OXFORD

Unveiling the structural and functional implications of uncharacterized NSPs and variations in the molecular toolkit across arteriviruses

Siuli Rakesh [ID], Kshitij Behera, Arunkumar Krishnan [ID] *

Department of Biological Sciences, Indian Institute of Science Education and Research Berhampur (IISER Berhampur), Berhampur 760010, India

*To whom correspondence should be addressed. Email: akrishnan@iiserbpr.ac.in

Abstract

Despite considerable scrutiny of mammalian arterivirus genomes, their genomic architecture remains incomplete, with several unannotated non-structural proteins (NSPs) and the enigmatic absence of methyltransferase (MTase) domains. Additionally, the host range of arteriviruses has expanded to include seven newly sequenced genomes from non-mammalian hosts, which remain largely unannotated and await detailed comparisons alongside mammalian isolates. Utilizing comparative genomics approaches and comprehensive sequence-structure analysis, we provide enhanced genomic architecture and annotations for arterivirus genomes. We identified the previously unannotated C-terminal domain of NSP3 as a winged helix-turn-helix domain and classified NSP7 as a new small β-barrel domain, both likely involved in interactions with viral RNA. NSP12 is identified as a derived variant of the N7-MTase-like Rossmann fold domain that retains core structural alignment with N7-MTases in Nidovirales but likely lacks enzymatic functionality due to the erosion of catalytic residues, indicating a unique role specific to mammalian arteriviruses. In contrast, non-mammalian arteriviruses sporadically retain a 2′-O-MTase and an exonuclease (ExoN) domain, which are typically absent in mammalian arteriviruses, highlighting contrasting evolutionary trends and variations in their molecular toolkit. Similar lineage-specific patterns are observed in the diversification of papain-like proteases and structural proteins. Overall, the study extends our knowledge of arterivirus genomic diversity and evolution.

Introduction

Arteriviruses are positive-sense, single-stranded RNA viruses with relatively small genomes and distinct virion structures, belonging to the order Nidovirales [1–3]. They predominantly infect boreoeutherian mammals such as horses, pigs, rodents, and non-human primates, causing critical illnesses like hemorrhagic fever and severe respiratory diseases, particularly in porcine and equine hosts [4–6]. Their high virulence and contagious nature pose significant challenges to the porcine and equine industries, leading to extensive research into their various molecular aspects to identify druggable targets and develop effective control measures. Within the Nidovirales order, arteriviruses have followed a distinct evolutionary path, establishing themselves as one of the four primary families alongside Mesoniviridae, Roniviridae, and Coronaviridae [3, 7, 8]. This unique trajectory positions arteriviruses as an essential model for investigating the broader evolutionary patterns, especially in comparison to other Nidovirale families.

Arterivirus genomes typically range from 13 to 15 kb in size, consisting of 10–15 Open Reading Frames (ORFs), with the two longest ORFs, ORF1a and ORF1b, positioned sequentially from the 5′ end and encoding a series of non-structural proteins (NSPs) that are crucial for viral genome replication. These ORFs are translated into two large replicase polyproteins, polyprotein1a (pp1a) and polyprotein1b (pp1b). Current knowledge suggests that the genomic structure of ORF1a is characterized by one or more papain-like proteases (PLPs), followed by a series of transmembrane (TM) segments and a 3C-like protease (3CL-Pro). Upon translation, PLPs process the ORF1a-encoded pp1a polyprotein into seven NSPs and thereby facilitate the maturation and assembly of the replication-transcription complex (RTC) [9–11]. While PLPs process ORF1a into seven NSPs, the 3CL-pro (NSP4) encoded in ORF1a primarily processes the downstream NSPs within ORF1b into four NSPs (NSP9–NSP12). These NSPs are key for RNA replication and processing [12, 13] and include the following: the RNA-dependent RNA polymerase (RdRp) in NSP9 [14], a zinc-binding and helicase domain in NSP10 [15], an endoribonuclease domain (NendoU) in NSP11 [16], and finally, NSP12, which encodes a globular domain [17], whose structural and functional roles remain largely unexplored. Further downstream of ORF1b, a variable number of shorter ORFs encode several structural proteins, including viral membrane (M) proteins, nucleocapsid (N) proteins, and multiple envelope glycoproteins (GP) that aid in virion assembly and release from the host cell [18, 19]. These minor ORFs exhibit considerable variation in both number and sequence-structure synapomorphies, with some likely arising from ancient duplication events and subsequently following distinct evolutionary trajectories shaped by viral-host coevolution [20], contributing to the overall diversity in arterivirus genome size.

Despite notable research progress on arteriviruses from mammalian hosts (hereafter mammalian arteriviruses), there are still significant gaps in our understanding of their complete genomic architecture and protein products. While many NSPs and their constituent domains have been studied, others—such as NSP3C (the C-terminal end of NSP3), NSP7 (within ORF1a), and NSP12 (the C-terminal end of ORF1b)—remain poorly understood and insufficiently characterized. This lack of detailed structural and functional information on these elusive ORFs impedes a complete understanding of arterivirus genomic architecture and their functional implications. Moreover, in 2018, Shi Mang *et al.* published 214 vertebrate-associated RNA viral genomes from a wide range of vertebrate hosts, including several from non-mammalian sources, for the first time. These included six different arterivirus genomes: one from ray-finned fish, one from cartilaginous fish, and four from reptilian (three snakes and one turtle) hosts [21], all of which remain unannotated and largely unstudied. These genomes, along with an independently sequenced arterivirus genome from a reptilian host (Chinese softshell turtle; *Trionyx sinensis* hemorrhagic syndrome arterivirus—TSHSA) [22], formed the core set of seven arterivirus genomes from non-mammalian hosts (referred to as non-mammalian arteriviruses hereafter) that we analyzed and compared with mammalian arteriviruses for the first time in our study. For mammalian arteriviruses, we compiled a set of 22 full-length arterivirus genomes from the NCBI RefSeq database, covering all major mammalian host groups (primates, porcine, rodents, equine, and marsupials). This inclusive set of arteriviruses from both mammalian and non-mammalian hosts (a total of 29 genomes) enabled us to analyze complete protein domain compositions, annotate previously uncharacterized segments, and perform in-depth comparative genomics analysis to illuminate broader aspects of arterivirus evolution and diversification.

In this study, we provide a detailed comparative genomic analysis of arteriviruses, uncovering key evolutionary trends and molecular toolkits that subtly distinguish mammalian and non-mammalian arteriviruses. By annotating previously uncharacterized domains, we identify genomic similarities and differences that likely influence viral replication, immune evasion, and host adaptation. Our findings show that non-mammalian arteriviruses, despite their diverse host range, share certain genomic traits, such as the presence of 2′-O-MTase and ExoN domains, which are absent in mammalian arteriviruses. In contrast, mammalian arteriviruses uniquely encode a conserved but catalytically inactive N7-MTase-like domain within NSP12 of ORF1b, which adopts an atypical Rossmann fold. Furthermore, we provide the first characterization of NSP3C and NSP7 within ORF1a, linking them to known nucleic acid binding domains. Our analysis also reveals host-specific divergences in glycoproteins across arteriviruses. Overall, through comparative genomics and sequence-structure analysis, we classify several previously overlooked functional domains, enhancing our understanding of arterivirus genomic architecture and evolution.

Materials and methods

Genomic datasets and sequence analysis

To enable a comprehensive comparison of arterivirus genomes from both mammalian and non-mammalian hosts, we first curated a dataset of 22 complete mammalian arterivirus genomes from the NCBI RefSeq database, excluding any not listed in RefSeq. These genomes encompass a range of boreoeutherian mammals, including both primates and non-primates. Alongside the mammalian genomes, we incorporated all currently available non-mammalian arterivirus genome sequences. This includes a total of seven genomes from various hosts such as snakes, turtles, ray-finned fish (Japanese halfbeak), and cartilaginous fish (ghost shark). All encoded protein FASTA sequences from these 29 genomes were obtained and clustered using the BLASTCLUST program (ftp://ftp.ncbi.nih.gov/blast/documents/blastclust.html) (RRID:SCR_016641; version 2.2.26). The parameters, including the length of pairwise alignments (L, 0.2–0.5) and the bit-score (S, 20–50), were adjusted to achieve the desired level of clustering by empirically modifying the alignment length and bit-score density threshold. Divergent sequences or smaller clusters were merged with larger ones when supported by supplementary evidence, including shared sequence motifs, structural synapomorphies, reciprocal BLAST search results, and/or associations in the genome context. To accurately define the encoded protein domains and their boundaries for each ORF, sequences from all analyzed genomes were subjected to sensitive profile–profile searches using HH-PRED against hidden Markov model (HMM) profiles derived from the Protein Data Bank (PDB) [23] and Pfam models [24]. For each query or seed sequence analyzed through HH-PRED, HHblits was utilized with default parameters (-e value 0.001, -number of iterations 2, and -minimum probability for inclusion 20%) to retrieve homologs from the UniRef30 database, subsequently creating three iterations of multiple sequence alignments (MSA) and corresponding HMM models for profile–profile comparison [25–27]. Additionally, RPS-BLAST [28] with default parameters was used to perform searches against our custom in-house database of diverse domains. MSAs for all domains analyzed in this study were generated using MAFFT [29] or Kalign (V3) [30], with manual adjustments based on profile–profile searches against PDB, structural alignments, and predicted 3D structural models (see below). Transmembrane regions were predicted using Deep-TMHMM with default parameters, applying a minimum probability threshold of 0.5 [31].

Structure analysis

3D structure predictions for all individual domains analyzed in this study were generated using AF3 [32]. Each predicted structure was then subjected to structural similarity searches with the DALI-lite program (RRID:SCR_003047) [33] against the PDB database clustered at 75% identity. The reference MSAs constructed for each analyzed domain were used to predict secondary structure topologies with the JPred (V4) program (RRID:SCR_016504) [34], and the predicted secondary structure boundaries from JPred4 were cross-validated with the AF3 models. Structural homology among domains was evaluated based on DALI Z-scores and 3D structural superimposition. A DALI Z-score of 3 and above were set as a minimum threshold, followed by manual structural and topological comparison to validate their homology. Structures were rendered, compared, and superimposed using PyMOL (https://www.pymol.org/) (RRID:SCR_000305).

Comparative genomics and phylogenetic analysis

Structural glycoproteins of arteriviruses from both mammalian and non-mammalian hosts were clustered by calcu-

lating all-against-all pairwise sequence similarities using the CLANS software [35]. Host-specific divergence and phylogenetic relationships of each structural glycoprotein were inferred with an approximate maximum likelihood (ML) approach in the FastTree program (RRID:SCR_015501) [36], with local support values estimated accordingly. To enhance the topology's accuracy, the number of minimum-evolution subtree-prune-regraft (SPR) rounds in FastTree was set to 4 (-spr 4), and the options "-mlacc" and "-slownni" were applied for more exhaustive ML nearest-neighbor interchanges. Phylogenetic tree topologies were also generated using ML methods based on the edge-linked partition model in the IQ-TREE software [37, 38], with branch support obtained via the ultrafast bootstrap method (1000 replicates) in IQ-TREE [39]. The FigTree program (RRID:SCR_008515) (http://tree.bio.ed.ac.uk/software/figtree/) was used to render phylogenetic trees.

Results and discussion

Evolutionary diversification of N-terminal PLPs and distinguishing features of newly characterized domains in ORF1a

In mammalian arteriviruses, the N-terminal region of ORF1a is characterized by several PLPs, with NSP1 encoding two to three of these proteases—PLP1α, PLP1β, and PLP1γ (Fig. 1A). PLP1α and PLP1β are conserved across all mammalian arteriviruses, while PLP1γ appears to be specific to primate arteriviruses. In addition, NSP2 encodes a conserved PLP, known as PLP2, which is recognized for its deubiquitinase activity and plays a key role in disrupting host immune signaling [11, 40, 41]. These PLP paralogs were earlier implied to have likely emerged via early duplication events followed by adaptive divergence, leading to their structural and functional variations [41]. In non-mammalian arteriviruses, we found a reduction in the number of PLPs within ORF1a compared to their mammalian counterparts (Fig. 1B). Our structure-informed searches could not detect the presence of PLP1β or PLP1γ, suggesting that they are likely absent, while PLP1α is exclusively identified in the genome of Hainan *Oligodon formosanus* arterivirus (HOFA—Reptilia). Notably, PLP2 was found in all analyzed non-mammalian arterivirus genomes except for Nanhai ghost shark arterivirus (NGSA—Chondrichthyes) and Hainan *Hebius popei* arterivirus (HHPA—Reptilia) (Fig. 1B). Sequence-structure analysis reveals that these non-mammalian PLPs, particularly PLP2, subtly differ from their mammalian analogs, exhibiting slightly distinct features. Nevertheless, all five PLP2s share a conserved catalytic motif, DGxCGhH [41], in the first α-helix, as well as a conserved catalytic histidine residue in the fourth β-strand. These features are consistently present across all analyzed arterivirus PLP2s (Supplementary Data S1 and S2).

Though no additional PLPs were found, most non-mammalian arteriviruses possess at least one α + β globular domain, typically located at the N-terminus or in genomic locations analogous to NSP1/NSP2-encoded domains found in mammalian arteriviruses (Fig. 1B and Supplementary Fig. S1). Predictions from AlphaFold3 (AF3) suggest an $\alpha\beta\alpha$-sandwich-like architecture for most of these domains; however, DALI structure similarity searches revealed no clear homology to known proteins. Besides these segments, several of these genomes encode highly divergent intervening regions within the ORF1a, typically 120–300 amino acids, with some

extending beyond 400–600 residues, as in TSHSA (Fig. 1B). Sequence and structural analysis revealed no associations with PLPs or other functional domains, with most regions being unalignable. AF3 predictions indicate these regions are predominantly disordered, with only a few minor segments demonstrating ordered secondary structure elements such as beta-hairpins and helical bundles embedded within the largely disordered structure. Additionally, no transmembrane segments were identified by Deep-TMHMM in these regions. For now, we provisionally designate these regions as probable DIVs, though their characterization and potential functional implications may become clearer with improved genome assemblies. (Fig. 1B and Supplementary Fig. S2).

Besides these, ORF1a encodes several transmembrane segments within NSP2 and NSP3, and the C-terminal end of NSP3 (NSP3C) contains a globular domain that we have now categorized as a wHTH domain (detailed in the next section) (Fig. 1A). Following NSP3, NSP4 encodes the conserved 3CL-pro in mammalian arteriviruses, and these are also consistently retained in similar genomic locations within ORF1a in non-mammalian arteriviruses, preserving their key sequence and structural features (Fig. 1A and B). Non-mammalian arteriviruses also include extra transmembrane segments adjacent to 3CL-pro, mirroring those found in mammalian arteriviruses. At the end of the C-terminal region, ORF1a in mammalian arteriviruses contains NSP7, which we have demonstrated to encode a small β-barrel (SBB) domain (see next sections). Beyond ORF1a, Fig. 1 illustrates the complete genomic architecture of arteriviruses, emphasizing all newly identified domains and offering a comprehensive comparative view of the genomic structure in both mammalian and non-mammalian arteriviruses (Supplementary Table S1).

NSP3C is a wHTH domain that likely interacts with the viral RNA

Our analysis reveals that the previously unannotated C-terminal domain of NSP3 in arteriviruses is a conserved wHTH domain featuring a tri-helical core and two antiparallel strands that form the nucleic acid-binding interface [42, 43] (Fig. 2A). This wHTH module is present across all mammalian arteriviruses as well as at least three non-mammalian arterivirus genomes (Fig. 1). Comparative analysis across the Nidovirales reveals that arterivirus NSP3 is homologous to coronavirus NSP4, with both encoding multiple transmembrane segments (typically 4–6) followed by a C-terminal wHTH domain [44, 45]. DALI searches using crystal structures of coronavirus NSP4C and AF3-predicted arterivirus NSP3C recovered several canonical wHTH modules (DALI Z-scores $\geq$ 3), including GnTR-wHTH and MarR-wHTH, highlighting their structural overlaps with known wHTH domains. Likewise, DALI searches also ranked coronavirus NSP4C among the top hits for arterivirus NSP3C, reinforcing their shared structural features (Supplementary Data S3). The core structures of NSP3C wHTH in arteriviruses and NSP4C wHTH in coronaviruses share a conserved βS1-αH1-αH2-αH3-βS2 arrangement, with minimal structural deviation observed upon superimposition (Fig. 2A). Previous studies on coronaviruses have shown that the TM segments of NSP3, along with other TM regions of ORF1a, contribute to the formation of double-membrane vesicles (DMVs) that act as scaffolds for the RTC [44–46]. In coronaviruses, the TM regions of NSP4 are embedded in the ER-derived membranes of DMVs, positioning the C-terminal wHTH domain near the
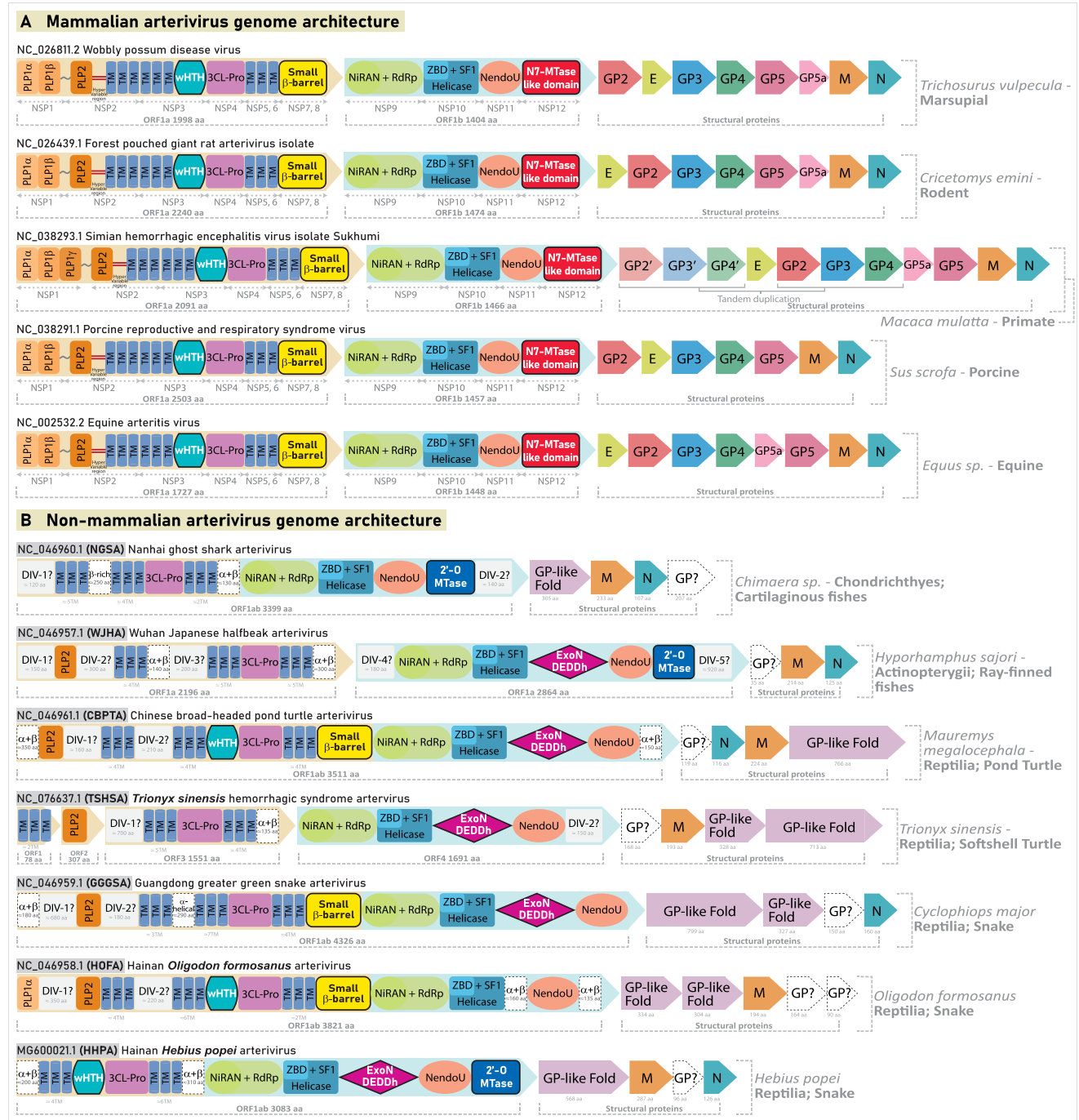
Figure 1. Genomic architecture and proteome composition of arterivirus genomes. (**A**) Genomic structure and protein domain organization of representative mammalian arteriviruses. ORFs are depicted as colored arrows, with their encoded domains labeled and highlighted. The color scheme is consistently applied across the figure for homologous regions and domains. Newly identified functional domains—the C-terminal winged helix-turn-helix (wHTH) domain of NSP3, the small β-barrel (SBB) domain of NSP7, and the N7-MTase-like domain of NSP12—are outlined with black borders. (**B**) Genomic structure and protein domain organization of arterivirus genomes from non-mammalian hosts. In cases where ORF1a and ORF1b regions are merged into a single, larger ORF [as observed in NGSA, Guangdong greater green snake arterivirus (GGGSA), HOFA, HHPA, and Chinese broad-headed pond turtle arterivirus (CBPTA) genomes] or split into multiple shorter ORFs (as in TSHSA genome), the corresponding regions of ORF1a and ORF1b are highlighted using the same color scheme as their mammalian counterparts (light brown background for ORF1a and light blue for ORF1b) to aid easier comparison. Highly divergent genomic regions predicted to be largely disordered with few minor secondary structural elements are provisionally marked as "disordered intervening regions (DIVs)?" with gray background. Likely functional domains with predicted secondary structural elements and ordered structures, as inferred from AF3, but not yet annotated in ORF1a and ORF1b due to a lack of homology with known domains, are displayed with a white background and dotted borders. Similarly, GP-like domains that lack clear structural homology for precise annotation are labeled as GP? and shown with a white background and dotted borders. Newly identified and classified functional domains, including the wHTH, SBB, 2'-O-MTase, and ExoN domains, are outlined in black. Each genome shown in both panels is labeled with its NCBI RefSeq accession number, source organism name, and common name. Abbreviations are provided for non-mammalian genomes for easy usage in the manuscript.
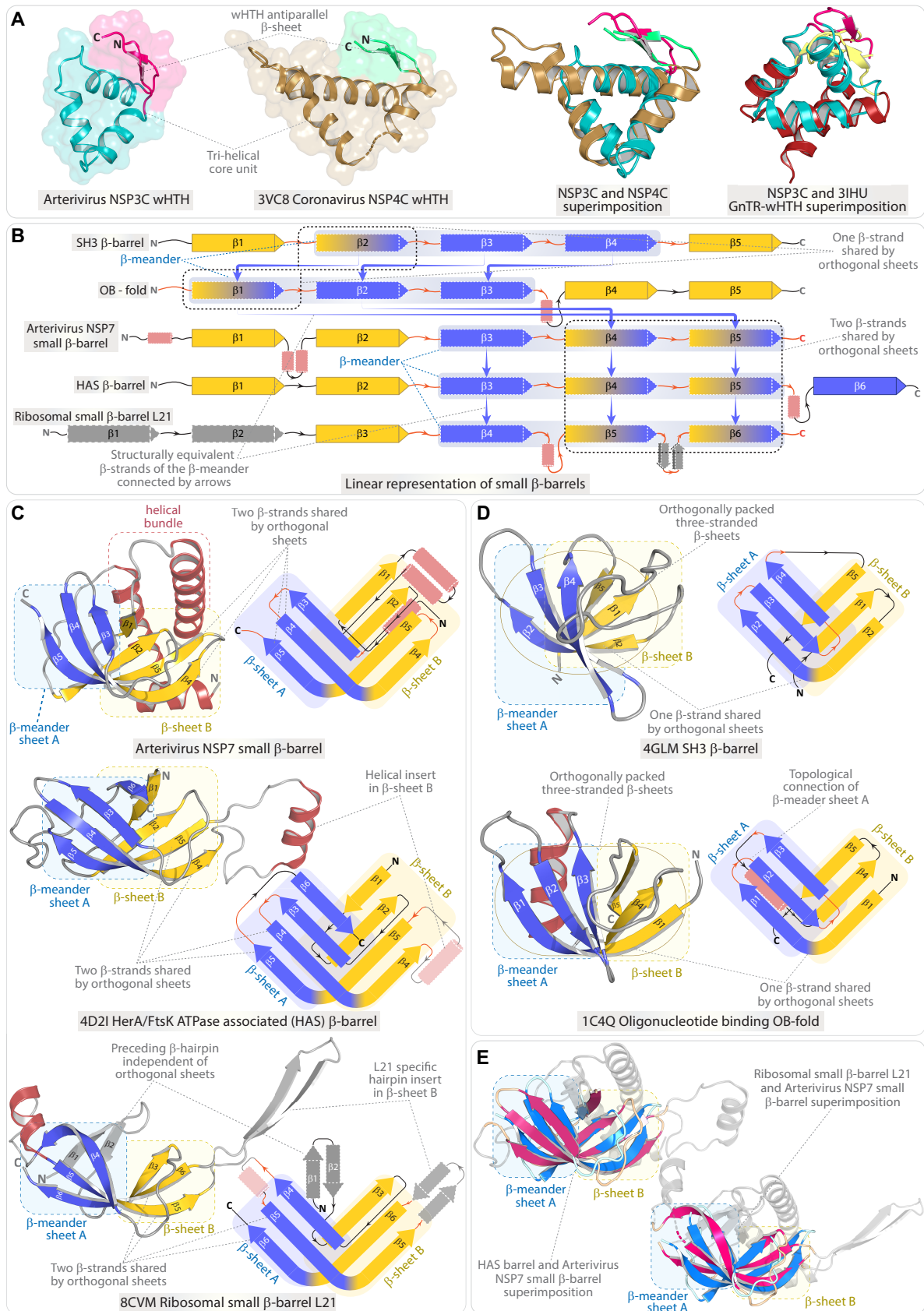
Figure 2. Comparative analysis and structural characteristics of the arterivirus NSP3C wHTH and NSP7 SBB domains. (**A**) The first two 3D structures in panel (A) illustrate the AF3-predicted C-terminal wHTH domain of NSP3 (NSP3C) from arteriviruses, alongside the experimentally resolved and corresponding wHTH domain in the C-terminal region of NSP4 (NSP4C) from coronaviruses (PDB ID: 3VC8). The structural superpositions on the right display the alignment of NSP3C with NSP4C, as well as NSP3C with the canonical wHTH domain in GnTRs (PDB ID: 3IHU). Panel (**B**) displays a linear

viral RNA to facilitate interactions. This is further supported by STED microscopy, which reveals that the NSP4 C-terminus, along with NSP7, NSP8, and NSP9, is consistently closer to the viral RNA, while the NSP3 N-terminus is positioned further away [47, 48]. Based on these findings, we propose that arterivirus NSP3C, a homolog of coronavirus NSP4C, with its TM segments and C-terminal wHTH domain, likely serves a similar function, with the wHTH domain potentially interacting with the viral RNA.

NSP7: a new member of the SBB assemblage with potential roles in viral RNA interaction

NSP7, located at the 3′-end of ORF1a, encodes a globular domain whose function remains poorly understood. Similar to the NSP3C module, NSP7 is conserved across all mammalian arteriviruses, with homologs also identified in at least three non-mammalian arteriviruses at comparable genomic locations (Fig. 1). Although the 3D structure of NSP7 has been resolved by NMR, the encoded domain remains unannotated, and its function in arteriviruses is unclear, with studies suggesting that it may represent a novel fold of unknown function [49]. Using AF3 models, we reveal that all identified NSP7 domains in arteriviruses (Fig. 1B) consistently maintain a SBB-like structural scaffold composed of five core β-strands. These strands fold into two "nearly" orthogonal and flexible β-sheets, forming a SBB domain. Topologically, NSP7 is organized as αH1-βS1-(αH2-αH3)-βS2-βS3-βS4-βS5, with the N-terminal αH1 extension and the bihelical hairpin insert following βS1 forming a helical bundle that is spatially separated and positioned behind the five-stranded β-barrel (Fig. 2C). Structurally, the core β-barrel of NSP7 affiliates with those of previously classified SBB domains [50–52] (Fig. 2B–D). These SBBs include several distinct yet closely related protein folds, which can be grouped together at a broader "superfold" level based on their shared structural framework, geometric similarities, and roles in nucleic acid binding and protein interactions [52, 53].

Using DALI-based searches, we found that NSP7 shares close structural similarities with several SBB members (DALI Z-scores ≥ 4.5), including the HAS-barrel domains of HerA/FtsK ATPases, ribosomal protein β-barrels, and the OB-fold domains (Supplementary Data S3). Like NSP7, these SBB domains typically feature a five- or occasionally six-stranded β-barrel, structurally divided into two orthogonally packed β-sheets that share at least one, and sometimes two, β-strands (Fig. 2B–D) [52]. The shared strand(s), subtly elongated, exhibit(s) a signature curvature with a kink in the β-strand, typically introduced by a conserved glycine or, occasionally, a proline. This feature defines the shared strand(s) as a "structural signature" in SBBs, with each half (N and C-termini) forming stabilizing interactions with β-strands of both orthogonally packed β-sheets, facilitating the forma-

tion of a semi-open barrel-like structure (Fig. 2B–D) [52]. For example, in a typical SH3-like SBB, βS2 is shared between the two sheets (Fig. 2D). Additional hallmarks of SBB-like domains include β-sheet A, which is sequentially and spatially contiguous, forming a characteristic β-meander, while β-sheet B is non-contiguous (e.g. in SH3, β-sheet B consists of βS1, the shared βS2, and βS5). The classification of the SBB group is anchored in the topological connections and structural arrangements of these orthogonally packed sheets, with each fold exhibiting distinct topological features and subtle secondary structure insertions while preserving the broader framework of a SBB [52].

A detailed comparison of NSP7 with established SBB members highlights its distinct structural relationships. While its three-stranded β-meander, topological connections, and β-strand orientations set it apart from the OB and SH3-like fold, NSP7 closely aligns with the HAS-barrel domain and the SBB of the 50S ribosomal subunit L21 (Fig. 2B and C). In all three domains, the β-meander is structurally and topologically equivalent, and they share two β-strands between their orthogonal β-sheets—a key distinction from the OB and SH3-like folds, which share only a single strand (Fig. 2B and C). Indeed, the core of the HAS-barrel domain and L21 closely superimpose with NSP7, showing an RMSD below 4 Å (Fig. 2E). A subtle distinction lies in βS1 orientation: in NSP7, it runs parallel to βS2, whereas, in the HAS-barrel, it is antiparallel, though the rest of the core barrel remains identical (Fig. 2C). Despite their core structural overlap, these three domains also possess unique insert regions and N-terminal/C-terminal extensions (Fig. 2C). NSP7 features an N-terminal helical bundle, while L21 usually has an N-terminal β-hairpin slightly tilted away from the core sheets and another β-hairpin insert extending from the shared core β-strands βS5 and βS6 of the β-barrel. In contrast, the HAS-barrel lacks both the N-terminal helical bundle and β-hairpin inserts but usually has an additional strand, βS6, as a C-terminal extension, stacked antiparallel to βS3 through a long loop extending from βS5. The HAS barrel also has a helical insert in the same position as the β-hairpin insert between βS5 and βS6 in L21 (Fig. 2B and C).

While earlier studies misclassified the L21 ribosomal protein as an SH3-like SBB domain [53], our comparative analysis elucidates the structural and topological distinctions between SH3-like β-barrels and the L21 ribosomal protein, as well as its structural homologs—the HAS-barrel and arterivirus NSP7 (Fig. 2B and D). Our findings show that NSP7 and its closest structural homologs—L21 and the HAS-barrel—share the same β-meander and core β-barrel structure while also retaining distinct, domain-specific insert regions (Fig. 2C). Together, we classify NSP7 as a novel and definitive member of the broader SBB assemblage, marked by both conserved and unique features. Given that the HAS-barrel and the ri-

topological representation and comparison of the arterivirus NSP7 SBB domain with closely related SBB domains that, while representing distinct folds, share a broader structural framework. These domains are collectively classified within the SBB assemblage, with NSP7 being identified as a new member in this study. (**C**) 3D structure and topological comparisons of the arterivirus NSP7 SBB (AF3-predicted), HerA/FtsK ATPase-associated (HAS) β-barrel (PDB ID: 4D2I), and ribosomal SBB L21 (PDB ID: 8CVM). As shown in panel (B), these three domains are closely related and share two β-strands that are common between the nearly orthogonally oriented β-sheet A and β-sheet B, unlike the single shared strand seen in Src Homology 3 (SH3) and oligonucleotide-binding (OB) folds. β-strands forming β-sheet A (β-meander) and β-sheet B are highlighted in blue and yellow, with the shared strands (common to both sheets) colored accordingly for each half. The schematic topology, displayed alongside the 3D structure, is presented in a simplified format to precisely illustrate the nearly orthogonal β-sheets and the connectivity between the β-strands. (**D**) Representative 3D structures and schematic topology of the SH3-like β-barrel (PDB ID: 4GLM) and OB fold domains (PDB ID: 1C4Q). (**E**) Structural superposition of the arterivirus NSP7 SBB with (i) the HAS-barrel and (ii) the ribosomal SBB L21.

bosomal β-barrel L21 are known to stabilize large protein-protein or protein-nucleic acid complexes [53–56], we propose that NSP7 may play a role in viral RNA binding and(or) nucleoprotein complex stabilization, which may be important to the arterivirus replication cycle. Though current knowledge on specific pathways, localization, and interaction mechanisms remains limited, earlier works [57, 58] supports our inferences, suggesting that NSP7 is essential for interactions with both host and viral proteins and potentially plays a significant role in the overall assembly of the RTC. The classification of NSP7 as an SBB domain opens avenues for future investigations into its interactions with viral RNA and other proteins.

Comprehensive characterization of ORF1b—illustrating newly identified domains across mammalian and non-mammalian arteriviruses

Analyses of arterivirus genomes have shown that ORF1b encodes at least three conserved domains: NiRAN + RdRp, Zinc-binding + Helicase, and NendoU [12–16]. In mammalian arteriviruses, both the overall length and domain organization of ORF1b are conserved, featuring these three domains followed by the NSP12 domain, which is unique to mammalian arteriviruses (Fig. 1). Although NSP12 is recognized as encoding a globular domain [17], its structural and functional roles are not yet fully understood. The following sections explore NSP12 in greater detail. In contrast, non-mammalian arteriviruses display greater variability in ORF1b domain composition, including the sporadic presence of ExoN and a canonical 2′-O-MTase—both identified for the first time in this study. Additionally, the equivalent of ORF1b is sometimes fused with ORF1a, leading to inconsistencies in nomenclature and the number of ORFs for non-mammalian arteriviruses. To maintain clarity, we refer to these regions as ORF1b and highlight their relationship to mammalian arteriviruses, as illustrated in Fig. 1.

Identification of ExoN and canonical 2′-O-MTase domains in ORF1b of non-mammalian arteriviruses

ExoN domain: While coronaviruses and other Nidovirales encode an ExoN domain essential for proofreading and replication fidelity [59, 60], earlier studies indicated the absence of this domain in mammalian arteriviruses. Our comparative analysis reaffirms this finding in mammalian arteriviruses, but uncovers an ExoN domain in five of seven non-mammalian arterivirus genomes (Figs 1B and 3A). Structure similarity searches using AF3-predicted domains as search seeds in DALI recovered homologs from coronaviruses (PDB ID—7EGQ) and various other exonucleases (DALI Z-scores $\geq$ 10) from prokaryotes and eukaryotes. Sequence-structure comparisons with the recovered homologs demonstrated the retention of RNase-H-like fold and conservation of the five catalytic residues (DEDDH) that coordinate Mg^{2+} ions, suggesting that ExoN domains in non-mammalian arteriviruses are likely enzymatically active [61, 62] (Fig. 3A; Supplementary Data S2 and S3). Notably, unlike coronaviruses, where NSP14 universally includes both the ExoN domain and a guanine N7-MTase, non-mammalian arteriviruses completely lack an N7-MTase. Their ExoN domain is instead located between the helicase and NendoU domains, with no intervening regions that could potentially encode a functional N7-MTase (Fig. 1B).

2′-O-MTase domains: Although all non-arterivirus members/families of the Nidovirales possess a 2′-O-MTase domain, its presence in arteriviruses remains uncertain. A previous study identified MTase-like signatures in the HHPA and NGSA genomes [63], but it did not decisively validate their presence through detailed sequence-structure analysis. In this study, through AF3 predictions and structural comparisons, we identify the presence of 2′-O-MTase exclusively in three non-mammalian arterivirus genomes (Wuhan Japanese half-beak arterivirus—WJHA, HHPA, and NGSA) and found no evidence of its presence in mammalian counterparts. All 2′-O-MTase domains in arteriviruses retain the defining features as observed in the coronavirus NSP16-encoded SAM-dependent 2′-O-MTase, including the canonical Rossmann fold and key residues for SAM and m7GpppA binding [64], and these features are also conserved across all 2′-O-MTases in Nidovirales (Fig. 3B–D and Supplementary Data S2). Structurally, these domains retain the core 2′-O-MTase topology (αH1-βS1-αH2-βS2-βS3-βS4-αH3-βS5-αH4-βS6-βS7-αH5). From a comparative and evolutionary perspective, coronaviruses encode both the NSP14 N7-MTase and NSP16 2′-O-MTase, which together construct the 5′-end cap structure of viral mRNA—critical for evading host immune defenses and maintaining viral genome stability [65–70]. However, the retention of both methyltransferases is not a conserved feature across Nidovirales, as lineage-specific losses of MTases were observed earlier [63, 71].

Structural and functional insights into NSP12: an inactive N7-MTase-like domain exclusive to mammalian arteriviruses

The presence of an MTase-encoding NSP in mammalian arteriviruses remains uncertain, with studies reporting inconsistent findings regarding their mRNA capping mechanisms. While one report suggests the presence of methylated nucleosides [72], another study indicates that the mRNA remains uncapped [73]. Strengthening the case against MTase activity, NSP12—identified as a potential MTase candidate based on its genomic position compared to other nidoviruses and as the only unannotated domain in ORF1b—showed no enzymatic activity *in vitro*, even when tested with potential cofactors [17]. However, its suppression resulted in a complete loss of viral progeny, suggesting its essential role in proliferation [17]. These findings underscore the need for further analysis to determine whether NSP12 encodes an MTase-like function or serves a distinct role in viral proliferation.

To address this uncertainty, we analyzed AF3-predicted structures of NSP12 across arteriviruses and performed structural similarity assessments using DALI and HHPRED. Our results reveal that NSP12 consistently retains a conserved six-stranded β-sheet forming an atypical Rossmann fold MTase-like domain. However, the key catalytic features required for enzymatic activity are absent, reinforcing earlier observations that mammalian arteriviruses uniquely lack the mRNA capping machinery (Fig. 4A) [17]. Here, we present a detailed structural and sequence analysis of NSP12, examining its deviations from typical Rossmann fold MTases and the implications for its functional role.

In typical Rossmann fold MTases, the central β-sheet comprises seven β-strands arranged in the sequence: 6↑-7↓-5↑-4↑-1↑-2↑-3↑. Key features of this central β-sheet include: (i) a characteristic topological crossover where βS3 at the rightmost end leads to βS4 in the centre; (ii) a β-hairpin at the C-terminal end, formed by βS6 and βS7, with βS7 inserted
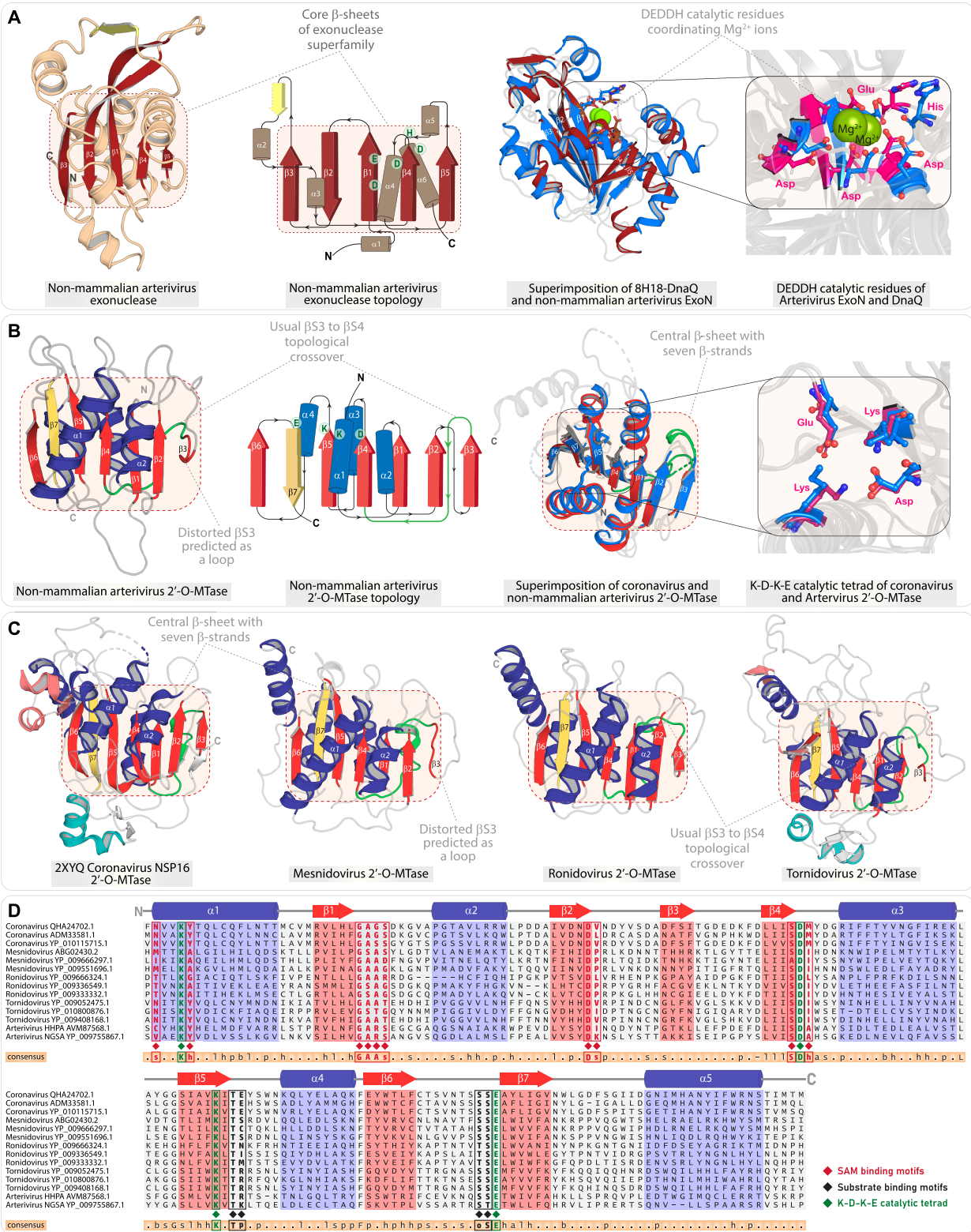
Figure 3. Conserved structural and sequence features of ExoN and 2′-O-MTase domains in non-mammalian arteriviruses. (**A**) The first two images in the panel display the representative AF3-predicted 3D structure and the corresponding schematic topology of the ExoN domain from non-mammalian arteriviruses. The structural superpositions on the right show the alignment of the arterivirus ExoN domain with the DnaQ exonuclease (PDB ID: 8H18), with a zoomed-in view (far right) highlighting the alignment of the DEDDH catalytic residues. (**B**) The first two illustrations in the panel depict the representative AF3-predicted 3D structure and the corresponding schematic topology of the 2′-O-MTase domain from non-mammalian arteriviruses. The structural superpositions on the right illustrate the alignment of the 2′-O-MTase domains from arteriviruses and coronaviruses (NSP16 2′-O-MTase), with a zoomed-in inset (far right) emphasizing the catalytic K-D-K-E tetrad. (**C**) Representative 2′-O-MTase (PDB: ID 2XYQ) from coronavirus, and AF3-predicted 3D structure of 2′-O-MTases from other Nidovirales. (**D**) Representative MSA of 2′-O-MTases across all Nidovirales, highlighting key conserved motifs for S-adenosyl methionine (SAM) binding, substrate binding, and the K-D-K-E catalytic tetrad.
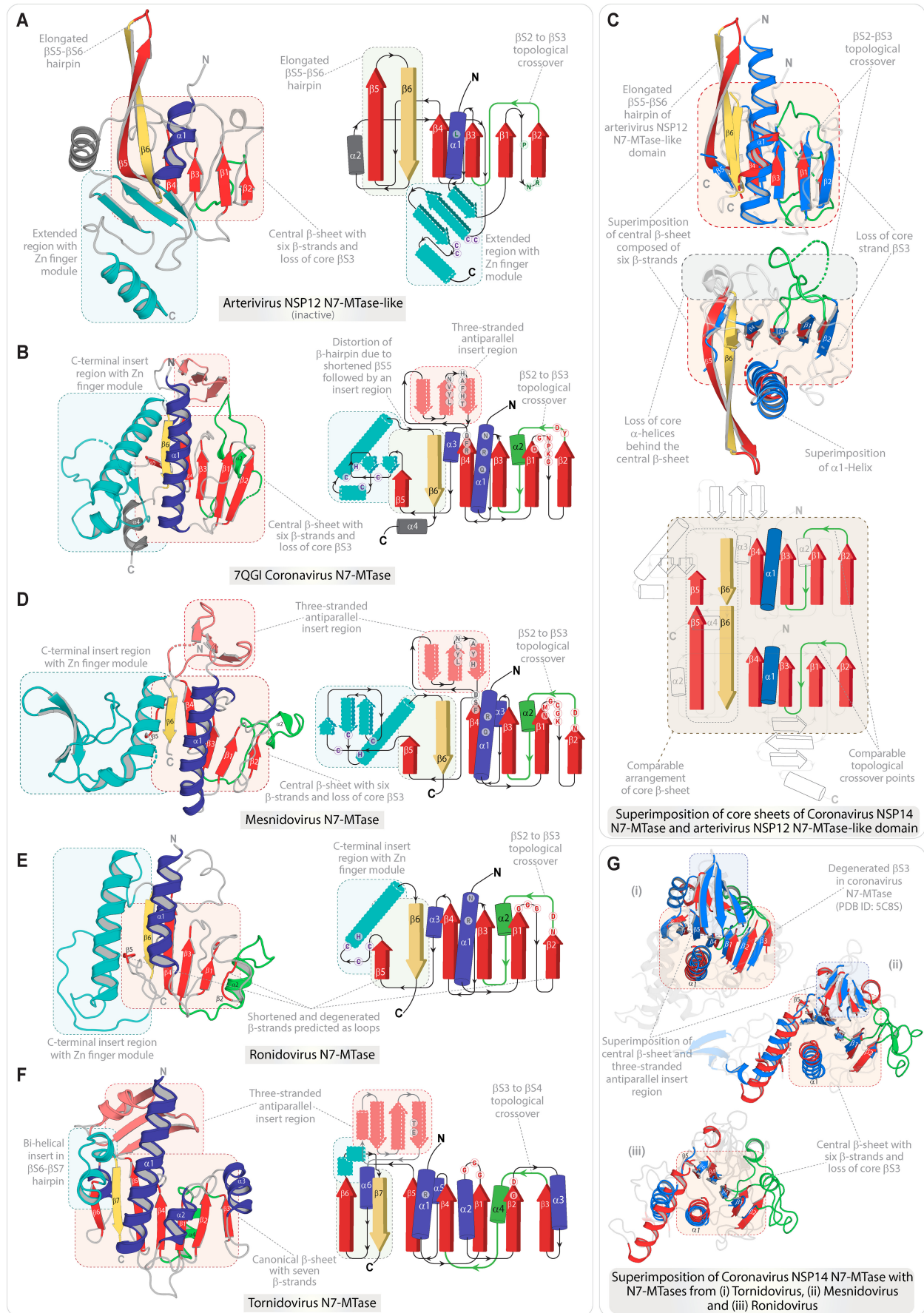
Figure 4. Structural features of the arterivirus NSP12 N7-MTase-like domain and its comparison with Nidovirale N7-MTases. Panels (**A**) and (**B**) display the 3D structure and corresponding topology of the arterivirus NSP12 N7-MTase-like domain (AF3 predicted) and the coronavirus NSP14 N7-MTases (PDB
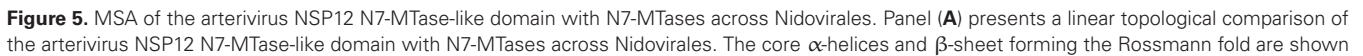
anti-parallel between βS5 and βS6, which serves as a key synapomorphy of Rossmann fold MTases [74–79]. In NSP12, the core sheet consists of six strands (5↑-6↓-4↑-3↑-1↑-2↑), with the loss of the typical βS3 causing βS2 to become the rightmost strand, and the topological crossover occurs from βS2 to βS3 at the centre (Fig. 4A). Aside from the absence of βS3, the core sheet of NSP12 precisely aligns with the central β-sheet architecture of canonical MTases, retaining all other structural synapomorphies, including the C-terminal β-hairpin. Besides the core region, NSP12 contains a C-terminal α-helix positioned parallel to the core sheet, followed by an extended loop that leads to a Zn finger (ZnF) module stacked in front of the core sheet (Fig. 4A). In some instances, a small β-strand insert is observed between αH1 and βS1, which is stacked together with the β-hairpin of the ZnF module, forming a three-stranded sheet that ends with a C-terminal helix (Fig. 4A).

Conserved core elements and unique variations relative to other Nidovirale N7-MTases: We conducted an extensive survey and comparisons to determine whether similar domains are present in other Nidovirales. Our structural and sequence analyses revealed several notable findings:

(i) At the structural level, we found that both the arterivirus NSP12 and coronavirus NSP14 N7-MTase domains exhibit the same structural framework, which can be summarized as a derived variant of the Rossmann fold. This is primarily defined by the absence of the core βS3 strand, resulting in a βS2-to-βS3 crossover and a core β-sheet composed of four parallel β-strands, followed by the C-terminal β-hairpin [80–85] (Fig. 4A–C). In some experimentally determined structures of coronavirus NSP14 N7-MTase (PDB IDs: 5C8S and 7N0D), the core β-sheet appears to include seven β-strands, but βS3 is small and degenerated, containing only two residues. While this technically forms the typical βS3-to-βS4 topological crossover, βS3 is not consistently resolved across all structures, and sequence alignment further supports its degeneration (Fig. 5A and B; Supplementary Fig. S3). Additionally, both domains show a reduction in core helices anterior to the core sheet, retaining only αH1, deviating from the typical and packed Rossmann fold αβα sandwich structure. Arterivirus NSP12 lacks helices behind the sheet, whereas coronavirus NSP14 retains two short helices (αH2 and αH3).

(ii) Expanding this analysis to other nidovirale clades, we observed that mesnidoviral and ronidoviral N7-MTases share the same core topology, featuring a six-stranded architecture, the absence of βS3, and a reduction in core helices, with only minimal deviations (see Fig. 4D and E and Table 1). In contrast, among all nidoviral N7-MTases, only the tornidovirus N7-MTase retains the canonical features of Rossmann fold MTases and is firmly established as one [63, 71]. Unlike other clades, it preserves the βS3, the characteristic βS3-to-βS4

crossover, and a core β-sheet with five parallel strands followed by a C-terminal β-hairpin (Fig. 4F and Table 1). Additionally, it maintains three helices in front of the core sheet and two short helices at the rear, making its structure the closest to a canonical Rossmann fold MTase (Fig. 4F). These findings highlight a clear structural divergence between nidoviral N7-MTases, where tornidovirus retain a near-canonical architecture, while other clades, including arteriviruses, coronaviruses, mesnidoviruses, and ronidoviruses, exhibit a derived atypical Rossmann fold with a reduced core topology and uncommon structural features.

(iii) Beyond these structural features, nidoviral N7-MTases are characterized by two key regions that, while positionally conserved, are only sporadically retained across different clades (Fig. 4A–G): (i) a ZnF module at the C-terminal end and (ii) a three-stranded antiparallel insertion extending outward from βS4 (βS5 in tornidovirus), forming a cap-like structure and a small pocket region above the core β-sheet. Previous studies have referred to this insertion as a "hinge" domain, which plays a role in the interaction between the N7-MTase and ExoN domains in the coronavirus N7-MTase + ExoN heterodimer, contributing additional residues for substrate binding [80–82, 86]. However, these insertions are not consistently present across Nidovirales. The ZnF module is found in arteriviruses (though not fully conserved), coronaviruses, mesnidoviruses, and ronidoviruses, whereas the three-stranded antiparallel β-sheet insert is present in all Nidovirales except arteriviruses and ronidoviruses (Fig. 4A–G). This suggests that the ZnF module may play a context-dependent, clade-specific role in RNA binding or protein-protein interactions, potentially dispensable in certain groups. Likewise, the three-stranded β-sheet insert does not appear to be a universally conserved feature across Nidovirales.

(iv) Each of these domains also exhibits subtle, clade-specific variations (Fig. 4 and Table 1). For instance, in NSP12, the βS5-βS6 hairpin is notably longer, spanning 12–13 amino acids, whereas, in the coronavirus N7-MTase, βS5 is shorter and followed by an insertion leading to an antiparallel βS6, typically six residues long (Fig. 4A and B). In mesnidoviral and ronidoviral N7-MTases, βS5 is often shorter or predicted as a loop in AF3 models. Additionally, the ZnF module at the C-terminal region shows clade-specific variations, making it not fully superimposable. Notably, arteriviruses chelate Zn^{2+} ions using C4-type zinc fingers, whereas coronaviruses, mesnidoviruses, and ronidoviruses display C3H-type zinc fingers (Figs 4 and 5 and Table 1).

At the sequence level, all Nidovirale N7-MTases, except for arterivirus NSP12, retain the key catalytic residues for SAM and substrate binding (Fig. 5B, Table 1 and Supplementary Data S4) [64, 80, 82, 87], though some require additional

ID: 7QGI). (**C**) The illustration on the top shows the structural superposition of the arterivirus NSP12 N7-MTase-like domain with the coronavirus NSP14 N7-MTase (front view), while the middle unit presents a top view, focusing primarily on the superposition of the core sheet and the anterior core αH1. The illustration at the bottom of the panel shows the topological comparison of the core β-sheet of the arterivirus NSP12 N7-MTase-like domain with the coronavirus NSP14 N7-MTase. The shared core sheet and the αH1 are colored. (**D-F**) 3D structure and corresponding topology diagrams of N7-MTases from mesnidovirus, ronidovirus, and tornidovirus. (**G**) Structural superposition of coronavirus NSP14 N7-MTase with N7-MTases from (i) tornidovirus, (ii) mesnidovirus, and (iii) ronidovirus.

Figure 5. MSA of the arterivirus NSP12 N7-MTase-like domain with N7-MTases across Nidovirales. Panel (**A**) presents a linear topological comparison of the arterivirus NSP12 N7-MTase-like domain with N7-MTases across Nidovirales. The core α-helices and β-sheet forming the Rossmann fold are shown

NSPs as co-factors for MTase activity [80, 81, 88–90]. In contrast, NSP12 from arterivirus lacks sequence conservation in the SAM-binding and substrate-binding sites, marking it as a unique, inactive MTase-like domain specific to mammalian arteriviruses (Fig. 5B). The absence of conserved and key catalytic residues indicates no enzymatic activity, and the key regions exhibit no significant selective pressure to retain these residues (Fig. 5B). The only conserved residues are a leucine located in the middle of αH1, a proline found in the center of the loop leading to βS2, and an intriguing asparagine-arginine dyad at the end of the loop just before the start of βS2 (Fig. 5B). However, these residues are positioned away from the typical SAM and substrate-binding sites, further supporting the notion that NSP12 does not function as a traditional MTase.

Together, we propose that arterivirus NSP12 encodes an atypical, inactive N7-MTase-like domain. While this inactive domain appears unique to arteriviruses within Nidovirales, similar domains are not entirely uncommon, and they exist in other viruses, such as the inactive MTase domain at the C-terminal end of NSP2 in alphaviruses [91–93] (Supplementary Fig. S4). Although catalytically inactive due to the loss of SAM-binding motifs, the alphavirus NSP2 MTase domain retains the canonical Rossmann fold. Previous studies suggested that it forms an RNA-binding interface with the N-terminal RNA helicase and protease domains of NSP2, as well as the downstream NSP3, which includes an ADP-ribose-binding macro domain and a zinc-binding domain [77, 93–95]. NSP2 has also been shown to degrade the RNA polymerase II subunit RPB1, thereby inhibiting host transcription and reducing the expression of interferon-stimulated genes (ISGs) [96, 97]. Recent studies also highlight the critical role of its C-terminal inactive MTase domain in immune evasion. This domain has been shown to selectively inhibit the JAK/STAT signaling pathway by promoting CRM-1 (chromosome region maintenance 1 export receptor)-mediated nuclear export of the STAT1 transcription factor, preventing ISG induction [97, 98]. This function remains independent of other NSP2 domains, emphasizing its specialized role in alphaviruses [98]. Moreover, prior studies suggest that the domain may act as a dominant-negative, interfering with host MTases (SETD2) to disrupt STAT1 methylation, which is essential for its phosphorylation and activation [98, 99]. While its precise mechanisms require further elucidation, the evidence suggests that this domain is retained for immune evasion in alphaviruses. Similarly, these findings raise the possibility that the retention of the inactive MTase domain in arterivirus NSP12, with its atypical Rossmann fold, serves a specialized function—potentially in immune evasion by modulating the host immune response. Alternatively, it may simply play a role in RNA binding within the viral replication machinery. The observation that NSP12 suppression inhibits viral proliferation further supports this hypothesis.

Divergence of structural proteins in arteriviruses

Though the C-terminal genomic organization and the number of ORFs are largely conserved among mammalian arteriviruses, our analysis reveals that the structural proteins they encode—particularly the envelope glycoproteins (GPs) and membrane (M) proteins—exhibit signatures of rapid evolution and considerable variability in sequence and structure. This trend also applies to structural proteins from non-mammalian arteriviruses, which demonstrate significant sequence diversity across different host species. Indeed, HH-PRED searches using these sequences and DALI searches utilizing AF3-predicted non-mammalian structural proteins found no reliable hits from other viruses. By analyzing datasets that include structural proteins from both mammalian and non-mammalian arteriviruses, we observed a strong pattern of host-specific divergence through CLANS-based clustering and phylogenetic analysis (Fig. 6). In mammalian arteriviruses, this divergence is particularly evident between primate and non-primate hosts, with structural proteins from these groups forming well-defined and distinct clusters. Phylogenetic trees for each of the analyzed structural proteins consistently validate this observation (Fig. 6 and Supplementary Data S5). Furthermore, studies have demonstrated that primate arteriviruses possess two copies each of GP2, GP3, and GP4, resulting from a duplication event, and the tree reconstruction shows that the ancestral and duplicate copies form distinct clusters [20, 100, 101]. The observed differences in sequence and AF3 predicted structures underscore previous findings that the ancestral and duplicated glycoproteins, unique to primate arteriviruses, fulfill independent and non-redundant functions [100].

Unlike mammalian arteriviruses, non-mammalian arteriviruses show no conservation in either genomic organization or the number of ORFs encoding structural proteins. While their genome sizes generally range from 13 to 15 kb, similar to their mammalian counterparts, some non-mammalian arterivirus genomes extend beyond this, reaching up to 18 kb (Fig. 1B). Notably, these include the arteriviruses from ray-finned fish—WJHA, GGGSA, and HHPA. Despite their large size, these viruses have fewer ORFs than mammalian arteriviruses. This reduction in ORF count is partly due to ORF fusion, such as the frequent merging of ORF1b with ORF1a, which leads to a decrease in the overall number of ORFs (Fig. 1B). We also predict that similar fusions have likely occurred in ORFs encoding structural proteins, as some ORFs are unusually long compared to those found in mammalian arteriviruses (Fig. 1B). For instance, certain GP-encoding ORFs from GGGSA, HHPA, CBPTA, and TSHSA are significantly larger, ranging between 568 and 799 amino acids. AF3 predictions suggest that these ORFs contain at least one well-defined GP-like domain, along with additional segments that may encode a second GP-like domain (Supplementary Fig. S5). However, many of these segments are too disordered to accurately define their domain

---

in blue and red, respectively, with the three-stranded insert in yellow and the C-terminal insert in teal. Secondary structural elements positioned outside of the αβα sandwich is colored gray. (**B**) Representative MSAs of N7-MTase domains from all five distinct Nidovirales families, with secondary structural elements colored as described. In the coronavirus N7-MTase, a short, inconsistently predicted βS3, which occasionally causes the structure to resemble and portray a seven-strand sheet, is highlighted with a yellow border. A consensus sequence is indicated at the bottom of each alignment. SAM-binding motifs, substrate-binding motifs, and Zn-chelating residues are highlighted in red, black, and magenta, respectively. Excluding the Zn-chelating residues, the SAM and substrate-binding (guanine-binding) residues exhibit erosion and lack conservation in arteriviruses. Arterivirus-specific conserved residues are highlighted in green.

Table 1. Synapomorphies and distinctive variations of N7-MTase-like domains in Nidovirales

|	Tornidovirus N7-MTase	Coronavirus N7-MTase	Mesnidovirus N7-MTase	Ronidovirus N7-MTase	Arterivirus NSP12 N7-MTase-like
Core topology, excluding inserts	αH1-βS1-αH2-βS2-αH3-βS3-αH4-βS4-αH5-βS5-αH6-βS6-βS7.	αH1-βS1-βS2-αH2-βS3-βS4-αH3-βS5-βS6.	αH1-βS1-βS2-αH2-βS3-βS4-αH3-βS5-βS6.	αH1-βS1-βS2-αH2-βS3-βS4-βS5-βS6.	αH1-βS1-βS2-βS3-βS4-βS5-βS6.
Genomic location	ORF1a [63, 71].	NSP14 of ORF1b along with an ExoN, followed by NSP15 NendoU and NSP16 2'-O-MTase domains [80–82].	C-terminal end of ORF1b, preceded by an ExoN and followed by 2'-O-MTase [7].	C-terminal end of ORF1b and followed by 2'-O-MTase [8].	NSP12 located at the C-terminal end of ORF1b, preceded by NSP11 NendoU domain [1].
Core helices of N7-MTases	Three helices, αH1, αH2, and αH3, are on the anterior side of central β-sheet; αH4, αH5, and αH6 are on the posterior side of central β-sheet.	Only one helix (αH1) on the anterior side of central β-sheet; lacks other helices on the anterior side; two short helices αH2 and αH3 on the posterior side of central β-sheet [80–82].	Only one helix (αH1) on the anterior side of central β-sheet; lacks other helices on the anterior side; two short helices αH2 and αH3 on the posterior side of central β-sheet.	Only one helix (αH1) on the anterior side of central β-sheet; lacks other helices on the anterior side; αH2 and αH3 on the posterior side of central β-sheet.	Only one helix (αH1) on the anterior side of central β-sheet; lacks other helices on the anterior side; complete loss or presence of short helices on the posterior side of the central β-sheet.
Central β-sheet architecture	Canonical seven-stranded core β-sheet with a typical Rossmann fold as observed in MTases [63, 71]; usual βS3-βS4 topological crossover.	Six-stranded core β-sheet architecture; βS2-βS3 topological crossover due to degeneration of typical βS3; degenerated core βS3 and shortened βS5; βS3 is occasionally represented by a small strand in PDB ID 5C8S and 7N0D [80, 83], forming a seven β-stranded central sheet (not consistent)	Six-stranded core β-sheet architecture; βS2-βS3 topological crossover due to degeneration of core βS3; shortened βS5 is predicted as a loop in a few AF3 structures.	Six-stranded core β-sheet architecture; βS2-βS3 topological crossover due to degeneration of core βS3; shortened βS5 is predicted as a loop in a few AF3 structures.	Six-stranded core β-sheet architecture; βS2-βS3 topological crossover due to degeneration of core βS3. Instead of shortened βS5, the βS5 is extended, forming an elongated β-hairpin with βS6.
Three-stranded anti-parallel β-sheet insert region	Three-stranded antiparallel β-sheet insert region after the core βS5; includes an additional small helix flanking the antiparallel β-sheet.	Three-stranded antiparallel β-sheet insert region after core βS4 [80–82].	Three-stranded antiparallel β-sheet insert region after core βS4.	Absence of insert region. βS4 directly leads to a shortened αH3 and βS5 with no β-sheet insert region.	Absence of insert region; βS4 directly leads to βS5-βS6 hairpin with no insert region.
C-terminal insert region with ZnF module	Bi-helical hairpin insert region between the core βS6 and βS7; absence of ZnF module within the bi-helical hairpin insert.	Insert region with two small β-strands and two helices between core βS5 and βS6; contains a conserved ZnF module (C3H type). The insert after βS5 distorts the typical C-terminal β-hairpin of Rossmann fold MTases.	Insert region with three antiparallel β-sheet followed by a helix: insert region occurs between core βS5 and βS6; contains a conserved ZnF module. The insert after βS5 distorts the typical C-terminal β-hairpin of Rossmann fold MTases.	Insert region with one helix between core βS5 and βS6, yet the loop region connecting the shortened βS5 and insert helix encodes the C3H type ZnF module; shortened βS5 distorts the typical C-terminal β-hairpin of Rossmann fold MTases.	Elongated βS5-βS6 hairpin with no insert region; the ZnF module occurs as a C-terminal extension (C4 type). An additional strand as an insert between αH1 and βS1 is stacked against the strands of the C-terminal extension.
Key SAM binding regions	Glycine-rich loop with a characteristic motif GxGxG, on top of βS1; conserved D in motif G[h]D, on top of βS2 (h: hydrophobic residue); Conserved motif SWH[Y/F], on top of βS4 [63, 71].	Glycine-rich loop with a subtly distinct motif: DI followed by GNPKG, on top of βS1; conserved D in motif [F/C]YD, on top of βS2; conserved motif WNCNV on top of βS3 [64, 80–83].	Glycine-rich loop with a subtly distinct motif NMGCGK, on top of βS1; conserved D in motif N[I/V]D, on top of βS2; conserved motif DSHWY on top of βS3.	Glycine-rich loop with a distinct motif GGG, on top of βS1; conserved D in motif N[h]D, on top of βS2 (h: hydrophobic residue); conserved motif D[h]Y[x]D on top of βS3 (h: hydrophobic residue, x: any residue).	No conservation of residues in structurally superimposable region.
Key substrate binding regions	Conserved R in αH1; conserved motif TE[T/K] in the three-stranded β-sheet insert region [63, 71].	Conserved N, R, and [Q/E] in αH1; conserved motif R[F/Y]D on top of βS4; LY[V/I]N and HAF[H/L]T in the three-stranded β-sheet insert region [64, 80–83].	Conserved R and Q in αH1; conserved motif F[S/Y]D on top of βS4; LYLN and A[b]Y[x]H motifs in the structurally superimposable three-stranded β-sheet insert region (b: basic residue, x: any residue).	Partially conserved N and a conserved R in αH1; loss of conserved motif on top of βS4 and loss of three-stranded β-sheet insert region.	No conservation of residues in structurally superimposable region. The only conserved residues include a L in αH1, P and an "NR" dyad in the loop connecting βS1-βS2.
Additional comments and remarks	Only N7-MTase domain within Nidovirales that exhibit seven-stranded Rossmann fold with βS3-βS4 crossover. However, it shares the three-stranded β-sheet insert region and multiple key binding sites with atypical N7-MTases of other Nidovirales.	This atypical six-stranded version is retained in all Nidovirales except for Tornidovirus, which exhibits a more canonical version. However, the domain retains catalytic motifs. Shown to be an active MTase domain [80–82].	This atypical six-stranded version is retained in all Nidovirales except for Tornidovirus, which exhibits a more canonical version. However, the domain retains catalytic motifs.	This atypical six-stranded version is retained in all Nidovirales except for Tornidovirus, which exhibits a more canonical version. However, the domain retains catalytic motifs.	While retaining the atypical six-stranded version, the domain exhibits further variations with an unusually extended C-terminal β-hairpin and no conservation of catalytic residues. Likely inactive domain.
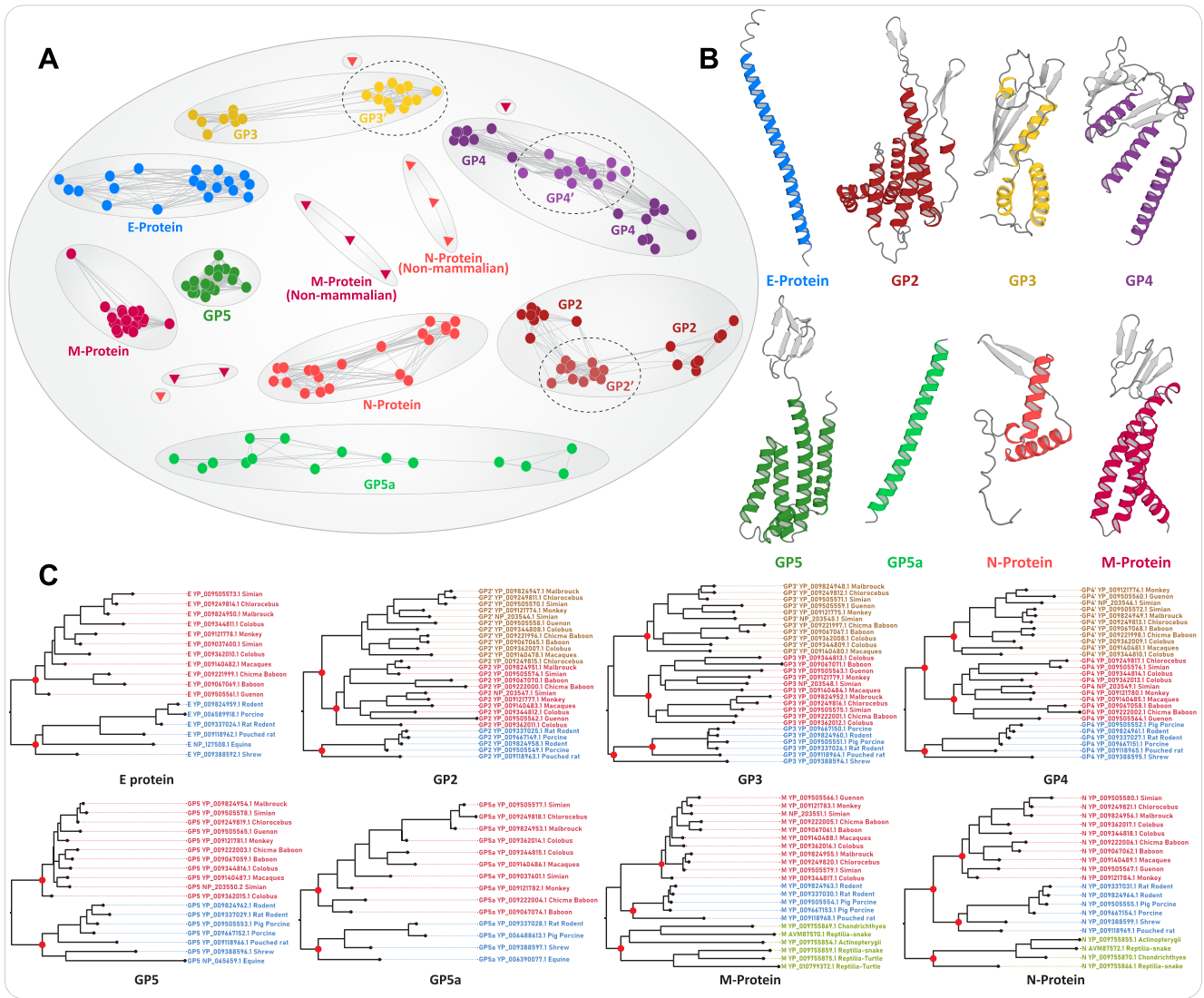
Figure 6. CLANS-based clustering and phylogenetic analysis of arterivirus structural glycoproteins. (**A**) CLANS-based clustering patterns of structural glycoproteins from both mammalian and non-mammalian arteriviruses. Using CLANS, all-versus-all pairwise sequence similarity-based clustering was conducted, with each glycoprotein cluster shown in distinct colors. (**B**) Representative AF3-predicted structures of each glycoprotein, with colors consistent with their respective CLANS clusters. (**C**) ML trees of each glycoprotein. The ML trees show clear clustering patterns that reflect host-specific divergence, especially between primate and non-primate hosts—primate clusters are highlighted in red and brown, non-primate clusters in blue. Likewise, the non-mammalian GPs are distinct, and their clusters are highlighted in green. Nodes with high confidence bootstrap values are marked in red dots.

compositions. Moreover, five other GP-encoding ORFs from these viruses exceed 300 amino acids, which is in contrast to mammalian arteriviruses, where the longest structural GPs (typically GP2 and GP5) range between 200 and 300 amino acids (Fig. 1B and Supplementary Fig. S5).

Despite their diversity, we were able to classify the majority of these proteins. For instance, all non-mammalian arterivirus genomes, with the exception of WJHA, contain at least two ORFs encoding GP-like structures that maintain key features, such as an extracellular ectodomain, one or more transmembrane segments, and an extended intracellular endodomain [102, 103] (Supplementary Fig. S5). Through sequence alignments and structural topology analysis, we identified M proteins in at least six non-mammalian arteriviruses (Figs 1B and 6). Similarly, the N protein features a disordered, positively charged RNA-binding region at the N-terminus and

a C-terminal domain comprised of α-helices and β-strands forming a four-stranded antiparallel β-sheet [104–106]. Anchoring on these features, we identified N proteins in at least five non-mammalian arteriviruses. Both M and N proteins in these viruses have experienced significant divergence, leading to small, isolated clusters or singlets in CLANS network analysis (Fig. 6). Phylogenetically, non-mammalian M and N proteins are distinct from their mammalian counterparts, with a few exceptions. While we could not find ORFs encoding structural homologs of M proteins in GGGSA isolates or N proteins in two genomes (HOFA and TSHSA), we advise caution in interpreting their absence as gene loss, as this may be due to poor genome assemblies. This is further illustrated by the TSHSA genome, which deviates from the typical genomic architecture and includes two atypical ORFs [22] located before the canonical ORF1a, encoding a transmembrane seg-

ment and PLP2 (Fig. 1B). Along with the significantly larger GP-like ORFs at its C-terminus, these features may represent artifacts resulting from suboptimal genome assemblies. Additionally, there are eight other structural proteins, ranging from 35 to 205 residues. AF3 predictions suggest these are mostly disordered or consist of helical segments, lacking clear structural homology for precise annotation. While these may represent GPs, we provisionally label them with a question mark as "GP?" (Fig. 1B). Nevertheless, the overall comparisons of the structural proteins from arteriviruses highlight the dynamic evolution of these viruses and emphasize the influence of host-specific factors in shaping the divergence of structural proteins across different host species [103, 107, 108].

Overall functional and evolutionary implications

The comparative analysis of arterivirus genomes, including seven from non-mammalian hosts, highlights key evolutionary patterns and molecular features that subtly differentiate mammalian and non-mammalian arteriviruses. Although non-mammalian arteriviruses span diverse hosts—one from a ray-finned fish, one from a cartilaginous fish, and five from reptiles (three snakes and two turtles)—they do not exhibit notably distinct genomic organizations specific to each host group. While all arterivirus genomes share a broadly unifying genomic architecture, pairwise comparisons across diverse hosts highlight specific differences that distinguish mammalian and non-mammalian arteriviruses as a whole, even amid the greater host diversity of the latter group.

While ORF1a in mammalian arteriviruses contains multiple PLP paralogs suggesting roles in polyprotein processing and immune evasion [9–11, 40, 41], non-mammalian arteriviruses have fewer PLPs, with PLP1β and PLP1γ likely absent and only PLP2 consistently conserved, indicating a distinct evolutionary trajectory. In addition, structural predictions reveal newly characterized ORF1a domains in non-mammalian arteriviruses, including both globular α + β topology domains and potentially disordered regions, though their functional significance is uncertain. Whether these α + β domains exhibit enzymatic functions similar to PLPs or play a role in host-specific interactions remains uncertain and requires further investigation. Improved genome assemblies, expanding sequencing of non-mammalian arteriviruses across additional hosts, and experimentally validating these predicted domains are crucial for addressing these uncertainties. Beyond PLPs, our findings reveal that ORF1a also encodes NSP3C, a conserved wHTH domain, and NSP7—a SBB domain, both of which provide insights into arterivirus replication. The structural alignment between NSP3C and the C-terminal region of coronavirus NSP4 suggests a shared function in RNA binding within RTCs, likely facilitating interactions with viral RNA essential for viral genome replication [47, 48]. Meanwhile, NSP7, part of the SBB domain family, exhibits structural homology with the HAS-barrel and L21 ribosomal protein, both known for their roles in nucleic acid binding and stabilizing protein-nucleic acid complexes [53–56]. The retention of both domains across most analyzed arteriviruses underscores their essential roles in RNA interactions, potentially crucial for arterivirus replication.

The evolutionary divergence of ORF1b between mammalian and non-mammalian arteriviruses underscores distinct replication and immune evasion strategies shaped by lineage-specific adaptations. Mammalian arteriviruses lack both the ExoN and 2′-O-MTase domains, suggesting that the NSP12

N7-MTase-like domain has been repurposed for an alternative function due to its likely loss of enzymatic activity. In contrast, non-mammalian arteriviruses exhibit sporadic retention of the 2′-O-MTase domains in certain genomes, while the ExoN domain—typically fused with N7-MTases in coronaviruses— is occasionally preserved, though the N7-MTase domain itself is absent. Overall, arteriviruses display a distinctive pattern within Nidovirales, marked by widespread loss of methylases, which contrasts sharply with other nidoviruses, such as coronaviruses that utilize both the N7-MTase + ExoN heterodimer and 2′-O-MTase domains for mRNA capping. This suggests that most arteriviruses may not depend on this process for immune evasion, aside from rare instances where 2′-O-MTase domains are retained. Comparatively, members of Tobaniviridae also lack conserved N7-MTases but consistently retain 2′-O-MTase domains, making arteriviruses a distinct case within Nidovirales [63]. The overall phyletic distribution, characterized by the sporadic retention of 2′-O-MTase, ExoN, and the NSP12 N7-MTase-like domain, suggests that the ancestral lineage leading to arteriviruses likely possessed these elements. Subsequently, mammalian arteriviruses appear to have lost both ExoN and the enzymatic function of NSP12, while non-mammalian arteriviruses have retained 2′-O-MTase and ExoN in a scattered pattern. The conservation of the inactive NSP12 N7-MTase-like domain in mammalian arteriviruses suggests it remains essential, potentially for immune evasion by modulating the host response or alternatively for RNA binding within the viral replication machinery. Additionally, we demonstrate that arterivirus glycoproteins and structural proteins evolve rapidly, with non-mammalian genomes exhibiting potential ORF fusions and larger GP-like ORFs, while mammalian genomes vary even between primate and non-primate hosts. Overall, these findings underscore the significance of domain co-option in the evolution of arteriviruses and the impact of host-virus co-evolution on shaping the diversity of arterivirus proteins between mammalian and non-mammalian arteriviruses.

Conclusions

The comparative genomics and sequence-structure analysis presented here reveals previously unannotated domains in arteriviruses, offering new insights into their evolution. By analyzing all sequenced non-mammalian genomes and reconstructing protein domain architectures, we identify key differences between mammalian and non-mammalian arteriviruses, including unique domains in each group. We classify NSP3C as a wHTH domain and NSP7 as a SBB domain and identify NSP12 in ORF1b as an inactive N7-MTase-like domain exclusive to mammalian arteriviruses. While mammalian arteriviruses lack any domains for mRNA capping, non-mammalian counterparts sporadically retain 2′-O-MTase and ExoN domains, suggesting that some may utilize epigenetic modifications for immune evasion. Additionally, non-mammalian arteriviruses show reductions in PLPs and glycoproteins, underscoring host-specific structural adaptations. These findings refine our understanding of arterivirus evolution and provide testable predictions for future research.

Acknowledgements

and A.K.: Visualization: S.R. and A.K. Validation: S.R. and A.K. Supervision: A.K. Resources: A.K. Software: S.R. and A.K. Project administration: A.K. Funding acquisition: A.K. Writing—original draft: S.R. and A.K. Writing—review and editing: A.K. with contributions from S.R.

Supplementary data

Supplementary data is available at NAR Genomics & Bioinformatics online.

Conflict of interest

None declared.

Funding

Data availability

The data underlying this article are available in the article and in its online supplementary material. These data are also available at https://doi.org/10.5281/zenodo.14172364.

References

1. Brinton MA, Gulyaeva AA, Balasuriya UBR *et al.* ICTV virus taxonomy profile: Arteriviridae 2021. *J Gen Virol* 2021;**102**:001632. https://doi.org/10.1099/jgv.0.001632

2. Snijder EJ, Kikkert M, Fang Y. Arterivirus molecular biology and pathogenesis. *J Gen Virol* 2013;**94**:2141–63. https://doi.org/10.1099/vir.0.056341-0

3. Liao Y, Wang H, Liao H *et al.* Classification, replication, and transcription of Nidovirales. *Front Microbiol* 2024;**14**:1291761. https://doi.org/10.3389/fmicb.2023.1291761

4. Balasuriya UB, Carossino M. Reproductive effects of arteriviruses: equine arteritis virus and porcine reproductive and respiratory syndrome virus infections. *Curr Opin Virol* 2017;**27**:57–70. https://doi.org/10.1016/j.coviro.2017.11.005

5. Balasuriya UBR, Carossino M, Timoney PJ. Equine viral arteritis: a respiratory and reproductive disease of significant economic importance to the equine industry. *Equine Vet Educ* 2018;**30**:497–512. https://doi.org/10.1111/eve.12672

6. Guo Z, Chen XX, Li R *et al.* The prevalent status and genetic diversity of porcine reproductive and respiratory syndrome virus in China: a molecular epidemiological perspective. *Virol J* 2018;**15**:2. https://doi.org/10.1186/s12985-017-0910-6

7. Lauber C, Ziebuhr J, Junglen S *et al.* Mesoniviridae: a proposed new family in the order Nidovirales formed by a single species of mosquito-borne viruses. *Arch Virol* 2012;**157**:1623–8. https://doi.org/10.1007/s00705-012-1295-x

8. Walker PJ, Cowley JA, Dong X *et al.* ICTV virus taxonomy profile: Roniviridae. *J Gen Virol* 2021;**102**:jgv001514. https://doi.org/10.1099/jgv.0.001514

9. Osipiuk J, Azizi S-A, Dvorkin S *et al.* Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors. *Nat Commun* 2021;**12**:743. https://doi.org/10.1038/s41467-021-21060-3

10. Mielech AM, Chen Y, Mesecar AD *et al.* Nidovirus papain-like proteases: multifunctional enzymes with protease, deubiquitinating and deISGylating activities. *Virus Res* 2014;**194**:184–90. https://doi.org/10.1016/j.virusres.2014.01.025

11. Ziebuhr J, Snijder EJ, Gorbalenya AE. Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J Gen Virol* 2000;**81**:853–79. https://doi.org/10.1099/0022-1317-81-4-853

12. Lauber C, Goeman JJ, Parquet MdC *et al.* The footprint of genome architecture in the largest genome expansion in RNA viruses. *PLoS Pathog* 2013;**9**:e1003500. https://doi.org/10.1371/journal.ppat.1003500

13. van Dinten LC, Wassenaar AL, Gorbalenya AE *et al.* Processing of the equine arteritis virus replicase ORF1b protein: identification of cleavage products containing the putative viral polymerase and helicase domains. *J Virol* 1996;**70**:6625–33. https://doi.org/10.1128/jvi.70.10.6625-6633.1996

14. Beerens N, Selisko B, Ricagno S *et al.* De novo initiation of RNA synthesis by the arterivirus RNA-dependent RNA polymerase. *J Virol* 2007;**81**:8384–95. https://doi.org/10.1128/JVI.00564-07

15. Deng Z, Lehmann KC, Li X *et al.* Structural basis for the regulatory function of a complex zinc-binding domain in a replicative arterivirus helicase resembling a nonsense-mediated mRNA decay helicase. *Nucleic Acids Res* 2014;**42**:3464–77. https://doi.org/10.1093/nar/gkt1310

16. Zhang M, Li X, Deng Z *et al.* Structural biology of the arterivirus nsp11 endoribonucleases. *J Virol* 2016;**91**:e01309–16. https://doi.org/10.1128/jvi.01309-01316

17. Lehmann KC, Hooghiemstra L, Gulyaeva A *et al.* Arterivirus nsp12 versus the coronavirus nsp16 2′-O-methyltransferase: comparison of the C-terminal cleavage products of two nidovirus pp1ab polyproteins. *J Gen Virol* 2015;**96**:2643–55. https://doi.org/10.1099/vir.0.000209

18. Veit M, Matczuk AK, Sinhadri BC *et al.* Membrane proteins of arterivirus particles: structure, topology, processing and function. *Virus Res* 2014;**194**:16–36. https://doi.org/10.1016/j.virusres.2014.09.010

19. Snijder Eric J, van Tol H, Pedersen Ketil W *et al.* Identification of a novel structural protein of arteriviruses. *J Virol* 1999;**73**:6335–45. https://doi.org/10.1128/JVI.73.8.6335-6345.1999

20. Lauck M, Sibley Samuel D, Hyeroba D *et al.* Exceptional simian hemorrhagic fever virus diversity in a wild African primate community. *J Virol* 2013;**87**:688–91. https://doi.org/10.1128/JVI.02433-12

21. Shi M, Lin X-D, Chen X *et al.* The evolutionary history of vertebrate RNA viruses. *Nature* 2018;**556**:197–202. https://doi.org/10.1038/s41586-018-0012-7

22. Lyu S, Yuan X, Zhang H *et al.* Complete genome sequence and analysis of a new lethal arterivirus, Trionyx sinensis hemorrhagic syndrome virus (TSHSV), amplified from an infected Chinese softshell turtle. *Arch Virol* 2019;**164**:2593–7. https://doi.org/10.1007/s00705-019-04324-8

23. Berman HM, Westbrook J, Feng Z *et al.* The Protein Data Bank. *Nucleic Acids Res* 2000;**28**:235–42. https://doi.org/10.1093/nar/28.1.235

24. Mistry J, Chuguransky S, Williams L *et al.* Pfam: the protein families database in 2021. *Nucleic Acids Res* 2021;**49**:D412–19. https://doi.org/10.1093/nar/gkaa913

25. Zimmermann L, Stephens A, Nam S-Z *et al.* A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol* 2018;**430**:2237–43. https://doi.org/10.1016/j.jmb.2017.12.007

26. Gabler F, Nam S-Z, Till S *et al.* Protein sequence analysis using the MPI bioinformatics toolkit. *Curr Protoc Bioinformatics* 2020;**72**:e108. https://doi.org/10.1002/cpbi.108

27. Remmert M, Biegert A, Hauser A *et al.* HHblits: lightning-fast iterative protein sequence searching by HMM–HMM alignment. *Nat Methods* 2012;**9**:173–5. https://doi.org/10.1038/nmeth.1818

28. Marchler-Bauer A, Panchenko AR, Shoemaker BA *et al*. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. *Nucleic Acids Res* 2002;**30**:281–3. https://doi.org/10.1093/nar/30.1.281

29. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;**30**:772–80. https://doi.org/10.1093/molbev/mst010

30. Lassmann T. Kalign 3: multiple sequence alignment of large datasets. *Bioinformatics* 2020;**36**:1928–9. https://doi.org/10.1093/bioinformatics/btz795

31. Hallgren J, Tsirigos KD, Pedersen MD *et al*. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. bioRxiv, https://doi.org/10.1101/2022.04.08.487609, 10 April 2022, preprint: not peer reviewed.

32. Abramson J, Adler J, Dunger J *et al*. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 2024;**630**:493–500. https://doi.org/10.1038/s41586-024-07487-w

33. Holm L, Laiho A, Törönen P *et al*. DALI shines a light on remote homologs: one hundred discoveries. *Protein Sci* 2023;**32**:e4519. https://doi.org/10.1002/pro.4519

34. Drozdetskiy A, Cole C, Procter J *et al*. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res* 2015;**43**:W389–94. https://doi.org/10.1093/nar/gkv332

35. Frickey T, Lupas A. CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* 2004;**20**:3702–4. https://doi.org/10.1093/bioinformatics/bth444

36. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 2010;**5**:e9490. https://doi.org/10.1371/journal.pone.0009490

37. Minh BQ, Schmidt HA, Chernomor O *et al*. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020;**37**:1530–4. https://doi.org/10.1093/molbev/msaa015

38. Chernomor O, von Haeseler A, Minh BQ. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol* 2016;**65**:997–1008. https://doi.org/10.1093/sysbio/syw037

39. Hoang DT, Chernomor O, von Haeseler A *et al*. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 2018;**35**:518–22. https://doi.org/10.1093/molbev/msx281

40. Sun Z, Chen Z, Lawson SR *et al*. The cysteine protease domain of porcine reproductive and respiratory syndrome virus nonstructural protein 2 possesses deubiquitinating and interferon antagonism functions. *J Virol* 2010;**84**:7832–46. https://doi.org/10.1128/JVI.00217-10

41. Gulyaeva A, Dunowska M, Hoogendoorn E *et al*. Domain organization and evolution of the highly divergent 5′ coding region of genomes of arteriviruses, including the novel possum nidovirus. *J Virol* 2017;**91**:e02096-16. https://doi.org/10.1128/JVI.02096-16

42. Aravind L, Anantharaman V, Balaji S *et al*. The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev* 2005;**29**:231–62. https://doi.org/10.1016/j.femsre.2004.12.008

43. Brennan RG. The winged-helix DNA-binding motif: another helix-turn-helix takeoff. *Cell* 1993;**74**:773–6. https://doi.org/10.1016/0092-8674(93)90456-Z

44. Clementz MA, Kanjanahaluethai A, O'Brien TE *et al*. Mutation in murine coronavirus replication protein nsp4 alters assembly of double membrane vesicles. *Virology* 2008;**375**:118–29. https://doi.org/10.1016/j.virol.2008.01.018

45. Oostra M, te Lintelo EG, Deijs M *et al*. Localization and membrane topology of coronavirus nonstructural protein 4: involvement of the early secretory pathway in replication. *J Virol* 2007;**81**:12323–36. https://doi.org/10.1128/JVI.01506-07

46. Snijder EJ, van Tol H, Roos N *et al*. Non-structural proteins 2 and 3 interact to modify host cell membranes during the formation of the arterivirus replication complex. *J Gen Virol* 2001;**82**:985–94. https://doi.org/10.1099/0022-1317-82-5-985

47. Chen A, Lupan A-M, Quek RT *et al*. A coronaviral pore-replicase complex links RNA synthesis and export from double-membrane vesicles. *Sci Adv* 2024;**10**:eadq9580. https://doi.org/10.1126/sciadv.adq9580

48. Manolaridis I, Wojdyla JA, Panjikar S *et al*. Structure of the C-terminal domain of nsp4 from feline coronavirus. *Acta Crystallogr D Biol Crystallogr* 2009;**65**:839–46. https://doi.org/10.1107/S0907444909018253

49. Manolaridis I, Gaudin C, Posthuma CC *et al*. Structure and genetic analysis of the arterivirus nonstructural protein 7alpha. *J Virol* 2011;**85**:7449–53. https://doi.org/10.1128/JVI.00255-11

50. Murzin AG, Lesk AM, Chothia C. Principles determining the structure of β-sheet barrels in proteins I. A theoretical analysis. *J Mol Biol* 1994;**236**:1369–81. https://doi.org/10.1016/0022-2836(94)90064-7

51. Murzin AG, Lesk AM, Chothia C. Principles determining the structure of β-sheet barrels in proteins II. The observed structures. *J Mol Biol* 1994;**236**:1382–400. https://doi.org/10.1016/0022-2836(94)90065-5

52. Youkharibache P, Veretnik S, Li Q *et al*. The small β-barrel domain: a survey-based structural analysis. *Structure* 2019;**27**:6–26. https://doi.org/10.1016/j.str.2018.09.012

53. Klein DJ, Moore PB, Steitz TA. The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J Mol Biol* 2004;**340**:141–77. https://doi.org/10.1016/j.jmb.2004.03.076

54. Iyer LM, Makarova KS, Koonin EV *et al*. Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. *Nucleic Acids Res* 2004;**32**:5260–79. https://doi.org/10.1093/nar/gkh828

55. Murphy EL, Singh KV, Avila B *et al*. Cryo-electron microscopy structure of the 70S ribosome from *Enterococcus faecalis*. *Sci Rep* 2020;**10**:16301. https://doi.org/10.1038/s41598-020-73199-6

56. Lomakin IB, Devarkar SC, Patel S *et al*. Sarecycline inhibits protein translation in cutibacterium acnes 70S ribosome using a two-site mechanism. *Nucleic Acids Res* 2023;**51**:2915–30. https://doi.org/10.1093/nar/gkad103

57. Li H, Luo Q, Jing H *et al*. Research progress on Porcine reproductive and Respiratory syndrome virus NSP7 protein. *Animals* 2023;**13**:2269. https://doi.org/10.3390/ani13142269

58. Chen J, Xu X, Tao H *et al*. Structural analysis of porcine reproductive and respiratory syndrome virus non-structural protein 7α (NSP7α) and identification of its interaction with NSP9. *Front Microbiol* 2017;**8**:853. https://doi.org/10.3389/fmicb.2017.00853

59. Gribble J, Stevens LJ, Agostini ML *et al*. The coronavirus proofreading exoribonuclease mediates extensive viral recombination. *PLoS Pathog* 2021;**17**:e1009226. https://doi.org/10.1371/journal.ppat.1009226

60. Eckerle LD, Becker MM, Halpin RA *et al*. Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS Pathog* 2010;**6**:e1000896. https://doi.org/10.1371/journal.ppat.1000896

61. Dürr SL, Bohuszewicz O, Berta D *et al*. The role of conserved residues in the DEDDh motif: the proton-transfer mechanism of HIV-1 RNase H. *ACS Catal* 2021;**11**:7915–27. https://doi.org/10.1021/acscatal.1c01493

62. Robson F, Khan KS, Le TK *et al*. Coronavirus RNA proofreading: molecular basis and therapeutic targeting. *Mol Cell* 2020;**79**:710–27. https://doi.org/10.1016/j.molcel.2020.07.027

63. Ferron F, Debat HJ, Shannon A *et al*. A N7-guanine RNA cap methyltransferase signature-sequence as a genetic marker of large genome, non-mammalian Tobaniviridae. *NAR Genom Bioinform* 2020;**2**:lqz022. https://doi.org/10.1093/nargab/lqz022

64. Nencka R, Silhan J, Klima M *et al*. Coronaviral RNA-methyltransferases: function, structure and inhibition.

Nucleic Acids Res 2022;**50**:635–50.
https://doi.org/10.1093/nar/gkab1279

65. Rehwinkel J, Tan CP, Goubau D *et al.* RIG-I detects viral genomic RNA during negative-strand RNA virus infection. *Cell* 2010;**140**:397–408. https://doi.org/10.1016/j.cell.2010.01.020

66. Poeck H, Bscheider M, Gross O *et al.* Recognition of RNA virus by RIG-I results in activation of CARD9 and inflammasome signaling for interleukin 1 beta production. *Nat Immunol* 2010;**11**:63–9. https://doi.org/10.1038/ni.1824

67. Loo YM, Fornek J, Crochet N *et al.* Distinct RIG-I and MDA5 signaling by RNA viruses in innate immunity. *J Virol* 2008;**82**:335–45. https://doi.org/10.1128/JVI.01080-07

68. Züst 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**:137–43. https://doi.org/10.1038/ni.1979

69. Decroly E, Ferron F, Lescar J *et al.* Conventional and unconventional mechanisms for capping viral mRNA. *Nat Rev Micro* 2012;**10**:51–65. https://doi.org/10.1038/nrmicro2675

70. Ramanathan A, Robb GB, Chan S-H. mRNA capping: biological functions and applications. *Nucleic Acids Res* 2016;**44**:7511–26. https://doi.org/10.1093/nar/gkw551

71. Shannon A, Sama B, Gauffre P *et al.* A second type of N7-guanine RNA cap methyltransferase in an unusual locus of a large RNA virus genome. *Nucleic Acids Res* 2022;**50**:11186–98. https://doi.org/10.1093/nar/gkac876

72. Sagripanti JL, Zandomeni RO, Weinmann R. The cap structure of simian hemorrhagic fever virion RNA. *Virology* 1986;**151**:146–50. https://doi.org/10.1016/0042-6822(86)90113-3

73. 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**:925–30. https://doi.org/10.1099/0022-1317-75-4-925

74. Martin JL, McMillan FM. SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr Opin Struct Biol* 2002;**12**:783–93. https://doi.org/10.1016/S0959-440X(02)00391-3

75. Schubert HL, Blumenthal RM, Cheng X. Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem Sci* 2003;**28**:329–35. https://doi.org/10.1016/S0968-0004(03)00090-2

76. Fauman EB, Blumenthal RM, Cheng X. *S-Adenosylmethionine-Dependent Methyltransferases*. Singapore: World Scientific, 1999, 1–38. https://doi.org/10.1142/9789812813077_0001

77. Medvedev KE, Kinch LN, Grishin NV. Functional and evolutionary analysis of viral proteins containing a Rossmann-like fold. *Protein Sci* 2018;**27**:1450–63. https://doi.org/10.1002/pro.3438

78. Cheng X, Blumenthal RM. Mammalian DNA Methyltransferases: A Structural Perspective. *Structure* 2008;**16**:341–350. https://doi.org/10.1016/j.str.2008.01.004

79. Iyer LM, Abhiman S, Aravind L. Natural history of eukaryotic DNA methylation systems. *Prog Mol Biol Transl Sci* 2011;**101**:25–104. https://doi.org/10.1016/B978-0-12-387685-0.00002-0

80. Ma Y, Wu L, Shaw N *et al.* Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. *Proc Natl Acad Sci USA* 2015;**112**:9436–41. https://doi.org/10.1073/pnas.1508686112

81. Ferron F, Subissi L, Silveira De Morais AT *et al.* Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proc Natl Acad Sci USA* 2018;**115**:E162–71. https://doi.org/10.1073/pnas.1718806115

82. Ogando NS, El Kazzi P, Zevenhoven-Dobbe JC *et al.* Structure–function analysis of the nsp14 N7–guanine methyltransferase reveals an essential role in *Betacoronavirus*

replication. *Proc Natl Acad Sci USA* 2021;**118**:e2108709118. https://doi.org/10.1073/pnas.2108709118

83. Kottur J, Rechkoblit O, Quintana-Feliciano R *et al.* High-resolution structures of the SARS-CoV-2 N7-methyltransferase inform therapeutic development. *Nat Struct Mol Biol* 2022;**29**:850–3. https://doi.org/10.1038/s41594-022-00828-1

84. Becares M, Pascual-Iglesias A, Nogales A *et al.* Mutagenesis of coronavirus nsp14 reveals its potential role in modulation of the innate immune response. *J Virol* 2016;**90**:5399–414. https://doi.org/10.1128/JVI.03259-15

85. Chen Y, Tao J, Sun Y *et al.* Structure-function analysis of severe acute respiratory syndrome coronavirus RNA cap guanine-N7-methyltransferase. *J Virol* 2013;**87**:6296–305. https://doi.org/10.1128/JVI.00061-13

86. Imprachim N, Yosaatmadja Y, Newman JA. Crystal structures and fragment screening of SARS-CoV-2 NSP14 reveal details of exoribonuclease activation and mRNA capping and provide starting points for antiviral drug development. *Nucleic Acids Res* 2023;**51**:475–87. https://doi.org/10.1093/nar/gkac1207

87. Byszewska M, Śmietański M, Purta E *et al.* RNA methyltransferases involved in 5′ cap biosynthesis. *RNA Biol* 2014;**11**:1597–607. https://doi.org/10.1080/15476286.2015.1004955

88. Rosas-Lemus M, Minasov G, Shuvalova L *et al.* The crystal structure of nsp10-nsp16 heterodimer from SARS-CoV-2 in complex with S-adenosylmethionine. bioRxiv, https://doi.org/10.1101/2020.04.17.047498, 26 April 2020, preprint: not peer reviewed.

89. Vithani N, Ward MD, Zimmerman MI *et al.* SARS-CoV-2 Nsp16 activation mechanism and a cryptic pocket with pan-coronavirus antiviral potential. *Biophys J* 2021;**120**:2880–9. https://doi.org/10.1016/j.bpj.2021.03.024

90. Zeng C, Wu A, Wang Y *et al.* Identification and characterization of a ribose 2′-O-methyltransferase encoded by the ronivirus branch of Nidovirales. *J Virol* 2016;**90**:6675–85. https://doi.org/10.1128/JVI.00658-16

91. Russo AT, White MA, Watowich SJ. The crystal structure of the Venezuelan equine encephalitis alphavirus nsP2 protease. *Structure* 2006;**14**:1449–58. https://doi.org/10.1016/j.str.2006.07.010

92. Sawicki Dorothea L, Perri S, Polo John M *et al.* Role for nsP2 proteins in the cessation of alphavirus minus-strand synthesis by host cells. *J Virol* 2006;**80**:360–71. https://doi.org/10.1128/JVI.80.1.360-371.2006

93. Shin G, Yost SA, Miller MT *et al.* Structural and functional insights into alphavirus polyprotein processing and pathogenesis. *Proc Natl Acad Sci USA* 2012;**109**:16534–9. https://doi.org/10.1073/pnas.1210418109

94. Eckei L, Krieg S, Bütepage M *et al.* The conserved macrodomains of the non-structural proteins of Chikungunya virus and other pathogenic positive strand RNA viruses function as mono-ADP-ribosylhydrolases. *Sci Rep* 2017;**7**:41746. https://doi.org/10.1038/srep41746

95. Law Y-S, Utt A, Tan YB *et al.* Structural insights into RNA recognition by the chikungunya virus nsP2 helicase. *Proc Natl Acad Sci USA* 2019;**116**:9558–67. https://doi.org/10.1073/pnas.1900656116

96. Fros JJ, van der Maten E, Vlak JM *et al.* The C-terminal domain of chikungunya virus nsP2 independently governs viral RNA replication, cytopathicity, and inhibition of interferon signaling. *J Virol* 2013;**87**:10394–400. https://doi.org/10.1128/JVI.00884-13

97. Akhrymuk I, Kulemzin SV, Frolova EI. Evasion of the innate immune response: the Old World alphavirus nsP2 protein induces rapid degradation of Rpb1, a catalytic subunit of RNA polymerase II. *J Virol* 2012;**86**:7180–91. https://doi.org/10.1128/JVI.00541-12

98. Göertz GP, McNally KL, Robertson SJ *et al.* The methyltransferase-like domain of Chikungunya virus nsP2

inhibits the interferon response by promoting the nuclear export of STAT1. *J Virol* 2018;**92**: e01008-18. https://doi.org/10.1128/JVI.01008-18

99. Chen K, Liu J, Liu S *et al.* Methyltransferase SETD2-mediated methylation of STAT1 is critical for interferon antiviral activity. *Cell* 2017;**170**:492–506. https://doi.org/10.1016/j.cell.2017.06.042

100. Vatter HA, Di H, Donaldson EF *et al.* Each of the eight simian hemorrhagic fever virus minor structural proteins is functionally important. *Virology* 2014;**462–463**:351–62. https://doi.org/10.1016/j.virol.2014.06.001

101. 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**:e19056. https://doi.org/10.1371/journal.pone.0019056

102. Dokland T. The structural biology of PRRSV. *Virus Res* 2010;**154**:86–97. https://doi.org/10.1016/j.virusres.2010.07.029

103. Banerjee N, Mukhopadhyay S. Viral glycoproteins: biological role and application in diagnosis. *Virusdisease* 2016;**27**:1–11. https://doi.org/10.1007/s13337-015-0293-5

104. Yoo D, Wootton Sarah K, Li G *et al.* Colocalization and interaction of the porcine arterivirus nucleocapsid protein with the small nucleolar RNA-associated protein fibrillarin. *J Virol* 2003;**77**:12173–83. https://doi.org/10.1128/JVI.77.22.12173-12183.2003

105. Doan DN, Dokland T. Structure of the nucleocapsid protein of porcine reproductive and respiratory syndrome virus. *Structure* 2003;**11**:1445–51. https://doi.org/10.1016/j.str.2003.09.018

106. Deshpande A, Wang S, Walsh MA *et al.* Structure of the equine arteritis virus nucleocapsid protein reveals a dimer-dimer arrangement. *Acta Crystallogr D Biol Crystallogr* 2007;**63**:581–6. https://doi.org/10.1107/S0907444907008372

107. Le Pendu J, Nyström K, Ruvoën-Clouet N. Host–pathogen co-evolution and glycan interactions. *Curr Opin Virol* 2014;**7**:88–94. https://doi.org/10.1016/j.coviro.2014.06.001

108. Singh K, Mehta D, Dumka S *et al.* Quasispecies nature of RNA viruses: lessons from the past. *Vaccines* 2023;**11**:308. https://doi.org/10.3390/vaccines11020308