Evolutionary analysis of whole-genome sequences confirms inter-farm transmission of Aleutian mink disease virus

Hagberg, Emma Elisabeth; Pedersen, Anders Gorm; Larsen, Lars Erik; Krarup, Anders

Published in:
Journal of General Virology

Link to article, DOI:
10.1099/jgv.0.000777

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Evolutionary analysis of whole-genome sequences confirms inter-farm transmission of Aleutian mink disease virus

Emma E. Hagberg,1,2*† Anders G. Pedersen,2 Lars E. Larsen3 and Anders Krarup1

Abstract

Aleutian mink disease virus (AMDV) is a frequently encountered pathogen associated with mink farming. Previous phylogenetic analyses of AMDV have been based on shorter and more conserved parts of the genome, e.g. the partial NS1 gene. Such fragments are suitable for detection but are less useful for elucidating transmission pathways while sequencing entire viral genomes provides additional informative sites and often results in better-resolved phylogenies. We explore how whole-genome sequencing can benefit investigations of AMDV transmission by reconstructing the relationships between AMDV field samples from a Danish outbreak. We show that whole-genome phylogenies are much better resolved than those based on the partial NS1 gene sequences extracted from the same alignment. Well-resolved phylogenies contain more information about the underlying transmission trees and are useful for understanding the spread of a pathogen. In the main case investigated here, the transmission path suggested by the tree structure was supported by epidemiological data. The use of molecular clock models further improved tree resolution and provided time estimates for the viral ancestors consistent with the proposed direction of spread. It was however impossible to infer transmission pathways from the partial NS1 gene tree, since all samples from the case farms branched out from a single internal node. A sliding window analysis showed that there were no shorter genomic regions providing the same phylogenetic resolution as the entire genome. Altogether, these results suggest that phylogenetic analyses based on whole-genome sequencing taking into account sampling dates and epidemiological data is a promising set of tools for clarifying AMDV transmission.

INTRODUCTION

Aleutian mink disease (AMD), also referred to as plasmacytosis, is the most important disease in the mink farming industry worldwide. The disease affects mink of all ages and is caused by Aleutian mink disease virus (AMDV), a single-stranded DNA virus belonging to the family Parvoviridae [1]. Like other parvoviruses, the AMDV genome consists of two large ORFs and two smaller ones, which by alternative splicing encode three non-structural (NS1, 2 and 3) and two structural viral proteins (VP1 and 2) [2, 3]. Infection results in a harmful activation of the immune system, leading to hypergammaglobulinaemia and systemic vascular diseases such as glomerulonephritis. Animal welfare is reduced and infected animals either die due to organ failure or become persistently infected carriers, transmitting the virus within and between herds [4].

In Denmark, AMDV is monitored by a mandatory national control programme [5], which briefly requires all farms to conduct serology-based screening at regular intervals according to the region’s disease status. Positive farms undergo more intensive monitoring and are encouraged to depopulate followed by thorough cleaning and disinfection of the farm. Given these regulations and due to the fact that parvoviruses are highly contagious and very resistant to environmental factors, AMDV protection and prevention imposes large costs on the mink farmers [4].
Previous molecular studies of AMDV strains circulating in Denmark have primarily been based on Sanger sequencing of a part of the NS1 gene [6–9]. Such short and relatively conserved regions are useful for diagnostic purposes and for exploring more distant relationships between strains; however, due to the small number of informative sites, studies based on the partial NS1 gene have resulted in phylogenetic trees with low resolution with limited use for exploring outbreaks and discerning the transmission routes of AMDV between farms.

In this paper we demonstrate the strength and applicability of using whole-genome sequence for reconstructing phylogenetic relationships in a case of AMDV transmission among three Danish mink farms. The increased genetic information from the entire viral genomes may be used to investigate routes of viral transmission in more detail [10, 11]. Whole-genome data were obtained using next-generation sequencing (NGS), which has previously been a useful tool for this purpose [12, 13]. In connection with the phylogenetic analysis, we further explored the use of molecular clock models, which allowed us to estimate the age of the ancestors of the viruses from individual farms – thus generating information crucial for tracking the source of new outbreaks.

**RESULTS**

**Sequence analysis**

The data quality was overall high with approximately 99% of the reads mapping to the AMDV-G reference genome (data not shown), and as previously observed, the homopolymeric region between nt 2470–2520 caused a dip in read coverage, but did not affect downstream analysis [3, 14]. The average read depth for each sample is shown in Table 1. The nucleotide diversity (average number of nucleotide differences per site) between the recent Danish isolates was relatively low: \( \pi = 0.0062, SE = 0.00032 \) for the partial NS1 gene and \( \pi = 0.0043, SE = 0.00014 \) for the whole-genome sequences. The individual mean pairwise differences between the sequences collected at each farm are reported in Table 1.

All full-length AMDV field strains were analysed for the presence of recombination using SimPlot [15] and all methods implemented in the RDP4 software package [16]. No recombination was detected between the sequences from the case farms (A, B and C), the remaining Danish isolates or between the Danish and international sequences. The SimPlot analysis reflected the overall low pairwise distances in the alignment and in line with previous studies indicated a slightly higher variability in the first ORF [3, 17].

**Phylogenetic results**

**Evaluating the use of whole-genome sequences for reconstructing phylogenies**

The two alignments, based on the partial NS1 gene and whole-genome sequences, were used as the input for the phylogenetic analyses. The best-fitting nucleotide substitution model for the partial-gene dataset was the so-called HKY model that distinguishes between transition and transversion rates as well as allowing for unequal base frequencies [18], while for the whole-genome dataset it was the HKY model with a proportion of invariant nucleotide sites and a gamma-rate distribution (HKY+IG). The phylogeny based on the partial-gene dataset showed all farm A, B and C sequences branching our from a single node (a polytomy) together with farm I; it was therefore unhelpful for estimating phylogenetic relationships and inferring transmission routes (see Fig. 1). The same conclusion applied to the remaining farms.

Analysis of the whole-genome dataset resulted in a better-resolved phylogeny with fewer polytomies and displayed high posterior clade probabilities (Fig. 1b). Sequences originating from farms B and C formed sub-trees within the tree of farm A. This is consistent with the hypothesis that farm A was infected first and that the infection then spread to farms B and C – an idea further supported by the clock-model analysis and the prevalence data (see below). The remaining Danish outbreak-derived sequences mainly clustered according to the farm from which they were sampled. The overall tree topology resembled that seen in previous studies [7, 19], with the Danish strains forming a clade of their own and the global strains all being placed as out-groups (Fig. 2). Our results illustrate that the whole-genome sequences contain additional important additional genetic information that improves the phylogenetic signal compared to the partial NS1 gene fragment used in Denmark. The considerably higher tree resolution obtained when using the whole-genome alignment enables us to begin understand these outbreaks in greater detail and investigate the route by which the infection spreads within a country.

Although the partial NS1 gene region seems to be insufficient for a robust phylogenetic analysis, it is possible that other short genomic regions might be useful for phylogenetic analysis. This could potentially be advantageous due to the somewhat easier workflow involved in Sanger sequencing and subsequent analysis of a single PCR fragment compared to the several steps involved in obtaining and analysing large collections of NGS reads. We therefore performed a sliding window analysis with the purpose of quantifying the phylogenetic information content in different subsections of the AMDV genome. Manual inspection of the whole genome alignment showed that nucleotide changes were located over the entire genome (suggesting that all the data are necessary to obtain full resolution), however with a slightly higher diversity in its first half. In order to stringently quantify the phylogenetic signal in different sub-sections of the alignment, 400 bp windows spaced at 25 bp intervals were extracted from the whole genome alignment (i.e. the windows started at positions 1, 25, 50, 75, 100, etc., with the first covering 1–400, the second 25–424, etc.). A window size of 400 was chosen as this corresponds to a typical PCR fragment size and could easily be Sanger-sequenced. For each of the resulting 179 sub-alignments, a tree was reconstructed using a full Bayesian phylogenetic analysis, and subsequently, the relative resolution for each window was measured in two different, but related, ways:
Table 1. Overview of sequences generated in this study

Sampling dates are indicated as yyyy-mm-dd, sequence lengths are either full length (4369 bp) or nearly full length (3198 bp), average read depth is the number of reads per nucleotide position, and the intra-farm diversity is presented as the mean pairwise sequence difference (pi) and its standard error (se).

<table>
<thead>
<tr>
<th>Farm</th>
<th>Mink no.</th>
<th>Sampling date (yyyy-mm-dd)</th>
<th>Length (bp)</th>
<th>Average read depth</th>
<th>GenBank acc. no.</th>
<th>Intra-farm diversity (mean pi, se)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>2014-11-21</td>
<td>4369</td>
<td>10959</td>
<td>KU856560</td>
<td>0.001068, 0.000257</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2014-11-21</td>
<td>4369</td>
<td>12912</td>
<td>KU856561</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2014-11-21</td>
<td>4369</td>
<td>12627</td>
<td>KU856562</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>2014-11-14</td>
<td>4369</td>
<td>12768</td>
<td>KU856563</td>
<td>0.00305, 0.000062</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2014-11-14</td>
<td>4369</td>
<td>6008</td>
<td>KU856564</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2014-11-14</td>
<td>4369</td>
<td>14170</td>
<td>KU856565</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>2014-11-14</td>
<td>4369</td>
<td>14211</td>
<td>KU856566</td>
<td>0.001373, 0.000285</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2014-11-14</td>
<td>4369</td>
<td>14238</td>
<td>KU856567</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2014-11-14</td>
<td>4369</td>
<td>8015</td>
<td>KU856568</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>2014-11-19</td>
<td>4369</td>
<td>3267</td>
<td>KX404887</td>
<td>0.000870,</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2014-11-19</td>
<td>4369</td>
<td>1001</td>
<td>KX404888</td>
<td>0.001137</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2014-02-10</td>
<td>4369</td>
<td>1496</td>
<td>KU856571</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2014-02-10</td>
<td>4369</td>
<td>1339</td>
<td>KU856572</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2014-02-10</td>
<td>4369</td>
<td>765</td>
<td>KU856573</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>2014-05-08</td>
<td>4369</td>
<td>1283</td>
<td>KX404889</td>
<td>0.00458,</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2014-05-08</td>
<td>3198</td>
<td>1967</td>
<td>KX404890</td>
<td>0.001187</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2014-05-08</td>
<td>4369</td>
<td>1438</td>
<td>KX404891</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>1</td>
<td>2014-11-11</td>
<td>3198</td>
<td>341</td>
<td>KX404892</td>
<td>0.00458,</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2014-11-11</td>
<td>3198</td>
<td>1040</td>
<td>KX404893</td>
<td>0.001187</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2014-11-11</td>
<td>4369</td>
<td>1174</td>
<td>KX404894</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>2014-11-10</td>
<td>3198</td>
<td>1290</td>
<td>KX404895</td>
<td>0.006610,</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2014-11-10</td>
<td>3198</td>
<td>1277</td>
<td>KX404896</td>
<td>0.001125</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2014-11-10</td>
<td>4369</td>
<td>2263</td>
<td>KX404897</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>1</td>
<td>2014-11-11</td>
<td>4369</td>
<td>2337</td>
<td>KX404898</td>
<td>0.001465,</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2014-11-11</td>
<td>3198</td>
<td>1155</td>
<td>KX404899</td>
<td>0.000220</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2014-11-11</td>
<td>3198</td>
<td>19014</td>
<td>KU856580</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2014-11-21</td>
<td>4369</td>
<td>12655</td>
<td>KU856569</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2014-11-21</td>
<td>4369</td>
<td>13965</td>
<td>KU856570</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>1</td>
<td>2014-03-26</td>
<td>4369</td>
<td>2801</td>
<td>KX404800</td>
<td>0.00305,</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2014-03-26</td>
<td>4369</td>
<td>4338</td>
<td>KX404801</td>
<td>0.001125</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2014-03-26</td>
<td>4369</td>
<td>2030</td>
<td>KX404802</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>1</td>
<td>2014-03-26</td>
<td>4369</td>
<td>2967</td>
<td>KX404803</td>
<td>0.000763,</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2014-03-26</td>
<td>4369</td>
<td>2546</td>
<td>KX404804</td>
<td>0.000125</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2014-03-26</td>
<td>4369</td>
<td>3257</td>
<td>KX404805</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>1</td>
<td>2014-11-25</td>
<td>4369</td>
<td>2708</td>
<td>KX404806</td>
<td>0.001221,</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2014-11-25</td>
<td>4369</td>
<td>239</td>
<td>KX404807</td>
<td>0.000311</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2014-11-25</td>
<td>4369</td>
<td>338</td>
<td>KX404808</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>1</td>
<td>2014-02-24</td>
<td>4369</td>
<td>23798</td>
<td>KU856574</td>
<td>0.000381,</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2014-02-24</td>
<td>4369</td>
<td>24082</td>
<td>KU856575</td>
<td>0.000080</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2014-02-24</td>
<td>4369</td>
<td>11615</td>
<td>KU856576</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2014-02-24</td>
<td>4369</td>
<td>11903</td>
<td>KX404809</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2014-02-24</td>
<td>4369</td>
<td>27274</td>
<td>KX404810</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>2014-02-24</td>
<td>4369</td>
<td>8565</td>
<td>KX404811</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>1</td>
<td>2014-06-19</td>
<td>4369</td>
<td>2118</td>
<td>KU856577</td>
<td>0.000458,</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2014-06-19</td>
<td>4369</td>
<td>1341</td>
<td>KX404812</td>
<td>0.000072</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2014-09-19</td>
<td>4369</td>
<td>2203</td>
<td>KX404813</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2014-09-19</td>
<td>4369</td>
<td>1974</td>
<td>KX404814</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2014-09-19</td>
<td>4369</td>
<td>1587</td>
<td>KX404815</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1</td>
<td>2004-10-15</td>
<td>4369</td>
<td>8888</td>
<td>KX404816</td>
<td>–</td>
</tr>
<tr>
<td>O</td>
<td>1</td>
<td>2004-05-27</td>
<td>4369</td>
<td>10419</td>
<td>KU856578</td>
<td>–</td>
</tr>
</tbody>
</table>
(1) by counting the number of internal nodes in the tree – the better the resolution of a tree, the fewer polytomies and hence the more internal nodes the tree will have; and (2) by counting the number of fully resolved internal nodes (i.e. internal nodes having exactly two offspring) – a fully resolved tree will have two offspring for all internal nodes. These numbers were then compared to the corresponding measurements made on the whole-genome phylogeny (i.e. we computed the ratio between the values based on each individual sub-alignment and the value in the tree made from the full alignment) and the two measurements gave very similar results (data not shown). Fig. 3 shows the relative resolution measured using the

---

**Fig. 1.** Comparison between the partial NS1 gene and whole-genome phylogenies. Phylogenetic trees based on the partial NS1 gene dataset (a) and the whole-genome dataset (b) constructed in MrBayes version 3.2 applying the HKY model and estimating the number of invariable sites and gamma-rate distributions from the data. The Markov-chain Monte Carlo (MCMC) simulations were run for 50 million iterations. Branch labels represent posterior probabilities for each clade (Bayesian support values) and branch lengths represent substitutions per site, as indicated by the scale bars.
number of internal nodes as a function of the start of the 400 bp window. Windows in the first half of the genome result in higher resolution, which is in agreement with the observed higher variability in this region. In particular, a region around approximately nt 1100–1700 (corresponding to the 3’ end of the NS1 gene) displayed good resolution. The segment typically used for partial NS1 gene sequencing is in this alignment located between nt 605–932, and corresponds to a dip in the curve of resolving power (Fig. 3). It should be noted that this gene segment was not originally selected for the purpose of maximizing phylogenetic signal content, but rather for reliable PCR amplification and diagnostic purposes, and thus our results emphasizes that the partial NS1 sequence is not well suited for investigating the spread of infection within Denmark. From our analysis, it is also clear that there is no single window of the genome that provides a resolution close to that of whole-genome sequencing, and even the best windows obtained at most 60 % of the resolution compared to using the whole-genome.

**Estimating divergence times using sampling dates**

The whole-genome dataset presented here was characterized by low diversity and a phylogeny with a shallow root and low levels of rate variation between its branches (mean nt difference=0.0033, se=0.0001). Such datasets are often well described by a strict molecular clock and simple coalescent population growth, especially when the dataset represents a population subsample as in the present study [20, 21]. We used the BEAST2 software package to simultaneously estimate divergence times and absolute rates of molecular
evolution by including sampling dates in the phylogenetic analysis. BEAST2 works within a Bayesian paradigm and produce a probability distribution over possible values of the estimated times and rates.

The results of the time-stamped analysis strongly indicated that the most recent common ancestor (MRCA) for all sequences isolated from farm A was older than the MRCA for sequences isolated from farms B and C (Fig. 4). Specifically, the Bayesian analysis indicated with a posterior probability of 99.9% that the farm A ancestor was the oldest. The estimated median age of the farm A MRCA was 1.4 years old (95% credible interval=0.74–2.26 years), while the median age of the MRCA for farms B and C was 0.92 years (95% credible interval=0.49–1.5 years). The sequences from farms B and C furthermore formed a sub-tree within the tree spread out by the farm A sequences. These observations were all consistent with the hypothesis that the viruses were
transmitted from farm A to farms B and C, but not the opposite, and further supported the conclusions from the undated analyses above.

**DISCUSSION**

In this study, whole-genome sequencing and phylogenetic analyses were investigated and demonstrated to be valuable tools for determining AMDV transmission routes between mink farms. Previous phylogenetic analyses of AMDV in Denmark have been based on partial NS1 gene sequences. While a short and relatively conserved genomic region like this is suitable for diagnostic purposes, whole-genome sequences, which are longer and include more informative regions of the genome, are better suited for obtaining high phylogenetic resolution. We demonstrated this benefit by directly comparing whole-genome and partial NS1 gene phylogenies constructed from the very same data set, and plotted as a function of the starting position of the 400 bp window relative to the AMDV-G genome (x-axis).

**Fig. 3.** Quantification of phylogenetic resolution across the AMDV genome. Bayesian phylogenetic trees were created using extractions of 400 bp partitions spaced at 25 bp intervals across the whole-genome alignment. The relative resolution was measured by comparing the number of fully resolved internal nodes in each partition to the full genome (y-axis), and plotted as a function of the starting position of the 400 bp window relative to the AMDV-G genome (x-axis).

It is important to realize that viral phylogenies and transmission trees are two different things, and that their structures can be quite dissimilar. This is because the physical transmission of a progeny virus must happen some time after it has genetically split from its parental lineage, and thus the internal nodes in the phylogeny (which correspond to the event at which two viral lineages split) will be further back in time than the equivalent internal nodes in the transmission tree (which correspond to transmission of a viral lineage to another animal). The branching order of the trees can also differ, since a single animal host contains many related viruses, and the order in which different lineages are transmitted to other animals is not necessarily the same order in which these lineages split on the viral phylogeny [22]. However, if samples are collected close enough in time to the transmission event, it should be possible to establish the direction of spread with more certainty [11, 23]. Furthermore, if sampling rates are low (i.e. if relatively few animals are sampled from each farm), then the timing of coalescent events in the phylogeny is very similar to the timing of transmission events and the structures of the two trees will be quite similar [24]. Under all circumstances, better resolution of the viral phylogeny will provide more information about the underlying transmission tree. In the present study, it should be taken into consideration that the farms in this enzootic area of Denmark have a history of culling their entire animal populations at the end of every production season and therefore a linear rate of viral evolution cannot be assumed. Despite the fact that adding sampling dates improved the resolution of the tree, it should be kept in mind that the inferred clock-rate, and hence ages of the MRCAs, does not necessarily reflect the real underlying mutation rate. The results presented here (based on samples from 2014) outlined this challenge as the MRCAs of farms A, B, and C were estimated to be between 1.2 and 5.2 years, despite that the mink populations in these farms were culled at the end of the previous season (i.e. 2013). It cannot be ruled out that the virus infecting farm A was not a new virus, but originated from a previous outbreak and had persisted in the enzootic area of Denmark.
Fig. 4. Time-calibrated phylogenetic tree. The tree was constructed in BEAST2 using the HKY model and estimating the number of invariant sites and the gamma-rate distribution from the data, and applying a strict molecular clock and an exponential coalescent population.
environment and later re-infected the new animals at the farm. Overall, the results from this study illustrate that phylogenetic analyses in relation to AMDV outbreaks will be most valuable in recently infected farms and that the time-stamped model is useful for inferring relative time points for the transmission events.

In addition to carefully choosing the parameters for the phylogenetic reconstruction process (e.g. substitution model and priors), the input data – that is the samples and sequences – are important. Sequencing of PCR-amplified DNA might not necessarily capture the full intra-individual viral variation due to clonal amplification. The use of a PCR pre-amplification step is common practice and due to e.g. abundant host DNA and low amounts of viral DNA, a feasible approach for the purpose of detecting and typing AMDV [3]. Possible within-farm variation could theoretically be addressed by sequencing additional samples per farm and by including a contemporary test population in the analysis, as was done in the present study. However, due to practical circumstances in regard to farm operations, it will rarely be possible to get more than one or two samples per farm, but the low intra-farm diversity presented in the present study suggests it is nevertheless possible to tease out the inter-farm spread by including additional farms and adding sampling dates in the analysis.

The practical operations of a mink farm, e.g. open barns allowing for plentiful ventilation, possible wildlife access, external feed-supply transportation routes and the proximity between the farms illustrated by the case farms A and B (Fig. 5), can impede biosafety and should be considered carefully in the daily routines at the farm. A study of avian influenza in livestock showed that the wind direction on the date of transmission was correlated to the between-farm viral spread [25]. However, as mentioned above, the specific time points for the transmission events could not be determined in this study and therefore the impact of external factors such as the wind will remain a speculation. Presumably, neither a constant or exponential tree population model, nor a strict molecular clock, is the best way to describe the dynamics in a viral population. But on the other hand, to successfully infer relevant parameters using, e.g., the more complex birth-death models, additional knowledge about population parameters and a larger dataset would be required. The number of publicly available whole-genome AMDV sequences is currently very limited, and it was just recently that Canuti et al. [19], in a study of field strains originating from Canada, China and Germany, demonstrated inconsistent phylogenies due to large variation between tree structures over the viral genome and the presence of potential recombination events. Thus, we suggest that future studies should aim to investigate larger test populations sampled from broader timespans and from different countries in order to map the entire AMDV genomic diversity at a global level, thereby generating data that would benefit all mink-farming countries regardless of prior genomic surveillance strategy.

In conclusion, this study illustrates that whole-genome sequencing is better suited for reconstructing high-resolution phylogenetic relationships between AMDV isolates compared to shorter gene fragments such as the partial NS1 gene fragment currently used for AMDV typing in Denmark. Furthermore, by confirming an epidemiological transmission route hypothesis between three case farms, we show that whole-genome phylogenies supplement epidemiological data, such as AMDV prevalence and test history of the farms, to indicate the direction of transmission, thus suggesting a framework that could become an important tool to identify inter-farm spread of AMDV.

**METHODS**

**Farms A, B and C**

We investigated a case of AMDV transmission in a small Danish AMDV outbreak where there was a strong *a priori* hypothesis about the route of viral transmission based on epidemiological data (Kopenhagen Fur, personal communication). Farms A, B and C all tested serologically positive for AMDV in November 2014. In addition, farm A had tested positive in August 2014 with an estimated prevalence of 21.4 %, in September the prevalence of farm C was estimated to 0.4 %, while farm B tested negative for AMDV in 2014 (overview in Fig. 5). These prevalence estimates were based on serological testing of a percentage of the breeding animals at certain intervals according to the Danish control programme [5]. The much higher AMDV prevalence on farm A compared to farms B and C at the time of testing in November 2014 formed the basis of the initial hypothesis that farm A transmitted AMDV to farms B and C.

To put these case farms into context, a contemporary test population consisting of mink sampled from other Danish farms tested positive for AMDV in 2014, was included in the analysis. Potential within-farm variation was addressed by, when possible, sequencing more than one sample per farm. However, due to practical circumstances in regards to farm operations, it is rarely possible to access more than one or two samples per farm, but this number should be sufficient to estimate between-farm variation if the dataset is large enough. The geographic locations of the farms and their AMDV prevalence at their most recent AMD serology test is shown in Fig. 5. Sequences from farms N and O, both sampled in 2004, were included based on the hypothesis of being sufficiently evolutionarily distant to root a time-stamped tree, but not being so different that they would introduce...
Fig. 5. Map of Denmark and overview of the included farms. Insert shows close-up of farms A, B and C. For each farm the number of breeding animals during the season of 2014 and the AMDV prevalence as a percentage of positive animals in the most recent antibody test are indicated. Map created using Tableau Desktop version 9.2 and insert from Google DigitalGlobe 2015.
Chimera, SiScan, 3Seq and manufacturer PGM (Life Technologies, Carlsbad, CA) according to the instruction and sequencing on a 318 chip using the Ion Torrent Multi-Assay Core (Lyngby, Denmark) for library preparation. DNA was PCR-amplified as described previously [3] and submitted to the Technical University of Denmark (DTU). The full-length field strains were analysed using SimPlot [15] and all methods implemented in the RDP4 software package [16], i.e. RDP, GENECONV, BootScan, MaxChi, Chimera, SiScan, 3Seq and LARD, to investigate recombination and to describe variability across the genome.

Table 2. Overview of sequences retrieved from NCBI GenBank

<table>
<thead>
<tr>
<th>Viral strain</th>
<th>Country</th>
<th>Length (bp)</th>
<th>GenBank accession no.</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMDV-M173</td>
<td>Canada</td>
<td>4200</td>
<td>KT878958</td>
<td>[19]</td>
</tr>
<tr>
<td>AMDV-M195</td>
<td>Canada</td>
<td>4161</td>
<td>KT878959</td>
<td>[19]</td>
</tr>
<tr>
<td>AMDV-M228</td>
<td>Canada</td>
<td>4164</td>
<td>KT878960</td>
<td>[19]</td>
</tr>
<tr>
<td>AMDV-WM25</td>
<td>Canada</td>
<td>4169</td>
<td>KT878961</td>
<td>[19]</td>
</tr>
<tr>
<td>AMDV-Beijing</td>
<td>China</td>
<td>4802</td>
<td>KT329428</td>
<td>[33]</td>
</tr>
<tr>
<td>AMDV-LN1</td>
<td>China</td>
<td>4543</td>
<td>GU183264</td>
<td>[34]</td>
</tr>
<tr>
<td>AMDV-LN2</td>
<td>China</td>
<td>4566</td>
<td>GU183265</td>
<td>[33]</td>
</tr>
<tr>
<td>AMDV-LN3</td>
<td>China</td>
<td>4566</td>
<td>GU268982</td>
<td>[33]</td>
</tr>
<tr>
<td>AMDV-SL3</td>
<td>Germany</td>
<td>4718</td>
<td>X97629</td>
<td>[35]</td>
</tr>
<tr>
<td>AMDV-Utah</td>
<td>Antigen</td>
<td>4369</td>
<td>KU513988</td>
<td>[3]</td>
</tr>
<tr>
<td>AMDV-G1</td>
<td>Antigen</td>
<td>4369</td>
<td>KU513985</td>
<td>[3]</td>
</tr>
<tr>
<td>GFAV</td>
<td>USA</td>
<td>4441</td>
<td>JN202450</td>
<td>[36]</td>
</tr>
</tbody>
</table>

Viral samples and preparation

One blood sample per animal was collected from a total of 48 animals originating from farms A–M, in addition to one archive blood sample from one animal from farms N and O, respectively (Table 1). Total DNA was extracted from each blood sample using QIAmp MinElute Virus Spin Kit (Qiagen, Hildren, Germany) and eluted in 50 μL low TE-buffer according to the manufacturer’s instructions. The viral DNA was PCR-amplified as described previously [3] and submitted to the Technical University of Denmark (DTU) Multi-Assay Core (Lyngby, Denmark) for library preparation and sequencing on a 318 chip using the Ion Torrent PGM (Life Technologies, Carlsbad, CA) according to the manufacturer’s instructions. Raw data in fastq format were processed as described previously [3]. Briefly, the steps prior to sequence assembly included QC, trimming, error correction and mapping to the AMDV-G reference with accession no. NC_001662, followed by naming each sequence according to Table 1. For a global comparison, previously published complete coding sequences for AMDV field strains originating from Canada, China and Germany, the complete coding sequences for the antigen strains AMDV-Utah and AMDV-G, in addition to the closely related Gray fox ampodarvovirus (GFAV), were retrieved from NCBI GenBank (overview and references in Table 2).

Sequence analysis

Intra-farm diversity was estimated by calculating the mean pairwise nucleotide distance and its corresponding standard error (SE) between the sequences collected from each farm [27]. The full-length field strains were analysed using SimPlot [15] and all methods implemented in the RDP4 software package [16], i.e. RDP, GENECONV, BootScan, MaxChi, Chimera, SiScan, 3Seq and LARD, to investigate recombination and to describe variability across the genome.

Phylogenetic reconstructions

Evaluating the use of whole-genome sequences for reconstructing phylogenies

The whole-genome sequences, including GFAV for rooting, were aligned at nucleotide level using MAFFT V.7.205 [28] and converted to nexus format. The 327 bp region corresponding to the DNA sequence flanked by the primers used for ‘conventional PCR’ amplification [8] was extracted from the alignment, thereby creating two datasets: ‘partial NS1 gene’ and ‘whole genome’. The best fitting substitution model for each dataset was selected using the program jModeltest [29]. The phylogenetic relationships were inferred in a Bayesian framework with Markov-chain Monte Carlo (MCMC) sampling in MrBayes version 3.2.3 [30], applying an HKY model with an estimated proportion of invariable sites, four gamma distribution rate categories and Dirichlet priors. The MCMC was run for 50 million generations for each of the datasets. The first 25 % of the samples was discarded as burn-in, and effective sample size (ESS) values above 400 for all parameters and standard deviation of split frequencies below 0.001 were considered as indications that the MCMC had converged successfully. FigTree version 1.4.2 was used for tree manipulations such as rooting and for visualization. To facilitate visualization, the outgroup sequence (GFAV) was removed from the maximum clade credibility (MCC) trees.

Estimating divergence times using sampling dates

The feasibility of using viral sampling dates to perform divergence time dating of the ancestors of the isolates involved in the outbreak was investigated in a Bayesian framework implemented in BEAST version 2.4.1 [31]. Often a simple model such as a strict molecular clock and a coalescent constant population growth prior size model is a useful starting point for analysis [20], especially if the dataset represents a subsample of the population as in the present study. In addition, both a strict and relaxed molecular clock, a coalescent constant and an exponentially growing tree population model was tested, and sampling was done with MCMC simulations run for 100 million generations to obtain estimates of the posterior distributions. The whole-genome dataset, excluding the GFAV isolate in order to avoid introducing long branch-lengths, was aligned as described above and used for this analysis. The MCMC log files were inspected for convergence based on the magnitude of ESS-values, and shapes of the traces and marginal posterior probabilities using Tracer version 1.6 [32]. Treeannotator version 1.8.2 was used for summarizing the tree log files, and FigTree version 1.4.2 (both distributed with the BEAST package [31]) were used for tree manipulations such as rooting and visualization. The model was also run without data to confirm that the priors did not influence the results unduly.

Funding information

This work was funded by the Research Foundation of the Danish Fur Breeder’s Association and the ‘Innovation Fund Denmark’ (award number: 16479).
Acknowledgements
The co-workers at Kopenhagen Diagnostics, Kopenhagen Fur, are thanked for collecting the samples and for providing valuable background information about the farms’ history.

Conflicts of interest
The authors declare there are no conflicts of interest.

Ethical statement
The work presented in the paper does not consist of experiments performed on humans or animals, but were performed using tissue samples.

References