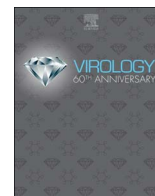




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



New insights about the regulation of Nidovirus subgenomic mRNA synthesis

Han Di, Ayisha A. McIntyre, Margo A. Brinton*

Department of Biology, Georgia State University, P.O. Box 4010, Atlanta, GA 30303, USA



ARTICLE INFO

Keywords:

Nidovirus
Coronavirus
Arterivirus
Transcription regulatory sequences
Subgenomic mRNAs
Next generation sequencing
Discontinuous RNA synthesis
Leader-body junction sequences

ABSTRACT

The members of the *Order Nidovirales* share a similar genome organization with two overlapping nonstructural polyproteins encoded in the 5' two-thirds and the structural proteins encoded in the 3' third. They also express their 3' region proteins from a nested set of 3' co-terminal subgenomic messenger RNAs (sg mRNAs). Some but not all of the Nidovirus sg mRNAs also have a common 5' leader sequence that is acquired by a discontinuous RNA synthesis mechanism regulated by multiple 3' body transcription regulating sequences (TRSs) and the 5' leader TRS. Initial studies detected a single major body TRS for each 3' sg mRNA with a few alternative functional TRSs reported. The recent application of advanced techniques, such as next generation sequencing and ribosomal profiling, in studies of arteriviruses and coronaviruses has revealed an expanded sg mRNA transcriptome and coding capacity.

1. Subgenomic mRNA production by members of the Order *Nidovirales*

The order *Nidovirales* consists of enveloped, single-stranded positive-sense RNA viruses that are currently classified into four virus families, *Coronaviridae*, *Arteriviridae*, *Mesoniviridae*, and *Roniviridae* (Gorbalenya et al., 2006; Zirkel et al., 2013). Due to the large number of new Nidoviruses recently identified by next generation sequencing (NGS) efforts, additional revisions to the *Nidovirales* taxonomy classification are currently being discussed. Among the four families, the *Coronaviridae* and *Arteriviridae* mainly infect mammalian hosts while the *Mesoniviridae* infect insects and the *Roniviridae* infect crustaceans. Although nidovirus genomes range in size from ~12 to ~30 kb, they share similarities in gene organization (two large, overlapping, non-structural protein ORFs encoded in the 5' two thirds and multiple structural protein ORFs encoded in the 3' third) and produce a nested set of 3' co-terminal subgenomic messenger RNAs (sg mRNA) in infected cells (Cong et al., 2017). Among the arteriviruses, simian hemorrhagic fever virus (SHFV) has the largest genome (15.7 kb) compared to those of equine arteritis virus (EAV), porcine reproductive and respiratory syndrome virus (PRRSV) and lactate dehydrogenase-elevating virus (LDV) (Snijder et al., 2013).

An early sucrose gradient sedimentation analysis of coronavirus murine hepatitis virus (MHV) viral RNA extracted from late harvest, sucrose gradient-purified virus detected several smaller, polyadenylated RNA fragments in addition to the full length genomic RNA (Lai and Stohman, 1978). These RNAs were initially thought to be

genomic degradation products but subsequent studies identified them as members of a nested set of six 3' co-terminal, subgenomic mRNAs that also possessed a 5' leader sequence, identical to that of the 5' region of the genome (Lai et al., 1981). Equine arteritis virus (EAV), the prototype of the family *Arteriviridae*, was later found to also produce six sg mRNAs that were 3' co-terminal and contained a common 5' leader sequence (de Vries et al., 1990b; van Berlo et al., 1982, 1986). A subsequent study showed that although most coronavirions contained a single copy of genome RNA, a few virions also contained a sg mRNA, suggesting that sg mRNAs can be packaged but inefficiently (Zhao et al., 1993).

The synthesis of sg mRNAs during an infection is not exclusive to the *Nidovirales*. However, the use of a discontinuous transcription mechanism for their generation and the resulting 5–3' co-terminal nested sg mRNAs produced are unique features of the coronaviruses and arteriviruses (Miller and Koev, 2000). Two models were proposed to explain the acquisition of the 5' leader sequence by coronavirus sg mRNAs. The initial model, a leader-primed transcription model (Lai, 1986) suggested that discontinuous transcription occurred during the synthesis of plus strand RNA with a single, full-length, minus strand anti-genome RNA serving as the template for synthesis of nascent genomic and sg mRNAs (Fig. 1A). It was hypothesized that short leader sequences synthesized from the 3' region of the anti-genome by the viral RNA dependent RNA polymerase served as primers to initiate RNA synthesis at multiple complementary intergenic sequences (internal promoters later called anti-transcription regulating sequences) located within the 5' region of the anti-genomic template. Elongation from

* Corresponding author.

E-mail address: mbrinton@gsu.edu (M.A. Brinton).

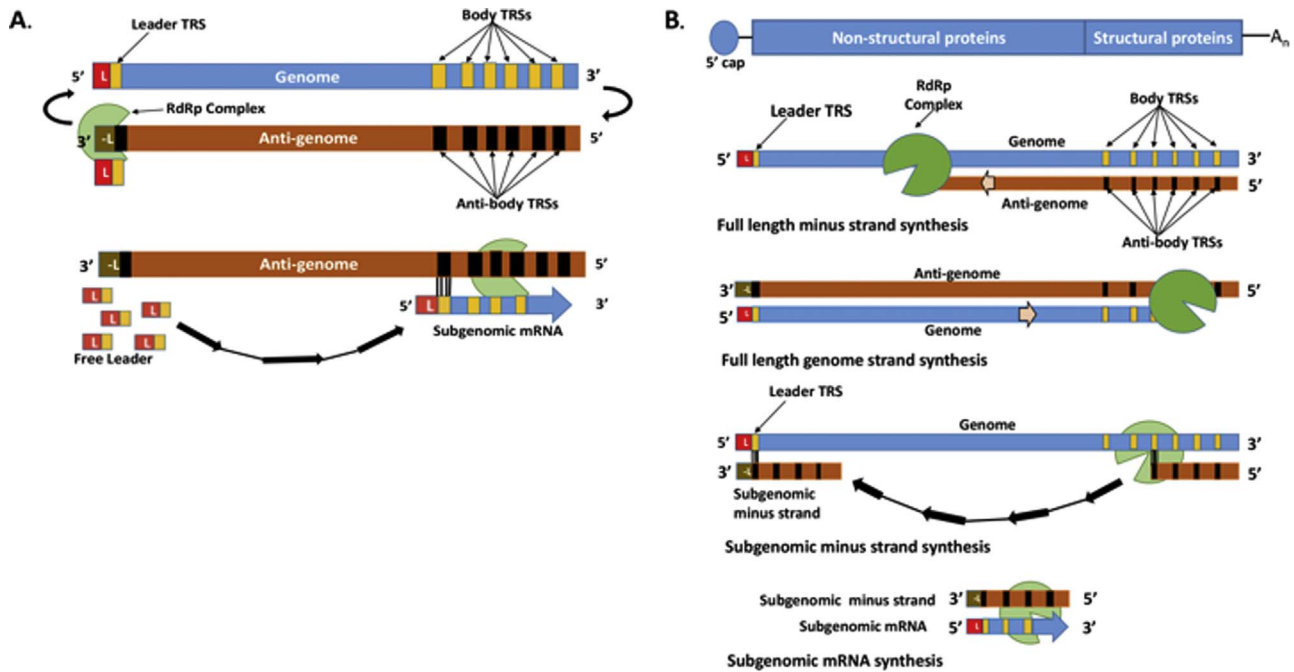


Fig. 1. Diagrams of models of nidovirus sg RNA generation. A. Leader-primed transcription model. Short leader fragments synthesized from the 3' region of the anti-genomic RNA function as primers for the synthesis of the sg mRNAs at multiple complementary anti-TRSs in the 5' region of the anti-genome. B. Discontinuous minus strand synthesis model. Minus strand synthesis from the genomic RNA template terminates at various genomic body TRSs and after the nascent minus strand anneals to the complementary genomic leader TRS, minus strand sg RNA synthesis is completed. The minus sg RNAs are then used as templates for sg mRNA synthesis.

these primers produced a set of 5–3' co-terminal sg mRNAs (Lai, 1986).

An alternate model later proposed suggested that the discontinuous step occurred during the synthesis of minus strand RNA from the genome (Sawicki and Sawicki, 1990; Sethna et al., 1989). The discovery of multiple subgenomic length minus strand RNAs as well as of multiple sg RNA replicative intermediates in cells infected with the coronavirus, porcine transmissible gastroenteritis virus (TGEV) or MHV, provided evidence supporting a discontinuous minus strand synthesis model (Sawicki and Sawicki, 1990; Sethna et al., 1989) (Fig. 1B). According to this model, full-length genomic minus strand RNAs as well as a 5' nested set of subgenomic length minus strand RNAs that have a common 3' anti-leader sequence are first synthesized from the genomic RNA and then serve as templates for genomic and sg mRNA synthesis, respectively. Subgenomic replicative intermediate and replicative form viral RNAs were subsequently identified in cells infected with the arterivirus EAV indicating that the same mechanism of sg RNA production is also used by a member of the family *Arteriviridae* (den Boon et al., 1996). Body transcription regulating sequences (TRSs) in the 3' region of the genome RNA are located at each of the premature termination sites for sg RNAs during minus strand RNA synthesis and were identified by amplifying and sequencing the leader-body TRS junction regions in the sg mRNAs produced. Each junction region consists of a 3' sequence fused to the 5' leader (van Marle et al., 1999a). In the EAV genome, the single 5' leader TRS 6 nt core sequence was shown to be located on the loop of a stem loop structure with two 10–15 nt flanking sequences forming the stem (Van Den Born et al., 2004; van den Born et al., 2005). Whether the leader regions of other arteriviruses form similar structures is not known. However, the leader TRS sequence of multiple coronavirus genomes has been shown to be located in the loop of a short RNA stem loop structure (Masters and Perlman, 2013) (Masters, 2013) and a *cis*-acting 5' proximal enhancer element was reported to be located in the leader region of the MHV genome (Wang and Zhang, 2000). The results of a study using a transfected *in vitro* synthesized sg mRNA indicated that long sg mRNAs which contain multiple TRSs can also function as templates for the synthesis of prematurely terminated minus strand RNAs (Wu and Brian, 2010).

The subsequent identification of multiple 5'–3' co-terminal sg mRNAs for additional coronavirus and arteriviruses led to the adoption of the view that all Nidoviruses utilize the same discontinuous minus strand synthesis strategy to produce a nested set of 3' co-terminal sg mRNAs with a common 5' leader in infected cells. However, data from studies on equine torovirus (EToV) (previously called Berne virus) and gill-associated virus (GAV), showed that not all nidoviruses produce 3' co-terminal sg mRNAs with a common 5' leader (Pasternak et al., 2006). EToV, produces a 3' nested set of four sg mRNAs, three of which (RNA 3, 4, and 5) do not possess a 5' leader sequence while RNA 2, the longest EToV sg mRNA, has a short 5' leader sequence (Smits et al., 2005; van Vliet et al., 2002). These findings suggest that EToV uses both discontinuous and non-discontinuous RNA synthesis mechanisms to produce its sg mRNAs. A study on GAV, a crustacean virus in the family *Roniviridae*, identified two 3' co-terminal sg mRNAs of different lengths that are initiated from the same intergenic sequence site. Neither of these sg mRNAs possesses a 5' leader sequence (Cowley et al., 2002).

2. Identification of functional alternative TRSs for known structural proteins

The TRSs for individual coronavirus and arterivirus sg mRNAs were initially identified by RT-PCR amplification and sequencing of leader-body junction regions of the sg mRNAs produced in infected cells. Typically a single major TRS was identified for each sg mRNA and it was assumed that the majority of the sg mRNA generated for each structural gene was regulated by the major TRS. However, two alternative body TRSs (also known as non-canonical TRSs) for sg mRNA3 were identified for EAV by amplification, cloning and sequencing of sg mRNA leader-body junctions (den Boon et al., 1996). A later EAV study not only confirmed the use of alternative TRSs for sg mRNA3, but also identified one alternative TRS for sg mRNA4 and one for sg mRNA5 (Pasternak et al., 2000). One alternative TRS was also identified for sg mRNA4 and one for sg mRNA7 of the PRRSV strain VR2332 (Nelsen et al., 1999). A different alternative TRS for sg mRNA7 was later

identified in the PRRSV strain tw91 (Lin et al., 2002). Mutation of the major TRS for sg mRNA4 in the EAV genome showed that the sg mRNAs produced from alternative TRSs although less abundant were still sufficient to generate infectious progeny virus but at a reduced titer, suggesting that alternative body TRSs can function as “back-ups” when the major TRS is not functional (Pasternak et al., 2000). Studies on coronaviruses have also shown that alternative body TRSs can function to produce sg mRNAs when the major body TRS of a 3' region ORF is mutated (Ozdarendeli et al., 2001; Schelle et al., 2005; Zuniga et al., 2004).

Although amplification, cloning and sequencing of sg mRNA leader-body junctions is an effective means of identifying TRSs that produce abundant sg mRNAs, it was less effective in identifying the TRSs that produce low abundance sg mRNAs due to the difficulty of detecting the RT-PCR product bands after electrophoresis on gels or to the limited number of clones screened. Although only one alternative EAV TRS was identified for sg mRNA4 and one for sg mRNA5, when both the known major and alternative TRSs for sg mRNA4 or sg mRNA5 were mutated, the mutant virus produced was still infectious and able to spread to neighboring cells, suggesting the possible existence of additional unidentified functional alternative TRSs producing sg mRNAs for each of these genes (Pasternak et al., 2000). Because NGS detects all sg mRNAs produced in RNA samples harvested from infected cells, including those at very low abundance, it can be used to identify all of the functional TRSs in a nidovirus genome and assess the relative abundance of each sg mRNA produced. The first NGS analysis of an arterivirus transcriptome identified a total of 96 functional body TRSs in the SHFV genome. Most of these were TRSs for alternative sg mRNAs of the SHFV structural proteins, with 8 identified for GP4 and 10 for GP5. (Di et al., 2017) (Fig. 2). A mass spectrometry analysis showed that for the majority of the SHFV structural proteins, the relative protein levels produced for each ORF correlated well with the combined sg mRNA abundance produced from the major TRS and all of the alternative TRSs. These results suggested that instead of just being “back-ups” for the major TRS, the alternative body TRSs are an integral part of the viral expression strategy and essential for the production of the optimal amount of each viral protein (Di et al., 2017). Because similar NGS analyses have not yet been performed for other arteriviruses it is

currently not known whether the SHFV genome is unique in having a high number of alternative TRSs for some of its ORFs or whether this is also a characteristic feature of other arteriviruses.

3. Identification of sg mRNAs for in-frame, C-terminal fragments of larger proteins

In addition to the alternative TRSs for PRRSV sg mRNAs for GP4 and N, a TRS located downstream of the AUG start codon of GP5 was identified in the PRRSV, strain VR2332, genome (Nelsen et al., 1999). Compared to sg mRNA5, this TRS generated a shorter sg mRNA (sg mRNA5-1) that is predicted to encode an in-frame, C-terminal peptide of GP5. A sg mRNA band migrating between sg mRNA5 and sg mRNA6 was also detected on SHFV Northern blots (Di et al., 2017; Vatter et al., 2014). Amplification, cloning and sequencing of the leader-body junctions revealed that this band was composed of seven different sg mRNAs of similar sizes that were generated from seven unique body TRSs located between TRS5 and TRS6. All seven of these sg mRNAs encode the same in-frame, C-terminal peptide of GP5 (Di et al., 2017) (Fig. 2). Sg mRNAs encoding an in-frame, C-terminal peptide from each of the SHFV structural proteins, GP2', GP4', GP2 and M, were also identified in infected cells (Di et al., 2017) (Fig. 2). Two previous studies on the coronaviruses, severe acute respiratory syndrome virus and infectious bronchitis virus, identified a novel sg mRNA encoding an in-frame, C-terminal peptide of the virion spike protein (Bentley et al., 2013; Hussain et al., 2005). Similarly, both a previous study and a recent NGS and ribosome profiling study of the coronavirus MHV detected translation initiation at an internal AUG within ORF5 (Irigoyen et al., 2016; Thiel and Siddell, 1994). Although both coronaviruses and arteriviruses appear to produce C-terminal fragments of larger viral proteins, the functions of these peptides are currently not known. Mutagenesis of the AUGs of five of the SHFV C-terminal peptides in an infectious clone provided preliminary evidence for the functional relevance of the GP5 and M C-terminal peptides. Both of these mutant viruses displayed a reduced viral yield and plaque size in cell culture (Di et al., 2017).

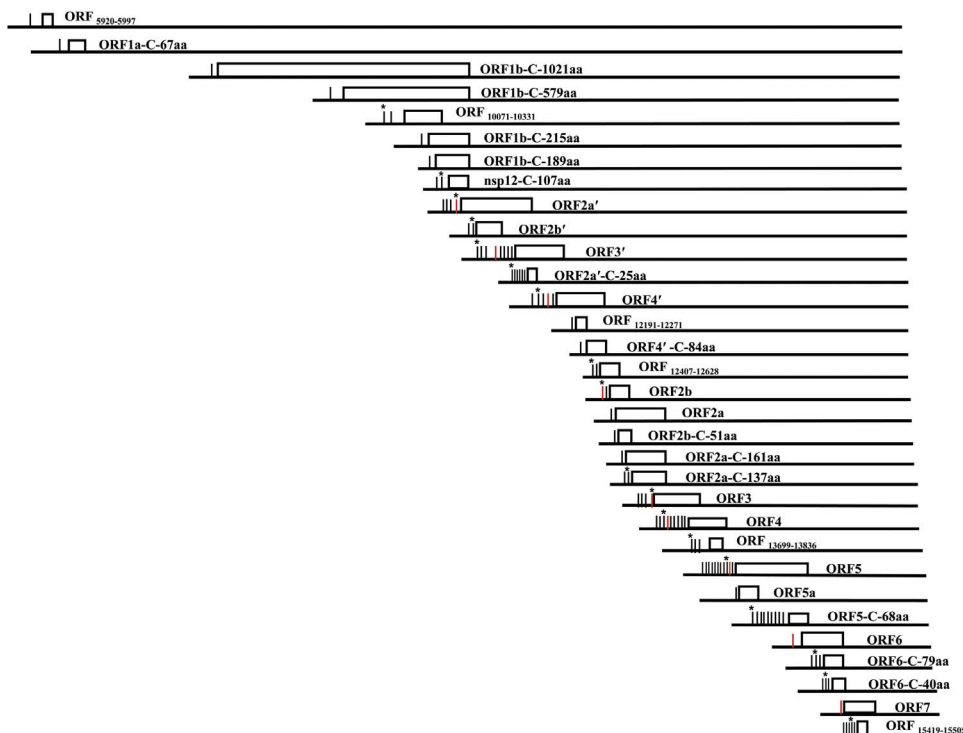


Fig. 2. Diagram of the SHFV transcriptome. The ORFs are represented by white boxes. Identified functional TRSs are indicated by vertical lines to the left each ORF. Red vertical lines represent the nine major SHFV body TRSs published prior to the recent SHFV NGS study. For ORFs with multiple functional TRSs, an asterisk indicates the TRS that produces the highest abundance of sg mRNA for that ORF. (Published in Di et al., 2017. Proc Natl Acad Sci USA 114(42): E8895-E8904).

4. Identification of sg mRNAs for alternative frame proteins

Knowledge of the coding capacity of arteriviruses and coronaviruses has expanded over the years with the discovery of additional ORFs. Arteriviruses were initially thought to encode three minor structural proteins, GP2, GP3 and GP4 and three major structural proteins GP5, membrane protein (M) and N (Snijder et al., 1999). An early Northern blot analysis of SHFV sg mRNA production in infected MA104 cells identified six strong sg mRNA bands. Based on their sizes, these sg mRNAs were predicted to encode the six known arterivirus minor and major structural proteins (Zeng et al., 1995). An extra set of minor structural protein ORFs for GP2', GP3' and GP4' was later shown to be encoded in the 3' region of the SHFV genome (Smith et al., 1997). A subsequent sg mRNA leader-body junction region amplification, cloning and sequencing study identified body TRSs for GP2' and GP4' but not for GP3' and it was proposed that GP3' was also expressed from the sg mRNA encoding GP2' (Godeny et al., 1998). A recent reanalysis of the SHFV sg mRNAs by Northern blotting using a more sensitive 5' leader probe detected ten sg mRNA bands including a sg mRNA3' for GP3' (Vatter et al., 2014). Additional structural protein ORFs, E and ORF5a, were later discovered in the genomes of EAV and PRRSV that were predicted to be expressed bicistronically from the sg mRNAs for GP2 and GP5, respectively (Firth et al., 2011; Johnson et al., 2011; Snijder et al., 1999). Bicistronic expression for the GP2' and E' proteins and for the E and GP2 proteins of SHFV was also initially proposed. However, separate sg mRNAs for both E' and GP2 were later identified in the NGS study of SHFV transcriptome (Di et al., 2017). Comparison of the relative sg mRNA and protein abundances suggested that the majority of E' is expressed from the 5' ORF of multiple E' sg mRNAs. In contrast, the abundance of the sg mRNAs with GP2 as the 5' ORF was much lower than the relative abundance of this protein suggesting that the majority of GP2 is translated as a second ORF from the E sg mRNAs. Some coronavirus proteins initially thought to be expressed only bicistronically from the second 5' ORF of a longer sg mRNA were later shown to also be translated from the 5' ORF of a different sg mRNA, for example, the mouse hepatitis virus (MHV) E protein (O'Connor and Brian, 2000; Zhang and Liu, 2000).

An additional alternative frame ORF overlapping ORF5 (GP5) was identified in a PRRSV genome and shown to express a protein designated ORF5a (Firth et al., 2011; Johnson et al., 2011). Although a separate sg mRNA encoding ORF5a as the 5' proximal ORF was not identified in that study, the NGS study of the SHFV transcriptome identified a sg mRNA for ORF5a at very low abundance (Di et al., 2017). An early study of PRRSV found that four USA PRRSV isolates, including VR2385 and ISU79, produced an additional sg mRNA3-1 that encodes an alternative frame ORF (ORF3-1) that overlaps ORF3 (GP3) (Meng et al., 1996). The production of this protein has not yet been confirmed. A recent bioinformatic study suggested the existence of a small ORF7a protein encoded in the +2 reading frame inside ORF7 (N) of type I and type II PRRSV genomes (Olasz et al., 2016). The recent SHFV NGS study identified five functional body TRSs that each generated sg mRNAs encoding ORF_{15419–15505} in the +2 reading frame located inside SHFV ORF7 (N) (Fig. 1) (Di et al., 2017). The SHFV NGS study also identified sg mRNAs encoding five additional alternative frame ORFs (Fig. 1). The predicted sizes of the proteins that would be translated from these alternative frame ORFs are 86aa, 25aa, 28aa, 73aa, 26aa, and 45aa. The 86 aa protein expressed from the largest of these alternative frame ORFs (ORF_{10071–10331}) was detected by mass spectrometry in infected MA104 extracts (Di et al., 2017). The low abundance of the sg mRNAs of each these proteins and the technical challenges of detecting small proteins from gels by MassSpec may be the reasons the rest of these small proteins were not detected. Studies on coronaviruses have also identified alternative frame proteins such as an alternative ORF located inside the N gene that encodes an accessory protein (Fischer et al., 1997; Liu et al., 2014; Meier et al., 2006). A recent analysis of coronavirus MHV protein ORF translation by

ribosome profiling and NGS detected the translation of multiple previously unidentified small alternative frame ORFs (Irigoyen et al., 2016). Many of these MHV alternative frame ORFs were initiated from non-canonical start codons such as CUG, suggesting an even larger coding potential for Nidovirus genomes. The authors proposed that some of these small alternative frame ORFs may function as regulators of the translation of downstream ORFs in the same sg mRNA (Irigoyen et al., 2016). Alternatively, increasing evidence from multiple host species suggests that small, functional proteins, some as short as 11 aa, translated from in frame or alternative frame ORFs in cellular mRNAs are involved in regulating a variety of cellular processes (Landry et al., 2015; Pueyo et al., 2016).

A coronavirus was recently found to also produce short non-coding RNAs. Three 18–22 nt small noncoding viral RNAs (svRNAs) were identified in SARS-CoV infected cells that were derived from the nsp3 (svRNA-nsp3.1 and -nsp3.2) and N (svRNA-N) genomic regions (Morales et al., 2017). Although similar in size to miRNAs, the generation of these RNAs was found to be RNase III, cell type, and host species independent. The functional relevance of one of these RNAs was indicated by the observed reduction in lung pathology and pro-inflammatory cytokine expression when svRNA-N was inhibited with an antagomir.

5. Identification of sg mRNA production from the genomic 5' nonstructural region

Although functional body TRSs were thought to be located only in the 3' structural protein region of Nidovirus genomes, a previous study on the arterivirus LDV identified a functional body TRS in the non-structural protein ORF1b region by amplifying, cloning and sequencing the leader-body junction of a long sg mRNA (Chen et al., 1993). This sg mRNA encoded an in-frame, C-terminal portion of ORF1b (200 aa) but translation of this truncated ORF1b polyprotein was not confirmed. A study on the coronavirus MHV identified three functional body TRSs in the ORF1a region by amplifying, cloning and sequencing the leader-body junctions (La Monica et al., 1992). The recent SHFV NGS study identified two functional body TRSs in ORF1a and eight in ORF1b that generated ten long sg mRNAs. Among these sg mRNAs, six encode different lengths of in-frame, C-terminal portions of ORF1b with the shortest one encoding only the C-terminus of nsp12 and the longest one encoding the C-terminus of nsp9 plus full-length nsp10, nsp11 and nsp12 (Fig. 2) (Di et al., 2017). ORF1b has been reported to be translated at a much lower efficiency (15–20%) than ORF1a due to the requirement of a -1 frameshift near the 3' end of ORF1a for ORF1a/1b translation to occur (den Boon et al., 1991; Snijder et al., 2013). Analysis of SHFV protein production by mass spectrometry detected 3 times more nsp12 and 1.2 times more nsp11 than nsp8, which is encoded at the 3' end of the ORF1a, suggesting that the long sg mRNAs produced from TRSs in the ORF1b region provide additional sources of nsp10, nsp11 and especially nsp12. However, the relative half-lives of the individual SHFV nonstructural proteins are not known and could also affect the relative abundances of these proteins. Previous studies on PRRSV detected the production of atypical sg mRNAs (designated heteroclitc sg mRNAs) in the ORF1a region. These sg mRNAs differed from typical sg mRNAs in having their 3' genomic sequences fused to various sites within ORF1a, rather than to a common 5' leader (Yuan et al., 2000, 2004). Production of heteroclitc sg mRNAs has not been reported for other arteriviruses.

6. Relative abundance of individual sg mRNAs and junction sequence heterogeneity

Northern blotting studies of virus infected cell extracts suggested that although the relative abundances of the individual viral sg mRNAs detected differed, the abundance of each sg mRNA was consistent at different times after infection (Di et al., 2017). The recent SHFV NGS

analysis allowed calculation of the relative abundance of each viral sg mRNA. The data indicated that the relative sg mRNA abundance produced from each identified body TRS, regardless of whether it is a major or alternative TRS, was consistent at early and later times after infection in both primary macrophages and MA104 cells, suggesting that sg mRNA production is primarily regulated by viral not cellular factors (Di et al., 2017).

Two previous PRRSV leader-body junction sequence analysis studies identified heterogeneous junction sequences (Lin et al., 2002; Meulenberg et al., 1993). The junction sequence of a sg mRNA is determined by the location within the particular body TRS where the polymerase disassociates and the location within the leader TRS where the nascent minus strand anneals. The SHFV NGS data indicated that about one-third of the total body TRSs identified, including both major and alternative TRSs, produced sg mRNAs with two or three heterogeneous junction sequences (Di et al., 2017). Heterogeneity in sg mRNA junction sequences produced from the same body TRS has also been detected for coronaviruses by amplification, cloning and sequencing the junction sequences or by NGS (Irigoyen et al., 2016; Makino et al., 1988; Schaad and Baric, 1993). The heterogeneity of sg mRNA junction sequences suggests that the fusion position between the leader and body TRSs can vary but the observation that the abundances of the sg mRNAs with each of the alternative junction sequences are consistent suggests that this variation is controlled by the characteristics of the viral elements regulating discontinuous minus strand synthesis at these sites.

7. Possible modes of sg mRNA regulation

The means by which discontinuous RNA synthesis is regulated by Nidoviruses is not well understood. Previous studies on both coronaviruses and arteriviruses have suggested that shorter sg mRNAs generated from TRSs located closer to the 3' end of the genome are usually in higher abundance (de Vries et al., 1990a; Krishnan et al., 1996; Pasternak et al., 2004). Studies of two closely located body TRSs showed that the downstream TRS could suppress the function of the upstream one (Joo and Makino, 1995; Pasternak et al., 2004). However, data from the SHFV NGS study clearly showed that across the genome there is not a linear correlation between the distance of a TRS from the 3' end of the genome and the corresponding sg mRNA abundance produced from that TRS. Instead, multiple genomic regions containing TRSs regulating high levels of discontinuous minus strand synthesis were flanked by regions containing TRSs regulating low levels of discontinuous synthesis (Di et al., 2017). Previous studies also observed that when mutations in a body TRS increased the duplex stability between the leader and body TRS sequences, the corresponding sg mRNA abundance increased with a few exceptions (Pasternak et al., 2001, 2003; Sola et al., 2005; Zuniga et al., 2004). However, some sequence motifs were found in the viral genomes with 100% similarity to the leader TRS core sequence that were not utilized as TRSs, suggesting that duplex stability is necessary but not sufficient for regulating discontinuous RNA synthesis. The duplex stability between the leader TRS and each of the seven closely located functional body TRSs for SHFV ORF5C was calculated and found not to show a linear correlation with the corresponding sg mRNA abundance (Di et al., 2017). These findings support previous suggestions that the local RNA secondary structure of each body TRS, which may be dynamic, as well as long distance genomic RNA-RNA interactions play important roles in regulating discontinuous RNA synthesis (Mateos-Gomez et al., 2013; Moreno et al., 2008; Pasternak et al., 2000; Sola et al., 2005).

EAV nsp1 plays an essential role in sg RNA generation but little is known about how protein-protein and protein-RNA interactions regulate sg RNA synthesis (Tijms et al., 2007). It was suggested that the interaction of nsp1 with a particular body TRS could stall a viral polymerase complex copying minus RNA from the genome at that site and that nsp1 may also facilitate targeting and base pairing between the

complementary body TRS and the leader TRS. The EAV nsp1 consists of the regions equivalent to nsp1 α and nsp1 β of other arteriviruses due to mutation of the nsp1 α cleavage site. The zinc finger located in the nsp1 α region was shown to be critical for sg RNA production. However, a subsequent study found that viral sg mRNA abundance was regulated by an intricate interaction network involving all of the EAV nsp1 subdomains (Nedialkova et al., 2010). EAV nsp10 was also shown to play an important role in sg RNA synthesis (van Marle et al., 1999b).

8. Conclusions

Following the initial discovery that coronaviruses produce a nested set of 3' co-terminal sg mRNAs with a common 5' leader sequence, the main emphasis of research was focused on determining the mechanism by which these sg mRNAs were generated. Minus strand sg RNAs were shown to be generated from the genomic template by a discontinuous minus strand synthesis mechanism. These sg RNAs then function as templates for sg mRNA synthesis. The amplification, cloning and sequencing of the leader-body junction sequences of the sg mRNAs produced for each viral gene typically identified a single major TRS that had sequence homology to the leader TRS. However, the detection of alternative regulatory sequences mediating discontinuous RNA synthesis (heteroclitte sg mRNAs) and of 3' co-terminal sg mRNAs without a 5' leader for some nidoviruses indicates the use of additional mechanisms for the generation of sg mRNAs among viruses in this order. Recent studies employing NGS and ribosome profiling have provided a greatly expanded view of the transcriptomes and coding capacities of both an arterivirus and a coronavirus. The finding that an arterivirus and a coronavirus may encode a number of small previously unidentified proteins opens new avenues for investigation of the cell-virus interaction.

Acknowledgements

Support for the work described was provided by Public Health Service research grant AI073824 to M. Brinton and by a Georgia State University Molecular Basis of Disease seed grant to H. Di. H. Di was also supported by a Georgia State University Molecular Basis of Disease fellowship.

References

- Bentley, K., Keep, S.M., Armesto, M., Britton, P., 2013. Identification of a noncanonically transcribed subgenomic mRNA of infectious bronchitis virus and other gammacoronaviruses. *J. Virol.* 87, 2128–2136.
- Chen, Z., Kuo, L., Rowland, R.R., Even, C., Faaberg, K.S., Plagemann, P.G., 1993. Sequences of 3' end of genome and of 5' end of open reading frame 1a of lactate dehydrogenase-elevating virus and common junction motifs between 5' leader and bodies of seven subgenomic mRNAs. *J. General Virol.* 74 (Pt 4), 643–659.
- Cong, Y., Verlhac, P., Reggiori, F., 2017. The Interaction between Nidovirales and autophagy components. *Viruses* 9.
- Cowley, J.A., Dimmock, C.M., Walker, P.J., 2002. Gill-associated nidovirus of Penaeus monodon prawns transcribes 3'-coterminal subgenomic mRNAs that do not possess 5'-leader sequences. *J. Gen. Virol.* 83, 927–935.
- de Vries, A.A., Chirnside, E.D., Bredenbeek, P.J., Gravestine, L.A., Horzinek, M.C., Spaan, W.J., 1990a. All subgenomic mRNAs of equine arteritis virus contain a common leader sequence. *Nucleic Acids Res.* 18, 3241–3247.
- de Vries, A.A., Chirnside, E.D., Bredenbeek, P.J., Gravestine, L.A., Horzinek, M.C., Spaan, W.J., 1990b. All subgenomic mRNAs of equine arteritis virus contain a common leader sequence. *Nucleic Acids Res.* 18, 3241–3247.
- den Boon, J.A., Kleijnen, M.F., Spaan, W.J., Snijder, E.J., 1996. Equine arteritis virus subgenomic mRNA synthesis: analysis of leader-body junctions and replicative-form RNAs. *J. Virol.* 70, 4291–4298.
- den Boon, J.A., Snijder, E.J., Chirnside, E.D., de Vries, A.A., Horzinek, M.C., Spaan, W.J., 1991. Equine arteritis virus is not a togavirus but belongs to the coronavirus like superfamily. *J. Virol.* 65, 2910–2920.
- Di, H., Madden Jr., J.C., Morantz, E.K., Tang, H.Y., Graham, R.L., Baric, R.S., Brinton, M.A., 2017. Expanded subgenomic mRNA transcriptome and coding capacity of a nidovirus. *Proc. Natl. Acad. Sci. USA* 114 (E8895–e8904).
- Firth, A.E., Zevenhoven-Dobbe, J.C., Wills, N.M., Go, Y.Y., Balasuriya, U.B.R., Atkins, J.F., Snijder, E.J., Posthuma, C.C., 2011. Discovery of a small arterivirus gene that overlaps the GP5 coding sequence and is important for virus production. *J. General Virol.* 92, 1097–1106.

- Fischer, F., Peng, D., Hingley, S.T., Weiss, S.R., Masters, P.S., 1997. The internal open reading frame within the nucleocapsid gene of mouse hepatitis virus encodes a structural protein that is not essential for viral replication. *J. Virol.* 71, 996–1003.
- Godeny, E.K., de Vries, A.A., 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.
- Hussain, S., Pan, J., Chen, Y., Yang, Y., Xu, J., Peng, Y., Wu, Y., Li, Z., Zhu, Y., Tien, P., Guo, D., 2005. Identification of novel subgenomic RNAs and noncanonical transcription initiation signals of severe acute respiratory syndrome coronavirus. *J. Virol.* 79, 5288–5295.
- Irigoyen, N., Firth, A.E., Jones, J.D., Chung, B.Y., Siddell, S.G., Brierley, I., 2016. High-resolution analysis of coronavirus gene expression by RNA sequencing and ribosome profiling. *PLoS Pathog.* 12, e1005473.
- Johnson, C.R., Griggs, T.F., Gnanandarajah, J., Murtaugh, M.P., 2011. Novel structural protein in porcine reproductive and respiratory syndrome virus encoded by an alternative ORF5 present in all arteriviruses. *J. Gen. Virol.* 92, 1107–1116.
- Joo, M., Makino, S., 1995. The effect of two closely inserted transcription consensus sequences on coronavirus transcription. *J. Virol.* 69, 272–280.
- Krishnan, R., Chang, R.Y., Brian, D.A., 1996. Tandem placement of a coronavirus promoter results in enhanced mRNA synthesis from the downstream-most initiation site. *Virology* 218, 400–405.
- La Monica, N., Yokomori, K., Lai, M.M., 1992. Coronavirus mRNA synthesis: identification of novel transcription initiation signals which are differentially regulated by different leader sequences. *Virology* 188, 402–407.
- Lai, M.M., 1986. Coronavirus leader-RNA-primed transcription: an alternative mechanism to RNA splicing. *BioEssays: News Rev. Mol., Cell. Dev. Biol.* 5, 257–260.
- Lai, M.M., Brayton, P.R., Armen, R.C., Patton, C.D., Pugh, C., Stohman, S.A., 1981. Mouse hepatitis virus A59: mRNA structure and genetic localization of the sequence divergence from hepatotropic strain MHV-3. *J. Virol.* 39, 823–834.
- Lai, M.M., Stohman, S.A., 1978. RNA of mouse hepatitis virus. *J. Virol.* 26, 236–242.
- Landry, C.R., Zhong, X., Nielly-Thibault, L., Roucou, X., 2015. Found in translation: functions and evolution of a recently discovered alternative proteome. *Curr. Opin. Struct. Biol.* 32, 74–80.
- 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.
- Liu, D.X., Fung, T.S., Chong, K.K., Shukla, A., Hilgenfeld, R., 2014. Accessory proteins of SARS-CoV and other coronaviruses. *Antivir. Res.* 109, 97–109.
- Makino, S., Soe, L.H., Shieh, C.K., Lai, M.M., 1988. Discontinuous transcription generates heterogeneity at the leader fusion sites of coronavirus mRNAs. *J. Virol.* 62, 3870–3873.
- Masters, P.S., Perlman, S., 2013. Coronaviridae. In: Knipe, D.M., Howley, P.M. (Eds.), *Fields Virology Sixth Edition*, pp. 825–858.
- Mateos-Gomez, P.A., Morales, L., Zuniga, S., Enjuanes, L., Sola, I., 2013. Long-distance RNA-RNA interactions in the coronavirus genome form high-order structures promoting discontinuous RNA synthesis during transcription. *J. Virol.* 87, 177–186.
- Meier, C., Aricescu, A.R., Assenberg, R., Aplin, R.T., Gilbert, R.J., Grimes, J.M., Stuart, D.I., 2006. The crystal structure of ORF-9b, a lipid binding protein from the SARS coronavirus. *Structure* 14, 1157–1165.
- Meng, X.J., Paul, P.S., Morozov, I., Halbur, P.G., 1996. A nested set of six or seven subgenomic mRNAs is formed in cells infected with different isolates of porcine reproductive and respiratory syndrome virus. *J. Gen. Virol.* 77 (Pt 6), 1265–1270.
- Meulenbergh, J.J., de Meijer, E.J., Moormann, R.J., 1993. Subgenomic RNAs of Lelystad virus contain a conserved leader-body junction sequence. *J. Gen. Virol.* 74 (Pt 8), 1697–1701.
- Miller, W.A., Koev, G., 2000. Synthesis of subgenomic RNAs by positive-strand RNA viruses. *Virology* 273, 1–8.
- Morales, L., Oliveros, J.C., Fernandez-Delgado, R., tenOever, B.R., Enjuanes, L., Sola, I., 2017. SARS-CoV-encoded small RNAs contribute to infection-associated lung pathology. *Cell Host Microbe* 21, 344–355.
- Moreno, J.L., Zuniga, S., Enjuanes, L., Sola, I., 2008. Identification of a coronavirus transcription enhancer. *J. Virol.* 82, 3882–3893.
- Nedialkova, D.D., Gorbalenya, A.E., Snijder, E., 2010. Arterivirus nsp1 modulates the accumulation of minus-strand templates to control the relative abundance of viral mRNAs. *PLOS Pathog.* 6, 31000772.
- Nelsen, C.J., Murtaugh, M.P., Faaberg, K.S., 1999. Porcine reproductive and respiratory syndrome virus comparison: divergent evolution on two continents. *J. Virol.* 73, 270–280.
- O'Connor, J.B., Brian, D.A., 2000. Downstream ribosomal entry for translation of coronavirus TGEV gene 3b. *Virology* 269, 172–182.
- Olasz, F., Denes, B., Balint, A., Magyar, T., Belak, S., Zadori, Z., 2016. Immunological and biochemical characterisation of 7ap, a short protein translated from an alternative frame of ORF7 of PRRSV. *Acta Vet. Hung.* 64, 273–287.
- Ozdarendeli, A., Ku, S., Rochat, S., Williams, G.D., Senanayake, S.D., Brian, D.A., 2001. Downstream sequences influence the choice between a naturally occurring non-canonical and closely positioned upstream canonical heptameric fusion motif during bovine coronavirus subgenomic mRNA synthesis. *J. Virol.* 75, 7362–7374.
- Pasternak, A.O., Gultyaev, A.P., Spaan, W.J., 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., Snijder, E.J., 2004. Regulation of relative abundance of arterivirus subgenomic mRNAs. *J. Virol.* 78, 8102–8113.
- Pasternak, A.O., Spaan, W.J., Snijder, E.J., 2006. Nidovirus transcription: how to make sense...? *J. Gen. Virol.* 87, 1403–1421.
- Pasternak, A.O., van den Born, E., Spaan, W.J., Snijder, E.J., 2001. Sequence requirements for RNA strand transfer during nidovirus discontinuous subgenomic RNA synthesis. *EMBO J.* 20, 7220–7228.
- Pasternak, A.O., van den Born, E., Spaan, W.J., Snijder, E.J., 2003. The stability of the duplex between sense and antisense transcription-regulating sequences is a crucial factor in arterivirus subgenomic mRNA synthesis. *J. Virol.* 77, 1175–1183.
- Pueyo, J.I., Magny, E.G., Couso, J.P., 2016. New peptides under the s(ORF)ace of the genome. *Trends Biochem. Sci.* 41, 665–678.
- Sawicki, S.G., Sawicki, D.L., 1990. Coronavirus transcription: subgenomic mouse hepatitis virus replicative intermediates function in RNA synthesis. *J. Virol.* 64, 1050–1056.
- Schaad, M.C., Baric, R.S., 1993. Evidence for new transcriptional units encoded at the 3' end of the mouse hepatitis virus genome. *Virology* 196, 190–198.
- Schelle, B., Karl, N., Ludewig, B., Siddell, S.G., Thiel, V., 2005. Selective replication of coronavirus genomes that express nucleocapsid protein. *J. Virol.* 79, 6620–6630.
- Sethna, P.B., Hung, S.L., Brian, D.A., 1989. Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons. *Proc. Natl. Acad. Sci. USA* 86, 5626–5630.
- Smith, S.L., Wang, X., Godeny, E.K., 1997. Sequence of the 3' end of the simian hemorrhagic fever virus genome. *Gene* 191, 205–210.
- Smits, S.L., van Vliet, A.L., Segeren, K., el Azzouzi, H., van Essen, M., de Groot, R.J., 2005. Torovirus non-discontinuous transcription: mutational analysis of a subgenomic mRNA promoter. *J. Virol.* 79, 8275–8281.
- Snijder, E.J., Kikkert, M., Fang, Y., 2013. Arterivirus molecular biology and pathogenesis. *J. Gen. Virol.* 94, 2141–2163.
- Snijder, E.J., van Tol, H., Pedersen, K.W., Raamsman, M.J., de Vries, A.A., 1999. Identification of a novel structural protein of arteriviruses. *J. Virol.* 73, 6335–6345.
- Sola, I., Moreno, J.L., Zuniga, S., Alonso, S., Enjuanes, L., 2005. Role of nucleotides immediately flanking the transcription-regulating sequence core in coronavirus subgenomic mRNA synthesis. *J. Virol.* 79, 2506–2516.
- Thiel, V., Siddell, S.G., 1994. Internal ribosome entry in the coding region of murine hepatitis virus mRNA 5. *J. Gen. Virol.* 75 (Pt 11), 3041–3046.
- Tijms, M.A., Nedialkova, D.D., Zevenhoven-Dobbe, J.C., Gorbalenya, A.E., Snijder, E.J., 2007. Arterivirus subgenomic mRNA synthesis and virion biogenesis depend on the multifunctional nsp1 autoprotease. *J. Virol.* 81, 10496–10505.
- van Berlo, M.F., Horzinek, M.C., van der Zeijst, B.A., 1982. Equine arteritis virus-infected cells contain six polyadenylated virus-specific RNAs. *Virology* 118, 345–352.
- van Berlo, M.F., Rottier, P.J., Horzinek, M.C., van der Zeijst, B.A., 1986. Intracellular equine arteritis virus (EAV)-specific RNAs contain common sequences. *Virology* 152, 492–496.
- Van Den Born, E., Gultyaev, A.P., Snijder, E.J., 2004. Secondary structure and function of the 5'-proximal region of the equine arteritis virus RNA genome. *RNA* 10, 424–437 (New York, N.Y.).
- van den Born, E., Posthuma, C.C., Gultyaev, A.P., Snijder, E.J., 2005. Discontinuous subgenomic RNA synthesis in arteriviruses is guided by an RNA hairpin structure located in the genomic leader region. *J. Virol.* 79, 6312–6324.
- van Marle, G., Dobbe, J.C., Gultyaev, A.P., Luytjes, W., Spaan, W.J., Snijder, E.J., 1999a. Arterivirus discontinuous mRNA transcription is guided by base pairing between sense and antisense transcription-regulating sequences. *Proc. Natl. Acad. Sci. USA* 96, 12056–12061.
- van Marle, G., van Dinten, L.C., Spaan, W.J., Luytjes, W., Snijder, E.J., 1999b. Characterization of an equine arteritis virus replicase mutant defective in subgenomic mRNA synthesis. *J. Virol.* 73, 5274–5281.
- van Vliet, A.L., Smits, S.L., Rottier, P.J., de Groot, R.J., 2002. Discontinuous and non-discontinuous subgenomic RNA transcription in a nidovirus. *EMBO J.* 21, 6571–6580.
- Vatter, H.A., Di, H., Donaldson, E.F., Baric, R.S., Brinton, M.A., 2014. Each of the eight simian hemorrhagic fever virus minor structural proteins is functionally important. *Virology* 462–463C, 351–362.
- Wang, Y., Zhang, X., 2000. The leader RNA of coronavirus mouse hepatitis virus contains an enhancer-like element for subgenomic mRNA transcription. *J. Virol.* 74, 10571–10580.
- Wu, H.Y., Brian, D.A., 2010. Subgenomic messenger RNA amplification in coronaviruses. *Proc. Natl. Acad. Sci. USA* 107, 12257–12262.
- Yuan, S., Murtaugh, M.P., Faaberg, K.S., 2000. Heteroclitte subgenomic RNAs are produced in porcine reproductive and respiratory syndrome virus infection. *Virology* 275, 158–169.
- Yuan, S., Murtaugh, M.P., Schumann, F.A., Mickelson, D., Faaberg, K.S., 2004. Characterization of heteroclitte subgenomic RNAs associated with PRRSV infection. *Virus Res.* 105, 75–87.
- Zeng, L., Godeny, E.K., Methven, S.L., Brinton, M.A., 1995. Analysis of simian hemorrhagic fever virus (SHFV) subgenomic RNAs, junction sequences, and 5' leader. *Virology* 207, 543–548.
- Zhang, X., Liu, R., 2000. Identification of a noncanonical signal for transcription of a novel subgenomic mRNA of mouse hepatitis virus: implication for the mechanism of coronavirus RNA transcription. *Virology* 278, 75–85.
- Zhao, X., Shaw, K., Cavanagh, D., 1993. Presence of subgenomic mRNAs in virions of coronavirus IBV. *Virology* 196, 172–178.
- Zirkel, F., Roth, H., Kurth, A., Drosten, C., Ziebuhr, J., Junglen, S., 2013. Identification and characterization of genetically divergent members of the newly established family Mesoniviridae. *J. Virol.* 87, 6346–6358.
- Zuniga, S., Sola, I., Alonso, S., Enjuanes, L., 2004. Sequence motifs involved in the regulation of discontinuous coronavirus subgenomic RNA synthesis. *J. Virol.* 78, 980–994.