A novel variant of the infectious salmon anaemia virus (ISAV) haemagglutinin gene suggests mechanisms for virus diversity.

Carey O. Cunningham¹*, Alison Gregory¹, Julia Black¹, Ian Simpson² & Rob S. Raynard¹.

¹FRS Marine Laboratory, PO Box 101, 375 Victoria Road, Aberdeen AB11 9DB, UK. ²FRS Freshwater Fisheries Laboratory, Faskally, Pitlochry, Perthshire, PH16 5LB

Abstract
A novel variant of the highly polymorphic region of the infectious salmon anaemia virus (ISAV) haemagglutinin gene was obtained from a wild adult salmon returning to a river on the East coast of Scotland. The predicted amino acid sequence contained all motifs previously found in ISAV isolates. This discovery provides further evidence for a pool of the virus in the marine environment and suggests that deletion may be employed to generate diversity of ISAV. The prospect of large scale variation in this gene has implications for the methods employed to analyse relationships between ISAV isolates.

Introduction
Infectious salmon anaemia disease has caused serious losses in Atlantic salmon farming in the Northern Hemisphere. The causative agent, infectious salmon anaemia virus (ISAV), has been identified as an orthomyxovirus (Falk et al., 1997; Mjaaland et al., 1997), but is different to the influenza and thogoto viruses (Krossøy et al., 1999). Only one segment of ISAV, encoding a polymerase PB1, shares significant homology with other members of the Orthomyxoviridae, and the classification of a separate genus Isavirus is proposed in the International Committee on Taxonomy of Viruses Index of Viruses version 3 (available at http://www.ncbi.nlm.nih.gov/ICTVdb/Ictv/index.htm).

The haemagglutinin gene of influenza viruses is known to be highly variable and important in antigenicity (Nicholson et al., 1998). The haemagglutinin gene of ISAV was identified recently and a highly polymorphic region (HPR) within the gene identified (Clouthier et al., 2002; Krossøy et al., 2001a; Rimstad et al., 2001). A large number of different HPR sequences have been found, and it has been suggested that this region could account for antigenic variation or be a result of recombination (Devold et al., 2001; Kibenge et al., 2001). Nucleotide sequence analysis of ISAV is an important tool for epizootiological study (Blake et al., 1999; Cunningham & Snow, 2000; Devold et al., 2001; Stagg et al., 2001). The HPR is the most variable part of the ISAV genome identified to date, and therefore appears ideal for this purpose. Here, we present novel variants of the ISAV haemagglutinin HPR, one of which is the longest identified so far. The discovery of this variant suggests that novel strains may arise through deletion and that many strains may co-circulate. Because of this variation and the potential mechanisms re-
 responsible for it, great care is required when selecting appropriate methods for phylogenetic and epizootiological analysis of the HPR.

**Materials & Methods**

Atlantic salmon (*Salmo salar* L.) were sampled from farms in the Faroe Islands during an outbreak of ISA and adult wild fish returning to rivers in Scotland, netted at the east coast. Kidney tissue from the Faroe islands and gill tissue from the wild fish was sampled into RNA Later (Ambion Inc), and kidney from some Faroes samples used to culture virus. The Norwegian Glesvaer isolate was received from the National Veterinary Institute, Oslo. ISA was cultured in salmon head kidney cells as described by Dannevig *et al.* (1995).

RNA was extracted from the tissues and cell supernatant using TRIzol (Invitrogen) according to the manufacturer’s instructions. Reverse transcription was performed from 2 µg of RNA according to Mjaaland *et al.* (1997) using MMLV-RT (Invitrogen). One aliquot of 5 ml cDNA was used in RT-PCR detection of ISA infected material by the method of Mjaaland *et al.* (1997). An aliquot of 5 µl cDNA from ISA infected samples was added to 45 µl PCR mix that contained 1 x reaction buffer (Bioline), 1.5 mM MgCl₂, 0.2 mM dNTPs (each), 2 U Taq polymerase (Bioline), 28 pmol primers HA FP For (5’ CAGACATTGACTGGAGTAGAATTG 3’) and HA FP Rev (5’ CTGCAATCCAAATACATGCCT 3’), and diethyl pyrocarbonate (DEPC) treated distilled water. All reactions were overlaid with mineral oil. PCR was performed by 30 cycles of 94°C for 60 seconds, 55°C for 60 seconds and 72°C for 90 seconds, followed by a final extension step of 72°C for 5 minutes. PCR products were visualised by electrophoresis on a 1.5% (w/v) agarose gel stained with ethidium bromide and viewed under UV transillumination.

PCR products were purified using Wizard PCR Preps (Promega) and were quantified on a 1% agarose gel alongside mass markers of known molecular weight. The HPR was sequenced twice in both the forward and reverse direction, using the same primers used in PCR with the dRhodamine Terminator Ready Reaction Mix and electrophoresis on an ABI 377 automated DNA sequencer (Applied Biosystems). Analysis of resulting sequences was carried out using Sequencher software (Gene Codes Corporation).

Predicted amino acid sequences were aligned with other ISA HPR sequences available on the SWALL or TrEMBL sequence databases. TrEMBL supplements the SWISS-PROT Protein Sequence Data Bank at http://www.ebi.ac.uk/swissprot/ and contains the translations of all coding sequences (CDS) present in the EMBL Nucleotide Sequence Database not yet integrated in SWISS-PROT (Bairoch & Apweiler, 1997).

**Results**

Tissue and cell culture material from ISA infected samples was used as template to amplify the region of the haemagglutinin gene encompassing the highly polymorphic region (HPR). PCR products were approximately 285 bp. Nucleotide sequences were deposited in the EMBL nucleotide database under accession numbers AJ440971 (Scottish), AJ440970 and AJ440972 (Faroe Islands). Independently, we obtained a third sequence from the Faroe Islands that had already been deposited in the database by Nylund *et al.* (AF536263).
Comparison of amino acid sequences is more easily visualised, so predicted amino acid sequences were determined and used for comparisons. Two different sequences (f, g) were found in the same farm in the Faroes. Sequences f, g and h from the Faroe islands and a from the Scottish wild fish are shown in Table 1.

A BLAST search of the SWALL database identified many different HPR sequences, as reported previously (Devold et al., 2001; Kibenge et al., 2001; Krossøy et al., 2001; Rimstad et al., 2001). One representative of each HPR type from each country is shown alongside the sequences identified in this study in Table 1. The HPR sequence from the Scottish wild fish reported here was remarkable in that it contained all amino acid motifs previously seen in ISAV in a single sequence. All other HPR sequences appear to be truncated versions of this.

The predicted amino acid sequence of the Norwegian Glesvaer isolate was found to be identical to accession number AAF32361 (sequence m), determined by (Rimstad et al., 2001). Another predicted sequence from this isolate; sequence o from (Kibenge et al., 2001), showed unusual variation from the other HPR types. Alignment of the nucleotide sequences from which these amino acid sequences were predicted (AF220607 and AF283998, respectively), along with sequences from another isolate, 7/92 (AF364898), revealed that the AF283998 sequence was more similar to that from isolate 7/92 than to the other Glesvaer sequence AF220607 (Figure 1). Sequence AF283998 contains 3 nucleotide insertions that result in frameshifts relative to sequence AF364898 and predict the amino acid sequence reported by Kibenge et al. (2001).

## Discussion

The novel ISAV HPR sequence obtained here is the longest found to date and contributes valuable data for the exploration of the
mechanisms employed by this novel orthomyxovirus to generate diversity and thus evade the host immune system. The HPR region of the ISAV HA is thought to form the first part of the ectodomain of the HA protein (Krossøy et al., 2001a; Rimstad et al., 2001), and variation in the length of the HPR could significantly affect structure, and therefore the antigenicity or virulence of the virus. Before exploring this in more detail, it is worth examining some of the unusual features of the sequences used in this study.

Sequences such as b, c and e in Table 1 all contain NN at positions where all other sequences have KI. This could easily result from 2 nucleotide substitutions during PCR or cloning, or miscalls by an automated sequencer; AAG ATC encoding KI and AAC ACC encoding NN. Sequence z is shown in Table 1 aligned with a single E in the middle. Examination of the nucleotide sequences reveals that only a single substitution, from GAA to AAA, could result in this sequence encoding a K instead. This would then produce the same sequence as x.

Sequences a and f determined during this study contain the motif NSVL, rather than TSVL found in other strains. The sequences obtained here were of good quality, and Asparagine (N) has the same physicochemical characteristics as Threonine (T), so it may be that substitution such as from ACC to AAC, or ACT to AAT, has resulted in substitution of the amino acids as well. As the two amino acids have the same character, the substitution may be functionally neutral and thus not selected against.

Sequence e is unusual in that it contains LGV where other sequences have PPQL. In this case, there is no obvious frameshift or single nucleotide change that could have resulted in CTT GGT GTA (LGV) instead of CCA CCT CAG CTG (PPQL). The amino acids LGV have the same character as PQL, so this variation could conceivably have arisen during a long separation of this isolate from others.

Sequence o is highly unusual. It seems likely that 3 nucleotide insertions in sequence AF283998 from isolate HI/92 have produced frameshifts and this unusual result.

These comparisons and examination of unusual motifs illustrate well that care must always be taken when using sequences from other laboratories or databases, as single nucleotide variations can have significant effects on the interpretation of results. In this study, the striking differences seen in HPR amino acid sequences did not require sequences to be confirmed, other than sequence m from the Norwegian Glesvaer isolate.
Recombination, reassortment, insertion and deletion are all found to generate diversity in influenza viruses (Luo et al., 1993; McCullers et al., 1999; Nerome et al., 1998). Unlike the haemagglutinin of influenza viruses, the ISAV haemagglutinin does not appear to undergo any posttranslational cleavage (Krossøy et al., 2001a; Rimstad et al., 2001; Skehel & Wiley, 2000). Recombination has been suggested as a possible mechanism for generating the diversity of ISAV HPR types (Devold et al., 2001). This mechanism is more common in RNA viruses than was once thought, and can be extremely important in generating novel influenza subtypes (Gibbs et al., 2001; Worobey & Holmes, 1999). Recombination requires the presence of more than one form of the virus in the same cell. The discovery of two distinct types of HPR in the same farm in the Faroes during this study, and two or three different types in samples from Norway (Mjaaland et al., in press) suggests that there is potential for this to occur. This coinfection could also facilitate reassortment, the other common type of genetic exchange in RNA viruses. It is likely that the lack of farms with more than one ISAV isolate, noted by Devold et al. (2001), may be due to the use of screening methods that have not identified various HPR types. In the case of the Faroese sample used here, 7 pools of tissue from 5 fish gave the AJ440972 sequence (sequence g) and 1 pool gave the AJ440970 sequence (sequence f). Sequences that are amplified during early rounds of the polymerase chain reaction are likely to form the vast majority of the PCR product formed, although they may not necessarily have been present at a greater concentration in the original sample (Kidd & Ruano, 1995). It is possible that some variants of HPR might be amplified more efficiently, and thus other HPR types, or those present at low concentrations in samples, may not be detected as frequently. The reduced efficiency of amplification of the HPR compared to part of segment 8 may exaggerate the uneven detection of some HPR variants.

Kibenge et al. (2001) suggested that sequence e represented the archetypal ISAV, as it was the longest sequence then discovered. No mechanism for the derivation of amino acids PPQL NQT FNT from this proposed archetype was suggested, however. These amino acids, and all other motifs previously reported from ISAV, are present in the sequence reported here. It is tempting to speculate that the long form of HPR found in this study is more likely to be an archetype, and that truncation of the haemagglutinin molecule results in the other forms of the virus. All the shorter sequences shown in Table 1 could potentially have arisen independently from a progenitor via single deletion events. This occurs in feline infectious peritonitis virus, where truncation results in pathogenic forms whereas the longer original form is avirulent (Vennema et al., 1998).

Some of the unusual patterns or motifs found in the HPR can be explained if deletion is a mechanism used to create alternative variants. Sequence o is distinctive in that it contains the motifs PPR and NI whereas other sequences have PPQL and SNI motifs. Alignment of the nucleotide sequences reveals that this could have arisen through deletion occurring at unusual points in the sequence (Figure 2). As this sequence was obtained from a farm, this variant is likely to be pathogenic, like the majority of the sequences shown here. In vitro coinfection experiments may shed some light on the relative probability of recombination,
reassortment, insertion and deletion in ISAV. If deletions are found to be important for generating virus diversity, they are of a much greater magnitude than the few amino acid changes found in the influenza haemagglutinin.

However the variation in this HPR occurs, the use of the haemagglutinin sequence in epizootiology or phylogeney reconstruction requires careful choice of suitable methods. Conventional phylogenetic reconstruction methods are not ideal where recombination has occurred (Schierup & Hein, 2000). Phylogenies demonstrating significant distance between two sequences such as $l$ and $y$ could be generated, when in fact these variants may have arisen through only 2 separate splicing or deletion events from a common ancestor such as the long sequence found in this study. Similarly, comparison of sequences might give a false impression of apparently wide divergence. Analysis of the sequences using split decomposition (Huson, 1998) does indeed show radiation of all types from a single point, rather than a network connecting different HPR groups. However, these patterns are not well supported by high bootstrap values and therefore have not been shown here.

Previous studies have identified 2 or 3 major groupings of HPR types (Devold et al., 2001; Kibenge et al., 2001). The virus isolates used by Kibenge et al. (2001) include several duplicate strains that have a common origin. The Scottish isolates 1, 2, 3, 4 and 6 all stemmed from the same point source and are identical (Stagg et al., 2001), and some of the Canadian isolates used are also from sites that have direct epizootiological links and therefore the same strain of ISAV. Thus the two groups found by Kibenge et al. are not surprising and do not reflect the true diversity of HPR types. Devold et al. (2001) presented a more rigorous analysis of HPR variants and a more realistic phylogeny. Deletions might produce certain types of HPR such as the blue and red groups of Devold et al. (2001), from which other variants arise via subsequent mutation, recombination or deletion. These possible explanations for links between different sequence types emphasise the importance of traditional epidemiological investigation alongside molecular analysis before conclusions can be drawn on potential links between outbreaks of ISA (Stagg et al., 2001). Reliance on this HPR alone might be particularly undesirable, as we have found isolates 1490/98 from Scotland and Glesvaer 2/90 from Norway to contain the same HPR sequence, but to vary at several positions in segment 2. This is consistent with either an independent origin for these isolates, their separation from a common ancestor for sufficient time to allow this genetic drift, or reassortment of ISAV segments from different strains.

Influenza A strains have been found to develop from a central trunk lineage, and those
strains that show greatest change in the HA1 domain of the haemagglutinin are most likely to give rise to future epidemics (Bush et al., 1999). The development and testing of this system of predictive evolution required analysis of several hundred isolates. Much more sampling, particularly of asymptomatic and wild fish, must be carried out before it can be shown if a similar strategy operates for ISAV.

The fish from which the long ISAV HPR sequence was obtained was an adult, returning to the river to spawn, with no obvious clinical signs of ISA disease. Evidence of ISAV infection without clinical signs has previously been found in farms and wild fish (Raynard, 2000; Raynard et al., 2001; 2002; Devold et al., 2000; Nylund et al., 1999). In our experience, it has often not been possible to amplify the HPR from wild fish, possibly due to the lower abundance of haemagglutinin gene transcripts compared to those from segment 8 or to the use of kidney tissue samples. In this case, ISAV may have been present at a relatively high concentration in the gill tissue.

The discovery of the ISAV strain reported here reinforces evidence for a pool of ISAV in the aquatic environment (Raynard et al., 2001; 2002), some strains of which may be less virulent than others. The vast majority of ISAV HPR types have been obtained from outbreaks of ISA disease. Sequence a from Scotland, and sequence b from Canada were obtained from apparently healthy fish. Sequence b has also been obtained from outbreaks of ISA, so it is clear that this strain can be pathogenic under certain conditions. It is not known if the novel variant found in this study can be pathogenic. As the fish was sampled from a commercial netting station, it was not possible to obtain more tissue or samples that could be used to inoculate cell cultures and isolate ISAV. Although the virus from which sequence a was obtained has not been isolated, it may be examined for pathogenicity once systems for site-directed mutagenesis and reverse genetics have been developed for ISAV.

Given the high degree of variation found in this HPR sequence within Norway (Devold et al., 2001), the fact that Chilean sequence c is identical to a Canadian sequence, b, is completely unexpected. It is possible that Atlantic salmon that mingle in feeding grounds in the North Atlantic might contribute to the spread of ISAV strains across the Atlantic. However, the probability of natural spread of one variant from the northern to the southern Hemisphere and to a different host species, without mutation or genetic drift in a highly polymorphic region of a molecule that is under considerable selection pressure from the host immune system, must be extremely remote.

Recombination and reassortment of genetic material also presents problems for applying the concept of a molecular clock. Large rearrangements of sequence will disrupt the gradual genetic drift on which the concept relies and gap stripping is usually employed to remove such hypervariable regions from analysis (Sala & Wain-Hobson, 2000). Krossøy et al. (2001b) estimated the divergence of North American and European strains of ISAV to have occurred approximately 100 years ago. These estimates were based on segments 2 and 8 of ISAV and no major recombination or reassortment has yet been found in these segments. However, sequences from the HPR of the haemagglutinin gene indicate that it is not suitable for this type of analysis.
At present, genetic evidence reveals that many different variants of the HPR exist in pathogenic forms of the virus. If this domain is important in antigenicity, the variability may pose problems for future vaccine development. The long sequence described in this study could provide a useful starting point for the analysis of structure, pathogenicity and antigenicity of ISAV.

References


