Transmission of SARS-CoV-2 in domestic cats imposes a narrow bottleneck

Katarina M. Braun[1]*, Gage K. Moreno[2]*, Peter J. Halfmann[1,3], Emma B. Hodcroft[4], David A. Baker[2], Emma C. Boehm[1], Andrea M. Weiler[1,5], Amelia K. Haj[2], Masato Hatta[1,3], Shiho Chiba[1,3], Tadashi Maemura[1,3], Yoshihiro Kawaoka[1,3], Katia Koelle[5], David H. O'Connor[2], Thomas C. Friedrich[1,5]

*These authors contributed equally

[1]Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, WI, United States of America

[2]Department of Pathology and Laboratory Medicine, University of Wisconsin-Madison, Madison, WI, United States of America

[3]Influenza Research Institute, School of Veterinary Sciences, University of Wisconsin-Madison, Madison, WI, United States

[4]Institute of Social and Preventative Medicine, University of Bern, Bern, Switzerland

[5]Wisconsin National Primate Research Center, University of Wisconsin-Madison, Madison, WI, United States of America

[6]Department of Biology, Emory University, Atlanta, GA, United States of America

Abstract

The evolutionary mechanisms by which SARS-CoV-2 viruses adapt to mammalian hosts and, potentially, undergo antigenic evolution depend on the ways genetic variation is generated and selected within and between individual hosts. Using domestic cats as a model, we show that SARS-CoV-2 consensus sequences remain largely unchanged over time within hosts, while dynamic sub-consensus diversity reveals processes of genetic drift and weak purifying selection. We further identify a notable variant at amino acid position 655 in Spike (H655Y), which was previously shown to confer escape from human monoclonal antibodies. This variant arises rapidly and persists at intermediate frequencies in index cats. It also becomes fixed following transmission in two of three pairs. These dynamics suggest this site may be under positive selection in this system and illustrate how a variant can quickly arise and become fixed in parallel across multiple transmission pairs. Transmission of SARS-CoV-2 in cats involved a narrow bottleneck, with new infections founded by fewer than ten viruses. In RNA virus evolution, stochastic processes like narrow transmission bottlenecks and genetic drift typically act to constrain the overall pace of adaptive evolution. Our data suggest that here, positive selection in index cats followed by a narrow transmission bottleneck may have instead accelerated the fixation of S H655Y, a potentially beneficial SARS-CoV-2 variant. Overall, our study suggests species- and context-specific adaptations are likely to continue to emerge. This underscores the importance of continued genomic surveillance for new SARS-CoV-2 variants as well as heightened scrutiny for signatures of SARS-CoV-2 positive selection in humans and mammalian model systems.

Author summary

Through ongoing human adaptation, spill-back events from other animal intermediates, or with the distribution of vaccines and therapeutics, the landscape of SARS-CoV-2 genetic variation is certain to change. The evolutionary mechanisms by which SARS-CoV-2 will continue to adapt to mammalian hosts depend on genetic variation generated within and between hosts. Here, using domestic cats as a model, we show that within-host SARS-CoV-2 genetic variation is predominantly influenced by genetic drift and purifying selection. Transmission of SARS-CoV-2 between hosts is defined by a narrow transmission bottleneck, involving 2-5 viruses. We further identify a notable variant at amino acid position 655 in Spike (H655Y), which arises rapidly and is transmitted in cats. Spike H655Y has been previously shown to confer escape from human monoclonal antibodies and is currently found in over 1000 human sequences. Overall, our study suggests species- and context-specific adaptations are likely to continue to emerge, underscoring the importance of continued genomic surveillance in humans and non-human mammalian hosts.

Introduction

Understanding the forces that shape genetic diversity of RNA viruses as they replicate within, and are transmitted between, hosts may aid in forecasting the future evolutionary trajectories of viruses on larger scales. The level and duration of protection provided by vaccines, therapeutics, and natural immunity against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) will depend in part on the amount of circulating viral variation and the rate at which adaptive mutations arise within hosts, are transmitted between hosts, and become widespread. Here, to model the evolutionary capacity of SARS-CoV-2 within and between

hosts, we characterize viral genetic diversity arising, persisting, and being transmitted in domestic cats.

A translational animal model can serve as a critical tool to study within- and between-host genetic variation of SARS-CoV-2 viruses. SARS-CoV-2 productively infects Syrian hamsters, rhesus macaques, cynomolgus macaques, ferrets, cats, and dogs in laboratory experiments. Natural infection with SARS-CoV-2 has also been documented in ferrets, mink, dogs, and small and large cats. This makes each of these potentially viable animal models, apart from large cats which are not typically used in biomedical research [1–5]. Among these species, natural transmission has only been observed in mink, cats, and ferrets [1, 6, 7]. Transmission from humans to mink and back to humans has also recently been documented [8]. Infectious virus has been recovered from various upper- and mid-respiratory tissues in cats and ferrets, including nasal turbinates, soft palate, tonsils, and trachea [1, 6]. However, only in cats has infectious virus been recovered from lung parenchyma, where infection is most commonly linked to severe disease in humans [1, 6, 9, 10].

Transmission bottlenecks, dramatic reductions in viral population size at the time of transmission, play an essential role in the overall pace of respiratory virus evolution [11–20]. For example, in humans airborne transmission of seasonal influenza viruses appears to involve a narrow transmission bottleneck, with new infections founded by as few as 1-2 genetically distinct viruses [12, 13, 16–18]. In the absence of selection acting during a transmission event, the likelihood of a variant being transmitted is equal to its frequency in the index host at the time of transmission (e.g. a variant at 5% frequency, has a 5% chance of being transmitted) [21]. When transmission involves the transfer of very few variants and selection is negligible, even beneficial variants present at low frequencies in the transmitting host are likely to be lost. Accordingly, although antigenic escape variants can sometimes be detected at very low levels

in individual human hosts, transmission of these variants has not been observed in nature [22, 22, 23]. In this way, narrow transmission bottlenecks are generally expected to slow the pace of seasonal influenza virus adaptation [11, 24] and may have similar effects on SARS-CoV-2.

Accurate estimates of the SARS-CoV-2 transmission bottleneck size will therefore aid in forecasting future viral evolution. Previous studies have reported discordant estimates of SARS-CoV-2 transmission bottleneck sizes in humans, ranging from “narrow” bottlenecks involving 1-8 virions to “wide” bottlenecks involving 100-1,000 virions [25–28]. However, studies of natural viral transmission in humans can be confounded by uncertainties regarding the timing of infection and directionality of transmission, and longitudinal samples that can help resolve such ambiguities are rarely available. Animal models overcome many of these uncertainties by providing access to longitudinal samples in well-defined index and contact infections with known timing.

Here we use a cat transmission model to show that SARS-CoV-2 genetic diversity is largely shaped by genetic drift and purifying selection, with the notable exception of a single variant in Spike at residue 655 (H655Y). These findings are in broad agreement with recent analyses of evolutionary forces acting on SARS-CoV-2 in humans, suggesting human SARS-CoV-2 isolates are relatively well-adapted to feline hosts [25–32]. While estimates of the size of the SARS-CoV-2 transmission bottleneck remain highly discordant in humans, we find very narrow transmission bottlenecks in cats, involving transmission of only 2-5 viruses. Our findings show cat models recapitulate key aspects of SARS-CoV-2 evolution in humans and we posit that the cat transmission model will be useful for investigating within- and between-host evolution of SARS-CoV-2 viruses.

Results

Within-host diversity of SARS-CoV-2 in cats is limited

Recently, members of our team inoculated three domestic cats with a second-passage SARS-CoV-2 human isolate from Tokyo (hCoV-19/Japan/UT-NCGM02/2020) [33]. Each index cat was co-housed with a naive contact cat beginning on day 1 post-inoculation (DPI). No new cat infections were performed for this study. Nasal swabs were collected daily up to 10 days post-inoculation, **Fig 1**. Viral RNA burden is plotted in **Supplementary Fig 1A** and infectious viral titers are shown in **Supplementary Fig 1B**.

Using conservative frequency thresholds previously established for tiled-amplicon sequencing, we called within-host variants (both intrahost single-nucleotide variants “iSNVs” and short insertions and deletions “indels”) throughout the genome against the inoculum SARS-CoV-2 reference (Genbank: MW219695.1) [34, 35]. Variants were required to be present in technical replicates at ≥3% and ≤97% of sequencing reads [36] (all within-host variants detected at >97% frequency were assumed to be fixed; see Methods for details). iSNVs were detected at least once at 38 different genome sites. Of the 38 unique variants, 14 are synonymous changes, 23 are nonsynonymous changes, and one occurs in an intergenic region; this distribution is broadly similar to recent reports of SARS-CoV-2 variation in infected humans [30]. Similarly, we detected indels occurring at 11 different genome sites across all animals and timepoints. We identified 6-19 distinct variants per cat, of which 4-7 were observed on two or more days over the course of the infection within each cat (**Supplementary Fig 2**). All variants (iSNVs and indels) are plotted by genome location and frequency in **Fig 2A.**

Genetic drift and purifying selection shape within-host diversity

To probe the evolutionary pressures shaping SARS-CoV-2 viruses within hosts, we first evaluated the proportion of variants shared between cats. Eighty-six percent of variants (34 of 38 iSNVs and 8 of 11 indels) were found in a single cat (42/49), 8% of variants were found in 2-5 cats (4/49), and the remaining 6% of variants were found in all 6 cats (3/49).

Purifying selection, which acts to purge deleterious mutations from a population, is known to result in an excess of low-frequency variants. In contrast, positive selection results in the accumulation of intermediate- and high-frequency variation [37]. Especially in the setting of an acute viral infection, exponential population growth is also expected to result in an excess of low-frequency variants [38]. To determine the type of evolutionary pressure acting on SARS-CoV-2 in cats, we plotted these distributions against a simple “neutral model” (light grey bars in **Fig 2B**), which assumes a constant population size and the absence of selection [37]. This model predicted that ~43% of polymorphisms would fall in the 3-10% frequency bin, ~25% into the 10-20% bin, ~14% into the 20-30% bin, ~10% into the 30-40% bin, and ~8% into the 40-50% bin. The frequency distribution of variants detected in each index cat across all available timepoints did not differ significantly from this “neutral” expectation ($p=0.265$, $p=0.052$, $p=0.160$, respectively; Mann Whitney U test).

Next we compared nonsynonymous (πN) and synonymous (πS) pairwise nucleotide diversity to further evaluate the evolutionary forces shaping viral populations in index and contact animals [39]. Broadly speaking, excess nonsynonymous polymorphism ($\pi N/\pi S > 1$) points toward diversifying or positive selection while excess synonymous polymorphism ($\pi N/\pi S < 1$) indicates purifying selection. When $\pi N / \pi S$ is approximately 1, genetic drift, i.e., stochastic changes in the frequency of viral genotypes over time, can be an important force shaping genetic diversity.

We observe that πS exceeds or is approximately equal to πN in most genes, although there is substantial variation among genes and cats (**Supplementary Table 1**). πS is significantly higher than πN in all 3 index cats in Spike (p=0.005, p=0.004, p=0.019, unpaired t-test) and ORF1ab (p=2.11e-05, p=1.84e-06, p=1.99e-06, unpaired t-test) and in index cats 2 and 3 in ORF8 (p=0.03, p=0.04, unpaired t-test). πS and πN are not significantly different in at least one index cat in ORF3a, envelope, and nucleocapsid. There was not enough genetic variation to measure nucleotide diversity in the remaining four genes (**Supplementary Table 1**). Taken together, these results suggest longitudinal genetic variation within feline hosts is principally shaped by genetic drift with purifying selection acting on individual genes, particularly ORF1ab and Spike.

Longitudinal sampling reveals few consensus-level changes within hosts

The consensus sequence recovered from all three index cats on the first day post-inoculation was identical to the inoculum or “stock” virus. This consensus sequence remained largely unchanged throughout infection in all index cats with the notable exception of two variants: H655Y in Spike (nucleotide site 23,525) and a synonymous change at amino acid position 67 in envelope (nucleotide site 26,445; S67S), which arose rapidly in all 3 index cats and rose to consensus levels (≥50% frequency) at various timepoints throughout infection in all index cats. Neither of these iSNVs was detected above 3% frequency in the inoculum, but when we mined all sequencing reads, S H655Y and E S67S could be detected at 0.85% and 0.34%, respectively. S H655Y was the consensus sequence on days 2-5 and days 7-8 in index cat 1, as well as on days 4 and 8 in index cat 2, and remained detectable above our 3% variant threshold throughout infection (**Fig 3**). Similarly, envelope S67S (E S67S) was the consensus

sequence on day 8 in index cat 1 and day 1 in index cat 2. S H655Y and E S67S were detectable on days 1-7 in cat 3 but stayed below consensus level.

Interestingly, S H655Y and E S67S became fixed together following transmission in two transmission pairs (contact cats 4 and 6) and were lost together during transmission to contact animal 5. In cat 5, however, two different variants in ORF1ab, G1756G and L3606F, became fixed after transmission. ORF1ab G1756G was not detected above 3% and L3606F was found at 17.2% in the day 5 sample from the index cat 2 (the cat transmitting to cat 5); it was not found in the inoculum at any detectable frequency. The categorical loss or fixation of these variants immediately following transmission, and in particular the fixation following transmission of a variant that was undetectable before, are highly suggestive of a narrow bottleneck [40].

In addition, a synonymous variant in an alanine codon at amino acid position 1,222 in Spike (nucleotide site 25,174) was found at >50% frequencies on days 4 and 8 in index cat 3, but was not detected above 3% on any other days. All iSNVs over time are shown in **Supplementary Fig 2** and all indels over time are shown in **Supplementary Fig 3**. These within-host analyses show that genetic drift appears to play a prominent role in shaping low-frequency genetic variation within hosts.

SARS-CoV-2 transmission in domestic cats is defined by a narrow transmission bottleneck

To estimate the size of SARS-CoV-2 transmission bottlenecks, we investigated the amount of genetic diversity lost following transmission in cats. We observed a reduction in the cumulative number of variants detected in each contact cat compared to its index: 7 fewer variants in cat 4 (n=9) compared to cat 1 (n=16), 9 fewer in cat 5 (n=10) than cat 2 (n=19), and 10 fewer in cat 6

(n=16) than cat 3 (n=6). Likewise, the frequency distribution of variants in all three contact cats following transmission differed from the distribution of variants in all three index cats prior to transmission (p-value=0.052, Mann Whitney U test). Following transmission, variant frequencies became more bimodally distributed than those observed in index cats, i.e., in contacts, most variants were either very low-frequency or fixed (**Supplementary Fig 2**).

To quantitatively investigate the stringency of each transmission event, we compared the genetic composition of viral populations immediately before and after viral transmission. We chose to use the first timepoint when infectious virus was recovered in the contact cat coupled with the timepoint immediately preceding this day in the index cat, as has been done previously [17]. We used days 2 (index) and 3 (contact) in pair 1, days 5 and 6 in pair 2, and days 4 and 5 in pair 3 (these sampling days are outlined in red in **Fig 1**). We applied the beta-binomial sampling method developed by Sobel-Leonard et al. to compare the shared set of variants (≥3%, ≤97%) in the pre/post-transmission timepoints for each pair [21]. Maximum-likelihood estimates determined that a mean effective bottleneck size of 5 (99% CI: 1-10), 3 (99% CI: 1-7), and 2 (99% CI: 1-3) best described each of the three cat transmission events evaluated here (**Fig 4**). This is in line with previous estimates for other respiratory viruses, including airborne transmission of seasonal influenza viruses in humans [40]. It is important to note, however, that the cat transmission pairs evaluated here shared physical enclosure spaces so the route of transmission could be airborne, direct contact, fomite, or a combination of these. Additionally, it has been shown that the route of influenza transmission can directly impact the size of the transmission bottleneck; for example in one study airborne transmission of influenza viruses resulted in a narrow bottleneck, whereas contact transmission resulted in a wider bottleneck [16].

Discussion

At the time of writing, the vast majority of humans remain immunologically naive to SARS-CoV-2. Whether through ongoing human adaptation, spill-back events from other animal intermediates, or with the distribution of vaccines and therapeutics, the landscape of SARS-CoV-2 variation is certain to change. Understanding the forces that shape genetic diversity of SARS-CoV-2 viruses within hosts will aid in forecasting the pace of genetic change as the virus faces shifting population-level immunity. Additionally, this baseline allows researchers to more easily identify a shift in the forces shaping within- and between-host diversity; for example, identification of signatures of positive selection might highlight rapidly-adapting, and therefore higher-risk, viruses.

Using domestic cats as a model system, we show stochastic processes like narrow transmission bottlenecks and genetic drift are major forces shaping SARS-CoV-2 genetic diversity within and between mammalian hosts. These stochastic forces typically act to constrain the overall pace of RNA virus evolution [12]. Despite this, we observe the rapid outgrowth of S H655Y in all three index cats, suggesting that this site may be under positive selection in this system. This variant achieved rapid fixation following transmission in two of three transmission pairs.

Our finding of narrow transmission bottlenecks is at odds with some recent studies in humans, which have estimated wide and variable SARS-CoV-2 transmission bottlenecks [25–28], but it is in line with other estimates suggesting that few SARS-CoV-2 viruses are transmitted between humans [25]. These discordant estimates are likely due to a combination of factors, including variable routes of transmission, uncertain sources of infection, difficulty collecting samples which closely bookend the transmission event, and inaccurate variant calls [25–28]. Human

studies have commonly identified transmission pairs using intrahousehold infections diagnosed within a defined timeframe. A major weakness with this approach is the possibility that some of these cohabiting individuals will share an alternative source of exposure. Furthermore, without fine-scale epidemiological and clinical metadata, pinpointing the time of likely transmission is challenging, so even samples collected before and after a real transmission event may be several days removed from the time of transmission. Here we were able to circumvent many of these challenges by taking advantage of domestic cats experimentally infected with SARS-CoV-2 arranged in defined transmission pairs with clinical monitoring and daily sample collection, making for a useful model system.

The size of the transmission bottleneck may have additional implications for individual infections. The total number of founding virions, or the inoculum dose, has been posited to play a role in coronavirus disease 2019 (COVID-19) clinical severity and outcomes [41, 42]. The transmission bottleneck can be parsed into two interdependent components: the population bottleneck, or the number of virus particles that found infection (similar to inoculation dose); and the genetic bottleneck, or the amount of viral diversity lost during transmission. For example, an infection founded by 1,000 genetically identical viruses would be categorized as resulting from a narrow genetic bottleneck (a single genotype initiates the infection) and a relatively large population bottleneck. The beta-binomial method used here measures the population bottleneck [21]. Our data are consistent with a narrow population bottleneck and therefore a low inoculum dose in these cats. The extent to which feline hosts experience symptoms when infected with SARS-CoV-2 is unclear, but the cats involved in this study remained afebrile throughout the study, did not lose body weight, and experienced no respiratory signs. Viral genetic diversity has been linked to pathogenesis and clinical outcomes in the context of other viruses (e.g., influenza A virus, polio, and respiratory syncytial virus) and because narrow transmission bottlenecks often reduce viral genetic diversity, bottlenecks may play an essential role in the outcome of

267 individual infections in this way as well [43–47]. The relationship between SARS-CoV-2 viral
268 genetic diversity and COVID-19 clinical severity remains unclear. Some have proposed a direct
269 relationship between particular viral lineages and COVID-19 severity [48], while others postulate
270 that host factors, like age and lymphocytopenia, are more likely to explain variable clinical
271 outcomes [49].

272 Although within-host diversity was limited in the cats evaluated here, we identify two notable
273 variants. S H655Y and E S67S were found at 0.85% and 0.34% in the stock, but were
274 preferentially amplified in all three index cats and were detectable at intermediate frequencies at
275 the first-day post-inoculation. Interestingly, S H655Y is not found in any of the 18 full-genome
276 domestic cat, tiger, and lion SARS-CoV-2 sequences available on GISAID (**Supplementary Fig**
277 **4**). S H655Y has, however, been reported in a variety of other settings, including transmission
278 studies in a hamster model, SARS-CoV-2 tissue culture experiments [50–53], and in a stock
279 virus passaged on Vero E6 cells [BioProject PRJNA645906, experiment numbers SRX9287152
280 (p1), SRX9287151 (p2), SRX9287154 (p3a); BioProject PRJNA627977]. S H655Y additionally
281 persisted in vivo in rhesus macaques challenged with one of these stock viruses [BioProject
282 PRJNA645906, experiment number SRX9287155]. As of 28 December, 2020, S H655Y has
283 been detected in 1,070 human SARS-CoV-2 viruses across 18 different countries in sequences
284 deposited in GISAID. The majority of these sequences come from the United Kingdom (n=886)
285 (**Supplementary Fig 5b and 5c**). It is important to note, however, that sampling of SARS-CoV-
286 2 sequences is heavily biased and sequences from the COVID-19 Genomics UK consortium
287 (COG-UK) are currently overrepresented in GISAID. Additionally, S H655Y is the 16th most
288 common variant detected in Spike among publicly-available SARS-CoV-2 sequences [54].
289 Sequences containing S H655Y variant are found in two distinct European clusters, EU1 and
290 EU2, suggesting it has arisen more than once (33269368, **Supplementary Fig 5a**).

Relatively little is known about the phenotypic impact of S H655Y in cats, humans, and other host species. Amino acid residue 655 is located near the polybasic cleavage site, residing between the receptor binding domain (RBD) and the fusion peptide, and therefore has been hypothesized to play a role in regulating Spike glycoprotein fusion efficiency [50, 51, 55]. In spite of its location outside of the RBD, S H655Y has been shown to arise on the background of a vesicular stomatitis virus (VSV) pseudotyped virus expressing various SARS-CoV-2 spike variants and confer escape from multiple monoclonal human antibodies in cell culture [50]. It is unlikely S H655Y represents a site of antibody escape in these cats because they were specific pathogen-free and had undetectable IgG antibody titers against SARS-CoV-2 Spike and Nucleocapsid proteins on the day of infection [33]. We did not do any experiments to elucidate the functional impact of this variant, but we speculate S H655Y could have improved Spike fusion efficiency and therefore host-cell entry in cats. It is possible S H655Y offers a similar advantage in human hosts and/or confers escape from some antibodies.

E S67S has not been documented elsewhere. Based on iSNV frequencies, S H655Y and E S67S appear to be in linkage with each other (see mirrored iSNV frequencies in cat 2 and cat 5 in **Fig 3** in particular), however with short sequence reads and sequencing approaches relying on amplicon PCR, we cannot rigorously assess the extent of linkage disequilibrium between these variants. It may be that S H655Y arose on the genetic background of an existing S67S variant in envelope. If S H655Y facilitates viral entry or replication in cats, viruses with this variant in linkage with E S67S might have been positively selected in all index cats.

Our data alone cannot resolve the precise mechanisms by which SARS-CoV-2 diversity is reduced during transmission, but the trajectories of S H655Y and E S67S raise some interesting possibilities. Although our sample size is small, the rise of S H655Y with E S67S in all index cats, and the fixation of these variants in 2 of 3 contact cats, suggest that selection for one or

both of these variants could have played a role in shaping genetic diversity recovered from contact cats. Viruses bearing these mutations could be preferentially amplified prior to, during, and/or after transmission.

If the transmission bottleneck is narrow and random, a variant's likelihood of being transmitted is equal to its frequency in the viral population at the time of transmission. If selection acts primarily within index hosts prior to transmission, S H655Y could have achieved a high enough frequency to be randomly drawn at the time of transmission. In this case, even a random, narrow transmission bottleneck could have facilitated the rapid fixation of a putatively beneficial variant. Next, suppose that viruses bearing S H655Y are shed more efficiently from index animals. In this case, evidence of selection in index animals would be limited and we would observe a small founding population in contact hosts where the beneficial variant is dominant. Alternatively, suppose viruses bearing S H655Y preferentially found infection in the recipient. In this case where selection is acting primarily in the contact host, transmission may involve transfer of a larger virus population after which beneficial variants may rapidly be swept to fixation. These scenarios are not mutually exclusive and it is possible selection to act in concert before, during, and after transmission. In any of these scenarios, we would observe a low-diversity virus population in contact animals in which the putatively beneficial variants had been enriched. Notably, S H655Y and E S67S are absent from contact cat 5 (pair 2), despite being detectable and even reaching consensus levels in the associated index animal. While these variants are lost during transmission in this pair, a variant in ORF1ab (Gly1756Gly), which was undetectable in index cat 2, became fixed in contact cat 5 following transmission. The dramatic shifts in iSNV frequency we observe in all 3 pairs are characteristic of a narrow transmission bottleneck [12]. Because narrow transmission bottlenecks can result in the loss of even beneficial variants, the fact that S H655Y and E S67S failed to be transmitted in pair 2 does not exclude the possibility that these variants enhance viral fitness. Altogether our data therefore

support the conclusion that SARS-CoV-2 transmission bottlenecks are narrow in this system, and may sometimes involve selection.

Large SARS-CoV-2 outbreaks in mink have been reported recently, some with “concerning” mutations that may evade human humoral immunity [56]. These mink outbreaks have resulted in the Danish authorities’ decision to cull 17 million mink as a safeguard against spill-back transmission into humans [56]. Similarly, the emergence of the B.1.1.7 SARS-CoV-2 lineage has brought to light the importance of detecting and characterizing novel variants which might confer increased transmissibility, infectiousness, clinical severity, or other phenotypic change. The precise origins of the defining B.1.1.7 variants are unknown. It has been speculated that it may have arisen from a chronically infected patient or through sub-curative doses of convalescent plasma [57]. While S H655Y has not been found in mink and is not one of the defining B.1.1.7 mutations, another one of the defining B.1.1.7 mutations, Spike N501Y, has emerged independently in mouse models [58]. This suggests that mammalian models can facilitate the detection of novel mutations and signatures of positive selection, which might highlight adaptive mutations. We observe one variant that arises early and is transmitted onward in cats, a potential reservoir and model species. Little has been specifically documented about this variant, but it was very interesting to note it confers escape from various human monoclonal antibodies and has been detected in more than 1,000 human viruses [50, 59]. Our study and the mink example show that species- and context-specific adaptations are likely as SARS-CoV-2 explores new hosts. Further investigation and ongoing surveillance for such variants is warranted. It is also important to prevent the reintroduction of such newly formed variants, of which we do not know the potential phenotypic impacts, by limiting the spread and evolution of SARS-CoV-2 in non-human reservoir species

As SARS-CoV-2 continues to spread globally, we must have models in place to recapitulate key evolutionary factors influencing SARS-CoV-2 transmission. With the imminent release of SARS-CoV-2 vaccines and therapeutics and increasing prevalence of natural exposure-related immunity, these models can help us forecast the future of SARS-CoV-2 variation and population-level genetic changes. Continued efforts to sequence SARS-CoV-2 across a wide variety of hosts, transmission routes, and spatiotemporal scales will be necessary to determine the evolutionary and epidemiological forces responsible for shaping within-host genetic diversity into global viral variation.

Methods

Nucleic acid extraction

For each sample, approximately 140 µL of viral transport medium was passed through a 0.22µm filter (Dot Scientific, Burton, MI, USA). Total nucleic acid was extracted using the Qiagen QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany), substituting carrier RNA with linear polyacrylamide (Invitrogen, Carlsbad, CA, USA) and eluting in 30 µL of nuclease-free H_2O.

Complementary DNA (cDNA) generation

Complementary DNA (cDNA) was synthesized using a modified ARTIC Network approach [34, 35]. Briefly, RNA was reverse transcribed with SuperScript IV Reverse Transcriptase (Invitrogen, Carlsbad, CA, USA) using random hexamers and dNTPs. Reaction conditions were as follows: 1µL of random hexamers and 1µL of dNTPs were added to 11 µL of sample RNA, heated to 65°C for 5 minutes, then cooled to 4°C for 1 minute. Then 7 µL of a master mix (4 µL

5x RT buffer,1 µL 0.1M DTT, 1µL RNaseOUT RNase Inhibitor, and 1 µL SSIV RT) was added and incubated at 42°C for 10 minutes, 70°C for 10 minutes, and then 4°C for 1 minute.

Multiplex PCR for SARS-CoV-2 genomes

A SARS-CoV-2-specific multiplex PCR for Nanopore sequencing was performed, similar to amplicon-based approaches as previously described [34, 35]. In short, primers for 96 overlapping amplicons spanning the entire genome with amplicon lengths of 500bp and overlapping by 75 to 100bp between the different amplicons were used to generate cDNA. Primers used in this manuscript were designed by ARTIC Network and are shown in **Supplementary Table 3**. cDNA (2.5 µL) was amplified in two multiplexed PCR reactions using Q5 Hot-Start DNA High-fidelity Polymerase (New England Biolabs, Ipswich, MA, USA) using the following cycling conditions; 98ºC for 30 seconds, followed by 25 cycles of 98ºC for 15 seconds and 65ºC for 5 minutes, followed by an indefinite hold at 4ºC [34, 35]. Following amplification, samples were pooled together before TrueSeq Illumina library prep.

TrueSeq Illumina library prep and sequencing

Amplified cDNA was purified using a 1:1 concentration of AMPure XP beads (Beckman Coulter, Brea, CA, USA) and eluted in 30µL of water. PCR products were quantified using Qubit dsDNA high-sensitivity kit (Invitrogen, USA) and were diluted to a final concentration of 2.5 ng/µl (150 ng in 50 µl volume). Each sample was then made compatible with deep sequencing using the Nextera TruSeq sample preparation kit (Illumina, USA). Specifically, each sample was enzymatically end repaired. Samples were purified using two consecutive AMPure bead cleanups (0.6x and 0.8x) and were quantified once more using Qubit dsDNA high-sensitivity kit (Invitrogen, USA). A non-templated nucleotide was attached to the 3′ ends of each sample,

406 followed by adaptor ligation. Samples were again purified using an AMPure bead cleanup (1x)
407 and eluted in 25µL of resuspension buffer. Lastly, samples were amplified using 8 PCR cycles,
408 cleaned with a 1:1 bead clean-up, and eluted in 30µL of RSB. The average sample fragment
409 length and purity was determined using the Agilent High Sensitivity DNA kit and the Agilent
410 2100 Bioanalyzer (Agilent, Santa Clara, CA). After passing quality control measures, samples
411 were pooled equimolarly to a final concentration of 4 nM, and 5 µl of each 4 nM pool was
412 denatured in 5 µl of 0.2 N NaOH for 5 min. Sequencing pools were denatured to a final
413 concentration of 10 pM with a PhiX-derived control library accounting for 1% of total DNA and
414 was loaded onto a 500-cycle v2 flow cell. Average quality metrics were recorded, reads were
415 demultiplexed, and FASTQ files were generated on Illumina's BaseSpace platform.

416 Processing of the raw sequence data, mapping, and variant 417 calling

418 Raw FASTQ files were analyzed using a workflow called "SARSquencer". Briefly, reads are
419 paired and merged using BBMerge (https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-
420 guide/bbmerge-guide/) and mapped to the reference (MW219695.1) using BBMap
421 (https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbmap-guide/). Mapped reads
422 were imported into Geneious (https://www.geneious.com/) for visual inspection. Variants were
423 called using callvariants.sh (contained within BBMap) and annotated using SnpEff
424 (https://pcingola.github.io/SnpEff/). The complete "SARSquencer" pipeline is available in the
425 GitHub accompanying this manuscript in `code/SARSquencer` as well as in a separate GitHub
426 repository – https://github.com/gagekmoreno/SARS_CoV-2_Zequencer. BBMap's output VCF
427 files were cleaned using custom Python scripts, which can be found in the GitHub
428 accompanying this manuscript

(https://github.com/katarinabraun/SARSCoV2_transmission_in_domestic_cats) [60]. Variants were called at ≥0.01% in reads that were ≥100 bp in length and supported by a minimum of 10 reads. Only variants at ≥3% frequency in both technical replicates were used for downstream analysis. In addition, all variants occurring in ARTIC v3 primer-binding sites were discarded before proceeding with downstream analysis.

Quantification of SARS-CoV-2 vRNA

Plaque forming unit analysis was performed on all nasal swabs as published in Halfmann et al. 2019 [33]. Viral load analysis was performed on all of the nasal swab samples described above after they arrived in our laboratory. RNA was isolated using the Viral Total Nucleic Acid kit for the Maxwell RSC instrument (Promega, Madison, WI) following the manufacturer's instructions. Viral load quantification was performed using a sensitive qRT-PCR assay developed by the CDC to detect SARS-CoV-2 (specifically the N1 assay) and commercially available from IDT (Coralville, IA). The assay was run on a LightCycler 96 or LC480 instrument (Roche, Indianapolis, IN) using the Taqman Fast Virus 1-stepMaster Mix enzyme (Thermo Fisher, Waltham, MA). The limit of detection of this assay is estimated to be 200 genome equivalents/ml saliva or swab fluid. To determine the viral load, samples were interpolated onto a standard curve consisting of serial 10-fold dilutions of in vitro transcribed SARS-CoV-2 N gene RNA.

Pairwise nucleotide diversity calculations

Nucleotide diversity was calculated using π summary statistics (**Supplementary Table 2**). π quantifies the average number of pairwise differences per nucleotide site among a set of sequences and was calculated per gene using SNPGenie

(https://github.com/chasewnelson/SNPgenie) [61]. SNPGenie adapts the Nei and Gojobori method of estimating nucleotide diversity (π), and its synonymous (πS) and nonsynonymous (πN) partitions from next-generation sequencing data [62]. When πN = πS, this indicates neutral evolution or genetic drift, with neither strong purifying nor positive selection playing a large role in the evolution of the viral population. πN < πS indicates purifying selection is acting to remove deleterious mutations, and πN > πS shows positive or diversifying selection acting on nonsynonymous variation [63]. We tested the null hypothesis that πN = πS within each gene using an unpaired t-test (**Supplementary Table 1**). The code to replicate these results can be found in the `diversity_estimates.ipynb` Jupyter Notebook in the `code` directory of the GitHub repository [60].

SNP Frequency Spectrum calculations

To generate SNP Frequency Spectrums (SFS), we binned all variants detected across timepoints within each index cat into six bins – 3-10%, 10-20%, 20-30%, 30-40%, 40-50%, 50-60%. We plotted the counts of variants falling into each frequency bin using Matplotlib 3.3.2 (https://matplotlib.org). We used code written by Dr. Louise Moncla to generate the distribution of SNPs for a given population assuming no selection or change in population size, which is expected to follow a 1/x distribution [37]. The code to replicate this can be found in the GitHub accompanying this manuscript, specifically in the `code/SFS.ipynb` Jupyter Notebook. This model predicts 42.8% of variants will fall within the 3-10% frequency range, 24.6% will fall within the 10-20% frequency range, 14.4% of variants will fall within the 20-30% frequency range, 10.2% of variants will fall within the 30-40% frequency range, and 7.9% of variants will fall within the 40-50% frequency range. We used a Mann-Whitney U test to test the null hypothesis that the distribution of variant frequencies for each index cat was equal to the neutral distribution.

The code to replicate these results can be found in the `SFS.ipynb` Jupyter Notebook in the `code` directory of the GitHub repository [60].

Focal Nextstrain build of S H655Y sequences

The focal H655Y build (**Supplementary Fig 5**) was prepared as described in Hodcroft et al. (2020), with different mutations targeted for the S:655 mutation [64]. Briefly: sequences with a mutation at nucleotide position 23525 (corresponding to a change at the 655 position in the spike glycoprotein) were selected from all available sequences on GISAID as of 29th December 2020. These sequences were included as the 'focal' set for a Nextstrain phylogenetic analysis, to which 'context' sequences were added, with the most genetically similar sequences given priority.

Data availability

Source data after mapping have been deposited in the Sequence Read Archive (SRA) under bioproject PRJNA666926[https://www.ncbi.nlm.nih.gov/bioproject/666926]. Derived data, analysis pipelines, and figures have been made available for easy replication of these results at a publically-accessible GitHub repository: https://github.com/katarinabraun/SARSCoV2_transmission_in_domestic_cats [60].

Code availability

Code to replicate analyses and re-create most figures is available at https://github.com/katarinabraun/SARSCoV2_transmission_in_domestic_cats [60]. **Figure 1** was created by hand in Adobe Illustrator and **Supplementary Figures 6** and **7** were created using samtools command line tools, were visualized in JMP Pro 15

(https://www.jmp.com/en_in/software/new-release/new-in-jmp-and-jmp-pro.html), and were then edited for readability in Adobe Illustrator. Code to process sequencing data is available at https://github.com/gagekmoreno/SARS_CoV-2_Zequencer and dependencies are available through Docker [65]. Results were visualized using Matplotlib 3.3.2(https://matplotlib.org), Seaborn v0.10.0 (https://github.com/mwaskom/seaborn), and Baltic v0.1.0 (https://github.com/evogytis/baltic).

Acknowledgements

This project was funded in part through a COVID-19 Response grant from the Wisconsin Partnership Program at the University of Wisconsin School of Medicine and Public Health to TCF and DHO. We would like to acknowledge Genetic Services (https://ehr.primate.wisc.edu/project/WNPRC/WNPRC_Units/Research_Services/Genetics_Services/Public/begin.view?) who sequenced the challenge stock viruses discussed in the discussion section of this paper (BioProject: PRJNA627977). GKM is supported by an NLM training grant to the Computation and Informatics in Biology and Medicine Training Program (NLM 5T15LM007359). We thank Chelsea Crooks for her careful reading of and comments on this manuscript.

Author Contributions

K.M.B. contributed conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, visualization, writing – original draft preparation, writing – review and editing.

G.K.M. contributed conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, visualization, writing – original draft preparation, writing – review and editing.

P.J.H. contributed conceptualization, data curation, investigation, methodology, project administration, resources, writing – review and editing.

E.B.H contributed – formal analysis, visualization, writing – review and editing.

E.B. contributed – visualization, writing – original draft preparation, writing – review and editing.

A.M.W. contributed investigation, writing – review and editing.

A.K.H. contributed formal analysis, writing – review and editing.

M.H. contributed conceptualization, resources, writing – review and editing.

S.C. contributed conceptualization, resources, supervision, writing – review and editing.

T.M. contributed conceptualization, resources, supervision, writing – review and editing.

Y.K. contributed conceptualization, resources, supervision, writing – review and editing.

K.K. contributed conceptualization, methodology, software, supervision, writing – review and editing.

D.H.O. contributed conceptualization, funding acquisition, methodology, supervision, writing – review and editing.

T.C.F. contributed conceptualization, funding acquisition, methodology, supervision, writing – review and editing.

Competing Interests

The authors declare no competing interests.

References

1. Shi J, Wen Z, Zhong G et al. Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. Science. 2020;368:1016-1020.
2. Rockx B, Kuiken T, Herfst S et al. Comparative pathogenesis of COVID-19, MERS, and SARS in a nonhuman primate model. Science. 2020;368:1012-1015.
3. Imai M, Iwatsuki-Horimoto K, Hatta M et al. Syrian hamsters as a small animal model for SARS-CoV-2 infection and countermeasure development. Proc Natl Acad Sci U S A. 2020;117:16587-16595.
4. Shan C, Yao YF, Yang XL et al. Infection with novel coronavirus (SARS-CoV-2) causes pneumonia in Rhesus macaques. Cell Res. 2020;30:670-677.
5. Oreshkova N, Molenaar RJ, Vreman S et al. SARS-CoV-2 infection in farmed minks, the Netherlands, April and May 2020. Euro Surveill. 2020;25
6. Lakdawala SS, Menachery VD. The search for a COVID-19 animal model. Science. 2020;368:942-943.
7. Richard M, Kok A, de Meulder D et al. SARS-CoV-2 is transmitted via contact and via the air between ferrets. Nat Commun. 2020;11:3496.
8. Oude Munnink BB, Sikkema RS, Nieuwenhuijse DF et al. Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. Science. 2020
9. Park SE. Epidemiology, virology, and clinical features of severe acute respiratory syndrome -coronavirus-2 (SARS-CoV-2; Coronavirus Disease-19). Clin Exp Pediatr. 2020;63:119-124.

10. Yang X, Yu Y, Xu J et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med. 2020;8:475-481.
11. McCrone JT, Lauring AS. Genetic bottlenecks in intraspecies virus transmission. Curr Opin Virol. 2018;28:20-25.
12. McCrone JT, Woods RJ, Martin ET, Malosh RE, Monto AS, Lauring AS. Stochastic processes constrain the within and between host evolution of influenza virus. Elife. 2018;7
13. Valesano AL, Fitzsimmons WJ, McCrone JT et al. Influenza B Viruses Exhibit Lower Within-Host Diversity than Influenza A Viruses in Human Hosts. J Virol. 2020;94
14. Wilker PR, Dinis JM, Starrett G et al. Selection on haemagglutinin imposes a bottleneck during mammalian transmission of reassortant H5N1 influenza viruses. Nat Commun. 2013;4:2636.
15. Moncla LH, Zhong G, Nelson CW et al. Selective Bottlenecks Shape Evolutionary Pathways Taken during Mammalian Adaptation of a 1918-like Avian Influenza Virus. Cell Host Microbe. 2016;19:169-180.
16. Varble A, Albrecht RA, Backes S et al. Influenza A virus transmission bottlenecks are defined by infection route and recipient host. Cell Host Microbe. 2014;16:691-700.
17. Zaraket H, Baranovich T, Kaplan BS et al. Mammalian adaptation of influenza A(H7N9) virus is limited by a narrow genetic bottleneck. Nat Commun. 2015;6:6553.
18. Xue KS, Bloom JD. Reconciling disparate estimates of viral genetic diversity during human influenza infections.[letter]. Nat Genet 2019;51(9):1298-1301.
19. Bergstrom CT, McElhany P, Real LA. Transmission bottlenecks as determinants of virulence in rapidly evolving pathogens. Proc Natl Acad Sci U S A. 1999;96:5095-5100.
20. Elena SF, Sanjuán R, Bordería AV, Turner PE. Transmission bottlenecks and the evolution of fitness in rapidly evolving RNA viruses. Infect Genet Evol. 2001;1:41-48.

583 21. Sobel Leonard A, Weissman DB, Greenbaum B, Ghedin E, Koelle K. Transmission
584 Bottleneck Size Estimation from Pathogen Deep-Sequencing Data, with an Application to
585 Human Influenza A Virus. J Virol. 2017;91
586 22. Dinis JM, Florek NW, Fatola OO et al. Deep Sequencing Reveals Potential Antigenic
587 Variants at Low Frequencies in Influenza A Virus-Infected Humans. J Virol. 2016;90:3355-
588 3365.
589 23. Kiso M, Iwatsuki-Horimoto K, Yamayoshi S et al. Emergence of Oseltamivir-Resistant
590 H7N9 Influenza Viruses in Immunosuppressed Cynomolgus Macaques. J Infect Dis.
591 2017;216:582-593.
592 24. Zwart MP, Elena SF. Matters of Size: Genetic Bottlenecks in Virus Infection and Their
593 Potential Impact on Evolution. Annu Rev Virol. 2015;2:161-179.
594 25. Wang D, Wang Y, Sun W et al. Population Bottlenecks and Intra-host Evolution during
595 Human-to-Human Transmission of SARS-CoV-2. bioRxiv. 20202020.06.26.173203.
596 26. Lythgoe KA, Hall M, Ferretti L et al. Shared SARS-CoV-2 diversity suggests localised
597 transmission of minority variants. bioRxiv. 20202020.05.28.118992.
598 27. Popa A, Genger J-W, Nicholson M et al. Mutational dynamics and transmission properties
599 of SARS-CoV-2 superspreading events in Austria. bioRxiv. 20202020.07.15.204339.
600 28. Graudenzi A, Maspero D, Angaroni F, Piazza R, Ramazzotti D. Mutational Signatures and
601 Heterogeneous Host Response Revealed Via Large-Scale Characterization of SARS-
602 COV-2 Genomic Diversity. bioRxiv. 20202020.07.06.189944.
603 29. Sapoval N, Mahmoud M, Jochum MD et al. Hidden genomic diversity of SARS-CoV-2:
604 implications for qRT-PCR diagnostics and transmission. bioRxiv. 2020
605 30. Karamitros T, Papadopoulou G, Bousali M, Mexias A, Tsiodras S, Mentis A. SARS-CoV-2
606 exhibits intra-host genomic plasticity and low-frequency polymorphic quasispecies. J Clin
607 Virol. 2020;131:104585.

608 31. Shen Z, Xiao Y, Kang L et al. Genomic diversity of SARS-CoV-2 in Coronavirus Disease
609 2019 patients. Clin Infect Dis. 2020
610 32. Ramazzotti D, Angaroni F, Maspero D et al. VERSO: a comprehensive framework for the
611 inference of robust phylogenies and the quantification of intra-host genomic diversity of
612 viral samples. bioRxiv. 20202020.04.22.044404.
613 33. Halfmann PJ, Hatta M, Chiba S et al. Transmission of SARS-CoV-2 in Domestic
614 Cats.[letter]. N Engl J Med 2020;383(6):592-594.
615 34. Quick J, Grubaugh ND, Pullan ST et al. Multiplex PCR method for MinION and Illumina
616 sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc.
617 2017;12:1261-1276.
618 35. Quick J. nCoV-2019 sequencing protocol. protocolsio. 2020
619 36. Grubaugh ND, Gangavarapu K, Quick J et al. An amplicon-based sequencing framework
620 for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol.
621 2019;20:8.
622 37. Moncla LH, Bedford T, Dussart P et al. Quantifying within-host diversity of H5N1 influenza
623 viruses in humans and poultry in Cambodia. PLoS Pathog. 2020;16:e1008191.
624 38. Hahn MW. Molecular population genetics. New York: Sinauer Associates.; 2019
625 39. Zhao L, Illingworth CJR. Measurements of intrahost viral diversity require an unbiased
626 diversity metric. Virus Evol. 2019;5:vey041.
627 40. Rausch JW, Capoferri AA, Katusiime MG, Patro SC, Kearney MF. Low genetic diversity
628 may be an Achilles heel of SARS-CoV-2. Proc Natl Acad Sci U S A. 2020;117:24614-
629 24616.
630 41. Guallar MP, Meiriño R, Donat-Vargas C, Corral O, Jouvé N, Soriano V. Inoculum at the
631 time of SARS-CoV-2 exposure and risk of disease severity. Int J Infect Dis. 2020;97:290-
632 292.

633 42. Gandhi M, Beyrer C, Goosby E. Masks Do More Than Protect Others During COVID-19:
634 Reducing the Inoculum of SARS-CoV-2 to Protect the Wearer. J Gen Intern Med.
635 2020;35:3063-3066.
636 43. Tanaka Y, Mizokami M. Genetic diversity of hepatitis B virus as an important factor
637 associated with differences in clinical outcomes. J Infect Dis. 2007;195:1-4.
638 44. Tahamtan A, Askari FS, Bont L, Salimi V. Disease severity in respiratory syncytial virus
639 infection: Role of host genetic variation. Rev Med Virol. 2019;29:e2026.
640 45. Xiao Y, Dolan PT, Goldstein EF et al. Poliovirus intrahost evolution is required to
641 overcome tissue-specific innate immune responses. Nat Commun. 2017;8:375.
642 46. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. Quasispecies diversity
643 determines pathogenesis through cooperative interactions in a viral population. Nature.
644 2006;439:344-348.
645 47. Memoli MJ, Bristol T, Proudfoot KE, Davis AS, Dunham EJ, Taubenberger JK. In vivo
646 evaluation of pathogenicity and transmissibility of influenza A(H1N1)pdm09 hemagglutinin
647 receptor binding domain 222 intrahost variants isolated from a single
648 immunocompromised patient. Virology. 2012;428:21-29.
649 48. Biswas SK, Mudi SR. Genetic variation in SARS-CoV-2 may explain variable severity of
650 COVID-19.[letter]. Med Hypotheses 2020;143:109877.
651 49. Zhang X, Tan Y, Ling Y et al. Viral and host factors related to the clinical outcome of
652 COVID-19. Nature. 2020;583:437-440.
653 50. Baum A, Fulton BO, Wloga E et al. Antibody cocktail to SARS-CoV-2 spike protein
654 prevents rapid mutational escape seen with individual antibodies. Science.
655 2020;369:1014-1018.
656 51. Dieterle ME, Haslwanter D, Bortz RH et al. A replication-competent vesicular stomatitis
657 virus for studies of SARS-CoV-2 spike-mediated cell entry and its inhibition. bioRxiv. 2020

52. Chan JF, Zhang AJ, Yuan S et al. Simulation of the clinical and pathological manifestations of Coronavirus Disease 2019 (COVID-19) in golden Syrian hamster model: implications for disease pathogenesis and transmissibility. Clin Infect Dis. 2020
53. YANG X, Dong N, CHAN W-C, CHEN S. Identification of super-transmitters of SARS-CoV-2. medRxiv. 20202020.04.19.20071399.
54. Tze Chuen Lee R. Spike glycoprotein mutation surveillance.
55. Huang Y, Yang C, Xu XF, Xu W, Liu SW. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. Acta Pharmacol Sin. 2020;41:1141-1149.
56. Organization WH. SARS-CoV-2 mink-associated variant strain – Denmark. Disease Outbreak News. 2020
57. Rambaut A, Loman N, Pybus O et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. 2020
58. Gu H, Chen Q, Yang G et al. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. Science. 2020;369:1603-1607.
59. Hodcroft EB, Neher RA. Phylogenetic analysis of SARS-CoV-2 clusters in their international context - cluster S.H655Y.
60. Braun KM. katarinabraun/SARSCoV2_transmission_in_domestic_cats. GitHub. 2020
61. Nelson CW, Moncla LH, Hughes AL. SNPGenie: estimating evolutionary parameters to detect natural selection using pooled next-generation sequencing data. Bioinformatics. 2015;31:3709-3711.
62. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol. 1986;3:418-426.
63. Hughes AL. Adaptive evolution of genes and genomes. 1999

64. Hodcroft EB, Zuber M, Nadeau S et al. Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. medRxiv. 2020

65. Moreno GK. gagekmoreno/SARS_CoV-2_Zequencer: Zequencer to accompany “Transmission of SARS-CoV-2 in domestic cats imposes a narrow bottleneck”. GitHub. 2020

Figures

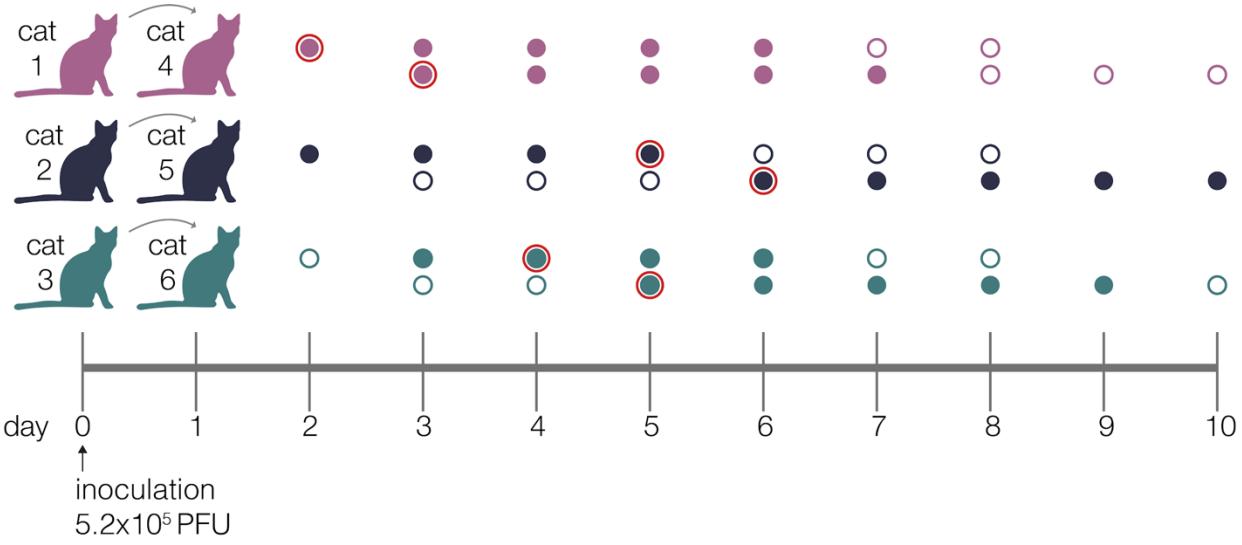
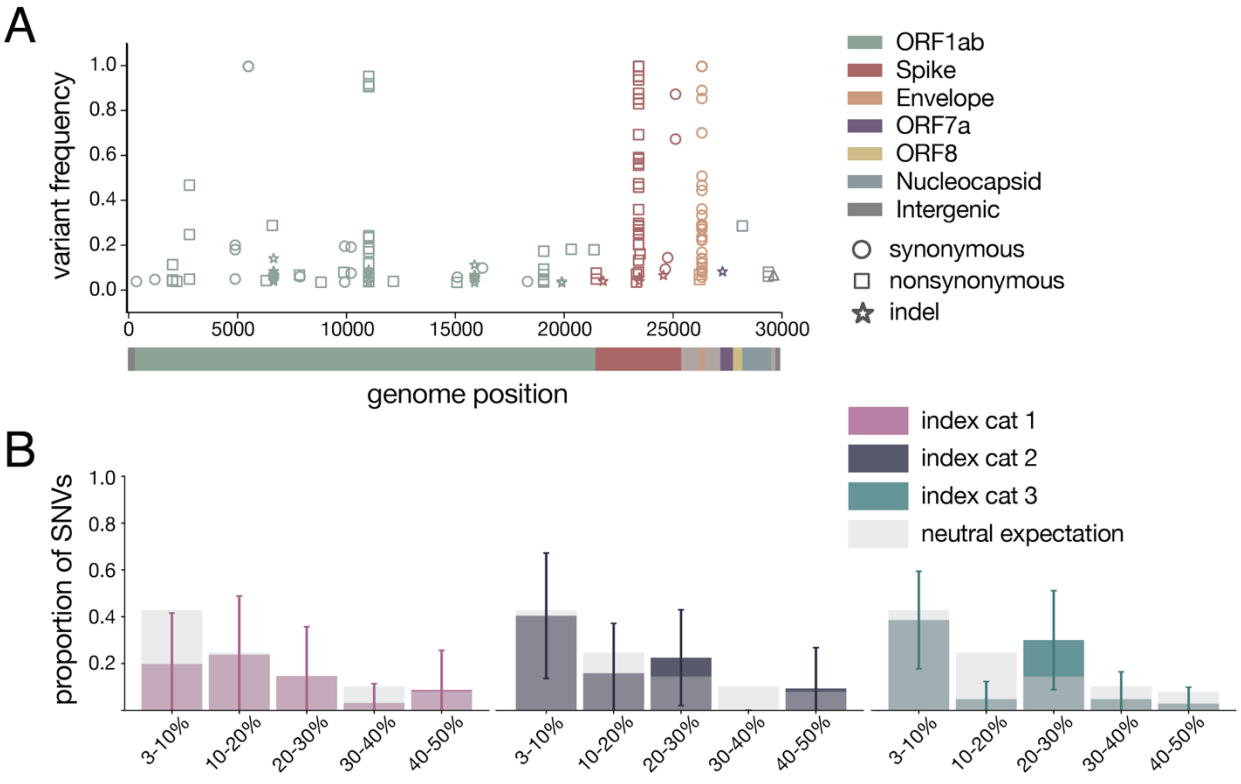


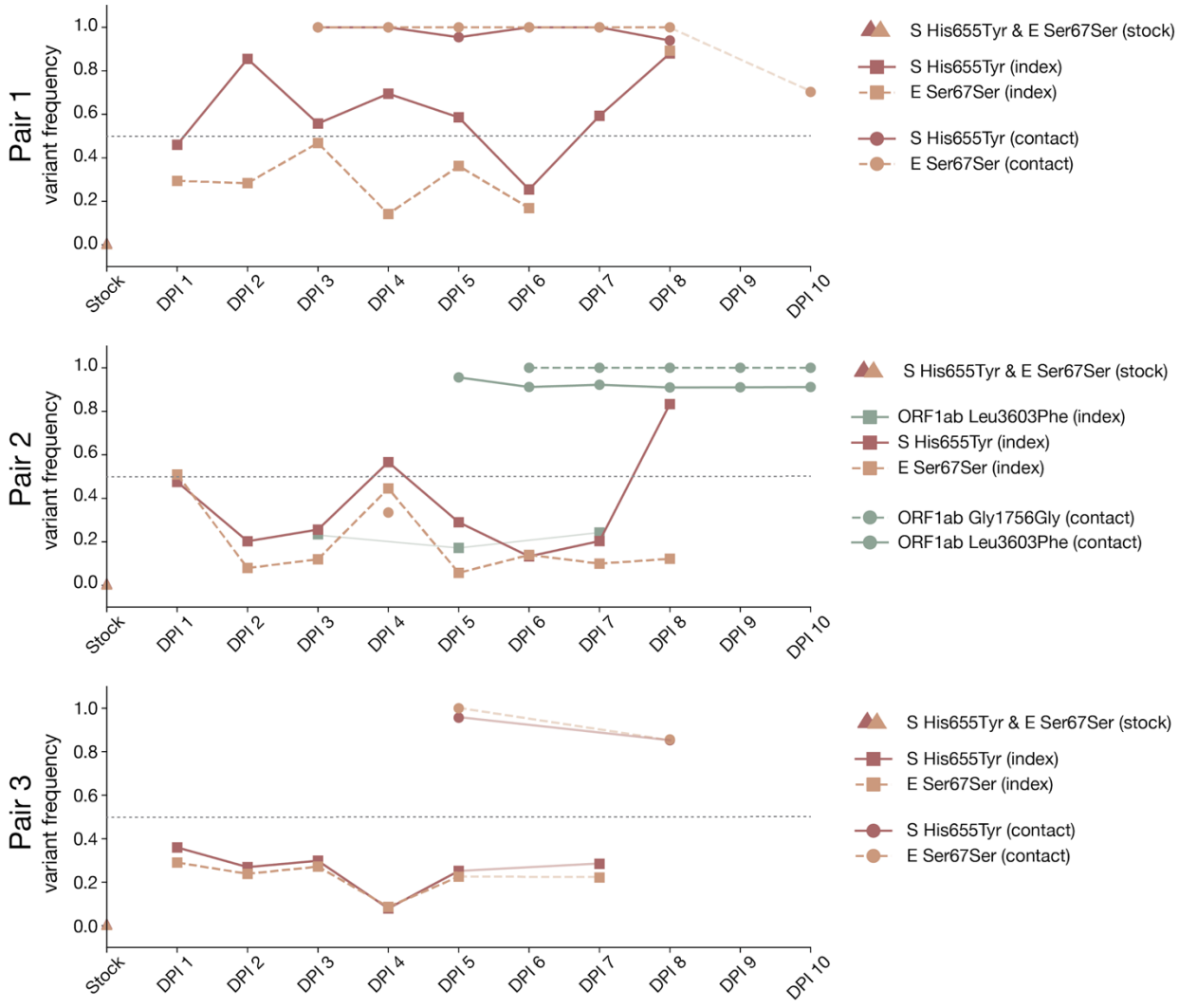
Figure 1. Experimental timeline. Schematic representing the sampling timeline for the three transmission pairs. Index cats were inoculated on day 0 with 5.2e5 PFU of a human isolate (hCoV-19/Japan/UT-NCGM02/2020) and were co-housed with a naive cat starting on day 1. Within each transmission pair, the top row of circles represent the index cat and the bottom row represents the contact cat. Open circles represent days on which there was no detectable infectious virus as indicated by plaque assay, and closed circles highlight days when live virus was recovered. Circles with a red outline indicate timepoints which were used in the beta-binomal estimate to calculate transmission bottleneck sizes.



705

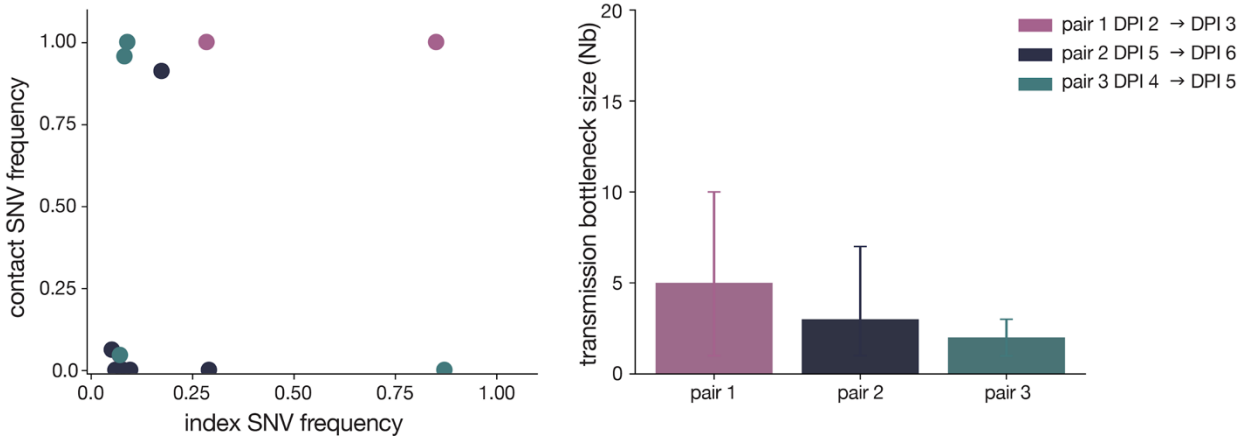
706 **Figure 2. Within-host diversity of SARS-CoV-2 viruses in domestic cats.** A) Plot representing all
707 variants (iSNVs and indels) detected in any cat at any timepoint. Variant frequencies are plotted by
708 genome location and are colored by gene. Circles represent synonymous iSNVs, squares represent
709 nonsynonymous iSNVs, and stars represent indels. B) iSNV frequency spectrums with error bars showing
710 standard deviation for index cats plotted against a “neutral model” (light gray bars) which assumes a
711 constant population size and the absence of selection.

712



713

714 **Figure 3. Frequency of iSNVs over time in each index and contact cat.** The frequency of iSNVs
715 discussed in the results over time in all six cats are shown. All iSNVs over time are shown in
716 **Supplementary Figure 2** and all indels over time are shown in **Supplementary Figure 3**. Each variant is
717 colored by gene location. Nonsynonymous variants are plotted with solid lines and synonymous variants
718 are plotted with dashed lines. Variants detected in index cats are denoted with squares and variants
719 detected in contact cats are denoted with circles. Timepoints with viral loads too low to yield high quality
720 sequences are shown by the gaps in data, but iSNVs are connected across these gaps using light lines
721 for readability (i.e. cat 1 day 9). The dotted line at 50% frequency represents the consensus threshold.



722

723 **Figure 4. SARS-CoV-2 transmission is defined by a narrow bottleneck.** Variant frequencies in the
724 index cats (x-axis) compared with frequencies of the same variants in the corresponding contact cats (y-
725 axis) that were used in the beta-binomial estimate are shown on the left. Estimates of SARS-CoV-2
726 transmission bottleneck with 99% confidence intervals shown on the right.