diseases

MDPI

*Review*

# Distribution and Evolutionary History of the Mobile Genetic Element s2m in Coronaviruses

**Torstein Tengs [1,*] and Christine M. Jonassen [2]**

[1] Norwegian Veterinary Institute, Ullevaalsveien 68, 0454 Oslo, Norway
[2] Centre for Laboratory Medicine, Østfold Hospital Trust, Kalnesveien 300, 1714 Grålum, Norway; chrjon@so-hf.no
[*] Correspondence: torstein.tengs@vetinst.no; Tel.: +47-23-21-63-76 or +47-47-31-48-50; Fax: +47-23-21-62-02

Academic Editor: Ding Xiang Liu
Received: 31 May 2016; Accepted: 25 July 2016; Published: 28 July 2016

**Abstract:** The mobile genetic element s2m has been described in several families of single-stranded RNA viruses. The function remains elusive, but an increasing number of s2m-containing sequences are being deposited in publicly available databases. Currently, more than 700 coronavirus sequences containing s2m can be found in GenBank, including the severe acute respiratory syndrome (SARS) coronavirus genome. This is an updated review of the pattern of s2m in coronaviruses, the possible functional implications and the evolutionary history.

**Keywords:** s2m; coronavirus; secondary structure; mobile genetic element

## 1. Introduction

The genomes of single-stranded RNA (ssRNA) viruses have complex secondary structures, maintained both through conventional and non-Watson-Crick base-pairing [1,2]. Coronaviruses have the largest genomes of RNA viruses and have been shown to have several highly conserved secondary structures both in the 5′ and 3′ end of their genomes [1]. These structural motifs mainly function through protein binding or through direct RNA-RNA interactions. Some of these elements function merely as 'spacer elements', where neither the primary nor the secondary structure are critical [3]. In other cases, the stem-loop structure itself seems essential and base-pairing nucleotides may be substituted [4]. More complex structures, such as the 54-nucleotide hairpin-type pseudoknot, appear to be conserved both in structure and location, but not in sequence [5]. For most of these elements, the exact mechanism of function is unknown, but virus replication, viability and transcription are the main functional categories.

A 43-nucleotide genetic element with a highly conserved secondary (as well as primary and tertiary) structure has been described in four different families of positive-sense ssRNA viruses, *Astroviridae*, *Caliciviridae*, *Picornaviridae* and *Coronaviridae* [6]. The presence of this element in four distantly related groups of viruses and the scattered distribution within these groups has led to the conclusion that viruses have the ability to acquire it horizontally [6]. The 3′ stem-loop structure, referred to as s2m, was originally described in astroviruses where it appears to have been present in the last common ancestor [6–8], but the element has been most thoroughly studied in the SARS coronavirus (SARS-CoV) where the three-dimensional crystal structure has been resolved to 2.7 A resolution [9]. The function remains obscure, but current hypotheses include hijacking of host protein synthesis through interactions with ribosomal proteins [9] and RNA interference (RNAi) via processing of the s2m elements into a mature microRNA [6]. In coronaviruses, the highly conserved nature of the element has also allowed the development of a PCR-based virus discovery strategy [10].

The presence of s2m near the 3′ end of some coronavirus genomes has been documented previously [6,11], but due to the exponential growth of gene sequence data available through public databases, this review is an update on the current status of s2m in this virus family. The evolutionary history is discussed and different hypotheses for the possible function of s2m are presented in light of recent progress in virus genomics.

## 2. Coronavirus Phylogeny and Distribution of s2m

A total of 20,068 coronavirus nucleotide sequences have been deposited in the current version of Genbank, and when translated into amino acid sequences, the complete ORF1ab polyprotein sequence could be found for 1113 of these entries. A phylogenetic analysis was performed using aligned ORF1ab sequences (Figure 1), and numerous clusters of nearly identical accessions could be identified. The largest group comprised 328 sequences from the porcine epidemic diarrhea virus (328 OTUs). Many smaller clusters were also found, encompassing two to 183 sequences, and all serogroups were monophyletic with 100% bootstrap support (Figure 1).

To get an overview of the current status for s2m in coronaviruses, all coronavirus nucleotide sequences were screened for the presence of the s2m motif using the strategy described by Tengs et al. [6]. The consensus sequence CGNGG(N)CCACGNNGNGT(N)ANNANCGAGGGT(N)ACAG was used as bait, allowing for possible insertions (N) and/or a single mismatch. Further, 708 of 20,068 (3.5%) of the sequences were found to contain the s2m motif (Table 1). However, s2m was not found in alpha coronaviruses and also appears to be absent in the *Torovirus* genus as well as the recently described *Bafinivirus* genus [12], albeit only a small number of sequences were available from the two latter groups. In all instances, the s2m sequence was found in the non-coding, 3′ end of the genome. Looking at coronavirus accessions annotated as 'complete genome', more than a third of the genomes appeared to be s2m-containing (Table 1). Many sequences have been generated using targeted (PCR-based) approaches that might not include the non-coding parts of the genome. This could explain why s2m appears to be approximately 10 times more common in complete genome sequences than in shorter GenBank submissions (Table 1).

**Table 1.** Coronavirus sequences in GenBank.

| | **Total** | **Containing s2m** |
|---|---|---|
| Coronavirus sequences | 20,068 | 706 (3.5%) |
| Alpha coronavirus * | 7190 | 0 |
| Beta coronavirus * | 4947 | 342 (6.9%) |
| Delta coronavirus * | 141 | 60 (42.6%) |
| Gamma coronavirus * | 6360 | 281 (4.4%) |
| Bafinivirus * | 12 | 0 |
| Torovirus * | 307 | 0 |
| Complete genomes | 1507 | 523 (34.7%) |

* GenBank taxonomy database annotation.

Three phylogenetic clusters were found to contain s2m (Figure 1). All of these groups were supported by 100% bootstrap values and thus are genetically distinct from related viruses not containing the element. Multiple hypothesis can be proposed in order to explain the distribution of s2m within the coronaviruses, including a model where s2m was present in the last common ancestor of this virus group and subsequently lost multiple times. However, given the mobile nature of the motif, a more parsimonious explanation includes two (or more) independent gains and a small number of losses. The gain of s2m at the base of the SARS-cluster (edge 1 in Figure 1) would not assume any losses in the beta coronaviruses, but for the gamma/delta coronavirus cluster, the evolutionary history seems to be more complicated. A gain at edge 4 would require multiple independent losses. There was no significant bootstrap support separating the Wigeon coronavirus and the Night-heron coronavirus sequences, so a monophyletic origin for these two strains could not be excluded. Loss in the common ancestor of this group combined with loss prior to the divergence of the Bottlenose dolphin coronavirus and the Beluga Whale coronavirus (edge 3) would thus indicate one gain and two independent losses. One could also hypothesize two independent gains (edge 2 and 5) in order to explain the s2m distribution in the gamma/delta cluster (Figure 1).

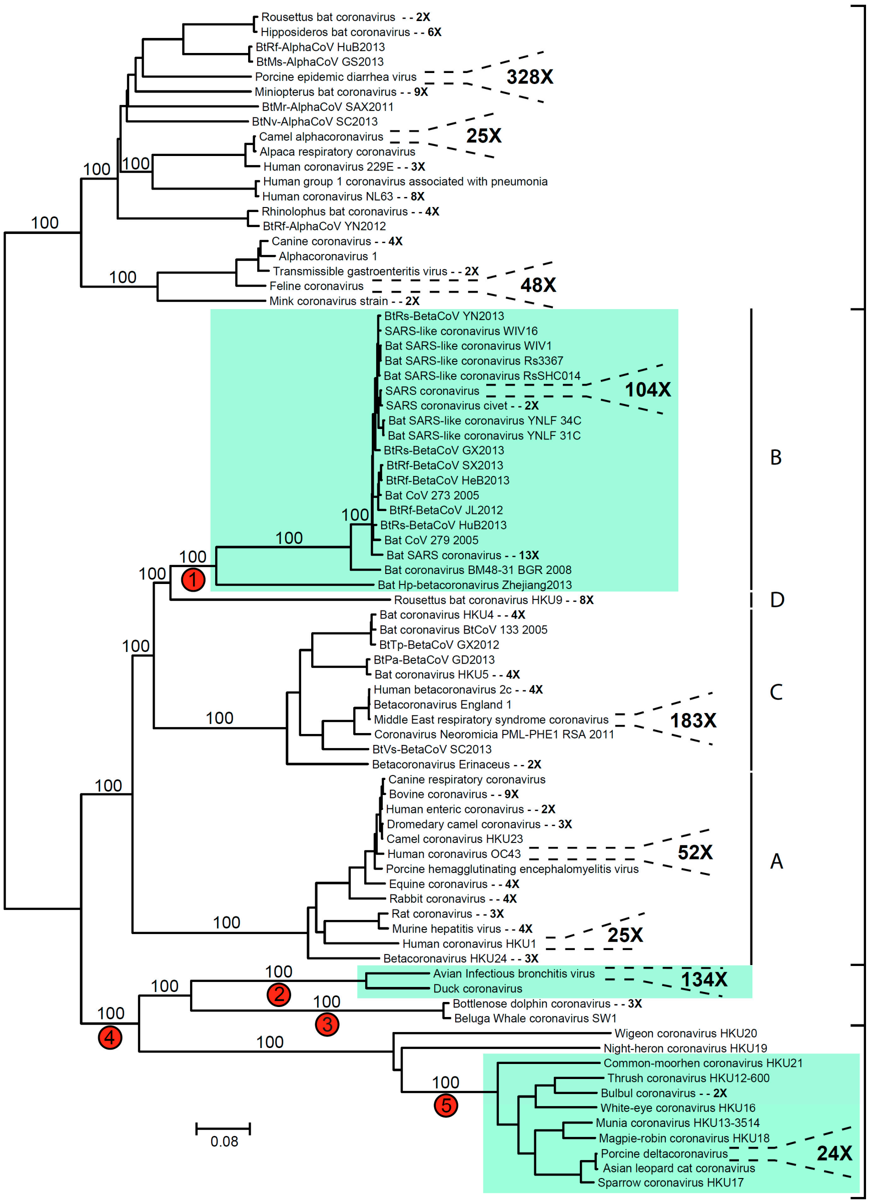
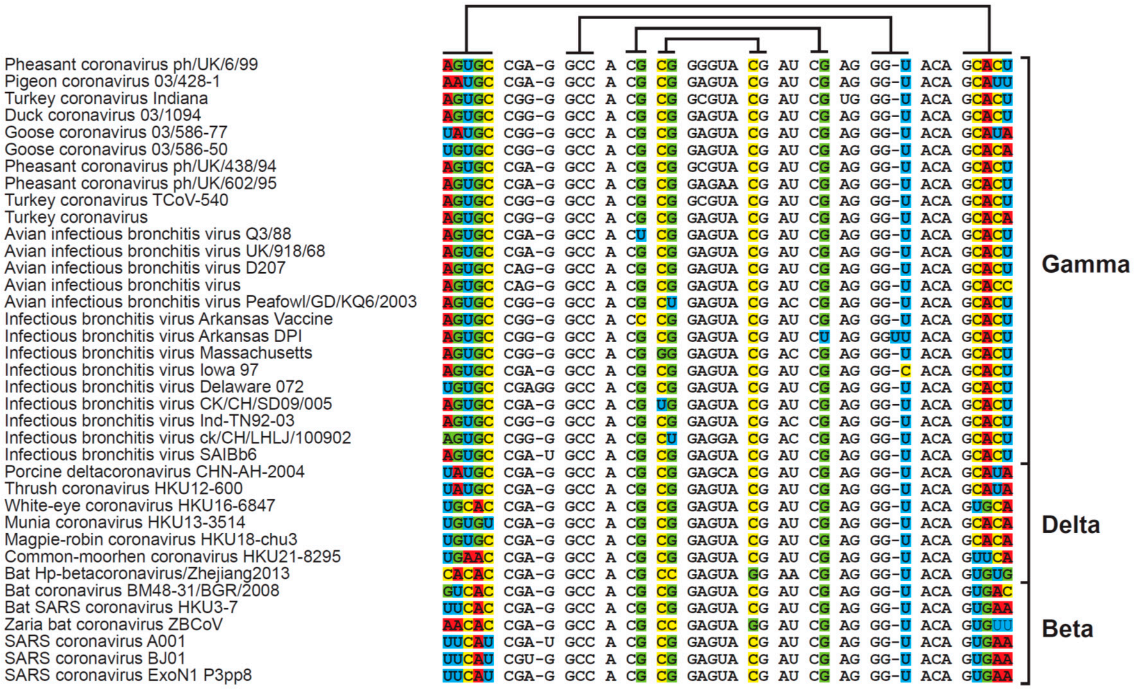

Rousettus bat coronavirus - - 2X
Hipposideros bat coronavirus - - 6X
BtRf-AlphaCoV HuB2013
BtMs-AlphaCoV GS2013
Porcine epidemic diarrhea virus
328X
Miniopterus bat coronavirus - - 9X
BtMr-AlphaCoV SAX2011
BtNv-AlphaCoV SC2013
Camel alphacoronavirus
25X
Alpaca respiratory coronavirus
Human coronavirus 229E - - 3X
Human group 1 coronavirus associated with pneumonia
Human coronavirus NL63 - - 8X
Rhinolophus bat coronavirus - - 4X
BtRf-AlphaCoV YN2012
Canine coronavirus - - 4X
Alphacoronavirus 1
Transmissible gastroenteritis virus - - 2X
Feline coronavirus
48X
Mink coronavirus strain - - 2X
BtRs-BetaCoV YN2013
SARS-like coronavirus WIV16
Bat SARS-like coronavirus WIV1
Bat SARS-like coronavirus Rs3367
Bat SARS-like coronavirus RsSHC014
SARS coronavirus
104X
SARS coronavirus civet - - 2X
Bat SARS-like coronavirus YNLF 34C
Bat SARS-like coronavirus YNLF 31C
BtRs-BetaCoV GX2013
BtRf-BetaCoV SX2013
BtRf-BetaCoV HeB2013
Bat CoV 273 2005
BtRf-BetaCoV JL2012
BtRs-BetaCoV HuB2013
Bat CoV 279 2005
Bat SARS coronavirus - - 13X
Bat coronavirus BM48-31 BGR 2008
Bat Hp-betacoronavirus Zhejiang2013
Rousettus bat coronavirus HKU9 - - 8X
Bat coronavirus HKU4 - - 4X
Bat coronavirus BtCoV 133 2005
BtTp-BetaCoV GX2012
BtPa-BetaCoV GD2013
Bat coronavirus HKU5 - - 4X
Human betacoronavirus 2c - - 4X
Betacoronavirus England 1
Middle East respiratory syndrome coronavirus
183X
Coronavirus Neoromicia PML-PHE1 RSA 2011
BtVs-BetaCoV SC2013
Betacoronavirus Erinaceus - - 2X
Canine respiratory coronavirus
Bovine coronavirus - - 9X
Human enteric coronavirus - - 2X
Dromedary camel coronavirus - - 3X
Camel coronavirus HKU23
Human coronavirus OC43
52X
Porcine hemagglutinating encephalomyelitis virus
Equine coronavirus - - 4X
Rabbit coronavirus - - 4X
Rat coronavirus - - 3X
Murine hepatitis virus - - 4X
Human coronavirus HKU1
25X
Betacoronavirus HKU24 - - 3X
Avian Infectious bronchitis virus
134X
Duck coronavirus
Bottlenose dolphin coronavirus - - 3X
Beluga Whale coronavirus SW1
Wigeon coronavirus HKU20
Night-heron coronavirus HKU19
Common-moorhen coronavirus HKU21
Thrush coronavirus HKU12-600
Bulbul coronavirus - - 2X
White-eye coronavirus HKU16
Munia coronavirus HKU13-3514
Magpie-robin coronavirus HKU18
Porcine deltacoronavirus
24X
Asian leopard cat coronavirus
Sparrow coronavirus HKU17
100
1
2
3
4
5
0.08
Alpha
Beta
Gamma
Delta
B
D
C
A

**Figure 1.** Coronavirus phylogeny with s2m-containing sequences highlighted. ORF1ab polyprotein amino acid sequences were aligned using the program MUSCLE [8] and default parameters. The phylogenetic analysis was performed using the program SeaView [9] and the neighbor joining clustering method with Kimura two-parameter distances. In order to avoid large clades of closely related sequences, operational taxonomic units (OTUs) with similar GenBank taxonomical annotation and almost identical sequences were identified and basal members of these monophyletic groups chosen to represent such sequence clusters. For instance, there are 183 complete ORF1ab polyprotein sequences available from different strains of the Middle East respiratory syndrome (MERS) coronavirus. These sequences are represented by a single accession (in this case, GenBank accession number ALB08298; isolate KOREA/Seoul/035-1-2015). Based on visual inspection of the alignment it was determined that the sequences belonging to the Torovirinae subfamily could not be reliably aligned and were excluded from the analysis. Brackets show serogroups as well as betacoronavirus lineages and key branches with 100% bootstrap support (100 pseudoreplicates) have been indicated. Red circles indicate possible losses/gains of s2m (see discussion in text).

Looking at topology and branch lengths, branch 1 (Figure 1) represents the shortest edge connecting an s2m-containing cluster with a virus strain not containing s2m. s2m is absent in Rousettus bat coronavirus HKU9, which is a cluster of eight nearly identical sequences basal to the SARS(-like) group. Rousettus bat coronavirus is thus far the only member of lineage D within the betacoronaviruses [13]. The most likely mechanism for transfer of the s2m element from one viral genome to another is through genetic recombination [6]. Recombination events are very common in single-stranded RNA viruses and have played an important role in the evolution of coronaviruses [14]. Nucleotide dot plot analyses of the Rousettus bat coronavirus versus the closest relatives in the SARS(-like) group revealed a high degree of sequence similarity in central parts of the genome, but did not identify any specific locus for recombination.

## 3. Sequence and Secondary Structure of s2m

Excluding sequences with ambiguous bases, a total of 37 different s2m genotypes were found within the coronavirus sequences (Figure 2). The great majority of the sequences could be folded into the canonical s2m stem-loop structure (Figure 3), albeit a single sequence had an insertion in the second stem-forming motif (Figure 1; Infectious bronchitis virus DPI) and some of the sequences require non-Watson-Crick base-pairing, such as the wobble base-pair G-U [15], in order to obtain the correct secondary structure. All clusters of identical s2m sequence groups were derived from viruses representing the same serogroups (beta, gamma or delta), indicating no recent transfers of the s2m element between these coronavirus clusters.



| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pheasant coronavirus ph/UK/6/99 | AGUGC | CGA-G | GCC | A | CG | CG | GGGUA | CG | AU | CG | AG | GG-U | ACA | GCACU |
| Pigeon coronavirus 03/428-1 | AAUGC | CGA-G | GCC | A | CG | CG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GCAUU |
| Turkey coronavirus Indiana | AGUGC | CGG-G | GCC | A | CG | CG | GCGUA | CG | AU | CG | UG | GG-U | ACA | GCACU |
| Duck coronavirus 03/1094 | AGUGC | CGG-G | GCC | A | CG | CG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GCACU |
| Goose coronavirus 03/586-77 | UAUGC | CGG-G | GCC | A | CG | CG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GCAUA |
| Goose coronavirus 03/586-50 | UGUGC | CGG-G | GCC | A | CG | CG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GCACA |
| Pheasant coronavirus ph/UK/438/94 | AGUGC | CGA-G | GCC | A | CG | CG | GCGUA | CG | AU | CG | AG | GG-U | ACA | GCACU |
| Pheasant coronavirus ph/UK/602/95 | AGUGC | CGA-G | GCC | A | CG | CG | GAGAA | CG | AU | CG | AG | GG-U | ACA | GCACU |
| Turkey coronavirus TCoV-540 | AGUGC | CGG-G | GCC | A | CG | CG | GCGUA | CG | AU | CG | AG | GG-U | ACA | GCACU |
| Turkey coronavirus | AGUGC | CGG-G | GCC | A | CG | CG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GCACA |
| Avian infectious bronchitis virus Q3/88 | AGUGC | CGA-G | GCC | A | CU | CG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GCACU |
| Avian infectious bronchitis virus UK/918/68 | AGUGC | CGA-G | GCC | A | CG | CG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GCACU |
| Avian infectious bronchitis virus D207 | AGUGC | CAG-G | GCC | A | CG | CG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GCACU |
| Avian infectious bronchitis virus | AGUGC | CAG-G | GCC | A | CG | CG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GCACC |
| Avian infectious bronchitis virus Peafowl/GD/KQ6/2003 | AGUGC | CGG-G | GCC | A | CG | CU | GAGUA | CG | AC | CG | AG | GG-U | ACA | GCACU |
| Infectious bronchitis virus Arkansas Vaccine | AGUGC | CGG-G | GCC | A | CC | CG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GCACU |
| Infectious bronchitis virus Arkansas DPI | AGUGC | CGG-G | GCC | A | CG | CG | GAGUA | CG | AU | CU | AG | GGUU | ACA | GCACU |
| Infectious bronchitis virus Massachusetts | AGUGC | CGG-G | GCC | A | CG | GG | GAGUA | CG | AC | CG | AG | GG-U | ACA | GCACU |
| Infectious bronchitis virus Iowa 97 | AGUGC | CGA-G | GCC | A | CG | CG | GAGUA | CG | AU | CG | AG | GG-C | ACA | GCACU |
| Infectious bronchitis virus Delaware 072 | UGUGC | CGAGG | GCC | A | CG | CG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GCACU |
| Infectious bronchitis virus CK/CH/SD09/005 | AGUGC | CGA-G | GCC | A | CG | UG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GCACU |
| Infectious bronchitis virus Ind-TN92-03 | AGUGC | CGG-G | GCC | A | CG | CG | GAGUA | CG | AC | CG | AG | GG-U | ACA | GCACU |
| Infectious bronchitis virus ck/CH/LHLJ/100902 | AGUGC | CGG-G | GCC | A | CG | CU | GAGGA | CG | AC | CG | AG | GG-U | ACA | GCACU |
| Infectious bronchitis virus SAIBb6 | AGUGC | CGA-U | GCC | A | CG | CG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GCACU |
| Porcine deltacoronavirus CHN-AH-2004 | UAUGC | CGA-G | GCC | A | CG | CG | GAGCA | CG | AU | CG | AG | GG-U | ACA | GCAUA |
| Thrush coronavirus HKU12-600 | UAUGC | CGA-G | GCC | A | CG | CG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GCAUA |
| White-eye coronavirus HKU16-6847 | UGCAC | CGA-G | GCC | A | CG | CG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GUGCA |
| Munia coronavirus HKU13-3514 | UGUGU | CGA-G | GCC | A | CG | CG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GCACA |
| Magpie-robin coronavirus HKU18-chu3 | UGUGC | CGA-G | GCC | A | CG | CG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GCACA |
| Common-moorhen coronavirus HKU21-8295 | UGAAC | CGA-G | GCC | A | CG | CG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GUUCA |
| Bat Hp-betacoronavirus/Zhejiang2013 | CACAC | CGA-G | GCC | A | CG | CC | GAGUA | GG | AA | CG | AG | GG-U | ACA | GUGUG |
| Bat coronavirus BM48-31/BGR/2008 | GUCAC | CGA-G | GCC | A | CG | CG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GUGAC |
| Bat SARS coronavirus HKU3-7 | UUCAC | CGA-G | GCC | A | CG | CG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GUGAA |
| Zaria bat coronavirus ZBCoV | AACAC | CGA-G | GCC | A | CG | CC | GAGUA | GG | AU | CG | AG | GG-U | ACA | GUGUU |
| SARS coronavirus A001 | UUCAU | CGA-U | GCC | A | CG | CG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GUGAA |
| SARS coronavirus BJ01 | UUCAU | CGU-G | GCC | A | CG | CG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GUGAA |
| SARS coronavirus ExoN1 P3pp8 | UUCAU | CGA-G | GCC | A | CG | CG | GAGUA | CG | AU | CG | AG | GG-U | ACA | GUGAA |

**Figure 2.** s2m coronavirus sequence motifs. For each genotype, one representative accession was chosen to illustrate the conserved nature of s2m both on a primary ('sequence') level and on a secondary (stem-loop structure) level. Lines above alignment indicate co-varying/stem-forming elements and columns with non-conserved bases for these nucleotide positions have been color coded.

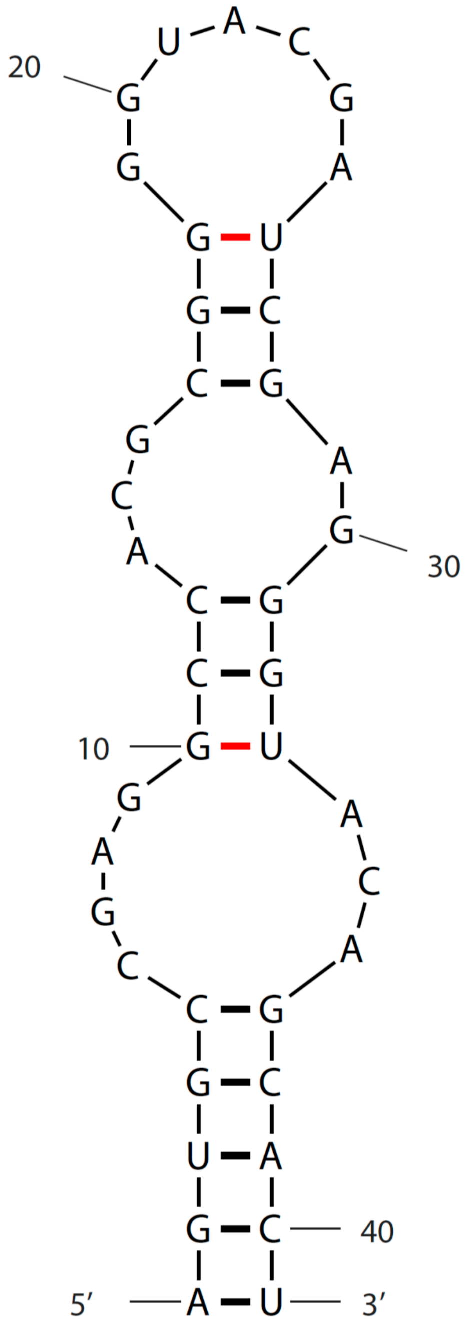


**Figure 3.** s2m secondary structure. The s2m element from Pheasant coronavirus strain ph/UK/6/99 was folded using mfold [16]. Non-Watson-Crick base-pairings are shown in red. For a more detailed folding with tertiary interactions and long-range contacts indicated, see [9].

## 4. Function of s2m

For a mobile genetic element to be successful, especially in the context of rapidly evolving genomes such as those of RNA viruses, it must offer an immediate selective advantage. s2m viruses infect higher vertebrates, making the permissive host species a lot more closely related than the viruses. The observation that s2m can apparently be transferred between unrelated viruses and remain functional (under selection to maintain sequence and structure) suggests strongly that the target for s2m is host-specific and not viral. As there are no conserved elements flanking the s2m motif in viral genomes, it is also plausible that the element has an 'autonomous' function, independent of other viral genes, transcripts or genome secondary/tertiary structures. Somewhat surprisingly, the acquisition of s2m only has a subtle effect; exchanging the (non-s2m-containing) 3′-end of a murine coronavirus (MCV) with the 3′-end of a SARS virus did not appear to have any dramatic consequences [17], but 3′-ends from other coronaviruses were not able to replace the 3′-end of MCV and render it viable. There appears to be only a single sequence in GenBank with a deleted version of s2m (Infectious bronchitis virus strain ck/CH/LHLJ/07VII; accession number JF274479). This strain was discovered as an escape mutant in a vaccine development project, but did not appear phenotypically different from closely related viruses in culturing experiments [18].

The folding of s2m is quite similar to the hairpin structures formed by microRNAs that are involved in RNAi-associated gene regulation, and s2m could potentially be involved in gene silencing. In vertebrates, pre-microRNA transcripts are generated in the nucleus and processed by the nuclear protein DGCR8 and the enzyme Drosha [19]. Processed pre-microRNAs are then exported from the nucleus in a process involving Exportin-5. The RNA is subsequently incorporated into the RNA-induced silencing complex (RISC) and eventually a functional RISC-bound small interfering RNA (siRNA) can serve as template for base-pairing recognition of messenger RNA (mRNA). The target mRNA is degraded, leading to a reduction in gene activity. ssRNA viruses generally replicate in the cytoplasm and earlier hypotheses on s2m function have thus been based on the assumption that the cellular components involved must be available in the cytosol. Recently, a paper was published documenting the presence of a non-canonical machinery for microRNA processing in the cytoplasm of human cells [20]. It has been well described that DNA viruses replicating in the nucleus can hijack the cellular machinery for RNAi activity for the regulation of both viral and host genes [21], but these new results using a recombinant Sindbis virus show that this is also possible for RNA viruses. The presence of multiple copies of s2m in some virus genomes [6] can be explained by there being an additive effect: more copies of the RNA will enable the formation of more siRNA/RISC complexes and give a more profound effect on target genes. A hypothesis where s2m is involved in RNAi-based gene regulation is also compatible with the target being host-specific (homologous genes in the infected species) and s2m functionality being independent of the rest of the virus genome.

Observations arguing against an RNAi-like function would be that the stem region of s2m is a bit short compared to standard processed RNAi molecules (albeit mRNA binding regions as short as 17 base pairs have been reported [22]). We were also unable to find any good candidate target genes using sequence similarity searches to look for potential microRNA binding sites, and even when looking at viruses infecting relatively closely related species (such as bats, see Figure 2), there was significant variability in the s2m stem regions that could be predicted to be involved in mRNA binding. These observations make it less likely that s2m is involved in conventional RNAi-based gene silencing.

Another hypothesis that would be compatible with the general assumptions regarding s2m functionality is that s2m is involved in the protection of the virus genome from being degraded by host ribonucleases. Coronaviruses are positive-sense ssRNA viruses with non-segmented genomes that are polyadenylated and capped. The genomes may serve as mRNA for the translation of viral polyproteins and contain two overlapping reading frames that encode precursor polyproteins pp1a and pp1ab through a frameshifting mechanism [23]. After infection, the viral genome is replicated and subgenomic RNAs are generated via negative-strand RNA intermediates [1,24]. In flaviviruses, another group of positive-sense ssRNA viruses, the presence of secondary structure elements has been shown to inhibit RNA decay through the formation of nuclease-resistant noncoding RNAs [25,26]. The small stem-loop structures described confer resistance to enzymatic degradation by the XRN1 5′-3′ exoribonuclease. As the s2m element is near the 3′ end of the coronavirus genome, inhibition of, for instance, XRN1 would only render the negative-strand versions of the genome (or subgenomic elements originating from the 3′ flank) protected from decay, but resistance could also be gained through other mechanisms. Full-length coronavirus messenger RNAs share many features with endogenous transcripts and any interference with mechanisms involved in cellular mRNA turnover could also affect virus RNA stability. The exosome contains multiple proteins involved in 3′-5′

degradation of RNA [27], and inhibition or stalling of, for instance, the Dis3 or Rrp6 polypeptides by s2m could also protect full-length sense copies of the genome.

For some positive-strand ssRNA viruses, genome circularization has been shown to be essential for virus replication [28]. RNA circularization would require s2m to bind either directly or indirectly (mediated by RNA binding proteins) to the 5′ end of the genome. The primary binding 5′ site could either be a five-prime cap (5′ cap) or an RNA structure (see discussion in [29]). It is possible that there are conserved structural RNA elements with such a binding affinity near the 5′ end in all s2m viruses, but it seems unlikely given the complete lack of sequence similarity when comparing data from the 5′ flank of these genomes. The direct involvement of a 5′ cap also seems implausible, as, for instance, s2m-containing picornaviruses lack 5′ capping.

In order to elucidate the function of s2m, it seems imperative to establish a reverse genetics system for an s2m-containing viral strain. A good system could be an astrovirus-derived infectious construct, as there already are protocols described for the design of astrovirus cDNA clones that lead to the production of infectious viral particles [30]. Targeted mutations may be introduced and comparative studies performed using different permissive cell lines. Any changes in transcriptional activity, virus replication rates and/or viral RNA stability can be correlated with mutations interfering with the structure of s2m. Though s2m is present in some coronaviruses associated with severe infections in humans, such as SARS-CoV, there are other highly pathogenic viruses where s2m is absent (for instance the Middle East respiratory syndrome coronavirus; MERS-CoV). The significance of having s2m for a viral strain remains to be resolved, but this mobile element still seems to be in play and it is likely to be found in newly emerging coronaviruses in the future.

**Acknowledgments:** This work was funded internally by the Norwegian Veterinary Institute.

**Author Contributions:** T.T. performed the sequence analyses and C.M.J. helped write the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## References

1. Yang, D.; Leibowitz, J.L. The structure and functions of coronavirus genomic 3′ and 5′ ends. *Virus Res.* **2015**, *206*, 120–133. [CrossRef] [PubMed]
2. Nicholson, B.L.; White, K.A. Exploring the architecture of viral RNA genomes. *Curr. Opin. Virol.* **2015**, *12*, 66–74. [CrossRef] [PubMed]
3. Yang, D.; Liu, P.; Giedroc, D.P.; Leibowitz, J. Mouse hepatitis virus stem-loop 4 functions as a spacer element required to drive subgenomic RNA synthesis. *J. Virol.* **2011**, *85*, 9199–9209. [CrossRef] [PubMed]
4. Liu, P.; Li, L.; Millership, J.J.; Kang, H.; Leibowitz, J.L.; Giedroc, D.P. A U-turn motif-containing stem-loop in the coronavirus 5′ untranslated region plays a functional role in replication. *RNA* **2007**, *13*, 763–780. [CrossRef] [PubMed]
5. Williams, G.D.; Chang, R.Y.; Brian, D.A. A phylogenetically conserved hairpin-type 3′ untranslated region pseudoknot functions in coronavirus RNA replication. *J. Virol.* **1999**, *73*, 8349–8355. [PubMed]
6. Tengs, T.; Kristoffersen, A.B.; Bachvaroff, T.R.; Jonassen, C.M. A mobile genetic element with unknown function found in distantly related viruses. *Virol. J.* **2013**, *10*, 132. [CrossRef] [PubMed]
7. Monceyron, C.; Grinde, B.; Jonassen, T.O. Molecular characterisation of the 3′-end of the astrovirus genome. *Arch. Virol.* **1997**, *142*, 699–706. [CrossRef] [PubMed]
8. Jonassen, T.O.; Monceyron, C.; Lee, T.W.; Kurtz, J.B.; Grinde, B. Detection of all serotypes of human astrovirus by the polymerase chain reaction. *J. Virol. Methods* **1995**, *52*, 327–334. [CrossRef]
9. Robertson, M.P.; Igel, H.; Baertsch, R.; Haussler, D.; Ares, M., Jr.; Scott, W.G. The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biol.* **2005**, *3*, e5. [CrossRef] [PubMed]
10. Jonassen, C.M. Detection and sequence characterization of the 3′-end of coronavirus genomes harboring the highly conserved RNA motif s2m. *Methods Mol. Biol.* **2008**, *454*, 27–34. [PubMed]

11. Jonassen, C.M.; Jonassen, T.O.; Grinde, B. A common RNA motif in the 3′ end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus. *J. Gen. Virol.* **1998**, *79 Pt 4*, 715–718. [CrossRef] [PubMed]
12. Schutze, H.; Ulferts, R.; Schelle, B.; Bayer, S.; Granzow, H.; Hoffmann, B.; Mettenleiter, T.C.; Ziebuhr, J. Characterization of white bream virus reveals a novel genetic cluster of nidoviruses. *J. Virol.* **2006**, *80*, 11598–11609. [CrossRef] [PubMed]
13. RAPID RISK ASSESSMENT. Severe Respiratory Disease Associated with a Novel Coronavirus. Available online: http://ecdc.europa.eu/en/publications/Publications/novel-coronavirus-rapid-risk-assessment-update.pdf (accessed on 26 July 2016).
14. Su, S.; Wong, G.; Shi, W.; Liu, J.; Lai, A.C.; Zhou, J.; Liu, W.; Bi, Y.; Gao, G.F. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol.* **2016**, *24*, 490–502. [CrossRef] [PubMed]
15. Varani, G.; McClain, W.H. The G × U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep.* **2000**, *1*, 18–23. [CrossRef] [PubMed]
16. The mfold Web Server. Available online: http://unafold.rna.albany.edu/?q=mfold (accessed on 26 July 2016).
17. Goebel, S.J.; Taylor, J.; Masters, P.S. The 3′ cis-acting genomic replication element of the severe acute respiratory syndrome coronavirus can function in the murine coronavirus genome. *J. Virology* **2004**, *78*, 7846–7851. [CrossRef] [PubMed]
18. Liu, S.; State Key Laboratory of Veterinary Biotechnology, Harbin Veterinary Research Institute, Harbin, China. Personal communication, 2012.
19. Bartel, D.P. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* **2004**, *116*, 281–297. [CrossRef]
20. Shapiro, J.S.; Varble, A.; Pham, A.M.; Tenoever, B.R. Noncanonical cytoplasmic processing of viral microRNAs. *RNA* **2010**, *16*, 2068–2074. [CrossRef] [PubMed]
21. Boss, I.W.; Renne, R. Viral miRNAs and immune evasion. *Biochim. Biophys. Acta* **2011**, *1809*, 708–714. [CrossRef] [PubMed]
22. Hackenberg, M.; Rodriguez-Ezpeleta, N.; Aransay, A.M. Miranalyzer: An update on the detection and analysis of microRNAs in high-throughput sequencing experiments. *Nucleic Acids Res.* **2011**, *39*, W132–W138. [CrossRef] [PubMed]
23. Bredenbeek, P.J.; Pachuk, C.J.; Noten, A.F.; Charite, J.; Luytjes, W.; Weiss, S.R.; Spaan, W.J. The primary structure and expression of the second open reading frame of the polymerase gene of the coronavirus MHV-A59; a highly conserved polymerase is expressed by an efficient ribosomal frameshifting mechanism. *Nucleic Acids Res.* **1990**, *18*, 1825–1832. [CrossRef] [PubMed]
24. Sawicki, S.G.; Sawicki, D.L. A new model for coronavirus transcription. *Adv. Exp. Med. Biol.* **1998**, *440*, 215–219. [PubMed]
25. Silva, P.A.G.C.; Pereira, C.F.; Dalebout, T.J.; Spaan, W.J.M.; Bredenbeek, P.J. An RNA pseudoknot is required for production of yellow fever virus subgenomic RNA by the host nuclease XRN1. *J. Virol.* **2010**, *84*, 11395–11406. [CrossRef] [PubMed]
26. Pijlman, G.P.; Funk, A.; Kondratieva, N.; Leung, J.; Torres, S.; van der Aa, L.; Liu, W.J.; Palmenberg, A.C.; Shi, P.Y.; Hall, R.A.; et al. A highly structured, nuclease-resistant, noncoding RNA produced by flaviviruses is required for pathogenicity. *Cell Host Microbe* **2008**, *4*, 579–591. [CrossRef] [PubMed]
27. Chlebowski, A.; Lubas, M.; Jensen, T.H.; Dziembowski, A. RNA decay machines: The exosome. *Biochim. Biophys. Acta* **2013**, *1829*, 552–560. [CrossRef] [PubMed]
28. Villordo, S.M.; Gamarnik, A.V. Genome cyclization as strategy for flavivirus RNA replication. *Virus Res.* **2009**, *139*, 230–239. [CrossRef] [PubMed]
29. Fricke, M.; Dunnes, N.; Zayas, M.; Bartenschlager, R.; Niepmann, M.; Marz, M. Conserved RNA secondary structures and long-range interactions in hepatitis C viruses. *RNA* **2015**, *21*, 1219–1232. [CrossRef] [PubMed]
30. Geigenmuller, U.; Ginzton, N.H.; Matsui, S.M. Construction of a genome-length cDNA clone for human astrovirus serotype 1 and synthesis of infectious RNA transcripts. *J. Virol.* **1997**, *71*, 1713–1717. [PubMed]