for each alignment to identify the best model for distance estimation, and phylogenetic trees were constructed with MEGA 6.06 (Tamura et al. 2013) using the maximum-likelihood method (Felsenstein 1981). Information about numbers of sequences, the genomic regions investigated, and models used are provided in the figure legends. Bootstrap analyses (Felsenstein 1985) with 1,000 replicates were performed to test the robustness of the analyses, and only clusters supported by  $\geq 70$  per cent were considered valid.

Accession numbers of all sequences used in this study are available in the [Supplementary Sequence Details File](#).

## 2.4 Amino acid sequence predictions

Splicing sites were determined following what was experimentally demonstrated (Qiu et al. 2006) and using the AMDV-G virus (accession number JN040434) as a reference. Donor and acceptor splicing sites were also predicted using the splicing prediction algorithm NNSPLICE from [fruitfly.org](http://fruitfly.org) and scores were assigned to each site (Reese et al. 1997). Splicing events were reproduced

in silico to determine the complete coding sequences of NS1, NS2, NS3, VP1, and VP2, which were then translated into aa sequences.

## 2.5 Recombination analyses

Alignments of partial NS1 nucleotide sequences and of the complete genomes were screened for the presence of chimeric sequences by all of the different methods included in the RDP 4.55 software package (Martin et al. 2015), specifically RDP (R, Martin and Rybicki 2000), GENECONV (G, Padidam, Sawyer, and Fauquet 1999), BootScan (B, Martin et al. 2005), MaxChi (M, Smith 1992), Chimera (C, Posada and Crandall 2001), SiScan (S, Gibbs, Armstrong, and Gibbs 2000), 3Seq (T, Bon, Posada, and Feldman 2007), and LARD (L, Holmes, Worobey, and Rambaut 1999). Only phylogenetically supported events identified by at least three different methods ( $P < 0.05$ ) were considered. Each recombination hypothesis was refined following detection order with breakpoint positions manually modified when considered adequate, and events were accepted or discarded based on probability scores, phylogenetic evidence, and sequence analysis. In detail, each identified event characterized by an average probability above the threshold for at least three different methods was evaluated independently. The phylogenetic placements of the putative recombinant strain (together with sequences showing evidence for the same recombination event) and the two putative parental strains (major and minor parents) were evaluated within the RDP framework. Events not associated with clear inconsistencies between different genomic regions were discarded. Inspection by eye of the triplets involved in every detected event allowed assessment of the chimeric nature of the sequences and determination of whether the putative breakpoint was properly located (between the two genetic regions, each derived from a different parental strain). If the position of the breakpoint had to be moved, new associated phylogenetic trees were again evaluated for confirmation. After modifying, accepting, or rejecting an event, the whole alignment was rescanned for new recombination hypotheses. Once all putative recombination breakpoints were identified, separate phylogenetic trees were built with separate genomic regions included between breakpoints using MEGA, as described above.

The analyses included only strains for which the entire genomic coding sequences were available and involved four sequences from Newfoundland (M228, M195, M173, WM25, all from 2014), three sequences from China (LN-1, LN-2 and LN-3, all from 2009) (Li et al. 2012), two sequences from USA (AMDV-G from the 1970s and UtahI from 1963) (Bloom et al. 1988), one sequence from Germany (SL-3, from the 1980s) (Schuierer et al. 1997), and five outgroup sequences (belonging to RFAV and GFAV) (Li et al. 2011; Shao et al. 2014). Outgroup sequences were used within the RDP framework only in tree reconstruction and in determining informative sites.

BootScan plots (Salminen et al. 1995) were built with SimPlot software 3.5.1 (Lole et al. 1999) using a window of 100 nucleotides and a step of twenty nucleotides with an F84 model (Felsenstein and Churchill 1996), the maximum-likelihood method (Felsenstein 1981), and 1,000 replicates.

Shimodaira–Hasegawa and Robinson–Foulds recombination matrices (Shimodaira and Hasegawa 2001; Simmonds and Welch 2006) were also constructed with the RDP software and used to visualize the impact of recombination on the phylogenetic relationships of the strains. For the construction of both

matrices, the default settings were used (window size: 400, step size: 100).

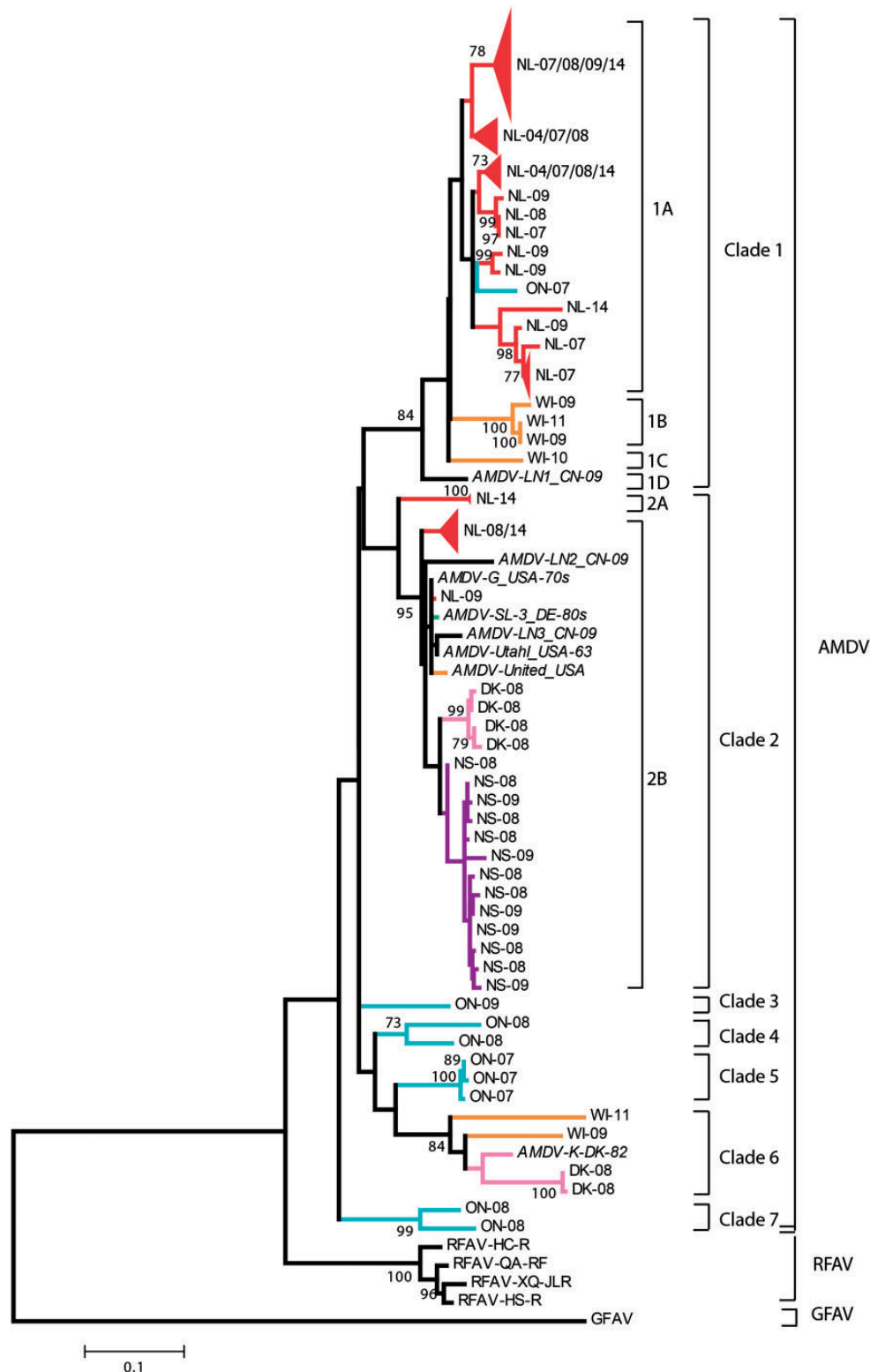
## 2.6 Selection pressure analysis

The Z-test of selection implemented in MEGA 6.06 (Tamura et al. 2013) was used to calculate the overall average synonymous (dS) and non-synonymous (dN) substitutions for each alignment. The Nei–Gojobori method (Nei and Gojobori 1986) was used to test both hypotheses,  $dN < dS$  (purifying selection pressure) and  $dN > dS$  (positive selection pressure), of deviation from strict neutrality (null hypothesis). The variance was estimated with the bootstrap method and 1,000 replicates. Sites under positive and purifying selection were assessed with six different methods: Single-Likelihood Ancestor Counting (SLAC), Fixed Effect Likelihood (FEL), Random Effects Likelihood (REL) (Pond and Frost 2005), Internal FEL (IFEL) (Kosakovsky Pond et al. 2006), Fast Unconstrained Bayesian Approximation for inferring selection (FUBAR) (Murrell et al. 2013), and Mixed Effects Model of Evolution (MEME) (Murrell et al. 2012). While SLAC evaluates the deviation of dN and dS from an expected neutral model, FEL calculates the probability for dN to be different from dS with a likelihood ratio test and REL uses Bayes Factors to evaluate the posterior probability of rejecting the null hypothesis of neutrality considering the distribution of dS and dN across the whole alignment. IFEL and MEME are extensions of FEL but FEL assumes dN/dS ratios apply to all branches, IFEL to interior branches, and MEME allows the ratios to be variable across lineages and sites, allowing the identification of episodic selection, which affects only a subset of lineages. Finally, FUBAR is a hierarchical Bayesian method that uses an MCMC approach; it works similarly to REL but relaxes some of its restrictions. All of these methods are phylogeny based, were used with HKY as the substitution model (Hasegawa, Kishino, and Yano et al. 1985), and are available online on the Datamonkey server (www.datamonkey.org) (Delpont et al. 2010). It is a commonly used practice to accept a site of being under selection only if predicted by more than one method (e.g., Boyle et al. 2014; Azarian et al. 2015) because some methods, like SLAC, are less sensitive and others, like REL, are more prone to overestimate the number of sites under selection. Furthermore, some methods are based on similar algorithms and likely to produce similar results. Therefore, we accepted a site to be under diversifying or negative selection only if predicted by at least three different methods. Finally, MEME can predict sites under episodic diversifying selection, and therefore, additional sites predicted by this method alone were also considered. Sites under selection were considered acceptable only when statistically significant ( $P < 0.1$  for SLAC, FEL, IFEL and MEME; Bayes Factor  $> 50$  for REL; posterior probability  $> 0.9$  for FUBAR).

## 3. Results

### 3.1 AMDV phylogeography

A total of 131 sequences from the NS1 region (nt 1207–1690) of AMDV genomes that originated from samples collected between 2004 and 2014 from ten different farms and wild animals in Newfoundland ( $N = 97$ ), Nova Scotia ( $N = 13$ ), Ontario ( $N = 9$ ), Wisconsin ( $N = 6$ ), and Denmark ( $N = 6$ ) were obtained and used to study AMDV molecular epidemiology. All sequences originated from viruses of mink because none of the other tested wild species was AMDV-positive. The resulting phylogenetic analysis of these sequences (Fig. 1) and the calculated identities



**Figure 1.** Phylogenetic analysis of AMDV partial NS1 sequences obtained during this study and originating from different areas of the world. The evolutionary history of the partial NS1 region (nt 1207–1690) was inferred using the maximum-likelihood method (Felsenstein 1981) based on the HKY model (Hasegawa, Kishino, and Yano 1985), identified as the best fitting model after the model test analysis, using MEGA6 (Tamura et al. 2013). A discrete Gamma distribution was used to model evolutionary rate differences among sites (+G=0.4098). The rate variation model allowed for some sites to be evolutionary invariable ([+I], 32.514% sites). The outcome of the bootstrap analysis (Felsenstein 1985) is shown next to the nodes, and branch lengths are proportional to genetic distances as indicated by the scale bar. Large groups of sequences originating from the same location and falling in the same clade have been collapsed at nodes into a triangle shape. Strains are labelled based on the original name (only for reference sequences, indicated in italics), sampling site (NL: Newfoundland; NS: Nova Scotia; ON: Ontario; WI: Wisconsin; USA: United States of America, state unknown; DK: Denmark; DE: Germany; CN: China) and year. Viral species, clades, and subclades are indicated by square brackets. Tree branches are colored based on sample origin (red: Newfoundland; purple: Nova Scotia; blue: Ontario; orange: USA; pink: Denmark; green: Germany; black: China).

between and within groups (available as [Supplementary Table S2](#)) were used to define clades and subclades, indicated by arbitrarily assigned numbers that are not meant as a classification but included to allow easier referencing in the text. We have defined clades as bootstrap-supported (>70%) phylogenetic clusters of sequences sharing >90 per cent identity with each other and <90 per cent identity with sequences from other clades. Subclades were defined in the same way but with a cut-off value of 96 per cent identity. Subclade 2A was an exception to this because it was >90 per cent identical to subclades 1A and 2B. Furthermore, this subclade was possibly responsible for difficulties in resolving the tree structure, resulting in bootstrap values below the threshold for clade 2 (bootstrap: 42) and subclade 1a (bootstrap: 43).

The seven clades resolved within the AMDV sequences were clearly separated from the other two amdoparvoviruses, GFAV and RFAV ( $\leq 85\%$  identity between species). Three of the AMDV clades (1, 2, and 6) also included reference sequences, whose complete genomes have been previously sequenced (except for AMDV-K), while the other clades comprised only sequences from this study. Most clades contained viruses from different geographic regions, but viruses collected from the same area tended to cluster together in separate groups within each larger clade. For example, clade 1 contained viruses from Newfoundland, Ontario, Wisconsin, and China and clade 2 included viruses from North America, Europe, and China, but sequences from the same region formed individual clusters within these clades in most cases. The other clades (3–7) were less diverse in terms of geographical origin of the sequences.

To confirm these results, we investigated whether sequences from other regions of the genome that have been used in other epidemiological studies showed a similar pattern. Three genomic regions were identified that included the majority of sequences available in the public databases. These were two regions of the NS1 ORF and one region of the VP1 ORF (nt 602–922, 1859–2208, and 2949–3228 of AMDV-G, respectively). Three separate phylogenetic trees were built with these different regions ([Fig. 2](#)), and we also included some sequences from Newfoundland for which complete coding sequences were obtained. The average identities between sequences used to build these trees were comparable to what was observed for the previously analyzed genomic region (87.8%, range 80.1–100%; 91.6%, range 82.3–100%; 91.7%, range 85–100%; vs. 91.5%, range 82–100%). Similar trends were observed in these trees. For example, viruses from Denmark formed three independent groups within larger groups that included viruses identified in other European regions and Canada ([Fig. 2A](#)). Similarly, the sequences from Newfoundland fell in different clusters but always formed independent groups ([Fig. 2A–C](#)). Lastly, viral sequences close to the reference strain AMDV-K that we identified in Ontario, Wisconsin, and Denmark were also observed interspersed among sequences obtained by others in Ontario, Denmark, Finland, Sweden, and Estonia ([Fig. 2A and B](#)).

### 3.2 AMDV in Newfoundland

Sequences from Newfoundland were found in only two of the seven identified clades ([Fig. 1](#)). Some strains in clade 1 were most similar to a virus from Ontario, while other Newfoundland sequences formed independent clusters that did not show such a close relationship to any other sequences.

We investigated the relationships among virus sequences from Newfoundland according to the year of sampling and the collection source ([Fig. 3](#)). A marked separation of viruses from

different farms was found. Closely related sequences were also observed on multiple farms, indicating that viruses in different farms originated from the same recent ancestor or that there was an exchange of viruses or infected animals between farms. In contrast, some farms were characterized by the presence of a distinct clade of viruses. However, with one exception, each farm contained viruses from only one clade.

Temporally, sequences from animals in 2007 were only found in clade 1, while those from samples collected in 2008 and 2009 were identified in both clades. This might indicate that viruses from the two clades were introduced during two separate events.

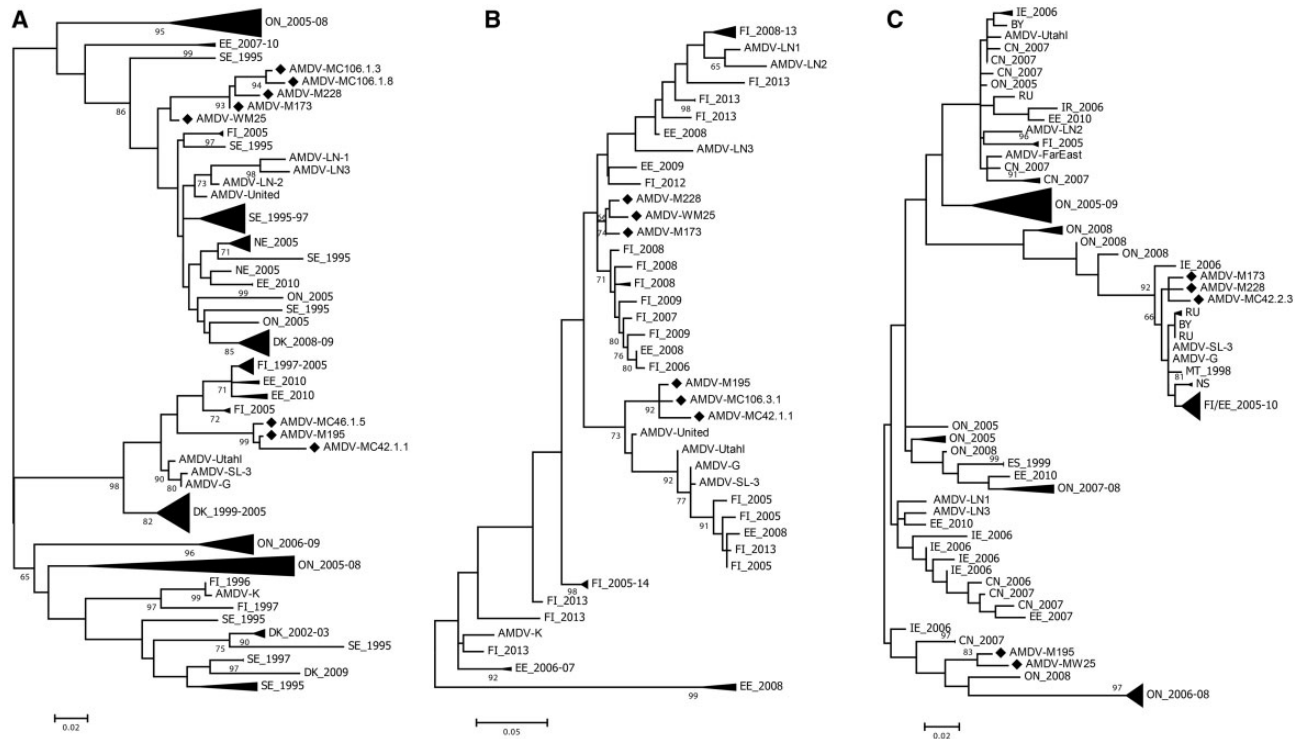
All viruses from wild animals belonged to clade 1. Approximately half of the wild mink contained viruses that were very close to those identified in farmed animals, as shown by the presence of clusters within lineage 1 that include sequences from both farmed and wild animals ([Fig. 3](#)). The other half of the viruses from wild animals formed independent clusters, where no sequences derived from farms were present and which therefore might represent a separate lineage of viruses unique to wild animals. This observation is further supported by the fact that the vast majority of sequences (7/8) identified in wild animals before the epidemic started (from 2004) are located in these clades. However, one sequence from 2004 could be identified that was closer to strains from 2007 circulating in farms.

### 3.3 Within-farm AMDV variation

All sequences from 2014 were derived from samples collected on the same farm, for which samples from an earlier time point (2007) were also available, and for each 2014 animal, eight to ten clones were sequenced. Analysis of the relationships of all 127 sequences identified from this farm showed a high level of genetic diversity ([Fig. 4](#)). The overall mean *p* distance was 7.2 per cent, and most of the sequences fell into two distinct clades. Comparison of sequences from the two clades showed an average identity of 89.3 per cent, while sequences within each clade were 95.4 per cent and 97.4 per cent identical for clades 1 and 2, respectively. Only one viral type was identified within the six sequences obtained from samples collected in 2007, while much higher variation was found in 2014 in samples obtained from twelve animals, suggesting that the introduction of viruses belonging to clade 1 into the farm occurred around 2007 and that the clade 2 viruses were introduced later. The presence of subclades within clade 1 that were not detected in 2007 likely reflects viral evolution over the 7-year period between sample collections.

A high co-infection rate was observed in samples collected from farmed animals in 2014 ([Fig. 4](#)). Remarkably, 41.7 per cent (5 out of 12) of the mink were positive for two or three different viruses. Furthermore, several polymorphic sites in each viral type within the same animal could be identified, revealing substantial intra-host mutation.

Finally, the presence of possible recombinant strains was observed ([Fig. 4](#)) and further investigated. With RDP we detected two possible recombination events with two different putative breakpoints, involving several clonal sequences identified in three animals. For both these events, a BootScan analysis was performed and two separate trees were built with partitions of the original alignment before and after the putative breakpoint ([Fig. 5](#)). Both events were well supported by *p* values calculated with all methods implemented in RDP (R:  $3.5 \times 10^{-2}$ ; B:  $3.4 \times 10^{-2}$ ; M:  $4.0 \times 10^{-3}$ ; C:  $1.1 \times 10^{-3}$ ; S:  $1.2 \times 10^{-5}$ ; L:  $3.5 \times 10^{-4}$ ;



**Figure 2.** Phylogenetic analyses of different genomic regions of AMDV strains identified worldwide. Trees were constructed with a 321-nt long portion of the NS1 genomic region (nt 602–922 of AMDV-G) of 179 different viruses (A), a 348-nt long portion of the NS1 genomic region (nt 1859–2208 of AMDV-G) of 56 different viruses (B) and a 280-nt long portion of the VP1 genomic region (nt 2949–3228 of AMDV-G) of 128 different viruses (C). Evolutionary histories were inferred with the maximum-likelihood method (Felsenstein 1981) based on the HKY model (Hasegawa, Kishino, and Yano 1985), identified as the best fitting model after the model test analysis, using MEGA6 (Tamura et al. 2013). A discrete Gamma distribution was used to model evolutionary rate differences among sites. The outcome of the bootstrap analysis (Felsenstein 1985) is shown next to the nodes, and branch lengths are proportional to genetic distances as indicated by the scale bars. Large groups of sequences originating from the same location and falling in the same clade have been collapsed at nodes into a triangle shape. Collection dates and sites (ON: Ontario; EE: Estonia; SE: Sweden; FI: Finland; NE: Netherlands; DK: Denmark; IE: Ireland; BY: Belarus; CN: China; RU: Russia; NS: Nova Scotia; MT: Montana; ES: Spain) are indicated. Sequences identified in this study from Newfoundland (NL) in 2014 are marked with a black diamond.

T:  $3.4 \times 10^{-5}$  for event in Fig. 5A–C  $1.5 \times 10^{-2}$ ; B:  $2.2 \times 10^{-2}$ ; M:  $2.3 \times 10^{-5}$ ; C:  $6.4 \times 10^{-5}$ ; S:  $1.2 \times 10^{-6}$ ; L:  $4.2 \times 10^{-2}$ ; T:  $7.3 \times 10^{-8}$  for event in Fig. 5D–F) and the sequences clearly clustered in different clades in the separate trees. In one case (Fig. 5A–C), one of the two putative parental strains (NL-14\_M106\_18) was present in the same animal where the recombinant strains were detected. The other recombinants were identified in two different animals (Fig. 5D–F), and close relatives to the two parental strains (NL-14\_M46\_12 and NL-14\_M10\_8) could be identified in the same animals where the recombinants were detected, with each parental strain found in a different animal. We cannot exclude that those chimeric sequences originated artificially during PCR, although the fact that two identical chimeric patterns were identified in two different animals is consistent with the presence of recombinant viruses in the farm.

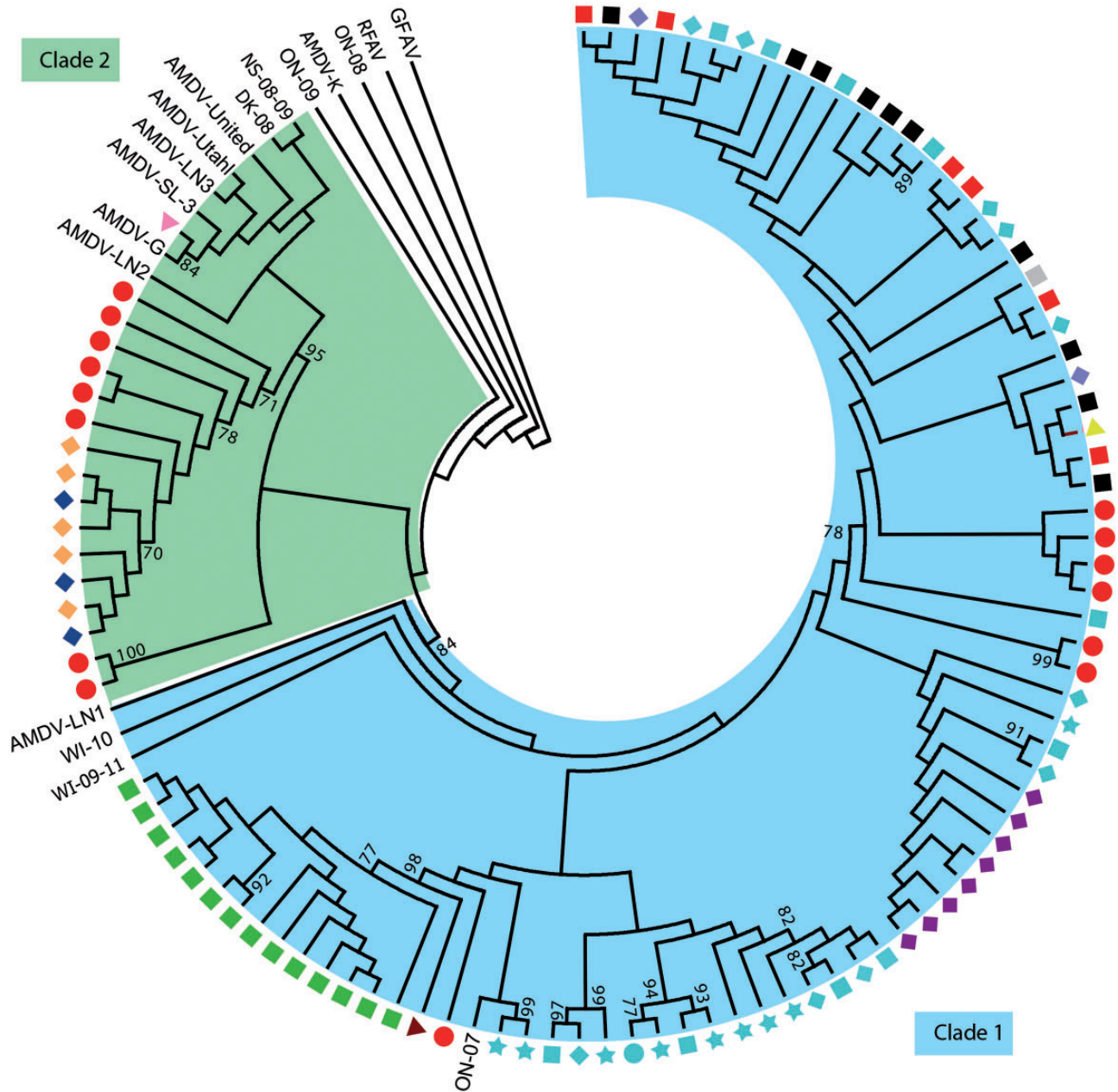
### 3.4 Complete genome sequence analysis

Attempts to obtain the complete genome sequences of viruses from samples with evidence of co-infection were not successful. The high within-host strain variability prevented the reliable assembly of cloned fragments. However, complete genome sequences were obtained for viruses from four samples collected in 2014 with single-strain infections. Three viruses were derived from farmed animals (M228, M173, and M195) and one from a wild mink (WM25). Additionally, several partial genomic fragments were obtained for other strains involved in co-infections.

These represented one complete and three nearly complete NS1 coding sequences (MC42.1.1 and MC106.1.3, MC106.1.8, and MC46.1.5) from three samples and two complete VP1 sequences (MC42.2.1 and MC42.2.3) from one sample.

The AMDV-G sequence was used as a reference to guide the splicing site search to identify exons, as they have been experimentally verified for that strain (Huang et al. 2012). The splicing pattern was resolved for all NS proteins and for VP1, and all donor and acceptor sites (except for the NS3 acceptor sites) showed high probability scores (scores are provided in the Supplementary Table S3). All AMDV and RFAV NS1 proteins were 641 aa long and the GFAV protein was 653 aa in length. The VP2 protein was 630 aa for the GFAV sequence and varied in length between 633 and 647 aa for AMDV and RFAV sequences. A glycine-rich region at the beginning of VP2, which varied between 3 and 13 aa was identified as the main cause for this variation (Supplementary Fig. S1). A one-aa deletion in VP2 was observed for strains FarEast, Rus17, M195, and WM25 (corresponding to AMDV-G VP2 residue T225), and the same deletion was also present in all RFAV sequences. The VP1u was 44 aa for GFAV and 43 aa for the other viruses. Finally, a 148-nt deletion was identified in one of the viral strains from Newfoundland (MC42.2.3), which most likely resulted in a non-functional capsid protein because of a consequent frameshift.

Phylogenetic analyses performed with all available complete NS1 aa sequences (Fig. 6A) showed three distinct AMDV clades, where Newfoundland strains formed two separate subclades



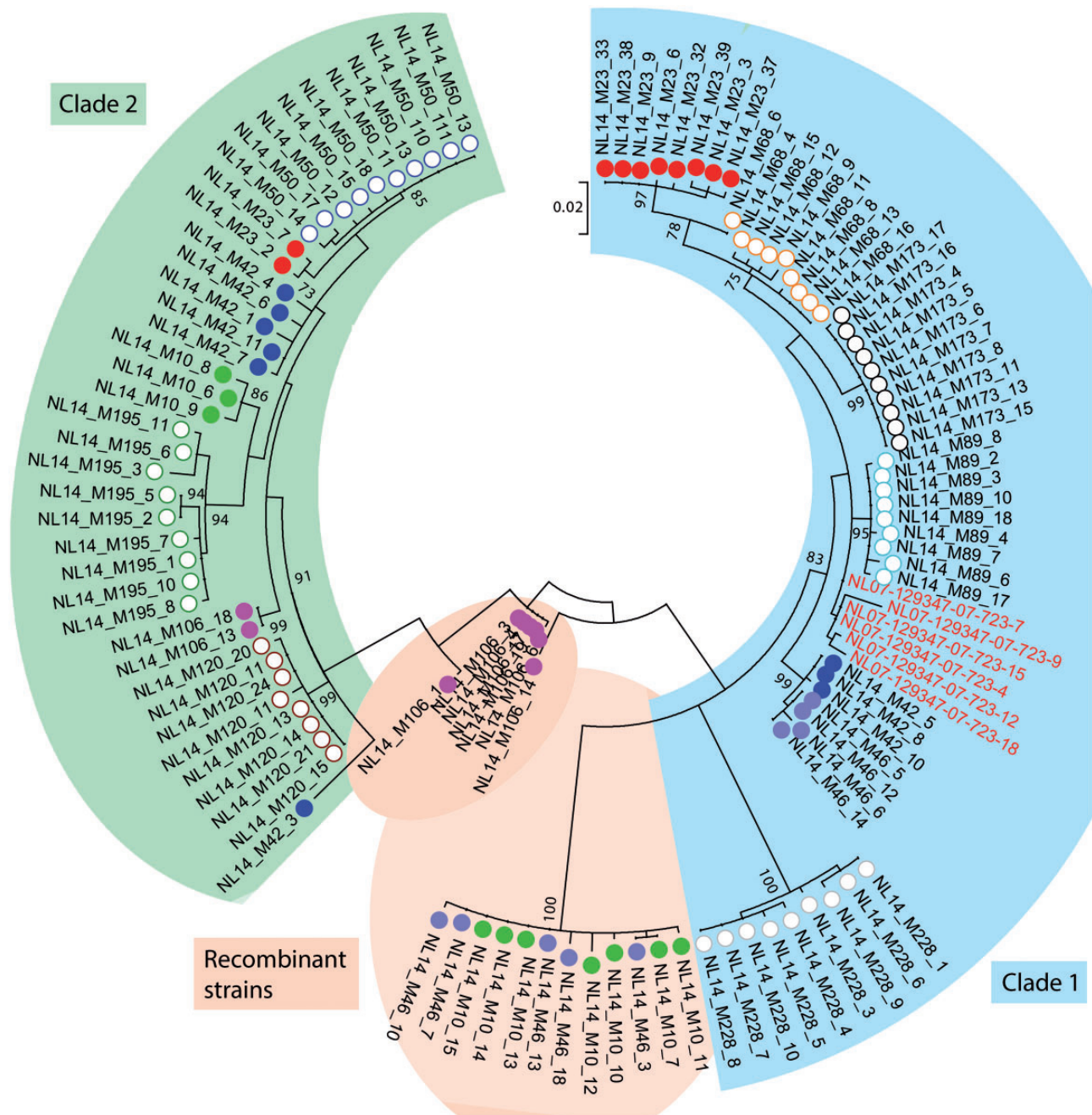
**Figure 3.** Phylogenetic relationship of AMDV strains from different locations in Newfoundland. The evolutionary history of the partial NS1 region (nt 1207–1690) of AMDV sequences was inferred using the maximum-likelihood method (Felsenstein 1981) based on the HKY model (Hasegawa, Kishino, and Yano 1985), identified as the best fitting model after the model test analysis, using MEGA6 (Tamura et al. 2013). A discrete Gamma distribution was used to model evolutionary rate differences among sites (+G = 0.4098). The rate variation model allowed for some sites to be evolutionary invariable ([+I], 32.514% sites). The outcome of the bootstrap analysis (Felsenstein 1985) is shown next to the nodes and only the tree topology is shown. Newfoundland strains are indicated by shapes and colors, where the shapes define the collection year (star: 2004; square: 2007; diamond: 2008; triangle: 2009; circle: 2014) and the colors indicate if the host was wild (light blue) or farmed (all other colors, with each farm represented by a different color). Strains identified in other areas are labelled according to the original name (only for reference sequences), sampling site (NS: Nova Scotia; ON: Ontario; WI: Wisconsin; DK: Denmark), and year.

within the first two main clades. A second analysis involving all available complete VP2 protein sequences (Fig. 6B) showed only two distinct major AMDV clades, revealed less variation with respect to NS1, and had no bootstrap-supported subclades. Names of clades and subclades are arbitrary and not related to those of Fig. 1. The AMDV-K strain, which in the NS phylogeny falls outside of the two main clades and forms a separate clade by itself, was originally sequenced from cloned dsDNA fragments obtained after restriction digestion from a viral isolate

containing multiple viral strains (Gottschalck et al. 1991). The NS1 and the VP2 sequences originated from two separate clones, and therefore, it is not known whether they belong to the same strain. This prevents us from drawing any conclusions regarding the AMDV-K clade identified in the NS1 tree with respect to the VP2 phylogeny.

Among the AMDV sequences analyzed, there were 289 parsimony-informative sites (PIS) out of 1,923 total sites in the NS1 alignment, which translated into 135 PIS out of 641 positions in



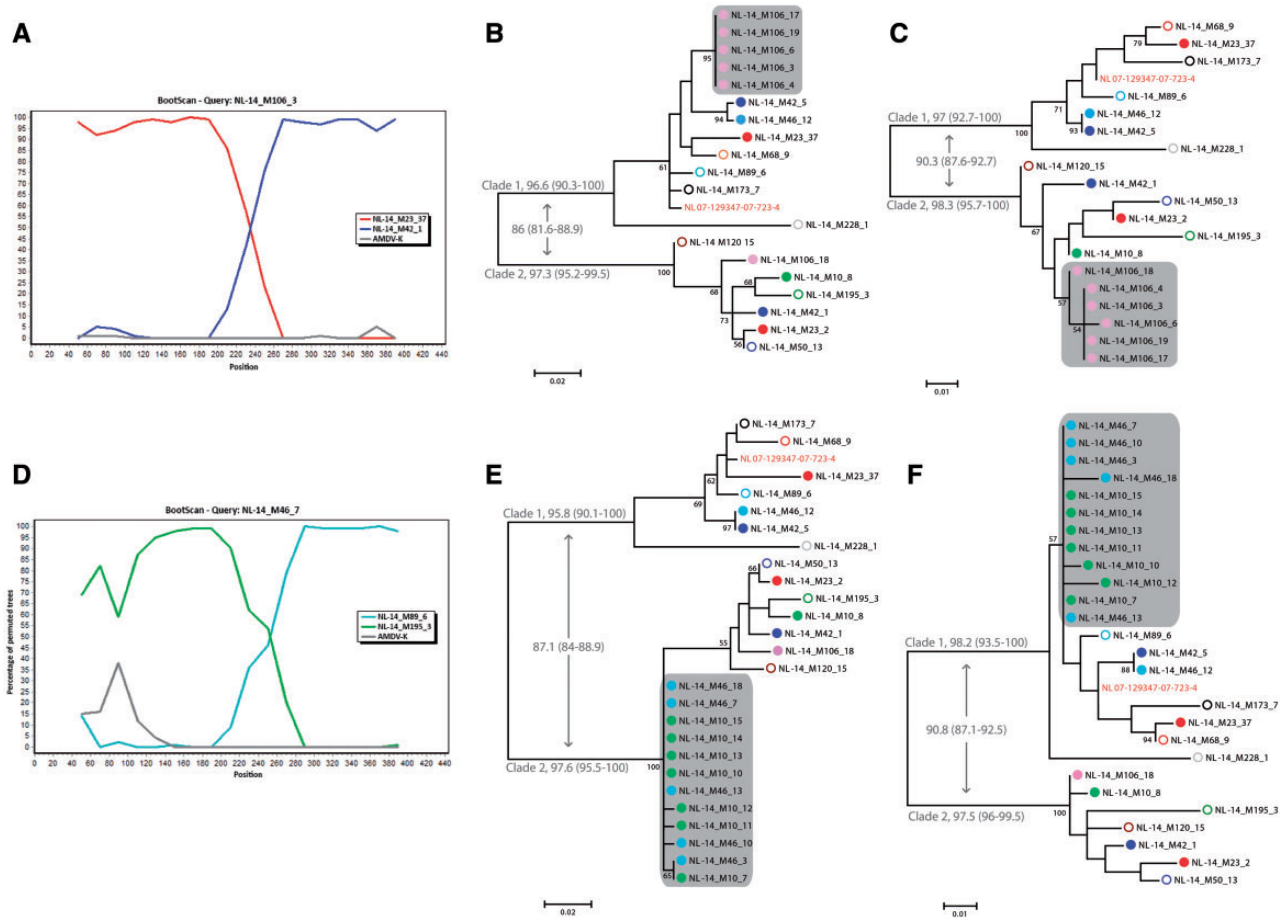


**Figure 4.** Phylogenetic relationships of AMDV strains from different mink within a single farm. The evolutionary history of the partial NS1 region (nt 1207–1690) was inferred using the maximum-likelihood method (Felsenstein 1981) based on the HKY model (Hasegawa, Kishino, and Yano 1985) using MEGA6 (Tamura et al. 2013). A discrete Gamma distribution was used to model evolutionary rate differences among sites (+G=0.1725). The outcome of the bootstrap analysis (Felsenstein 1985) is shown next to the nodes, and branch lengths are proportional to genetic distances as indicated by the scale bar. Strains identified in 2007 are indicated in red, while strains from 2014 are in black. Full circles represent viruses found in animals with co-infections, while empty circles represent single infections. Virus sequences with identical symbols and colors were from the same animal.

the aa alignment. Similarly, we identified 162 PIS out of 1,941 sites in the VP2 alignment, which translated into 47 sites out of 647 aa residues. Analysis of the identities (1–*p*-distances, Fig 6C) between groups indicated that AMDV-K, the only member of the third clade for which a complete coding sequence of at least one of the 2 ORFs is available, is distant enough from all other strains to be considered a different species according to the parvovirus classification criteria (ICTVdb 2014) because the NS1 protein sequences need to be  $\geq 85$  per cent identical to be considered the

same species. However, no appropriate classification can be proposed until the complete sequence is available. The distance analyses reflected what was observed with the phylogenies in that NS1 was less conserved than the structural protein VP2, both within and between species. For example, the RFAV sequences were approximately 72–76 per cent and 86–92 per cent identical to AMDV sequences in NS1 and VP2, respectively.

The clades identified on the NS1 tree did not correspond to those identified in the VP2 phylogeny. For example, the



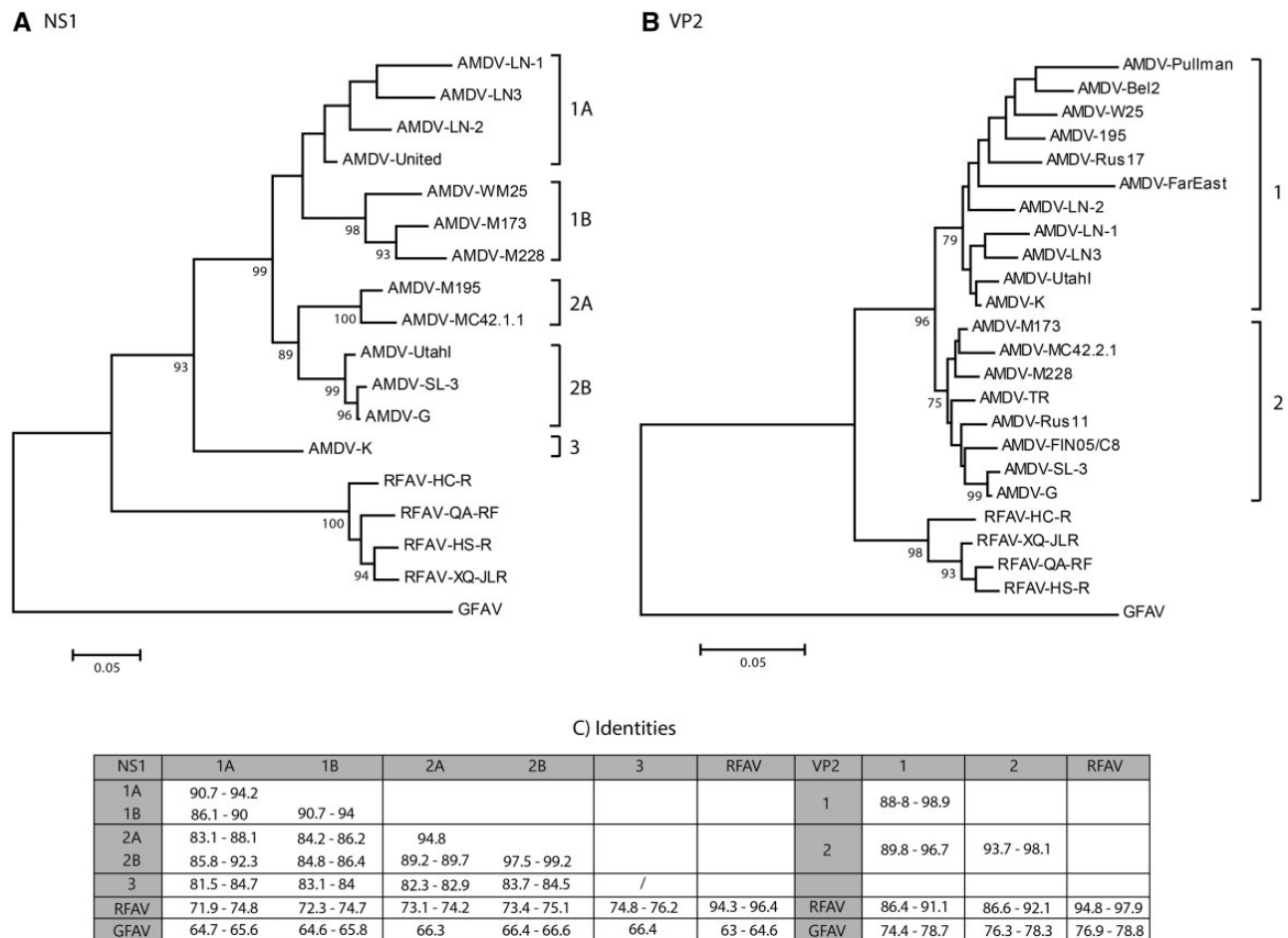
**Figure 5.** Recombination analysis of clonal AMDV strains from one mink farm. One event is displayed in panels A, B, and C, and the other event in panels D, E, and F. A BootScan analysis is shown for each event (A and D), and involved one of the recombinant sequences as query, sequences from relatives to the two parental strains, and AMDV-K as an outgroup. Trees built with the sequence partitions before the identified breakpoints are shown in panels B (nt 1207–1413) and E (nt 1207–1449), while trees built with the partitions after the breakpoints are shown in panels C (nt 1414–1690) and F (nt 1450–1690). The evolutionary histories were inferred using the maximum-likelihood method (Felsenstein 1981) based on the HKY model (Hasegawa, Kishino, and Yano 1985) using MEGA6 (Tamura et al. 2013). A discrete Gamma distribution was used to model evolutionary rate differences among sites. The outcome of the bootstrap analysis (Felsenstein 1985) is shown next to the nodes (only values above 50 are reported), and branch lengths are proportional to genetic distances as indicated by the scale bars. Full circles represent viruses found in animals with coinfections, while empty circles represent single infections. Virus sequences with identical symbols and colors were from the same animal. The phylogenetic placements of the recombinant strains are highlighted by shaded areas. Average identities (1–*p*-distance) in percentage values between and within clades (range) are reported in gray on each tree.

UtahI strain, which formed a 99 per cent bootstrap-supported cluster with strains G and SL-3 in the NS1 tree, was found in a different cluster together with the three Chinese sequences LN-1, LN-2, and LN-3 in the VP2 tree (Fig. 6A and B). Furthermore, the calculated *p*-distances within and between phylogenetically defined groups were not always coherent, because in some instances the distance between two sequences belonging to different groups was lower than the distance between two other sequences from within the same group (Fig. 6C). For example, the lowest identity between NS1 sequences in clade 1 was approximately 86 per cent, whereas sequences from clades 1 and 2 were up to 92.3 per cent identical. The same pattern was observed for VP2, where the lowest identity value between sequences within clade 1 was 88.8 per cent and the lowest identity when comparing sequences from clades 1 and 2 was 89.8 per cent. These inconsistencies between the phylogenetic trees and between phylogenetic grouping and *p*-distances suggested the presence of recombinant viruses.

### 3.5 Recombination breakpoint detection and investigation of chimeric sequences

The determination of recombination breakpoints was performed with viruses for which the complete coding sequence was available. Compatibility matrices were built to evaluate the impact of recombination indirectly by determining phylogenetic incongruities across the sequence alignments. The Robinson-Foulds and Shimodaira–Hasegawa compatibility matrices (Supplementary Fig. S2A and B) demonstrated that large variations exist between tree structures over the genome sequences for both adjacent and distant genomic regions, with very limited cold areas. This inconsistency in phylogeny suggested the possible occurrence of recombination during the evolutionary history of AMDV, and therefore, the location of putative recombination breakpoints was studied further with RDP.

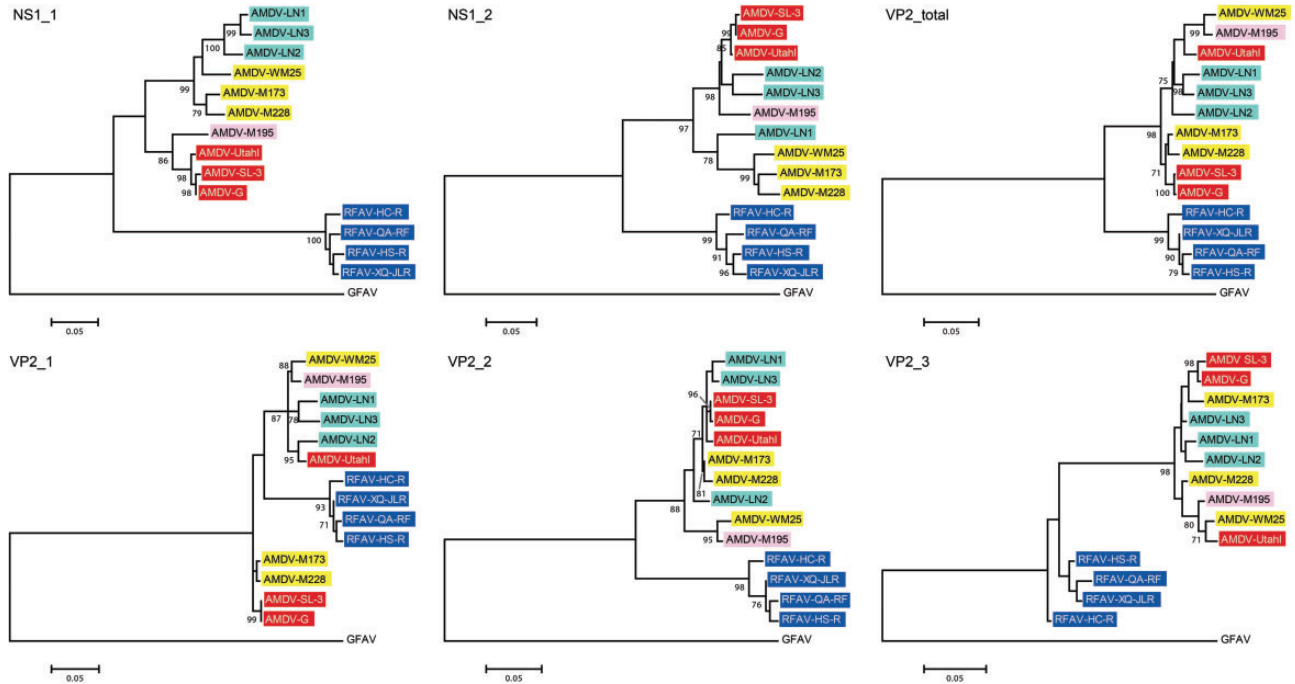
Several potential recombination events were detected (a recombination event map and probabilities for each event are provided as Supplementary Fig. S3) and four putative



**Figure 6.** Analyses of the complete AMDV coding regions. (A) Phylogenetic tree constructed with NS1 protein sequences. The evolutionary history was inferred using the maximum-likelihood method (Felsenstein 1981) based on the JTT model (Jones, Taylor, and Thornton 1992), identified as the best fitting model after the model test analysis, using MEGA6 (Tamura et al. 2013). A discrete Gamma distribution was used to model evolutionary rate differences among sites (+G, parameter = 0.6942). The outcome of the bootstrap analysis (Felsenstein 1985) is shown next to the nodes and branch lengths are proportional to genetic distances as indicated by the scale bar. (B) Phylogenetic tree based on VP2 protein sequences. The evolutionary history was inferred using the maximum-likelihood method (Felsenstein 1981) based on the General Reverse Transcriptase model (Dimmic et al. 2002), identified as the best fitting model after the model test analysis, using MEGA6 (Tamura et al. 2013). A discrete Gamma distribution was used to model evolutionary rate differences among sites (+G = 0.286). The outcome of the bootstrap analysis (Felsenstein 1985) is shown next to the nodes and branch lengths are proportional to genetic distances as indicated by the scale bar. (C) Identities (1 - *p*-distances) calculated within and between groups considering both NS1 and VP2 protein sequences. Values indicate the range of identities between pairs of sequences and are expressed as percentages. Clades correspond to those indicated on the trees displayed in panels A and B.

breakpoints, indicated by letters in [Supplementary Fig. S3](#), were evaluated. The first considered breakpoint (B) was located approximately 900nt from the beginning of the NS1 ORF (positions refer to strain AMDV-G), the second (between B and C) was located inside the VP1u region and after the end of NS1 (at approximately nt 2315 of the complete genome), the third (E) and the fourth (F) were located inside the VP1/2 ORF. The alignment was consequently split into five sub-alignments located between the putative breakpoints and new trees were constructed ([Fig. 7](#)). Similar phylogenetic trees based on the same genomic regions were also built including viruses whose genomes have not been completely sequenced to provide a wider view of the AMDV evolutionary history ([Supplementary Fig. S4](#)). Phylogenetic trees built with sub-portions of the VP2 ORF (VP2\_1, DE: nt 1-732; VP2\_2, EF: nt 733-1320; VP2\_3, FG: nt 1321-1875) did not strongly support recombination in this area, and therefore, we could not exclude that those phylogenetic incongruities originated from difficulties in resolving the tree structure. However,

well-supported phylogenetic incongruities could be observed between trees built with the two portions of the NS1 ORF (NS1\_1, AB: nt 1-903; NS1\_2, BC: 900-end) and between these and the tree built with the complete VP2 ORF (VP2\_total, DG). For example, the two Chinese sequences LN-2 and LN-3 clustered with the other Chinese sequence LN-1 and the Newfoundland strain WM25 in the first and third tree, while clustering with the two USA sequences UtahI and G and the German strain SL-3 in the second tree. A similar pattern in NS1 was observed for the USA strain United ([Supplementary Fig. S4](#)). Another example is the Newfoundland sequence M195 that, together with the UtahI strain, clustered with the strains SL-3 and G in the trees built with partial NS1 sequences, while clustering with the Newfoundland WM25 and the Chinese strains in the tree built with the VP2 ORF. However, the inconsistent phylogenies observed between the structural and non-structural genomic regions could also be the result of different evolutionary histories for the two genes (e.g., different evolutionary rates) as also supported by



**Figure 7.** Phylogenetic reconstruction of complete AMDV genomes based on different genomic regions located between putative recombination breakpoints. Evolutionary histories were inferred with the maximum-likelihood method (Felsenstein 1981) based on the HKY model (Hasegawa, Kishino, and Yano 1985), identified as the best fitting model after the model test analysis, using MEGA6 (Tamura et al. 2013). A discrete Gamma distribution was used to model evolutionary rate differences among sites. The outcome of the bootstrap analysis (Felsenstein 1985) is shown next to the nodes and branch lengths are proportional to genetic distances as indicated by the scale bars. Trees are based on genomic regions between nucleotides 1–903 (NS1\_1), 904–1926 (NS1\_2) of the NS1 ORF and between nucleotides 1–1875 (VP2\_total), 1–732 (VP2\_1), 733–1320 (VP2\_2), and 1321–1875 (VP2\_3) of the VP2 ORF; all positions refer to the AMDV-G sequence. Viruses are highlighted with the same color throughout all trees.

the different amounts of sequence divergence observed among strains in these two regions (Fig. 6).

When the two Newfoundland strains M228 and M195, which consistently segregated into different clusters in these three trees, were used as reference we could identify five different patterns but we did not observe an associated geographical distribution to go along with these patterns, as, for example, sequences identified in Newfoundland and China or in Germany and the USA showed the same genome composition (Supplementary Fig. S5).

### 3.6 Overall and site-by-site selection pressure evaluation

The non-synonymous/synonymous substitution rate ratios can be overestimated when recombination occurs, leading to the identification of sites falsely recognized as positively selected (Shriner et al. 2003). Therefore, each of the five sub-alignments of sequences between breakpoints (the same used for phylogenetic analyses in Fig. 7) was examined separately. The Z-test allowed us to reject the null hypothesis of strict neutral selection in favor of an alternative hypothesis of  $dN < dS$  in almost all cases. Two exceptions were the clades containing UtahI, SL-3, G, and M195 in NS1\_1 and WM25 and M195 in VP2\_2, for which the null hypothesis of neutrality could not be rejected. This showed the predominance of negative selection on the analyzed genomic sub-regions of both the entire dataset and separate clades. Mean values for  $dN/dS$  calculated with SLAC ranged between 0.51 and 0.62 for the NS1 ORF and between 0.24 and 0.33 for the VP2 ORF.

Six different methods were used to identify positively and negatively selected sites in both ORFs, and only sites identified by at least three different methods were accepted. In our system, considering the capability of each method to identify accepted sites under selection without overestimating the number of selected sites, FEL was the best to identify positively selected sites, whereas FUBAR was the best to identify sites under negative selection (Supplementary Table S4). Complete lists of all sites identified to be under selection by all methods are available in Supplementary Tables S5, S6, S7, and S8.

Nine sites under positive selection and forty under negative selection were identified in NS1 (Supplementary Table S5), while four sites under positive and thirty-four under negative selection pressure were observed in VP2 (Supplementary Table S6). This represented a similar percentage of sites under purifying selection in both proteins (6.2% and 5.4% for NS1 and VP2, respectively). Only 0.6 per cent of sites in the capsid protein were under diversifying selection, while 1.4 per cent of aa in the non-structural protein were under positive selection. Six (15%) of the sites under purifying selection identified in NS1 were located in the helicase domain, five of which (sites 432, 433, 434, 438, and 439) were localized in the ATP-binding loop (Walker et al. 1982) as illustrated in Canuti et al. (2014). Interestingly, MEME predicted eight additional sites to be under episodic diversifying selection, and one of these was also localized in the helicase domain, specifically in the Walker B domain, and involved the cluster formed by three Newfoundland strains (WM25, M173, and M228) in the NS1\_2 tree (Fig. 7). This domain is usually composed of a stretch of four hydrophobic residues, followed by two polar residues (hhhh(D/E)E). The alanine (codon GCT) present at the position of the fourth hydrophobic residue

in 80 per cent of the analyzed sequences was replaced by a cysteine (TGT) in M228 and by the polar residue threonine (ACT) in WM25 and represent changes that could have functional consequences for the domain.

Three of the four identified positively selected sites in VP2 (Supplementary Table S6) were localized in immunogenic loops (one in loop 1: 83; one in loop 2: 227; one in loop 4: 439), and 17.6 per cent (6/34) of sites under negative selection were also localized in immunogenic loops (two sites in loop 1: 80 and 93; two sites in loop 2: 239 and 240; two sites in loop 4: 452 and 481) (McKenna et al. 1999). An episodically positively selected residue was identified by MEME and this was located in antigenic loop 3 (residue 308).

Finally, we observed no differences between dN and dS rates when performing the Z-test on the NS2 and NS3 regions after the splicing site (where a frameshift causes the terminal portion of NS2 and NS3 to be different from NS1). The average mean values for dN/dS calculated with SLAC for these sequences were 1.3 and 3.6, respectively. A site-by-site exploration identified three positively selected sites in the NS2 unique region (Supplementary Table S7), and no sites under selection were identified in the NS3 unique region (Supplementary Table S8).

## 4. Discussion

AMDV is one of the most important pathogens of mink and other mustelids worldwide. It has been recognized for over 50 years as a major problem for farms, where it is often endemic. There is no available vaccine, and affected farms have to implement laborious eradication programs to establish an AMDV-free status (Cho and Greenfield 1978), that is, however, hard to achieve and to maintain (Canuti, Whitney, and Lang 2015), especially because of the environmental stability of the virus. The virus causes a persistent infection that leads to a slowly progressive wasting syndrome (Porter, Larsen, and Porter 1973), which is associated with high mortality, abortion, reduced reproduction rates and litter sizes in farms (Broll and Alexandersen 1996). Furthermore, the virus also represents a severe risk for wild animals and poses a serious threat to endangered species (Mañas et al. 2001; Fournier-Chambrillon et al. 2004). Farms represent one source of viruses for wild animals because accidental, or sometimes deliberate, mink release from farms is not a rare event (Northcott, Payne, and Mercer 1974; Nituch et al. 2011). It is therefore important to study the diversity and the molecular epidemiology of this virus to comprehend its evolution and transmission dynamics if we want to be able to contain the disease and prevent further spread.

### 4.1 Global epidemiology of AMDV

During this study, we determined 97 AMDV sequences from farmed and wild animals from the Island of Newfoundland to study the molecular epidemiology of AMDV in this region and in a global perspective. Although there was evidence for AMDV in Newfoundland since at least the 1960s–70s (H.G. Whitney, unpublished data), two studies documented the absence of AMDV from farms during the early 2000s. In 2001, 20 per cent of the farmed mink population of the province was tested and in 2004–5 approximately 50 per cent of the farmed mink population representing all farms in the province was screened again and only one positive animal was found in a small herd (thirty-five animals), which was entirely depopulated (Bradley 2005). These studies also showed that AMDV was already circulating in wild mink before the farm epidemic started in 2007. Over a

period of 11 years (2004, 2007–2009, and 2014), two of the three main well-characterized AMDV clades (Canuti, Whitney, and Lang 2015) were identified on the Island of Newfoundland, with viruses belonging to only one of the two clades found in the earliest samples. This suggests that the two variants may have been introduced during at least two separate events, although viruses from the second clade might have gone undetected in the earlier time point, and it is also possible that new introductions keep occurring. The oldest record of mink farming in Newfoundland is from 1934, when mink were imported from Nova Scotia for the establishment of several farms. However, the number of farms started to decline during the 1950s and the last farm pelted out in 1971 (Northcott, Payne, and Mercer 1974). Mink were then reintroduced for farming purposes in recent years, and mink farming has experienced rapid growth on the Island with farmers actively renovating their farms and importing new animals for stocking. Although it is theoretically possible that AMDV was introduced during the first farming wave, which was also the origin of the local wild mink population, the viruses identified in this study in farms seem to be from a more recent introduction. This is supported by the high genetic similarity of viruses from Newfoundland to viruses circulating in 2007–9 in Wisconsin and Ontario, although we cannot exclude, however unlikely, that viruses moved in those years from Newfoundland to other sites in North America. It is, however, uncertain whether the epidemic in farms originated from a new introduction from outside the island or from viruses already circulating within the wild mink population, because viruses similar to those identified during the epidemic were already present in wild animals. From 2008, the year following the beginning of the outbreak, viruses from a second clade started circulating in farms and these more likely represent a new introduction because we have no evidence for these strains in the wild.

We also obtained sequences from viruses from other locations in North America (Nova Scotia, Ontario, and Wisconsin) and from Europe (Denmark). Among these sequences, we identified several viral clades that are phylogenetically close to a third AMDV clade but for which a complete genome sequence is currently unavailable. These variants were only identified among sequences from Ontario, Wisconsin, and Denmark, and similar clades have been reported previously by others in samples collected in Ontario, Denmark, Finland, Sweden, and Estonia (Olofsson et al. 1999; Jensen et al. 2012; Nituch et al. 2012; Knuuttila et al. 2015). The absence of these variants from Newfoundland and Nova Scotia suggests a lower viral diversity in these areas compared to other parts of the world.

Phylogeographic analyses performed on four different genomic regions, involving both original sequences from this study and others from the public databases, and representing viruses circulating all over the world showed the existence of a partial but marked geographic distribution pattern for some strains. Although some viruses from different countries and continents are within the same clades, sequences tended to form independent country-specific subclades. These observations suggest that viruses are actively exchanged between different countries, presumably by trading of infected mink or contaminated equipment, but also that viruses, once reaching a new location, evolve rapidly within that population and produce independent lineages.

Approximately half of the sequences from Newfoundland wild mink were very close to those identified in farmed animals, suggesting a continuing exchange of viruses between wild and farmed populations, as reported in other locations (Nituch et al. 2011, 2012; Knuuttila et al. 2015). Although viral diversity was

higher in farms, we identified a cluster of viruses unique to wild animals and this included the majority of strains circulating in the wild before the epidemic started in farms. However, we cannot exclude the possibility that higher viral diversity remains undetected in farms or in the wild. These observations might indicate that farms represent a source of viruses for wild populations but also that the virus can transmit among wild animals, where it evolves independently, constituting a wild reservoir for farms. The movement of viruses between farms and the wild is also demonstrated by the increase in prevalence of AMDV in the wild around the time the epidemic started in farms (from 14% observed in 2004 to 45% in 2007–8). Furthermore, many AMDV-positive animals were trapped in areas surrounding affected farms, especially in the central part of the Avalon peninsula, where most of the samples collected in 2004 were negative (data not shown). The absence of detection of AMDV in Newfoundland in wild carnivores other than mink suggests that the virus has not transmitted among wild animals in this location as extensively as in other countries (Mañas et al. 2001; Fournier-Chambrillon et al. 2004; Farid et al. 2010). However, the number of animals tested was small and these results are therefore preliminary.

#### 4.2 Evolution of AMDV

A second objective of this work was to study the evolutionary dynamics of AMDV, and for this purpose, the genetic variability of viruses within a single farm was analyzed. Intensive farming provides a unique set of conditions that are considerably different from those of a natural environment that favor pathogen diffusion. Pathogen transmission is easier when the local host density is high and a fast turnover of individuals provides viruses the constant presence of a naïve population. In the case of AMDV, its exceptional environmental stability is also likely to favor an increase in  $R_0$  in farmed settings. These conditions facilitate viral spread and overall viral replication is therefore increased by the efficient host-to-host transmission, which are factors that correlate with faster viral evolution and higher diversity (Elena and Sanjuán 2005; Mennerat et al. 2010). Additionally, the continuous movement of animals between affected farms causes the continuous (re)introduction of viruses, which will be co-existing with previously circulating viruses whose evolution has already been shaped by a long period of sustained transmission within the farm, further amplifying viral richness. Our data are consistent with these hypotheses because we observed a very high viral genetic diversity within one farm. This was observed not only at the population level but also within single animals. However, these data need to be validated by the opposite observations in the wild to confirm that frequent co- and super-infections in chronically infected hosts are more frequently observed in high-density host populations.

The evolutionary dynamics of AMDV are further complicated by qualities intrinsic to this virus. The first is the ability of the virus to establish persistent infections, and the second is its exceptionally high environmental resistance (Canuti, Whitney, and Lang 2015). These two factors favor the occurrence of co-infections that create the condition necessary for viral recombination to take place, a mechanism that serves as positive feedback and further enhances viral diversity. We found the presence of multiple infections by two or three different strains in >40 per cent of the animals in 2014. Additionally, we identified various recombinant strains and the presence of multiple polymorphic sites in viruses within the same individuals, resulting from point mutations that occurred during chronic

infections. This exceptionally high, and possibly continuously increasing, within-host richness is a distinctive feature potentially associated with the increased chance of emergence of new characteristics, for example an extended host-tropism, in evolving pathogens (Shackelton et al. 2005; Duffy, Shackelton, and Holmes 2008). These might pose a threat not only for farmed animals but also for other species should these viruses escape from the farm.

After acquiring the complete genome sequences of several Newfoundland strains and comparing them to reference sequences, we observed that chimeric recombinant AMDV sequences appear to be recurrent. Parvoviruses are particularly recombination-prone because of their genome replication mechanism, which involves the continuous rearrangement of double-stranded intermediate DNA replication forms (Cotmore and Tattersall 2014), and the presence of recombination within this viral family has been frequently observed (Shackelton et al. 2007; Ohshima and Mochizuki 2009; Tyumentsev et al. 2014). Recombination might be additionally favored in the case of AMDV because of the high rate of co-infection in farms, and its role in viral evolution might be significant in these circumstances. However, we obtained clonal sequences from only one wild mink and our knowledge about AMDV evolutionary dynamics would benefit greatly from more sequences from viruses circulating in wild animals. It would also be very interesting to determine whether similarly high co-infection rates occur and if recombination takes place in farms where other amdoparvoviruses circulate.

Analyses of the direction of selective forces acting on AMDV identified a marked predominance of negative selection pressure on both structural and non-structural proteins. Surprisingly, a site-by-site analysis revealed a higher percentage of positively selected codons within the NS1 protein compared to the capsid protein VP2 (1.4% vs. 0.6%) and we could identify only four residues located in the immunodominant epitopes that were subjected to positive selection pressure. Our data are in agreement with previous literature because the predominance of purifying selection acting on non-structural proteins of parvoviruses has been previously reported (Pereira, Leal, and Durigon 2007), and the importance of negative selection in shaping the evolution of both ORFs has been documented for different parvoviruses (Lukashov and Goudsmit 2001). When recombination occurs, forces acting on one site can become independent from the action at another site and it is known that recombination can speed up the fixation of beneficial mutations (Felsenstein 1974). Therefore, although we analyzed each genomic area between breakpoints separately, potential mistakes in our site-by-site analysis can be expected because the entire spectrum of AMDV diversity is currently unknown, and therefore, some recombination will not have been documented. However, the low level of diversifying selection acting on the structural proteins is further confirmed by our observation that the degree of genetic variability is much higher in NS1. A possible explanation for this phenomenon can be identified in the context of the pathogenic mechanism of AMDV. The primary replication site for AMDV is in circulating macrophages and viral entry is mediated by cellular Fc receptors recognizing antibody-coated viral particles. Since antibodies enhance viral entry into cells, a phenomenon called antibody-dependent enhancement (Kanno, Wolfenbarger, and Bloom 1993), the host immune response contributes to the disease process by being fundamental for viral replication and in these circumstances escaping the immune response might not be beneficial for the virus. The lack of diversifying pressure at the level of the

immunoepitopes can also be partially explained by the continuous availability of naïve individuals in farms, as previously reported for the canine parvovirus (Pereira, Leal, and Durigon 2007). However, the first hypothesis is more compatible with the tendency of AMDV to establish persistent infections.

During our study, we mainly analyzed viruses from farmed animals and it appears many of the wild animals analyzed may have acquired the infection from an original farm source. Therefore, the evolutionary dynamics of AMDV in a more natural environment might be different compared to what we observed. Future studies should focus their attention on viruses identified in areas where mink farms are absent to elucidate how selection pressure shapes the evolution of AMDV in such natural environments. Finally, the acquisition of additional sequence information, especially the complete coding sequences of those variants for which only partial information is currently available, is a priority to obtain a full understanding of AMDV epidemiology, diffusion, and evolution.

In conclusion, the high prevalence of AMDV observed in farms can be explained both by virological factors, such as the ability to establish persistent infections and the high stability of viral particles, and environmental conditions, such as high host density. These conditions facilitate the establishment of co-infections that favor the occurrence of recombination, which enhances the extant AMDV diversity. These viruses can be transmitted to wild animals and exchanged between different farms and countries, where rapidly evolving viruses give rise to many novel parallel lineages.

## Data Availability

All sequences obtained in this study were deposited in GenBank under accession numbers GQ895066–GQ895128, GQ981388–GQ981397, and KT878893–KT878968 (strain details are available in the [Supplementary Sequence Details file](#)).

## Acknowledgements

This paper is dedicated to the memory of Bruce Hunter, who passed away suddenly on 19 October 2011. He was an exemplary scientist and friend to many. The authors thank John Reynolds and Dr. Edward Miller for providing ermine samples, Angela Blundon, Joanne Walsh, Hannah Munro, and Jennifer O'Reilly for help with spleen processing, and farmers and trappers from Newfoundland for their generous contribution of samples. This research was supported by the Natural Sciences and Engineering Research Council of Canada (EGP 469208–14), the Newfoundland and Labrador Provincial Agricultural Research and Development Program, and the Newfoundland and Labrador Forestry and Agrifoods Agency.

## Supplementary data

[Supplementary data](#) are available at [Virus Evolution](#) online.

*Conflict of interest:* None declared.

## References

Alexandersen, S. (1986) 'Acute Interstitial Pneumonia in Mink Kits: Experimental Reproduction of the Disease', *Veterinary Pathology*, 23: 579–88.

- Azarian, T. et al. (2015) 'Impact of Spatial Dispersion, Evolution, and Selection on Ebola Zaire Virus Epidemic Waves', *Scientific Reports*, 5: 10170.
- Bloom, M. E. et al. (1988) 'Nucleotide Sequence and Genomic Organization of Aleutian Mink Disease Parvovirus (ADV): Sequence Comparisons Between a Nonpathogenic and a Pathogenic Strain of ADV', *Journal of Virology*, 62: 2903–15.
- et al. (1997) 'Expression of Aleutian Mink Disease Parvovirus Capsid Proteins in Defined Segments: Localization of Immunoreactive Sites and Neutralizing Epitopes to Specific Regions', *Journal of Virology*, 71: 705–14.
- Bodewes, R. et al. (2014) 'Viral Metagenomic Analysis of Feces of Wild Small Carnivores', *Virology Journal*, 11: 89.
- Boni, M. F., Posada, D., and Feldman, M. W. (2007) 'An Exact Nonparametric Method for Inferring Mosaic Structure in Sequence Triplets' *Genetics*, 176: 1035–47.
- Boyle, D. B. et al. (2014) 'Evolution of Bluetongue Virus Serotype 1 in Northern Australia Over 30 Years', *Journal of Virology*, 88: 13981–9.
- Bradley, K. (2005) *Provincial Aleutian Disease Status Verification Project*. Newfoundland and Labrador Agriculture.
- Broll, S., and Alexandersen, S. (1996) 'Investigation of the Pathogenesis of Transplacental Transmission of Aleutian Mink Disease Parvovirus in Experimentally Infected Mink', *Journal of Virology*, 70: 1455–66.
- Canuti, M. et al. (2014) 'Persistent Viremia by a Novel Parvovirus in a Slow Loris (*Nycticebus coucang*) With Diffuse Histiocytic Sarcoma', *Frontiers in Microbiology*, 5: 655.
- , Whitney, H., and Lang, A. S. (2015) 'Amdoparvoviruses in Small Mammals: Expanding Our Understanding of Parvovirus Diversity, Distribution, and Pathology', *Frontiers in Microbiology*, 6: 1119.
- Cho, H. J., and Greenfield, J. (1978) 'Eradication of Aleutian Disease of Mink by Eliminating Positive Counterimmunoelectrophoresis Test Reactors', *Journal of Clinical Microbiology*, 7: 18–22.
- Christensen, L. S. et al. (2011) 'Diversity and Stability of Aleutian Mink Disease Virus During Bottleneck Transitions Resulting from Eradication in Domestic Mink in Denmark', *Veterinary Microbiology*, 149: 64–71.
- Cotmore, S. F., and Tattersall, P. (2014) 'Parvoviruses: Small Does not Mean Simple', *Annual Review of Virology*, 1: 517–37.
- Delpont, W. et al. (2010) 'Datamonkey 2010: A Suite of Phylogenetic Analysis Tools for Evolutionary Biology', *Bioinformatics*, 26: 2455–7.
- Dimmic, M. W. et al. (2002) 'rtREV: An Amino Acid Substitution Matrix for Inference of Retrovirus and Reverse Transcriptase Phylogeny', *Journal of Molecular Evolution*, 55: 65–73.
- Duffy, S., Shackelton, L. A., and Holmes, E. C. (2008) 'Rates of Evolutionary Change in Viruses: Patterns and Determinants', *Nature Reviews Genetics*, 9: 267–76.
- Dyer, N. W., Ching, B., and Bloom, M. E. (2000) 'Nonsuppurative Meningoencephalitis Associated With Aleutian Mink Disease Parvovirus Infection in Ranch Mink', *Journal of Veterinary Diagnostic Investigation*, 12: 159–62.
- Elena, S. F., and Sanjuán, R. (2005) 'Adaptive Value of High Mutation Rates of RNA Viruses: Separating Causes from Consequences', *Journal of Virology*, 79: 11555–8.
- Farid, A. H. et al. (2010) 'A Survey of Aleutian Mink Disease Virus Infection of Feral American Mink in Nova Scotia', *Canadian Veterinary Journal*, 51: 75–7.
- et al. (2012) 'Prevalence of the Aleutian Mink Disease Virus Infection in Nova Scotia', *Canadian Preventive Veterinary Medicine*, 106: 332–8.
- Felsenstein, J. (1974). 'The Evolutionary Advantage of Recombination', *Genetics*, 78: 737–56.

- (1981). 'Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach.', *Journal of Molecular Evolution*, 17: 368–76.
- (1985) 'Confidence Limits on Phylogenies: An Approach Using the Bootstrap', *Evolution*, 39: 783.
- , and Churchill, G. A. (1996) 'A Hidden Markov Model Approach to Variation Among Sites in Rate of Evolution', *Molecular Biology and Evolution*, 13: 93–104.
- Fournier-Chambrillon, C. et al. (2004) 'Antibodies to Aleutian Mink Disease Parvovirus in Free-Ranging European Mink (*Mustela lutreola*) and Other Small Carnivores from Southwestern France' *Journal of Wildlife Diseases*, 40: 394–402.
- Gibbs, M. J., Armstrong, J. S., and Gibbs, A. J. (2000) 'Sister-Scanning: A Monte Carlo Procedure for Assessing Signals in Recombinant Sequences', *Bioinformatics*, 16: 573–82.
- Gottschalck, E. et al. (1991) 'Nucleotide Sequence Analysis of Aleutian Mink Disease Parvovirus Shows That Multiple Virus Types are Present in Infected Mink', *Journal of Virology*, 65: 4378–86.
- et al. (1994) 'Sequence Comparison of the Non-Structural Genes of Four Different Types of Aleutian Mink Disease Parvovirus Indicates an Unusual Degree of Variability', *Archives of Virology*, 138: 213–31.
- Gunnarsson, E. (2001) 'Documenting Freedom from Disease and Re-Establishing a Free Status After a Breakdown Aleutian Disease (plasmacytosis) in Farmed Mink in Iceland', *Acta Veterinaria Scandinavica*, 42: 1.
- Hall, T. (1999) 'BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT', *Nucleic Acids Symposium Series*, 41: 95–8.
- Hasegawa, M., Kishino, H., and Yano, T. (1985) 'Dating of the Human-Ape Splitting by a Molecular Clock of Mitochondrial DNA', *Journal of Molecular Evolution*, 22: 160–74.
- Holmes, E. C., Worobey, M., and Rambaut, A. (1999) 'Phylogenetic Evidence for Recombination in Dengue Virus', *Molecular Biology and Evolution*, 16: 405–9.
- Huang, Q. et al. (2012) 'Internal Polyadenylation of Parvoviral Precursor mRNA Limits Progeny Virus Production', *Virology*, 426: 167–77.
- et al. (2014) 'Molecular Characterization of the Small Nonstructural Proteins of Parvovirus Aleutian Mink Disease Virus (AMDV) During Infection', *Virology*, 452–453: 23–31.
- ICTVdb. (2014) ICTVdb. The Universal Virus Database of the International Committee of Viruses. [www.ictvonline.org](http://www.ictvonline.org) (last accessed date 31 December 2015).
- Jahns, H. et al. (2010) 'Neuropathologic Features of Aleutian Disease in Farmed Mink in Ireland and Molecular Characterization of Aleutian Mink Disease Virus Detected in Brain Tissues', *Journal of Veterinary Diagnostic Investigation*, 22: 101–5.
- Jensen, T. H. et al. (2012) 'High Prevalence of Aleutian Mink Disease Virus in Free-Ranging Mink on a Remote Danish Island', *Journal of Wildlife Diseases*, 48: 497–502.
- Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992) 'The Rapid Generation of Mutation Data Matrices from Protein Sequences', *Computer Applications in the Biosciences*, 8: 275–82.
- Kanno, H., Wolfenbarger, J. B., and Bloom, M. E. (1993) 'Aleutian Mink Disease Parvovirus Infection of Mink Macrophages and Human Macrophage Cell Line U937: Demonstration of Antibody-Dependent Enhancement of Infection', *Journal of Virology*, 67: 7017–24.
- Knuutila, A. et al. (2009) 'Molecular Epidemiology of Aleutian Mink Disease Virus in Finland', *Veterinary Microbiology*, 133: 229–38.
- et al. (2015) 'Aleutian Mink Disease Virus in Free-Ranging Mustelids in Finland—A Cross-Sectional Epidemiologic and Phylogenetic Study', *Journal of General Virology*, 6: 1423–35.
- Kosakovsky Pond, S. L. et al. (2006) 'Adaptation to Different Human Populations by HIV-1 Revealed by Codon-Based Analyses', *PLoS Computational Biology*, 2: e62.
- Larkin, M. A. et al. (2007) 'Clustal W and Clustal X Version 2.0', *Bioinformatics*, 23: 2947–8.
- Leimann, A. et al. (2015) 'Molecular Epidemiology of Aleutian Mink Disease Virus (AMDV) in Estonia, and a Global Phylogeny of AMDV', *Virus Research*, 199: 56–61.
- Li, L. et al. (2011) 'Novel Amdovirus in Gray Foxes', *Emerging Infectious Diseases*, 17: 1876–8.
- Li, Y. et al. (2012) 'Genetic Characterization of Aleutian Mink Disease Viruses Isolated in China', *Virus Genes*, 45: 24–30.
- Lole, K. S. et al. (1999) 'Full-Length Human Immunodeficiency Virus Type 1 Genomes from Subtype C-Infected Seroconverters in India, With Evidence of Intersubtype Recombination', *Journal of Virology*, 73: 152–60.
- Lukashov, V. V., and Goudsmit, J. (2001) 'Evolutionary Relationships Among Parvoviruses: Virus-Host Coevolution Among Autonomous Primate Parvoviruses and Links Between Adeno-Associated and Avian Parvoviruses', *Journal of Virology*, 75: 2729–40.
- Mañas, S. et al. (2001) 'Aleutian Mink Disease Parvovirus in Wild Riparian Carnivores in Spain', *Journal of Wildlife Diseases*, 37: 138–44.
- Martin, D. P. et al. (2005) 'A Modified Bootscan Algorithm for Automated Identification of Recombinant Sequences and Recombination Breakpoints', *AIDS Research and Human Retroviruses*, 21: 98–102.
- et al. (2015) 'RDP4: Detection and Analysis of Recombination Patterns in Virus Genomes', *Virus Evolution*, 1: vev003.
- Martin, D., and Rybicki, E. (2000) 'RDP: Detection of Recombination Amongst Aligned Sequences', *Bioinformatics*, 16: 562–3.
- McKenna, R. et al. (1999) 'Three-Dimensional Structure of Aleutian Mink Disease Parvovirus: Implications for Disease Pathogenicity', *Journal of Virology*, 73: 6882–91.
- Mennerat, A. et al. (2010) 'Intensive Farming: Evolutionary Implications for Parasites and Pathogens', *Evolutionary Biology*, 37: 59–67.
- Moya, A., Holmes, E. C., and González-Candelas, F. (2004) 'The Population Genetics and Evolutionary Epidemiology of RNA Viruses', *Nature Reviews Microbiology*, 2: 279–88.
- Murrell, B. et al. (2012) 'Detecting Individual Sites Subject to Episodic Diversifying Selection', *PLoS Genetics*, 8: e1002764.
- et al. (2013) 'FUBAR: A Fast, Unconstrained Bayesian Approximation for Inferring Selection', *Molecular Biology and Evolution*, 30: 1196–205.
- Nei, M., and Gojobori, T. (1986) 'Simple Methods for Estimating the Numbers of Synonymous and Nonsynonymous Nucleotide Substitutions', *Molecular Biology and Evolution*, 3: 418–26.
- Nituch, L. A. et al. (2011) 'Mink Farms Predict Aleutian Disease Exposure in Wild American mink', *PLoS One*, 6: e21693.
- et al. (2012) 'Molecular Epidemiology of Aleutian Disease Virus in Free-Ranging Domestic, Hybrid, and Wild Mink', *Evolutionary Applications*, 5: 330–40.
- Northcott, T. H., Payne, N. F., and Mercer, E. (1974) 'Dispersal of Mink in Insular Newfoundland', *Journal of Mammalogy*, 55: 243–8.
- Ohshima, T., and Mochizuki, M. (2009) 'Evidence for Recombination Between Feline Panleukopenia Virus and



- Canine Parvovirus Type 2', *Journal of Veterinary Medical Science*, 71: 403–8.
- Oie, K. L. et al. (1996) 'The Relationship Between Capsid Protein (VP2) Sequence and Pathogenicity of Aleutian Mink Disease Parvovirus (ADV): A Possible Role for Raccoons in the Transmission of ADV Infections', *Journal of Virology*, 70: 852–61.
- Olofsson, A. et al. (1999) 'Unusual, High Genetic Diversity of Aleutian Mink Disease Virus', *Journal of Clinical Microbiology*, 37: 4145–9.
- Padidam, M., Sawyer, S., and Fauquet, C. M. (1999) 'Possible Emergence of New Geminiviruses by Frequent Recombination', *Virology*, 265: 218–25.
- Pereira, C. A. D., Leal, E. S., and Durigon, E. L. (2007) 'Selective Regimen Shift and Demographic Growth Increase Associated With the Emergence of High-Fitness Variants of Canine Parvovirus', *Infection, Genetics and Evolution*, 7: 399–409.
- Pond, S. L. K., and Frost, S. D. W. (2005) 'Not so Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection', *Molecular Biology and Evolution*, 22: 1208–22.
- Porter, D. D., Larsen, A. E., and Porter, H. G. (1973) 'The Pathogenesis of Aleutian Disease of Mink. 3. Immune Complex Arteritis', *American Journal of Pathology*, 71: 331–44.
- Posada, D., and Crandall, K. A. (2001) 'Evaluation of Methods for Detecting Recombination from DNA Sequences: Computer Simulations', *Proceedings of the National Academy of Sciences United States of America*, 98: 13757–62.
- Qiu, J. et al. (2006) 'The Transcription Profile of Aleutian Mink Disease Virus in CRFK Cells is Generated by Alternative Processing of Pre-mRNAs Produced from A Single Promoter', *Journal of Virology*, 80: 654–62.
- Reese, M. G. et al. (1997) 'Improved Splice Site Detection in Genie', *Journal of Computational Biology*, 4: 311–23.
- Salminen, M. O. et al. (1995) 'Identification of Breakpoints in Intergenotypic Recombinants of HIV Type 1 by Bootscanning', *AIDS Research and Human Retroviruses*, 11: 1423–25.
- Schuijver, S. et al. (1997) 'Sequence Analysis of the Lymphotropic Aleutian Disease Parvovirus ADV-SL3', *Archives of Virology*, 142: 157–66.
- Shackelton, L. A., and Holmes, E. C. (2006) 'Phylogenetic Evidence for the Rapid Evolution of Human B19 Erythrovirus', *Journal of Virology*, 80: 3666–9.
- et al. (2005) 'High Rate of Viral Evolution Associated With the Emergence of Carnivore Parvovirus', *Proceedings of the National Academy of Sciences United States of America*, 102: 379–84.
- et al. (2007) 'Comparative Analysis Reveals Frequent Recombination in the Parvoviruses', *Journal of General Virology*, 88: 3294–301.
- Shao, X-Q. et al. (2014) 'Novel Amdoparvovirus Infecting Farmed Raccoon Dogs and Arctic Foxes', *Emerging Infectious Diseases*, 20: 2085–8.
- Shimodaira, H., and Hasegawa, M. (2001) 'CONSEL: For Assessing the Confidence of Phylogenetic Tree Selection', *Bioinformatics*, 17: 1246–7.
- Shriner, D. et al. (2003) 'Potential Impact of Recombination on Sitewise Approaches for Detecting Positive Natural Selection', *Genetics Research*, 81: 115–21.
- Simmonds, P., and Welch, J. (2006) 'Frequency and Dynamics of Recombination Within Different Species of Human Enteroviruses', *Journal of Virology*, 80: 483–93.
- Smith, J. M. (1992) 'Analyzing the Mosaic Structure of Genes', *Journal of Molecular Evolution*, 34: 126–9.
- Tamura, K. et al. (2013) 'MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0', *Molecular Biology and Evolution*, 30: 2725–9.
- Tyumentsev, A. I. et al. (2014) 'Recombination in the Evolution of Human Bocavirus', *Infection, Genetics and Evolution*, 28: 11–4.
- Walker, J. E. et al. (1982) 'Distantly Related Sequences in the Alpha- and Beta-Subunits of ATP Synthase, Myosin, Kinases and Other ATP-Requiring Enzymes and a Common Nucleotide Binding Fold', *EMBO Journal*, 1: 945–51.
- Zehender, G. et al. (2010) 'Rapid Molecular Evolution of Human Bocavirus Revealed by Bayesian Coalescent Inference', *Infection, Genetics and Evolution*, 10: 215–20.