Discovery of a small arterivirus gene that overlaps the GP5 coding sequence and is important for virus production

Andrew E. Firth,[1] Jessika C. Zevenhoven-Dobbe,[2] Norma M. Wills,[3] Yun Young Go,[4] Udeni B. R. Balasuriya,[4] John F. Atkins,[3,5] Eric J. Snijder[2] and Clara C. Posthuma[2]

Correspondence
Eric J. Snijder
e.j.snijder@lumc.nl
Clara C. Posthuma
c.c.posthuma@lumc.nl

[1]Department of Pathology, University of Cambridge, Cambridge, UK

[2]Molecular Virology Laboratory, Department of Medical Microbiology, Center of Infectious Diseases, Leiden University Medical Center, Leiden, The Netherlands

[3]Department of Human Genetics, University of Utah, Salt Lake City, UT 84112-5330, USA

[4]Department of Veterinary Science, Maxwell H. Gluck Equine Research Center, University of Kentucky, Lexington, KY 40546-0099, USA

[5]BioSciences Institute, University College Cork, Cork, Ireland

Received 26 November 2010
Accepted 7 February 2011

The arterivirus family (order *Nidovirales*) of single-stranded, positive-sense RNA viruses includes porcine respiratory and reproductive syndrome virus and equine arteritis virus (EAV). Their replicative enzymes are translated from their genomic RNA, while their seven structural proteins are encoded by a set of small, partially overlapping genes in the genomic 3′-proximal region. The latter are expressed via synthesis of a set of subgenomic mRNAs that, in general, are functionally monocistronic (except for a bicistronic mRNA encoding the E and GP2 proteins). ORF5, which encodes the major glycoprotein GP5, has been used extensively for phylogenetic analyses. However, an in-depth computational analysis now reveals the arterivirus-wide conservation of an additional AUG-initiated ORF, here termed ORF5a, that overlaps the 5′ end of ORF5. The pattern of substitutions across sequence alignments indicated that ORF5a is subject to functional constraints at the amino acid level, while an analysis of substitutions at synonymous sites in ORF5 revealed a greatly reduced frequency of substitution in the portion of ORF5 that is overlapped by ORF5a. The 43–64 aa ORF5a protein and GP5 are probably expressed from the same subgenomic mRNA, via a translation initiation mechanism involving leaky ribosomal scanning. Inactivation of ORF5a expression by reverse genetics yielded a severely crippled EAV mutant, which displayed lower titres and a tiny plaque phenotype. These defects, which could be partially complemented in ORF5a-expressing cells, indicate that the novel protein, which may be the eighth structural protein of arteriviruses, is expressed and important for arterivirus infection.

INTRODUCTION

Arteriviruses are enveloped viruses (approx. 60 nm diameter) with a 13–16 kb positive-sense RNA genome. The family *Arteriviridae* (Snijder & Meulenberg, 1998; Snijder & Spaan, 2006) currently comprises four members: equine arteritis virus (EAV, the family prototype), lactate dehydrogenase-elevating virus (LDV) of mice, porcine reproductive and respiratory syndrome virus (PRRSV) and simian haemorrhagic fever virus (SHFV). The consequences of arterivirus infection can range from an asymptomatic, persistent or acute infection to abortion or lethal haemorrhagic fever (Maclachlan *et al.*, 2007). Three of the four arteriviruses (EAV, LDV and SHFV) were first isolated and characterized approximately 50 years ago. In the late 1980s, two distantly related PRRSV genotypes, European (PRRSV-EU) and North American (PRRSV-NA), sharing a genome-wide mean nucleotide identity of only ~63 % (Allende *et al.*, 1999; Suárez *et al.*, 1996), emerged simultaneously on both sides of the Atlantic (Collins *et al.*, 1992; Wensvoort *et al.*, 1991). Since then, the virus has been causing worldwide epidemics of a previously unknown reproductive and respiratory disease and has become one of the most economically important swine diseases. Recently, a large outbreak of highly virulent PRRSV affected the Asian swine industry, causing considerable economic losses (Zhou & Yang, 2010).

The unification of the previously unclassified members of the arterivirus family was the direct result of sequence analysis of their genomes, which also revealed an intriguing relationship with coronaviruses. Despite striking differences in genome size and virion structure, the genome organization and expression strategy of the two groups were found to be comparable and their key replicative enzymes were postulated to share common ancestry (den Boon *et al.*, 1991). A specific feature of their genome expression strategy is the generation of a nested set of subgenomic (sg) mRNAs to express a variety of genes from the 3′-proximal region of the genome (Pasternak *et al.*, 2006; Sawicki *et al.*, 2007). Arteriviruses and coronaviruses were united in the order *Nidovirales*, which was subsequently expanded by the addition of the roniviruses, toroviruses and bafiniviruses (Gorbalenya *et al.*, 2006).

The polycistronic arterivirus genome contains at least nine known ORFs (Fig. 1a). The replicative enzymes are encoded by the 5′-proximal replicase ORFs 1a and 1b, yielding two long non-structural polyproteins, pp1a and pp1ab, with the synthesis of the latter depending on a ribosomal frameshift near the 3′ end of ORF1a. The pp1a and pp1ab precursors are processed into at least 13 mature non-structural proteins (nsps) by three or four viral proteinases encoded by ORF1a (Ziebuhr *et al.*, 2000; Fang & Snijder, 2010). In EAV, PRRSV and LDV, seven relatively small genes have been identified in the 3′-proximal domain of the genome, positioned in the same relative order downstream of ORF1b (Fig. 1a). In SHFV, this region of the genome appears to contain three additional ORFs, which may have arisen from the duplication of ORFs 2–4 (Godeny *et al.*, 1998) but have not been characterized in any detail. In the other arteriviruses, ORFs 2a–7 encode seven structural proteins: the nucleocapsid (N) protein and six envelope proteins, the glycoproteins (GP) GP2, GP3, GP4 and GP5, and the non-glycosylated envelope (E) and membrane (M) proteins (Snijder & Meulenberg, 1998; Snijder & Spaan, 2006). By using reverse genetics to engineer gene knockouts, in both EAV and PRRSV the expression of each of these seven proteins was found to be critical for the production of infectious progeny (Molenkamp *et al.*, 2000; Wissink *et al.*, 2005). These experiments also revealed that cells transfected with mutants lacking expression of E, GP2, GP3 or GP4 release virus-like particles (VLPs) consisting of RNA and the GP5, M and N proteins (Wieringa *et al.*, 2004; Zevenhoven-Dobbe *et al.*, 2004). Together with the earlier characterization of arterivirus particles (de Vries *et al.*, 1992; Faaberg & Plagemann, 1995; Meulenberg *et al.*, 1995; Snijder *et al.*, 1999, 2003; Wieringa *et al.*, 2004), these data indicate that the N protein and the disulphide-linked GP5–M heterodimer form the structural core of the virion. The E protein and a GP2–GP3–GP4 heterotrimer are minor virus constituents that may play a role in receptor binding and/or virus entry (Van Breedam *et al.*, 2010). Still, many details of arterivirus assembly, maturation and structural protein function remain to be elucidated.

The genome organization of the arterivirus ORF2–7 region is extremely compact, with the termini of most ORFs overlapping with neighbouring genes and with RNA signals (transcription-regulatory sequences, TRSs) for sg RNA synthesis commonly being embedded in upstream ORFs (Fig. 1a). These TRSs are assumed to operate as attenuators of negative-sense RNA synthesis and to direct base pairing of the nascent negative-sense strand with a complementary sequence in the genomic 5′ end (leader TRS). This results in the generation of a 3′- and 5′-coterminal set of negative-sense templates that serve to produce the arterivirus sg mRNAs (Pasternak *et al.*, 2006; Sawicki *et al.*, 2007).

The arterivirus genome organization illustrates the strong selective pressure that RNA virus genomes are under to compress the maximum amount of coding and regulatory information into minimal sequence space. Overlapping coding sequences (CDSs) are particularly common in RNA viruses, but can be difficult to detect, especially when they are short. We have been using a number of complementary comparative computational approaches to systematically identify novel overlapping genes in RNA virus genomes (Firth & Brown, 2006; Chung *et al.*, 2008; Firth *et al.*, 2008; Firth & Atkins, 2009). When applied to arteriviruses, these methods detected a novel short ORF, hereafter termed ORF5a, overlapping the 5′ end of the GP5 CDS in the +2 reading frame. Here, we present bioinformatic evidence and experimental verification for the functional importance of ORF5a in arteriviruses.

RESULTS AND DISCUSSION

Computational analysis reveals the presence of a small ORF overlapping the arterivirus GP5 gene

The coding potential of ORF5a was first detected in an alignment of PRRSV-NA sequences by using the gene-finding software MLOGD (Firth & Brown, 2006; Chung *et al.*, 2008). MLOGD uses nucleotide and amino acid substitution matrices to model sequence evolution in dual-coding, single-coding and non-coding regions. It can be used to predict whether query ORFs are likely to be coding sequences via an approximate likelihood-ratio test, where the null model comprises any known CDSs and the alternative model comprises the known CDSs and the query ORF. When applied to the PRRSV-NA alignment, MLOGD detected a positive coding signature for the previously defined arterivirus CDSs and also for ORF5a (Fig. 1b). In the National Center for Biotechnology Information (NCBI) PRRSV-NA reference sequence, NC_001961 (15 428 nt), ORF5a covers nucleotides 13 778–13 930 (51 codons) and therefore overlaps the 5′ end of the GP5 CDS (nucleotides 13 788–14 390) by 48 codons (~24 % of the GP5 gene) in the +2 reading frame. ORF5a begins immediately 3′ of the ORF4 termination codon and 10 nt 5′ of the ORF5 initiation codon (Fig. 2b).

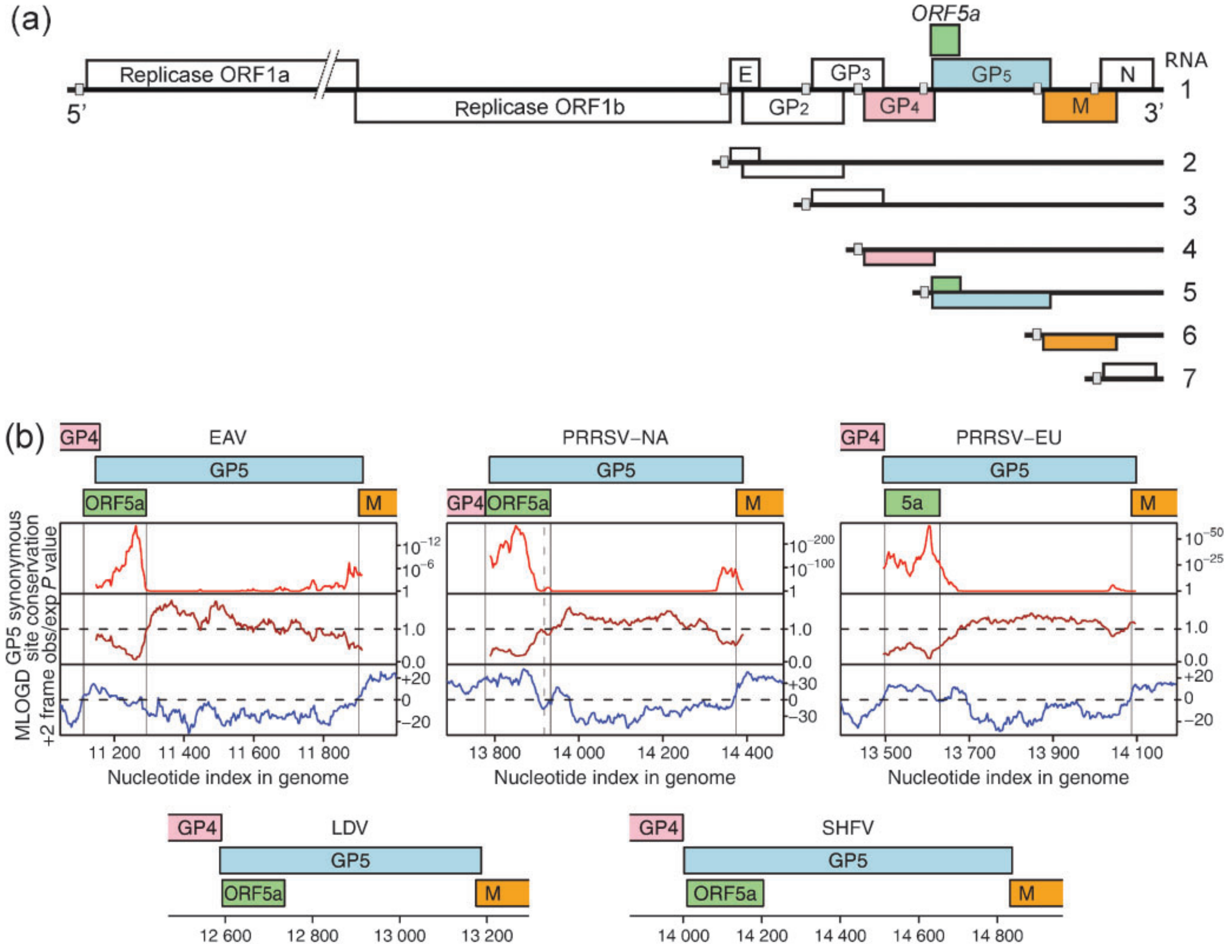

Fig. 1. (a) EAV genome organization and expression. The positions of the ORFs encoding the seven known structural proteins and the newly discovered ORF5a are indicated. ORF5a slightly overlaps with the GP4 gene and largely overlaps with the GP5 gene. Small grey boxes indicate the positions of the TRS that direct sg RNA synthesis and serve as junction sequences for the 5′ common leader sequence of arterivirus mRNAs. (b) Maps of the ORF5a/ORF5 region of the genomes of representative arterivirus sequences (see Fig. 2b for GenBank accession numbers). For EAV and the two PRRSV genotypes, plots are shown for several coding-potential statistics in a 15-codon sliding window. The top and middle panels illustrate the degree of conservation at synonymous sites within the GP5 CDS: the top panel depicts the probability that the degree of conservation within a given window could be obtained under a null model of neutral evolution at synonymous sites, while the middle panel depicts the absolute amount of conservation as represented by the ratio of the observed number of substitutions within a given window to the number expected under the null model. Scores below the dashed line correspond to more conserved regions and are indicative of overlapping functional elements, either coding or non-coding. Owing to the huge quantity of GP5 sequence data (8344 sequences) available for the PRRSV-NA plot, the formal *P* values for this analysis are extremely small for the conserved 5′ region. The bottom panel depicts the MLOGD coding potential score in the +2 frame (relative to the GP5 CDS). For illustrative purposes the M CDS, and the GP4 CDS in PRRSV-NA, were not included in the null-model CDS annotations so that these CDSs, like ORF5a, register positive coding potential (i.e. scores above the dashed line) in the MLOGD analysis. Note that, regardless of sign (positive or negative), the MLOGD signal tends to be weaker within regions of high conservation (e.g. within ORF5a) due to there being fewer substitutions with which to discriminate between the null model and the alternative model. A dashed vertical line indicates the 3′ end of the common ORF5a variant (46 codons instead of 51 codons) in PRRSV-NA.

A similar ORF5a was identified in PRRSV-EU and the other arteriviruses, EAV (Fig. 2a), LDV and SHFV, and again MLOGD detected a positive coding signature for ORF5a using sequence alignments of PRRSV-EU, EAV and LDV. Incidentally, the presence of this ORF in EAV, though not its bioinformatic coding signature, had

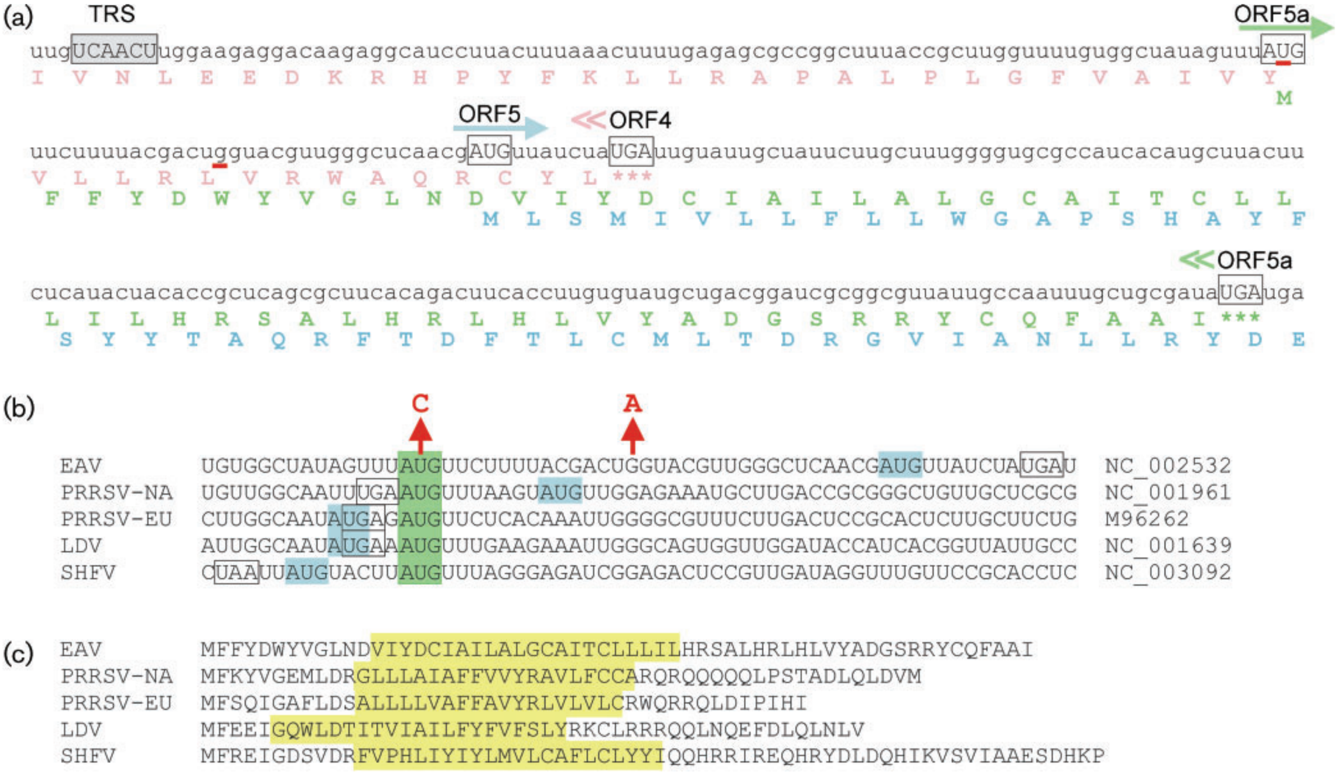
(a)
TRS
ORF5a
uugUCAACUuggaagaggacaagaggcauccuuacuuuaaacuuuugagagcgccggcuuuaccgcuugguuuuguggcuauaguuuAUG
I V N L E E D K R H P Y F K L L R A P A L P L G F V A I V Y
M
ORF5
ORF4
uucuuuuacgacugguacguugggcucaacgAUGuuaucuaUGAuuguauugcuauucuugcuuuggggugcgccaucacaugcuuacuu
V L L R L V R W A Q R C Y L ***
F F Y D W Y V G L N D V I Y D C I A I L A L G C A I T C L L
M L S M I V L L F L L W G A P S H A Y F
ORF5a
cucauacuacaccgcucagcgcuucacagacuucaccuuguguaugcugacggaucgcggcguuauugccaauuugcugcgauaUGAuga
L I L H R S A L H R L H L V Y A D G S R R Y C Q F A A I ***
S Y Y T A Q R F T D F T L C M L T D R G V I A N L L R Y D E
(b)
C
A
EAV UGUGGCUAUAGUUUAUGUUCUUUUACGACUGGUACGUUGGGCUCAACGAUGUUAUCUAUGAU NC_002532
PRRSV-NA UGUUGGCAAUUUGAAUGUUUAAGUAUGUUGGAGAAAUGCUUGACCGCGGGCUGUUGCUCGCG NC_001961
PRRSV-EU CUUGGCAAUAUGAGAUGUUCUCACAAAUUGGGGCGUUUCUUGACUCCGCACUCUUGCUUCUG M96262
LDV AUUGGCAAUAUGAAAUGUUUGAAGAAAUUGGGCAGUGGUUGGAUACCAUCACGGUUAUUGCC NC_001639
SHFV CUAAUUAUGUACUUAUGUUUAGGGAGAUCGGAGACUCCGUUGAUAGGUUUGUUCCGCACCUC NC_003092
(c)
EAV MFFYDWYVGLNDVIYDCIAILALGCAITCLLLILHRSALHRLHLVYADGSRRYCQFAAI
PRRSV-NA MFKYVGEMLDRGLLLAIAFFVVYRAVLFCCARQRQQQQQLPSTADLQLDVM
PRRSV-EU MFSQIGAFLDSALLLLVAFFAVYRLVLVLCRWQRRQLDIPIHI
LDV MFEEIGQWLDTITVIAILFYFVFSLYRKCLRRRQQLNQEFDLQLNLV
SHFV MFREIGDSVDRFVPHLIYIYLMVLCAFLCLYYIQQHRRIREQHRYDLDQHIKVSVIAAESDHKP

Fig. 2. Expression of arterivirus ORF5a. (a) RNA sequence and translation of the EAV ORF5a region, showing the slightly overlapping GP4 and GP5 genes, and the newly discovered ORF5a that overlaps with both. The TRS for sg RNA5 synthesis and relevant initiation and termination codons are indicated. Red underscores indicate residues mutated to knock out ORF5a in mutant Δ5a-2. (b) Context of the ORF5a (green) and ORF5 (blue) translation initiation codons in five representative arterivirus sequences, illustrating the potential for leaky scanning being the translation initiation mechanism for the second reading frame (see text for details). The figure shows that the ORF5a initiation codon is the more upstream in EAV and PRRSV-NA and the more downstream in the other arteriviruses. The ORF4 termination codon is boxed and the two residues mutated to knock out EAV ORF5a in mutant Δ5a-2 are indicated by arrows. (c) Amino acid sequences of the ORF5a proteins of five representative arterivirus sequences. The position of the putative transmembrane domain, as predicted by PHOBIUS (http://phobius.binf.ku.dk; Käll *et al.*, 2004), is indicated in yellow. The ORF5a protein is predicted to be a type III membrane protein (lumenal amino-terminal domain, central signal-anchor/transmembrane domain and cytoplasmic carboxy-terminal domain).

previously been noted by Pasternak *et al.* (2000). Since there is only a single SHFV sequence available, MLOGD could not be used to test the coding potential of ORF5a in that virus. Although this analysis was based on the application of MLOGD to specific ORFs, for illustrative purposes the MLOGD score in a 15-codon sliding window in the +2 reading frame (with respect to ORF5) is shown in Fig. 1(b) for alignments of 127 EAV, 232 PRRSV-NA and 34 PRRSV-EU sequences (i.e. the available sequences with coverage of both ORF5 and at least 150 nt of its 5′ and 3′ flanking sequences). Positive coding signatures are apparent for ORF5a (and also for the M and GP4 CDSs in cases where they fall in the +2 reading frame with respect to the GP5 CDS). Additional alignments were generated for 224 EAV, 8344 PRRSV-NA and 587 PRRSV-EU sequences that have full coverage of the GP5 CDS, and were analysed for conservation at synonymous sites as described in Firth & Atkins (2009) [see also Simmonds *et al.* (2008); see Shi *et al.* (2010) and Zhang *et al.* (2010) for recent PRRSV and EAV phylogenies]. Enhanced conservation at synonymous sites is indicative of overlapping functional elements – either coding or non-coding. In all three alignments, greatly enhanced conservation was apparent in the region where ORF5a overlaps the GP5 CDS (Fig. 1b).

The newly discovered ORF5a has 59 codons in EAV, either 51 or 46 codons in PRRSV-NA, 43 codons in PRRSV-EU, 47 codons in LDV and 64 codons in SHFV (Fig. 2c). When applied to the five representative sequences in Fig. 2(c), despite there being considerable sequence divergence, PHOBIUS and SIGNALP 3.0 (Käll *et al.*, 2004; Bendtsen *et al.*, 2004) consistently predict that the ORF5a protein is a type III membrane protein with a short (5–12 aa) lumenal domain at its amino terminus, a central signal-anchor/transmembrane sequence, and a cytosolic carboxy-terminal domain of 13–31 aa.

The ORF5a protein and GP5 are probably expressed from the same subgenomic mRNA via leaky scanning

The regions around the ORF5a initiation codon in five representative arterivirus sequences are shown in Fig. 2(b). In EAV and PRRSV-NA, the ORF5a AUG codon is 5′ of the ORF5 initiation codon and the TRS for the sg RNA5 is located sufficiently far upstream to include the ORF5a AUG codon within that sg mRNA, with no additional intervening AUG codons (Snijder & Meulenberg, 1998; Pasternak *et al.*, 2000; Lin *et al.*, 2002). On the other hand, in PRRSV-EU, LDV and SHFV, the ORF5a AUG codon is located just a few nucleotides 3′ of the ORF5 initiation codon.

In eukaryotes, the efficiency of translation initiation is modulated by the identity of the nucleotides surrounding the initiation codon, with the optimal context (known as the ‘Kozak consensus sequence’) in mammals involving mainly a guanine at the +4 position and a purine at the −3 position, where the presence of an adenine at the −3 position is the strongest single indicator of efficient initiation (Kozak, 2002). When the context of the first AUG codon on a transcript is suboptimal, efficient leaky scanning to alternative downstream AUG codons can occur (Kozak, 2002). Here, neither the ORF5 nor ORF5a initiation codons have a guanine at the +4 position, but all five arterivirus ORF5 initiation codons have an adenine at the −3 position. In contrast, the ORF5a initiation codons have either a pyrimidine (EAV, PRRSV-NA and SHFV) or a guanine (PRRSV-EU and LDV) at the −3 position. Thus, in PRRSV-NA and EAV, where the ORF5a AUG codon is 5′, there is the potential for efficient leaky scanning to the ORF5 initiation codon. In contrast, in PRRSV-EU, LDV and SHFV, where the ORF5a AUG is 3′, conventional leaky scanning for ORF5a translation may be expected to be less efficient. However, in these sequences, the ORF5 and ORF5a AUG codons are very closely spaced (separated by just 2 or 5 nt) and it has been demonstrated that, if two AUG codons are separated by less than ~10 nt, efficient initiation may occur at the downstream AUG codon, even when the upstream AUG codon is in a strong context (Matsuda & Dreher, 2006).

The analysis of initiation codons and their context, and ORF5a length, was extended to all available arterivirus sequences in GenBank. Some 10 800 sequences with full or partial coverage of the ORF5a/ORF5 region were analysed. In contrast to the above comparative phylogenetic analyses, simply counting sequences ignores potential phylogenetic bias (some clades may be greatly over-represented relative to others; indeed; clustering the sequences with BLASTCLUST or CLUSTAL X shows this to be the case). Thus, the following analysis is more meaningful with respect to the range of features present rather than to the relative abundances of those features. Many sequences (including ~90 % of the ~9600 available PRRSV-NA sequences) cover ORF5 entirely but have no 5′ flanking sequence. Therefore they do not allow analysis of the upstream ORF5a initiation codon or its context, but still allow analysis of the presence or absence of premature termination codons within the part of ORF5a that overlaps ORF5. Nonetheless 136 EAV, 445 PRRSV-NA, 622 PRRSV-EU, 1 SHFV and 5 LDV sequences have coverage of the ORF5a initiation site, and the above-mentioned initiation sites and contexts are very highly conserved (except that ~63 % of PRRSV-EU and ~9 % of PRRSV-NA sequences have a guanine instead of an adenine as the nucleotide at the −3 position of the ORF5 initiation codon context). Similarly, the above-mentioned ORF5a lengths are highly conserved in those sequences with full or partial coverage of the ORF5a region, with the notable exception that PRRSV-NA exhibits two main forms of differing lengths: a 51 codon version (~26 % of sequences) and a 3′-truncated 46-codon version (~74 % of sequences). A very small fraction of sequences were found to be ORF5a-defective as a result of either premature ORF5a-frame termination codons or substitutions within the ORF5a initiation codon. However, these do not necessarily represent sequences from viable viruses, but might instead be due to sequencing errors, sequencing of defective virus sequences or curation errors (e.g. five sequence records in which non-contiguous GP3 and GP5 sequences appear to have been fused *in silico*). Indeed, similar numbers of GP5-defective sequences were found during the analysis.

To summarize, four completely separate computational analyses – conservation of an ORF, conservation of a potential translation mechanism, positive MLOGD coding signature and a greatly suppressed frequency of substitutions at ORF5-frame synonymous sites – all pointed to the existence of a novel coding sequence, ORF5a.

Inactivation of EAV ORF5a expression dramatically affects virus production

To investigate the biological relevance of ORF5a in the arterivirus replicative cycle, we used an EAV reverse-genetics system (van Dinten *et al.*, 1997) to engineer two virus mutants in which ORF5a expression was inactivated. In both mutants, Δ5a-1 and Δ5a-2, the ORF5a translation initiation codon was mutated (AUG to ACG; U-11113→C), while in the second mutant a premature termination codon was engineered at codon six of ORF5a (G-11128→A). Care was taken that the mutations introduced were translationally silent with respect to the overlapping GP4 gene (Fig. 2b). Following electroporation of *in vitro*-synthesized full-length RNA into BHK-21 cells, which are also susceptible to EAV infection, the phenotype of the ORF5a knockout mutants was compared with that of the wild-type control. Dual-labelling immunofluorescence (IF) microscopy for replicase (nsp3) and N protein was performed to verify comparable transfection efficiency and monitor the synthesis of genome and subgenomic mRNA7, respectively (not shown). Viral RNA synthesis in transfected cells was also analysed by Northern blot hybridization (not shown). Finally, infectivity titres in the medium of transfected cells were determined by plaque assay, both after the first cycle (18 h post-transfection) and following subsequent virus amplification in initially non-transfected BHK-21 cells.

Both IF microscopy and hybridization analysis of intracellular viral RNA demonstrated that the two ORF5a knockout mutants were viable and that their RNA synthesis could not be distinguished from that of the wild-type control. However, the analysis of viral progeny released into the medium of transfected cells revealed that both ORF5a knockout mutants were seriously crippled. Compared with the wild-type virus, they showed a reduction in virus titre of approximately two logs, and a tiny plaque size (Fig. 3a). Upon analysis of the 40 h post-transfection harvest of mutant Δ5a-1, several large plaques were also observed, suggesting rapid reversion of its single point mutation to the wild-type sequence, which was subsequently confirmed by sequence analysis of the ORF5a region of the progeny virus. However, no reversion was observed for mutant Δ5a-2, in which a double point mutation had been engineered.

Confirmation of the importance of EAV ORF5a in a second, virulent EAV isolate and in natural host cells

EAV strains can differ significantly in the severity of the disease they induce in their natural host (Balasuriya *et al.*, 2007; Maclachlan *et al.*, 2007). The molecular clone used in the initial experiments described above (pEAV515, a derivative of pEAV030; van Dinten *et al.*, 1997) was based on an EAV isolate that was extensively adapted to cell culture (>300 passages) and causes very mild clinical signs upon experimental infection of horses (Balasuriya *et al.*, 1999). To independently confirm the effect of an ORF5a knockout in a second EAV isolate, the two ORF5a-inactivating nucleotide substitutions of mutant Δ5a-2 were transferred to a molecular clone (pEAVrVBS) of the virulent Bucyrus strain (VBS) of EAV, which can cause severe disease in horses (Balasuriya *et al.*, 2007). Furthermore, in order to use a more natural host cell than the BHK-21 cells used above, we transfected the *in vitro*-synthesized full-length RNAs into equine pulmonary artery endothelial cells (EEC; Hedges *et al.*, 2001). Again, in both virus backbones, the Δ5a-2 mutant displayed a severely crippled phenotype, with tiny plaques (Fig. 3b) and infectivity titres reduced by at least two logs. Thus, this comparative study independently confirmed that the expression of ORF5a is critical for efficient EAV amplification, irrespective of the background of the molecular clone and cell type used for analysis.

Partial complementation of the EAV ORF5a knockout defect in ORF5a-expressing cells

Despite the overwhelming bioinformatic evidence for the functionality of ORF5a, we wanted to exclude the possibility that the crippled phenotype of the Δ5a-2 mutant was due to damaging some unidentified yet important RNA signal in the EAV genome. To this end, we investigated whether the infectivity defect could be complemented in a cell line expressing the ORF5a protein. In view of previous success in rescuing EAV structural-protein knockouts in BHK-21 cells containing alphavirus RNA replicons (Zevenhoven-Dobbe *et al.*, 2004), ORF5a

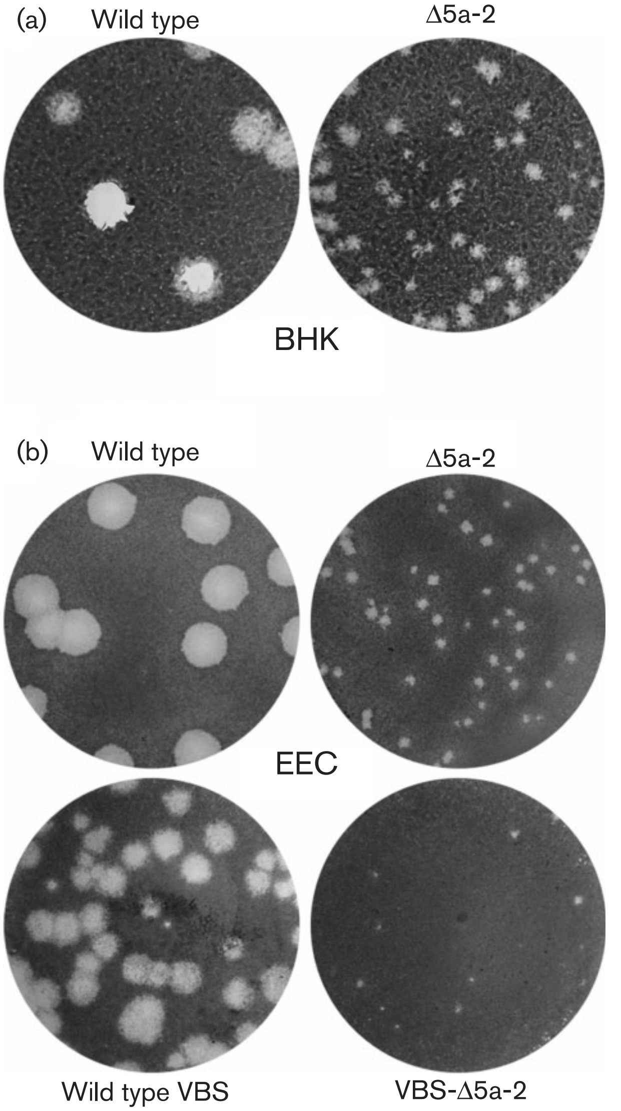

Fig. 3. (a) Comparison of plaque morphology on BHK-21 cells of wild-type EAV and ORF5a knockout mutant Δ5a-2, illustrating the severely crippled phenotype of the latter. In addition to the dramatically reduced plaque size, Δ5a-2 infectivity titres were reduced by approximately two logs. (b) Comparison of wild-type and Δ5a-2 plaque morphology on equine endothelial (EEC) cells using both a cell-culture-adapted EAV isolate (EAV515) and an isolate that is virulent in horses (VBS). In both virus backbones, Δ5a-2 plaque sizes and titres are severely reduced compared with those of the corresponding parental virus.

was cloned into pSinRep19, a Sindbis virus (SINV)-based expression vector (Agapov *et al.*, 1998). BHK-21 cells were transfected with the EAV ORF5a-expressing SinRep19 replicon and positive selection was applied by adding puromycin to the culture medium. BHK-21 cells containing a SinRep19 replicon that expresses GFP (SinRep19–GFP; Agapov *et al.*, 1998; Zevenhoven-Dobbe *et al.*, 2004) were used as a negative control.

Plaque assays of wild-type and Δ5a-2 virus were performed in both the ORF5a- and GFP-expressing cell cultures. While no difference in plaque size was observed for wild-type virus (not shown), the Δ5a-2 virus plaque size was clearly increased in ORF5a-expressing cell cultures (Fig. 4). A growth curve experiment in which GFP- and ORF5a-expressing cells were infected with Δ5a-2 mutant virus showed that the Δ5a-2 titre increased up to fivefold in the ORF5a-expressing cells (data not shown). Complete complementation to wild-type plaque size and titre could not be achieved, which may be due to a variety of technical or intrinsic factors such as expression level, but these experiments clearly demonstrated that the expression of the ORF5a protein directly influences the EAV life cycle.

The ORF5a protein: the eighth structural protein of arteriviruses?

In this study, we present evidence for the presence, conservation and biological significance of a previously unrevealed short CDS in the 3′-proximal domain of all arterivirus genomes. Much like the case of the arterivirus E and GP2 genes, whose translation is assumed to be initiated following leaky scanning of the 5′-proximal domain of sg mRNA2 (Snijder *et al.*, 1999), ORF5a and the overlapping GP5 gene are also likely to be expressed from a functionally bicistronic sg mRNA. Another striking similarity with the E–GP2 gene pair (Snijder *et al.*, 1999) is the fact that the ORF5a initiation codon precedes the GP5 gene in some arteriviruses (EAV and PRRSV-NA), but is just downstream of the ORF5 initiation codon in the others. This observation raises interesting questions about the evolution of this part of the arterivirus genome, in particular when evolving optimal sg mRNA synthesis and the expression levels of each of the viral structural proteins are taken into account. In EAV, a 10 nt segment starting with the ORF5 initiation codon can now be assumed to encode amino acids for three different proteins, GP4, the ORF5a protein and GP5, in three different reading frames, which again illustrates the pressure on RNA viruses to compress maximal coding information into a minimal sequence.

Previous attempts to engineer EAV-based vectors might, unknowingly, have been affected by the presence of ORF5a. For example, in one construct the overlap between ORF4 and ORF5 was removed to create an insertion site for foreign sequences (de Vries *et al.*, 2000). This was achieved by using a 24 nt insertion that, fortuitously, now proves to be exactly in-frame with ORF5a, although in the ORF5a protein it introduces the substitution of Asp-12 for Ala and Tyr-15 with the sequence Cys-Glu-Leu-Lys-Gln-His-Ala-Val-His. In view of the reported wild-type phenotype of this recombinant virus (de Vries *et al.*, 2000), this drastic change appears to be tolerated. Similarly, after passaging in cell culture 70 times, a spontaneous insertion of 12 nt into ORF5a (Ala-Arg-Ser-Leu inserted after the amino-terminal Met-Phe) of PRRSV-NA was tolerated (Han *et al.*, 2009). In contrast, derivatives of the EAV-based vector in which the 24 nt-sequence served to insert additional foreign sequences, encoding e.g. GFP (de Vries *et al.*, 2001) or ectodomains of other arterivirus GP5 proteins (Dobbe *et al.*, 2001), must have suffered because of the disruption of ORF5a. In fact, the instability and crippled or non-viable phenotypes reported for some of these constructs could (in part) be explained by a lack of expression of a functional ORF5a protein.

The lack of sequence similarity with the contents of protein databases, the protein sequence plasticity described above and the limited sequence conservation among arteriviruses (Fig. 2c) make it difficult to postulate a function for the ORF5a protein. Its expression from the 3′-proximal domain of the genome, and the apparent lack of an effect of ORF5a inactivation on viral RNA synthesis, would suggest that the protein is an additional virion component, and its hydrophobicity profile would place it in the virus envelope. On the other hand, knockouts of the other seven structural-protein genes were lethal to the virus, suggesting that the ORF5a protein belongs to a different functional category. Unfortunately, several attempts to raise antisera using synthetic peptides representing sequences from the

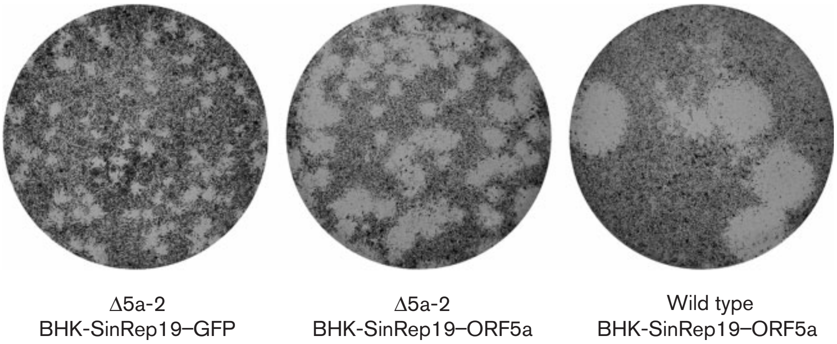

Fig. 4. Complementation of the EAV ORF5a knockout mutant Δ5a-2 in ORF5a-expressing cells. The Δ5a-2 small-plaque phenotype (left) was partially complemented when the plaque assay was performed using BHK-21 cells carrying an ORF5a-expressing SINV RNA replicon (middle; Agapov *et al.*, 1998). As a negative control, cells carrying a GFP-expressing replicon were used. The plaque morphology of wild-type virus on ORF5a-expressing cells is shown on the right.

non-hydrophobic domain of the EAV ORF5a protein were unsuccessful, thus obstructing our efforts to directly analyse virus particles for the presence of the ORF5a protein or analyse its subcellular localization in the infected cell. We used the SINV replicon system to express ORF5a proteins with epitope tags (haemagglutinin- or myc-tags at either the amino or carboxy terminus), but the Δ5a-2 mutant virus was no longer complemented by these fusion proteins. Finally, we were unable to demonstrate the presence of the ORF5a protein in virions by mass spectrometric analysis of purified EAV. However, a recent proteomics analysis of purified PRRSV-NA did identify several peptides from the ORF5a protein of the porcine arterivirus, and follow-up studies confirmed both the expression of the novel protein in infected cells and the production of antibodies recognizing the PRRSV ORF5a protein in infected swine (Johnson *et al.*, 2011). Thus, the combined data now obtained for PRRSV-NA and EAV strongly suggest that the ORF5a protein is the eighth structural protein of the arterivirus particle, although more extensive studies will be required to gain detailed insight into its function in arterivirus biology.

METHODS

Computational analysis. Virus sequences were obtained from GenBank initially in Feb 2007 and again in Nov 2010. Arterivirus nucleotide sequences with coverage of the ORF5 region were identified by applying NCBI TBLASTN (Altschul *et al.*, 1990) to the GP5 amino acid sequences from the five reference sequences shown in Fig. 2(b). In total one SHFV, five LDV, 435 EAV and 10 359 PRRSV sequences with full or partial coverage of ORF5 were retrieved. PRRSV sequences were separated into PRRSV-NA and PRRSV-EU genotypes (with respect to GP5), based on the TBLASTN results. Using a 76 % cut-off threshold on the TBLASTN 'maximum identity' score, all but eight of the 10 359 sequences were cleanly separated between the two genotypes. Within each group, ORF5 sequences were translated, aligned and back-translated to nucleotide sequence alignments using EMBOSS and CLUSTAL (Rice *et al.*, 2000; Larkin *et al.*, 2007). Flanking nucleotide sequences, where available, were similarly extracted, aligned and combined into the alignments. Patent sequences and constructs were removed prior to analysis, and sequences annotated as defective (found by using the search terms 'non-functional', 'truncated' and 'defective' in GenBank records) were inspected and removed from the analyses where appropriate. MLOGD statistics were calculated as described by Firth & Brown (2006). The synonymous site conservation statistic was calculated as described by Firth & Atkins (2009). Prior to these latter two analyses: (i) sequences with premature termination codons in ORF5 or ORF5a were also removed as being presumed-defective sequences (these amount to <1 % of sequences and are not obviously clustered phylogenetically), (ii) one EAV sequence with nine ambiguous nucleotide codes – indicative of poor sequence quality – was removed and (iii) since there was no shortage of PRRSV-NA sequences, PRRSV-NA sequences with any ambiguous nucleotide codes (~10 % of sequences) were removed.

Cells and plaque assays. EEC cells (Hedges *et al.*, 2001) were maintained in Dulbecco's modified Eagle's medium (DMEM; Mediatech) with 110 mg l^{-1} sodium pyruvate, 10 % FBS (HyClone; Thermo Scientific), 100 U penicillin ml^{-1} and 100 μg streptomycin ml^{-1} (Invitrogen), and 200 mM L-glutamine. BHK-21 cells (ATCC CCL10) were grown in Eagle's minimum essential medium with 10 % ferritin-supplemented bovine calf serum and 100 U penicillin ml^{-1} and 100 μg streptomycin ml^{-1}. Plaque assays on BHK-21 cells were performed for 3–5 days as described previously (Nedialkova *et al.*, 2010), whereas EEC cells were overlaid with DMEM containing 0.75 % carboxymethyl cellulose and incubated for 4 days.

Engineering of EAV ORF5a knockout mutants and reverse genetics. ORF5a knockout mutants Δ5a-1 and Δ5a-2 (Fig. 2b) were created as follows: point mutations were introduced in an appropriate shuttle vector by standard site-directed PCR mutagenesis (QuikChange Site-Directed Mutagenesis kit; Agilent Technologies). Using suitable restriction sites, the desired mutations in the ORF5a-coding sequence were transferred into EAV full-length cDNA clone pEAV515 or into pEAVrVBS, an infectious cDNA clone of the virulent Bucyrus strain of EAV (Balasuriya *et al.*, 2007). pEAV515 is a derivative of the EAV cDNA clone pEAV030 (van Dinten *et al.*, 1997), in which the *Bgl*II site at position 5217 in ORF1a was knocked out by using a translationally silent substitution. EAV515 virus has a wild-type phenotype, as determined by comparison with pEAV030-derived virus in growth-curve experiments and plaque assays. BHK-21 or EEC cells were electroporated with *in vitro*-synthesized full-length RNA transcripts of wild-type and mutant cDNA clones as described previously (van Dinten *et al.*, 1997). For IF assays, cells were seeded on coverslips, paraformaldehyde-fixed and double labelled for EAV nsp3 and N protein as described by van der Meer *et al.* (1998). Analysis of revertants was done by infecting fresh BHK-21 cells with supernatant harvested at 40 and 72 h post-transfection. Total RNA was isolated after 24 h by using TRIzol (Invitrogen) and the ORF5a region was amplified by RT-PCR and sequenced using standard protocols.

Generation of ORF5a-expressing BHK-21 cells. The coding sequence of ORF5a was PCR amplified using sense primer E1036 (CGTACGCGTCCATGGGGTTCTTTTACGACTGGTACGTTG; *Mlu*I site underlined) and antisense primer E1037 (CGTGCATGCTCA-TATCGCAGCAAATTGGCAATA; *Sph*I site underlined), and cloned between the *Mlu*I and *Sph*I sites of the pSinRep19 multiple cloning site (Agapov *et al.*, 1998). To obtain EAV ORF5a-expressing BHK-21 cells (BHK-SinRep19–ORF5a), *in vitro*-transcribed SinRep19–ORF5a RNA was transfected and puromycin selection (5 μg ml^{-1}) was applied from 24 h post-transfection onwards. A SinRep19 vector expressing GFP was used to produce a control cell line (BHK-SinRep19–GFP).

ACKNOWLEDGEMENTS

A. E. F. is supported by a grant from the Wellcome Trust (088789). J. F. A. is supported by Science Foundation Ireland grant 08/IN.1/B1889 and National Institutes of Health grant R01 GM079523. We acknowledge LUMC colleague Ali Tas for excellent technical assistance.

REFERENCES

Agapov, E. V., Frolov, I., Lindenbach, B. D., Prágai, B. M., Schlesinger, S. & Rice, C. M. (1998). Noncytopathic Sindbis virus RNA vectors for heterologous gene expression. *Proc Natl Acad Sci U S A* **95**, 12989–12994.

Allende, R., Lewis, T. L., Lu, Z., Rock, D. L., Kutish, G. F., Ali, A., Doster, A. R. & Osorio, F. A. (1999). North American and European porcine reproductive and respiratory syndrome viruses differ in non-structural protein coding regions. *J Gen Virol* **80**, 307–315.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

Balasuriya, U. B., Snijder, E. J., van Dinten, L. C., Heidner, H. W., Wilson, W. D., Hedges, J. F., Hullinger, P. J. & MacLachlan, N. J. (1999). Equine arteritis virus derived from an infectious cDNA clone is attenuated and genetically stable in infected stallions. *Virology* **260**, 201–208.

Balasuriya, U. B., Snijder, E. J., Heidner, H. W., Zhang, J., Zevenhoven-Dobbe, J. C., Boone, J. D., McCollum, W. H., Timoney, P. J. & MacLachlan, N. J. (2007). Development and characterization of an infectious cDNA clone of the virulent Bucyrus strain of Equine arteritis virus. *J Gen Virol* **88**, 918–924.

Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340**, 783–795.

Chung, B. Y., Miller, W. A., Atkins, J. F. & Firth, A. E. (2008). An overlapping essential gene in the *Potyviridae*. *Proc Natl Acad Sci U S A* **105**, 5897–5902.

Collins, J. E., Benfield, D. A., Christianson, W. T., Harris, L., Hennings, J. C., Shaw, D. P., Goyal, S. M., McCullough, S., Morrison, R. B. & other authors (1992). Isolation of swine infertility and respiratory syndrome virus (isolate ATCC VR-2332) in North America and experimental reproduction of the disease in gnotobiotic pigs. *J Vet Diagn Invest* **4**, 117–126.

den Boon, J. A., Snijder, E. J., Chirnside, E. D., de Vries, A. A. F., Horzinek, M. C. & Spaan, W. J. M. (1991). Equine arteritis virus is not a togavirus but belongs to the coronaviruslike superfamily. *J Virol* **65**, 2910–2920.

de Vries, A. A. F., Chirnside, E. D., Horzinek, M. C. & Rottier, P. J. M. (1992). Structural proteins of equine arteritis virus. *J Virol* **66**, 6294–6303.

de Vries, A. A. F., Glaser, A. L., Raamsman, M. J. B., de Haan, C. A. M., Sarnataro, S., Godeke, G. J. & Rottier, P. J. M. (2000). Genetic manipulation of equine arteritis virus using full-length cDNA clones: separation of overlapping genes and expression of a foreign epitope. *Virology* **270**, 84–97.

de Vries, A. A. F., Glaser, A. L., Raamsman, M. J. & Rottier, P. J. (2001). Recombinant equine arteritis virus as an expression vector. *Virology* **284**, 259–276.

Dobbe, J. C., van der Meer, Y., Spaan, W. J. M. & Snijder, E. J. (2001). Construction of chimeric arteriviruses reveals that the ectodomain of the major glycoprotein is not the main determinant of equine arteritis virus tropism in cell culture. *Virology* **288**, 283–294.

Faaberg, K. S. & Plagemann, P. G. W. (1995). The envelope proteins of lactate dehydrogenase-elevating virus and their membrane topography. *Virology* **212**, 512–525.

Fang, Y. & Snijder, E. J. (2010). The PRRSV replicase: exploring the multifunctionality of an intriguing set of nonstructural proteins. *Virus Res* **154**, 61–76.

Firth, A. E. & Atkins, J. F. (2009). A conserved predicted pseudoknot in the NS2A-encoding sequence of West Nile and Japanese encephalitis flaviviruses suggests NS1′ may derive from ribosomal frameshifting. *Virol J* **6**, 14.

Firth, A. E. & Brown, C. M. (2006). Detecting overlapping coding sequences in virus genomes. *BMC Bioinformatics* **7**, 75.

Firth, A. E., Chung, B. Y., Fleeton, M. N. & Atkins, J. F. (2008). Discovery of frameshifting in Alphavirus 6K resolves a 20-year enigma. *Virol J* **5**, 108.

Godeny, E. K., de Vries, A. A. F., Wang, X. C., Smith, S. L. & de Groot, R. J. (1998). Identification of the leader-body junctions for the viral subgenomic mRNAs and organization of the simian hemorrhagic fever virus genome: evidence for gene duplication during arterivirus evolution. *J Virol* **72**, 862–867.

Gorbalenya, A. E., Enjuanes, L., Ziebuhr, J. & Snijder, E. J. (2006). Nidovirales: evolving the largest RNA virus genome. *Virus Res* **117**, 17–37.

Han, W., Wu, J. J., Deng, X. Y., Cao, Z., Yu, X. L., Wang, C. B., Zhao, T. Z., Chen, N. H., Hu, H. H. & other authors (2009). Molecular mutations associated with the *in vitro* passage of virulent porcine reproductive and respiratory syndrome virus. *Virus Genes* **38**, 276–284.

Hedges, J. F., Demaula, C. D., Moore, B. D., McLaughlin, B. E., Simon, S. I. & MacLachlan, N. J. (2001). Characterization of equine E-selectin. *Immunology* **103**, 498–504.

Johnson, C. R., Griggs, T. F., Gnanandarajah, J. S. & Murtaugh, M. P. (2011). Novel structural protein in porcine reproductive and respiratory syndrome virus encoded in an alternative open reading frame 5 present in all arteriviruses. *J Gen Virol* **92**, 1107–1116.

Käll, L., Krogh, A. & Sonnhammer, E. L. (2004). A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* **338**, 1027–1036.

Kozak, M. (2002). Pushing the limits of the scanning mechanism for initiation of translation. *Gene* **299**, 1–34.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A. & other authors (2007). CLUSTAL W and CLUSTAL_X version 2.0. *Bioinformatics* **23**, 2947–2948.

Lin, Y. C., Chang, R. Y. & Chueh, L. L. (2002). Leader-body junction sequence of the viral subgenomic mRNAs of porcine reproductive and respiratory syndrome virus isolated in Taiwan. *J Vet Med Sci* **64**, 961–965.

Maclachlan, N. J., Balasuriya, U. B., Murtaugh, M. P., Barthold, S. W. & Lowenstine, L. J. (2007). Arterivirus pathogenesis and immune response. In *Nidoviruses*, pp. 325–337. Edited by S. Perlman, T. Gallagher & E. J. Snijder. Washington, DC: American Society for Microbiology.

Matsuda, D. & Dreher, T. W. (2006). Close spacing of AUG initiation codons confers dicistronic character on a eukaryotic mRNA. *RNA* **12**, 1338–1349.

Meulenberg, J. J. M., Petersen-den Besten, A., De Kluyver, E. P., Moormann, R. J. M., Schaaper, W. M. M. & Wensvoort, G. (1995). Characterization of proteins encoded by ORFs 2 to 7 of Lelystad virus. *Virology* **206**, 155–163.

Molenkamp, R., van Tol, H., Rozier, B. C. D., van der Meer, Y., Spaan, W. J. M. & Snijder, E. J. (2000). The arterivirus replicase is the only viral protein required for genome replication and subgenomic mRNA transcription. *J Gen Virol* **81**, 2491–2496.

Nedialkova, D. D., Gorbalenya, A. E. & Snijder, E. J. (2010). Arterivirus Nsp1 modulates the accumulation of minus-strand templates to control the relative abundance of viral mRNAs. *PLoS Pathog* **6**, e1000772.

Pasternak, A. O., Gultyaev, A. P., Spaan, W. J. M. & Snijder, E. J. (2000). Genetic manipulation of arterivirus alternative mRNA leader-body junction sites reveals tight regulation of structural protein expression. *J Virol* **74**, 11642–11653.

Pasternak, A. O., Spaan, W. J. M. & Snijder, E. J. (2006). Nidovirus transcription: how to make sense...? *J Gen Virol* **87**, 1403–1421.

Rice, P., Longden, I. & Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends Genet* **16**, 276–277.

Sawicki, S. G., Sawicki, D. L. & Siddell, S. G. (2007). A contemporary view of coronavirus transcription. *J Virol* **81**, 20–29.

Shi, M., Lam, T. T., Hon, C. C., Hui, R. K., Faaberg, K. S., Wennblom, T., Murtaugh, M. P., Stadejek, T. & Leung, F. C. (2010). Molecular epidemiology of PRRSV: a phylogenetic perspective. *Virus Res* **154**, 7–17.

Simmonds, P., Karakasiliotis, I., Bailey, D., Chaudhry, Y., Evans, D. J. & Goodfellow, I. G. (2008). Bioinformatic and functional analysis of RNA secondary structure elements among different genera of human and animal caliciviruses. *Nucleic Acids Res* **36**, 2530–2546.

Snijder, E. J. & Meulenberg, J. J. M. (1998). The molecular biology of arteriviruses. *J Gen Virol* **79**, 961–979.

Snijder, E. J. & Spaan, W. J. M. (2006). Arteriviruses. In *Fields Virology*, 5th edn, pp. 1337–1355. Edited by D. M. Knipe & P. M. Howley. Philadelphia, PA: Lippincott, Williams & Wilkins.

Snijder, E. J., van Tol, H., Pedersen, K. W., Raamsman, M. J. B. & de Vries, A. A. F. (1999). Identification of a novel structural protein of arteriviruses. *J Virol* **73**, 6335–6345.

Snijder, E. J., Dobbe, J. C. & Spaan, W. J. M. (2003). Heterodimerization of the two major envelope proteins is essential for arterivirus infectivity. *J Virol* **77**, 97–104.

Suárez, P., Zardoya, R., Martín, M. J., Prieto, C., Dopazo, J., Solana, A. & Castro, J. M. (1996). Phylogenetic relationships of European strains of porcine reproductive and respiratory syndrome virus (PRRSV) inferred from DNA sequences of putative ORF-5 and ORF-7 genes. *Virus Res* **42**, 159–165.

Van Breedam, W., Delputte, P. L., Van Gorp, H., Misinzo, G., Vanderheijden, N., Duan, X. & Nauwynck, H. J. (2010). Porcine reproductive and respiratory syndrome virus entry into the porcine macrophage. *J Gen Virol* **91**, 1659–1667.

van der Meer, Y., van Tol, H., Locker, J. K. & Snijder, E. J. (1998). ORF1a-encoded replicase subunits are involved in the membrane association of the arterivirus replication complex. *J Virol* **72**, 6689–6698.

van Dinten, L. C., den Boon, J. A., Wassenaar, A. L. M., Spaan, W. J. M. & Snijder, E. J. (1997). An infectious arterivirus cDNA clone: identification of a replicase point mutation that abolishes discontinuous mRNA transcription. *Proc Natl Acad Sci U S A* **94**, 991–996.

Wensvoort, G., Terpstra, C., Pol, J. M., ter Laak, E. A., Bloemraad, M., de Kluyver, E. P., Kragten, C., van Buiten, L., den Besten, A. & other authors (1991). Mystery swine disease in The Netherlands: the isolation of Lelystad virus. *Vet Q* **13**, 121–130.

Wieringa, R., de Vries, A. A. F., van der Meulen, J., Godeke, G. J., Onderwater, J. J. M., van Tol, H., Koerten, H. K., Mommaas, A. M., Snijder, E. J. & Rottier, P. J. (2004). Structural protein requirements in equine arteritis virus assembly. *J Virol* **78**, 13019–13027.

Wissink, E. H. J., Kroese, M. V., van Wijk, H. A. R., Rijsewijk, F. A. M., Meulenberg, J. J. M. & Rottier, P. J. M. (2005). Envelope protein requirements for the assembly of infectious virions of porcine reproductive and respiratory syndrome virus. *J Virol* **79**, 12495–12506.

Zevenhoven-Dobbe, J. C., Greve, S., van Tol, H., Spaan, W. J. M. & Snijder, E. J. (2004). Rescue of disabled infectious single-cycle (DISC) equine arteritis virus by using complementing cell lines that express minor structural glycoproteins. *J Gen Virol* **85**, 3709–3714.

Zhang, J., Timoney, P. J., Shuck, K. M., Seoul, G., Go, Y. Y., Lu, Z., Powell, D. G., Meade, B. J. & Balasuriya, U. B. (2010). Molecular epidemiology and genetic characterization of equine arteritis virus isolates associated with the 2006-2007 multi-state disease occurrence in the USA. *J Gen Virol* **91**, 2286–2301.

Zhou, L. & Yang, H. (2010). Porcine reproductive and respiratory syndrome in China. *Virus Res* **154**, 31–37.

Ziebuhr, J., Snijder, E. J. & Gorbalenya, A. E. (2000). Virus-encoded proteinases and proteolytic processing in the *Nidovirales. J Gen Virol* **81**, 853–879.