



Recurrent viral capture of cellular phosphodiesterases that antagonize OAS-RNase L

Stephen A. Goldstein^{a,b,1} and Nels C. Elde^{a,b,1}

Edited by Eugene Koonin, NIH, Bethesda, MD; received July 25, 2023; accepted November 20, 2023

Phosphodiesterases (PDEs) encoded by viruses are putatively acquired by horizontal transfer of cellular PDE ancestor genes. Viral PDEs inhibit the OAS-RNase L antiviral pathway, a key effector component of the innate immune response. Although the function of these proteins is well-characterized, the origins of these gene acquisitions are less clear. Phylogenetic analysis revealed at least five independent PDE acquisition events by ancestral viruses. We found evidence that PDE-encoding genes were horizontally transferred between coronaviruses belonging to different genera. Three clades of viruses within *Nidovirales*: merbecoviruses (MERS-CoV), embecoviruses (HCoV-OC43), and toroviruses encode independently acquired PDEs, and a clade of rodent alphacoronaviruses acquired an embecovirus PDE via recent horizontal transfer. Among rotaviruses, the PDE of rotavirus A was acquired independently from rotavirus B and G PDEs, which share a common ancestor. Conserved motif analysis suggests a link between all viral PDEs and a similar ancestor among the mammalian AKAP7 proteins despite low levels of sequence conservation. Additionally, we used ancestral sequence reconstruction and structural modeling to reveal that sequence and structural divergence are not well-correlated among these proteins. Specifically, merbecovirus PDEs are as structurally divergent from the ancestral protein and the solved structure of human AKAP7 PDE as they are from each other. In contrast, comparisons of rotavirus B and G PDEs reveal virtually unchanged structures despite evidence for loss of function in one, suggesting impactful changes that lie outside conserved catalytic sites. These findings highlight the complex and volatile evolutionary history of viral PDEs and provide a framework to facilitate future studies.

coronaviruses | evolution | horizontal gene transfer | innate immunity | OAS-RNase L

Horizontal gene transfer (HGT) is a major force in ancient and ongoing evolution among diverse viruses and hosts. Within the virosphere, the acquisition and re-purposing of host genes has shaped evolutionary history spanning from the origins of viruses themselves (1) to more specialized interactions with host immune pathways. Viruses in distinct unrelated orders, nidoviruses and rotaviruses, have independently acquired host mRNAs encoding eukaryotic LigT-like 2H-phosphoesterases, specifically 2'-5' phosphodiesterases (PDEs) with similarity to eukaryotic protein family members (2).

The role of PDEs in both coronavirus and rotavirus interactions with host innate antiviral immunity has been well characterized. Coronaviruses have large, single-stranded positive-sense RNA genomes of ~27 to 32 kilobases, the largest among all vertebrate viruses (3, 4), whereas rotaviruses have segmented double-stranded RNA (dsRNA) genomes. The two virus families are located at distant positions within the broader RNA virus phylogeny, which itself is difficult to confidently infer given great diversity among RNA virus families (5). As with all RNA viruses, dsRNA molecules are produced during transcription and replication of coronaviruses and rotaviruses and directly activate host immune signaling proteins including melanoma differentiation-associated protein 5 (MDA5) (6–8) and oligoadenylate synthetases (OAS) (9). These sensors activate interferon responses and RNase L, respectively, to restrict viral replication.

Viruses that encode 2'-5' PDEs such as the embecovirus and merbecovirus subgenera within the betacoronavirus genus, a rodent clade of alphacoronaviruses, and group A, B, and G rotaviruses can replicate unrestricted by the OAS-RNase L pathway. Their PDEs degrade the OAS-produced second messenger 2'-5' oligoadenylate, blocking it from inducing RNase L dimerization and cleavage of single-stranded viral and cellular RNA (Fig. 1). The role of viral PDEs during infection has been demonstrated for the prototypical betacoronavirus mouse hepatitis virus (10), MERS-CoV (11, 12), and group A rotavirus (13) and supported by evidence showing 2'-5' oligoadenylate degradation via biochemical assay for other viral PDEs (14, 15). Notably, selective pressure to encode a PDE may vary within an organism by cell type (16, 17), so viruses of the same host may face different pressures when infecting different cell types.

Significance

Horizontal gene transfer (HGT) of host genes to virus genomes is a common but underappreciated means of RNA virus evolution. We used a combination of sequence comparisons and protein structure predictions to reconstruct the evolutionary history of viral phosphodiesterase (PDE) genes. PDEs encoded by nidoviruses and rotaviruses appear to be descended from host AKAP7-like PDE genes and are potent antagonists of the antiviral OAS-RNase L pathway. By characterizing the evolutionary history of these genes, we shed light on how HGT shaped virus interactions with host immunity. By integrating evolutionary approaches, we overcome challenges posed by vast sequence divergence across viral and cellular PDE diversity to show that cellular AKAP7-like PDEs were recurrently acquired by two families of RNA viruses.

Author affiliations: ^aDepartment of Human Genetics, University of Utah, School of Medicine, Salt Lake City, UT 84112; and ^bHHMI, Chevy Chase, MD 20815

Author contributions: S.A.G. designed research; S.A.G. performed research; S.A.G. and N.C.E. analyzed data; and S.A.G. and N.C.E. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2024 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution License 4.0 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).

¹To whom correspondence may be addressed. Email: sgoldstein@genetics.utah.edu or nelde@genetics.utah.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2312691121/-/DCSupplemental>.

Published January 26, 2024.

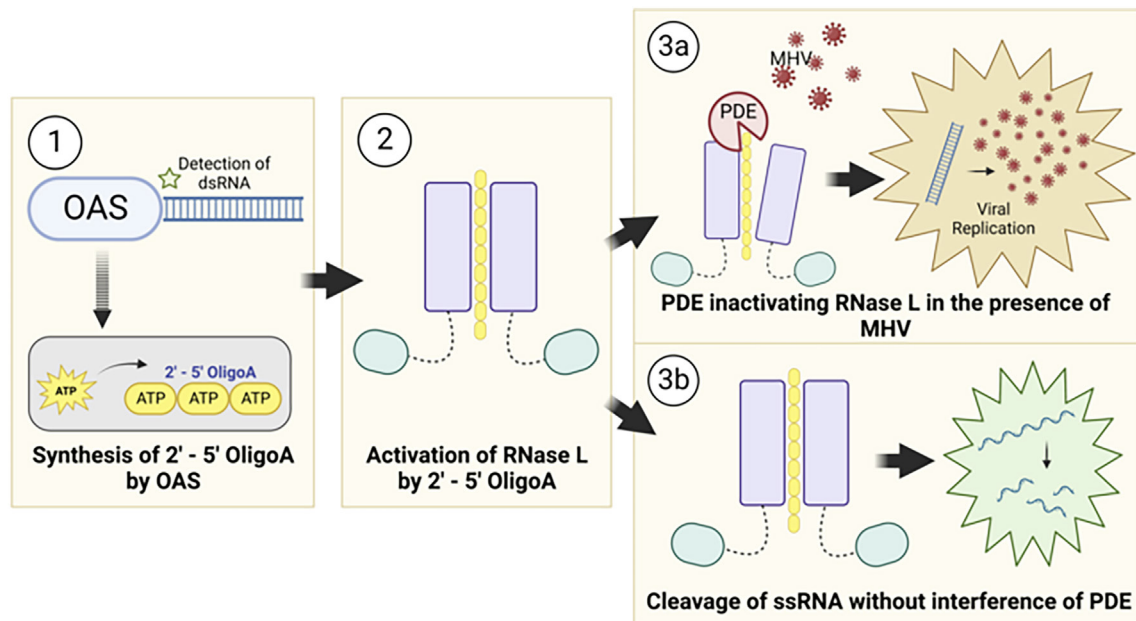


Fig. 1. Viral PDE antagonism of OAS-RNase L. (1) OAS proteins sense viral dsRNA upon infection and synthesize 2'-5' oligoadenylate (2-5A) by linking ATP molecules via the 2' and 5' carbons. (2) 2-5A binds RNase L monomers, inducing homodimerization and activation. (3a) Viral PDEs prevent RNase L activation by cleavage of 2-5A, facilitating viral replication. (3b) In the absence of a functional PDE, RNase L is activated and cleaves both viral and cellular ssRNA, restricting viral replication, inducing translational arrest, and ultimately resulting in cell death.

In contrast to the extensive research on PDE function, the origin and evolutionary history of viral PDEs remain obscure. Based on structural similarity, the PDEs encoded by mouse hepatitis virus (18) and group A rotaviruses (15, 19) have been linked to the PDE domain of mammalian AKAP7 proteins. Previously predicted structures for PDEs encoded by MERS-CoV (11), toroviruses (14) and group B and G rotaviruses (15) reveal an AKAP7 PDE-like fold and retain catalytic histidine (HxS/T) motifs.

The interface between viral PDEs and the OAS-RNase L pathway is an attractive opportunity to study evolutionary pressures related to the acquisition, fixation, and evolution of host-derived viral genes. The restriction of rotavirus (13) and coronavirus (10, 12, 17) replication in the absence of an active PDE along with the presence of these genes across diverse viruses provide strong evidence for potential fitness advantages conferred by PDE acquisition.

The capacity for viral genomes to accommodate new genomic material is a key determinant of their ability to evolve via HGT. Considered broadly, the genomic architecture of viral PDEs is of two types. PDEs encoded by rodent alphacoronaviruses (20), embecoviruses (14), and merbecoviruses (11) are ~240 to 300 amino acids long, translated off of a discrete viral mRNA, presumably exist in the cell as autonomous proteins, and exhibit high amino acid sequence divergence relative to conserved regions of the replicase gene. However, PDEs may also act as functional protein sub-domains of as few as ~115 amino acids. The rotavirus A PDE consists of 142 amino acids, rotavirus B PDE 115 amino acids, and rotavirus G PDE 116 amino acids, with all organized as the C-terminal subdomain of the >750 amino acid VP3 protein. The torovirus PDE is a 141 amino acid C-terminal domain of the >4,000 amino acid polyprotein 1a, though whether it is proteolytically processed into an autonomous protein is unclear.

In this study, we identified several striking features at play in the evolution of viral PDEs. We used a combination of phylogenetics, conserved motif analysis, and AlphaFold-based structural predictions to characterize the origin of viral PDEs, their diversification following horizontal transfer, and the relationship of divergent viral PDEs to each other and mammalian AKAP7. Despite similarity

in function and predicted structure with each other and mammalian AKAP7, we exclude the possibility of a single horizontal transfer event giving rise to viral PDEs and instead propose a more complex and volatile process of diversification. These findings advance our understanding of the history of viral PDEs as diverse and potent immune antagonists and provide a detailed characterization of how host-derived viral genes originate and evolve.

Results

Viral PDEs Derive from Several AKAP-Like Horizontal Transfer Events. We aligned 173 viral PDE amino acid sequences (18 torovirus, 10 alphacoronavirus, 55 embecovirus, 27 merbecovirus, 8 rotavirus G, 20 rotavirus B, and 35 rotavirus A) using MAFFT (version 7), which due to high sequence divergence involved intensive manual refinement (see *Methods* for details). Next, we constructed an unrooted maximum-likelihood (ML) phylogenetic tree (Fig. 2A). The PDE tree contains five major branches reflective of five putative acquisition events: rotavirus A (RVA) VP3, the rotavirus B (RVB) and rotavirus G (RVG) VP3 single common ancestor, toroviruses, embecoviruses, and merbecoviruses. The arrangement of some branches suffered low bootstrap support likely owing to sequence divergence, but the topology was sufficiently robust to reveal major events in viral PDE evolution.

We found evidence for three independent acquisitions of PDEs by ancestral nidoviruses. The nidovirus PDE phylogeny is incongruent with the nidovirus RdRp phylogeny (Fig. 2B) and the PDEs are not syntenic between virus groups, with one exception. Consequently, we found no support for a recent common ancestor of all nidovirus PDEs. In contrast to the three acquisitions of an ancestral cellular PDE, the rodent alphacoronavirus PDE and the embecovirus PDE are syntenic and have high sequence identity. This synteny suggests minimal homology-guided recombination played a role in the transfer of a PDE gene from one to the other. Embecovirus PDEs exhibit higher within-clade diversity, suggesting an older origin and providing support for the hypothesis that alphacoronavirus PDEs were acquired via recombination with embecoviruses (20), in contrast to

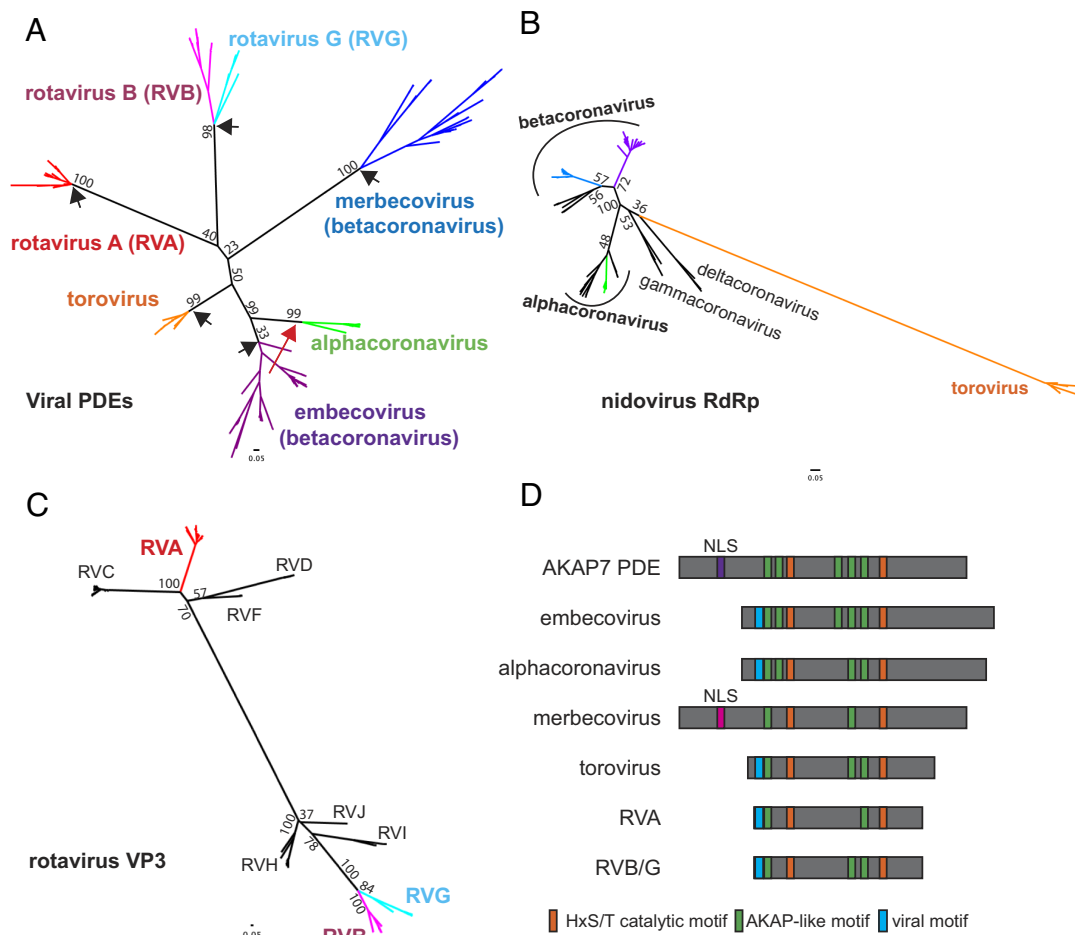


Fig. 2. Phylogenetic and sequence analysis of viral PDEs reveals multiple independent acquisitions and conservation of ancestral motifs. (A) Unrooted ML Tree of viral PDE amino acid sequences showing five PDE acquisitions from a cellular ancestor (black arrows) and one acquisition via recombination (red arrow). (B) Unrooted ML tree of coronavirus and torovirus full-length RdRp amino acid sequences. (C) Unrooted ML tree of rotavirus VP3 amino acid sequences, with the C-terminal domain PDEs trimmed from the RVA, RVB, and RVG sequences. (D) Schematic showing the conservation or absence of AKAP-like motifs in viral PDEs.

the idea that they were acquired independently (Fig. 2A) (21). Given the overlapping ecological niches of rodent betacoronaviruses and alphacoronaviruses, we cannot exclude the possibility that ancestral viruses in each clade independently acquired a similar cellular PDE. Given the greater intra-clade divergence among embecovirus which supports an older origin; however, it is unlikely independently acquired PDEs would retain such high identity.

The rotavirus PDE phylogeny supports with high confidence two independent PDE gene captures—one on the RVA VP3 branch and one into an RVB/RVG VP3 common ancestor. We constructed a ML tree of 137 rotavirus VP3 genes representing all known rotavirus genotypes (RVA, RVB, RVC, RVG, RVH, RVI, and RVJ) with the PDE domain trimmed from the RVA, RVB, and RVG VP3s (Fig. 2C). The VP3 (with the PDE trimmed off) phylogeny matches the PDE phylogeny where RVA VP3 is extremely distant from the RVB and RVG VP3s, which derive from a recent common ancestor. This supports a parsimonious scenario in which that ancestor acquired a PDE which diversified at the same rate as the rest of the gene. Surprisingly, despite the synteny of the RVA and RVB/RVG PDEs, their origins are clearly independent of each other.

Previous work identified mammalian AKAP7 central domains (CD) as structural and functional homologs of viral PDEs (22, 23). However, low sequence conservation leaves the evolutionary link between the viral and cellular PDEs uncertain. While generating the viral PDE phylogeny (Fig. 2A), we identified several AKAP7-like motifs largely conserved in the viral PDEs (Fig. 2D)

that proved critical to anchoring the low-identity alignment and conclusively link the viral PDEs to a cellular AKAP7-like ancestor, as no other eukaryotic PDEs share these motifs. One short motif (blue in Fig. 2D) is shared among all viral PDEs with the exception of merbecoviruses but is not present in any AKAP7. Though convergent evolution is formally possible, it is less parsimonious than this being another motif derived from the AKAP7-like ancestor. The motif sequence is AD(D/K/Q), amino acids that are enriched in the mammalian full-length AKAP7 AA 1–48 region that includes the nuclear localization sequence (NLS) (AA 40–47), suggesting the conserved viral motif is a remnant of N-terminal truncation that was repeatedly selected for during optimization of the OAS-RNase L suppression activity (22).

Finally, a PDE encoded by ORF4 of a shrew alphacoronavirus was recently reported (21), raising the possibility of an additional independent PDE acquisition, or virus-to-virus horizontal transfer. Its phylogenetic positioning remains uncertain on the end of a long branch with very low support due to sparse sampling (SI Appendix, Fig. S1A). Several of the key motifs, including the second catalytic motif, are degraded, indicating that this protein is almost certainly non-functional (SI Appendix, Fig. S1B). This observation further highlights the prevalence of PDE gene capture by HGT.

Ancestral Sequence Reconstruction of Nidovirus PDEs. Our phylogenetic analysis revealed clear distinctions between nidovirus PDE clades, supporting independent acquisition of PDEs by

merbecoviruses, embecoviruses, and toroviruses. Nevertheless, confidence is clouded by the high sequence divergence among PDEs and the shared, recombination-driven history of embecovirus and alphacoronavirus PDEs. Consequently, we performed additional analysis combining ancestral sequence reconstruction with AlphaFold2 structural modeling to further investigate relationships among viral PDEs (24). We initially focused on embecovirus NS2 and merbecovirus NS4b proteins given the higher degree of within-clade sequence diversity among these proteins, which likely indicates a combination of more thorough sampling and an older acquisition. We aligned 55 embecovirus NS2 sequences with the *HsAKAP7* PDE domain. We then constructed a ML tree of embecovirus NS2 protein sequence with an *HsAKAP7* outgroup rooting and used this tree as the basis for ancestral reconstruction in FastML (Fig. 3A). We aligned the *HsAKAP7* and ancestral embecovirus NS2 sequences from major nodes and calculated amino acid identity. The deepest ancestral node exhibited substantially higher amino acid identity to *HsAKAP7* than shallower nodes or observed extant sequences.

We conducted the same analysis with 28 merbecovirus NS4b sequences, which was of particular interest due to the wide range in size of these proteins (246–285 AA), largely due to a highly variable N-terminus. Ancestral sequence reconstruction produced an AncMerbecovirus (AncMerbeco) NS4b with 21.7% identity to *HsAKAP7*, again substantially higher than for ancestral sequences at reconstructed nodes or among observed sequences (Fig. 3B). The AncMerbeco NS4b is 246 AA long, the same as the NS4b encoded by human MERS-CoV isolates and significantly shorter than the 285 AA NS4b encoded by most HKU4 isolates. We also reconstructed ancestral sequences of alphacoronavirus and torovirus PDEs. In the case of alphacoronaviruses, this allowed comparisons not only with *HsAKAP7* but also with ancestral embecovirus PDEs, critical for defining the relationship between them (SI Appendix, Fig. S2).

Nidovirus PDE Structural Models Show Similar Divergence from *HsAKAP7*. Using proposed ancestral sequences, we generated nidovirus PDE structural models and compared them to the

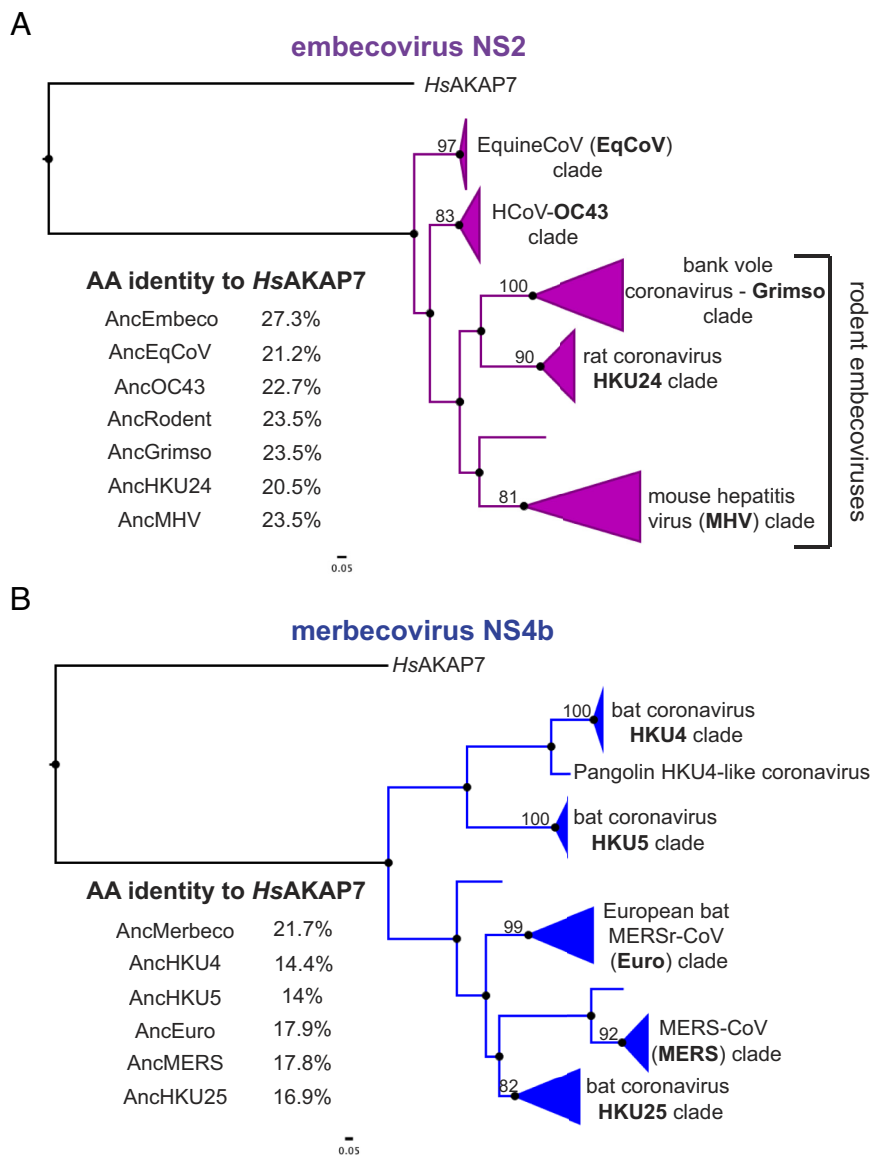


Fig. 3. Ancestral sequence reconstruction (ASR) of embecovirus and merbecovirus PDEs. (A) ML tree of 55 embecovirus PDE sequences, collapsed into major clades, rooted with the *HsAKAP7* PDE domain sequence (AA 82–287) and sequence identity to *HsAKAP7* of reconstructed sequences at major nodes. (B) ML tree of 27 merbecovirus PDE sequences, collapsed into major clades, rooted with the *HsAKAP7* PDE domain sequence (AA 82–287) and sequence identity to *HsAKAP7* of reconstructed sequences at major nodes.

structure of human (*Hs*)AKAP7 PDE to determine whether there were differences in RMSD, a quantitative estimate of the similarity of structures. Overall amino acid identity was the same between AncEmbeco (Fig. 4A) and AncAlphaCoV NS2 proteins (Fig. 4B) compared to *Hs*AKAP7 (~27%) and RMSDs relative to the *Hs*AKAP7 structure were nearly identical, indicating similar divergence from a cellular ancestor and suggesting that they may have a common origin, supporting a scenario of HGT between these clades. In contrast, AncMerbeco NS4b has the lowest amino acid identity to *Hs*AKAP7 (~21%) and highest RMSD (Fig. 4C), supporting its independent origin which, given the substantial within-clade diversity of these proteins (Fig. 2A), indicates an ancient acquisition event. In contrast, the torovirus PDE encoded on the 3' end of ORF1a has relatively high identity to *Hs*AKAP7 PDE (~39%) and the lowest RMSD value (Fig. 4D). Coupled with the limited within-clade diversity among torovirus PDEs (Fig. 2A), this suggests that the torovirus PDE was acquired independently and more recently than the other nidovirus PDEs.

A potentially complicating feature of the analysis is indels in viral PDE sequences. Relative to the *Hs*AKAP7 PDE structure, which comprises just the central 206 AA of the 315 AA *Hs*AKAP7 protein, nidovirus PDEs have additional domains at the N-termini (merbecovirus) or C-termini (embecovirus and alphacoronavirus). To account for the possibility that RMSD values might be inflated by poor modeling of these domains, we predicted the structure of the core PDE domains of nidovirus PDEs. The AncMerbeco core

PDE was not successfully predicted by AlphaFold2, so we used the sequence from a more recent node, the AncMERS core PDE. Overall, similarity to *Hs*AKAP7 PDE was generally unchanged regardless of whether the full-length or core viral PDE structural model was analyzed (SI Appendix, Fig. S3), bolstering the value of comparing full-length PDE structural models despite variable and disordered termini.

Motif and Structural Analysis Support Embecovirus-to-Alpha-coronavirus Horizontal Transfer of NS2. Our prior phylogenetic and structural analyses were consistent with a relationship between embecovirus and alphavirus PDEs, such that one was derived from the other via recombination. Because these viruses fall into different coronavirus genera, a single introduction event is implausible, leaving the possibility of a recombination event or acquisition of a similar cellular ancestor. Given the longer branch lengths in the embecovirus PDE phylogeny, we hypothesize the acquisition event occurred in ancestral embecoviruses and that ORF2 was subsequently transferred to the alphacoronavirus clade. To refine our understanding of this history more deeply, we conducted additional sequence and structural analysis.

We first analyzed conservation of key motifs between AncEmbeco NS2 and the reconstructed ancestral PDE sequences from the other viral clades. Compared to AncEmbeco NS2, AncAlphaCoV NS2 had the fewest amino acid differences, with only three substitutions outside the catalytic His motif in which the second residue

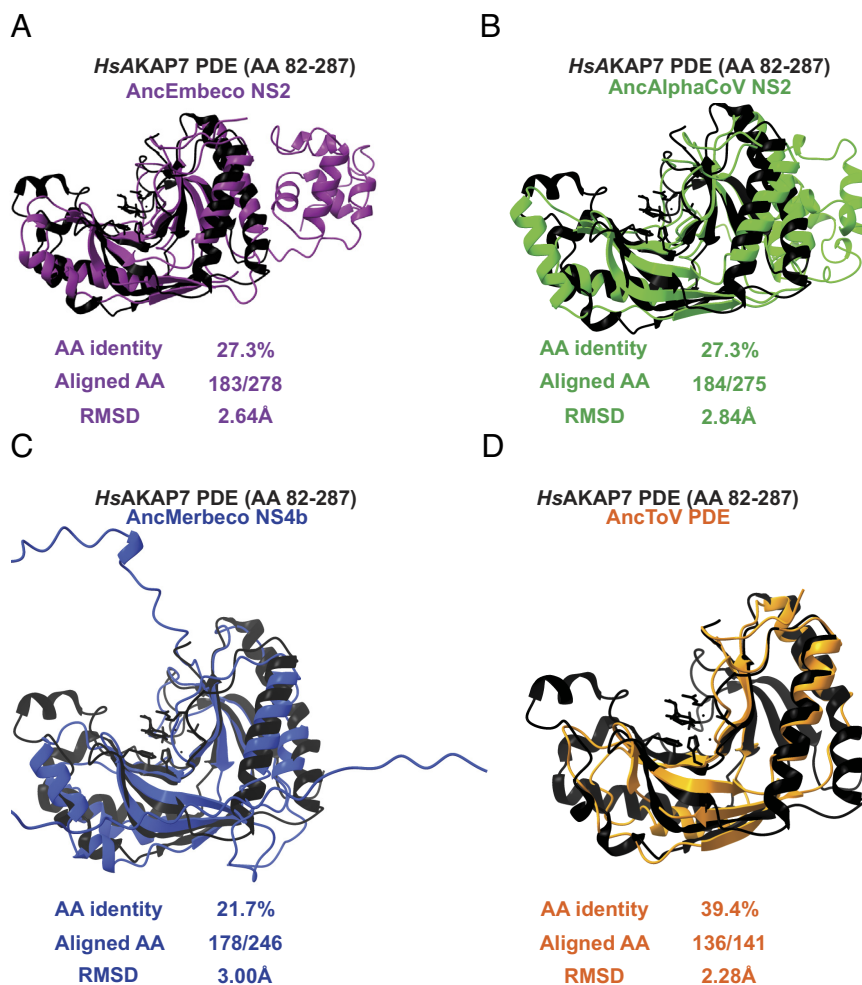


Fig. 4. Structural similarity between *Hs*AKAP7 and ancestral nidovirus PDEs. (A–D) Overlay of the previously determined *Hs*AKAP7 structure with the indicated ancestral nidovirus PDE predicted structures. The number of aligned amino acids (from the query sequence) and the RMSD were calculated using FatCat.

frequently toggles among PDEs (Fig. 5A), supporting the phylogenetic inference that the embecovirus and alphacoronavirus NS2 proteins are closely related. We then conducted a series of overlays of AlphaFold2-predicted structural models of reconstructed PDEs as well as the known *HsAKAP7* PDE structure. Specifically, we used a core PDE domain structural model to capture similarities/differences in the PDE fold itself and avoid disordered regions or N-terminal/C-terminal domains that AlphaFold2 failed to model. Modeling the core region better approximates the *HsAKAP7* PDE structure which was solved for just the 206 AA PDE central domain, not the entire protein.

The RMSD recovered by comparing the AncEmbeco and AncAlphaCoV NS2 models was comparable to the RMSD of AncEmbeco and AncHKU24, from a more recent inferred node in the embecovirus PDE phylogeny (Fig. 5B and C), although amino acid identity was substantially lower. Comparing the AncEmbeco and AncAlphaCoV NS2 models to the *HsAKAP7* PDE structure

revealed that the viral PDEs are much closer to each other in amino acid identity and RMSD than either is to *HsAKAP7* PDE (Fig. 5D and E). Although it's possible that this is due to convergence in their structural evolution, the combination of higher amino acid identity and structural model similarity strongly support a common origin, which resulted from recombination. Finally, we compared the AncEmbeco NS2 model to the AncMERS NS4b model (AlphaFold2 failed to predict the AncMerbeco core PDE structure), showing that their divergence is comparable to that between the NS2 PDEs and *HsAKAP7* PDE (Fig. 5F) and supporting a single origin of the embecovirus and alphacoronavirus PDEs.

Merbecovirus PDEs Exhibit Significant Intra-Clade Diversification.

In addition to untangling the origin of viral PDEs, we sought to understand how they evolved after acquisition as new virus-encoded proteins, compared to their eukaryotic ancestors. Merbecovirus NS4b stands out among other viral PDEs for its

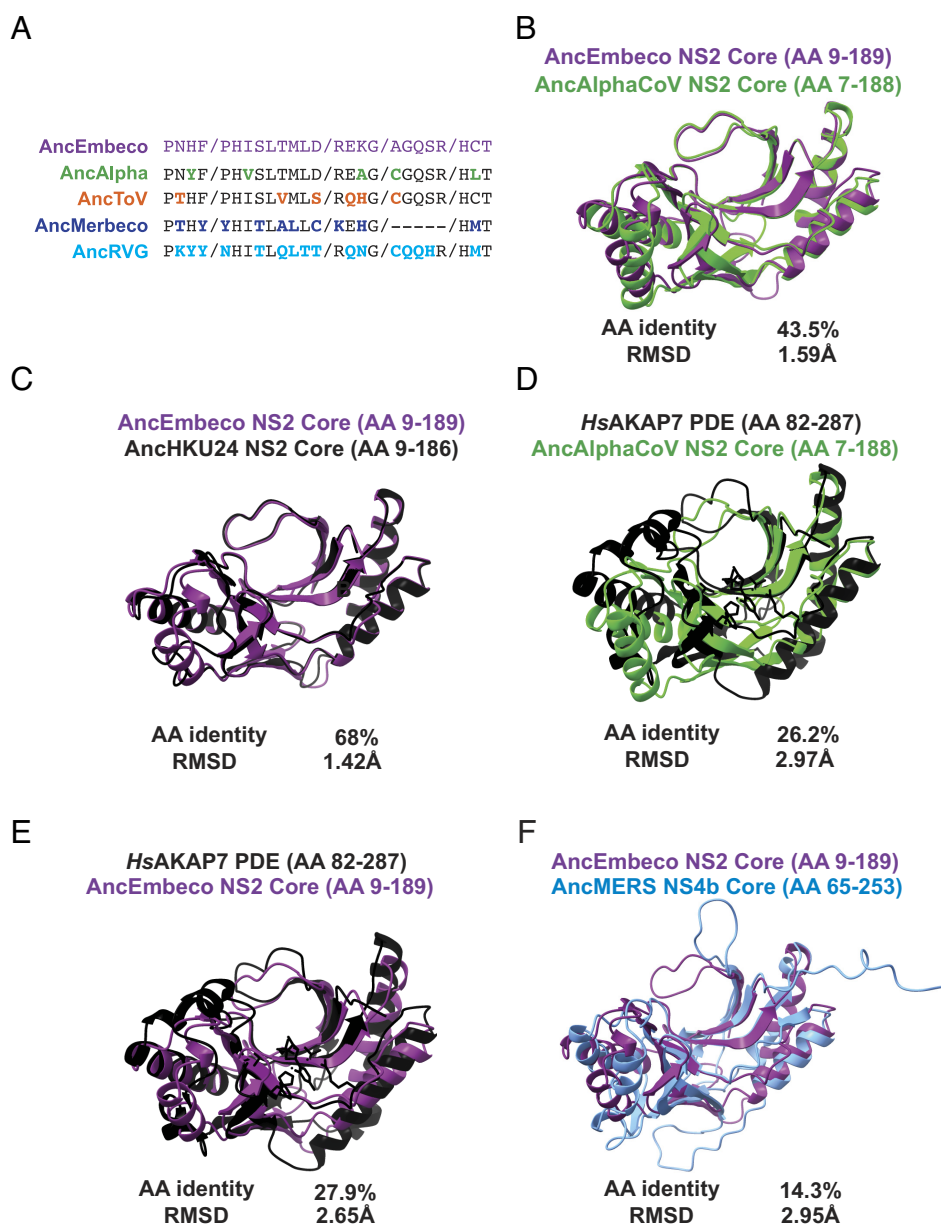


Fig. 5. Comparative analysis of AncEmbeco and AncAlphaCoV PDEs. (A) Motif analysis with AncEmbeco as the reference. Amino acid differences in key motifs are bolded and in the color corresponding to the viral PDE. (B) Overlay and RMSD comparison of AncEmbeco NS2 and AncAlphaCoV NS2. (C) Overlay and RMSD comparison of AncEmbeco NS2 and AncHKU24 NS2. (D) Overlay and RMSD comparison of *HsAKAP7* PDE and AncAlphaCoV NS2. (E) Overlay and RMSD comparison of *HsAKAP7* PDE and AncEmbeco NS2. (F) Overlay and RMSD comparison of AncEmbeco NS2 and AncMERS NS4b.

striking intra-clade diversification and structural divergence. In contrast, the embecovirus PDE phylogeny exhibits shorter branch lengths (Fig. 2B) and embecovirus PDEs have higher sequence identity with each other, even when accounting for greater sampling in the dataset. This supports their use as a comparator with merbecovirus NS4b. Relative to the AncEmbeco NS2, two reconstructed sequences at recently inferred nodes have high sequence conservation (Fig. 6A) and the modeled structures have a low RMSD value, suggesting minimal structural change throughout the course of embecovirus NS2 evolution (Fig. 6B). We also compared how the radiation of embecovirus NS2 altered structure across the phylogeny by overlaying the modeled structures of the reconstructed NS2 PDEs ancestral to clades defined by the prototypical betacoronavirus mouse hepatitis virus (AncMHV) and bank vole (*Myodes glareolus*) betacoronavirus isolate-Grimso (AncGrimso) (Fig. 6C). This confirmed that PDE

structural divergence is relatively constrained even among diverse embecoviruses.

In contrast, the merbecovirus NS4b has undergone significantly greater sequence divergence and structural variation. There has been notable extension, truncation, insertion, and deletion in the N terminus of the protein (Fig. 6D), likely facilitated by the fact that NS4b is encoded on a bicistronic ORF4ab viral subgenomic mRNA, which enables it to tolerate mutations that move the start codon upstream or downstream of the ancestral site within the overlapping region with ORF4a. However, even accounting for sequence divergence, the structural variation among merbecovirus NS4b proteins is remarkable. AncMerbeco and AncHKU25 NS4b (Fig. 6E); reconstructed at a less basal node) exhibit 69.4% amino acid identity, comparable to AncEmbeco and AncGrimso (Fig. 6B), yet their RMSD is 2.71Å compared to 1.52Å for the embecovirus NS2 proteins and the deviation of reconstructed

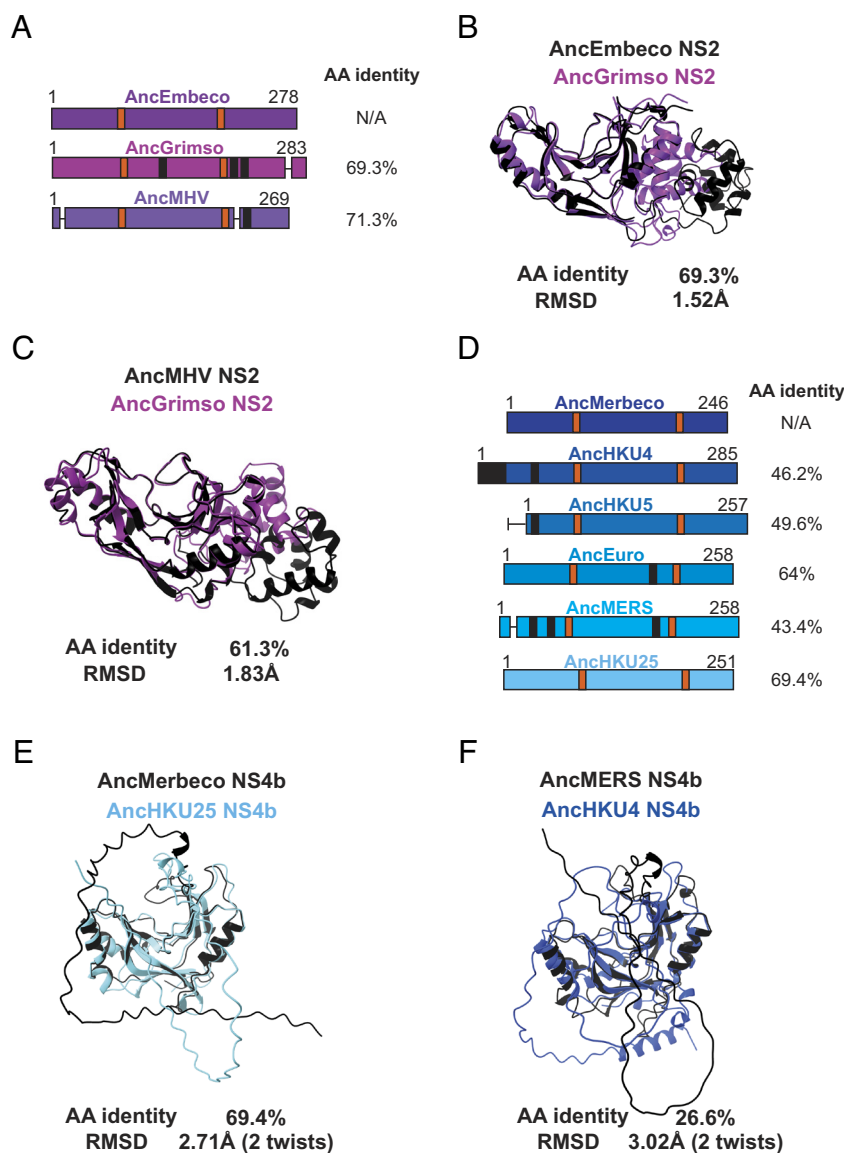


Fig. 6. Variable sequence and structural PDE diversity within Nidovirus subgenera. (A) Schematic of embecovirus NS2 at key ancestral nodes with catalytic motifs (orange), deletions (horizontal lines), and insertions (black) indicated along with protein length and amino acid identity to the AncEmbeco sequence. (B) Overlay of AlphaFold predicted structures of the reconstructed AncEmbeco and derived AncGrimso sequences with amino acid identity and the RMSD as calculated in FatCat. (C) Overlay of AlphaFold predicted structures of the AncMHV and AncGrimso sequences with amino acid identity and the RMSD as calculated in FatCat. (D) Schematic of merbecovirus NS4b at key ancestral nodes with catalytic motifs (orange), deletions (horizontal lines), and insertions (black) indicated along with protein length and amino acid identity to the AncMerbeco sequence. (E) Overlay of AlphaFold predicted structures of the AncMerbeco sequence and a shallower node (AncHKU25) with similar AA identity to each other as between the sequences in panel B, with RMSD from a FatCat comparison. (F) Overlay of AlphaFold predicted structures of shallower ancestral nodes with different AA sequence identity and RMSDs from FatCat comparisons.

sequences at recently inferred nodes of the embecovirus phylogeny is still 1.83Å (Fig. 6C). The reconstructed sequences of AncMERS and AncHKU4 NS4b exhibit just 26.6% amino acid identity and an average distance of deviation >3 Å, reflecting marked diversification among merbecovirus PDEs (Fig. 6F).

Sequence and Structural Analysis of Rotavirus PDEs Supports Two Horizontal Transfer Events. Rotavirus PDE and VP3 phylogenies are consistent with independent acquisitions of PDEs in the RVA VP3 ancestor and in a common ancestor of RVB and RVG VP3 (Fig. 2 A and C). As with the nidovirus PDEs, ancestral sequence reconstruction was successful in producing sequences with the highest amino acid identity to *HsAKAP7* at the deepest node of the tree (Fig. 7 A and C).

The AncRVA and AncRVB/G PDEs exhibited similar amino acid identity and RMSDs to *HsAKAP7* of 34 to 36% and ~2.4 to 2.8Å (Fig. 7 B and D). However, their amino acid identity to each other was considerably lower than to *HsAKAP7* PDE and their modeled structures were also relatively dissimilar, with RMSD >3Å (Fig. 7E). This result strongly suggests that these PDEs are similarly diverged from *HsAKAP7* and each other, consistent with multiple origins of the rotavirus PDEs despite their synteny. A comparison of modeled structures of reconstructed RVB and RVG sequences reveal a stark contrast (Fig. 7F). These reconstructed PDE sequences had nearly 50% amino acid identity and an RMSD <1Å, consistent with the phylogenetic inference that the RVB and RVG PDEs arose from a single horizontal transfer event in a common ancestral VP3.

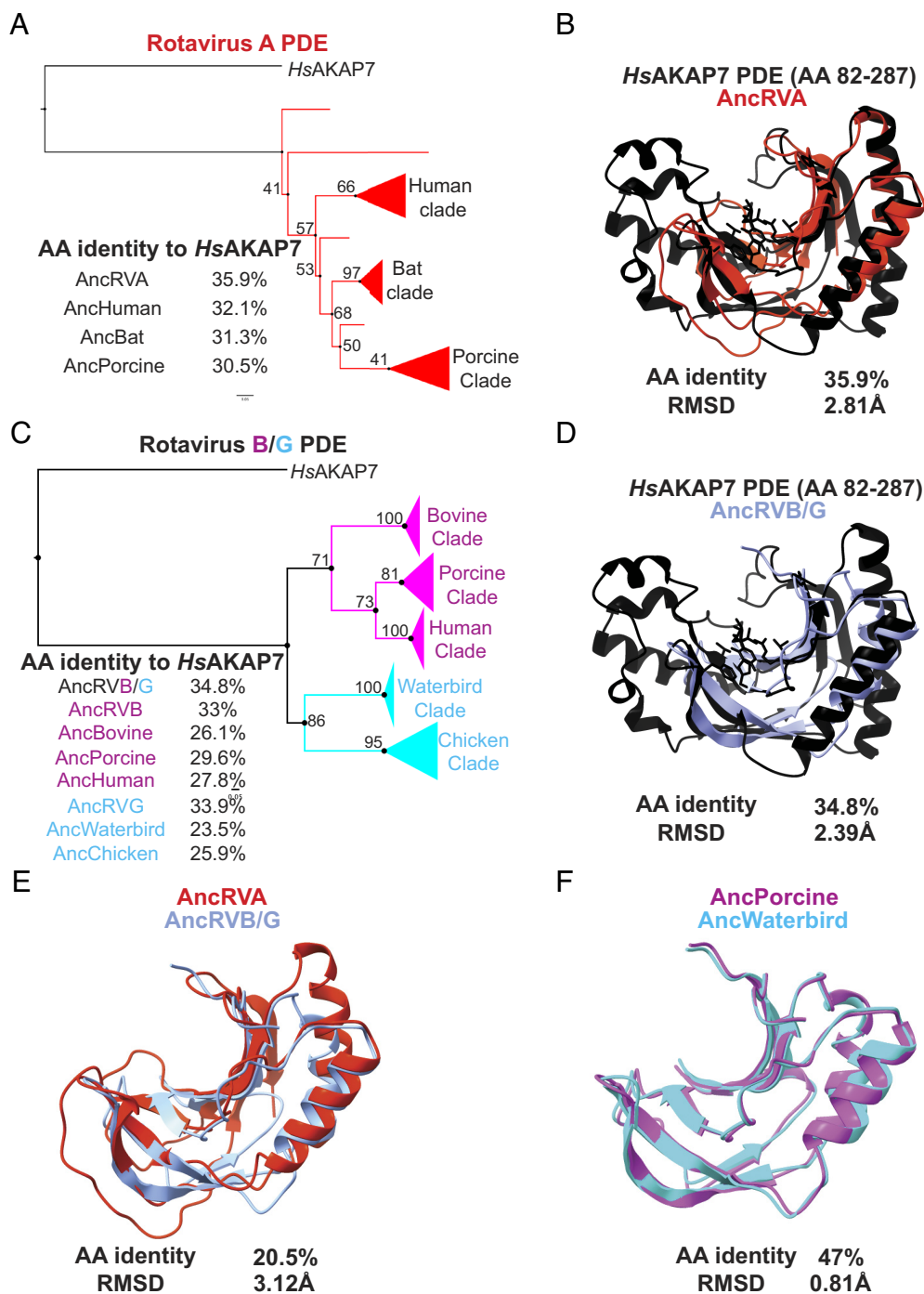


Fig. 7. Ancestral sequence reconstruction and structural analysis of rotavirus PDEs. (A) Ancestral sequence reconstruction (ASR) of RVA PDEs at major nodes with AA identity to *HsAKAP7*. (B) Overlay of the AncRVA PDE structure predicted in AlphaFold with the solved structure of the *HsAKAP7* PDE. (C) ASR of RVB/G PDEs at major nodes with AA identities to the *HsAKAP7* PDE sequence. (D) Overlay of the AncRVB/G PDE structure predicted in AlphaFold with the solved structure of the *HsAKAP7* PDE. (E) Overlay of the AncRVA PDE structure modeled in AlphaFold with the structural model of the AncRVB/G PDE. (F) Overlay of the AncRVB Porcine clade PDE model structure and AncRVG Waterbird clade PDE structural model.

Loss of PDE Function in Rotavirus G Is Not Correlated with Major Structural Changes.

A previous study compared enzymatic activity of PDEs encoded by RVA, RVB, and RVG and found that the PDE of the RVG tested was inactive (15). Although a T90N amino acid substitution in the second catalytic motif is associated with loss of enzymatic activity, reversion of the substitution did not rescue function. These findings point to cryptic functional determinants outside the catalytic sites, as previously reported for MHV NS2 (25). The AncRVG and AncChicken RVG PDEs contain identical key motifs and threonines at position 90 (T90; Fig. 8A). In contrast, RVG^{inactive} and other closely related viruses encode substitutions at position 90 and in other conserved motifs (Fig. 8A). To determine whether structural divergence might underlie loss-of-function, we compared structural models of RVB^{active} and RVG^{inactive} PDEs, which are similarly diverged from the AncRVB/G model in both sequence identity and RMSD (Fig. 8B and C). We found that modeled structures of RVB^{active} and RVG^{inactive} PDEs are similar (Fig. 8D) and therefore no obvious structural changes can explain the RVG PDE loss of function. Whether amino acid changes in the key conserved motifs are involved is unclear and an important topic for future experimental study.

Discussion

In this study, we traced and characterized the acquisition and evolution of viral PDEs. These proteins are among the clearest examples of acquisition of a host gene by RNA viruses, in contrast to more abundant examples among DNA viruses. Previous work examining the role of gene flow in RNA virus evolution has largely considered conserved genes such as capsids, RNA-dependent RNA polymerases, and helicases (26). In contrast, we characterize an example of gene transfer shaping virus interactions with host immunity, with potential implications for host-range and virulence. A recent report proposing bacterial origins for eukaryotic OAS genes, suggests that both sides of the interface between viral PDE and OAS-RNase L arose via HGT (27). Understanding how HGT has influenced virus-innate immune interfaces benefits from a better understanding of the emergence of newly acquired viral genes and their subsequent evolution. Here, we took advantage of recent advances in protein structure prediction to study the evolution of viral PDEs.

The amino acid sequence of viral PDEs can undergo extensive substitution without loss of function so identity between viral and host sequences is often very low. This complicates phylogeny-only approaches to understanding the history of these proteins, which motivated our integration of sequence and structural prediction-based comparisons. Previous work has linked these proteins to an AKAP7-like ancestor (2, 23) based on similarity to solved *HsAKAP7*, MHV NS2, and RVA PDE structures (15, 18, 19, 28), but these analyses provide only a limited view of PDE evolution.

Our phylogenetic analysis is a major expansion on prior efforts (23), incorporating 173 PDE sequences from embecoviruses, merbecoviruses, toroviruses, alphacoronaviruses, RVAs, RVBs, and RVGs. The structure of the nidovirus PDE tree, including several long branches incongruent with the RdRp phylogeny, suggests multiple independent horizontal transfer events and one virus-to-virus transfer via recombination (Fig. 2B). Because of the phylogenetic incongruence between the nidovirus PDE and RdRp trees, there is no support to infer a single introduction into the nidovirus or coronavirus ancestor followed by loss on some branches. The relatively large RMSD values between embecovirus and merbecovirus PDEs and the absence of synteny further argue against a single introduction into even the betacoronavirus common ancestor. Comparison of the rotavirus PDE and C-terminal

truncated VP3 phylogenies supports two introductions, one into an ancestral RVA VP3 and one into an RVB/G VP3 common ancestor (Fig. 2C). The syntenic but independent acquisition of PDEs by rotaviruses suggests that the VP3 segment is particularly tolerant of gaining new genetic material. Future studies could test this idea by experimentally recapitulating horizontal transfer events using the established rotavirus A molecular clone (13, 29).

Our characterization of several horizontal transfer events of orthologous ancestral AKAPs and their common use in antagonism of the OAS-RNase L pathway has intriguing biological implications. The cellular role of the AKAP7 PDE remains mysterious, so whether these proteins were evolutionarily repurposed after viral acquisition is currently unclear. Importantly, these horizontal transfer events are recurrent but phylogenetically restricted, occurring multiple times only in rotaviruses and nidoviruses despite the fact that the OAS-RNase L pathway restricts diverse viral families (9). Among RNA viruses, constraint on genome length due to error rate and packaging capacity may be a limiting factor (30), which is alleviated among coronaviruses by their proofreading exonuclease and resultant large genomes.

The ability of some coronaviruses to tolerate insertions of PDE genes raises the question of why other coronavirus subgenera, including sarbecoviruses such as SARS-CoV-2, lack PDEs. Two possibilities can be proposed. One is that OAS-RNase L activity, and thus the selective pressure to encode an antagonist is cell-type specific (17, 31). The cellular tropism of sarbecoviruses in their natural rhinolophid hosts is unknown, but it is possible that they primarily infect cells that lack a potent OAS-RNase L pathway, similar to MHV and liver hepatocytes (17), or cells that specifically do not express the prenylated p46 isoform of OAS1 that has been linked to restriction of SARS-CoV-2 replication (32–36). Alternatively, rhinolophid bats may have acquired loss-of-function mutations in OAS genes, as seen with primate OAS1 (37), eliminating the pressure for sarbecoviruses to antagonize the pathway.

Merbecovirus NS4b highlights the potential for highly variable evolutionary trajectories following acquisition of host PDE genes. They exhibit striking sequence divergence and structural deviation from both *HsAKAP7* and other viral PDEs and extensive divergence between each other. Relative to other viral PDEs and like AKAP7 proteins, merbecovirus PDEs contain a NLS and appear in the nucleus in abundance (12, 38, 39). However, the merbecovirus NLS is bipartite in contrast to the monopartite AKAP7 NLS and shows no evidence of being derived from the cellular ancestral NLS, suggesting it arose *de novo* early after acquisition of the PDE. Concordant with NS4b localization to the nucleus, this PDE has additional functions in immune antagonism distinct from disruption of RNase L activation (12, 38–43). This multi-functionality distinguishes NS4b from all other viral PDEs and likely subjects NS4b to unique selective pressures, presumably accounting for its volatile evolutionary history.

As the evolutionary history of viral PDEs comes into better focus, our understanding of the roles for PDEs in viral replication and impacts on fitness remains limited. With the exception of MHV, any fitness benefits provided by coronavirus PDEs has not been explored, especially during infection and transmission of viruses in natural hosts. When bat coronaviruses move into new hosts the loss of some accessory protein function is common, suggesting the advantages provided are host specific. The development of more sophisticated cell culture systems from a variety of species, such as primary epithelial cells (44, 45) and bat intestinal organoids (46) opens the possibility of unlocking new insights into how the combination of HGT and natural selection shapes the evolution of RNA viruses and their hosts.

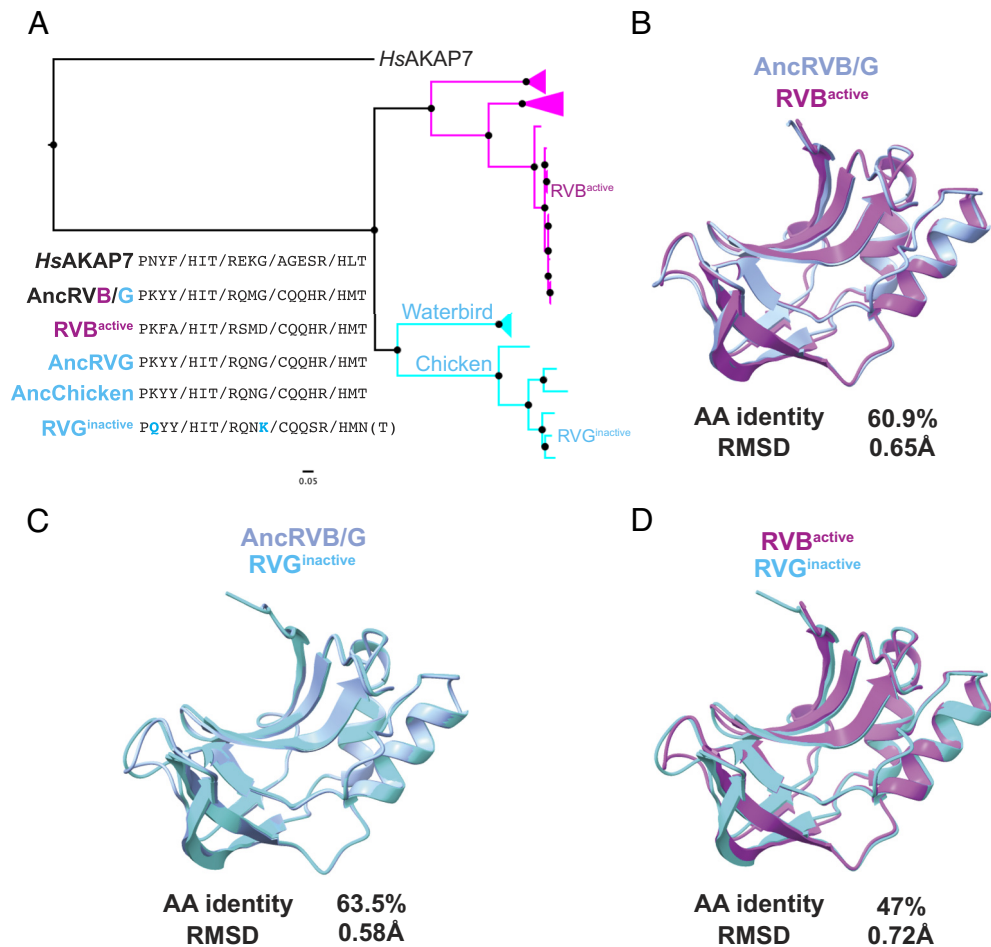


Fig. 8. Sequence motif and structural analysis of RVB and RVG functional divergence. (A) ML tree of RVB and RVG PDEs, with previously assayed PDEs (15) indicated. (B) Overlay and similarity metrics for the predicted structures of the AncRVB/G PDE and the active RVB PDE tested by Ogden et al. (15). (C) Overlay and similarity metrics for the predicted structures of the AncRVB/G PDE and the inactive RVG PDE tested by Ogden et al. (15). (D) Overlay and similarity metrics for the predicted structures of the active RVB PDE and the inactive RVG PDE tested by Ogden et al. (15). RMSDs for all overlays were calculated in FatCat.

Methods

Phylogenetic Analysis.

Viral PDEs. All alignments (*SI Appendix*) were generated using the MAFFT plug-in (47) with default parameters in Geneious Prime v2022.2.1. Alignments were manually refined as necessary. ML trees with 100 replicates were generated using IQ-Tree 2 (48) on the Los Alamos National Laboratory server. Different substitution models (with 20 rate categories) were necessary for different alignments to maximize tree stability and branch support. The all-viral PDE tree is unrooted, as no clear outgroup exists to all viral PDEs. ML trees for individual clades of viral PDEs are rooted with HsAKAP7. Alignments (.fasta*) and IQ-Tree logs (.log*) containing detailed parameters and associated with every PDE tree can be found in the Supplementary Files. The ViralPDEs.fasta file contains accessions for all viral PDE sequences analyzed.

Nidovirus RdRps. To generate the nidovirus RdRp tree (Fig. 1B), we aligned 145 nidovirus RdRp sequences representing toroviruses, alphacoronaviruses, betacoronaviruses, gammacoronaviruses, and deltacoronaviruses using MAFFT (default parameters) and generated a ML tree in IQ-Tree 2 using a GTR20 substitution model, 20 rate categories, and 100 bootstrap replicates. Raw files associated with the analysis are in *SI Appendix*.

Rotavirus VP3. We aligned 137 rotavirus VP3 amino acid sequences using MAFFT and trimmed the PDE domains from the RVA, RVB, and RVG sequences in the alignment. We then used IQ-Tree 2 to generate a ML tree using a JTT substitution model, 20 rate categories, and 100 bootstrap replicates.

ML tree visualization. All .treedata outputs from IQ-Tree 2 were imported into FigTree v1.4.4 for visualization and coloring. Branch support values and labels were added manually in Adobe Illustrator 2023.

Sequence Motif Analysis. We used previously defined motifs (14, 23) to identify motifs and guide manual scanning of the viral PDE/AKAP alignments.

Ancestral Sequence Reconstruction and Analysis. We generated alignments of each viral PDE group with the HsAKAP7 PDE domain (described above) and fed these alignments into FastML (49, 50) along with the ML tree for each alignment generated with IQ-Tree 2. Ancestral sequences were then reconstructed using a JTT substitution model and other default parameters (optimize branch length, Gamma distribution, marginal reconstruction) and exported for analysis. The raw files (outFile_seq_marginal.txt*, outTreefileAncestor.txt*, outTreefileNewick.txt*) are available on FigShare. After downloading the inferred ancestral sequences, we identified major nodes and aligned them to HsAKAP7 to confirm that deeper nodes have higher amino acid percent identity to the eukaryotic reference PDE. We also analyzed evolutionary trajectory in conserved AKAP-like motifs.

Structural Modeling. We exported major ancestral nodes from each viral PDE family as .fasta files and analyzed these using AlphaFold2 (v2.1.2) 9 as implemented at the University of Utah Center for High Performance Computing (CHPC) with the following parameters:

```
# explicit run of run_alphafold pointing to the appropriate databases
on the RAM disk, for file passed in as $1singularity exec --nv /uufs/chpc.
utah.edu/sys/installdir/alphafold/2.1.2/alphafold_2.1.2.sif /app/
run_alphafold.sh --data_dir=/scratch/general/lustre/app-repo/alpha-
fold --uniref90_database_path=/scratch/general/lustre/app-repo/
alphafold/uniref90/uniref90.fasta --mgnify_database_path=/scratch/
```

```

general/lustre/app-repo/alphafold/mgnify/mgy_clusters_2018_12.fa
--uniclust30_database_path=/dev/shm/af/uniclust30_2018_08 --
bfd_database_path=/dev/shm/af/bfd_metaclust_clu_complete_
id30_c90_final_seq.sorted_opt --pdb70_database_path=/dev/shm/
af/pdb70 --template_mmcif_dir=/scratch/general/nfs1/app-repo/
alphafold/pdb_mmcif/mmcif_files --obsolete_pdb_path=/scratch/
general/nfs1/app-repo/alphafold/pdb_mmcif/obsolete.dat --fasta_
paths=$1 --max_template_date=2020-05-14 --output_dir=temp_
module --use_gpu_relax

```

We then extracted the ranked_0.pdb model for each sequence for further analysis. Models were imported into and analyzed with UCSF ChimeraXv1.5–developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases (51, 52). Our primary analysis in ChimeraX involved overlays of predicted and solved structures to generate images that were then exported to Adobe Illustrator 2023.

- V. V. Dolja, E. V. Koonin, Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. *Virus Res.* **244**, 36–52 (2018).
- R. Mazumder, L. M. Iyer, S. Vasudevan, L. Aravind, Detection of novel members, structure–function analysis and evolutionary classification of the ²H phosphoesterase superfamily. *Nucleic Acids Res.* **30**, 5229–5243 (2002).
- P. T. Nga *et al.*, Discovery of the first insect nidovirus, a missing evolutionary link in the emergence of the largest RNA virus genomes. *PLoS Pathog.* **7**, e1002215 (2011).
- A. Saber, A. A. Gulyaeva, J. L. Brubacher, P. A. Newmark, A. E. Gorbalenya, A planarian nidovirus expands the limits of RNA genome size. *PLoS Pathog.* **14**, e1007314 (2018).
- Y. I. Wolf *et al.*, Origins and evolution of the global RNA virome. *mBio* **9**, e02329-18 (2018).
- Z. B. Zolinger, R. Elliott, K. M. Rose, S. R. Weiss, MDA5 is critical to host defense during infection with murine coronavirus. *J. Virol.* **89**, 12330–12340 (2015).
- E. Kindler *et al.*, Early endonuclease-mediated evasion of RNA sensing ensures efficient coronavirus replication. *PLoS Pathog.* **13**, e1006195 (2017).
- D.-M. Yang, T.-T. Geng, A. G. Harrison, P.-H. Wang, Differential roles of RIG-I like receptors in SARS-CoV-2 infection. *Mil. Med. Res.* **8**, 49 (2021).
- Y. Li *et al.*, Activation of RNase L is dependent on OAS3 expression during infection with diverse human viruses. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2241–2246 (2016).
- L. Zhao *et al.*, Antagonism of the interferon-induced OAS-RNase L pathway by murine coronavirus ns2 protein is required for virus replication and liver pathology. *Cell Host Microbe* **11**, 607–616 (2012).
- J. M. Thornbrough *et al.*, Middle east respiratory syndrome coronavirus NS4b protein inhibits host RNase L activation. *mBio* **7**, e00258 (2016).
- C. E. Comar *et al.*, Antagonism of dsRNA-induced innate immune pathways by NS4a and NS4b accessory proteins during MERS coronavirus infection. *mBio* **10**, e00319-19 (2019).
- Y. Song *et al.*, Reverse genetics reveals a role of rotavirus VP3 phosphodiesterase activity in inhibiting RNase L signaling and contributing to intestinal viral replication in vivo. *J. Virol.* **94**, e01952-19 (2020).
- S. A. Goldstein *et al.*, Lineage A betacoronavirus NS2 Proteins and the homologous torovirus berne pp1a carboxy-terminal domain are phosphodiesterases that antagonize activation of RNase L. *J. Virol.* **91**, e02201-16 (2017).
- K. M. Ogden *et al.*, Structural basis for 2′–5′-oligoadenylate binding and enzyme activity of a viral RNase L antagonist. *J. Virol.* **89**, 6633–6645 (2015).
- L. Zhao, L. D. Birdwell, A. Wu, R. Elliott, K. M. Rose, Cell-type-specific activation of the oligoadenylate synthetase–RNase L pathway by a murine coronavirus. *J. Virol.* **87**, 8408–8418 (2013).
- Y. Li, S. R. Weiss, Antagonism of RNase L is required for murine coronavirus replication in Kupffer cells and liver sinusoidal endothelial cells but not in hepatocytes. *J. Virol.* **90**, 9826–9832 (2016).
- B. Sui *et al.*, Crystal structure of the mouse hepatitis virus ns2 phosphodiesterase domain that antagonizes RNase L activation. *J. Gen. Virol.* **97**, 880–886 (2016).
- T. Brandmann, M. Jinek, Crystal structure of the C-terminal 2′,5′-phosphodiesterase domain of group A rotavirus protein VP3. *Proteins* **83**, 997–1002 (2015).
- W. Wang *et al.*, Extensive genetic diversity and host range of rodent-borne coronaviruses. *Virus Evol.* **6**, veaa078 (2020).
- D. Forni *et al.*, Homology-based classification of accessory proteins in coronavirus genomes uncovers extremely dynamic evolution of gene content. *Mol. Ecol.* **31**, 3672–3692 (2022).
- E. Gusheo *et al.*, Murine AKAP7 has a 2′,5′-phosphodiesterase domain that can complement an inactive murine coronavirus ns2 gene. *mBio* **5**, e01312-14 (2014).
- A. Asthana, C. Gaughan, B. Dong, S. R. Weiss, R. H. Silverman, Specificity and mechanism of coronavirus, rotavirus, and mammalian two-histidine phosphoesterases that antagonize antiviral innate immunity. *mBio* **12**, e0178121 (2021).
- J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
- S. M. Sperry *et al.*, Single-amino-acid substitutions in open reading frame (ORF) 1b–nsp14 and ORF 2a proteins of the coronavirus mouse hepatitis virus are attenuating in mice. *J. Virol.* **79**, 3391–3400 (2005).
- D. Bäckström *et al.*, Virus genomes from deep sea sediments expand the ocean megavirome and support independent origins of viral gigantism. *mBio* **10**, e02497-18 (2019).
- E. M. Culbertson, T. C. Levin, Eukaryotic antiviral immune proteins arose via convergence, horizontal transfer, and ancient inheritance. *PLoS Biol.* **21**, e3002436 (2023).
- M. G. Gold, F. D. Smith, J. D. Scott, D. Barford, AKAP18 contains a phosphoesterase domain that binds AMP. *J. Mol. Biol.* **375**, 1329–1343 (2008).
- S. Komoto *et al.*, Generation of infectious recombinant human rotaviruses from just 11 cloned cDNAs encoding the rotavirus genome. *J. Virol.* **93**, e02207-18 (2019).
- E. C. Holmes, Error thresholds and the constraints to RNA virus evolution. *Trends Microbiol.* **11**, 543–546 (2003).
- L. Zhao *et al.*, Cell-type-specific activation of the oligoadenylate synthetase–RNase L pathway by a murine coronavirus. *J. Virol.* **87**, 8408–8418 (2013).
- A. Wickenhagen *et al.*, A prenylated dsRNA sensor protects against severe COVID-19. *Science* **374**, eabj3624 (2021).
- J. He *et al.*, Association of SARS susceptibility with single nucleic acid polymorphisms of OAS1 and MxA genes: A case-control study. *BMC Infect. Dis.* **6**, 106 (2006).
- S. Zhou *et al.*, A Neanderthal OAS1 isoform protects individuals of European ancestry against COVID-19 susceptibility and severity. *Nat. Med.* **27**, 659–667 (2021).
- O. Danziger, R. S. Patel, E. J. DeGrace, M. R. Rosen, B. R. Rosenberg, Inducible CRISPR activation screen for interferon-stimulated genes identifies OAS1 as a SARS-CoV-2 restriction factor. *PLoS Pathog.* **18**, e1010464 (2022).
- F. W. Soveg *et al.*, Endomembrane targeting of human OAS1 p46 augments antiviral activity. *eLife* **10**, e71047 (2021).
- C. M. Carey *et al.*, Recurrent loss-of-function mutations reveal costs to OAS1 antiviral activity in primates. *Cell Host Microbe* **25**, 336–343.e4 (2019).
- K. L. Matthews, C. M. Coleman, Y. van der Meer, E. J. Snijder, M. B. Frieman, The ORF4b-encoded accessory proteins of Middle East respiratory syndrome coronavirus and two related bat coronaviruses localize to the nucleus and inhibit innate immune signalling. *J. Gen. Virol.* **95**, 874–882 (2014).
- T. S. Munasinghe *et al.*, MERS-CoV ORF4b employs an unusual binding mechanism to target IMPα and block innate immunity. *Nat. Commun.* **13**, 1604 (2022).
- Y. Yang *et al.*, Middle East respiratory syndrome coronavirus ORF4b protein inhibits type I interferon production through both cytoplasmic and nuclear targets. *Sci. Rep.* **5**, 17554 (2015).
- Y. Kitagawa, T. Tsukamoto, M. Itoh, B. Gotoh, Middle East respiratory syndrome coronavirus ORF4b protein inhibits TLR7- and TLR9-dependent alpha interferon induction. *FEBS Lett.* **596**, 2538–2554 (2022).
- M. Bello-Perez *et al.*, MERS-CoV ORF4b is a virulence factor involved in the inflammatory pathology induced in the lungs of mice. *PLoS Pathog.* **18**, e1010834 (2022).
- C. E. Comar *et al.*, MERS-CoV endoribonuclease and accessory proteins jointly evade host innate immunity during infection of lung and nasal epithelial cells. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2123208119 (2022).
- M. Gultom *et al.*, Establishment of well-differentiated camelid airway cultures to study Middle East respiratory syndrome coronavirus. *Sci. Rep.* **12**, 10340 (2022).
- M. Gultom *et al.*, Susceptibility of well-differentiated airway epithelial cell cultures from domestic and wild animals to severe acute respiratory syndrome coronavirus 2. *Emerg. Infect. Dis.* **27**, 1811–1820 (2021).
- J. Zhou *et al.*, Infection of bat and human intestinal organoids by SARS-CoV-2. *Nat. Med.* **26**, 1077–1083 (2020).
- K. Katoh, K. Misawa, K.-I. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
- B. Q. Minh *et al.*, IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
- H. Ashkenazy *et al.*, FastML: A web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Res.* **40**, W580–W584 (2012).
- A. Moshe, T. Pupko, Ancestral sequence reconstruction: Accounting for structural information by averaging over replacement matrices. *Bioinformatics* **35**, 2562–2568 (2019).
- T. D. Goddard *et al.*, UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
- E. F. Pettersen *et al.*, UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
- Z. Li, L. Jaroszewski, M. Iyer, M. Sedova, A. Godzik, FATCAT 2.0: Towards a better understanding of the structural diversity of proteins. *Nucleic Acids Res.* **48**, W60–W64 (2020).
- S. A. Goldstein, N. C. Elde, Supporting data for Recurrent viral capture of cellular phosphodiesterases that antagonize OAS–RNase L. Figshare. https://figshare.com/projects/Viral_PDE_Evolution/164854. Deposited 14 April 2023.

To quantify the similarities (or differences) between different model structures, we conducted pairwise alignments in FATCAT 2.0 (53) to calculate RMSD values.

Figure Design. Figure 1 was produced using BioRender. All other figures were produced using Adobe Illustrator 2023.

Data, Materials, and Software Availability. Phylogenetic data, FastML data, AlphaFold files .txt files with raw data from ancestral sequence reconstruction raw data from phylogenetic analysis of viral PDEs .fasta multiple sequence alignments iqtree log files iqtree .treefile files Raw output from AlphaFold structural modeling .fasta file of amino acid sequences .pdb files of the highest confidence prediction for each sequence data have been deposited in FigShare (https://figshare.com/projects/Viral_PDE_Evolution/164854) (54). All other data are included in the manuscript and/or supporting information.

ACKNOWLEDGMENTS. We thank Zoe A. Hilbert for advice and critical feedback. We thank Ian Boys for assistance implementing AlphaFold2 and FatCat. We thank Kristina M. Babler for help designing the Figure 1 schematic. This work was supported by NIH grants F32AI152341 to S.A.G. and R35GM134936 to N.C.E