# Transmission of SARS-CoV-2 from humans to animals and potential host adaptation

Cedric C.S. Tan[1,2], Su Datt Lam[3,4], Damien Richard[1,5], Christopher Owen[1], Dorothea Berchtold[1], Christine Orengo[4], Meera Surendran Nair[6,7], Suresh V. Kuchipudi[6,7], Vivek Kapur[7,8], Lucy van Dorp[1], François Balloux[1]

1) UCL Genetics Institute, University College London, United Kingdom.
2) Genome Institute of Singapore, A*STAR, Singapore.
3) Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Malaysia.
4) Institute of Structural and Molecular Biology, University College London, United Kingdom.
5) Infection, Immunity and Inflammation, University College London, United Kingdom.
6) Animal Diagnostics Lab, Department of Veterinary and Biomedical Sciences, The Pennsylvania State University, USA.
7) Huck Institutes of the Life Sciences, The Pennsylvania State University, USA.
8) Department of Animal Science, The Pennsylvania State University, USA.

Correspondence: cedriccstan@gmail.com (C.C.S. Tan).

L.v.D and F.B. are co-last authors.

**Keywords:**

SARS-CoV-2, COVID-19, human-to-animal transmission, mink, white-tailed deer, host-adaptation

## Abstract

SARS-CoV-2, the agent of the COVID-19 pandemic, can infect a wide range of mammals. Since its spread in humans, secondary host jumps of SARS-CoV-2 from humans to a variety of domestic and wild populations of mammals have been documented. The evolution of SARS-CoV-2 in different host species is of fundamental interest while also providing indication of how SARS-CoV-2 may have adapted to human hosts soon after the initial host jump, a time window for which there are no genome sequences available. Moreover, the study of SARS-CoV-2 circulating in animals is critical to assess the risk that the transmission of animal-adapted viral lineages back into humans (i.e., spillback) may pose. Here, we compared the genomic landscapes of SARS-CoV-2 isolated from animal species relative to that in humans, profiling the mutational biases indicative of potentially different selective pressures in animals. We focused on viral genomes collected in infected mink (*Neovison vison*) and white-tailed deer (*Odocoileus virginianus*) for which reports of multiple independent spillover events and subsequent animal-to-animal transmission are available. We identified six candidate mutations for animal-specific adaptation in mink (NSP9_G37E, Spike_F486L, Spike_N501T, Spike_Y453F, ORF3a_T229I, ORF3a_L219V), and one in deer (NSP3a_L1035F), though these mutations appear to confer minimal advantage for circulation in humans. Additionally, circulation of SARS-CoV-2 in mink and deer has not caused considerable changes to the evolutionary trajectory of SARS-CoV-2 thus far. Finally, our results suggest that minimal adaptation was required for human-to-animal spillover and subsequent onward transmission in mink and deer, highlighting the 'generalist' nature of SARS-CoV-2 as a pathogen of mammalian hosts.

## Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is able to infect multiple mammalian host species (1), which is a characteristic seen in other coronaviruses (2). SARS-CoV-2 emerged in humans in late 2019 (3), presumably after animal-to-human transmission (i.e., zoonotic spillover) of an ancestral viral lineage belonging to the subgenus *Sarbecovirus* that circulated in bats (4,5). The evolutionary events after the zoonotic host jump but preceding the initial COVID-19 outbreak in Wuhan, China (late December 2019) remain poorly characterised due to the lack of genomic sequences collected during that period. Further, SARS-CoV-2 has likely been circulating in humans for some time before it was formally detected. This is expected given the remarkably high rate of asymptomatic infections in humans (~40%) (6). A critical mass of cases presenting with the severe disease must be reached before alerting infectious disease surveillance efforts.

For a successful host jump of a pathogen and its subsequent emergence, several traits must be acquired. One key prerequisite is the ability to infect cells of the novel host, which depends on the presence of compatible host cell receptors. SARS-CoV-2 can infect cells of multiple mammalian host species (7–9), primarily due to the conservation of the angiotensin-converting enzyme 2 (ACE2), the primary host cell receptor used for viral entry, across mammals (10–12). Another essential trait is the ability to transmit efficiently within populations of the novel host. Infections of host populations that do not efficiently transmit the pathogen further, also known as 'dead-end' hosts, may quickly lead to pathogen extinction within that population. Dogs, which are susceptible to SARS-CoV-2 infection but do not efficiently transmit the virus (1) are a possible example of a dead-end host. On the other hand, human-to-human transmission is rapid, with early estimates of the mean number of subsequent infections produced by an infectious person in a totally naïve population (i.e., basic reproductive number, $R_0$) ranging from 1.5-6.5 (13).

Evolutionary analyses suggest that both efficient human-to-human transmission and ACE2 usage was not acquired recently, but in ancestral bat-associated lineages (4,5), suggesting that SARS-CoV-2 was well pre-adapted for circulation in humans prior to its emergence. Consistent with this, early efforts to identify mutations associated to SARS-CoV-2's transmissibility failed to identify obvious candidates for adaptation to its human host (14,15). However, with the recent emergence of more transmissible Variants of Concern (VoC) such as Alpha, Delta and Omicron, it is generally accepted that SARS-CoV-2 is still adapting to its human host (16), maintaining its fitness in the face of increasing vaccine coverage and infection-acquired immunity in the human population.

After its initial zoonotic host jump into humans, multiple secondary host jumps of SARS-CoV-2 from humans into animals (i.e., human-to-animal spillover) and significant transmission have been reported for domestic and wild mammals. This offers potential insights into the early evolutionary dynamics leading to and following host jumps. As of 16th November 2021, a total of 1,060 high quality SARS-CoV-2 genomes associated with natural or experimental infection of 15 animal species have been deposited on GISAID (17,18) (**Table 1**). The first animal-associated outbreaks

seeded by human-to-animal spillover events emerged in mink farms in the Netherlands in April 2020 (19), and subsequently in Denmark in June 2020 (20,21), where transmission was rapid. Indeed, initial testing found that 65% of mink (*Neovison vison*) in Danish mink farms had been infected by late June 2020 (22). Further, SARS-CoV-2 in minks were found to transmit readily back into humans (i.e., spillback) (23). These findings prompted culls of minks in Dutch mink farms in early June 2020 (24). Separately, in November 2020, an initial report from Denmark raised concerns about the emergence of a mink-associated SARS-CoV-2 lineage circulating in humans and minks of farms in Northern Jutland, Denmark (25). This 'mink-derived' lineage, touted as the 'cluster 5 variant', possessed five mutations in the Spike protein (H69/V70 deletion, Y453F, D614G, I692V, M1229I) and showed evidence of immune escape (25–27), which led to the subsequent decision to cull approximately 17 million Danish minks (28).

**Table 1. Summary of high-quality animal-associated SARS-CoV-2 genomes.** Animal taxa that were associated with less than 10 isolates were excluded from this table for brevity.

| Host taxon | Common name | Population type | No. of isolates | No. of countries | Onward transmission |
|---|---|---|---|---|---|
| *Neovison vison* | American mink | Farmed | 789 | 9 | Yes (19–21,23,29,30) |
| *Odocoileus virginianus* | White-tailed deer | Wild | 73 | 1 | Yes (31–34) |
| *Felis catus domesticus* | Domestic cat | Pet/stray | 65 | 13 | Yes (35–38) |
| *Canis lupus familiaris* | Domestic dog | Pet/stray | 32 | 4 | Unlikely (39) |
| *Panthera spp.* | Big cats | Captive | 73 | 5 | Yes (40,41) |

In the last quarter of 2021, studies reporting human-to-animal spillover into wild white-tailed deer (*Odocoileus virginianus*) in the USA began to surface (31,32,34,42). White-tailed deer are one of the most abundant wild ruminants in the USA, and some of these spillover events were associated with the start of the regular deer hunting season (42). Significant onward transmission was observed, with ~30% of sampled deer being SARS-CoV-2-positive in Iowa (42) and Ohio (31), and a reported 40% seroprevalence across four US states (34).

Of fundamental interest is whether SARS-CoV-2 required host-adaptive mutations to jump into animal hosts, the extent of host-specific adaptation following its host jumps, and how the introduction of SARS-CoV-2 into animals impacts the evolutionary trajectory of the virus. Given the rapid and extensive onward transmission in mink and deer, there was likely ample opportunity

for the virus to adapt to circulation in these host populations. Further, the rapid testing and intensive sequencing efforts early into these outbreaks offer a glimpse of key evolutionary events surrounding spillovers and the establishment of new host reservoirs. To this end, we focused on published and publicly available sequences isolated from mink and deer, analysing these animal-associated sequences relative to carefully curated subsamples of human SARS-CoV-2. In particular, we looked for changes to the mutational biases, genomic composition, and mutational rates in animal SARS-CoV-2 clusters relative to human associated counterparts. Additionally, we screened for mutations that may have arisen due to host-specific adaptation and subsequently assessed the potential impact of these mutations bioinformatically. Our results shed light on changes to the evolutionary and genomic landscape of a pandemic virus following its human-to-animal host jumps.

## Results

*Multiple human-to-animal spillover events of SARS-CoV-2*

Following the global spread of SARS-CoV-2, spillover of the virus from humans into domestic and wild animal species have been documented. Placement of the animal-associated genomes shown in **Table 1** recapitulates these multiple independent human-to-animal spillover events (**Figure 1a**). The clustering of animal isolates on the global phylogeny correlates well with the different species-specific transmission potentials and the extent of transmission amongst animal populations, though this could also in part be due to differential sampling efforts. Cat and dog isolates appear as highly polyphyletic singletons reflecting the poor animal-to-animal transmission in companion animals in addition to sparse sequencing efforts. Separately, we find small clusters of isolates from big cats (i.e., *Panthera spp.*), reflecting outbreaks of SARS-CoV-2 in multiple species of captive zoo animals around the world (40,41,43,44).

Manual inspection of the global phylogeny supports a minimum number of 24 and 20 phylogenetically distinct clusters of SARS-CoV-2 in mink and deer due to independent spillover events of multiple human SARS-CoV-2 lineages. Several large clusters were observed in mink (**Figure 1b**), with the largest mink cluster in Denmark (20) reaching >300 sequenced infections. This reflects the efficient mink-to-mink transmission of SARS-CoV-2 in intensive farming settings. Additionally, we find multiple moderately sized clusters of SARS-CoV-2 in deer that represent frequent spillover events due to the geographical overlap of deer and human habitats, followed by substantial deer-to-deer transmission.

Finally, we find that the animal outbreaks were seeded by 71 of the 1,508 PANGO lineages (45) that have been circulating in humans prior to 16 Nov 2021, including the Alpha, Delta and Mu variant lineages. The 71 PANGO lineages found in animals do not seem restricted to particular clades of the global diversity of SARS-CoV-2 and instead appear to be broadly representative of the different lineages circulating in humans. This suggests that efficient onward transmission to animals is generally not restricted to a particular subset of SARS-CoV-2 lineages in circulation in humans.

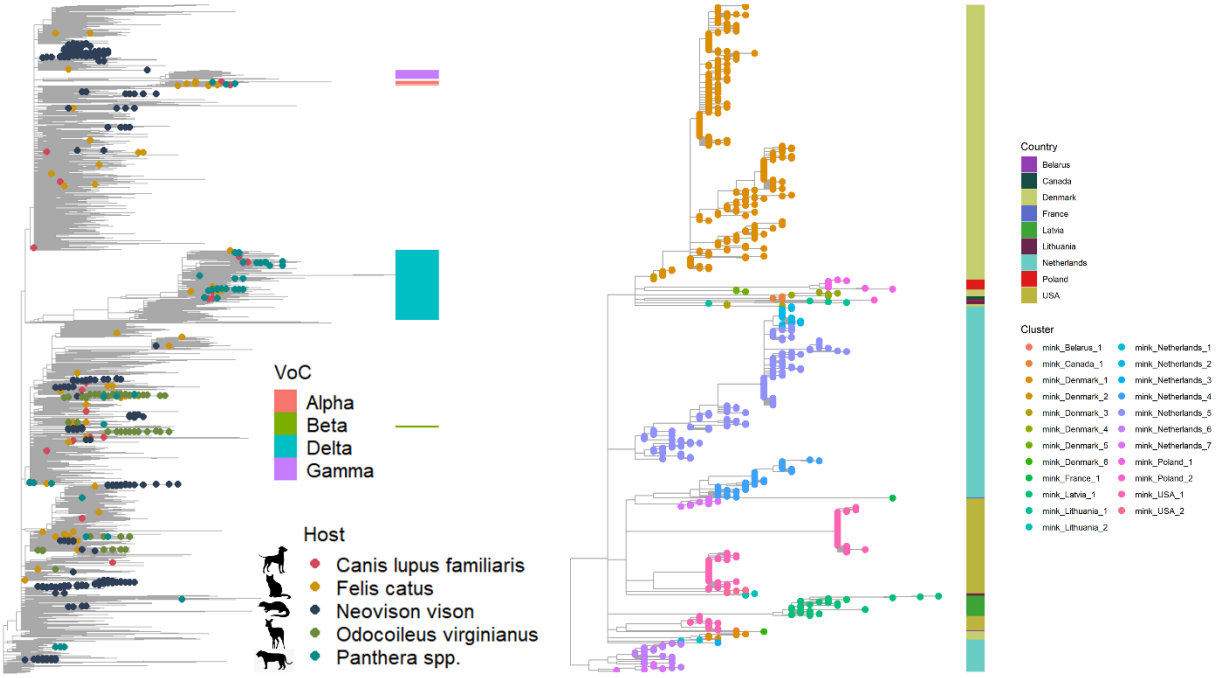


**Figure 1. Multiple emergences and onward transmission of SARS-CoV-2 in animals.** (a) Subsampled Audacity tree ($n$ = 15,846) comprising 10 human isolates per PANGO lineage, and all animal isolates shown in **Table 1**, illustrating the global context of SARS-CoV-2 infections in animals. (b) Maximum likelihood tree of all 789 mink isolates, with manually curated cluster names (see '**Methods**') and country of isolation annotated.

*Homoplasy and allele frequency analyses identify candidate mutations for host-specific adaptation*

To identify candidates for host adaptation, we compared mink SARS-CoV-2 sequences to a roughly similar number of human isolates with matching PANGO lineage, range of sampling dates and country origin (human background 1, see '**Methods**'). This allowed us to identify 18 and 31 candidate mutations which may be the result of mink or deer-specific adaptation, respectively. These mutations were (A) at a two-fold higher allele frequency in animal than human isolates and (B) had an animal allele frequency > 0.1, or (C) have emerged at least thrice independently in each animal host-only phylogeny (**Figure 2a** and **b**). Since spillover events involve only a subset of human viral lineages, selectively neutral mutations that were already present in these lineages may appear homoplastic following spillover into independent animal populations. As such mutations that (D) were not inherited from the parent human lineage are more likely to be adaptive. This can be determined by visually inspecting the animal isolates in the context of human background 1 (**Figure S1a**). The genomic and residue positions, allele frequencies and number of emergences for the 18 putative mink- and 31 deer-specific candidate mutations are shown in **Table S1**.

Of the identified mutations, five non-synonymous changes in minks (NSP9_G37E, Spike_F486L, Spike_N501T, ORF3a_T229I, ORF3a_L219V) and one in deer (NSP3_L1035F) fulfilled all four criteria in addition to (E) being present in at least three independent clusters (**Figure 2c** and **d**), are the strongest candidates for putative host adaptation. Three synonymous changes satisfying criteria (A)-(E) (NSP2_C1348T, NSP3a_C7303T, NSP4_C9430T) were also found in deer-associated SARS-CoV-2 but may have a more cryptic relationship with protein function and host-adaptation. Interestingly, none of the strong candidate mutations satisfying criteria (A)-(E) in deer were found in the Spike protein.

Notably, though Spike_Y453F has been shown to improve Spike:mink-ACE2 interactions and suggested to be mink-adaptive (26,27), its frequency in mink-associated virus is comparable to those considered in human background 1 (see '**Methods'**). Inspection of a subsampled Audacity tree comprising all mink and human isolates collected in Denmark prior to 1 December 2020 (**Figure S1b**) found that the mink and human isolates in the mink_Denmark_1 cluster are interspersed, suggesting complex back-and-forth transmission patterns between minks and humans. This makes it difficult to interpret whether the mutation first arose in human lineages and spilled over into minks, or the inverse. Nevertheless, exclusion of the mink_Denmark_1 cluster results in Y453F being at greater than two-fold frequency in minks relative to humans, satisfying criterion (A). We therefore consider Y453F to also be a strong candidate mink-adaptive mutation.

Separately, we did not find any mutations that were fixed in the animal populations and at a considerably lower frequency in humans. Under a scenario where key host-specific mutations must be acquired for an expansion of host tropism and subsequent spillover, we expect such mutations to be fixed in viruses isolated from the novel animal host, but at a lower frequency in the primary

host. As such, the absence of fixed mutations suggest that host-specific adaptation was not necessary for human-to-animal spillover of SARS-CoV-2 into mink and deer.

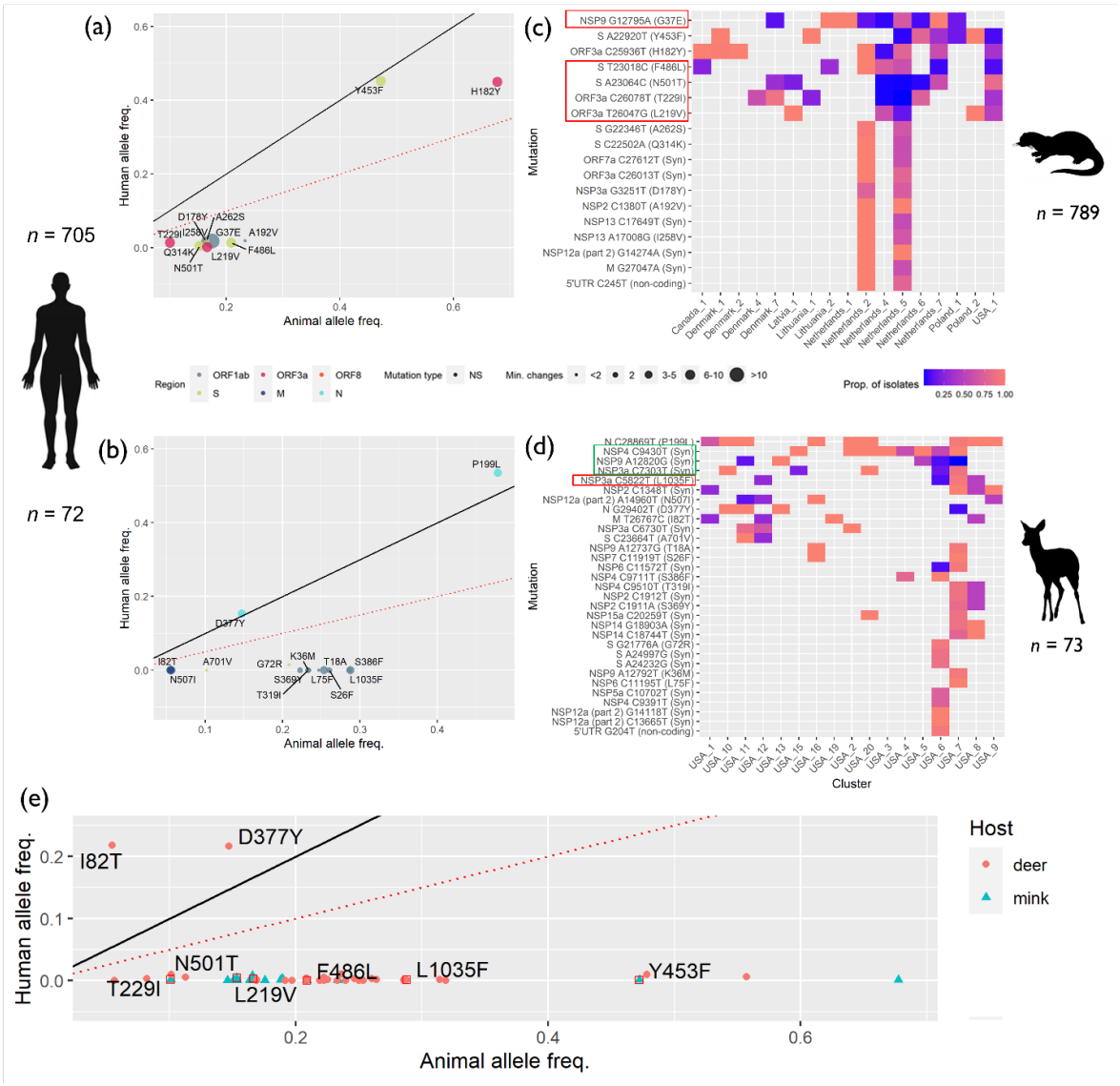


**Figure 2. Homoplasy and allele frequency analysis.** Scatter plot of putatively adaptive non-synonymous mutations in (a) mink and (b) deer. Point size represents the minimum number of independent emergences for each mutation in a phylogeny reconstructed from 789 mink or 73 deer isolates. Human isolates with matching PANGO lineages, from the same countries, and that were sampled within the range of sampling dates of mink ($n$ = 705) or deer isolates ($n$ = 72), were used to compute the human background allele frequencies (human background 1). The dotted red lines and solid black lines, indicate where the allele frequencies in each animal host are two-fold that in humans, and where the human and animal allele frequencies are equal, respectively. Heatmap visualising the proportions of mutation-carrying SARS-CoV-2 isolates within manually curated phylogenetic clusters in (c) mink and (d) deer. The strongest candidate mutations are indicated by red boxes. (e) Allele frequencies of 18 mink and 31 deer candidate mutations in human background 2. Strong candidate non-synonymous and synonymous mutations satisfying criteria (A)-(E) are indicated by red and green boxes, respectively. The protein associated to each mutation is given by the colour in panel (b).

Finally, we compared the frequencies of candidate mutations in animals relative to those in all human lineages within the same country regardless of sampling time (human background 2) to infer host-specific selective pressures acting on these mutations. We find that most of the candidate mutations prevalent in animal isolates are almost non-existent in human isolates (**Figure 2e**), suggesting that while these mutations may be tolerated/adaptive in animals, they may be selected against in humans. The only two exceptions are mutations in the membrane and nucleocapsid proteins, M_I82T and N_D377Y.

*Emergence of candidate animal-adaptive mutations largely predates documented human-to-animal spillovers*

We placed the range of sampling dates of animal isolates in the context of the broader COVID-19 pandemic timeline (**Figure 3a**). Three of the seven strongest candidates (ORF3a_T229I, Spike_N501T, NSP3_L1035F) emerged in humans early in the initial wave of the pandemic, predating the first documented emergences in animal lineages. This is also the case for M_I82T and N_D377Y, which are at a higher frequency in human lineages than in animals (**Figure 2e**). Further, the earliest sequenced human isolates carrying either of these mutations did not originate from the countries where the associated animal outbreaks occurred (T229I-Italy; N501T-Italy; I82T-Canada; D377Y-India), except for L1035F (USA). These mutations were also found in human isolates distributed across a large timespan and across 57-166 countries, even where mink or deer populations are not present (**Figure 3b**). Since we find no evidence of strong positive selection acting on these mutations in humans (**Figure 2e**; see the previous section), carriage of these alleles across many countries is not expected. These observations support the hypothesis that these mutations may have emerged during the evolutionary history of SARS-CoV-2 in humans independent of human-to-animal-to-human transmission. On the other hand, the strong candidate mink-adaptive mutations (NSP9_G37E, ORF3a_L219V, Spike_F486L, Spike_Y453F) emerged in humans after the first mink outbreaks in the Netherlands. Some of the earliest human isolates carrying these mutations were first sequenced in the Netherlands and Denmark. Further, the human isolates carrying these mutations tend to originate from countries where human-to-mink SARS-CoV-2 transmission has been reported (46) (**Figure 3c**). These findings suggest an association of these mutations with human-to-mink spillover and subsequent spillback.

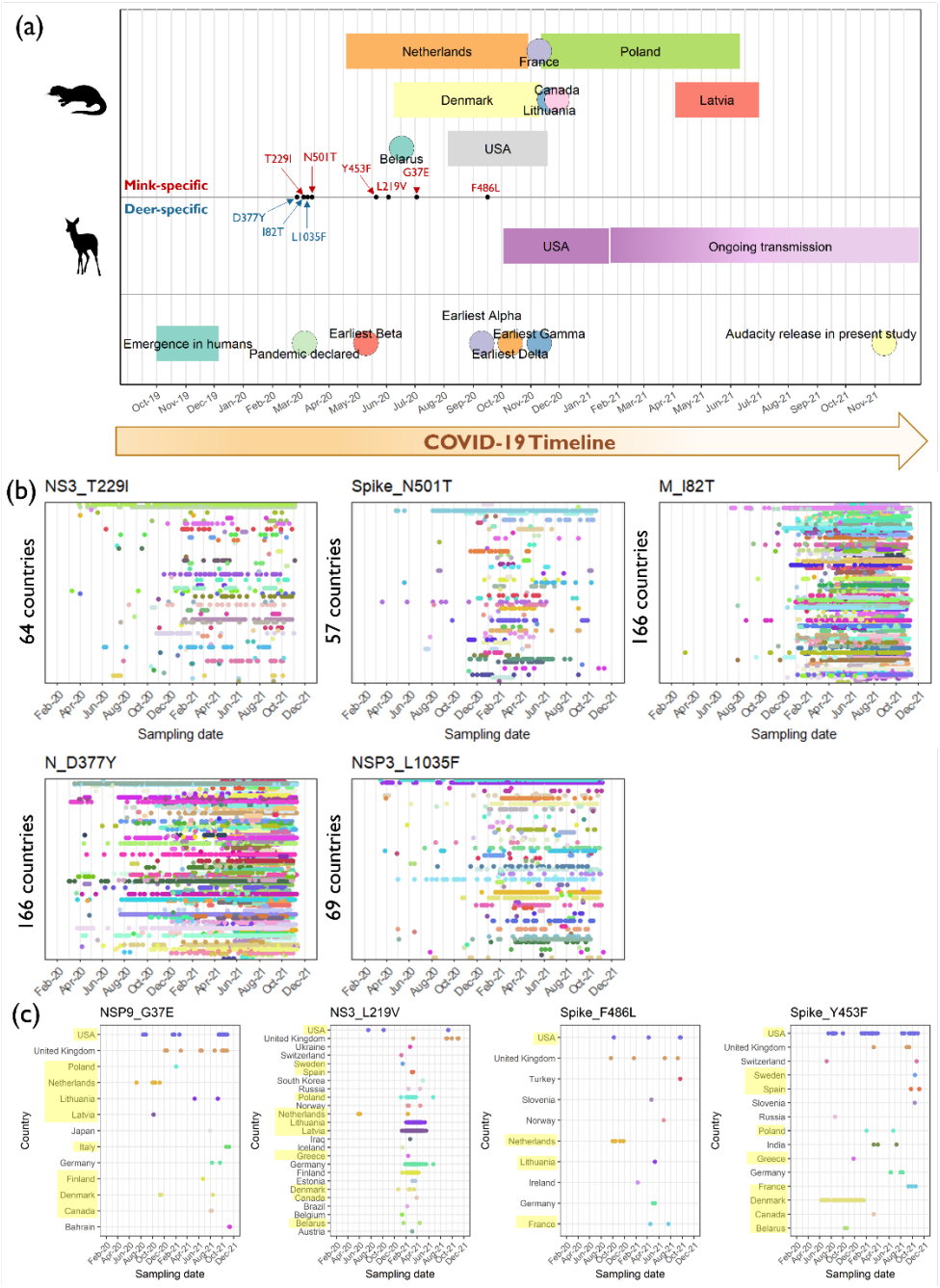


**Figure 3. Timeline of the COVID-19 pandemic.** (a) The key events of the pandemic from the estimated emergence of SARS-CoV-2 in humans (3) to the sampling dates of the first isolates for each VoC are annotated in the lowest panel. The coloured rectangles in the upper first and second panels indicate the range of sampling dates of animal-associated SARS-CoV-2 sequences in the different countries. The sampling dates of the earliest human isolates carrying each candidate mutation are annotated along the timeline are indicated by black points. Panels (b) and (c) show the temporal distributions of candidate mutations in human SARS-CoV-2 isolates collected prior to 16 November 2021. For panel (b), country names were omitted, and the number of countries where the candidate mutations were found in human isolates are annotated. For panel (c), countries where human-to-mink transmission has been documented are highlighted in yellow (46).

*Immediate changes to genomic composition in animal isolates*

To investigate changes to the genomic landscape of SARS-CoV-2 immediately following a host jump, we analysed the nucleotide-nucleotide transitions and dinucleotide frequencies of animal isolates relative to human background 1. The proportions of nucleotide-nucleotide transitions differed between mink ($\chi^2 = 182.5$, $p < 0.001$) and deer ($\chi^2 = 48.1$, $p < 0.001$) relative to those in human isolates (**Figure 3a**). However, the overall mutational profiles are similar with C→U transitions dominating. Consistently, a principal components analysis of dinucleotide frequencies shows highly overlapping host clusters, indicating that the genome composition of SARS-CoV-2 infecting different hosts do not differ considerably (**Figure 3b**). Of note, A→G transitions appear to occur less frequently in mink than humans (permutation test, $p = 0.057$), though this change is subtle compared to the overrepresentation of C→U mutations (**Figure 3a**). Direct comparisons between mink and deer, or between the two human backgrounds could not be made due to imbalanced representation of PANGO lineages.

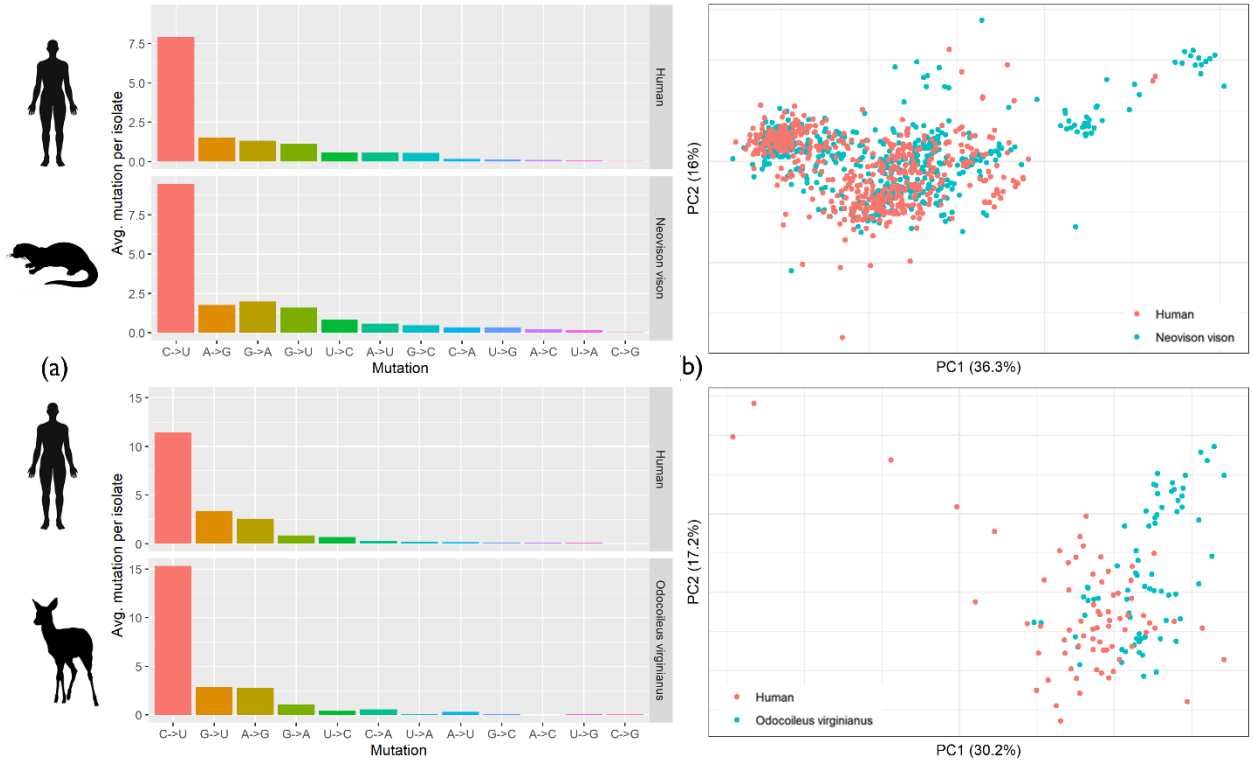


**Figure 3. Host-specific genomic landscapes.** (a) Nucleotide-nucleotide transition frequencies (x-axis) against average mutations observed per isolate in human and animal hosts (as indicated by symbols), and (b) principal components analysis of all dinucleotide frequencies, stratified by host.

*Spillovers into novel animal hosts did not lead to inflated substitution rates*

We attempted to tip-calibrate animal-human maximum-likelihood phylogenies, comprising either mink or deer isolates with their corresponding human backgrounds (background 1). Root-to-tip regressions for isolates from each country suggest that only isolates from Denmark, Latvia, Netherlands, and Poland had sufficient temporal signal in the data to reliably calibrate a time tree ($r^2$ = 0.31-0.83). Tip-calibration of a phylogeny comprising all mink and human background 1 isolates from these countries estimated the time to most recent common ancestor (tMRCA) to be 22nd November 2019 (90% maximum posterior interval: 5th October 2019-15th January 2020), and the substitution rate to be (5.54 ± 0.4 s.d.) x $10^{-4}$ substitutions/site/year. These estimates are highly consistent with previous estimates (3), suggesting that our reconstructed time-scaled phylogenies are reliable. To determine host-specific rate variation, we estimated mink-human time trees for each country and then visualised the terminal branch lengths of isolates corresponding to each host (**Figure 4**). With the exception of Denmark, the substitution rates of SARS-CoV-2 in humans appear to systematically exceed those in minks.

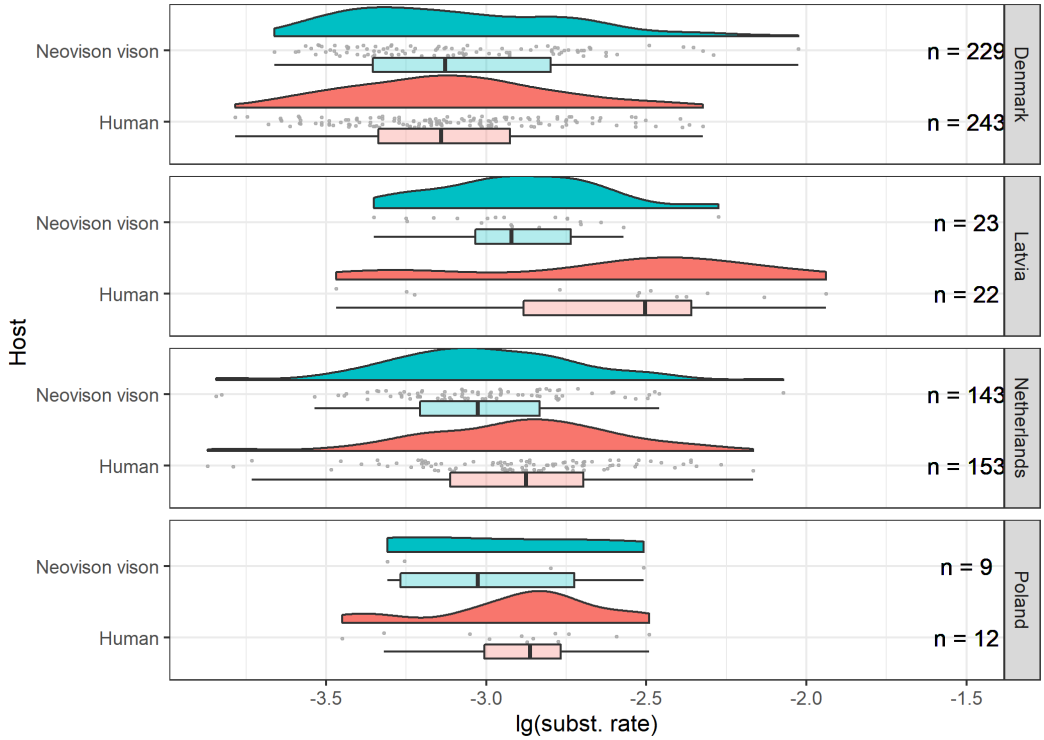


**Figure 4. Host-specific substitution rate variation.** Raincloud plots (47) of terminal branch lengths stratified by host, comprising Gaussian kernel probability density, scatter and box-and-whisker plots. Multiple mink-human maximum-likelihood phylogenies of mink and human background 1 isolates were reconstructed and used for tip-calibration. Isolates that did not have complete dates or that were duplicate sequences were removed prior to analysis. The final number of isolates in each stratum that were used for tip-calibration are annotated.

*Predicted impact of candidate host-adaptive mutations on viral proteins*

We attempted to bioinformatically assess the impact of non-synonymous candidate mutations on protein function using *PROVEAN* scores (48) and their putative impact on viral fitness in a novel host using structural analyses. *PROVEAN* scores have been shown to correlate with how deleterious a mutation is to protein function (48). They are computed based on the BLOSUM62 substitution matrix (49) whose substitution scores loosely reflect how biochemically conservative a mutation is (50), with positive scores implying more conservative mutations. Interestingly, most of the strong candidate mutations analysed are predicted to be conservative and functionally neutral (**Table 2**), including all mink-associated candidate mutations in the Spike protein. This is also the case for human-adaptive mutations implicated in immune escape and that have emerged recurrently in more transmissible viral lineages (16), suggesting that adaptive mutations, at least in the Spike protein, may not necessarily strongly impact protein function as assessed by these metrics.

**Table 2. *PROVEAN* scores of strong candidate mutations adaptive to human and non-human hosts.** Mutations that were predicted by *PROVEAN* to be deleterious to protein function are in red.

| **Protein** | **Mutation** | **Potentially adaptive to** | **Reference** | ***PROVEAN* score** |
|---|---|---|---|---|
| ORF1ab | NSP9_G37E | mink | present study | -5.108 |
| | NSP3_L1035F | deer | | -0.808 |
| Spike | N501T | mink | | 0.746 |
| | F486L | | | -0.035 |
| | Y453F | | (26,27), present study | -0.393 |
| | N501Y | Rodents, Humans | (16,51) | -0.090 |
| | L452R | Humans (16) | (16) | 0.559 |
| | E484K | | | 0.128 |
| | D614G | | | 0.598 |
| | P681R | | | 0.741 |
| ORF3a | L219V | mink | present study | 0.276 |
| | T229I | | | -4.276 |

Further, since N501T, F486L and Y453F fall within the SARS-CoV-2 Spike RBD, we considered their role in receptor binding affinity as putative sites of adaptation to a mink host. The Spike RBD (codon positions 319-541 (52)) provides a critical region for SARS-CoV-2 to attach to host cells via docking to ACE2 receptors, thereby allowing subsequent SARS-CoV-2 entry into host cells and eventual replication (53,54). Specific residues within the RBD have been identified as critical for receptor binding (12,55,56), with potential to modulate both infectivity and antigenicity (52). All three candidate Spike mutations (Y453F, F486L and N501T) suggested by the evolutionary analyses are in residues directly involved in contacts in the Spike: ACE2 interface and are therefore relevant to the binding affinity and stability of the complex (**Figure S2**).

We modelled various structures of mink or human ACE2 bound to the wild-type (WT, i.e., Wuhan-Hu-1 reference sequence) Spike protein, or to mutant Spike proteins carrying either N501T, F486L or Y453F. We then used the protein docking prediction protocols *HADDOCK* (57) and *mCSM-PPI2* (58) to analyse the change in stability of the Spike:ACE2 complexes due to each of these mutations (See '**Methods**'). We used this approach as previous work showed that it gave results that correlated well with experimental data on susceptibility to infection (12). Interestingly, the stability predictions of both methods are somewhat conflicting, and indicate marginal changes in the stability of the complex (**Figure 5**). Further, candidate mutations are that are predicted to stabilise (or destabilise) the Spike: human-ACE2 complex are also predicted to stabilise (or destabilise) the Spike: mink-ACE2 complex (**Figure 5b**). Overall, the *PROVEAN* and protein docking analyses are consistent with the hypothesis that SARS-CoV-2 mutations tend to be conservative and any small changes to structure caused by the candidate Spike mutations do not significantly affect the stability of the Spike: ACE2 complex.

(a) **HADDOCK**

| Spike structure | Human ACE2 | Mink ACE2 |
|---|---|---|
| WT (Wuhan-Hu-1) | -137.0 | -152.5 |
| Y453F | -145.0 | -155.1 |
| F486L | -139.8 | -159.5 |
| N501T | -137.6 | -148.7 |

(b) **mCSM-PPI2**

| Substitution performed | Human ACE2 | Mink ACE2 |
|---|---|---|
| Y453F | -0.083 | -0.496 |
| F453Y | 0.359 | 0.759 |
| F486L | -0.936 | -0.504 |
| L486F | 0.401 | 0.224 |
| N501T | 0.882 | 1.223 |
| T501N | -0.674 | -0.812 |

**Figure 5. Predicted effects of candidate mutations.** (a) *HADDOCK* predicted energies for the Spike: ACE2 complexes. More negative values relative to the WT-Spike: ACE2 complexes (highlighted in grey) indicate stronger binding energy of the complex. (b) *mCSM-PPI2* predicted changes in complex stabilities. Negative DDG values are associated with destabilisation of the complex following mutation of the residue and positive values with stabilisation of the complex. Values in green and red indicate predicted increases or decreases in complex stability respectively.

## Discussion

Coronaviruses have placed an enormous burden on public health globally in recent years, including four endemic (human coronavirus HKU1, OC43, 229E and NL63), two epidemic (SARS, MERS), and most recently one pandemic species (SARS-CoV-2). There is no doubt that novel coronaviruses will continue to emerge in humans. Therefore, understanding the cross-species transmission of SARS-CoV-2 and associated host adaptation is highly relevant to outbreak mitigation and future prevention. In this work, we analysed published and publicly available SARS-CoV-2 sequences isolated from animals in comparison with carefully selected human-associated sequences to understand the evolutionary events surrounding a host jump event.

Secondary host jumps of SARS-CoV-2 into animals have been documented for a variety of species, including cats and dogs, tigers and lions in zoos, farmed mink and wild deer in the US. While in all cases, host range expansion arose through multiple independent spillover events, only those in mink and deer have led to the observation of extensive subsequent animal-to-animal transmission to date. Irrespective of the transmissibility potential of SARS-CoV-2, in different hosts, this is most likely due to companion animals and zoo animals having limited contact with congeners. While mink and deer spillovers were identified early, it is likely that SARS-CoV-2 has already established itself in other animal reservoirs that are less well-documented. For example, a recent study of wild mustelids found three wild martens (*Martes martes*) and two badgers (*Meles meles*) to be seropositive for SARS-CoV-2 (59). Given the virus' prevalence in the human population and its ability to infect a broad range of mammalian hosts, it may be surprising if the number of non-human reservoir species did not increase.

Our analysis of animal SARS-CoV-2 isolates points to differing patterns of onward transmission in different sampled animal systems. We focused on deer and mink associated viral lineages for which phylogenetic transmission clusters have been well sampled and documented. Our analyses, focusing on a set of criteria applied to recurrent mutations, identify putative signatures of host adaptation following onward transmission of SARS-CoV-2 in mink and deer. The impact of mutations on animal-to-animal transmission remains largely uncharacterised. Indeed, only one of our candidate mutations has previously been suggested to be animal adaptive. Phylodynamic analyses of Dutch mink farm outbreaks have previously shown that viruses in minks that carry the Spike_F486L mutation may evolve and transmit at a faster rate (60). Meanwhile, our functional prediction analyses using bioinformatic approaches suggest a minimal impact of all strong candidate Spike mutations, including Spike_F486L, on Spike: ACE2 interactions. Together, these findings highlight the obscure and complex relationships between mutations and viral fitness. Additionally, the absence of strong candidate deer-adaptive mutations in the Spike protein, together with the presence of strong candidates in ORF1ab and ORF3a highlight the likely importance of mutations in non-Spike proteins. Further experimental investigations, particularly on the relationships between mutations and viral fitness, are warranted.

White-tailed deer present as the best animal models for understanding the natural transmission of SARS-CoV-2 and constitute the first known animal reservoir of the virus, with locally high prevalence as documented by seropositivity of 30-40% (31,34,61). Moreover, white-tailed deer populations are large, interconnected and distributed over a wide geographic range, including most of North America, Central America and parts of South America. Given the difficulties encountered by most worldwide governments to control the transmission of SARS-CoV-2 in humans, any attempt to eradicate the virus in white-tailed deer would be highly challenging if even possible.

The culling of farmed minks in Denmark in late 2020, and the more recent speculation that Omicron may have evolved in rodents (62), highlight recurring concerns over the emergence and accumulation of mutations while circulating in novel animal hosts following human-to-animal spillover, subsequently leading to the back-jump of more transmissible viral lineages into humans. Our results indicate that the putatively animal-adaptive mutations, for instance in mink lineages, likely confer minimal or no evolutionary advantage in humans, and as a result have been maintained at low frequencies. Additionally, our results suggest that the mutations accumulated while circulating in minks and deer have not caused drastic changes to the genomic landscape of SARS-CoV-2, since the relative proportions of nucleotide-nucleotide transitions occurring and the genomic composition in animal isolates largely mirror those in humans. Instead, we find a similar overrepresentation of C→U mutations in both human and animal hosts. Additionally, the most abundant transitions after C→U are G→U, A→G and G→A. Some of these mutation types are consistent with systematic mutational pressures exerted by host-editing processes, involving APOBEC and ADAR proteins, and reactive oxygen species (C→U, A→G, and G→U, respectively) (63). Of note is the subtle depletion of A→G mutations in minks vis-à-vis humans, which may reflect the differing activity of host ADAR in these species, though this would need to be experimentally validated. Nevertheless, these findings hint that similar mutagenic pressures in humans, mink and deer, greatly overshadow those of host adaptation.

The current minimal levels of host-specific adaptation in mink and deer are reminiscent of our previous work early in the first wave of the COVID-19 pandemic, which failed to identify mutations in SARS-CoV-2 associated with increased transmissibility in humans (14). The emergence of more transmissible VoCs driving the subsequent pandemic waves, highlight the strong collective, likely epistatic, phenotypic effects of multiple mutations. As such, while our analyses have not identified analogical 'animal-VoCs', this does not preclude the potential for new, more transmissible lineages to emerge in animal reservoirs in the future.

We could not find any crucial, prerequisite mutations for the secondary spillover of SARS-CoV-2 into mink and deer and observed no inflation of the substitution rates relative to that in its primary human host. These findings confirm that not only does human SARS-CoV-2 have the ability to infect multiple host species (i.e., broad host range), but it is also well pre-adapted to circulation in

mink and deer despite significant ongoing adaptation to humans. This reinforces previous suggestions of SARS-CoV-2 as a 'generalist' virus (15). This 'generalist' property may stem, in part, from the use of ACE2 as the primary host receptor for viral entry since the sequence and structure of ACE2 is fairly conserved across a broad range of mammals (10,12). Other host pathways exploited by viral proteins, which determine transmission efficiency, may similarly be conserved. However, further experimental work identifying such host-viral interactions needs to be done.

A virus circulating in its natural host continues to evolve, indefinitely so, largely due to the pressure exerted by its host's immunity. Though, a faster rate of evolution may be expected soon after a successful jump into a novel host. By the time of sampling, human-associated SARS-CoV-2 lineages are still adapting to their human hosts, and their rate of evolution might still be inflated relative to their long-term future quasi-equilibrium. As such, the fact that we did not observe a higher rate of evolution of viral lineages circulating in mink and deer, should not necessarily be interpreted as an absence of selective pressure in its novel animal hosts, but rather as heightened selection on viruses circulating in humans not having yet relaxed.

We note several limitations of our present study. The phylogenetically distinct clusters that we manually curated do not necessarily correspond to discrete spillover events between an individual and a single animal. In fact, as demonstrated by the mink_Denmark_1 cluster, complex transmission patterns are difficult to disentangle solely based on sequence information alone. This is further exacerbated by the difficulty of identifying and sequencing every human or non-human host within any transmission chain (64). Transmission chain reconstruction (i.e., 'who-infected-whom') using *SeqTrack* (64) or *TransPhylo* (65) may provide a more reliable estimate of the number of individual spillover events, but is beyond the scope of our study. Separately, SARS-CoV-2 surveillance in animals early on in the pandemic was minimal or absent so we cannot rule out the possibility that some early animal outbreaks were left undetected, and that some animal-specific mutations may have been introduced into the global diversity of SARS-CoV-2 circulating in humans during this period. As such, our claim that the emergence of animal-adaptive mutations in humans largely predate human-to-animal transmission is restricted to documented spillover events. Additionally, our approach to identify putatively adaptive alleles may not be able to detect these animal-specific mutations as it relies on a comparison of animal-associated allele frequencies against that from a human background. For our bioinformatic functional analyses, the performance of PROVEAN on assessing the functional impact of mutations has not been specifically validated on viral sequence datasets, so it remains unclear whether the default score threshold can be used to reliably identify putatively 'deleterious' mutations. Additionally, while our PROVEAN and structural analyses attempt to assess the effects of mutations on protein structure and function, it is difficult to interpret whether these effects (or lack thereof) directly affect fitness and the mechanisms for doing so. Mutational studies *in vitro* or *in vivo* are key in elucidating such

mechanisms and may shed light on the broader strategies that SARS-CoV-2 employ to adapt for circulation in novel host species.

Overall, our findings indicate that the mutational load required for efficient SARS-CoV-2 transmission in novel hosts is low, highlighting the 'generalist' nature of SARS-CoV-2 as a mammalian pathogen. In light of this, human-to-animal and spillback events are both a realised and likely outcome of widespread SARS-CoV-2 transmission in human populations. The establishment of SARS-CoV-2 in animal reservoirs further challenge adoption of a suppression/elimination strategy to pandemic mitigation since back-spill to human populations, as seen in association with Danish and Dutch mink farms, seem to be inevitable. Our results indicate that putatively animal-adaptive mutations have emerged in the short time that SARS-CoV-2 was circulating in mink and deer, but that these mutations do not appear to confer a significant advantage for circulation in humans. Nevertheless, mutational surveillance of SARS-CoV-2 in human and animal populations remain important to document the adaptive potential of the virus and its consequences in human and animal hosts.

## Methods

*Data acquisition*

All animal SARS-CoV-2 isolates that were present in the 16 November 2021 release of the Audacity (UShER (66)) tree on GISAID (17,18) were retrieved (**Table 1**). Additionally, human accessions were subsampled from the Audacity tree based on various inclusion criteria depending on the analysis performed. The inclusion criteria used for each analysis are described in the '**Human backgrounds***'* section. The alignments of human and animal genomes (to WIV04; EPI_ISL_402124) corresponding to these accessions were then extracted from the Audacity masked multiple sequence alignment (8$^{th}$ December 2021) on GISAID using the *subseq* utility of *Seqtk* (https://github.com/lh3/seqtk).

*Maximum likelihood and Audacity phylogenies*

Maximum likelihood trees were inferred from the masked genomic sequence alignments using the *Augur* wrapper (67) for *IQ-Tree2* (68), specifying a GTR + Γ substitution model. All trees were either visualised using *Dendroscope 3* (69) or *ggtree* (70), and manipulated using the *Ape* package (71) in *R*. Where the number of isolates considered is large, we extracted subtrees from the Audacity tree for further analysis using the *drop.tip* function in the *R* package, *Ape v5.5* (71). This was to avoid the excessive computational overhead of phylogenetic reconstruction.

*Animal SARS-CoV-2 cluster annotation*

To place animal SARS-CoV-2 isolates in the context of human infections, we visualised a subsampled Audacity tree, representing the global genomic diversity of SARS-CoV-2 (**Figure 1a**). A total of 15,846 isolates, comprising 10 human SARS-CoV-2 isolates per country per lineage, in addition to all isolates shown in **Table 1** were included in this subsampled tree. Separately, we visually inspected a subsampled Audacity tree comprising animal isolates and all human isolates collected prior to the most recent animal isolates for either host (mink: 819,813 isolates, 6 July 2021; deer: 167,967 isolates, 28 January 2021). The accessions considered in these analyses are provided in **Table S2a-c**. This was to identify phylogenetically distinct clusters of animal isolates representing independent spillover events. Monophyletic clades of animal SARS-CoV-2 isolates that were assigned the same PANGO lineage (45) were initially designated as separate clusters. These preliminary clusters were manually inspected, and subsequently merged or separated based on their phylogenetic placement. Cluster information of all animal accessions included in this study is provided in **Table S3**.

*Identifying recurring mutations*

The maximum likelihood trees and corresponding alignments of SARS-CoV-2 isolates associated with a single host species (i.e., mink or deer) were screened for homoplasies using *HomoplasyFinder v0.0.0.9* (72). Homoplasies are mutations that have emerged recurrently and

independently throughout a taxon's evolutionary history and may be indicative of host adaptation. *HomoplasyFinder* employs the method first described by Fitch (73), providing, for each site, the site specific consistency index and the minimum number of independent emergences in the phylogenetic tree. All nucleotide positions with a consistency index <0.5 are considered homoplastic.

*Human backgrounds*

In our analyses, we compared mink or deer-associated SARS-CoV-2 isolates to different subsamples of human isolates. Selection of appropriate human backgrounds to identify patterns of host-specific adaptation is crucial to minimise the risk of artefactual results. Depending on the inclusion criteria of human isolates, the inferences that can be made differ greatly. In this study, the main human background (referred to as 'human background 1') comprises human isolates with countries of isolation, PANGO lineages, and range of sampling dates matching those for animal isolates. Additionally, human isolates that fulfilled these criteria were randomly subsampled to match the number of viral isolates per PANGO lineage in animals. This human background controls for biases in the relative sizes of SARS-CoV-2 lineages, genomic diversity, and sequencing effort. A second human background (referred to as 'background 2') comprising 10 human isolates for each PANGO lineage present within the countries of isolation, regardless of sampling date, was also used. This background allows us to compare animal-specific vis-à-vis human-specific adaptation of SARS-CoV-2 in a wider evolutionary context.

*Allele frequency and mutational biases*

Allele frequencies and nucleotide-nucleotide transitions (e.g. number of C→U mutations) were computed for all positions in the animal or human SARS-CoV-2 masked sequence alignment using the *base.freq* function from the *Ape* package in custom R scripts. We tested whether the frequency of nucleotide-nucleotide transitions in human and animal genomes differed using a Monte Carlo simulation of the $\chi 2$ statistic with fixed margins (2000 iterations) (74,75). This was implemented using the *chisq.test* function in R with the *simulate.p.value* flag. Dinucleotide frequencies were computed using the *dinucleotideFrequency* in the *Biostrings* (76) package in *R*. A permutation test for 1000 iterations was performed to determine if the average number of A→G transitions differed between human and mink-associated isolates. Briefly, for each iteration, we randomised the host labels of mink and human SARS-CoV-2 isolates and computed the change in log10-transformed ratio of the proportion of A→G transitions in animal to that for human isolates. We then calculated the *p*-value as the proportion of iterations where the computed metric was lesser than that observed without permutation. Separately, ordination of host-specific dinucleotide frequencies was performed via a principal components analysis with the *prcomp* function in *R*. Dinucleotide frequencies were zero-centred and scaled to unit variance prior to ordination. The accessions used for these analyses are provided in **Table S2d-g**.

*Estimating host-specific substitution rates*

Animal isolates, stratified by country, were analysed relative to human isolates from the same country and isolation time span. Phylogenies of human and animal SARS-CoV-2 isolates were informally assessed for temporal signal via linear regression of root-to-tip distances against time, using *TreeTime* (77). These phylogenies were then tip-calibrated using *TreeTime* under an uncorrelated relaxed clock model, with a normal prior on rate heterogeneity across branches. Additionally, tip-calibration was run using a Kingman coalescent tree prior with an effective population size estimated using a skyline (78). The terminal branch lengths, stratified by host, of the inferred divergence trees were divided by those of the time-scaled trees, to obtain estimates of the host-specific mean substitution rates in substitutions per site per year. To determine if rate estimates differed across the SARS-CoV-2 genome, separate phylogenies were inferred and tip-calibrated using different partitions of the alignments, namely ORF1ab, Spike, non-Spike structural proteins, and accessory proteins. Isolates with ambiguous sampling dates were excluded from this analysis. Identical sequences were randomly removed using the *rmdup* utility of *SeqKit* (79). The final accessions used in these substitution rate analyses are provided in **Table S2h** and **i**.

*Predicting changes in the stability of viral proteins following mutation*

We used the *PROVEAN web server* (80) to bioinformatically assess the functional impact of candidate adaptive mutations on viral proteins. The *PROVEAN* score is an alignment-based metric that determines the change in sequence similarity of a protein given a single amino acid substitution, which was shown to correlate well with the functional impact of that mutation (48). *PROVEAN* scores that are less than -2.5 are classified as 'deleterious' mutations.

Additionally, we modelled various versions of the Spike: ACE2 complex to determine the change in stability of the Spike: ACE2 complex due to mutation. The structure of the wild-type (WT; i.e., Wuhan-Hu-1 reference sequence) SARS-CoV-2 Spike protein bound to human ACE2 has been solved at 2.45Å resolution (81) (PDB ID 6M0J). We visualised this structure using *PyMOL v2.4.1* (82). We used this as the template to model various structures of ACE2 bound to the SARS-CoV-2 Spike protein. In particular, we modelled structures of mink ACE2 bound to the WT Spike protein, and human or mink ACE2 bound to mutant Spike proteins carrying either of the candidate mutations Y453F, F486L or N501T. We generated query–template alignments using *HH-suite* (83) and predicted 3D models using *MODELLER v.9.24* (84). We used the 'very_slow' schedule for model refinement to optimise the geometry of the complex and interface. We generated 10 models for each Spike:ACE2 complex and selected the model with the lowest nDOPE (85) score, which reflects the quality of the model. Positive scores are likely to be poor models, while scores lower than -1 are likely to be native-like. The sequence similarity of the human ACE2 and the mink ACE2 is fairly high (83% amino acid sequence identity), and all generated models were of high quality (nDOPE < -1).

Following successful modelling of the various Spike: ACE2 complexes, two independent methods were used to assess changes to complex stability. The first, *HADDOCK* (57), is one of the top-performing protein-protein docking servers in the CAPRI competition (86). The *HADDOCK* score is a weighted sum of various predicted energy values (i.e., van der Waals, electrostatics and desolvation). We used the *HADDOCK v2.4* web server to score all complexes (**Figure 5a**). We then compared the scores of WT-Spike: human/mink-ACE2 to mutant-Spike: human/mink-ACE2 complexes. We also calculated the change in the stability of the Spike:ACE2 complexes using *mCSM-PPI2* (58) (**Figure 5b**). This program assigns a graph-based signature vector to each mutation, which is then used within machine learning models to predict the change in binding energy following an amino acid substitution. The signature vector is based upon atom-distance patterns in the protein, pharmacophore information and available experimental information, evolutionary information, and energetic terms. We used the *mCSM-PPI2* server (http://biosig.unimelb.edu.au/mcsm_ppi2/) for the simulations. In particular, we simulated the mutation of the WT Spike, and the back-mutation of mutant Spike to the WT (i.e., F453Y, L486F or T501N), while bound to human or mink ACE2. For *HADDOCK*, a value that is more negative than for the reference WT-Spike: ACE2 complex suggests stabilisation of the complex. Meanwhile for *mCSM-PPI-2,* negative and positive DDG values reflect destabilisation and stabilisation of the complex by the mutation, respectively. These two methods were used because we found in a previous study that the reported stability changes following mutations in the Spike: ACE2 complex correlated well with the available *in vivo* and *in vitro* experimental data on susceptibility to infection (12).

## Data availability

All data used in this study are publicly available on registration at GISAID (https://www.gisaid.org/). A full list of originating and submitting laboratories is provided in **Table S4**.

## Code availability

All custom code used to perform the analyses are hosted on GitHub (https://github.com/cednotsed/ditto.git). For all nucleotide transitions, the corresponding amino acid residue positions and changes were determined using an association table generated using a custom *Python* 3.7.11 script hosted on GitHub (https://github.com/cednotsed/SARS-CoV-2-hookup).

## Ethics declarations

The authors declare no competing interests.

## References


1. Abdel-Moneim AS, Abdelwhab EM. Evidence for SARS-CoV-2 infection of animal hosts. Pathogens. 2020;9(7):529.
2. Chan JF-W, To KK-W, Tse H, Jin D-Y, Yuen K-Y. Interspecies transmission and emergence of novel viruses: lessons from bats and birds. Trends Microbiol [Internet]. 2013;21(10):544–55. Available from: https://www.sciencedirect.com/science/article/pii/S0966842X13000899
3. van Dorp L, Acman M, Richard D, Shaw LP, Ford CE, Ormond L, et al. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. Infect Genet Evol [Internet]. 2020;104351. Available from: http://www.sciencedirect.com/science/article/pii/S1567134820301829
4. Boni MF, Lemey P, Jiang X, Lam TT-Y, Perry B, Castoe T, et al. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nat Microbiol. 2020;5(11):1408–17.
5. MacLean OA, Lytras S, Weaver S, Singer JB, Boni MF, Lemey P, et al. Natural selection in the evolution of SARS-CoV-2 in bats created a generalist virus and highly capable human pathogen. PLoS Biol. 2021;19(3):e3001115.
6. Ma Q, Liu J, Liu Q, Kang L, Liu R, Jing W, et al. Global Percentage of Asymptomatic SARS-CoV-2 Infections Among the Tested Population and Individuals With Confirmed COVID-19 Diagnosis: A Systematic Review and Meta-analysis. JAMA Netw Open. 2021;4(12):e2137257–e2137257.
7. Chu H, Chan JF-W, Yuen TT-T, Shuai H, Yuan S, Wang Y, et al. Comparative tropism, replication kinetics, and cell damage profiling of SARS-CoV-2 and SARS-CoV with implications for clinical manifestations, transmissibility, and laboratory studies of COVID-19: an observational study. The Lancet Microbe [Internet]. 2020;1(1):e14–23. Available from: https://www.sciencedirect.com/science/article/pii/S2666524720300045
8. Schlottau K, Rissmann M, Graaf A, Schön J, Sehl J, Wylezich C, et al. SARS-CoV-2 in fruit bats, ferrets, pigs, and chickens: an experimental transmission study. The Lancet Microbe [Internet]. 2020;1(5):e218–25. Available from: https://www.sciencedirect.com/science/article/pii/S2666524720300896
9. Shi J, Wen Z, Zhong G, Yang H, Wang C, Huang B, et al. Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS–coronavirus 2. Science (80- ). 2020;368(6494):1016–20.
10. Damas J, Hughes GM, Keough KC, Painter CA, Persky NS, Corbo M, et al. Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. Proc Natl Acad Sci U S A. 2020 Sep;117(36):22311–22.
11. Luan J, Lu Y, Jin X, Zhang L. Spike protein recognition of mammalian ACE2 predicts the host range and an optimized ACE2 for SARS-CoV-2 infection. Biochem Biophys Res Commun. 2020;526(1):165–9.
12. Lam SD, Bordin N, Waman VP, Scholes HM, Ashford P, Sen N, et al. SARS-CoV-2 spike protein predicted to form complexes with host receptor protein orthologues from a broad range of mammals. Sci Rep [Internet]. 2020;10(1):16471. Available from: https://doi.org/10.1038/s41598-020-71936-5
13. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. J Travel Med. 2020;

14. van Dorp L, Richard D, Tan CCS, Shaw LP, Acman M, Balloux F. No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. Nat Commun [Internet]. 2020;11(1):5986. Available from: https://doi.org/10.1038/s41467-020-19818-2
15. MacLean OA, Lytras S, Weaver S, Singer JB, Boni MF, Lemey P, et al. Natural selection in the evolution of SARS-CoV-2 in bats, not humans, created a highly capable human pathogen. bioRxiv. 2020 Jul;2020.05.28.122366.
16. Peacock TP, Penrice-Randal R, Hiscox JA, Barclay WS. SARS-CoV-2 one year on: evidence for ongoing viral adaptation. J Gen Virol. 2021;102(4):1584.
17. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data–from vision to reality. Eurosurveillance. 2017;22(13):30494.
18. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Challenges. 2017;1(1):33–46.
19. Oreshkova N, Molenaar RJ, Vreman S, Harders F, Munnink BBO, Hakze-van Der Honing RW, et al. SARS-CoV-2 infection in farmed minks, the Netherlands, April and May 2020. Eurosurveillance. 2020;25(23):2001005.
20. Hammer AS, Quaade ML, Rasmussen TB, Fonager J, Rasmussen M, Mundbjerg K, et al. SARS-CoV-2 transmission between mink (Neovison vison) and humans, Denmark. Emerg Infect Dis. 2021;27(2):547.
21. Larsen HD, Fonager J, Lomholt FK, Dalby T, Benedetti G, Kristensen B, et al. Preliminary report of an outbreak of SARS-CoV-2 in mink and mink farmers associated with community spread, Denmark, June to November 2020. Eurosurveillance. 2021;26(5):2100009.
22. Boklund A, Hammer AS, Quaade ML, Rasmussen TB, Lohse L, Strandbygaard B, et al. SARS-CoV-2 in Danish mink farms: course of the epidemic and a descriptive analysis of the outbreaks in 2020. Animals. 2021;11(1):164.
23. Munnink BBO, Sikkema RS, Nieuwenhuijse DF, Molenaar RJ, Munger E, Molenkamp R, et al. Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. Science (80- ). 2021;371(6525):172–7.
24. Enserink M. Coronavirus rips through Dutch mink farms, triggering culls. American Association for the Advancement of Science; 2020.
25. Lassaunière R, Fonager J, Rasmussen M, Frische A, Polacek Strandh C, Rasmussen TB, et al. SARS-CoV-2 spike mutations arising in Danish mink and their spread to humans. Statens Serum Inst. 2021;
26. Ren W, Lan J, Ju X, Gong M, Long Q, Zhu Z, et al. Mutation Y453F in the spike protein of SARS-CoV-2 enhances interaction with the mink ACE2 receptor for host adaption. PLoS Pathog. 2021;17(11):e1010053.
27. Bayarri-Olmos R, Rosbjerg A, Johnsen LB, Helgstrand C, Bak-Thomsen T, Garred P, et al. The SARS-CoV-2 Y453F mink variant displays a pronounced increase in ACE-2 affinity but does not challenge antibody neutralization. J Biol Chem. 2021;296.
28. Denmark M of E and F of. COVID-19: All mink in Denmark Must Be culled. 2020;
29. Shriner SA, Ellis JW, Root JJ, Roug A, Stopak SR, Wiscomb GW, et al. SARS-CoV-2 exposure in escaped mink, Utah, USA. Emerg Infect Dis. 2021;27(3):988.
30. Rabalski L, Kosinski M, Smura T, Aaltonen K, Kant R, Sironen T, et al. Severe Acute Respiratory Syndrome Coronavirus 2 in Farmed Mink (Neovison vison), Poland. Emerg Infect Dis. 2021;27(9):2333.
31. Hale VL, Dennis PM, McBride DS, Nolting JM, Madden C, Huey D, et al. SARS-CoV-2 infection in free-ranging white-tailed deer. Nature. 2021;1–8.

32. Palmer M V, Martins M, Falkenberg S, Buckley A, Caserta LC, Mitchell PK, et al. Susceptibility of white-tailed deer (Odocoileus virginianus) to SARS-CoV-2. J Virol. 2021;95(11):e00083-21.
33. Cool K, Gaudreault NN, Morozov I, Trujillo JD, Meekins DA, McDowell C, et al. Infection and transmission of ancestral SARS-CoV-2 and its alpha variant in pregnant white-tailed deer. Emerg Microbes Infect. 2021;(just-accepted):1–39.
34. Chandler JC, Bevins SN, Ellis JW, Linder TJ, Tell RM, Jenkins-Moore M, et al. SARS-CoV-2 exposure in wild white-tailed deer (Odocoileus virginianus). Proc Natl Acad Sci. 2021;118(47).
35. Deng J, Liu Y, Sun C, Bai J, Sun J, Hao L, et al. SARS-CoV-2 Serological Survey of Cats in China before and after the Pandemic. Virol Sin. 2020;35(6):846–8.
36. Zhang Q, Zhang H, Gao J, Huang K, Yang Y, Hui X, et al. A serological survey of SARS-CoV-2 in cat in Wuhan. Emerg Microbes Infect. 2020;9(1):2013–9.
37. Villanueva-Saz S, Giner J, Tobajas AP, Pérez MD, González-Ramírez AM, Macías-León J, et al. Serological evidence of SARS-CoV-2 and co-infections in stray cats in Spain. Transbound Emerg Dis. 2021;
38. Patterson EI, Elia G, Grassi A, Giordano A, Desario C, Medardo M, et al. Evidence of exposure to SARS-CoV-2 in cats and dogs from households in Italy. Nat Commun [Internet]. 2020;11(1):6231. Available from: https://doi.org/10.1038/s41467-020-20097-0
39. Bosco-Lauth AM, Hartwig AE, Porter SM, Gordy PW, Nehring M, Byas AD, et al. Experimental infection of domestic dogs and cats with SARS-CoV-2: Pathogenesis, transmission, and response to reexposure in cats. Proc Natl Acad Sci. 2020;117(42):26382–8.
40. McAloose D, Laverack M, Wang L, Killian ML, Caserta LC, Yuan F, et al. From people to Panthera: Natural SARS-CoV-2 infection in tigers and lions at the Bronx Zoo. MBio. 2020;11(5):e02220-20.
41. Bartlett SL, Diel DG, Wang L, Zec S, Laverack M, Martins M, et al. SARS-CoV-2 infection and longitudinal fecal screening in Malayan tigers (Panthera tigris jacksoni), Amur tigers (Panthera tigris altaica), and African lions (Panthera leo krugeri) at the Bronx Zoo, New York, USA. J Zoo Wildl Med. 2021;51(4):733–44.
42. Kuchipudi S V, Surendran-Nair M, Ruden RM, Yon M, Nissly RH, Vandegrift KJ, et al. Multiple spillovers from humans and onward transmission of SARS-CoV-2 in white-tailed deer. Proc Natl Acad Sci [Internet]. 2022 Feb 8;119(6):e2121644119. Available from: http://www.pnas.org/content/119/6/e2121644119.abstract
43. Mishra A, Kumar N, Bhatia S, Aasdev A, Kanniappan S, Sekhar AT, et al. Sars-cov-2 delta variant among asiatic lions, india. Emerg Infect Dis. 2021;27(10):2723.
44. Fernández-Bellon H, Rodon J, Fernández-Bastit L, Almagro V, Padilla-Solé P, Lorca-Oró C, et al. Monitoring natural SARS-CoV-2 infection in lions (Panthera leo) at the Barcelona Zoo: Viral dynamics and host responses. Viruses. 2021;13(9):1683.
45. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. 2020;5(11):1403–7.
46. Control EFSA and EC for DP and, Boklund A, Gortázar C, Pasquali P, Roberts H, Nielsen SS, et al. Monitoring of SARS-CoV-2 infection in mustelids. EFSA J. 2021;19(3):e06459.
47. Allen M, Poggiali D, Whitaker K, Marshall TR, Kievit RA. Raincloud plots: a multi-platform tool for robust data visualization. Wellcome open Res. 2019;4.

48. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the Functional Effect of Amino Acid Substitutions and Indels. PLoS One. 2012;7(10).
49. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci. 1992;89(22):10915–9.
50. Pearson WR. Selecting the right similarity-scoring matrix. Curr Protoc Bioinforma. 2013;43(1):3–5.
51. Shuai H, Chan JF-W, Yuen TT-T, Yoon C, Hu J-C, Wen L, et al. Emerging SARS-CoV-2 variants expand species tropism to murines. EBioMedicine. 2021;73:103643.
52. Li Q, Wu J, Nie J, Li X, Huang W, Correspondence YW, et al. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. 2020;
53. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020 Apr;181(2):271-280.e8.
54. Letko M, Marzi A, Munster V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat Microbiol. 2020 Apr;5(4):562–9.
55. Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, et al. Structural basis of receptor recognition by SARS-CoV-2. Nature. 2020 May;581(7807):221–4.
56. Brielle ES, Schneidman-Duhovny D, Linial M. The SARS-CoV-2 exerts a distinctive strategy for interacting with the ACE2 human receptor. bioRxiv. 2020 Mar;2020.03.10.986398.
57. van Zundert GCP, Rodrigues JPGLM, Trellet M, Schmitz C, Kastritis PL, Karaca E, et al. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J Mol Biol. 2016 Feb;428(4):720–5.
58. Rodrigues CHM, Myung Y, Pires DE V, Ascher DB. mCSM-PPI2: predicting the effects of mutations on protein–protein interactions. Nucleic Acids Res. 2019 Jul;47(W1):W338–44.
59. Davoust B, Guérin P, Orain N, Fligny C, Flirden F, Fenollar F, et al. Evidence of antibodies against SARS-CoV-2 in wild mustelids from Brittany (France). bioRxiv [Internet]. 2022 Jan 1;2022.01.20.477038. Available from: http://biorxiv.org/content/early/2022/01/20/2022.01.20.477038.abstract
60. Lu L, Sikkema RS, Velkers FC, Nieuwenhuijse DF, Fischer EAJ, Meijer PA, et al. Adaptation, spread and transmission of SARS-CoV-2 in farmed minks and associated humans in the Netherlands. Nat Commun. 2021;12(1):1–12.
61. Kuchipudi S V, Surendran-Nair M, Ruden RM, Yon M, Nissly RH, Nelli RK, et al. Multiple spillovers and onward transmission of SARS-CoV-2 in free-living and captive white-tailed deer. bioRxiv. 2021;
62. Wei C, Shan K-J, Wang W, Zhang S, Huan Q, Qian W. Evidence for a mouse origin of the SARS-CoV-2 Omicron variant. J Genet Genomics. 2021;
63. Mourier T, Sadykov M, Carr MJ, Gonzalez G, Hall WW, Pain A. Host-directed editing of the SARS-CoV-2 genome. Biochem Biophys Res Commun. 2021;538:35–9.
64. Jombart T, Eggo RM, Dodd PJ, Balloux F. Reconstructing disease outbreaks from genetic data: a graph approach. Heredity (Edinb). 2011;106(2):383–90.
65. Didelot X, Fraser C, Gardy J, Colijn C. Genomic infectious disease epidemiology in partially sampled and ongoing outbreaks. Mol Biol Evol. 2017;34(4):997–1007.
66. Turakhia Y, Thornlow B, Hinrichs AS, De Maio N, Gozashti L, Lanfear R, et al. Ultrafast Sample placement on Existing tRees (UShER) enables real-time phylogenetics for the

SARS-CoV-2 pandemic. Nat Genet [Internet]. 2021;53(6):809–16. Available from: https://doi.org/10.1038/s41588-021-00862-7
67. Huddleston J, Hadfield J, Sibley TR, Lee J, Fay K, Ilcisin M, et al. Augur: a bioinformatics toolkit for phylogenetic analyses of human pathogens. J Open Source Softw. 2021;6(57):2906.
68. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol [Internet]. 2020 May 1;37(5):1530–4. Available from: https://doi.org/10.1093/molbev/msaa015
69. Huson DH, Scornavacca C. Dendroscope 3: An Interactive Tool for Rooted Phylogenetic Trees and Networks. Syst Biol [Internet]. 2012 Dec 1;61(6):1061–7. Available from: https://doi.org/10.1093/sysbio/sys062
70. Yu G, Smith DK, Zhu H, Guan Y, Lam TT. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017;8(1):28–36.
71. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2019;35(3):526–8.
72. Manni F, Rotola A, Caselli E, Bertorelle G, Luca D Di. Detecting Recombination in TT Virus: A Phylogenetic Approach . J Mol Evol [Internet]. 2002;55(5):563–72. Available from: https://doi.org/10.1007/s00239-002-2352-y
73. Fitch WM. Toward defining the course of evolution: minimum change for a specific tree topology. Syst Biol. 1971;20(4):406–16.
74. Hope ACA. A Simplified Monte Carlo Significance Test Procedure. J R Stat Soc Ser B. 1968 Sep;30(3):582–98.
75. Patefield WM. Algorithm AS 159: An Efficient Method of Generating Random R × C Tables with Given Row and Column Totals. Appl Stat. 1981;30(1):91.
76. Pages H, Aboyoun P, Gentleman R, DebRoy S, Pages MH, IRanges L, et al. Package 'Biostrings.' 2013;
77. Sagulenko P, Puller V, Neher RA. TreeTime: Maximum-likelihood phylodynamic analysis. Virus Evol [Internet]. 2018 Jan 1;4(1):vex042. Available from: https://doi.org/10.1093/ve/vex042
78. Strimmer K, Pybus OG. Exploring the Demographic History of DNA Sequences Using the Generalized Skyline Plot. Mol Biol Evol [Internet]. 2001 Dec 1;18(12):2298–305. Available from: https://doi.org/10.1093/oxfordjournals.molbev.a003776
79. Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One [Internet]. 2016 Oct 5;11(10):e0163962. Available from: https://doi.org/10.1371/journal.pone.0163962
80. Choi Y, Chan AP. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics. 2015;31(16):2745–7.
81. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature. 2020 May;581(7807):215–20.
82. Version 2.4.1 Schrödinger L. The PyMOL Molecular Graphics System.
83. Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, Söding J. HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinformatics. 2019 Dec;20(1):473.
84. Webb B, Sali A. Comparative Protein Structure Modeling Using MODELLER. Curr Protoc

Bioinforma. 2016;54:5.6.1-5.6.37.
85. Shen M-Y, Sali A. Statistical potential for assessment and prediction of protein structures. Protein Sci. 2006 Nov;15(11):2507–24.
86. Koukos PI, Roel-Touris J, Ambrosetti F, Geng C, Schaarschmidt J, Trellet ME, et al. An overview of data-driven HADDOCK strategies in CAPRI rounds 38-45. Proteins Struct Funct Bioinforma. 2020 Aug;88(8):1029–36.