Hidden evolutionary constraints dictate the retention of coronavirus accessory genes

Stephen A. Goldstein[1], Teagan M. Feeley[1], Kristina M. Babler[1], Zoë A. Hilbert[1,2], Diane M. Downhour[1,2], Niema Moshiri[3], Nels C. Elde[1,2]

[1]Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT, USA

[2]Howard Hughes Medical Institute, 4000 Jones Bridge Road, Chevy Chase, MD 20815, USA

[3]Department of Computer Science and Engineering, University of California, San Diego, La Jolla, CA 92093,

Summary

Coronaviruses exhibit many mechanisms of genetic innovation[1–5], including the acquisition of accessory genes that originate by capture of cellular genes or through duplication of existing viral genes[6,7]. Accessory genes influence viral host range and cellular tropism, but little is known about how selection acts on these variable regions of virus genomes. We used experimental evolution of mouse hepatitis virus (MHV) encoding a cellular AKAP7 phosphodiesterase and an inactive native phosphodiesterase, NS2 (ref 8) to simulate the capture of a host gene and analyze its evolution. After courses of serial infection, the gene encoding inactive NS2, ORF2, unexpectedly remained intact, suggesting it is under cryptic constraint uncoupled from the function of NS2. In contrast, AKAP7 was retained under strong selection but rapidly lost under relaxed selection. Guided by the retention of ORF2 and similar patterns in related betacoronaviruses, we analyzed ORF8 of SARS-CoV-2, which arose via gene duplication[6] and contains premature stop codons in several globally successful lineages. As with MHV ORF2, the coding-defective SARS-CoV-2 ORF8 gene remains largely intact, mirroring patterns observed during MHV experimental evolution, challenging assumptions on the dynamics of gene loss in virus genomes and extending these findings to viruses currently adapting to humans.

32 **Introduction**

33 Coronaviruses have unusually large RNA genomes, ranging from ~27-32 kilobases[9,10]. The
34 stability of these large genomes is facilitated by the incorporation of a proofreading exonuclease
35 into the RNA polymerase complex, unique among RNA viruses, that enhances replication
36 fidelity[11–13]. Consequently, coronaviruses can accommodate addition of genetic material in the
37 form of accessory genes acquired via horizontal gene transfer[14–16] and gene duplication[6,17].
38 Evolutionary theory on the streamlining of RNA virus genomes[1,3,4] predicts that new genes, such
39 as coronavirus accessory genes, should encode near-immediate fitness benefits to avoid
40 deletion.

41

42 Experimental studies of tobacco etch virus (TEV) support the assumption that exogenous genetic
43 material is rapidly purged under relaxed selection[2,3]. However, exogenous genes engineered into
44 TEV, a small single-stranded positive sense RNA virus, substantially increase genome size, which
45 might drive loss regardless of a potential fitness benefit offered by the gene product. To date, no
46 comparable studies using animal RNA viruses to test the interplay between gain of beneficial
47 coding sequence and selection on genome size have been reported. Therefore, our
48 understanding of how viral genome architecture and content evolve is poor, relative to the intense
49 focus on how viruses evolve by nucleotide substitution. Given the major evolutionary leaps
50 afforded by gene rearrangements, gains, and losses this represents a significant gap in our
51 fundamental understanding of viral evolution.

52

53 To explore how selection acts on newly acquired viral genes we used experimental evolution of
54 mouse hepatitis virus (MHV), a prototypical betacoronavirus, the genus that includes SARS and
55 MERS-related coronaviruses. ORF2 of MHV and related viruses encodes a 2'-5'
56 phosphodiesterase (PDE), NS2, that antagonizes the OAS-RNase L antiviral pathway. We used
57 recombinant MHV encoding a cellular PDE, the AKAP7 central domain (AKAP7), as a functional

58 replacement for an inactive NS2 (ref 8) with a substitution at position 126 (His126Arg), that results
59 in restricted viral replication[18]. Viral PDEs were derived via capture of cellular AKAP7-like
60 genes[19,20] so using this virus, MHV^{AKAP7}, simulates a horizontal transfer event giving rise to a novel
61 viral gene. Critically, the AKAP7 PDE is inserted in place of ORF4 (ref 8), allowing us to probe
62 the relationship between selection and gene retention without substantial expansion of the
63 genome.
64
65 We found that the retention of coronavirus accessory gene sequences in different selective
66 environments does not conform to expectations of genome streamlining. Under relaxed selection
67 in the absence of active OAS-RNase L immune pressure, AKAP7 was rapidly lost, consistent with
68 the impact of selective constraints on genome size. However, in striking contrast, MHV ORF2 was
69 retained under all conditions tested, despite encoding an inactive PDE. In parallel, we found that
70 ORF8 of SARS-CoV-2 is largely intact even in lineages where it has lost coding capacity[21]. These
71 results suggest fundamental and widespread constraints on streamlining of viral genomes due to
72 genetic information with hidden functions beyond protein coding capacity.
73
74 **Results**
75 **Cell type-dependent selective pressure on viral phosphodiesterases**
76 To test the impacts of distinct selective pressures on coronavirus genome composition, we used
77 a recombinant mouse hepatitis virus (MHV) that has an inactivating H126R amino acid change in
78 NS2, its native PDE, which is functionally complemented by insertion of the coding sequence for
79 the cellular AKAP7 central domain PDE (AKAP7)[8] (**Figure 1A-B**). The AKAP7 gene inserted into
80 MHV is 618 nucleotides and replaces the defective 321 nucleotide ORF4, so expansion of the
81 genome is minimal (~1%). MHV with inactive NS2 (MHV^{NS2Mut}) is restricted in mouse primary bone
82 marrow-derived macrophages (BMDMs) by OAS-RNase L (ref 18), while AKAP7 inserted into
83 MHV (MHV^{AKAP7}) compensates for the inactive NS2 and restores replication to wild-type virus

84 levels in BMDMs and mouse liver. MHV^{NS2mut} is restricted in immortalized primary macrophages[22],
85 but replicates to wild-type levels in L2 mouse fibroblasts (**Figure 1C**), providing differential
86 selection for courses of experimental evolution[18]. Similarly, MHV^{AKAP7} replicated to high titers
87 whereas MHV encoding an inactive AKAP7 PDE (MHV^{AKAP7mut}) was significantly restricted
88 (**Figure 1D**), demonstrating that these cells have an active OAS-RNase L pathway, which we
89 hypothesized imposes selective pressure to retain an active PDE. In contrast, MHV^{AKAP7} and
90 MHV^{AKAP7mut} replicated to an equivalent level in L2 fibroblasts, establishing these cells as a model
91 for MHV^{AKAP7} evolution under relaxed selection (**Figure 1E**).
92
93 **Deletion of AKAP7 occurs rapidly under relaxed selection**
94 Our experimental evolution workflow uses serial passage in macrophages and L2 cells to model
95 evolution of a horizontally acquired PDE under different conditions (**Figure 1E**). We plaque
96 purified passage 0 (p0) MHV^{AKAP7} and passaged this virus in triplicate in each cell type at a
97 multiplicity of infection (MOI) of 0.01. At each passage we titrated virus replication by plaque assay
98 to calculate the MOI for the subsequent passage and collected total RNA for analysis of AKAP7
99 by PCR and Sanger sequencing. In macrophages, full-length AKAP7 was retained through ten
100 passages (**Figure 1F, Figure S1A-B**), whereas in L2 cells, smaller PCR amplicons appeared by
101 passage three or four (**Figure 1G, Figure S1C-D**). The smaller bands became increasingly
102 prominent through p10, while the full length AKAP7 band faded in intensity, indicating viruses with
103 the full-length gene were disappearing from the population (**Figure 1G, Figure S1C-D**). We
104 cloned and sequenced full-length bands from macrophages and L2 cells, and smaller bands from
105 L2 cells at each passage.

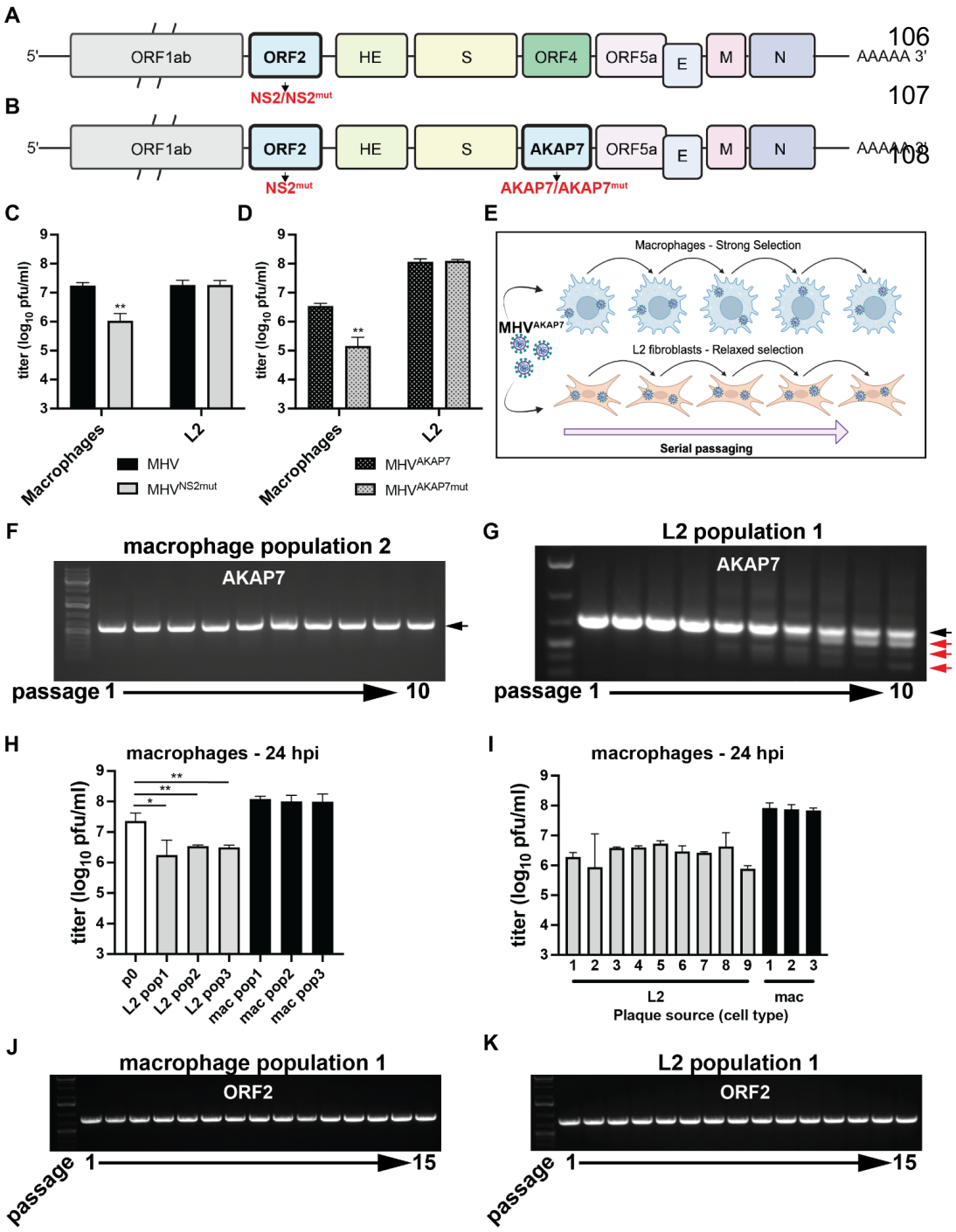

Figure 1. Experimental evolution of mouse hepatitis virus reveals hidden selective constraints. A) Schematic of the MHV genome with NS2, the protein encoded by ORF8 indicated. B) Schematic of the MHV^{AKAP7} genome, with AKAP7 inserted in place of ORF4. C) Replication of wild-type MHV and MHV^{NS2mut} in macrophages (p=0.002) and L2 fibroblasts (p=0.968) 24 hpi, infected at an MOI of 0.01. One representative experiment of three is shown. D) Replication of MHV^{AKP7} and MHV^{AKAP7mut} in macrophages (p=0.002) and L2 cells (p=0.681) 24 hpi, infected at an MOI of 0.01. One representative experiment of three is shown. E) Schematic of the experimental evolution protocol. F-G) PCR analysis of AKAP7 at passage 1 to 10 in macrophages and L2 fibroblasts, respectively. The black arrow indicates full-length intact AKAP7, while red arrows indicate AKAP7 amplicons with deletions. H) Replication in macrophages 24 hpi (MOI=0.01) of p0 MHVAKAP7 and p10 virus populations from L2 fibroblasts and macrophages. Population 1 p-value=0.0257, Population 2 p-value=0.0058, Population 3 p-value=0.0054. Statistical significance determined by unpaired t-test. I) Replication of plaque purified p10 isolates from L2 fibroblasts (gray bars) and macrophages (black bars), in macrophages 24 hpi (MOI=0.01). J-K) PCR analysis of ORF2 in macrophages and L2 fibroblasts, respectively.

109 AKAP7 from both cell types contained no mutations (**Supplementary File 1**). The smaller bands
110 replacing full-length AKAP7 during L2 passage featured substantial deletions (**Supplementary**
111 **File 1**). The AKAP7 deletion variants did not undergo mutation, consistent with high fidelity
112 coronavirus replication and revealed a lower barrier to genetic change by deletion than
113 substitution, which has been described in SARS-CoV-2 as well[23]. Virus populations recovered at
114 p10 from L2s exhibited significantly restricted replication relative to macrophage-passaged virus
115 populations (**Figure 1H, Figure S2A-B**), demonstrating that despite the heterogeneity in AKAP7
116 deletion patterns between replicates, all three MHV^{AKAP7} populations lost the ability to effectively
117 suppress OAS-RNase L.
118
119 We then plaque purified virus isolates from replicate three p10 virus populations from both cell
120 types and used these isolates to infect macrophages and L2 cells. Consistent with the loss of
121 AKAP7 during L2 passage, p10 plaque purified isolates from these cells exhibited reduced
122 replication in macrophages but replicated to wild-type titers in L2 fibroblasts (**Figure 1I, Figure**
123 **S2C-E)**, with some experiments showing a growth advantage for AKAP7-deleted isolates, which
124 warrants further study (**Figure S2C-E**). The rapid loss of AKAP7 under relaxed selection is
125 consistent with an evolutionary model wherein newly encoded viral proteins viral genes must
126 provide a near-immediate advantage for their parent gene to escape deletion.
127
128 **Inactive MHV ORF2 is retained under strong and relaxed selection**
129 Like the loss of AKAP7 under relaxed selection, there is an *a priori* prediction that ORF2, which
130 encodes an inactive PDE in MHV^{AKAP7}, would be also rapidly lost from the MHV genome. ORF2,
131 encoding the H126R inactivating substitution in NS2 should be dispensable in both macrophages
132 and L2 fibroblasts. Surprisingly, PCR of ORF2 (**Figure 1J-K**) at each serial passage in
133 macrophages and L2s, and sequencing of ORF2 after 10 passages, revealed no deletions novel
134 mutations, or reversion of the H126R substitution (**Supplementary File 2**). Given the unexpected

135 retention of inactive ORF2, we conducted five additional passages in both cell types to provide
136 additional opportunity for deletion. We sequenced the macrophage-passaged viruses again at
137 p15 to confirm there was no reversion at amino acid position 126 that might suggest ORF2
138 retention in these cells, even out to passage 15, was due to OAS-RNase L-mediated selection
139 (**Figure 1I-J; Supplementary File 2**). This unpredicted retention of ORF2 raised the possibility
140 that protein-coding coronavirus genes might be regularly retained under constraint distinct from
141 protein function.
142
143 To investigate whether similar patterns of constraint exist in the wild, we analyzed ORF2 from
144 other betacoronaviruses in the same subgenus as MHV. Rabbit coronavirus HKU14 (ref 24) has
145 a premature stop codon mutation truncating NS2 to 43 amino acids and additional downstream
146 stop codons in the same frame. However, the entire nucleotide sequence encoding full-length
147 NS2 is present in 4/5 HKU14 sequences in NCBI, while the other has ~100 bp of missing
148 sequence (**Supplementary FIle 3**). In contrast, ORF2 in porcine hemagglutinating
149 encephalomyelitis virus (PHEV) appears much more evolutionarily malleable[25–28], with various
150 isolates containing ORF2 with premature stop codons or deletions (**Supplementary File 3**).
151 These findings are broadly consistent with constraints on deletion of ORF2, with the heterogeneity
152 in PHEV suggesting there are dispensable genomic regions.
153
154 **Long read sequencing confirms retention of ORF2 after experimental evolution**
155 To more extensively analyze ORF2 and AKAP7 genes after experimental evolution we performed
156 Oxford Nanopore direct cDNA sequencing of plaque purified MHV^{AKAP7} p0, five isolates after 10
157 passages in L2 cells and p10 isolate from macrophage infections. PCR analysis of the plaque
158 purified isolates showed AKAP7 was uniformly intact in macrophage-derived isolates but deleted
159 in L2-derived plaque isolates (**FigureS3A-B**). Mean read length across direct cDNA Nanopore
160 sequencing runs ranged from 1,065 to 1,528 bases and average coverage was 2,603x to 27,296x

161 per nucleotide of the MHV^{AKAP7} genome with a mean of 8,943x. Across all sequencing runs ~70-
162 80% of reads aligned to
163 the MHV^{AKAP7} genome
164 (**Supplementary File 5**).
165 To quantify changes in
166 ORF2 and AKAP7 gene
167 content after ten
168 passages and control for
169 variable overall
170 sequencing depth across
171 runs, we normalized
172 average coverage of
173 ORF2 and AKAP7 to
174 average coverage of
175 ORF1ab for each isolate.
176 Consistent with PCR

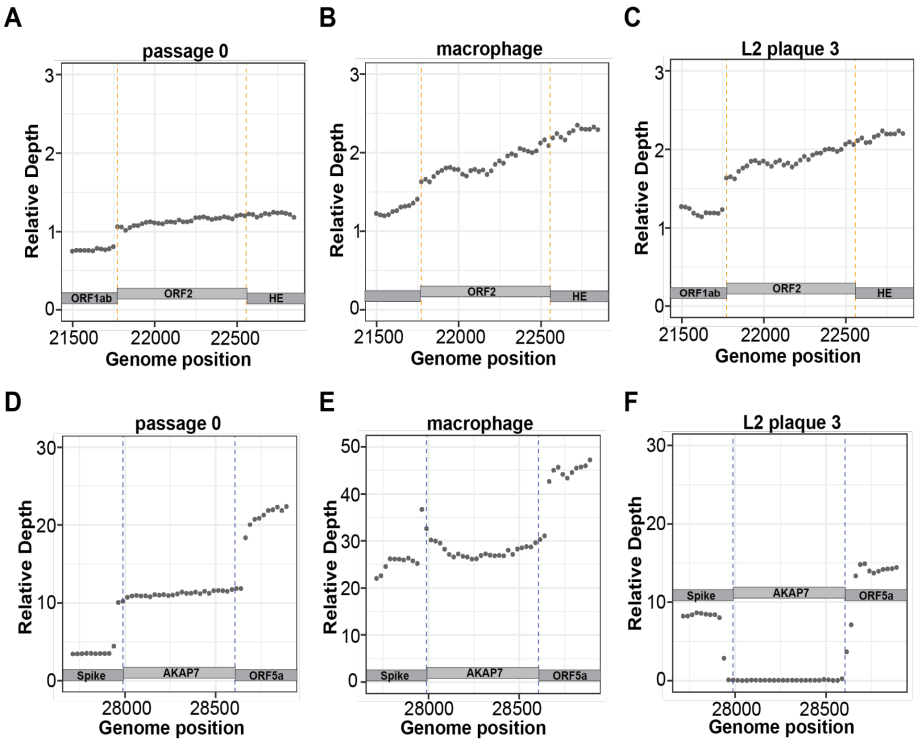

Figure 2. Nanopore direct cDNA sequencing shows retention of ORF2 and near-complete loss of AKAP7 following serial passage. A-C) Relative coverage depth plots of ORF2 in purified plaque isolates from passage 0 MHV^{AKAP7}, and passage 10 macrophage and L2 fibroblast purified plaques. D-E) Relative coverage depth plots of AKAP7 in purified plaque isolates from passage 0 MHV^{AKAP7}, and passage 10 macrophage and L2 fibroblast purified plaques.

177 analysis, average relative coverage of ORF2 was the same in plaque purified passage 10 isolates
178 from macrophages and L2 fibroblasts (**Figure 2A-C; Figure S3C-F**), demonstrating that this
179 genomic region is under unanticipated constraint. The average relative AKAP7 coverage for p0
180 and macrophage p10 were 11.2 and 27.8, respectively (**Figure 2D-E**). In contrast, the average
181 relative AKAP7 coverage for L2 p10 isolates was 0.332, a 33-fold decrease from p0 consistent
182 with PCR analysis. (**Figure 2F; S3G-J)**.
183

184 **SARS-CoV-2 ORF8 is retained despite loss of coding capacity**

185 Our analysis of ORF2 in MHV-like viruses suggested the constraints we identified experimentally
186 also exist in virus populations in the wild. However, the strength of this analysis was limited by

187 the relative paucity of genomic data available for animal viruses. In contrast, the SARS-CoV-2
188 dataset in the Global Initiative on Sharing All Influenza Data (GISAID) database contains more
189 than 16 million genomes, offering an unprecedented window into viral evolution. Analogous to the
190 inactivation of MHV ORF2 in our system, SARS-CoV-2 ORF8 has acquired premature stop
191 codons in numerous globally successful lineages, including B.1.1.7, XBB.1, XBB.1.5, XBB.1.9,
192 and XBB.1.16 (ref 21), and a mutation in the ORF8 transcriptional regulatory sequence (TRS) in
193 the BA.5 lineage that substantially reduces ORF8 subgenomic RNA and protein synthesis[29].
194 ORF8 has been of interest since early in the SARS-CoV-2 pandemic due to a small deletion that
195 emerged in its SARS-CoV ortholog in 2003 [30], an early cluster of infections in Singapore involving
196 an ORF8 deletion variant[31,32], as well as debate over its role during SARS-CoV-2 infection[6,32–39].
197 Despite the intense focus on ORF8 function and loss of full-length protein synthesis, very little
198 attention has been devoted to the evolution of the gene itself.
199
200 We downloaded all sequences assigned to the B.1.1.7, BA.5, XBB.1, XBB.1.5, XBB.1.9, and
201 XBB.1.16 lineages from GISAID and applied a cutoff date to exclude any samples without
202 metadata or with metadata suggesting they are incorrectly assigned (**Figure S4**). For each
203 lineage, we further defined a start-date for circulation and calculated the date at which 50% and
204 90% of sequences assigned to this lineage were collected, as of November 29, 2023. We further
205 excluded any sequences with any ambiguous nucleotides in the ORF8 gene to eliminate
206 confounding of our analysis by poor-quality sequences, which might obscure *bona fide* deletions.
207 We calculated a deletion length distribution of ORF8 in B.1.1.7 (1,097,307 sequences), BA.5
208 (20,833 sequences), XBB.1 (22,031 sequences), XBB.1.5 (166,068 sequences), XBB.1.9.1
209 (19.990 sequences), and XBB.1.16 (19,119 sequences) (**Figure 3C**). Across all lineages 96% of
210 deletions were smaller than ten nucleotides, ranging from a high of 98.15% of deletions smaller
211 than ten nucleotides for B.1.1.7 to a low of 86.6% of deletions smaller than ten nucleotides for
212 XBB.1.16. (**Figure 3C, Supplementary File 5**). We then analyzed all sequences in the dataset

to determine whether any regions of ORF8 were particularly deletion-tolerant or resistant. While XBB.1.5 ORF8 deletions were enriched in the 5' half of the gene (**Figure 3D, S5B)**, deletions were uniformly rare at all positions across the other lineages we analyzed (**Figure 3D-F, S5A-D)** except for a single site at the 3' end that we speculate may impact sgRNA synthesis of downstream ORFs N and 9b.

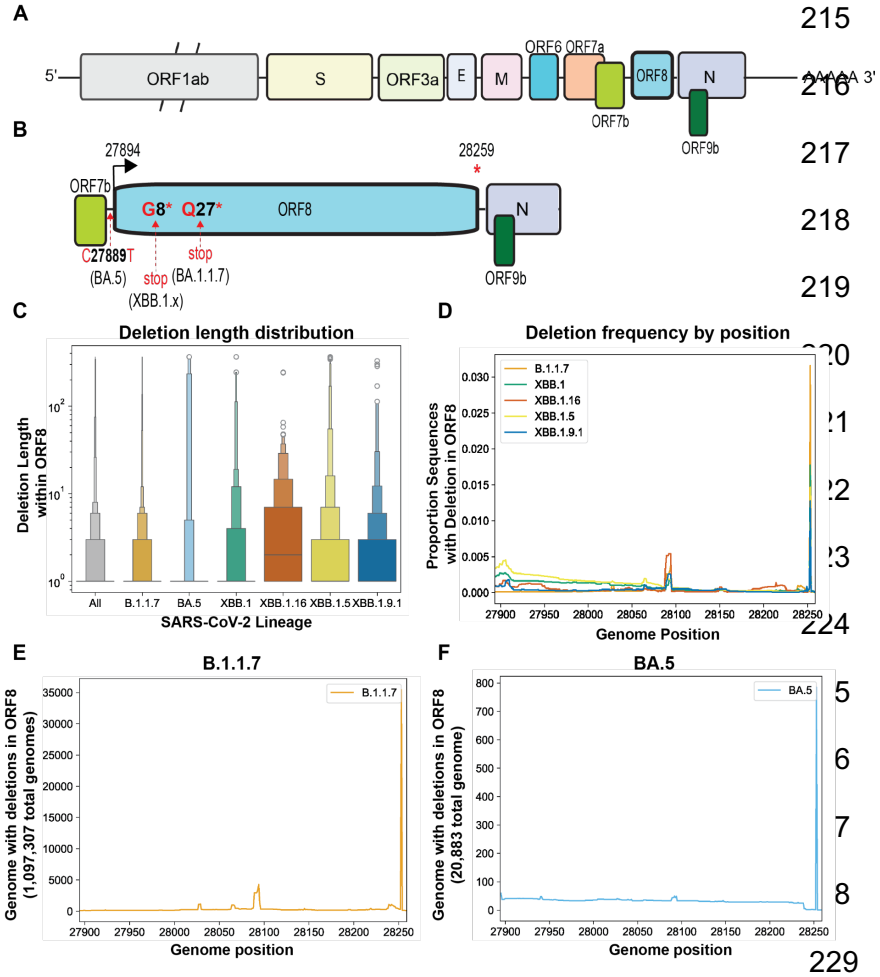

Figure 3. ORF8 is retained in SARS-CoV-2 lineages despite mutations that disrupt protein coding. A) Schematic of SARS-CoV-2 genome. B) Schematic (with ORF8 expanded, not to scale) of ORF8, with mutations resulting in loss of subgenomic mRNA synthesis (BA.5) or an early stop codon (B.1.1.7 and XBB.1.x) indicated. C) Deletion length distribution - this displays the size of all deletions in ORF8 in each lineage from 0 to >100 nucleotides. D-F) Proportion of genomes from lineages with a deletion at each position in SARS-CoV-2 ORF8.

To determine whether ORF8 in these lineages degraded over time, we calculated the percent of ORF8 gene content in these lineages, sorted by collection date, and plotted these values for all genomes of all six lineages (**Figure 4A-F**). To determine how the appearance of ORF8 deletions in these lineages related to their epidemiological rise and fall, we calculated the dates at which 50% (dashed vertical line) and 90% (solid vertical line) of sequences assigned to each lineage had been collected. For all lineages, genomes with deletions were rare outside this window, generally occurring early after the appearance of a lineage and going extinct during lineage growth or emerging very late when the lineage was already in steep decline.

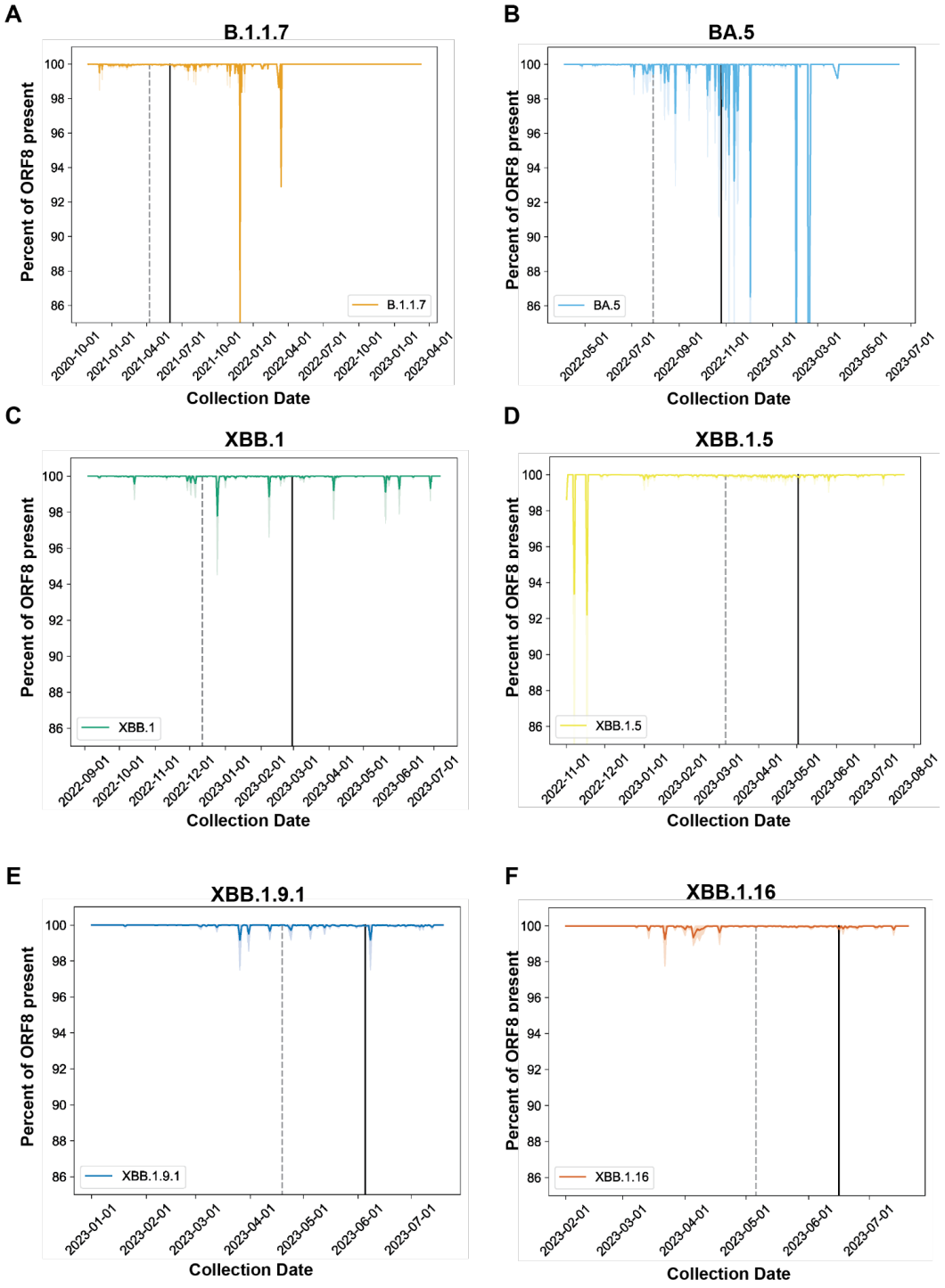

Figure 4. SARS-CoV-2 ORF8 is recurrently retained following acquisition of an early stop codon. A-F) Line plots of percent nucleotide content for the indicated SARS-CoV-2 lineages over time throughout the course of their circulation in humans. Dark lines are the mean percentage for all sequences collected on that date, and faded lines are the 95% confidence interval. Dashed vertical line is the date by which 50% of sequence assigned to the designated lineage were collected, and the solid line is the 90% cutoff.

240 **Discussion**

241 Findings from this study challenge a simple evolutionary model wherein the fate of protein coding
242 genes hinges only on protein function. The sustained, unexpected retention of ORF2 in both
243 macrophages and L2 fibroblasts through fifteen passages suggests a complex interplay of
244 selective pressure acting on viral coding sequence. In L2s the PDE activity of ORF2 is
245 dispensable, and the stability of the H126R substitution during experimental evolution in
246 macrophages is consistent with AKAP7 fully counteracting OAS-RNase L activity. While it is
247 formally possible that an unknown secondary role for the ORF2-encoded protein NS2 antagonizes
248 innate immunity, no such function has been identified and the protein contains no identifiable
249 domains other than the PDE.

250

251 The difference in genomic location between ORF2 and AKAP7 may be relevant to their respective
252 fates. AKAP7 is located upstream of ORF5a, which encodes an accessory gene that promotes
253 resistance to interferon signaling[41], whereas ORF2 immediately precedes the hemagglutinin-
254 esterase (HE) and spike structural genes. HE subgenomic mRNA is not expressed from the MHV
255 A59 genome[42,43], so it is unlikely that the selection against ORF2 deletion relates to transcription
256 of the immediate downstream gene.

257

258 Under a conventional model of selection favoring genome streamlining, viruses with ORF8
259 deleted would quickly dominate populations soon after the potential to encode full-length protein
260 was lost. The exact nature of the evolutionary constraint on MHV ORF2 and SARS-CoV-2 ORF8
261 is not yet clear. A leading possibility is that at least some coronavirus genes contain regulatory
262 functions in addition to their protein coding capacity that tune the transcription of other genes.
263 MHV ORF2 is upstream of spike, albeit with the defective HE gene intervening, and SARS-CoV-
264 2 ORF8 immediately precedes nucleocapsid, which has multiple functions key to promoting viral
265 fitness[44]. If accessory genes promote transcription of spike, nucleocapsid, and other subgenomic

266 RNAs, loss of such genes might reduce viral fitness and lineages containing such deletions would
267 go extinct. Increased transcription of subgenomic RNAs encoding innate immune antagonists has
268 been linked to enhanced suppression of innate immunity in major SARS-CoV-2 variants of
269 concern[45], which is consistent with our hypothesis and analysis of SARS-CoV-2 ORF8 retention.
270
271 The loss of ORF2 in HKU1 and the observation of some ORF8 deletion SARS-CoV-2 viruses
272 indicates that patterns of constraint may be subtle or absent in some genomic regions. In the case
273 of SARS-CoV-2, the apparent decreased fitness of ORF8-deletion viruses might be balanced by
274 adaptations that compensate for loss of ORF8, affording fitness benefits of genome compression
275 while shedding the deleterious impact of losing the gene. Continued surveillance of SARS-CoV-
276 2 might reveal such variants over and provide more insight into fundamental factors driving
277 coronavirus evolution.
278
279 Identifying additional patterns of evolutionary constraint may reveal other sequence features
280 impacting viral fitness. Patterns of selection can be prioritized for mechanistic studies using
281 reverse genetic strategies established for MHV and SARS-CoV-2, opening a bridge between
282 experimentation and computational analysis of viral evolution. Experimentally dissecting fitness
283 impacts will benefit from future competition studies. For example, the loss of ORF2 in HKU1
284 results in viable viruses and fitness defects of $MHV^{\Delta ORF2}$ might appear only in direct competition
285 with wild-type viruses and not in parallel courses of experimental evolution. Overall, this work
286 demonstrates the power of merging experimental approaches with computational analysis of
287 massive datasets to identify patterns of constraint influencing the evolution of emerging and
288 pandemic viruses.
289
290
291

292 **Materials and Methods**

293 ***Cells and Viruses***

294 Immortalized primary macrophages were provided by Sunny Shin[22], 17-Clone 1 and L2 cells were
295 provided by Susan Weiss. Macrophages were cultured in RPMI-1640 supplemented with 10%
296 FBS, 1% L-glutamine, and 1% Penicillin/Streptomycin, while 17-Clone 1 and L2 cells were
297 cultured in DMEM supplemented the same way.

298

299 All viruses used in this study were provided by Susan Weiss at the University of Pennsylvania.
300 Wild-type MHV and MHV^{NS2mut} were previously described[18] as were MHV^{AKAP7} and $MHV^{AKAP7mut8}$.
301 Viruses were received as seed aliquots and 100 µl of each virus was added to one T75 flask
302 containing a confluent monolayer of 17-Clone 1 cells in 2 ml volume. Flasks were incubated for
303 one hour at 37° C, after which 10 ml of fresh media was added. Once widespread cytopathic effect
304 was present (18-24 hours post-infection) the flasks were put through three freeze-thaw cycles
305 and the supernatant was clarified via centrifugation and aliquoted for later use.

306

307 ***Virus infections and plaque assays***

308 One day after seeding in 12-well plates macrophages or L2 cells were infected at an MOI of 0.01
309 in 200 µl total volume and incubated for one hour at 37 ^{o} C with rocking every 15 minutes. After
310 one hour cells were washed 3 times with PBS and 1 ml of fresh media (2% FBS) was added to
311 each well. 24 hours post-infection 300 µl of supernatant was collected and stored at -80° C. We
312 also harvested total RNA for additional analyses (described below). Initial infections, all serial
313 passages, and endpoint replication comparisons were done at an MOI of 0.01. Viral titers were
314 quantified by plaque assay on confluent L2 cell monolayers. Supernatants collected 24 hours
315 post-infection were serially diluted 1:10 to 10^{-8}, inoculated onto L2 monolayers and incubated for
316 1 hour at 37 ^{o} C with rocking every 15 minutes. After one hour 3 mL of semi-solid agarose overlay
317 was added to each well and the plates returned to 37 ^{o} C for 24 hours. 24 hours post-infection the

318 agarose overlay was removed and the monolayers were washed with PBS and stained with 0.5%
319 crystal violet plus 20% methanol. The stain was washed off with water and plaques counted and
320 recorded for analysis.
321
322 ***Statistical testing***
323 Statistical testing of viral titers (**Figure 1C-D, H-I; Figure S2**) was conducted in GraphPad Prism
324 10, via unpaired t-test.
325
326 ***RNA extraction and cDNA synthesis***
327 Total RNA was harvested 24 hours post-infection on each serial passage with the Zymo Research
328 Quick-RNA Miniprep Kit (Cat. #R1057) following the manufacturer's protocol. To increase yield
329 we eluted the RNA in a reduced volume of 25 µl nuclease-free water. RNA quantification and
330 quality assessment was conducted using a Synergy HT BioTek plate reader. cDNA was
331 synthesized using the Thermo Scientific Maxima First Strand cDNA Synthesis Kit with dsDNase
332 (Lot. 2664799) with 1 µg of input RNA.
333
334 ***PCR and Sanger sequencing***
335 At each passage the AKAP7 central domain and ORF2 genes from MHV^{AKAP7} were analyzed by
336 PCR. We used Phusion Flash HiFi Master Mix (Cat # F548S) in 20 µl/reaction with 2 µl of cDNA
337 template. Thirty cycles of PCR were performed. Cycling conditions were: **1)** Denaturation, 98°C
338 for 10 seconds **2) a)** Denaturation (98°C for 2 seconds) **b)** Annealing (63°C for 10 seconds) **c)**
339 Extension (72°C for 30 seconds) **d)** Final extension (72°C for 2 minutes) **3)** Hold (4°C).
340
341 PCR products were visualized by agarose gel (1%) electrophoresis then extracted and purified
342 using Zymoclean Gel DNA Recovery Kit (Zymo Research Lot. #213587) following manufacturer
343 protocols. DNA was eluted in 10 µl. Purified PCR products were prepared for Sanger Sequencing

344 by TOPO cloning using the Invitrogen TA Cloning™ Kit, with pCR™2.1 Vector (Cat. #K202020)
345 following manufacturer protocols. Plasmid DNA was extracted using the Zippy Plasmid Miniprep
346 Kit (Zymo Research Cat. # 11-308) following manufacturer protocol with elution 75 µl nuclease-
347 free water. Plasmid DNA was quantified with a Synergy HT BioTek plate reader and sent for
348 sequencing to the University of Utah DNA sequencing core or Genewiz.
349
350 *Sequencing primers*

Primers	Forward	Reverse
AKAP7	5'-**ATTGCTACCTGGCCCCG**-3'	5'-**CTAGGGTCTTAGGCCCAAATG**-3'
MHV ORF2	5'-**ATGGCCTTTGCTGACAAGCCTAAT**-3'	5'-**TCAACACATACAACCCTTCATTCT**-3'
M13	5'-**GTAAAACGACGGCCAG**-3'	5'-**CAGGAAACAGCTATGAC**-3'

351
352 ***Sequence analysis***
353 Sequence files were imported into Geneious and aligned to either the AKAP7 or ORF2 coding
354 sequences using the MAFFT Geneious plug-in [46]. HKU1, HKU14, and PHEV multiple sequence
355 alignments were likewise generated using the MAFFT Geneious plug-in with default parameters.
356
357 ***Oxford Nanopore direct cDNA sequencing***
358 1 mL of media containing a plaque isolate was used to infect 2x 10 cm dishes of desired cell type
359 (macrophages or L2). 18-20 hours post-infection total RNA was collected using Zymo Research
360 Quick-RNA Miniprep Kit (Cat. #R1057). The same lysis buffer (1 ml) was used on both dishes/per
361 infection to maximize yield. Polyadenylated RNA was extracted using the Invitrogen Poly(A)Purist
362 MAG Kit (Cat # AM1922) following the manufacturer protocol. 100 ng PolyA+ RNA was prepared
363 for sequencing using the Direct cDNA Sequencing Kit SQK-DCS109 from Oxford Nanopore
364 Technologies, according to the manufacturer's instructions. The prepared cDNA libraries were

365 sequenced using the MinION Mk1C with FLO-MIN106D flow cells, version 9.4.1. Fastq files were
366 generated from fast5 files using the high-accuracy model of Guppy basecaller, version 6.5.7 with
367 the following parameters hac_guppy -c dna_r9.4.1_450bps_hac.cfg -x auto --compress_fastq.

368

369 ***Direct cDNA sequencing coverage analysis***

370 Fastq files from Nanopore sequencing runs were individually aligned to the MHV^{AKAP7} (passage
371 0) reference genome. Alignments were performed using minimap2 (v.2.23) and subsequently
372 sorted and indexed with samtools (v.1.16). Coverage at every position across the genome for
373 each sample was calculated using the bedtools (v.2.26.0) genomecov command. Coverage was
374 then normalized for each sample to the average coverage across ORF1ab (nucleotide positions
375 211-21746 in the reference genome), which contains the non-amplified reading frame, allowing
376 for comparison across sequencing runs and samples. A rolling average of this normalized read
377 depth was calculated across the genome in 25 bp windows and plotted with ggplot2 in R. For
378 AKAP7 and ORF2 analysis, rolling averages were recalculated for the region containing the gene
379 sequence ± 300 bp of flanking sequence on either side and plotted identically to the genome wide
380 coverage plots. Fastq sequencing files are available in the Sequence Read Archive at BioProject
381 ID PRJNA1031673.

382

383 ***SARS-CoV-2 ORF8 Deletion Analysis***

384 All methods and code used to conduct this analysis are available at
385 https://github.com/niemasd/SC2-Deletion-Analysis.

386

387

Acknowledgements

This work was supported by NIH grants F32AI152341 to SAG and R35GM134936 to NCE. We thank Susan Weiss and Sunny Shin at the University of Pennsylvania for reagents. We thank Tom Peacock at the Pirbright Institute and David Bauer from the Francis Crick Institute for helpful discussions.

Author Contributions

SAG: project conceptualization, experimental design, experimentation, data analysis, manuscript writing, manuscript editing

TMF: experimentation, data analysis

KMB: experimentation, data analysis, schematic design, manuscript editing

ZAH: AKAP7 and ORF2 coverage quantification and analysis

DMD: sequencing library preparation, direct cDNA sequencing and analysis

NM: SARS-CoV-2 ORF8 deletion analysis

NCE: project conceptualization, data analysis, manuscript writing, manuscript editing

References

1. Willemsen, A. & Zwart, M. P. On the stability of sequences inserted into viral genomes. *Virus Evol* **5**, vez045 (2019).
2. Zwart, M. P., Willemsen, A., Daròs, J.-A. & Elena, S. F. Experimental evolution of pseudogenization and gene loss in a plant RNA virus. *Mol. Biol. Evol.* **31**, 121–134 (2014).
3. Willemsen, A., Zwart, M. P., Ambrós, S., Carrasco, J. L. & Elena, S. F. 2b or Not 2b: Experimental Evolution of Functional Exogenous Sequences in a Plant RNA Virus. *Genome Biol. Evol.* **9**, 297–310 (2017).
4. Belshaw, R., Pybus, O. G. & Rambaut, A. The evolution of genome compression and genomic novelty in RNA viruses. *Genome Res.* **17**, 1496–1504 (2007).

413 5. Simon-Loriere, E., Holmes, E. C. & Pagán, I. The effect of gene overlapping on the rate of
414 RNA virus evolution. *Mol. Biol. Evol.* **30**, 1916–1928 (2013).
415 6. Tan, Y., Schneider, T., Leong, M., Aravind, L. & Zhang, D. Novel Immunoglobulin Domain
416 Proteins Provide Insights into Evolution and Pathogenesis of SARS-CoV-2-Related
417 Viruses. *MBio* **11**, (2020).
418 7. Gorbalenya, A. E., Enjuanes, L., Ziebuhr, J. & Snijder, E. J. Nidovirales: evolving the
419 largest RNA virus genome. *Virus Res.* **117**, 17–37 (2006).
420 8. Gusho, E. *et al.* Murine AKAP7 has a 2',5'-phosphodiesterase domain that can complement
421 an inactive murine coronavirus ns2 gene. *MBio* **5**, e01312–e1314 (2014).
422 9. Debat, H. J. Expanding the size limit of RNA viruses: Evidence of a novel divergent
423 nidovirus in California sea hare, with a ~35.9 kb virus genome. *bioRxiv* 307678 (2018)
424 doi:10.1101/307678.
425 10. Saberi, A., Gulyaeva, A. A., Brubacher, J. L., Newmark, P. A. & Gorbalenya, A. E. A
426 planarian nidovirus expands the limits of RNA genome size. *PLoS Pathog.* **14**, e1007314
427 (2018).
428 11. Smith, E. C. & Denison, M. R. Coronaviruses as DNA wannabes: a new model for the
429 regulation of RNA virus replication fidelity. *PLoS Pathog.* **9**, e1003760 (2013).
430 12. Eckerle, L. D., Lu, X., Sperry, S. M., Choi, L. & Denison, M. R. High Fidelity of Murine
431 Hepatitis Virus Replication Is Decreased in nsp14 Exoribonuclease Mutants. *J. Virol.* **81**,
432 12135–12144 (2007).
433 13. Smith, E. C. & Denison, M. R. Implications of altered replication fidelity on the evolution and
434 pathogenesis of coronaviruses. *Curr. Opin. Virol.* **2**, 519–524 (2012).
435 14. Goldstein, S. A. & Elde, N. C. Recurrent Viral Capture of a Host-Derived Immune
436 Antagonist. *bioRxiv* 2023.05.12.540623 (2023) doi:10.1101/2023.05.12.540623.
437 15. Mazumder, R., Iyer, L. M., Vasudevan, S. & Aravind, L. Detection of novel members,
438 structure-function analysis and evolutionary classification of the 2H phosphoesterase

439 superfamily. *Nucleic Acids Res.* **30**, 5229–5243 (2002).
440 16. Asthana, A., Gaughan, C., Dong, B., Weiss, S. R. & Silverman, R. H. Specificity and
441 Mechanism of Coronavirus, Rotavirus, and Mammalian Two-Histidine Phosphoesterases
442 That Antagonize Antiviral Innate Immunity. *MBio* **12**, e0178121 (2021).
443 17. Neches, R. Y., Kyrpides, N. C. & Ouzounis, C. A. Atypical Divergence of SARS-CoV-2 Orf8
444 from Orf7a within the Coronavirus Lineage Suggests Potential Stealthy Viral Strategies in
445 Immune Evasion. *MBio* **12**, (2021).
446 18. Zhao, L. *et al.* Antagonism of the interferon-induced OAS-RNase L pathway by murine
447 coronavirus ns2 protein is required for virus replication and liver pathology. *Cell Host*
448 *Microbe* **11**, 607–616 (2012).
449 19. Goldstein, S. A. & Elde, N. C. Recurrent viral capture of cellular phosphodiesterases that
450 antagonize OAS-RNase L. *Proc. Natl. Acad. Sci. U. S. A.* **121**, e2312691121 (2024).
451 20. Nomburg, J., Price, N. & Doudna, J. A. Birth of new protein folds and functions in the
452 virome. *bioRxiv* 2024.01.22.576744 (2024) doi:10.1101/2024.01.22.576744.
453 21. Wagner, C. *et al.* Positive selection underlies repeated knockout of ORF8 in SARS-CoV-2
454 evolution. *bioRxiv* (2023) doi:10.1101/2023.09.21.23295927.
455 22. Blasi, E., Radzioch, D., Merletti, L. & Varesio, L. Generation of macrophage cell line from
456 fresh bone marrow cells with a myc/raf recombinant retrovirus. *Cancer Biochem. Biophys.*
457 **10**, 303–317 (1989).
458 23. McCarthy, K. R. *et al.* Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive
459 antibody escape. *Science* **371**, 1139–1142 (2021).
460 24. Lau, S. K. P. *et al.* Isolation and characterization of a novel Betacoronavirus subgroup A
461 coronavirus, rabbit coronavirus HKU14, from domestic rabbits. *J. Virol.* **86**, 5481–5496
462 (2012).
463 25. He, W.-T. *et al.* Newly identified lineages of porcine hemagglutinating encephalomyelitis
464 virus exhibit respiratory phenotype. *Virus Evol* **9**, vead051 (2023).

465 26. Vijgen, L. *et al.* Evolutionary history of the closely related group 2 coronaviruses: porcine
466 hemagglutinating encephalomyelitis virus, bovine coronavirus, and human coronavirus
467 OC43. *J. Virol.* **80**, 7270–7274 (2006).
468 27. Shi, J. *et al.* Genomic characterization and pathogenicity of a porcine hemagglutinating
469 encephalomyelitis virus strain isolated in China. *Virus Genes* **54**, 672–683 (2018).
470 28. Rho, S. *et al.* Detection and genetic analysis of porcine hemagglutinating encephalomyelitis
471 virus in South Korea. *Virus Genes* **42**, 90–96 (2011).
472 29. Bouhaddou, M. *et al.* SARS-CoV-2 variants evolve convergent strategies to remodel the
473 host response. *Cell* **186**, 4597–4614.e26 (2023).
474 30. Muth, D. *et al.* Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus
475 acquired during the early stages of human-to-human transmission. *Sci. Rep.* **8**, (2018).
476 31. Young, B. E. *et al.* Effects of a major deletion in the SARS-CoV-2 genome on the severity
477 of infection and the inflammatory response: an observational cohort study. *Lancet* **396**,
478 603–611 (2020).
479 32. Su, Y. C. F. *et al.* Discovery and Genomic Characterization of a 382-Nucleotide Deletion in
480 ORF7b and ORF8 during the Early Evolution of SARS-CoV-2. *MBio* **11**, e01610–20 (2020).
481 33. Flower, T. G. *et al.* Structure of SARS-CoV-2 ORF8, a rapidly evolving immune evasion
482 protein. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
483 34. Gamage, A. M. *et al.* Infection of human Nasal Epithelial Cells with SARS-CoV-2 and a
484 382-nt deletion isolate lacking ORF8 reveals similar viral kinetics and host transcriptional
485 profiles. *PLoS Pathog.* **16**, e1009130–e1009130 (2020).
486 35. Bello-Perez, M. *et al.* SARS-CoV-2 ORF8 accessory protein is a virulence factor. *MBio*
487 e0045123 (2023).
488 36. Wang, X. *et al.* SARS-CoV-2 ORF8 Protein Induces Endoplasmic Reticulum Stress-like
489 Responses and Facilitates Virus Replication by Triggering Calnexin: an Unbiased Study. *J.*
490 *Virol.* **97**, e0001123 (2023).

491 37. Tamura, T. *et al.* Virological characteristics of the SARS-CoV-2 XBB.1.5 variant. *bioRxiv*
492 2023.08.16.553332 (2023) doi:10.1101/2023.08.16.553332.
493 38. Rodriguez-Rodriguez, B. A. *et al.* A neonatal mouse model characterizes transmissibility of
494 SARS-CoV-2 variants and reveals a role for ORF8. *Nat. Commun.* **14**, 3026 (2023).
495 39. Ponde, N. O. *et al.* SARS-CoV-2 ORF8 Mediates Signals in Macrophages and Monocytes
496 through MyD88 Independently of the IL-17 Receptor. *J. Immunol.* **211**, 252–260 (2023).
497 40. Fehr, A. R. & Perlman, S. Coronaviruses: An Overview of Their Replication and
498 Pathogenesis. in 1–23 (Springer New York, 2015).
499 41. Koetzner, C. A. *et al.* Accessory protein 5a is a major antagonist of the antiviral action of
500 interferon against murine coronavirus. *J. Virol.* **84**, 8262–8274 (2010).
501 42. Shieh, C. K. *et al.* Identification of a new transcriptional initiation site and the corresponding
502 functional gene 2b in the murine coronavirus RNA genome. *J. Virol.* **63**, 3729–3736 (1989).
503 43. Kazi, L., Lissenberg, A., Watson, R., de Groot, R. J. & Weiss, S. R. Expression of
504 hemagglutinin esterase protein from recombinant mouse hepatitis virus enhances
505 neurovirulence. *J. Virol.* **79**, 15064–15073 (2005).
506 44. Johnson, B. A. *et al.* Nucleocapsid mutations in SARS-CoV-2 augment replication and
507 pathogenesis. *PLoS Pathog.* **18**, e1010627 (2022).
508 45. Thorne, L. G. *et al.* Evolution of enhanced innate immune evasion by SARS-CoV-2. *Nature*
509 **602**, 487–495 (2022).
510 46. Katoh, K., Misawa, K., Kuma, K.-I. & Miyata, T. MAFFT: a novel method for rapid multiple
511 sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066
512 (2002).

513