Multiple introductions of salmonid alphavirus from a wild reservoir have caused independent and self-sustainable epizootics in aquaculture

Marius Karlsen,1 Britt Gjerset,2 Tove Hansen1 and Andrew Rambaut3

Correspondence
Marius Karlsen
marius.karlsen@pharmaq.no

1PHARMAQ AS, Harbitzalléen 2a, 0275 Oslo, Norway
2Norwegian Veterinary Institute, Pb 750 Sentrum, 0106 Oslo, Norway
3University of Edinburgh, Edinburgh, UK

Salmonid alphavirus (SAV) causes infections in farmed Atlantic salmon and rainbow trout in Europe. Genetic diversity exists among SAV strains from farmed fish and six subtypes have been proposed based on genetic distance. Here, we used six full-genome sequences and 71 partial sequences of the structural ORF to estimate the evolutionary rate of SAV. The rate, 2.13×10−4 nt substitutions per site per year, was further used to date evolutionary events in a Bayesian phylogenetic framework. The comparison of these dates with known historical events suggested that all six subtypes diverged prior to the twentieth century, earlier than the first attempts to introduce and farm rainbow trout in Europe. The subtypes must therefore have existed in a wild reservoir, as yet unidentified. The strains of each subtype, with the exception of subtype 2, have a common ancestor that existed after the 1970s – the start of modern farming of Atlantic salmon. These ancestors are likely to represent the independent introductions to farmed fish populations from the wild reservoir. The subtypes have developed subsequently into self-sustainable epizootics. The most parsimonious phylogeographic reconstruction suggested that the location of the wild reservoir is in or around the North Sea. After the initial introductions to aquaculture, further transmission of SAV was likely related to the industry infrastructure. This was exemplified by the finding of genetically identical subtype 2 and 3 strains separated by large geographical distances, as well as genetically distinct co-circulating lineages within the same geographical area.

INTRODUCTION

Atlantic salmon (Salmo salar) and rainbow trout (Oncorhynchus mykiss) are the most extensively farmed salmonid species in Europe. Atlantic salmon is a native species in Europe, and was developed into a farmed animal during the 1960s in Norway and the UK. Rainbow trout is a non-native species in Europe and was first introduced for farming purposes from western North America around 1880. It was developed into an intensively farmed animal during the 1960s using similar farming systems to those used for the Atlantic salmon (Hershberger, 1992).

A major pathogen of the European farmed salmon and trout, salmon pancreas disease virus (family Togaviridae), commonly named salmonid alphavirus (SAV), causes pancreas disease or sleeping disease (McLoughlin & Graham, 2007; Nelson et al., 1995; Weston et al., 2002). Diseased fish show lethargic behaviour, necrosis of pancreatic tissue and inflammation of heart and skeletal muscle (McLoughlin et al., 2002). The initial descriptions of pancreas disease originate from the late 1970s and early 1980s in Irish and Scottish Atlantic salmon farms (Munro et al., 1984). During the 1980s, outbreaks with similar histopathological lesions were reported from Norwegian salmon farms (Poppe et al., 1989) and French rainbow trout farms (then referred to as sleeping disease) (Boucher & Baudin Laurencin, 1994).

The genome of SAV is an 11.9 kb positive-sense ssRNA molecule with two ORFs, encoding non-structural and structural polyproteins (Weston et al., 2002). The molecule is capped at the 5′ end and polyadenylated at the 3′ end, and functions as an mRNA molecule once it enters a host cell (Welsh et al., 2000; Weston et al., 2002). SAV also transcribes a subgenomic mRNA that is translated subsequently as a polyprotein containing the structural proteins capsid, E3, E2, 6K/TF and E1 (Villoing et al., 2000; Welsh et al., 2000; Weston et al., 1999, 2002). The capsid contains a putative serine protease, and is likely to be released by autoproteolytic cleavage to the cytoplasm where it binds to viral RNA genomes and assembles into nucleocapsids.

The GenBank/EMBL/DDBJ accession numbers for the new sequences of salmonid alphavirus are KF668057–KF668084.

One supplementary table is available with the online version of this paper.
Transmission patterns of SAV in aquaculture are not clear. Viral RNA has been detected in the water during viraemia in tank trials with salmon, and cohabitant salmon and trout are infected readily (Andersen et al., 2010; Nelson et al., 1995). It therefore appears likely that the virus transmits by water contact once it has entered a farm. It has been suggested that transmission also occurs from farm to farm via water currents (Kristoffersen et al., 2009; Viljugrein et al., 2009) and that transportation of fish is responsible for transmission across larger distances (Karlsen et al., 2006). Viral RNA has also been reported from eggs and fry, and the possibility of transgenerational transmission cannot be ruled out (Bratland & Nylund, 2009). Phylogenetic analyses of the partial coding region of E2 from SAV strains isolated from farmed Atlantic salmon and rainbow trout in Europe have suggested the existence of six genetic subtypes, SAV1–6 (Fringuelli et al., 2008). Five of these subtypes (1, 2, 4, 5 and 6) have been reported in Ireland or the UK (Graham et al., 2012), while the remaining subtype 3 has so far only been found in Norway. All strains that have been sequenced from Continental Europe to date belong to subtype 2 and this subtype was also found recently in Norway (Hjortaas et al., 2013). In Norway, Ireland and the UK, most outbreaks have been reported from marine farms, while outbreaks in other European countries are common in freshwater farms of rainbow trout (Boucher & Baudin Laurencin, 1994).

Wild reservoirs of SAV are largely unknown, but strains belonging to subtype 5 have been sequenced from wild caught common dab (Limanda limanda), long rough dab (Hippoglossoides platessoides) and plaice (Pleuronectes platessa) in the Shetland Islands (Snow et al., 2010). The high genetic similarity of these strains to subtype 5 strains from farmed fish in Scotland suggests a recent epidemiological link between these populations, but demonstrates that SAV is not exclusively a pathogen of salmonids, broadening the possibilities for natural reservoirs.

SAV, like other RNA viruses, has a rate of evolution that is considerably higher than most other organisms (Karlsen et al., 2006). This high rate means that sampling of viral sequences over even a relatively modest span of time and space can reveal sufficient genetic diversity to disclose information about transmission pathways and important evolutionary events. Here, we used sequence data to measure the rate of evolution of SAV in order to obtain a timescale on SAV evolution. Based on these data, we proposed that a diverse wild reservoir of SAV is likely to exist in or around the North Sea, and that at least six introductions from this reservoir have occurred to the farming industry to create independent and self-sustainable epizootics.

**RESULTS**

**SAV evolves at a rate similar to other alphaviruses**

First, six nearly complete genome sequences and 71 sequences covering the partial structural protein-coding region (all with known date of sampling) were used to calculate an evolutionary rate for SAV (Table S1, available in JGV Online). The estimated rate of evolution was $2.24 \times 10^{-4}$ nt substitutions per site per year based on pairwise distances computed from four full-length subtype 3 genome sequences that were separated in time by 5 years. This represents the maximum possible rate of evolution as the method assumes that all the observed genetic distance was generated in these 5 years. The same analysis done on the 71 sequences covering only the partial structural ORF gave a similar rate, indicating that rate estimates made from this genetic region are representative of the whole genome. A Bayesian statistical approach using BEAST (Drummond et al., 2012) was then implemented on a dataset containing the 71 partial sequences (total alignment length: 1872 nt). The estimated evolutionary rate obtained by this method was $2.13 \times 10^{-4}$ nt substitutions per site per year with a 95% highest posterior density (HPD) interval of 1.33–2.89. These rates are similar to the rates estimated for other alphaviruses (Table 1).

**Divergence of all SAV subtypes occurred prior to the development of modern aquaculture**

The rate obtained from Bayesian statistical analyses was used to estimate dates of divergence between SAV lineages (Fig. 1). Subtypes 1–5 all share a common ancestor that existed between 1557 and 1813. This ancestor diverged into one lineage containing subtypes 1, 4 and 5, and another lineage containing subtypes 2 and 3. The first lineage shared a common ancestor between 1808 and 1921, while subtypes 2 and 3 shared an ancestor between 1642 and 1857. The dates significantly precede modern salmonid farming and imply that these evolutionary divergences occurred in wild host species.

**Evidence of at least six introductions to aquaculture from the wild reservoir**

The most recent common ancestor (MRCA) in each subtype was dated, and compared to the historical events of (i) the introduction of farmed rainbow trout to Europe and (ii) the start of Atlantic salmon farming in Europe (Fig. 2). The 95% HPD intervals of the MRCA of all subtypes,
except subtype 2, span a period after the start of Atlantic salmon farming. The 95% HPD interval of subtype 2 located to the period after the introduction of farmed rainbow trout to Europe, but before farming of Atlantic salmon started. Therefore, each subtype is likely to represent a single and separate introduction to farmed fish from a wild reservoir. Subtype 6 was not included in this analysis due to insufficient sequence data.

All currently known SAVs are likely to originate from a common ancestor that existed in or around the North Sea

The premise that each subtype most likely represents independent emergence in farmed fish from a wild reservoir allowed us to better characterize the geographical location of the original wild reservoir. A maximum-likelihood phylogenetic analysis of 80 sequences coding for SAVs

### Table 1. Evolutionary rates of selected alphaviruses

Rates obtained for SAV in the present study are compared with published rates for other alphaviruses. Where an interval is reported in parentheses, this is the 95% interval. Where a range is reported, the study gave more than one estimate for different types of virus.

<table>
<thead>
<tr>
<th>Virus</th>
<th>Gene</th>
<th>Rate [nt substitutions per site per year (×10⁻⁴)]</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salmonid alphavirus C–E3–E2</td>
<td>Genome</td>
<td>2.13 (1.33, 2.89)</td>
<td>Present study</td>
</tr>
<tr>
<td>Salmonid alphavirus Genome</td>
<td></td>
<td>2.24</td>
<td>Present study</td>
</tr>
<tr>
<td>Eastern equine encephalitis virus</td>
<td>Structural polyprotein</td>
<td>2.1 (1.7, 2.6)</td>
<td>Arrigo et al. (2010)</td>
</tr>
<tr>
<td>Chikungunya virus Genome</td>
<td>Genome</td>
<td>2.30–8.41</td>
<td>Volk et al. (2010)</td>
</tr>
<tr>
<td>Buggy Creek virus C–E3–E2–E1</td>
<td></td>
<td>1.66–2.64</td>
<td>Padhi et al. (2008)</td>
</tr>
<tr>
<td>Venezuelan equine encephalitis virus</td>
<td>Structural polyprotein</td>
<td>1.28 (0.21, 2.53)</td>
<td>Auguste et al. (2010)</td>
</tr>
<tr>
<td>Highland J virus Genome</td>
<td>Genome</td>
<td>1.22</td>
<td>Allison &amp; Stallknecht (2009)</td>
</tr>
</tbody>
</table>

### Fig. 1. Dated phylogeny of SAV strains using BEAST.

The phylogenetic tree was obtained using a HKY + G model for nucleotide substitution with a GMRF skyride model for changes in population size. The tree was calibrated to a timescale inferred from the sequence information. The posterior probability is indicated for relevant nodes. Clades corresponding to subtypes 1–5 are colour coded.
for partial E2 sequences (357 nt) of SAV was undertaken in order to correlate geographical data to the different subtypes/introductions. The analysis identified clades corresponding to the previously suggested subtypes, with limited genetic diversity within each clade (Fig. 3). It further confirmed that subtype 6, found in Ireland, is the most divergent SAV lineage known to date. The most parsimonious geographical location of each introduction was then studied by identifying and comparing the location of the most divergent sequence in each subtype to the geographical location of the most ancestral clade. This analysis located all the introductions of subtypes 1, 2, 4, 5 and 6 to the UK or Ireland. Subtype 3 was located to Western Norway.

‘Sleeping disease’/subtype 2 epizootic has its origin around Scotland and spread extensively throughout Europe after 1958

The maximum-likelihood analysis addressing the intra-subtype genetic diversity and geographical location of isolates (Fig. 3) was correlated with divergence dates from the Bayesian analysis (Fig. 1). All subtype 2 sequences shared a common ancestor that was likely located to the UK between 1882 and 1958, and that gave rise to a lineage that has been found in Atlantic salmon sites in Scotland and Norway, and a second lineage that has been spread widely to farmed rainbow trout in Continental Europe. Insufficient sequence data preclude dating of events after 1958 in this subtype.

Subtype 3 epizootic was introduced to the Norwegian industry in the 1980s and later developed into two independently circulating lineages

The Bayesian phylogenetic analysis suggested that all known subtype 3 isolates from farmed Norwegian fish originated from a common ancestor that existed between

---

**Fig. 2.** Estimated dates for the existence of the MRCA of each subtype (introductions to European farmed fish). The dates were inferred from the phylogenetic tree in Fig. 1, and compared to the dates of (i) the introduction of farmed rainbow trout to Europe and (ii) the beginning of modern Atlantic salmon farming. Boxes represent 95% HPD intervals, with the line indicating the mean.

---

**Fig. 3.** Geographical location of sampled SAV strains and estimated ancestors. A phylogram was constructed using maximum likelihood with a GTR + G + I model for nucleotide substitution. The most likely tree was bootstrapped using 500 replicates in MEGA5. Relevant bootstrap values are shown. Branches are coloured according to the most parsimonious geographical location in Europe. Branch colours correspond to colours given in the map. Bar, number of base substitutions per site. Maps were obtained from d-map.com.
1985 and 1996 (Fig. 2), corresponding well with the first descriptions of pancreas disease in Norway. Since 1996, subtype 3 has diverged into two co-circulating lineages, SAV3A and SAV3B, each exhibiting strong statistical support (Fig. 4).

The spatial and temporal dispersal of isolates in SAV3A and SAV3B was compared (Fig. 4). Isolates in SAV3A were reported in the period 1997–2011, while those in SAV3B were reported between 2003 and 2008. Comparison of geographical distribution revealed that strains belonging to SAV3A were isolated in the counties of Rogaland, Hordaland, Sogn og Fjordane, Møre og Romsdal, Sør-Trøndelag, Nordland, Troms and Finnmark. Although only five available sequences belong to SAV3B, they originated from geographically widespread locations including the counties of Hordaland, Sogn og Fjordane and Finnmark. The lineages are not distinguished by association with specific geographical regions, and viruses belonging to both lineages appeared to have been transmitted over large distances, suggesting a human influence on dispersal.

**DISCUSSION**

Previous studies have identified considerable genetic diversity among strains of SAV that are isolated from farmed fish (Fringuelli *et al.*, 2008; Hodneland *et al.*, 2005). By using two different approaches we have obtained a robust evolutionary rate for SAV and have thus been able to set a timescale for the evolutionary history of the virus. Based on this timescale, we suggest that each of the subtypes represents separate introductions to aquaculture from a wild reservoir. This implies that SAV has likely been introduced from a wild reservoir at least six times in the history of modern aquaculture. The subtypes have been or are still active as self-sustainable epizootics in populations of farmed fish. The presence of genetically distinct strains in the same time and space, combined with genetically identical strains separated by large geographical distances, suggests that human transport of virus played a significant role in dispersal after the introduction to aquaculture.

Transmissions to farmed fish from a wild reservoir are not common events, but appear to have occurred more frequently around the UK/Ireland than in other parts of Europe. In this region, five of the six known subtypes, including the more divergent subtype 6, have been found (Graham *et al.*, 2012) and their introduction from wild hosts to the farming industry is likely to have occurred here. The last subtype, SAV3, was more likely introduced in Western Norway, also an area in direct association with the North Sea. The only available sequences from wild fish are from Scotland (Snow *et al.*, 2010). These sequences differ from known contemporary salmonid subtype 5 strains by four to eight mutations in a total alignment length of 298 nt, which equals a genetic identity of 97.3–98.7 %.

Salmonid subtype 5 sequences used for the dating analysis all share a recent common ancestor (95 % HPD interval: 1994–2003). Unfortunately, the sequences from wild flatfish were too short for inclusion in this analysis. Their degree of genetic identity to known sequences from salmon suggests, however, that a host-switch event between wild
flatfish and farmed salmonid fish is likely to have occurred recently. It is possible that this represents the authentic introduction of subtype 5 to farmed fish from the wild reservoir, considering the young age of the MRCA of this subtype. The high genetic diversity of SAVs in the area around the North Sea, combined with the finding of SAV in marine flatfish (Snow et al., 2010), could thus be indicative of a marine wild reservoir. SAVs belonging to subtypes 1, 2 and 3 have, however, also been reported from Atlantic salmon and rainbow trout kept in freshwater environments (Bratland & Nylund, 2009; Lester et al., 2011; Villoing et al., 2000).

The subtype 2 epizootic in rainbow trout in Continental Europe, traditionally referred to as sleeping disease, appears to share a relatively recent ancestor with Scottish SAV strains found in Atlantic salmon in the sea. This, combined with the genetic homogeneity among SAV2 strains from Italy, France and Spain, suggests that SAV in Continental Europe was introduced from the British Isles.

In the history of fish farming in Norway, we find evidence for only one introduction from wild fish. This introduction occurred around the time of the first reports of pancreas disease in Norway (Poppe et al., 1989) and resulted in the epizootic caused by subtype 3. Recent outbreaks of subtype 2 strains (Hjortaas et al., 2013) represent a second introduction to Norway, but the genetic identity of these strains to those sequenced from Scottish farmed salmon in 2006 makes it probable that a virus was introduced first to farmed fish in Scotland and later transported to Norway with biological material.

Within subtype 3, we found statistical support for the existence of two distinct lineages, SAV3A and SAV3B. These lineages have previously been observed, but their significance has not been understood (Jansen et al., 2010; Karlsen et al., 2006). Our findings that these lineages have strong statistical support and have evolved separately at least since 2003 suggest that they are of value for epizootiological studies. Both lineages have been reported from geographically distant locations such as the counties of Hordaland and Finnmark, and within the same area at the same time. This co-circulation of distinct lineages within the same temporal and spatial area is not consistent with a model of transmission where virus is spread solely by passive transport in water currents. Human influence on the dispersal of SAV has been argued for previously, both in Norway and around the British Isles (Graham et al., 2012; Karlsen et al., 2006). This transmission mode may be important not only when the virus is transmitted over large distances, but also within the geographical areas that are considered to be enzootic. The ability to link these two SAV3 lineages to factors in the industrial infrastructure could give useful information about such transmission patterns of SAV in the Norwegian aquaculture industry.

Knowledge of the diversity and origin of SAV epizootics can be used for understanding transmission pathways, identification of potential wild reservoirs and designing effective vaccination and mitigation strategies. The apparent lack of reseeding of virus from the wild reservoir to farmed fish in Norway and Continental Europe suggests that it should be possible to implement an effective mitigating strategy in these areas. Around the British Isles, introduction from wild hosts is more common and might complicate mitigation strategies.

**METHODS**

**RNA extraction, reverse transcriptase (RT)-PCR and sequencing.** RNA was extracted from mixed heart and mid-kidney tissue as described previously (Jansen et al., 2010). Briefly, extracted RNA was reverse transcribed using random primers and SuperScript III RT (Invitrogen); 2.5 μl cDNA with 0.15 μM each primer was added in a final PCR volume of 25 μl (HotStarTaq PCR; Qiagen) under the following conditions: denaturation for 15 min at 95°C, followed by 40 amplification cycles of 94°C for 30 s, 59°C for 30 s and 72°C for 90 s, and finally 72°C for 10 min. The RT-PCR products were examined by agar gel electrophoresis and purified using the ExoSAP-IT protocol (USB) prior to sequencing with BigDye Terminator v3.1 Cycle Sequencing kits (Applied Biosystems). Consensus sequences were generated using Sequencher (Gene Codes). The 1871 bp region within the capsid–E3–E2–6K genes was amplified using partial overlapping sequences with the primer pairs F1600, R2357, F2234, SAV20R, E2666F and E2YR described in Karlsen et al. (2006) and Jansen et al. (2010). New sequences were deposited in GenBank and accession numbers are given in Table S1.

**Estimates of genetic distance based on full-length genomes.** Six full-length SAV genomes were retrieved from GenBank (Table S1) and aligned using MUSCLE (Edgar, 2004), and the genetic distance between strains was calculated using a General Time Reversible model (Tavaré, 1986) for nucleotide substitution in conjunction with a discrete gamma model of rate heterogeneity amongst sites. The mean genetic distance between the four SAV3 genomes was 1.03 × 10^{-3} (range 0.34–1.12 × 10^{-3}) and the greatest distance was divided by the time that separated the strains (5 years) to obtain an upper bound on the possible rate of evolution. The same analysis was done on an alignment covering only the partial structural ORF (the same region that was used for subsequent Bayesian phylogenetic analyses) in order to ensure that the obtained rate was not sensitive to the choice of genomic region.

**Bayesian phylogenetic analysis.** A total of 71 sequences covering the partial structural ORF were aligned using MUSCLE and used for a Bayesian phylogenetic analysis in BEAST 1.6.1 (Drummond et al., 2012) (Table S1). The total alignment length was 1872 nt. Each sequence was labelled with the date of sampling to the precision known from available information. When only the year of sampling was available, the age of the sequence was introduced as a parameter and sampled uniformly from within that year (Shapiro et al., 2011). A Hasegawa–Kishino–Yano model of nucleotide substitution (Hasegawa et al., 1985) was used with rates amongst sites modelled by a discrete gamma distribution. The ratio of transition rate to transversion rate and the shape parameter of the gamma distribution were estimated jointly with the other parameters of the model. A highly flexible model of changing population size (GMRF skyride) (Minin et al., 2008) was employed as a prior on the evolutionary tree. Other priors were left as the default setting in BEAST. The Markov chain Monte Carlo algorithm was run for 10^7 generations and all effective sample size values were confirmed to be >200 when inspected in the software Tracer.
Phylogeographic analysis. A maximum-likelihood analysis was performed in MEGA5 (Tamura et al., 2011) on an alignment that was identical to that used for Bayesian analysis, except that nine shorter sequences with geographically relevant information connected to them were included (Table S1). A GTR model for nucleotide substitution with a discrete gamma distribution (Yang, 1994) and with a proportion of invariable sites (GTR+G+I) to model rate differences between sites was used. The obtained tree was bootstrapped with 500 replicates and branches were coloured according to their estimated geographical origin. A neighbour-joining analysis using the Tamura–Nei model of nucleotide substitution (Tamura & Nei, 1993) with a discrete gamma distribution for rate differences between sites was also conducted on the same dataset to confirm that tree topology could be reproduced with different methods.

ACKNOWLEDGEMENTS

The authors are grateful to Marianne Bordevik (PHARMAQ AS, Oslo, Norway) for assisting with PCR analyses.

REFERENCES


