J. Gen. Virol. 2013. **94**, 2515-2523. DOI:10.1099/vir.0.055673-0

A novel spliced gene in alcelaphine herpesvirus 1 encodes a glycoprotein which is secreted *in vitro*

George C. Russell[1#], Helen Todd[1], David Deane[1], Ann Percival[1], Mark P. Dagleish[1], David M. Haig[2] and James P. Stewart[3#]

[1] Moredun Research Institute, Pentlands Science Park, Penicuik, EH26 0PZ, UK
[2] School of Veterinary Medicine and Science, Nottingham University, Sutton Bonington, Leicestershire LE12 5RD, UK
[3] Department of Infection Biology, Institute of Infection and Global Health, University of Liverpool, Liverpool Science Park IC2, L3 5RF, UK

Corresponding authors:
George C. Russell, george.russell@moredun.ac.uk; Tel: +44 131 445 6149; Fax: +44 1313 445 6111
James P. Stewart , j.p.stewart@liv.ac.uk ; Tel: +44 151 795 0221

Footnote: Nucleotide sequences determined in this work have been submitted to the European Nucleotide Archive and have been assigned accession numbers as follows: cDNA clone 001-D10, HE855448; cDNA clone 01-B01, HE855449; cDNA clone 002-H06, HE855450; A9.5*01 allele, HE855451; A9.5*02 allele, HE855452; synthetic expression cassette MCS-HA, FN553440.

1

Summary

Herpesviruses often contain cryptic, spliced genes that are not obvious from the initial *in silico* annotation. Alcelaphine herpesvirus 1 (AlHV-1) contains 72 annotated open-reading frames but there are also a number of gaps between these that may have protein coding potential. Comparative analysis of coding potential between AlHV-1 and the related ovine herpesvirus 2 (OvHV-2) revealed a putative novel spliced gene that we have termed *A9.5*. Analysis of cDNA clones from AlHV-1-infected cells revealed three overlapping clones corresponding to *A9.5* and the coding sequence was confirmed by RT-PCR of RNA from AlHV-1 infected cattle tissues. The *A9.5* gene was predicted to encode a secreted glycoprotein with Mr 19 kDa. Empirical analysis showed that a recombinant haemagglutinin-tagged A9.5 fusion protein was secreted from transfected cells and had an Mr of 45 kDa, which was reduced to 20 kDa by endoglycosidase F treatment, confirming that A9.5 was a secreted glycoprotein. *In situ* RNA hybridisation showed that *A9.5* was expressed in cells associated with MCF lesions in infected cattle. Detailed analysis of the available OvHV-2 sequences revealed an homologous gene (*Ov9.5*) with conserved splicing signals and predicted amino acid sequence features in both sequenced isolates of this related virus. We have therefore identified a novel spliced gene in two related macaviruses that is expressed in malignant catarrhal fever lesions. Future work will determine its importance for the pathogenesis of disease.

INTRODUCTION

Alcelaphine herpesvirus 1 (AlHV-1) is a ruminant gammaherpesvirus of the genus *Macavirus* (subfamily *Gammaherpesvirinae*; family *Herpesviridae*; order *Herpesvirales*) that naturally infects wildebeest and which causes malignant catarrhal fever (MCF) in susceptible in-contact species such as cattle and deer. The natural infection of wildebeest with AlHV-1 is efficient and sub-clinical, in contrast to MCF which is generally sporadic and is almost always fatal in affected animals (reviewed by Russell *et al.*, 2009). Ovine herpesvirus 2 (OvHV-2) is a related virus that naturally infects sheep worldwide and also causes MCF in susceptible species like cattle, bison and deer. Analysis of PCR-amplified DNA sequences and serological determinants from ruminant gammaherpesviruses suggests that there are currently about 10 recognized MCF viruses (Li *et al.*, 2005), of which only AlHV-1 and OvHV-2 have been completely sequenced (Ensser *et al.*, 1997; Hart *et al.*, 2007; Taus *et al.*, 2007).

The genomes of AlHV-1 and OvHV-2 are co-linear with other gammaherpesviruses in their genetic organisation and contain about 72 genes each (Ensser *et al.*, 1997; Hart *et al.*, 2007; Taus *et al.*, 2007), many of which are homologous to genes in other herpesviruses. Bioinformatic analysis of AlHV-1 and OvHV-2 led to the annotation of about 10 unique genes that are characteristic of MCF viruses, numbered *A1* - *A10* in AlHV-1 and *Ov2* - *Ov10* in OvHV-2 (reviewed in Russell *et al.*, 2009). These unique genes are likely to have roles in the unique pathology of MCF and include viral immunomodulators and receptors for host immune molecules (Boudry *et al.*, 2007; Coulter *et al.*, 2001; Jayawardane *et al.*, 2008; Russell *et al.*, 2009). Of the ten unique genes originally described in AlHV-1, potential functions have been assigned to seven by bioinformatic and functional analysis. The remaining three genes (*A1*, *A4* and *A10*) have no significant similarity to any known protein, have not been detected as full-length transcripts and therefore remain putative genes. The

potential for discovery of further unique MCF virus genes was demonstrated by the annotation of *Ov4.5* in the OvHV-2 genome and the identification of its homologue *A4.5* in AlHV-1 (Hart *et al.*, 2007; Mills *et al.*, 2003). *Ov4.5* and *A4.5* share similarity to the *Bcl-2* gene family (Mills *et al.*, 2003).

The genomes of both AlHV-1 and OvHV-2 contain several regions within which no genes have currently been annotated, such as a 7.5 kbp gap between *ORF11* and *ORF17*, a 7.3 kbp gap between *ORF69* and *ORF73* and a 2.5 kbp gap between *A9* and *A10* (Ensser *et al.*, 1997). These gene-sparse areas are found adjacent to the terminal repeats and at the boundaries between conserved blocks of genes (Hart *et al.*, 2007). Bioinformatic analysis of these intergenic regions suggests that they contain no open reading frames of significant size and have no obvious sequence similarity to known genes. The recent description of microRNA transcripts in two of these regions in OvHV-2 (Levy *et al.*, 2012), however, suggests that the full complement of genes and transcripts in these viruses remains to be clearly defined, and that molecular analyses may be required to complement bioinformatic gene assignments. We therefore hypothesised that detailed analysis of intergenic gaps in MCF viruses could identify novel genes.

Here we describe the analysis of one intergenic region at the right hand end of the unique region of AlHV-1 and the consequent identification, expression and analysis of a novel spliced gene that encodes a secreted glycoprotein. The related virus OvHV-2 encodes an homologous gene and in both viruses this gene appears to be polymorphic.

RESULTS

Identification of the *A9.5* gene and sequence analysis.

Previous analysis of the genomic sequences of AlHV-1 and OvHV-2 had revealed 72 open-reading frames (Ensser *et al.*, 1997; Hart *et al.*, 2007; Taus *et al.*, 2007). We surmised that there might be cryptic, spliced genes in at least some of the gaps between these ORFs. The coding potential in the intergenic regions was initially examined by comparing the sequence of AlHV-1 with the two published genomic sequences of OvHV-2 using the TBLASTX algorithm (Altschul *et al.*, 1990). This analysis revealed three conserved blocks of coding sequence on the reverse strand between the *A9* and *A10* ORFs (Fig. 1). There were in-frame translation termination codons that meant these did not form part of a continuous ORF. However, analysis of potential splice/donor acceptor sites using Netgene 2 (www.cbs.dtu.dk/services/NetGene2; Brunak *et al.*, 1991; Hebsgaard *et al.*, 1996) suggested the presence of consensus splice donor/acceptor sequences between these blocks of sequence. From this we inferred that there could be a novel spliced gene in this region.

To confirm the presence of this gene empirically and to determine the precise coding sequence we constructed two cDNA libraries using mRNA extracted from AlHV-1-infected BT cells. Assembly of virus-specific cDNA sequences onto the AlHV-1 C500 genome sequence (Ensser *et al.*, 1997) identified three overlapping cDNA clones in the region of the AlHV-1 genome between the *A9* and *A10* genes (nt 126374-128455), which represented a novel spliced transcript of up to 6 exons (Fig. 1). The three clones had similar polyadenylation sites (nt 126374, 126376 and 126433; supplemental material, Fig. S1) and all contained the same three 3' exons but had different 5' ends. Analysis of the protein coding potential of the consensus cDNA sequence showed the longest open reading frame spanned exons 2-6, encoded 169 amino acids but had no start codon. Three of these exons corresponded almost exactly to the conserved blocks of sequence found in our TBLASTX

analysis of this region (Fig. 1). Examination of the genomic sequence and the observation of alternative splicing at the start of exons 3 and 4 (Fig. 1) suggested that a longer version of clone 002-H06 could encode a 170-residue protein on exons 3-6 (supplemental material, Fig. S1). This putative protein-coding gene was named *A9.5* to reflect its position in the genome. Bioinformatic analysis predicted that the A9.5 protein had an N-terminal signal sequence (D-score 0.759; SignalP 4.0; www.cbs.dtu.dk) and 10 predicted N-glycosylation sites (NetNGlyc 1.0; www.cbs.dtu.dk). The predicted mature polypeptide also contained 10 cysteine residues, some of which could be important for the tertiary structure of the protein. The A9.5 polypeptide sequence was checked for functional motifs using InterProScan Sequence Search (www.ebi.ac.uk/Tools/pfa/iprscan) but this analysis only confirmed the detection of an N-terminal signal peptide.

Cloning and expression of the *A9.5* coding region.

To determine whether the predicted *A9.5* coding region was expressed as RNA, RT-PCR was performed on RNA from AlHV-1-infected cattle tissues and on RNA from AlHV-1-infected BT cells. Infected tissues and cultured cells both contained *A9.5*-specific transcripts as defined by amplification of the *A9.5*ex4-6 RT-PCR product (Table 1), but the complete coding region (*A9.5*-CDS, Table 1) was only amplified from AlHV-1-infected cattle tissue (data not shown). A longer 5-exon RT-PCR product (*A9.5*ex2-6(a), *A9.5*ex2-6(b); Table 1), could not be amplified by either of two primer pairs from any of the RNA sources tested, despite being present in the cDNA library made from AlHV-1 infected BT cells (Fig. 1, clone 01-B01). This may indicate low abundance of this transcript or relative inefficiency of the PCR primers. Of note is the fact that all of the PCR products in Table 1 could be amplified from viral genomic DNA prepared from virulent AlHV-1 strain C500 but not from the high-passage attenuated AlHV-1 C500 virus. This is in accord with the published analysis of the

attenuated virus genome that suggests that deletions and rearrangements have occurred at the right end of the unique portion of the AlHV-1 genome (Wright *et al.*, 2003).

The A9.5-CDS RT-PCR product, amplified using RNA from MCF affected cattle tissue, was inserted into the mammalian expression vector pVR1255 (Norman *et al.*, 1997), to generate a plasmid that would express a recombinant A9.5 protein, fused at its C-terminus to a triple haemagglutinin (HA) tag. This construct, pVR-A9.5-HA, was transfected into HEK293T cells and samples were taken at 24 h and 48 h of culture. A9.5-HA protein was purified by affinity chromatography from culture medium and from extracts of the transfected cells. Proteins were resolved by SDS-PAGE and western blotting. The results (Fig. 2) revealed that the HA-tagged A9.5 fusion protein from the culture medium migrated with apparent Mr of approximately 45 kDa (Lane a) , while a band of around 40 kDa was purified from cellular extracts (Lane b).

Post-translational modification of A9.5.

The presence of multiple N-linked glycosylation sites in the A9.5 polypeptide was supported by the significant retardation of the protein on SDS-PAGE compared to the predicted Mr of 24 kDa for the full-length A9.5-HA fusion polypeptide and 21 kDa for mature A9.5-HA without the signal peptide. In order to confirm that A9.5 was glycosylated, affinity-purified A9.5-HA fusion protein was treated with the endoglycosidase Peptide: N-Glycosidase F (PNGase F) that cleaves between the innermost N-acetyl-glucosamine and asparagine residues of N-linked glycoproteins. Comparison of PNGase F-treated and untreated A9.5-HA samples by western blotting (Fig. 2) showed that the deglycosylated A9.5-HA migrated with an Mr of approximately 20 kDa (Lane d) compared with ca. 40kDa for the untreated A9.5-HA fusion protein (Lane c). The large change in SDS-PAGE migration caused by PNGase F treatment of A9.5-HA suggests that at least some of the predicted N-linked glycosylation sites in the A9.5 sequence were glycosylated.

Expression of A9.5 *in vivo.*

To identify cells that expressed *A9.5* mRNA in tissues from MCF-affected animals, we used RNA *in situ* hybridisation (RNA-ISH) on tissue sections from cattle infected intranasally with pathogenic AlHV-1, as described previously (Haig *et al.*, 2008). The *A9.5*ex4-6 PCR product (Table 1) was used as the source of sense and antisense RNA probes for RNA-ISH of kidney sections from MCF-affected and control cattle. This probe was selected as it contained two introns, making it more likely to detect only spliced A9.5 mRNA and not cross-react with virus genomic DNA. Kidney was used because lymphocyte infiltration in this tissue was only observed in animals with clinical MCF. Initial tests of both A9.5 probes with control, mock-infected cattle tissues showed no staining, while *A9.5* mRNA-positive cells could be identified in MCF-affected tissues only with the antisense probe (Fig. 3). The *A9.5*-expressing cells were mononuclear inflammatory cells, as denoted by morphology with haematoxylin and eosin (HE) stained semi-serial sections, and were comparable in frequency and morphology to T-cells infiltrating the kidney of AlHV-1 infected rabbits (Anderson *et al.*, 2007) and cattle (G.C. Russell and H. Todd, unpublished results). This shows that the *A9.5* gene is transcribed in MCF-affected tissue and demonstrates that a significant number of infected cells are present in MCF tissue lesions.

Homologues of *A9.5*.

BlastX and Psi-Blast analysis of the A9.5 amino-acid sequence revealed segments of similarity within the OvHV-2 genome. No other similar proteins were identified in the Genbank non-redundant protein database using these algorithms. The OvHV-2 homologue had a position analogous to that of *A9.5*, between the *Ov9* and *Ov10* genes (Fig. 1), and was therefore named *Ov9.5*. Analysis of the putative *Ov9.5* gene sequence demonstrated conserved splice donor and acceptor sites that could produce the same exon structure as *A9.5* and a continuous open reading frame within the putative *Ov9.5* mRNA with similarities of

size and sequence to *A9.5*. Noteably, analysis of the two OvHV-2 genome sequences (Hart *et al.*, 2007; Taus *et al.*, 2007) revealed that the Ov9.5 amino-acid sequences were less similar (52 % identity) than the products of all other known ORFs in OvHV-2 (84 – 100 % identity). To distinguish these *Ov9.5* variants we have termed them *Ov9.5*01* (Hart *et al.*, 2007) and *Ov9.5*02* (Taus *et al.*, 2007).

Having observed the polymorphism of *Ov9.5* we investigated whether *A9.5* might be polymorphic by PCR amplification of the *A9.5* gene from an AlHV-1-positive DNA sample from an Ankole cow that had died of MCF (Whitaker *et al.*, 2007). Sequencing of the 839 bp *A9.5*-gene PCR product (Table 1) showed that this AlHV-1 isolate contained an *A9.5* gene with the same coding potential and splice site positions as the C500 strain but the predicted *A9.5* protein encoded by this gene was only 80 % identical to the C500 polypeptide. We have therefore termed the C500 variant *A9.5*01* and the Ankole variant *A9.5*02*.

Alignment of the *A9.5*01* and *A9.5*02* predicted polypeptide sequences with those of *Ov9.5*01* (Hart *et al.*, 2007) and *Ov9.5*02* (Taus *et al.*, 2007) revealed 24 % amino acid identity between the OvHV-2 and AlHV-1 proteins. Despite this low level of similarity at the amino-acid level, all four proteins were predicted to have N-terminal signal sequences and shared six conserved N-linked glycosylation sites and six conserved cysteine residues (Fig. 4) suggesting conserved function and lending weight to the assignment of *A9.5/Ov9.5* as unique MCF virus genes.

Sequence-structure similarities between the A9.5/Ov9.5 sequences and other proteins of known structure were searched using the profile comparison tool HHpred (Soding, 2005; Soding *et al.*, 2005) by submission of the aligned full-length A9.5/Ov9.5 protein sequences to the HHpred server (http://toolkit.tuebingen.mpg.de/hhpred). Only two proteins among the identified hits were of a similar size to A9.5/Ov9.5. These were the cytokines IL-21 (p value 0.002) and IL-4 (p value 0.004), which share a compact 4-helix bundle structure stabilised by

3 disulphide bonds. Structure prediction using QUARK and I-Tasser (Xu & Zhang, 2012; Zhang, 2008) suggests that A9.5 also contains four helices with conserved cysteine residues at or near the ends of each helix (Fig. 4). This may imply structural similarities between the A9.5 and Ov9.5 proteins and four-helix bundle cytokines, such as IL-4 and IL-21.

DISCUSSION

In this work we have identified a novel spliced gene (*A9.5*) in AlHV-1 that appears to be conserved in at least one other MCF virus – OvHV-2. The *A9.5* complete coding sequence was found to be transcribed only in tissues of AlHV-1 infected cattle, and this transcript was predicted to encode a secreted glycoprotein that has no clear homologue outside the MCF viruses. Recent microarray analysis of virus transcription in AlHV-1 infected cells has confirmed the detection of a transcript between *A9* and *A10* that terminated close to the polyadenylation sites for A9.5, although no further analysis was done (Palmeira *et al*., 2013).

When expressed as an HA-tagged fusion, the A9.5 glycoprotein was found both in the culture medium and within the cells. These two forms of the A9.5-HA protein had distinct migration in SDS-PAGE. This may reflect incomplete glycosylation or partial processing of the intracellular protein due to over-expression. Treatment of the intracellular form with PNGase F led to a reduction in Mr that was in accord with the removal of N-linked carbohydrate side chains.

Alignment of the predicted primary translation products of *A9.5*01* and *Ov9.5*01* showed that 56 of 150 aligned residues (37 %) were identical and that these conserved residues included six cysteine residues and six predicted N-linked glycosylation sites (Fig. 4). The conservation of these sequence features between the AlHV-1 and OvHV-2-encoded A9.5 homologues, despite the low overall sequence identity, suggests that the function of this protein in the two viruses is likely to be similar. The detection of weak structure-sequence

similarity between the A9.5 homologues and four-helix-bundle cytokines may be notable but functional analysis of this novel gene product will be required to confirm or refute the bioinformatic predictions.

RNA-ISH demonstrated that *A9.5* was expressed in MCF affected tissues but it will be of interest to study the expression of this gene further to define its role in infection of the natural reservoir host and in the pathogenesis of MCF. Currently there is no *in situ* marker of virus-infected cells in MCF. However, the detection of numerous *A9.5*-expressing cells in lesions in the kidneys of cattle with MCF suggests that this gene may be a good marker of infection and may have a functional role in MCF pathogenesis. Of note is the fact that re-analysis of proteomic datasets from AlHV-1 virions (Dry *et al.*, 2008) showed no evidence of A9.5 protein in AlHV-1 virus particles propagated in BT cells, in accord with our inability to amplify the *A9.5* CDS from these cells. Thus, *A9.5* may only be properly expressed during *in vivo* infection rather than in culture. It is also of interest to note that this region of the virus genome appears to be affected during the attenuation process (Wright *et al.*, 2003) and that *A9.5* could not be amplified from the avirulent 'high passage' variant of AlHV-1. Thus, *A9.5* may encode an immunomodulatory molecule whose function is intimately involved in the pathogenesis of MCF. The high degree of polymorphism within this gene in both MCF viruses also suggests the *A9.5* and *Ov9.5* genes may be under immunological selection. This selection would only operate in the natural hosts where the virus is propagated so analysis of *A9.5/Ov9.5* expression and function in these hosts will be required to address this issue.

The discovery of distinct *A9.5* and *Ov9.5* sequences in a small number of isolates of MCF viruses demonstrates that this gene is highly polymorphic. The provenance of these isolates may, however, influence their sequences. For example, virus strains found in clinical MCF cases may reflect a specific subset of the virus strains circulating in the reservoir species; virus strains from different regions of the globe may be distinct due to their different

evolutionary histories and maintenance of MCF viruses by *in vivo* or *in vitro* passage may affect both sequence and structure. The *A9.5* and *Ov9.5* genes sequenced so far were from a lab strain of AlHV-1 that had been passaged *in vivo* in the UK for over 30 years (Plowright *et al.*, 1975); from a natural case of AlHV-1 MCF in an Ankole cow in the UK (Whitaker *et al.*, 2007); from an *in vitro* passaged OvHV-2 cell line derived from an MCF-affected cow in the UK (Hart *et al.*, 2007; Schock *et al.*, 1998); and from sheep nasal secretions obtained in the USA (Taus *et al.*, 2007). Thus, the sequence differences between these *A9.5* and *Ov9.5* alleles may be influenced by a number of factors. It should also be noted that for OvHV-2 at least, the DNA sequence of the regions immediately flanking the Ov9.5 putative coding region exhibit a far higher degree of sequence similarity (> 90 % identity) than the coding region (60 % identity), suggesting that sequence variation is a specific feature of the *A9.5/Ov9.5* coding region. Analysis of a larger number of virus isolates from a range of locations should be performed to determine the true degree of polymorphism and the range of strains in circulation. Analysis of sequence variation in other MCF virus genes has suggested that the AlHV-1 and OvHV-2 genomes are generally much less divergent. For example, the *Ov2.5* gene sequenced from seven UK isolates exhibited greater than 99 % identity (Jayawardane *et al.*, 2008), while an analysis of sequence divergence in three genomic regions of AlHV-1 from about 45 clinical cases of malignant catarrhal fever in South Africa showed a range of translated amino acid identity from about 90 % to 97 % (Sole, Bremer and Bastos; unpublished data; accession numbers EU847312 - EU847407).

Several publications have described a range of presentations of sheep-associated MCF, including cases with alimentary or neurological signs in addition to the more common head and eye form (Holliman *et al.*, 2007; Mitchell & Scholes, 2009; Twomey *et al.*, 2002). It will therefore be of interest to compare *Ov9.5* sequences from a range of MCF cases to

determine whether distinct strains of OvHV-2 are associated with different clinical presentations of this complex disease.

METHODS

Viruses and infected cattle. AlHV-1 C500 strain was used here exactly as described previously (Haig *et al.*, 2008). Briefly, virulent AlHV-1, derived from peripheral blood lymphoid cell suspensions of rabbits infected with AlHV-1 C500 and succumbing to MCF, was propagated in cultures of bovine turbinate (BT) cells. Infected BT cell cultures exhibiting clear cytopathic effects were passaged onto fresh BT cells by a 1:4 split no more than four times and were used as a source of infected cells and cell-free virus.

Disease-free and MCF seronegative male Friesian-Holstein calves of approximately 7 months of age were used to provide AlHV-1-infected tissue for this work. The animal experiments were carried out with the approval of the Moredun Research Institute's experiments and ethics committee and complied fully with the Home Office of Great Britain and Northern Ireland "Animals (Scientific Procedures) Act 1986". Following onset of clinical MCF caused by intranasal challenge with approximately 10^5 $TCID_{50}$ of pathogenic AlHV-1, animals were euthanized with an overdose of intravenous sodium pentobarbitone. Confirmation of MCF was by histopathology and PCR as described (Haig *et al.*, 2008). Tissue samples for subsequent expression analysis were cut into 5mm cubes using sterile disposable scalpels and were stored in RNAlater (Ambion) at -20°C.

cDNA library construction. Total RNA from control BT cells and from BT cells infected with virulent cell-free AlHV-1 C500 was used for construction of full-length and normalised infected cell cDNA libraries. The cells were harvested three days after infection and RNA was extracted using Trizol reagent (Life Technologies) and purified using the RNeasy mini kit (Qiagen). RNA yield and quality were assayed by spectrophotometry and Agilent Bioanalyser. Only RNA with RNA integrity number > 7.5 was used. Three independent high quality infected cell RNA preparations were pooled to ensure good representation of virus transcripts in the cDNA libraries.

RNA preparations (300 µg) derived from infected and uninfected cells were used for library production (GATC Biotech,Konstanz, Germany). Two libraries were constructed in the vector pBluescript II SK(+): a subtracted (normalised) library, enriched for transcripts specific to infected cells, with average insert size of 450 bp; and a full length infected cell library, with average insert size of 1200 bp. Quality control sequencing of 192 clones randomly selected from the subtracted cDNA library showed 78 different sequences of which 68 were encoded by AlHV-1; and 48 clones sequenced from the full-length library represented 26 sequences of which 11 were from AlHV-1.

DNA sequence analysis. Approximately 500 clones from each library were subjected to sequencing of one DNA strand and the resulting data were analysed using the DNASTAR Lasergene software suite (DNASTAR, Inc. Madison, WI, USA). The finished clone sequences were aligned with the AlHV-1 genomic DNA sequence using Splign (www.ncbi.nlm.nih.gov/sutils/splign/splign.cgi; (Kapustin *et al.*, 2008) to define the start and end points of exons. Similarity searches were done using BLAST tools (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi; (Altschul *et al.*, 1990; Altschul *et al.*, 1997)); protein topology was predicted using HMMTOP (www.enzim.hu/hmmtop/server/hmmtop.cgi; (Tusnády & Simon, 2001); and N-linked glycosylation sites were predicted using NetNGlyc 1.0 Server (www.cbs.dtu.dk/services/NetNGlyc/). Multiple sequence alignments were prepared using the ClustalW algorithm within Lasergene; the MUSCLE program at the University of Wageningen Bioinformatics laboratory (www.bioinformatics.nl/tools/muscle.html); and the MAFFT server at the European Bioinformatics Institute (www.ebi.ac.uk/Tools/msa/mafft). Structure prediction for the mature A9.5 polypeptide was performed using QUARK and I-Tasser (zhanglab.ccmb.med.umich.edu/; Xu & Zhang, 2012; Zhang, 2008).

Cloning and transfection of *A9.5* cDNA into mammalian cells. RNA from infected cells or tissues was prepared as described above and 5 µg samples from MCF-affected kidney and lymph node were used to make oligo-dT-primed first-strand cDNA (Superscript III, Life Technologies). Aliquots of cDNA equivalent to 0.5 µg total RNA were used in subsequent polymerase chain reactions (PCR) with primers described in Table 1 and KOD hot-start DNA polymerase. The 513 bp spliced RT-PCR product encoding the *A9.5* CDS was inserted into pCR4-blunt-TOPO (Life Technologies) and sequenced using flanking primers. A verified A9.5 fragment was excised using the restriction enzyme *Nco*I and was ligated into the mammalian expression vector pVR-MCS-HA, derived from pVR1255 (a kind gift from Vical Inc.; Norman *et al.*, 1997). To generate pVR-MCS-HA, a synthetic 218 bp double-stranded cassette encoding a ribosome binding site with start codon, a multiple cloning site, a TEV protease cleavage site and an in-frame triple-HA epitope tag (TOP Gene Technologies, Montreal Canada) was inserted between the NotI and BamHI sites of pVR1255. Analysis confirmed the correct sequence and orientation of one clone (pVR-A9.5-HA) which could express *A9.5* polypeptide with a triple-HA C-terminal tag. DNA for transient transfection was prepared from 500 ml LB-kanamycin (50 µg/ml) cultures using the Qiagen endofree-maxi kit.

HEK293T cells were grown in IMDM (Sigma), containing 5 % foetal calf serum and 2 mM L-glutamate, to 70 % confluence before being transfected with plasmid DNA (1µg per 10^6 cells) using lipofectamine 2000 (Life Technologies) according to the manufacturer's protocol. The transfection reagent was replaced with fresh medium after four hours and transfected cells were allowed to grow for up to 72 h before harvesting cells and medium.

RNA *in situ* hybridisation. The primers *A9.5*ex4F and *A9.5*ex6R (Table 1) were used to amplify the 366 bp *A9.5*ex4-6 fragment from cDNA equivalent to 0.5 µg total RNA from AlHV-1 infected bovine tissue. This PCR product was inserted into the vector pGEM-T-

Easy (Promega) and one clone with the predicted sequence was selected for production of RNA probes. Strand-specific DIG-UTP-labelled RNA probes were generated using the DIG RNA labelling kit SP6/T7 (Roche Applied Science). RNA *in situ* hybridisation (RNA-ISH) of DIG-labelled RNA probes was performed on 5 µm sections of 4 % paraformaldehyde-fixed kidney, using an Omnislide thermal cycler and wash module (Hybaid) as described (Anderson *et al.*, 2001). Proteinase K was used at an optimal concentration of 5 µg/ml and the *A9.5* probes were at 2 ng/µl. Hybridised, DIG-labelled sections were visualised using an anti-DIG Fab-alkaline phosphatase antibody conjugate (Roche Applied Science). Sections were counterstained with haematoxylin to improve tissue contrast and aid cell identification. Semi-serial sections were stained with haematoxylin and eosin (HE) to enable more accurate morphological assessment of any *A9.5*-positive cells.

Nucleotide sequences. All of the novel sequences described in this article have been submitted to the European Nucleotide Archive and have been assigned accession numbers as follows: cDNA clone 001-D10, HE855448; cDNA clone 01-B01, HE855449; cDNA clone 002-H06, HE855450; the RT-PCR product *A9.5*-CDS, encoding the *A9.5*01* allele, HE855451; the PCR product *A9.5*-gene, encoding the *A9.5*02* allele, HE855452; and the synthetic expression cassette MCS-HA, FN553440.

ACKNOWLEDGEMENTS

This work was supported by grants from the Scottish Government's Rural and Environment Science and Analytical Services Division and from the Biotechnology and Biological Sciences Research Council (BBSRC UK) (26/S16844 and BBS/B/00468).

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology* **215**, 403-410.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389-3402.

Anderson, I. E., Buxton, D., Campbell, I., Russell, G., Davis, W. C., Hamilton, M. J. & Haig, D. M. (2007). Immunohistochemical study of experimental malignant catarrhal fever in rabbits. *Journal of Comparative Pathology* **136**, 156-166.

Anderson, I. E., Reid, H. W., Nettleton, P. F., McInnes, C. J. & Haig, D. M. (2001). Detection of cellular cytokine mRNA expression during orf virus infection in sheep: differential interferon-[gamma] mRNA expression by cells in primary versus reinfection skin lesions. *Veterinary Immunology and Immunopathology* **83**, 161-176.

Boudry, C., Markine-Goriaynoff, N., Delforge, C., Springael, J. Y., de, L. L., Drion, P., Russell, G., Haig, D. M., Vanderplasschen, A. F. & Dewals, B. (2007). The A5 gene of alcelaphine herpesvirus 1 encodes a constitutively active G-protein-coupled receptor that is non-essential for the induction of malignant catarrhal fever in rabbits. *Journal of General Virology* **88**, 3224-3233.

Brunak, S., Engelbrecht, J. & Knudsen, S. (1991). Prediction of Human Messenger-Rna Donor and Acceptor Sites from the Dna-Sequence. *Journal of Molecular Biology* **220**, 49-65.

Coulter, L. J., Wright, H. & Reid, H. W. (2001). Molecular genomic characterization of the viruses of malignant catarrhal fever. *Journal of Comparative Pathology* **124**, 2-19.

Dry, I., Haig, D. M., Inglis, N. F., Imrie, L., Stewart, J. P. & Russell, G. C. (2008). Proteomic analysis of pathogenic and attenuated alcelaphine herpesvirus 1. *Journal of Virology* **82**, 5390-5397.

Ensser, A., Pflanz, R. & Fleckenstein, B. (1997). Primary structure of the alcelaphine herpesvirus 1 genome. *Journal of Virology* **71**, 6517-6525.

Haig, D. M., Grant, D., Deane, D., Campbell, I., Thomson, J., Jepson, C., Buxton, D. & Russell, G. C. (2008). An immunisation strategy for the protection of cattle against alcelaphine herpesvirus-1-induced malignant catarrhal fever. *Vaccine* **26**, 4461-4468.

Hart, J., Ackermann, M., Jayawardane, G., Russe, G., Haig, D. M., Reid, H. & Stewart, J. P. (2007). Complete sequence and analysis of the ovine herpesvirus 2 genome. *Journal of General Virology* **88**, 28-39.

Hebsgaard, S. M., Korning, P. G., Tolstrup, N., Engelbrecht, J., Rouze, P. & Brunak, S. (1996). Splice site prediction in Arabidopsis thaliana pre-mRNA by combining local and global sequence information. *Nucleic Acids Research* **24**, 3439-3452.

Holliman, A., Daniel, R., Twomey, D. F., Barnett, J., Scholes, S., Willoughby, K. & Russell, G. (2007). Malignant catarrhal fever in cattle in the UK. *Veterinary Record* **161**, 494-495.

Jayawardane, G., Russell, G. C., Thomson, J., Deane, D., Cox, H., Gatherer, D., Ackermann, M., Haig, D. M. & Stewart, J. P. (2008).A captured viral interleukin 10 gene with cellular exon structure. *Journal of General Virology* **89**, 2447-2455.

Kapustin, Y., Souvorov, A., Tatusova, T. & Lipman, D. (2008). Splign: algorithms for computing spliced alignments with identification of paralogs. *Biology Direct* **3**, 20.

Levy, C. S., Hopkins, J., Russell, G. C. & Dalziel, R. G. (2012). Novel virus-encoded microRNA molecules expressed by ovine herpesvirus 2-immortalized bovine T-cells. *Journal of General Virology* **93**, 150-154.

Li, H., Gailbreath, K., Flach, E. J., Taus, N. S., Cooley, J., Keller, J., Russell, G. C., Knowles, D. P., Haig, D. M., Oaks, J. L., Traul, D. L. & Crawford, T. B. (2005).A novel subgroup of rhadinoviruses in ruminants. *Journal of General Virology* **86**, 3021-3026.

Mills, R., Rozanov, M., Lomsadze, A., Tatusova, T. & Borodovsky, M. (2003). Improving gene annotation of complete viral genomes. *Nucleic Acids Research* **31**, 7041-7055.

Mitchell, E. & Scholes, S. (2009). Unusual presentation of malignant catarrhal fever involving neurological disease in young calves. *Veterinary Record* **164**, 240-242.

Norman, J. A., Hobart, P., Manthorpe, M., Felgner, P. & Wheeler, C. (1997). Development of improved vectors for DNA-based immunization and other gene therapy applications. *Vaccine* **15**, 801-803.

Palmeira, L., Sorel, O., Van Campe, W., Boudry, C., Roels, S., Myster, F., Reschner, A., Coulie, P. G., Kerkhofs, P., Vanderplasschen, A. & Dewals, B. G. (2013). An essential role for γ-herpesvirus latency-associated nuclear antigen homolog in an acute lymphoproliferative disease of cattle. *Proc. Natl. Acad. Sci. USA*. (in press) doi:10.1073/pnas.1216531110.

Plowright, W., Herniman, K. A., Jessett, D. M., Kalunda, M. & Rampton, C. S. (1975). Immunisation of cattle against the herpesvirus of malignant catarrhal fever: failure of inactivated culture vaccines with adjuvant. *Research in Veterinary Science* **19**, 159-166.

Russell, G. C., Stewart, J. P. & Haig, D. M. (2009). Malignant catarrhal fever: A review. *Veterinary Journal* **179**, 324-335.

Schock, A., Collins, R. A. & Reid, H. W. (1998). Phenotype, growth regulation and cytokine transcription in Ovine Herpesvirus-2 (OHV-2)-infected bovine T-cell lines. *Veterinary Immunology and Immunopathology* **66**, 67-81.

Soding, J. (2005). Protein homology detection by HMM–HMM comparison. *Bioinformatics* **21**, 951-960.

Soding, J., Biegert, A. & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research* **33**, W244-W248.

Taus, N. S., Herndon, D. R., Traul, D. L., Stewart, J. P., Ackermann, M., Li, H., Knowles, D. P., Lewis, G. S. & Brayton, K. A. (2007). Comparison of ovine herpesvirus 2 genomes isolated from domestic sheep (Ovis aries) and a clinically affected cow (Bos bovis). *Journal of General Virology* **88**, 40-45.

Tusnídy, G. E. & Simon, I. (2001). The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**, 849-850.

Twomey, D. F., Holt, G. J. & Reid, H. W. (2002). Malignant catarrhal fever in cattle with suspected bracken poisoning. *Veterinary Record* **151**, 486-487.

Whitaker, K. A., Wessels, M. E., Campbell, I. & Russell, G. C. (2007). Outbreak of wildebeest-associated malignant catarrhal fever in Ankole cattle. *Veterinary Record* **161**, 692-695.

Wright, H., Stewart, J. P., Ireri, R. G., Campbell, I., Pow, I., Reid, H. W. & Haig, D. M. (2003). Genome re-arrangements associated with loss of pathogenicity of the [gamma]-herpesvirus alcelaphine herpesvirus-1. *Research in Veterinary Science* **75**, 163-168.

Xu, D. & Zhang Y. (2012) Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins*, **80**, 1715-1735

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, **9**, 40.

Table 1. PCR primers and predicted product sizes. Restriction enzyme recognition sites added to primers to facilitate cloning are shown in bold type.

PCR product	Primer name	Primer sequence	Product size Genomic (bp)	Product size cDNA (bp)
*A9.5*ex2-6(a)	*A9.5*ex2F	AGCCTGCAGAGGTGTGTCTT	1169	505
	*A9.5*cdsR	**CCATGG**GCAGAACGCCTGACACC		
*A9.5*ex2-6(b)	*A9.5*ex2F	AGCCTGCAGAGGTGTGTCTT	1208	545
	*A9.5*ex6R	GCACAAACCGAGTTTGACAG		
A9.5-CDS	*A9.5*cdsF	**CCATGG**AGGGCAGGCGAGTA	761	513
	*A9.5*cdsR	**CCATGG**GGCAGAACGCCTGACACC		
*A9.5*ex4-6	*A9.5*ex4F	TGGTTGTGAGAATGAAACACTTG	529	366
	*A9.5*ex6R	GCACAAACCGAGTTTGACAG		
A9.5-gene	*A9.5*geneF	CGGGTAAGCTGCAGATTTTT	839	N/A
	*A9.5*ex6R	GCACAAACCGAGTTTGACAG		

FIGURE LEGENDS

Fig. 1. Schematic diagram of the right end of the unique portion of the AlHV-1 genome (thick line), showing the ends of the *A9* and *A10* ORFs (filled boxes) and an annotated repetitive sequence element (striped box) (Ensser *et al.*, 1997). The nt positions of the ends of the *A9* and *A10* ORFs are shown above the genome. The position of three regions found to be conserved between AlHV-1 and OvHV-2 on the reverse strand by TBLASTX analysis are shown as grey blocks with broken outlines and the relevant segments of amino acid similarity between AlHV-1 and OvHV-2 are aligned above the genome. Between the aligned sequences, identical residues are shown while similar residues are represented by plus (+) signs.

Beneath the depiction of the AlHV-1 genome, the genomic position and coding potential of *A9.5* cDNA clones are illustrated. The relative positions of cDNA clones 001-D01, 002-H06 and 001-B01 are given. Open boxes represent sequences present in the cDNA clones while intervening sequences, absent from the clones, are shown as thin lines. PolyA sequences at the end of each cDNA clone are shown as grey filled boxes. Sequences present in one cDNA clone but spliced out in larger clones are shown as grey outlined boxes. The position and extent of the *A9.5* open reading frame is shown in the bottom line as a set of grey boxes connected by a dashed line (labelled *Protein*). The N-terminal 18 residues of the predicted *A9.5* open reading frame were not encoded by any of these cDNA clones and are shown as a dashed box.

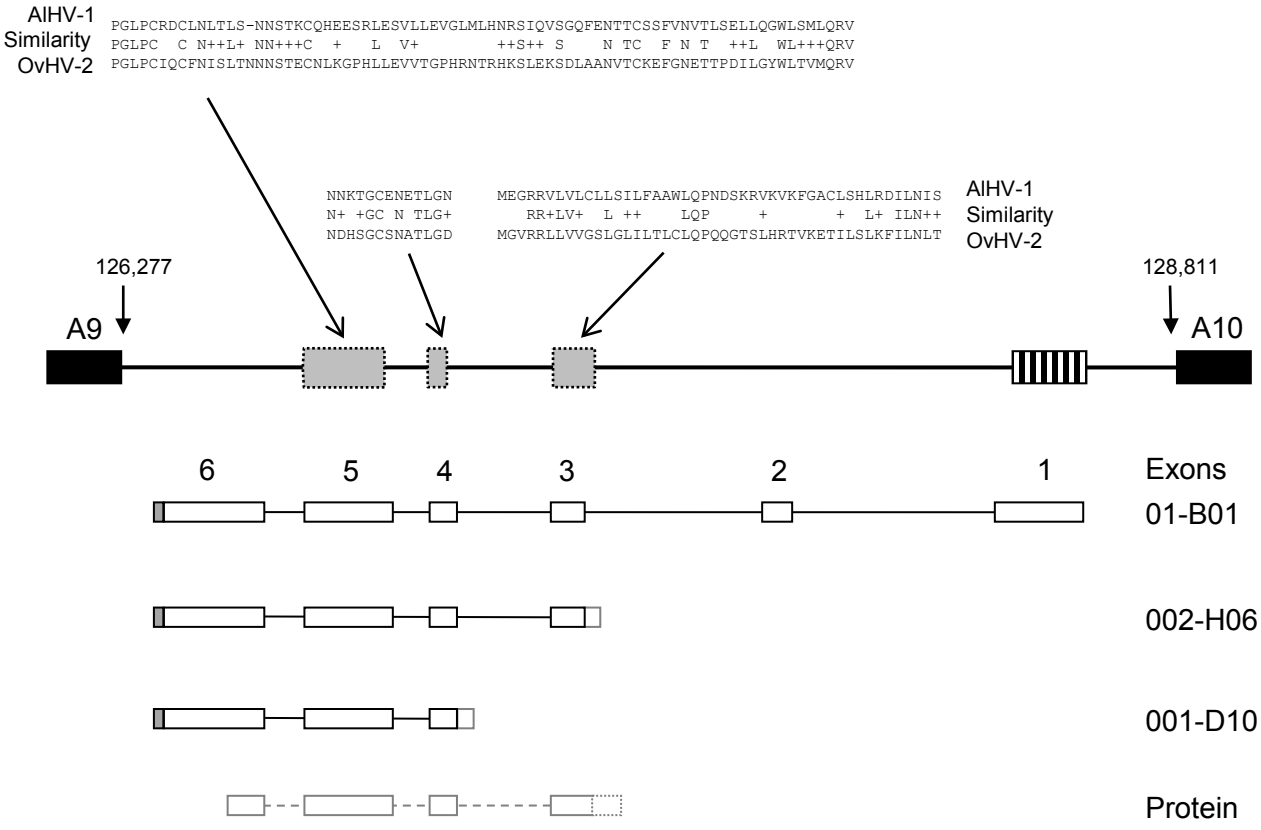
Fig. 2. Western blotting of A9.5-HA polypeptide with anti-HA-HRP, visualised with Supersignal West Pico reagent (Pierce) using an ImageQuant LAS4000 instrument. A9.5-HA protein was purified from culture supernatant (a) or whole cell lysates (b), (c), (d) of HEK293T cells transiently-transfected with pVR-A9.5-HA by HA-specific affinity

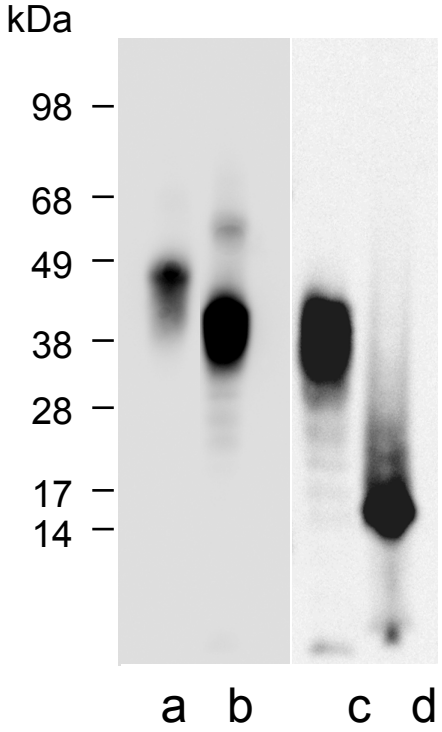
chromatography. Protein fractions were resolved in NuPAGE® 12 % Bis-Tris Gels (Life Technologies). Lane (a), soluble extracellular A9.5-HA; lane (b) detergent-soluble intracellular A9.5-HA; lane (c), A9.5-HA, as (b), incubated in PNGase buffer 3h; lane (d), A9.5-HA, as (c), treated with 2.5 units PNGase F for 3h. The positions and Mr (kDa) of protein size markers (SeeBlue® Plus2, Life Technologies) are given on the left of the figure.

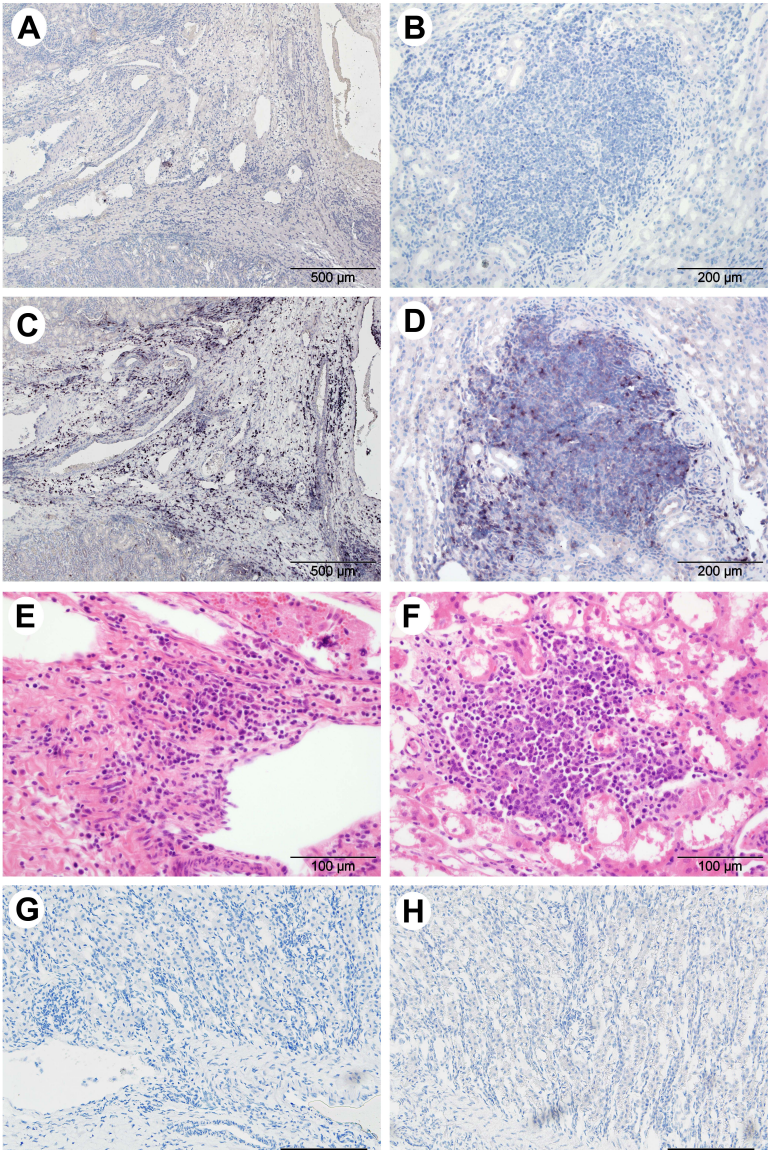
Fig.3. *A9.5* mRNA detected by RNA *in situ* hybridisation in MCF-affected kidney. Cells expressing the *A9.5* transcript were visualised in sections taken from bovine kidney from either MCF-affected or negative control animals. Panels A, C, E. MCF case 1. (A) sense RNA probe negative control (×40 magnification) for (C) antisense RNA probe detection of *A9.5*-expressing lymphoid cells (brown/black pigment) in kidney interstitial tissue at the level of the cortico-medullary junction (×40 magnification). (E) HE stained section, semi-serial to (A) and (C), illustrating the morphology of the infiltrating lymphoid cells (mononuclear inflammatory cells typical of lymphocytes) (×200 magnification). Panels B, D, F. MCF Case 2. (B) sense RNA probe negative control (×100 magnification) for (D) antisense RNA probe detection of *A9.5*-expressing cells (brown/black pigment) in an inflammatory focus (×100 magnification). (F) HE stained section, semi-serial to (B) and (D), illustrating morphology of infiltrating mononuclear inflammatory (lymphoid) cells (×200 magnification). Panels G, H. Negative control animal. (G) sense RNA probe, (H) antisense RNA probe (×100 magnification).

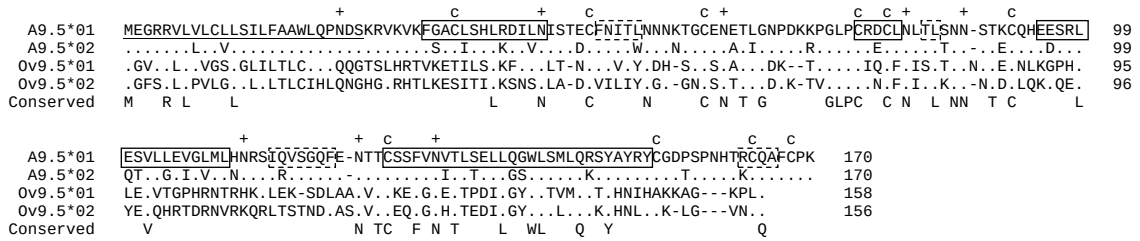
Fig.4. Features of the A9.5 polypeptide sequence and conservation in the putative Ov9.5 homologues. The predicted polypeptide sequences of the *A9.5* and *Ov9.5* alleles were aligned using MUSCLE and MAFFT. Residues identical to the A9.5*01 polypeptide sequence are replaced by dots (.). The predicted signal sequence of A9.5*01 is underlined; the positions of predicted N-linked glycosylation sites are marked by plus signs (+); and cysteine residues in the mature A9.5 polypeptide are indicated by the letter c above the

alignment. The positions of secondary structural elements predicted in the A9.5 structure by I-Tasser are show by boxes in the A9.5 sequence line: solid boxes for α–helix and dashed boxes for β-sheet. Where residues were conserved in all four sequences they are shown in the 'Conserved' line beneath the alignment.


AlHV-1
Similarity
OvHV-2
PGLPCRDCLNLTLS-NNSTKCQHEESRLESVLLEVGLMLHNRSIQVSGQFENTTCSSFVNVTLSELLQGWLSMLQRV
PGLPC C N++L+ NN+++C + L V+ ++S++ S N TC F N T ++L WL+++QRV
PGLPCIQCFNISLTNNNSTECNLKGPHLLEVVTGPHRNTRHKSLEKSDLAANVTCKEFGNETTPDILGYWLTVMQRV
NNKTGCENETLGN MEGRRVLVLCLLSILFAAWLQPNDSKRVKVKFGACLSHLRDILNIS
N+ +GC N TLG+ RR+LV+ L ++ LQP + + L+ ILN++
NDHSGCSNATLGD MGVRRLLVVGSLGLILTLCLQPQQGTSLHRTVKETILSLKFILNLT
AlHV-1
Similarity
OvHV-2
126,277
128,811
A9
A10
6
5
4
3
2
1
Exons
01-B01
002-H06
001-D10
Protein


kDa
98
68
49
38
28
17
14
a
b
c
d


A
500 µm
B
200 µm
C
500 µm
D
200 µm
E
100 µm
F
100 µm
G
200 µm
H
200 µm


A9.5*01 MEGRRVLVLCLLSILFAAWLQPNDSKRVKVKFGACLSHLRDILNISTECFNITLNNNKTGCENETLGNPDKKPGLPCRDCLNLTLSNN-STKCQHEESRL 99
A9.5*02 .....L..V.....S..I...K..V....D.....W...N.....A.I.....R.....E.....T..-..E....D... 99
Ov9.5*01 .GV..L..VGS.GLILTLC...QQGTSLHRTVKETILS.KF...LT-N...V.Y.DH-S..S.A...DK--T.....IQ.F.IS.T..N..E.NLKGPH. 95
Ov9.5*02 .GFS.L.PVLG..L.LTLCIHLQNGHG.RHTLKESITI.KSNS.LA-D.VILIY.G.-GN.S.T...D.K-TV.....N.F.I..K..-N.D.LQK.QE. 96
Conserved M R L L L N C N C N T G GLPC C N L NN T C L
A9.5*01 ESVLLEVGLMLHNRSIQVSGQFE-NTTCSSFVNVTLSELLQGWLSMLQRSYAYRYCGDPSPNHTRCQAFCPK 170
A9.5*02 QT..G.I.V..N....R.....-.....I..T...GS.....K.....T.....K..... 170
Ov9.5*01 LE.VTGPHRNTRHK.LEK-SDLAA.V..KE.G.E.TPDI.GY..TVM..T.HNIHAKKAG---KPL. 158
Ov9.5*02 YE.QHRTDRNVRKQRLTSTND.AS.V..EQ.G.H.TEDI.GY...L...K.HNL..K-LG---VN.. 156
Conserved V N TC F N T L WL Q Y Q