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

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

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

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

[5]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, escape human immunity 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, but dynamic sub-consensus diversity reveals processes of genetic drift and weak purifying selection. Transmission bottlenecks in this system appear narrow, with new infections being founded by fewer than ten viruses. We identify a notable variant at amino acid position 655 in Spike (H655Y) which arises rapidly in index cats and becomes fixed following transmission in two of three pairs, suggesting this site may be under positive selection in feline hosts. We speculate that narrow transmission bottlenecks and the lack of pervasive positive selection combine to constrain the pace of ongoing SARS-CoV-2 adaptive evolution in mammalian hosts.

Introduction

Understanding the forces that shape genetic diversity of RNA viruses as they replicate within, and are transmitted between, hosts may help us to forecast 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, persist 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 which makes each of these viable animal models [1–5]. Among these species, natural transmission has only been observed in mink, cats, and ferrets [1,6,7]. Very recently, transmission from humans to mink back to humans has 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 lower respiratory tract lung parenchyma, where infection is most commonly linked to severe disease in humans [1,6,9,10].

In a recent study, members of our team experimentally infected three index cats with a SARS-CoV-2 human isolate and introduced one naive direct contact cat per index one day following index inoculation [11]. Each index cat transmitted SARS-CoV-2 to the corresponding contact cat, and infectious virus was recovered from all six cats over multiple timepoints [11]. This study was designed to address whether SARS-CoV-2 could infect and transmit between cats so viruses collected from this study were not characterized beyond determining viral titers. Here, we use deep sequencing to define patterns of SARS-CoV-2 genetic variation over time within the index cats and following transmission to the contact cats.

Transmission bottlenecks – dramatic reductions in viral population size at the time of transmission – play an essential role in the overall pace of evolution of respiratory viruses [12–21]. 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 [13,14,17–19]. When transmission involves the transfer of very few variants, even

63 beneficial variants present at low frequencies below consensus in the transmitting host are likely
64 to be lost. Accordingly, antigenic escape variants can sometimes be detected in acute influenza
65 virus infections, and selective transmission of such variants has not been observed in nature
66 [22,22,23]. Narrow transmission bottlenecks in which natural selection is weak are expected to slow
67 the pace of seasonal influenza virus adaptation [12,24] and may have similar effects on SARS-
68 CoV-2.

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

78 Here we use a cat transmission model to show that SARS-CoV-2 genetic diversity is largely
79 shaped by genetic drift and purifying selection. This finding is in broad agreement with recent
80 analyses of evolutionary forces acting on SARS-CoV-2 in humans [25,27–32]. Our results further
81 suggest that human SARS-CoV-2 isolates are relatively well-adapted to feline hosts, and that
82 cat models recapitulate key aspects of SARS-CoV-2 evolution in humans. Notably, we estimate
83 narrow bottlenecks in cats, involving transmission of only 2-6 viruses, consistent with the subset
84 of studies in humans that have reported small bottleneck sizes [25–28]. We posit that the cat
85 transmission model will be useful for investigating within- and between-host evolution of SARS-
86 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) [11]. 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 variant-calling 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) [33,34]. Variants were required to be present in technical replicates at ≥3% and ≤97% of sequencing reads [35] (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 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 6**). 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 (42 out of 38 iSNVs and 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 [36]. Importantly, especially in the setting of an acute viral infection, exponential population growth is also expected to result in an excess of low-frequency variants [37]. To determine the type of evolutionary pressure acting on SARS-CoV-2 in cats, we plotted these distributions against a simple "neutral model" (transparent grey bars in **Fig 2B**), which assumes a constant population size and the absence of selection [36]. 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 [38]. 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 – stochastic changes in the frequency of viral genotypes over time – can be an important force shaping genetic diversity.

We observe that πS exceeds or 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) 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 were detected above 3% frequency in the inoculum, but when we mined all sequencing reads, S H655Y and E S67S can be detected at 0.85% and 0.34% in the inoculum, 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 at all

timepoints in cat 3 on days that SARS-CoV-2 was detectable ≥ 10^4 copies/mL and day 8 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), and interestingly 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 [39].

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 6** and all indels over time are shown in **Supplementary Fig 7**. 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 associated index: 7 fewer

variants in cat 4 (n=9) compared to cat 1 (n=16), 10 fewer in cat 5 (n=19) than cat 2 (n=10), 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 near-fixed (**Supplementary Fig 6**).

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 [18]. 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 [40]. 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 [39]. Additionally, it has been shown that the route of influenza transmission directly impacts the size of the transmission bottleneck [17]. 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.

Discussion

While SARS-CoV-2 continues to spread globally at an alarming rate, the vast majority of humans remain immunologically naive to this virus. 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 SARS-CoV-2 evolution within and between humans will facilitate our ability to uncover the evolutionary trajectories accessible to this virus as it faces changes in population-level immunity. Using domestic cats as a translational model, we show that genetic drift appears to be a major force shaping SARS-CoV-2 evolution. Selection within hosts is weak, and transmission bottlenecks, even with the potential for contact transmission, appear narrow. These observations suggest that SARS-CoV-2 may already be well adapted to mammalian hosts. The strong role of genetic drift may combine with the relatively slow mutation rate and narrow transmission bottlenecks to slow the overall pace of viral evolution.

Here we use deep viral sequencing to carefully uncover within-host variants in 6 domestic cats grouped into three defined transmission pairs. We find genetic drift and purifying selection shape SARS-CoV-2 genetic diversity within feline hosts, and a stringent bottleneck defines viral transmission. This latter finding is at odds with some recent studies in humans, which have estimated wide and variable SARS-CoV-2 transmission bottlenecks [25–28]. Our data, however, is in line with other narrow estimates [25]. These discordant estimates are likely due to a combination of factors including uncertain sources of infection, difficulty collecting samples which closely bookend the transmission event, and inaccurate variant calls [25–28]. These 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 both 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.

Reports of natural transmission from humans to cats lends credence to the idea that cats can serve as a viable model for studying transmission [11,41–43]. Like in humans, infectious SARS-CoV-2 virus can be recovered from cat nasal turbinates, soft palates, tonsils, tracheas, and lungs [1]. Respiratory droplet transmission, which is thought to drive SARS-CoV-2 infection in humans, has been repeatedly reported in cats [44]. However, it is still likely that cat-specific processes and characteristics impact patterns of SARS-CoV-2 evolution and transmission in this model system. Our study found that overall genetic diversity in cats was low and signatures support prominent roles for genetic drift and purifying selection. The combination of low genetic diversity, genetic drift, and purifying selection are in line with recent estimates of human within-host diversity [25,27–32]. Importantly, our study only evaluated three transmission pairs with an undefined mode of transmission in the context of a single human isolate. Future studies will need to further explore the role of respiratory droplet transmission among a larger cohort of cats [43,45].

As has been has demonstrated for other viruses, a narrow transmission bottleneck may act as a brake on the pace with which SARS-CoV-2 will be able to evade humoral immunity and antiviral therapeutics through the stochastic loss of beneficial variants before they become fixed [12]. 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 [46,47]. The relationship

between the size of the genetic bottleneck and inoculum dose is not entirely clear, but these are not synonymous concepts. For example, an infection founded by 1,000 genetically identical viruses would be categorized as resulting from a narrow genetic bottleneck but a relatively large inoculation dose, or population bottleneck. Furthermore, levels of viral genetic diversity have been linked to pathogenesis or clinical outcomes in the context of other viruses (e.g., influenza A virus, polio, and respiratory syncytial virus) and because narrow transmission bottlenecks reduce viral genetic diversity, bottlenecks may play an essential role in the outcome of individual infections in this way as well [48–52]. The relationship between SARS-CoV-2 viral genetic diversity and COVID-19 severity has been discussed, but remains unclear [53,54].

Although within-host diversity was limited in the cats evaluated here, we identify two notable variants. S H655Y and E S67S arose rapidly, despite being found at 0.85% and 0.34% in the stock, and were detectable at intermediate frequencies at the first-day post-inoculation in all three index cats. S H655Y has been previously reported in other settings as well – natural SARS-CoV-2 infections in humans, transmission studies in a hamster model, as well as SARS-CoV-2 tissue culture experiments [55–58]. S H655Y is near the polybasic cleavage site, residing between the receptor binding domain and the fusion peptide, and is thought to modulate Spike glycoprotein fusion efficiency [55,56,59]. 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 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 replication in cats, viruses with this variant in linkage with E S67S might have been positively selected in all index cats.

Furthermore, S H655Y with E S67S were transmitted and fixed in contact cats in two of the three transmission events evaluated here. Although our sample size is small, the convergent origination of S H655Y with E S67S in all index cats and the fixation of these variants following transmission in two contact cats signals a potentially important functional role for one or both of these variants in feline hosts, and points towards a potential selective bottleneck. Although we cannot easily test this, if the transmission bottleneck were large and S H655Y, in linkage with E S67S, were rapidly selected in contact hosts immediately following transmission we might see a similar pattern to what we observe in cats 4 and 6. More generally, one challenge in interpreting these data is that we would expect to see a similar pattern of iSNVs in contact hosts following a narrow transmission bottleneck (as we suspect is the case here) as well as a wide transmission bottleneck in combination with a rapid selective sweep.

Large SARS-CoV-2 outbreaks in mink have been reported recently, some with 'concerning' mutations that may evade human humoral immunity [60]. This resulted in the Danish authorities' decision to cull 17 million mink as a safeguard against spill-back transmission into humans [60]. Humans are currently the primary reservoir for SARS-CoV-2, but the mink example shows that SARS-CoV-2 is able to infect and transmit among other mammals with the potential for ongoing zoonosis and anthroponosis. This exemplifies the need to understand the evolutionary mechanisms and pace at which SARS-CoV-2 is able to adapt to, and transmit between a broad range of host species. In our study we see variants arising early and being transmitted onward in cats, a potential reservoir species. Our study and the mink example show that species- and context-specific adaptations are inevitable as SARS-CoV-2 explores new hosts. While we do not know the phenotypic impacts of these variants, the rapid rise of variants in potential reservoir species may significantly impact humans if exposed to these new species-specific SARS-CoV-2 adaptations.

As more than 300,000 new SARS-CoV-2 cases occur each day worldwide, 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 incidence of natural exposure-related immunity, these models can help us forecast the future of SARS-CoV-2 variation and population-level genetic changes. Here, we use six domestic cats to show how SARS-CoV-2 genetic variation is predominantly influenced by genetic drift and purifying selection within individual hosts. Additionally, we find a role for narrow transmission bottlenecks shaping founding diversity in all three contact cats. 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 [33,34]. 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 [33,34]. 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 [33,34]. 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, followed by adaptor ligation. Samples were again purified using an AMPure bead cleanup (1x) and eluted in 25µL of resuspension buffer. Lastly, samples were amplified using 8 PCR cycles, cleaned with a 1:1 bead clean-up, and eluted in 30µL of RSB. The average sample fragment length and purity was determined using the Agilent High Sensitivity DNA kit and the Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA). After passing quality control measures, samples were pooled equimolarly to a final concentration of 4 nM, and 5 µl of each 4 nM pool was denatured in 5 µl of 0.2 N NaOH for 5 min. Sequencing pools were denatured to a final concentration of 10 pM with a PhiX-derived control library accounting for 1% of total DNA and was loaded onto a 500-cycle v2 flow cell. Average quality metrics were recorded, reads were demultiplexed, and FASTQ files were generated on Illumina's BaseSpace platform.

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

Raw FASTQ files were analyzed using a workflow called "SARSquencer". Briefly, reads are paired and merged using BBMerge (https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbmerge-guide/) and mapped to the reference (MW219695.1) using BBMap (https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbmap-guide/). Mapped reads were imported into Geneious (https://www.geneious.com/) for visual inspection. Variants were called using callvariants.sh (contained within BBMap) and annotated using SnpEff (https://pcingola.github.io/SnpEff/). The complete "SARSquencer" pipeline is available in the GitHub accompanying this manuscript in `code/SARSquencer` as well as in a separate GitHub

repository – https://github.com/gagekmoreno/SARS_CoV-2_Zequencer. BBMap's output VCF files were cleaned using custom Python scripts, which can be found in the GitHub accompanying this manuscript (https://github.com/katarinabraun/SARSCoV2_transmission_in_domestic_cats). 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 [11]. 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.

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 [36]. 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.

Data availability

Source data after have been deposited in the Sequence Read Archive (SRA) under bioproject PRJNA666926[https://www.ncbi.nlm.nih.gov/bioproject/666926]. The consensus genome sequences for national and international genomes are available from GISAID (www.gisaid.org; see Supplementary Table 3). 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.

Code availability

Code to replicate analyses and re-create most figures is available at https://github.com/katarinabraun/SARSCoV2_transmission_in_domestic_cats. **Figure 1** was created by hand in Adobe Illustrator and **Supplementary Figures 2** and **3** 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 [64]. 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. KMB is supported by F30 AI145182-01A1 from the National Institute of Allergy and Infectious Disease. 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. 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. et al. Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Science* **368**, 1016-1020 (2020).
2. Rockx, B. et al. Comparative pathogenesis of COVID-19, MERS, and SARS in a nonhuman primate model. *Science* **368**, 1012-1015 (2020).

3. Imai, 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* **117**, 16587-16595 (2020).
4. Shan, C. et al. Infection with novel coronavirus (SARS-CoV-2) causes pneumonia in Rhesus macaques. *Cell Res* **30**, 670-677 (2020).
5. Oreshkova, N. et al. SARS-CoV-2 infection in farmed minks, the Netherlands, April and May 2020. *Euro Surveill* **25**, (2020).
6. Lakdawala, S. S. & Menachery, V. D. The search for a COVID-19 animal model. *Science* **368**, 942-943 (2020).
7. Richard, M. et al. SARS-CoV-2 is transmitted via contact and via the air between ferrets. *Nat Commun* **11**, 3496 (2020).
8. Oude Munnink, B. B. et al. Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* (2020).
9. Park, S. E. Epidemiology, virology, and clinical features of severe acute respiratory syndrome -coronavirus-2 (SARS-CoV-2; Coronavirus Disease-19). *Clin Exp Pediatr* **63**, 119-124 (2020).
10. Yang, X. 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* **8**, 475-481 (2020).
11. Halfmann, P. J. et al. Transmission of SARS-CoV-2 in Domestic Cats. *N Engl J Med* **383**, 592-594 (2020).
12. McCrone, J. T. & Lauring, A. S. Genetic bottlenecks in intraspecies virus transmission. *Curr Opin Virol* **28**, 20-25 (2018).
13. McCrone, J. T. et al. Stochastic processes constrain the within and between host evolution of influenza virus. *Elife* **7**, (2018).

14. Valesano, A. L. et al. Influenza B Viruses Exhibit Lower Within-Host Diversity than Influenza A Viruses in Human Hosts. *J Virol* **94**, (2020).

15. Wilker, P. R. et al. Selection on haemagglutinin imposes a bottleneck during mammalian transmission of reassortant H5N1 influenza viruses. *Nat Commun* **4**, 2636 (2013).

16. Moncla, L. H. et al. Selective Bottlenecks Shape Evolutionary Pathways Taken during Mammalian Adaptation of a 1918-like Avian Influenza Virus. *Cell Host Microbe* **19**, 169-180 (2016).

17. Varble, A. et al. Influenza A virus transmission bottlenecks are defined by infection route and recipient host. *Cell Host Microbe* **16**, 691-700 (2014).

18. Zaraket, H. et al. Mammalian adaptation of influenza A(H7N9) virus is limited by a narrow genetic bottleneck. *Nat Commun* **6**, 6553 (2015).

19. Xue, K. S. & Bloom, J. D. Reconciling disparate estimates of viral genetic diversity during human influenza infections. *Nat Genet* **51**, 1298-1301 (2019).

20. Bergstrom, C. T., McElhany, P. & Real, L. A. Transmission bottlenecks as determinants of virulence in rapidly evolving pathogens. *Proc Natl Acad Sci U S A* **96**, 5095-5100 (1999).

21. Elena, S. F., Sanjuán, R., Bordería, A. V. & Turner, P. E. Transmission bottlenecks and the evolution of fitness in rapidly evolving RNA viruses. *Infect Genet Evol* **1**, 41-48 (2001).

22. Dinis, J. M. et al. Deep Sequencing Reveals Potential Antigenic Variants at Low Frequencies in Influenza A Virus-Infected Humans. *J Virol* **90**, 3355-3365 (2016).

23. Kiso, M. et al. Emergence of Oseltamivir-Resistant H7N9 Influenza Viruses in Immunosuppressed Cynomolgus Macaques. *J Infect Dis* **216**, 582-593 (2017).

24. Zwart, M. P. & Elena, S. F. Matters of Size: Genetic Bottlenecks in Virus Infection and Their Potential Impact on Evolution. *Annu Rev Virol* **2**, 161-179 (2015).

25. Wang, D. et al. Population Bottlenecks and Intra-host Evolution during Human-to-Human Transmission of SARS-CoV-2. *bioRxiv* 2020.06.26.173203 (2020).

514 26. Lythgoe, K. A. et al. Shared SARS-CoV-2 diversity suggests localised transmission of
515 minority variants. *bioRxiv* 2020.05.28.118992 (2020).
516 27. Popa, A. et al. Mutational dynamics and transmission properties of SARS-CoV-2
517 superspreading events in Austria. *bioRxiv* 2020.07.15.204339 (2020).
518 28. Graudenzi, A., Maspero, D., Angaroni, F., Piazza, R. & Ramazzotti, D. Mutational
519 Signatures and Heterogeneous Host Response Revealed Via Large-Scale
520 Characterization of SARS-COV-2 Genomic Diversity. *bioRxiv* 2020.07.06.189944 (2020).
521 29. Sapoval, N. et al. Hidden genomic diversity of SARS-CoV-2: implications for qRT-PCR
522 diagnostics and transmission. *bioRxiv* (2020).
523 30. Karamitros, T. et al. SARS-CoV-2 exhibits intra-host genomic plasticity and low-frequency
524 polymorphic quasispecies. *J Clin Virol* **131**, 104585 (2020).
525 31. Shen, Z. et al. Genomic diversity of SARS-CoV-2 in Coronavirus Disease 2019 patients.
526 *Clin Infect Dis* (2020).
527 32. Ramazzotti, D. et al. VERSO: a comprehensive framework for the inference of robust
528 phylogenies and the quantification of intra-host genomic diversity of viral samples. *bioRxiv*
529 2020.04.22.044404 (2020).
530 33. Quick, J. et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and
531 other virus genomes directly from clinical samples. *Nat Protoc* **12**, 1261-1276 (2017).
532 34. Quick, J. nCoV-2019 sequencing protocol. *protocols.io* (2020).
533 35. Grubaugh, N. D. et al. An amplicon-based sequencing framework for accurately
534 measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol* **20**, 8 (2019).
535 36. Moncla, L. H. et al. Quantifying within-host diversity of H5N1 influenza viruses in humans
536 and poultry in Cambodia. *PLoS Pathog* **16**, e1008191 (2020).
537 37. Hahn, M. W. *Molecular population genetics* (New York: Sinauer Associates., 2019).

538 38. Zhao, L. & Illingworth, C. J. R. Measurements of intrahost viral diversity require an
539 unbiased diversity metric. *Virus Evol* **5**, vey041 (2019).
540 39. Rausch, J. W., Capoferri, A. A., Katusiime, M. G., Patro, S. C. & Kearney, M. F. Low
541 genetic diversity may be an Achilles heel of SARS-CoV-2. *Proc Natl Acad Sci U S A* **117**,
542 24614-24616 (2020).
543 40. Sobel Leonard, A., Weissman, D. B., Greenbaum, B., Ghedin, E. & Koelle, K.
544 Transmission Bottleneck Size Estimation from Pathogen Deep-Sequencing Data, with an
545 Application to Human Influenza A Virus. *J Virol* **91**, (2017).
546 41. Segalés, J. et al. Detection of SARS-CoV-2 in a cat owned by a COVID-19-affected
547 patient in Spain. *Proc Natl Acad Sci U S A* **117**, 24790-24793 (2020).
548 42. Bosco-Lauth, A. M. et al. Experimental infection of domestic dogs and cats with SARS-
549 CoV-2: Pathogenesis, transmission, and response to reexposure in cats. *Proc Natl Acad*
550 *Sci U S A* **117**, 26382-26388 (2020).
551 43. Barrs, V. R. et al. SARS-CoV-2 in Quarantined Domestic Cats from COVID-19
552 Households or Close Contacts, Hong Kong, China. *Emerg Infect Dis* **26**, (2020).
553 44. Prather, K. A. et al. Airborne transmission of SARS-CoV-2. *Science* **370**, 303-304 (2020).
554 45. Hosie, M. J. et al. Respiratory disease in cats associated with human-to-cat transmission
555 of SARS-CoV-2 in the UK. *bioRxiv* 2020.09.23.309948 (2020).
556 46. Guallar, M. P. et al. Inoculum at the time of SARS-CoV-2 exposure and risk of disease
557 severity. *Int J Infect Dis* **97**, 290-292 (2020).
558 47. Gandhi, M., Beyrer, C. & Goosby, E. Masks Do More Than Protect Others During COVID-
559 19: Reducing the Inoculum of SARS-CoV-2 to Protect the Wearer. *J Gen Intern Med* **35**,
560 3063-3066 (2020).
561 48. Tanaka, Y. & Mizokami, M. Genetic diversity of hepatitis B virus as an important factor
562 associated with differences in clinical outcomes. *J Infect Dis* **195**, 1-4 (2007).

563 49. Tahamtan, A., Askari, F. S., Bont, L. & Salimi, V. Disease severity in respiratory syncytial
564 virus infection: Role of host genetic variation. *Rev Med Virol* **29**, e2026 (2019).
565 50. Xiao, Y. et al. Poliovirus intrahost evolution is required to overcome tissue-specific innate
566 immune responses. *Nat Commun* **8**, 375 (2017).
567 51. Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E. & Andino, R. Quasispecies
568 diversity determines pathogenesis through cooperative interactions in a viral population.
569 *Nature* **439**, 344-348 (2006).
570 52. Memoli, M. J. et al. In vivo evaluation of pathogenicity and transmissibility of influenza
571 A(H1N1)pdm09 hemagglutinin receptor binding domain 222 intrahost variants isolated
572 from a single immunocompromised patient. *Virology* **428**, 21-29 (2012).
573 53. Biswas, S. K. & Mudi, S. R. Genetic variation in SARS-CoV-2 may explain variable
574 severity of COVID-19. *Med Hypotheses* **143**, 109877 (2020).
575 54. Zhang, X. et al. Viral and host factors related to the clinical outcome of COVID-19. *Nature*
576 **583**, 437-440 (2020).
577 55. Baum, A. et al. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational
578 escape seen with individual antibodies. *Science* **369**, 1014-1018 (2020).
579 56. Dieterle, M. E. et al. A replication-competent vesicular stomatitis virus for studies of SARS-
580 CoV-2 spike-mediated cell entry and its inhibition. *bioRxiv* (2020).
581 57. Chan, J. F. et al. Simulation of the clinical and pathological manifestations of Coronavirus
582 Disease 2019 (COVID-19) in golden Syrian hamster model: implications for disease
583 pathogenesis and transmissibility. *Clin Infect Dis* (2020).
584 58. YANG, X., Dong, N., CHAN, W.-C. & CHEN, S. Identification of super-transmitters of
585 SARS-CoV-2. *medRxiv* 2020.04.19.20071399 (2020).

59. Huang, Y., Yang, C., Xu, X. F., Xu, W. & Liu, S. W. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol Sin* **41**, 1141-1149 (2020).

60. Organization, W. H. SARS-CoV-2 mink-associated variant strain – Denmark. *Disease Outbreak News* (2020). https://www.who.int/csr/don/06-november-2020-mink-associated-sars-cov2-denmark/en/

61. Nelson, C. W., Moncla, L. H. & Hughes, A. L. SNPGenie: estimating evolutionary parameters to detect natural selection using pooled next-generation sequencing data. *Bioinformatics* **31**, 3709-3711 (2015).

62. Nei, M. & Gojobori, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**, 418-426 (1986).

63. Hughes, A. L. & Hughes, P. B. S. A. L. *Adaptive Evolution of Genes and Genomes* (Oxford University Press, 1999).

64. Moreno, G. K. gagekmoreno/SARS_CoV-2_Zequencer: Zequencer to accompany "Transmission of SARS-CoV-2 in domestic cats imposes a narrow bottleneck". *GitHub* (2020).

Figures

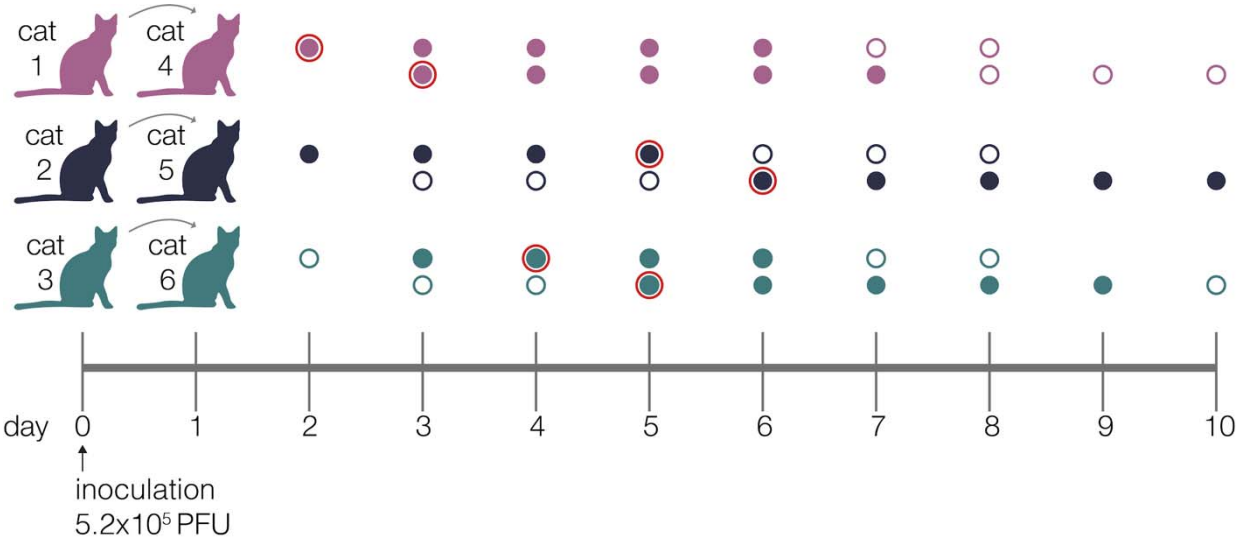

Figure 1. Experimental timeline. Schematic representing the sampling timeline for the three transmission pairs. Index cats were inoculated on day 0 with 5.2e5 PFU of a human isolate (hCoV-19/Japan/UT-NCGM02/2020) and were co-housed with a naive cat starting on day 1. Within each transmission pair, the 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 betabinomal estimate to calculate transmission bottleneck sizes.

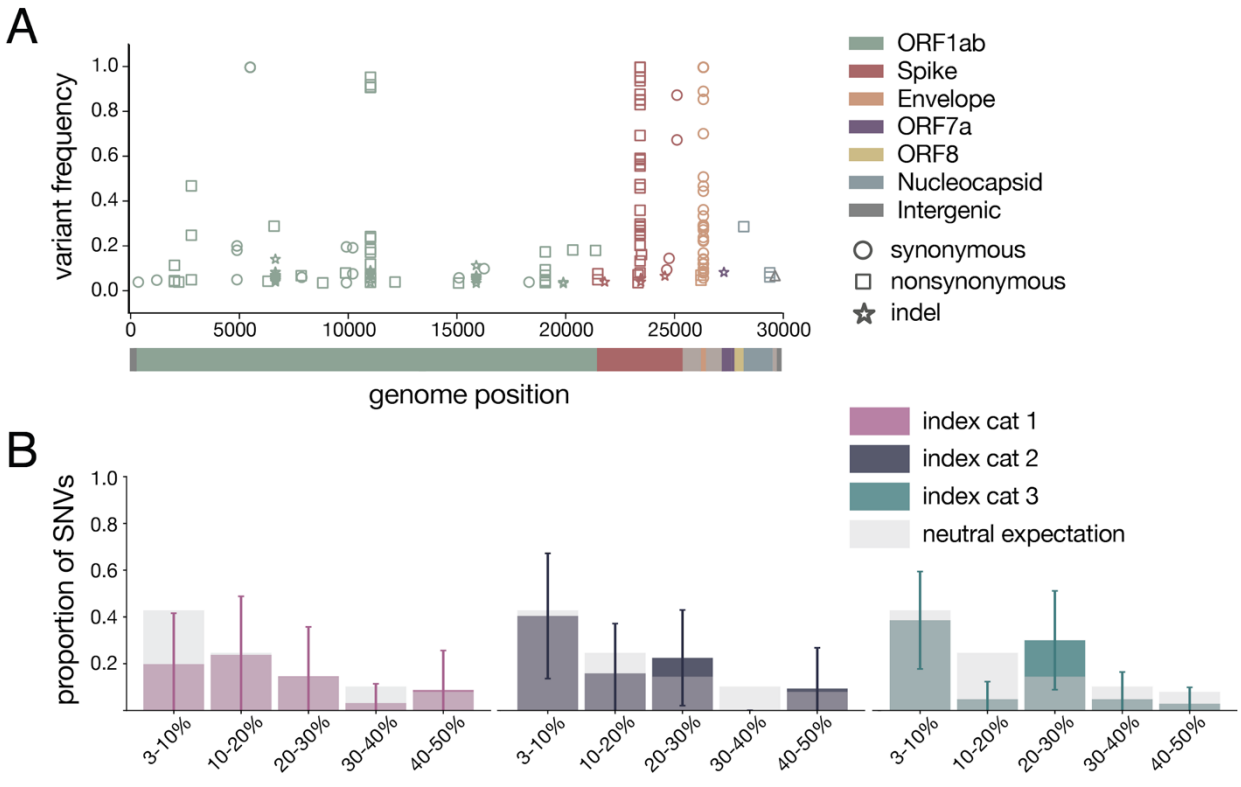

Figure 2. Within-host diversity of SARS-CoV-2 viruses in domestic cats. A) Plot representing all variants (iSNVs and indels) detected in any cat at any timepoint. Variant frequencies are plotted by genome location and are colored by gene. Circles represent synonymous iSNVs, squares represent nonsynonymous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

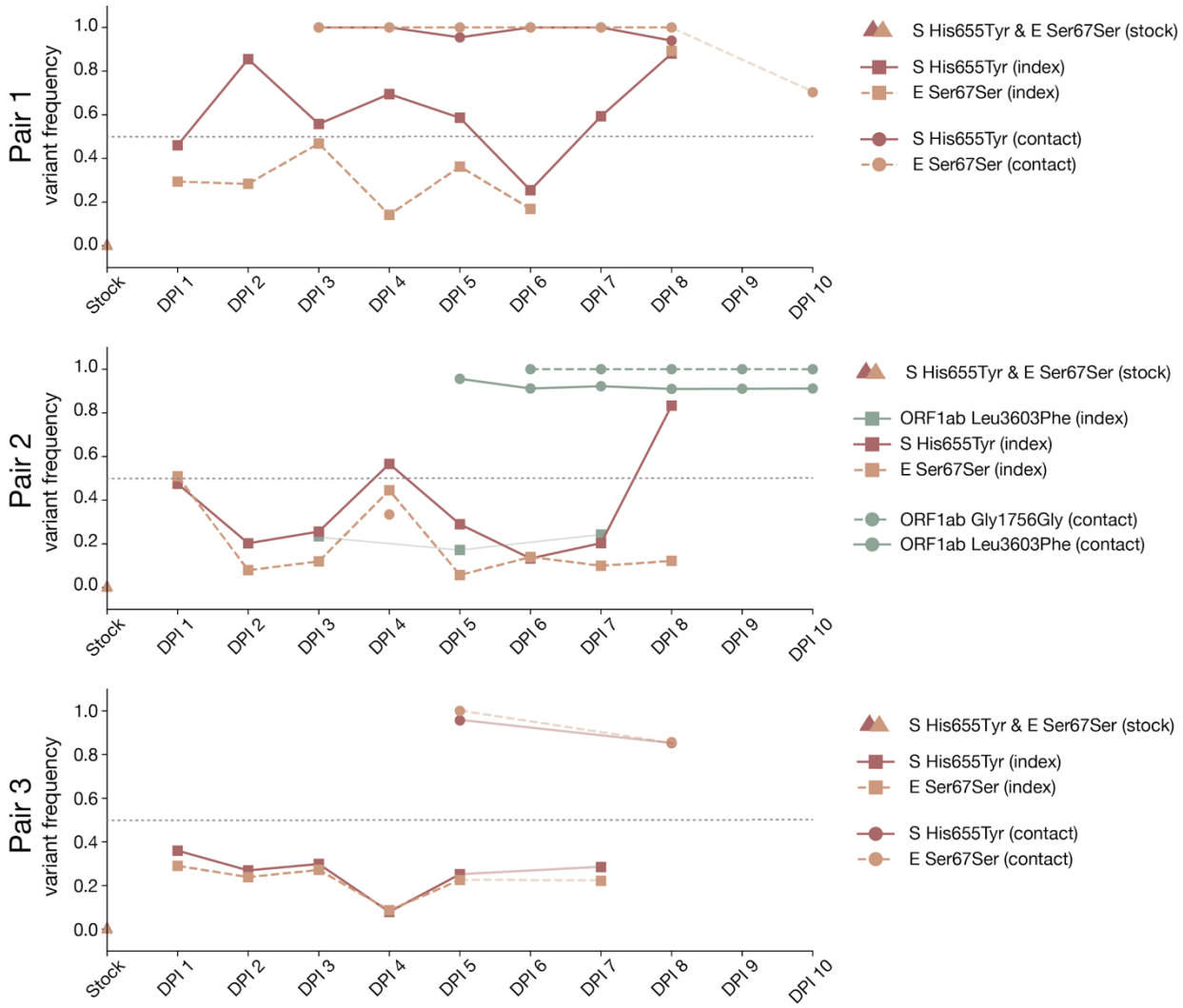

Figure 3. Frequency of iSNVs over time in each index and contact cat. The frequency of iSNVs discussed in the results over time in all six cats are shown. All iSNVs over time are shown in **Supplementary Figure 6** and all indels over time are shown in **Supplementary Figure 7**. Each variant is colored by gene location. Nonsynonymous variants are plotted with solid lines and synonymous variants are plotted with dashed lines. Variants detected in index cats are denoted with squares and variants detected in contact cats are denoted with circles. Timepoints with viral loads too low to yield high quality sequences are shown by the gaps in data, but iSNVs are connected across these gaps using light lines for readability (i.e. cat 1 day 9). The dotted line at 50% frequency represents the consensus threshold.

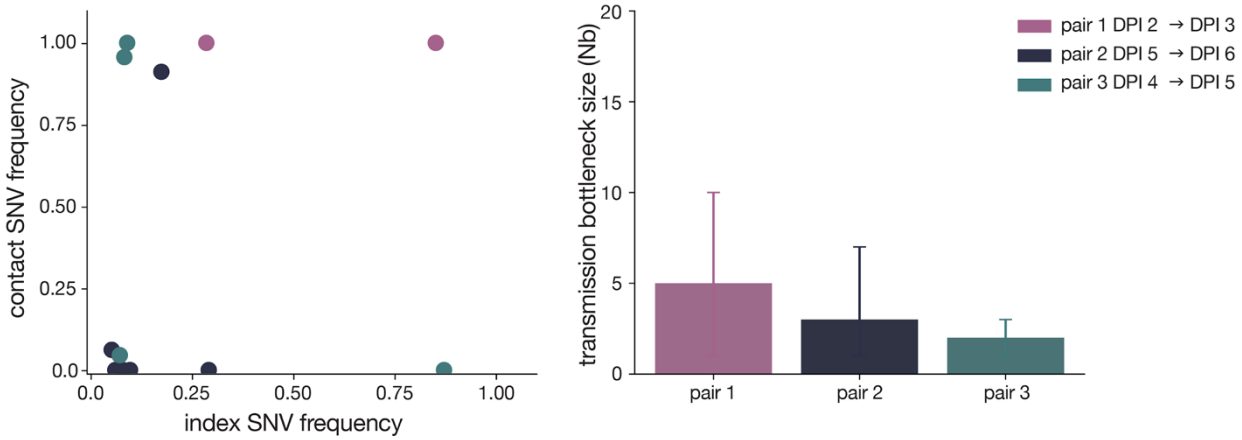

Figure 4. SARS-CoV-2 transmission is defined by a narrow bottleneck. Variant frequencies in the index cats (x-axis) compared with frequencies of the same variants in the corresponding contact cats (y-axis) that were used in the beta-binomial estimate are shown on the left. Estimates of SARS-CoV-2 transmission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