CHAPTER 6

Arterivirus nsp12 versus the coronavirus nsp16 2′-O-methyltransferase: comparison of the C-terminal cleavage products of two nidovirus pp1ab polyproteins

Kathleen C. Lehmann
Lisa Hooghiemstra
Anastasia Gulyaeva
Dmitry Samborskiy
Jessika C. Zevenhoven-Dobbe
Eric J. Snijder
Alexander E. Gorbalenya
and Clara C. Posthuma

Submitted

ABSTRACT

The 3′-terminal domain of the most conserved open reading frame 1b (ORF1b) in three of the four families of the order *Nidovirales* (except the *Arteriviridae*) encodes a (putative) 2′-O-methyltransferase (O-MT), known as nonstructural protein (nsp) 16 in coronaviruses and implicated in methylation of the 5′ cap structure of nidoviral mRNAs. Like coronavirus transcripts, arterivirus mRNAs are assumed to possess a 5′ cap although no candidate methyltransferases (MTases) were identified thus far. To address this knowledge gap, we analyzed the uncharacterized nsp12 of arteriviruses, which occupies the ORF1b position equivalent to that of coronavirus nsp16. In our in-depth bioinformatics analysis of nsp12, the protein was confirmed to be family-specific while having diverged much farther than other nidovirus ORF1b-encoded proteins, including those of the *Coronaviridae*. Only one invariant and several partially conserved, predominantly aromatic residues were identified in nsp12, which may adopt a structure with alternating α-helices and β-strands, an organization also found in known MTases. However, no statistically significant similarity was found between nsp12 and the two-fold larger coronavirus nsp16, nor could we detect MTase activity in biochemical assays using recombinant equine arteritis virus nsp12. Our further analysis established that this subunit is essential for replication of this prototypic arterivirus. Using reverse genetics, we assessed the impact of 25 substitutions at 14 positions, yielding virus phenotypes ranging from wild-type-like to nonviable. Notably, replacement of the invariant phenylalanine 109 with tyrosine was lethal. We conclude that nsp12 plays an essential role during EAV replication, possibly by acting as a co-factor for another enzyme.

INTRODUCTION

Arteriviruses (family *Arteriviridae*) are positive-stranded RNA viruses with genome sizes ranging from 13 to 16 kilobases. The family currently comprises a single genus that includes four species: *Equine arteritis virus* (EAV), *Simian hemorrhagic fever virus* (SHFV), *Lactate dehydrogenase-elevating virus* (LDV), and *Porcine reproductive and respiratory syndrome virus* (PRRSV) (1;2). Among those, the latter is the economically most relevant species causing annual losses to the American swine industry alone of about $800 million (3). Additionally, several recently identified arteriviruses remain to be formally classified, but are likely to prototype multiple novel species or even higher order taxa (4-7). Arterivirus genomes are polycistronic and contain 10 to 15 (known) open reading frames (ORFs). The 5'-proximal ORFs 1a and 1b are expressed as polyproteins (pps) 1a and 1ab that are autoproteolytically processed into the nonstructural proteins (nsps) required for genome replication and transcription (Figure 1) (8). The remaining ORFs mostly encode structural proteins that are expressed from a set of subgenomic (sg)

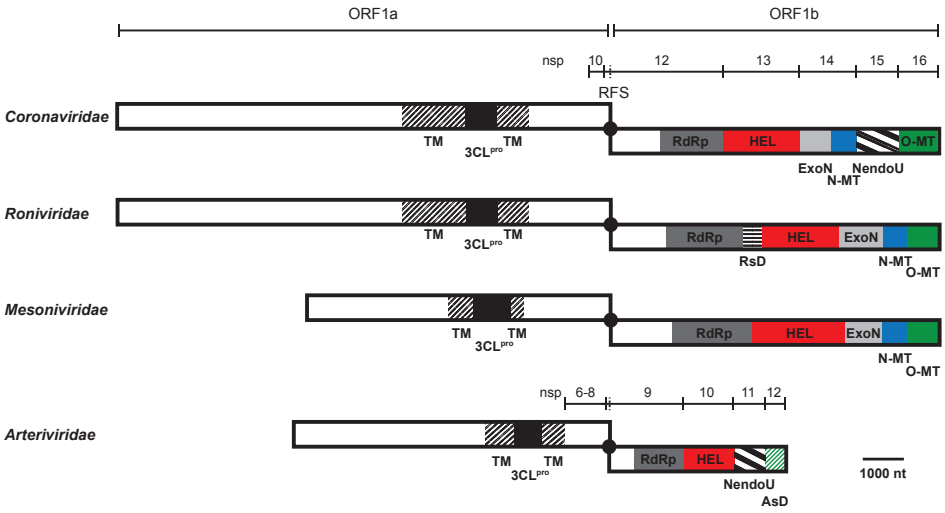

Figure 1: Organization of key replicase domains encoded by nidovirus open reading frames (ORFs) 1a and 1b. Proteolytic cleavage products described in the text for the *Corona-* and *Arteriviridae* are indicated. Matching colors/patterns indicate domain conservation between families. Domains (putatively) involved in capping (HEL, N-MT, O-MT, AsD) are depicted in bright colors. nsp, nonstructural protein; TM, transmembrane domain; $3CL^{pro}$, 3C-like protease; black dot and RFS, ribosomal frameshift site; RdRp, RNA-dependent RNA polymerase; HEL, helicase/RNA triphosphatase; ExoN, exoribonuclease; N-MT, N7-methyltransferase; NendoU, endoribonuclease; O-MT, 2'-O-methyltransferase; RsD, Ronivirus-specific domain; AsD, arterivirus-specific domain (nsp12). Genomic organizations are shown for Beluga whale coronavirus SW1 (*Coronaviridae*), gill-associated virus (*Roniviridae*), Nam Dinh virus (*Mesoniviridae*), and porcine respiratory and reproductive syndrome virus, North American genotype (*Arteriviridae*). Depicted is a simplified domain organization since most enzymes are multidomain proteins. Note that viruses of the *Coronaviridae* family that do not belong to the subfamily of *Coronavirinae* encode a truncated version of N-MT. Adapted from (61).

mRNAs (9). Based on overall similarities in terms of genome expression and organization as well as synteny and homology of key replicase domains, arteriviruses were united in the order *Nidovirales* with the families *Mesoniviridae*, *Roniviridae*, and *Coronaviridae*, the latter including two distantly related subfamilies, *Coronavirinae* and *Torovirinae* (10;11). In the nidovirus tree, the arteriviruses form a basal lineage next to the one that combines the three other families, which have substantially larger genomes (12).

ORF1b is the most conserved part of the nidovirus genome, and all ORF1b-encoded proteins characterized thus far are enzymes conserved in two or more nidovirus families. The RNA-dependent RNA polymerase and a zinc-binding domain (ZBD) fused with a superfamily 1 helicase (HEL1) are conserved in all nidoviruses. In contrast, six other domains are lineage specific. Four of these are conserved in two or three nidovirus families only: exoribonuclease (ExoN), N7-methyltransferase (N-MT), nidovirus uridylate-specific endoribonuclease (NendoU), and 2'-O-methyltransferase (O-MT). Two other domains are yet uncharacterized and unique to either roniviruses (RsD, ronivirus-specific domain) or arteriviruses (AsD, arterivirus-specific domain). Since five of the six lineage-specific domains occupy a unique position in the genome, the pattern of their conservation could be explained by loss or acquisition of a single domain during nidovirus evolution (12). The exception is AsD, which resides in the most C-terminal subunit of the arterivirus ORF1b polyprotein (nsp12), the position occupied by the O-MT protein in all other nidoviruses (nsp16 in coronaviruses, Figure 1). If these positionally equivalent proteins are unrelated, as reported 14 years ago based on the analysis of only a few genome sequences and prior to the identification of the O-MT (13), their emergence would require the consideration of complex evolutionary hypotheses. Thus, the relation of AsD with the O-MT and other proteins must be re-evaluated while taking advantage of the increased availability of sequences and improved techniques.

Unlike AsD, the coronavirus O-MT has been experimentally characterized (14-17) and was found to provide one of the four activities required for the formation of a so-called type I cap (cap-1) (mGpppNm) structure at the 5' end of coronaviral mRNAs (18;19). Two other coronavirus enzymes, HEL1 (nsp13) (20;21) and the N-MT (nsp14) (14;22), are also known to be involved in capping, whereas the fourth enzyme required (guanylyltransferase) remains to be identified. *In vitro* the coronavirus N-MT and O-MT were found to cooperate during cap formation. The latter enzyme also requires the ORF1a-encoded nsp10 as a co-factor (14). Although arteriviruses were not characterized in detail, the SHFV genome was reported to be capped (23), and they do encode a HEL1 (24), which could contribute to capping. Thus, the discovery of arteriviral N-MT and/or O-MT activities could be readily accommodated in a functionally sensible manner.

Based on the above evolutionary and functional considerations, we sought to characterize nsp12 of arteriviruses by testing the hypothesis that it may be a methyltransferase. We show that, unlike the coronavirus O-MT, nsp12 is poorly conserved among known arteriviruses compared to the proteins carrying the endoribonuclease (nsp11) and helicase (nsp10) activities, and that it contains only one evolutionary invariant residue. No statistically significant similarity was found between arterivirus nsp12 and coronavirus nsp16 or other proteins although the two nidovirus proteins may belong to the same α/β fold class. Likewise, no MTase activity was detected in carefully controlled assays using recombinant EAV nsp12 in the absence or presence of several other nsps that were included as potential co-factors. Using reverse genetics, a large set of EAV nsp12 mutants was generated and tested for replication, revealing phenotypes ranging from wild-type-like to replication-deficient, which broadly correlated with the natural variation of the probed residues. We conclude that nsp12 plays an essential role in EAV replication and discuss possible directions to elucidate its enigmatic function.

RESULTS

Sequence similarities and dissimilarities between arterivirus nsp12 and (putative) methyltransferases of the *Coronaviridae*

We first analyzed the conservation of nsp12 in comparison with that of other proteins deriving from the C-terminal portion of pp1ab of arteriviruses and the *Coronavirinae* and *Torovirinae*. Starting at the ZBD, the region analyzed included the three proteins implicated in 5' cap formation in coronaviruses. We found that nsp12 is conserved in all established and provisional arterivirus species, including the most distantly related wobbly possum disease virus (WPDV). Inspection of the arterivirus conservation profile showed that the entire nsp12 sequence exhibits similarity values that are below average for this pp1ab region (0.320 on a -0,1-1 scale; Figure 2A). Only the C-terminal domain of nsp10 and to some extent the ZBD were similarly divergent while the similarity of the nsp10 helicase core and particularly nsp11 were above average. This remarkably low conservation distinguishes arterivirus nsp12 also from all proteins in this region of the *Coronavirinae* (average conservation 0.491) and *Torovirinae* (0.270), including nsp16 (Figures 2B and C). Accordingly, arterivirus nsp12 contains the smallest number of conserved residues among the analyzed proteins, with only a single phenylalanine (F109 in EAV) being evolutionarily invariant (Figure 3). Other notable conserved nsp12 residues (out of 18 in total) are an asparagine, a serine/threonine and six aromatic residues. We also noted the presence of four conserved cysteines in a pattern typical for zinc-fingers in the C-terminal part of nsp12 in the five simian arteriviruses, which con-

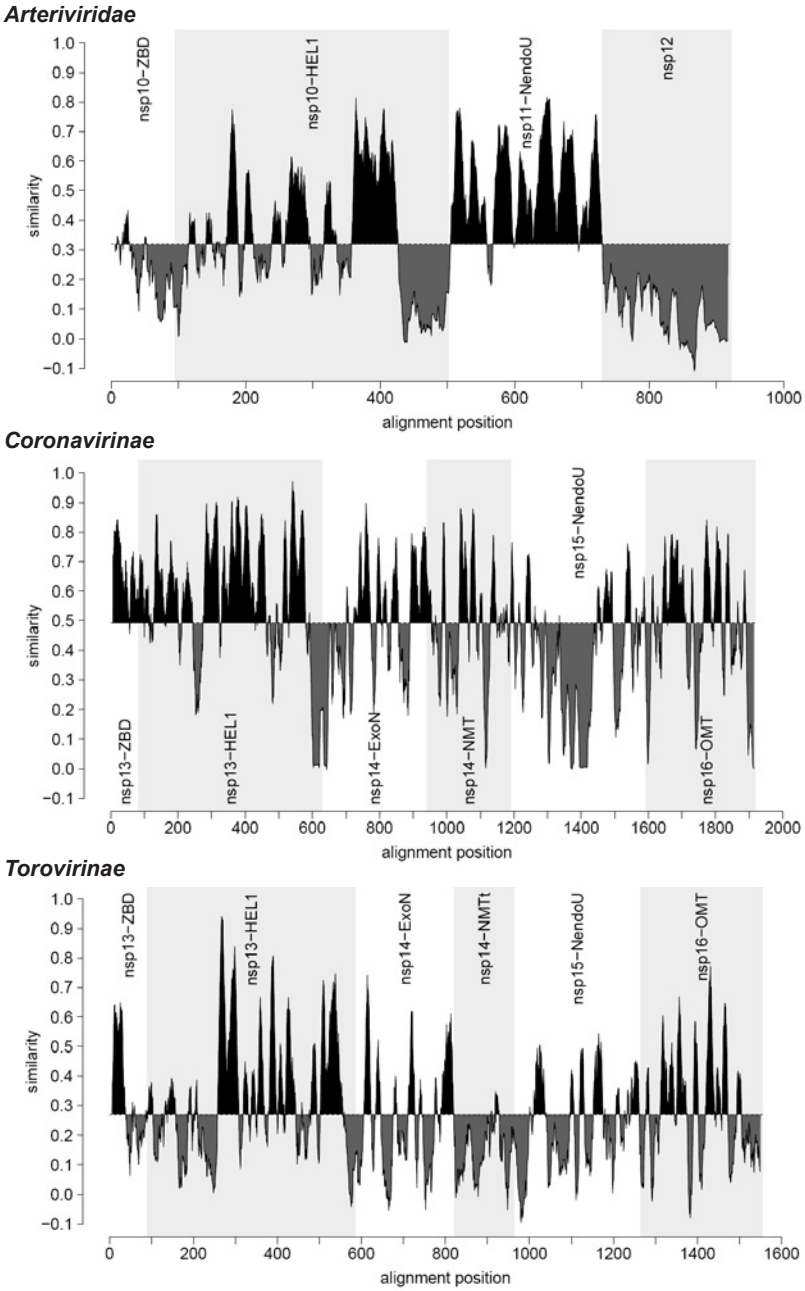

Figure 2: Similarity density plots of the C-terminal region of polyprotein 1ab of different nidovirus (sub) families. Values above and below average similarities are indicated in black and gray, respectively. nsp, nonstructural protein; ZBD, zinc-binding domain; Hel1, helicase core domain; ExoN, exoribonuclease; NMT, N7-methyltransferase (t, truncated); NendoU, endoribonuclease; OMT, 2′-O-methyltransferase. For the sake of simplicity, we have applied the nsp nomenclature of the *Coronavirinae* subfamily also to the orthologous torovirus domains for which the processing of pp1a/pp1ab is yet to be fully described.

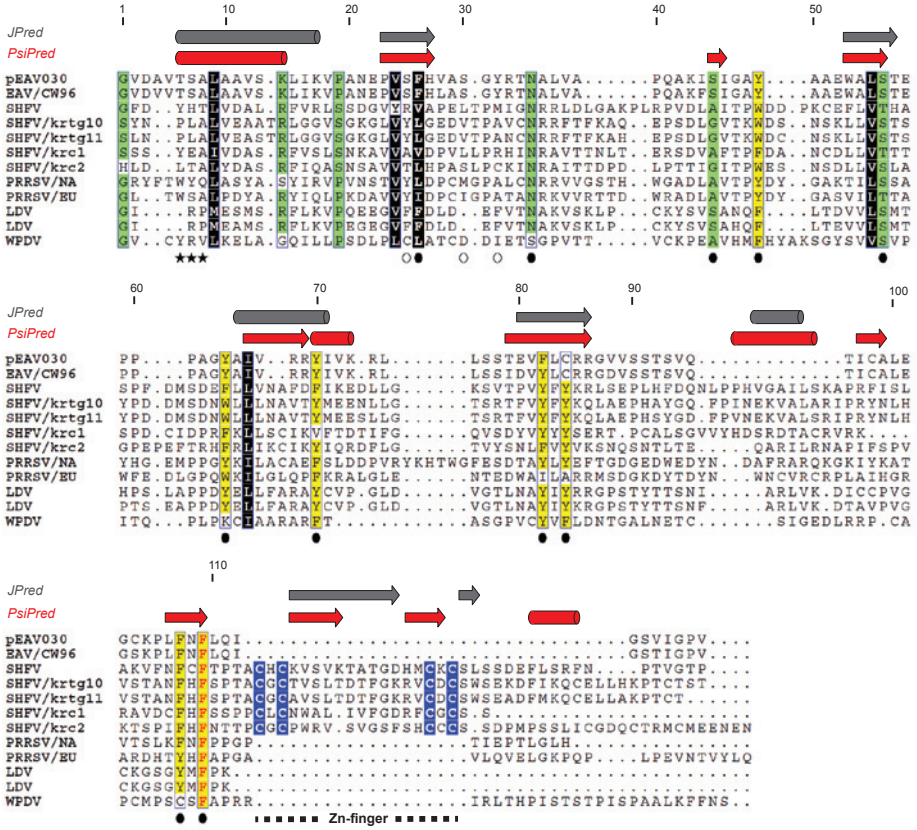

Figure 3: Multiple sequence alignment and secondary structure predictions of representative arterivirus nsp12 sequences. Partially and fully conserved amino acids are highlighted in colored boxes. Colors represent residues with similar biophysical properties; yellow, aromatic; black, hydrophobic; blue, (putatively) zinc-binding; green, other. Secondary structures (barrel, α-helix; arrow, β-strand) were predicted with JPred (50) (gray) or PsiPred (51) (red) based on the multiple sequence alignment. Residue numbers are indicated for nsp12 of the EAV-Bucyrus isolate (pEAV030) (57), the parental strain of pEAV211 used for the reverse genetics experiments. Replaced residues are indicated below the alignment; black stars, positions where stop codons were introduced; empty circles, control residues; filled circles, conserved residues. A putative zinc-finger in simian arterivirus nsp12 sequences is indicated by a dashed line. EAV, equine arteritis virus (GenBank accession number AY349167); SHFV, simian hemorrhagic fever virus (AF180391, JX473847, JX473848, HQ845737, HQ845738); PRRSV, porcine reproductive and respiratory syndrome virus (JX138233, JF802085); LDV, lactate dehydrogenase-elevating virus (L13298, U15146); WPDV, wobbly possum disease virus (JN116253).

stitute a phylogenetically compact cluster. Patristic pair-wise distances (PPDs) of nsp12 compared to those of ZBD, HEL1, and NendoU were consistently larger while PPDs of (putative) O-MTs were comparable on average with those of five other domains in the *Coronavirinae* and *Torovirinae* (Figure S1). These results showed that, in comparison to the coronavirus O-MT, nsp12 must have evolved under unusually relaxed constraints or

in a changing molecular environment. Secondary structure predictions using JPred and PsiPred consistently indicated the alternation of α-helices and β-strands in arterivirus nsp12 (Figure 3). Interestingly, also the coronavirus MTases belong to the α/β structural class and contain conserved aromatic residues (15;17). Nevertheless, HH-suite profile-profile comparison did not reveal sequence similarity above the background between nsp12 and the O-MT of corona- or toroviruses, E=0.41 and 0.53, respectively (Figure S2). Furthermore, these proteins are also of different sizes: 119-178 aa (arterivirus nsp12) versus 263-312 aa (coronavirus nsp16), with the arterivirus proteins being also smaller than MTases of other origins. The above HH-based negative result contrasted with the strong similarity signal observed in (control) comparisons between arteriviruses and corona- or toroviruses for HEL1 and NendoU (E=3.5e-17 or better), or in the control comparison between corona- and torovirus nsp16, E=2.3e-32 (Figure S2). No statistically significant similarity was observed between nsp12 and other proteins in an HMM-based scan of the PFAM-A database (top hit: PF12581, E=1.0). We thus concluded that nsp12 has diverged beyond recognition from its homologs and differs considerably from the O-MT of large nidoviruses. Nevertheless, the obtained results did not rule out the possibility that it could be a deviant MTase, and we therefore set out to test this hypothesis experimentally by biochemical and molecular virological methods.

Purification of recombinant EAV nsp12 and several ORF1a-encoded proteins

We engineered vectors encoding recombinant EAV nsp12 derivatives carrying either an N-terminal or a C-terminal hexahistidine tag and expressed them in *E. coli*. Only the N-terminally tagged protein was successfully expressed and purified by metal affinity chromatography using Co^{2+} (Talon) beads (Figure 4A). The protein appeared to be reasonably stable at all conditions tested, including a pH range from 6.0 to 7.5 and protein concentrations of up to 500 μM. Yet upon storage the protein increasingly formed dimers and higher order multimers, even in the presence of 1 mM DTT. In gel filtration experiments with fresh protein these oligomers were not evident. Instead a single peak was observed (not shown) that corresponded well to the expected size of an nsp12 monomer (calculated weight 13 kDa vs. predicted weight based on Stokes radius 16 kDa).

In addition to nsp12, we also expressed five small mature proteins and cleavage intermediates from the nsp7 region of pp1a (nsp6-7, nsp6-7-8, nsp7α, nsp7β, and nsp7 (i.e., nsp7α-7β)) (25;26) (Figure 4B). In coronaviruses, the corresponding part of ORF1a encodes nsp10, an essential co-factor for the O-MT (14). Consequently, we added these purified recombinant proteins to nsp12 in MTase activity assays (see below).

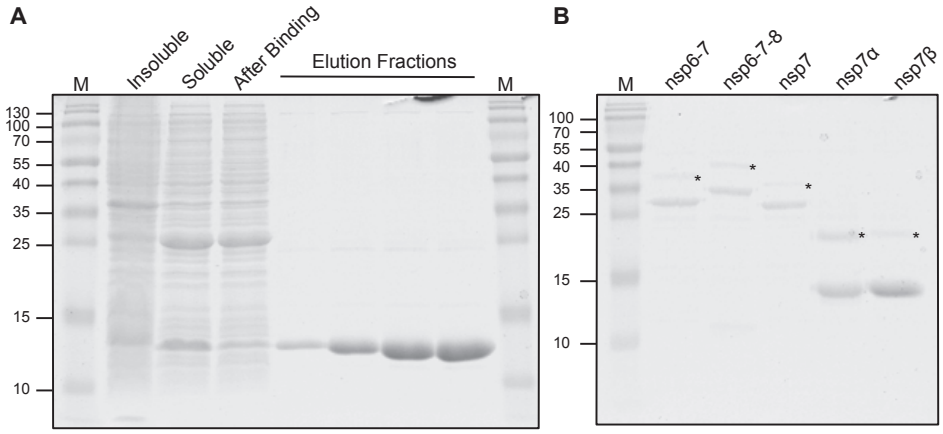

Figure 4: SDS-PAGE analysis of purified EAV nonstructural proteins. (**A**) The progression of metal-ion chromatography of EAV nsp12-containing (MW 13 kDa) E. coli lysates was monitored by Coomassie brilliant blue staining. Insoluble and soluble: proteins retained in pellet or supernatant, respectively, after cell lysis and ultracentrifugation; after binding: proteins in supernatant after removal of Talon beads. (**B**) Elution fractions of EAV ORF1a proteins and intermediates (MWnsp6-7 29 kDa, MWnsp6-7-8 34 kDa, MWnsp7 26 kDa, MWnsp7α 15 kDa, MWnsp7ß 13 kDa). Products marked with an asterisk are remaining ubiquitin-nsp fusion proteins. Size markers are indicated on the left in kDa.

Recombinant nsp12 does not display *in vitro* MTase activity using a variety of substrates

Using purified arterivirus proteins, we proceeded to test for MTase activity in the presence of different methyl acceptors by employing an *in vitro* assay similar to that previously established for SARS-CoV nsp14 and nsp16 (14). In agreement with published results (14;22;27), both SARS-CoV MTases (kindly provided by Dr. Etienne Decroly, Marseille), which were used as positive controls, transferred the radioactive methyl group from the universal methyl donor S-adenosylmethionine to non-methylated or N7-methylated cap analogs (Figure 5). Likewise, also vaccinia virus capping enzyme, obtained from a commercial source and known to harbor N-MT activity, demonstrated the expected activity. Based on these activities and the results of two negative control reactions (assays using BSA and no acceptor, respectively), we defined an incorporation threshold of 1000 cpm to distinguish the enzyme activity in this assay. According to this definition, EAV nsp12 did not display activity with any of the methyl acceptors in the absence or presence of any of the potential ORF1a-encoded co-factors described above (nsp6-7, nsp6-7-8, nsp7α, nsp7β, and nsp7).

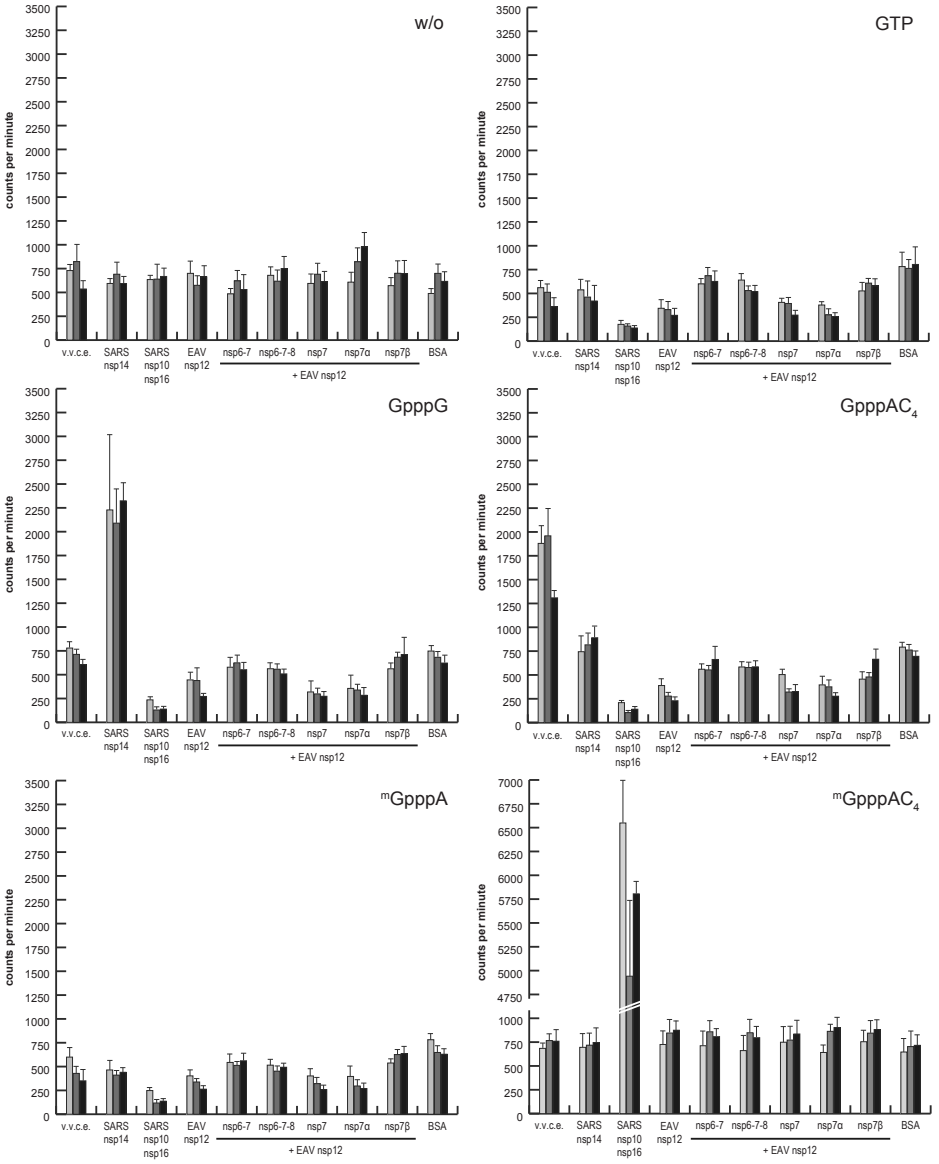

Figure 5: Methyltransferase (MTase) activity assays using recombinant EAV nsp12 in the presence and absence of possible co-factors. Recombinant EAV nsp12 (1 µM) and equimolar amounts of the indicated possible co-factors were incubated for 30 (light gray), 60 (dark gray), or 180 min (black) with S-[methyl-^{3}H]-adenosylmethionine and the indicated methyl acceptor. Proteins with known MTase activity served as positive controls. v.v.c.e., vaccinia virus capping enzyme (0.1 U/µl, N-MT); SARS nsp14 (75 nM, N-MT); SARS nsp10/nsp16 (2 µM complex, O-MT); SARS, SARS coronavirus; BSA served as negative control; error bars indicate the standard deviation of the mean of two independent experiments. The background variation evident for several of the protein combinations using GTP, GpppG, GpppAC$_4$, or ^{m}GpppA most likely represents an artifact originating from a position effect, which was observed repeatedly in the employed 96-well format.

The tolerance of EAV replication to nsp12 mutagenesis correlates with the natural variation of probed residues

To establish the general importance of nsp12 for EAV replication, we used reverse genetics to assess whether EAV tolerates replacements at conserved positions, including the single absolutely (F109) and ten partially (F26, N35, S45, Y49, S56, Y64, Y70, F82, C84, and F107) conserved residues (Figure 3). We also tested replacements of three poorly conserved residues (S25, S30, and Y32) that served as controls. Furthermore, we also abolished nsp12 expression by replacing its codons 6 to 8 with three consecutive translation termination codons (STOP mutant). The engineered cDNA clones were used for *in vitro* transcription, yielding full-length RNA that was subsequently electroporated into BHK-21 cells. The effects of the replacements were first assessed on the level of viral protein expression by immunofluorescence microscopy utilizing antibodies against nsp3 and the structural nucleocapsid (N) protein. Furthermore, we monitored the production of virus progeny by harvesting transfected cell culture supernatants and performing plaque assays (Table 1).

For the STOP mutant neither protein expression nor progeny production was observed, indicating that nsp12 performs an indispensable function during virus replication. Alternatively, the truncation of nps12 may have affected virus viability indirectly, e.g. by impairing proteolytic cleavage of the nsp11/nsp12 junction, which might be detrimental to the activity of the nsp11 endoribonuclease. This concern was addressed by replacing individual nsp12 residues.

The 14 residues probed by making 25 mutants could be classified into four groups based on the impact of their replacement. The first group included residues F107 and F109, with the four mutants carrying alanine or (more conservative) tyrosine substitutions at these positions not producing any virus progeny. Interestingly, in contrast to both alanine mutants and F109Y, which also did not produce viral proteins, immunofluorescence signal for nsp3 and N protein was detected for F107Y at 24 h and 48 h post transfection (p.t.), with a stronger signal being observed at the earlier time point. Collectively, these results show that F107 or F109 are most strongly constrained in EAV and indicate a vital role of these residues in virus viability.

The second group comprised residues F26, N35, and C84, which appeared to be only slightly less important than the aforementioned F107 and F109, based on the phenotype of five mutants. Alanine substitutions at position F26 and N35 were either lethal (F26A) or severely detrimental (N35A), whereas tyrosine or aspartate substitutions of these residues (F26Y and N35D) were compatible with at least some residual replica-

Table 1: EAV nsp12 mutants and their phenotypes

group[‡]	mutant	observed amino acid variation[§]	wild-type sequence	mutated sequence	immunofluorescence assay				titer (PFU/ml)		nsp12 sequence of P1 virus[†]
					14 h p.t.	24 h p.t.	48 h p.t.	68 h p.t.	14 h p.t.	48 h p.t.	
	wt				+	+	+	+	$3{\cdot}10^{6}$	$2{\cdot}10^{8}$	n.d.
4	S25A	FYTSARC	UCA	GCU	+	+	+	+	$3{\cdot}10^{6}$	$2{\cdot}10^{7}$	mutation retained
2	F26A	FILV	UUC	GCA	-	-	-	-	<20	<20	n.d.
	F26Y	FILV	UUC	UAU	-	-	+	+	<20	$4{\cdot}10^{7}$	reversion
4	S30A	SLVIMD	UCA	GCU	+	+	+	+	$6{\cdot}10^{5}$	$6{\cdot}10^{7}$	mutation retained
4	Y32A	IMAFRCY	UAC	GCA	-	+	+	+	20	$5{\cdot}10^{5}$	Y32V
	Y32F	IMAFRCY	UAC	UUU	+	+	+	+	$5{\cdot}10^{8}$	$2{\cdot}10^{8}$	mutation retained
2	N35A	NS	AAC	GCU	-	-	-	+	<20	<20	reversion
	N35D	NS	AAC	GAU	+	+	+	+	$1{\cdot}10^{3}$	$1{\cdot}10^{8}$	reversion
4	S45A	SAG	UCA	GCU	+	+	+	+	$2{\cdot}10^{7}$	$5{\cdot}10^{7}$	mutation retained
	S45T	SAG	UCA	ACC	+	+	+	+	$1{\cdot}10^{3}$	$1{\cdot}10^{8}$	reversion
3	Y49A	YFW	UAC	GCA	-	-	-	-	<20	<20	n.d.
	Y49F	YFW	UAC	UUU	+	+	+	+	$6{\cdot}10^{5}$	$4{\cdot}10^{7}$	mutation retained
4	S56A	ST	UCA	GCU	+	+	+	+	$2{\cdot}10^{7}$	$4{\cdot}10^{7}$	mutation retained
	S56T	ST	UCA	ACC	+	+	+	+	$4{\cdot}10^{7}$	$2{\cdot}10^{7}$	mutation retained
3	Y64A	YFWK	UAU	GCA	-	-	-	-	<20	<20	n.d.
	Y64F	YFWK	UAU	UUC	+	+	+	+	$6{\cdot}10^{7}$	$2{\cdot}10^{8}$	mutation retained
3	Y70A	YFV	UAU	GCA	-	-	-	-	<20	<20	n.d.
	Y70F	YFV	UAU	UUC	+	+	+	+	$4{\cdot}10^{7}$	$4{\cdot}10^{7}$	mutation retained
3	F82A	FYI	UUC	GCA	-	-	-	-	<20	<20	n.d.
	F82Y	FYI	UUC	UAU	+	+	+	+	$1{\cdot}10^{8}$	$1{\cdot}10^{8}$	mutation retained
2	C84Y	CYAF	UGC	UAU	+	+	+	+	$3{\cdot}10^{2}$	$3{\cdot}10^{7}$	reversion

Table 1: EAV nsp12 mutants and their phenotypes (continued)

group[‡]	mutant	observed amino acid variation[§]	wild-type sequence	mutated sequence	immunofluorescence assay				titer (PFU/ml)		nsp12 sequence of P1 virus[†]
					14 h p.t.	24 h p.t.	48 h p.t.	68 h p.t.	14 h p.t.	48 h p.t.	
1	F107A	FYC	UUC	GCA	-	-	-	-	<20	<20	n.d.
	F107Y	FYC	UUC	UAU	-	+*	+*	-	<20	<20	n.d.
1	F109A	F	UUC	GCA	-	-	-	-	<20	<20	n.d.
	F109Y	F	UUC	UAU	-	-	-	-	<20	<20	n.d.
	STOP		UCA GCA CUA	UGA UGA UGA	-	-	-	-	<20	<20	n.d.

[‡]groups as defined in the text; [§]see Figure 3; *non-spreading, [†]P1 virus was generated by infection of fresh BHK-21 cells with supernatant harvested at 68 h p.t. (for stable mutants) or the earliest positive time point in immunofluorescence microscopy (for reverting mutants and Y32V); n.d. not done; representative results of two independent data sets obtained by two different researchers.

tion, which allowed early reversion of these mutants. Similarly, also the C84Y mutant reverted, which is notable given the presence of a tyrosine at this position in most other arteriviruses.

In contrast to the above results, EAV tolerated replacements by another aromatic residue at four other partially conserved aromatic residues, Y49, Y64, Y70, and F82, which form group 3. These virus mutants were stable and yielded progeny titers up to 1 log below that of the wild-type control. Interestingly, although the titer of Y49F was not very different from that of the parental virus, this mutant exhibited a small-plaque phenotype (Figure 6). In contrast alanine substitutions at these positions were again lethal.

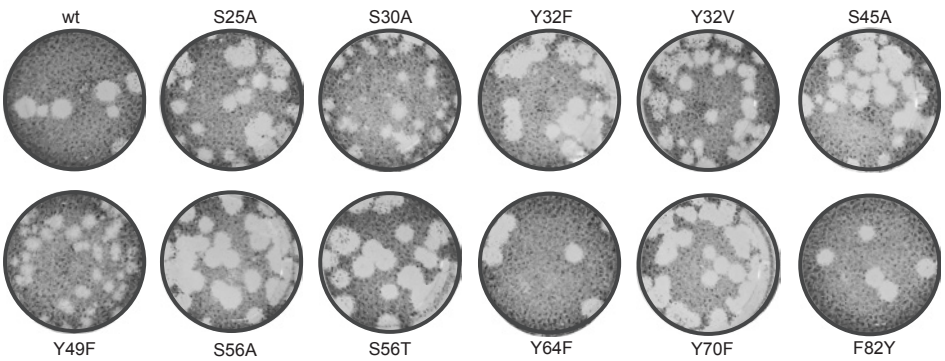

Figure 6: Plaque phenotypes of viable EAV nsp12 mutants. Virus-containing supernatants obtained 48 h post transfection were serially diluted and used to infect BHK-21 cells. After 72 h cells were fixed with 4% formaldehyde and stained with crystal violet.

The replacement – more or less conservative – of all residues mentioned thus far had a moderate to severe impact on virus replication. In contrast, the fourth group included five residues whose replacement did neither affect viral protein production nor progeny titers. As expected this group included the three poorly conserved control residues (S25, S30, and Y32). Nevertheless, S30A exhibited a small-plaque phenotype (Figure 6). Unexpectedly, we also repeatedly observed the pseudo-reversion of Y32A to Y32V, which required only a single nucleotide change. Although valine is not among the naturally occurring amino acid residues at this position (Figure 3), a hydrophobic residue is observed in several arteriviruses other than EAV. Besides substitutions of these control residues, EAV also tolerated the substitution of S56 with alanine or threonine. Given the strict conservation of serine and threonine, this lack of impact was the expected outcome for S56T, but was rather surprising for S56A. Finally, S45A was stable and indistinguishable from the parental virus, while S45T reverted. Together with the sequence variation at this position, which is limited to the small amino acids glycine, alanine, and serine, this

probably indicates a certain degree of steric hindrance by any residue larger than serine. Overall the observed mutant phenotypes were compatible with the natural variation observed at the respective positions, with the possible exception of the C84Y mutant. These correlations support the multiple sequence alignment of the highly variable nsp12 and suggest that EAV replication in BHK-21 is a faithful model system for probing nsp12 function by mutagenesis.

Both mutants displaying a small-plaque phenotype (S30A and Y49F), as well as the unexpected Y32V pseudo-revertant, were further investigated in terms of growth kinetics and accumulation of intracellular viral RNA (not shown). Compared to the wild-type control, S30A and Y49F demonstrated a slight delay in replication early during infection (8 h post infection (p.i.)) but eventually reached comparable titers by 24 h p.i.. In line with this finding, the amounts of genomic and sg mRNA at 8 h p.i. were reduced for both mutants. Whether this was due to a decreased synthesis or lower stability of their RNAs remains to be investigated. In contrast, the stable Y32V mutant was essentially indistinguishable from the wild-type control both in growth kinetics and amounts of RNA produced.

DISCUSSION

The most conserved ORF1b of nidoviruses encodes only two proteins that have not been studied before in any virus. Our study aimed to address this knowledge gap for one of these proteins, arterivirus nsp12. It established (i) the exceptional divergence of nsp12, (ii) the lack of strong bioinformatics and biochemical support for nsp12 being an MTase, and (iii) the fact that nsp12 is essential for arterivirus replication.

So far, none of the four enzymatic activities required for conventional cap-1 synthesis, or any of the known alternative capping strategies, was uncovered for arteriviruses although arteriviral mRNAs are presumed to be capped. In the conserved relative arrangement of replicative enzymes within nidovirus polyproteins 1a and 1ab, the unique arterivirus nsp12 is encoded in a genome position equivalent to that of the coronavirus O-MT, which is conserved also in invertebrate nidoviruses (Figure 1). We thus asked whether this so far uncharacterized subunit may represent an MTase, potentially capable to perform both methylation reactions as, for example, the flavivirus NS5 MTase domain is (28). Upon our bioinformatics analysis of nsp12 sequences, we found that this subunit, similar to the N-MT residing in coronavirus nsp14, is enriched with (partially) conserved aromatic amino acids and is predicted to fold in alternating α-helices and β-strands

(Figure 3). Nevertheless, no statistically significant similarity was found between nsp12 and other MTases of viral or cellular origin.

When we subsequently sought to verify our hypothesis using an *in vitro* MTase assay, we could not detect any activity for recombinant EAV nsp12, whereas our positive controls clearly confirmed the functionality of the assay. To explain this lack of activity, we argued that, as for coronavirus nsp16, a second EAV protein may be required to form a functional MTase complex. By analogy with the coronavirus nsp10 co-factor, we tested the possibility that this second protein might be encoded just upstream of the ORF1a/1b ribosomal frameshift site. We thus expressed and purified nsp7α and nsp7β, as well as three polyprotein cleavage intermediates containing these two proteins, and included them in our assays (Figure 5). However, also in these extended assays we could not detect any MTase activity. This could have multiple reasons. First, the proteins tested here may not be the correct co-factors or may be unable to properly associate with nsp12 under the conditions employed. Second, more than one co-factor may be needed to spur nsp12's MTase activity, or different RNA substrates containing specific sequences may be required. Finally, our results are compatible with a scenario in which nsp12, which is smaller than other viral MTases, does not possess MTase activity, in which case other hypotheses about its function should be considered (see below).

To explore nsp12's relevance for arterivirus replication, we engineered one truncation and 25 point mutations of EAV nsp12 and launched the corresponding mutant genomes in BHK-21 cells. Reflecting the conservation of several aromatic residues in arteriviruses, substitution with alanine was tolerated in none of the cases, whereas more conservative substitutions maintaining the residue's aromatic nature were tolerated in most of the partially conserved positions (Table 1). The only exception was F107Y, which interestingly showed a certain level of protein expression but did not produce infectious progeny. Since two arteriviruses distantly related to EAV, LDV and PRRSV genotype 1, naturally encode a tyrosine at this position (Figure 3), this result suggests an epistatic interaction between residue 107 and other unknown residue(s). EAV also did not tolerate a block of nsp12 expression (STOP mutant) or the replacement of its single absolutely conserved nsp12 residue, F109, with alanine or tyrosine. This phenotype could be explained by a trans-dominant negative effect of the nsp12 substitutions on an interaction partner of nsp12, if this partner is essential for EAV replication. This explanation is also compatible with the nonviable phenotype of several other mutants and suggests a particularly important role of the most constrained and proximal F107 and F109 in such a putative interaction.

The fact that EAV does not tolerate substitution of its single invariant nsp12 residue stands in remarkable contrast to phenotypes described for mutants of the invariant residues of the NendoU or O-MT of nidoviruses (29-32), which are both more strongly conserved than nsp12. In these studies alanine substitutions of absolutely conserved putative active site residues resulted in lower virus progeny titers and in part in small-plaque phenotypes in cell culture but did not entirely abolish virus replication.

In conclusion, our combined results may be most compatible with the notion that nsp12 is not an MTase and possibly not even an enzyme but rather a co-factor of an essential component of the arterivirus replicase. In this context, a future in-depth analysis of the nsp12 interaction network could be most informative. If nsp12 is not an MTase, this activity must be provided by another protein, but it is unlikely to be one of the three other ORF1b proteins, which are known to possess different enzymatic domains. This implies that arteriviruses may be (very) different from other nidoviruses with respect to either the nature of the 5′ end of their mRNAs and/or the mechanism generating it. We note that the presence of a 5′-terminal cap-1 structure was reported for the SHFV genome (23), but that monophosphates were claimed to present at the 5′ end of LDV mRNAs (33), calling for additional studies to resolve the apparent conflict. Finally, the possibility of cap-snatching, the strategy employed by some families of negative-stranded RNA viruses (34-36), may be explored for arteriviruses. This mechanism might accommodate the nsp11 NendoU as endoribonuclease and nsp12 as a cap-binding protein, which would connect coronavirus nsp16 and arterivirus nsp12 to a common target in an unorthodox way.

MATERIAL AND METHODS

Bioinformatics

Genomes of members of the *Arteriviridae* and *Coronaviridae* families were retrieved from GenBank (37) and RefSeq (38) using the Homology-Annotation hYbrid retrieval of GENetic Sequences (Haygens) tool http://veb.lumc.nl/HAYGENS. Codon-based multiple sequence alignments (MSAs) of virus genomes were produced using the Viralis platform (39) and assisted by the HMMER 3.1 (40), Muscle 3.8.31 (41), and ClustalW 2.012 (42) programs. Only one virus per established or tentative species, which were defined with the help of DEmARC1.3 (43), was retained for bioinformatics analyses. To retrieve information about genomes, the SNAD program (44) was used. To reveal the full extent of similarity between pairs of alignments, they were converted into HMM profiles, which were compared and visualized in a dot-plot fashion using a routine in HH-suite 2.0.15

(45;46). Distribution of similarity density in alignments was plotted using R package Bio3D (47) under the conservation assessment method "similarity", substitution matrix Blosum62 (48), and a sliding window of 11 alignment columns. To search for homologs among profiles in the PFAM A database (49), the HH-suite 2.0.15 software (45;46) was used. Secondary structure of proteins was predicted by applying JPred 3 (50) and PsiPred (51) to MSAs, with the prediction being applied to the top sequence in the MSA. The MSAs were converted into figures using ESPript (52). Reconstruction of phylogenetic trees was performed using PhyML 3.0, with the WAG amino acid substitution matrix, allowing substitution rate heterogeneity among sites (4 categories), and 1000 iterations of non-parametric bootstrapping (53). Pairwise patristic distances (PPDs) between viruses were calculated from protein trees using R package "ape" (54). Linear regression was calculated using R package "stats" (55).

Reverse genetics of EAV

Mutations specifying alanine and conservative replacements of (partially) conserved and control residues in nsp12 were generated using the QuikChange protocol. In all cases translationally silent marker mutations were introduced to allow discrimination between (partial) reversion of mutants after transfection and (possible) contamination with wild-type virus. Mutated gene fragments were introduced into full-length cDNA clone pEAV211 (56), a pEAV030 derivative (57), using appropriate shuttle vectors and restriction enzymes. The presence of the mutations was confirmed by sequencing. pEAV211 DNA was *in vitro* transcribed and RNA was purified by LiCl precipitation. RNA was transfected into BHK-21 cells as described previously (58). Transfected cells were monitored by immunofluorescence microscopy until 68 h post transfection (p.t.), using antibodies directed against EAV nsp3 and N protein as described (59). To monitor the production of viral progeny, plaque assays were performed with supernatants collected at 14 and 48 h p.t. or during the first 24 hours p.i. to determine growth kinetics, as described (58). To verify the presence of the introduced mutations or reversions in viable mutants, fresh BHK-21 cells were infected with supernatants harvested at time points at which transfected cells were positive in immunofluorescence microscopy. RNA was isolated after 18 h or when cytopathic effect was detected. Finally, the nsp12-coding region was amplified by RT-PCR using random hexameric primers in the RT step and EAV-specific primers for the PCR. PCR fragments were purified and sequenced.

Protein expression and purification

N-terminal and C-terminal His-tag fusion proteins of wild-type nsp12 were expressed from a pDEST vector. Plasmids were transformed into *E. coli* BL21 (DE3) and cells were

grown in Luria Broth with 100 µg/ml ampicillin at 37°C until OD_{600} reached 0.7. Expression was induced after addition of 0.5 mM IPTG and cells were grown for further 4 h at 37°C.

EAV ORF1a-encoded proteins were expressed with N-terminal ubiquitin and C-terminal His tags from pASK vectors (60). Plasmids were transformed into *E. coli* C2523 containing the pCG1 plasmid, which leads to constitutive expression of the ubiquitin-specific protease UBP1. Cells were grown in Luria Broth with 100 µg/ml ampicillin and 34 µg/ml chloramphenicol at 37°C until OD_{600} reached 0.7. Expression was induced after addition of 200 ng/ml anhydrotetracycline and cells were grown for another 18 h at 20°C. All pellets were harvested by centrifugation and stored at -20°C until further use.

Proteins were batch purified by metal affinity chromatography using Co^{2+} (Talon beads). All steps were performed at 4°C or on ice. Cells were resuspended in nsp12 resuspension buffer (20 mM HEPES, pH 7.5, 5 mM β-mercaptoethanol) or co-factor resuspension buffer (20 mM HEPES, pH 7.5, 10% glycerol (v/v), 5 mM β-mercaptoethanol) containing 500 mM NaCl and Roche complete EDTA-free protease inhibitor cocktail. Lysis was achieved by 30 min incubation with lysozyme (0.1 mg/ml). Genomic DNA was sheared during four sonication cycles of 10 s with intermittent cooling. Cell debris was removed by centrifugation at 20.000 g for 20 min. Cleared supernatants were incubated with an appropriate amount of Talon beads for 1 h under slow rolling. Beads were collected and washed four times for 15 min with a 20-fold volume of the respective resuspension buffer supplemented with 10 mM imidazole and first 500 mM, then 250 mM, and finally twice 100 mM NaCl. Proteins were eluted with the respective resuspension buffer containing 300 mM imidazole and 100 mM NaCl. Elution fractions were examined by SDS-PAGE, pooled, and dialyzed against 20 mM HEPES, pH 7.5, 100 mM NaCl, 25% glycerol, 1 mM DTT. All proteins were stored at -20°C. Typical yields were 1-2 mg/l culture for all proteins. Protein concentrations were calculated based on theoretical extinction coefficients and absorption at 280 nm.

Gel filtration of nsp12 was performed on a Superdex 75 10/300 GL gel filtration column with 10 mM Na-phosphate buffer, pH 6.0, 100 mM NaCl, 1 mM DTT at 4°C and a flow rate of 0.5 ml/min.

Methyltransferase assay

Methyltransferase assays were performed essentially as described previously (14). Proteins at the indicated final concentrations were incubated at 30°C for 30, 60, or 180 min in a buffer containing 20 mM HEPES, pH 7.5, 5 mM DTT, 0.5 mM $MgCl_2$, 0.5 mM $MnCl_2$,

10 µM S-adenosylmethionine, 2 µM capping substrate, and 1×10^3 Bq/µl S-[methyl-^{3}H]-adenosylmethionine. Additionally 7.5 mM NaCl were carried over from the protein storage buffer. Vaccinia virus capping enzyme (New England Biolabs) was incubated in the buffer supplied by the vendor. To stop the reaction a 10-fold volume of ice-cold S-adenosylhomocysteine (100 µM) was added. Samples were spotted on DEAE filtermats (Perkin Elmer), which were subsequently washed twice with 10 mM ammonium formate, pH 8.0, then twice with water, and finally with ethanol. Filtermats were cut, and radioactivity was measured by scintillation counting.

ACKNOWLEDGEMENTS

This work was supported by the European Union's Seventh Framework program (FP7/2007-2013) through the EUVIRNA project (European Training Network on (+) RNA virus replication and antiviral drug development, grant agreement no. 264286) and the SILVER project (grant agreement no. 260644), by the Collaborative Agreement on Bioinformatics between Leiden University Medical Center and Moscow State University (MoBiLe), and by the Leiden University Fund. The authors wish to acknowledge Dr. Etienne Decroly (AFMB, Marseille, France) for helpful discussions and for providing purified SARS-CoV nsp14 and nsp10/16 as well as MTase substrates; Alexander Kravchenko and Igor Sidorov for maintaining and advancing the Viralis platform and its databases; and Linda Boomaars and Irina Albulescu for technical assistance.

SUPPLEMENTARY DATA

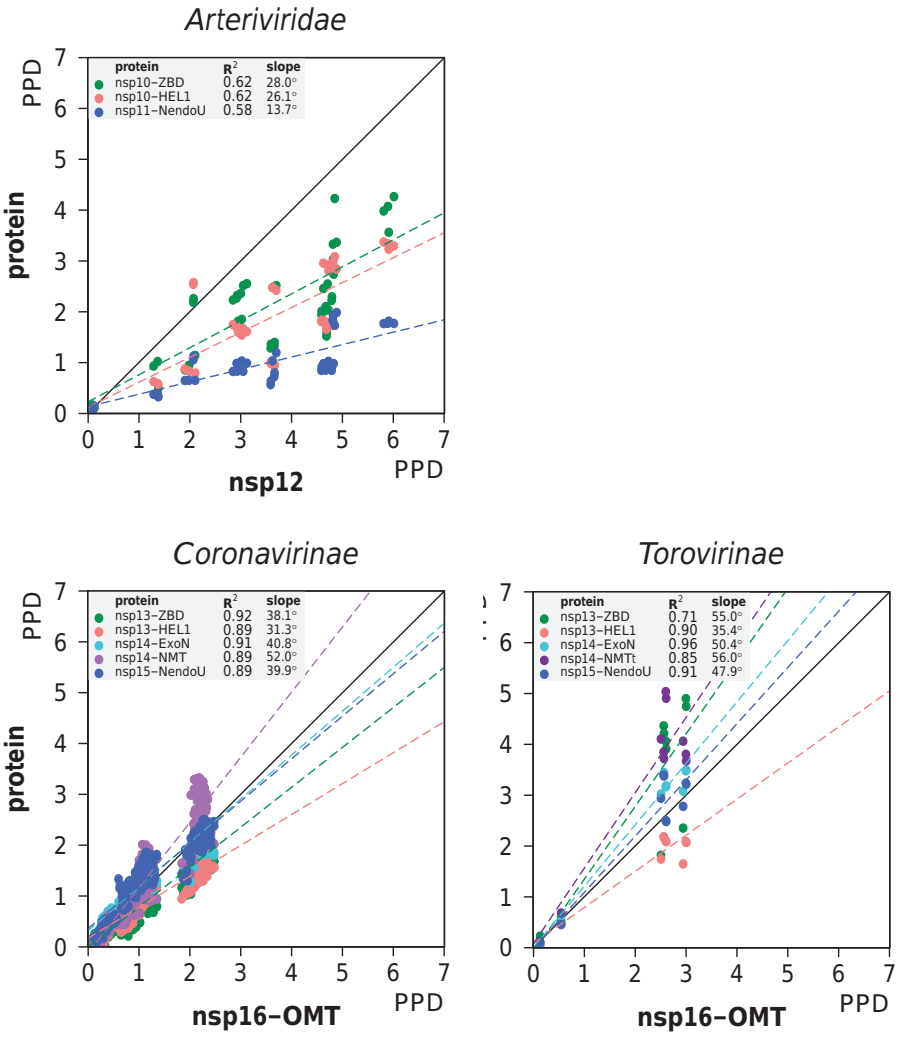

Figure S1. Relative scale of divergence of nsp12 of the *Arteriviridae* and (putative) nsp16 of the *Coronavirinae* and *Torovirinae*. Shown are three (sub)family-specific two-dimensional scatter plots that compare PPDs of the most C-terminal protein (nsp12 or nsp16-OMT, x-axis) versus PPDs of other proteins/domains of ORF1b starting from the ZBD (detailed in inset, y-axis). PPDs were calculated from PhyML trees for separate proteins. Dashed lines, linear regressions fit in respective (color matching) dot distributions with R^2 and slope values being detailed in the inset panels.

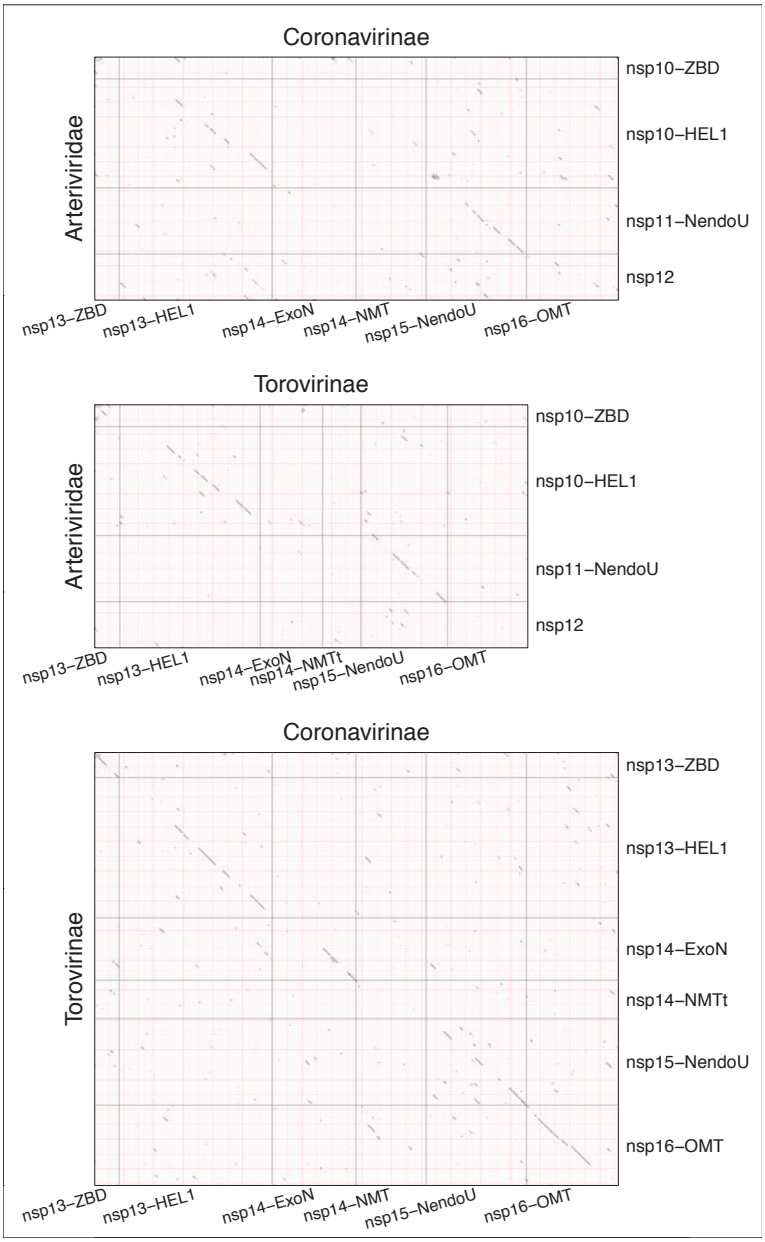

Figure S2. Analysis of co-conservation of C-terminal parts of ORF1b of the *Arteriviridae*, *Coronavirinae,* and *Torovirinae*. Shown are three pair-wise MSA-based HMM-HMM plots comparing parts of ORF1b starting from the ZBD of three origins. The position of proteins and some domains are indicated. Each MSA was converted to an HMM profile, three possible pairs of obtained HMMs were aligned with the help of HH-suite 2.0.15 software (45,46). The presence of similarity above the threshold of 0.3 is recorded with a dot. Diagonal persistence of dots is strong evidence for statistically significant similarity (homology) of a protein pair.

REFERENCE LIST

1. Snijder EJ, Kikkert M, Fang Y. Arterivirus molecular biology and pathogenesis. J.Gen.Virol. 2013; 94(Pt 10):2141-2163.

2. Faaberg KS, Balasuriya UB, Brinton MA, *et al.* Family *Arteriviridae*. In King AMQ, Adams MJ, Carstens EB *et al.* editors, Virus taxonomy. Ninth report of the international committee on taxonomy of viruses, Amsterdam, Elsevier Academic Press, 2012;796-805.

3. Sang Y, Rowland RR, Blecha F. Antiviral regulation in porcine monocytic cells at different activation States. J.Virol. 2014;88(19):11395-11410.

4. Dunowska M, Biggs PJ, Zheng T, *et al.* Identification of a novel nidovirus associated with a neurological disease of the Australian brushtail possum (Trichosurus vulpecula). Vet.Microbiol. 2012; 156(3-4):418-424.

5. Bailey AL, Lauck M, Sibley SD, *et al.* Two novel simian arteriviruses in captive and wild baboons (Papio spp.). J.Virol. 2014;88(22):13231-13239.

6. Lauck M, Hyeroba D, Tumukunde A, *et al.* Novel, divergent simian hemorrhagic fever viruses in a wild Ugandan red colobus monkey discovered using direct pyrosequencing. PLoS.One. 2011;6(4): e19056.

7. Lauck M, Sibley SD, Hyeroba D, *et al.* Exceptional simian hemorrhagic fever virus diversity in a wild African primate community. J.Virol. 2013;87(1):688-691.

8. Molenkamp R, Greve S, Spaan WJ, *et al.* Efficient homologous RNA recombination and requirement for an open reading frame during replication of equine arteritis virus defective interfering RNAs. J.Virol. 2000;74(19):9062-9070.

9. Pasternak AO, Spaan WJ, Snijder EJ. Nidovirus transcription: how to make sense…? J.Gen.Virol. 2006;87(Pt 6):1403-1421.

10. de Groot RJ, Baker SC, Baric R, *et al.* Family *Coronaviridae*. In King AMQ, Adams MJ, Carstens EB *et al.* editors, Virus taxonomy. Ninth report of the international committee on taxonomy of viruses, Amsterdam, Elsevier Academic Press, 2012;806-828.

11. de Groot RJ, Cowley JA, Enjuanes L, *et al.* Order *Nidovirales*. In King AMQ, Adams MJ, Carstens EB *et al.* editors, Virus taxonomy. Ninth report of the international committee on taxonomy of viruses, Amsterdam, Elsevier Academic Press, 2012.

12. Nga PT, Parquet MC, Lauber C, *et al.* Discovery of the first insect nidovirus, a missing evolutionary link in the emergence of the largest RNA virus genomes. PLoS.Pathog. 2011;7(9):e1002215.

13. Gorbalenya AE. Big nidovirus genome. When count and order of domains matter. Adv.Exp.Med. Biol. 2001;49:41-17.

14. Bouvet M, Debarnot C, Imbert I, *et al. In vitro* reconstitution of SARS-coronavirus mRNA cap methylation. PLoS.Pathog. 2010;6(4):e1000863.

15. Chen Y, Su C, Ke M, *et al.* Biochemical and structural insights into the mechanisms of SARS coronavirus RNA ribose 2'-O-methylation by nsp16/nsp10 protein complex. PLoS.Pathog. 2011;7(10): e1002294.

16. Decroly E, Imbert I, Coutard B, *et al.* Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2'O)-methyltransferase activity. J.Virol. 2008;82(16):8071-8084.

17. Decroly E, Debarnot C, Ferron F, *et al.* Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex. PLoS.Pathog. 2011;7(5): e1002059.

18. Lai MM, Stohlman SA. Comparative analysis of RNA genomes of mouse hepatitis viruses. J.Virol. 1981;38(2):661-670.

19. van Vliet AL, Smits SL, Rottier PJ, *et al.* Discontinuous and non-discontinuous subgenomic RNA transcription in a nidovirus. EMBO J. 2002;21(23):6571-6580.

20. Ivanov KA, Ziebuhr J. Human coronavirus 229E nonstructural protein 13: characterization of duplex-unwinding, nucleoside triphosphatase, and RNA 5'-triphosphatase activities. J.Virol. 2004; 78(14):7833-7838.

21. Ivanov KA, Thiel V, Dobbe JC, *et al.* Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. J.Virol. 2004;78(11):5619-5632.

22. Chen Y, Cai H, Pan J, *et al.* Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. Proc.Natl.Acad.Sci.U.S.A 2009;106(9):3484-3489.

23. Sagripanti JL, Zandomeni RO, Weinmann R. The cap structure of simian hemorrhagic fever virion RNA. Virology 1986;151(1):146-150.

24. Seybert A, van Dinten LC, Snijder EJ, *et al.* Biochemical characterization of the equine arteritis virus helicase suggests a close functional relationship between arterivirus and coronavirus helicases. J.Virol. 2000;74(20):9586-9593.

25. van Aken D, Zevenhoven-Dobbe J, Gorbalenya AE, *et al.* Proteolytic maturation of replicase polyprotein pp1a by the nsp4 main proteinase is essential for equine arteritis virus replication and includes internal cleavage of nsp7. J.Gen.Virol. 2006;87(Pt 12):3473-3482.

26. Wassenaar AL, Spaan WJ, Gorbalenya AE, *et al.* Alternative proteolytic processing of the arterivirus replicase ORF1a polyprotein: evidence that NSP2 acts as a cofactor for the NSP4 serine protease. J.Virol. 1997;71(12):9313-9322.

27. Jin X, Chen Y, Sun Y, *et al.* Characterization of the guanine-N7 methyltransferase activity of coronavirus nsp14 on nucleotide GTP. Virus Res. 2013;176(1-2):45-52.

28. Zhou Y, Ray D, Zhao Y, *et al.* Structure and function of flavivirus NS5 methyltransferase. J.Virol. 2007;81(8):3891-3903.

29. Kang H, Bhardwaj K, Li Y, *et al.* Biochemical and genetic analyses of murine hepatitis virus Nsp15 endoribonuclease. J.Virol. 2007;81(24):13587-13597.

30. Menachery VD, Yount BL, Jr., Josset L, *et al.* Attenuation and restoration of severe acute respiratory syndrome coronavirus mutant lacking 2'-o-methyltransferase activity. J.Virol. 2014;88(8):4251-4264.

31. Posthuma CC, Nedialkova DD, Zevenhoven-Dobbe JC, *et al.* Site-directed mutagenesis of the Nidovirus replicative endoribonuclease NendoU exerts pleiotropic effects on the arterivirus life cycle. J.Virol. 2006;80(4):1653-1661.

32. Zust R, Cervantes-Barragan L, Habjan M, *et al.* Ribose 2'-O-methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. Nat. Immunol. 2011;12(2):137-143.

33. Chen Z, Faaberg KS, Plagemann PG. Determination of the 5' end of the lactate dehydrogenase-elevating virus genome by two independent approaches. J.Gen.Virol. 1994;75 (Pt 4):925-930.

34. Fujimura T, Esteban R. Cap-snatching mechanism in yeast L-A double-stranded RNA virus. Proc. Natl.Acad.Sci.U.S.A 2011;108(43):17667-17671.

35. Mir MA, Duran WA, Hjelle BL, *et al.* Storage of cellular 5' mRNA caps in P bodies for viral cap-snatching. Proc.Natl.Acad.Sci.U.S.A 2008;105(49):19294-19299.

36. Reich S, Guilligay D, Pflug A, *et al.* Structural insight into cap-snatching and RNA synthesis by influenza polymerase. Nature 2014;516(7531):361-366

37. Benson DA, Cavanaugh M, Clark K, *et al.* GenBank. Nucleic Acids Res. 2013;41(Database issue): D36-D42.

38. Pruitt KD, Brown GR, Hiatt SM, *et al.* RefSeq: an update on mammalian reference sequences. Nucleic Acids Res. 2014;42(Database issue):D756-763.

39. Gorbalenya AE, Lieutaud P, Harris MR, *et al.* Practical application of bioinformatics by the multidisciplinary VIZIER consortium. Antiviral Res. 2010;87(2):95-110.

40. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011;39(Web Server issue):W29-37.

41. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792-1797.

42. Larkin MA, Blackshields G, Brown NP, *et al.* Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947-2948.

43. Lauber C, Gorbalenya AE. Partitioning the genetic diversity of a virus family: approach and evaluation through a case study of picornaviruses. J.Virol. 2012;86(7):3890-3904.

44. Sidorov IA, Reshetov DA, Gorbalenya AE. SNAD: Sequence Name Annotation-based Designer. BMC.Bioinformatics. 2009;10:251.

45. Soding J. Protein homology detection by HMM-HMM comparison. Bioinformatics. 2005;21(7): 951-960.

46. Remmert M, Biegert A, Hauser A, *et al.* HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat.Methods 2012;9(2):173-175.

47. Grant BJ, Rodrigues AP, ElSawy KM, *et al.* Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics. 2006;22(21):2695-2696.

48. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proc.Natl.Acad. Sci.U.S.A 1992;89(22):10915-10919.

49. Finn RD, Bateman A, Clements J, *et al.* Pfam: the protein families database. Nucleic Acids Res. 2014;42(Database issue):D222-230.

50. Cole C, Barber JD, Barton GJ. The Jpred 3 secondary structure prediction server. Nucleic Acids Res. 2008;36(Web Server issue):W197-201.

51. Buchan DW, Minneci F, Nugent TC, *et al.* Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res. 2013;41(Web Server issue):W349-357.

52. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. 2014;42(Web Server issue):W320-324.

53. Guindon S, Dufayard JF, Lefort V, *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst.Biol. 2010;59(3):307-321.

54. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004;20(2):289-290.

55. R Development Core Team. R: A language and environment for statistical computing. 2011. Vienna, Austria, R Foundation for statistical computing.

56. van den Born E, Gultyaev AP, Snijder EJ. Secondary structure and function of the 5'-proximal region of the equine arteritis virus RNA genome. RNA. 2004;10(3):424-437.

57. van Dinten LC, Den Boon JA, Wassenaar AL, *et al.* An infectious arterivirus cDNA clone: identification of a replicase point mutation that abolishes discontinuous mRNA transcription. Proc.Natl. Acad.Sci.U.S.A 1997;94(3):991-996.

58. Nedialkova DD, Gorbalenya AE, Snijder EJ. Arterivirus Nsp1 modulates the accumulation of minus-strand templates to control the relative abundance of viral mRNAs. PLoS.Pathog. 2010; 6(2):e1000772.

59. van der Meer Y, van TH, Locker JK, *et al.* ORF1a-encoded replicase subunits are involved in the membrane association of the arterivirus replication complex. J.Virol. 1998;72(8):6689-6698.

60. Gohara DW, Ha CS, Kumar S, *et al.* Production of "authentic" poliovirus RNA-dependent RNA polymerase (3D(pol)) by ubiquitin-protease-mediated cleavage in Escherichia coli. Protein Expr. Purif. 1999;17(1):128-138.

61. Lauber C, Goeman JJ, Parquet MC, *et al.* The footprint of genome architecture in the largest genome expansion in RNA viruses. PLoS.Pathog. 2013;9(7):e1003500.