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

Sana Naderi[1], Peter Chen[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 and 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 considerable 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 mammalian species, both non-human and human, to identify mutations associated with each species. We found the highest frequency of animal-to-human transmission from mink, compared with negligible transmission from other sampled species (cat, dog, and deer). Although inferred transmission events could be limited by potential 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; however, three SNVs strongly associated with mink and 26 with deer. These SNVs are candidates for future experiments to identify their role in differential pathogenesis, immune escape and host response modulation. Together, our results are consistent with sustained deer-to-deer transmission and highlight the importance of studying animal-associated SARS-CoV-2 mutations to assess their potential impact on 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 ("Evolutionary Rate of SARS-CoV-2 Increases during Zoonotic Infection of Farmed Mink" 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. Alternatively, it could permit significant genetic drift, opening up new peaks in the human-adaptive fitness landscape. For example, 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). Although this scenario remains hypothetical and is not exclusive of other hypotheses involving evolution in unsampled human populations or a chronically infected individual, it illustrates the potential dramatic consequences of evolution in animal reservoirs. 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 individual species of animals, 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). Evidence suggests that certain mutations may improve replication fitness of SARS-CoV-2 in animal hosts. For instance, experimental infections of mink (*Neovison vison*) and ferret *(Mustela furo*) identified mutations within the viral spike protein that increase binding to the ACE2 receptor in these animals, while decreasing infection of human airway cells (Zhou et al. 2022). One of these mutations (N501T) was also found to be prevalent in infected mink sampled in the United States (Cai and Cai 2021; Eckstrand et al. 2021).

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 broad dataset. Using phylogenetic methods, we inferred transmission events between humans and four other frequently infected animals and quantified their relative frequencies. 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 also identified mutations associated with specific animal species. We recovered the 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 (**Supplementary Table S1**). After applying sequence quality filters and considering only animals with 30 or more sequences, we were left with four species: mink, deer, dog, and cat. Our filters excluded common experimental animals such as mice and ferrets. 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 close human-derived relatives, and this percentage is slightly lower for cat (85.6%), dog (89%), and mink (91.7%). The variance in these percentages could be explained by sampling bias and the varying levels of sampling effort. For example, cats and dogs might be undersampled relative to mink and deer.

To investigate potential animal-to-human transmission events in greater detail, including the direction of transmission, we used ancestral state reconstruction on viral phylogenetic trees. We are aware that such ancestral state reconstruction can be biased by differential sampling across species. Our goal is, therefore, not to infer absolute rates of cross-species transmission, but rather to provide a consistent comparative framework for interspecies transmission. For each candidate species, the animal-derived sequences, and their closest relative human-derived sequence were combined with ten random sub-samplings of human-derived sequences ($n \approx 50$ 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 1a-d**) and human-to-animal transitions on each tree (**Supplementary Table S2**). To determine a 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 27.74, 1.29, and 0.52 respectively) compared to the permutations (266.06, 2.84, and 1.87) while the observed counts in dog (standard deviation of 1.10) is similar to the permutations (0.91). In the human-to-animal direction, the observed standard deviations in transmission counts are 0.48, 0.73, 2.17, and 2.17 for cat, dog, mink and deer, respectively, compared to permuted values of 6.66, 1.70, 186.05, and 16.46. These results show that, in general, our inference of transmission events converges on a non-random value, but that the estimate for dog-to-human transmission might be less reliable.

Based on the bootstrap-filtered counts, we inferred less than one transmission event from animals back to humans on average from cats, dogs, and deer (**Figure 1e, Table 1**). In contrast, there were an average of 12 or more transmission events inferred from mink to human. 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 human to animal, was generally higher and much more uniform across species (**Figure 1f, Table 1**), with lower bounds in the range of 4.6 to 12.8 events. However, the higher sampling of human-derived compared with animal-derived sequences may have inflated the human-to-animal events relative to the animal-to-human events. Mink are also better sampled than the other animal species (**Table 1**), and further sampling of other animals could identify more transmission events. 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 higher (or at least 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	12.5 – 108.3	0.3 – 1.1	0 – 4.4	0.1 – 1.6
Human-to-animal	12.8 – 65.4	11.6 – 59.5	8.5 – 68.3	4.6 – 35.0

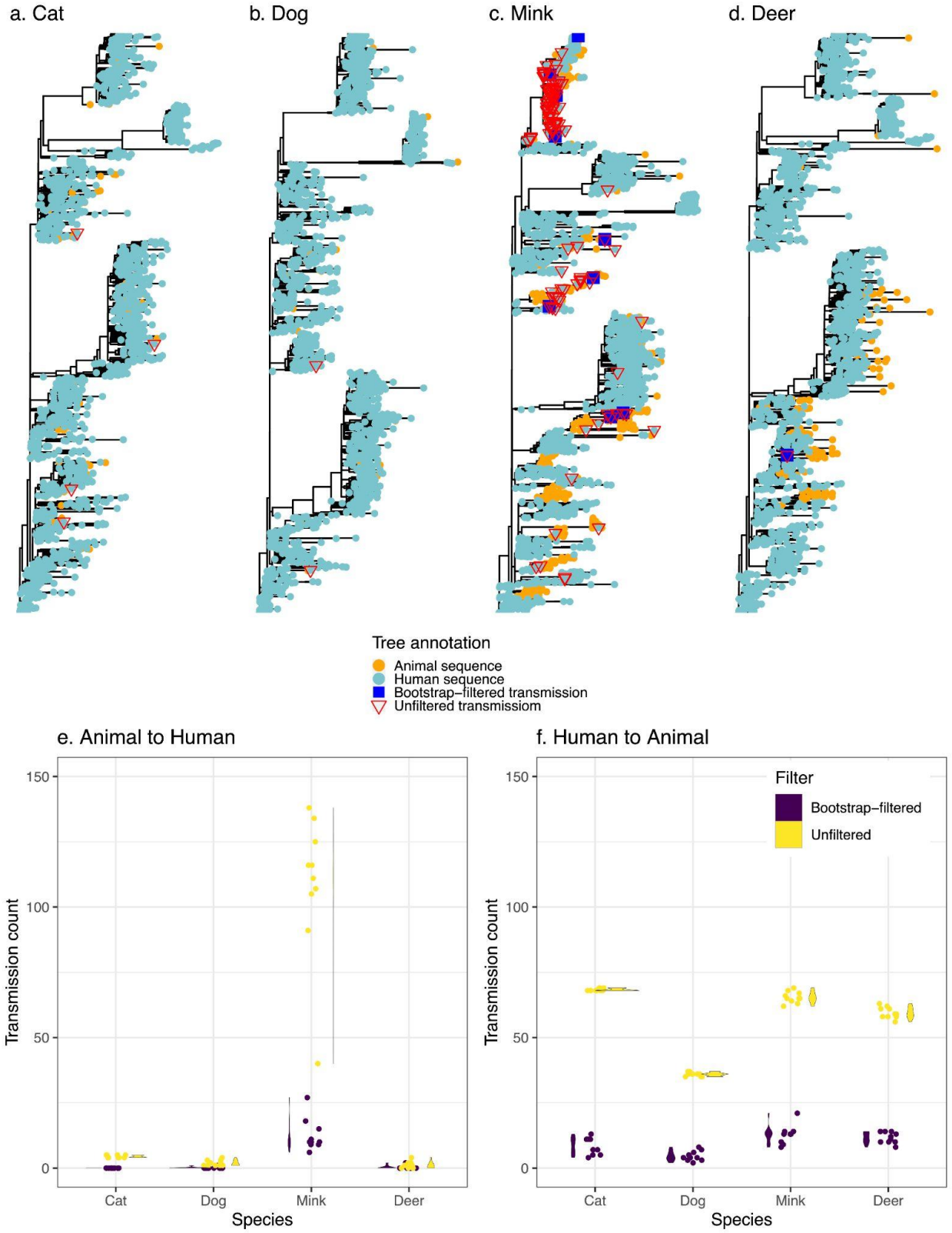

Figure 1. Transmission events inferred between humans and animals. Panels a-d display a representative tree for every species with animal to human transmissions marked on the tree. Trees are

rooted with the Wuhan reference genome. Panels e and f display the distribution (violin plots alongside points plotted with jitter to avoid overlap) of inferred transmission counts (across 10 replicate trees) in each animal species, in both bootstrap-filtered and unfiltered trees.

Next, we sought to identify mutations associated with particular animal species compared to humans. These mutations could be candidates for species-specific adaptations. To do so, we conducted GWAS using POUTINE, a method which implicitly controls for 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 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 2**). In all cases, we used a significance cutoff of a family-wise $p < 0.05$ to correct for multiple hypothesis testing. The mink GWAS revealed three unique SNVs across all ten runs (**Table 2**). One of these hits appears in all ten replicates and the remaining two appear in at least half of the replicates. All three of these mutations are non-synonymous, including N501T which was previously associated with a mink outbreak in the United States (Cai and Cai 2021). The deer GWAS revealed 26 unique statistically significant SNVs across all ten runs, of which seven appear in all ten replicates, and five in at least half the runs (**Table 3**). Out of these 26 hits, five are intergenic (within the 5′ and 3′ UTRs) and 12 are synonymous mutations. Notably, 21 of the hits are C>U transition mutations.

Table 2. **Single-nucleotide variants associated with mink by GWAS.** "Pos" refers to the nucleotide position in the reference genome. Homoplasy counts in focal animals (cases), humans (controls), and *P*-values are averaged across replicates in which the site's family-wise *P*-values was < 0.05. Where applicable, amino acid positions refer to the polyprotein with mature protein positions in parenthesis.

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
26047	U	G	L219V	ORF3a	6	0	0.0014	0.0365	10
12795	G	A	G4177E (nsp9 G37E)	ORF1ab/pp1ab/nsp9/replicase	6	0	0.0015	0.0368	6
23064	A	C	N501T	Spike/S1/RBD/binds ACE2	6.4	0	0.0010	0.0258	5

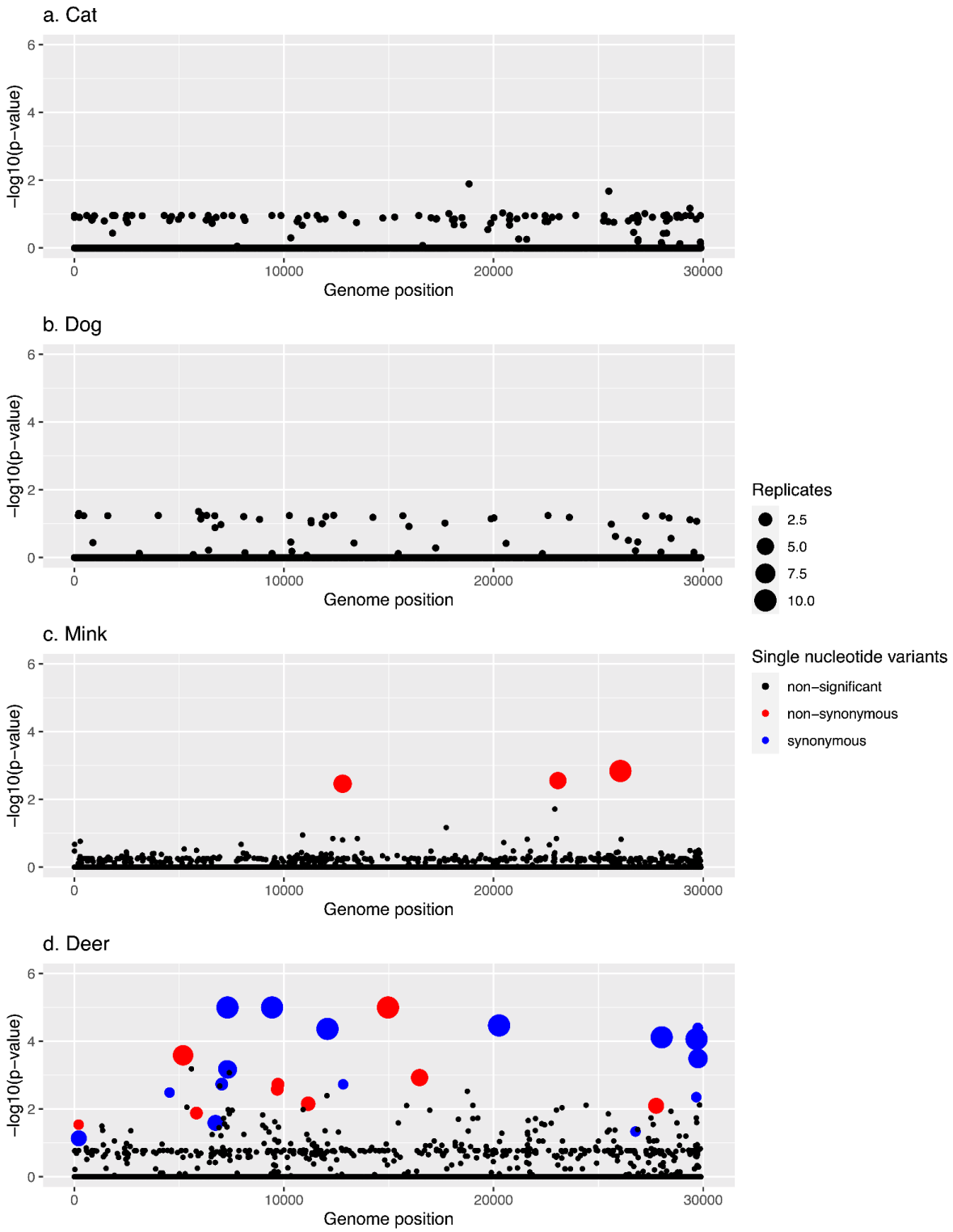

Figure 2: Manhattan plots summarizing GWAS hits in each animal species. In every panel, the x-axis represents the nucleotide position in the SARS-CoV-2 reference genome and the y-axis represents the -log10 of the pointwise *p*-values averaged over replicates. Statistically significant hits with family-wise corrected *p*-values of lower than 0.05 are shown in red (non-synonymous) or blue (synonymous), while non-statistically significant p-values are in black.

Table 3. **Single-nucleotide variants associated with deer by GWAS**. "Pos" refers to the nucleotide position in the reference genome. Homoplasy counts in focal animals (cases), humans (controls), and *P*-values are averaged across replicates in which the site's family-wise *P*-values was < 0.05. Where applicable, amino acid positions refer to the polyprotein with mature protein positions in parenthesis.

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
7303	C	U	I2346I (nsp3 1524)	ORF1a/pp1ab/ pp1a/nsp3	17.8	1.2	9.99E-06	9.99E-06	10
9430	C	U	I3055I (nsp4 I292I)	ORF1a/pp1ab/ pp1a/nsp4	15.2	6.2	9.99E-06	9.99E-06	10
14960	A	U	N4899I (nsp12 N507I)	ORF1ab/pp1ab /nsp12/RdRp	7.8	0.1	9.99E-06	1.09E-05	10
20259	C	U	F6665F (nsp15 F213F)	ORF1ab/pp1ab /nsp15	4.8	0.1	3.39E-05	0.0013	10
28016	C	U	F41F	ORF8	4	0	7.59E-05	0.0061	10
12073	C	U	D3936D (nsp7 D67D)	ORF1a/pp1ab/ pp1a/nsp7	5.2	1.1	4.29E-05	0.0025	10
29679	C	U	IG	3'UTR	5	1.8	8.59E-05	0.0055	10
5184	C	U	P1640L (nsp3 P822L)	ORF1a/pp1ab/ pp1a/nsp3	4.6	1.6	0.0002	0.0115	8
29750	C	U	IG	3'UTR/S2M	5	2.6	0.0002	0.0103	7
7318	C	U	F2351F (nsp3 F1533F)	ORF1a/pp1ab/ pp1a/nsp3	4	0.3	0.0001	0.0114	6
16466	C	U	P5401L (nsp13 P77L)	ORF1ab/pp1ab /nsp13/Hel	5	1	4.99E-05	0.0019	5
7267	C	U	F2334F (nsp3 F1516F)	ORF1a/pp1ab/ pp1a/nsp3	4.4	0.8	9.39E-05	0.0079	5
210	G	U	IG	5'UTR/SL5a	4	0.5	0.0001	0.0136	4
6730	C	U	N2155N (nsp3 N1337N)	ORF1a/pp1ab/ pp1a/nsp3	4	0.75	0.0002	0.0168	4
27752	C	U	T120I	ORF7a	4	0.75	0.0002	0.0169	4
11152	C	U	V3629V (nsp6 V60V)	ORF1a/pp1ab/ pp1a/nsp6	4	0.7	0.0002	0.0153	3
5822	C	U	L1853F (nsp3 L1035F)	ORF1a/pp1ab/ pp1a/nsp3	4	0.5	0.0001	0.0118	2
9711	C	U	S3149F (nsp4 S386F)	ORF1a/pp1ab/ pp1a/nsp4	4	0.5	8.49E-05	0.0118	2
9679	C	U	F3138F (nsp4 F375F)	ORF1a/pp1ab/ pp1a/nsp4	4	0	9.49E-05	0.0067	2
7029	C	U	S2255F (nsp3 S1437F)	ORF1a/pp1ab/ pp1a/nsp3	4	0.5	0.0002	0.0149	2
29738	C	A	IG	3'UTR/S2M	4	0	3.99E-05	0.0059	1

26767	U	C	I82T	ORF5/M	4	0	8.99E-05	0.0057	1
203	C	U	IG	5'UTR/SL5a	4	1	0.0003	0.0191	1
12820	A	G	L4185L (nsp9 L45L)	ORF1a/pp1ab/ pp1a/nsp9	5	1	3.99E-05	0.0009	1
4540	C	U	Y1425Y (nsp3 Y607Y)	ORF1a/pp1ab/ pp1a/nsp3	4	1	0.0002	0.0239	1
29666	C	U	L37F	ORF10	4	1	0.0002	0.0219	1

Discussion

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 largely in experimental studies or region-specific sampling efforts that are difficult to generalize. Our goal in this study was to apply standardized methods to identify animal-to-human transmission events and to discover animal-associated mutations using viral genomic data readily available to date. The results are necessarily biased by sampling effort. Nevertheless, our approach provides a 'level playing field' upon which to assess the results of previous animal-specific studies. Our study also provides a pipeline for researchers to assess the transmission and evolution of SARS-CoV-2 within animals and between animals and humans.

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, essentially making them "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 a prospective household transmission study. Nonetheless, transmission from humans to animals was remarkably uniform across species and small differences across species may be explained by sampling effort. 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).

Despite the large sample size of mink-derived viral sequences, we only detected three mink-associated SNVs using GWAS, including the previously identified N501T spike mutation. This is consistent with relatively little time for SARS-CoV-2 to adapt to mink between transmission cycles in humans. In contrast, we detected many more SNVs associated with deer despite a smaller sample size. This suggests a greater number of deer-adapted SNVs compared to mink-adapted SNVs, perhaps due to multiple uninterrupted cycles of deer-deer transmission. Sustained deer-deer transmission is also supported by previous studies (Kuchipudi et al. 2022), including the observation of relatively divergent SARS-CoV-2 genomes

(Pickering et al. 2022). Asymptomatic or mild disease in deer relative to mink (Oreshkova et al. 2020) might also allow more opportunity for evolution and transmission during prolonged infections. Despite pruning out highly divergent branches, including several of the Canadian sequences involved in a potential deer-to-human transmission event (Pickering et al. 2022), some relatively long deer-associated branches are notable in our study (**Figure 1d**). Together, these results point to deer as an important reservoir of novel SARS-CoV-2 mutations. Ongoing monitoring of deer-to-human transmission events is therefore warranted.

Several viral mutations have been previously associated with cat and dog (Elaswad et al. 2020), but we found no statistically significant mutations associated with either of these species. This could be due to limited viral adaptation, which would be expected if cats and dogs are effectively dead-end hosts for the virus, with little time for cycles of intra-species transmission. Alternatively, the limited sample size for these species could have limited GWAS power to identify adaptive mutations. Given the higher viral load and shedding in cats compared to dogs (Bosco-Lauth et al. 2020; Shi et al. 2020), we expect greater adaptation and transmission in cats; however further data will be needed to test this expectation.

We identified three statistically significant mutations associated with mink. The substitution L219V, which appears in all ten GWAS replicates, has been previously detected as a substitution associated with mink, in the ORF3a accessory gene (Elaswad et al. 2020). However, the previous detection was not statistically tested for significance. Similarly, the G4177E (nsp9 G37E) mutation has been identified previously in mink-derived sequences (Eckstrand et al. 2021). The N501T substitution has also been previously associated with mink (Lu et al. 2021; Elaswad et al. 2020) and ferret (Zhou et al. 2022), and the same site, has also been reported to be adaptive in mice, with an N501Y amino acid substitution. The mink-associated N501T substitution has been associated with increased binding to human ACE2 (Starr et al. 2020). The N501Y substitution has been detected in several human SARS-CoV-2 VOCs with higher transmissibility, such as the Alpha variant (Y. Liu et al. 2021; Vöhringer et al. 2021). Interestingly, other mutations that have been widely associated with mink such as Y453F (Zhou et al. 2022; Lu et al. 2021; Elaswad et al. 2020) and L260F (Adney et al. 2022) that were suggested to arise as a result of rapid adaptation, do not appear as statistically significant mutations in our GWAS.

Many of the deer-associated mutations are synonymous and occur broadly across the genome outside of the relatively well-studied spike protein. Notably, half of the mutations identified by GWAS were synonymous mutations, primarily transition mutations (92.3%, 12/13), which are typically generated at a higher frequency than transversions. The deer-associated synonymous mutations showed no particular bias toward codons that are preferred in deer relative to humans; therefore selection for codon usage optimization does not easily explain these associations (data not shown). We also note that the vast majority of mutations were C>U transitions, which may be a reflection of APOBEC-mediated RNA editing, which results in deamination of cytosine to uracil in single-stranded RNA (Harris and Dudley 2015; Salter and Smith 2018). Consistent with this hypothesis, C>U transitions in deer contained the consensus sequence [U/A][U/A]C[A/U][A/U], which resembles that observed for human

APOBEC1-mediated deamination [AU]C[AU]; however, it remains to be seen whether the deer APOBEC1 isoform has the same substrate specificity or whether there is increased APOBEC expression or activity in deer tissue (Di Giorgio et al. 2020; Rosenberg et al. 2011)). Alternatively, the C>U transitions may be related to RNA secondary structure or nucleotide composition biases required for genome condensation for viral replication organelle biogenesis or virion assembly, respectively, in the deer host.

Our GWAS revealed several mutations in deer that localized to distinct RNA secondary structures in the 5´ and 3´ UTRs of the viral RNA. Specifically, in the 5´ UTR, both mutations (C203T and G210T) localized to stem loop 5a (SL5a), a highly conserved stem-loop structure previously implicated in virion assembly in related coronaviruses (Yang and Leibowitz 2015; Morales Lucia et al. 2013; Miao et al. 2021). Similarly, in the 3´ UTR, all the mutations localized to the 3´ terminal stem-loop structure, with two specifically localized to the S2M region (C29738A and C29750T) (Yang and Leibowitz 2015; Gilbert and Tengs 2021; Wacker et al. 2020). Interestingly, while the 3´ terminal stem-loop structure is known to be hypervariable, the S2M region is highly conserved in sequence and structure across a wide range of RNA viruses, including members of the *Astroviridae*, *Caliciviridae*, *Picornaviridae,* and *Coronaviridae* families (Tengs et al. 2013; Gilbert and Tengs 2021). While the role of this S2M region is poorly understood, its strict conservation across a range of positive-strand RNA viruses suggests that it may play an important role in the viral life cycle. Notably, both the mutations identified by GWAS are predicted to maintain the overall S2M consensus fold, so may reflect differences in species-specific interactions between S2M and host proteins or RNA molecules.

The deer GWAS also revealed a few nonsynonymous mutations in viral proteins important in RNA binding and host antiviral responses. Specifically, N4899I (a.k.a. N507I) lies in the nsp12 RNA-dependent RNA polymerase, within the highly conserved motif G which is important in positioning the 5´ template strand during viral RNA synthesis (Sheahan et al. 2020). While N507 is known to make contact with the +2 nucleotide of the template strand, experimental investigations will be needed to understand how the N507I mutation impacts the active site structure and/or viral RNA synthesis (Hillen et al. 2020). Additionally, the P5401L (a.k.a. P77L) mutation in the nsp13 helicase is predicted to be a solvent-exposed residue within the N-terminal Zinc-binding domain (Newman et al. 2021). Given that this is distant from the helicase active site, we predict that it is unlikely to affect helicase enzymatic activity; but rather that it may be important for protein folding or protein-protein interactions. Finally, we identified two mutations that may have implications for host adaptation as these were identified in proteins known to interact with the host antiviral response. Specifically, P1460L (a.k.a. P822L) in the nsp3 protease localizes to the deubiquitinating site in the PLpro domain which overlaps with the ISG15 binding site, suggesting it may modulate host antiviral responses (Yang and Leibowitz 2015; Shin et al. 2020; G. Liu et al. 2021). Interestingly, the deer GWAS also revealed a T120I mutation within the ORF7a accessory protein. This residue is adjacent to K119, which is implicated in the inhibition of the antiviral response in human cells (Redondo et al. 2021; Cao et al. 2021). Specifically, K119 polyubiquitination has been shown to block STAT2 phosphorylation, leading to inhibition of type I IFN. Thus, it is possible that the T120I mutation modulates ubiquitination at the K119 residue in deer; however, this will require experimental validation.

Upon finalization of this manuscript, another analysis of SARS-CoV-2 transmission and potential host adaptation appeared, using a similar dataset from GISAID (Tan et al. 2022). This work focused only on transmission from humans to other animals, while we also considered the animal-to-human direction. Their approach to identifying animal-associated mutations was conceptually similar to ours – focusing on homoplasic mutations and a set of reasonable but arbitrary filters for allele frequencies and effect sizes – whereas we took a more formal statistical GWAS approach. Our three mink GWAS hits are a subset of the four identified in the other study, and neither study identified any deer-associated mutations in Spike (Tan et al. 2022). Our study identified more significantly deer-associated mutations (including the single hit reported by Tan et al. at nsp3 L1035F), which could be due to different data filtering and significance testing approaches. Overall, the two studies are broadly concordant, complementary, and pave the way for future studies on larger datasets.

In conclusion, while the dynamics of anthroponosis and zooanthroponosis for SARS-CoV-2 is still unclear, cross-species transmission events are likely to continue to occur given the continued geographically widespread infections, high transmission rates, and emergence of new variants. While our study suggests that cats and dogs may be “dead-end” hosts, we observed several significant substitutions in mink and deer suggestive of sustained animal-animal transmission and perhaps reflective of host adaptation. This suggests that continuous molecular surveillance of SARS-CoV-2 animal isolates is likely to reveal new insights into SARS-CoV-2 host range and adaptation, and may contribute to our understanding of risk for spillback of new variants. This also highlights the need to monitor for similar patterns of susceptibility to infection and sustained intra-species transmission in related species. Our study draws attention to several specific, statistically significant nucleotide and amino acid substitutions that may play a role in host adaptation, pathogenesis, and/or transmission and are candidates for experimental study.

Methods

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

From the 7.6 million human-derived viral sequences present in GISAID's human alignment dated February 28, 2022, a set of closely related sequences for every animal species was extracted. To do so, we used Nextstrain (Hadfield et al. 2018) to calculate a proximity matrix based on pairwise substitutions between every animal-derived sequence and all ~8 million human-derived sequences. The query alignment was generated using MAFFT (Katoh and Standley 2013) in which animal sequences were aligned to the gapped version of the Wuhan

Iv04 reference sequence present in GISAID's human-host alignment dated February 28, 2022, with the "keeplength" option of the software enabled to maintain the length of context alignment in the query alignment.

Using this matrix and noting that some sequences from a given species might share a number of close relatives, we extracted the closest human-derived sequences for every sequence of a given 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. 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.

Ancestral state reconstruction: Discrete ancestral state reconstruction was performed on all 40 trees in the study, with states set as "human" and "animal". The reconstruction was done

using the maximum likelihood method “ace” implemented in the “ape” R package (Paradis and Schliep 2019). Ancestral state reconstruction was done using the “equal rates model” which allows for an equal probability of transition in both directions of “animal to human” and “human to animal” on the tree.

Estimating transmission events: Once the states were labeled, the set of all branches along which a transition from an “animal” node state to a “human” node state has occurred was identified, and the human end of the branch, namely the most recent common ancestor of the introduced human clade or the human tip in case of singletons, was marked as a transition node. In order to identify an independent set of transitions and to avoid the redundancy of reporting nested spilled-back clades as separate events, the most basal of these transition nodes were identified. In order to provide a well-rounded record of spillovers, the same analysis was done in the human to animal direction as well. For setting a lower bound to the transmission counts, in another set of analyses, all transitions in the desired direction were identified, then the branches whose parent side had a bootstrap support of lower than 75% were discarded, and subsequently, the most basal transitions were identified. To validate the ancestral state reconstruction and the transmission counts, a permutation test was done in which the tips of every 10 trees for all candidate species were randomly shuffled in 1000 permutations, and the same algorithm for ancestral state reconstruction and subsequently counting transitions in both directions (animal to human and human to animal) was applied to find the most basal transitions, unfiltered counts on the shuffled trees.

Genome-wide association studies: We used POUTINE (Chen and Shapiro 2021) to scan the genome for mutations that are statistically associated with each animal species. POUTINE relies on the viral phylogeny to identify homoplasic mutations associated with a phenotype of interest, in this case human vs. animal hosts. By considering only homoplasic mutations at the tips of the tree, POUTINE implicitly accounts for population structure. POUTINE was run on the 40 tree replicates (10 per species) and in each run, animal-host sequences were set as “cases” and 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.

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 Tables

Table S1. GISAID accession numbers of all sequences used in this study.

Table S2. 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	Animal to human (Bootstrap-filtered)	Animal to human (Un-filtered)	Human to animal (Bootstrap-filtered)	Human to animal (Un-filtered)
Mink	1	15	138	21	63
Mink	2	9	111	14	66
Mink	3	6	116	13	69
Mink	4	10	107	9	62
Mink	5	18	40	13	68
Mink	6	9	134	8	65
Mink	7	10	125	10	67
Mink	8	27	105	14	65
Mink	9	10	91	13	65
Mink	10	11	116	13	64
Deer	1	0	0	10	58
Deer	2	1	1	14	58
Deer	3	0	0	11	59
Deer	4	0	1	10	59
Deer	5	0	0	13	58
Deer	6	2	2	14	56
Deer	7	0	1	8	63
Deer	8	0	2	10	61
Deer	9	0	0	14	62
Deer	10	0	4	12	61
Cat	1	0	5	7	69
Cat	2	0	4	11	68
Cat	3	0	5	5	68
Cat	4	0	4	11	69
Cat	5	0	5	5	68
Cat	6	0	5	7	68

Cat	7	0	4	13	69
Cat	8	0	4	11	68
Cat	9	0	4	11	68
Cat	10	0	4	4	68
Dog	1	0	4	5	35
Dog	2	0	1	8	36
Dog	3	0	1	7	37
Dog	4	0	2	3	37
Dog	5	0	3	4	36
Dog	6	1	3	2	36
Dog	7	0	1	4	35
Dog	8	0	1	3	36
Dog	9	0	1	6	36
Dog	10	0	2	4	35

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.

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.

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.

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.

Hadfield, James, Colin Megill, Sidney M. Bell, John Huddleston, Barney Potter, Charlton Callender, Pavel Sagulenko, Trevor Bedford, and Richard A. Neher. 2018. "Nextstrain: Real-Time Tracking of Pathogen Evolution." *Bioinformatics* 34 (23): 4121–23.

Harris, Reuben S., and Jaquelin P. Dudley. 2015. "APOBECs and Virus Restriction." *Virology* 479-480 (May): 131–45.

Hillen, Hauke S., Goran Kokic, Lucas Farnung, Christian Dienemann, Dimitry Tegunov, and Patrick Cramer. 2020. "Structure of Replicating SARS-CoV-2 Polymerase." *Nature* 584 (7819): 154–56.

Katoh, Kazutaka, and Daron M. Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution* 30 (4): 772–80.

Kotwa, Jonathon D., Ariane Massé, Marianne Gagnier, Patryk Aftanas, Juliette Blais-Savoie, Jeff Bowman, Tore Buchanan, et al. 2022. "First Detection of SARS-CoV-2 Infection in Canadian Wildlife Identified in Free-Ranging White-Tailed Deer (Odocoileus Virginianus) from Southern Québec, Canada." *bioRxiv*. https://doi.org/10.1101/2022.01.20.476458.

Kuchipudi, Suresh V., Meera Surendran-Nair, Rachel M. Ruden, Michele Yon, Ruth H. Nissly, Kurt J. Vandegrift, Rahul K. Nelli, et al. 2022. “Multiple Spillovers from Humans and Onward Transmission of SARS-CoV-2 in White-Tailed Deer.” *Proceedings of the National Academy of Sciences of the United States of America* 119 (6). https://doi.org/10.1073/pnas.2121644119.

Liu, Guanqun, Jung-Hyun Lee, Zachary M. Parker, Dhiraj Acharya, Jessica J. Chiang, Michiel van Gent, William Riedl, et al. 2021. “ISG15-Dependent Activation of the Sensor MDA5 Is Antagonized by the SARS-CoV-2 Papain-like Protease to Evade Host Innate Immunity.” *Nature Microbiology* 6 (4): 467–78.

Liu, Yang, Jianying Liu, Kenneth S. Plante, Jessica A. Plante, Xuping Xie, Xianwen Zhang, Zhiqiang Ku, et al. 2021. “The N501Y Spike Substitution Enhances SARS-CoV-2 Infection and Transmission.” *Nature*, November. https://doi.org/10.1038/s41586-021-04245-0.

Lu, Lu, Reina S. Sikkema, Francisca C. Velkers, David F. Nieuwenhuijse, Egil A. J. Fischer, Paola A. Meijer, Noortje Bouwmeester-Vincken, et al. 2021. “Adaptation, Spread and Transmission of SARS-CoV-2 in Farmed Minks and Associated Humans in the Netherlands.” *Nature Communications* 12 (1): 6802.

Miao, Zhichao, Antonin Tidu, Gilbert Eriani, and Franck Martin. 2021. “Secondary Structure of the SARS-CoV-2 5’-UTR.” *RNA Biology* 18 (4): 447–56.

Morales Lucia, Mateos-Gomez Pedro A., Capiscol Carmen, del Palacio Lorena, Enjuanes Luis, and Sola Isabel. 2013. “Transmissible Gastroenteritis Coronavirus Genome Packaging Signal Is Located at the 5′ End of the Genome and Promotes Viral RNA Incorporation into Virions in a Replication-Independent Process.” *Journal of Virology* 87 (21): 11579–90.

Murall, Carmen Lía, Eric Fournier, Jose Hector Galvez, Arnaud N’Guessan, Sarah J. Reiling, Pierre-Olivier Quirion, Sana Naderi, et al. 2021. “A Small Number of Early Introductions Seeded Widespread Transmission of SARS-CoV-2 in Québec, Canada.” *Genome Medicine* 13 (1): 169.

Newman, Joseph A., Alice Douangamath, Setayesh Yadzani, Yuliana Yosaatmadja, Antony Aimon, José Brandão-Neto, Louise Dunnett, et al. 2021. “Structure, Mechanism and Crystallographic Fragment Screening of the SARS-CoV-2 NSP13 Helicase.” *Nature Communications* 12 (1): 4848.

Nguyen, Lam-Tung, Heiko A. Schmidt, Arndt von Haeseler, and Bui Quang Minh. 2015. “IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies.” *Molecular Biology and Evolution* 32 (1): 268–74.

Oreshkova, Nadia, Robert Jan Molenaar, Sandra Vreman, Frank Harders, Bas B. Oude Munnink, Renate W. Hakze-van der Honing, Nora Gerhards, et al. 2020. “SARS-CoV-2 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.
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.
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.