NUCLEOTIDE SEQUENCE COMPARISON OF THE ENTIRE CODING REGIONS OF VIRULENT AND AVIRULENT VIRAL HAEMORRHAGIC SEPTICAEMIA VIRUS.

by Alan M Betts and David M Stone.

Introduction

Viral haemorrhagic septicaemia virus (VHSV) is a fish rhabdovirus of the genus *Novirhabdovirus* and the causative agent of viral haemorrhagic septicaemia, a disease causing considerable losses in farmed trout throughout Europe. The virus usually causes skin haemorrhages and haemorrhaging of the kidney and liver, with mortality rates as high as 90 % (**Figure 1**). To date the virus has been isolated from freshwater salmonid species and from non-salmonid species in both the marine and freshwater environments. This has led to the theory that marine fish may act as a reservoir for VHSV and it is possible that the use of untreated 'trash' marine fish as a food source for trout farms in Europe may have led to the introduction of the virus.

Figure 1: Rainbow trout infected with VHSV (Courtesy of Dr. G. Bovo)

VHSV has a non-segmented negative sense RNA genome of approximately 11,500 nucleotides coding for five structural proteins: the nucleocapsid (N), the RNA polymerase associated protein (M1 or P), the matrix protein (M2), the transmembrane glycoprotein (G) and an RNA polymerase (L). In addition, a non-structural protein (termed the Nv protein) of unknown function is also encoded and each gene is separated by a non-coding region of approximately 100bp that contains characteristic transcription initiation and polyadenylation signals (**Figure 2**). The complete nucleotide sequence of the N, M1, M2, G and Nv genes has previously been determined and over the course of this study the L gene sequence has also been completed. However, to date sequence data has been derived from several strains of VHSV and prior to this study the sequence of the entire coding region for a single strain had not been carried out.

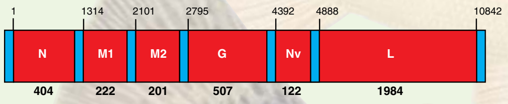

Figure 2: Schematic representation of the VHSV genome. Numbers above the bar represent the number of nucleotides from the N gene initiation codon and numbers below the bar represent the size of each protein in amino acid residues. Coding regions are coloured red and non-coding regions blue.

Materials and Methods

The VHSV isolates used in this study are described in **Table 1**. The Hededam isolate is virulent for rainbow trout under experimental conditions (Jorgensen 1980) and the 14-58 isolate was characterised from a virulent VHSV outbreak in rainbow trout (J. Castric Pers. Comm.). Both the cod ulcus and 96-43 strains have been demonstrated to be avirulent for rainbow trout experimentally (Jorgensen 1992 and Dixon *et al*., 1997).

Viruses were grown in BF-2 cells and genomic RNA extracted as described previously (Stone *et al*., 1997). For the viral genes N, M1, M2, G and Nv, primers for first strand cDNA synthesis and subsequent PCR were designed from published sequences (Genbank accession number x73873) to give an amplified product size of approximately 1 kb. Each product was designed to overlap the proceeding one by at least 100 bp to allow for accurate overlap analysis of sequences generated. In the case of the L gene, primers were designed from preliminary sequence information (Genbank accession number AJ009814) and the remaining sequences were generated using identical approaches to those used for other structural genes. First strand cDNA synthesis and PCR were carried out in duplicate using standard protocols (Sambrook *et al*., 1989). Products were cloned into the pGEM-T vector according to the manufacturer's protocol (Promega) and several clones were sequenced in both directions using an Applied Biosystems 310 automated sequencer. Sequence data were analysed using the DNA Navigator software (Applied Biosystems).

Virus Isolate	Year of isolation	Location	Species	Code
Virulent strains				
14-58	1990	France	Rainbow trout	14-58
Hededam	1972	Denmark	Rainbow trout	Hed
Avirulent strains				
96-43	1996	England	Atlantic herring	96-43
Cod Ulcus	1979	Denmark	Atlantic cod	Cod

Table 1. Isolates of VHSV used in this study

Results and Discussion

Nucleotide sequence of the VHSV RNA polymerase gene

The VHSV L gene comprises a coding region between nucleotides 4888 and 10842, encoding a protein of 1984 amino acids, with an approximate molecular mass of 225kDa. At the termini of the open reading frame, located 43 nucleotides from the stop codon is the characteristic rhabdovirus polyadenylation signal, AGAT AG$(A)_7$C/T GGC. A comparison between the L genes of VHSV and the closely related rhabdovirus IHNV shows a homology of 60% at the amino acid level, suggesting a high degree of evolutionary constraint. In addition, the amino acid sequence contains the highly conserved motifs A, B, C and D that are located in the proposed catalytic domain of the L protein (**Figure 3**). These motifs are characteristic of the unsegmented negative strand RNA viruses (Poch *et al*., 1990 and Schutz *et al*.,1995). In addition to these motifs, another highly conserved motif consisting of basic amino acids can be found commencing at amino acid 480 consisting of the sequence KEXEXXXXGRXF. This motif has been proposed to play a role in ribonucleotide binding (Poch *et al*., 1990).

```
Virus       Motif A               Motif B           Motif C       Motif D

VHSV     yihisks1DinKfctsqR  afsglmGGiEGlcQyvWticlllrv  vmaqGDNviin  hcpqhltlavKka
IHNV     fihvnkslDinKfctsqR  vfsglkGGiEGlcQyvWticlllrv  ilaqGDNviit  hcpqhltlaiKka
HRSV     iskcsiitDlsKfnqafR  lyryhmGGiEGwcQklWtieaisll  alinGDNqsid  hnGvyypasiKkv
Marbrg   vrgasfvtDleKynlafR  ayhyllGGiEGlqQklWtciscaqi  ssvmGDNqcit  lnGvqlpqslKtm
Rabies   rvtyafhlDyeKwnnhqR  cwngqdGGlEGlrQkgWslvsllmi  vlaqGDNqvlc  frGnilvpesKrw
Sendai   tlscflttDleKyclnwR  fihnprGGiEGycQklWtlisisia  amvqGDNqaia  ydGkilpqclKal
SV5      laasflttDlkKyclqwR  fivsprGGiEGlcQkaWtmisiavi  smvqGDNqaia  yqGriltqalKna
VSV      aicianhiDyeKwnnhqR  cwqgqeGGlEGlrQkgWsilnllvi  vlaqGDNqvic  frGvirgletKrw
```

Figure 3: Conserved motifs within the catalytic domain of the RNA polymerase of unsegmented negative strand viruses, located between amino acid residues 550-850. Conserved amino acid residues are shown in bold type with grey shading. Data are illustrated for human respiratory syncytial virus (HRSV), infectious haematopoietic necrosis virus (IHNV), marburg virus (Marbrg), rabies virus (Rabies), sendai virus (Sendai), simian virus 5 (SV5), vesicular stomatitis virus (VSV) and viral haemorrhagic septicaemia virus (VHSV). Adapted from Schutze et al., (1995).

Comparison of virulent and avirulent strains of VHSV at the nucleotide and amino acid level

The entire coding region of VHSV was found to span 10842 nucleotides, including intergenic non-coding regions. A comparison of virulent and avirulent strains of VHSV at the nucleotide level indicated that the virulent freshwater strains under study (14-56 and Hededam) were 2.3% divergent and the avirulent marine strains (96-43 and Cod ulcus) were 0.7% divergent. All four strains under study had a combined divergence of 3.6%, clearly indicating that they all belong to the same genogroup of the novirhabdoviruses. However, nucleotide substitutions between virulent and avirulent strains were generally scattered throughout the genome and a comparison at the amino acid level was required to deduce structural changes that may affect virulence.

A summary of the comparison of virulent and avirulent strains at the amino acid level is given in **Table 2**. Many of the mutations seen at the nucleotide level were silent and the overall divergence between the four strains at the amino acid level was only 1.9%. Between strains the L protein is the most conserved of all the VHSV proteins and the Nv protein the least. This could be explained by the high evolutionary pressure on viruses to maintain an efficient active RNA polymerase, compared to the low constraint on the Nv protein, which may not be required either structurally or functionally. Indeed, the fish rhabdovirus spring viremia of carp virus (SVCV) does not contain an Nv gene at the G-L gene junction and yet is a virulent fish pathogen (Kurath *et al*., 1997).

Both N and G proteins had a characteristic pattern of amino acid substitutions between strains. The N protein contained substitutions in the first and last third of the protein, avoiding the conserved RNA binding domain of the middle region (Said *et al*., 1998). When replacements in the G gene were compared to a recently published model for novirhabdovirus G protein structure, a cluster of 6 replacements (from a total of 17) fell within the major antigenic site II (Walker and Kongsuwan, 1999). Mutations in this region have also been identified using monoclonal antibody escape mutants and thus may represent a site involved in virulence (Benmansour *et al*., 1997 and Gaudin *et al*., 1999).

When the coding regions of the strains under study were compared to the pathogenic reference strain 07-71 a number of conserved amino acid changes were identified (**Figure 4**). Interestingly, these were located mainly within the N and M1 coding regions and may represent functional changes, which effect the efficiency of viral replication or assembly. For instance, substitutions in the N protein may affect the efficiency of the N protein to bind viral RNA and those in the M1/P protein may alter the efficiency of the viral RNA polymerase. Thus substitutions in either or both of these regions could affect virulence. We are currently undertaking a project to construct an infectious clone of VHSV in order to determine whether these mutations play a direct role in the virulence of VHSV.

Compared viral strains	Number of amino acid substitutions per viral protein. Numbers in brackets indicate substitutions as a percentage of the total number of amino acids per protein.						
	N	M1	M2	G	Nv	L	Genome
96-43 vs 14-58	11 (2.7)	5 (2.3)	2 (1.0)	10 (2.0)	8 (6.6)	13 (0.7)	49 (1.4)
96-43 vs Hed	11 (2.7)	3 (1.4)	1 (0.5)	13 (2.6)	3 (2.5)	10 (0.5)	41 (1.1)
Cod vs 14-58	9 (2.2)	5 (2.3)	3 (1.5)	7 (1.4)	6 (4.9)	11 (0.6)	41 (1.1)
Cod vs Hed	9 (2.2)	3 (1.4)	2 (1.0)	9 (1.8)	3 (2.5)	8 (0.4)	43 (1.2)
96-43 vs Cod	2 (0.5)	0 (0.0)	1 (0.5)	4 (0.8)	2 (1.6)	4 (0.2)	13 (0.4)
1458 vs Hed	7 (1.7)	2 (0.9)	1 (0.5)	12 (2.4)	9 (7.4)	12 (0.6)	34 (1.0)

Table 2: Comparison of the predicted amino acid sequence of individual structural proteins from virulent and avirulent strains of VHSV. The avirulent strains are labelled in blue and the virulent strains in red.

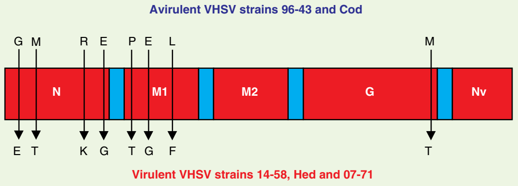

Figure 4: Schematic representation of the conserved amino acid substitutions between avirulent and virulent strains of VHSV. Conserved amino acid substitutions and their approximate location within each gene are represented, with avirulent residues (from the sequence of 96-43 and cod ulcus strains) located above the schematic genome and virulent residues (from reference strain 07-71, 14-58 and Hededam) located below.

Summary

We have completed the sequencing of the entire VHSV structural genome for both virulent and avirulent strains of VHSV. The marine and freshwater strains of VHSV are very closely related genetically, having less than 3% divergence at the nucleotide level and less than 1.5% divergence at the amino acid level. These figures indicate that under a high selective pressure it may possible for an avirulent strain originating in a marine species to be adapted into a virulent strain. Thus we are currently undertaking a project to determine the molecular basis for the virulence of VHSV by constructing an infectious clone.

Acknowledgement

This work was funded by MAFF: Contract no. F1118

References

Benmansour, A., Bascuro, B., Monnier, A. F., Vende, P., Winton, J. R. and de Kinkelin, P. (1997). *J. Gen. Virol.* 78, 2837-2846

Dixon, P. F., Feist, S., Kehoe, E., Parry, L., Stone, D. M. and Way, K. (1997). *Diseases of Aquatic Organisms* 30, 81-89

Gaudin, Y., de Kinkelin, P. and Benmansour, A. (1999). *J. Gen. Virol.* 80, 1221-1229

Jorgensen P. E.V. (1980). In: Ahne W. (Ed) Fish Diseases. Third COPRAQ-Session. Springer Verlag. Berlin, p3-7

Jorgensen P. E.V. (1992)P roceedings of the OJI International Symposium of Salmonid Diseases. Hokkaido University Press, Sapporo, Japan, p 60-71

Kurath, G., Higman, K. H. and Bjorklund, H.V. (1997). *J. Gen. Virol.* 78, 113-117

Poch, O., Blumberg, B. M., Bougueleret, L. and Tordo, N. (1990). *J. Gen. Virol.* 71, 1153-1162

Said, T., Bruley, H., Lamoureux, A. and Bremont, M. (1998). *J. Gen. Virol.* 79, 47-50

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbour, NY: Cold Spring Harbour Laboratory.

Schutz, H., Enjmann, P. J., Kuchling, R., Mundt, E., Niemann, H. and Mettenleiter, T. C. (1995). *J. Gen. Virol.* 76, 2519-2527

Stone, D. M., Way, K. and Dixon, P. F. (1997). *J. Gen. Virol.* 78, 1319-1326

Walker, P. and Kongsuwan, K. (1999). *J. Gen. Virol.* 80, 1211-1220

The Centre for Environment, Fisheries and Aquaculture Science (CEFAS), Weymouth Laboratory, Barrack Road, The Nothe, Weymouth, Dorset DT4 8UB, UK http://www.cefas.co.uk

CEFAS
The Centre for Environment, Fisheries & Aquaculture Science