Zooanthroponotic transmission of SARS-CoV-2 and host-specific viral mutations revealed by genome-wide phylogenetic analysis

Sana Naderi[1], Peter E. Chen[1,2], Carmen Lía Murall[1,3], Raphael Poujol[4], Susanne Kraemer[3], Bradley S. Pickering[5,6,7], Selena M. Sagan[1,8*], B. Jesse Shapiro[1,9*]

1. Department of Microbiology & Immunology, McGill University, Montreal, QC, Canada
2. Département de sciences biologiques, Université de Montréal, Montreal, QC, Canada
3. Public Health Agency of Canada, Winnipeg, MB, Canada
4. Research Centre, Montreal Heart Institute, Montreal, QC, Canada
5. National Centre for Foreign Animal Disease, Canadian Food Inspection Agency, Winnipeg, MB, Canada
6. Department of Veterinary Microbiology and Preventative Medicine, College of Veterinary Medicine, Iowa State University, Ames, Iowa, USA
7. Department of Medical Microbiology and Infectious Diseases, University of Manitoba, Winnipeg
8. Department of Biochemistry, McGill University, Montreal, QC, Canada
9. McGill Genome Centre, Montreal, QC, Canada

*for correspondence: selena.sagan@mcgill.ca; jesse.shapiro@mcgill.ca

Abstract

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is a generalist virus, infecting and evolving in numerous mammals, including captive and companion animals, free-ranging wildlife, and humans. Transmission among non-human species poses a risk for the establishment of SARS-CoV-2 reservoirs, makes eradication difficult, and provides the virus with opportunities for new evolutionary trajectories, including selection of adaptive mutations and emergence of new variant lineages. Here we use publicly available viral genome sequences and phylogenetic analysis to systematically investigate transmission of SARS-CoV-2 between human and non-human species and to identify mutations associated with each species. We found the highest frequency of animal-to-human transmission from mink, compared with lower transmission from other sampled species (cat, dog, and deer). Although inferred transmission events could be limited by sampling biases, our results provide a useful baseline for further studies. Using genome-wide association studies, no single nucleotide variants (SNVs) were significantly associated with cats and dogs, potentially due to small sample sizes. However, we identified three SNVs statistically associated with mink and 26 with deer. Of these SNVs, ~⅔ were plausibly introduced into these animal species from local human populations, while the remaining ~⅓ were more likely derived in animal populations and are thus top candidates for experimental studies of species-specific adaptation. Together, our results highlight the importance of studying animal-associated SARS-CoV-2 mutations to assess their potential impact on human and animal health.

Importance. SARS-CoV-2, the causative agent of COVID-19, can infect many animal species, making eradication difficult because it can be reseeded from different reservoirs. When viruses replicate in different species, they may be faced with different evolutionary pressures and acquire new mutations, with unknown consequences for transmission and virulence in humans. Here we analyzed SARS-CoV-2 genome sequences from cats, dogs, deer, and mink to estimate transmission between each of these species and humans. We found several transmission events from humans to each animal, but relatively few detectable transmissions from animals back to humans, with the exception of mink. We also identified three mutations more likely to be found in mink than humans, and 26 in deer. These mutations could help the virus adapt to life in these different species. Ongoing surveillance of SARS-CoV-2 from animals will be important to understand their potential impacts on both human and animal health.

Introduction

Coronaviruses can have broad animal host ranges, and severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is no exception. Although the animal reservoir of ancestral SARS-CoV-2 remains unknown, SARS-CoV-2 has close relatives in bats and an ancestral variant likely spilled over into humans via an intermediate animal host in a seafood market in Wuhan, China (Worobey et al. 2022). Although SARS-CoV-2-related coronaviruses from animals in the Wuhan market were not sampled, there have been several subsequent reports of SARS-CoV-2 transmission ("spillback") from humans to animals including in farmed mink (Oude Munnink et al. 2021) and wild white-tailed deer (*Odocoileus virginianus*) (Kuchipudi et al. 2022; Kotwa et al. 2022). Consequently, transmission among potential animal reservoirs is a key feature of the past and future evolution of coronaviruses. In addition to making SARS-CoV-2 elimination highly unlikely, evolution in animal reservoirs could transiently increase evolutionary rates (Porter et al. 2022) and potentially select for novel mutations with effects on transmission and virulence in humans (Otto et al. 2021). Viral adaptation to one host species might result in the tradeoff of reduced transmission in other species. For example, the Middle East respiratory syndrome coronavirus appears primarily adapted to camels, with relatively short-lived transmission chains in humans (Dudas et al. 2018). Alternatively, evolution in different species could potentially create new paths to peaks in the human-adaptive fitness landscape. It is speculated that the SARS-CoV-2 Omicron variant of concern (VOC) might have evolved in a non-human animal (possibly rodent) before transmitting widely among humans (Wei et al. 2021). This scenario remains hypothetical and is not exclusive of other hypotheses, such as evolution in unsampled human populations, and/or in immunocompromised/chronically infected individuals. Nevertheless, ongoing transmission among humans and non-human animals and its implications for viral evolution and host adaptation thus deserves further study.

There have been several reports of SARS-CoV-2 transmission from humans to other animal species, and in some cases back to humans. In addition to farmed mink (Oude Munnink et al. 2021), there have also been reports of transmission from pet hamsters (Yen et al. 2022) and possible transmission events from white-tailed deer to humans in North America (Pickering et al. 2022). Certain mutations may improve replication fitness of SARS-CoV-2 in animal hosts. For instance, experimental infections of ferrets *(Mustela furo*) identified a Y453F spike (S) mutation that increases viral binding to the ACE2 receptor in these animals, while decreasing infection of human airway cells (Zhou et al. 2022). Another potentially mustelid-adapted mutation (S:N501T) was also found to be prevalent in infected mink sampled in the United States (Cai and Cai 2021;

Eckstrand et al. 2021). Yet this mutation appears not to suffer a fitness tradeoff in humans; rather, it caused stronger binding to human ACE2 in an *in vitro* screen (Starr et al. 2020).

While such case studies have been valuable in highlighting patterns of transmission and adaptation across individual animal species since the beginning of the pandemic, a standardized, global analysis of the available data is lacking. In this study, we comprehensively compared SARS-CoV-2 sequences derived from animal hosts using a publicly available dataset. Using phylogenetic methods, we inferred transmission events between humans and four other well-sampled animals and quantified their relative frequencies. Compared to earlier studies which tended to focus on one species at a time and much smaller datasets we use an automated approach to avoid manual browsing of the tree. This not only increases the efficiency of the analysis but also reduces the potential for human error. Secondly, rather than simply identifying all branches involving a transition from one host species to another, our method scans the tree for such branches and retains only the most basal nodes (closest to the root) on the descendant end of a given branch. This approach overcomes the possible overcounting of nested clades as separate transmission events and is thus conservative in reporting possible cross-species transmission events.

Our analysis revealed a relatively high number of mink-to-human transmission events, while instances of animal-to-human transmission from cats (*Felis catus domesticus)*, dogs (*Canis lupus familaris),* or deer (*Odocoileus virginianus*) were rare. Using genome-wide association studies (GWAS), we identified mutations associated with specific animal species. We recovered the S:N501T mutation previously associated with mink, along with two other amino acid changes in other SARS-CoV-2 genes. We also identified several novel mutations associated with deer, including both synonymous and nonsynonymous substitutions. Together, our work provides a quantitative framework for tracking SARS-CoV-2 transmission across animals from available genomic sequences and points to several candidate animal-adaptive mutations for experimental follow-up.

Results

Although secondary spillover events of SARS-CoV-2 transmission from animals back to humans have been reported, their frequency has yet to be quantified and compared across animal species. To this end, we retrieved all available SARS-CoV-2 genome sequences sampled from non-human animals from GISAID (**Figure 1a**, **Supplementary Table S1**). After applying

sequence quality filters (**Supplementary Table S2**) and considering only animals with 50 or more sequences, we were left with four species: cat, dog, mink and deer. Our filters excluded common experimental animals such as mice and ferrets (**Figure 1a**). For each of the four animal species, we extracted a similar number of closely-related human-derived sequences (Methods). If these closely-related animal-human pairs represent recent transmission chains, we would expect them to come from the same geographic region. Consistent with this expectation, we found that 95.4% of deer-derived sequences share the same sampling location as their closest human-derived relatives, and this percentage is slightly lower for cat (85.6%), dog (89%), and mink (91.7%). The high percentage for deer could be due to higher sampling effort in North America, the origin of all deer sequences. More generally, the variation in these percentages could be explained by other sampling biases. For example, cats and dogs might be undersampled relative to mink and deer.

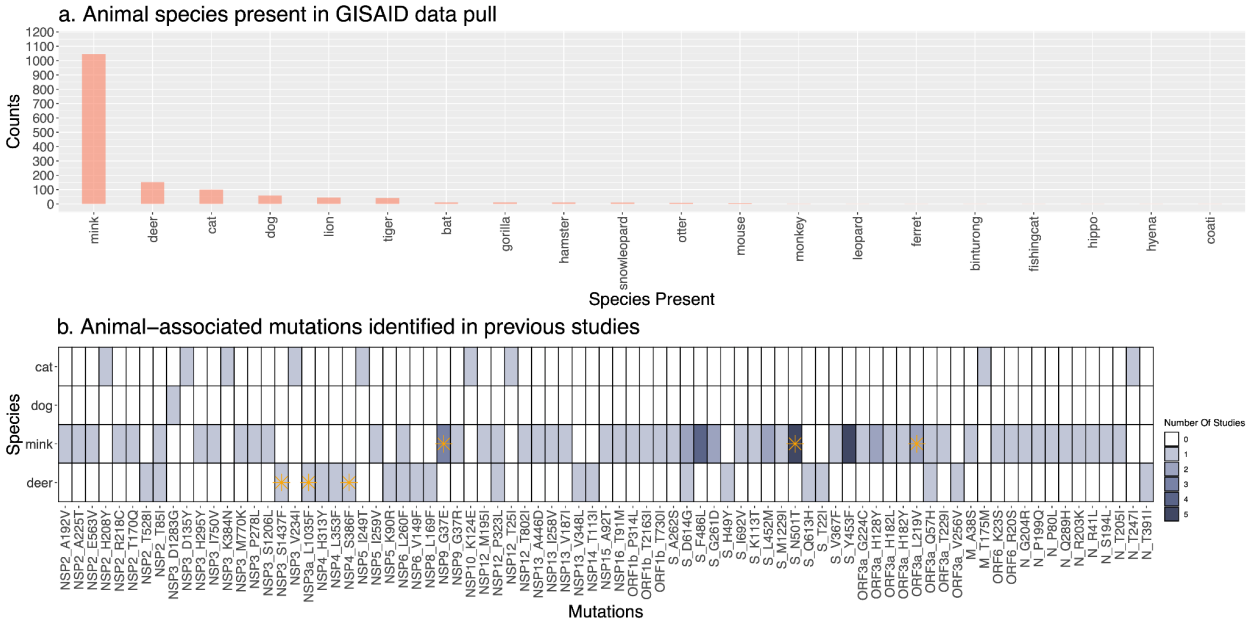

Figure 1. Overview of available SARS-CoV-2 genome sequences from different animal species. A. Barplot of the number of genome sequences available in GISAID (on Feb. 28, 2022) sampled from each animal species.. Only species with 50 or more sequences were included in the current study: cat, dog, mink, and deer. **B.** Heatmap of animal-associated mutations identified in previous publications. Darker colors indicate mutations found in a greater number of studies. Each row corresponds to one of the species included in this study, and columns correspond to mutations along the SARS-CoV-2 reference genome. Mutations identified as family-wise significant ($P < 0.05$) in our genome-wide association studies are indicated with an orange asterisk. A detailed list of publications reporting these mutations is in Supplementary table S8. Only single nucleotide variants, not insertions or deletions, are included. The heatmap illustrates the results of several key prior studies, but does not represent a comprehensive meta-analysis.

To investigate potential animal-to-human transmission events in greater detail, including the direction of transmission, we analyzed viral phylogenetic trees. For each species, the animal-derived sequences and their closest relative human-derived sequence were combined with ten random sub-samplings of human-derived sequences (n ≈ 50 per subtree per month of the pandemic), from which we inferred ten replicate phylogenies per species, providing an assessment of phylogenetic uncertainty. Using ancestral state reconstruction as previously described (Murall et al. 2021), we counted the most basal animal-to-human (**Figure 2a-d**) and human-to-animal transitions on each tree (**Supplementary Table S3**). Representative detailed trees are available in **Figure 2 supplements 1-4.** Ancestral state reconstruction can be biased by differential sampling across species. Such bias tends to be more severe close to the root of the tree (De Maio et al. 2015); but, in our case, the root is known to be human (as we excluded more divergent animal outgroups). Most of the inferred transmission events are close to the tips and far from the root (**Figure 2**), which we expect to minimize (but not entirely eliminate) bias. Our goal is, therefore, not to infer absolute rates of cross-species transmission, but rather to provide a consistent comparative framework for interspecies transmission.

To determine a more conservative lower bound for the transmission counts, we excluded transition branches with <75% bootstrap support (Methods). As further validation, we performed a permutation test to determine whether the estimated transmission counts converged on a non-random value. We performed the same ancestral state reconstruction on 1000 permutations of each of the 40 phylogenies whose tip labels (animal or human) were shuffled randomly, and the number of transmissions in both directions were recorded. This permutation test revealed that in both directions (animal-to-human and human-to-animal) the observed transmission counts in mink, deer, and cat falls within a narrow range (standard deviations of 10.32, 0.96, and 0.52 respectively) compared to the permutations (68.83, 2.85, and 1.90) while the observed counts in dog (standard deviation of 0.95) is similar to the permutations (0.90). In the human-to-animal direction, the observed standard deviations in transmission counts are 0.48, 0.48, 2.15, and 5.2 for cat, dog, mink and deer, respectively, compared to permuted values of 8.91, 1.68, 67.37, and 18.44. These results show that, in general, our inferences of transmission events converge upon stable values, but that the estimate for dog-to-human transmission might be less reliable.

Based on the bootstrap-filtered counts, we inferred fewer than two transmission events on average to humans from dogs and deer, and around four transmissions on average from cats to

humans (**Figure 3.a, Table 1**). In contrast, there were an average of 38 or more transmission events inferred from mink back to humans. The upper bound (bootstrap unfiltered) estimates are higher, but the pattern of much higher transmission from mink is retained. The inferred number of transmissions in the opposite direction, from humans to other animals, was generally higher and much more uniform across species (**Figure 3.b, Table 1**), with lower bounds in the range of 31.5 to 58.5 events. The relatively higher sampling of sequences from humans than other animals may have inflated the number of inferred human-to-animal events relative to animal-to-human events. Mink are also better sampled than the other animal species, and further sampling of other animals could identify more transmission events. On the other hand, the number of inferred transmission events is not clearly related to sample size. For example, despite their smaller sample size compared to deer, cats tend to be involved in more transmission events to and from humans (**Table 1**). Despite these caveats, it is notable that human-to-animal transmission events are relatively constant across animal species, while mink-to-human transmission is much more frequently detected compared with other animals.

Table 1: Average inferred transmission events between humans and animals.

Average inferred number of transitions (filtered – unfiltered)	Mink (n=1038)	Deer (n=134)	Cat (n=78)	Dog (n=39)
Animal-to-human	38.0 – 112.3	0.7 – 1.4	4.4 – 4.4	1.4 – 1.7
Human-to-animal	42.3 – 65.2	38.3 – 55.2	58.5 – 68.3	31.5 – 35.7

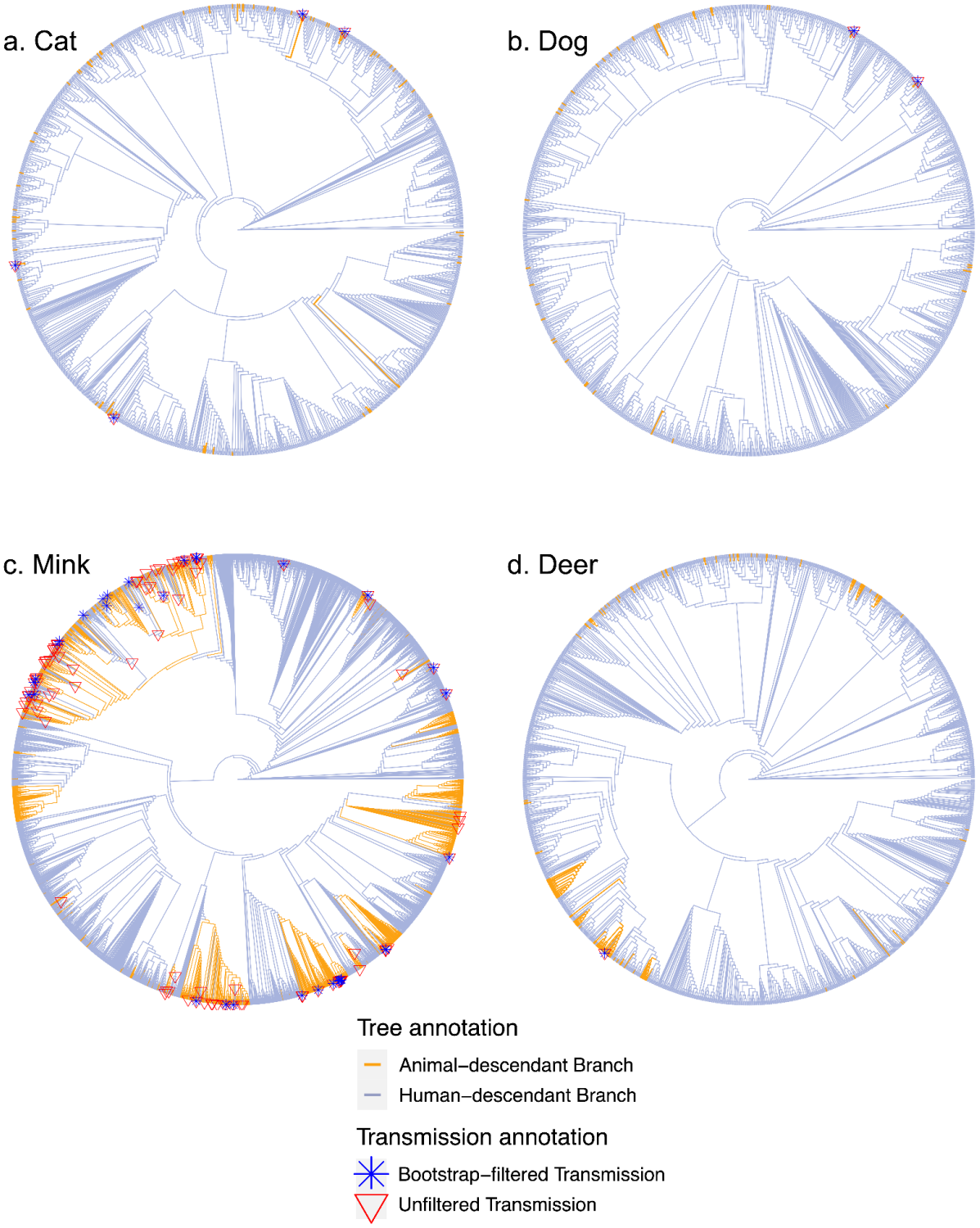

Figure 2. Transmission events inferred between humans and animals. Panels a-d display a representative tree for every species with animal to human transmissions marked on the tree. More detailed versions of these trees are in Figure 2 Supplements 1-4. Trees are rooted with the Wuhan reference genome (from one of the first sampled human COVID-19 patients).

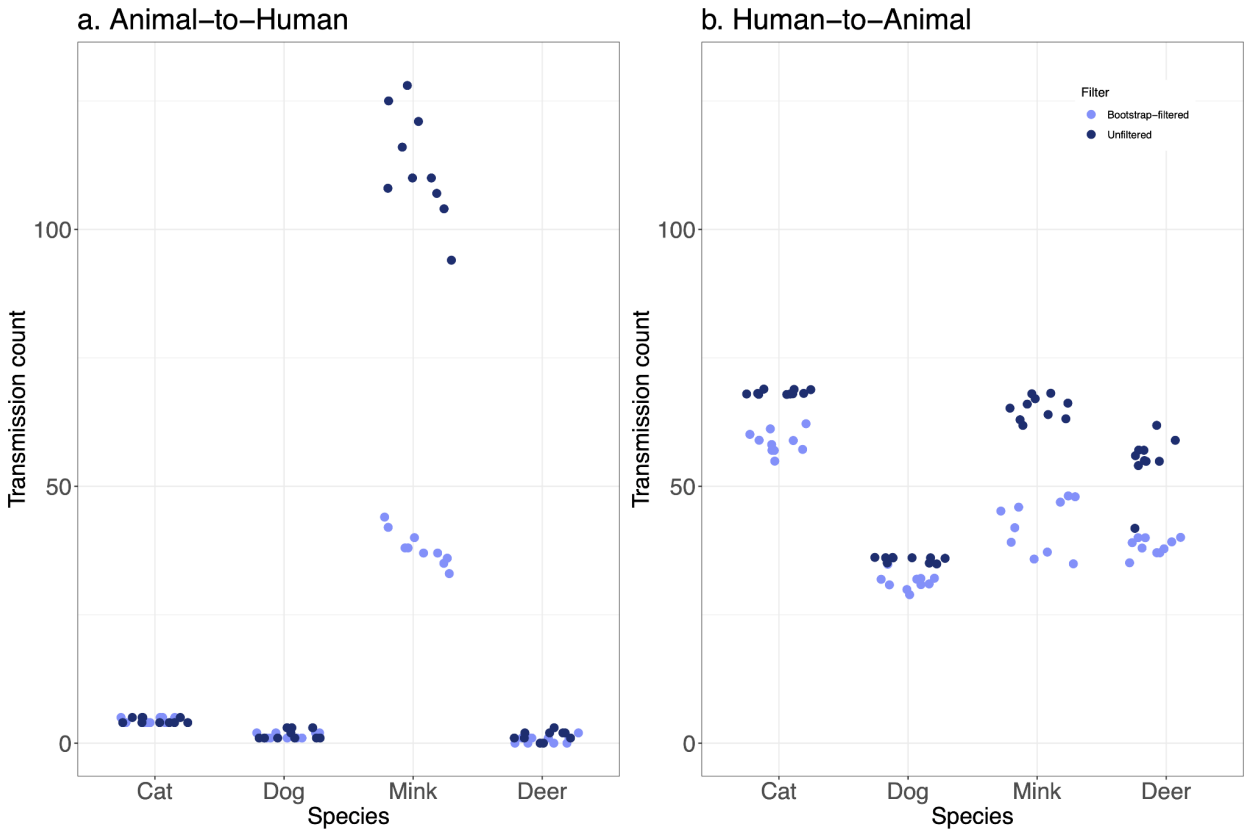

Figure 3. Transmission events from animals to humans are rarely detected, except from mink. The distribution of inferred transmission counts (across 10 replicate trees) in each animal species, in both bootstrap-filtered and unfiltered trees are shown in **A)** the animal to human direction, and **B)** the human to animal direction. Points are plotted with jitter to avoid overlap.

Next, we sought to identify mutations associated with each animal species compared to humans. These mutations could be candidates for species-specific adaptations. Dozens of mutations in the SARS-CoV-2 genome have been reported in association with different animal species (particularly mink; **Fig 1b**) but it is unclear how many of these associations are statistically significant. To test for significant associations, we conducted GWAS using POUTINE, a method which implicitly controls for viral population structure and linkage (non-independence) between mutations by considering only homoplasic mutations that are identical by state, but not identical by descent, and that have occurred independently multiple times in the SARS-CoV-2 phylogeny (Chen and Shapiro 2021). We performed a separate GWAS to identify mutations associated with each species, and replicated the GWAS on each of the ten replicate trees described above.

We identified numerous single nucleotide variants (SNVs) with high statistical significance associated with mink and deer, but none in cats or dogs (**Figure 4**). In all cases, we used a

224 significance cutoff of a family-wise $p < 0.05$ to correct for multiple hypothesis testing. The mink
225 GWAS revealed three unique SNVs (**Table 2**). One of these hits appears in all ten replicates and
226 the remaining two appear in at least half of the replicates. All three of these mutations are non-
227 synonymous, including S:N501T which was previously associated with a mink outbreak in the
228 United States (Cai and Cai 2021). Inspecting the distributions of these GWAS hits across the tree
229 reveals several independent origins (**Figure 2-Figure Supplement 3**). For example, S:N501T
230 occurred independently in The Netherlands, Denmark, Latvia, Lithuania, Spain, France, and the
231 USA (**Figure 2-Figure Supplement 3**), explaining the strong association detected by POUTINE.
232 The deer GWAS revealed a total of 26 unique statistically significant SNVs, of which seven appear
233 in all ten replicates, and five in at least half (**Table 3**). Out of these 26 hits, five are intergenic
234 (within the 5′ and 3′ UTRs) and 12 are synonymous mutations. Notably, 21 of the hits are C>U
235 transition mutations. The seven hits found in all ten replicates clearly occur multiple times
236 independently in different branches of the tree; for example, ORF1ab:N4899I (which affects amino
237 acid 507 of the mature RNA-dependent RNA polymerase protein, nsp12) occurred at least twice
238 independently in both the states of New York and Iowa (**Figure 2-Figure Supplement 4**).
239
240 We next asked whether mutations identified by GWAS plausibly occurred in an animal host or if
241 they were more likely circulating in a local human population before being transmitted to a different
242 animal species. While both these categories of mutations are statistically associated with a
243 particular animal species, the former are particularly good candidates for species-specific
244 adaptations. To address this question, we obtained the global frequency of the minor allele of
245 each significant GWAS hit in Cov-Spectrum (Chen et al. 2021) – a significantly larger database
246 than our downsampled trees used for GWAS. The minor alleles were generally rare (<1%
247 frequency; **Supplementary Tables S4 & S5**), as expected for non-human animal-associated
248 mutations in a large database dominated by human sequences. To test the hypothesis that
249 animal-associated mutations arose not in animals but in a local human population that then
250 transmitted to animals in the same region, for each GWAS hit we first defined the "in" region in
251 which the animals containing the mutation were sampled and the "out" regions including all other
252 regions. We then performed a Fisher's exact test to determine if human-derived sequences
253 containing the animal-associated mutation were enriched in the "in" region, resulting in an odds
254 ratio (OR) significantly higher than 1 (Methods). For mink, two GWAS hits had OR significantly
255 greater than 1 ($P < 0.0001$, **Table 2**). Only the S:N501T mutation had OR significantly less than
256 1, making it the best candidate for having arisen in a mink host, rather than in a human who later
257 transmitted the mutant virus to a mink. For deer, the majority of GWAS hits (18/26) could be

258 explained by transmission from the local human population (OR > 1) while the remaining eight
259 could not (OR < 1 or not significantly different than 1, **Table 3**).
260
261 Finally, we asked if mutations identified as significantly associated with a specific animal by
262 GWAS tended to occur on branches with inferred human-to-animal transitions, as would be
263 expected for species-specific adaptive mutations. To do so, using ancestral sequence
264 reconstruction, we identified mutations that occurred along these transition branches. As
265 expected, all three significant mink-associated mutations occurred multiple times on human-to-
266 mink transition branches, with the human parent node carrying the reference allele and the animal
267 descendant node carrying the mink-associated alternate allele (**Supplementary table S6**). Of the
268 26 deer-associated mutations, 22 occurred multiple times on human-to-deer branches, of which
269 21 carry the exact nucleotide substitution identified by GWAS (**Supplementary table S7**). By
270 contrast, none of the substitutions identified by GWAS occurred on animal-to-human branches,
271 consistent with species-specific adaptation.
272
273 **Table 2**. **Single-nucleotide variants associated with mink by GWAS.** "Pos" refers to the nucleotide
274 position in the reference genome. Homoplasy counts in focal animals (cases), humans (controls), and *P*-
275 values are averaged across replicates in which the site's family-wise *P*-values was < 0.05. Where
276 applicable, amino acid positions refer to the polyprotein with mature protein positions in parenthesis. The
277 'local transmission odds ratio' is the result of a Fisher's exact test of the likelihood that the alternate base
278 (animal-associated minor allele) was enriched in the local human population where the mink sequences
279 bearing the alternate base were sampled (Methods). n.s., not significant. Odds ratio *P*-values: * < 0.05, **
280 < 0.01, *** < 0.001.

Pos.	Ref. base	Alt. base	Amino acid change	Gene	Homoplasy count in focal animal	Homoplasy count in humans	*P*-value (pointwise)	*P*-value (familywise)	Significant in N replicates	Local transmission odds ratio
26047	U	G	L219V	ORF3a	6	0	0.0014	0.0365	10	3.93***
12795	G	A	G4177E (nsp9 G37E)	ORF1ab/pp1ab/nsp9/replicase	6	0	0.0015	0.0368	6	7.53***
23064	A	C	N501T	Spike/S1/RBD/binds ACE2	6.4	0	0.0010	0.0258	5	0.48***

281

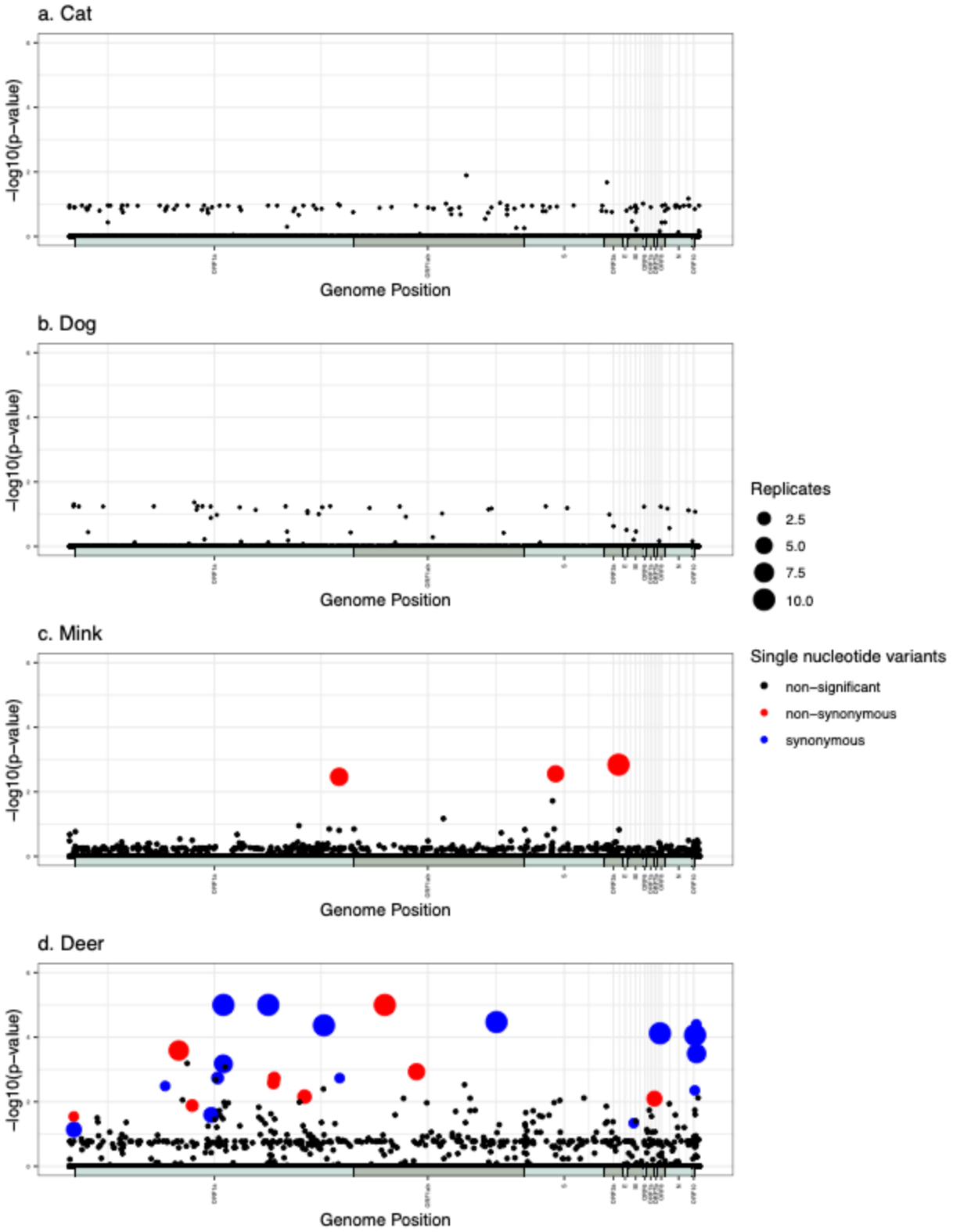

Figure 4: Manhattan plots summarizing GWAS hits in each animal species. In every panel, the x-axis represents the nucleotide position in the SARS-CoV-2 reference genome and the y-axis represents the -log10 of the pointwise *p*-values averaged over replicates. ORFs are shown as alternating blue and pink horizontal bars along the x-axis. Statistically significant hits with family-wise corrected *p*-values of lower

287 than 0.05 are shown in red (non-synonymous) or blue (synonymous), while non-statistically significant p-
288 values are in black.
289
290 **Table 3**. **Single-nucleotide variants associated with deer by GWAS**. "Pos" refers to the nucleotide
291 position in the reference genome. Homoplasy counts in focal animals (cases), humans (controls), and *P*-
292 values are averaged across replicates in which the site's family-wise *P*-values was < 0.05. Where
293 applicable, amino acid positions refer to the polyprotein with mature protein positions in parenthesis. IG,
294 Intergenic. The 'local transmission odds ratio' is the result of a Fisher's exact test of the likelihood that the
295 alternate base (animal-associated minor allele) was enriched in the local human population where the deer
296 sequences bearing the alternate base were sampled (Methods). n.s., not significant. Odds ratio *P*-values:
297 * < 0.05, ** < 0.01, *** < 0.001.
298

Pos.	Ref. base	Alt. base	Amino acid change	Gene	Homoplasy count in focal animal	Homoplasy count in humans	*P*-value (pointwise)	*P*-value (familywise)	Significant in N replicates	Local transmission odds ratio
7303	C	U	I2346I (nsp3 I1524I)	ORF1a/pp1ab/pp1a/nsp3	17.8	1.2	9.99E-06	9.99E-06	10	2.51***
9430	C	U	I3055I (nsp4 I292I)	ORF1a/pp1ab/pp1a/nsp4	15.2	6.2	9.99E-06	9.99E-06	10	2.20***
14960	A	U	N4899I (nsp12 N507I)	ORF1ab/pp1ab/nsp12/RdRp	7.8	0.1	9.99E-06	1.09E-05	10	0**
20259	C	U	F6665F (nsp15 F213F)	ORF1ab/pp1ab/nsp15	4.8	0.1	3.39E-05	0.0013	10	n.s.
28016	C	U	F41F	ORF8	4	0	7.59E-05	0.0061	10	6.09***
12073	C	U	D3936D (nsp7 D67D)	ORF1a/pp1ab/pp1a/nsp7	5.2	1.1	4.29E-05	0.0025	10	n.s.
29679	C	U	IG	3'UTR	5	1.8	8.59E-05	0.0055	10	3.17***
5184	C	U	P1640L (nsp3 P822L)	ORF1a/pp1ab/pp1a/nsp3	4.6	1.6	0.0002	0.0115	8	2.61***
29750	C	U	IG	3'UTR/S2M	5	2.6	0.0002	0.0103	7	3.12***
7318	C	U	F2351F (nsp3 F1533F)	ORF1a/pp1ab/pp1a/nsp3	4	0.3	0.0001	0.0114	6	3.80***
16466	C	U	P5401L (nsp13 P77L)	ORF1ab/pp1ab/nsp13/Hel	5	1	4.99E-05	0.0019	5	4.09***
7267	C	U	F2334F (nsp3 F1516F)	ORF1a/pp1ab/pp1a/nsp3	4.4	0.8	9.39E-05	0.0079	5	2.79***

210	G	U	IG	5'UTR/SL5a	4	0.5	0.0001	0.0136	4	3.98***
6730	C	U	N2155N (nsp3 N1337N)	ORF1a/pp1ab/pp1a/nsp3	4	0.75	0.0002	0.0168	4	1.81**
27752	C	U	T120I	ORF7a	4	0.75	0.0002	0.0169	4	4.03***
11152	C	U	V3629V (nsp6 V60V)	ORF1a/pp1ab/pp1a/nsp6	4	0.7	0.0002	0.0153	3	0.80**
5822	C	U	L1853F (nsp3 L1035F)	ORF1a/pp1ab/pp1a/nsp3	4	0.5	0.0001	0.0118	2	n.s.
9711	C	U	S3149F (nsp4 S386F)	ORF1a/pp1ab/pp1a/nsp4	4	0.5	8.49E-05	0.0118	2	0.56**
9679	C	U	F3138F (nsp4 F375F)	ORF1a/pp1ab/pp1a/nsp4	4	0	9.49E-05	0.0067	2	2.32***
7029	C	U	S2255F (nsp3 S1437F)	ORF1a/pp1ab/pp1a/nsp3	4	0.5	0.0002	0.0149	2	0.22***
29738	C	A	IG	3'UTR/S2M	4	0	3.99E-05	0.0059	1	n.s.
26767	U	C	I82T	ORF5/M	4	0	8.99E-05	0.0057	1	4.09***
203	C	U	IG	5'UTR/SL5a	4	1	0.0003	0.0191	1	5.94***
12820	A	G	L4185L (nsp9 L45L)	ORF1a/pp1ab/pp1a/nsp9	5	1	3.99E-05	0.0009	1	4.52***
4540	C	U	Y1425Y (nsp3 Y607Y)	ORF1a/pp1ab/pp1a/nsp3	4	1	0.0002	0.0239	1	2.80***
29666	C	U	L37F	ORF10	4	1	0.0002	0.0219	1	1.54***

Discussion

Summary and caveats. SARS-CoV-2 can be considered a generalist virus, able to infect a range of different species – perhaps more so than other viruses, and even other coronaviruses such as MERS with more restricted host preferences (Dudas et al. 2018). Yet SARS-CoV-2 may still replicate and transmit better in certain species than others, providing opportunities for species-specific adaptive mutations to arise. SARS-CoV-2 transmission between humans and animals has typically been studied in individual species in isolation, and using heterogeneous methods and datasets. Similarly, viral mutations associated with particular species have been reported, but

310 largely in experimental studies or region-specific sampling efforts that are difficult to generalize.
311 Our goal in this study was to apply standardized methods to identify animal-to-human
312 transmission events and to discover animal-associated mutations using available viral genomic
313 data. The results are likely biased by uneven sampling effort across species. Specifically,
314 oversampling of human-derived sequences could bias the ancestral state reconstruction toward
315 human-to-animal rather than animal-to-human transmission. In the future, such biases might be
316 accounted for using a Bayesian structured coalescent approximation (De Maio et al. 2015).
317 However, we do not expect this bias to be severe, for three reasons. First, most of our inferred
318 transmission events are close to the tips, while bias towards the most common state appears
319 worst close to the root (De Maio et al. 2015). Second, the number of inferred transmission events
320 involving each animal species is not clearly associated with its sample size. Third, we also used
321 permutations to flag certain inferences (*e.g.* dog-to-human transmission) as unreliable in our
322 current sample. Overall, our transmission analysis provides a 'level playing field' upon which to
323 assess the results of previous animal-specific studies. Our genome-wide association studies
324 considered only homoplasic mutations that occur multiple times independently in the SARS-CoV-
325 2 phylogeny. This provides strong protection against false-positive associations that could occur
326 due to population structure in clonal pathogens (Chen and Shapiro 2021) at the expense of
327 missing non-homoplasic true-positives. Like most GWAS approaches, we test each nucleotide
328 site independently, implicitly assuming additive fitness effects. This may be a reasonable
329 assumption, as non-additive epistatic fitness effects have been inferred to be negligible in SARS-
330 CoV-2 evolution to date, at least in the spike receptor-binding domain (Rochman et al. 2022). As
331 genome sequences continue to be sampled from different animal species, approaches described
332 here will allow researchers to further study the evolution and transmission of SARS-CoV-2 and
333 other pathogens across animal species. Sampling more sequences from diverse animals
334 (including the four analyzed here, as well as others) will enable more robust estimates of
335 secondary zoonotic and zooanthroponotic transmission events and species-specific adaptations.
336
337 ***What explains variation in transmission and evolution across species?*** Methodological and
338 sampling considerations aside, there are two major factors affecting rates of transmission across
339 animal species: structural similarity in viral receptor proteins (or more general phylogenetic
340 similarity) and rates of contact between species. These factors have been used to predict cross-
341 species viral transmission, with habitat overlap playing a dominant role over phylogenetic
342 similarity at least on a global scale (Carlson et al. 2022). On more local scales, contact with pets
343 such as cats and dogs will be high with a large number of pet owners, while contact with minks

may be high transiently, or for a subset of the population working on farms. Similarly, contact with deer may be restricted to hunters, or may involve indirect contact with human waste. A recent study compared ACE2 orthologs across mammalian species and classified deer ACE2 as having "high" structural similarity to human ACE2; meanwhile cat similarity was "medium", dog was "low" and mink species were "very low" (Damas et al. 2020). Clearly, large mink outbreaks have occurred despite a relatively dissimilar ACE2 structure – and the 'mismatch' between SARS-CoV-2 spike and mink ACE2 could impose strong selection for mink-adapted spike mutations (discussed below). Meanwhile, if deer ACE2 was already structurally similar to human ACE2, this could explain the lack of deer-associated mutations in spike. While it is tempting to speculate that the many other deer-associated mutations outside of spike could provide deer-specific adaptations that reduce transmission back to humans, it is equally possible that the rarity of deer-to-human transmission is explained by relatively low contact rates. As SARS-CoV-2 continues to evolve and transmit in both deer and humans, ongoing sequencing efforts might allow us to test the relative importance of these hypotheses. In summary, while global-scale predictions of zoonotic risk is possible based on contact rates and phylogenetic similarity, it is more difficult to predict the precise risk for a particular species of interest (*e.g.* mink vs. deer), which depends on detailed life-history and genetic factors. However, combining estimates of habitat overlap with the genomic analyses described here could enable more fine-grained predictions.

Consistent with previous reports (Oude Munnink et al. 2021), mink had the largest number of inferred transmission events to humans. This could be because mink have more opportunities to interact with humans on mink farms, whereas contact between deer and humans is more limited and potentially seasonal (Kuchipudi et al. 2022). However, mink outbreaks could be more frequently reported due to higher surveillance of mink farms and noticeable symptomatic disease in these animals (Oreshkova et al. 2020). As for cats and dogs, we inferred much lower transmission to humans, suggesting that they might be "dead-ends" for the virus. This does not mean that SARS-CoV-2 transmission from dogs or cats to humans is not possible, and it may be more readily detected with deeper sampling or in prospective household transmission studies. Transmission from humans to animals was relatively uniform across species, with somewhat higher inferred transmission to cats than to other species. Overall, our results support the previous reports of relatively high rates of mink to human transmission, and only rare or anecdotal transmission from other animals, such as deer (Pickering et al. 2022).

377 Despite the large sample size of mink-derived viral sequences, we only detected three mink-
378 associated SNVs using GWAS, including the previously identified S:N501T mutation. This is
379 consistent with relatively little time for SARS-CoV-2 to adapt to mink between transmission cycles
380 in humans. In contrast, we detected many more SNVs associated with deer despite a smaller
381 sample size. Even if we only consider the eight SNVs less likely to have arisen in human
382 population before transmitting to deer, or the seven SNVs detected in all replicate GWAS runs,
383 there are still more than twice as many deer-associated than mink-associated GWAS hits. This
384 could suggest a greater number of deer-adapted SNVs compared to mink-adapted SNVs,
385 perhaps due to multiple uninterrupted cycles of deer-deer transmission. Sustained deer-deer
386 transmission is also supported by previous studies (Kuchipudi et al. 2022), including the
387 observation of relatively divergent SARS-CoV-2 genomes (Pickering et al. 2022). Asymptomatic
388 or mild disease in deer relative to mink (Oreshkova et al. 2020) might also allow more opportunity
389 for evolution and transmission during prolonged infections. Despite pruning out highly divergent
390 branches, including several of the Canadian sequences involved in a potential deer-to-human
391 transmission event (Pickering et al. 2022), some relatively long deer-associated branches are
392 notable in our study, suggesting rapid evolution in deer (**Figure 2d**). Together, these results point
393 to deer as an important reservoir of novel SARS-CoV-2 mutations. Ongoing monitoring of deer-
394 to-human transmission events is therefore warranted.
395
396 ***Lack of GWAS hits in cats or dogs.*** Several viral mutations have been previously associated
397 with cats or dogs (Elaswad et al. 2020), but we found no statistically significant mutations
398 associated with either of these species (**Figure 1b**). This could be due to limited viral adaptation,
399 which would be expected if cats and dogs are effectively dead-end hosts for the virus, with little
400 time for cycles of intra-species transmission. Alternatively, the limited sample size for these
401 species could have limited GWAS power to identify adaptive mutations. Given the higher viral
402 load and shedding in cats compared to dogs (Bosco-Lauth et al. 2020; Shi et al. 2020), we expect
403 greater adaptation and transmission in cats; however, further data will be needed to test this
404 expectation.
405
406 ***Mink GWAS hits.*** We identified three statistically significant mutations associated with mink. The
407 substitution ORF3a:L219V, which appears in all ten GWAS replicates, has been previously
408 detected as a substitution associated with mink, in the ORF3a accessory gene (Elaswad et al.
409 2020). However, the previous detection was not statistically tested for significance. Similarly, the
410 ORF1ab:G4177E (nsp9:G37E) mutation has been identified previously in mink-derived

411 sequences (Eckstrand et al. 2021). The S:N501T substitution has also been previously associated
412 with mink (Lu et al. 2021; Elaswad et al. 2020) and ferret (Zhou et al. 2022), and the same site,
413 has been reported to be adaptive in mice, with an S:N501Y amino acid substitution. The mink-
414 associated S:N501T substitution increases binding to human ACE2 (Starr et al. 2020), and the
415 S:N501Y substitution has been detected in several human SARS-CoV-2 VOCs with higher
416 transmissibility, such as the Alpha variant (Y. Liu et al. 2021; Vöhringer et al. 2021). Certain other
417 mutations that have been previously associated with mink, including S:Y453F (Zhou et al. 2022;
418 Lu et al. 2021; Elaswad et al. 2020) and S:L260F (Adney et al. 2022) that were suggested to arise
419 as a result of rapid adaptation, are not statistically significant in our GWAS. While such mutations
420 could be truly animal-associated and were not picked up in GWAS due to limited power or
421 sampling, others may be anecdotal reports that do not survive the scrutiny of rigorous statistical
422 testing, or associations identified in laboratory conditions that are not currently observed in nature.
423
424 Our GWAS approach was designed to identify species-specific mutations, which could increase
425 fitness in one species at the expense of others. It is less likely to identify mutations that increase
426 fitness across species. One possible example is the S:Y453F mutation, associated with mink in
427 several studies but not in our GWAS. In all replicate trees, the mutation occurred four times as a
428 homoplasy in mink but never or only once in humans, resulting in a pointwise p-value < 0.05 that
429 did not survive correction for multiple hypothesis testing (family-wise $p > 0.05$). In cell culture
430 experiments, S:Y453F decreased infection of human airway cells (Zhou et al. 2022). The mutation
431 has been observed in human-derived sequences, but they are at quite low frequency in Cov-
432 Spectrum. The non-significant GWAS result for this mutation could therefore be explained by its
433 occurrence in mink but also (more rarely) in humans. We performed ancestral state reconstruction
434 of this mutation on the ten replicate mink trees, and observed that S:Y453F occurred 17 times on
435 human-to-human transition branches, 30 times on human-to-mink branches, 85 times on mink-
436 to-mink branches, and never on mink-to-human branches. The frequent occurrence of S:Y453F
437 on mink-to-mink and human-to-mink branches is consistent with its adaptive value in mink, but its
438 occasional occurrence on human-to-human branches suggests it does not completely prevent
439 transmission among humans. This and other mutations that could potentially increase host range
440 – or at least those that increase fitness in one species without a strong tradeoff in another species
441 – are of particular relevance but might require new or adapted GWAS methods to detect.
442
443 ***Deer GWAS hits.*** Many of the deer-associated mutations are synonymous and occur broadly
444 across the genome outside of the relatively well-studied spike (S) protein (**Figure 4**). These are

445 primarily transition mutations, which are typically generated at a higher frequency than
446 transversions. The deer-associated synonymous mutations showed no particular bias toward
447 codons that are preferred in deer relative to humans; therefore selection for codon usage
448 optimization does not easily explain these associations (data not shown). Nonetheless, the vast
449 majority of mutations were C>U transitions, which may be a reflection of APOBEC1-mediated
450 RNA editing, which results in deamination of cytosine to uracil in single-stranded RNA (Harris and
451 Dudley 2015; Salter and Smith 2018). Consistent with this hypothesis, C>U transitions in deer
452 contained the consensus sequence [U/A][U/A]C[A/U][A/U], which resembles that observed for
453 human APOBEC1-mediated deamination [AU]C[AU]; however, it remains to be seen whether the
454 deer APOBEC1 isoform has the same substrate specificity or whether there is increased
455 APOBEC expression or activity in deer tissue (Di Giorgio et al. 2020; Rosenberg et al. 2011).
456 Alternatively, the C>U transitions may be related to RNA secondary structure or nucleotide
457 composition biases required for genome condensation during viral replication organelle
458 biogenesis or virion assembly in the deer host.

459 Our GWAS revealed several mutations in deer that localized to distinct RNA secondary structures
460 in the 5´ and 3´ UTRs of the viral RNA. Specifically, in the 5´ UTR, both mutations (C203U and
461 G210U) localized to stem loop 5a (SL5a), a highly conserved stem-loop structure previously
462 implicated in virion assembly in related coronaviruses (Yang and Leibowitz 2015; Morales Lucia
463 et al. 2013; Miao et al. 2021). Similarly, in the 3´ UTR, all the mutations localized to the 3´ terminal
464 stem-loop structure, with two specifically localized to the S2M region (C29738A and C29750U)
465 (Yang and Leibowitz 2015; Gilbert and Tengs 2021; Wacker et al. 2020). Interestingly, while the
466 3´ terminal stem-loop structure is known to be hypervariable, the S2M region is highly conserved
467 in sequence and structure across a wide range of RNA viruses, including members of the
468 *Astroviridae*, *Caliciviridae*, *Picornaviridae,* and *Coronaviridae* families (Tengs et al. 2013; Gilbert
469 and Tengs 2021). While the role of this S2M region is poorly understood, its strict conservation
470 across a range of positive-strand RNA viruses suggests that it may play an important role in the
471 viral life cycle. Notably, both the mutations identified by GWAS are predicted to maintain the
472 overall S2M consensus fold, so these may reflect differences in species-specific interactions
473 between S2M and host proteins or RNA molecules.

474 The deer GWAS also revealed a few nonsynonymous mutations in viral proteins important in RNA
475 binding and host antiviral responses. Specifically, ORF1ab:N4899I (nsp12:N507I) lies in the
476 nsp12 RNA-dependent RNA polymerase, within the highly conserved motif G which is important
477 in positioning the 5´ template strand during viral RNA synthesis (Sheahan et al. 2020). While N507

478 is known to make contact with the +2 nucleotide of the template strand, experimental
479 investigations will be needed to understand how the N507I mutation impacts the active site
480 structure and/or viral RNA synthesis (Hillen et al. 2020). The ORF1ab:P5401L (nsp13:P77L)
481 mutation in the nsp13 helicase is predicted to be a solvent-exposed residue within the N-terminal
482 Zinc-binding domain (Newman et al. 2021). However, given that this is distant from the helicase
483 active site, we predict that it is unlikely to affect helicase enzymatic activity. Finally, we identified
484 two mutations that may have implications for host adaptation as they are identified in proteins
485 known to interact with the host antiviral response. Specifically, ORF1a:P1640L (nsp3:P822L) in
486 the nsp3 protease localizes to the deubiquitinating site in the PLpro domain which overlaps with
487 the ISG15 binding site, suggesting it may modulate host antiviral responses (Yang and Leibowitz
488 2015; Shin et al. 2020; G. Liu et al. 2021). Interestingly, the deer GWAS also revealed the
489 ORF7a:T120I mutation within the ORF7a accessory protein. This residue is adjacent to K119,
490 which is implicated in the inhibition of the antiviral response in human cells (Redondo et al. 2021;
491 Cao et al. 2021). Specifically, K119 polyubiquitination has been shown to block STAT2
492 phosphorylation, leading to inhibition of type I IFN. Thus, it is possible that the ORF7a:T120I
493 mutation modulates ubiquitination at the K119 residue; however, this will require experimental
494 validation.

495 ***Concluding perspectives.*** Upon finalization of this manuscript, another analysis of SARS-CoV-
496 2 transmission and potential host adaptation was published, using a similar dataset from GISAID
497 (Tan et al. 2022). Notably, this work focused only on transmission from humans to other animals,
498 while we also considered the animal-to-human direction. The approach used to identify animal-
499 associated mutations was conceptually similar to ours – focusing on homoplasic mutations and a
500 set of reasonable but arbitrary filters for allele frequencies and effect sizes – whereas we took a
501 more formal statistical GWAS approach. Our three mink GWAS hits are a subset of the four
502 identified in the other study, and neither study identified any deer-associated mutations in Spike
503 (Tan et al. 2022). Our study identified more significantly deer-associated mutations (including the
504 single hit reported by Tan et al. in nsp3, ORF1a:L1853F (nsp3:L1035F), which could be due to
505 different data filtering and significance testing approaches. Overall, the two studies are
506 complementary, and pave the way for future studies on larger datasets.

507 In conclusion, while the dynamics of anthroponosis and zooanthroponosis for SARS-CoV-2 are
508 still unclear, cross-species transmission events are likely to continue given continued
509 geographically widespread infections, high transmission rates, and the emergence of new
510 variants. We identified several statistically significant animal-associated substitutions in mink and

511 deer, suggestive of sustained animal-to-animal transmission and perhaps reflective of host
512 adaptation. Continuous molecular surveillance of SARS-CoV-2 animal isolates is likely to reveal
513 new insights into SARS-CoV-2 host range and adaptation, and may contribute to our
514 understanding of risk for spillback of new variants. This also highlights the need to monitor for
515 similar patterns of susceptibility to infection and sustained intra-species transmission in related
516 species. Our study draws attention to several specific, statistically significant nucleotide and
517 amino acid substitutions that may play a role in host adaptation, pathogenesis, and/or
518 transmission and are candidates for experimental study.

519

520 **Methods**

521

522 **Data**: On February 28, 2022, we downloaded from GISAID all SARS-CoV-2 consensus genome
523 sequences derived from non-human animals. These host species include several mammalian
524 species such as whitetail deer, mink, cats, dogs, lions, and monkeys among others. The raw non-
525 human dataset obtained from GISAID was filtered for low-quality sequences, sequences with a
526 length of less than 29k base pairs, and an ambiguous nucleotide (N) count above 500. Sequences
527 with incomplete dates recorded in the metadata were discarded and excluded from the study.
528 This study focuses on species with at least 30 sequences in the dataset, namely mink (n=1038),
529 deer (n=134), cat (n=78), and dog (n=39).

530

531 From the 7.6 million human-derived viral sequences present in GISAID's human alignment dated
532 February 28, 2022, a set of closely related sequences for every animal species was extracted. To
533 do so, we used Nextstrain (Hadfield et al. 2018) to calculate a proximity matrix based on pairwise
534 substitutions between every animal-derived sequence and all ~8 million human-derived
535 sequences. The query alignment was generated using MAFFT (Katoh and Standley 2013) in
536 which animal sequences were aligned to the gapped version of the Wuhan Iv04 reference
537 sequence present in GISAID's human-host alignment dated February 28, 2022, with the
538 "keeplength" option of the software enabled to maintain the length of context alignment in the
539 query alignment.

540

541 Using this matrix and noting that some sequences from a given species might share a number of
542 close relatives, we extracted the closest human-derived sequences for every sequence of a given
543 species such that the unique set of best-hit human-host sequences for a particular species would

have roughly the same count as the sequences from that animal species. To provide greater and more representative phylogenetic context, 500 human-derived sequences were subsampled randomly from every month of the pandemic, from January 2020 to February 2022, resulting in 13,000 human-derived sequences, distributed uniformly over time. Here again, the same quality filtering criteria were applied and the sample was drawn from sequences that were at least 29000 base pairs long and had fewer than 500 ambiguous nucleotides ("N" characters). The GISAID identifiers for all sequences used in this study are reported in **Supplementary Table S1**.

Alignment: All the above sequences, that is, the animal-host sequences along with their closely related human host sequences and the context random human subsample, and the Wuhan IV04 reference sequence were aligned using MAFFT (version 7.471). We used the NW-NS-2 setting, which is a speed-oriented, progressive method without FFT approximation. The alignment was done in two steps; first, a preliminary alignment was done and problematic sequences causing almost invariant insertions due to stretches of ambiguous nucleotides (N's) were removed from the dataset, and the remaining sequences were aligned again with the exact same algorithm, resulting in the final alignment, including a total of 14787 sequences, comprised of 1038 mink sequences, 134 deer sequences, 39 dog sequences, and 78 cat sequences. As for the closely related human-host sequences, this dataset included 852 close relatives for mink, 61 for cat, 31 for dog, and 123 for deer. The remaining sequences were human-host sequences randomly subsampled.

Phylogeny: Ten replicate trees for each candidate species were generated, each of which included the focal species' sequences, its closely related human-host sequences, and one-tenth of the randomly subsampled human-host sequences which were again split uniformly over time to ensure temporal heterogeneity, resulting in an overall count of 40 trees. Maximum-likelihood divergence trees were inferred using IQ-TREE (Nguyen et al. 2015) with a general time-reversible model and with 1000 bootstrap iterations using the ancestral sequence reconstruction (-ASR) option. The resulting trees were pruned for branches that are unreasonably divergent, and tips whose lengths were considered outliers according to the interquartile criterion and were longer than $q_{0.75} + 1.5IQR$ were pruned. $q_{0.75}$ in the above term refers to the third quartile, and IQR refers to the difference between the first and the third quartiles. This criterion discarded a maximum of 5 deer sequences out of 134, a maximum of 2 mink sequences out of 1038, and a maximum of 1 dog sequence out of 39, while no cat sequences were pruned.

578 **Ancestral state reconstruction**: Discrete ancestral state reconstruction was performed on all 40
579 trees in the study, with states set as “human” and “animal”. The reconstruction was done using
580 the maximum likelihood method “ace” implemented in the “ape” R package (Paradis and Schliep
581 2019). Ancestral state reconstruction was done using the “equal rates model” which allows for an
582 equal probability of transition in both directions of “animal to human” and “human to animal” on
583 the tree.
584
585 **Estimating transmission events**: Once the states were labeled, the set of all branches along
586 which a transition from an “animal” node state to a “human” node state has occurred was
587 identified, and the human end of the branch, namely the most recent common ancestor of the
588 introduced human clade or the human tip in case of singletons, was marked as a transition node.
589 In order to identify an independent set of transitions and to avoid the redundancy of reporting
590 nested spilled-back clades as separate events, the most basal of these transition nodes were
591 identified. In order to provide a well-rounded record of spillovers, the same analysis was done in
592 the human to animal direction as well. For setting a lower bound to the transmission counts, in
593 another set of analyses, all transitions in the desired direction were identified, then the branches
594 whose parent side had a bootstrap support of lower than 75% were discarded, and subsequently,
595 the most basal transitions were identified. To validate the ancestral state reconstruction and the
596 transmission counts, a permutation test was done in which the tips of every 10 trees for all
597 candidate species were randomly shuffled in 1000 permutations, and the same algorithm for
598 ancestral state reconstruction and subsequently counting transitions in both directions (animal to
599 human and human to animal) was applied to find the most basal transitions, unfiltered counts on
600 the shuffled trees.
601
602 **Genome-wide association studies:** We used POUTINE (Chen and Shapiro 2021) to scan the
603 genome for mutations that are statistically associated with each animal species. POUTINE relies
604 on the viral phylogeny to identify homoplasic mutations associated with a phenotype of interest,
605 in this case human vs. animal hosts. By considering only homoplasic mutations at the tips of the
606 tree, POUTINE implicitly accounts for population structure. POUTINE was run on 40 tree
607 replicates: 10 per species, with each replicate containing the same sequence alignment used in
608 the transmission analysis, though trees were inferred without the -ASR option in IQ-TREE.
609 (Ancestral sequence reconstruction was unnecessary because POUTINE only tests mutations
610 that occur on terminal branches). In each run, animal-host sequences were set as “cases” and
611 human-host sequences as “controls.” We chose to treat animals as cases because the root of the

tree is human, and initial transmission events are human-to-animal with more recent and rare animal-to-human events. The dataset is, therefore, better suited to identify animal-associated mutations than human-associated mutations. In each of the ten replicates for every species, any sites with a minor allele familywise p-value of less than 0.05 were recorded as a hit for that run. The unique collective set of hits for every species across all 10 runs was retained and the number of replicates in which each hit appeared was recorded.

Fisher's exact test for geographic bias: In order to check for geographical bias we performed a Fisher's exact test on every SNV identified by POUTINE to compare its frequency in human hosts inside and outside regions where animal-host sequences bearing these mutations are found. For each SNV, we obtained its geographical distribution of human-host sequence counts broken down by geographical division from CoV-Spectrum. We partitioned global divisions into regions that animal-host sequences bearing the specific mutation are found in–namely, "in" regions, and regions in which such sequences are not found, or "out" regions. We then created the 2-by-2 contingency table in which rows correspond to wildtype allele and alternate (animal-associated) allele counts, and columns correspond to "in" or "out" region counts. We then calculated the pointwise estimate and confidence interval for the odds ratio, and flagged mutations in which the two frequencies were significantly different. Counts and frequencies for this step are recorded in Supplementary tables S4 and S5.

Code availability: Scripts used to perform all analyses are available at https://github.com/Saannah/Animal.SARS-CoV-2.git

Acknowledgments. We thank all the authors, developers, and contributors to the GISAID database for making their SARS-CoV-2 sequences publicly available. We are grateful to CoVaRR-Net colleagues Sally Otto, Art Poon, Caroline Colijn, Will Hsiao, Fiona Brinkman, Paul Gordon, Arinjay Banerjee, Jason Kindrachuk, Angie Rasmussen, and Samira Mubareka for useful discussions that helped improve the manuscript.

Funding. This study was supported by the Canadian Institutes for Health Research (CIHR) operating grant to the Coronavirus Variants Rapid Response Network (CoVaRR-Net). Data analyses were enabled by computing and storage resources provided by Compute Canada and Calcul Québec.

Supplementary Figures

Figure 2-Figure Supplement 1. Detailed representative phylogeny of cat- and human-derived SARS-CoV-2 sequences. In order to make the tree topology clear, branch lengths are not to scale.

Figure 2-Figure Supplement 2. Detailed representative phylogeny of dog- and human-derived SARS-CoV-2 sequences. In order to make the tree topology clear, branch lengths are not to scale.

Figure 2-Figure Supplement 3. Detailed representative phylogeny of mink- and human derived SARS-CoV-2 sequences. In order to make the tree topology clear, branch lengths are not to scale. The coloured boxes to the right of the tree show the allelic state of the three mink-associated GWAS hits in each terminal branch of the phylogeny, with dark red indicating the animal-associated alternate allele.

Figure 2-Figure Supplement 4. Detailed representative phylogeny of deer- and human derived SARS-CoV-2 sequences. In order to make the tree topology clear, branch lengths are not to scale. The coloured boxes to the right of the tree show the allelic state of the seven deer-associated GWAS hits that appeared in all ten replicate GWAS runs, with dark red indicating the animal-associated alternate allele.

667 **Supplementary Tables**
668
669 **Table S1**. GISAID accession numbers of all sequences used in this study.
670
671 **Table S2.** Number of viral sequences passing quality filters. The counts show the initial number
672 of sequences downloaded from GISAID from each animal species, and the remaining number
673 after each consecutive quality filter. The 'quality control' count shows the number of sequences
674 after removing those with incomplete sampling dates and/or >500 ambiguous bases (Ns). The
675 'post-alignment pruning' shows the count after removing sequences shorter than 29,000 bases
676 and/or with an insertion absent in all other sequences (introducing a gap in the alignment). The
677 'divergent tree branches' shows the count after removing sequences that introduce long branches
678 into the phylogeny (Methods). Ranges of counts indicate variation across tree replicates.
679

Species	*Raw downloaded (N sequences)*	*Post-quality control*	*Post-alignment pruning*	*Post-divergent tree branch pruning*
mink	1339	1046	1038	1036
deer	156	153	134	129-133
cat	120	100	78	78
dog	76	59	39	38-39

680

Table S3. Table of transmission counts for all candidate species, in both animal-to-human and human-to-animal direction, for both bootstrap-filtered and unfiltered cases.

species	replicate	filtered count (animal to human)	unfiltered count (animal to human)	filtered count (human to animal)	unfiltered count (human to animal)
Cat	1	5	5	59	69
Cat	2	4	4	60	68
Cat	3	5	5	58	68
Cat	4	4	4	57	69
Cat	5	5	5	57	68
Cat	6	5	5	57	68
Cat	7	4	4	59	69
Cat	8	4	4	61	68
Cat	9	4	4	62	68
Cat	10	4	4	55	68
Dog	1	2	3	32	35
Dog	2	1	1	31	36
Dog	3	1	1	31	36
Dog	4	2	2	30	36
Dog	5	2	3	32	35
Dog	6	2	3	32	35
Dog	7	1	1	32	36
Dog	8	1	1	29	36
Dog	9	1	1	35	36
Dog	10	1	1	31	36
Mink	1	35	121	48	67
Mink	2	40	110	36	63
Mink	3	37	107	35	66
Mink	4	37	108	48	66
Mink	5	38	104	45	68

Mink	6	33	94	42	65
Mink	7	42	125	37	63
Mink	8	36	116	39	68
Mink	9	38	110	46	64
Mink	10	44	128	47	62
Deer	1	0	2	37	57
Deer	2	1	1	39	56
Deer	3	1	1	40	59
Deer	4	2	3	35	57
Deer	5	0	0	38	55
Deer	6	1	1	40	54
Deer	7	0	2	37	55
Deer	8	0	0	40	62
Deer	9	1	2	39	55
Deer	10	1	2	38	42

Table S4. Human-derived sequence counts bearing each of the significant GWAS hits identified in deer inside and outside regions where deer sequences containing each mutation are found. Odds ratio and the p-values are reported following a Fisher's exact test. GWAS hits with OR < 1 or not significantly different from 1 are highlighted in green.

Position	Alternate allele count in deer regions	Wildtype allele count in deer regions	Alternate allele count outside deer regions	Wildtype allele count outside deer regions	Frequency of alternate allele in deer regions	Frequency of alternate allele outside deer regions	p_value	Conclusion
7303	433	838982	2226	10804361	0.000516101656531368	0.0002	8.6862e-56	ratio higher in deer regions
9430	1812	946458	9313	10688419	0.0019145065074203	0.0009	1.2416e-172	ratio higher in deer regions
14960	0	268814	266	11376922	0	2.3381e-05	0.0035	ratio lower in deer regions
20259	21	446625	417	11198939	4.70193115029387e-05	3.7236e-05	0.3171	ratio not significantly different

28016	20	213258	176	11432548	9.37831171632483e-05	1.5395e-05	8.6845e-10	ratio higher in deer regions
12073	210	321652	7184	11316956	0.000652879509532041	0.0006	0.6703	ratio not significantly different
29679	180	289240	2231	11354351	0.000622320564237312	0.0002	2.80664e-37	ratio higher in deer regions
5184	7012	238828	126835	11273327	0.0293600415361683	0.0112	0	ratio higher in deer regions
29750	1442	577400	8856	11058304	0.00249740214755802	0.0008	2.2500e-267	ratio higher in deer regions
7318	131	272041	1440	11372390	0.000481545061222389	0.0001	9.5330e-35	ratio higher in deer regions
16466	316970	860090	864781	9604161	0.368531200223232	0.0900	0	ratio higher in deer regions
7267	181	206553	3597	11435671	0.000876288410238534	0.0003	2.1368e-31	ratio higher in deer regions
210	307476	869584	854780	9614162	0.353589762461131	0.0889	0	ratio higher in deer regions
6730	32	20464	10034	11615472	0.00156372165754496	0.0007	0.0018	ratio higher in deer regions
27752	306844	870216	842927	9626015	0.35260670913888	0.0876	0	ratio higher in deer regions
11152	196	301370	9246	11335190	0.000650363340743936	0.0008	0.0013	ratio lower in deer regions
5822	7	39079	3456	11603460	0.00017912433787968	0.0003	0.2361	ratio not significantly different
9711	26	391615	1331	11253030	6.63917367822989e-05	0.0001	0.0020	ratio lower in deer regions
9679	79	561506	673	11083744	0.000140693064722372	6.0719	2.1058e-10	ratio higher in deer regions
7029	7	157363	2276	11486356	4.44831377134396e-05	0.0002	4.4009e-07	ratio lower in deer regions
29738	17	20479	11278	11614228	0.000830118658137604	0.0001	0.6520	ratio not significantly different

26767	318642	858418	871184	9597758	0.371196782919277	0.0908	0	ratio higher in deer regions
203	726	72179	19565	11553532	0.0100583272142867	0.0017	6.16135e-297	ratio higher in deer regions
12820	10	20486	1255	11624251	0.00048813824074978	0.0001	0.0001	ratio higher in deer regions
4540	322	195956	6718	11443006	0.00164322603033334	0.0006	7.7575e-55	ratio higher in deer regions
29666	1940	513027	27323	11103712	0.00378147738812967	0.0025	3.2392e-66	ratio higher in deer regions

Table S5. Human-derived sequence counts bearing each of the significant GWAS hits identified in mink inside and outside regions where mink sequences containing each mutation are found. Odds ratio and the p-values are reported following a Fisher's exact test. GWAS hits with OR < 1 or not significantly different from 1 are highlighted in green.

Position	Alternate allele count in mink regions	Wildtype allele count in mink regions	Alternate allele count outside mink regions	Wildtype allele count outside mink regions	Frequency of alternate allele in mink regions	Frequency of alternate allele outside mink regions	p_value	Conclusion
26047	23	239184	279	11406516	9.6160e-05	2.44597035589132e-05	1.11877702252856e-07	ratio higher in mink regions
12795	34	152594	340	11493034	0.0002	2.95831370550196e-05	2.75263696173362e-18	ratio higher in mink regions
23064	67	296468	5363	11344104	0.0002	0.000472756596730777	1.43477105492256e-11	ratio lower in mink regions

Table S6. Number of times Mink GWAS hits appear along human-to-mink transmission branches. The counts are summed across all branches and all 10 tree replicates.

species	Position	Number of times site is mutated on human-to-mink transition branches	Number of times the nucleotide substitution is identical to the GWAS hit
mink	26047	47	47
mink	12795	51	51
mink	23064	15	15

Table S7. Number of times Deer GWAS hits appear along human-to-deer transmission branches. The counts are summed across all branches and all 10 tree replicates. Yellow colored rows are mutations that never appear on a human-to-animal transmission branch over all deer replicates. The orange colored row corresponds to a nucleotide position mutated along human-to-animal transmission branches, but the substitution was never identical to the animal-associated allele identified by GWAS at that position.

species	Position	Number of times site is mutated on human-to-mink transition branches	Number of times the nucleotide substitution is identical to the GWAS hit
deer	7303	147	138
deer	9430	133	133
deer	14960	10	10
deer	20259	47	47
deer	28016	30	30
deer	12073	21	21
deer	29679	40	40
deer	5184	6	6
deer	29750	40	20
deer	7318	36	27
deer	16466	0	0
deer	7267	39	30
deer	210	0	0

deer	6730	0	0
deer	27752	7	7
deer	11152	26	26
deer	5822	11	11
deer	9711	26	26
deer	9679	15	15
deer	7029	14	14
deer	29738	3	3
deer	26767	0	0
deer	203	10	10
deer	12820	10	0
deer	4540	10	10
deer	29666	40	40

Table S8. SARS-CoV-2 mutations previously associated with non-human animal species in the literature. Insertions and deletions are not considered. For each of the studies, mutations reported in the main text or main tables are included. For the Pickering et al. (2022) study, substitutions found in deer in their study and also appear at least once in the database of previously reported deer sequences (listed in the paper's supplementary table S2) are included.

Mutation	Species	Study
S_N501T	**mink**	https://www.cell.com/cell-reports/pdf/S2211-1247(22)00060-2.pdf
S_Y453F	**mink**	https://www.cell.com/cell-reports/pdf/S2211-1247(22)00060-2.pdf
S_F486L	**mink**	https://www.cell.com/cell-reports/pdf/S2211-1247(22)00060-2.pdf
S_L452M	**mink**	https://www.nature.com/articles/s41467-021-27096-9
S_Y453F	**mink**	https://www.nature.com/articles/s41467-021-27096-9
S_F486L	**mink**	https://www.nature.com/articles/s41467-021-27096-9
S_N501T	**mink**	https://www.nature.com/articles/s41467-021-27096-9

NSP2_T85I	deer	https://www.nature.com/articles/s41564-022-01268-9#Sec2
NSP2_T528I	deer	https://www.nature.com/articles/s41564-022-01268-9#Sec2
NSP3_S1437F	deer	https://www.nature.com/articles/s41564-022-01268-9#Sec2
NSP4_H313Y	deer	https://www.nature.com/articles/s41564-022-01268-9#Sec2
NSP4_L353F	deer	https://www.nature.com/articles/s41564-022-01268-9#Sec2
NSP4_S386F	deer	https://www.nature.com/articles/s41564-022-01268-9#Sec2
NSP5_K90R	deer	https://www.nature.com/articles/s41564-022-01268-9#Sec2
NSP6_V149F	deer	https://www.nature.com/articles/s41564-022-01268-9#Sec2
NSP6_L260F	deer	https://www.nature.com/articles/s41564-022-01268-9#Sec2
NSP8_L169F	deer	https://www.nature.com/articles/s41564-022-01268-9#Sec2
NSP12_P323L	deer	https://www.nature.com/articles/s41564-022-01268-9#Sec2
NSP13_V348L	deer	https://www.nature.com/articles/s41564-022-01268-9#Sec2
NSP14_T113I	deer	https://www.nature.com/articles/s41564-022-01268-9#Sec2
S_T22I	deer	https://www.nature.com/articles/s41564-022-01268-9#Sec2
S_H49Y	deer	https://www.nature.com/articles/s41564-022-01268-9#Sec2
DEL	deer	https://www.nature.com/articles/s41564-022-01268-9#Sec2
S_Q613H	deer	https://www.nature.com/articles/s41564-022-01268-9#Sec2
S_D614G	deer	https://www.nature.com/articles/s41564-022-01268-9#Sec2
ORF3a_Q57H	deer	https://www.nature.com/articles/s41564-022-01268-9#Sec2
ORF3a_V256V	deer	https://www.nature.com/articles/s41564-022-01268-9#Sec2
N_T391I	deer	https://www.nature.com/articles/s41564-022-01268-9#Sec2
NSP9_G37E	mink	https://www.nature.com/articles/s41467-022-30698-6
S_F486L	mink	https://www.nature.com/articles/s41467-022-30698-6
S_N501T	mink	https://www.nature.com/articles/s41467-022-30698-6
S_Y453F	mink	https://www.nature.com/articles/s41467-022-30698-6
ORF3a_L219V	mink	https://www.nature.com/articles/s41467-022-30698-6
NSP3a_L1035F	deer	https://www.nature.com/articles/s41467-022-30698-6
NSP2_T170Q	mink	https://peerj.com/articles/10609/

NSP2_A192V	mink	https://peerj.com/articles/10609/
NSP2_H208Y	cat	https://peerj.com/articles/10609/
NSP2_R218C	mink	https://peerj.com/articles/10609/
NSP2_A225T	mink	https://peerj.com/articles/10609/
NSP2_E563V	mink	https://peerj.com/articles/10609/
NSP3_D135Y	cat	https://peerj.com/articles/10609/
NSP3_V234I	cat	https://peerj.com/articles/10609/
NSP3_P278L	mink	https://peerj.com/articles/10609/
NSP3_H295Y	mink	https://peerj.com/articles/10609/
NSP3_K384N	cat	https://peerj.com/articles/10609/
NSP3_I750V	mink	https://peerj.com/articles/10609/
NSP3_M770K	mink	https://peerj.com/articles/10609/
NSP3_D1283G	dog	https://peerj.com/articles/10609/
NSP5_I249T	cat	https://peerj.com/articles/10609/
NSP5_I259V	mink	https://peerj.com/articles/10609/
NSP9_G37E	mink	https://peerj.com/articles/10609/
NSP9_G37R	mink	https://peerj.com/articles/10609/
NSP10_K124E	cat	https://peerj.com/articles/10609/
NSP12_T25I	cat	https://peerj.com/articles/10609/
NSP12_M195I	mink	https://peerj.com/articles/10609/
NSP12_T802I	mink	https://peerj.com/articles/10609/
NSP13_I258V	mink	https://peerj.com/articles/10609/
NSP13_A446D	mink	https://peerj.com/articles/10609/
NSP15_A92T	mink	https://peerj.com/articles/10609/
S_G261D	mink	https://peerj.com/articles/10609/
S_G261D	mink	https://peerj.com/articles/10609/
S_A262S	mink	https://peerj.com/articles/10609/
S_V367F	mink	https://peerj.com/articles/10609/

S_L452M	mink	https://peerj.com/articles/10609/
S_Y453F	mink	https://peerj.com/articles/10609/
S_F486L	mink	https://peerj.com/articles/10609/
S_N501T	mink	https://peerj.com/articles/10609/
M_T175M	cat	https://peerj.com/articles/10609/
N_R41L	mink	https://peerj.com/articles/10609/
N_P80L	mink	https://peerj.com/articles/10609/
N_P199Q	mink	https://peerj.com/articles/10609/
N_T247I	cat	https://peerj.com/articles/10609/
ORF3a_H182L	mink	https://peerj.com/articles/10609/
ORF3a_H182Y	mink	https://peerj.com/articles/10609/
ORF3a_L219V	mink	https://peerj.com/articles/10609/
ORF3a_G224C	mink	https://peerj.com/articles/10609/
ORF3a_T229I	mink	https://peerj.com/articles/10609/
ORF6_R20S	mink	https://peerj.com/articles/10609/
ORF6_K23S	mink	https://peerj.com/articles/10609/
NSP2_T85I	mink	https://journals.plos.org/plospathogens/article/figure?id=10.1371/journal.ppat.1009952.t001
NSP3_S1206L	mink	https://journals.plos.org/plospathogens/article/figure?id=10.1371/journal.ppat.1009952.t001
NSP9_G37E	mink	https://journals.plos.org/plospathogens/article/figure?id=10.1371/journal.ppat.1009952.t001
NSP12_P323L	mink	https://journals.plos.org/plospathogens/article/figure?id=10.1371/journal.ppat.1009952.t001
NSP13_V187I	mink	https://journals.plos.org/plospathogens/article/figure?id=10.1371/journal.ppat.1009952.t001
NSP16_T91M	mink	https://journals.plos.org/plospathogens/article/figure?id=10.1371/journal.ppat.1009952.t001
S_N501T	mink	https://journals.plos.org/plospathogens/article/figure?id=10.1371/journal.ppat.1009952.t001
S_D614G	mink	https://journals.plos.org/plospathogens/article/figure?id=10.1371/journal.ppat.1009952.t001
S_K113T	mink	https://journals.plos.org/plospathogens/article/figure?id=10.1371/journal.ppat.1009952.t001

ORF3a_H128Y	**mink**	https://journals.plos.org/plospathogens/article/figure?id=10.1371/journal.ppat.1009952.t001
ORF3a_Q57H	**mink**	https://journals.plos.org/plospathogens/article/figure?id=10.1371/journal.ppat.1009952.t001
M_A38S	**mink**	https://journals.plos.org/plospathogens/article/figure?id=10.1371/journal.ppat.1009952.t001
N_T205I	**mink**	https://journals.plos.org/plospathogens/article/figure?id=10.1371/journal.ppat.1009952.t001
N_Q289H	**mink**	https://journals.plos.org/plospathogens/article/figure?id=10.1371/journal.ppat.1009952.t001
NSP6_L260F	**mink**	https://insight.jci.org/articles/view/159573
ORF1b_P314L	**mink**	https://www.frontiersin.org/articles/10.3389/fmicb.2021.698944/full
ORF1b_T730I	**mink**	https://www.frontiersin.org/articles/10.3389/fmicb.2021.698944/full
ORF1b_T2163I	**mink**	https://www.frontiersin.org/articles/10.3389/fmicb.2021.698944/full
S_Y453F	**mink**	https://www.frontiersin.org/articles/10.3389/fmicb.2021.698944/full
S_D614G	**mink**	https://www.frontiersin.org/articles/10.3389/fmicb.2021.698944/full
S_I692V	**mink**	https://www.frontiersin.org/articles/10.3389/fmicb.2021.698944/full
S_M1229I	**mink**	https://www.frontiersin.org/articles/10.3389/fmicb.2021.698944/full
ORF3a_H128Y	**mink**	https://www.frontiersin.org/articles/10.3389/fmicb.2021.698944/full
N_S194L	**mink**	https://www.frontiersin.org/articles/10.3389/fmicb.2021.698944/full
N_R203K	**mink**	https://www.frontiersin.org/articles/10.3389/fmicb.2021.698944/full
N_G204R	**mink**	https://www.frontiersin.org/articles/10.3389/fmicb.2021.698944/full

References

Adney, Danielle R., Jamie Lovaglio, Jonathan E. Schulz, Claude Kwe Yinda, Victoria A. Avanzato, Elaine Haddock, Julia R. Port, et al. 2022. “Severe Acute Respiratory Disease in American Mink (Neovison Vison) Experimentally Infected with SARS-CoV-2.” *bioRxiv*. https://doi.org/10.1101/2022.01.20.477164.

Bosco-Lauth, Angela M., Airn E. Hartwig, Stephanie M. Porter, Paul W. Gordy, Mary Nehring, Alex D. Byas, Sue VandeWoude, Izabela K. Ragan, Rachel M. Maison, and Richard A. Bowen. 2020. “Experimental Infection of Domestic Dogs and Cats with SARS-CoV-2: Pathogenesis, Transmission, and Response to Reexposure in Cats.” *Proceedings of the National Academy of Sciences of the United States of America* 117 (42): 26382–88.

Cai, Hugh Y., and Allison Cai. 2021. “SARS-CoV2 Spike Protein Gene Variants with N501T and G142D Mutation-Dominated Infections in Mink in the United States.” *Journal of Veterinary Diagnostic Investigation: Official Publication of the American Association of Veterinary Laboratory Diagnosticians, Inc* 33 (5): 939–42.

Cao, Zengguo, Hongjie Xia, Ricardo Rajsbaum, Xianzhu Xia, Hualei Wang, and Pei-Yong Shi. 2021. “Ubiquitination of SARS-CoV-2 ORF7a Promotes Antagonism of Interferon Response.” *Cellular & Molecular Immunology* 18 (3): 746–48.

Carlson, Colin J., Gregory F. Albery, Cory Merow, Christopher H. Trisos, Casey M. Zipfel, Evan A. Eskew, Kevin J. Olival, Noam Ross, and Shweta Bansal. "Climate change increases cross-species viral transmission risk." *Nature* 607, no. 7919 (2022): 555-562.

Chen, Chaoran, Sarah Nadeau, Michael Yared, Philippe Voinov, Ning Xie, Cornelius Roemer, and Tanja Stadler. 2021. “CoV-Spectrum: Analysis of Globally Shared SARS-CoV-2 Data to Identify and Characterize New Variants.” *Bioinformatics* , December. https://doi.org/10.1093/bioinformatics/btab856.

Chen, Peter E., and B. Jesse Shapiro. 2021. “Classic Genome-Wide Association Methods Are Unlikely to Identify Causal Variants in Strongly Clonal Microbial Populations.” *bioRxiv*. https://doi.org/10.1101/2021.06.30.450606.

Damas, Joana, Graham M. Hughes, Kathleen C. Keough, Corrie A. Painter, Nicole S. Persky, Marco Corbo, Michael Hiller et al. "Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates." *Proceedings of the National Academy of Sciences* 117, no. 36 (2020): 22311-22322.

De Maio, Nicola, Chieh-Hsi Wu, Kathleen M. O’Reilly, and Daniel Wilson. 2015. “New Routes to Phylogeography: A Bayesian Structured Coalescent Approximation.” *PLoS Genetics* 11 (8): e1005421.

Di Giorgio, Salvatore, Filippo Martignano, Maria Gabriella Torcia, Giorgio Mattiuz, and Silvestro G. Conticello. 2020. “Evidence for Host-Dependent RNA Editing in the Transcriptome of SARS-CoV-2.” *Science Advances* 6 (25): eabb5813.

Dudas, Gytis, Luiz Max Carvalho, Andrew Rambaut, and Trevor Bedford. "MERS-CoV spillover at the camel-human interface." *elife* 7 (2018): e31257.

Eckstrand, Chrissy D., Thomas J. Baldwin, Kerry A. Rood, Michael J. Clayton, Jason K. Lott, Rebecca M. Wolking, Daniel S. Bradway, and Timothy Baszler. 2021. “An Outbreak of SARS-CoV-2 with High Mortality in Mink (Neovison Vison) on Multiple Utah Farms.” *PLoS Pathogens* 17 (11): e1009952.

Elaswad, Ahmed, Mohamed Fawzy, Shereen Basiouni, and Awad A. Shehata. 2020. “Mutational Spectra of SARS-CoV-2 Isolated from Animals.” *PeerJ* 8 (December): e10609.

“Evolutionary Rate of SARS-CoV-2 Increases during Zoonotic Infection of Farmed Mink.” 2022. Virological. April 7, 2022. https://virological.org/t/evolutionary-rate-of-sars-cov-2-increases-during-zoonotic-infection-of-farmed-mink/792.

Gilbert, Clément, and Torstein Tengs. 2021. “No Species-Level Losses of s2m Suggests Critical Role in Replication of SARS-Related Coronaviruses.” *Scientific Reports* 11 (1): 16145.

775 Hadfield, James, Colin Megill, Sidney M. Bell, John Huddleston, Barney Potter, Charlton
776 Callender, Pavel Sagulenko, Trevor Bedford, and Richard A. Neher. 2018. "Nextstrain:
777 Real-Time Tracking of Pathogen Evolution." *Bioinformatics* 34 (23): 4121–23.
778 Harris, Reuben S., and Jaquelin P. Dudley. 2015. "APOBECs and Virus Restriction." *Virology*
779 479-480 (May): 131–45.
780 Hillen, Hauke S., Goran Kokic, Lucas Farnung, Christian Dienemann, Dimitry Tegunov, and
781 Patrick Cramer. 2020. "Structure of Replicating SARS-CoV-2 Polymerase." *Nature* 584
782 (7819): 154–56.
783 Katoh, Kazutaka, and Daron M. Standley. 2013. "MAFFT Multiple Sequence Alignment
784 Software Version 7: Improvements in Performance and Usability." *Molecular Biology and*
785 *Evolution* 30 (4): 772–80.
786 Kotwa, Jonathon D., Ariane Massé, Marianne Gagnier, Patryk Aftanas, Juliette Blais-Savoie,
787 Jeff Bowman, Tore Buchanan, et al. 2022. "First Detection of SARS-CoV-2 Infection in
788 Canadian Wildlife Identified in Free-Ranging White-Tailed Deer (Odocoileus Virginianus)
789 from Southern Québec, Canada." *bioRxiv*. https://doi.org/10.1101/2022.01.20.476458.
790 Kuchipudi, Suresh V., Meera Surendran-Nair, Rachel M. Ruden, Michele Yon, Ruth H. Nissly,
791 Kurt J. Vandegrift, Rahul K. Nelli, et al. 2022. "Multiple Spillovers from Humans and Onward
792 Transmission of SARS-CoV-2 in White-Tailed Deer." *Proceedings of the National Academy*
793 *of Sciences of the United States of America* 119 (6).
794 https://doi.org/10.1073/pnas.2121644119.
795 Liu, Guanqun, Jung-Hyun Lee, Zachary M. Parker, Dhiraj Acharya, Jessica J. Chiang, Michiel
796 van Gent, William Riedl, et al. 2021. "ISG15-Dependent Activation of the Sensor MDA5 Is
797 Antagonized by the SARS-CoV-2 Papain-like Protease to Evade Host Innate Immunity."
798 *Nature Microbiology* 6 (4): 467–78.
799 Liu, Yang, Jianying Liu, Kenneth S. Plante, Jessica A. Plante, Xuping Xie, Xianwen Zhang,
800 Zhiqiang Ku, et al. 2021. "The N501Y Spike Substitution Enhances SARS-CoV-2 Infection
801 and Transmission." *Nature*, November. https://doi.org/10.1038/s41586-021-04245-0.
802 Lu, Lu, Reina S. Sikkema, Francisca C. Velkers, David F. Nieuwenhuijse, Egil A. J. Fischer,
803 Paola A. Meijer, Noortje Bouwmeester-Vincken, et al. 2021. "Adaptation, Spread and
804 Transmission of SARS-CoV-2 in Farmed Minks and Associated Humans in the
805 Netherlands." *Nature Communications* 12 (1): 6802.
806 Miao, Zhichao, Antonin Tidu, Gilbert Eriani, and Franck Martin. 2021. "Secondary Structure of
807 the SARS-CoV-2 5'-UTR." *RNA Biology* 18 (4): 447–56.
808 Morales Lucia, Mateos-Gomez Pedro A., Capiscol Carmen, del Palacio Lorena, Enjuanes Luis,
809 and Sola Isabel. 2013. "Transmissible Gastroenteritis Coronavirus Genome Packaging
810 Signal Is Located at the 5′ End of the Genome and Promotes Viral RNA Incorporation into
811 Virions in a Replication-Independent Process." *Journal of Virology* 87 (21): 11579–90.
812 Murall, Carmen Lía, Eric Fournier, Jose Hector Galvez, Arnaud N'Guessan, Sarah J. Reiling,
813 Pierre-Olivier Quirion, Sana Naderi, et al. 2021. "A Small Number of Early Introductions
814 Seeded Widespread Transmission of SARS-CoV-2 in Québec, Canada." *Genome Medicine*
815 13 (1): 169.
816 Newman, Joseph A., Alice Douangamath, Setayesh Yadzani, Yuliana Yosaatmadja, Antony
817 Aimon, José Brandão-Neto, Louise Dunnett, et al. 2021. "Structure, Mechanism and
818 Crystallographic Fragment Screening of the SARS-CoV-2 NSP13 Helicase." *Nature*
819 *Communications* 12 (1): 4848.
820 Nguyen, Lam-Tung, Heiko A. Schmidt, Arndt von Haeseler, and Bui Quang Minh. 2015. "IQ-
821 TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood
822 Phylogenies." *Molecular Biology and Evolution* 32 (1): 268–74.
823 Oreshkova, Nadia, Robert Jan Molenaar, Sandra Vreman, Frank Harders, Bas B. Oude
824 Munnink, Renate W. Hakze-van der Honing, Nora Gerhards, et al. 2020. "SARS-CoV-2
825 Infection in Farmed Minks, the Netherlands, April and May 2020." *Euro Surveillance:*

Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin 25 (23). https://doi.org/10.2807/1560-7917.ES.2020.25.23.2001005.

Otto, Sarah P., Troy Day, Julien Arino, Caroline Colijn, Jonathan Dushoff, Michael Li, Samir Mechai, et al. 2021. “The Origins and Potential Future of SARS-CoV-2 Variants of Concern in the Evolving COVID-19 Pandemic.” *Current Biology: CB* 0 (0). https://doi.org/10.1016/j.cub.2021.06.049.

Oude Munnink, Bas B., Reina S. Sikkema, David F. Nieuwenhuijse, Robert Jan Molenaar, Emmanuelle Munger, Richard Molenkamp, Arco van der Spek, et al. 2021. “Transmission of SARS-CoV-2 on Mink Farms between Humans and Mink and back to Humans.” *Science* 371 (6525): 172–77.

Paradis, Emmanuel, and Klaus Schliep. 2019. “Ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R.” *Bioinformatics* 35 (3): 526–28.

Pickering, Bradley, Oliver Lung, Finlay Maguire, Peter Kruczkiewicz, Jonathan D. Kotwa, Tore Buchanan, Marianne Gagnier, et al. 2022. “Highly Divergent White-Tailed Deer SARS-CoV-2 with Potential Deer-to-Human Transmission.” *bioRxiv*. https://doi.org/10.1101/2022.02.22.481551.

Porter, Ashleigh F., Damian FJ Purcell, Benjamin P. Howden, and Sebastian Duchene. "Evolutionary rate of SARS-CoV-2 increases during zoonotic infection of farmed mink." *Virus Evolution* (2023).

Redondo, Natalia, Sara Zaldívar-López, Juan J. Garrido, and Maria Montoya. 2021. “SARS-CoV-2 Accessory Proteins in Viral Pathogenesis: Knowns and Unknowns.” *Frontiers in Immunology* 12 (July): 708264.

Rochman, Nash D., Guilhem Faure, Yuri I. Wolf, Peter L. Freddolino, Feng Zhang, and Eugene V. Koonin. "Epistasis at the SARS-CoV-2 receptor-binding domain interface and the propitiously boring implications for vaccine escape." *MBio* 13, no. 2 (2022): e00135-22.

Rosenberg, Brad R., Claire E. Hamilton, Michael M. Mwangi, Scott Dewell, and F. Nina Papavasiliou. 2011. “Transcriptome-Wide Sequencing Reveals Numerous APOBEC1 mRNA-Editing Targets in Transcript 3’ UTRs.” *Nature Structural & Molecular Biology* 18 (2): 230–36.

Salter, Jason D., and Harold C. Smith. 2018. “Modeling the Embrace of a Mutator: APOBEC Selection of Nucleic Acid Ligands.” *Trends in Biochemical Sciences* 43 (8): 606–22.

Sheahan, Timothy P., Amy C. Sims, Shuntai Zhou, Rachel L. Graham, Andrea J. Pruijssers, Maria L. Agostini, Sarah R. Leist, et al. 2020. “An Orally Bioavailable Broad-Spectrum Antiviral Inhibits SARS-CoV-2 in Human Airway Epithelial Cell Cultures and Multiple Coronaviruses in Mice.” *Science Translational Medicine* 12 (541). https://doi.org/10.1126/scitranslmed.abb5883.

Shi, Jianzhong, Zhiyuan Wen, Gongxun Zhong, Huanliang Yang, Chong Wang, Baoying Huang, Renqiang Liu, et al. 2020. “Susceptibility of Ferrets, Cats, Dogs, and Other Domesticated Animals to SARS-Coronavirus 2.” *Science* 368 (6494): 1016–20.

Shin, Donghyuk, Rukmini Mukherjee, Diana Grewe, Denisa Bojkova, Kheewoong Baek, Anshu Bhattacharya, Laura Schulz, et al. 2020. “Papain-like Protease Regulates SARS-CoV-2 Viral Spread and Innate Immunity.” *Nature* 587 (7835): 657–62.

Starr, Tyler N., Allison J. Greaney, Sarah K. Hilton, Daniel Ellis, Katharine H. D. Crawford, Adam S. Dingens, Mary Jane Navarro, et al. 2020. “Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding.” *Cell* 182 (5): 1295–1310.e20.

Tan, Cedric C. S., Su Datt Lam, Damien Richard, Christopher J. Owen, Dorothea Berchtold, Christine Orengo, Meera Surendran Nair, et al. 2022. “Transmission of SARS-CoV-2 from Humans to Animals and Potential Host Adaptation.” *Nature Communications* 13 (1): 1–13.

Tengs, Torstein, Anja Bråthen Kristoffersen, Tsvetan R. Bachvaroff, and Christine Monceyron Jonassen. 2013. “A Mobile Genetic Element with Unknown Function Found in Distantly

Related Viruses." *Virology Journal* 10 (April): 132.
Vöhringer, Harald S., Theo Sanderson, Matthew Sinnott, Nicola De Maio, Thuy Nguyen, Richard Goater, Frank Schwach, et al. 2021. "Genomic Reconstruction of the SARS-CoV-2 Epidemic in England." *Nature*, October. https://doi.org/10.1038/s41586-021-04069-y.
Wacker, Anna, Julia E. Weigand, Sabine R. Akabayov, Nadide Altincekic, Jasleen Kaur Bains, Elnaz Banijamali, Oliver Binas, et al. 2020. "Secondary Structure Determination of Conserved SARS-CoV-2 RNA Elements by NMR Spectroscopy." *Nucleic Acids Research* 48 (22): 12415–35.
Wei, Changshuo, Ke-Jia Shan, Weiguang Wang, Shuya Zhang, Qing Huan, and Wenfeng Qian. 2021. "Evidence for a Mouse Origin of the SARS-CoV-2 Omicron Variant." *bioRxiv*. https://doi.org/10.1101/2021.12.14.472632.
Worobey, Michael, Joshua I. Levy, Lorena M. Malpica Serrano, Alexander Crits-Christoph, Jonathan E. Pekar, Stephen A. Goldstein, Angela L. Rasmussen, et al. 2022. *The Huanan Market Was the Epicenter of SARS-CoV-2 Emergence*. https://doi.org/10.5281/zenodo.6299116.
Yang, Dong, and Julian L. Leibowitz. 2015. "The Structure and Functions of Coronavirus Genomic 3' and 5' Ends." *Virus Research* 206 (August): 120–33.
Yen, Hui-Ling, Thomas H. C. Sit, Christopher J. Brackman, Shirley S. Y. Chuk, Haogao Gu, Karina W. S. Tam, Pierra Y. T. Law, et al. 2022. "Transmission of SARS-CoV-2 Delta Variant (AY.127) from Pet Hamsters to Humans, Leading to Onward Human-to-Human Transmission: A Case Study." *The Lancet* 399 (10329): 1070–78.
Zhou, Jie, Thomas P. Peacock, Jonathan C. Brown, Daniel H. Goldhill, Ahmed M. E. Elrefaey, Rebekah Penrice-Randal, Vanessa M. Cowton, et al. 2022. "Mutations That Adapt SARS-CoV-2 to Mink or Ferret Do Not Increase Fitness in the Human Airway." *Cell Reports* 38 (6): 110344.